



# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

## FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

## ÚSTAV MANAGEMENTU

INSTITUTE OF MANAGEMENT

## NÁVRH REPORTOVACÍHO SYSTÉMU PRO POTŘEBY ŘÍZENÍ PREDIKCE SPOTŘEBY ELEKTRICKÉ ENERGIE.

DESIGN OF A REPORTING SYSTEM FOR THE NEEDS OF MANAGING ELECTRICITY CONSUMPTION  
PREDICTION

### DIPLOMOVÁ PRÁCE

MASTER'S THESIS

### AUTOR PRÁCE

AUTHOR

Bc. Jan Buš

### VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Kříž, Ph.D.

BRNO 2025

# Zadání diplomové práce

Ústav: Ústav managementu  
Student: **Bc. Jan Buš**  
Vedoucí práce: **Ing. Jiří Kříž, Ph.D.**  
Akademický rok: 2024/25  
Studijní program: Strategický rozvoj podniku

Garant studijního programu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává diplomovou práci s názvem:

## **Návrh reportovacího systému pro potřeby řízení predikce spotřeby elektrické energie.**

### **Charakteristika problematiky úkolu:**

Úvod  
Cíle práce, metody a postupy zpracování  
Teoretická východiska práce  
Analýza současného stavu  
Vlastní návrhy řešení  
Závěr  
Seznam použité literatury  
Přílohy

### **Cíle, kterých má být dosaženo:**

Cílem práce je navrhnout efektivní reportní systém, který umožní kontrolu, analýzu a vizualizaci dat spojených s predikcí a skutečnou spotřebou elektrické energie. Tento systém bude podporovat procesy řízení predikce elektrické energie na různých úrovních agregace (od jednotlivých spotřebičů po celou společnost) a měl by zlepšit přesnost, rychlost a kvalitu rozhodování.

### **Základní literární prameny:**

DECKLER, Greg a POWELL, Brett. Mastering Microsoft Power BI: Expert techniques for building actionable, interactive reports and dashboards. Birmingham, Velká Británie: Packt Publishing, 2022, 712 s. ISBN 978-1801811482.

CHMELÁR, Michal. Reporting v Power BI, PowerPivot a jazyk DAX. Pezinok, Slovenská republika: Smart People, 2018, 557 s. ISBN 978-80-973078-0-6.

KOLEKTIV AUTORŮ. Úvod do liberalizované energetiky: Trh s elektřinou, druhé aktualizované vydání. Praha: Asociace Energetických Manažerů, 2016, 553 s. ISBN 978-80-260-9212-4.

LABERGE, Robert. Datové sklady: agilní metody a business intelligence. Brno: Computer Press, 2012, ISBN 978-802-5137-291.

POUR, Jan, Miloš MARYŠKA a NOVOTNÝ, Ota. Business intelligence v podnikové praxi. Praha: Professional Publishing, 2012, 276 s. ISBN 978-80-7431-065-2.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2024/25

V Brně dne 9.2.2025

L. S.

---

doc. Ing. Vít Chlebovský, Ph.D.  
garant

---

prof. Ing. et Ing. Stanislav Škapa, Ph.D.  
děkan

## **Abstrakt**

Diplomová práce se zaměřuje na návrh reportního systému pro predikci spotřeby elektrické energie ve vybrané společnosti. Práce kombinuje teoretický úvod do Business Intelligence, specifika trhu s elektřinou v ČR a přehled použitých statistických metod. Analytická část hodnotí přesnost stávajících predikcí, obchodní chování na krátkodobém trhu a současná řešení reportingu predikcí. Závěrečná část představuje návrh reportního řešení v Power BI, včetně datového modelu a toku dat. Navrhnutý reportní systém by měl vést k finančním úsporám spojených s obchodem na krátkodobých trzích s elektřinou.

## **Klíčová slova**

Business Intelligence, datový sklad, proces ETL, dimenzionální modelování, reporting, vizualizace dat, statistická analýza, trh s elektřinou

## **Abstract**

The master's thesis focuses on the design of a reporting system for electricity consumption prediction in a selected company. It combines a theoretical introduction to Business Intelligence, an overview of the Czech electricity market, and a summary of the applied statistical methods. The analytical part evaluates the accuracy of current forecasts, trading behavior on the short-term electricity market, and existing reporting solutions. The final part presents a proposed reporting system in Power BI, including a data model and data flow. The proposed reporting system should lead to financial savings related to trading on short-term electricity markets.

## **Key words**

Business Intelligence, data warehouse, ETL process, dimensional modeling, reporting, data visualization, statistical analysis, electricity market

### **Bibliografická citace**

BUŠ, Jan. *Návrh reportovacího systému pro potřeby řízení predikce spotřeby elektrické energie*. [online]. Brno, 2025 [cit. 2025-05-19]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/165321>. Diplomová práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav managementu. Vedoucí práce Ing. Jiří Kříž, Ph.D.

### **Čestné prohlášení**

Prohlašuji, že předložená diplomová práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 19. 5. 2025

---

Bc. Jan Buš

autor

## **Poděkování**

Tímto bych rád vyjádřil své upřímné poděkování vedoucímu diplomové práce panu Ing. Jiřímu Křížovi, Ph.D., za cenné rady a doporučení, které mi byly velkým přínosem pro vypracování závěrečné práce. Dále děkuji svým kolegům v zaměstnání za ochotu, podnětné připomínky a pomoc při realizaci praktické části. V neposlední řadě děkuji své rodině a přátelům za podporu, motivaci a trpělivost po celou dobu studia.

# OBSAH

ÚVOD.....	11
CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ.....	12
1 TEORETICKÁ VÝCHODISKA PRÁCE.....	14
1.1 Představení Business Intelligence .....	14
1.1.1 Stručná historie Business Intelligence .....	14
1.1.2 Současné trendy v Business Intelligence.....	15
1.1.3 Funkce Business Intelligence a jeho využití.....	16
1.2 Procesy a architektura systémů Business Intelligence .....	17
1.2.1 Datové sklady .....	17
1.2.2 Dimenzionální modelování .....	22
1.2.3 Proces ETL – Integrovaní a transformační úlohy .....	28
1.3 Uživatelská aplikace Business Intelligence – Power BI.....	30
1.3.1 Komponenty Power BI .....	31
1.3.2 Datové zdroje, modelování a transformace v Power BI.....	31
1.3.3 Tvorba vizualizací a reportů.....	35
1.4 Trh s elektřinou.....	37
1.4.1 Účastníci trhu s elektřinou.....	38
1.4.2 Obchod s elektřinou.....	39
1.4.3 Odchyly a jejich zúčtování .....	42
1.5 Použité statistické metody .....	44
1.5.1 Deskriptivní statistika.....	44
1.5.2 Testování statistických hypotéz.....	46
1.5.3 Analýza závislostí.....	49
1.5.4 Metriky pro hodnocení přesnosti predikčních modelů' .....	51
2 ANALÝZA SOUČASNÉHO STAVU.....	53

2.1	Softwarová řešení společnosti reporting a administrativní procesy .....	53
2.1.1	Reportingová a analytická řešení.....	53
2.2	Predikce spotřeby elektřiny .....	54
2.2.1	Reporting predikce elektřiny .....	57
2.3	Datová struktura a logistika aplikace .....	58
2.4	Zhodnocení obchodní činnosti na krátkodobých trzích ve vztahu k odchylkám	59
2.4.1	Základní přehled zkoumaného datového souboru .....	59
2.4.2	Deskriptivní statistika .....	60
2.4.3	Hodnocení přesnosti predikce .....	67
2.4.4	Korelace ceny odchylky a velikosti odchylky .....	68
2.5	Zhodnocení analýzy současného stavu.....	69
3	VLASTNÍ NÁVRHY ŘEŠENÍ.....	72
3.1	Datový model a proces ETL.....	72
3.1.1	Integrace a transformace zdrojových dat.....	72
3.1.2	Výsledná podoba datového tržiště.....	80
3.2	Vlastní návrh reportního systému v prostředí Power BI .....	82
3.2.1	Datový tok .....	83
3.2.2	Uživatelské rozhraní sestavy .....	83
3.2.3	Definování metrik a KPIs .....	88
3.2.4	Struktura sestavy a popis stránek reportu .....	90
3.3	Zhodnocení návrhové části.....	100
3.3.1	Přínosy návrhové části.....	100
3.3.2	Ekonomické zhodnocení.....	101
3.3.3	Omezení návrhové části.....	103
	ZÁVĚR.....	105
	SEZNAM POUŽITÝCH ZROJŮ.....	108
	SEZNAM POUŽITÝCH ZKRATEK .....	112

SEZNAM POUŽITÝCH VZORCŮ.....	115
SEZNAM POUŽITÝCH OBRÁZKŮ.....	116
SEZNAM POUŽITÝCH GRAFŮ .....	118
SEZNAM POUŽITÝCH TABULEK .....	119
SEZNAM PŘÍLOH .....	120

## ÚVOD

Spotřeba elektrické energie patří ve výrobních podnicích mezi klíčové nákladové položky, zejména pokud jde o energeticky náročnou výrobu. V prostředí, kde se obchod s elektřinou odehrává i na krátkodobých trzích a ceny podléhají dynamickým výkyvům, hraje přesné plánování a predikce spotřeby zásadní roli. Pro efektivní rozhodování je proto nezbytné mít k dispozici kvalitní, aktuální a přehledně prezentovaná data.

Tato diplomová práce se zaměřuje na návrh reportovacího systému pro podporu řízení predikce spotřeby elektrické energie ve výrobní společnosti, která významnou část své spotřeby pokrývá nákupem elektřiny na krátkodobém trhu. Cílem práce je navrhnout reportovací nástroj, který využije různé datové zdroje a umožní jejich integraci a transformaci do analyticky využitelné podoby a poskytne managementu přehledné reporty pro podporu operativního, ale i strategického rozhodování.

Navrhované řešení vychází z principů Business Intelligence, které se touto problematikou zabývá. Součástí řešení je návrh logického rámce architektury datového tržiště, integračních a transformačních úloh a samotnému návrhu reportního nástroje v ekosystému Power BI od společnosti Microsoft, který vizualizuje klíčové ukazatele do uživatelsky přívětivé podoby.

Práce je členěna do tří částí. První část představuje teoretická východiska v oblasti Business Intelligence, problematiky fungování trhů s elektřinou v České republice, popis využitých statistických metod aplikovaných v analytické a částečně návrhové části práce. Druhá část práce se věnuje analýze aktuálního stavu ve zkoumané společnosti – konkrétně popisu obchodní činnosti na trhu s elektřinou a jejího aplikačního řešení pro obchodování na krátkodobých trzích. Součástí je také rozbor predikčního systému a jeho reportovacího řešení a následná statistická analýza dostupných dat s využitím deskriptivní statistiky, metrik pro hodnocení predikcí a korelační analýzy. Zjištění z analytické části slouží jako výchozí podklad pro třetí část práce, která se zaměřuje na návrh vlastního reportovacího řešení.

# CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ

Hlavním cílem práce je navrhnout efektivní reportní systém, který umožní kontrolu, analýzu a vizualizaci dat spojených s predikcí a skutečnou spotřebou elektrické energie. Tento systém bude podporovat procesy řízení predikce elektrické energie na různých úrovních agregace (od jednotlivých spotřebičů po celou společnost) a měl by zlepšit přesnost, rychlost a kvalitu rozhodování.

Díličními cíli, které podpoří dosažení hlavního cíle, jsou:

- Zhodnotit současný predikční systém a jeho reportingový nástroj pro hodnocení přesnosti predikcí spotřeby elektřiny a identifikovat hlavní slabiny.
- Analyzovat obchodní činnost společnosti na krátkodobých trzích s elektřinou ve vztahu k obchodované elektřině a cenám za vzniklé odchylky (v letech 2020-2023).
- Navrhnout konceptuální logický datový model a popsat hlavní integrační a transformační požadavky.
- Navrhnout strukturu a obsah reportního nástroje.
- Vyhodnotit tři fiktivní ekonomické scénáře zlepšení přesnosti predikcí spotřeby elektřiny a zjistit potenciální úporu Kč/MWh.

Pro realizaci cílů práce byla aplikována kombinace několika výzkumných metod, které zahrnovaly dokumentovou analýzu, kvantitativní výzkum sekundárních dat, návrh datových struktur a simulační modelování.

Hodnocení stávajícího predikčního systému a reportního nástroje bylo provedeno prostřednictvím dokumentové analýzy, přičemž pozornost byla věnována především uživatelskému rozhraní, vizuálním výstupům reportů, možnostem zobrazení dat a datové struktuře implementované v rámci používaného predikčního systému.

Analýza obchodní činnosti společnosti na krátkodobých trzích byla založena na kvantitativním výzkumu sekundárních dat získaných z dostupného reportingu operátora trhu ve formátu .xlsx. Tato data byla následně zpracována v prostředí Microsoft Excel pro základní operace, a dále v prostředí PyCharm pomocí jazyka Python a knihovny Pandas pro pokročilejší transformace. Statistické analýzy byly prováděny s využitím knihovny SciPy, zatímco vizualizace výsledků byly realizovány pomocí knihoven Matplotlib a Sunburst. Mezi aplikované statistické metody patřila deskriptivní statistika,

Kolmogorovův–Smirnovův test pro ověření normality dat, Spearmanova korelace a testování statistických hypotéz.

Návrh konceptuálního logického datového modelu vycházel z principů datového a dimenzionálního modelování, s přihlédnutím k požadavkům na ETL procesy (tj. extrakci, transformaci a načtení dat). Pro vizualizaci návrhu architektury a toku dat byly využity nástroje dbdraw.io a Miro, které sloužily pro modelování struktury dat a znázornění jednotlivých fází zpracování.

V další fázi byl vypracován návrh struktury a obsahu reportního nástroje, který vycházel z teoretických poznatků o možnostech vizualizace dat v prostředí Power BI a ze zjištěných nedostatků současného řešení. V návrhu byly využity možnosti samotného prostředí Power BI a jeho skriptovacího jazyka DAX pro výpočty metrik a klíčových ukazatelů výkonnosti.

Pro ekonomické vyhodnocení dopadů na zlepšení predikce byly sestaveny tři fiktivní ekonomické scénáře, které simulovaly zlepšení přesnosti predikcí na základě historických dat z let 2020–2023. Tato data byla zpracována v prostředí Microsoft Excel.

# 1 TEORETICKÁ VÝCHODISKA PRÁCE

## 1.1 Představení Business Intelligence

Business Intelligence (dále jen BI) představuje soubor znalostí, procesů, metod a technologií, jejichž cílem je usnadnit a zkvalitnit rozhodování v podnicích. Základem BI jsou především historická data podniku, která jsou zpravidla interní a strukturovaná<sup>1</sup>. Jejich analýza a zpracování ve spojení s dalšími analytickými metodami a nástroji umožňuje podporovat taktické i strategické rozhodování.

Velkou předností BI řešení je snadná a rychlá dostupnost přesných a aktuálních informací podnikovým uživatelům, kteří nemají hlubší znalosti v oblasti informačních technologií (dále jen IT). V tradičním prostředí jsou podniková data často uložena v komplexních databázových systémech, ke kterým mají přístup pouze IT specialisté. Ti však naopak nemusí rozumět obchodním procesům, strategickému řízení či klíčovými metrikám, které jsou pro efektivní rozhodování nezbytné.

BI systémy tento problém překonávají tím, že fungují jako most mezi světem IT a světem byznysu. Umožňují manažerům, analytikům a dalším rozhodovacím pracovníkům snadný přístup k relevantním datům s minimální časovou prodlevou (Laberge, 2012, str. 24-26).

### 1.1.1 Stručná historie Business Intelligence

Historie BI sahá do konce 50. let 20. století, kdy pojem poprvé použil výzkumník společnosti IBM, Hans Peter Luhn. Ve své definici vnímá podnik (Business) jako soubor aktivit, přičemž klíčovou roli hraje tok informací, který tvoří základ "intelligence system" (systému inteligence). Podle Luhna spočívá podstata BI ve schopnosti identifikovat vzájemné souvislosti mezi fakty a využít je k efektivnímu rozhodování a dosažení strategických cílů (Alexe a kol., 2014).

Reálné využití konceptu BI přinesl v roce 1989 analytik americké společnosti Gartner Group Howard Dresner, využitím dat informačních systémů pomocí centralizace v datových skladech. Datové sklady umožnily snadnější a rychlejší přístup k podnikovým

---

<sup>1</sup> Přestože tradiční BI systémy pracují s interními a strukturovanými daty, stále častěji pracují i s externími a nestrukturovanými daty.

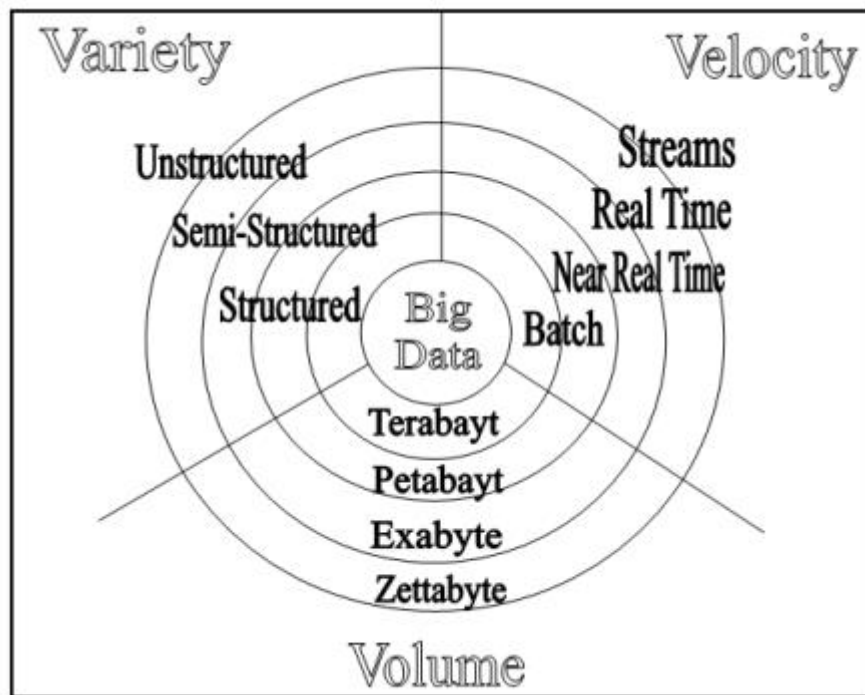
datům, která byla uložena na různých místech. S tímto přístupem přichází tzv. první generace BI (BI 1.0), kdy došlo k rozvoji procesů získávání a zpracování dat – proces ETL (*Extract, Transform, Load*) a softwarů na architektuře OLAP (*Online Analytical Processing*). BI se stává reálným softwarovým řešením a na trhu se objevují dodavatelé poskytující služby BI, které v té době primárně sloužilo k organizaci dat a jejich vizualizaci. Nabízená řešení byla nicméně složitá, pomalá a jejich tvorba a obsluha vyžadovala odborná školení. Na začátku 21. století přichází 2. generace – BI 2.0, která přinesla zásadní vylepšení analytických nástrojů, a především vznik cloudových BI platform, které umožnily jednodušší nasazení, širší využití a snížení závislosti na IT oddělení. Klíčovým průlomem bylo zpracování dat v reálném čase, což umožnilo firmám přijímat rozhodnutí na základě aktuálních informací, nikoli historických reportů. Tento posun přispěl k efektivnějšímu řízení podniků, rychlejšími reakcemi na změny na trhu a lepšímu porozumění chování zákazníků. Kromě toho se objevily samoobslužné BI nástroje, které umožnily běžným zaměstnancům bez pokročilých analytických znalostí pracovat s daty a vytvářet reporty bez nutnosti zásahu IT specialistů. Exponenciální růst internetu a cloud computingu hrál v tomto vývoji zásadní roli. Díky němu se BI technologie staly dostupnější, levnější a umožnily firmám využívat analytiku kdykoliv a odkudkoliv (Pavkov a kol., 2016).

### **1.1.2 Současné trendy v Business Intelligence**

V posledních letech prošla oblast BI zásadní transformací díky rychlému rozvoji umělé inteligence, strojového učení, Big Data a cloud computingu. Moderní BI se neomezuje pouze na reportování historických dat, ale zahrnuje real-time analýzu, prediktivní modelování a samoobslužné BI platformy, které usnadňují práci i netechnickým uživatelům (Rane a kol., 2024)

S expanzí technologií za poslední desetiletí roste i množství dat, která jsou v digitálním prostoru generována. Vzhledem k jejich obrovskému množství se označují jako „Big Data“ (*velká data nebo veledata, pozn. autora*). Tato data obsahují cenné informace, které podnikům mohou poskytovat konkurenční výhodu. Hlavními zdroji Big data jsou především online transakce, e-maily, audio soubory, obrázky, příspěvky na sociálních sítích, vyhledávací dotazy, senzorová data, mobilní aplikace apod. Big Data jsou definována 3 hlavními charakteristikami tzv. 3V:

- **Objem** (*Volume*), tedy velikost dat, která přesahuje petabajty (1 petabajt =  $10^{15}$  bajtů) a exabajty (1 exabajt =  $10^{18}$  bajtů), což vyžaduje nové přístupy zpracování dat.
- **Rychlost** (*Velocity*) generování a zpracování dat může být různorodé. Může probíhat v reálné čase, dávkovaně, streamováním atd.
- **Rozmanitost** (*Variety*) dat, která neumožňuje skladování dat v klasických databázových systémech, protože data nejsou pouze strukturovaná, ale i polostrukturovaná a nestruturovaná (využití tzv. datových jezer).



**Obrázek 1: Big Data – 3V**

(Zdroj: Sagioglu a Sinanc, 2013)

Big Data vyžaduje nové přístupy ke zpracování dat, jelikož tradiční databázové systémy nejsou schopny zvládnout jejich objem a složitost (Sagioglu a Sinanc, 2013).

### 1.1.3 Funkce Business Intelligence a jeho využití

Vhodné řešení BI by se dle R. Laberge mělo vyznačovat 5 vlastnostmi (2012, str. 25):

- **Použitelnost poznatků** – umožňuje okamžité a efektivní využití informací,
- **Rychlost** – rychle reaguje na požadavky uživatelů,
- **Aktuálnost** – poskytuje náhled na aktuální data,
- **Přesnost** – použitá data jsou přesná,
- **Užitečnost** – poskytuje hodnotu.

Řešení a výstupy řešení BI se mohou u různých společností lišit a mohou být realizovány mnoha metodami. Nicméně výběr vhodného řešení pro daný podnik je důležitý, protože vhodné řešení šetří zdroje. V ideálním případě by se měl podnik seznámit se všemi dostupnými využitelnými technologiemi. Pro technologii BI je nezbytně nutná dostupnost relevantních dat na jejichž základě bude podnik provádět rozhodnutí (Laberge, 2012, str. 28).

Technologie BI lze využít takřka ve všech oblastech. Nejen tedy pro podnikové činnosti, u kterých je nutné sledovat a analyticky vyhodnocovat hodnoty daných ukazatelů. Mezi klíčové oblasti aplikace BI patří finance, marketing, výroba, logistika, řízení vztahů s dodavateli, lidské zdroje, řízení výkonnosti, zákaznické vztahy, webová analytika, konkurenční zpravodajství a monitorování podnikových aktivit (Pour et al., 2012, str. 177).

## **1.2 Procesy a architektura systémů Business Intelligence**

Klíčovým prvkem systémů BI je zajištění kvality, správy a vzájemné kompatibility dat s nástroji určenými pro jejich prezentaci uživatelům. Datová úložiště a vizualizační nástroje by měly být navrženy tak, aby tvořily jednotný, harmonicky propojený systém. Struktura ukládání dat musí umožnit jejich efektivní archivaci, zároveň však zajistit snadný a rychlý přístup pro analytické a reportovací účely (Laberge, 2012, str. 33).

### **1.2.1 Datové sklady**

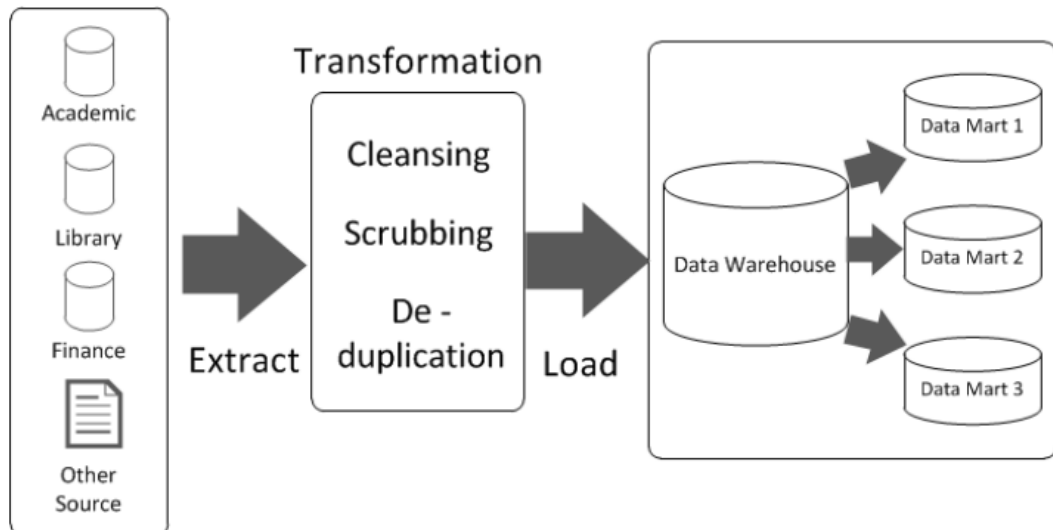
Systémy BI (s výjimkou predikčních modelů) klasicky nevytváří nová data, ale využívají data vytvořená klasickými transakčními aplikacemi (ERP, CRM apod.), jejich databáze se z pohledu BI považují za zdrojové. Tyto transakční systémy (architektura OLTP – Online Transactional Processing) primárně plní účel přístupu k detailním informacím, ukládání a aktualizaci dat. BI aplikace fungují na analytických systémech (OLAP – Online Analytical Processing), které jsou optimalizované na efektivní poskytování

analytických informací, kdy data v takovýchto systémech musejí být optimalizovány pro analytické úlohy (Pour et al., 2012, str. 17). Tyto analytické databáze se označují jako datové sklady.

Datový sklad (*data warehouse*) je systém, který umožňuje shromažďovat, organizovat, uchovávat a sdílet historická data. Datový sklad může být jedna velká centrální databáze, ale i sada různých databází. Společným jmenovatelem je nicméně centrální řízení. (Laberge, 2012, str. 37).

Architektura datového skladu vychází z požadavků na jeho konkrétní účel – tedy jaké informace má poskytovat a jakým způsobem budou tato data využívána. Klíčovým aspektem je tok dat, jehož cílem je získat z často různorodých zdrojů sjednocené, vyčištěné a transformované informace, které jsou vhodné pro další analytické zpracování. V oblasti datových skladů a toku dat se využívají dva způsoby přístupu pro architekturu datových skladů a tedy i toku dat – *top-down* (shora dolů) a *bottom-up* (zdola nahoru) (Laberge, 2012, str. 39-40).

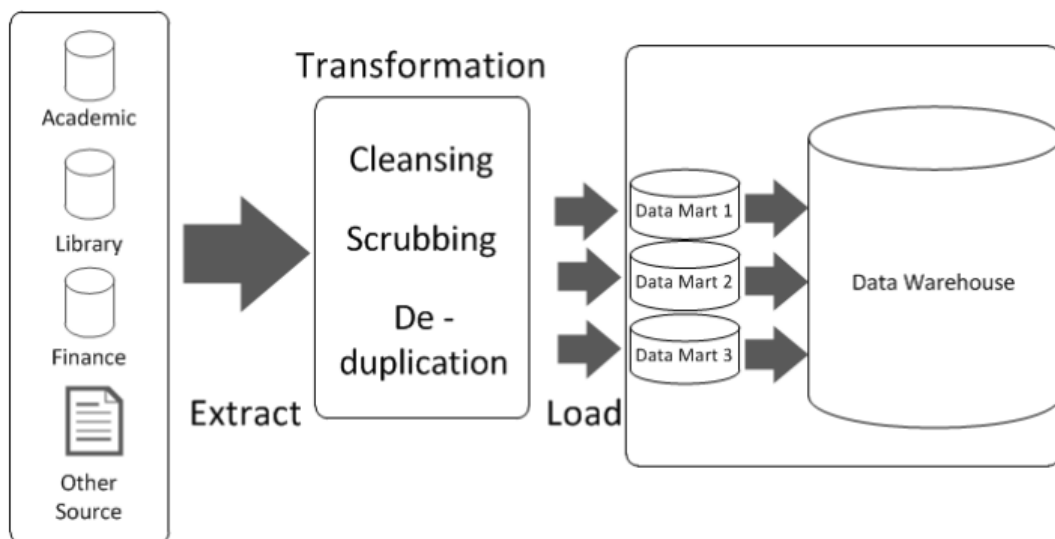
Přístup *top-down* je přístup Williama H. Inmona, který považuje datový sklad za centrální, integrované úložiště veškerých podnikových dat určených pro analytické účely. V rámci tohoto přístupu je nejprve vytvořen centrální datový sklad, do něhož jsou integrována a transformována data ze všech relevantních zdrojových systémů. Teprve následně se na základě tohoto centrálního skladu budují tematicky zaměřená datová tržiště, která slouží specifickým potřebám jednotlivých oddělení nebo uživatelských skupin (Moscoso-Zea et al., 2018).



**Obrázek 2: Inmonův přístup top-down**

(Zdroj: Moscoso-Zea et al., 2018)

Přístup *bottom-up*, který prosazuje Ralph Kimball, vychází z opačné logiky než Inmonův model. V rámci tohoto přístupu se nejprve vytvářejí datová tržiště, která pokrývají konkrétní analytické potřeby jednotlivých oddělení (např. finance, marketing, prodej). Tato tržiště jsou navržena tak, aby obsahovala vysoce dostupná a snadno analyzovatelná data v dimenzionální struktuře (např. hvězdicový model). Jednotlivá datová tržiště jsou následně propojena do logicky integrovaného celku, který lze vnímat jako distribuovaný datový sklad. Kimball klade důraz na rychlý přínos pro koncové uživatele, postupné rozšiřování a iterativní vývoj, což umožňuje pružně reagovat na změny požadavků (Moscoso-Zea, Paredes-Gualtor, a Luján-Mora, 2018).



**Obrázek 3: Kimballův přístup bottom-up**

(Zdroj: Moscoso-Zea et al., 2018)

Oba přístupy dosahují stejného cíle, tedy centralizace dat. Přičemž oba přístupy mají své výhody i nevýhody. V praxi bývá častěji uplatňován přístup Ralpha Kimballa, který díky své modularitě, nižší počáteční náročnosti a orientaci na rychlé doručení přínosů lépe odpovídá požadavkům většiny komerčních organizací. Přístup Williama H. Inmona je častěji využíván ve velkých nebo vysoce regulovaných organizacích, kde je kladen důraz na centralizaci, integritu dat a dlouhodobou udržitelnost řešení (Laberge, 2012, str. 40-41).

### 1.2.1.1 Komponenty datového skladu

Architektura datového skladu se skládá z několika navzájem propojených systémů, které jsou organizovány do vrstev. Tyto vrstvy lze chápat jako jednotlivé komponenty datového skladu a zároveň jako součásti datového toku – od získávání dat ze zdrojových (produkčních) systémů, přes jejich zpracování a uložení, až po finální využití v nástrojích BI. Každý systém plní specifickou roli v rámci této architektury a společně zajišťují spolehlivý, konzistentní a efektivní tok dat od zdroje ke koncovému uživateli. Z pohledu množství vrstev jsou děleny na (Laberge, 2012, str. 220-221):

- **Architektura s jednou vrstvou**
- **Architektura se dvěma vrstvami**
- **Architektura se třemi vrstvami**

**Architektura s jednou vrstvou** je nedoporučovanou a v současné době nevyžívanou architekturou. Její podstava spočívá v tom, že se jedná o přímý analytický náhled do produkčních systémů bez využití samostatné analytické vrstvy nebo datového skladu. Tento přímý přístup k produkčním datům je problematický, protože výrazně zatěžuje systémy určené primárně pro každodenní provoz a transakční zpracování.

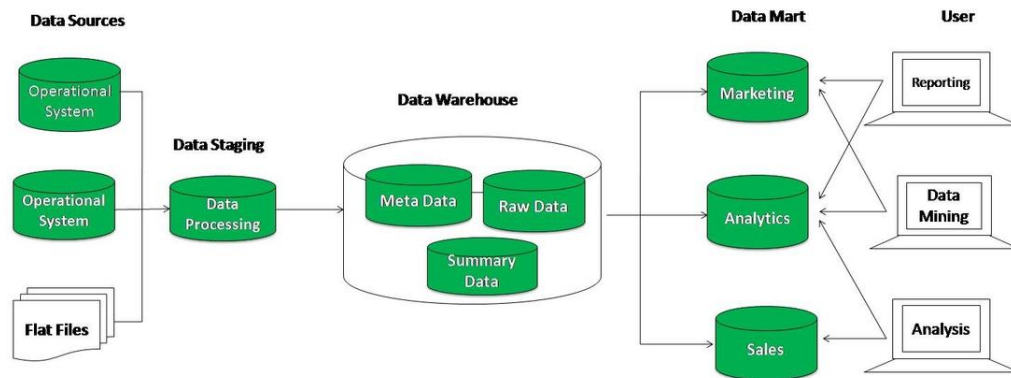
Produkční systémy mají sloužit k operativnímu fungování organizace, tedy ke zpracování objednávek, evidenci zákazníků nebo skladové evidenci. Pokud jsou však současně využívány k náročnému analytickému zpracování, dochází k jejich přetížení, zpomalování a v krajním případě i k narušení provozní stability. Z těchto důvodů je jednovrstvá architektura v současné praxi považována za nevhodnou a jsou využívány modely s oddělením transakčních a analytických systémů (Laberge, 2012, str. 220-221).

**Architektura se dvěma vrstvami** je klasickou architekturou datového skladu, kdy dochází k rozlišení mezi zdrojovými systémy a systémy datového skladu. Díky tomuto rozdělení je možné se zaměřit na analytický systém datového skladu samostatně a rozlišit tak dílčí komponenty, resp. vrstvy datového skladu, kterými jsou (Laberge, 2012, str. 220-221):

- Vrstva získávání dat
- Vrstva centralizovaných dat
- Distribuční vrstva
- Výkonnostní vrstva
- Vrstva uživatelské prezentace

V současné době je nejvyžívanější **třívrstvá (či vícevrstvá) architektura** datového skladu, která kromě zdrojových systémů a centrálního datového skladu doplňuje mezivrstvou úložiště provozních dat (*Data Staging Area*), která představuje dočasné úložiště provozních dat. Tato vrstva slouží jako přechodný prostor, do kterého jsou data nejprve načtena ve své původní podobě, a teprve následně procházejí transformačními úlohami (Laberge, 2012, str. 220-222).

Uložiště provozních dat umožňuje oddělit zdrojové systémy od zátěže spojené s datovým zpracováním, což výrazně zvyšuje stabilitu a výkon celého řešení. Tato vrstva poskytuje prostor pro validaci, čištění, mapování a slučování dat. Po dokončení transformačních úloh jsou data přenesena do centrálního datového skladu, kde jsou uložena ve strukturované a analyticky optimalizované podobě (Pour et al., 2012, str. 109).



**Obrázek 4: Architektura datového skladu (včetně vrstev)**

(Zdroj: GeeksforGeeks, 2023)

## 1.2.2 Dimenzionální modelování

Dimenzionální modelování představuje klíčový přístup k návrhu databázových struktur pro účely analytického zpracování dat. Na rozdíl od normalizovaných modelů, které jsou typické pro transakční systémy (OLTP), se dimenzionální modely zaměřují na jednoduchost, přehlednost a vysoký výkon při čtení dat v analytických systémech (OLAP). Jejich cílem je vytvořit takovou strukturu, která umožňuje rychlé agregace, filtrování a řazení nad velkým objemem historických dat uložených v datovém skladu a datových tržištích. Základním prvkem dimenzionálního modelování je rozdělení dat na dvě hlavní části – tabulky faktů a tabulky dimenzí (Pour et al., 2012, str. 67-68).

### 1.2.2.1 Tabulky faktů

Centrem daného systému je tabulka faktů (může jich být i více), která obsahuje kvantitativní a měřitelné podnikové údaje (ekonomické, výrobní, logistické atd.), které jsou identifikovatelné často velkým množstvím cizích klíčů dimenzionálních tabulek a často vlastním umělým primárním klíčem (*Surrogate key*). Metriky mají

granularitu (tj. úroveň detailu) primární dimenze (např. dimenze data) (Laberge, 2012, str. 155).

Určení úrovně granularity je klíčové z toho důvodu, že např. při zvolení nízké úrovně granularity (tedy nízké úrovně detailu) neumožňuje práci s detailními daty, než je úroveň granularity. Naopak vysoká úroveň detailu dat umožňuje analýzu na detailní úrovni, ale znamená i větší množství záznamů a tím i vyšší náročnost na výpočetní techniku. Tabulky faktů zabírají obvykle 90 % celkové kapacity datového skladu (zbylých 10 % zabírají tabulky dimenzí) (Pour et al., 2012, str. 109).

Podle toho, jakým způsobem lze s fakty pracovat v analytických dotazech, rozlišují se tři základní typy faktů: aditivní, semiaditivní a neaditivní fakty.

Aditivní fakty lze plně sčítat napříč všemi dimenzemi – časem, zákazníkem, produktem, regionem apod. Jsou to nejběžnější a nejjednodušejí zpracovatelné typy faktů, které tvoří základ většiny analýz. Semiaditivní fakty lze agregovat (např. sčítat) pouze přes některé dimenze, nikoliv přes všechny. Nejčastěji je omezením dimenze času – to znamená, že fakt lze sčítat například přes produkty nebo regiony, ale ne přes čas. Neaditivní fakty nelze sčítat přes žádnou dimenzi. Obvykle se jedná o procentuální, poměrové nebo vypočítané hodnoty, které ztrácí význam při přímé agregaci (Laberge, 2012, str. 156-157).

Dále se rozlišují typy tabulek faktů ve vztahu k jejich granularitě dat – transakční, periodické snímkové a akumulární snímkové tabulky.

Transakční faktová tabulka zaznamenává jednotlivé události (transakce) v systému – každá řádka odpovídá jedné události nebo akci (např. nákup, rezervace, přihlášení). Tento typ tabulky obsahuje časové razítko a je typicky velmi objemný, protože se do něj přidává záznam pokaždé, když nastane sledovaná akce. Periodická snímková tabulka zachycuje souhrnný (agregovaný) stav nebo výkon v pravidelných časových intervalech – např. denně, týdně nebo měsíčně. Akumulární snímková tabulka je specifickým typem, který zachycuje vícestupňový proces v čase – tedy záznam události, která prochází několika stavy. Na rozdíl od předchozích typů se záznam v tabulce postupně aktualizuje – doplňují se data o tom, kdy daný stav nastal (Pour et al., 2012, str. 72).

### 1.2.2.2 Tabulky dimenzí

Tabulky dimenzí tvoří klíčový prvek dimenzionálního modelu datového skladu. Na rozdíl od tabulek faktů, které obsahují kvantitativní, měřitelné údaje (např. obrat, počet objednávek), slouží dimenzionální tabulky k popisu kontextu těchto měření. Umožňují uživatelům analyzovat data z různých pohledů, jako je například zákazník, produkt, časové období nebo region. Zatímco u faktových tabulek je typické malé množství sloupců (atributů) a obrovské množství řádků (záznamů), u tabulek dimenzí je logika obrácená – typicky obsahují menší množství záznamů, ale širší škálu popisných atributů, které zvyšují srozumitelnost a analytickou využitelnost dat. Hodnoty atributů jsou převážně textové a diskrétní a měly by tak poskytovat co nejširší kontext a charakteristiky (Pour et al., 2012, str. 74-75).

Každý záznam v tabulce dimenzí představuje unikátní prvek určité dimenze (např. jednoho konkrétního zákazníka či produktu) a je jednoznačně identifikován pomocí vlastního primárního klíče, který je v tabulce faktů cizím klíčem a existuje tedy vztah, resp. relace 1:N, který zajišťuje podmínku spojení tabulky faktů a tabulky dimenzí. Tyto záznamy poskytují kontext pro číselné hodnoty ve faktové tabulce. Např. umožňují analyzovat výši tržeb podle typu zákazníka, produktové kategorie nebo země (Pour et al., 2012, str. 75).

Primární klíč dimenzionální tabulky by měl být tvořen tzv. náhradním klíčem (*Surrogate Key*), což je uměle vytvořený identifikátor, který je nezávislý na zdrojovém systému. Tento přístup zajišťuje konzistenci napříč datovým skladem a umožňuje efektivní správu historizace záznamů (např. v případě změn u zákazníka nebo produktu) (Laberge, 2012, str. 168-169).

**Potvrzená dimenze** je dimenze, která je sdílená napříč více datovými tržišti nebo faktovými tabulkami a má jednotnou strukturu a význam. Díky tomu je možné analyzovat různá fakta podle společných dimenzionálních atributů. Typickým příkladem jsou dimenze zákazníka, obchodu, produktu apod., které mohou být použity jak v prodejních datech, tak v jiných datových sadách např. reklamace (Laberge, 2012, str. 171).

**Dimenze data** je speciální typ dimenze, který popisuje časový kontext faktů. Obsahuje atributy jako den, týden, měsíc, kvartál, rok, pracovní den, státní svátek apod. Využívá se k časovým analýzám, trendům nebo porovnávání mezi obdobími.

**Dimenze času** se na rozdíl od dimenze data, která se zaměřuje na kalendářní členění, zabývá detailními časovými úseky v rámci jednoho dne – např. hodina, minuta, směna, časové pásmo (Laberge, 2012, str. 172-173).

**Degenerovaná dimenze** je atribut, který má charakter dimenze, ale není uložen v samostatné tabulce. Typicky jde o kód nebo číslo objednávky, které se ukládá přímo ve faktové tabulce, protože k němu neexistují další popisné informace. Zachovává jednoduchost modelu bez vytváření zbytečných dimenzí (Laberge, 2012, str. 174-175)

**Směsná dimenze** (*Junk Dimension*) vzniká sloučením více nízko-kardinálních atributů, které by samostatně netvořily plnohodnotné dimenze a jejich ponechání ve faktové tabulce by vedlo k jejímu zbytečnému nafukování a zpomalení dotazů. Typicky jde o různé příznaky, stavy nebo kategorie, které mají omezený počet hodnot. Například atributy „typ platby“ (hotově, kartou) a lze spojit do jedné směsné dimenze. Tím se sníží počet sloupců ve faktové tabulce, zachová se logická organizace dat a zlepší se přehlednost modelu. Směsné dimenze jsou běžnou technikou pro optimalizaci dimenzionálního modelu, zejména pokud nelze nebo není efektivní historizaci těchto atributů samostatně sledovat (Pour et al., 2012, str. 83-84).

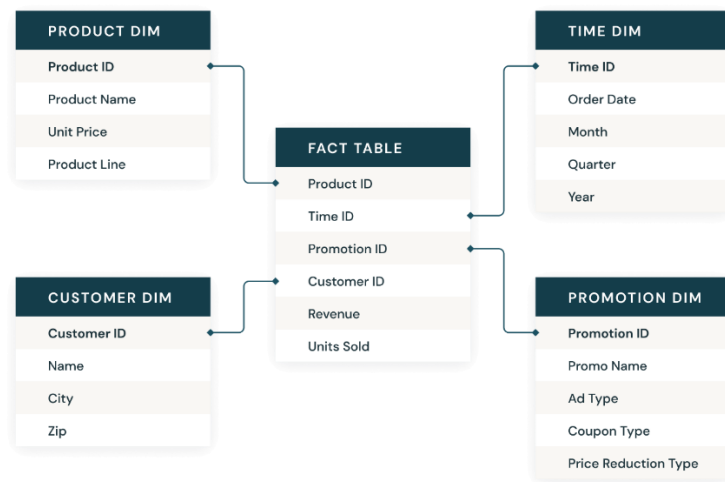
### 1.2.2.3 Struktura datového modelu

Na základě vztahů mezi tabulkami faktů a dimenzí lze datový model v datovém skladu nebo v datovém tržišti uspořádat různými způsoby. Volba konkrétní struktury závisí na požadavcích na výkonnost, srozumitelnost, udržitelnost a typy dotazů, které mají být na datech prováděny. Mezi nejběžnější modely patří hvězdicový model (*Star Schema*) a model sněhové vločky (*Snowflake Schema*) dimenzí (Laberge, 2012, str. 148-149).

**Hvězdicový model** je nejjednodušší a nejpřehlednější dimenzionální struktura, která se skládá z jedné centrální tabulky faktů a několika okolních tabulek dimenzí. Tyto dimenze nejsou dále normalizovány – obsahují veškeré atributy v jedné tabulce, což usnadňuje dotazování a podporuje vysoký výkon BI nástrojů. Nicméně u nadřazených úrovní hierarchie dané dimenze se hodnoty atributů vícenásobně opakují a dochází tak k redundanci. Pro praktické využití představuje nejlepší řešení především pro uživatelské aplikace. Schéma hvězdice je rychlejší v době odezvy pro poskytování výstupů díky omezenému výskytu spojovacích operací (*Join*) mezi tabulkami dimenze a faktovou tabulkou. Model je uživatelsky přehlednější, ale při častých změnách v hierarchii dimenze

je nutné jednu změnu promítnout do více řádků tabulky. Struktura připomíná hvězdu – tabulka faktů tvoří střed, od kterého „vyzařují“ jednotlivé dimenze. Díky jednoduchosti je hvězdicový model vhodný pro většinu analytických scénářů a je široce využíván v praxi. (Pour et al., 2012, str. 76-77).

### Star schema



**Obrázek 5: Schéma hvězdy**

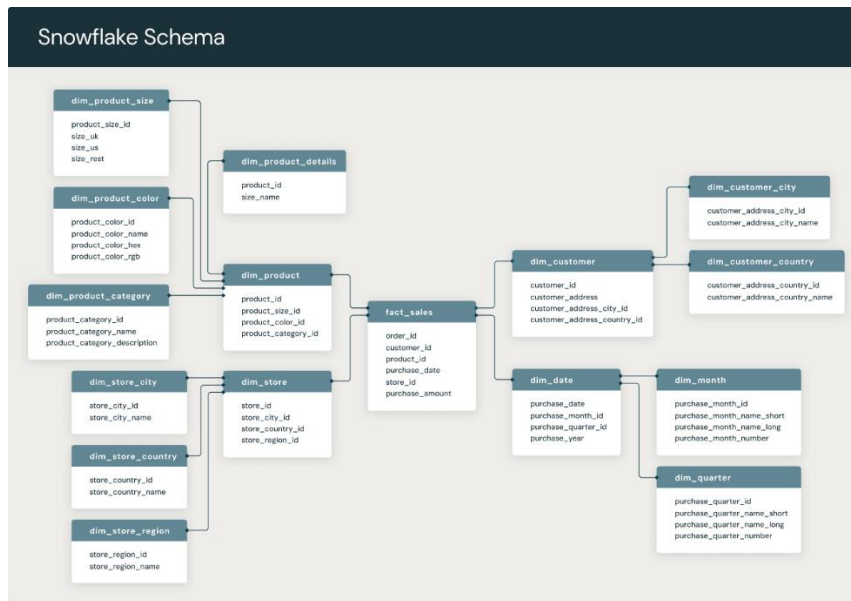
(Zdroj: Databricks, [n.d.])

#### 1.2.2.4 Model sněhové vločky

Schéma sněhové vločky je rozšířením hvězdicového modelu, které se vyznačuje normalizací dimenzionálních tabulek. Zatímco v hvězdicovém modelu má každá dimenze podobu jedné ploché tabulky s atributy, ve sněhové vločce jsou některé atributy vyčleněny do samostatných tabulek, se kterými jsou propojeny cizími klíči.

Tento přístup zároveň umožňuje přesnější modelování složitých hierarchií a vztahů mezi daty. Schéma sněhové vločky je také vhodné pro prostředí, kde dochází k častým aktualizacím dimenzí, protože díky normalizaci lze měnit referenční údaje efektivně a bez nutnosti rozsáhlých úprav. Výhodou je rovněž efektivita při práci s atributy s nízkou kardinalitou, které se často opakují – typicky např. kategorie produktů, regiony nebo stavy objednávek. Na druhou stranu má sněhové schéma i určité nevýhody. Patří mezi ně především složitější struktura modelu, která může být hůře pochopitelná zejména pro běžné uživatele nebo analytiku bez hlubších znalostí databázi. Z hlediska výkonu může být problematická i nutnost většího počtu spojovacích (join) operací mezi jednotlivými

tabulkami, což může zpomalit dotazy nad velkými datovými objemy (Pour et al., 2012, str. 77-79).

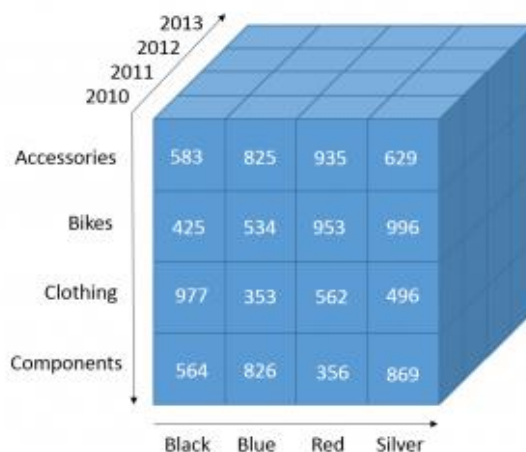


Obrázek 6: Schéma sněhové vločky

(Zdroj: Databricks, [n.d.]b)

### 1.2.2.5 OLAP Kostka

OLAP je technologie, která umožňuje rychlou analýzu multidimenzionálních dat. Klíčovým prvkem OLAP architektury je tzv. OLAP kostka (*OLAP cube*), která reprezentuje data v několika dimenzích a umožňuje uživatelům snadno provádět analytické operace. Zatímco tradiční relační databáze pracují s dvojrozměrným uspořádáním dat (tabulky), OLAP kostka poskytuje vícerozměrný model, který je přirozenější pro analytické dotazování, např. prodeje podle regionu, produktu a času (Pour et al., 2012, str. 22-23).



**Obrázek 7: OLAP kostka**

(Zdroj: Lutz, 2018)

Existují tři hlavní přístupy ke zpracování OLAP dat, které se liší v tom, kde a jak jsou data fyzicky uložena a jak se nad nimi provádějí analytické operace.

- **ROLAP (Relační OLAP)** pracuje s daty uloženými v klasických relačních databázích, přičemž OLAP logika (např. hierarchie, agregace) je implementována pomocí SQL dotazů. Tento přístup nevyžaduje speciální formát dat a může využívat již existující databázové systémy (např. SQL Server).
- **MOLAP (Multidimezionální OLAP)** ukládá data ve specializovaných multidimenzionálních databázích (např. Microsoft Analysis Services, IBM Cognos), kde jsou agregace předpočítány a uloženy ve formě OLAP kostky. Data jsou optimalizována pro velmi rychlé analytické dotazy.
- **HOLAP (Hybridní OLAP)** kombinuje přístupy ROLAP a MOLAP. Detailní data jsou uložena relačně (ROLAP), zatímco agregace jsou předpočítány a uloženy v multidimenzionální formě MOLAP (Pour et al., 2012, str. 23).

### 1.2.3 Proces ETL – Integrovační a transformační úlohy

Z důvodu různorodosti datových zdrojů musejí být data před vlastním nahráním do datového skladu upravena do jednotné podoby. Tento cíl plní proces datové pumpy, resp. ETL (*Extract, Transform, Load*), který představuje jednu z hlavních komponent architektury datového skladu. Proces ETL lze označit jako datový tok, kterým se dostávají zdrojová data do datového skladu. Zdrojová data musí být zpravidla transformována do požadovaného formátu a nahrána do cílové databáze, resp. datového skladu.

Transformaci je možné provádět jak na databázových systémech, tak na neformátových souborech, či na jejich kombinaci (Laberge, 2012, str. 241-243).

Tvorba procesu ETL, je často nejnáročnější částí daného projektu BI a představuje nejpracnější součást daného BI projektu (cca 60–80 % pracnosti). Transformační úlohy mají definovaný daný vstup a výstup. Je tedy nutné proces stavět na základě vstupních a výstupních systémů. Je tedy nutné provést dimenzionální analýzu a modelování.

Při tvorbě procesu ETL je nutné znát specifika zdrojových systémů (SQL databáze, Oracle, cloudové systémy, CRM, ERP apod.) nebo zdrojových souborů (dokumenty Excel, textové soubory) a jejich datové struktury, které jsou předmětem transformace do datového skladu. Je nutné si k zdrojovým systémům a dokumentům zajistit přístup (ODBC, OLE DB, API aj.) a identifikovat dílčí datové zdroje (tabulky v databázích a relevantní zdroje, API endpointy, listy v Excelu aj.). Pro identifikované zdrojové datové struktury je nutné určit cílové datové struktury (daný datový sklad či datové tržiště) a zajistit přístup a rozhraní k cílovým databázím (Pour et al., 2012, str. 111-112).

Transformační úlohy v rámci ETL procesu představují jednotlivé kroky, jejichž cílem je upravit, standardizovat či obohatit zdrojová data dle potřeb cílového systému. Rozsah těchto úloh je velmi široký a závisí na konkrétních požadavcích projektu i charakteru zpracovávaných dat. Mezi základní úpravy patří:

- Třídění dat
- Rozčlenění vstupních záznamů
- Sloučení dat
- Agregace dat
- Prosté kopírování datových struktur
- Změna kódování dat tabulky
- Změny formátu položek
- Změny datových typů
- Doplnění o nové atributy (například sloučením zdrojových atributů) aj. (Pour et al., 2012, str. 112-113)

Uvedení datového skladu do provozu zpravidla vyžaduje provedení počátečního načtení dat (*Initial Load*), které slouží k naplnění cílového úložiště historickými a aktuálními daty ze zdrojových systémů. Jedná se o jednorázový proces, při němž dochází ke kompletnímu

přenosu dat, nezávisle na době jejich vzniku nebo poslední změně. Počáteční načítání se typicky realizuje při nasazení nového datového modelu, migraci mezi systémy nebo změně datové architektury. Ve většině případů se načítá celý obsah relevantních zdrojových tabulek, případně agregované výřezy, které tvoří výchozí datový základ pro následné přírůstkové zpracování a analytické využití (Laberge, 2012, str. 247).

Po úvodním naplnění datového skladu následuje fáze pravidelné aktualizace dat, která je realizována formou přírůstkového (inkrementálního) načítání (*Delta Load*). Cílem tohoto procesu je přenášet do datového skladu pouze nové nebo změněné záznamy od posledního běhu ETL procesu. Tím se výrazně snižuje objem zpracovávaných dat, zkracuje celková doba zpracování a minimalizuje zátěž jak na zdrojové systémy, tak na cílovou databázi (Laberge, 2012, str. 247, 260).

Pro určení změn se nejčastěji využívá časové razítko, které zaznamenává datum a čas poslední změny záznamu. Alternativně lze použít také umělý primární klíč v kombinaci s hodnotou maximálního klíče načteného při předchozím běhu. Oba přístupy umožňují spolehlivě identifikovat přírůstky, aniž by bylo nutné zpracovávat celý objem zdrojových dat (Laberge, 2012, str. 260).

### **1.3 Uživatelská aplikace Business Intelligence – Power BI**

Pro zpracování a prezentaci dat z datových skladů a tržišť existuje široká škála nástrojů BI<sup>2</sup>, které umožňují tvorbu reportů, vizualizací a interaktivních dashboardů. Tyto aplikace tvoří poslední vrstvu BI architektury tzv. vrstvu uživatelské prezentace, která zprostředkovává přehledné a srozumitelné výstupy pro koncové uživatele bez nutnosti znalosti databázových jazyků.

Power BI je produktem společnosti Microsoft a představuje komplexní platformu pro tvorbu, správu a sdílení datových analýz – jde o ucelenou sadu služeb, aplikací a konektorů, které z různorodých zdrojů vytvářejí srozumitelné a vizuálně přitažlivé pohledy na data (Microsoft, 2024a).

---

<sup>2</sup> Pro účely této práce bude popsán právě systém Power BI, jelikož je i zvoleným nástrojem pro uživatelskou prezentaci dat v návrhové části práce.

### 1.3.1 Komponenty Power BI

Power BI není jen samostatným aplikačním řešením, ale kompletním ekosystémem propojených aplikací a služeb, které jsou schopny pokrýt všechny části datového toku – od jejich získání a transformace až po publikování a sdílení výstupů. Platforma byla původně navržena jako cloudová služba, nicméně dnes je dostupná desktopově a umožňuje i hybridní scénáře (Chmelár, 2018, str. 10).

**Power BI Desktop** je bezplatná desktopová aplikace sloužící jako hlavní vývojové prostředí pro návrh reportů a datových modelů. Umožňuje připojení ke zdrojům, transformaci dat a vytváření přehledných vizualizací (Chmelár, 2018, str. 10-11).

**Power BI Service** je cloudová platforma pro publikaci a sdílení reportů. Kromě správy reportů a dashboardů umožňuje i plánované aktualizace dat, notifikace, a tvorbu reportů přímo v prohlížeči bez nutnosti používat lokální aplikaci (Chmelár, 2018, str. 11-12).

**Power BI Data Gateway** propojuje cloudovou službu s on-premise zdroji dat a umožňuje bezpečný přenos dat. Podporuje automatickou aktualizaci bez zásahu uživatele. Existuje ve variantě pro jednotlivce (Personal) i organizace (Enterprise) (Chmelár, 2018, str. 13).

**Power BI Embedded** je nástroj pro vývojáře, kteří chtějí integrovat Power BI vizualizace do vlastních aplikací pomocí rozhraní REST API bez nutnosti přímého přístupu uživatelů do Power BI prostředí (Chmelár, 2018, str. 14).

**Power BI Premium** je placené řešení určené pro organizace s větším počtem uživatelů. Nabízí vyhrazené výpočetní prostředky, podporu rozsáhlých datových modelů a pokročilé funkce jako inkrementální načítání, dedikované refreshy nebo on-premise nasazení pomocí Power BI Report Serveru. Umožňuje zobrazování reportů i pro uživatele bez individuální licence, což jej činí nákladově efektivním pro větší firmy (Chmelár, 2018, str. 14-15).

### 1.3.2 Datové zdroje, modelování a transformace v Power BI

#### 1.3.2.1 Import dat a datové zdroje

Power BI umožňuje importovat data z více než 50 různých zdrojů (např. SQL Server, Oracle, Access, Excel, webové stránky, SharePoint, textové dokumenty apod.), která je dostupná skrze nástroj Power Query (Chmelár, 2018, str. 426).

Při importu dat z databázových systémů je dostupná v režimu „Import dat“ a „Direct Query“ (Deckler a Powell, 2022, str. 50). V režimu „Import dat“ dochází k tomu, že jsou data z daného datového zdroje – typicky relační databáze nebo datového skladu – načtena do paměti Power BI a uložena do sloupcového in-memory úložiště. Výsledkem je, že při interakci uživatele s reportem již nejsou dotazována externí data a před samotným načtením dat do modelu může uživatel provést různé transformační operace, jako jsou filtrování, výpočty nových sloupců nebo změny datových typů. Obnova dat v importním režimu probíhá plánovaně (dle nastavené periody obnovy v Power BI Service nejlépe s využitím inkrementální aktualizaci (delta nahrávání) (Deckler a Powell, 2022, str. 42-43,).

Vedle často používaného importního režimu nabízí Power BI také režim připojení Direct Query. Tento přístup umožňuje přímý dotaz na datový zdroj v reálném čase, bez načítání a ukládání dat do paměťového modelu Power BI. Místo toho jsou všechny požadované operace při interakci s reportem okamžitě překládány do nativních dotazů (např. SQL), které jsou odeslány na datový zdroj. Výsledky se pak zobrazují v reportu téměř okamžitě po provedení dotazu ve zdroji. Tento přístup se využívá při nutnosti přístupu k real-time zobrazování dat (Deckler a Powell, 2022, str. 53-55).

### **Power Query a jazyk M**

Power Query je nástroj společnosti Microsoft určený pro extrakci, transformaci a načítání dat. Umožňuje uživatelům připojit se k různým zdrojům dat, transformovat je a načítat do cílových systémů, jako jsou Excel, Power BI nebo Microsoft Dataverse. Power Query je integrován do několika produktů Microsoftu, včetně Excelu a Power BI (Chmelár, 2018, str. 426).

Editor Power Query je grafické uživatelské rozhraní, které umožňuje uživatelům provádět transformace dat bez nutnosti psaní kódu. Uživatelé mohou aplikovat různé transformace, jako je filtrování, seskupování, slučování nebo změna datových typů, prostřednictvím intuitivního rozhraní. Každý krok transformace je zaznamenán a lze jej upravit nebo zrušit. Editor také automaticky generuje kód v jazyce M, který lze zobrazit a upravit v pokročilém editoru. (Microsoft, 2025a).

Power Query využívá vlastní jazyk s názvem M, který je funkcionální, dynamicky typovaný a navržený pro práci s datovými dotazy. I když většina uživatelů pracuje

s Power Query prostřednictvím grafického rozhraní, pokročilí uživatelé mohou psát nebo upravovat M kód přímo pro složitější transformace nebo optimalizace (Microsoft, 2024b). Výhodou transformací dat v Power Query je, že veškeré transformace probíhají před samotným nahráním do datového modelu, čímž se eliminují problémy spojené s dodatečným zpracováním přímo v modelu nebo ve vizualizační vrstvě. Power Query a jazyk M nabízejí přes 500 funkcí pro transformaci a práci s daty. Editor dovoluje přidávat nové transformace, přejmenovávat je nebo dočasně zakazovat bez jejich odstranění. Kroky se dají jednoduše přeuspořádat přetažením a pravým tlačítkem lze vložit nový krok před či za vybranou transformaci (Chmelár, 2018, str. 426-427).

Nejlepší praxí je provádět transformační kroky co nejbližší zdroji dat, tedy v rámci datového skladu, a minimalizovat objem logiky realizované v Power Query a jazyku M (Deckler a Powell, 2022, str. 52).

### **1.3.2.2 Modelování dat – Zobrazení modelu**

Plátno Zobrazení modelu (*Model View*) v Power BI slouží jako interaktivní databázový diagram, který zpřehledňuje veškeré tabulky načtené do datové sady a jejich vzájemné vazby. Každá relace je zobrazena čarou doplněnou ikonami, jež okamžitě prozrazují kardinalitu (například 1:N u vztahu dimenze a faktů či 1:1 a N:M v méně obvyklých situacích), stav relace (plná čára znamená aktivní vazbu, přerušovaná čára pasivní vazbu). Úpravy relací probíhají přímo v diagramu, kde lze kdykoliv změnit kardinalitu, směr cross-filtrování i aktivitu vazby. Tím lze bezprostředně ovlivnit, jak se filtruje datový model, aniž by musel opouštět vizuální kontext. (Deckler a Powell, 2022, str. 124-126)

### **1.3.2.3 Data Analysis Expressions – DAX**

Data Analysis Expressions (dále jen DAX) je jazyk výrazu vzorce používaný ve službě Analysis Services, Power BI a Power Pivotu v Excelu. DAX vzorce zahrnují funkce, operátory a hodnoty pro provádění pokročilých výpočtů a dotazů na data v souvisejících tabulkách a sloupcích v tabulkových datových modelech. DAX umožňuje vytvářet výpočty v mírách, počítaných sloupcích, počítaných tabulkách a zabezpečení na úrovni řádků. Míry jsou vzorce dynamického výpočtu, ve kterých se výsledky mění v závislosti na kontextu. Počítané sloupce jsou sloupce přidáné do existující tabulky s definovanými

hodnotami pomocí DAX vzorce. Počítané tabulky jsou objekty založené na výrazu vzorce odvozeného ze všech nebo částí jiných tabulek ve stejném modelu. Zabezpečení na úrovni řádků umožňuje definovat, které řádky mohou být vráceny členy konkrétní role prostřednictvím výsledků dotazu. (Microsoft, 2025b).

V profesionálním prostředí by měly být všechny výpočty zabudovány právě do DAX výrazů, protože to podporuje opakovanou použitelnost, čitelnost a správu kódu. Kvalitní vývojář Power BI by měl být schopen přepisovat běžnou logiku z SQL nebo Excelu do DAXu. Díky schopnosti pracovat s proměnnými, referencemi na jiné metriky a funkcím pro práci s časem, texty a souhrny, je DAX robustním jazykem, který umožňuje vývoj škálovatelných a přesných modelů pro datové analýzy (Deckler a Powell, 2022, str. 177-182).

Vyhodnocení DAX měr v Power BI probíhá ve čtyřech hlavních krocích, které určují, jak jsou hodnoty metrik vypočítávány v rámci vizualizací. Každá metrika je počítána nezávisle, a to podle následujícího postupu:

1. **Počáteční filtr** (*Initial Filter Context*) - tento kontext zahrnuje veškeré filtry definované autorem reportu – ať už se jedná o vizuální výběry (slicery, řádky a sloupce v maticích) nebo o filtry nastavené na úrovni stránky, vizuálu či celého reportu.
2. **Úprava kontextu pomocí DAX** - základní metriky obvykle ponechávají filtr beze změny, ale složitější výrazy využívají funkci `CALCULATE()`, která umožňuje filtr rozšířit, nahradit nebo zcela předefinovat. Pokud dojde ke konfliktu mezi filtrem z reportu a filtrem v metrice, přednost má ten definovaný v DAX výrazu.
3. **Šíření filtrů přes relace** – aplikovaný filtr se dále propaguje do souvisejících tabulek podle směru definovaných relací. Standardně filtr z dimenzí ovlivňuje faktové tabulky, ale v některých případech se používá i obousměrné filtrování (např. při složitějších modelových vztazích).
4. **Výpočet logiky metriky** – samotný výraz metriky se poté vyhodnotí vůči aktuálnímu datovému výřezu. To může být součet, počet řádků, průměr nebo jiný výpočet podle funkce jako `SUMX()`, `COUNTROWS()` apod. (Deckler a Powell, 2022, str. 182-186).

Tento proces se opakuje zvláště pro každou hodnotu v reportu, což znamená, že složitější tabulky a vizualizace s velkým počtem hodnot mohou vyžadovat vyšší výpočetní výkon. V DAXu se rozlišují dvě základní skupiny funkcí, a to skalární funkce, které vracejí jednu hodnotu (např. SUM(), DISTINCTCOUNT(), LOOKUPVALUE()) a tabulkové funkce, které vracejí celou tabulku (např. FILTER(), ALL(), SUMMARIZECOLUMNS()), kterou lze využít jako vstup pro další výrazy, často prostřednictvím CALCULATE().

Metriky samotné nemohou vracet tabulky, ale mohou je využívat jako mezivýsledek v rámci výpočtu. Výstup tabulkových funkcí lze navíc využít i při tvorbě vypočtených tabulek nebo pokročilých vizualizací (např. stránkované reporty) (Deckler a Powell, 2022, str. 187-191).

Další důležitou součástí DAX jsou proměnné, které (klíčové slovo VAR) zvyšují přehlednost a efektivitu výpočtů. Proměnné jsou vyhodnoceny jen jednou a jejich hodnota může být znovu použita. Používají se k ukládání mezivýsledků (skalárních i tabulkových), což snižuje počet dotazů na datový model a zlepšuje výkon. Složité výrazy lze tímto způsobem rozdělit na části a výrazně zjednodušit jejich správu a údržbu.

Příklad použití ukazuje složitou metriku pro výpočet počtu zákazníků, kteří splňují více podmínek (např. rodinný stav nebo nákup určitého typu zboží v daném roce). Použití funkcí jako SUMMARIZE(), UNION() a DISTINCT() umožňuje kombinovat různé množiny zákazníků a upravovat filtrační kontext metriky (Deckler a Powell, 2022, str. 192-197).

### **1.3.3 Tvorba vizualizací a reportů**

Vizualizační vrstva patří k nejvýraznějším přednostem platformy Power BI, protože umožňuje prezentovat data v mnoha grafických formách a tím zásadně usnadňuje jejich interpretaci. Smyslem práce s daty není pouze provést výpočet, ale porozumět vypočteným výsledkům prostřednictvím vhodně zvoleného a správně nakonfigurovaného grafického zobrazení. Vizualizace představuje obecný typ grafického prvku, který má Power BI k dispozici ve svém katalogu – například sloupcový, koláčový či čárový graf, mapu, tabulku, matici nebo průřez. Lze si ji představit jako předem definovanou šablonu či konstrukční plán, který určuje, jak má výsledný prvek vypadat a jaké datové vstupy od něj očekávat. Tvorba zcela nových typů vizualizací je vyhrazena vývojářům a běžný uživatel Power BI Desktop pracuje s nabídkou, kterou mu platforma standardně

poskytuje. Naproti tomu vizuál je konkrétní instance vizualizace vykreslená na plátně reportu. Uživatel do její konfigurace přiřazuje příslušné sloupce či metriky, upravuje formátování a nastavuje interakce, čímž instanci přizpůsobuje potřebám analytického scénáře (Chmelár, 2018, str. 218-219)

V Power BI je vše, co se na stránce reportu nachází – grafy, tabulky, mapy, ale i průřezy nebo další ovládací prvky (např. tlačítka). Všechny tyto objekty zpracovává vykreslovací modul Power View, který zajišťuje jednotné chování napříč různými typy zobrazení průřezy (*Slicers*), které uživatelé často vnímají spíše jako interaktivní filtry, jsou z technologického hlediska rovněž vizuály (Chmelár, 2018, str. 219-220).

Při tvorbě reportů v Power BI je vhodné nepřistupovat k práci nahodile, ale řídit se ověřenými principy návrhu uživatelského rozhraní a datové prezentace. Důležitým krokem v tomto procesu je plánování rozvržení stránky reportu, které pomáhá zachovat logiku a přehlednost zobrazených dat.

Jednoduchým, ale velmi účinným nástrojem pro plánování reportu, je náčrt rozložení vizuálů. Takový náčrt definuje:

- Umístění hlavních metrik (např. tržby, marže) v levém horním rohu reportu.
- Prostor pro průřezy (*slicers*) pod klíčovými ukazateli pro snadné filtrování.
- Větší prostor uprostřed stránky pro vizuály s nejvyšší analytickou hodnotou (Deckler a Powell, 2022, str. 376-380).

Důležitou částí tvorby reportu je definování Klíčových ukazatelů výkonu (KPI), které představují základní nástroje pro měření výkonnosti procesů, oddělení i celé organizace. Jedná se o kvantifikovatelné metriky, které slouží ke sledování plnění strategických, taktických i operativních cílů. Správně zvolené KPI poskytují podložené informace, které pomáhají organizaci řídit se daty a přijímat informovaná rozhodnutí. Mezi nejběžnější KPIs patří například celkový obrat, zisková marže, produktivita zaměstnanců, míra reklamací nebo zákaznická spokojenost. Cílem KPIs však není pouze měřit,

ale především řídit. Umožňují sledovat pokrok vůči stanoveným cílům, včas identifikovat odchylky a potenciální problémy, a tvoří základ pro strategická rozhodnutí (Deckler a Powell, 2022, str. 199-203).

Jednou z pokročilých funkcí Power BI, která významně rozšiřuje možnosti prezentace a práce s reporty, je systém záložek (Bookmarks). Záložky umožňují uložit konkrétní stav stránky reportu, včetně výběru filtrů, viditelnosti vizuálů a dalších nastavení, a tuto konfiguraci pak kdykoliv znovu vyvolat. Díky tomu lze vytvářet aplikaci podobné scénáře nebo prezentace dat, ve kterých se uživatel interaktivně pohybuje mezi předdefinovanými stavy reportu (Deckler a Powell, 2022, str. 267-269).

## **1.4 Trh s elektřinou**

Elektrická energie (dále jen elektřina) je z fyzikálního hlediska vyráběna výrobcem, přenášena přenosnou sítí z místa výroby do distribuční sítě až ke spotřebiteli. V tržním pojetí je elektřina vnímána jako hromadná komodita. Elektřina nespĺňuje charakteristiky normálního zboží, kdy v momentě dodání zboží nelze objednávku měnit. Elektřina je jako síťová komodita zákazníkem odebírána podle toho kolik ji zákazník v daném čase potřebuje. Dodávání elektřiny přes přenosnou či dále přes distribuční síť má de facto nulové zpoždění, což je její zásadní vlastností. Elektřina tak musí být vyráběna v momentě, kdy dochází k její spotřebě. Tento faktor vyžaduje složité mechanismy systémového a dispečerského řízení, které zajišťuje neustálou rovnováhu mezi výrobou a spotřebou (Kolektiv autorů, 2016, str. 13–15).

Trh s elektřinou v České republice (dále jen ČR) podléhá regulaci evropského a českého práva. Evropská legislativa pro trh s elektřinou propojuje energetickou politiku s ochranou hospodářské soutěže, spotřebitelů a životního prostředí. Lisabonská smlouva posílila pravomoci Evropské unie (dále jen EU) v regulaci trhu, zavádí sdílené a výlučné pravomoci a podporuje volný pohyb elektřiny a služeb. Energetická politika EU usiluje o propojení sítí, dekarbonizaci a ochranu spotřebitelů, přičemž národní legislativa musí být v souladu s evropským právem (Kolektiv autorů, 2016, str. 35–41).

ČR má regulaci trhu s elektřinou ukotvenou především v Energetickém zákoně a ve vyhlášce o Pravidlech trhu s elektřinou. Český energetický zákon (zákon č. 458/2000 Sb.) je hlavním právním rámcem regulace trhu s elektřinou zahrnující licencování, ochranu spotřebitelů a řízení elektrizační soustavy. Klíčovou rolí v oblasti cenové regulace a dohledu nad trhem hraje Energetický regulační úřad (dále jen ERÚ). Zákon podporuje nezávislost přenosových a distribučních soustav a rozvoj obnovitelných zdrojů energie prostřednictvím finančních pobídek a přednostního připojení. Spotřebitelé mají

zajištěnou ochranu díky přísnějším pravidlům pro transparentnost a cenovou regulaci. Legislativa se pravidelně přizpůsobuje evropským standardům pro liberalizovaný trh s energií (Kolektiv autorů, 2016, str. 55–58).

### 1.4.1 Účastníci trhu s elektřinou

Jak již bylo nastíněno výše, trh s elektřinou je silně regulovaný, a tím pádem je regulace uplatňována na účastníky tohoto trhu. Z obchodního hlediska je základní vztah výrobce a spotřebitele, které spojuje zprostředkovatel (obchodník s elektřinou). Elektřina je jako komodita obchodována na burze, kde se klasicky potkává nabídka s poptávkou a probíhají tak organizované obchody. Nejméně častou možností obchodního vztahu je obchod zákazníka přímo s výrobcem. Tato situace není čítná a probíhá v situacích u dlouhodobých dodávek mezi větším výrobcem a velkým spotřebitelem (Kolektiv autorů, 2016, str. 55–58).

Každý subjekt (účastník) na energetickém trhu je definován jako registrovaný účastník trhu v informačním systému operátora trhu s elektřinou a je identifikován výjimečným registračním číslem přiděleným operátorem trhu. Registrovaný účastník má právo přístupu k sítím. Subjekt zúčtování je registrovaný účastník trhu, který má přístup sítím a na rozdíl od běžného registrovaného účastníka zodpovídá za odchylky sjednané a naměřené energie v jednotlivých zúčtovacích intervalech (hodinách) daného obchodního dne a má přístup k trhům s elektřinou, účasti na přeshraničních transakcích a neomezeného obchodování s ostatními subjekty zúčtování v rámci jeho finančního zajištění u OTE. Stává se subjektem zúčtování uzavřením smlouvy s operátorem trhu, který následně zajišťuje vyhodnocení, zúčtování a vypořádání jeho odchylek (Kolektiv autorů, 2016, str. 78–79).

Role účastníků lze rozdělit na výrobce, provozovatele přenosových soustav, provozovatele distribučních soustav, obchodníky s elektřinou, spotřebitele, organizátory trhu (burzy). Další významnou roli hrají regulační a správní orgány (Kolektiv autorů, 2016, str. 78).

**Provozovatel přenosové soustavy – ČEPS, a.s.** je výhradní provozovatel přenosové soustavy v ČR dle licence na přenos elektřiny, kterou uděluje ERÚ (podle Energetického zákona). Je zodpovědný za spolehlivý provoz a rozvoj přenosové soustavy o napětí 400 kV a 220 kV, vyrovnává výrobu a spotřebu v reálném čase a poskytuje systémové služby.

Kromě toho umožňuje přeshraniční přenosy elektřiny pro export, import, tranzit a aktivně přispívá k rozvoji liberalizovaného trhu s elektřinou v ČR i Evropě (ČEPS, n.d.).

**Operátor trhu s elektřinou – OTE, a.s.** (dále jen OTE) je společnost, jejímž jediným akcionářem je ČR. Jeho hlavní úlohou je registrace účastníků trhu a jejich obchodních diagramů, vyhodnocování a zúčtování odchylek (rozdíl mezi sjednaným a skutečně odebraným nebo vyrobeným množstvím elektřiny) velkoobchodních účastníků. OTE také spravuje data o nabídkách a poptávkách elektřiny, zpracovává měřené hodnoty a typové diagramy a poskytuje podklady pro zúčtování a fakturaci. OTE plní funkci organizátora krátkodobého trhu s elektřinou (denní, vnitrodenní a vyrovnávací) (OTE, 2024a, str. 6-7). Další účastníci na energetickém trhu jsou obchodníci a výrobci.

**Obchodník s elektřinou** je subjekt s licenci na obchodování s elektřinou, který ji nakupuje a prodává. Má práva na přístup k síti, účast na trhu, obchodování s elektřinou a získávání relevantních informací. Zároveň má povinnosti vůči OTE (registrace, zúčtování, poskytování dat), provozovatelům soustav (např. dodávání informací pro provoz) a konečným zákazníkům, pokud působí jako dodavatel (ČR, zákon č. 458/2000 Sb.).

#### **1.4.2 Obchod s elektřinou**

Obchod s elektřinou lze rozdělit na maloobchod a velkoobchod. Na maloobchodním trhu je dodávána elektřina koncovému spotřebiteli (domácnosti a podniky). Na velkoobchodním trhu probíhají transakce především za účelem dalších transakcí a obchody tedy nejsou primárně určeny koncovému spotřebiteli (což není nutné pravidlo). Pro potřeby problematiky bude dále rozebírán pouze velkoobchodní trh, který se dělí podle délky kontraktů nebo jejich charakterů (Kolektiv autorů, 2016, str. 81-82).

**Trhy s dlouhodobými produkty** jsou koncipovány na obchody s elektřinou na dobu delší než jeden měsíc. Klasicky se provádí formou bilaterálních dohod (termínové trhy – forwards), tržních produktů s fyzickou dodávkou (v době dodávky se elektřina odebírá/dodává přímo fyzicky z burzy) nebo formou produktů s finančním vypořádáním (futures) (Kolektiv autorů, 2016, str. 82).

**Krátkodobé trhy** jsou v ČR organizovány operátorem trhu (OTE a.s.). Na tyto trhy mají přístup registrovaní účastníci trhu, kteří jsou zároveň subjekty zúčtování (zodpovědnost za odchylky). Fungování krátkodobých trhů upravuje vyhláška č. 541/205 Sb.,

o Pravidlech trhu s elektřinou. Pravidla obchodování zahrnují podávání nabídek, uzavírání obchodů, zúčtování a vypořádání transakcí na jednotlivých segmentech trhu. Obchodovaná elektřina se předává v elektrizační soustavě ČR nebo v zahraniční soustavě zapojené do společného krátkodobého trhu pokrývajícího tržní oblast ČR i zahraniční oblasti. Krátkodobé trhy se dělí podle charakteru a délky kontraktů na denní, vnitrodenní a vyrovnávací trh.

**Denní trh** často označovaný jako spotový trh zahrnuje obchody na dobu jednoho dne, tzn. obchod proběhne den před dodáním elektřiny. Denní trh funguje formou anonymní aukce, kde se na základě nabídek a poptávek v eurech stanovují výsledky pro 24 obchodních hodin následujícího dne (Kolektiv autorů, 2016, str. 91). Výsledky zahrnují:

- Přenos elektřiny do/ze zahraničí při propojení české tržní oblasti s okolními oblastmi.
- Konečnou cenu elektřiny na denním trhu.
- Objem zobchodované elektřiny.

Finanční vypořádání obchodů probíhá následující pracovní den po dni dodávky podle sjednaných smluvních podmínek.

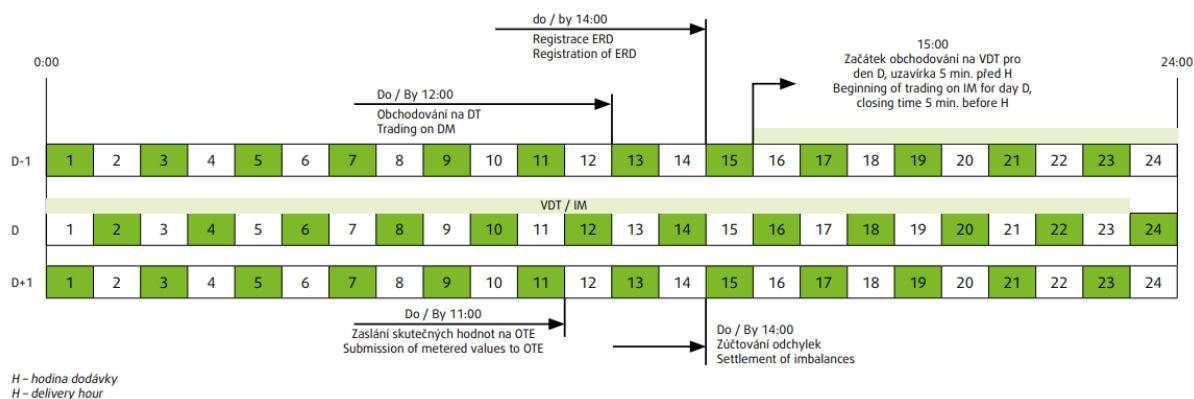
**Vnitrodenní trh** umožňuje obchodníkům flexibilně vyrovnávat svou obchodní pozici těsně před hodinou dodávky, což pomáhá řešit přebytky nebo nedostatky elektřiny a přispívá k optimálnímu fungování elektrizační soustavy. Tento trh je organizován pro jednotlivé hodiny daného obchodního dne a otevírá se již od 15 hodin předcházejícího dne. Nabídky a poptávky lze podávat až do jedné hodiny před konkrétní obchodní hodinou. Obchody na vnitrodenním trhu jsou zahrnuty do vyhodnocení a zúčtování odchylek operátorem trhu (Kolektiv autorů, 2016, str. 92).

Krátkodobé trhy představují významný způsob obchodování s elektřinou. V roce 2023 bylo prostřednictvím denního trhu zobchodováno 25,917 TWh a vnitrodenního trhu 5,598 TWh elektřiny, což představuje 33,21 % veškerých obchodů prostřednictvím operátora trhu OTE (OTE, 2024c).

**Trh s regulační energií** je klíčovým nástrojem pro vyrovnávání nerovnováhy mezi výrobou a spotřebou elektřiny v elektrizační soustavě. Tento proces zajišťuje provozovatel přenosové soustavy, společnost ČEPS, která působí jako iniciátor trhu a jedna z protistran. Regulační energie může být opatřena třemi způsoby:

aktivací podpůrných služeb, nákupem na vyrovnávacím trhu organizovaném OTE a dovozem elektřiny ze zahraničí. Regulační energie z podpůrných služeb je poskytována certifikovanými výrobními bloky, které mají oprávnění na jejich poskytování. Dodávky této energie jsou zúčtovány operátorem trhu na základě smlouvy mezi poskytovatelem podpůrných služeb a operátorem (Kolektiv autorů, 2016, str. 93).

**Vyrovnávací trh** umožňuje účastníkům nabízet kladnou i zápornou regulační energii s cílem pomoci vyrovnat systémovou odchylku. Nabídky mohou být podány až hodinu před začátkem dodávky, přičemž trh se uzavírá 30 minut předem. Na trhu se mohou účastnit i registrované uživatelské tarify, pokud mají souhlas subjektu zúčtování, který za ně přebírá odpovědnost, a pokud mají smlouvu o přístupu na vyrovnávací trh s OTE. Přijaté nabídky na regulační energii jsou následně zahrnuty do standardního systému vyhodnocování a zúčtování odchylek (Kolektiv autorů, 2016, str. 93).



**Obrázek 8: Realizační diagram (ERD)**

(Zdroj: OTE, 2024c)

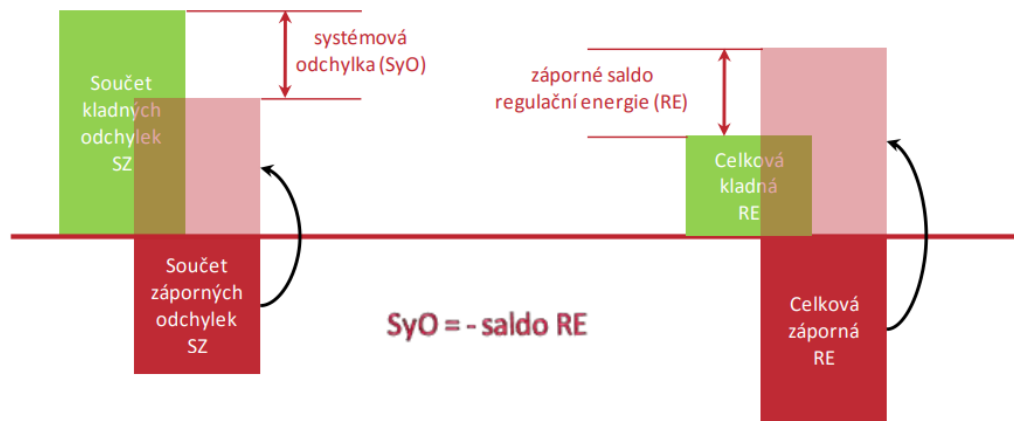
Na krátkodobých trzích jsou registrovány všechny dvoustranné obchody účastníků trhu v systému OTE prostřednictvím tzv. realizačního diagramu (ERD). Součástí systému je i zúčtování odchylek a vyrovnávacího trhu. ERD (viz obrázek 8) poskytuje přehled o časovém rozvrhu obchodování na krátkodobých trzích s elektřinou, který zahrnuje fáze obchodování (denní a vnitrodenní trhy) a zúčtování odchylek. Každý den má specifické časové intervaly pro podávání nabídek, obchodování a následné zúčtování skutečně spotřebované či vyrobené elektřiny (OTE, 2024c, str. 14-15).

### 1.4.3 Odchylyky a jejich zúčtování

Jak již bylo nastíněno v předchozí kapitole, kolik elektřiny se vyrábí, tolik se jí spotřebovává, ale výrobu ani spotřebu není možné naprosto přesně naplánovat. Vznikají tak odchylyky z důvodů poruch, počasí a skutečnosti, že spotřeba elektřiny je náhodná veličina, kterou jednoduše nelze přesně odhadovat. Kvůli odchylkám se musí elektrizační soustava neustále vyrovnávat, k čemuž slouží regulační energie, která je do sítě dodávána nebo odebírána k vyrovnání odchylek (Kolektiv autorů, 2016, str. 235-236).

Regulační energie ale není bezplatná, protože subjekty, které se na vyrovnání podílejí, mají nárok na finanční odměnu za poskytování kladné nebo záporné regulační energie. Tyto náklady v liberalizovaném trhu s elektřinou hradí účastníci trhu, kteří nesou zodpovědnost za odchylku (subjekty zúčtování), a to dle jejich podílu na celkové odchylce vzniklé v elektrizační soustavě (systémová odchylka). Je důležité zmínit, že ne všichni účastníci trhu nesou zodpovědnost za odchylku. Takový účastník uzavírá smlouvu s jedním subjektem zúčtování o přenesení zodpovědnosti a jeho dodávky, odběry nebo obchodní pozice zaneseny do bilance daného subjektu zúčtování (Kolektiv autorů, 2016, str. 231-233). Způsob zúčtování regulační energie a odchylek je legislativně ukotven ve vyhlášce č. 408/2015 Sb., o Pravidlech trhu s elektřinou v Příloze č. 8.

Vypořádání odchylek provádí operátor trhu (OTE), který na základě pozic účastníků trhu (závazky dodat nebo odebrat elektřinu ve stanoveném čase) a jejich skutečně dodané nebo odebrané energii vyhodnocuje individuální odchylku účastníka trhu. V reálném čase, tedy v obchodním intervalu (1 hodina), dochází k dodávce regulační energie (kladná nebo záporná). Na základě obchodních pozic regulační energie a odchylek je operátorem trhu stanovena cena individuální odchylky účastníkům trhu, kteří nesou zodpovědnost za odchylku ve vyhodnocovací periodě (V ČR 1 hodina) (Vyhláška č. 408/2015).



**Obrázek 9: Vztah systémové odchylky a regulační elektřiny**

(Zdroj: Kolektiv autorů, 2016)

Individuální odchylky mohou být kladné (vyrobilo se více nebo spotřebovalo méně, než bylo plánováno.), nebo záporná (vyrobilo se méně nebo spotřebovalo více, než bylo plánováno.). Agregace všech individuálních odchylek vytváří systémovou odchylku. V případě, že individuální odchylka subjektu napomáhá vyrovnat systémovou odchylku (má opačný směr než systémová odchylka) jedná se o tzv. protiodchylku (Kolektiv autorů, 2016, str. 233-234).

Zúčtování cen odchylek a protiodchylek funguje dle směru odchylky. Pokud individuální odchylka zhoršuje situaci (je ve směru systémové odchylky), subjekt zúčtování je penalizován, pokud odchylka situaci zlepšuje dostane finanční odměnu. Výše odměn a penalizací je stanovena vyhláškou č 408/2015 sb. Zúčtovací cena odchylky (ZC) odpovídá nejdražší nabídkové ceně regulační energie použité k vyrovnání systémové odchylky, přičemž pokud je tato cena příliš nízká, použije se minimální zúčtovací cena stanovená ERÚ. Zúčtovací cena protiodchylky (ZCproti) odpovídá průměrné ceně regulační energie aktivované v opačném směru než systémová odchylka (Vyhláška č. 408/2015). Cílem tohoto určení cen odchylek a protiodchylek je motivace subjektů zúčtování k samoregulaci, tedy k minimalizaci odchylek, případně vedení individuální odchylky do pozice protiodchylky. Důvodem je především skutečnost, že regulační energie a její poskytovatelé nemají neomezené možnosti a při neudržení systémové rovnováhy elektrizační soustavy dochází k vyhlášení předcházení stavu nouze a nařízení omezování spotřeby. Nejhorší scénář představuje výpadek elektřiny nebo

přímo dlouhý rozsáhlý výpadek v elektrizační síti tzv. „blackout“ (Kolektiv autorů, 2016, str. 242).

## 1.5 Použité statistické metody

V této části práce je představen teoretický základ statistických ukazatelů a metod, které jsou využity v analytické a návrhové části. Cílem je vysvětlit principy a předpoklady jednotlivých metod a poskytnout tak potřebný kontext pro pochopení jejich aplikace v praktické části práce.

### 1.5.1 Deskriptivní statistika

Deskriptivní (popisná) statistika představuje soubor základních statistických ukazatelů, grafů a tabulek, které slouží k sumarizaci, popisu a první orientaci ve zkoumaném výběrovém souboru (Hendl, 2022, str. 318). Tento soubor se skládá z hodnot náhodných veličin, které při opakovaném měření nebo pozorování mohou nabývat různých hodnot s určitou pravděpodobností (Janáček, 2022, str. 13).

#### 1.5.1.1 Míry centrální tendence

Míry centrální tendence představují jednu z charakteristik náhodné veličiny a vyjadřují centrum a polohu kardinálních (číselných) proměnných, tedy v jaké poloze datového souboru se nachází střed. Nejčastěji se používá aritmetický průměr, medián a modus (Hendl, 2012, str. 99).

Nejčastěji využívanou charakteristikou využívanou pro odhad střední hodnoty kardinálních proměnných je **aritmetický průměr** ( $\bar{x}$ ), kdy jsou všechny hodnoty datového souboru vyděleny jejich počtem. Vztah aritmetického průměru je následující (Hendl, 2012, str. 99):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Kde:

$n$  – počet hodnot ve výběrovém, resp. datovém souboru

$x_i$  –  $i$ -tá hodnota náhodné veličiny

**Medián** ( $Me$ ) představuje další charakteristiku centrální tendence, která rozděluje výběrový soubor uspořádaný podle velikosti na dvě stejné poloviny. Jedná se o hodnotu,

kteřá se nachází uprostřed souboru. V případě lichého počtu prvků odpovídá medián prostřední hodnotě. Je-li počet prvků sudý, mediánem je aritmetický průměr dvou prostředních hodnot. Matematický vztah je následující (Hendl, 2012, str. 100):

$$Me = x_{(n+1)/2} \quad (2)$$

Jestliže  $n$  sudé číslo, pak:

$$Me = 0,5(x_{n/2} + x_{n/2+1}) \quad (3)$$

Míry centrální tendence – aritmetický průměr, medián a modus – slouží k vyjádření toho, kolem jakých hodnot se data koncentrují, avšak žádná z nich není univerzálně nejlepší. Volba vhodného ukazatele musí vždy vycházet ze tří faktorů: úrovně měření dat, tvaru jejich rozdělení a cíle analýzy. Aritmetický průměr je smysluplný zejména u intervalové či poměrové škály, pokud je rozdělení přibližně symetrické a analytik plánuje využít parametrické statistické testy; v těchto situacích průměr reflektuje plnou numerickou informaci a dobře se pojí se standardními testovacími postupy. Naproti tomu medián je vhodnější pro ordinální data nebo pro případy se silně vychýleným rozdělením a s přítomností extrémních hodnot, protože jako robustní statistika odolává jejich vlivu a lépe vystihuje typické pozorování. Modus nachází uplatnění tam, kde rozdělení vykazuje více vrcholů, kdy potřebujeme jen rychlý přehled, případně když se pracuje s nominální škálou, pro niž průměr ani medián nelze definovat (Hendl, 2012, str. 99-101).

### 1.5.1.2 Míry rozptýlenosti

Při statistickém popisu dat nestačí uvést pouze jejich střední hodnotu, neboť ta nepostihuje míru variability, s jakou se jednotlivé hodnoty od této střední hodnoty odchyľují. Data se stejnou střední hodnotou mohou mít různou míru rozptýlenosti, což významně ovlivňuje interpretaci výsledků. K zachycení proměnlivosti dat slouží míry rozptýlenosti, přičemž výběr konkrétní míry závisí na typu rozdělení, účelu analýzy i odolnosti vůči extrémním hodnotám. Základní charakteristiky představují variační rozpětí rozptyl a směrodatná odchylka (Hendl, 2012, str. 101).

Rozptyl a z něj odvozená směrodatná odchylka měří, jak moc se jednotlivé hodnoty ve výběrovém souboru liší od průměru – tedy jak jsou rozptýlené kolem střední hodnoty.

Je důležité upozornit na omezení použití rozptylu a směrodatné odchylky, která vyplývají ze skutečnosti, že tyto míry jsou vhodné pouze tehdy, pokud je aritmetický průměr

relevantním ukazatelem střední hodnoty. Obě statistiky, aritmetický průměr i směrodatná odchylka, jsou totiž citlivé na odlehlé hodnoty, která mohou významně zkreslit výsledek. V případech, kdy má datové rozdělení výrazně zešikmený tvar nebo obsahuje extrémní hodnoty, nejsou rozptyl a směrodatná odchylka vhodnými ukazateli rozptýlenosti, protože neposkytují věrný obraz o variabilitě dat (Hendl, 2012, str. 102).

V takových případech jsou vhodnější míry rozptýlenosti založené na empirických kvantilech. Tyto míry jsou robustní vůči extrémním hodnotám a vycházejí z rozdělení dat podle pořadí. Empirický kvantil je taková hodnota, pod kterou leží definovaná část dat. Kvantily se označují symbolem  $x_q$ , přičemž  $q$  je parametr z intervalu  $0 < q < 1$  určující podíl hodnot, které leží pod tímto kvantilem. Hladiny kvantilů lze udávat i v procentech. Například:

- 25% kvantil se označuje jako první (dolní) kvartil  $Q_1$
- 50% kvantil je medián  $Q_2$
- 75% kvantil je třetí (horní) kvartil  $Q_3$  (Hendl, 2012, str. 103).

**Interkvartilové rozpětí (IQR)** Interkvartilové rozpětí představuje rozdíl mezi třetím a prvním kvantilem:

$$Q = Q_3 - Q_1 \quad (4)$$

IQR zachycuje středních 50 % datového rozdělení a je užitečné zejména tehdy, když chceme charakterizovat rozptýlenost bez vlivu extrémních hodnot. Na rozdíl od směrodatné odchylky není interkvartilové rozpětí ovlivněno odlehlými hodnotami, a proto se častěji používá při šikmých nebo nepravidelných rozděleních (Hendl, 2012, str. 103).

### 1.5.2 Testování statistických hypotéz

Testy statistické významnosti jsou základním nástrojem inferenční statistiky. Umožňují rozhodovat o hypotézách na základě výběrových dat. Jejich cílem je vyhodnotit, zda je pozorovaný rozdíl v datech pouze náhodný, nebo zda poukazuje na skutečný jev v populaci (Hendl, 2012, str. 171).

Statistické testování významnosti se zakládá na ověřování hypotéz. Ve výzkumu se často setkáváme se situací, kdy chceme na základě dat rozhodnout mezi dvěma možnostmi –

např. „účinek existuje“ vs. „účinek neexistuje“. K tomu slouží následující metodický postup (Hendl, 2012, str. 181)

Statistické testování probíhá ve čtyřech základních krocích (Hendl, 2012, str. 182):

1. Vyjádření výzkumné otázky ve formě statistických hypotéz ( $H_0$  a  $H_1$ )
2. Stanovení akceptovatelné úrovně chyb rozhodování
3. Vypočtení testovací statistiky
4. Vyslovení doporučení

**Určení statistických hypotéz** spočívá ve formulaci nulové hypotézy ( $H_0$ ) a alternativní hypotézy ( $H_1$ ). Nulová hypotéza ( $H_0$ ) předpokládá, že mezi zkoumanými skupinami neexistuje rozdíl nebo že pozorovaný jev je důsledkem náhody, resp. přirozené variability dat. Naproti tomu alternativní hypotéza ( $H_1$ ) představuje protiklad nulové hypotézy. Tvrdí, že mezi skupinami existuje statisticky významný rozdíl nebo že mezi sledovanými proměnnými existuje vztah, který nelze připsat náhodě (Hendl, 2012, str. 182).

**Určení hladiny významnosti  $\alpha$**  představuje pravděpodobnost chybného zamítnutí  $H_0$ , tedy že zamítneme nulovou hypotézu, přestože je ve skutečnosti pravdivá. Hladina  $\alpha$  se volí velmi malá – klasicky 0,05 nebo 0,01 (Hendl, 2012, str. 183).

**Výpočet testovací statistiky**, který slouží jako základ pro rozhodnutí o platnosti či neplatnosti nulové hypotézy. Tento výpočet vychází z empirických dat a předpokladů dané hypotézy. Testovací statistika vyjadřuje rozdíl mezi bodovým odhadem parametru a jeho hypotetickou hodnotou, vztahovaný ke směrodatné chybě odhadu (Hendl, 2012, str. 183).

V závěrečné fázi testování statistických hypotéz je cílem interpretovat výsledky a učinit rozhodnutí ohledně platnosti nulové hypotézy. K tomu slouží tzv. p-hodnota, která udává pravděpodobnost, že i při platnosti nulové hypotézy ( $H_0$ ) by došlo k výskytu stejně nebo ještě více extrémního výsledku, než jaký byl skutečně pozorován. Na základě porovnání p-hodnoty s předem stanovenou hladinou významnosti  $\alpha$  se rozhoduje následovně :

- Pokud  $p < \alpha$ , nulovou hypotézu zamítáme.
- Pokud  $p \geq \alpha$ , nulovou hypotézu nezamítáme (Hendl, 2012, str. 183-184).

Tímto způsobem p-hodnota kvantifikuje sílu důkazu proti nulové hypotéze. Čím menší p-hodnota, tím větší nesoulad mezi daty a předpokladem, že  $H_0$  platí.

Při interpretaci p-hodnoty je klíčové si uvědomit, že testování se nezaměřuje na potvrzení  $H_0$ , ale na hledání důkazů proti ní. Nízká p-hodnota značí, že za předpokladu platnosti  $H_0$  je pozorovaný výsledek velmi nepravděpodobný, a tudíž existuje důvod o této hypotéze pochybovat (Hendl, 2012, str. 184).

### 1.5.2.1 Testy hypotéz o rozdělení základního souboru

Znalost rozdělení pravděpodobnosti základního souboru hraje klíčovou roli při volbě vhodných statistických metod pro analýzu výběrových dat. Mnoho statistických postupů, zejména parametrických metod, předpokládá určité specifické rozdělení dat – nejčastěji normální rozdělení. Pokud jsou však tyto metody aplikovány na výběrový soubor, který dané předpoklady nesplňuje, může dojít k významnému zkreslení výsledků a nesprávné interpretaci. Proto je před aplikací těchto metod nezbytné ověřit, zda rozdělení dat odpovídá teoretickému předpokladu, například pomocí testů shody rozdělení (Neubauer et al., 2021, str. 224).

Rozdělení dat je možné ověřovat buď grafickými metodami (např. histogram, Q–Q diagram), nebo pomocí statistických testů, které poskytují formální rozhodnutí o shodě dat s předpokládaným teoretickým rozdělením (Shapiro-Wilkův test, Kolmogorovův-Smirnovův, C-test aj.) (Neubauer et al., 2021, str. 227, 233).

**Kolmogorovův-Smirnovův test** slouží ke srovnání empirického rozdělení výběrového souboru s teoretickým rozdělením, které musí být předem zcela specifikováno. Umožňuje testovat, zda náhodný výběr pochází z konkrétního spojitého rozdělení, typicky například z normálního rozdělení. Nulová hypotéza  $H_0$  testu stanovuje, že data pocházejí z rozdělení, které odpovídá předpokládané teoretické distribuční funkci. Testová statistika (s normálním rozdělením) je definována jako maximální absolutní rozdíl mezi těmito dvěma funkcemi (Neubauer et al., 2021, str. 233):

$$D = \sup_x |F_n(x) - F(x)| \quad (5)$$

Kde:

$F_n(x)$  - empirická distribuční funkce dat,

$F(x)$  – teoretická distribuční funkce normálního rozdělení

$\sup_x$  – je supremum (největší rozdíl)

### 1.5.3 Analýza závislostí

Statistická analýza se jen zřídka zaměřuje na jednu izolovanou proměnnou. Často je důležité znát srovnání více rozdělení, změnu proměnné v čase nebo vztahy mezi více proměnnými. Uvedené vztahy obvykle nejsou čistě deterministické. Proto se pro jejich popis používají statistické metody z oblasti korelační a regresní analýzy. Cílem je určit, zda mezi dvěma proměnnými existuje vztah, případně odhadnout, jak silná tato závislost je a zda může být využita. Vztah dvou proměnných lze zobrazit graficky (korelační diagram) (Hendl, 2012, str. 184).

#### 1.5.3.1 Korelační analýza

V nejširším smyslu označuje korelace míru závislosti, resp. asociace mezi dvěma proměnnými. Pokud je u určitých hodnot jedné proměnné pozorována tendence výskytu určitých hodnot proměnné druhé, proměnné jsou korelované. V případě silné korelace mají hodnoty proměnných tendenci se společně měnit – buď stejným směrem (kladná korelace), nebo opačně (záporná korelace) (Hendl, 2012, str. 250-251).

Je nutné zmínit, že korelace neimplikuje kauzalitu, tedy příčinný vztah sledovaných proměnných a pokud je mezi proměnnými zjištěna korelace, může mít několik možných interpretací (Janáček, 2022, str. 36-38):

- Proměnná X ovlivňuje proměnnou Y – existuje kauzální vztah ve směru X k Y
- Proměnná Y ovlivňuje proměnnou X – existuje kauzální vztah ve směru Y k X
- Obě proměnné se vzájemně ovlivňují – existuje obousměrná kauzalita
- Mezi X a Y není přímý vztah, ale obě proměnné jsou ovlivněny třetí (skrytou) proměnnou, či více proměnnými
- Korelace je náhodná – statisticky významnou, ale věcně bezvýznamnou souvislost

Pro výpočet korelačního koeficientu se využívají různé metody, např. Pearsonův korelační koeficient nebo Spearmanův korelační koeficient pořadí.

**Spearmanův korelační koeficient pořadí** ( $r_s$ ) je neparametrický koeficient korelace, který určuje míru síly a směru monotónního vztahu mezi dvěma proměnnými. Byl zaveden anglickým psychologem Charlesem Edwardem Spearmanem (1863–1945) jako alternativa k Pearsonovu korelačnímu koeficientu v případech, kdy nelze předpokládat lineární závislost ani normální rozdělení dat. Základním principem

Spearmanovy metody je nahrazení původních hodnot jejich pořadím v rámci výběrového souboru. Korelace se pak měří mezi těmito pořadími, nikoli mezi hodnotami samotnými. Tím je dosaženo vyšší odolnosti vůči odlehlým hodnotám a možností odhalení monotónních (ne nutně lineárních) vztahů mezi proměnnými. Spearmanův koeficient je užitečný zejména (Hendl, 2012, str. 268-269):

- Pokud je potřebný rezistentní odhad korelace, méně citlivý na odlehlá data
- Při testování schopnosti osob nebo systémů správně řadit objekty podle určitého hlediska
- Pro analýzu časových řad s možným monotónním trendem
- V případech, kdy data nejsou měřena na intervalové škále, ale jako pořadová

Při výpočtu koeficientu je nutné brát v potaz výskyt duplicitních měření obou sledovaných proměnných. V případě, že se ve výběrovém souboru vyskytuje velké množství duplicit, je nutné použít definiční vztah vycházející z Pearsonova korelačního koeficientu, nikoliv zjednodušený vztah (Holčík a Komenda, 2015):

$$r_s = \frac{\sum_{i=1}^n (x_{ri}y_{ri} - n\bar{x}_r\bar{y}_r)}{(n-1)s_{x_r}s_{y_r}} \quad (6)$$

Kde:

$x_{ri}$  – Pořadí  $i$ -té hodnoty proměnné  $X$

$y_{ri}$  – Pořadí  $i$ -té hodnoty proměnné  $Y$

$\bar{x}_r$  – Průměrné pořadí všech hodnot proměnné  $X$

$\bar{y}_r$  – Průměrné pořadí všech hodnot proměnné  $Y$

$s_{x_r}$  – Směrodatná odchylka pořadí proměnné  $X$

$s_{y_r}$  – Směrodatná odchylka pořadí proměnné  $Y$

$n$  – Počet dvojic pozorování

Hodnota  $r_s$  se pohybuje v intervalu  $[-1;1]$ , kde:

$r_s = 1$  = perfektní pozitivní monotónní vztah

$r_s = -1$  = perfektní negativní monotónní vztah

$r_s = 0$  = žádný monotónní vztah

### 1.5.4 Metriky pro hodnocení přesnosti predikčních modelů

Hodnocení přesnosti predikčních modelů je zásadním krokem při návrhu a optimalizaci predikčního systému. V praxi však neexistuje žádná univerzální metrika, která by byla vhodná ve všech situacích. Různé indikátory se liší svou citlivostí vůči chybám, směrovým zkreslením nebo výskytem odlehlých hodnot, a proto je volba vhodné metriky vždy závislá na konkrétním kontextu a cílech predikce (Vandepuut, 2019).

**Průměrná absolutní chyba (MAE)** je metrika udávající průměrnou absolutní velikost chyby. Každá odchylka mezi predikcí a skutečnou hodnotou je brána v absolutní hodnotě, čímž se eliminuje vliv směru chyby. MAE tedy představuje intuitivní a přímo interpretovatelnou míru přesnosti. Její nevýhodou je, že není škálována vůči velikosti hodnocené veličiny. Matematický vztah je následující (Vandepuut, 2019):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

Kde:

$n$  – počet datových bodů

$y_i$  – skutečná hodnota pro každý  $i$ -tý datový bod

$\hat{y}_i$  – predikovaná hodnota pro každý  $i$ -tý datový bod

**Odmocněná střední kvadratická chyba (RMSE)** se počítá jako odmocnina průměrné druhé mocniny chyb a je citlivější na velké chyby než MAE. Zatímco MAE dává všem chybám stejnou váhu, RMSE zvýrazňuje větší odchylky – čím je chyba větší, tím více se promítne do hodnoty RMSE. Tak poskytuje důraz na celkovou přesnost modelu a minimalizuje aritmetický průměr chyb, což znamená, že optimalizace podle RMSE vede ke snížení biasu. Na druhé straně vysoká citlivost na odlehlé hodnoty může vést k nežádoucímu zkreslení modelu. Její matematický vztah je následující (Vandepuut, 2019):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

Kde:

$n$  – počet datových bodů

$y_i$  – skutečná hodnota pro každý  $i$ -tý datový bod

$\hat{y}_i$  – predikovaná hodnota pro každý  $i$ -tý datový bod

**Bias** udává, zda má predikční model tendenci systematicky přeceňovat nebo podceňovat skutečné hodnoty. Je definován jako průměrná chyba v průběhu všech sledovaných období. Pozitivní bias znamená, že model obecně nadhodnocuje poptávku, zatímco negativní bias indikuje podhodnocení. Důležité je poznamenat, že nízký bias neznamená automaticky vysokou přesnost – kladné a záporné chyby se totiž mohou navzájem vyrušit, což zakrývá skutečný rozsah chyb (Vandeput, 2019):

$$Bias = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (9)$$

Kde:

$n$  – počet datových bodů

$y_i$  – skutečná hodnota pro každý  $i$ -tý datový bod

$\hat{y}_i$  – predikovaná hodnota pro každý  $i$ -tý datový bod

## **2 ANALÝZA SOUČASNÉHO STAVU**

Tato kapitola analyzuje procesy predikce, řízení spotřeby elektřiny a odchylky predikcí blíže nespecifikované výrobní společnosti XYZ, a.s. (dále jen společnost). Společnost provozuje desítky výrobních závodů. Výrobní program společnosti je energeticky náročný a podíl spotřeby elektrické energie se výrazně podílí na cenové složce finálního produktu.

### **2.1 Softwarová řešení společnosti reporting a administrativní procesy**

Hlavní páteří funkčnosti celého podniku v oblasti podnikových procesů představuje ERP systém SAP EEC, který je využíván pro finance, nákup, prodej, logistiku, řízení zásob apod.

Další klíčovou technologickou platformou využívanou ve společnosti je ekosystém společnosti Microsoft, který zahrnuje jak klasické kancelářské aplikace, tak pokročilé cloudové a vývojové nástroje. Základním prvkem je Microsoft 365 (dříve Office 365), který poskytuje komplexní sadu aplikací pro kancelářskou práci a týmovou spolupráci. Mezi nejpoužívanější patří Word, Excel, PowerPoint, Outlook, Microsoft Teams, SharePoint atd.

Kromě kancelářských a kolaborativních nástrojů společnost využívá také řešení Microsoft Power Platform, především Power Apps a Power Pages pro tvorbu vlastních aplikací a webových stránek.

Společnost disponuje vlastní serverovou infrastrukturou, která umožňuje centralizovanou správu dat a přístup ke sdíleným souborům pro zaměstnance. V rámci této infrastruktury jsou využívány síťové disky, hostované na firemním souborovém serveru, který se nachází v serverovně společnosti. Dále společnost provozuje vlastní SQL Server, který slouží jako centrální databázový systém pro různá podniková aplikační řešení.

#### **2.1.1 Reportingová a analytická řešení**

Pro potřeby reportingu se na celopodnikové úrovni (především pro potřeby vyššího a vrcholného managementu) využívá reportní BI řešení od společnosti SAP, které využívá data z ERP systému a jeho databáze.

Pro účely reportingu a analýzy na úrovni jednotlivých uživatelů je v organizaci využíván Microsoft Excel, který je nejčastěji u periodických reportů doplněn o automatizaci

pomocí programovacího jazyka Visual Basic for Applications (VBA). Automatizace je implementována zejména u časově náročných procesů, jako jsou ETL operace (*Extract, Transform, Load*), integrace s externími systémy a soubory či opakovaná zpracování velkých objemů dat.

Přestože VBA umožňuje částečnou eliminaci manuální práce, proces zpracování reportů stále vyžaduje zásahy uživatelů, což s sebou nese nevýhody. U sdílených řešení, kde s reporty pracuje více uživatelů současně, může docházet ke kolizím a neúmyslným změnám obsahu, což může negativně ovlivnit kvalitu dat. Dalším rizikem je možnost vzniku chyb v důsledku lidského faktoru, zejména při manuálním vyplňování dat, což může vést ke zkreslení výsledných reportů a analýz.

Společnost má k dispozici i pokročilé nástroje Microsoft Platform – Power BI a Power Automate, ale jejich využívání je velmi sporadické (v případě Power Automate) až takřka neexistující (Power BI).

## **2.2 Predikce spotřeby elektřiny**

Společnost má nastavenou strategii nákupu elektřiny, která kombinuje dlouhodobé obchodní nástroje (s platností delší než 1 měsíc) prostřednictvím dlouhodobých dodavatelských smluv a forwardových kontraktů a krátkodobé obchodní nástroje prostřednictvím krátkodobého trhu s elektřinou (denní a vnitrodenní trh). Strategie je nastavená tak, aby dlouhodobé obchody představovaly 60 % spotřeby a krátkodobé 40 % spotřeby.

Spotřeba je v průběhu času proměnlivá. A aby docházelo k efektivnímu využití zdrojů je variabilita spotřeby vyrovnávána na krátkodobých trzích, které poskytují značnou míru časové flexibility pro uzavírání obchodů.

Pro obchodování elektřiny na krátkodobých trzích je používán predikční systém, který je softwarovým řešením vyvinutým na míru třetí stranou. Tento predikční systém je založený na expertních odhadech spotřeby pověřenými pracovníky. Predikce jsou zpravidla mírně nadhodnocovány (cca 2-3 %), to především z důvodu preference kladné odchylky, které jsou při měsíčním zúčtování OTE spíše protiodchylkou a tedy nejsou

penalizovány, naopak jsou finančně odměněny<sup>3</sup>. Tento jev nicméně není pravidlem a při celkovém zúčtování mohou být kladné odchylky stále penalizovány.

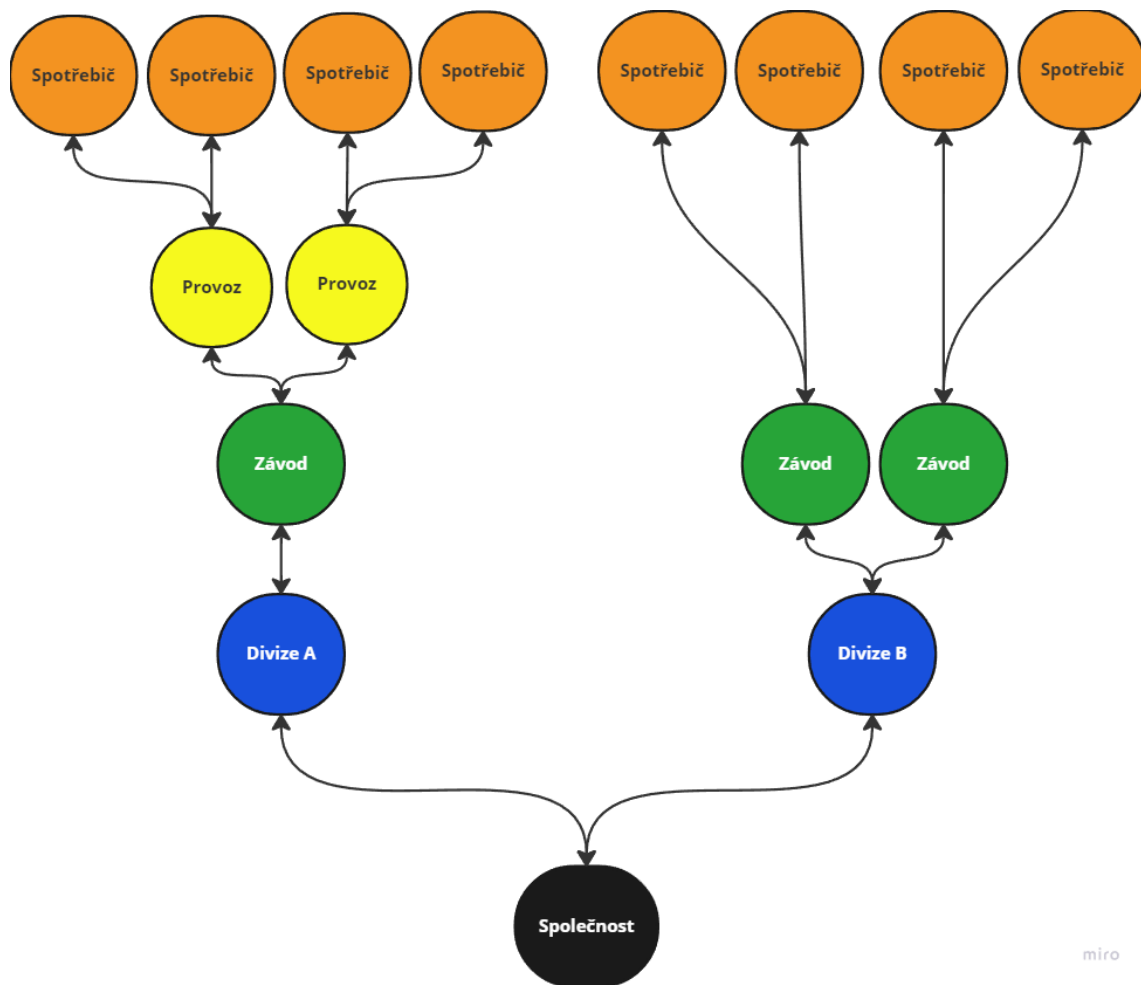
Hlavním nástrojem pro predikci je realizační diagram (ERD), který je v systému přiřazený ke každému aktivnímu logovanému spotřebiči. ERD je založený na způsobu obchodování elektřiny na krátkodobém trhu, který obsahuje obchodní intervaly (15 minut) v jednom dni (96 intervalů).

Každý spotřebič má určenou normovanou spotřebu (v MWh) na jednotku času. Za určování norem spotřeby jsou zodpovědní energetici. Pověření techničtí pracovníci jednotlivých závodů a provozů do ERD diagramů vyplňují plán denní spotřeby. Pro každý spotřebič jsou určeny časové intervaly, během nichž bude spotřebič v provozu, nebo naopak mimo provoz. U spotřebičů jsou nastaveny výchozí plány spotřeby (mohou být i nulové). V případě, že technický pracovník neprovede zásah v predikčním systému, je použit výchozí plán spotřeby.

Rozhraní aplikace umožňuje zobrazení agregací predikované spotřeby dle organizačního členění, kde nejnižší úroveň agregace představuje „spotřebič“ a nejvyšší „společnost“ (souhrn všech predikcí v daném čase). Členění je rozdílné pro jednotlivé výrobní divize (viz obrázek 10), kde „divize A“ má mezičlánek „provoz“. Finálním výstupem je agregace predikcí za celou společnost v jednotlivých obchodních intervalech.

---

<sup>3</sup> Tato skutečnost není potvrzena pokročilou metodou, ale jde spíš o expertní pohled energetiků a nákupčích.



**Obrázek 10: Agregace predikcí**

(Zdroj: Vlastní zpracování v Miro)

Od celkové denní spotřeby rozdělené do časových intervalů je odečten objem elektřiny nakoupený přes dlouhodobé kontrakty (objem je konstantní a v čase se mění jen skokově). Výsledkem je objem elektřiny, který se buď nakoupí nebo prodá na denním trhu.

Proces nákupu na denním trhu je automatizovaný. V aplikaci je nastaven čas, přibližně 45 minut před uzávěrkou obchodů na denním trhu ve 12:00 D-1, kdy systém automaticky odešle příkaz k obchodu ve formátu XML. Tento příkaz je odeslán prostřednictvím komunikačního scénáře v rozhraní automatické komunikace OTE.

Po úspěšném či neúspěšném zadání příkazu k obchodu systém zašle odpovědným osobám (administrátorům, nákupčím a energetikům) konfirmační email. Tento email obsahuje výsledek příkazu a ERD diagram obchodů s příslušnými hodnotami pro příslušný

obchodní den. Proces predikce a nákupu elektřiny je doplněn procesním diagramem (viz Příloha 1).

### 2.2.1 Reporting predikce elektřiny

Predikční systém obsahuje i rozhraní reportingu, jehož úkolem je kontrola predikcí elektřiny s porovnáním s reálnou spotřebou. V reportingu jsou dostupné dva nástroje:

- *Srovnávací tabulka*
- *Sloupcový graf*

*Srovnávací tabulka* umožňuje volbu agregace dat podle času pomocí interaktivního kalendáře, kde je možné zvolit interval sledovaných dní (nejvyšší úroveň granularity je jeden den) a úrovně měření/predikce pouze na úroveň provozu nebo závodu a postrádá tak hlubší detail (viz obrázek 10) na úrovni spotřebičů. Výsledkem jsou hodnoty predikcí (první sloupec) a reálné spotřeby (druhý sloupec) a jejich rozdíl (třetí sloupec). Rozhraní nenabízí možnost exportu dat.

*Sloupcový graf* umožňuje volbu agregace pouze podle úrovně měření/predikce (opět pouze na úrovni provoz/závod) a zobrazení času je pevně dané na jeden den, který je možné zvolit pomocí interaktivního kalendáře. Jednotlivé sloupce představují obchodní interval (1 hodina), přičemž obchodní intervaly byly zkráceny na 15 minut<sup>4</sup> a vizualizace tuto skutečnost nereflektuje. V grafu jsou sloupce pro predikované množství elektřiny a reálnou spotřebu.

Tento reporting není významně pracovníky používán, protože jeho výstupy nejsou často dostatečné pro účely analýz a reportingu. Navíc u některých spotřebičů jsou zobrazovány nulové nebo jiné hodnoty, než byly predikovány, což naznačuje nevyřešený problém při zpracování dat programem.

Pro účely reportingu (na měsíční a roční bázi) je využíván především systém OTE, který umožňuje export vlastních dat<sup>5</sup> společnosti o aktivitách na krátkodobých trzích (na úrovni celé společnosti) a měřené spotřebě elektřiny v obchodních intervalech (na úrovni závodů). Data se zpracovávají v prostředí MS Excel (kombinace manuálního zpracování

---

<sup>4</sup> Od 1. 7. 2024

<sup>5</sup> Možnosti formátu jsou: .xlsx, .txt, .xlm

a skriptu ve VBA). Tento reporting slouží primárně pro potřeby finančního účetnictví. Toto řešení má své limity, především skutečnost, že proces zpracování vyžaduje velké množství manuálních zásahů pracovníků (což zvyšuje možnost vzniku chyby způsobené lidským faktorem).

Dalším omezením je agregace na úrovni společnosti a závodů, která neumožňuje hlubší porozumění. Například v případech vzniku velké odchylky nejsou pracovníci schopni jasně a rychle identifikovat příčinu vzniku. Tyto situace mohou být způsobeny i nevhodným predikcím technických pracovníků a prakticky tak neexistuje systém agilní kontroly predikční činnosti.

### 2.3 Datová struktura a logistika aplikace

Databázový systém aplikace je implementován na platformě Microsoft SQL Server, což je relační databázový systém spravovaný prostřednictvím jazyka SQL (Structured Query Language).

V databázi aplikace jsou data měření a predikcí organizována v několika klíčových tabulkách, které slouží k ukládání predikovaných a měřených hodnot a jejich vzájemnému propojení.

Tabulka predikcí ukládá data na úrovni jednotlivých spotřebičů, zatímco tabulky měření zaznamenávají naměřené hodnoty na úrovni elektroměrů (názvy tabulek jsou změněné):

- *PredictionTable* – tabulka s historickými daty predikcí spotřeby
- *ConsumptionTable1* – tabulka s historickými daty měření spotřeby (první)
- *ConsumptionTable2* – tabulka s historickými daty měření spotřeby (druhá)
- *ConsumptionTable3* – tabulka s historickými daty měření spotřeby (třetí)

Tabulky měření jsou tři, a to z důvodu, že ve společnosti se využívají tři energetické řídicí systémy. Do tabulek měření se ukládají průběhová data měření všech elektroměrů. Tabulky měření a predikce jsou propojené s mapovací tabulkou *ApplianceMeasurementMap*, která obsahuje ty identifikátory elektroměrů, které jsou přiřazené ke konkrétnímu spotřebiči<sup>6</sup>. Přiřazení elektroměrů lze provést přímo

---

<sup>6</sup> Ne všechny elektroměry jsou přiřazeny ke konkrétnímu spotřebiči, i když s ním mají souvislost. To především z toho důvodu, že některé elektroměry jsou podružné jiným.

v nastavení aplikace – název spotřebiče (identifikátor) se vloží do kolonky *číslo měření*. Toto nastavení spravují energetici. Každý záznam má svůj unikátní číselný identifikátor (primární klíč), identifikátor spotřebiče (cizí klíč) a identifikátor elektroměru (cizí klíč). V tomto relačním modelu platí logika, kdy spotřebič má alespoň jeden přiřazený elektroměr a elektroměr má přiřazený právě jeden spotřebič. Tato struktura umožňuje databázovému systému přesně přiřazovat predikované hodnoty k reálným měřením, což je nezbytné pro následnou analýzu odchylek a reportování. Relační model nejdůležitějších součástí databáze pro potřeby reportingu je zobrazen v Příloze 7. Pro pokročilé analytické potřeby není klasický relační databázový systém vhodný a ideálním řešením by bylo vytvoření analytického dimenzionálního datového úložiště.

## **2.4 Zhodnocení obchodní činnosti na krátkodobých trzích ve vztahu k odchylkám**

V následující části bude popsána a zhodnocena obchodní činnost společnosti na krátkodobých trzích s elektřinou ve vztahu k odchylkám predikce spotřeby v letech 2021 až 2023. Analýza zahrnuje nejen objemové, ale také cenové aspekty obchodů a odchylek, aby bylo možné identifikovat hlavní příčiny a důsledky těchto rozdílů.

### **2.4.1 Základní přehled zkoumaného datového souboru**

Zkoumaný datový soubor (tabulka 1) obsahuje 13 souborů (časových řad) podstatných pro potřeby této tohoto výzkumu. Pro společnost, zejména na vnitrodenním trhu, je běžné, že neobchoduje nepřetržitě, proto záznamy z časových intervalů, kdy k obchodům nedocházelo (mají hodnotu nula nebo jsou prázdná), nebyly zahrnuty do popisné statistiky. Tento přístup byl zvolen, aby se předešlo zkreslení dat časovými periodami, kdy obchod neprobíhal. Jiná situace nastává u odchylek, které vznikají nezávisle na aktivitě na krátkodobých trzích, ale v danou sledovanou časovou periodu mohou být právě záporné, nebo kladné. Při analýze dat se ukázalo, že žádný záznam o celkových odchylkách neměl prázdnou hodnotu, a i v případech nulové odchylky byl zaznamenán obchod.

Počet prvků zkoumaných souborů (především krátkodobé trhy) se liší podle četnosti uskutečnění obchodu. Z vyjádření počtu uskutečněných obchodů je jasná převaha denního nad vnitrodenním trhem, kdy na denním trhu dochází k obchodu téměř každou

hodinu, zatímco na vnitrodenním každou 9. hodinu. U odchylek je patrná převaha kladných nad zápornými.

**Tabulka 1: Přehled datových souborů**

Soubor	Celkem časových period	Počet obchodů	Průměrný interval mezi obchody
Denní trh – saldo (MWh)	26277	25685	1 hodina a 1 minuta
Denní trh – saldo (CZK/MWh)	26277	25685	1 hodina a 1 minuta
Vnitrodenní trh – saldo (MWh)	26277	2931	8 hodin a 58 minut
Vnitrodenní trh – Saldo (CZK/MWh)	26277	2931	8 hodin a 58 minut
Obchod – saldo (MWh)	26277	26152	1 hodina
Obchod – saldo (CZK/MWh)	26277	26152	1 hodina
Odchylka kladná (MWh)	26277	15808	1 hodina a 40 minut
Odchylka záporná (MWh)	26277	10469	2 hodiny a 31 minut
Odchylka celkem (MWh)	26277	26277	1 hodina
Odchylka kladná (CZK/MWh)	26277	15808	1 hodina a 40 minut
Odchylka záporná (CZK/MWh)	26277	10469	2 hodiny a 31 minut
Odchylka celkem (CZK/MWh)	26277	26277	1 hodina

(Zdroj: Vlastní zpracování)

## 2.4.2 Deskriptivní statistika

Prvním krokem v analýze dat byla deskriptivní (popisná) statistika. Před samotnými výpočty statistik byla data podrobena parametrickému testu normálního rozdělení, konkrétně Kolmogorovu-Smirnovu testu normality.

### 2.4.2.1 Formulace hypotéz normality dat

Pro každý sledovaný soubor bude vyslovena hypotéza  $H_0$  (nulová hypotéza) a  $H_1$  (alternativní hypotéza):

- $H_0$ : Data v příslušném souboru mají normální rozdělení ( $p > 0,05$ )
- $H_1$ : Data v příslušném souboru nemají normální rozdělení ( $p < 0,05$ )

Výsledky Kolmogorovova-Smirnovova testu (tabulka 2) jednoznačně potvrzují, že žádný ze zkoumaných datových souborů nesplňuje podmínky normálního rozdělení, neboť u všech analyzovaných souborů byla p-hodnota nižší než 0,05. Tato skutečnost má zásadní dopad na výběr vhodných statistických metod a na interpretaci výsledků. Zejména aritmetický průměr může být u nenormálních rozdělení zavádějící, protože není robustní

vůči extrémním hodnotám. Absence normálního rozdělení je viditelná i při grafickém zobrazení rozdělení pomocí histogramů viz příloha 2 a 3.

Proto byl kromě aritmetického průměru vypočítán i robustní ukazatel centrální tendence (medián) a robustní ukazatel variability (interkvartilové rozpětí – IQR). Tyto ukazatele jsou vhodnější pro popis nenormálně rozdělených dat, protože lépe odolávají vlivu asymetrie a extrémních hodnot.

**Tabulka 2: Výsledky Kolmogorova-Smirnova testu normality a hypotéz**

Soubor	Stat ( <i>D</i> )	p-hodnota ( <i>p</i> )	H0	H1
Denní trh - saldo (MWh)	0,8628	0,000	hypotéza zamítnuta	hypotéza přijata
Denní trh - saldo (CZK/MWh)	0,9186	0,000	hypotéza zamítnuta	hypotéza přijata
Vnitrodenní trh - Saldo (MWh)	0,4179	0,000	hypotéza zamítnuta	hypotéza přijata
Vnitrodenní trh - Saldo (CZK/MWh)	0,5003	0,000	hypotéza zamítnuta	hypotéza přijata
Obchod - saldo (MWh)	0,8639	0,000	hypotéza zamítnuta	hypotéza přijata
Obchod - saldo (CZK)	0,9224	0,000	hypotéza zamítnuta	hypotéza přijata
Odchylka kladná (MWh)	0,5004	0,000	hypotéza zamítnuta	hypotéza přijata
Odchylka záporná (MWh)	0,5006	0,000	hypotéza zamítnuta	hypotéza přijata
Odchylka celkem (MWh)	0,2091	0,000	hypotéza zamítnuta	hypotéza přijata
Odchylka kladná (CZK/MWh)	0,5874	0,000	hypotéza zamítnuta	hypotéza přijata
Odchylka záporná (CZK/MWh)	0,7502	0,000	hypotéza zamítnuta	hypotéza přijata
Odchylka celkem (CZK/MWh)	0,5535	0,000	hypotéza zamítnuta	hypotéza přijata

(Zdroj: Vlastní zpracování)

#### 2.4.2.2 Obchody na krátkodobých trzích

Rozlišení obchodů má následující logiku – záporné hodnoty znamenají nákup, kladné prodej (nákup představuje náklad, prodej výnos).

Saldo objemu na denním trhu (tabulka 3) jednoznačně ukazuje na převahu nákupů, zatímco vnitrodenní trh je více vyrovnaný a vykazuje vyšší volatilitu. Výraznou odchylku a největší nestabilitu lze pozorovat v roce 2022, kdy byla variabilita salda na obou trzích nejvyšší.

**Tabulka 3: Popisná statistika souborů Denní trh – saldo (MWh) a Vnitrodenní trh – Saldo (MWh)**

	Rok	Průměr	Medián	Interkvartilové rozpětí (IQR)
Denní trh – saldo (MWh)	2021	-9,159	-8,900	7,900
	2022	-13,343	-13,600	12,400
	2023	-9,124	-10,100	8,800
	<b>Celkem</b>	<b>-10,501</b>	<b>-10,500</b>	<b>9,600</b>
Vnitrodenní trh – Saldo (MWh)	2021	1,901	3,000	8,400
	2022	-4,380	-1,800	12,225
	2023	-0,532	0,200	6,600
	<b>Celkem</b>	<b>-1,865</b>	<b>0,300</b>	<b>8,000</b>

(Zdroj: Vlastní zpracování)

Saldo ceny na obou trzích (tabulka 4) vykazuje záporné hodnoty, což značí převahu nákupů nad prodeji. Nejvýraznější odchylkou je rok 2022, který byl charakteristický vysokou variabilitou a extrémním chováním na obou trzích. Naopak roky 2021 a 2023 vykazují stabilnější chování trhu. Volatilitu lze pozorovat při porovnání cen a objemu na vnitrodenním trhu. Dle hodnot mediánu je viditelné, že ceny jsou levnější na vnitrodenním trhu, který je rizikovější, a to především podle interkvartilového rozpětí, které často více než dvojnásobně oproti dennímu trhu.

**Tabulka 4: Popisná statistika souborů Denní trh – saldo (CZK/MWh) a Vnitrodenní trh – Saldo (CZK/MWh)**

	Rok	Průměr	Medián	Interkvartilové rozpětí (IQR)
Denní trh – saldo (CZK/MWh)	2021	-2026,13	-1869,24	1312,32
	2022	-5173,38	-5058,45	3997,56
	2023	-1863,16	-2313,73	1215,12
	<b>Celkem</b>	<b>-2990,30</b>	<b>-2510,23</b>	<b>2662,07</b>
Vnitrodenní trh – Saldo (CZK/MWh)	2021	800,94	1696,48	4896,77
	2022	-5049,87	-3265,12	8135,84
	2023	-1291,50	-61,80	2553,42
	<b>Celkem</b>	<b>-2650,30</b>	<b>-42,75</b>	<b>3707,74</b>

(Zdroj: Vlastní zpracování)

Po sečtení celkového salda obchodů (nákupů a prodejí na denním a vnitrodenním trhu – tabulka 5) je viditelná podoba se saldem na denním trhu, což lze vysvětlit nižší četností obchodů na vnitrodenním než na denním trhu.

**Tabulka 5: Popisná statistika souborů Obchod – saldo (MWh) a Obchod – saldo (CZK/MWh)**

	Rok	Průměr	Medián	Interkvartilové rozpětí (IQR)
<b>Obchod - saldo (MWh)</b>	<b>2021</b>	-9,05	-8,80	7,90
	<b>2022</b>	-13,49	-13,90	12,20
	<b>2023</b>	-9,09	-10,10	9,00
	<b>Celkem</b>	<b>-10,54</b>	<b>-10,50</b>	<b>9,70</b>
<b>Obchod - saldo (CZK/MWh)</b>	<b>2021</b>	-2019,50	-1864,29	1306,59
	<b>2022</b>	-5889,14	-5264,61	4439,07
	<b>2023</b>	-2112,37	-2342,38	1275,68
	<b>Celkem</b>	<b>-3332,93</b>	<b>-2570,60</b>	<b>2899,72</b>

(Zdroj: Vlastní zpracování)

### 2.4.2.3 Odchytky

Kladná odchytky představuje množství elektřiny, které bylo nakoupeno navíc oproti skutečné spotřebě. V průběhu sledovaných let (tabulka 6) lze pozorovat postupný pokles mediánu, průměru a interkvartilového rozpětí (IQR), což naznačuje klesající tendenci variability těchto odchylek, a tedy i přesnější predikci. Záporná odchytky představuje situace, kdy se nakoupilo méně elektřiny, než bylo potřeba. Tyto hodnoty jsou menší (bližší nule) ve srovnání s kladnými odchylkami, což může indikovat, že případy podhodnocení jsou lépe kontrolované, nebo preferenci pro nadhodnocení. Celková odchytky představuje kombinaci kladných a záporných odchylek. Z kladných hodnot mediánů a průměrů vyplývá, že kladná odchytky převažuje nad zápornou.

**Tabulka 6: Popisná statistika Odchylka kladná (MWh), Odchylka záporná (MWh) a Odchylka celkem (MWh)**

	Rok	Průměr	Medián	Interkvartilové rozpětí (IQR)
Odchylka kladná (MWh)	2021	2,29	1,64	2,61
	2022	1,96	1,30	2,16
	2023	1,68	1,14	1,72
	<b>Celkem</b>	<b>1,97</b>	<b>1,32</b>	<b>2,13</b>
Odchylka záporná (MWh)	2021	-1,34	-0,89	1,18
	2022	-1,72	-0,84	1,21
	2023	-1,14	-0,72	1,17
	<b>Celkem</b>	<b>-1,41</b>	<b>-0,83</b>	<b>1,19</b>
Odchylka celkem (MWh)	2021	0,75	0,33	2,68
	2022	0,45	0,34	2,26
	2023	0,66	0,44	1,92
	<b>Celkem</b>	<b>0,62</b>	<b>0,37</b>	<b>2,27</b>

(Zdroj: Vlastní zpracování)

Pro interpretaci cen odchylek (tabulka 7) je nutné zmínit, že určení ceny nezávisí pouze na jejich velikosti nebo směru, ale především na ceně regulační energie, kterou se vyrovnává systémová odchylka a na tom, jestli je v dané periodě odchylka protiodchylkou či nikoliv. U cen odchylek platí stejná logika jako u obchodů na krátkodobých trzích – kladná cena je výnos (za odchylku, resp. protiodchylku vzniká finanční odměna) a záporná cena je náklad (za odchylku vzniká finanční penalizace). Z hodnot mediánu a IQR vyplývá, že kladné odchylky představovaly spíše kladné protiodchylky, za které vznikaly finanční odměny a záporné odchylky byly spíše penalizovány. Penalizace u záporných odchylek je v porovnání vyšší než odměna u kladných odchylek, což potvrzuje vyšší penalizaci než odměňování pro motivaci subjektů nevytvářet odchylky. Z pohledu celkových odchylek je patrné vyrovnání blízko nule (především rok 2021), u které převažuje penalizace (díky vyššímu IQR záporných odchylek).

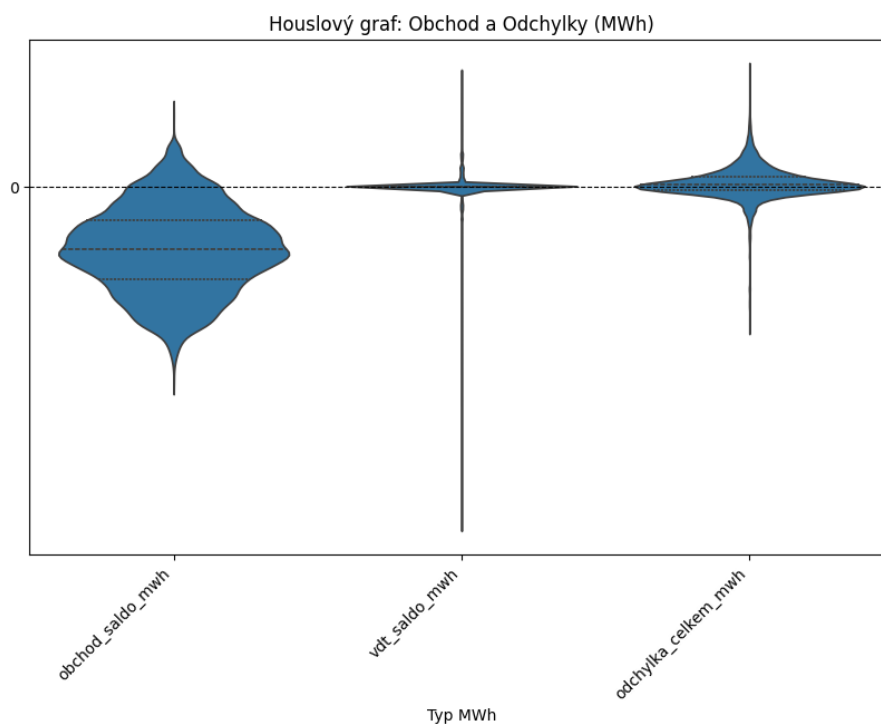
**Tabulka 7: Popisná statistika Odchylka kladná (CZK/MWh), Odchylka záporná (CZK/MWh) a Odchylka celkem (CZK/MWh)**

	Rok	Průměr	Medián	Interkvartilové rozpětí (IQR)
<b>Odchylka kladná (CZK/MWh)</b>	<b>2021</b>	1571,07	2350,00	2427,72
	<b>2022</b>	3805,54	1733,26	8334,09
	<b>2023</b>	846,79	0,00	3546,08
	<b>Celkem</b>	<b>2043,52</b>	<b>1934,81</b>	<b>3264,78</b>
<b>Odchylka záporná (CZK/MWh)</b>	<b>2021</b>	-2484,09	-2484,05	2795,49
	<b>2022</b>	-6347,44	-4057,00	11916,86
	<b>2023</b>	-3524,53	-2613,56	3172,59
	<b>Celkem</b>	<b>-4128,35</b>	<b>-2574,85</b>	<b>4674,49</b>
<b>Odchylka celkem (CZK/MWh)</b>	<b>2021</b>	-141,76	-1,81	4759,46
	<b>2022</b>	-369,51	-156,69	7367,52
	<b>2023</b>	-732,26	-147,20	4786,09
	<b>Celkem</b>	<b>-414,54</b>	<b>-44,41</b>	<b>4858,92</b>

(Zdroj: Vlastní zpracování)

Před provedením dalších analýz je nutné více pochopit zkoumaná data (salda obchodů a odchylky). Pro tyto účely jsou zkoumané soubory vizualizovány houslovými grafy (graf 1 a 2)

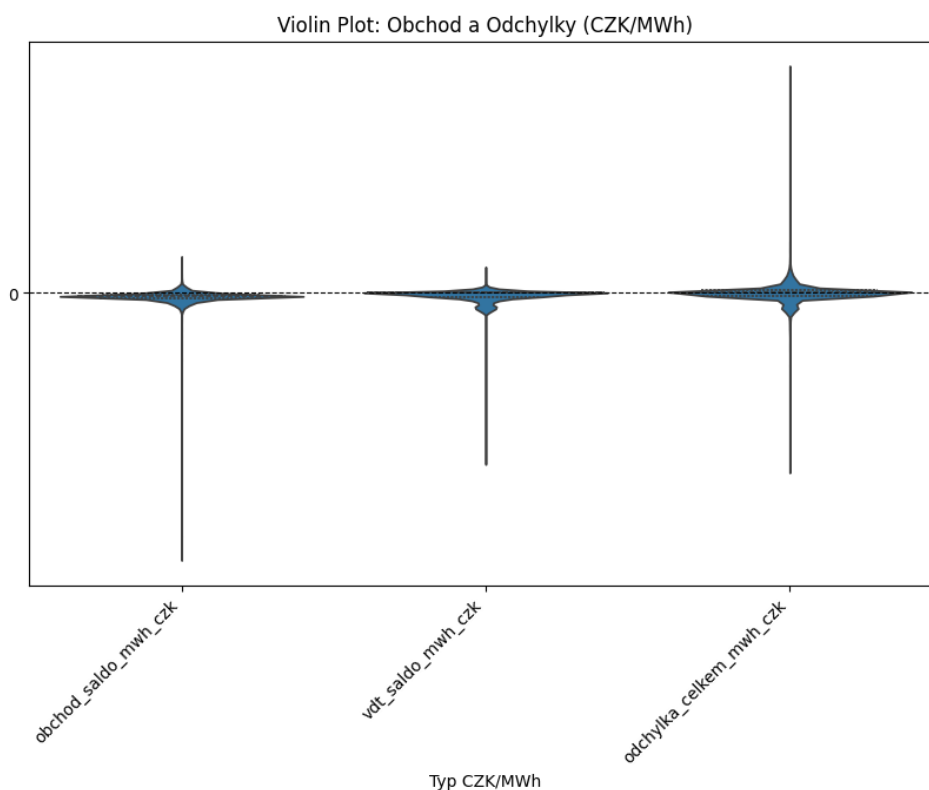
Houslový graf pro položky v MWh (graf 1) potvrzuje převahu nákupu elektřiny. V případě obchodů na vnitrodenním trhu, které slouží pro vyrovnávání spotřeb v krátkých časových intervalech před reálnou spotřebou (u zkoumaného datové souboru je nejzazší možným časem 1 hodina), je vidět spíše vyrovnaný podíl nákupů a prodejů, což může potvrzovat vyrovnávací funkci. Viditelné jsou odlehlé extrémní hodnoty, jejichž hustota je velmi malá. U odchylek je patrný vyšší podíl kladných odchylek, kdy záporné odchylky mají vyšší hodnotu odlehlých hodnot.



**Graf 1: Houslový graf – Obchod a odchylky (MWh)**

(Zdroj: Vlastní zpracování)

U cen za MWh (graf 2) je viditelná koncentrace většiny dat kolem hodnoty nula, ale viditelný je výskyt vyšších cen (náklady) za odchylky a nákupy na vnitrodenním trhu. Viditelný je i výskyt výrazných odlehých u odchylek v obou směrech a nákupních cenách.



**Graf 2: Houslový graf – Obchod a odchylky (CZK/MWh)**

(Zdroj: Vlastní zpracování)

### 2.4.3 Hodnocení přesnosti predikce

Nyní bude kvantifikována přesnost predikcí ve zkoumaném datovém setu. Přesnost predikce odchylek byla hodnocena pomocí vybraných statistických metrik, které umožňují kvantifikovat rozdíly mezi predikovanými hodnotami a skutečnými daty. Pro hodnocení přesnosti predikcí byly použity následující metriky: střední absolutní chyba (*MAE*), odmocněná střední kvadratická chyby (*RMSE*) a systematická chyba (*Bias*).

Výsledná hodnota střední absolutní chyby (*MAE*) se snižuje (tabulka 9) ukazuje, že průměrná odchylka predikcí se ve zkoumaných rocích snižuje, což může naznačovat zlepšování predikcí. Toto potvrzuje i směrodatná odchylka chyb (*RMSE*), která se v průběhu let (s výjimkou roku 2022) snižuje. Hodnota systematické chyby (*Bias*). Jasně naznačuje nadhodnocování, tedy preferenci kladné odchylky (výsledky jsou uvedeny v tabulce 8).

Odchylka v rozmezí 1,5 – 2 MWh představuje cca 16 % (10,54 MWh je medián obchodu v MWh), což je poměrně vysoké číslo a značí, že predikce spotřeby elektřiny nejsou příliš přesné.

**Tabulka 8: Metriky pro hodnocení přesnosti predikcí (MAE, RMSE a Bias)**

	MAE (MWh)	RMSE (MWh)	Bias
2021	1,89	2,71	0,75
2022	1,86	3,13	0,45
2023	1,49	2,20	0,66
<b>Celkem</b>	1,75	2,71	0,62

(Zdroj: Vlastní zpracování)

#### 2.4.4 Korelace ceny odchylky a velikosti odchylky

Při zkoumání výsledků ukazatelů popisné statistiky byla zjištěna kladná hodnota mediánu ceny (odměny) u kladných odchylek a záporná hodnota mediánu ceny (penalizace) u záporných odchylek. Dále byl pozorován jev systematického nadhodnocování predikcí, který potvrzuje predikční strategii preference nadhodnocování predikcí. Tento jev je tedy zkoumán dál pomocí výpočtu Spearmanova korelačního koeficientu ( $r_s$ ) (neparametrický koeficient) a vizualizací dat v korelačním diagramu.

Spearmanova korelace je neparametrický statistický nástroj, který se používá k určení síly a směru monotónního vztahu mezi dvěma proměnnými. Tato metoda je vhodná zejména pro data, která nemají normální rozdělení, protože se nezakládají na hodnotách, ale na pořadí dat.

Vypočet možnosti existence korelace a umožní vyjádření existence závislosti cen odchylek a jejich směru. Určení statistických hypotéz je následující:

*Test hypotéz o existenci závislosti:*

- $H_0$ : Mezi velikostí odchylek (MWh) a jejich cenou (CZK/MWh) neexistuje monotónní vztah, ( $r_s = 0$ ).
- $H_1$ : Mezi velikostí (MWh) a jejich cenou (CZK/MWh) existuje monotónní, ( $r_s \neq 0$ ).

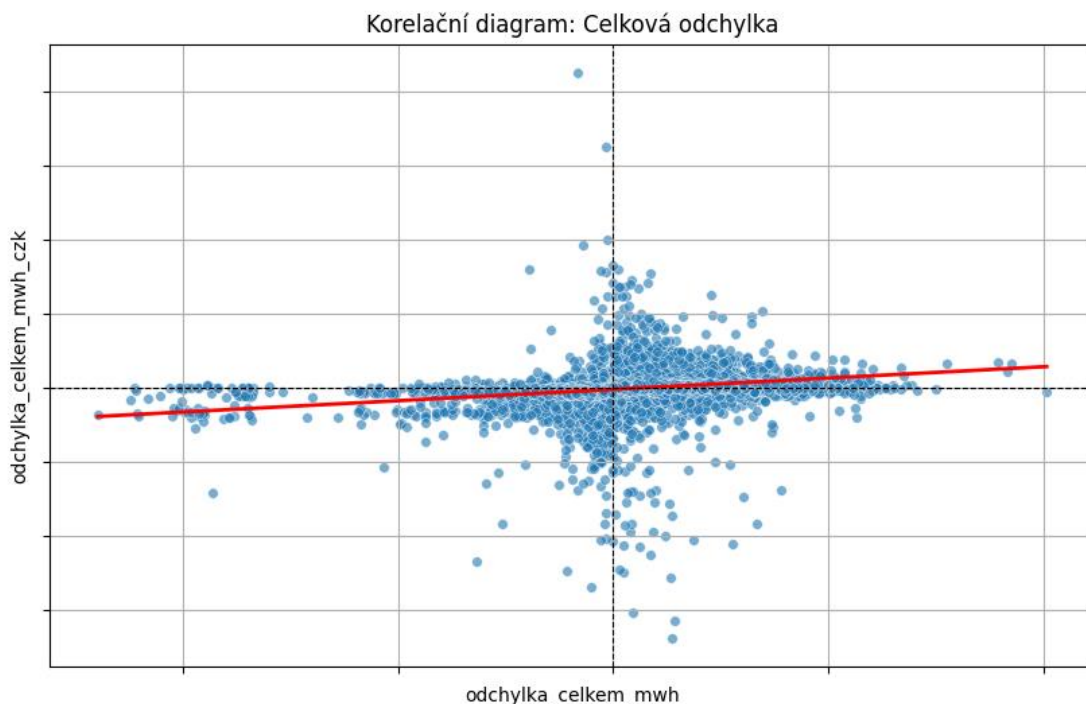
*Test hypotéz statistické významnosti:*

- $H_0$ : Korelace mezi velikostí odchylek a jejich cenou není statisticky významná ( $p > 0,05$ )

- *H1: Korelace mezi velikostí odchylek a jejich cenou je statisticky významná ( $p \leq 0,05$ ).*

Výsledná hodnota Spearmanova korelačního koeficientu činí  $r_s = 0,471817$  a p-hodnota činí  $p = 0,0$ .

U testu *hypotézy o existenci monotónního vztahu proměnných* se přijme alternativní hypotéza H1 a u testu *hypotéz statistické významnosti* se přijme alternativní hypotéza H1. Hodnota vypočteného koeficientu představuje středně silný monotónní vztah mezi velikostí odchylek s jejich cenou. Což může naznačovat, že kladné odchylky mohou být spíše protiodchylkou, než odchylkou a naopak. Nicméně je nutné dodat, že při pohledu na klasický korelační diagram (graf 3) je patrná vysoká rozptýlenost bodů kolem průniku osy  $x$  a  $y$ . Tato rozptýlenost se jeví jako náhodná. Výsledek korelační analýzy je tedy nutné brát se značnou rezervou.



**Graf 3: Korelační diagram odchylky (MWh) a jednotkové ceny odchylky (CZK/MWh)**

(Zdroj: Vlastní zpracování)

## 2.5 Zhodnocení analýzy současného stavu

Společnost disponuje rozvinutou infrastrukturou pro správu podnikových procesů, přičemž klíčovou roli zde hrají systémy SAP ERP a nástroje z ekosystému Microsoft. Tyto technologie poskytují robustní základ pro běžné operace i podporu administrativních

činností. Nicméně v oblasti reportingu a pokročilé analytiky jsou využívány zejména nástroje, které vyžadují manuální zásahy, jako je Microsoft Excel. To s sebou přináší zvýšené riziko chyb, omezenou škálovatelnost řešení a nízkou efektivitu při práci s větším množstvím dat.

V oblasti predikce spotřeby elektrické energie je implementován vlastní predikční systém vyvinutý třetí stranou. Ten využívá expertní odhady technických pracovníků a norem spotřebičů k vytvoření realizačních diagramů (ERD).

Automatizovaný proces nákupu elektřiny na denním trhu přes komunikační rozhraní OTE zajišťuje určitou úroveň efektivity, nicméně v oblasti analytického zhodnocení predikcí a spotřeby zůstává značný prostor pro zlepšení. Stávající reportingové nástroje predikčního systému nenabízejí dostatečnou hloubku, flexibilitu ani uživatelský komfort. Absence možnosti exportu dat a nedostatečná granularita grafických výstupů brání efektivní analýze a zpětné kontrole přesnosti predikcí.

Důležitým zjištěním je také roztržitost datové základny – historická měření jsou uložena ve třech separátních tabulkách odpovídajících různým energetickým řídicím systémům. Datová integrita je sice zajištěna pomocí mapovací tabulky, ale současná struktura neumožňuje efektivní datovou konsolidaci a hlubší analýzy bez pokročilejšího ETL procesu a vytvoření vlastního analytického úložiště dat.

Z analytické části vyplývá, že ve společnosti sice existuje rámec pro řízení spotřeby elektřiny a obchodování na krátkodobých trzích, avšak chybí ucelené a uživatelsky přívětivé řešení pro komplexní reporting a podporu rozhodování.

Zkoumání odchodů na krátkodobých trzích s elektřinou ukázalo výraznou převahu nákupu nad prodejem na denním trhu. Medián u vnitrodenního trhu byl spíše blízko nule, což potvrdilo jeho vyrovnávací funkci. Ceny elektřiny na trzích vykazují vyšší ceny na denním trhu s výrazně nižší variabilitou než na vnitrodenním trhu, který je více volatilní.

Analýza obchodní činnosti společnosti na krátkodobých trzích s elektřinou potvrdila výraznou převahu nákupů nad prodeji na denním trhu. Naopak vnitrodenní trh vykazuje výrazně vyrovnanější bilanci, přičemž medián objemového salda se pohyboval blízko nuly, což potvrzuje jeho funkci jako vyrovnávacího mechanismu v rámci operativního řízení spotřeby.

Z pohledu odchylek mezi predikovanou a skutečnou spotřebou bylo zjištěno, že kladné odchylky (nadhodnocení) převládají nad zápornými. Tato skutečnost naznačuje systematickou tendenci predikcí nadhodnocovat spotřebu. Ukazatele variability i střední hodnoty (mediány) u všech typů odchylek v průběhu sledovaného období klesaly, což může být vnímáno jako zlepšující se přesnost predikčního systému. Finanční vyjádření odchylek ukázalo, že kladné odchylky častěji představují protiodchylky, které generují výnosy, zatímco záporné odchylky jsou obvykle penalizovány a představují náklad. Z hlediska velikosti těchto finančních dopadů však penalizace převažují nad odměnami, což lze interpretovat jako motivační nástroj pro co nejpřesnější predikce a prevenci vzniku odchylek.

Statistické metriky přesnosti predikcí – střední absolutní chyba (MAE), odmocněná střední kvadratická chyba (RMSE) a systematická chyba (Bias) – potvrdily zlepšující se přesnost predikcí v čase, ačkoli i nadále zůstává průměrná odchylka v rozsahu 1,5–2 MWh, což představuje přibližně 16 % mediánu objemu obchodů a může mít významný dopad na celkovou bilanci společnosti.

Za účelem hlubšího porozumění vztahu mezi velikostí a finančním dopadem odchylek byla provedena Spearmanova korelační analýza. Výsledná hodnota korelačního koeficientu ( $r_s = 0,4718$ ;  $p = 0,0$ ) ukazuje na statisticky významný, středně silný monotónní vztah mezi velikostí odchylky a její cenou. Tento vztah však nevypovídá přímo o kauzalitě. Vizualizace dat pomocí korelačního diagramu naznačuje, že zejména menší odchylky nevykazují výrazný strukturální trend, což svědčí o náhodném charakteru menších nepřesností. Z toho vyplývá, že zatímco větší odchylky mohou mít finančně predikovatelné důsledky, menší odchylky nemusí být systematického původu.

Celkově lze říct, že současný systém predikcí a jejich vyhodnocení ukazuje na určité pokroky ve zlepšování přesnosti, ale zároveň poukazuje na potřebu modernizace a zpřesnění analytických nástrojů, které by umožnily efektivněji řídit predikce spotřeby elektřiny.

## 3 VLASTNÍ NÁVRHY ŘEŠENÍ

Primárním účelem reportního systému je porovnávat predikovanou spotřebu s reálnou spotřebou podle spotřebičů, provozů, závodů a divizí a odhalovat tak nepřesnosti predikcí způsobených špatným nastavením nebo lidským faktorem, identifikovat příčiny, zdroje a místa vzniku odchylek a docílit minimalizaci odchylek a s ním spojeného finančního rizika penalizací operátorem trhu s elektřinou (OTE). Na základě výsledků statistické analýzy je patrná obchodní strategie nadhodnocování predikce spotřeby elektřiny, která souvisí s existencí středně silné pozitivní kolorace směru odchylky a její ceny. Cílená kladná odchylka by však neměla být příliš vysoká, protože odměna za odchylku nemusí převyšovat cenu elektřiny nakoupené na krátkodobém trhu, která byla zaplacená, ale nebyla spotřebována. Dále existuje riziko, že i kladná odchylka může stále přispívat k systémové odchylce a byla by tak penalizována vyšší cenou (v absolutní hodnotě).

### 3.1 Datový model a proces ETL

Prvním krokem k implementaci reportního systému je návrh dimenzionálního datového modelu datového tržiště a procesu ETL. K současnému problému bude systém využívat data z podnikového Microsoft SQL Serveru, a to konkrétně ze dvou databází – relační databáze aplikace pro predikci elektřiny a relační databáze se zdrojovými tabulkami dat měření z různých energetických systémů. V této kapitole nebudou plně popisována technická řešení architektury datového tržiště, ale pouze jeho logický rámec a nutné transformace pro správného fungování.

#### 3.1.1 Integrace a transformace zdrojových dat

Jak již bylo nastíněno v předchozí kapitole, pro vytvoření využitelného multidimenzionálního datového modelu s architekturou OLAP bude nutné provést integraci a transformaci zdrojových dat.

Datový model bude mít tři hlavní dimenze:

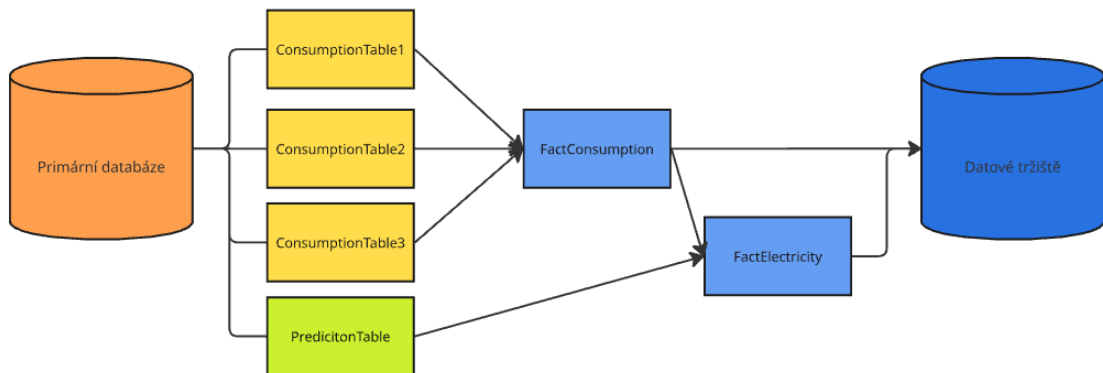
- Dimenze času – nejnižší úroveň detailu 1 hodina
- Dimenze uživatelů
- Dimenze spotřebičů

Z pohledu dimenze spotřebičů příslušnosti jsou data ve zdrojových tabulkách agregována v rozdílné granularitě:

- Zdrojové tabulky *ConsumptionTable1*, *ConsumptionTable2*, *ConsumptionTable3* jsou agregovány na nejnižší úrovni detailu – hodnoty měření spotřeby jednotlivých elektroměrů
- Zdrojová tabulka *PredicitonTable* je na úrovni spotřebičů

Pro kýžený datový model by měla být vytvořena faktová tabulka s názvem *FactElectricity*, která bude plnit primární funkci datového uložení pro reportní systém a bude obsahovat data predikcí spotřeby a reálných spotřeb.

Pro zachování granularity na úrovni elektroměrů<sup>7</sup> bude v datovém tržišti existovat i tabulka *FactConsumption*, která bude integrovat tři zdrojové tabulky z primární databáze. Z tabulek *PredictionTable* a *FactConsumption* bude vytvořena finální tabulka *FactElectricity*.



**Obrázek 11: Datový tok faktů z produkční databáze do datového tržiště**

(Zdroj: Vlastní zpracování v Miro)

### 3.1.1.1 Proces integrace a transformace faktové tabulky *FactConsumption*

Vzhledem ke skutečnosti, že každý energetický systém je jiný, tak i zdrojové tabulky jsou rozdílné. Identifikované rozdíly spočívají především v:

---

<sup>7</sup> Především pro potřeby energetiků.

- Časových pásmech sloupců časových razítek měření (identifikátory datum a čas)
- Identifikátorech elektroměrů
- Jednotkách spotřeby (MWh/kWh)

Prvním krokem je transformace a sjednocení identifikátorů elektroměrů. Tabulky *ConsumptionTable2* a *ConsumptionTable3* mají klasické celočíselné identifikátory, ale tabulka *ConsumptionTable1* má identifikátor ve formě textových řetězců, které však vykazovaly jednotný vzor — každý řetězec zahrnoval ve své struktuře unikátní číselný kód, například ve formátu *EM2001\_QS*, *EM2002\_QS*, *EM2003\_QS* <sup>8</sup> apod. V rámci transformačního kroku je nutná extrakce čtyřciferného číselného údaje ze středu textového identifikátoru, čímž je možné jej převést na jednotný číselný formát.

Při porovnání všech tří zdrojových tabulek se ukázalo, že ačkoliv v rámci jednotlivých tabulek nedochází k duplicitám, napříč tabulkami se mohou stejné hodnoty identifikátorů vyskytovat opakovaně. To představuje riziko kolize při spojování dat do jediné faktové tabulky. Pro zajištění jednoznačnosti identifikátorů proto zvoleno řešení, kdy se při transformaci přidává k identifikátorům z každé tabulky specifický číselný prefix. Každá tabulka má přiřazený jiný rozsah identifikátorů, například 1 000 000, 2 000 000 a 3 000 000. Tyto prefixy jsou dostatečně vysoké na to, aby pokryly veškeré existující hodnoty a zároveň zajistily, že identifikátory z jednotlivých tabulek budou vždy unikátní i v rámci společné faktové tabulky *FactConsumption*.

**Tabulka 9: Sjednocené identifikátory tabulky FactConsumption**

Tabulka	Příklad původního identifikátoru	Příklad nového identifikátoru
<b>ConsumptionTable1</b>	EM2001_QS	1002001
<b>ConsumptionTable2</b>	50001	2050001
<b>ConsumptionTable3</b>	11	3000011

(Zdroj: Vlastní zpracování)

Další důležitou součástí procesu transformace je unifikace časových razítek měření, která se ve zdrojových tabulkách liší nejen svým formátem, ale i použitou časovou zónou. Aby

---

<sup>8</sup> Toto jsou pouze ukázkové textové řetězce jejichž struktura a hodnota je pozměněna z důvodu citlivosti dat.

je bylo možné správně analyzovat a porovnávat napříč všemi zdroji, je potřeba převést časová razítka do jednotného standardu.

Ve dvou ze tří zdrojových tabulek jsou časové údaje uloženy v koordinovaném světovém čase (UTC), což zajišťuje vysokou přesnost a konzistenci. Naproti tomu třetí tabulka pracuje s časovými razítky ve středoevropském čase (CET), který se během roku mění na středoevropský letní čas (CEST). Tato rozdílnost vyžaduje, aby se časové údaje převedly do jednotného formátu, nejlépe do obou časových zón – jak do UTC, tak do CET/CEST. Časová razítka budou ve dvou formátech datového typu – datum a čas a celé číslo: *yyyy-mm-dd hh:mm:ss* a *yymmddhhmm* (2024-01-01 01:00:00 a 2401010100). Celé číslo bude sloužit jako identifikátor pro propojení s dimenzí datumu a času (DimDateTime).

Posledním krokem pro transformaci dat faktové tabulky je sjednocení jednotek měření z kWh (*ConsumptionTable2*) na MWh (*ConsumptionTable1*, *ConsumptionTable3*) a přidání nových sloupců *SourceName* a *SourceId*, které budou obsahovat název zdrojové tabulky z primární databáze a vlastní identifikátor (první celé číslo stanoveného prefixu) pro snadné rozlišení a filtraci dat.

### **3.1.1.2 Proces integrace a transformace faktové tabulky FactElectricity**

Faktová tabulka *FactElectricity* bude hlavním zdrojem dat datového modelu pro reportingový nástroj. Hlavní úlohou pro získání kýžené podoby tabulky bude propojení tabulky predikovaných spotřeb v hodinových intervalech spotřebičů společnosti – *PredictionTable* a nové tabulky *FactConsumption*.

Tabulka *PredictionTable* obsahuje důležité sloupce nezbytné pro správné fungování datového modelu:

- *PredictionTableId* – sloupec celočíselných primárních klíčů záznamů predikcí
- *ApplianceId* – sloupec celočíselných identifikátorů spotřebičů (cizí klíč) - tedy ke kterému spotřebiči byla predikce vytvořena.
- *Created* – sloupec s časovými razítky, kdy byly záznamy nahrány do tabulky (zároveň i čas odeslání souhrnné denní predikce spotřeby pro nákup elektřiny na denním trhu)
- *UserId* – sloupec celočíselných identifikátorů (cizí klíč) uživatele (predikujícího), který predikci vytvořil. Obsahuje i prázdné (*null*) hodnoty v případech, kdy predikce nebyla vytvořena a bylo použito výchozí nastavení spotřeby spotřebiče.
- *Modified* – sloupec s časovými razítky, kdy byla predikce uživatelem odeslána do systému aplikace.
- *Date* – sloupec s daty, ke kterým byla predikce vytvořena.
- *PowerConsumptionMWH0* – *PowerConsumptionMWH23* – 24 sloupců predikovanými hodinovými hodnotami spotřeby pro daný záznam, resp. den<sup>9</sup>.

Její původní struktura má formát, kde každému záznamu (řádku) odpovídá jeden kalendářní den a spotřeba je vyjádřena ve 24 samostatných sloupcích (*PowerConsumptionMWH0* až *PowerConsumptionMWH23*), odpovídajících hodinám dne. Tato struktura není vhodná pro časové modelování a analytické operace, které vyžadují jednotnou časovou granularitu a možnost přímého spojení s dimenzemi času, spotřebiče nebo uživatele.

Tato transformace bude probíhat procesem pivotování (odstranění sloupců 0–23 a jejich převedení na hodnoty v novém sloupci *Hour*) a vytvoření nového sloupce *DateTime*, který vznikne kombinací *Date* a *Hour*. Po vzoru tabulky *FactConsumption* bude vytvořen sloupec celočíselných identifikátorů *DateTimeId*.

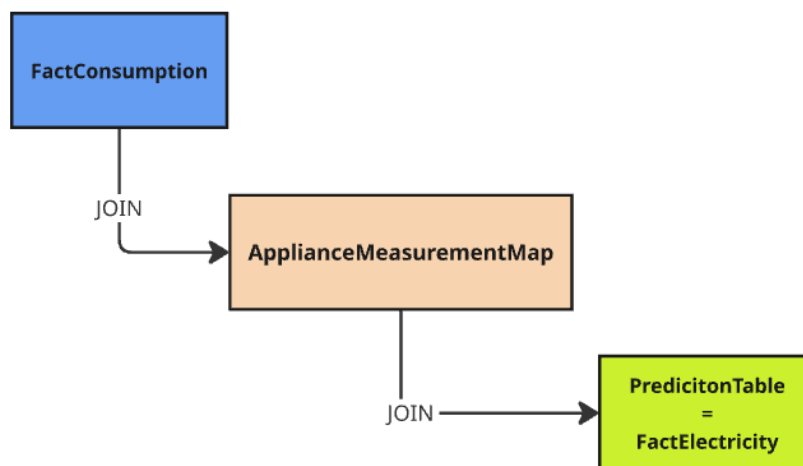
Dále bude nutná transformace sloupce s identifikátory uživatelů *UserId*, kde musí být prázdné hodnoty (*null*) nahrazeny číselnou hodnotou, která bude reprezentovat výchozí nastavení predikce spotřebičů. Vzhledem k možným změnám (přidání nových uživatelů)

---

<sup>9</sup> Horizontální rozložení hodnot predikce v čase je založeno na bázi predikčního ERD diagramu, který je základním nástrojem pro obchodování na denním trhu s elektřinou.

je nutné zvolit hodnotu, která nebude narušovat rostoucí číselnou posloupnost identifikátorů. Z tohoto důvodu budou všechny hodnoty nahrazovány specifickou hodnotou -1, která zachovává konzistenci číselného typu sloupce.

Nejkomplexnějším krokem v procesu transformace a integrace dat bude připojení tabulky *FactConsumption* na tabulku predikcí. Zdrojové tabulky spotřeby jsou s aplikací predikce napojeny přes spojovací tabulku *ApplianceMeasurementMap*, která obsahuje identifikátory spotřebičů (*ApplianceId*) a k nim v aplikačním rozhraní uživateli (administrátoři aplikace) přiřazené identifikátory elektroměrů ze všech zdrojových tabulek, které mají datový typ textového řetězce (především kvůli textovým identifikátorům tabulky *ConsumptionTable1*). Spojovací tabulka bude využita pro stejný účel i v datovém tržišti, s tím rozdílem, že původní identifikátory elektroměrů budou nahrazeny novými, podle stejného řešení tabulky *FactConsumption*.



**Obrázek 12: Integrace faktů predikcí a reálné spotřeby elektřiny**

(Zdroj: Vlastní zpracování)

Tímto spojením je možné agregovat hodnoty spotřeby na úrovni elektroměrů na úroveň spotřebičů, kdy platí pravidlo, že jeden spotřebič může mít jeden a více elektroměrů, a že jeden elektroměr má přiřazený právě jeden spotřebič. Jak již bylo zmíněno v analytické části ne všechny elektroměry jsou relevantní, a tedy přímo přiřazené v predikční aplikaci k danému spotřebiči<sup>10</sup>. Touto agregací jsou vyfiltrována data

---

<sup>10</sup> Jedná se o elektroměry podružné, které například měří část spotřeby komplexního spotřebiče, který v realitě je například soustavou více spotřebičů.

nepřirazených elektroměrů, a právě z tohoto důvodu je tabulka *FactConsumption* jako další faktová tabulka, byť bez využití pro reportní systém.

### 3.1.1.3 Proces integrace a transformace dimenzionálních tabulek

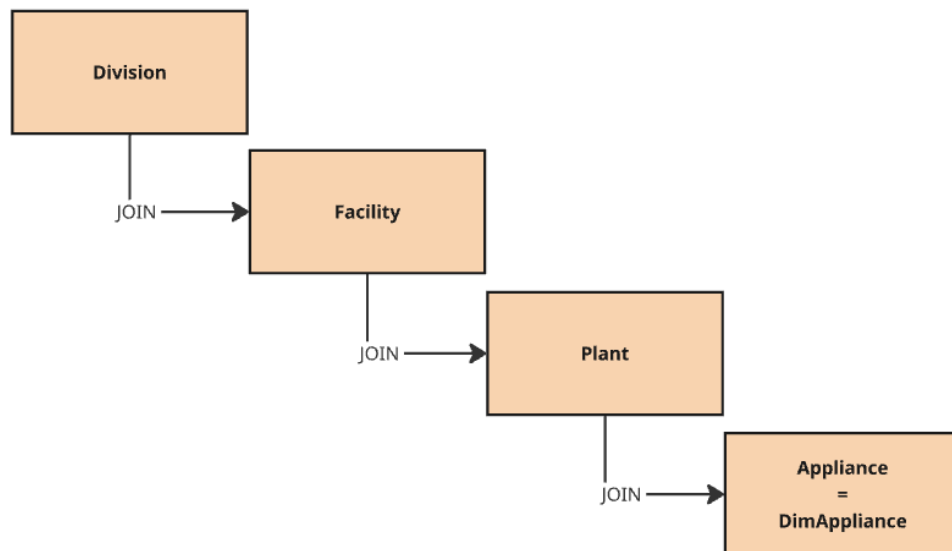
Jak již bylo nastíněno na začátku kapitoly 3.1.1 datový model bude mít tři dimenze – datum a čas (*DimDateTime*), dimenze spotřebičů (*DimAppliance*) a dimenzi uživatelů (*DimUser*).

Dimenzionální tabulka *DimDateTime* bude vycházet z faktové tabulky *FactElectricity*, kdy časové rozpětí záznamů bude tuto tabulku reflektovat (od data nejstaršího záznamu po nejnovější). Nejnižší granularita bude na úrovni jedné hodiny ve formátu datum a čas, resp. jeho celočíselného aliasového sloupce *DateTimeId*, který bude primárním klíčem tabulky. Vzhledem k analytickému využití modelu bude sloupec obsahovat i další sloupce:

- *Date* – sloupec s příslušným datumem záznamu
- *FiscalYear* – sloupec s příslušným kalendářním rokem záznamu
- *FiscalQuarter* – sloupec s příslušným číslem kvartálu záznamu (1-4)
- *FiscalMonth* – sloupec s příslušným číslem měsíce záznamu (1-12)
- *FiscalDayOfWeek* – sloupec s příslušným číslem dne v týdnu (1-7)
- *FiscalHour* – sloupec s příslušným číslem hodiny dne (0-23)
- *WeekdayWeekend* – sloupec s rozlišením záznamů na „pracovní den“ a „víkend“
- *NameQuarter* – sloupec s textovým názvem kvartálu (Q1 až Q4)
- *NameMonth* – sloupec s textovým názvem měsíce (leden až prosinec)
- *NameDayOfWeek* – sloupec s textovým názvem dne v týdnu (pondělí až neděle)
- *WeekOfYear* – sloupec s textovým názvem týdnu v roce a rokem (např. 2024-T01)

Dimenzionální tabulka *DimAppliance* bude popisovat místní a organizační příslušnost daných spotřebičů. Tabulka bude denormalizovanou podobou tabulek z databáze aplikace – *Appliance*, *Plant*, *Facility* a *Division*, kdy budou všechny tabulky přímo spojeny do tabulky *Appliance*, resp. *DimAppliance*. Tabulka bude obsahovat především identifikátory spotřebičů (primární klíč), provozů, závodů a divizí a sloupce s jejich názvy:

- *ApplianceId* – celočíselný identifikátor spotřebiče
- *PlantId* – celočíselný identifikátor provozu
- *FacilityId* – celočíselný identifikátor závodu
- *DivisionId* – celočíselný identifikátor divize
- *NameAppliance* – názvy spotřebičů
- *NamePlant* – názvy provozů
- *NameFacility* – názvy závodů
- *NameDivision* – názvy divizí
- *MeasurementComparasionTolerancePercentPlant* – sloupec s celočíselnými hodnotami, které vyjadřují hodnotu povoleného procentuálního podílu odchylky predikce od reálné spotřeby



**Obrázek 13: Relační diagram integrace organizační dimenze**

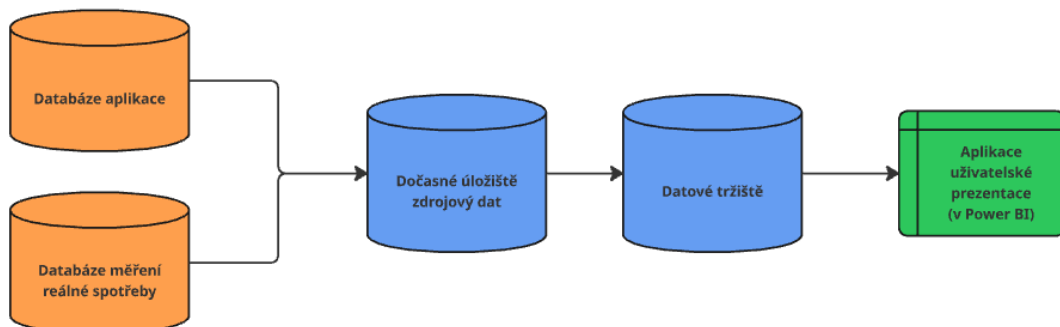
(Zdroj: Vlastní zpracování)

Dimenzionální tabulka *DimUser* bude umožňovat pohled na data z pozice uživatelů, resp. prediktorů. Velká část predikcí je prováděna přímo prediktory, kteří v aplikačním rozhraní vyplňují, ve kterých hodinách dne jaké spotřebiče budou v provozu. Jedná se o lidské rozhodování, ve kterém může docházet k chybám, a proto je důležité sledovat i počínání uživatelů aplikace. Pro dimenzi uživatele je využita zdrojová tabulka *User*. Využitelné pro datový model jsou sloupce *UserId*, *FirstName* a *SecondName*, ze kterých se vytvoří dodatečný sloupec *FullName*, který bude obsahovat celé jméno uživatele.

### 3.1.2 Výsledná podoba datového tržiště

Cílem procesu ETL je vytvoření stabilního, optimalizovaného a škálovatelného datového tržiště, které poskytuje klíčové informace o predikci a spotřebě elektrické energie v rámci celé organizace. Takto navržené řešení umožňuje jak detailní analýzy, tak i přehled o spotřebě na celopodnikové úrovni. Důležitou vlastností datového tržiště je jeho škálovatelnost, která umožňuje nejen zpracování rostoucího objemu dat, ale také rozšiřování o nové oblasti. V budoucnu se počítá s integrací dalších typů dat, například finančních údajů, které umožní rozšířit analytické možnosti systému o pohledy spojené s náklady, výnosy či ekonomickým hodnocením spotřeby.

Architektura datového toku bude obsahovat i dočasné úložiště zdrojových dat (*Staging Area*), která bude umístěna přímo v datovém tržišti a bude obsahovat kopie zdrojových tabulek, do kterých se při každé aktualizaci nahrají netransformovaná data ze zdrojových databází a k náročným transformačním úlohám bude docházet přímo v datovém tržišti, což zajistí minimální zatížení zdrojových transakčních databází.



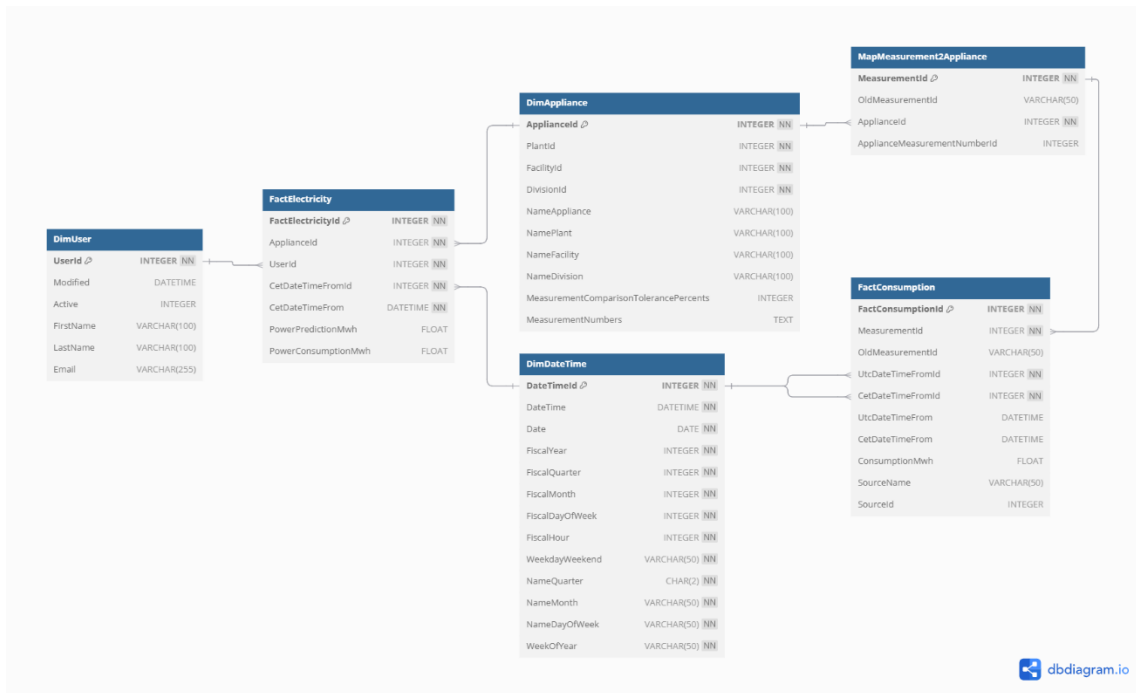
**Obrázek 14: Architektura datového toku vlastního návrhu**

(Zdroj: Vlastní zpracování)

Díky své standardizované struktuře a optimalizované podobě je datové tržiště vhodné pro široké spektrum nástrojů určených k datové analýze a reportingu. Uživatelé mohou data snadno načítat nejen pomocí pokročilých BI nástrojů (např. Power BI), ale také prostřednictvím běžně dostupných aplikací, jako je Microsoft Excel, který je ve společnosti stále preferovaným nástrojem pro ad-hoc analýzy i pravidelný reporting.

V rámci provozu datového tržiště se předpokládá automatická denní aktualizace dat, která zajistí pravidelný přísun nových měření a predikcí. Takto nastavený frekvenční režim představuje vyvážený kompromis mezi aktuálností dat a efektivitou zpracování, přičemž

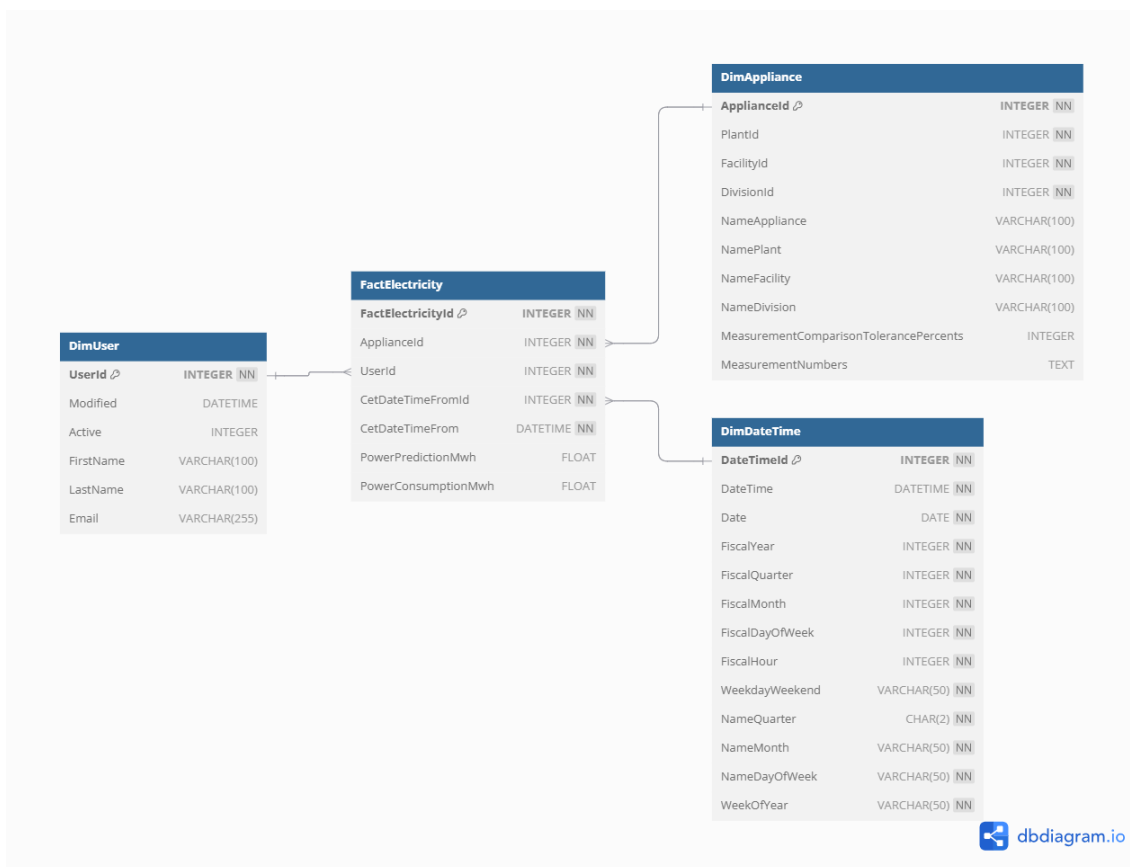
umožňuje vytvářet každodenní reporty a analýzy s minimální prodlevou. Tím se datové tržiště stává spolehlivým základem pro pravidelný reporting, analýzu a včasné rozhodování.



**Obrázek 15: Relační návrh datového tržiště**

(Zdroj: Vlastní zpracování v dbdiagram.io)

Datový model, který bude využíván nástrojem BI pro účely vizualizace dat, je navržen ve formě hvězdicového schématu (star schema). Tento model vychází z datového tržiště (bez tabulky *FactConsumption*) a představuje strukturu vhodnou pro analytické dotazování. V jeho středu se nachází faktová tabulka *FactElectricity*, která obsahuje kvantitativní údaje o spotřebě a predikci elektrické energie a ta je obklopena několika dimenzionálními tabulkami, které poskytují kontextové informace.



Obrázek 16: Návrh datového modelu pro uživatelskou aplikaci Power BI

(Zdroj: Vlastní zpracování v dbdiagram.io)

### 3.2 Vlastní návrh reportního systému v prostředí Power BI

Pro implementaci reportního systému je nutné zvolit vhodné softwarové řešení, které umožní efektivní zpracování dat, jejich vizualizaci a analytickou interpretaci.

Firma využívá většinu kancelářských aplikací v rámci ekosystému Microsoft 365 a Microsoft Power Platform, a kromě ERP systému SAP je většina interních procesů postavena právě na nástrojích od společnosti Microsoft. Součástí licenčního balíčku, který má společnost již aktivně zakoupený, je také nástroj Power BI, a je tedy přirozenou volbou pro vizualizaci a prezentaci dat.

Díky existujícím licencím, známému uživatelskému prostředí a snadné integraci s ostatními nástroji Microsoftu (například Microsoft SQL Server, Excel, SharePoint) je Power BI plně kompatibilní s technologickým zázemím firmy. Jeho nasazení proto nevyžaduje žádné dodatečné investice ani změny ve stávající infrastruktuře, což z něj činí praktickou a ekonomicky výhodnou volbu.

Na základě navrženého datového modelu a dostupnosti nástroje Power BI v rámci stávající infrastruktury společnosti je dalším krokem návrh vlastního reportního systému, který umožní efektivní vizualizaci a interpretaci klíčových dat o spotřebě a predikci elektřiny.

Cílem této části je vytvořit přehledné, interaktivní a uživatelsky srozumitelné dashboardy, které pokrývají jak detailní analytické pohledy, tak i agregované přehledy na úrovni celé společnosti. Návrh reportů vychází z požadavků cílových uživatelů (především nákupčí elektřiny a prediktoři) a klade důraz na intuitivní ovládání, responzivní vizualizace a možnost filtrování dat podle různých dimenzí.

Reportní systém bude navržen modulárně, s důrazem na:

- Přehled denní a historické spotřeby.
- Sledování predikované vs. skutečné spotřeby.
- Sledování odchylek s nimi spojenými ukazateli.
- Filtrování podle uživatele, lokality, spotřebiče nebo časového úseku.

### **3.2.1 Datový tok**

Prvním krokem v nástroji Power BI je připojení na datový zdroj, import dat, možná úprava datového modelu a nastavení automatické aktualizace dat – jednou denně.

Pro zajištění aktuálnosti reportovaných dat je Power BI nutné napojit přímo na SQL Server, kde se nachází datové tržiště vytvořené v rámci první fáze ETL procesu. Napojení je možné prostřednictvím přímého dotazu (*Direct Query*) nebo importu dat (*Import Mode*). V tomto případě je zvolen režim importu dat, který umožňuje rychlé načítání a plynulé zobrazení vizualizací i při práci s většími objemy dat. Výhodou tohoto přístupu je také možnost nastavení denní inkrementální (*Delta*) aktualizace, která při pravidelném osvěžování načítá pouze nově přidané nebo změněné záznamy v datovém tržišti namísto kompletního datasetu. Tím se výrazně snižuje zatížení systému, zkracuje doba aktualizace a zároveň se zvyšuje celková výkonnost a škálovatelnost řešení.

### **3.2.2 Uživatelské rozhraní sestavy**

Před samotným návrhem vizualizací a výpočtových měř je nezbytné definovat základní koncepci uživatelského rozhraní (UI) celé sestavy. V této fázi nejde o konkrétní grafy či

tabulky, ale o promyšlené uspořádání a logiku prezentace dat, tedy jakým způsobem budou uživatelé se sestavou pracovat a jak snadno v ní naleznou požadované informace. Cílem je vytvořit přehledné, intuitivní a vizuálně konzistentní rozhraní, které podpoří uživatelskou zkušenost (UX) a zároveň bude odpovídat potřebám cílových skupin uživatelů. Rozložení jednotlivých prvků, použití barev, typů vizualizací i navigačních prvků přímo ovlivňuje funkčnost a použitelnost.

### **3.2.2.1 Barvy a typografie**

Pro barevné řešení vizualizací a jednotlivých prvků byly zvoleny umírněné a kontrastní barvy, které zajišťují dobrou čitelnost jak na světlém, tak i na tmavším podkladu. Jako základní barevná paleta byla zvolena kombinace modré a bílé, přičemž modrá barva se používá v různých odstínech a sytostech pro odlišení kategorií nebo hodnot. Bílá barva slouží jako výchozí pozadí vizualizací, zatímco samotné stránky sestavy využívají světle šedý podklad, který vytváří jemný kontrast a přispívá k lepší vizuální oddělitelnosti jednotlivých prvků.

V rámci samotných vizualizací jsou barvy vybírány s ohledem na jejich význam. Kontrastní barevné dvojice (například modrá a červená) jsou použity pro zobrazení porovnávaných hodnot, aby rozdíly byly na první pohled patrné. Naopak u dat, která spolu souvisejí nebo se doplňují, jsou použity jemnější odstíny (např. modrá a světle modrá), které vizuálně signalizují jejich propojenost bez nadměrného zvýraznění.

Pro typografii sestavy byla zvolena fontová rodina *Segoe UI*, konkrétně *varianty Segoe UI, Segoe UI Semibold a Segoe UI Bold*. Jedná se o bezpatkové a vizuálně čisté písmo. Díky dostupnosti různých stupňů tučnosti umožňuje Segoe UI snadné vizuální odlišení jednotlivých textových prvků, jako jsou nadpisy, popisky nebo klíčové hodnoty.

### **3.2.2.2 Rozložení stránky sestavy**

Výchozí rozložení sestavy se dá rozdělit na čtyři části (viz příloha 9):

- **Zobrazovací oblast**
- **Horizontální fixní ovládací panel**
- **Vertikální fixní ovládací panel**
- **Rozbalovací ovládací panel**

**Horizontální fixní ovládací panel** (umístěný podél horní hrany sestavy) obsahuje název stránky, tlačítka a průřezy pro filtraci dat podle dimenze místa (*DimAppliance*) a umožňují tak filtraci podle divizí, závodů a provozů.

**Vertikální fixní ovládací panel** (umístěný podél levé hrany sestavy) obsahuje piktogramy, které po stisknutí přepínají zobrazované sestavy.

**Rozbalovací ovládací panel** umožňuje filtraci podle času (*DimDateTime*) a je použit především z důvodu čistosti a přehlednosti sestavy tak, aby průřezy tvořily jen zlomek plochy sestavy a nezabíraly místo vizuálům dat. Ovládací panel se otevírá tlačítkem s ikonou „trychtýře“ a lze ho zavřít buď ikonou „křížku“ nebo kliknutím mimo panel průřezů.

Vzhledem k rozdílné organizační struktuře jednotlivých divizí vznikají specifické požadavky na způsob vizualizace dat. Tři z divizí jsou členěny hierarchicky ve struktuře divize → závody → spotřebiče, zatímco čtvrtá divize využívá podrobnější členění ve formátu divize → závody → provozy → spotřebiče. Tento rozdíl se promítá i do způsobu, jakým je třeba data prezentovat v reportech.

Zvláštní situace nastává u divize č. 2, kde predikce spotřeby sice existují na úrovni jednotlivých spotřebičů, ale ve skutečnosti je celková hodnota predikce za celý závod přiřazena k jedinému spotřebiči. Z toho důvodu nejsou standardní výpočty a vizualizace na úrovni spotřebičů pro tuto divizi metodicky správné, protože by vedly ke zkreslení výsledků. Je proto nutné použít specifické metriky a vizualizační prvky, které zohledňují toto omezení a umožňují prezentovat data bez ztráty jejich interpretovatelnosti.

Z výše uvedených důvodů je průřez podle divizí implementován tak, aby umožňoval dynamickou změnu obsahu sestavy. Využita je k tomu funkcionalita záložek (Bookmarks) v prostředí Power BI, která umožňuje tzv. layering, tedy vrstvení vizuálů na sebe. Při přepínání mezi záložkami – například pomocí ovládacího tlačítka – se zobrazí pouze relevantní vizualizace pro danou divizi, zatímco ostatní zůstávají skryté.

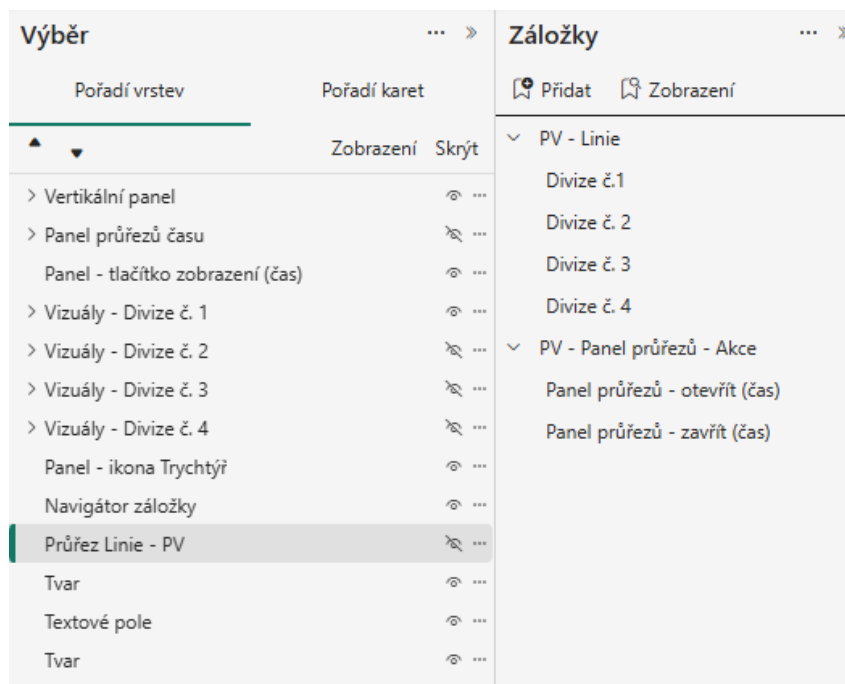
Tento přístup zajišťuje, že každá divize vidí data odpovídající její specifické struktuře, aniž by docházelo k zahlcení nebo dezinterpretaci dat.

Stejný princip je využit také u rozbalovacího ovládacího panelu pro výběr časového období, kde se pomocí záložek zobrazuje či skrývá příslušný ovládací prvek pro filtrování podle času. Tento přístup zvyšuje přehlednost celé sestavy a umožňuje uživateli zaměřit se vždy pouze na relevantní část dat.

### **3.2.2.3 Implementace vrstvení vizuálů pomocí záložek**

Použití vrstvení vizuálů v prostředí Power BI nepředstavuje přímočarou ani intuitivní funkcionalitu, ale spíše pokročilou techniku, která slouží k tvorbě interaktivních a uživatelsky přizpůsobených sestav. Vzhledem k tomu, že tato metoda umožňuje efektivně řešit zobrazení odlišných struktur jednotlivých divizí v rámci jediné sestavy, je vhodné detailně popsat logiku jejího fungování a konkrétní postup implementace. Následující část se proto zaměřuje na praktickou realizaci tohoto přístupu na příkladu jednoho vizuálu.

Pro využití vrstvení vizuálů v Power BI je nezbytné pracovat s podoknem „Výběr“, které umožňuje správu jednotlivých prvků na stránce – konkrétně jejich viditelnosti, pořadí vrstev a seskupování. Současně je využíváno podokno „Záložky“, jehož hlavní funkcí je ukládání aktuálního stavu sestavy. Každá záložka uchovává informaci o tom, které vizualizace jsou viditelné, jak jsou nastavené filtry, průřezy a další prvky. Pomocí těchto záložek je následně možné přepínat mezi různými pohledy na data nebo ovládat zobrazení vrstvených vizuálů prostřednictvím interaktivních tlačítek.



**Obrázek 17: Řešení uživatelského rozhdarní přes funkce „Výběr“ a Záložky“ v prostředí Power BI**

(Zdroj: Vlastní zpracování v Power BI)

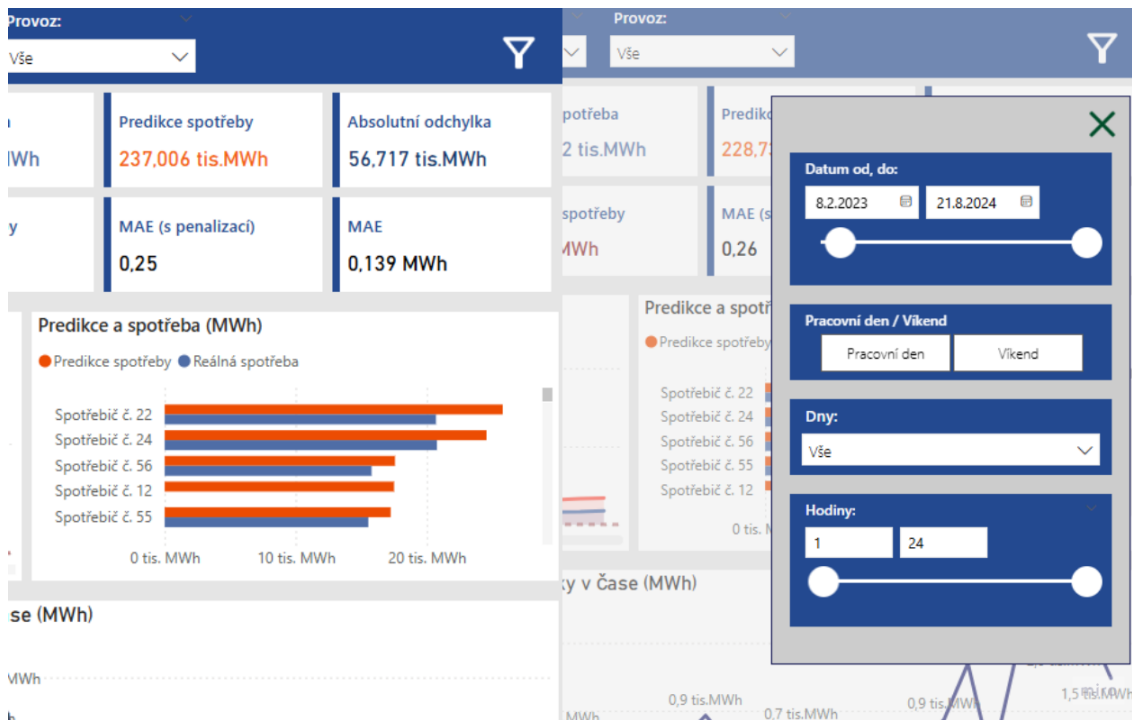
Podokno „Výběr“ tedy umožňuje upravovat pořadí, viditelnost. Podokno „Záložky“ umožňuje nastavení zobrazení. Pro přehlednost a snadnější orientaci jsou jednotlivé vizualizace i záložky seskupeny do logických celků – tak, aby odražely funkční části sestavy a jejich vzájemné souvislosti.

V případě filtrování dat podle divizí je použit tlačítkový prvek typu „Navigátor záložek“, který nahrazuje klasický průřez. Jelikož navigátor záložek sám o sobě nefunguje jako filtr, ale pouze přepíná mezi záložkami s přednastaveným stavem, je v pozadí sestavy neviditelně přítomen průřez divizí, který je trvale skrytý a nastaven tak, aby umožňoval pouze jediný výběr divize. Výběr tlačítka tedy aktivuje odpovídající záložku, která zviditelní konkrétní vizualizace a zároveň nastaví potřebný filtr přes ukrytý průřez. Toto řešení umožňuje dynamické zobrazování rozdílných vizualizací včetně rozdílných průřezů. U divizí, které nemají organizační úroveň provozů, by byl takový průřez zbytečný a pro uživatele potenciálně matoucí.

Dalším prvkem, který využívá princip vrstvení vizuálů, je rozbalovací panel pro výběr časového období. Jeho otevírání a zavírání je řešeno pomocí tlačítek s navázanými akcemi, které mění viditelnost celého panelu. Panel samotný se skládá z několika vrstev – vlastních průřezů pro výběr data, obdélníkového tvaru tvořícího pozadí panelu,

ikony křížku pro manuální zavření a také překryvného tlačítka přes celou stránku s průhledným šedým odstínem.

Toto překryvné tlačítko slouží ke zvýraznění aktivního stavu panelu a zároveň plní funkci zavíracího prvku – pokud uživatel klikne kamkoli mimo samotný obsah panelu (vyjma uzavíracího křížku), panel se automaticky skryje. Tímto způsobem je zajištěno jak intuitivní ovládání, tak i přehlednost celého uživatelského rozhraní bez nutnosti trvalého zobrazování filtrů.



**Obrázek 18: Porovnání rozhraní s aktivním a neaktivním rozbalovacím panelem průřezů**

(Zdroj: Vlastní zpracování v Power BI)

### 3.2.3 Definování metrik a KPIs

Pro efektivní vyhodnocování výsledků reportního systému je nezbytné stanovit soubor metrik a klíčových indikátorů výkonu (Key Performance Indicators, dále jen KPIs), které budou sloužit jako základ pro analytické a rozhodovací procesy. Vzhledem k tomu, že dostupná datová sada neobsahuje finanční informace, jsou navržené metriky orientovány především na hodnocení přesnosti predikce spotřeby a sledování odchylek mezi predikovanými a skutečnými hodnotami.

Výpočty těchto ukazatelů jsou převážně realizovány pomocí jazyka DAX (Data Analysis Expressions), který je nativní součástí Power BI a umožňuje tvorbu výkonných

výpočtových měř a dynamických analytických funkcí (Skripty měř v jazyku DAX jsou uvedeny v příloze 8). DAX poskytuje dostatečnou flexibilitu pro práci s dimenzemi, podmíněnými výpočty i pokročilými metrikami porovnávacími jednotlivé scénáře nebo období.

### 3.2.3.1 Celková reálná spotřeba, predikce spotřeby a odchylky

Základní metriky představující celkový objem reálně zaznamenané spotřeby elektřiny, predikované spotřeby a absolutních odchylek reality od predikce za dané období slouží jako referenční základy pro výpočet relativních metrik a pro jednoduché zobrazování v tabulkách, grafech nebo kartách, kde se využívají pouze jako sloupce faktové tabulky bez výpočtu míry v DAX

### 3.2.3.2 Relativní odchylka

Vyjadřuje, o kolik procent se celková predikce liší od reálné spotřeby, kdy jsou poměřovány celkové sumy. Umožňuje základní srovnání v čase i mezi objekty s různou absolutní spotřebou. Matematický vztah výpočtu je následující:

$$RO = \frac{\sum_{i=1}^n \hat{Y}_i - \sum_i Y_i}{\sum_{i=1}^n \hat{Y}_i} \times 100 [\%] \quad (10)$$

Kde:

$\hat{Y}$  – hodnota predikce [MWh]

$Y$  – hodnota spotřeby [MWh]

$i$  – časový záznam [hod., den, týden atd.]

$n$  – počet prvků

### 3.2.3.3 Střední absolutní chyba – MAE

Průměrná absolutní odchylka mezi predikcí a skutečností, která zobrazuje průměrnou odchylku bez ohledu na její směr. Metrika MAE je odolná vůči extrémním hodnotám a je jednoduše pochopitelná.

### 3.2.3.4 Podíl predikcí v toleranci

Další důležitou metrikou použitou pro hodnocení kvality predikčního modelu je podíl predikcí v toleranci (Within Tolerance Ratio, WTR). Tato metrika udává, jaký podíl všech predikcí se nachází v tolerančním pásmu  $\pm 5\%$  od skutečné spotřeby. Jinými slovy hodnotí, kolik predikcí bylo dostatečně přesných z pohledu praktického použití.

Na rozdíl od metrik typu MAE, které pracují s průměrnou velikostí chyby, WTR neposuzuje velikost odchylky jako takovou, ale spíše četnost přesných predikcí v rámci definované akceptovatelné odchylky. Díky tomu je tato metrika velmi dobře interpretovatelná i pro běžné uživatele, kteří potřebují jednoduchý ukazatel spolehlivosti modelu. Zavedení tolerance zároveň odstraňuje vliv nepodstatných drobných odchylek, které z pohledu provozu nejsou kritické, ale u jiných metrik by zbytečně ovlivňovaly celkový výsledek. Matematický vztah je následující:

$$WTR = \frac{|\{i: |\hat{Y}_i - Y_i| \leq 0,05 \cdot Y_i\}|}{n} \quad (11)$$

Kde:

$\hat{Y}$  – hodnota predikce [MWh]

$Y$  – hodnota spotřeby [MWh]

$i$  – časový záznam [hod., den, týden atd.]

$n$  – počet prvků

### 3.2.4 Struktura sestavy a popis stránek reportu

V této části jsou detailně popsány jednotlivé stránky (dashboards), které tvoří výslednou reportní sestavu v prostředí Power BI. Každá stránka je navržena tak, aby zobrazovala určitý typ informací nebo pohledu na data z různých pohledů. Stránky jsou navrženy s ohledem na přehlednost, uživatelskou přívětivost a především na specifické potřeby různých typů uživatelů – od technických pracovníků, nákupčích až po střední managementu. Důraz je kladen na modularitu, interaktivitu a možnost filtrování dat podle různých dimenzí (čas, divize, lokalita, zařízení, prediktor apod.). Power BI zároveň podporuje interaktivní filtrování prostřednictvím vizuálů, což znamená, že uživatel může kliknutím na konkrétní prvek (např. sloupec, datový bod nebo legendu) automaticky

filtrovat obsah ostatních vizualizací na stránce. Tato funkcionality výrazně zvyšuje flexibilitu a intuitivnost práce s reportem.

Každá stránka sestavy obsahuje kombinaci grafických a tabulkových vizualizací, doplněnou o definované metriky, KPIs a ovládací prvky. Finální podoba reportního systému v Power BI je tvořena pěti samostatnými stránkami (dashboardsy), z nichž každá je zaměřena na jinou úroveň detailu a analytického pohledu. Tento modulární přístup umožňuje uživatelům snadno přistupovat k informacím podle jejich konkrétních potřeb.

#### **3.2.4.1 Úvodní stránka**

První stranou v sestavě je úvodní stránka (viz příloha 10) s názvem sestavy – Spotřeba a predikce elektřiny, Reportní systém. Úvodní stránka neobsahuje žádné vizualizace pouze grafické prvky a vertikální panel pro přepínání stránek sestavy. Úvodní stránka může dále obsahovat doplňující informace a kontaktní údaje na správce sestavy.

#### **3.2.4.2 Energetický přehled**

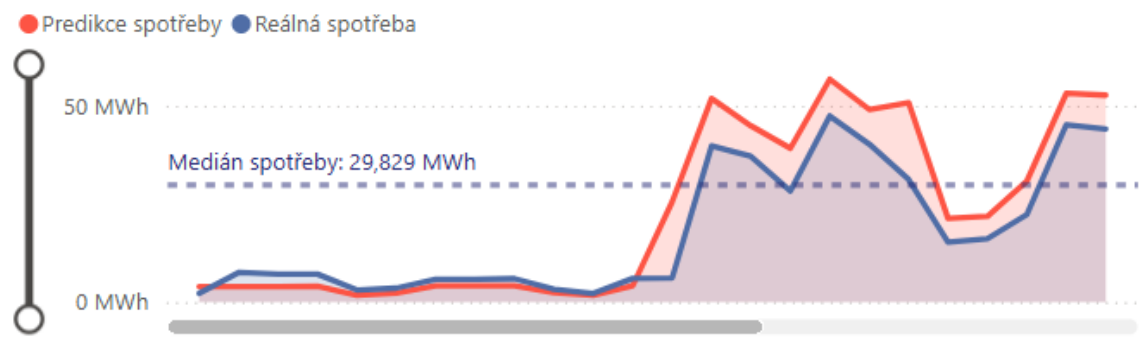
Druhá stránka sestavy (viz příloha 11) obsahuje pohled na spotřebu a predikci z celopodnikové úrovně s možností filtrace na jednotlivé divize pomocí průřezu (jediná stránka sestavy, které pro filtraci podle divizí nevyužívá tlačítko navigátoru záložek). Stránka obsahuje celkem pět vizualizací, které kombinují časový vývoj spotřeby a predikce, kumulativní srovnání a KPIs:

- **Vývoj spotřeby a predikce elektřiny (MWh)** – spojnicový graf
- **Vývoj odchylky (MWh)** – spojnicový graf
- **Spotřeba (MWh)** – mapa stromové struktury
- **Odchylky a spotřeba (MWh)** – spojnicový a skládaný sloupcový graf
- **Matice s metrikami a KPIs**

**Vývoj spotřeby a predikce elektřiny (MWh)** umožňuje uživateli pohled na základní časový vývoj spotřeby a predikce. Čára predikce je zvýrazněna světle červenou barvou a čára spotřeby modrou barvou. Tato barevná kombinace vytváří dostatečný kontrast pro jednoduché rozlišení. Na ose Y jsou položeny sloupce z faktové tabulky *PowerPredictionMwh* a *PowerConsumptionMwh*, které mají vlastní formát zaokrouhlení na tři desetinná místa (pro zachování kWh) s jednotkou MWh. Na ose X je položena hierarchie sloupců časové dimenze – rok, kvartál, měsíc, týden a den (funkce vnořování

na různé úrovni v hierarchii). Graf je také doplněn o referenční linku mediánu spotřeby, která je zobrazena jako přerušovaná světle modrá čára s popiskem a hodnotou.

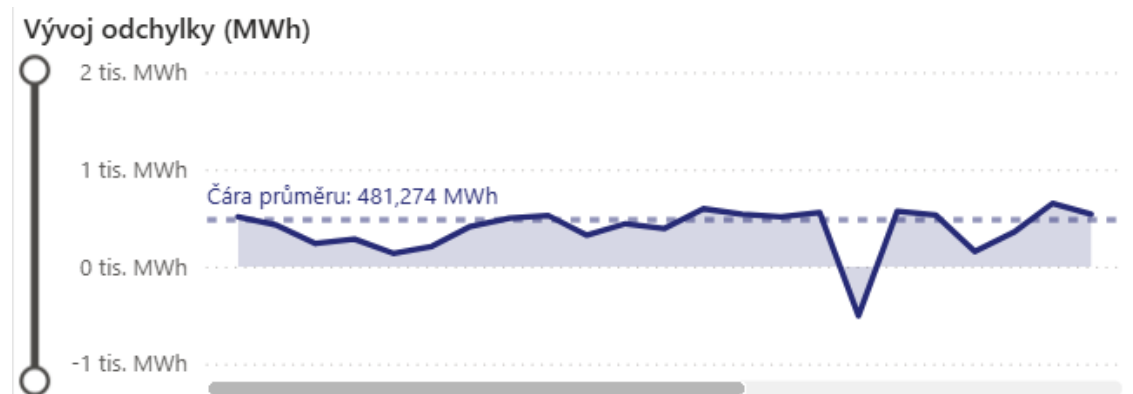
#### Vývoj spotřeby a predikce elektřiny (MWh)



Obrázek 19: Ukázka vizuálu Vývoj spotřeby a predikce elektřiny (MWh)

(Zdroj: Vlastní zpracování v Power BI)

**Vývoj odchyly (MWh)** je spojnicový graf, který podobně jako přechodí vizualizace umožňuje uživateli zobrazení vývoje odchyly v čase. Na ose Y je položený sloupec z faktové tabulky *DeviationMwh*, Na ose X je položena hierarchie sloupců časové dimenze – rok, kvartál, měsíc, týden a den (funkce vnořování na různé úrovni v hierarchii). Graf je opět doplněn o referenční linku, tentokrát s aritmetický průměrem.

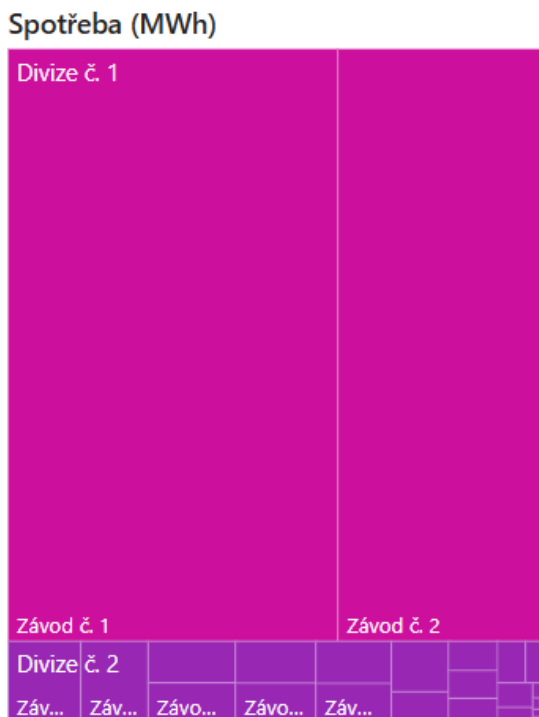


Obrázek 20: Ukázka vizuálu Vývoj odchyly (MWh)

(Zdroj: Vlastní zpracování v Power BI)

**Spotřeba (MWh)** je vizuál mapy stromové struktury, který zobrazuje rozložení spotřeby elektřiny mezi jednotlivé závody v rámci celé společnosti. Barevné odlišení obdélníků podle divizí umožňuje snadnou vizuální orientaci a zároveň přispívá k porovnání spotřeby mezi jednotlivými organizačními jednotkami. Uživatel tak na první pohled získá přehled o tom, které divize a závody jsou nejvýznamnějšími odběrateli elektřiny, a to jak v

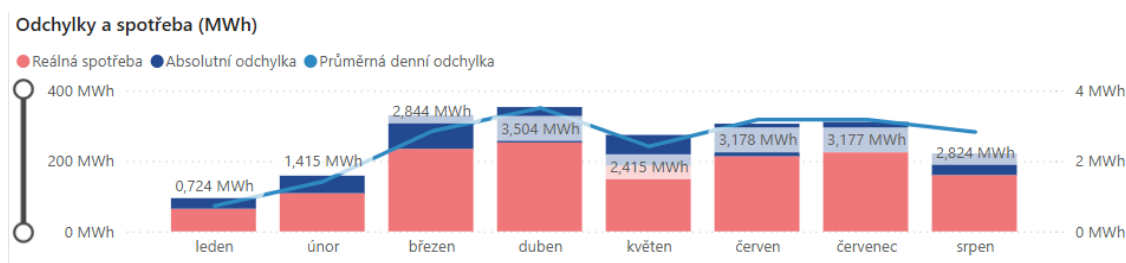
absolutních hodnotách, tak v relativním poměru k ostatním (v popisku dat je po i vyjádřená relativní spotřeba).



**Obrázek 21: Ukázka vizuálu Spotřeba (MWh)**

(Zdroj: Vlastní zpracování v Power BI)

**Odchytky a spotřeba (MWh)** je vizuál, který spojuje skládaný sloupcový a spojnicový graf, který je časově orientovaný a využívá hierarchické vnořování, díky kterému si uživatel může zvolit úroveň detailu – roky, kvartály nebo měsíce v roce. Sloupcová část grafu zobrazuje kumulativní reálnou spotřebu společně s odchylkami predikce, což umožňuje sledovat přesnost predikcí. I přesto, že predikovaná spotřeba není explicitně zobrazena jako samostatná hodnota, je z vizuálu logicky patrná díky vztahu mezi sloučenou odchylkou a reálnou spotřebou.



**Obrázek 22: Ukázka vizuálu Odchytky a spotřeba (MWh)**

(Zdroj: Vlastní zpracování v Power BI)

Spojnicová linie znázorňuje průměrnou denní odchylku, která slouží jako doplňující prvek pro sledování dlouhodobého vývoje predikční přesnosti v průběhu roku.

Ten je z hlediska času zaměřený (lze vnořováním zvolit roky, kvartály nebo měsíce v roce) na kumulativní spotřebu, odchylky a průměr denní odchylky. Složená odchylka a reálná spotřeba graficky zobrazí nepredikovanou spotřebu, přestože není explicitně zobrazena. Průměr denní odchylky je vhodným doplňkem pro zobrazení vývoje odchylek v roce.

**Matice s relevantními metrikami a KPIs je** poslední vizualizací na stránce sestavy, která slouží pro celkový přehled. Struktura matice je hierarchická – nejvyšší úroveň řádků je divize, kterou je možné rozbalit na úroveň jednotlivých závodů. Bezprostředně pod záhlavím je umístěn souhrnný řádek, který zobrazuje agregované hodnoty za celou společnost. Matice zahrnuje všechny v této práci definované metriky, včetně absolutních a relativních odchylek, MAE a podílu predikcí v toleranci. Tento typ vizualizace poskytuje nejen detailní pohled, ale i možnost srovnání napříč organizačními jednotkami v jednotném rozhraní.

### 3.2.4.3 Potenciální výkony

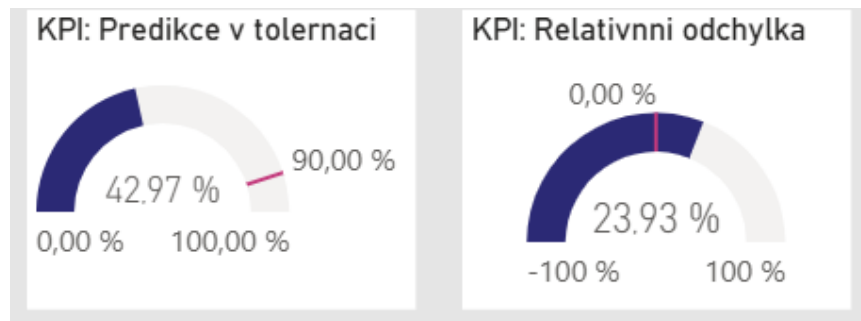
Třetí stránka sestavy s názvem Potenciální výkony (Příloha 12) je zaměřena na detailní hodnocení přesnosti predikcí na úrovni závodů, provozů či jednotlivých spotřebičů. Její hlavní přínos spočívá v možnosti analyzovat přesnost predikcí a následně řešit nepřesnosti.

Vzhledem k tomu, že organizační struktura divizí se liší (některé zahrnují provozy, jiné nikoli), je zde využít „záložkový průřez“ – interaktivní ovládací prvek, který pomocí vrstvení vizuálů zobrazuje pro každou divizi vlastní sadu odpovídajících vizualizací. Tento přístup umožňuje přizpůsobit uživatelské rozhraní realitě datové struktury bez nutnosti přepínání mezi samostatnými stránkami.

Stránka obsahuje celkem osm vizualizací, které se vzájemně doplňují a poskytují jak souhrnný, tak detailní pohled:

- Dvojice **ukazatelů KPIs** (WTR, relativní odchylka) – měřidlo
- **Ostatní metriky a KPIs** – karty (nové)
- **Vývoj predikce a spotřeby v hodinách (MWh)** – spojnicový graf
- **Predikce a spotřeba (MWh)** – pruhový graf
- **Odchylky v hodinách (MWh)**
- **Vývoj absolutní odchylky (MWh)**

**Dvojice ukazatelů KPIs (WRT a relativní odchylka)** jsou zobrazeny ve formě měřidel, která vizuálně odlišují hodnotu metriky od požadovaného stavu. Slouží k rychlé orientaci, zda daný závod či spotřebič dosahuje požadované přesnosti predikce. Každý ukazatel reprezentuje jiný aspekt predikční kvality – přesnost a poměrové srovnání.



Obrázek 23: Ukázka vizuálů KPI Predikce v toleranci a Relativní odchylka

(Zdroj: Vlastní zpracování v Power BI)

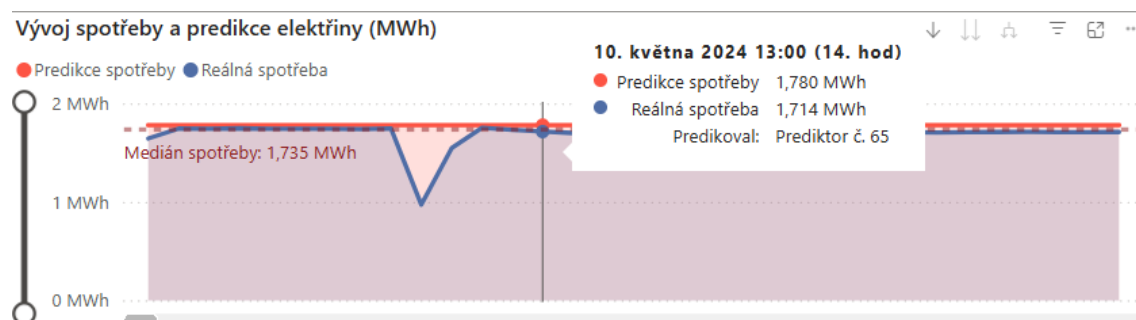
**Ostatní metriky a KPIs** jsou zobrazeny v sadě karet s hodnotami všech klíčových ukazatelů včetně MAE, absolutní odchylky nebo průměrné reálné spotřeby. Poskytují kompaktní souhrn pro každý výběr a doplňují měřidla o přesnější číselné informace a další kontext.

Reálná spotřeba 14,147 tis.MWh	Predikce spotřeby 14,909 tis.MWh	Absolutní odchylka 762,062 MWh
Medián spotřeby 1,058 MWh	MAE (s penalizací) 0,54	MAE 0,255 MWh

Obrázek 24: Ukázka vizuálu Karta metrik a KPIs

(Zdroj: Vlastní zpracování v Power BI)

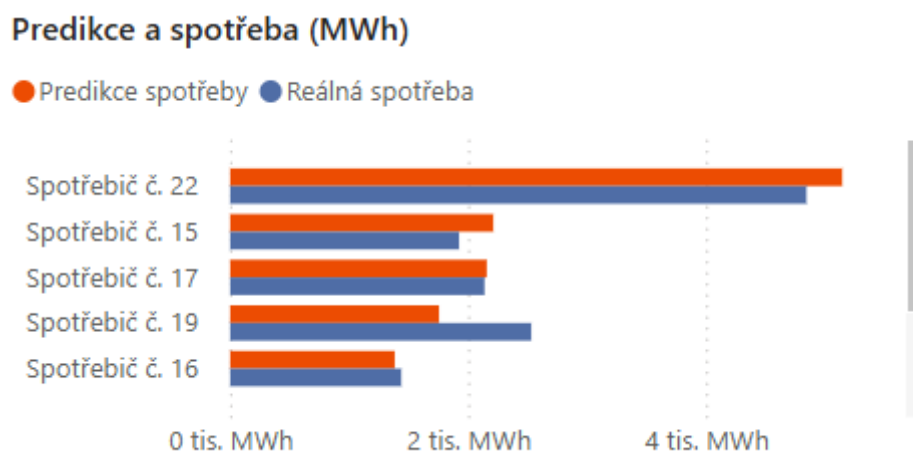
**Vývoj predikce a spotřeby v hodinách (MWh)** je spojnicový graf, který znázorňuje časový průběh predikované a skutečné spotřeby na úrovni hodin. Tento vizuál slouží k detailnímu porovnávání predikce a reálné spotřeby a pomáhá tak identifikovat časové úseky s výraznou odchylkou. Graf je doplněn referenční linkou, která znázorňuje medián spotřeby bez nuly<sup>11</sup>. Dále je v popisících dat (body na křivce) doplněn údaj, kdo v daný časový interval predikoval plánovanou spotřebu elektřiny.



**Obrázek 25: Ukázka vizuálu Vývoj spotřeby a predikce v hodinách (MWh)**

(Zdroj: Vlastní zpracování v Power BI)

**Predikce a spotřeba (MWh)** je skupinový pruhový graf, který zobrazuje porovnání sumy predikce a reálné spotřeby především pro jednoduchou orientaci o objemu spotřeby a přesnosti a také slouží jako filtrační prvek, kdy při kliknutí se zobrazí na ostatních vizuálech vyfiltrují pouze daný spotřebič.



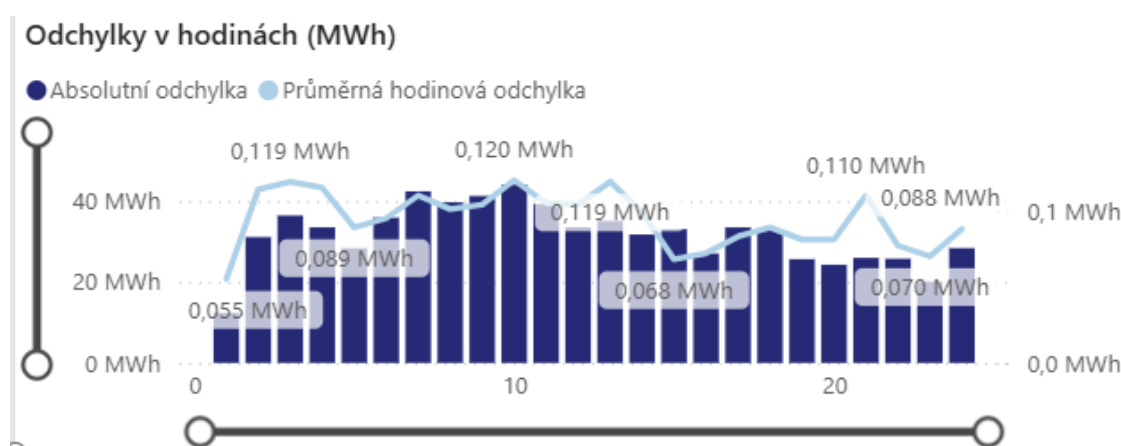
**Obrázek 26: Ukázka vizuálu Predikce a spotřeba (MWh)**

(Zdroj: Vlastní zpracování v Power BI)

<sup>11</sup> Toto řešení je nutné především kvůli zobrazení na úrovni spotřebičů, kdy nulová hodnota spotřeby znamená, že spotřebič je vypnutý a reálně tak spotřeba neprobíhá.

**Odchylky v hodinách (MWh)** je vizuál, který spojuje skládaný sloupcový a spojnicový graf pro zobrazení vývoje predikčních odchylek v průběhu dne. Na vodorovné ose jsou vyznačeny jednotlivé hodiny dne (1–24), přičemž každý sloupec znázorňuje součet absolutních odchylek mezi predikovanou a skutečnou spotřebou pro danou hodinu v rámci vybraného časového období.

Součástí vizuálu je také spojnicová křivka, která reprezentuje průměrnou hodinovou odchylku, čímž doplňuje kumulativní pohled o srovnatelné relativní zhodnocení přesnosti v jednotlivých hodinách. Tato kombinace umožňuje snadno identifikovat konkrétní hodiny dne, ve kterých dochází k systematickému selhávání modelu.



**Obrázek 27: Ukázka vizuálu Odchylky v hodinách (MWh)**

(Zdroj: Vlastní zpracování v Power BI)

**Vývoj absolutní odchylky (MWh)** je poslední vizualizací této stránky sestavy, která znázorňuje časový vývoj absolutní hodnoty rozdílu mezi predikovanou a skutečnou spotřebou. Každý bod na křivce odpovídá jedné časové jednotce (např. den nebo hodina) a reprezentuje velikost odchylky bez ohledu na její směr.

Součástí vizualizace je také referenční čára, která znázorňuje průměrnou absolutní odchylku za zvolené období. Tato linka poskytuje uživateli kontext pro posouzení, zda jsou aktuální hodnoty nadprůměrné, či naopak odpovídají očekávané přesnosti.

### 3.2.4.4 Hodnocení prediktorů

Čtvrtá stránka sestavy (příloha 13) je zaměřena na hodnocení přesnosti predikcí z pohledu jednotlivých uživatelů, kteří jsou v systému za tvorbu predikcí zodpovědní. Umožňuje porovnávat výkonnost jednotlivých prediktorů.

Cílem této stránky je nabídnout transparentní nástroj pro interní vyhodnocování toho, jak kvalitně jednotliví uživatelé (prediktoři) predikují spotřebu. Na základě dostupných metrik (např. MAE, relativní odchylka, WTR) lze vyhodnotit, kdo z prediktorů dosahuje nejlepší přesnosti, a kde je naopak prostor pro zlepšení.

Stránka obsahuje 3 vizualizace:

- **HeatMap MAE prediktorů (hodiny, dny)** – matice
- **Přehled KPIs** – matice
- **Predikce dle spotřebičů, závodů a prediktorů (MWh)** – rozkladový strom

**HeatMap MAE prediktorů (hodiny, dny)** je vizuál matice s podmíněným formátováním, která zobrazuje hodnoty střední absolutní chyby (MAE) v jednotlivých časových jednotkách – konkrétně po hodinách v rámci dne nebo po jednotlivých dnech. Úroveň časového detailu lze měnit pomocí vnořené hierarchie.

Barevné rozlišení buněk (tzv. heatmapa) reprezentuje intenzitu odchylky:

světle zelená barva značí nízké hodnoty MAE, tedy mírné odchylky mezi predikcí a skutečností, sytě červená barva signalizuje naopak výrazné chyby v predikci.

Tento efekt je dosažen pomocí barevné škály v podmíněném formátování, která se dynamicky přizpůsobuje hodnotám v daném výřezu dat. Nejnížší a nejvyšší hodnoty jsou automaticky zvýrazněny na základě relativní pozice v rámci aktuálního výběru.

Heatmapa slouží jako velmi intuitivní a vizuálně přehledný nástroj pro rychlou identifikaci časových vzorců, ve kterých mají prediktory tendenci chybovat – např. systematicky vyšší odchylky v určitých hodinách nebo dnech týdne.

Prediktor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Prediktor č. 1	0,10	0,10	0,09	0,09	0,09	0,10	0,11	0,16	0,18	0,21	0,34	0,39	0,34	0,17	0,15	0,13	0,13	0,09	0,13	0,13	0,11	0,09	0,10	0,10
Prediktor č. 11	0,61	0,56	0,56	0,04	0,05	0,58	0,62	0,63	0,89	0,76	0,75	0,79	0,91	1,15	1,25	0,70	0,91	1,11	0,63	0,62	0,65	0,66	0,79	1,04
Prediktor č. 13	0,00	0,00	0,00	0,00	0,00	0,10	0,10	0,10	0,10	0,10	0,10	0,05	0,05	0,10	0,10	0,10	0,10	0,10	0,10	0,10	0,10	0,00	0,00	0,00
Prediktor č. 14	0,00	0,00	0,00	0,00	0,00	0,01	0,08	0,06	0,06	0,06	0,14	0,19	0,23	0,13	0,04	0,03	0,03	0,03	0,02	0,03	0,03	0,03	0,00	0,00
Prediktor č. 60	0,57	0,60	0,67	0,74	0,56	0,54	0,57	0,60	0,51	0,58	0,49	0,45	0,70	0,74	0,68	0,54	0,69	0,69	0,46	0,36	0,37	0,58	0,53	0,56
Prediktor č. 61	0,10	0,09	0,10	0,11	0,11	0,10	0,14	0,16	0,15	0,16	0,14	0,16	0,16	0,13	0,11	0,11	0,12	0,11	0,10	0,11	0,12	0,11	0,10	0,13
Prediktor č. 63	0,06	0,05	0,05	0,05	0,05	0,04	0,10	0,10	0,11	0,11	0,09	0,11	0,10	0,08	0,08	0,09	0,07	0,06	0,07	0,08	0,07	0,06	0,05	0,05
Prediktor č. 65	0,72	0,71	0,67	0,62	0,57	0,66	0,75	0,73	0,73	0,73	1,29	1,39	1,03	0,84	0,91	0,98	0,93	0,81	0,81	0,73	0,64	0,66	0,68	0,76
Prediktor č. 67	0,16	0,16	0,17	0,16	0,17	0,15	0,19	0,25	0,26	0,25	0,22	0,25	0,26	0,19	0,18	0,18	0,16	0,13	0,15	0,15	0,15	0,12	0,12	0,15
Výchozí nastavení spotřebiče	1,20	1,21	1,26	1,24	1,23	1,38	1,61	2,05	2,13	2,09	4,17	4,64	3,88	1,65	1,61	1,57	1,58	1,57	1,57	1,59	1,59	1,51	1,10	1,10

Obrázek 28: Ukázka vizuálu Heatmapy MAE

(Zdroj: Vlastní zpracování v Power BI)

**Přehled KPIs** je vizuál matice, který podobně jako v stránce energetického přehledu zahrnuje v této práci definované metriky relativních odchylek, MAE a podílu predikcí v toleranci. Metriky jsou zobrazovány v členění podle uživatelů (prediktorů). Matice umožňuje rychlé srovnání výkonnosti jednotlivých prediktorů na základě různých aspektů predikční kvality.

**Predikce dle spotřebičů, závodů a prediktorů (MWh)** je vizualizací rozkladového stromu, která zobrazuje podíly predikcí z různých dimenzí – konkrétně podle spotřebičů, závodů a prediktorů. Velkou výhodou této vizualizace je její modularita a interaktivita. Uživatel si může dynamicky zvolit pořadí dimenzí rozkladu podle vlastních preferencí nebo aktuálních analytických potřeb. Díky tomu lze jednoduše zkoumat, které spotřebiče nebo organizační jednotky přispívají nejvíce k celkové predikci a tím i spotřebě, a kteří prediktoři vytvářejí nejvyšší podíl predikcí apod. Vizuál také může fungovat jako filtr pro další vizuály.

### 3.2.4.5 Odchyly

Poslední stránka reportní sestavy (Příloha 14) je zaměřena na analýzu odchylek mezi predikovanou a skutečnou spotřebou elektrické energie a na s nimi související klíčové ukazatele výkonnosti (KPIs). Konkrétně se jedná o metriky jako jsou střední absolutní chyba (MAE), které byly podrobněji popsány v předchozích částech práce.

Vizualizace na této stránce sledují časový vývoj odchylek v členění podle spotřebičů, provozů nebo závodů. Důraz je kladen na identifikaci opakujících se vzorců nebo výrazných výkyvů v přesnosti predikce, které mohou signalizovat problémy v modelu, datovém vstupu nebo provozním chování zařízení.

Tato stránka slouží jako důležitý podklad pro hlubší diagnostiku predikčního výkonu a umožňuje cílené zásahy v oblastech, kde dochází k největším odchylkám. Stránka obsahuje následující vizuály:

- **Pás karet s metrikami a KPIs**
- **Heatmapa dle provozů (hodiny, dny) – matice**
- **Predikce a spotřeba (MWh) – skládaný pruhový graf**
- **MAE (MWh) – spojnicový graf**

**Pás karet s metrikami a KPIs** je prostý vizuál karet, které, podobně jako u ostatních stánek dávají jasný přehled o metrikách jako spotřeba predikce, odchylky, predikce v toleranci.

**Heatmapa dle provozů (hodiny, dny)** je v podstatě stejným vizuálem jako u čtvrté stánky sestavy *Hodnocení prediktorů*, tedy zobrazuje průměrnou absolutní odchylku (MAE) v barevné škále s tím rozdílem, že není zaměřený na prediktory, ale na závody, provozy a spotřebiče. Matice umožňuje hierarchické rozložení zobrazovaných dat. Nejvyšší úrovní zobrazení jsou jednotlivé závody, které je možné rozbalit na úroveň provozů a následně až na konkrétní spotřebiče. Díky tomuto vnořenému zobrazení má uživatel možnost přirozeně procházet organizační strukturou a sledovat přesnost predikce ve stále větším detailu. Tato funkcionalita významně usnadňuje identifikaci konkrétních míst s největšími odchylkami a zároveň umožňuje porovnávat výkonnost jednotlivých úrovní v rámci celé struktury.

**Predikce a spotřeba (MWh)** je stejný skládaný pruhový graf jako na stránce *Potenciální výkony*, který slouží především pro základní rozlišení velikosti spotřeby a predikce k jednotlivým spotřebičům a také jako možný filtr na jednotlivé spotřebiče.

### **3.3 Zhodnocení návrhové části**

#### **3.3.1 Přínosy návrhové části**

Navržené řešení přináší společnosti řadu přínosů jak v oblasti nákupního procesu, tak i v rovině provozního řízení. Zásadním přínosem je vytvoření centrálního reportního systému, který umožňuje efektivně sledovat a vyhodnocovat spotřebu, predikci a odchylky elektrické energie na různých úrovních organizační struktury, od celopodnikové úrovně až po jednotlivé spotřebiče.

Díky zavedení přehledného datového modelu a sjednocení vstupních dat do podoby hvězdnicového schématu se významně zvyšuje transparentnost a srozumitelnost informací, které jsou prezentovány ve vizualizacích Power BI. Uživatelé tak získávají jednotný nástroj pro sledování spotřeby, porovnávání predikcí a hodnocení výkonnosti v různých částech podniku.

Dalším významným přínosem je zpřístupnění dat širokému spektru uživatelů – od energetiků, nákupčích přes výrobní plánovače až po střední management. Díky interaktivnímu uživatelskému rozhraní je možné přizpůsobit pohled na data konkrétním potřebám jednotlivých uživatelů a podporovat tak jejich rozhodování.

Zavedení metrik a klíčových ukazatelů výkonu (KPIs) umožňuje objektivní vyhodnocování přesnosti predikcí a tím i cílené zlepšování plánovacích procesů. Současně s tím vzniká prostor pro odhalování anomálií a včasnou identifikaci problémových oblastí (např. špatně nastavené predikce, nevhodně nastavené náběhy spotřebičů, nepečlivost prediktorů apod.).

### **3.3.2 Ekonomické zhodnocení**

V době vyhotovení této práce nebyl reportní systém ani jeho jednotlivé komponenty dosud nasazen do produkčního prostředí. Z tohoto důvodu není možné provést empirické zhodnocení přínosu návrhu na základě reálného provozu a sledovaných zlepšení v oblasti přesnosti predikcí nebo finančních úspor.

Pro účely ekonomického zhodnocení se proto vychází z historických dat o odchylkách mezi predikovanou a skutečnou spotřebou elektřiny a z obchodních dat z krátkodobého trhu s elektřinou (viz kapitola 2). Na základě těchto údajů byly sestaveny scénáře možného zlepšení přesnosti predikce, které vycházejí z reálných odchylek v minulosti a předpokladu, že implementace systému umožní jejich snížení.

Pro vyhodnocení dopadu na náklady spojené s odchylkami byly definovány tři základní scénáře zlepšení:

- Mírně pesimistický scénář: zlepšení přesnosti o 10 %
- Mírně optimistický scénář: zlepšení o 25 %
- Velmi optimistický scénář: zlepšení o 50 %

Každý ze scénářů předpokládá, že zpřesněním predikce dojde ke snížení objemu energie nakupované na denním a vnitrodenním trhu a omezení sankcí, ale tím i odměn za odchylky, resp. protiodchylky.

Výpočet ekonomického přínosu je proveden přímo na zdrojových průběhových datech (viz kapitola 2), kde je v rámci simulace aplikována úprava jednotlivých záznamů podle definovaných scénářů zpřesnění predikce. U každého časového záznamu je upravena hodnota odchylky – a to tak, že se její velikost redukuje o příslušné procento (10 %, 25 %, 50 %), přičemž je zachován její původní směr (kladná/záporná) a objem reálně spotřebované elektřiny.

Takto upravená data následně slouží pro zpětný výpočet nových nákladů na odchylky a obchodovanou elektřinu. Každý scénář tak reflektuje jinou míru predikční přesnosti a umožňuje kvantifikovat její dopad na výsledné náklady.

**Tabulka 10: Scénáře zlepšení predikcí**

Scénář	Cena za obchod	Úspora na MWh (obchod)	Cena za odchylku	Ztráta na MWh (odchylka)	Cena celkem	Úspora na MWh (celkem)
<b>Výchozí</b>	2 559,44 Kč		57,80 Kč		2 501,64 Kč	
<b>10%</b>	2 536,57 Kč	22,87 Kč	52,02 Kč	5,78 Kč	2 484,55 Kč	17,09 Kč
<b>25%</b>	2 502,27 Kč	57,17 Kč	43,35 Kč	14,45 Kč	2 458,92 Kč	42,72 Kč
<b>50%</b>	2 445,10 Kč	114,33 Kč	28,90 Kč	28,90 Kč	2 416,21 Kč	85,44 Kč

(Zdroj: Vlastní zpracování)

Z výsledků ekonomické simulace, která pracuje s upravenými daty podle různých scénářů zpřesnění predikce, je patrné, že celkové náklady na elektřinu (tj. náklady na nákup plus náklady na odchylky) klesají se zlepšující se přesností predikce.

Ve výchozím stavu činí průměrná celková cena na jednu MWh 2 501,64 Kč. Již při mírném zpřesnění predikce o 10 % dochází ke snížení celkové ceny na 2 484,55 Kč/MWh, což odpovídá úspoře 17,09 Kč/MWh. S rostoucí přesností se úspora výrazně zvyšuje – při 25% zlepšení je úspora již 42,72 Kč/MWh a při 50% zlepšení činí 85,44 Kč/MWh.

Z analýzy jednotlivých složek nákladů je zřejmé, že největší úspora vzniká na straně nákupů elektřiny – tedy snížením potřeby nadbytečného obchodovaného množství. Např. ve scénáři 50% zpřesnění došlo k úspoře 114,33 Kč/MWh pouze v nákladech na nákup. Sice došlo k částečné ztrátě na straně odchylek, konkrétně snížením výnosů z protiodchylek. Tyto ztráty však rostou pomaleji (např. 28,90 Kč/MWh při 50% zlepšení) a nepřevyšují dosaženou úsporu.

Závěrem lze říci, že každé procento zlepšení predikční přesnosti přináší přímou ekonomickou úsporu, přičemž efekt je v simulovaných scénářích nelineární – čím větší zpřesnění, tím výraznější úspora.

Tato bilance potvrzuje, že pozitivní dopad zpřesnění predikce převyšuje ztrátu z ušlých příležitostí. V praxi to znamená, že i když podnik přichází o část výnosů z protiodchylek, snížení nákladů na nepřesně obchodovanou elektřinu a penalizace za odchylky je výrazně převažuje.

### **3.3.3 Omezení návrhové části**

Při hodnocení navrženého řešení je nezbytné zohlednit několik důležitých skutečností a omezení, která vyplývají ze zvoleného rozsahu této práce. V rámci návrhu není pokryto kompletní časové okno celého životního cyklu systému, tedy fáze návrhu, realizace, implementace a finálního zhodnocení úspěšnosti v reálném provozu. Práce se soustředí zejména na návrh architektury modelu a sestavy v prostředí Power BI, bez přímé návaznosti na dlouhodobý provozní monitoring nebo ekonomické vyhodnocení v čase.

Součástí návrhu není ani detailní analýza kvality a úplnosti zdrojových dat, která může výrazně ovlivnit výslednou přesnost a spolehlivost predikcí. Tato oblast bude muset být řešena v dalších fázích implementace a je nutné zavést kontinuální datovou kontrolu, která bude sledovat úplnost, konzistenci a aktuálnost datových vstupů.

Další potenciální komplikací, která přesahuje rámec této práce, je možné chybné nebo zastaralé přiřazení elektroměrů ke konkrétním spotřebičům. Je proto doporučeno, aby při implementaci systému došlo k odbornému ověření správnosti těchto vazeb přímo v jednotlivých závodech, ideálně ve spolupráci s pracovníky zodpovědnými za energetický management a vytvořit tak aktuální a co nejvíce reprezentativní určení vazeb mezi elektroměry a spotřebiči.

Současně není v práci podrobně řešena kvalita historických dat – zejména jejich úplnost, konzistence či možnost výskytu chyb při přiřazení měřicích bodů ke spotřebičům. Tyto aspekty budou muset být v reálném nasazení sledovány a ideálně i validovány odborníky přímo ve výrobních závodech.

Omezení lze spatřovat také v oblasti uživatelského rozhraní a návrhu reportních stránek a zvolených KPIs. Sestava Power BI byla navržena na základě obecných požadavků, bez hlubší analýzy konkrétních potřeb různých skupin uživatelů. Nebyla provedena validace pomocí reálné uživatelské zpětné vazby, a proto může být nutné rozhraní dále upravovat podle způsobu používání v praxi a preferencí uživatelů.

## ZÁVĚR

Hlavním cílem práce bylo navrhnout efektivní reportní systém, který umožní kontrolu, analýzu a vizualizaci dat spojených s predikcí a skutečnou spotřebou elektrické energie, který byl rozdělen do pěti dílčích cílů. Tento cíl byl naplněn prostřednictvím ucelené analýzy současného stavu a následného návrhu řešení, které kombinuje datové modelování a uživatelsky přívětivý reportingový nástroj v prostředí Power BI.

V první fázi byl zhodnocen současný predikční systém a jeho reportingové kapacity, prostřednictvím detailní analýzy rozhraní systému i používaných datových struktur. Zjištění ukázala, že stávající reporting je funkčně omezený, postrádá granularitu, možnosti exportu a nereflektuje reálné obchodní intervaly, což omezuje jeho využití v praxi. Výstupy jsou málo přehledné, některé hodnoty chybné a pracovníci proto upřednostňují externí datové zdroje a ruční zpracování dat v Excelu, které je časově náročné a náchylné k chybám. Data jsou navíc roztržštěná ve více tabulkách a jejich propojení vyžaduje pokročilý ETL proces. Tato omezení jasně ukazují potřebu komplexnějšího, uživatelsky přívětivého a automatizovaného reportního řešení.

Druhý dílčí cíl zaměřený na analýzu obchodní činnosti na krátkodobých trzích s elektřinou byl naplněn vyhodnocením odchylek mezi predikovanou a skutečnou spotřebou. Bylo zjištěno, že kladné odchylky, tedy nadhodnocení spotřeby, převažují nad zápornými, což naznačuje systematickou tendenci predikčního systému přeceňovat spotřebu. Přestože tyto kladné odchylky mohou vést k výnosům z protiodchylek, finančně stále vznikají penalizace za odchylky. Statistické ukazatele přesnosti (MAE, RMSE, Bias) signalizují zlepšení v čase, ale průměrná odchylka stále představuje značný podíl objemu obchodů (cca 16 %). Spearmanova korelační analýza navíc potvrdila středně silný vztah mezi velikostí odchylek a jejich cenou, zejména u větších hodnot, zatímco menší odchylky nesou spíše náhodný charakter. Tato zjištění ukazují, že přesnost predikcí má vliv na ekonomické výsledky a současně podtrhují potřebu zpřesnění analytických nástrojů pro efektivní řízení spotřeby.

Třetím dílčím cílem bylo navrhnout konceptuální logický datový model včetně popisu integračních a transformačních požadavků. Tento cíl byl naplněn vytvořením datového modelu ve formě hvězdicového schématu, který sjednocuje historická měření z různých systémů a umožňuje efektivní realizaci ETL procesů. Navržený model tvoří základ pro

vznik stabilního, škálovatelného a optimalizovaného datového tržiště, které integruje klíčová data o predikci a spotřebě napříč organizací. Díky standardizované struktuře je tržiště snadno přístupné jak pokročilými analytickými nástroji (např. Power BI), tak běžnými prostředky jako je Microsoft Excel. Automatizovaná denní aktualizace dat navíc zajišťuje aktuálnost reportů a podporuje včasné rozhodování. Model je připraven i na budoucí rozšiřování o další datové domény, včetně finančního pohledu.

Následně byl splněn čtvrtý dílčí cíl, kterým bylo navržení struktury a obsahu reportního nástroje. Výsledkem je sestava reportů v prostředí Power BI, která poskytuje uživatelsky přívětivé rozhraní pro sledování predikcí, spotřeby a jejich odchylek na různých úrovních – od jednotlivých spotřebičů po podnikový celek. Součástí řešení je sada klíčových ukazatelů výkonnosti (KPIs) a vizuálů, které umožňují sledování přesnosti predikcí v čase, a také nástroje pro detekci anomálií a identifikaci problematických oblastí.

Pátý dílčí cíl, tedy vyhodnocení ekonomického přínosu zpřesnění predikcí, byl naplněn sestavením tří simulačních scénářů (10 %, 25 % a 50% zlepšení). Výsledky simulací ukázaly, že každé zlepšení predikční přesnosti vede k přímé ekonomické úspoře. Například při 50% zlepšení dochází k celkové úspoře 85,44 Kč/MWh, zejména díky omezení potřeby nadbytečných nákupů elektřiny. Ztráty vzniklé snížením výnosů z protiodchylek nebyly vysoké a ani v jednom scénáři nepřevýšily úsporu ze zpřesnění predikcí.

Lze konstatovat, že hlavní cíl a dílčí cíle byly úspěšně naplněny. Analýza současného stavu odhalila významná omezení stávajícího systému, zejména v oblasti granularity, automatizace a přesnosti výstupů. Následná analýza odchylek potvrdila, že nadhodnocení spotřeby je častější a ekonomicky nevýhodné, čímž byla zdůrazněna nutnost zpřesnění predikčního procesu. Navržený datový model ve formě hvězdicového schématu umožňuje centralizované zpracování dat pomocí efektivního procesu ETL a vytváří základ pro škálovatelné datové tržiště. Výsledný reportní nástroj v prostředí Power BI poskytuje přehledné, interaktivní a uživatelsky přívětivé rozhraní pro sledování spotřeby, predikcí i odchylek a je doplněn o klíčové ukazatele výkonnosti. Ekonomická simulace navíc prokázala, že každé zpřesnění predikce přináší měřitelné úspory. Celkově tedy práce přináší praktický návrh, který může významně přispět ke zlepšení predikční přesnosti a ekonomické efektivity řízení spotřeby elektřiny.

Je nezbytné poukázat na určitá omezení této práce, která je nutné vzít v úvahu při interpretaci výsledků. Analýza obchodní činnosti společnosti vychází výhradně z historických dat, přičemž nereflektuje konkrétní tržní události, jako jsou mimořádné výkyvy nebo sezónní vlivy, které mohly v daném období ovlivnit chování na trhu s elektřinou. Podobně i zjištěná korelace mezi velikostí odchylek a jejich finančním dopadem nemusí mít přímý kauzální charakter, protože analýza nezahrnuje další potenciálně ovlivňující faktory, např. lidský faktor, změny v provozním řízení nebo vývoj cenových křivek. Tato omezení je nutné promítnout i do simulačních scénářů ekonomických přínosů, které vycházejí z předpokladu zlepšení přesnosti predikcí, ale nezohledňují reálné provozní prostředí, jelikož navržený reportní systém nebyl v době zpracování práce nasazen do produkčního provozu. Dále je třeba zmínit, že návrh struktury a obsahu reportního nástroje vycházel primárně z obecných požadavků pracovníků zodpovědných za nákup elektřiny a reflektuje především současné potřeby zjištěné v rámci predikčního systému. Je pravděpodobné, že při reálném nasazení bude nutné reporty dále přizpůsobit specifickým požadavkům dalších uživatelských skupin, zejména technických pracovníků či energetických specialistů, jejichž potřeby se mohou významně lišit.

## SEZNAM POUŽITÝCH ZROJŮ

ALEXE, Cătălina Monica; SIMON, Cristina Petronela a ALEXE, Cătălin George, 2014. *Business intelligence – past, present and future*. Online. Network Intelligence Studies, roč. 2, č. 1, s. 7–14. ISSN-L 2344-1712. Dostupné z: IDEAS/RePEc, <https://ideas.repec.org/a/cmj/networ/y2014i3p7-14.html>, [citováno 2025-05-09].

ČEPS, n.d. *O nás*. Online. ČEPS, a.s. Dostupné z: <https://www.ceps.cz/cs/o-nas>. [citováno. 2025-05-09].

Databricks, [n.d.]a. *Star Schema*. Online. In: Databricks Inc. Dostupné z: <https://www.databricks.com/glossary/star-schema>. [citováno 2025-05-09].

Databricks, [n.d.]b. *Snowflake Schema*. Online. In: Databricks Inc. Dostupné z: <https://www.databricks.com/glossary/snowflake-schema>. [citováno 2025-05-09].

DECKLER, Greg a POWELL, Brett, 2022. *Mastering Microsoft Power BI: expert techniques to create interactive insights for effective data analytics and business intelligence*. Second edition. Expert insight. Birmingham: Packt. ISBN 978-1-80181-148-4.

GeeksforGeeks, 2023. *Data warehouse development life cycle model*. Online. In: GeeksforGeeks, Sanchlaya Education Private Limited. 25.4.2023. Dostupné z: <https://www.geeksforgeeks.org/data-warehouse-development-life-cycle-model/>. [citováno 2025-05-09].

HENDL, Jan, 2012. *Přehled statistických metod: analýza a metaanalýza dat*. 4. rozš. vyd. Praha: Portál. ISBN 978-80-262-0200-4.

HENDL, Jan, 2022. *Základy matematiky, logiky a statistiky pro sociologii a ostatní společenské vědy v příkladech*. 3. doplněné vyd. Praha: Karolinum. ISBN 978-80-246-5400-3.

HOLČÍK, Jiří a KOMENDA, Martin 2015. *Matematická biologie: e-learningová učebnice*. Online. Brno: Masarykova univerzita. ISBN 978-80-210-8095-9. Dostupné z: <https://portal.matematickabiologie.cz/>. [citováno 2025-05-09].

CHMELÁR, Michal, 2018. *Reporting v Power BI, PowerPivot a jazyk DAX*. Pezinok: Smart People. ISBN 978-80-973078-0-6.

- JANÁČEK, Julius, 2022. *Statistika jednoduše: průvodce světem statistiky*. Praha: Grada Publishing. ISBN 978-80-271-1738-3.
- KOLEKTIV AUTORŮ, 2016. *Úvod do liberalizované energetiky: Trh s elektřinou*. 2. aktualiz. vyd. Praha: Asociace Energetických Manažerů. ISBN 978-80-260-9212-4.
- LABERGE, Robert, 2012. *Datové sklady: agilní metody a business intelligence*. Brno: Computer Press. ISBN 978-80-251-3729-1.
- LUTZ, Greg, 2018. *Working with OLAP Cubes*. Online. In: MESCIUS USA, Inc. 18. 12. 2018. Dostupné z: <https://developer.mescius.com/blogs/working-with-olap-cubes>. [citováno 2025-05-09].
- MICROSOFT, 2024a. *Co je Power BI?*. Online. In: Microsoft Learn, 22. 3. 2024 Dostupné z: <https://learn.microsoft.com/cs-cz/power-bi/fundamentals/power-bi-overview>. [citováno 2025-05-09].
- MICROSOFT, 2024b. *Specifikace jazyka M – Úvod*. Online. In: Microsoft Learn, 17. 12. 2024 Dostupné z: <https://learn.microsoft.com/cs-cz/powerquery-m/m-spec-introduction>. [citováno 2025-05-09].
- MICROSOFT, 2025a. *Co je Power Query?*. Online. In: Microsoft Learn, 29. 1. 2025 Dostupné z: <https://learn.microsoft.com/cs-cz/power-query/power-query-what-is-power-query>. [citováno 2025-05-09].
- MICROSOFT, 2025b. *Přehled jazyka DAX*. Online. In: Microsoft Learn, 9. 4. 2025 Dostupné z: <https://learn.microsoft.com/cs-cz/dax/dax-overview>. [citováno 2025-05-09].
- MOSCOSO-ZEA, Oswaldo; PAREDES-GUALTOR Joel a LUJÁN-MORA, Sergio, 2018. *A holistic view of data warehousing in education*. Online. IEEE Access, roč. 6, s. 64659–64673. ISSN 2169-3536. Dostupné z: IEEE Xplore, <https://doi.org/10.1109/ACCESS.2018.2876753> [citováno 2025-05-09].
- NEUBAUER, Jiří; SEDLÁČÍK, Marek a KŘÍŽ, Oldřich, 2021. *Základy statistiky: Aplikace v technických a ekonomických oborech*. 3. rozšíř. vyd. Praha: Grada Publishing. ISBN 978-80-271-3421-2.

OTE, 2024a. *Výroční zpráva 2023*. Online. PDF. OTE, a.s. Dostupné z: <https://www.ote-cr.cz/cs/o-spolecnosti/vyrocní-zpravy>. [citováno 2025-05-09].

OTE, 2024b. *Trh s elektřinou: Organizovaný krátkodobý trh*. Online. PDF. OTE, a.s. Dostupné z: <https://www.ote-cr.cz/cs/kratkodobe-trhy/elektrina/trh-s-elektrinou-2023.pdf>. [cit. 2025-05-09].

OTE, 2024c. *Roční zpráva o trhu s elektřinou a plynem v ČR 2023*. Online. PDF. OTE, a.s. Dostupné z: <https://www.ote-cr.cz/cs/o-spolecnosti/vyrocní-zpravy>. [citováno 2025-05-09].

PAVKOV, Sanja; POŠČIĆ, Patrizia a JAKŠIĆ, Danijel, 2016. *Business intelligence systems yesterday, today and tomorrow – an overview*. Online. Zbornik Veleučilišta u Rijeci, roč. 4, č. 1, s. 97–108. ISSN 1848-3526. Dostupné z: Portal of Croatian scientific and professional journals – HRČAK, <https://hrcak.srce.hr/160236>, [citováno 2025-05-09].

POUR, Jan; MARYŠKA, Miloš a NOVOTNÝ Ota, 2012. *Business intelligence v podnikové praxi*. Praha: Professional Publishing. ISBN 978-80-7431-065-2.

RANE, Nitin; PARAMESHA, Mallikarjuna; CHOUDHARY, Saurabh a RANE, Jayesh, 2024. *Business Intelligence and Business Analytics With Artificial Intelligence and Machine Learning: Trends, Techniques, and Opportunities* Online. Dostupné z: SSRN, <http://dx.doi.org/10.2139/ssrn.4831920>, [citováno 2025-05-09].

SAGIROGLU, Seref a SINANC, Duygu, 2013. *Big data: A review*. Online. In: 2013 International Conference on Collaboration Technologies and Systems (CTS). San Diego, 20-24 May 2013. IEEE, s. 42–47. ISBN 978-1-4799-0652-4. Dostupné z: <https://doi.org/10.1109/CTS.2013.6567202>, [citováno 2025-05-09].

VANDEPUT, Nicolas, 2019. *Forecast KPI: RMSE, MAE, MAPE & Bias*. Online. In: LinkedIn Corporation, 28. 6. 2019 Dostupné z: <https://www.linkedin.com/pulse/forecast-kpi-rmse-mae-mape-bias-nicolas-vandepu/>. [citováno 2025-05-09].

Vyhláška 408/2015 Sb., o pravidlech trhu s elektřinou, ve znění pozdějších předpisů.

Vyhláška č. 541/2005 Sb., o pravidlech trhu s elektřinou, ve znění pozdějších předpisů.

Zákon č. 458/2000 Sb., o podmínkách podnikání a výkonu státní správy v energetických odvětvích a o změně některých zákonů (energetický zákon), ve znění pozdějších předpisů.

## SEZNAM POUŽITÝCH ZKRATEK

.xlsx	Přípona souboru Microsoft
API	Application Programming Interface
BI	Business Intelligence
CEST	Středoevropský letní čas
CET	Středoevropský čas
CRM	Řízení vztahů se zákazníky (Customer Relationship Management)
ČEPS	Provozovatel přenosové soustavy – ČEPS, a.s.
ČR	Česká republika
<i>D</i>	Kolmogorovovův-Smirnovův test
DAX	Data Analysis Expressions
ERD	Realizační diagram
ERP	Plánování podnikových zdrojů (Enterprise Resource Planning)
ERÚ	Energetický regulační úřad
ETL	Extract-Transform-Load
EU	Evropská Unie
$F_n(x)$	Empirická distribuční funkce
$H_0$	Nulová hypotéza
$H_1$	Alternativní hypotéza
HOLAP	Hybridní OLAP
IQR	Interkvartilové rozpětí
IT	Informační technologie
Kč	Korun českých
KPI	Klíčový ukazatel výkonu (Key Performance Indicator)
MAE	Průměrná absolutní chyba
<i>Me</i>	Medián
MOLAP	Multidimenzionální OLAP
MS	Microsoft

MWh	megawatthodina
$n$	Počet prvků
ODBC	Open Database Connectivity
OLAP	Online Analytical Processing
OLTP	Online Transactional Processing
OTE	Operátor trhu s elektřinou – OTE, a.s.
$p$	p-hodnota
$q$	Kvantil
$Q_1$	Dolní kvartil
$Q_3$	Horní kvartil
RMSE	Odmocněná střední kvadratická chyba
RO	Relativní odchylka
ROLAP	Relační OLAP
$r_s$	Spearmanův korelační koeficient
SQL	Structured Query Language
$sup_x$	Supremum (největší rozdíl)
$s_{xr}$	Směrodatná odchylka pořadí proměnné X
$s_{yr}$	Směrodatná odchylka pořadí proměnné Y
UI	Uživatelské rozhraní (User Interface)
UTC	Koordinovaný světový čas
UX	Uživatelská zkušenost (User Experience)
VBA	Visual Basic for Applications
WTR	Podíl predikcí v toleranci (Within Tolerance Ratio)
$\bar{x}$	Aritmetický průměr
XML	Extensible Markup Language
$\bar{x}_r$	Průměrné pořadí všech hodnot proměnné X
$x_{ri}$	Pořadí i-té hodnoty proměnné X
$\bar{y}_r$	Průměrné pořadí všech hodnot proměnné Y
$y_{ri}$	Pořadí i-té hodnoty proměnné Y

$ZC$	Zúčtovací cena odchylky
$ZC_{\text{proti}}$	Zúčtovací cena protiodchylky
$\alpha$	Hladina významnosti
$F(x)$	Teoretická distribuční funkce

## SEZNAM POUŽITÝCH VZORCŮ

Aritmetický průměr (1).....	44
Medián lichý (2).....	45
Medián sudý (3).....	45
Interkvartilové rozpětí (4).....	46
Kolmogorovův-Smirnovův test (5).....	48
Spearmanův korelační koeficient (6).....	50
Průměrná absolutní chyba (7).....	51
Odmocněná střední kvadratická chyba (8).....	51
Bias (9).....	52
Relativní odchylka (10).....	89
Podíl predikcí v toleranci (11).....	90

## SEZNAM POUŽITÝCH OBRÁZKŮ

Obrázek 1: Big Data – 3V.....	16
Obrázek 2: Inmonův přístup top-down.....	19
Obrázek 3: Kimballův přístup bottom-up .....	20
Obrázek 4: Architektura datového skladu .....	22
Obrázek 5: Schéma hvězdy.....	26
Obrázek 6: Schéma sněhové vločky .....	27
Obrázek 7: OLAP kostka.....	28
Obrázek 8: Realizační diagram (ERD) .....	41
Obrázek 9: Vztah systémové odchylky a regulační elektřiny.....	43
Obrázek 10: Agregace predikcí .....	56
Obrázek 11: Datový tok faktů z produkční databáze do datového tržiště .....	73
Obrázek 12: Integrace faktů predikcí a reálné spotřeby elektřiny .....	77
Obrázek 13: Relační diagram integrace organizační dimenze.....	79
Obrázek 14: Architektura datového toku vlastního návrhu .....	80
Obrázek 15: Relační návrh datového tržiště .....	81
Obrázek 16: Návrh datového modelu pro uživatelskou aplikaci Power BI.....	82
Obrázek 17: Řešení uživatelského rozhraní přes funkce „Výběr“ a „Záložky“ v prostředí Power BI .....	87
Obrázek 18: Porovnání rozhraní s aktivním a neaktivním rozbalovacím panelem průřezů .....	88
Obrázek 19: Ukázka vizuálu Vývoj spotřeby a predikce elektřiny (MWh) .....	92
Obrázek 20: Ukázka vizuálu Vývoj odchylky (MWh).....	92
Obrázek 21: Ukázka vizuálu Spotřeba (MWh).....	93
Obrázek 22: Ukázka vizuálu Odchylky a spotřeba (MWh).....	93
Obrázek 23: Ukázka vizuálů KPI Predikce v toleranci a Relativní odchylka .....	95
Obrázek 24: Ukázka vizuálu Karta metrik a KPIs.....	95
Obrázek 25: Ukázka vizuálu Vývoj spotřeby a predikce v hodinách (MWh).....	96

Obrázek 26: Ukázka vizuálu Predikce a spotřeba (MWh) .....	96
Obrázek 27: Ukázka vizuálu Odchylky v hodinách (MWh) .....	97
Obrázek 28: Ukázka vizuálu Heatmapy MAE .....	99

## SEZNAM POUŽITÝCH GRAFŮ

Graf 1: Houslový graf – Obchod a odchylky (MWh).....	66
Graf 2: Houslový graf – Obchod a odchylky (CZK/MWh).....	67
Graf 3: Korelační diagram odchylky (MWh) a jednotkové ceny odchylky (CZK/MWh) .....	69

## SEZNAM POUŽITÝCH TABULEK

Tabulka 1: Přehled datových souborů .....	60
Tabulka 2: Výsledky Kolmogorovova-Smirnova testu normality a hypotéz .....	61
Tabulka 3: Popisná statistika souborů Denní trh - saldo (MWh) a Vnitrodenní trh – Saldo (MWh) .....	62
Tabulka 4: Popisná statistika souborů Denní trh – saldo (CZK/MWh) a Vnitrodenní trh – Saldo (CZK/MWh) .....	62
Tabulka 5: Popisná statistika souborů Obchod – saldo (MWh) a Obchod – saldo (CZK/MWh) .....	63
Tabulka 6: Popisná statistika Odchylka kladná (MWh), Odchylka záporná (MWh) a Odchylka celkem (MWh) .....	64
Tabulka 7: Popisná statistika Odchylka kladná (CZK/MWh), Odchylka záporná (CZK/MWh) a Odchylka celkem (CZK/MWh) .....	65
Tabulka 8: Metriky pro hodnocení přesnosti predikcí (MAE, RMSE a Bias) .....	68
Tabulka 9: Sjednocené identifikátory tabulky FactConsumption.....	74
Tabulka 10: Scénáře zlepšení predikcí .....	102

## SEZNAM PŘÍLOH

Příloha 1: Procesní graf – predikce a nákup elektřiny na denním trhu.....	I
Příloha 2: Histogramy rozdělení datových souborů (část I.) .....	II
Příloha 3: Histogramy rozdělení datových souborů (část II.).....	III
Příloha 4: Skript v programovacím jazyku Python pro výpočet Kolmogorovova-Smirnova testu normality .....	IV
Příloha 5: Skript v programovacím jazyku Python pro výpočet ukazatelů deskriptivní statistiky .....	V
Příloha 6: Skript v programovacím jazyku Python pro výpočet Spermanova korelačního koeficientu .....	VI
Příloha 7: Relační schéma databáze aplikace .....	VII
Příloha 8: Skrip pro výpočet KPIs v prostředí DAX .....	VIII
Příloha 9: Rozložení stránky sestavy .....	IX
Příloha 10: Úvodní strana reportního nástroje v Power BI .....	X
Příloha 11: Strana Energetický přehled reportního nástroje v Power BI.....	XI
Příloha 12: Strana Potenciální výkony reportního nástroje v Power BI.....	XII
Příloha 13: Strana Hodnocení prediktorů reportního nástroje v Power BI.....	XIII
Příloha 14: Strana Odchylky reportního nástroje v Power BI .....	XIV

# PŘÍLOHY

## Příloha 1: Procesní graf – predikce a nákup elektřiny na denním trhu

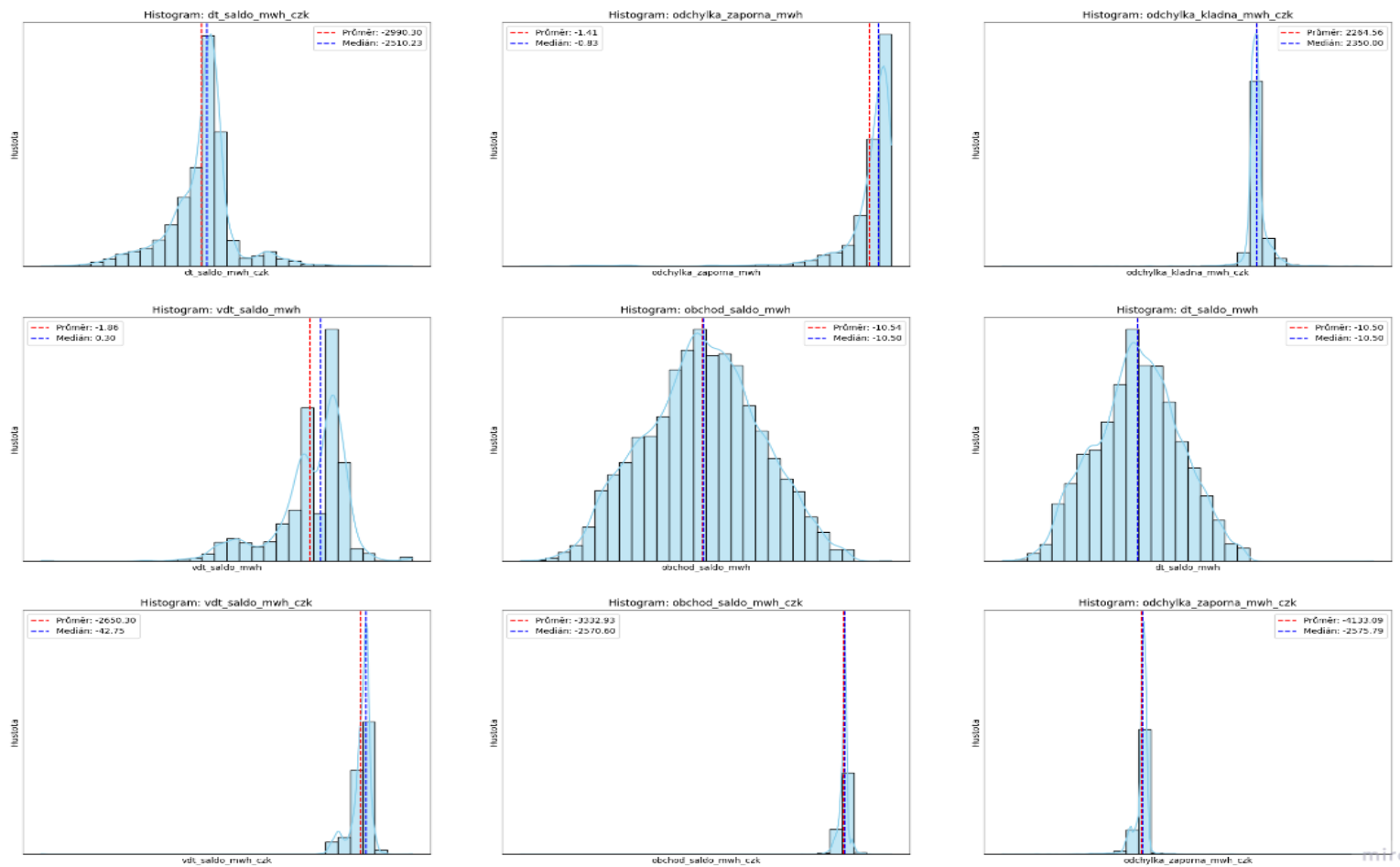
(Zdroj: Vlastní zpracování)



Kde:  
den D - den dodávky elektřiny  
den D-1 - den před dodávkou elektřiny<sub>ro</sub>  
den D-2 - 2 dny před dodávkou elektřiny

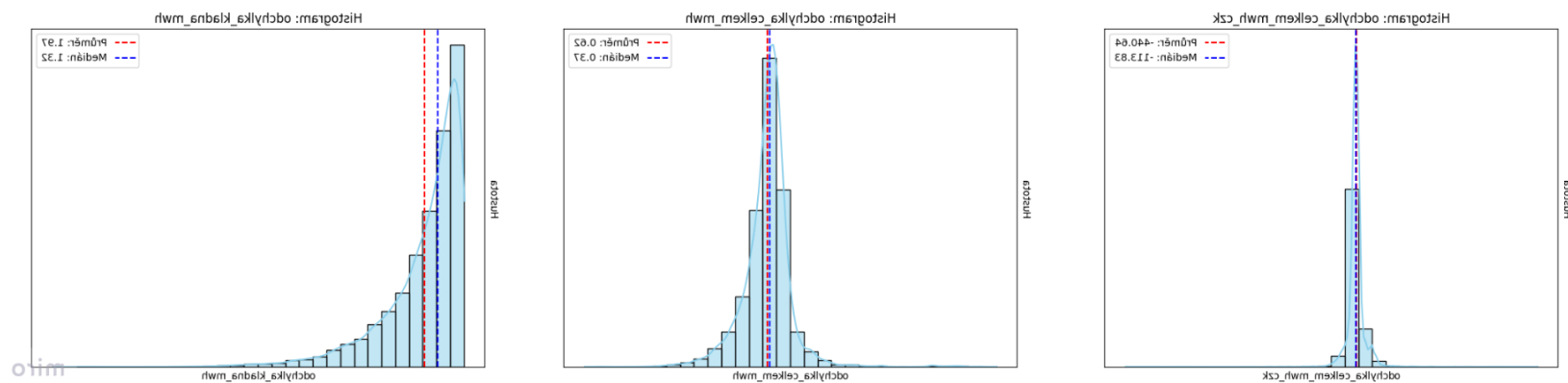
## Příloha 2: Histogramy rozdělení datových souborů (část I.)

(Zdroj: Vlastní zpracování podle interních dat)



### Příloha 3: Histogramy rozdělení datových souborů (část II.)

(Zdroj: Vlastní zpracování podle interních dat)



## Příloha 4: Skript v programovacím jazyku Python pro výpočet Kolmogorovova-Smirnovova testu normality

(Zdroj: Vlastní zpracování)

```
def kolmogorov_smirnov_pro_vsechny_sloupce(dataframe):  
  
    results = []  
    # Iterace přes všechny sloupce v DataFrame (od třetího dál)  
    for column in dataframe.columns[3:]:  
        # Filtr: odstranění NaN a hodnot 0  
        filtered_data = dataframe[column].dropna()  
        filtered_data = filtered_data[filtered_data != 0]  
        # Pokud je sloupec prázdný po filtraci, ignoruje  
        if filtered_data.empty:  
            print(f"Sloupec '{column}' je po filtraci prázdný, přeskočeno.")  
            continue  
        # Kolmogorov-Smirnov test  
        stat, p = kstest(filtered_data, 'norm')  
        # Výsledek hypotéz  
        if p > 0.05:  
            h0_result = "nulová hypotéza přijata"  
            h1_result = "alternativní hypotéza zamítnuta"  
        else:  
            h0_result = "nulová hypotéza zamítnuta"  
            h1_result = "alternativní hypotéza přijata"  
  
        # Přidání výsledků do seznamu  
        results.append({  
            "Sloupec": column,  
            "Stat": stat,  
            "p-value": p,  
            "H0": h0_result,  
            "H1": h1_result  
        })  
    # Převod seznamu výsledků do DataFrame  
    results_df = pd.DataFrame(results)  
  
    return results_df
```

## Příloha 5: Skript v programovacím jazyku Python pro výpočet ukazatelů deskriptivní statistiky

(Zdroj: Vlastní zpracování)

```
def descriptive_statistics(dataframe):

    # Přidání sloupce 'Year' na základě sloupce 'datetime'
    dataframe['Year'] = pd.to_datetime(dataframe['datetime']).dt.year
    # Příprava výsledného DataFramu
    results = []
    # Iterace přes sloupce DataFramu (od třetího dál)
    for column in dataframe.columns[3:]:
        for year in dataframe['Year'].unique().tolist() + ['Celkem']:
            # Filtrace podle roku nebo všechna data
            if year == 'Celkem':
                filtered_data = dataframe[column].dropna()
            else:
                filtered_data = dataframe[dataframe['Year'] == year][column].dropna()
            # Odstranění nulových hodnot
            if column in config.Config.exceptions:
                filtered_data = filtered_data
            else:
                filtered_data = filtered_data[filtered_data != 0]

            if filtered_data.empty:
                # Přidání prázdných statistik, pokud nejsou není
                stats = {
                    "Průměr": None,
                    "Medián": None,
                    "Modus": None,
                    "Směrodatná odchylka": None,
                    "Rozptyl": None,
                    "Interkvartilové rozpětí (IQR)": None,
                    "25. percentil": None,
                    "75. percentil": None,
                    "Min": None,
                    "Max": None
                }
            else:
                # Výpočet modusu pouze pokud existuje jeden
                modes = filtered_data.mode()
                modus = modes.iloc[0] if len(modes) == 1 else None

                # Výpočet statistik
                stats = {
                    "Průměr": filtered_data.mean(),
                    "Medián": filtered_data.median(),
                    "Modus": modus,
                    "Směrodatná odchylka": filtered_data.std(),
                    "Rozptyl": filtered_data.var(),
                    "Interkvartilové rozpětí (IQR)": filtered_data.quantile(0.75) -
                    filtered_data.quantile(0.25),
                    "25. percentil": filtered_data.quantile(0.25),
                    "75. percentil": filtered_data.quantile(0.75),
                    "Min": filtered_data.min(),
                    "Max": filtered_data.max()
                }

            results.append((column, year, stats))
    final_results = pd.DataFrame(
        {(column, year): stats for column, year, stats in results}
    ).T
    final_results.index.names = ['Sloupec', 'Rok']

    return final_results
```

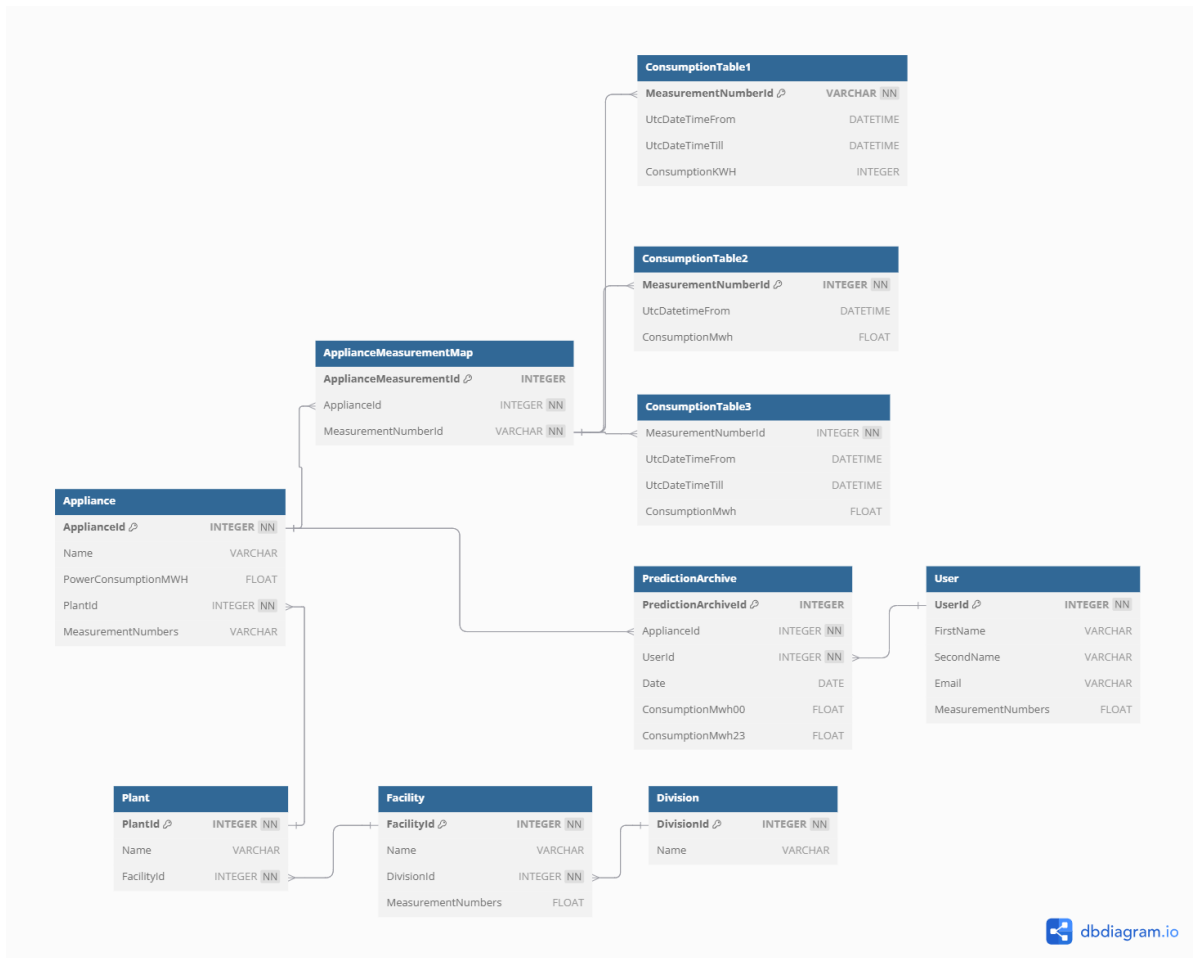
## Příloha 6: Skript v programovacím jazyku Python pro výpočet Spearmanova korelačního koeficientu

(Zdroj: Vlastní zpracování)

```
def odchylky_spearmanova_korelace(dataframe):  
  
    # Definice požadovaných dvojic sloupců  
    column_pairs = {  
        "Odchylka celkem": ['odchylka_celkem_mwh', 'odchylka_celkem_mwh_czk']  
    }  
    results = {}  
  
    for typ, columns in column_pairs.items():  
        # Kontrola, jestli sloupce existují  
        if not all(col in dataframe.columns for col in columns):  
            raise ValueError(f"DataFrame musí obsahovat sloupce: {columns}")  
        filtered_data = dataframe[columns].dropna()  
        # Výpočet Spearmanovy korelace  
        korelace, p_value = stats.spearmanr(filtered_data[columns[0]], filtered_data[columns[1]])  
        results[typ] = {  
            "Spearmanova korelace": korelace,  
            "p-hodnota": p_value  
        }  
  
    return results
```

## Příloha 7: Relační schéma databáze aplikace (část pro reporting)

(Zdroj: Vlastní zpracování)



## Příloha 8: Skript pro výpočet KPIs v prostředí DAX

(Zdroj: Vlastní zpracování)

```
--// WTR:
WithinToleranceRatio =
DIVIDE(
    CALCULATE(
        COUNTROWS(FactElectricity),
        FILTER(
            FactElectricity,
            ABS(FactElectricity[PowerPredictionMWh] - FactElectricity[PowerConsumptionMWh])
            <= 0.05 * FactElectricity[PowerConsumptionMWh]
        )
    ),
    COUNTROWS(FactElectricity)
)

--// RO:
RelativniOdchylka =
VAR ConsumptionMwh = SUM(FactElectricity[PowerConsumption])
VAR PredicitonMwh = SUM(FactElectricity[PowerPredictionMWh])
RETURN
IF(
    NOT ISBLANK(PredicitonMwh) && PredicitonMwh <> 0,
    DIVIDE(PredicitonMwh - ConsumptionMwh, PredicitonMwh),
    BLANK()
)

-- Hodinové MAE:
MAE_hourly =
AVERAGEX(
    FactElectricity,
    ABS(
        FactElectricity[PowerPredictionMWh] - FactElectricity[PowerConsumption]
    )
)

-- Hodinové MAE pro heatmapu (agregace podle úrovně hodina ve dni/den v týdnu)
MAE_Dynamic_HeatMap =
SWITCH(
    TRUE(),
    // Hodinová úroveň
    ISINSCOPE('DimDateTime'[HourOfDay]),
    AVERAGEX(
        VALUES(FactElectricity[CetDateTimeFrom]),
        ABS(
            CALCULATE(SUM(FactElectricity[PowerPredictionMWh])) -
            CALCULATE(SUM(FactElectricity[PowerConsumptionMWh]))
        )
    ),
    // Denní úroveň
    ISINSCOPE('DimDateTime'[WeekDay]),
    AVERAGEX(
        VALUES('DimDateTime'[Date].[Date]),
        ABS(
            CALCULATE(SUM(FactElectricity[PowerPredictionMWh])) -
            CALCULATE(SUM(FactElectricity[PowerConsumptionMWh]))
        )
    ),
    BLANK()
)
```

## Příloha 9: Rozložení stránky sestavy

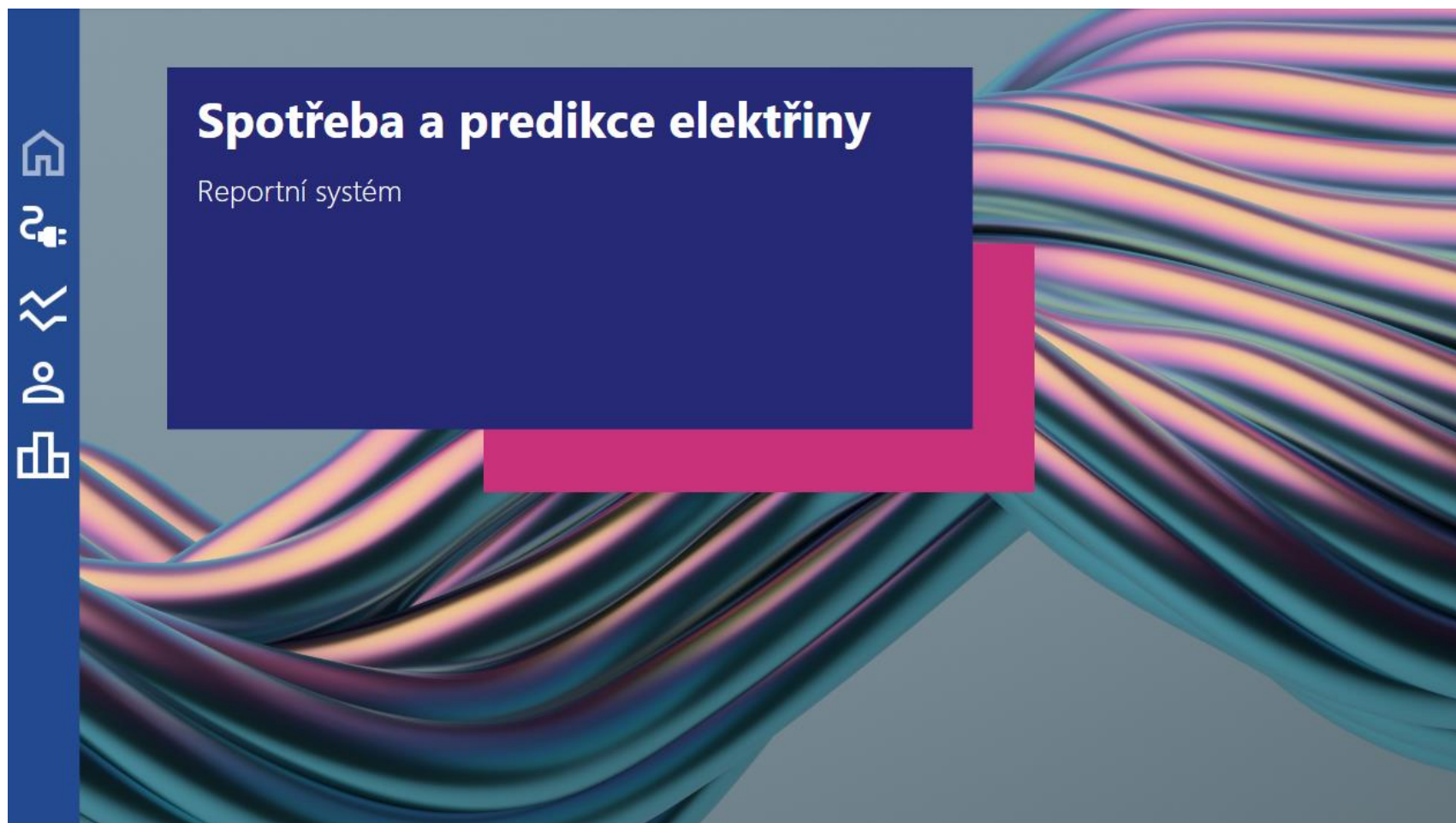
(Zdroj: Vlastní zpracování)

The screenshot displays a software interface for a dashboard layout. The interface is divided into several sections:

- Top Navigation Bar:** Contains the text "Název sestavy" (Dashboard Name), four buttons labeled "Divize č.1", "Divize č.2", "Divize č.3", and "Divize č.4", two dropdown menus labeled "Závod:" and "Provoz:" both set to "Vše", and a search icon.
- Left Sidebar:** Contains four navigation icons: a home icon, a help icon, a refresh icon, and a user profile icon.
- Main Content Area:** A large, empty grey rectangular area.
- Right-Hand Panel:** A panel with a close button (X) in the top right corner, containing four filter sections:
  - Datum od, do:** Two date input fields showing "8.2.2023" and "21.8.2024" with calendar icons, and a range slider below them.
  - Pracovní den / Víkend:** Two radio buttons labeled "Pracovní den" and "Víkend".
  - Dny:** A dropdown menu currently set to "Vše".
  - Hodiny:** Two input fields showing "1" and "24", and a range slider below them.
- Bottom Right:** A small "miro" logo.

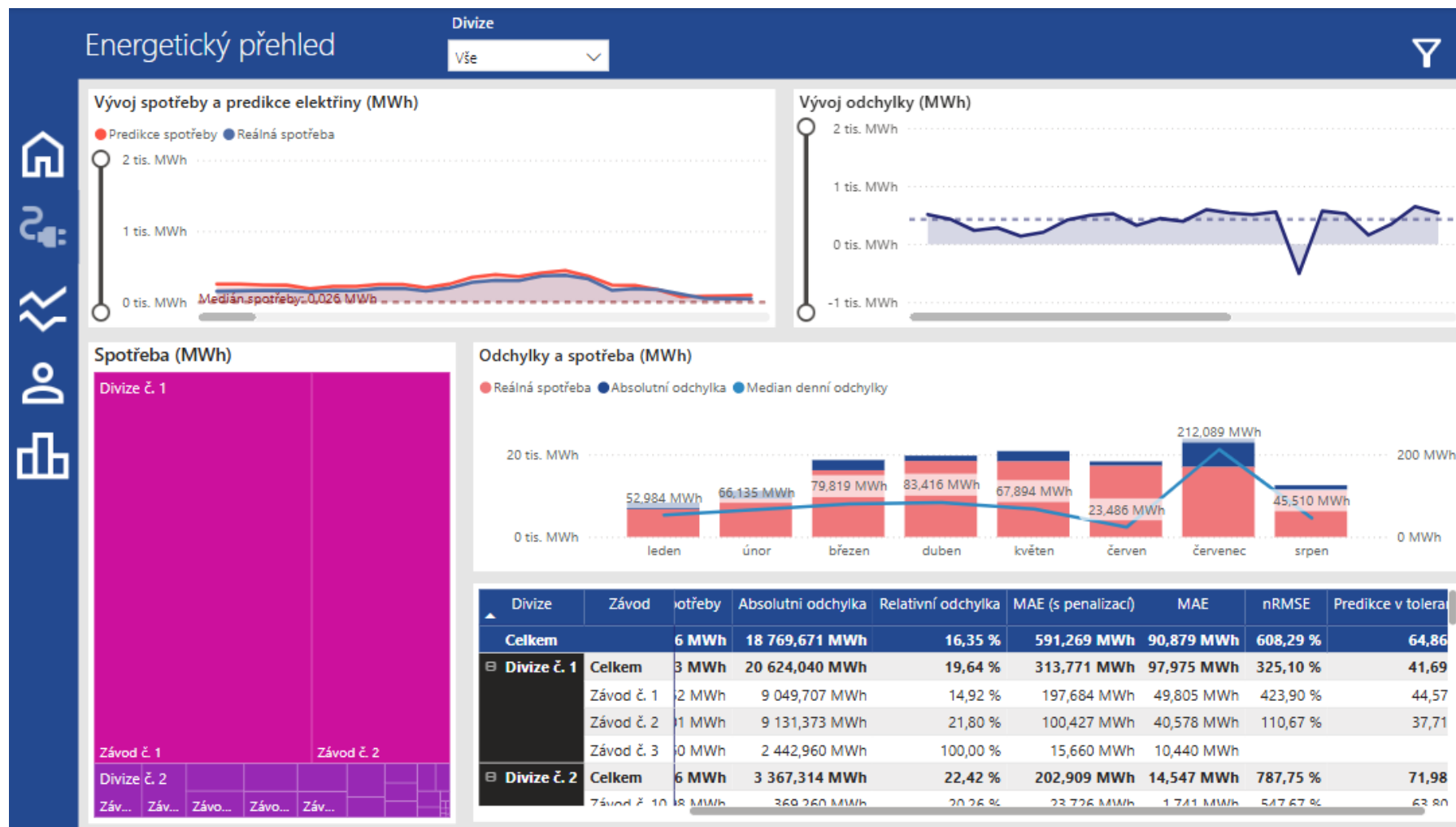
## Příloha 10: Úvodní strana reportního nástroje v Power BI

(Zdroj: Vlastní zpracování)



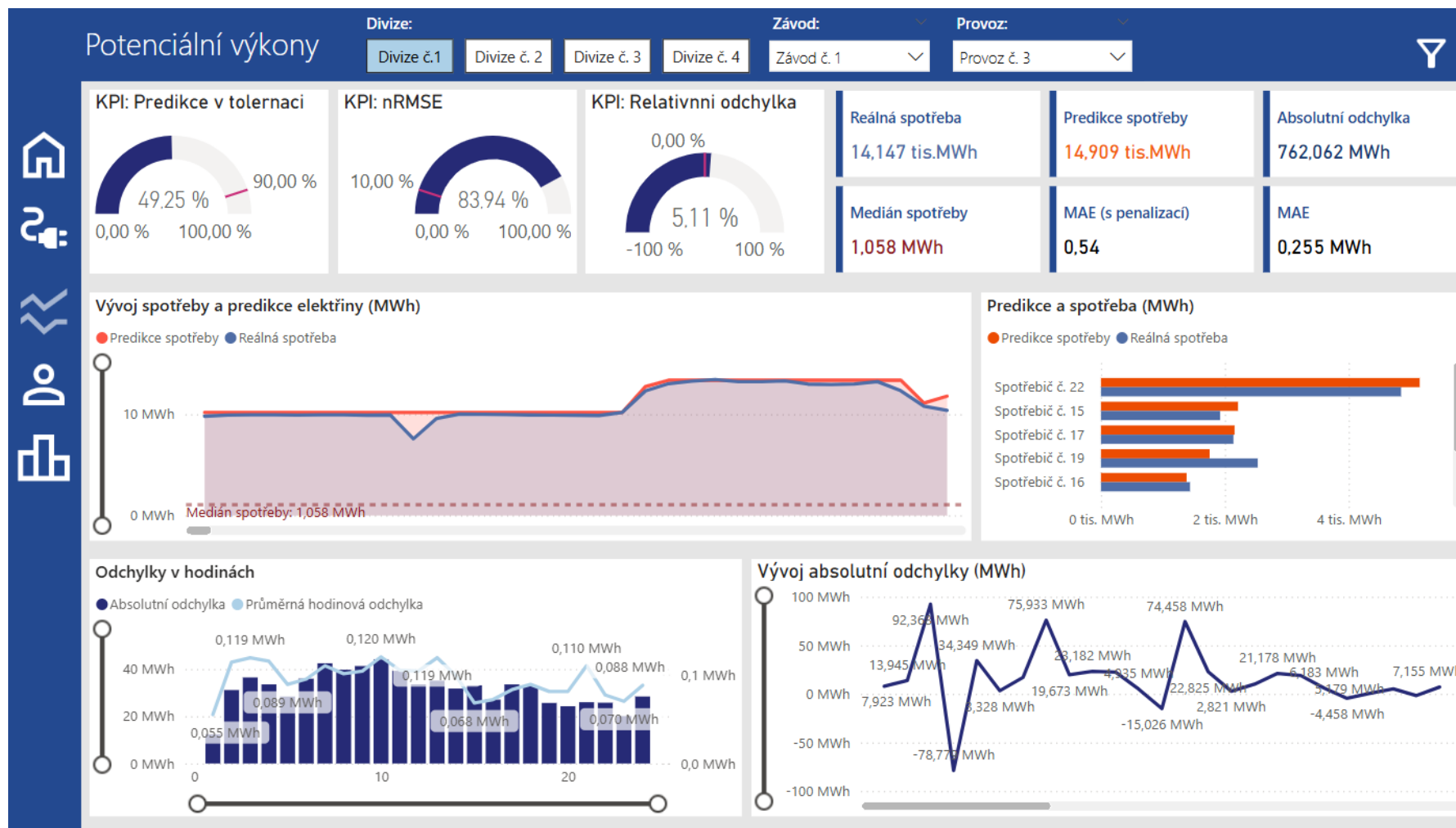
## Příloha 11: Strana Energetický přehled reportního nástroje v Power BI

(Zdroj: Vlastní zpracování)



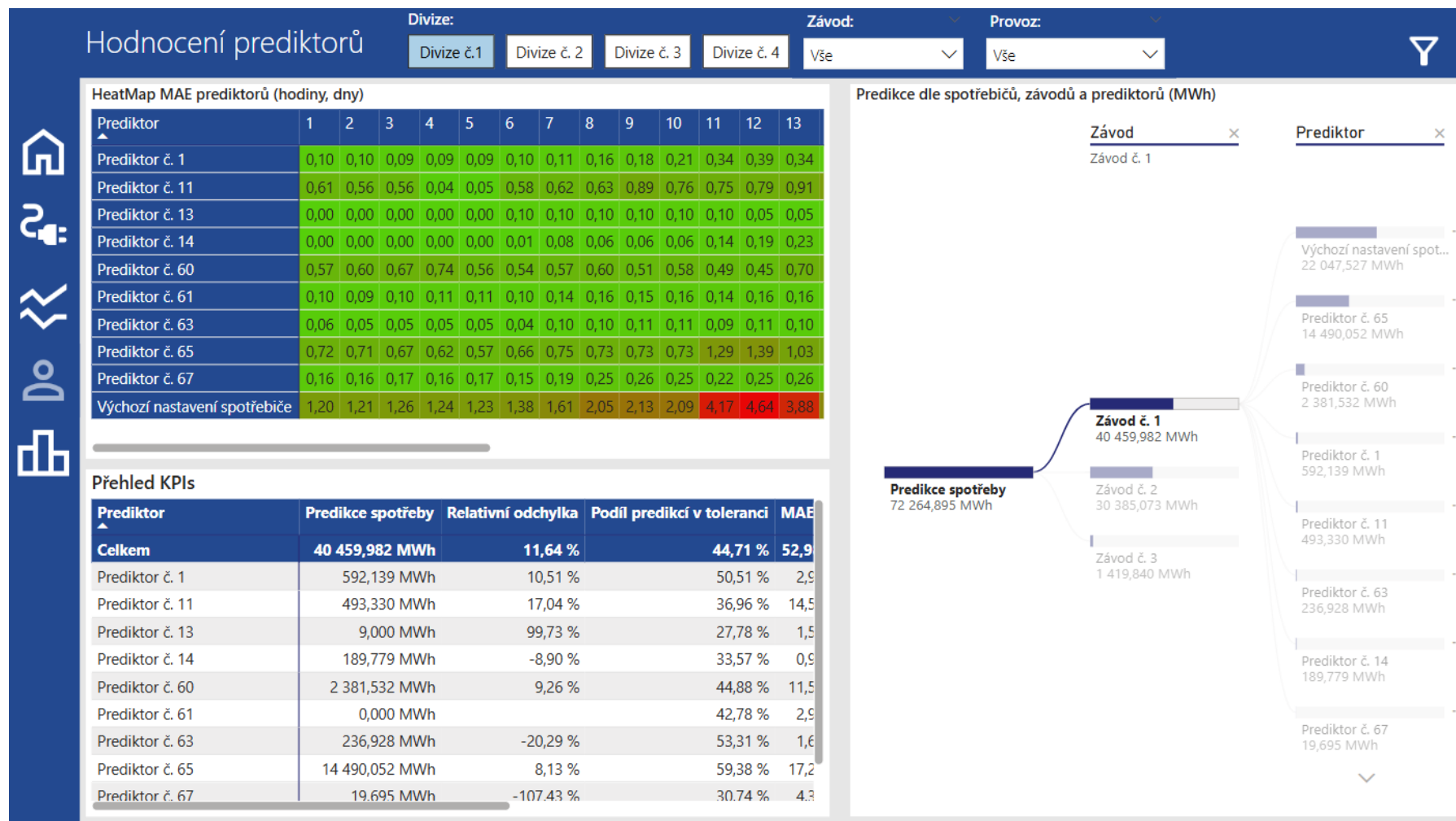
## Příloha 12: Strana Potenciální výkony reportního nástroje v Power BI

(Zdroj: Vlastní zpracování)



## Příloha 13: Strana Hodnocení prediktorů reportního nástroje v Power BI

(Zdroj: Vlastní zpracování)



## Příloha 14: Strana Odchylky reportního nástroje v Power BI

(Zdroj: Vlastní zpracování)

