



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

MATEMATICKÉ METODY V EKONOMII

MATHEMATICAL METHODS IN ECONOMICS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Martin Pohanka

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Mgr. Veronika Novotná, Ph.D.

BRNO 2025

Zadání bakalářské práce

Ústav:	Ústav informatiky
Student:	Martin Pohanka
Vedoucí práce:	doc. Mgr. Veronika Novotná, Ph.D.
Akademický rok:	2024/25
Studijní program:	Manažerská informatika

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 bakalářskou práci s názvem:

Matematické metody v ekonomii

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 analyzovat současnou situaci podniku pomocí statistických metod, přičemž výsledky budou využity k predikci chování zákazníků a návrhu optimalizovaných marketingových strategií zaměřených na cílenou reklamu.

Základní literární prameny:

BAG, D., 2017. Business Analytics. London: Routledge. ISBN 9781138916111. Dostupné z: <https://doi.org/10.4324/9781315464695>.

KROPÁČ, J., 2009. Statistika B. 2. dopl. vyd. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, ISBN 978-80-214-3295-6.

KUBANOVÁ, J., 2008, Statistické metody pro ekonomickou a technickou praxi. 3. vyd. Bratislava: STATIS, ISBN 978-80-85659-474.

SEDLÁČEK, J.,2007, Finanční analýza podniku. 1. vyd. Brno: Computer Press, ISBN 978-80-251-1830-6.

SIEGEL, E.,2016, Predictive analytics: the power to predict who will click, buy, lie, or die. Hoboken, NJ: Wiley, 2016. ISBN 978-1-119-14567-7.

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

V Brně dne 9.2.2025

L. S.

Ing. Jiří Kříž, Ph.D.
garant

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

Abstrakt

Bakalářské práce se zabývá zkoumáním nákupního chování zákazníků e-shopu specializovaného na cyklistické vybavení s cílem analyzovat současnou situaci podniku pomocí statistických metod, přičemž výsledky budou využity k predikci chování zákazníků a návrhu optimalizovaných marketingových strategií zaměřených na cílenou reklamu. Během výzkumu byly provedeny analýzy nákupních košíků, shlukovací analýza a sledování dalších nákupů pomocí SQL dotazů a jazyka Python. S pomocí shlukové analýzy bylo určeno pět hlavních skupin zákazníků. Na tyto výsledky navazuje návrh marketingových strategií pro cílení na různé skupiny včetně křížového prodeje (cross-selling), nabídnutí lepších produktů (up-selling) a osobním přístupem ke komunikaci. Práce dále obsahuje implementační plán se čtyřmi fázemi realizace marketingových strategií. Výsledky poskytují základ pro rozvoj datově řízených marketingových procesů s cílem zvýšit retenci zákazníků, hodnotu průměrného nákupu a celkový obrát e-shopu.

Klíčová slova

analýza dat, python, cílená reklama, analýza nákupních košíků, shluková analýza, segmentace zákazníků, produktové strategie, cross-selling, marketingové strategie

Abstract

This bachelor's thesis examines customer purchasing behavior in an e-commerce store specializing in cycling equipment. The research aims to analyze the company's current situation using statistical methods, with results being utilized to predict customer behavior and design optimized marketing strategies focused on targeted advertising. The research included shopping cart analysis, cluster analysis, and tracking of subsequent purchases using SQL queries and Python. Through cluster analysis, five main customer segments were identified. Based on these results, marketing strategies were developed for targeting different groups, including cross-selling, up-selling, and personalized communication approaches. The thesis also includes an implementation plan with four phases for executing marketing strategies. The findings provide a foundation for developing data-driven marketing processes aimed at increasing customer retention, average purchase value, and overall e-commerce turnover.

Keywords

data analysis, python, targeted advertising, market basket analysis, cluster analysis, customer segmentation, product strategies, cross-selling, marketing strategies

Bibliografická citace

POHANKA, Martin. *Matematické metody v ekonomii*. Brno, 2025. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/165326>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav informatiky. Vedoucí práce doc. Mgr. Veronika Novotná, Ph.D.

Čestné prohlášení

Prohlašuji, že předložená bakalářská 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 16. 5. 2025

Martin Pohanka

autor

Poděkování

Tímto bych chtěl vyjádřit své upřímné poděkování doc. Mgr. Veronice Novotné, Ph.D. za její odborné vedení, cenné rady a inspirativní připomínky při zpracování mé bakalářské práce. Velmi si vážím času, který mi věnovala, její ochoty a trpělivosti během konzultací.

OBSAH

ÚVOD	11
CÍL PRÁCE	12
1 TEORETICKÁ VÝCHODISKA	13
1.1 Ekonomie	13
1.1.1 Mikroekonomie.....	14
1.1.2 Makroekonomie	15
1.2 Matematická ekonomie	17
1.3 Business analytika.....	17
1.3.1 Analytika zákazníků.....	18
1.3.2 Cíle business analytiky	19
1.4 Cílená reklama	19
1.4.1 Analýza dat pro cílenou reklamu	20
1.4.2 GDPR problematika při analýze dat	21
1.4.3 Proces analýzy dat pro cílenou reklamu	22
1.4.4 Předzpracování dat.....	23
1.4.5 Shluková analýza a segmentace zákazníků.....	27
1.4.6 Analýza následujících produktů (cross-selling).....	28
2 ANALYTICKÁ ČÁST	29
2.1 Představení společnosti	29
2.2 Úvod do analytické části.....	29
2.3 Metody analýzy	29
2.3.1 SQL dotazy	29
2.3.2 Python	30
2.4 Popis datového souboru	30
2.4.1 Data kategorií.....	30
2.4.2 Data prodeje.....	30
2.4.3 Data zákazníků.....	31
2.4.4 Data PSČ	32
2.4.5 SQL databáze	32
2.5 Analýza dat.....	34
2.5.1 Analýza počtu prodeje v jednotlivých kategoriích	34
2.5.2 Analýza prodeje v jednotlivých kategoriích podle celkové ceny	35

2.5.3 Analýza nákupních košíků	36
2.5.4 Analýza zákazníků	39
2.5.5 Analýza následujících produktů a skupin	40
2.6 Interpretace výsledků	47
2.6.1 Analýza nákupních košíků	47
2.6.2 Analýza zákazníků	48
2.6.3 Analýza následujících produktů a skupin	52
3 VLASTNÍ NÁVRHY ŘEŠENÍ.....	55
3.1 Shrnutí hlavních výsledků analýz	55
3.1.1 Analýza nákupních košíků	55
3.1.2 Analýza kategorií	55
3.1.3 Analýza následných nákupů	55
3.2 Návrh marketingových strategií.....	56
3.2.1 Strategie cílení na segmenty zákazníků	56
3.2.2 Co nabízet – produktové strategie.....	57
3.2.3 Komunikace se zákazníky.....	57
3.3 Možnosti implementace a očekávané přínosy	59
3.3.1 Scénář strategického implementačního plánu.....	59
3.3.2 Odhadovaný potenciální přínos	60
3.3.3 Scénář byznysového dopadu implementace	61
3.4 Náklady na implementaci	62
3.4.1 Technická implementace	62
3.4.2 Personální kapacity	62
3.4.3 Marketingové výdaje	63
3.5 Prioritizace návrhů	63
3.6 Návrh metrik pro měření úspěšnosti projektu	64
3.7 Kritická omezení současné analýzy	65
3.7.1 Datová omezení	65
3.7.2 Metodologická omezení.....	65
3.7.3 Technická omezení	65
3.8 Doporučení pro další analýzy	66
3.8.1 Pokročilé prediktivní modelování chování zákazníků	66
3.8.2 Analýza zákaznického životního cyklu	66
3.8.3 Systematické testování a optimalizace konverze	66
3.8.4 Hlubková analýza webových interakcí a nákupního procesu	67

3.8.5 Analýza sentimentu a zpětné vazby zákazníků.....	67
3.8.6 Implementace prediktivní analýzy nákupního košíku	67
4 ZÁVĚR	69
SEZNAM POUŽITÝCH ZDROJŮ	71

ÚVOD

Název mé bakalářské práce je Matematické metody v ekonomii. Matematika má důležitou úlohu v dnešní ekonomii a je potřebný nástroj pro pochopení a předpovídání vývoje hospodářských jevů. Spojení matematiky se statistikou dává silné základy pro zkoumání velkého množství dat, to dovoluje lépe rozhodování a plánování.

Práce prozkoumává možnosti využití matematických metod při analýze historických dat z prodeje, a to s ohledem na optimalizaci cílené reklamy. V dnešní době, kdy firmy mají velké množství dat je schopnost správně tato data analyzovat nezbytná pro zvýšení konkurenceschopnosti na trhu. Tato práce se zaměřuje na identifikaci vzorců a trendů historických nákupů pomocí statistických metod co mohou pomoci ke strategickému rozhodování co mohou pomoci ke strategickému rozhodování

Práce bude konkrétně zaměřena na využití matematických a statistických metod pro vyhodnocení historických prodejních dat a návrh optimálních strategií pro cílenou reklamu. Hlavní roli bude hrát regresní analýza a další pokročilé statistické postupy, které umožní pochopit historické chování zákazníků a predikovat budoucí trendy.

CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ

Cílem práce je analyzovat současnou situaci podniku pomocí statistických metod, přičemž výsledky budou využity k predikci chování zákazníků a návrhu optimalizovaných marketingových strategií zaměřených na cílenou reklamu. Metody a postupy se zaměřují především na analýzu historických dat z e-shopu s využitím SQL a Pythonu. Vybrané metody zahrnují analýzu nákupních košíků, shlukovou analýzu pro segmentaci zákazníků a sledování následných nákupů. V rámci práce jsou využity základní metody popisné statistiky spolu s vizualizací dat. Na základě provedených analýz jsou formulovány marketingové strategie spolu s jejich implementačním plánem.

1 TEORETICKÁ VÝCHODISKA

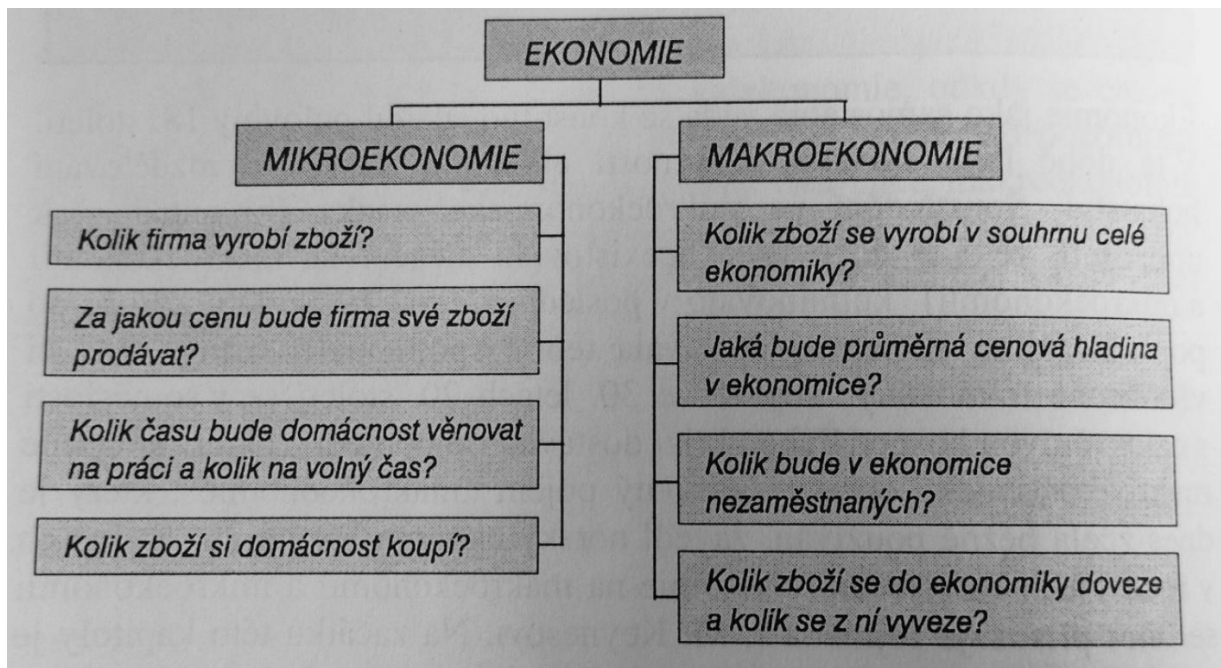
1.1 Ekonomie

Ekonomie není jen věda o číslech a grafech, jádro se nachází ve zkoumání lidského chování. Zajímá se o to jak lidé, firmy a celá společnost reagují na výzvy spojené s omezenými zdroji. Od rozhodování o rozpočtu domácnosti až po tvorbu národní hospodářské politiky. ekonomie analyzuje příčiny a důsledky těchto procesů, přičemž klade důraz na efektivní alokaci zdrojů co maximalizuje společenský blahobyť. Jedním z hlavních pojmů v ekonomii je racionální chování. To ale nelze vidět jako vždy správné nebo lepší rozhodování. Racionalita spočívá ve výběru cest, které co nejlépe vedou k dosažení osobních cílů, a každý člověk má své vlastní zájmy, cíle i hodnoty. Milovník hor nemusí chápat, že někdo má rád dovolenou na pláži a obráceně. Podobné rozdíly jsou jasné i mezi dalšími lidmi, například v jejich způsobu dopravy nebo nákupech. Ekonomie ale neposuzuje, či výběry jsou správné, ale dívá se na zásady, podle kterých se lidé rozhodují. [1][2]

Když se díváme na zvyšování cen nebo debaty o regulaci trhů, jde vždy o balancování zájmů různých skupin. Uživatelé, pracovníci a akcionáři a každý má své preference a cíle. Vláda jako regulátor musí zvažovat vlivy svých politik nejen na jednu ale na všechny strany, co jsou zapojené. Ekonomie proto pomáhá analyzovat tyto konflikty a identifikovat příležitosti k optimálním řešením. [1] [3]

Hospodářská politika má jasný dopad na národní ekonomiku, jako je práce lidí, růst platů nebo síla měny. Hlavními částmi ekonomiky jsou rozhodnutí jednotlivců jako podnikatelé, kupující, výrobci a investoři. Ekonomie nás učí, že většina hospodářských výsledků pramení z chování lidí a jejich rozhodnutí, a důraz klade na pochopení toho co je motivuje a jaké to má dlouhé následky. Bez ohledu na druh ekonomického systému kapitalistický, socialistický nebo jiný, ekonomické problémy se týkají rozdělení vzácných zdrojů. Tato nedostupnost vyžaduje efektivnost a pořadí při uspokojení lidských potřeb. Jako na bojišti lékař rozhoduje, jak nejlépe využít omezené léky a čas, aby maximalizoval šance na záchranu životů. Podobně vlády a firmy rozhodují, kam nasměrovat své prostředky pro dosažení největšího přínosu. [1][2]

Ekonomie je tedy mnohem více než jen soubor teorií, je to nástroj, který nám pomáhá činit informovanější rozhodnutí. Tato rozhodnutí ovlivňují nejen náš individuální blahobyť, ale také společnost jako celek. Ekonomie jako věda zahrnuje širokou škálu témat, která můžeme rozdělit do dvou hlavních oblastí: mikroekonomie, která si pokládá otázky pro jednotlivé tržní subjekty a makroekonomie, která si pokládá otázky z regionální, národní a celosvětové ekonomiky (Obrázek 1). Tyto oblasti se vzájemně doplňují a umožňují hlubší porozumění fungování ekonomiky na různých úrovních. [1]



Obrázek 1: Rozdíl makroekonomie a mikroekonomie[4]

1.1.1 Mikroekonomie

Mikroekonomie se věnuje zkoumání chování jednotlivých ekonomických aktérů jako jsou spotřebitelé, firmy a stát a jejich vzájemné interakce na trhu. Tento obor se soustředí na hlavní ekonomické činnosti, mezi které patří spotřeba, výroba a směna, a představuje základní rámec pro pochopení alokace vzácných zdrojů a rozhodovací procesy. [5]

1.1.1.1 Chování spotřebitelů: Užitek, preference a rozhodování

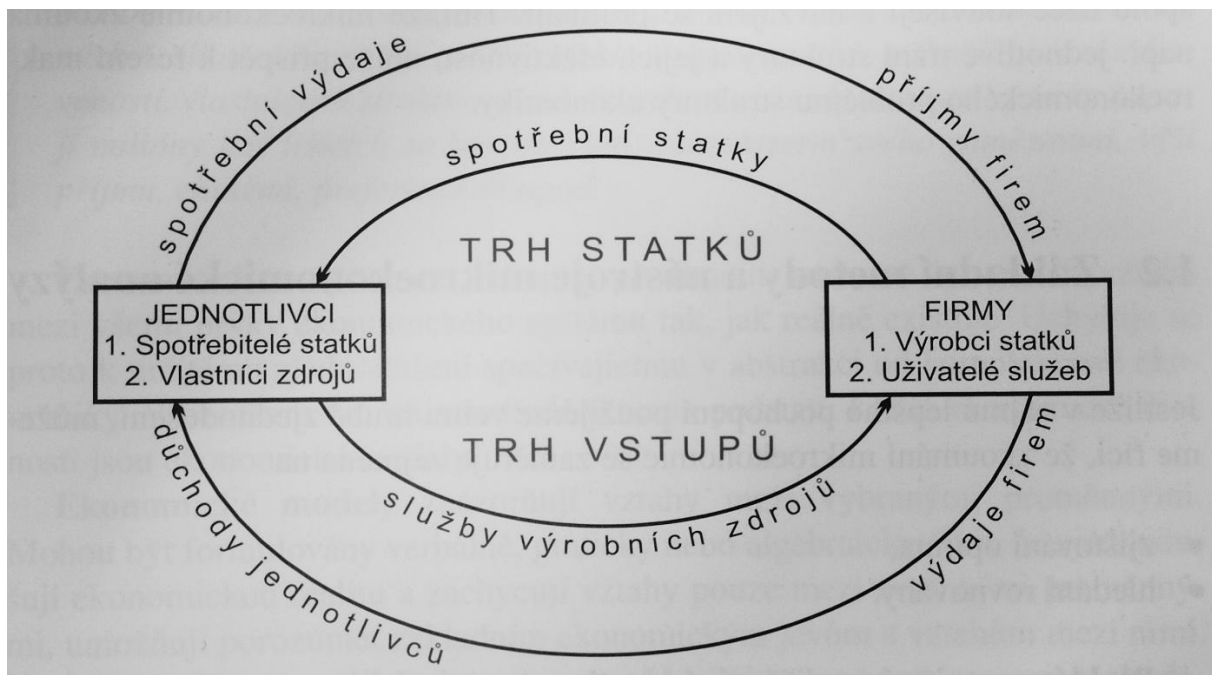
Lidé mají omezené množství zdrojů, a proto musí při svých volbách přemýšlet, jak nejlépe rozdělit své peníze na nákup různých statků (výrobků a služeb). Toto rozhodování se zakládá na srovnávání toho, co dostanou z používání jednotlivých věcí, s náklady spojenými s jejich koupí. Cíl rozumně přemýšlejícího kupujícího je zvětšit svůj prospěch, i když má omezené příjmy. Při rozhodování lidí hrají důležitou roli jejich preference, které ovlivňují jaké statky budou chtít. Tyto volby jsou důsledkem mnoha věcí, včetně tělesných, duševních, kulturních a společenských vlivů. Na nižších úrovních vývoje společnosti je hlavním cílem přežití, zatímco se vzrůstající životní úrovní se zájem lidí posouvá ke splnění vyšších potřeb. [5][6]

Tyto volby jsou výsledkem mnoha faktorů, včetně tělesných, duševních, kulturních a společenských vlivů. Na nízké úrovni rozvoje společnosti je hlavní cíl přežít. Zatímco se zvyšujícím životním standardem se zájem lidí posouvá k splnění větších potřeb. Racionální rozhodování spotřebitele má dvě hlavní otázky, jak získat peníze a jak tyto peníze rozdělit mezi různé zboží. První otázka je prozkoumána v analýze nabídky výrobních faktorů. Druhá otázka se týká formování poptávky na trhu a je důležitá pro chování spotřebitelů. Spotřeba dává podnět k výrobě, která mění zdroje na zboží. [5][6]

1.1.1.2 Chování firem: Výroba, náklady a zisk

Firmy jsou dalším důležitým prvkem mikroekonomie. Rozhodují o tom, co dělat, kolik toho udělat za jakou cenu a jaké nástroje použít. Hlavním cílem firmy je zvýšení zisku. Její rozhodování zahrnuje výběr nejlepšího výstupu a ceny. Výroba je charakterizována jako proces přeměny vstupů (zdrojů) na výstupy (produkty a služby). Tento proces může zahrnovat fyzickou přeměnu (například výrobu potravin), změnu v čase (například uskladnění plodin) nebo prostorovou přeměnu (například přeprava zboží) (Obrázek 2). Mikroekonomie taky zohledňuje náklady na transakci, co souvisí s chodem firem, které zahrnují náklady o dohodě ohledně používání výrobních faktorů. Další důležitou součástí finančního řízení podniku je horizontální a vertikální analýza, které poskytují přehled o finančním zdraví firmy. Zatímco horizontální analýza sleduje vývoj jednotlivých položek v čase, vertikální analýza umožňuje pochopit jejich podíl v rámci celkové finanční struktury podniku. [6][7]

Mikroekonomie má velký význam v chápání základních pravidel fungování hospodářských systémů. To je nutné pro jakoukoli použití pokročilejších metod analýzy dat. Dobrá znalost jednání ekonomických aktérů umožňuje nejenom objasnit současné události, ale může i vytvářet základy pro další modelování a předpovědi. Tady mikroekonomie těsně souvisí s matematickou ekonomikou, která pomocí matematických nástrojů zaznamenává ekonomické poznatky. [6]



Obrázek 2: Schéma mikroekonomického oběhu[5]

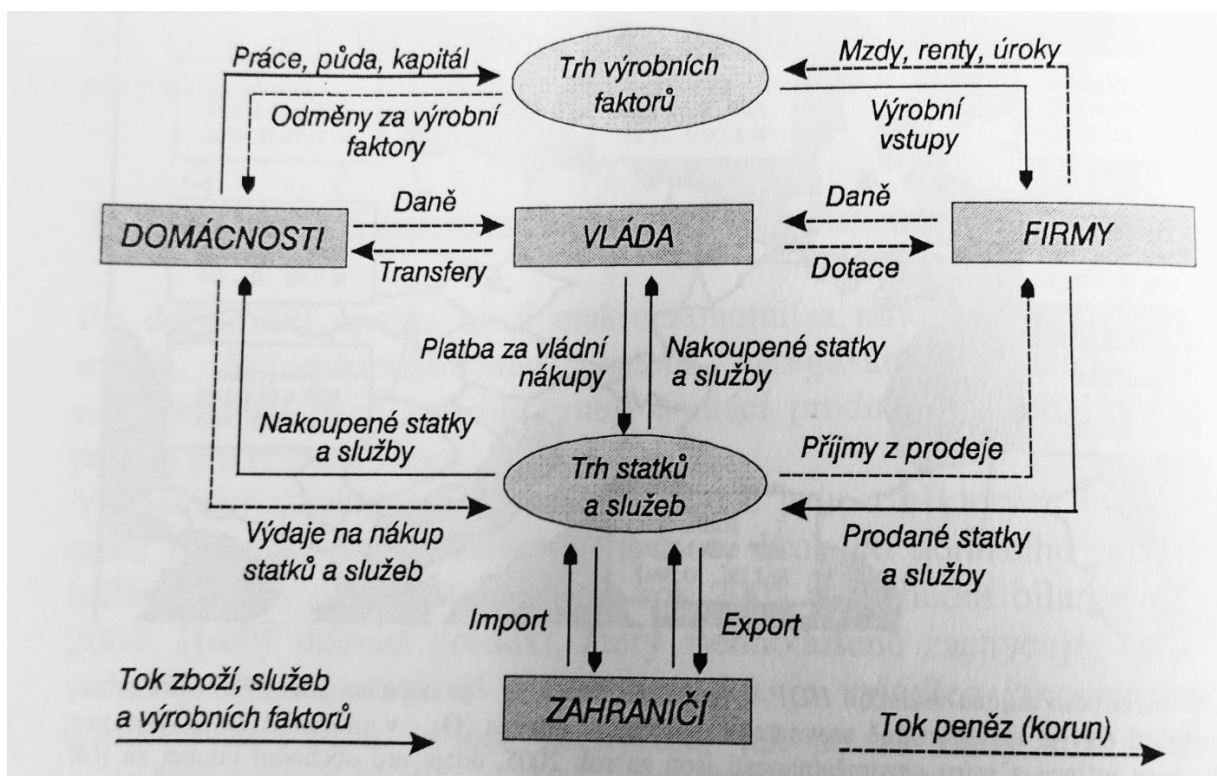
1.1.2 Makroekonomie

Makroekonomie je obor, který zkoumá ekonomiku jako celek. Na rozdíl od zaměření na jednotlivé části hospodářství se makroekonomie věnuje složitým otázkám, jako jsou národní produkt míra zaměstnanosti, cenová hladina, inflace, veřejné finance nebo obchodní

balance. Tyto oblasti mají velký vliv na fungování společnosti a jsou hlavním bodem debat mezi politiky a odborníky. Makroekonomie se také soustředí na celkovou výkonost ekonomiky. To zahrnuje zkoumání ekonomické úrovně, hospodářského růstu a jeho důvodů. Jakož i krátkých a dlouhých změn v hospodářství oproti jeho potenciálu. Zkoumá celkovou výrobu statků a služeb, spotřebu investice a chod trhů. Jakými jsou trh práce nebo trh s aktivy. Makroekonomie se také snaží najít důvody pro ekonomické problémy, jako inflace a nezaměstnanost a hledá způsoby k jejich řešení. Další důležitý prvek makroekonomie je zkoumání chování státních úřadů, například vlády a centrální banky, a jejich zásahů do ekonomiky. Důležitými nástroji státu jsou měnová a fiskální politika, jejichž prostřednictvím může stát ovlivňovat ekonomický vývoj. Makroekonomie také zahrnuje mezinárodní porovnání národních ekonomik a analýzu jejich vzájemných vztahů v rámci globálního hospodářství. [4][8][9]

1.1.2.1 Makroekonomický koloběh

Makroekonomický cyklus ukazuje zjednodušený vzor, který zachycuje hlavní ekonomické spojení mezi čtyřmi hlavními skupinami: domácnostmi, podniky, vládou a zahraničím. Tento vzor ukazuje tok věcí, služeb, výrobních faktorů a peněz v hospodářství. Jsou zde dva hlavní trhy. Na trhu výrobních faktorů domácnosti nabízejí práci, půdu a kapitál, které firmy chtějí a platí rodinám mzdy renty a úroky. Tyto příjmy pak rodiny používají k nákupu statků a služeb. Na trhu statků a služeb, firmy nabízejí zboží vytvořené z výrobních faktorů a domácnosti poptávají díky svým příjmům. Zahraničí zde funguje jako exportér i importér a vláda jako poptávající subjekt (Obrázek 3). [8]



Obrázek 3: Schéma makroekonomického koloběhu[8]

1.2 Matematická ekonomie

Matematická ekonomie je obor, který zkoumá vztahy v ekonomice s pomocí matematických nástrojů. Tento vědní obor vychází z poznatků o ekonomice, což zahrnuje popis ekonomických systémů a jejich terminologie, které pak formalizuje pomocí matematiky. Jako každá věda i matematická ekonomie chce dosáhnout tří hlavních cílů: popsat, vysvětlit a předpovědět. Hlavní myšlenka na, kterou se tato disciplína zaměřuje, je vytváření matematických modelů ekonomiky. Tyto modely pomáhají nejen k jasnému popisu ekonomických situací, ale také umožňují lepší porozumění základním vztahům uvnitř zkoumaných systémů. Díky těmto modelům můžeme prozkoumat nynější stav, tak předpovědět budoucí vývoj konkrétního systému a jeho reakce na změny různých parametrů. [10]

Pokud bychom se na ekonomii dívali hlavně jako na analytickou a většinou teoretickou oblast, můžeme matematiku označit za její jazyk. Tento přístup ale může být v některých případech problémový. Obzvláště když ekonomie bere inspiraci z různých jiných oblastí lidského poznání jako je filozofie, sociologie, psychologie, právo a další. Zavedení matematiky jako univerzálního jazyk může být těžké. Matematická ekonomie se ale zaměřuje na tu část ekonomie, kde použití matematiky není jen smysluplné, ale také poskytuje velké přínosy. Díky použití matematických metod je možné dobře prozkoumat složité ekonomické systémy a získané znalosti použít k lepšímu pochopení a předpovědi změn. Matematická ekonomie tedy ukazuje důležitý nástroj pro pokročilé ekonomické zkoumání, jejichž výsledky mohou být užity jak v teorii ale také v reálném prostředí. [10][11]

1.3 Business analytika

Oblasti business analytiky zahrnuje analýzu rozsáhlých datových souborů s cílem odhalit, pochopit a předat nové poznatky a informace., které se dějí v podniku. Business analytika se soustředí na různé oblasti měření mezi které patří (ale nejsou omezené na):

- Analytika lidských zdrojů
- Analytika dodavatelského řetězce
- Analytika zákazníků
- Analytika podnikových procesů
- Finanční analytika

Analytika lidských zdrojů se zabývá zkoumáním pracovníků a pokrývá celý jejich životní cyklus od nábora, přes hodnocení výkonu, až po motivaci a zapojení zaměstnanců. Správné ukazatele pomáhají zlepšit postupy a procesy, zvyšují spokojenost, udržení, školení a podporu pro zaměstnance, zlepšují morálku, snižují náklady a zvyšují produktivitu. Do této oblasti patří například nábor, vzdělávání, vztahy mezi zaměstnanci, jejich spokojenost a fluktuace. [12]

Analytika dodavatelského řetězce se soustředí na zkoumání částí dodávek, jako je výběr a řízení dodavatelů, analýza zásob nebo efektivita přepravy. Tento přístup dává možnost k vylepšení procesů a lepší dodávky zákazníkům. Působí jako nástroj ke zlepšení procesů a je pomocník pro programy na snížení nákladů. Nabízí přehled o přímých i nepřímých nákladech souvisejících se zbožím a službami. Současně identifikuje možnosti úspor prostřednictvím

konsolidace dodavatelů a komodit. Zároveň zlepšuje dodržování pravidel díky efektivnímu sledování výdajů a spolupráce s dodavateli. [12]

Analytika zákazníků se věnuje pochopení zákazníků, jejich životního cyklu, potřeb a spokojenosti. Jedná se o systematickou interpretaci informací o zákaznících, která podnikům pomáhá udržet ziskové zákazníky a aktivně s nimi budovat vztah. Analýza chování zákazníků identifikuje a hodnotí důležité faktory, na základě, kterých si zákazníci vybírají mezi produkty. Charakterizace zákazníků umožňuje podnikům lépe porozumět jejich potřebám, čímž mohou zvýšit prodeje a růst. Znalost faktorů ovlivňujících rozhodování zákazníků přispívá k jejich ovlivňování prostřednictvím cílené marketingové strategie. [12]

Analytika podnikových procesů se zabývá analýzou činností souvisejících s dodávkami produktů a služeb v rámci organizace, aby podpořila zlepšení efektivity a procesů. Tento přístup zahrnuje různé pohledy na systém a usnadňuje optimalizaci rozhodovacích procesů. [12]

Finanční analytika se zaměřuje na analýzu finančních dopadů v rámci business analytiky. Jedním z jejích aspektů je práce s čistými hodnotami, které jsou výsledkem po zohlednění daní, poplatků, odvodů či kapitálových nákladů. Finanční analytika se zabývá řízením peněžních toků, rozložením likvidity, řízením investic do nových produktů či pracovního kapitálu a plánováním prostředků. Díky tomu přispívá ke zlepšení finanční výkonnosti, kontroluje výdaje a zvyšuje ziskovost napříč prodejními, regiony, produkty nebo kanály. [12]

1.3.1 Analytika zákazníků

Současní zákazníci mají pocit, že mají větší míru kontroly a nezávislosti, a to firmy nemohou přehlédnout. Firmy vynakládají dostatečné úsilí, aby své zákazníky co nejlépe poznaly. Hlavní je vytvořit úplný a celkový pohled na zákazníka, který zahrnuje všechny komunikační cesty. Tato strategie umožňuje sběr dat o zákazníkovi a jejich efektivní spojení. Tento přístup se zaměřuje na předvídání chování zákazníků a aktivně navrhuje personalizované nabídky. Díky tomu umožňuje jak křížový prodej, tak přechod od základní nabídky k vyšší kvalitě a tím přispívá ke zlepšení zákaznické spokojenosti. Přesnost při cílení na zákazníka může zlepšit personalizace kampaní a interakcí stejně jako optimalizace načasování a obsah komunikace. [12][13]

Důležitou částí zákaznické analytiky je schopnost dodavatelského řetězce reagovat na změny trhu, co se dá hodnotit jeho rychlostí, účinností a rozsahem přizpůsobení. Například pružnost, když roste poptávka, se může posuzovat podle počtu dnů potřebných na úpravy. Přičemž také hraje důležitou roli to jak řešíme stížnosti zákazníků a návrh produktů pro zlepšení marketingových funkcí. Firmy nemohou přehlížet potřebu zajistit nevyšší spokojenost zákazníků. Životní cyklus výrobku hraje důležitou roli. Zahrnuje věci, jako je povědomí o značce, změny, aktivace zákazníků, věrnost k výrobku, zisk na klienta, náklady kanálů na klienta, prodej křížových produktů, spokojenost zákazníků, fluktuace a hodnota zákazníka během jeho životního cyklu. [12]

1.3.2 Cíle business analytiky

Business analytika zahrnuje všechny důležité informační a rozhodovací atributy podniku. Je nezbytné, aby se tato disciplína stala součástí celkové strategické vize každé organizace, protože hraje významnou roli v rozvoji a úspěchu podniků. Hlavní cíle business analytiky jsou následující:

- Poskytovat aktuální a aplikovatelné informace, které podporují kvalitní rozhodování v podnikání.
- Nabízet nástroje na všech organizačních úrovních k podpoře rozhodování týkajícího se zákaznických cílů a ziskovosti, včetně porovnání výkonnosti.
- Zajišťovat analýzy, které pomáhají přesněji a objektivněji předpovídat budoucí vývoj.
- Poskytovat náhledy a pochopení, jež podporují informovaná rozhodnutí a sebevědomé kroky, a nabízet zpětnou vazbu potřebnou k vytvoření organizace s učením. [12][14]

Business analytika je charakteristická několika jedinečnými vlastnostmi, Cíleností, Intuitivností a Efektivností.

Cílenost: Business analytika musí jasně identifikovat účel, pro který jsou dodávky realizovány a analýzy prováděny. Zjištění z analytiky musí odpovídat funkcím podniku (finance, marketing, prodej apod.) i prioritám a cílům managementu, jako jsou výkonnost, růst, riziko či ziskovost. [12]

Intuitivnost: Tato analytika je přínosná svou schopností odhalovat nové poznatky a souvislosti, které dříve nebyly zjevné. Identifikace příčin a důsledků je důležitým faktorem pro strategické rozhodování. Analýzy, které pouze potvrzují stávající stav nebo běžné znalosti, nepřinášejí žádnou přidanou hodnotu. [12]

Efektivnost: Výsledky a akční plány podnikové analýzy musí být proveditelné a užitečné, co znamená, že vedoucí by měli moci hned jednat na základě rad z analýzy. [12]

Oblastí podnikové analýzy patří do širokého spektra činnosti, které jsou důležité pro efektivní práci a rozvoj firem. Business analytika poskytuje nástroje a informace potřebné pro efektivní rozhodování ve firmě. Její výsledky navíc nacházejí přímé uplatnění v praktických oblastech, jako je například reklama. Dovednost zkoumat data a předvídat chování kupujících totiž tvoří jeden ze základů efektivní a personalizované marketingové strategie. [12][14]

Cílená reklama je příkladem, jak se poznatky získané pomocí obchodní analytiky promítají do praxe. Díky zkoumání starých i nových dat o chování zákazníků může být reklama nejen více aktuální, ale i účinnější, přičemž splňuje cíle, které obchodní analytika stanovuje

1.4 Cílená reklama

Tradiční způsoby inzerování, jako je televize, rádio, noviny a billboardy, mají jen malý potenciál zaměřování. To často vede k zbytečnému utrácení peněz na oslovení nezájemců. Naopak online inzerce dává lepší možnosti cílení. Umožňuje spojit reklamu s konkrétní činností uživatele (např. vyhledávání nebo prohlížení), což usnadňuje rozčlenění zákazníků. Navíc lze

měřit úspěšnost reklamy skrze analýzu dat o návštěvnicích přivedených reklamou což dovoluje lepší rozdělení financí. Hlavní význam internetové reklamy je dovednost přesného zaměření na uživatele, kterou zajišťují lidé. Ti zkoumají informace o uživateli a nabízejí pokročilé způsoby rozdělení a cílení, které se liší podle druhu internetové reklamy a požadavků na data. [15][16]

Cílení podle kontextu je postavené na sdělení navštívené webové stránky nebo na otázce vložené do vyhledávače. Například uživatel, který čte blog o kolech pravděpodobněji uvidí reklamy na jízdní kola. Tato metoda nevyžaduje uživatelská data a je oblíbená u klasifikovaných a vyhledávacích reklam. Chování cílené reklamy využívá informace o uživateli, jako je historie vyhledávání, aktivity na platformách nebo technické údaje o zařízení. Tato metoda je náročná na data, která se získávají pomocí cookies nebo jiných technologií sledování. Příkladem je re-targeting, který připomíná uživateli produkty, o něž projeví zájem, ale nezakoupili je. Segmentované cílení se zakládá na informacích, které uživatel sám sdílí, třeba při vytváření profilu na webu nebo zákaznické kartičky. Tyto údaje mohou být spojeny s jinými daty pro dosažení přesného cílení. [16][15]

Online cílení reklamy roste stále víc oblíbenost mezi inzerenty díky větší účinnosti a návratu investic. Behaviorální reklama, která používá uživatelská data, má mnohem vyšší míry prokliků než běžná reklama, zvláště u lidí, kteří už mají zájem o produkt. Ale moc osobní reklama může snížit efektivitu, pokud chybí důvěra uživatele v inzerenta nebo jasnost ohledně využití dat. Z pohledu ekonomického má chování cílení různé vlivy na poskytovatele reklamních míst a reklamní firmy. Na jedné straně to umožňuje lepší rozdělení, což může snížit soutěž mezi inzerenty, ale také zvyšuje šanci na kliknutí, tedy příjmy z úspěšných akcí. Celkový dopad však závisí od účinnosti metod, které jsou použity pro cílení. [16][15]

1.4.1 Analýza dat pro cílenou reklamu

V nynější době zkoumání dat je jeden z hlavních nástrojů, co pomáhají firmám dostat výhodu na trhu. Díky pokroku v oblasti sběru dat má firma přístup k cenným informacím o chování zákazníků jejich preferenčních a reakcích na reklamy. Tyto informace hrají důležitou roli při vytváření efektivních plánů a rozhodování v marketingu. Proces analýzy dat dává firmám šanci lépe pochopit potřeby jejich zákazníků, a také cílit reklamu přesněji na určité části trhu. Hledání vzorů v chování zákazníků dává manažerům možnost upravit nabídky a zlepšit marketingové akce. To může vést k vyšší účinnosti vynaložených prostředků a lepším návratům investic. [17][13]

Kromě cílení reklamy, pomocná analýza dat pomáhá při rozhodování na vyšší úrovni. Dává skutečné informace pro plánování a předpovídání budoucího vývoje. Firmy, co umí lépe používat tyto nástroje, mohou lépe reagovat na měnící se potřeby zákazníků. Zároveň mohou maximalizovat dopad jejich marketingových aktivit. [17]

1.4.2 GDPR problematika při analýze dat

Analýza zákaznických dat z e-shopů má velkou hodnotu pro cílenou reklamu, ale také nese mnoho právních a morálních otázek ohledně ochrany osobních údajů. Obecné nařízení o ochraně osobních údajů (GDPR), které začalo platit v roce 2018, zásadně změnilo přístup k používání osobních údajů v Evropě. A velmi ovlivnilo, jak mohou firmy používat zákaznická data. [18]

1.4.2.1 Základní principy GDPR relevantní pro analýzu dat

GDPR stanovuje zásadní pravidla, která je nutné dodržovat při práci s osobními údaji a při jejich využívání pro marketingové účely.

1. *Zákonnost a transparentnost*

Při práci s daty musí být právní základ, jako je souhlas osob, plnění smlouvy nebo oprávněný zájem správce. V kontextu online obchodů a cílených reklam se často používá souhlas nebo oprávněný zájem.

2. *Účelové omezení a minimalizace údajů*

Údaje mohou být sbírány jen pro jasné a vyjádřené cíle, a nemohou být zpracovávány způsobem, který není v souladu s těmito cíli. To znamená, že analýza dat od zákazníků například na reklamu musí být předem oznámena zákazníkům a je potřeba o tom informovat.

3. *Omezení uložení*

Osobní údaje by měly být uloženy jen po dobu nutnou k dosažení cíle zpracování, a musí být chráněny před neoprávněným nebo nezákonným zacházením.

4. *Anonymizace a pseudonymizace dat*

Jedním z hlavních způsobů, jak analyzovat data a chránit soukromí je anonymizace a pseudonymizace. Anonymizovaná data už nejsou považována za osobní údaje, zatímco pseudonymizovaná stále patří pod pravidla GDPR, ale dávají nějakou ochranu.

5. *Právo být zapomenut*

Za určitých podmínek má člověk právo požádat o smazání svých osobních dat. Například, když už nejsou potřebné pro původní cíl zpracování. Avšak firma není povinna této žádosti vyhovět ve všech situacích. [18][19]

1.4.2.2 Rizika v případě nedodržení pravidel

Při nedodržení pravidel je možné dostat pokutu, která je závislá na vážnosti provinění (Obrázek 4). Může dosahovat až stamilionů korun. Nejvyšší pokuta, která s GDPR zatím v české republice byla vydána činila 351 milionů Kč pro společnost Avast Software s.r.o. Společnost uváděla, že používá anonymizaci dat, ale prokázalo se, že údaje anonymizovány nebyly a mohlo dojít k identifikaci části osob. Zároveň společnost mylně uváděla, k čemu jsou tato data využívána. [20]



Obrázek 4: Rizika v případě Nedodržení pravidel [19]

S narůstajícím používáním umělé inteligence a strojového učení při zkoumání dat zákazníků se objevují nové problémy ohledně transparentnosti algoritmů, vysvětlitelnosti rozhodnutí a potenciální diskriminace skupin zákazníků. [18]

1.4.3 Proces analýzy dat pro cílenou reklamu

Analýza dat v kontextu cílené reklamy obsahuje pár důležitých kroků, které zajišťují dobré použití informací, co máme. První krok je sběr správných dat, která můžou pocházet z interních databází jako historie nákupů zákazníků nebo z externích zdrojů jak sociální sítě či sledování chování lidí na internetu. Kvalita získaných dat je důležitá, a proto následuje fáze jejich čištění co zajistí odstranění chybné nebo neúplné záznamy. Další krok je zkoumání dat. V této fázi se najdou vzory a trendy, které mohou být skryté v analýze souboru. Často se při tomto zkoumání používají techniky vizualizace dat, co pomáhají odhalit vztahy mezi proměnnými. Například pomocí metod jako je analýza hlavní prvků (Principal Component Analysis, PCA) lze dostat hlubší pohled do struktury dat. Tato metoda dovoluje snížit počet rozměrů datového souboru což znamená, že data se zjednoduší a stanou se přehlednějšími. Ale zachovává všechny důležité vlastnosti a informace, co jsou v původních datech. [17][21]

Díky tomu je snadnější pracovat s daty a dělat další analýzy bez ztráty jejich významu. Když je soubor s daty moc složitý, může být těžké ho analyzovat nebo vizualizovat. Tato metoda pomáhá najít a vybrat hlavní rysy dat, které jsou nejvíce důležité, a málo důležité informace odstraní nebo spojí. Díky tomu je práce s daty jasnější a výsledků analýz jsou často jednodušší na porozumění. Každý z těchto kroků pomáhá vytvářet jasnější obraz o tom, jak se zákazníci chovají a co mají rádi. To dělá cílení reklamních zpráv lepším. Účinný způsob analýzy dat zajistí, že výsledky mohou být snadno použity při rozhodování strategie a zlepšení marketingových kampaní. [17][13]

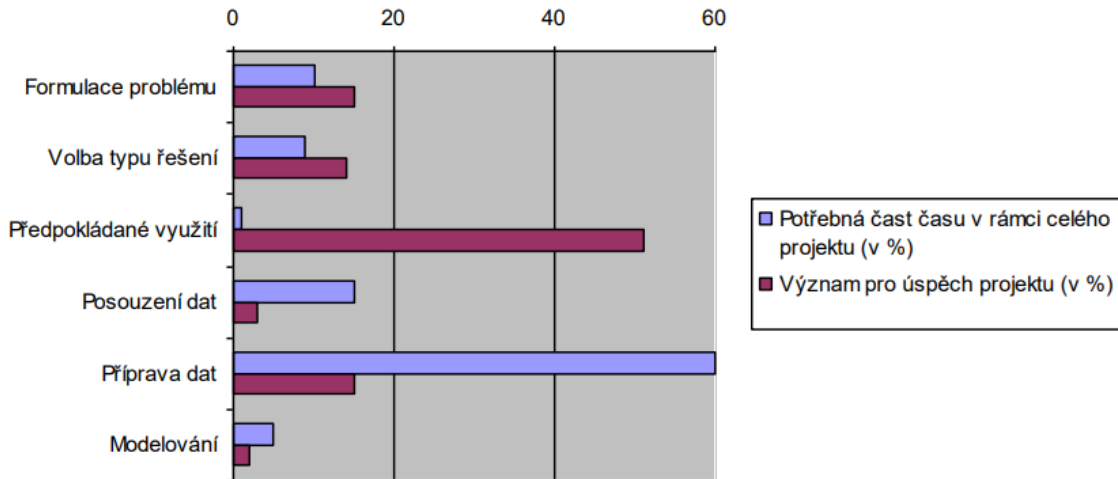
1.4.4 Předzpracování dat

Kvalitní data jsou důležitá pro jejich správné a účelné použití v každé analýze. Kvalita dat závisí na několika hlavních věcech, jako je přesnost, úplnost, konzistence, aktuálnost, důvěryhodnost a srozumitelnost.

Přesnost dat ukazuje, jak moc odpovídají pravdě. Chybná data mohou vzniknout z chyb při sběru nebo zpracování informací. Úplnost dat se týká toho, jestli jsou všechny potřebné atributy a hodnoty. Chybějící data mohou být třeba kvůli opomenutí nebo záznamům, které nebyly brány za důležité v době sběru. Konzistence dat znamená, že údaje nemají rozporuplné nebo nesouladné hodnoty jako v různých kódech či formátech. [13]

Jiný důležitý prvek, aktuálnost, mluví o tom, jestli data ukazují pravdivou situaci v reálném čase. Zpoždění s obnovou dat může špatně ovlivnit jejich použitelnost i když jsou později správná. Důvěryhodnost dat ukazuje, do jaké míry jsou uživateli považována za spolehlivá. To může být ovlivněno dřívějšími omyly nebo potížemi, i když data je formálně správná. Porozumění pak určuje, jak lehce lze data chápat a používat, na příklad ve vztahu k použitým kódům a názvoslovím. Kvalita informací je velmi závislá na tom, jak se mají používat. Různí lidé mohou mít jiné nároky na stejná data. Například sada dat, která je pro jeden cíl dost dobrá, může být pro jiný cíl nevyhovující. Při posuzování kvality je důležité vzít v úvahu okolnosti konkrétních požadavků a potřeb uživatele. Nízká kvalita dat může mít různé důvody. Patří sem chyby při vkládání, technologické limity, nesrovnalosti ve formátech nebo názvech, chybějící hodnoty, chybně zaznamenané údaje nebo neúplná historie úprav dat. Tyto problémy je často nutné najít a odstranit přes procesy zpracování dat jako čištění dat, doplnění chybějících hodnot nebo standardizace formátu. [13]

Příprava dat je často podceňovaná fáze, která má velký vliv na úspěšnost projektu. Jedná se o proces, který může být časově velmi náročný, protože zahrnuje několik kroků, aby byla použitelná pro následné analýzy a modelování. Přestože tato fáze může trvat dlouho, je možné ji provádět opakovaně (Obrázek 5). [22]



Obrázek 5: Časové nároky jednotlivých kroků při analýze dat [22]

1.4.4.1 Hlavní úkoly v předzpracování dat

Před chystání dat má několik důležitých kroků, mezi které spadá:

- čištění dat
- integrace dat
- redukce dat
- transformace dat

Tyto kroky jsou nezbytné pro zajištění kvality dat a jejich efektivní využití při analýzách a data miningu. [13]

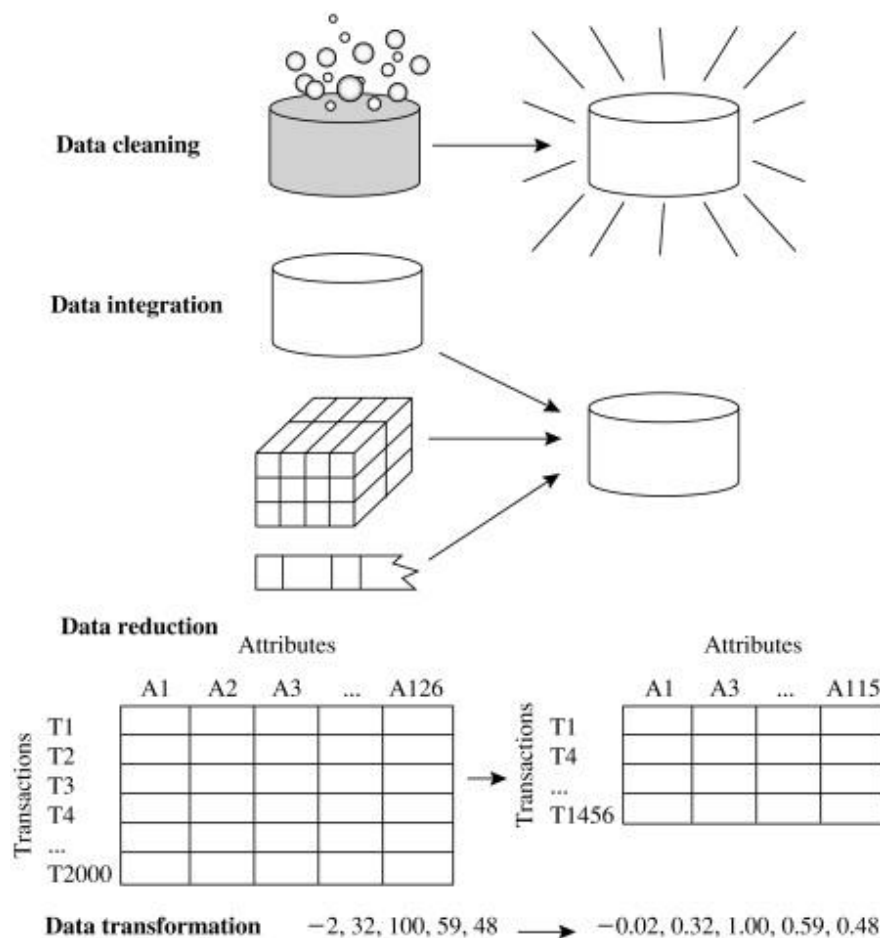
Čištění dat se soustředí na odstranění problémů v datech tím, že doplňuje chybějící hodnoty, vyhlazuje šumová data, identifikuje nebo odstraňuje odlehlé hodnoty a řeší nekonzistence. Když jsou data považována za "špinavá", lidé jim obvykle nevěří a výsledky dalších analýz mohou být nepřesné nebo nespolehlivé. I když některé metody pro analýzy a data mining obsahují způsoby jak pracovat s neúplnými nebo šumovými daty, nemusejí být vždy dost silné. Proto je čištění dat důležitým krokem, který předchází analýzám. [13][23]

Integrace dat je spojení informací z různých míst, jako jsou tabulky, soubory nebo skupiny dat. Tenhle postup může mít potíže s rozdíly. Například když různé databáze používají odlišné názvy pro stejné atributy (např. "customer_id" a "zakaznicke_id") nebo různé hodnoty u stejných atributů (např. "Jan", "Honza"). Redundantní nebo nekonzistentní data mohou zpomalit nebo narušit proces objevování znalostí. Aby se předešlo těmto problémům, je integrace dat často doprovázena dalšími procedurami pro odstranění redundance a sjednocení informací. [13]

Redukce dat pomáhá zmenšit rozsah datového souboru bez toho, aby došlo ke ztrátě důležitých informací pro analýzu. Tato redukce může zahrnovat techniky snižování dimenzionality, například kompresi dat, analýzu hlavních komponent (PCA), výběr podmnožin atributů nebo konstrukci nových atributů. Alternativně může být použita numerická redukce, která nahrazuje

původní data menšími reprezentacemi, jako jsou regresní modely, histogramy, klastrování nebo vzorkování. Díky těmto metodám je možné pracovat s menším množstvím dat při zachování analytické přesnosti. [13]

Transformace dat zahrnuje úkony, jako je normalizace, diskretizace a vytváření konceptuálních hierarchií. Normalizace mění hodnoty atributů do menšího rozmezí, což umožňuje, aby různé atributy (např. věk a roční příjem) měly při analýze stejnou váhu. Diskretizace nahrazuje původní hodnoty atributů kategoriemi nebo širšími koncepty, jako například přiřazením věkových skupin „mladý“, „dospělý“ a „senior“. Konceptuální hierarchie umožňují analyzovat data na různých úrovních abstrakce, což je přínosné při složitějších analýzách (Obrázek 6). [13]



Obrázek 6: funkce předzpracování dat [13]

Jednotlivé kroky předzpracování dat se často prolínají. Například odstranění redundantních dat může být považováno jak za součást čištění dat, tak za formu redukce dat. Tyto kroky jsou klíčové pro přípravu dat k dalším analytickým procesům a zajištění kvality a spolehlivosti výsledků.

1.4.4.2 Chybějící hodnoty

Při analýze dat se často potkáme se situací, kdy některé atributy mají chybějící hodnoty. Tohle je problém, protože chybějící hodnota může špatně ovlivnit kvalitu analýzy a výstupy následného dolování informací. Je pár způsobů, jak s chybějícími hodnotami pracovat.

1. Ignorování záznamů s chybějícími hodnotami: Tato metoda je běžně používána v případě, že chybí třída v klasifikační úloze. Obvykle je to účinné jen tehdy, když záznam má hodně chybějících čísel. Nevýhoda tohoto způsobu je ztráta informací v jiných attributech záznamu, co by mohli být užitečné pro analýzu.
2. Ruční doplnění chybějících hodnot: Tento způsob může být dost pomalý a těžký, hlavně pokud pracujeme s velkými datovými soubory, co mají hodně chybějících hodnot.
3. Nahrazení chybějící hodnoty globální konstantou: V tomto případě jsou všechny chybějící hodnoty atributu nahrazené stejnou konstantou třeba „Neznámý“. Tento přístup je jednoduchý, ale má limity a může vést ke špatné čtení v dolovacím procesu, protože všechny chybějící hodnoty budou myšleny jako jeden koncept.
4. Použití míry centrální tendence: Chybějící hodnoty se dají nahradit průměrnou nebo mediánovou hodnotou vlastnosti. Symetrické rozložení dat dovoluje použití průměru, zatímco u nesymetrických dat je lépe použít medián. Tento způsob pomáhá udržet základní znaky dat.
5. Nahrazení hodnoty podle příslušné třídy Chybějící hodnoty lze zaměnit průměrem nebo mediánem hodnot stejného atributu v rámci dané třídy Například při třídění zákazníků podle úvěrového rizika můžeme vzít průměrný příjem ve správné rizikové skupině. U nesymetrických distribucí je opět vhodnější medián.
6. Použití nejpravděpodobnější hodnoty: Tento přístup využívá statistické modely, jako jsou regrese, Bayesovské metody nebo indukce rozhodovacích stromů, k předpovědi chybějících hodnot na základě ostatních atributů. Tento přístup je oblíbený, protože bere v úvahu více informací a pomáhá zachovat vztahy mezi atributy. [13][23]

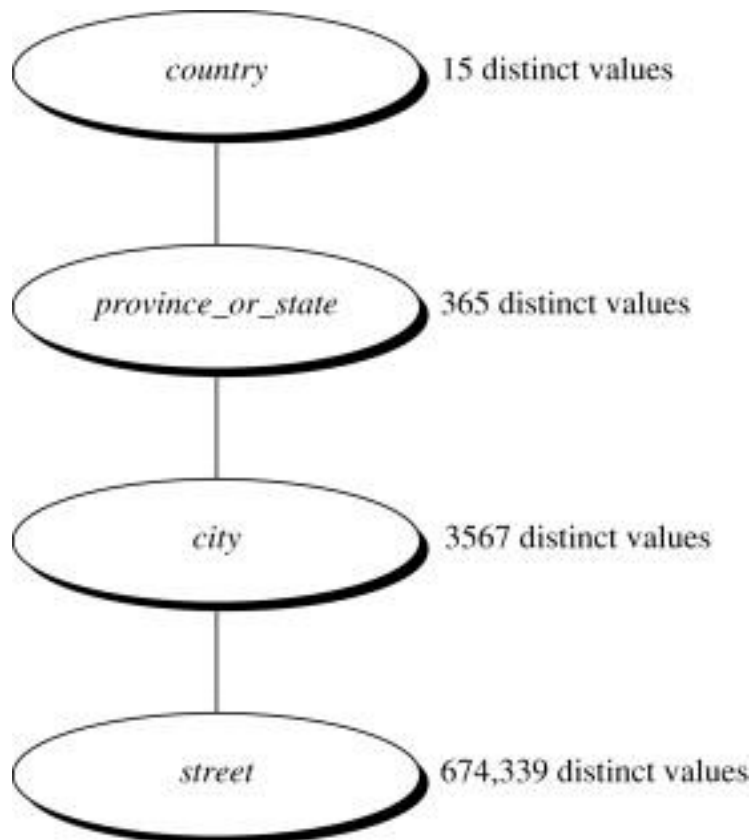
Je však důležité si uvědomit, že vyplňování chybějících hodnot vždy zavádí určitou míru zkreslení. I přesto může být zvolený přístup dostatečně přesný a užitečný.

V některých případech chybějící hodnota nemusí znamenat chybu v datech. Může to být úmyslně prázdné místo, například kdy si zákazník vytvoří kartu na prodejně a nechce uvádět, kde bydlí. Dobré návrhy formulářů by měly mít možnost označit tyto hodnoty jako „neaplikovatelné“. Systémy pro záznam mohou také automaticky najít další typy prázdných hodnot, jako „nevím“ nebo „žádné“. Je ideální, aby každý atribut měl jasná pravidla ohledně zpracování nulových hodnot. Dobře navržené metody sběru a zadávání dat mohou hodně snížit množství chybějících hodnot. [13]

1.4.4.3 Generování hierarchií

Transformace dat v případě nominálních atributů často zahrnuje vytváření hierarchií konceptů. Nominální atributy mají omezený (ale potenciálně velký) počet různých hodnot, které nejsou mezi sebou uspořádané. Mezi jmenné znaky patří například místo, druh práce nebo typ věci. Ruční určení hierarchií může být pro uživatele nebo odborníka na danou doménu časově náročná a složitá. Naštěstí mnoho hierarchií je nepřimo zahrnuto v databázovém schématu a mohou být automaticky generovány na úrovni definice schématu. Tyto uspořádání lze využít k přetváření dat na různé detailní úrovni. Třeba během zkoumání prodeje mohou vzory těžení dat ukázat spojení na úrovni specifických oblastí nebo států, stejně tak i na úrovni jednotlivých

poboček. Hierarchie konceptů u nominálních atributů obvykle zahrnuje skupinu atributů. Odborník nebo uživatel může snadno definovat hierarchii specifikováním částečného nebo úplného uspořádání atributů. V relační databázi mohou atributy zahrnovat „ulici“, „město“, „stát“ a „zemi“. Na úrovni schématu lze definovat hierarchii specifikováním uspořádání, například: ulice < město < stát < země (Obrázek 7). [13][23]



Obrázek 7: Hierarchie místa bydliště s počtem hodnot [13]

1.4.5 Shluková analýza a segmentace zákazníků

Shluková analýza patří mezi nejčastěji využívané nástroje v oblasti analýzy dat pro účely cílené reklamy, a to především díky své schopnosti segmentovat zákazníky. Tato technika umožňuje rozdělit zákazníky do homogenních skupin na základě jejich podobností, což zajišťuje lepší přizpůsobení marketingových kampaní jednotlivým segmentům. Efektivní segmentace zvyšuje pravděpodobnost úspěchu kampaní a přispívá k efektivnějšímu využití dostupných zdrojů. [17]

Jednou z nejběžnějších metod shlukové analýzy je metoda kmeans clustering, která se často používá k rozdělení zákazníků na základě jejich nákupního chování. Tato metoda uspořádá zákazníky podle blízkosti jejich charakteristik do předem definovaného počtu skupin. Díky tomu lze identifikovat odlišné nákupní vzorce a lépe porozumět preferencím jednotlivých segmentů. Kromě kmeans clusteringu existují i další přístupy k segmentaci zákazníků. Dvoufázová klasifikace umožňuje analyzovat větší objemy dat a přizpůsobit proces segmentace složitějším datovým strukturám. Kohonenovy mapy představují jiný typ analýzy, který poskytuje alternativní pohled na data a zohledňuje i nelineární vztahy mezi proměnnými. Tyto

metody umožňují zkoumat strukturu dat a odhalovat nové souvislosti, které mohou být pro marketingové strategie velmi cenné. Na rozdíl od klasifikace, která patří mezi metody supervised learning a vyžaduje trénovací data s předem označenými třídami, shlukování (clustering) je metoda unsupervised learning, která nepotřebuje předem definované kategorie. Zatímco klasifikace přiřazuje objekty do již známých tříd, shlukování objevuje přirozené skupiny objektů na základě jejich podobnosti bez předchozí znalosti struktury dat. [17][23]

Aplikace shlukovací analýzy v cílené reklamě výrazně usnadňuje pochopení potřeb a chování různorodých skupin zákazníků. Tímto způsobem mohou firmy vytvářet personalizované kampaně, které lépe odpovídají očekáváním cílových skupin, což vede k efektivnějšímu oslovování zákazníků a lepšímu dosažení obchodních cílů. [17]

1.4.6 Analýza následujících produktů (cross-selling)

Cross-selling představuje strategii, kdy se prodejci snaží zákazníkům nabídnout doplňkové produkty nebo služby k již zakoupenému nebo zamýšlenému nákupu. Tato marketingová technika je založena na předpokladu, že zákazník, který již projevil zájem o určitý produkt, může být nakloněn k zakoupení souvisejících nebo komplementárních položek. Z pohledu podniku jde o efektivní způsob, jak zvýšit tržby a zároveň uspokojit potřeby zákazníků poskytnutím komplexnějšího řešení. Identifikace vhodných produktů pro cross-selling vyžaduje systematickou analýzu nákupních dat a vzorců chování zákazníků. [24][25]

2 ANALYTICKÁ ČÁST

2.1 Představení společnosti

Název společnosti: BIKEMAX s.r.o.

Datum vzniku: 20. prosince 2012

Právní forma: Společnost s ručením omezeným

Sídlo: Rokytova 4259/1, 615 00 Brno – Židenice

Základní kapitál: 200 000 Kč

Společníci / Majitelé: Marek Doležal, Jakub Večeřa, Daniel Ballek

Společnost byla založena 20. prosince 2012 a sídlí v Brně, městské části Židenice, na adrese Rokytova 4259/1. Firma je vedena jako společnost s ručením omezeným s počátečním základním kapitálem ve výši 200 000 Kč. Podle dostupných informací jsou aktuálními společníky a skutečnými majiteli společnosti pánové Marek Doležal, Jakub Večeřa a Daniel Ballek.

Hlavní činností společnosti je prodej jízdních kol, elektrokol, cyklistického vybavení a doplňků prostřednictvím e-shopu a kamenných prodejen. Klade důraz na kvalitní zákaznický servis, odborné poradenství a širokou nabídku produktů.

2.2 Úvod do analytické části

V části vyhodnocování dat bude zaměřeno na data z e-shopu, která mají posloužit jako základ pro cílenou reklamu na internetu. Budou analyzována pomocí databázového jazyka SQL a programovacího jazyku Python. Konkrétně se zaměřím na historii nákupů, preferované kategorie, produkty kupované v jednom nákupu a na produkty kupované v následujících nákupech.

2.3 Metody analýzy

2.3.1 SQL dotazy

Structured Query Language (SQL) je nástrojem pro práci s relačními databázemi. V rámci této práce byly SQL dotazy použity k efektivnímu získávání, filtrování a manipulaci s daty. Pomocí operací, jako jsou SELECT, JOIN, GROUP BY a HAVING, bylo možné nejen získat relevantní datové sady pro další analýzu, ale také provádět základní analytické úkony přímo na úrovni databáze. Například pomocí agregačních funkcí, jako jsou COUNT, SUM, AVG, MIN a MAX, bylo možné vypočítat různé statistické ukazatele a shrnout rozsáhlé datové množiny do přehledných reportů.

2.3.2 Python

Python je univerzální programovací jazyk, který se ukázal jako velmi užitečný pro datovou analýzu. V této práci byly využity knihovny jako Pandas, NumPy a Matplotlib pro předzpracování dat, jejich analýzu a vizualizaci. Kromě toho byla využita knihovna MLExtend pro pokročilé techniky data miningu. MLExtend poskytuje množství metod pro analýzu dat, například algoritmy pro hledání asociací pravidel, klasifikační modely, metody shlukování a další techniky. Ve spojení s Pythonem umožňuje tato knihovna aplikaci algoritmů, jako je Apriori nebo FP-Growth, které odhalují skryté vazby a vzory v datových množinách. Tyto techniky byly použity ke zpracování velkých datových objemů, jejich analýze a identifikaci důležitých trendů či vztahů mezi datovými atributy.

2.4 Popis datového souboru

Tato kapitola se bude zabývat původními datovými soubory, které byli vyexportovány z databáze e-shopu a jejich následnou úpravou do formy použitelné a vhodné pro datovou analýzu.

2.4.1 Data kategorií

Data kategorií produktů byla exportována ve formátu, který byl již vhodný pro analýzu, a pro základní potřeby se nemusela nijak upravovat (Tabulka 1).

Tabulka 1: Ukázka z datového souboru kategorií produktů

ID skupiny	Název	ID nadřizené
1	Vše	
276	Adaptéry	96
2	Auto nosiče	20
148	Brzdové adaptéry	146

2.4.2 Data prodeje

Data prodeje byli kvůli kvantitě dat exportována pro každá rok zvlášť (Tabulka 2 a Tabulka 3). Jelikož datový soubor obsahoval několik pro mnou prováděnou analýzu nepotřebných sloupců, tak musela být data předpřipravena pro použití při analýze. Nejprve byly spojeny všechny soubory prodeje ve složce do jednoho pomocí Python kódu. A následně vytvořen nový datový soubor produktů, na který se bude následně odkazovat, kvůli ušetření paměti při analýzách (Tabulka 4). V tabulce nákupů, kde jsou ponechány neduplicitní čísla dokladů, datum, kdy byl prodej uskutečněn a ID zákazníka (Tabulka 5). A chyběla jen data položek v nákupech, kde byly nahrazeny názvy produktů jejich číselnými alternativami (Tabulka 6).

```

1. import csv
2. import os
3.
4. def combine_csv_files_csvlib(folder_path, output_file):
5.     csv_files = [file for file in os.listdir(folder_path) if file.endswith('.csv')]
6.
7.     with open(output_file, mode='w', encoding='utf-8', newline='') as outfile:
8.         writer = None
9.         for file in csv_files:
10.            with open(os.path.join(folder_path, file), encoding='utf-8') as infile:
11.                reader = csv.reader(infile, delimiter=';')
12.                if writer is None:
13.                    writer = csv.writer(outfile, delimiter=';')
14.                    writer.writerow(next(reader))
15.                for row in reader:
16.                    writer.writerow(row)
17.
18. combine_csv_files_csvlib("C://Users//Path//To//Folder", "prodeje.csv")

```

Tabulka 2: První polovina ukázky datového souboru exportovaných prodejů

doklad	polozka	datum_prodeje	id_zakaznika	id_skupiny	skupina	typ	id_karty	sklad
101234567	1	22.07.2013	1	93	SLU	30	1,0014	-
101234567	2	27.07.2013		110	ZAM	0	1,0015	101K
101234568	1	13.09.2024	45	26	GRI	0	1,0041	101Z

Tabulka 3: Druhá polovina ukázky datového souboru exportovaných prodejů

karta	nazev	atribut_1	atribut_2	Mena	mnozstvi_evid [Ks]	id_mj_evid	Cena [Kč]	cena_cel [Kč]
SLU000025	Košík	Bílá/červená	boční	CZK	1	KS	49	49
ZAM000012	Zámek			CZK	1	KS	56	56
GRI000089	Gripy	Černá		CZK	1	KS	59	59

Tabulka 4: Datový soubor produktů

id_produkту	nazev	id_skupiny
1	Košík	93
2	Zámek	110
3	Gripy	26

Tabulka 5: Datový soubor prodejů

doklad	datum_prodeje	id_zakaznika
101234567	22.07.2013	1
101234567	27.07.2013	
101234568	13.09.2024	45

Tabulka 6: Datový soubor položek v nákupech

doklad	polozka	Mnozstvi [Ks]	cena [Kč]	id_produkту
101234567	1	1	49	1
101234567	2	1	56	2
101234568	1	1	59	3

2.4.3 Data zákazníků

Data byla dodána anonymizovaná, se sloupcem dodavatelé, který určuje velkoodběratele, které byly z datového souboru vyloučeny kvůli zkreslení výsledků analýz. Data byla dodána v následujícím formátu, která obsahovala ještě město, ale to je zadáváno ručně zákazníkem

a obsahovalo výrazně větší chybovost než PSČ, tudíž bylo smysluplnější ponechat pouze PSČ (Tabulka 7).

Tabulka 7: Datový soubor Zákazníků

ID zákazníka	Dodavatel	PSČ
105	Ne	
104	Ne	54351
103	Ne	75661

2.4.4 Data PSČ

Většina Zákaznických dat obsahuje poštovní směrovací číslo (PSČ) pro lokalizaci zákazníků, nebo PSČ bylo dodáno podle města zákazníka v základním datovém souboru, pokud bylo dostupné. Následně byla stáhnuta volně dostupná data o PSČ ze stránek české pošty. Zde byly ještě kódy obcí/okresů/krajů, které byly odstraněny, takže po smazání přebytečných dat byl vytvořen datový soubor, použitelný pro účely analýzy (Tabulka 8). [26]

Tabulka 8: Datový soubor PSČ lokalit

psc	nazobce	nazokresu	nazevkraj
10000	Praha	Hlavní město Praha	Hlavní město Praha
10100	Praha	Hlavní město Praha	Hlavní město Praha
60200	Brno	Brno-město	Jihomoravský

2.4.5 SQL databáze

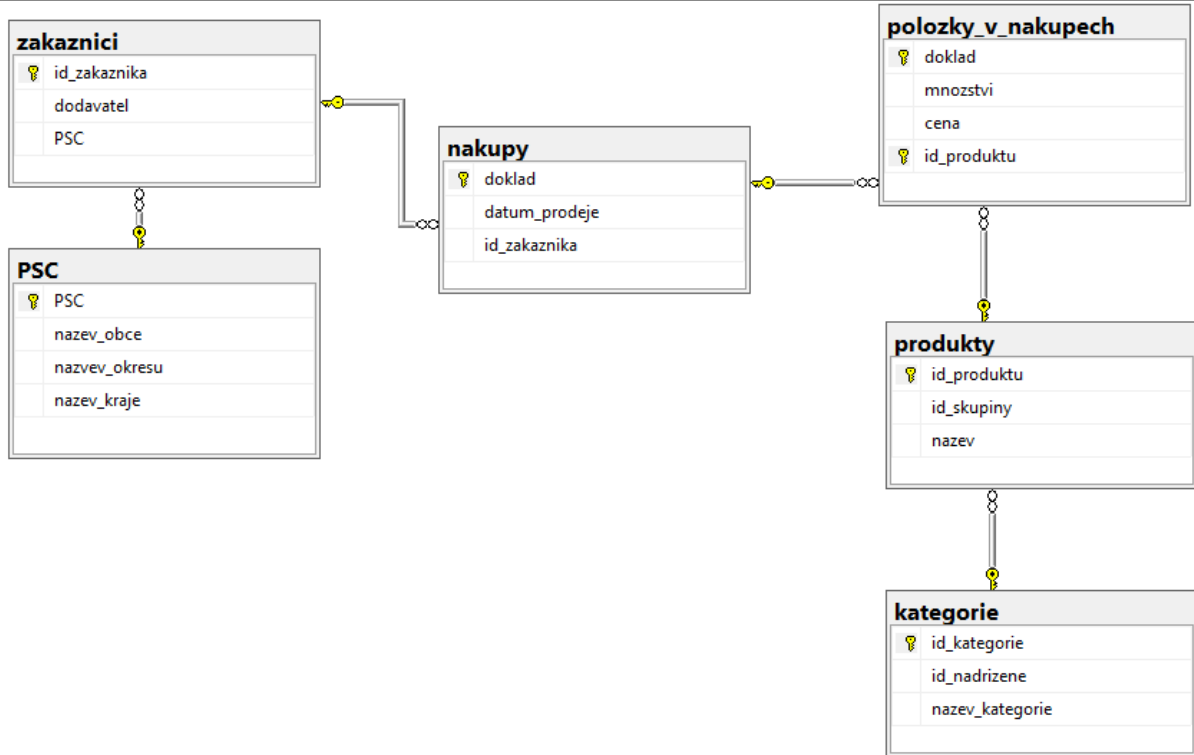
Pro prvotní analýzu a snadné připojení k ostatním nástrojům, které budou použity, byla zvolena možnost vytvoření si lokálního SQL serveru a následně se k němu připojovat pro získání dat. SQL kód použitý pro vytvoření tabulek v databázi a jejich diagram (Obrázek 8).

```
1. create table zakaznici(  
2.     id_zakaznika int primary key,  
3.     dodavatel varchar (10),  
4.     sidlo varchar (10),  
5.     PSC int references PSC(PSC) references PSC(PSC),  
6.     pocet_nakupu int  
7. );  
8. create table PSC(  
9.     PSC int primary key,  
10.    nazev_obce varchar (50),  
11.    nazvev_okresu varchar (50) null,  
12.    nazev_kraje varchar(50) null,  
13. );  
14.  
15. create table kategorie(  
16.     id_kategorie int primary key,  
17.     id_nadrizene int null,  
18.     nazev_kategorie varchar (100)  
19. );  
20. create table nakupy_bez_odberatelu(  
21.     doklad nvarchar(50) primary key,  
22.     den_prodeje tinyint,  
23.     mesic_prodeje tinyint,  
24.     rok_prodeje smallint,  
25.     datum_prodeje date,  
26.     id_partnera int references Zakaznici(id_zakaznika)
```

```

27. );
28. create table polozky_v_nakupech(
29.     doklad nvarchar(50) references nakupy_bez_odberatelu(doklad),
30.     id_produkty int,
31.     id_skupiny int
32.     primary key(doklad, id_produkty)
33. );
34. create table produkty(
35.     id_produktu int primary key,
36.     nazev varchar (100)
37. );

```



Obrázek 8: Diagram SQL databáze

2.5 Analýza dat

2.5.1 Analýza počtu prodejů v jednotlivých kategoriích

Nejprve byla provedena analýza kategorií, které jsou nejvíce kupované, aby mohla sloužit při dalších analýzách a případně zjistit, jestli má cenu se výsledky analýzy zabývat vůči přínosům, které pro společnost mají. Analýza byla provedena v Pythonu a vykreslena do grafu, který je snadno přehledný.

```
1. import pyodbc
2. import pandas as pd
3. import plotly.express as px
4.
5. server = r'localhost'
6. database = 'DB'
7.
8.         connection_string = f'DRIVER={{SQL
Server}};SERVER={server};DATABASE={database};Trusted_Connection=yes;'
9.
10. connection = pyodbc.connect(connection_string)
11.
12. query = """
13. SELECT
14.     id_kategorie,
15.     id_nadrizene,
16.     nazev_kategorie,
17.     pocet_prodeju
18. FROM SalesByCategory
19. """
20. df = pd.read_sql_query(query, connection)
21.
22. connection.close()
23.
24. def calculate_sales_with_hierarchy(df):
25.     sales_dict = df.set_index('id_kategorie')['pocet_prodeju'].to_dict()
26.     parent_dict = df.set_index('id_kategorie')['id_nadrizene'].to_dict()
27.
28.     def get_total_sales(category_id):
29.         total_sales = sales_dict.get(category_id, 0)
30.         for child_id, parent_id in parent_dict.items():
31.             if parent_id == category_id:
32.                 total_sales += get_total_sales(child_id)
33.         return total_sales
34.
35.     df['total_sales'] = df['id_kategorie'].apply(get_total_sales)
36.     return df
37.
38. df = calculate_sales_with_hierarchy(df)
39.
40. fig = px.treemap(
41.     df,
42.     path=['nazev_kategorie'],
43.     values='total_sales',
44. )
45.
46. fig.show()
```

Graf obrazuje, které kategorie jsou nejvíce kupované a nejvíce se podílejí na nákupech zákazníků. Služby se podílejí méně než zboží a jízdní kola více než elektrokola (Obrázek 9).



Obrázek 9: Graf podílů kategorií na počtu produktů v nákupech

2.5.2 Analýza prodejů v jednotlivých kategoriích podle celkové ceny

Zde byl upraven již použitý kód, aby se zjistili informace ohledně toho, jak moc se konkrétní kategorie podílejí na celkovém prodeji podle jejich ceny. Změny v Python kódu:

```

1. query = """
2. SELECT
3.     k.id_kategorie,
4.     k.id_nadrizene,
5.     k.nazev_kategorie,
6.     COALESCE(SUM(kc.cena), 0) AS total_cena
7. FROM
8.     kategorie k
9. LEFT JOIN
10.    kategorie_cena kc
11. ON
12.    kc.id_skupiny = k.id_kategorie
13. GROUP BY
14.    k.id_kategorie, k.id_nadrizene, k.nazev_kategorie
15. ORDER BY
16.    k.id_kategorie;
17. """

1. sales_dict = df.set_index('id_kategorie')['total_cena'].to_dict()

```

Graf ukazuje, že některé méně kupované kategorie počtem mohou být zásadnější pro obchod, jelikož obsahují průměrně dražší položky. Například kategorie Elektrokola a Jízdní kola (Obrázek 10).



Obrázek 10: Graf podílů kategorií na ceně produktů v nákupech

2.5.3 Analýza nákupních košíků

Zde bylo zaměřeno na položky a kategorie často kupované dohromady v posledních pěti letech, aby analýza nebyla ovlivněna starými a již neprodávanými produkty. Zároveň byly odstraněny produkty, které představují servisní zakázky, jelikož jsou pro analýzu nepotřebné.

```

1. import pandas as pd
2. from mlxtend.frequent_patterns import association_rules
3. from mlxtend.preprocessing import TransactionEncoder
4. from mlxtend.frequent_patterns import fpgrowth
5.
6. df = pd.read_csv("nakupy.csv", sep=";")
7. transactions = df.groupby("doklad")["id_kategorie"].apply(list).tolist()
8. #Pro produkty:
9. #transactions = df.groupby("doklad")["id_produkty"].apply(list).tolist()
10.
11. te = TransactionEncoder()
12. te_ary = te.fit(transactions).transform(transactions)
13. df_transformed = pd.DataFrame(te_ary, columns=te.columns_)
14.
15. freq_items = fpgrowth(df_transformed, min_support=0.003, use_colnames=True)
16. rules = association_rules(freq_items, metric="confidence", min_threshold=0.01)
17. rules = rules[rules["lift"] > 1]
18.
19. rules.to_csv("asociace_skupin.csv", index=False)
20. df_transformed.to_csv("binarni_format.csv", index=False)

```

Výsledky byly použity s knihovnami matplotlib a networkx v Pythonu pro vizualizaci nejvýraznějších vazeb mezi produkty/kategoriemi.

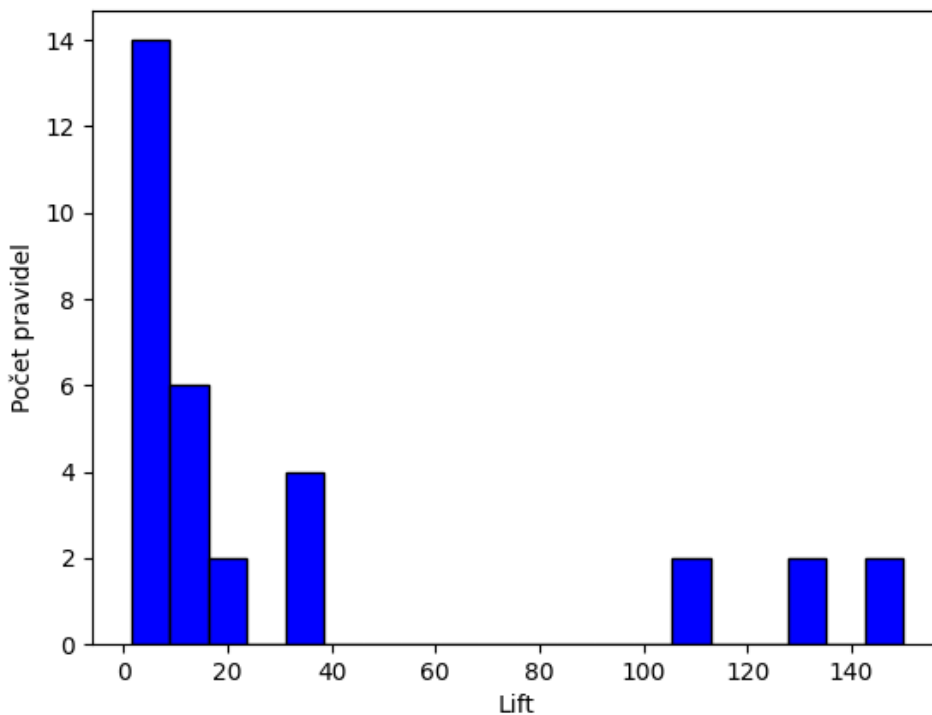
Matplotlib je komplexní knihovna pro tvorbu statických, animovaných a interaktivních vizualizací v Pythonu. Umožňuje snadnou tvorbu publikovatelných grafů, interaktivních prvků pro přibližování a posouvání, přizpůsobení vizuálního stylu a export do různých formátů. [27]

NetworkX je výkonná knihovna pro práci s grafy v Pythonu, která umožňuje jejich tvorbu, manipulaci a analýzu. Poskytuje širokou škálu algoritmů pro zpracování komplexních sítí, včetně hledání nejkratších cest, analýzy centrálnosti a generování různých typů grafů. Knihovna

podporuje práci s orientovanými i neorientovanými grafy a umožňuje přidávat k uzlům a hranám libovolná data. [28]

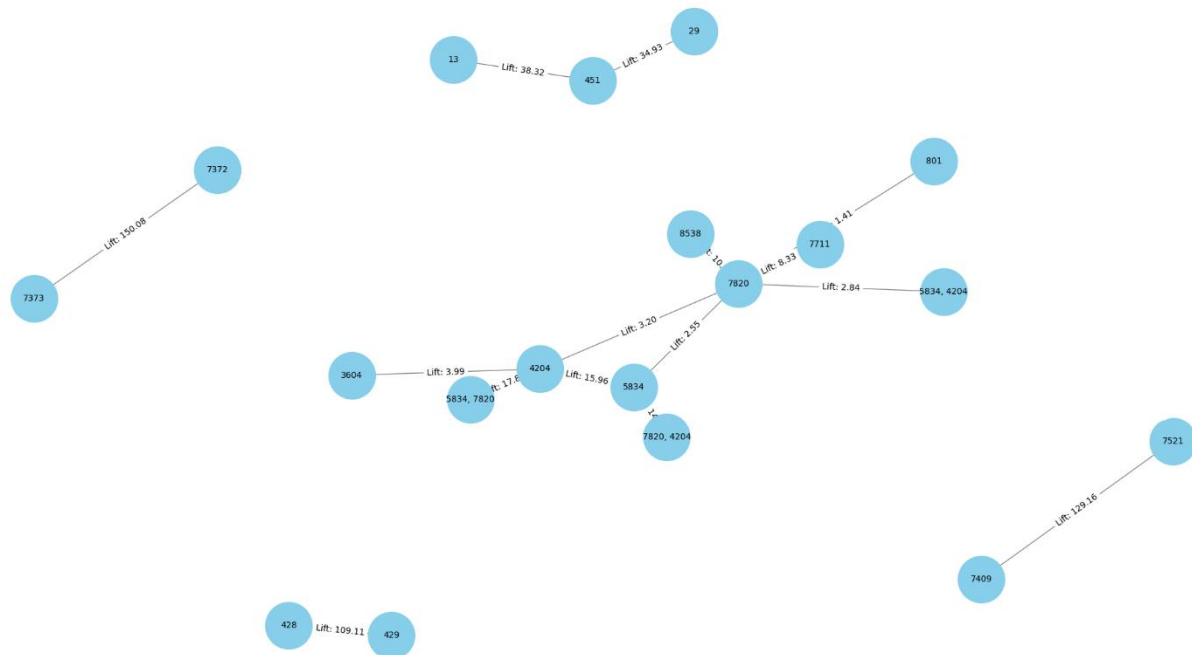
```
1. import pandas as pd
2. import matplotlib.pyplot as plt
3. import networkx as nx
4.
5. df = pd.read_csv("asociace_skupin.csv")
6.
7. df["antecedents"] = df["antecedents"].apply(lambda x: ", ".join(map(str, eval(x))))
8. df["consequents"] = df["consequents"].apply(lambda x: ", ".join(map(str, eval(x))))
9.
10. G = nx.Graph()
11. for _, row in df.iterrows():
12.     G.add_edge(row["antecedents"], row["consequents"], weight=row["lift"])
13.
14. plt.figure(figsize=(12, 8))
15. pos = nx.spring_layout(G, seed=42)
16. nx.draw(G, pos, with_labels=True, node_size=3000, node_color="skyblue", edge_color="gray",
17. font_size=10)
17. edge_labels = {(row["antecedents"], row["consequents"]): f"Lift: {row['lift']:.2f}" for _, row
18. in df.iterrows()}
18. nx.draw_networkx_edge_labels(G, pos, edge_labels=edge_labels, font_size=10)
19. plt.title("Síťová vizualizace asocičních pravidel")
20. plt.show()
21.
22. df["lift"].plot(kind="hist", bins=20, color="blue", edgecolor="black")
23. plt.xlabel("Lift")
24. plt.ylabel("Počet pravidel")
25. plt.title("Distribuce síly asociací mezi produkty")
26. plt.show()
```

V grafu je počet pravidel a jejich síla a je zde většina pravidel s menší silou, ale stále ukazující dostatečnou pravděpodobnost na hromadný nákup a následně 6, které mají velkou pravděpodobnost být koupené společně (Obrázek 11).



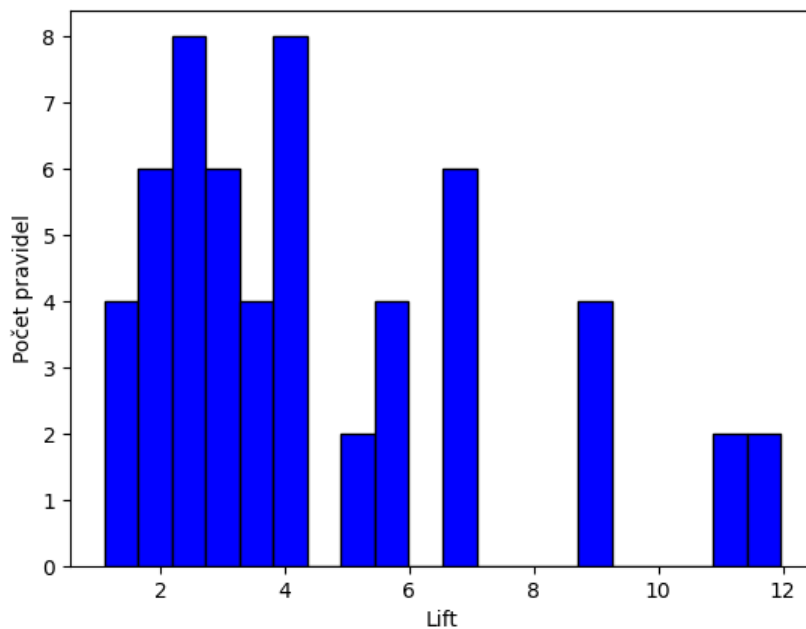
Obrázek 11: Distribuce síly asociací mezi produkty

V grafu je vidět, jak jsou společně produkty provázány a které spolu souvisejí více a které jsou osamocené dvojce nebo trojce produktů (Obrázek 12).



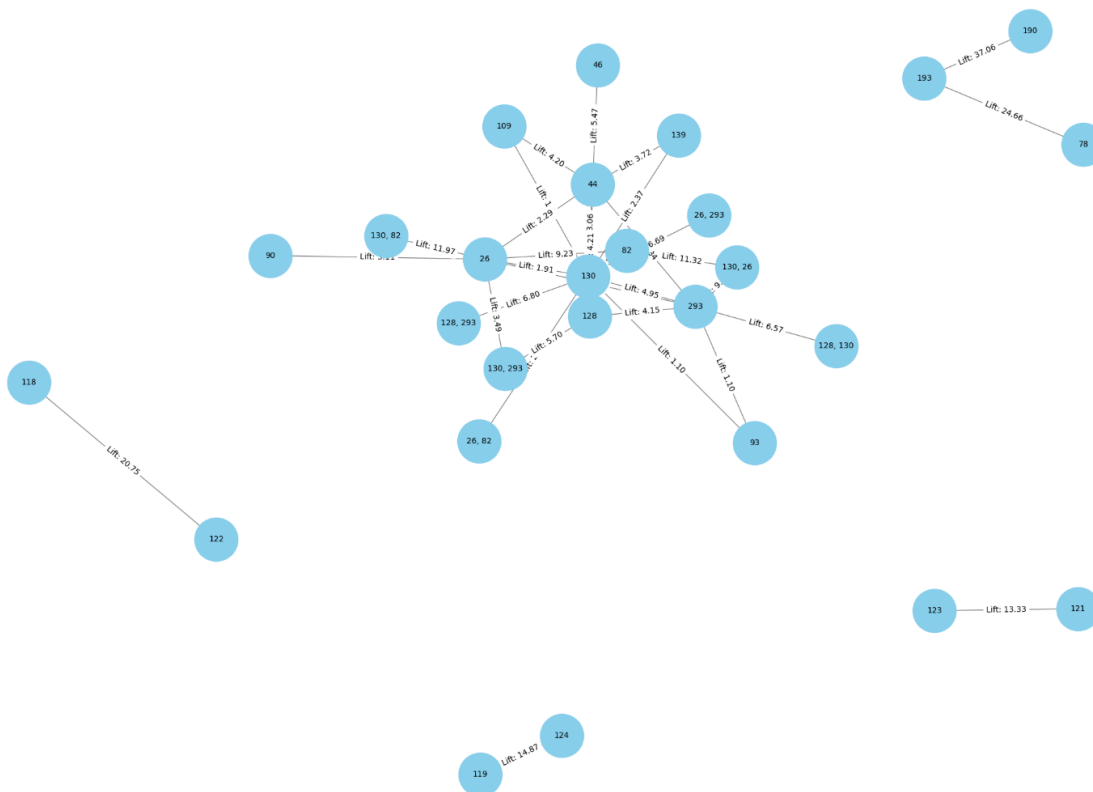
Obrázek 12: Síťový graf asociací mezi produkty

V grafu je počet pravidel a jejich síla a že s přibývajícím pravděpodobností výskytu společně v nákupu počet kategorií klesá (Obrázek 13).



Obrázek 13: Distribuce síly asociací mezi kategoriemi

V grafu je vidět, které kategorie jsou často kupované společně, které na sebe více navazují a které jsou více osamocené kategorie (Obrázek 14).



Obrázek 14: Síťový graf asociací mezi produkty

2.5.4 Analýza zákazníků

Zde byla analýza zaměřena na rozřazení zákazníků na základě jejich nákupního chování. K analýze byla využívána data o celkové ceně všech zakoupených produktů, průměrné ceně za produkt, počet nákupů, frekvenci nákupů a průměrný počet produktů v nákupu.

Kmeans je metoda shlukování dat, která se využívá pro rozdělení datových bodů do skupin na základě jejich podobnosti. V knihovně sklearn.cluster v Pythonu se tato metoda implementuje pomocí třídy KMeans, která umožňuje efektivní nalezení optimálního rozdělení dat. [29]

```

1. def apply_kmeans(metrics_df, num_clusters):
2.     features = ['total_spent', 'avg_price', 'num_purchases', 'purchase_frequency',
'avg_items_per_purchase']
3.     metrics_df_clean = metrics_df.dropna(subset=features)
4.     X = metrics_df_clean[features]
5.     scaler = StandardScaler()
6.     X_scaled = scaler.fit_transform(X)
7.
8.     kmeans = KMeans(n_clusters=num_clusters, random_state=42)
9.     metrics_df_clean['cluster'] = kmeans.fit_predict(X_scaled)
10.    metrics_df = pd.merge(metrics_df, metrics_df_clean[['id_partnera', 'cluster']],
11.                          on='id_partnera', how='left')
12.
13.    return metrics_df, kmeans, X_scaled, features, scaler
14.
15. def analyze_clusters(metrics_df):
16.    cluster_summary = metrics_df.groupby('cluster').agg({
17.        'id_partnera': 'count',
18.        'total_spent': ['mean', 'min', 'max'],
19.        'avg_price': ['mean', 'min', 'max'],

```

```

20.     'num_purchases': ['mean', 'min', 'max'],
21.     'purchase_frequency': ['mean', 'min', 'max'],
22.     'avg_items_per_purchase': ['mean', 'min', 'max']
23. })
24.
25.     return cluster_summary

```

2.5.5 Analýza následujících produktů a skupin

V této části bylo zaměřeno na nejtěžnější analýzu, která se zabývá návazností po sobě se prodávajících produktů, produktů dalších nakoupených produktů během celého zákaznického cyklu a návaznosti kategorií a jejich časové rozdíly.

Analýza začala SQL dotazem na navazující produkty a počtem, kolikrát se vyskytují.

```

1. SELECT
2.     z1.id_zakaznika,
3.     p1.id_produkту AS prvni_produkту,
4.     p2.id_produkту AS nasledny_produkту,
5.     COUNT(DISTINCT n1.doklad) AS pocet_vyskytu
6. FROM
7.     polozky_v_nakupech p1
8. JOIN
9.     nakupy n1 ON p1.doklad = n1.doklad
10. JOIN
11.     Zakaznici z1 ON n1.id_partnera = z1.id_zakaznika
12. JOIN
13.     nakupy n2 ON z1.id_zakaznika = n2.id_partnera
14. JOIN
15.     polozky_v_nakupech p2 ON n2.doklad = p2.doklad
16. WHERE
17.     n2.datum_prodeje > n1.datum_prodeje
18.     AND p1.id_produkту <> p2.id_produkту
19.     AND z1.id_zakaznika <> 0
20. GROUP BY
21.     z1.id_zakaznika, p1.id_produkту, p2.id_produkту
22. ORDER BY
23.     pocet_vyskytu DESC;

```

Kvůli potřebě jít více do hloubky byla použita následující funkce a Python knihovna mlxtend pro hlubší analýzu vzorců nakupování.

Mlxtend (Machine Learning Extensions) je knihovna pro Python, která poskytuje užitečné nástroje pro každodenní úlohy v oblasti datové vědy a strojového učení. Obsahuje rozšíření pro populární knihovny a nabízí širokou škálu funkcí, včetně ensemble metod, výběru rysů, asociačních pravidel. [30]

2.5.5.1 Nahrání dat z databáze

```

1. def load_data_from_db():
2.     try:
3.         import pyodbc
4.         server = r'localhost'
5.         database = 'DB'
6.
7.         connection_string=f'DRIVER={{SQLServer}};SERVER={server};DATABASE={database};Trusted_Connection=yes;'
8.         conn = pyodbc.connect(connection_string)
9.
10.        query = """
11.        SELECT
12.            z.id_zakaznika,
13.            n.datum_prodeje,

```

```

14.         p.id_produkту,
15.         p.id_skupiny
16.     FROM
17.         Zakaznici z
18.     JOIN
19.         nakupy_bez_odberatelu n ON z.id_zakaznika = n.id_partnera
20.     JOIN
21.         polozky_v_nakupech p ON n.doklad = p.doklad
22.     ORDER BY
23.         z.id_zakaznika, n.datum_prodeje
24.     """
25.
26.     df = pd.read_sql(query, conn)
27.     conn.close()
28.     return df
29. except Exception as e:
30.     print(f"Chyba při připojení k databázi: {e}")

```

2.5.5.2 Vytvoření párů produktů, které byly zakoupeny po sobě jedním zákazníkem

```

1.     def create_sequence_pairs(self, max_time_diff=None, chunk_size=25,
max_products_per_customer=100, temp_file=None):
2.         print(f"Zpracovávám sekvence pro {self.data['id_zakaznika'].nunique()} zákazníků po
skupinách po {chunk_size}...")
3.         pair_counts = {}
4.         pair_time_diffs = {}
5.
6.         customers = self.data['id_zakaznika'].unique()
7.         total_customers = len(customers)
8.
9.         for i in range(0, total_customers, chunk_size):
10.            chunk_customers = customers[i:min(i+chunk_size, total_customers)]
11.            print(f"Zpracovávám zákazníky {i+1} až {min(i+chunk_size, total_customers)} z
{total_customers}")
12.
13.            chunk_data = self.data[self.data['id_zakaznika'].isin(chunk_customers)].copy()
14.
15.            for customer, group in chunk_data.groupby('id_zakaznika'):
16.                group = group.sort_values('datum_prodeje')
17.
18.                if len(group) > max_products_per_customer:
19.                    print(f"Zákazník {customer} má {len(group)} produktů, omezují na
{max_products_per_customer}")
20.                    group = group.head(max_products_per_customer)
21.
22.                purchases_by_date = {}
23.                for _, row in group.iterrows():
24.                    date = row['datum_prodeje']
25.                    product = row['id_produkту']
26.
27.                    if date not in purchases_by_date:
28.                        purchases_by_date[date] = []
29.
30.                    if product not in purchases_by_date[date]:
31.                        purchases_by_date[date].append(product)
32.
33.                sorted_dates = sorted(purchases_by_date.keys())
34.
35.                for date_idx in range(len(sorted_dates) - 1):
36.                    date1 = sorted_dates[date_idx]
37.                    products1 = purchases_by_date[date1]
38.
39.                    for next_date_idx in range(date_idx + 1, min(date_idx + 4, len(sorted_dates))):
40.                        date2 = sorted_dates[next_date_idx]
41.
42.                        time_diff = (date2 - date1).days
43.                        if max_time_diff is not None and time_diff > max_time_diff:
44.                            continue

```

```

44.
45.             products2 = purchases_by_date[date2]
46.
47.             max_products = 10
48.             prods1 = products1[:max_products] if len(products1) > max_products
else products1
49.             prods2 = products2[:max_products] if len(products2) > max_products
else products2
50.
51.             for prod1 in prods1:
52.                 for prod2 in prods2:
53.                     if prod1 != prod2:
54.                         pair = (prod1, prod2)
55.
56.                         if pair not in pair_counts:
57.                             pair_counts[pair] = 0
58.                             pair_time_diffs[pair] = []
59.
60.                             pair_counts[pair] += 1
61.                             pair_time_diffs[pair].append(time_diff)
62.
63.             print("Vytvářím výsledný DataFrame...")
64.             pairs_data = []
65.             for (prod1, prod2), freq in pair_counts.items():
66.                 avg_time_diff = sum(pair_time_diffs[(prod1, prod2)]) / len(pair_time_diffs[(prod1,
prod2)])
67.                 pairs_data.append({
68.                     'first_product': prod1,
69.                     'next_product': prod2,
70.                     'frequency': freq,
71.                     'time_diff_days': avg_time_diff
72.                 })
73.
74.             pairs_df = pd.DataFrame(pairs_data)
75.             if not pairs_df.empty:
76.                 pairs_df = pairs_df.sort_values('frequency', ascending=False)
77.
78.             print(f"Hotovo. Vytvořeno {len(pairs_df)} unikátních párů produktů.")
79.             return pairs_df

```

2.5.5.3 Vytvoření párů produktů, které byly zakoupeny jedním zákazníkem mezi všemi následujícími nákupy

```

1.     def create_sequence_pairs(self, max_time_diff=None, chunk_size=25,
max_products_per_customer=100, temp_file=None):
2.         print(f"Zpracovávám sekvence pro {self.data['id_zakaznika'].nunique()} zákazníků po
skupinách po {chunk_size}...")
3.
4.         pair_counts = {}
5.         pair_time_diffs = {}
6.
7.         customers = self.data['id_zakaznika'].unique()
8.         total_customers = len(customers)
9.
10.        for i in range(0, total_customers, chunk_size):
11.            chunk_customers = customers[i:min(i+chunk_size, total_customers)]
12.            print(f"Zpracovávám zákazníky {i+1} až {min(i+chunk_size, total_customers)} z
{total_customers}")
13.
14.            chunk_data = self.data[self.data['id_zakaznika'].isin(chunk_customers)].copy()
15.
16.            for customer, group in chunk_data.groupby('id_zakaznika'):
17.
18.                group = group.sort_values('datum_prodeje')
19.
20.                if len(group) > max_products_per_customer:

```

```

21.             print(f"Zákazník {customer} má {len(group)} produktů, omezují na
{max_products_per_customer}")
22.             group = group.head(max_products_per_customer)
23.
24.             purchases_by_date = {}
25.             for _, row in group.iterrows():
26.                 date = row['datum_prodeje']
27.                 product = row['id_produkту']
28.
29.                 if date not in purchases_by_date:
30.                     purchases_by_date[date] = []
31.
32.                 if product not in purchases_by_date[date]:
33.                     purchases_by_date[date].append(product)
34.
35.             sorted_dates = sorted(purchases_by_date.keys())
36.
37.             for i in range(len(sorted_dates)):
38.                 date1 = sorted_dates[i]
39.                 products1 = purchases_by_date[date1]
40.
41.                 for j in range(i + 1, len(sorted_dates)):
42.                     date2 = sorted_dates[j]
43.
44.                     time_diff = (date2 - date1).days
45.                     if max_time_diff is not None and time_diff > max_time_diff:
46.                         continue
47.
48.                     products2 = purchases_by_date[date2]
49.
50.                     max_products = 15
51.                     prods1 = products1[:max_products] if len(products1) > max_products
else products1
52.                     prods2 = products2[:max_products] if len(products2) > max_products
else products2
53.
54.                     for prod1 in prods1:
55.                         for prod2 in prods2:
56.                             if prod1 != prod2:
57.                                 pair = (prod1, prod2)
58.
59.                                 if pair not in pair_counts:
60.                                     pair_counts[pair] = 0
61.                                     pair_time_diffs[pair] = []
62.
63.                                 pair_counts[pair] += 1
64.                                 pair_time_diffs[pair].append(time_diff)
65.
66.             print("Vytvářím výsledný DataFrame...")
67.             pairs_data = []
68.             for (prod1, prod2), freq in pair_counts.items():
69.                 avg_time_diff = sum(pair_time_diffs[(prod1, prod2)]) / len(pair_time_diffs[(prod1,
prod2)])
70.                 pairs_data.append({
71.                     'first_product': prod1,
72.                     'next_product': prod2,
73.                     'frequency': freq,
74.                     'time_diff_days': avg_time_diff
75.                 })
76.
77.             pairs_df = pd.DataFrame(pairs_data)
78.             if not pairs_df.empty:
79.                 pairs_df = pairs_df.sort_values('frequency', ascending=False)
80.
81.             print(f"Hotovo. Vytvořeno {len(pairs_df)} unikátních párů produktů.")
82.             return pairs_df
83.

```

2.5.5.4 Výpočet pravděpodobností přechodu mezi produkty pomocí Markovova modelu

Markovovy rozhodovací procesy představují matematický nástroj pro modelování rozhodovacích situací, v nichž výsledky závisí jak na náhodě, tak na řízených akcích uživatele. Jsou široce využívány k analýze a řešení různých optimalizačních úloh. [31]

Počítané přechodové pravděpodobnosti lze vyjádřit jako, kde P_{ij} je pravděpodobnost přechodu.

$$P_{ij} = \frac{\text{Počet přechodů z produktu } i \text{ na produkt } j}{\text{Celkový počet přechodů z produktu } i}$$

```
1. def calculate_markov_transitions(self, chunk_size=25):
2.     unique_products = self.data['id_produkту'].unique()
3.     n_products = len(unique_products)
4.
5.     product_to_idx = {product: i for i, product in enumerate(unique_products)}
6.     idx_to_product = {i: product for product, i in product_to_idx.items()}
7.
8.     transition_matrix = np.zeros((n_products, n_products))
9.
10.    customers = self.data['id_zakaznika'].unique()
11.    total_customers = len(customers)
12.
13.    print(f"Počítám pravděpodobnosti přechodů pro {total_customers} zákazníků po skupinách
14.    po {chunk_size}...")
15.    for i in range(0, total_customers, chunk_size):
16.        chunk_customers = customers[i:min(i+chunk_size, total_customers)]
17.        print(f"Zpracovávám zákazníky {i+1} až {min(i+chunk_size, total_customers)} z
18.        {total_customers}")
19.        chunk_data = self.data[self.data['id_zakaznika'].isin(chunk_customers)]
20.
21.        for customer, group in chunk_data.groupby('id_zakaznika'):
22.            products = group['id_produkту'].tolist()
23.
24.            for i in range(len(products) - 1):
25.                for j in range(i + 1, len(products)):
26.                    from_idx = product_to_idx[products[i]]
27.                    to_idx = product_to_idx[products[j]]
28.                    transition_matrix[from_idx, to_idx] += 1
29.
30.            row_sums = transition_matrix.sum(axis=1, keepdims=True)
31.
32.            row_sums[row_sums == 0] = 1
33.            transition_probs = transition_matrix / row_sums
34.
35.            transitions_df = pd.DataFrame(index=unique_products, columns=unique_products,
36.            data=transition_probs)
37.
38.            transitions_long = []
39.            for i in range(n_products):
40.                for j in range(n_products):
41.                    if transition_probs[i, j] > 0:
42.                        transitions_long.append({
43.                            'first_product': idx_to_product[i],
44.                            'next_product': idx_to_product[j],
45.                            'probability': transition_probs[i, j]
46.                        })
47.            transitions_long_df = pd.DataFrame(transitions_long)
48.            if not transitions_long_df.empty:
49.                transitions_long_df = transitions_long_df.sort_values('probability', ascending=False)
50.
51.            return transitions_long_df, transitions_df
```

2.5.5.5 Analýza přechodů mezi kategoriemi produktů

```
1. def analyze_category_transitions(self, chunk_size=25):
2.
3.     if 'id_skupiny' not in self.data.columns:
4.         print("Data neobsahují informace o kategoriích produktů.")
5.         return pd.DataFrame()
6.
7.     customers = self.data['id_zakaznika'].unique()
8.     total_customers = len(customers)
9.
10.    print(f"Analyzuji přechody mezi kategoriemi pro {total_customers} zákazníků po skupinách
11.    po {chunk_size}...")
12.
13.    all_category_pairs = []
14.
15.    for i in range(0, total_customers, chunk_size):
16.        chunk_customers = customers[i:min(i+chunk_size, total_customers)]
17.        print(f"Zpracovávám zákazníky {i+1} až {min(i+chunk_size, total_customers)} z
18.        {total_customers}")
19.
20.        chunk_data = self.data[self.data['id_zakaznika'].isin(chunk_customers)]
21.
22.        category_pairs = []
23.
24.        for customer, group in chunk_data.groupby('id_zakaznika'):
25.            categories = group['id_skupiny'].tolist()
26.
27.            for i in range(len(categories)):
28.                for j in range(i+1, len(categories)):
29.                    category_pairs.append((categories[i], categories[j]))
30.
31.            if category_pairs:
32.                chunk_pairs_df = pd.DataFrame(category_pairs, columns=['first_category',
33.                'next_category'])
34.                chunk_counts = chunk_pairs_df.groupby(['first_category',
35.                'next_category']).size().reset_index(name='frequency')
36.                all_category_pairs.append(chunk_counts)
37.
38.            if not all_category_pairs:
39.                return pd.DataFrame(columns=['first_category', 'next_category', 'frequency'])
40.
41.            combined_pairs = pd.concat(all_category_pairs, ignore_index=True)
42.
43.            category_counts = combined_pairs.groupby(['first_category',
44.            'next_category'])['frequency'].sum().reset_index()
45.            category_counts = category_counts.sort_values('frequency', ascending=False)
46.
47.            return category_counts
```

2.5.5.6 Export výsledků do Excel souboru s postupným zpracováním pro úsporu paměti

```
1. def export_to_excel(self, filename="analiza_navaznosti_produkту.xlsx", chunk_size=25):
2.     print("Začínám exportovat výsledky do Excelu...")
3.     self.prepare_data()
4.
5.     with pd.ExcelWriter(filename, engine='openpyxl') as writer:
6.         print("Exportuji základní statistiky...")
7.         stats = pd.DataFrame({
8.             'Metrika': ['Počet zákazníků', 'Počet produktů', 'Počet transakcí', 'Počet
9.             kategorií'],
10.            'Hodnota': [
11.                self.data['id_zakaznika'].nunique(),
12.                self.data['id_produkту'].nunique(),
13.                len(self.data),
```

```

13.         self.data['id_skupiny'].nunique() if 'id_skupiny' in self.data.columns
else 'N/A'
14.     ]
15.     })
16.     stats.to_excel(writer, sheet_name='Statistiky', index=False)
17.
18.     print("Provádím sekvenční analýzu...")
19.     try:
20.         sequence_pairs = self.create_sequence_pairs(
21.             max_time_diff=1000,
22.             chunk_size=chunk_size,
23.             max_products_per_customer=100
24.         )
25.
26.         export_limit = min(10000, len(sequence_pairs))
27.         sequence_pairs.head(export_limit).to_excel(writer,
sheet_name='Navaznosti_produkту', index=False)
28.         print(f"Exportováno {export_limit} nejčastějších párů produktů.")
29.     except Exception as e:
30.         print(f"Chyba při sekvenční analýze: {e}")
31.
32.         pd.DataFrame(columns=['first_product', 'next_product', 'frequency',
'next_time_diff_days']).to_excel(
33.             writer, sheet_name='Navaznosti_produkту', index=False
34.         )
35.
36.         del sequence_pairs
37.         import gc
38.         gc.collect()
39.
40.         print("Počítám pravděpodobnosti přechodů...")
41.         try:
42.             transitions, _ = self.calculate_markov_transitions(chunk_size=chunk_size)
43.
44.             export_limit = min(10000, len(transitions))
45.             transitions.head(export_limit).to_excel(writer,
sheet_name='Pravdepodobnosti_prechodu', index=False)
46.             print(f"Exportováno {export_limit} nejpravděpodobnějších přechodů.")
47.         except Exception as e:
48.             print(f"Chyba při výpočtu pravděpodobností přechodů: {e}")
49.         pd.DataFrame(columns=['first_product', 'next_product', 'probability']).to_excel(
50.             writer, sheet_name='Pravdepodobnosti_prechodu', index=False
51.         )
52.
53.         del transitions
54.         gc.collect()
55.
56.         if 'id_skupiny' in self.data.columns:
57.             print("Analyzuji přechody mezi kategoriemi...")
58.             try:
59.                 category_transitions =
self.analyze_category_transitions(chunk_size=chunk_size)
60.                 category_transitions.to_excel(writer, sheet_name='Navaznosti_kategorii',
index=False)
61.                 print(f"Exportováno {len(category_transitions)} přechodů mezi kategoriemi.")
62.             except Exception as e:
63.                 print(f"Chyba při analýze přechodů mezi kategoriemi: {e}")
64.                 pd.DataFrame(columns=['first_category', 'next_category',
'frequency']).to_excel(
65.                     writer, sheet_name='Navaznosti_kategorii', index=False
66.                 )
67.
68.                 del category_transitions
69.                 gc.collect()
70.
71.         print(f"Výsledky byly exportovány do souboru {filename}")
72.         return filename
73.

```

V tomto kódu bylo přidáno pár omezení, například aby se výpočty dělali po částech, kvůli úspoře paměti, kterou by bylo potřeba poskytnout, kdyby se celý proces prováděl najednou.

2.6 Interpretace výsledků

2.6.1 Analýza nákupních košíků

Při prvních pokusech o analýzu často kupovaných produktů dohromady byly dostávány převážně výsledky servisních zakázek (servis kol/elektrokol) společně se servisním materiálem. To ale není úplně důležité, protože lidé si nekupují přímo konkrétní produkty, když dávají například kolo na servis, ale chtějí pouze vyřešit problém nebo například zkontrolovat stav kola. Z analýzy tudíž byly odstraněny servisní položky, které obchod nabízí a bylo se zaměřeno na fyzické produkty (Tabulka 9).

Tabulka 9: Tabulka nejčastěji kupovaných produktů společně

Produkt 1	Produkt 2
801	7820
7820	4204
451	29
451	13
7302	23
5834	7820
5834	4204
4204	5834, 7820
4204	3604
428	429
7372	7373
7409	7521

Zde je 12 nejčastějších produktů, které jsou kupovány dohromady a jsou kupovány dostatečně často, aby mělo význam se s nimi zabývat. Vyskytuje se zde například často produkt 7820, což je stojánek na kolo od výrobce, kterým je obchod hlavním dodavatelem do ČR. Často je opět kombinována se stejnou značkou dalších produktů. Například s rohy na řídítka – 5834 a gripy na řídítka – 4204, které jsou často kombinovány i společně. Také zde máme produkty, které spolu velice souvisí jako brzdové – 29 a řadící lanka – 13 s koncovkou lanka – 451 nebo přední a zadní blatník.

Tato analýza může být použita pro nabídku produktů kupovaných dohromady na e-shopu při přidání nějakého produktu do košíku, nebo pro cílenou reklamu ihned po koupi produktu nebo pro masovou e-mailovou kampaň nedlouho po koupi daných produktů.

Při analýze kategorií byly zjištěny nejsilnější a nejčastější závislosti (Tabulka 10).

Tabulka 10: Tabulka kategorií nejčastěji kupovaných společně

Kategorie 1	Kategorie 2	Pravděpodobnost
139	44	0,049
139	82	0,085
139	109	0,108
130	139	0,127
26	139	0,079
139	293	0,136
130	44	0,127
26	82	0,004
26	139	0,079
82	26	0,728
128	130	0,547

V kategoriích kupovaných dohromady se nejčastěji v této analýze se objevují kategorie, které odpovídají produktům, již jsou kupované dohromady 82 – rohy, 26 – gripy, ale časté jsou i další kategorie 139 – světla a 44 – košíky. Zároveň se zde již objevují i v nejčastějších kategoriích kategorie, které obsahují dražší položky, jako například 109 – horská kola. Tyto výsledky se dají použít například při plánování e-mailových kampaní nebo při akčních nabídkách.

Tato analýza byla dobrým počátečním bodem pro zjišťování souvislostí mezi produkty/kategoriemi a již tyto samostatné výstupy se dají použít pro zlepšování cílené reklamy, obzvláště s kombinací s použitím pravděpodobnosti a následné analýzy po spuštění cílené reklamy.

2.6.2 Analýza zákazníků.

Při analýze zákazníků pomocí shlukování bylo vytvořeno 5 skupin, které mají své jasné charakteristiky podle měřených veličin, kterými jsou celková útrata, průměrná cena nákupu, počet nákupů, frekvence nákupů a průměrný počet produktů v nákupu (Tabulka 11 a Tabulka 12).

Tabulka 11: první polovina výsledků shlukování dat

cluster	zakazníci	celkově utraceno [Kč]			průměrná cena [Kč]			počet nákupů		
	Počet	mean	min	max	mean	min	max	mean	min	max
0	2266	11997,63	179	58141	2444,369	44,75	8366,75	1,086055	1	2
1	1798	31520,29	192	98886	5966,426	43,2	29138	2,460512	2	3
2	2146	49352,82	27858	98954	16600,52	5075	29269,5	1,041473	1	2
3	3485	17622,75	8390	35326	14550,86	7942	29244	1,004304	1	2
4	6721	6768,552	0	30394	3326,797	0	9192,667	1,02336	1	2

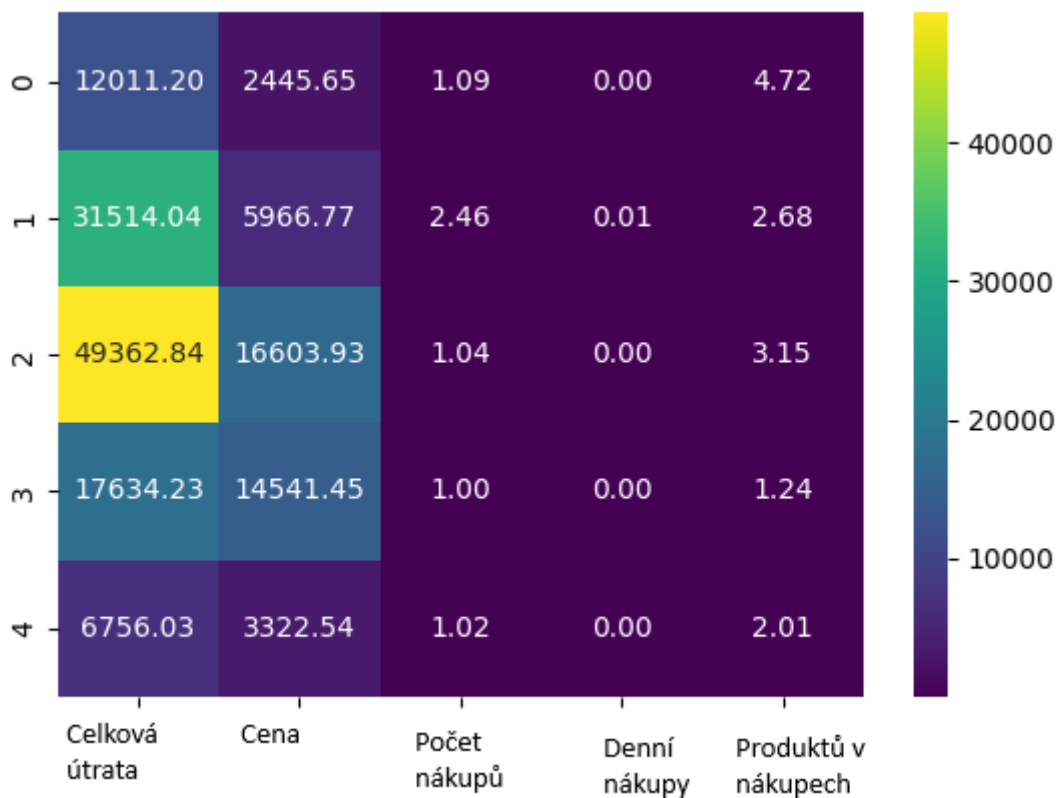
Tabulka 12: druhá polovina výsledků shlukování dat

cluster	zákazníci	Frekvence nákupů			Průměrný počet produktů v nákupu [Ks]		
	Počet	mean	min	max	mean	min	max
0	2266	0,000268	0	0,007937	4,72087379	3	6,5
1	1798	0,010648	0,00101	0,024793	2,68437152	1	6,66666667
2	2146	0,000156	0	0,010471	3,14631873	1	6,5
3	3485	1,01E-05	0	0,005102	1,24002869	1	3
4	6721	6,05E-05	0	0,006024	2,01487874	1	3

Výsledky shlukování rozdělily zákazníky do následujících shluků na základě maximálních, minimálních a průměrných dat (Obrázek 15):

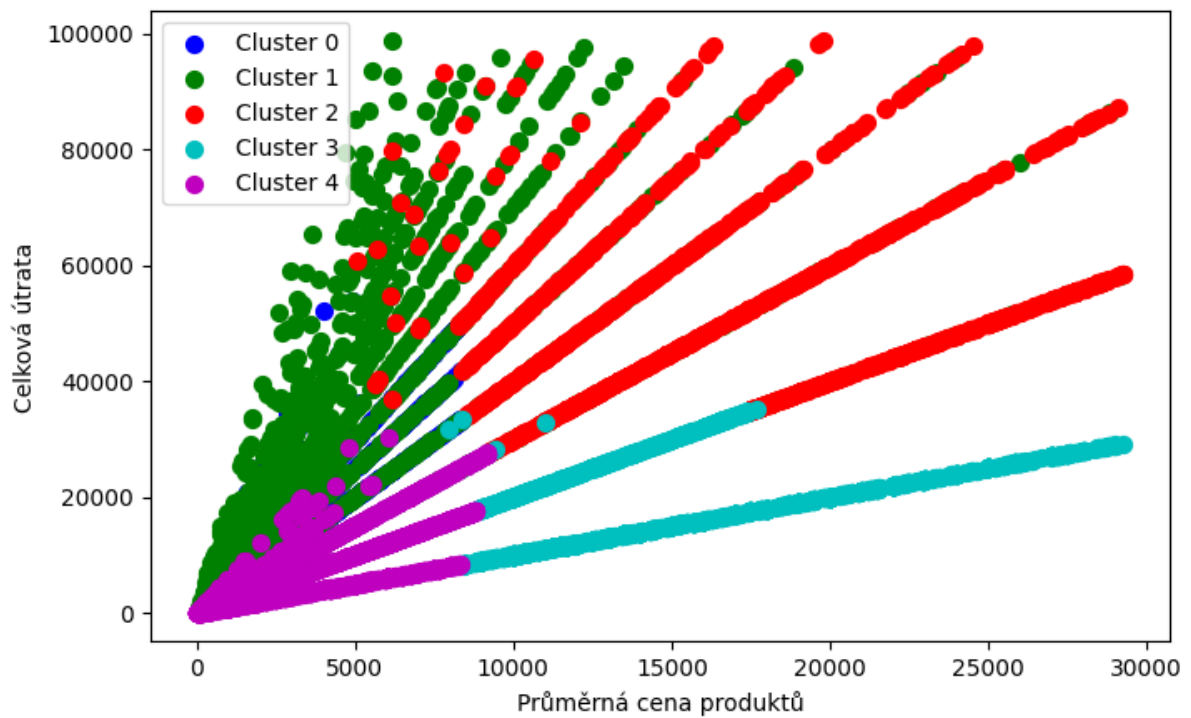
- Shluk 0 – Příležitostní nakupující s více položkami
 - Kupují více položek během jednoho nákupu
 - Nakupují zřídka
 - Volí produkty střední cenové kategorie
- Shluk 1 – Věrní zákazníci
 - Nejčastější nakupující
 - Vysoké výdaje s více nákupy
 - Projevují věrnost/opakované nákupy
 - Nejhodnotnější dlouhodobí zákazníci
- Shluk 2 – Premiový zákazníci s velkými nákupy a výdaji
 - Nejvyšší průměrné výdaje
 - Kupují velmi drahé produkty
 - Provádějí velké jednorázové nákupy
- Shluk 3 – Kupující jednotlivých drahých produktů
 - Zaměřují se na specifické drahé produkty
 - Minimalistický přístup k nakupování (nejméně položek na nákup)
 - Kupují drahé položky, ale celkově utrácejí méně než Cluster 2
 - Cílené, konkrétní nákupy
 - Kupující pravděpodobně jízdních kol a elektrokol, jako nejdražších produktů v nabídce

- Shluk 4 – Cenově citlivý zákazníci
 - Největší skupina (6 721 zákazníků)
 - Nejnižší průměrné výdaje
 - Potenciálně citliví na cenu



Obrázek 15: Průměrné hodnoty metrik pro každý shluk

Po vizualizaci zákazníků do grafu, je vidět rozložení zákazníků v shlucích na základě průměrné ceny produktů a celkové útraty, odkud je vidět, že se shluk 1, který obsahuje pravidelné zákazníky, je pro celkovou útratu spolu se shlukem 2 nejdůležitější (Obrázek 16).



Obrázek 16: Rozložení zákazníků v shlucích na základě průměrné ceny produktů a celkové útraty

Když se přiřadí jednotlivé shluky k daným zákazníkům, tak vytváří podklady pro každého zákazníka a může se předpokládat, jaké budou jeho následné nákupy a podle toho přizpůsobit i cílenou marketingovou kampaň (Tabulka 13).

Tabulka 13: přiřazené shluky k jednotlivým zákazníkům

id_zakaznika	cluster
1000012	0
1000014	4
1000027	4
1000164	4
1000378	3
1000435	3
1000534	4
1000601	0
1000614	4
1000654	4

2.6.3 Analýza následujících produktů a skupin

2.6.3.1 Analýza pomocí SQL dotazů

Při prvotní analýze přes SQL dotazy bylo opět naraženo na problém servisních položek, které jsou velmi časté a velmi často obsahují stejné produkty. Po odstranění servisů poskytla analýza mnohem přehlednější výsledky, ve které převládají produkty, které mají nízkou cenu a jsou kupovány velmi často i ve velkých počtech (Tabulka 14).

Tabulka 14: Analýza následujících produktů pomocí dotazů SQL

Produkt 1	Produkt 2	Počet
451	7348	394
451	504	394
451	125	394
13	2245	344
13	2480	344
13	7348	344
13	3380	344
13	2500	344
13	2478	344
130	2837	302
130	1976	302
130	4913	302
130	2686	302
130	3393	302
130	5225	302

Tento výstup může být užitečný s počtem, kolikrát se následující produkt vyskytuje, ale nebere v potaz kolikrát se produkt vyskytuje v nákupech celkově a jaká je tedy pravděpodobnost přechodu na daný produkt.

2.6.3.2 Analýza pomocí Pythonu

Pro pokročilejší analýzu bylo účinnější použití programovacího jazyku Python a zjistit, které produkty následují po sobě, jak často se vyskytují a s jakou pravděpodobností. Tyhle výsledky následně budou hlavním výstupem, který se bude používat při zlepšování cílené reklamy na internetu. Nejprve bylo zaměřeno na produkty v následujícím nákupu u zákazníků a uloženo 10000 nejčastějších (Tabulka 15).

Tabulka 15: Nejčastější produkty v následujícím nákupu

Produkt	Následující produkt	Počet výskytů	Rozdíl dní	Pravděpodobnost
7820	4204	231	83	0,027
4204	5834	188	125	0,099
5834	4204	171	123	0,03
4204	3267	73	108	<0,007
7820	3267	72	54	0,007
7820	8538	62	57	0,03
4204	3458	60	349	<0,007
7820	2632	45	103	<0,007
7820	5260	43	56	<0,007
7820	6601	42	97	<0,007
7820	5234	41	56	<0,007
4204	6461	41	114	0,011
4204	7348	37	331	<0,007
1142	3458	36	401	0,065

V tabulce jsou produkty, které se objevují v ihned následném nákupu a s jakou pravděpodobností se vyskytují. Objevují se zde často doplňky ke kolu jako 7820 – stojánek předchází 4204 – gripy, 4204 – gripy předcházejí 3267 – MTB duše, 1142 – řetěz na kolo předchází 3458 – brzdové destičky.

Z analýzy je zároveň patrné nejčastější následující kategorie pro příští nákup (Tabulka 16).

Tabulka 16: Nejčastější kategorie v následujícím nákupu

Kategorie	Následující kategorie	Počet výskytů
93	93	32762
130	93	5605
130	293	4428
44	93	3967
93	130	3824
93	5	3754
293	93	3593
128	93	3532
26	93	2914
93	26	2748

V následujících nákupech se často vyskytuje kategorie 93 – služby (servisní položky), což by mohlo znamenat dobrou retenci zákazníků a dá se na ně upozorňovat v nějakých kampaních, ale pro cílenou reklamu budou důležitější ne službové produkty. V tomhle případě například 130 – stojánky na kola a 293 – horská elektrokola. Zde je možnost vyčíst, že se hodně zákazníků vrací po zakoupení kola na servis.

Když analýza pokračuje na analýzu návaznosti produktů, které nepřicházejí pouze v následujícím nákupu, ale v dalších nákupech, po celý zákaznický cyklus, tak je vidět nejpraktičtější výstup pro cílenou reklamu na internetu (Tabulka 17).

Tabulka 17: Nejčastější produkty v následujících nákupech

Produkt	Následující produkt	Počet Výskytů	Rozdíl dní	Pravděpodobnost
7820	4204	246	87	0,026
4204	5834	206	159	0,098
4204	2571	199	425	0,013
5834	4204	184	139	0,03
7820	5834	161	83	0,017
4204	7820	152	144	0,059
2500	2571	138	429	0,048
5834	2571	129	416	0,016
7820	2571	106	216	0,009
4204	2500	105	394	<0,007
3604	4204	93	183	0,057
2571	2500	93	343	0,015
3458	2571	90	421	0,044
2571	3458	85	376	0,059
5834	7820	85	152	0,061

Zde jsou nejčastější produkty a produkty následující, podle kterých se bude nejčastěji zavádět cílená reklama. Jsou zde často opakující se produkty jako v předchozí analýze, ale zde je pro účely cílené reklamy přesnější pravděpodobnost a rozdíl dní, jelikož prozkoumává i další nákupy. V první položce je vidět, že se výrazně hodnoty nezměnily, ale u druhé je možnost zaznamenat, že rozdíl dní mezi nákupy vzrostl o 34 dní na 159 dní.

3 VLASTNÍ NÁVRHY ŘEŠENÍ

3.1 Shrnutí hlavních výsledků analýz

Provedené analýzy poskytly cenné informace o nákupním chování zákazníků, jejich segmentaci a vztazích mezi produkty v nabídce e-shopu.

3.1.1 Analýza nákupních košíků

Tato analýza odhalila významné vztahy mezi produkty kupovanými společně. Mezi nejčastější kombinace patří stojánky na kolo (7820) s gripy na řídítka (4204) a rohy na řídítka (5834). Další často kupované kombinace zahrnují brzdová lanka (29) s řadicími lanky (13) a koncovkami (451). Tyto vazby poskytují příležitosti pro cross-selling.

3.1.2 Analýza kategorií

Analýza ukázala silné vazby mezi kategoriemi jako světla (139), košíky (44), gripy (26) a rohy (82). V těchto kombinacích se objevují i dražší položky jako horská kola (109), což naznačuje potenciál pro cílené marketingové kampaně.

Segmentace zákazníků identifikovala pět klíčových segmentů:

1. **Příležitostní nakupující s více položkami** – nakupují zřídka, ale více položek najednou, preferují střední cenovou kategorii
2. **Věrní zákazníci** – nejčastěji nakupující s vysokými výdaji a opakovanými nákupy, představují nejhodnotnější dlouhodobé zákazníky
3. **Prémiový zákazníci** – realizují velké jednorázové nákupy s nejvyššími průměrnými výdaji, kupují velmi drahé produkty
4. **Kupující jednotlivých drahých produktů** – zaměřují se na specifické drahé položky (pravděpodobně jízdní kola a elektrokola), minimalistický přístup k nakupování
5. **Cenově citliví zákazníci** – největší skupina (6 721 zákazníků) s nejnižšími průměrnými výdaji

3.1.3 Analýza následných nákupů

Tato analýza poskytla informace o pravděpodobnosti a načasování následných nákupů produktů. Například po zakoupení stojánku na kolo (7820) následuje s 2,6% pravděpodobností nákup gripů (4204) v průměru po 87 dnech. Po nákupu gripů (4204) následuje s 9,8% pravděpodobností nákup rohů na řídítka (5834) průměrně po 159 dnech. Tato data mohou sloužit pro efektivní cílení reklamních kampaně.

Analýza také ukázala, že zákazníci se často vrací po zakoupení kola na servis, což indikuje dobrou retenci zákazníků a příležitost pro servisní marketing.

3.2 Návrh marketingových strategií

Cílem marketingových strategií je optimalizovat cílení na specifické segmenty zákazníků a vytvořit personalizované nabídky, které zohlední jejich potřeby a zvyklosti.

3.2.1 Strategie cílení na segmenty zákazníků

3.2.1.1 Segment 1: Příležitostní nakupující s více položkami

- **Strategie:** Zvýšit frekvenci nákupů pomocí sezónních nabídek a připomínek.
- **Taktiky:**
 - Vytvořit balíčky souvisejících produktů z jejich předchozích nákupů
 - Nabízet slevy na další nákup s časově omezenou platností (např. 60 dní)
 - Komunikovat sezónní nabídky související s cyklistikou (jarní údržba, letní výlety, podzimní příprava)

3.2.1.2 Segment 2: Věrní zákazníci

- **Strategie:** Maximalizovat hodnotu vztahu a posílit loajalitu.
- **Taktiky:**
 - Implementovat věrnostní program s výhodami pro pravidelné zákazníky
 - Nabízet exkluzivní předprodeje a dřívější přístup k novým produktům
 - Poskytovat personalizované doporučení produktů na základě historie nákupů
 - Nabízet VIP servisní podmínky a rozšířené záruky

3.2.1.3 Segment 3: Prémiový zákazníci

- **Strategie:** Udržet vysokou hodnotu nákupů a zvýšit frekvenci.
- **Taktiky:**
 - Cílit na prémiové doplňky k již zakoupeným produktům
 - Nabízet exkluzivní prémiové produkty s omezenou dostupností
 - Komunikovat novinky v prémiovém segmentu
 - Zajistit prémiový zákaznický servis a konzultace

3.2.1.4 Segment 4: Kupující jednotlivých drahých produktů

- **Strategie:** Konvertovat jednorázové nákupy na opakované.
- **Taktiky:**
 - Zaměřit se na doplňky ke konkrétním zakoupeným modelům kol/elektrokol
 - Nabízet pravidelné servisní prohlídky a údržbu
 - Vytvářet edukační obsah o údržbě a vylepšení zakoupeného produktu
 - Představovat nové modely a možnosti vylepšení

3.2.1.5 Segment 5: Cenově citliví zákazníci

- **Strategie:** Zvýšit hodnotu nákupů pomocí cenově výhodných nabídek.
- **Taktiky:**
 - Komunikovat akce, slevy a výprodeje
 - Nabízet cenově dostupné alternativy k dražším produktům
 - Vytvářet limitované časové nabídky pro zvýšení naléhavosti
 - Využívat skupinové nabídky pro zvýšení průměrné hodnoty nákupu

3.2.2 Co nabízet – produktové strategie

3.2.2.1 Cross-selling

Implementovat automatické nabídky souvisejících produktů na základě analýzy nákupních košíků:

- Při nákupu stojánku (7820) nabízet gripy (4204) a rohy (5834)
- Při nákupu gripů (4204) nabízet rohy (5834) a MTB duše (3267)
- Při nákupu horských kol nabízet doplňky jako světla, košíky a stojánky

3.2.2.2 Up-selling

- U cenově citlivých zákazníků zdůrazňovat výhody mírně dražších alternativ
- U prémiových zákazníků nabízet nadstandardní příslušenství a vylepšení

3.2.2.3 Kategoriální nabídky

Vytvářet tematické nabídky na základě silných vazeb mezi kategoriemi:

- Balíčky "bezpečnost a denní vybavenost" (světla, odrazky, helmy)
- Balíčky "komfort" (gripy, rohy, sedla)
- Balíčky "údržba" (řetězy, maziva, čisticí prostředky)

3.2.2.4 Servisní nabídky

Využít zjištění o časté návaznosti servisních služeb po nákupu kol:

- Nabízet servisní balíčky při nákupu kol a elektrokol
- Komunikovat preventivní servisní prohlídky před sezónou
- Vytvořit servisní předplatné pro pravidelné zákazníky

3.2.3 Komunikace se zákazníky

3.2.3.1 Personalizovaná e-mailová komunikace

- Využít segmentaci zákazníků pro cílené kampaně:
 - Pro věrné zákazníky: novinky, exkluzivní nabídky, VIP akce
 - Pro prémiové zákazníky: luxusní produkty, limitované edice
 - Pro cenově citlivé: akční nabídky, výprodeje, časově omezené slevy

- Frekvence:
 - Věrní zákazníci: častější komunikace (2 – 3x měsíčně)
 - Příležitostní nakupující: méně častá komunikace (1x měsíčně)
 - Cenově citliví: akční nabídky (1 – 2x měsíčně)

3.2.3.2 Webové prostředí

Implementovat personalizované produktové doporučení na základě historie nákupů, produktů v košíku a prohlížených kategorií.

Využít pop-up okna s personalizovanými nabídkami pro registrované zákazníky.

3.2.3.3 Remarketing

Využít data o pravděpodobnosti a časování následných nákupů pro cílené reklamní kampaně na sociálních sítích a v rámci Google Ads.

3.2.3.4 Sezónní strategie

Jaro (březen-květen): Zaměření na přípravu kol, servis, nové modely

Léto (červen-srpen): Doplnky pro jízdu, ochrana před sluncem, hydratace

Podzim (září-listopad): Příprava na horší počasí, osvětlení, zimní vybavení

Zima (prosinec-únor): Indoorové tréninky, dárkové poukazy, příprava na sezónu

3.2.3.5 Cílená komunikace podle analýzy následných nákupů

Implementovat automatizované marketingové kampaně podle zjištěných časových intervalů mezi nákupy:

- Po nákupu stojánku (7820) cílit na gripy (4204) po 80-90 dnech
- Po nákupu gripů (4204) cílit na rohy (5834) po 150-160 dnech
- Po nákupu řetězu (1142) cílit na brzdové destičky (3458) po 390-410 dnech

3.2.3.6 Servisní připomínky

Na základě typu zakoupeného produktu (kola/elektrokola) a sezónních faktorů nastavit automatické připomínky servisních prohlídek. První servis po 1-2 měsících používání, sezónní servis před zahájením cyklistické sezóny a pravidelný servis elektrokol po 500-1000 km.

3.2.3.7 Výroční akce

Implementovat systém slev a bonusů k výročí registrace či prvního nákupu.

Implementace těchto marketingových strategií založených na datových analýzách by měla vést k významnému zvýšení retence zákazníků, hodnoty průměrného nákupu a celkového obrátu e-shopu. Důležité pro úspěch bude kontinuální monitoring efektivity jednotlivých kampaní a jejich průběžná optimalizace na základě nově získávaných dat.

3.3 Možnosti implementace a očekávané přínosy

3.3.1 Scénář strategického implementačního plánu

Implementace navržených marketingových strategií by měla probíhat ve čtyřech navazujících fázích, které zajistí systematický přístup a možnost průběžného vyhodnocování úspěšnosti:

3.3.1.1 Technická příprava a infrastruktura (1-2 měsíce)

Integrace segmentačního modelu do systému pro řízení vztahů se zákazníky (CRM systému) zahrnuje několik kroků. Nejdříve je potřeba implementovat mechanismus automatického přiřazování zákazníků do předem definovaných shluků. Dále je nutné nastavit systém tak, aby pravidelně aktualizoval zákaznické shluky na základě nejnovějších dat a změn v chování zákazníků. Součástí integrace je rovněž vytvoření intuitivního rozhraní pro marketingové oddělení, které poskytne přehled o jednotlivých segmentech a usnadní využití těchto informací pro cílené marketingové kampaně.

Technické zajištění personalizace zahrnuje několik opatření, která umožňují přizpůsobit zákaznickou zkušenost na míru. Jedním z kroků je implementace dynamického obsahu na webových stránkách, který se mění v závislosti na shluku, do kterého zákazník patří. Dále je nutné zajistit nastavení personalizovaných notifikací na e-shopu, které budou reagovat na individuální potřeby či zájmy každého zákazníka. V neposlední řadě je nezbytná konfigurace marketingové automatizace, která podporuje vícekanálovou komunikaci a zajišťuje, že personalizovaná sdělení budou doručena správnou formou a v pravý čas.

3.3.1.2 Pilotní kampaně pro validaci konceptu (2-3 měsíce)

1. Cross-sellingové kampaně

- Implementace doporučovacího systému na produktových stránkách a v košíku
- Spuštění e-mailových kampaní s doporučeními top 5 kombinací produktů
- Nastavení remarketingových kampaní s produkty s nejvyšší pravděpodobností následného nákupu

2. Segmentové pilotní kampaně

- Spuštění pilotních kampaní pro každý segment s testováním obsahu
- Testování cenových strategií pro segment cenově citlivých zákazníků
- Evaluace odezvy VIP programu pro prémiové zákazníky na malém vzorku

3. Měření a optimalizace

- Určení hlavních ukazatelů výkonnosti pro jednotlivé kampaně
- Průběžná analýza výsledků a iterativní úpravy kampaní
- Vyhodnocení ROI pilotních kampaní a identifikace nejúspěšnějších strategií

3.3.1.3 Plnohodnotné nasazení strategií (3-6 měsíců)

1. Komplexní aktivace segmentů

- Spuštění plnohodnotných kampaní pro všechny segmenty
- Implementace personalizovaných landing pages podle segmentu
- Aktivace vícekanálové komunikace (e-mail, web, sociální sítě, SMS)

2. Automatizace marketingových procesů

- Implementace marketingové automatizace pro životní cyklus zákazníka
- Nastavení triggerů pro automatické spuštění kampaní na základě chování
- Vytvoření automatizovaných scénářů pro reaktivaci neaktivních zákazníků

3. Cross-selling a upselling program

- Implementace full-scale doporučovacího systému využívajícího data z analýzy košíků
- Nastavení personalizovaných upsell nabídek během nákupního procesu
- Implementace časovaných e-mailových kampaní pro následné produkty

3.3.1.4 Optimalizace a škálování (6-12 měsíců)

1. Analytický dashboard pro vyhodnocování

- Vytvoření komplexního marketingového dashboardu s real-time daty
- Implementace pokročilé atribuční analýzy pro vyhodnocování kampaní
- Nastavení systému alertů pro anomálie v chování zákazníků

2. Kontinuální optimalizace

- Pravidelná revize segmentace (kvartálně)
- Dynamická optimalizace cross-sellingových nabídek na základě aktuálních dat
- Zavádění nových hypotéz a jejich testování na základě zpětné vazby

3. Škálování úspěšných strategií

- Identifikace nejúspěšnějších kampaní a jejich škálování
- Rozšíření personalizace na další aspekty zákaznické zkušenosti
- Integrace offline a online marketingových aktivit

3.3.2 Odhadovaný potenciální přínos

3.3.2.1 Ekonomické přínosy (Tabulka 18):

- Zvýšení průměrné hodnoty objednávky
- Nárůst frekvence nákupů
- Zvýšení retence zákazníků

Tabulka 18: Kvantifikované ekonomické přínosy

Metrika	Současný stav	Očekávaný přínos	Procentuální zlepšení
Průměrná hodnota objednávky	3 400 Kč	3600 Kč	6 %
Míra konverze	2,10 %	2,50 %	19 %
Frekvence nákupů u Segmentu 1	1× za 215 dní	1× za 180 dní	16 %
Retence zákazníků (roční)	18 %	21 %	17 %

Uvedené hodnoty vycházejí z kvalifikovaných odhadů na základě výsledků datové analýzy a interní znalosti firmy získané konzultací s jejím společníkem. Nejde o výstupy prediktivního modelu, ale o prakticky orientovaný návrh očekávaných přínosů.

3.3.2.2 Operační přínosy (Tabulka 19):

- **Efektivita marketingových procesů:**
 - Zkrácení času potřebného na přípravu kampaní
 - Zvýšení efektivity remarketing kampaní
 - Redukce manuální práce při segmentaci zákazníků
- **Datová kvalita a využití:**
 - Zlepšení kvality zákaznických dat
 - Zvýšení míry využití zákaznických dat pro rozhodování

Tabulka 19: Kvantifikované operační přínosy

Metrika	Současný stav	Očekávaný přínos	Procentuální zlepšení
Čas na přípravu kampaní	12 hodin/kampaň	4 hodiny/kampaň	-67 %
Efektivita remarketingových kampaní	0,70 %	1,80 %	157 %
Využití dat pro rozhodování	25 % dostupných dat	75 % dostupných dat	200 %

Uvedené hodnoty vycházejí z kvalifikovaných odhadů na základě výsledků datové analýzy a interní znalosti firmy získané konzultací s jejím společníkem. Nejde o výstupy prediktivního modelu, ale o prakticky orientovaný návrh očekávaných přínosů.

3.3.2.3 Zákaznická zkušenost:

- **Personalizace a relevance:**
 - Zvýšení relevance nabídek měřené mírou otevření e-mailů
 - Zvýšení úspěšnosti personalizovaných nabídek

3.3.3 Scénář byznysového dopadu implementace

V krátkodobém horizontu je cílem dosáhnout konkrétních zlepšení pomocí dostupných nástrojů a strategií. Jednou z priorit je zvýšení průměrné hodnoty objednávek optimalizací cross-sellingových procesů, což povede ke zlepšení prodejních výsledků. Dalším krokem je

zaměření na výsledky z remarketingových kampaní, které budou přesněji cílené a efektivnější. Nakonec je důležité zlepšit využití již existujících zákaznických dat, a to bez nutnosti dalších finančních investic.

Ve střednědobém horizontu se aktivity zaměřují na dosažení strategických cílů v oblasti zákaznického chování a predikce. Důležité je systematické zvyšování míry úspěšnosti v rámci jednotlivých segmentů, což umožní efektivnější využití marketingových a prodejních strategií. Dále je potřeba stabilizovat zvýšenou frekvenci nákupů u příležitostných zákazníků, čímž se podpoří jejich loajalita a opakované nákupy. K těmto cílům pak přispěje vybudování plně funkčního prediktivního modelu, který bude schopný efektivně předvídat další nákupní chování a usnadní plánování budoucích aktivit.

V dlouhodobém horizontu je cílem vytvořit pevné základy pro udržitelný růst a konkurenční výhodu. Hlavním krokem je vybudování datově řízeného marketingu, který umožní přesnější rozhodování a lepší cílení kampaní, což poskytne významnou konkurenční výhodu. Dále je prioritou zvýšení podílu na trhu prostřednictvím vyšší loajality zákazníků, kdy se dlouhodobé vztahy stanou klíčem k udržení zákazníků i jejich doporučení dalším. Nakonec je třeba zaměřit se na transformaci jednorázových zákazníků na pravidelné a loajální, což zajistí stabilní příjmy a podporu značky v dlouhodobém měřítku.

3.4 Náklady na implementaci

3.4.1 Technická implementace

Technická implementace zahrnuje prvky pro efektivní správu zákaznických dat, automatizaci marketingových procesů (Tabulka 20).

Tabulka 20: Náklady technické implementace

Položka	Jednorázové náklady [Kč]	Pravidelné měsíční náklady [Kč]	Poznámka
Integrace segmentačního modelu do CRM	40 000	0	Externí vývoj
Implementace doporučovacího systému	50 000	5 000	Externí vývoj + provoz
Automatizovaný e-mailing	20 000	2 000	Implementace + provoz
Marketing automation nástroje	30 000	4 000	Implementace + licence
Analytický dashboard	15 000	1 000	Vývoj + provoz
Celkem technická implementace	155 000	12 000	–

3.4.2 Personální kapacity

Pro zajištění efektivního provozu a správu technických řešení je třeba personální kapacita rozdělena mezi více pozic (Tabulka 21).

Tabulka 21: Náklady na personální kapacitu

Pozice	Potřebný úvazek	Měsíční náklady [Kč]	Poznámka
Datový analytik	0,4	25 000	externí/interní
Specialista na digitální marketing	0,5	25 000	interní
Vývojář	0,2	18 000	externí
Celkem personální kapacity	1,1	68 000	–

3.4.3 Marketingové výdaje

Marketingové výdaje zahrnují aktivity pro zvýšení viditelnosti, zapojení cílového publika a optimalizaci výkonu kampaní (Tabulka 22).

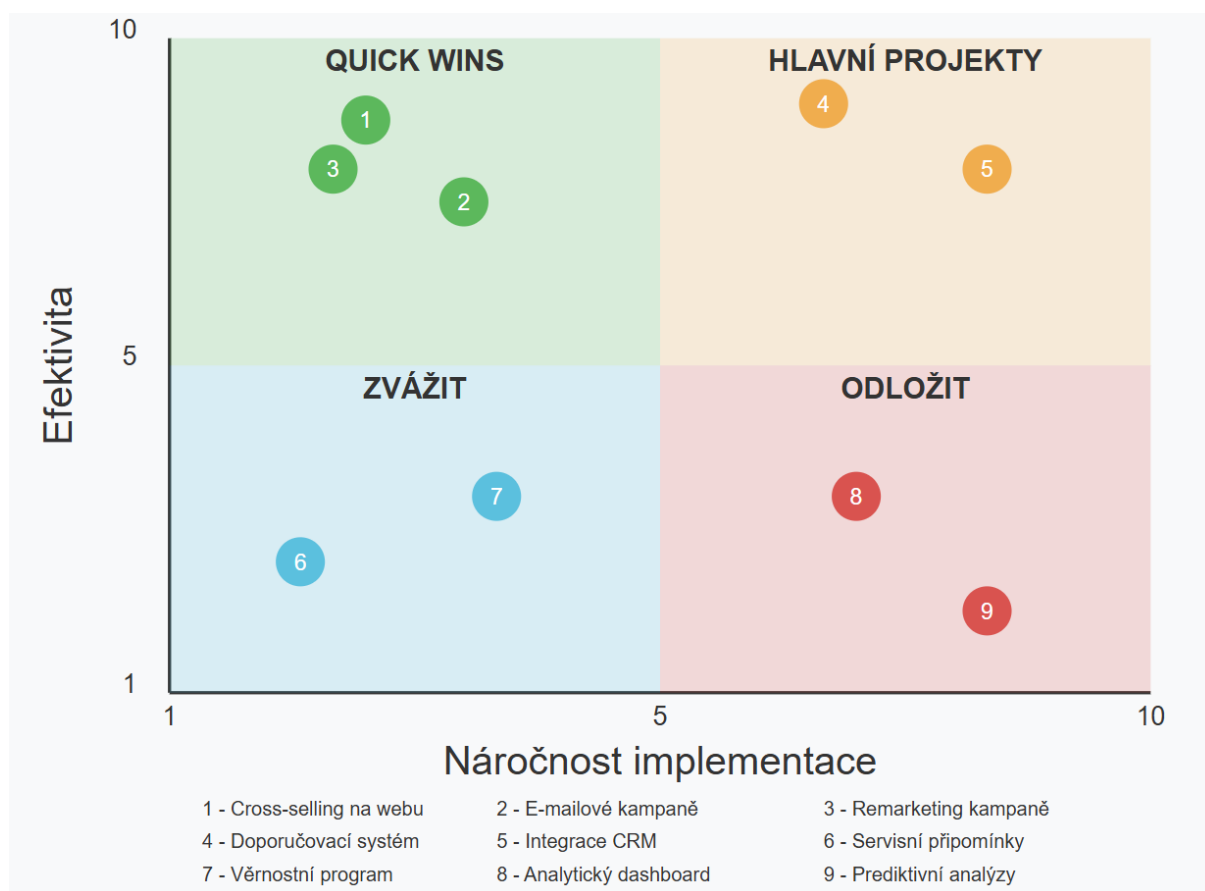
Tabulka 22: Náklady na marketingové aktivity

Položka	Roční náklady [Kč]	Poznámka
Remarketingové kampaně	120 000	Google, Facebook, Instagram
Obsahové kampaně	100 000	Tvorba obsahu + distribuce
Testování a optimalizace	60 000	A/B testování, UX
Celkové marketingové výdaje	280 000	–

3.5 Prioritizace návrhů

Pro efektivní implementaci navržených řešení byla vytvořena matice prioritizace založená na dvou parametrech (Obrázek 17).

- Efektivita – potenciální přínos pro business (1-10, kde 10 je nejvyšší)
- Náročnost – komplexita implementace a potřebné zdroje (1-10, kde 10 je nejnáročnější)



Obrázek 17: Graf prioritizace návrhů

3.6 Návrh metrik pro měření úspěšnosti projektu

1. Ekonomické metriky
 - Průměrná hodnota objednávky
 - Zisk na zákazníka
 - Míra konverze
2. Zákaznické metriky
 - Míra retence zákazníků
 - Frekvence nákupů
 - Doba mezi nákupy
3. Marketingové metriky
 - Efektivita remarketingových kampaní
 - Úspěšnost cross-sellingu
 - Úspěšnost predikovaných následných nákupů
4. Zákaznická zkušenost
 - Míra odhlášení z marketingových kampaní

3.7 Kritická omezení současné analýzy

Současná analýza identifikovala významné vzorce v nákupním chování zákazníků, implementace plnohodnotného prediktivního modelování je limitována absencí klíčových webových dat. Pro vytvoření predikcí budoucího nákupního chování jsou nezbytná data o interakcích zákazníků s webem, včetně sledování prohlížených produktů, času stráveného na jednotlivých stránkách, opuštěných košíků a cesty zákazníka napříč webem. Po získání těchto dat bude možné implementovat dva účinné prediktivní modely. Rozhodovací strom, který dokáže efektivně predikovat pravděpodobnost následného nákupu na základě komplexních interakcí mezi nákupní historií a webovým chováním, a logistickou regresi pro identifikaci zákazníků s vysokou pravděpodobností odchodu. Vzhledem k tomu, že webová data představují užitečný zdroj informací o zákaznických preferencích a záměrech, bylo rozhodnuto odložit vývoj prediktivních modelů do doby, kdy bude implementována pokročilá webová analytika schopná zachytit tyto interakce. Tento přístup zajistí, že budoucí prediktivní modely budou postaveny na kompletním obrazu zákaznického chování, nikoli pouze na transakčních datech, což výrazně zvýší jejich přesnost a praktickou využitelnost pro marketingové rozhodování.

3.7.1 Datová omezení

Absence demografických údajů – Analýza je založena výhradně na transakčních datech bez znalosti věku, pohlaví a dalších demografických charakteristik, které by mohly významně přispět k přesnější segmentaci a cílení. Taková data ale nejsou shromažďována a ukládána.

Neznámé komunikační preference – Absence dat o preferovaných komunikačních kanálech zákazníků znesnadňuje optimální nastavení komunikačního mixu.

Chybějící behaviorální data z webu – Analýza nepropojuje transakční data s chováním zákazníků na webu před nákupem (prohlížené produkty, čas strávený na stránkách, opuštěné košíky).

Absence dat o konkurenci – Nelze vyhodnotit relativní tržní pozici společnosti a identifikovat konkurenční výhody či nevýhody.

3.7.2 Metodologická omezení

Statičnost segmentace: Současná segmentace představuje pouze statický pohled, který nezachycuje přechody zákazníků mezi segmenty v čase.

Absence kontextuálních informací: Analýza nezohledňuje vnější faktory ovlivňující nákupní chování (sezónnost, počasí, ekonomické ukazatele).

Jednostranný pohled na zákazníka: Chybí kvalitativní data o motivacích, spokojenosti a bariérách, které by doplnily kvantitativní analýzu.

3.7.3 Technická omezení

Omezená podrobnost dat – Data nejsou dostatečně detailní pro některé typy analýz (např. typ zařízení použité k nákupu).

3.8 Doporučení pro další analýzy

3.8.1 Pokročilé prediktivní modelování chování zákazníků

Očekávané přínosy:

- Zvýšení přesnosti cílení marketingových aktivit
- Identifikace rizikových zákazníků s vysokou pravděpodobností odchodu s předstihem 1-3 měsíců
- Optimalizace rozpočtu na retenci díky proaktivnímu přístupu

3.8.2 Analýza zákaznického životního cyklu

Metodologický přístup:

- Vytvoření detailních zákaznických cest pro každý segment
- Identifikace klíčových momentů v zákaznickém životním cyklu
- Analýza délky vztahu a faktorů ovlivňujících loajalitu podle segmentů

Technické požadavky:

- Implementace pokročilých nástrojů pro vizualizaci zákaznické cesty
- Sběr dat o všech interakcích zákazníka napříč kanály
- Vytvoření systému pro sledování změn v zákaznickém chování v čase

Očekávané přínosy:

- Identifikace kritických bodů v zákaznické cestě s potenciálem pro optimalizaci
- Zvýšení retence zákazníků díky cíleným intervencím v kritických momentech
- Prodloužení průměrné doby zákaznického vztahu

3.8.3 Systematické testování a optimalizace konverze

Metodologický přístup:

- Vytvoření strukturovaného programu testování pro všechny marketingové kanály
- Testování různých typů sdělení a vizuálů pro jednotlivé segmenty
- Testování různých kombinací nabídek, slev a produktových doporučení

Technické požadavky:

- Implementace specializovaných nástrojů pro testování (např. Google Optimize)
- Nastavení statisticky validních experimentů s dostatečnou velikostí vzorku
- Vytvoření metodiky pro pravidelné vyhodnocování a implementaci výsledků testů

Očekávané přínosy:

- Zvýšení konverzního poměru na e-shopu
- Optimalizace ROI marketingových kampaní
- Vytvoření data-driven kultury v marketingovém oddělení

3.8.4 Hlubková analýza webových interakcí a nákupního procesu

Metodologický přístup:

- Implementace pokročilého webového analytického systému se sledováním uživatelských cest
- Analýza heat map a pro identifikaci problematických míst na webu
- Analýza opuštěných košíků a nedokončených transakcí včetně identifikace příčin

Technické požadavky:

- Propojení webové analytiky s CRM systémem pro komplexní pohled na zákazníka
- Zavedení pokročilého e-commerce sledování pro důkladnější analýzu nákupního procesu.

Očekávané přínosy:

- Snížení míry opuštění košíku
- Zvýšení míry dokončení nákupu
- Identifikace a odstranění překážek v nákupním procesu

3.8.5 Analýza sentimentu a zpětné vazby zákazníků

Metodologický přístup:

- Implementace systematického sběru zpětné vazby od zákazníků
- Analýza sentimentu z recenzí, komentářů a interakcí na sociálních sítích
- Identifikace hlavních faktorů ovlivňujících spokojenost v jednotlivých segmentech

Technické požadavky:

- Nasazení nástrojů pro analýzu textu a sentimentu
- Vytvoření automatizovaného systému sběru zpětné vazby napříč kanály
- Propojení dat o sentimentu s transakčními daty pro korelační analýzy

Očekávané přínosy:

- Zvýšení zákaznické spokojenosti
- Včasná identifikace a řešení problémů s produkty či službami
- Využití pozitivní zpětné vazby pro marketingové účely

3.8.6 Implementace prediktivní analýzy nákupního košíku

Metodologický přístup:

- Vytvoření prediktivního modelu pro doporučování produktů v reálném čase
- Analýza sekvenčních nákupních vzorců pro předpověď následujících nákupů
- Implementace personalizovaných doporučení na základě podobnosti zákazníků

Technické požadavky:

- Nasazení doporučovacího systému
- Vytvoření systému pro automatickou aktualizaci doporučení na základě nových dat
- Implementace testování různých algoritmů doporučování

Očekávané přínosy:

- Zvýšení průměrné hodnoty objednávky
- Zlepšení uživatelské zkušenosti díky relevantním doporučením
- Zvýšení konverzního poměru u doporučených produktů

Implementací těchto pokročilých analytických přístupů společnost významně prohloubí porozumění svým zákazníkům a vytvoří robustní datově řízený systém marketingového rozhodování. Systematické využívání těchto analýz povede k výraznému zvýšení efektivity marketingových aktivit, personalizaci zákaznické zkušenosti a v konečném důsledku k růstu tržeb a ziskovosti.

4 ZÁVĚR

Tato bakalářská práce se zaměřila na analýzu dat z e-shopu za účelem vytvoření podkladů pro cílenou reklamu na internetu. V rámci analytické části byly provedeny různé typy analýz, které odhalily zajímavé vzorce nákupního chování zákazníků a identifikovaly potenciální příležitosti pro zlepšení marketingových strategií.

Prvním krokem analýzy bylo zpracování a příprava datových souborů z databáze e-shopu. Tato fáze zahrnovala čištění dat, jejich strukturování do vhodného formátu a vytvoření SQL databáze pro efektivní dotazování. Následně byly aplikovány různé analytické metody s využitím jazyků SQL a Python, včetně knihoven pro data mining a analýzu asociačních pravidel.

V rámci analýzy nákupních košíků byly identifikovány produkty a kategorie, které jsou zákazníci často nakupovány společně. Mezi nejčastější kombinace patří stojánky na kola (produkt 7820) s gripy na řídítka (produkt 4204) a rohy na řídítka (produkt 5834). Tyto informace mohou být využity pro okamžitou nabídku souvisejících produktů v e-shopu při přidání určitého produktu do košíku, případně pro cílené e-mailové kampaně.

Hlavní částí práce byla analýza návaznosti produktů a kategorií v průběhu zákaznického cyklu. Zde bylo vytvořeno 10000 nejčastějších dvojic produktů, které je možné použít pro cílenou reklamu. Data obsahují informace o četnosti výskytu i o pravděpodobnosti přechodu na daný produkt. Bylo zjištěno, že někteří zákazníci se po nákupu kola často vrací na servis, což potvrzuje dobrou zákaznickou retenci. Dále byly identifikovány produkty, které zákazníci nakupují v následujících objednávkách, společně s průměrnou dobou mezi těmito nákupy a pravděpodobností přechodu mezi produkty.

Data ukazují, že po zakoupení stojánku na kolo (7820) následuje s pravděpodobností 2,6 % nákup gripů (4204) v průměrném časovém odstupu 87 dní. Podobně po nákupu gripů (4204) následuje s pravděpodobností 9,8 % nákup rohů na řídítka (5834) v průměrném odstupu 159 dní. Tyto informace jsou cenné pro načasování cílených reklamních kampaní.

Provedené analýzy poskytují komplexní pohled na nákupní chování jejich zákazníků a konkrétní údaje pro optimalizaci marketingových aktivit. Výsledky mohou být použity pro personalizaci nabídek produktů na základě předchozích nákupů, efektivnější načasování cílených reklamních kampaní, optimalizaci cross-sellingových strategií v e-shopu, zlepšení zákaznické zkušenosti díky relevantnějším doporučením.

Pro praktické využití těchto poznatků doporučuji implementovat automatizovaný systém, který bude na základě analytických výstupů spouštět cílené reklamní kampaně v optimálním čase po nákupu určitého produktu. Zároveň by bylo vhodné průběžně vyhodnocovat efektivitu těchto kampaní a případně upravovat parametry na základě nových dat.

Dalším krokem by mohlo být rozšíření analýzy o demografické údaje zákazníků a zohlednění sezónních trendů, což by umožnilo ještě přesnější cílení reklamních sdělení. Rovněž by bylo přínosné provést analýzu cenové elasticity poptávky pro různé kategorie produktů, která by mohla poskytnout podklady pro optimalizaci cenových strategií.

Závěrem lze konstatovat, že datová analýza nákupního chování zákazníků společnosti přinesla konkrétní a prakticky využitelné výstupy, které mohou významně přispět ke zlepšení marketingových strategií a v konečném důsledku i ke zvýšení tržeb společnosti.

SEZNAM POUŽITÝCH ZDROJŮ

- [1] HOLMAN, Robert. *Ekonomie*. 6. vydání. V Praze: C.H. Beck, 2016. ISBN 978-80-7400-278-6.
- [2] MUSIL, Petr. *Ekonomie*. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk, 2008. ISBN 978-80-7380-126-7.
- [3] FARKAČOVÁ, Lenka. *Neučebnice ekonomie: Pro každého na každý den*. Grada, 2021. ISBN 8027131936.
- [4] JUREČKA, Václav a MACHÁČEK, Martin. *Makroekonomie*. 4., aktualizované a rozšířené vydání. Praha: Grada Publishing, 2023. ISBN 978-80-271-3635-3.
- [5] JUREČKA, Václav. *Mikroekonomie*. 3., aktualizované a rozšířené vydání. Praha: Grada Publishing, 2018. ISBN 978-80-271-0146-7.
- [6] HOŘEJŠÍ, Bronislava. *Mikroekonomie*. 5., aktualizované vydání. Praha: Management Press, 2010. ISBN 978-80-7261-218-5.
- [7] SEDLÁČEK, Jaroslav. *Finanční analýza podniku*. Brno: Computer Press, 2007. ISBN 978-80-251-1830-6.
- [8] PAVELKA, Tomáš. *Makroekonomie: základní kurz*. 3. vyd. Slaný: Melandrium, 2007. ISBN 978-80-86175-58-4.
- [9] SOUKUP, Jindřich; POŠTA, Vít; NESET, Pavel a PAVELKA, Tomáš. *Makroekonomie*. 4. aktualizované vydání. Praha: Management Press, 2022. ISBN 978-80-7261-596-4.
- [10] STOKLASA, Jan. *Kapitoly z matematické ekonomie*. Univerzita Palackého v Olomouci, 2023. ISBN 9788024463640.
- [11] KUBANOVÁ, Jana. *Statistické metody pro ekonomickou a technickou praxi*. 3. Bratislava: STATIS, 2008. ISBN 978-80-85659-474.
- [12] BAG, Dinabandhu. *Business Analytics*. London: Routledge, 2017. ISBN 9781138916111. Dostupné z: <https://doi.org/10.4324/9781315464695>.
- [13] HAN, Jiawei; KAMBER, Micheline a PEI, Jian. *Data mining: concepts and techniques*. 2nd Edition. Waltham, Mass.; Tokyo: Morgan Kaufmann, 2012. ISBN 0123814790. Dostupné z: <https://doi.org/10.1016/C2009-0-61819-5>.
- [14] KRAUS, Mathias; FEUERRIEGEL, Stefan a OZTEKIN, Asil. Deep learning in business analytics and operations research: Models, applications and managerial implications.

- Online. *European journal of operational research*. 2020, roč. 281, č. 3, s. 628-641. ISSN 0377-2217. Dostupné z: <https://doi.org/10.1016/j.ejor.2019.09.018>. [cit. 2025-03-18].
- [15] GOLDFARB, Avi a TUCKER, Catherine. Online Display Advertising: Targeting and Obtrusiveness. Online. *Marketing science (Providence, R.I.)*. 2011, roč. 30, č. 3, s. 389-404. ISSN 0732-2399. Dostupné z: <https://doi.org/10.1287/mksc.1100.0583>. [cit. 2025-03-18].
- [16] FOURBERG, Niklas; TAŞ, Serpil; WIEWIORRA, Lukas; GODLOVITCH, Ilsa; DE STREEL, Alexandre et al. *Online advertising: the impact of targeted advertising on advertisers, market access and consumer choice*. Brussels: European Parliament, 2021. ISBN 9284681316.
- [17] POLÁČKOVÁ, Julie. Data Mining v praxi: segmentace zákazníků dle nákupního chování. Online. *Scientific Papers of the University of Pardubice. Series D, Faculty of Economics and Administration*. 2011, roč. 2011, č. 20, s. 135. ISSN 1211-555X. Dostupné z: <https://www.proquest.com/docview/2265541316>. [cit. 2025-03-07].
- [18] *Privacy Technologies and Policy*. 9857. Switzerland: Springer International Publishing, 2016. ISBN 9783319447599. ISSN 0302-9743.
- [19] *GDPR: nové příležitosti, nové povinnosti: co musí každý podnik vědět o obecném nařízení EU o ochraně osobních údajů*. Luxembourg: Publications Office, 2018. ISBN 9279794272.
- [20] ÚOOÚ uložil pokutu 351 mil. Kč za porušení GDPR. Online. In: Úřad pro ochranu osobních údajů. 2024. Dostupné z: <https://uouu.gov.cz/novinky/vse/uouu-ulozil-pokutu-351-mil-kc-za-poruseni-gdpr>. [cit. 2025-04-08].
- [21] KROPÁČ, Jiří. *Statistika B: jednorozměrné a dvourozměrné datové soubory, regresní analýza, časové řady*. 2., dopl. vyd. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2009. ISBN 978-80-214-3984-9.
- [22] BARTÍK, Vladimír. *4. Získávání znalostí z databází (Data Mining)* [interní přednáška], Datové sklady, fakulta podnikatelská; Brno: VUT v Brně, Czechia, 4. listopadu 2024.
- [23] LUENGO, Julián; GARCÍA-GIL, Diego; RAMÍREZ-GALLEGO, Sergio; GARCÍA, Salvador a HERRERA, Francisco. *Big Data Preprocessing: Enabling Smart Data*. Cham: Springer International Publishing, 2020. ISBN 3030391043. Dostupné z: <https://doi.org/10.1007/978-3-030-39105-8>.
- [24] LEE, Dongwon; PARK, Sung-hyuk a MOON, Songchun. Utility-based association rule mining: A marketing solution for cross-selling. Online. *Expert systems with applications*. 2013, vol. 40, no. 7, s. 2715-2725. ISSN 0957-4174. Dostupné z: <https://doi.org/10.1016/j.eswa.2012.11.021>. [cit. 2025-04-23].

- [25] WONG, Rcw; FU, Awc a WANG, K. Data mining for inventory item selection with cross-selling considerations. Online. *Data mining and knowledge discovery*. 2005, vol. 11, no. 1, s. 81-112. ISSN 1384-5810. Dostupné z: <https://doi.org/10.1007/s10618-005-1359-6>. [cit. 2025-04-23].
- [26] *Česká pošta* [online]. 2025 [cit. 2025-04-23]. Dostupné z: <https://www.ceskaposta.cz/ke-stazeni/zakaznicke-vystupy>
- [27] Matplotlib 3.10.1 documentation. *Matplotlib* [online]. 2012 [cit. 2025-04-23]. Dostupné z: <https://matplotlib.org/stable/index.html>
- [28] *NetworkX* [online]. 2014 [cit. 2025-04-23]. Dostupné z: <https://networkx.org/>
- [29] Sklearn.cluster. *Scikit-learn* [online]. 2025 [cit. 2025-04-23]. Dostupné z: <https://scikit-learn.org/stable/api/sklearn.cluster.html>
- [30] *Mlxtend* [online]. 2014, 26.1.2024 [cit. 2025-04-23]. Dostupné z: <https://pypi.org/project/mlxtend/>
- [31] STEWART, William J. *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton: Princeton University Press, 2009. ISBN 978-0-691-14062-9.