



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

VYUŽITÍ NÁSTROJŮ BUSINESS INTELLIGENCE PRO ANALÝZU A VIZUALIZACI DAT

USE OF BUSINESS INTELLIGENCE TOOLS FOR DATA ANALYSIS AND VISUALIZATION

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Tomáš Hrin

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jan Luhan, Ph.D., MSc

BRNO 2023

Zadání diplomové práce

Ústav: Ústav informatiky
Student: Bc. Tomáš Hrin
Vedoucí práce: Ing. Jan Luhan, Ph.D., MSc
Akademický rok: 2022/23
Studijní program: Informační management

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

Využití nástrojů Business Intelligence pro analýzu a vizualizaci dat

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:

Zaměření diplomové práce je do oblasti využití nástrojů Business Intelligence, jako podpory rozhodování v daném subjektu. Cílem práce je na základě poskytnutých zdrojových dat provést analýzu a vizualizace vybraných oblastí včetně efektivní přípravy dat.

Základní literární prameny:

HARINATH, S. et al. Professional Microsoft SQL Server 2012 Analysis Services with MDX and DAX. 1st ed. Indianapolis: Wiley, 2012. 1129 s. ISBN 978-1-118-10110-0.

CHMELÁR, M. Reporting v Power BI, PowerPivot a jazyk DAX. 1. vyd. Pezinok: Smart People, 2018. 557 s. ISBN 978-80-973078-0-6.

LABERGE, R. Datové sklady - Agilní metody a business intelligence. 1. vyd. Praha: Computer Press, 2012. 352 s. ISBN 978-80-251-3729-1.

NOVOTNÝ, O., J. POUR a D. SLÁNSKÝ. Business Intelligence: Jak využít bohatství ve vašich datech. 1. vyd. Praha: Grada, 2005. 256 s. ISBN 80-247-1094-3.7.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2022/23

V Brně dne 5.2.2023

L. S.

doc. Ing. Miloš Koch, CSc.
garant

doc. Ing. Vojtěch Bartoš, Ph.D.
děkan

Abstrakt

Diplomová práca sa zaoberá využitím nástrojov Business Intelligence na analýzu dát prihlášok a priebehu štúdia študentov Fakulty podnikateľskej. Náplňou práce je tvorba analýz, ktoré by mohli vedeniu univerzity pomôcť v rozhodovaní o budúcej stratégii a čistenie a modelovanie dát, ktoré sú využité na vizualizáciu vytvorených analýz. Na záver sú zhrnuté výsledky analýz s odporúčaním riešenia prípadných problémov.

Klíčová slova

VUT, Fakulta podnikateľská, Business Intelligence, Microsoft Power BI, čistenie a modelovanie dát, analýza dát, vizualizácia dát

Abstract

The diploma thesis deals with the use of Business Intelligence tools for the analysis of application data and the study progress of students of the Faculty of Business and Management. The content of the work is the creation of analyzes that could help the university management decide on the future strategy and the cleaning and modeling of data, which are used to visualize the created analyses. In the final part, the analysis results are summarized with recommendations for solving potential problems.

Keywords

BUT, Faculty of Business and Management, Business Intelligence, Microsoft Power BI, data cleaning and modeling, data analysis, data visualization

Bibliografická citace

HRIN, Tomáš. *Využití nástrojů Business Intelligence pro analýzu a vizualizaci dat* [online]. Brno, 2023 [cit. 2023-05-12]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/152078>. Diplomová práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav informatiky. Vedoucí práce Ing. Jan Luhan, Ph.D., MSc.

Čestné prohlášení

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

V Brně dne 12. 5. 2023

Bc. Tomáš Hrin

Autor

Pod'akovanie

Týmto spôsobom by som sa rád pod'akoval pánovi Ing. Janu Luhanovi, Ph.D., MSc. za odborné vedenie mojej práce. Za cenné rady a poskytnutie nápadu na náplň mojej diplomovej práce ďakujem pánovi Ing. Jiřímu Křížovi, Ph.D a Fakulte podnikateľskej.

Obsah

Úvod.....	11
Ciele práce, metódy a postupy spracovania.....	12
1 Teoretické východiská práce	13
1.1 Základné pojmy.....	13
1.1.1 Dáta.....	13
1.1.2 Informácia.....	13
1.1.3 Znalosť.....	14
1.1.4 Systém.....	14
1.2 Business Intelligence.....	15
1.2.1 Definícia.....	15
1.2.2 Vrstvy Business Intelligence	15
1.2.3 Databáza.....	16
1.2.4 Reporting, dolovanie dát a OLAP.....	18
1.2.5 Prezentácia dát a nástroje Business Intelligence.....	20
1.3 Microsoft Power BI.....	21
1.3.1 Čo je Power BI?.....	21
1.3.2 Jazyk DAX a jazyk M.....	23
1.4 Jazyk Python	24
1.5 Fuzzy párovanie	24
2 Analýza súčasného stavu	26

2.1	Predstavenie univerzity VUT a fakulty podnikateľskej.....	26
2.1.1	Univerzita VUT	26
2.1.2	Fakulta podnikateľská.....	26
2.2	Priebeh štúdia študentov	28
2.2.1	Prijímacie riadenie	28
2.2.2	Priebeh štúdia študentov od nástupu po ukončenie	29
2.3	Analýzy, čo by vedenie univerzity mohli zaujímať a ich prínosy	30
2.3.1	Analýza regionálneho pôvodu študentov a ich výber oborov.....	30
2.3.2	Analýza vývoja počtu prihlášok a úspešnosti študentov v čase.....	30
2.3.3	Analýza úspešnosti študentov v rôznych fázach štúdia v rôznych oboroch 31	
2.3.4	Analýza vzťahu medzi úspechom na prijímacích skúškach a vysokoškolským úspechom	31
2.3.5	Analýza dokončenia štúdia	31
2.3.6	Analýza stredných škôl podľa typu	32
2.3.7	Analýza vysokých škôl a fakúlt	32
3	Vlastné návrhy riešenia.....	33
3.1	Popis získaných dát	33
3.1.1	Popis prvého súboru.....	33
3.1.2	Popis druhého súboru.....	34
3.2	Čistenie dát.....	35

3.2.1	Čistenie prvého súboru	35
3.2.2	Čistenie druhého súboru	50
3.3	Spojenie dát	51
3.4	Vizualizácia analýz	53
3.4.1	Analýza regionálneho pôvodu študentov a ich výber oborov.....	53
3.4.2	Analýza vývoja počtu prihlášok a úspešnosti študentov v čase.....	56
3.4.3	Analýza úspešnosti študentov v rôznych fázach štúdia v rôznych oboroch	59
3.4.4	Analýza vzťahu medzi úspechom na prijímacích skúškach a vysokoškolským úspechom	60
3.4.5	Analýza dokončenia štúdia	62
3.4.6	Analýza stredných škôl podľa typu	63
3.4.7	Analýza vysokých škôl podľa názvov a fakúlt	65
3.5	Zhrnutie výsledkov analýz a návrh opatrení	66
3.5.1	Pokles prihlášok.....	66
3.5.2	Začaté a nedokončené štúdiá	67
3.5.3	Zlé študijné výsledky bakalárskych oborov a kombinovaných štúdií	68
	Záver	69
	Zoznam použitej literatúry	70
	Zoznam obrázkov	72

Úvod

V minulosti bol vysokoškolský titul vzácnosť, dnes je bežným javom ho mať pred menom. Po skončení strednej školy študenti môžu hneď začať pracovať, no mnoho z nich sa uchýli k štúdiu práve na vysokej škole. Môžu na to mať rôzne dôvody. Niektorí sa na vysokú školu prídu čo najviac naučiť, nech pri nástupe do zamestnania majú odborné znalosti. Niektorí sa chcú v budúcnosti venovať výskumu a tak sa rozhodnú ísť vyštudovať školu vrátane tretieho stupňa. Väčšina študentov však po skončení strednej školy ešte nevie, čo chcú v dnešnom svete neobmedzených možností vykonávať za povolanie, a tak sa rozhodnú na to prísť na vysokej škole, kde sa niečo naučia a ešte aj zažijú príjemný vysokoškolský život.

Študentov vysokých škôl je v dnešnej dobe naozaj mnoho a táto práca sa bude zaoberať práve študentmi na univerzite Vysokého učení technického v Brne, konkrétne na Fakulte podnikateľskej. Práca sa bude zaoberať analýzou dát tvorených štúdiami študentov na fakulte, a to konkrétne ich prihláškami a ich štúdiami. Mnoho študentov znamená mnoho dát, na základe ktorých sa dajú získať cenné informácie a zlepšiť tak manažérske rozhodovanie vedenia.

Ciele práce, metódy a postupy spracovania

Zameranie diplomovej práce je využiť nástroje Business Intelligence ako podporu rozhodovania pre vedenie Fakulty podnikateľskej. Cieľom práce je na základe poskytnutých zdrojových dát vytvoriť analýzy a vizualizácie vybraných oblastí vrátane efektívnej prípravy dát v Microsoft Power BI.

V tejto práci bude najprv v teoretickej časti pripravený teoretický základ nutný na pochopenie problematiky v tejto práci.

V analytickej časti bude predstavená univerzita VUT a fakulta podnikateľská, popísaný priebeh štúdia študentov a na záver budú zostrojené analýzy, ktoré by mohli vedeniu fakulty priniesť cenné informácie.

V návrhovej časti práce budú popísané získané dáta o študentoch a následne budú vyčistené a pripravené do formy na analýzu v Power BI. V závere budú vizualizované analýzy z analytickej časti a budú z nich vyvodené závery a rady, ako prípadné problémy či nedostatky napraviť.

1 Teoretické východiská práce

V teoretickej časti diplomovej práce bude vytvorený teoretický základ nutný pre pochopenie ďalších častí diplomovej práce. Z úvodu to budú základné pojmy a po ich vysvetlení sa prejde na pojmy síce komplexnejšie a odbornejšie, ale často využívané v ostatných častiach tejto práce.

1.1 Základné pojmy

Úvodná kapitola bude zameraná na vysvetlenie teoretických pojmov, ktoré sú základom pre pochopenie business Intelligence.

1.1.1 Dáta

Dáta sú fakty získané vnímaním okolitého sveta. Dáta môžeme čítať, vidieť, počuť, hmatat' či inak interpretovať a ďalej ich spracovávať.

Z hľadiska práce s dátami je možné ich rozdeliť na:

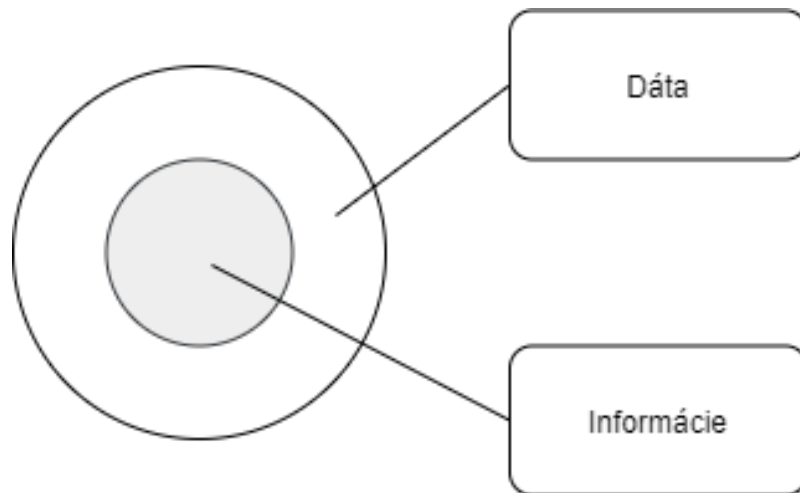
- **Štruktúrované dáta:** fakty alebo vnemy, ktoré disponujú určitou štruktúrou pri vytváraní a ukladaní dát, ktorá umožňuje ľahko vybrať tie dáta, ktoré sú potrebné na vyriešenie určitého problému
- **Neštruktúrované dáta:** nemajú štruktúru a tým pádom nie je jednoduché vybrať konkrétne dáta a využiť ich na riešenie určitého problému. Je to vlastne tok bitov bez ďalšieho rozlíšenia ako napríklad textový dokument alebo zvuková nahrávka či videozáznam (1)

1.1.2 Informácia

Informácie sú **dáta**, ktoré ich prijímateľ vie spracovať, rozumie im a vie ich aj využiť. Sú to teda dáta, ktoré si človek spracuje pomocou mozgu a majú pre neho určitý význam alebo kontext. Všetky informácie sú teda dáta, no nie všetky dáta sú informáciou. Podmienkou je, aby prijímateľovi dát priniesli niečo nové a užitočné. (2)

Ako príklad sa dá uviesť komunikácia s človekom v cudzom jazyku, ktorému nerozumieme. Opýtame sa ho na otázku, na ktorú potrebujeme odpoveď a čakáme od

neho reakciu v materinskom jazyku, ktorá pre nás bude užitočnou informáciou. Ak nám však odpovie v cudzom jazyku, ktorému nerozumieme, budú to pre nás len dáta, pretože si to nevieme interpretovať a nemá to žiaden význam.



Obrázok č. 1: Dáta a informácie

(Zdroj: Vlastné spracovanie)

1.1.3 Znalosť

Po správnej interpretácii dát vznikla informácia, ktorá bola uložená do nášho mozgu. V momente keď v praktickej situácii vieme informáciu využiť, stáva sa **znalosťou**. Informácia je výsledkom interpretácie dát na základe vlastných schopností, no znalosť je formou aktívneho učenia a sú vytvárané pomocou zložitého systému individuálne osvojovaných skúseností, faktov, vzťahov či myšlienkových procesov a významov.

1.1.4 Systém

Systém je usporiadaná množina prvkov, ktoré majú určité vlastnosti a vzťahy medzi sebou. Spolu tvoria celok, ktorý plní určitý účel. Ak každý prvok funguje správne, ale nespolupracujú spolu prvky, systém svoj účel neplní. Zmena jedného prvku zmení aj tie ostatné. (1)

1.2 Business Intelligence

1.2.1 Definícia

Business Intelligence je termín, ktorý zastrešuje znalosti, procesy, technológie, aplikácie a postupy, ktoré napomáhajú rozhodovaniu. Účelom business Intelligence je pracovať s dátami z minulosti a využiť ich tak, aby napomáhali rozhodovaniu v budúcnosti. BI je teda úzko prepojená s plánovaním budúcnosti. (3)

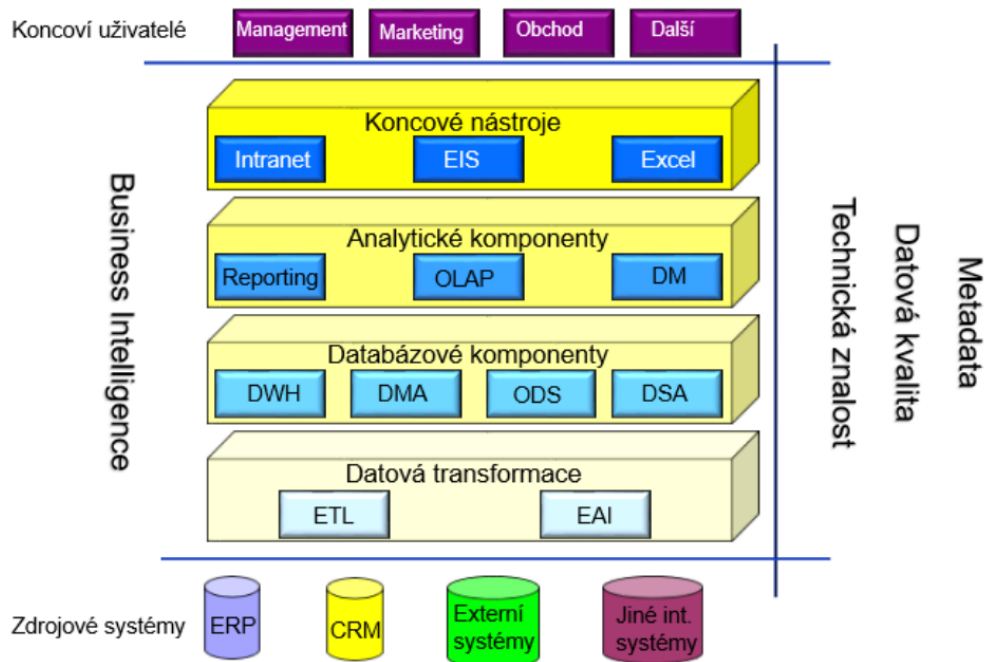
Hlavným prínosom Business Intelligence je uľahčiť rozhodovanie o smerovaní podniku, ale aj každodenné rozhodovanie o bežných činnostiach, ktoré robia aj nižší manageri. Aby si firma udržala prednú priečku v jej obore, BI je v dnešnej dobe nutnosť, pretože vďaka nej si firmy šetria čas, dokážu analyzovať seba či svojich zákazníkov a rozhodujú sa objektívne. (4)

1.2.2 Vrstvy Business Intelligence

Čo konkrétne teda Business Intelligence je? BI je súhrn prístupov a aplikácii ICT, ktoré sa dajú rozdeliť na vrstvy:

- **Vrstva na extrakciu, transformáciu a nahranie dát:** na túto vrstvu patria ETL a EAI systémy. **ETL** (Extraction, transformation, load) je trojstupňový proces, ktorý umožňuje zbierať dáta z rôznych zdrojov, Po zbere dát ich transformuje, čo znamená ich čistenie a úprava na požadovanú prehľadnú formu a na záver ich uloží do jedného centralizovaného miesta. **EAI** (Enterprise Application Integration) je proces pracujúci v reálnom čase vo vrstve zdrojových aplikácií, kde je integrovaný a využíva sa na prenos dát do dátových úložísk.
- **Vrstva na ukladanie dát:** vrstva zaisťuje procesy ukladania, aktualizácie s správou dát. Patria sem **dátové sklady**, dátové trhy, **databázy** a operatívne a dočasné úložiska dát. **Databáza** slúži na ukladanie štruktúrovaných a spojitých údajov pre bežné transakcie, zatiaľ čo **dátový sklad** sa zameriava na zbieranie a ukladanie veľkého množstva neštruktúrovaných dát, ktoré sa používajú pre analýzu a vizualizáciu – jednoducho povedané, je to akési úložisko dát.
- **Vrstva na analýzu dát:** vrstva zastrešuje komponenty, ktoré sa používajú na detailnú analýzu dát a to konkrétne reporting, dolovanie dát a OLAP

- **Vrstva na prezentovanie dát:** vrstva na komunikáciu užívateľov s jednotlivými komponentami BI. Na tejto vrstve používateľ zadá požiadavku systému BI na analýzu či prezentáciu určitých dát, napríklad pomocou EIS alebo Microsoft Excel. (5)



Obrázok č. 2: Vrstvy BI

Zdroj: (5)

1.2.3 Databáza

Databáza je základným pojmom v oblasti Business Intelligence a predstavuje organizovanú kolekciu štruktúrovaných informácií alebo dát, ktoré sú typicky uložené elektronicky pomocou počítačového softvéru. Databázy slúžia na uchovávanie a správu rôznych typov údajov, ktoré sú dôležité pre fungovanie organizácií a podnikov. Databáza je ako veľké úložisko dát, ktoré je možné porovnať s kartotékami alebo súborovými systémami. Kartotéky sú záznamy uložené na papieri, napríklad v lekárskech ambulanciách, a súborové systémy znamenajú usporiadané súbory a adresáre uložené na počítačovom médiu, ako je pevný disk. (3)

V dnešnej dobe je pojem databáza širší a zahŕňa niekoľko súčastí, ktoré spolu tvoria databázový systém. Tento systém sa skladá z troch hlavných častí. Prvou časťou je systém riadenia databázy (DBMS), ktorý je softvérový program schopný vytvárať, meniť a mazať informácie z databázy. DBMS je zodpovedný za správu a organizáciu databázy a manipuláciu s dátami pomocou špeciálneho jazyka, ako je napríklad jazyk SQL. Druhou časťou je databázová aplikácia, ktorá je zložená z programov, umožňuje komunikáciu medzi užívateľom a systémom riadenia databázy. Tieto aplikácie tiež využívajú jazyk SQL na prácu s dátami a získavanie požadovaných informácií. Tretou časťou sú samotné databázy, ktoré obsahujú skutočné dáta a informácie, ktoré sú relevantné pre konkrétnu organizáciu. Databázy môžu byť rôzneho typu, napríklad relačné, objektové, NoSQL alebo dátové sklady, a obsahujú tabuľky, stĺpce a riadky, ktoré uchovávajú údaje o entitách a ich vzťahoch. (6)

Pre správne fungovanie BI je dôležité mať k dispozícii kvalitné a spoľahlivé dáta. Preto je potrebné, aby databázy boli správne navrhnuté a organizované. Databázový model, ako napríklad relačný model, definuje logickú štruktúru databázy a spôsob, ako s dátami pracovať. Tabuľky v relačnej databáze obsahujú záznamy o entitách a ich vzťahy sú vytvárané pomocou kľúčov. Primárny kľúč je jedinečný identifikátor, ktorý slúži na jednoznačnú identifikáciu záznamov v tabuľke. V niektorých prípadoch môže byť potrebné použiť viacero stĺpcov na vytvorenie primárneho kľúča, čo sa nazýva zložený primárny kľúč. Cudzí kľúč je stĺpec alebo sada stĺpcov v jednej tabuľke, ktoré sa odkazujú na primárny kľúč inej tabuľky a slúžia na vytvorenie vzťahu medzi nimi, takzvanej **kardinality**. Sú rôzne typy kardinality:

- **1:1** – jednému záznamu v prvej tabuľke odpovedá jeden v druhej tabuľke
- **1:N** – jednému záznamu v prvej tabuľke odpovedá viac záznamov v druhej tabuľke
- **M:N** – viacero záznamov v jednej tabuľke odpovedá viacero záznamom v druhej tabuľke. Je vhodné tento vzťah zmeniť na 1:M a M:1 pomocou spojovacej tabuľky, no nemusí to byť nutné.

Pri analýze dát v BI sa využívajú rôzne techniky a nástroje. Jednou z hlavných techník je dotazovanie, ktoré umožňuje vyhľadávať a filtrovať dáta na základe určitých kritérií. Dotazovací jazyk SQL je často používaný na manipuláciu s dátami v databázach. (5)

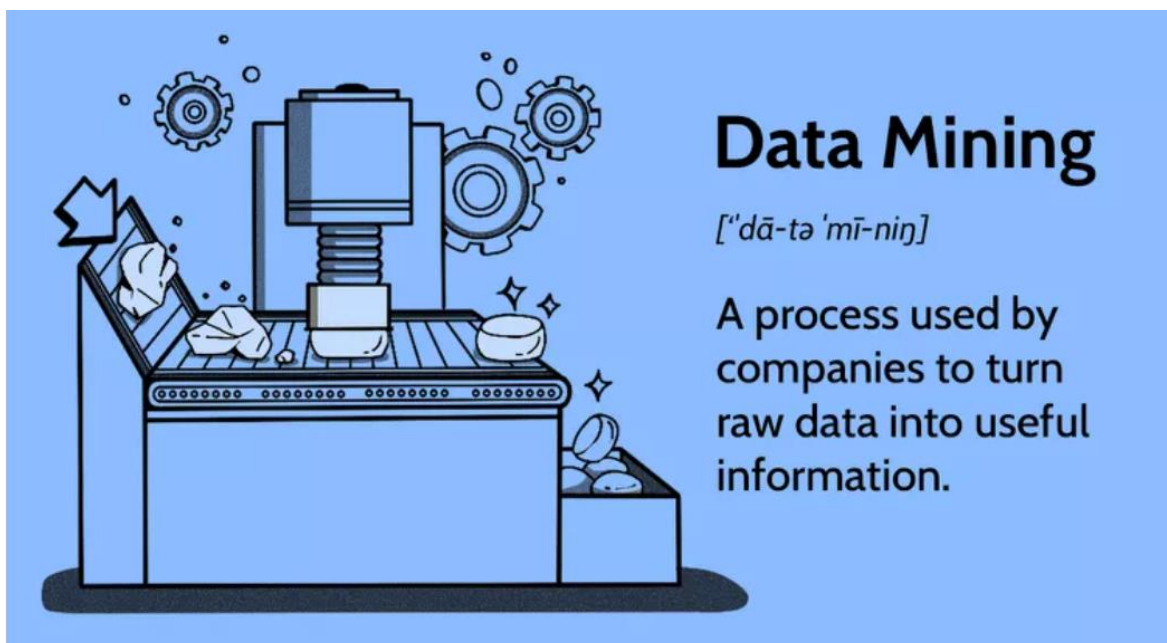
1.2.4 Reporting, dolovanie dát a OLAP

Dáta sú uložené a transformované, teraz je potrebné z nich získať podstatné informácie.

Prvým spôsobom ako to urobiť je **reporting**. Tento proces zabezpečuje spracovanie informácií z rôznych zdrojových systémov do prehľadnej a zrozumiteľnej formy, čo umožňuje zobrazit' výsledky podniku ako celku alebo jednotlivých častí, ako napríklad finančnú situáciu podniku, zamestnancov, alebo stav skladových zásob. Reporting môže byť členený na statický, dynamický a ad hoc reporting. Všeobecne platí, že obsah reportu závisí od cieľovej skupiny užívateľov, pričom užívateľov môžeme roztriediť do dvoch skupín: interných a externých. Interní užívatelia sú zvyčajne vlastníci firiem, manažéri a predstavitelia spoločnosti, ktorí sú zodpovední za výsledky podniku a potrebujú relevantné informácie pri rozhodovaní. Externí užívatelia sú odberatelia, dodávatelia, spolupracujúce firmy, banky v role veriteľa a štátne orgány, ktoré majú záujem o výkony a výsledky podniku z rôznych dôvodov. Pretože externí užívatelia majú rôzne záujmy a potreby, reporting musí byť navrhnutý tak, aby bol zrozumiteľný a relevantný pre každú skupinu užívateľov. Pri tvorbe reportov pre rôzne cieľové skupiny užívateľov je dôležité zohľadniť ich potreby a požiadavky. Každá skupina môže mať iné nároky na formát a obsah reportu. Napríklad manažéri podniku môžu potrebovať vysokú mieru detailov a presnosti, zatiaľ čo zamestnanci podniku môžu preferovať jednoduchšie a prehľadnejšie zobrazenie informácií. Ako príklad reportingu by mohol byť report o mesačných predajoch, ktorý by zohľadňoval celkové tržby za mesiac, tržby podľa produktov, podľa zákazníckych segmentov a podobne. (4),

Druhým spôsobom je **data mining**. Data mining je proces hľadania a analyzovania surových dát za účelom hľadania vzorov a spoločných znakov a tak extrahovať užitočné informácie. Pri dolovaní dát sa najprv dáta nahrajú do dátového skladu. Následne sa pomocou aplikácií na dolovanie dát hľadajú určité podobnosti a prepojenia medzi dátami, ktoré nie sú bežnému oku viditeľné. Sú rôzne prístupy ako tieto podobnosti hľadať, ako

napríklad neurónové siete alebo rozhodovacie stromy. Ako príklad by mohol slúžiť napríklad supermarket, ktorý chce z dát vydolovať niečo, čo mu pomôže pri predajoch potravín. Zhromaždí teda dáta o nákupoch v supermarkete a využije dolovací program, ktorý na základe zhľukovania vytvorí zhľuky potravín, ktoré najčastejšie zákazníci kupujú. Supermarket tak príde na to, že veľa zákazníkov si pri nákupe kečupu kúpi zároveň aj horčicu. Na základe tejto informácie, ak to doteraz neurobil, by mal supermarket umiestniť horčicu vedľa kečupu, aby svoje zisky zvýšil. Iným príkladom by mohla byť detekcia podvodov, kde data mining bude hľadať nekonzistencie v dátach. (7)

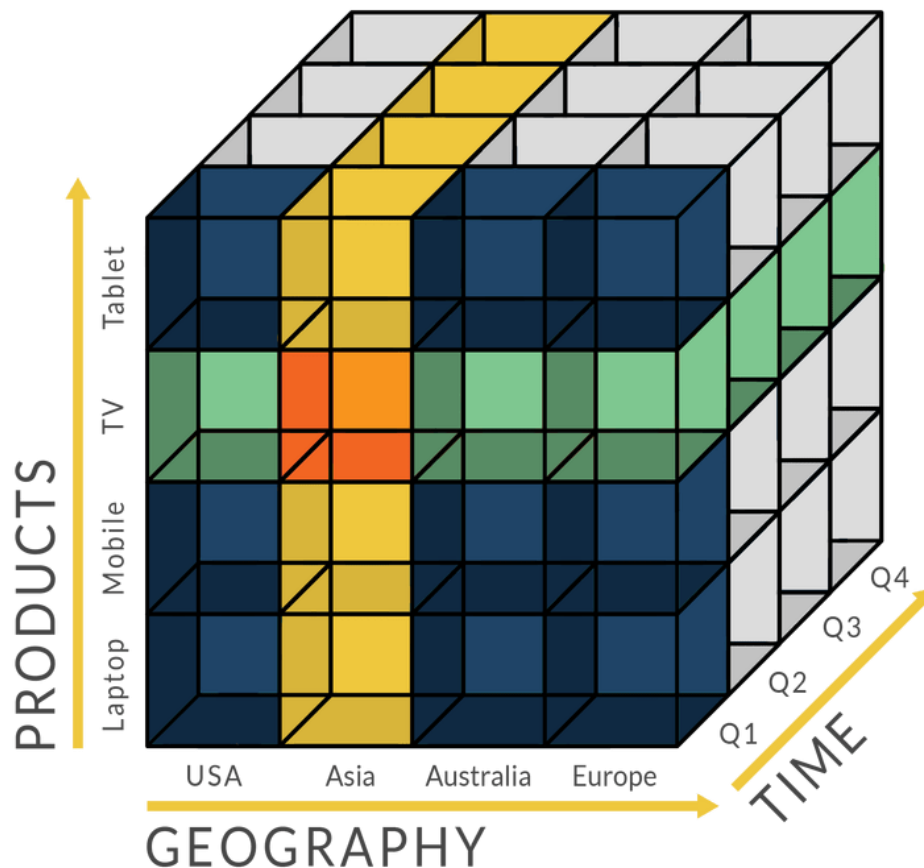


Obrázok č. 3: Data mining

(Zdroj: (7))

Tretí spôsob umožňujúci rýchlu analýzu dát je **OLAP kocka**. OLAP (Online Analytical Processing) kocka je multidimenzionálna databáza, ktorá sa používa na analytické spracovanie veľkého množstva dát. OLAP kocka je založená na koncepte viacrozmerných dátových kociek, ktoré zahrňujú dáta z rôznych faktorov a dimenzií. OLAP kocka poskytuje rýchly a efektívny prístup k analýze a vizualizácii dát, umožňuje užívateľom prehliadať dáta z rôznych perspektív a podľa rôznych kritérií. Napríklad,

môžu sa analyzovať predaje produktov z hľadiska regiónu, času, zákazníka alebo predajného kanálu. OLAP kocka sa skladá z dimenzií a faktov. Dimenzie predstavujú rôzne charakteristiky, podľa ktorých sa dáta môžu zoskupovať a analyzovať. Faktory sú numerické hodnoty, ktoré sa zobrazujú vo výsledkoch analýzy. OLAP kocky sú často používané v obchodnom a finančnom sektore, ale aj v iných oblastiach, kde je potrebné analyzovať a vizualizovať veľké množstvo dát. (8)



Obrázok č. 4: OLAP datová kocka

(Zdroj: (8))

1.2.5 Prezentácia dát a nástroje Business Intelligence

Po vyčistení a analýze ostáva už len dáta prezentovať. Pre lepšiu predstavu, pochopenie a efektívne prezentovanie sú vhodné vizualizácie ako grafy či mapy alebo aj základné tabuľky. Vizualizovať, ale aj modelovať, analyzovať a získavať dáta umožňujú nástroje

Business Intelligence, ako napríklad Microsoft Power BI, ktorý bude popísaný v ďalšej kapitole.

1.3 Microsoft Power BI

1.3.1 Čo je Power BI?

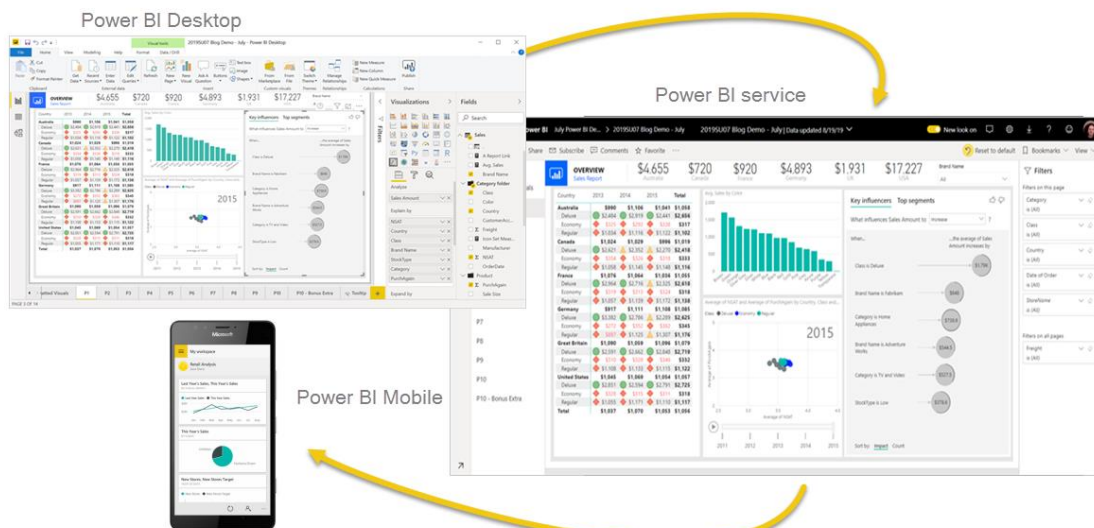
Microsoft Power BI je cloudová platforma od spoločnosti Microsoft, ktorá umožňuje užívateľom vizualizovať, analyzovať a zdieľať dáta pomocou interaktívnych dashboardov či reportov. Užívatelia môžu používať **Power BI Desktop** na vytváranie vizualizácií dát a následne ich publikovať do **Power BI Service**, kde ich môžu zdieľať so svojimi kolegami alebo klientmi.

Niektoré z najdôležitejších funkcií Power BI zahŕňujú:

- **Pripojenie k rôznym zdrojom dát:** Power BI umožňuje pripojiť sa k rôznym zdrojom dát a načítať ich do jedného miesta na vizualizáciu a analýzu. Medzi tieto zdroje patria napríklad Excel súbory, SQL databázy, cloudové úložiská a mnoho ďalších.
- **Vizualizácia dát:** Power BI ponúka mnoho spôsobov vizualizácie dát, od jednoduchých tabuliek a grafov po pokročilé heat mapy a geomapovanie. Užívatelia môžu vytvárať vlastné vizualizácie alebo používať predvolené šablóny.
- **Zdieľanie a spolupráca:** Užívatelia môžu zdieľať svoje vizualizácie a správy s kolegami alebo klientmi. Power BI umožňuje rôzne úrovne prístupu, od prezerania až po úpravu správ.
- **Interaktívne filtre a parametre:** Užívatelia môžu nastaviť filtre a parametre na základe ktorých sa mení zobrazenie dát v správach.
- **Plánované aktualizácie:** Užívatelia môžu plánovať automatizované aktualizácie a odosielanie správ prostredníctvom Power BI.
- **Prispôsobiteľné vizualizácie:** Power BI umožňuje užívateľom vytvárať vlastné vizualizácie pomocou JavaScriptu a iných programovacích jazykov.
- **Integrácia s inými nástrojmi:** Power BI sa integruje s inými nástrojmi od spoločnosti Microsoft, ako sú napríklad Excel, SharePoint a Teams, ako aj s nástrojmi tretích strán, ako sú napríklad Salesforce a Adobe Analytics.

- **Analytické nástroje:** Power BI umožňuje užívateľom používať pokročilé analytické nástroje, ako sú napríklad DAX a M, na manipuláciu s dátami a vytváranie výpočtov a ukazovateľov výkonnosti.
- **Mobilná aplikácia:** Power BI ponúka mobilnú aplikáciu pre iOS a Android, ktorá umožňuje užívateľom prístup k vizualizáciám dát a správam kedykoľvek a kdekoľvek.
- **Bezpečnosť a overovanie totožnosti:** Power BI ponúka rôzne úrovne bezpečnosti a overovania totožnosti pre užívateľov, vrátane integrovaného riešenia Azure Active Directory.
- **Power BI Embedded:** Power BI umožňuje vývojárom vytvárať vlastné aplikácie s vloženými vizualizáciami dát pomocou Power BI Embedded.
- **Power BI Report Server:** Power BI ponúka možnosť nainštalovať si vlastný Power BI Report Server pre správu a publikovanie správ v rámci firemnej siete.

Power BI je preto veľmi užitočný nástroj pre každého, kto pracuje s dátami a potrebuje rýchly a jednoduchý spôsob vizualizácie a analýzy týchto dát. S jeho mnohými funkciami a integráciami s inými nástrojmi je Power BI vhodný pre širokú škálu používateľov, od jednotlivcov po veľké organizácie. (9)



Obrázok č. 5: Prepojenie jednotlivých Power BI aplikácií

(Zdroj: (9))

1.3.2 Jazyk DAX a jazyk M

M a DAX sú dva jazyky používané v Power BI pre manipuláciu a modelovanie dát. Hoci majú obidva jazyky podobné funkcie, sú určené na rôzne účely a majú odlišnú syntax a využitie.

M jazyk je skriptovací jazyk používaný v Power Query, ktorý sa používa na transformáciu a modelovanie dát. M jazyk umožňuje užívateľom získať dáta z rôznych zdrojov, ako sú napríklad súbory CSV, databázy alebo webové stránky, a následne tieto dáta transformovať a formátovať pomocou rôznych nástrojov a funkcií, ako sú napríklad filtrovanie, zhlukovanie alebo spojovanie. M jazyk sa používa aj na tvorbu vlastných funkcií a výpočtov.

DAX (Data Analysis Expressions) je jazyk používaný na vytváranie výrazov a výpočtov v Power BI. DAX sa používa na modelovanie a analýzu dát v tabuľkách Power Pivot alebo v relačnej databáze. DAX umožňuje užívateľom vytvárať výpočty, ktoré zahŕňujú agregačné funkcie, ako sú suma, priemer, maximum a minimum, a umožňuje vytvárať vlastné funkcie pomocou jazyka. Syntax DAX je podobná jazyku Excelových funkcií.

Hoci M jazyk a DAX majú podobné funkcie a sú často používané spolu v Power BI, majú odlišné použitie a syntax. M jazyk sa používa na získavanie a transformáciu dát, zatiaľ čo DAX sa používa na vytváranie výpočtov a ukazovateľov výkonnosti. (6)

1.4 Jazyk Python

Python je vysokoúrovňový programovací jazyk, ktorý je považovaný za jeden z najpopulárnejších jazykov na svete.. Python je interpretovaný jazyk, čo znamená, že kód sa interpretuje v čase behu programu, nie vopred ako v prípade kompilovaných jazykov ako C++ alebo Java. Python je známy svojou jednoduchosťou a čitateľnosťou, čo robí zápis kódu intuitívnejším a rýchlejšim. Python má bohatú knižnicu modulov a nástrojov, ktoré uľahčujú programátorom rýchle vytváranie aplikácií. Jeho využitie sa rozšírilo na mnohé oblasti vrátane webových aplikácií, vedeckého výskumu, umelá inteligencia, strojové učenie, spracovanie dát a mnoho ďalších. Python je multiplatformový a beží na všetkých hlavných operačných systémoch, ako sú Windows, Linux a MacOS. Jeho vývoj a podpora sú zabezpečené komunitou programátorov, ktorí pravidelne pridávajú nové funkcionality a zlepšenia do jeho jadra a knižníc. (10)



Obrázok č. 6: Logo jazyku Python

(Zdroj: (10))

1.5 Fuzzy párovanie

Fuzzy párovanie je metóda, ktorá umožňuje identifikovať záznamy, ktoré sa líšia v rôznej miere, a teda sú len čiastočne zhodné v databáze. Pri fuzzy párovaní sa používajú algoritmy, ktoré využívajú rôzne techniky na meranie podobnosti dvoch záznamov. Tieto techniky môžu zahŕňať porovnávanie znakových reťazcov, výpočet Levenshteinovej vzdialenosti (miera vzdialenosti medzi dvoma reťazcami) a podobne. Pri použití techník fuzzy párovania, algoritmy na meranie podobnosti reťazcov poskytujú hodnoty, ktoré vyjadrujú mieru podobnosti medzi dvomi reťazcami. Tieto hodnoty sú zvyčajne normalizované na interval od 0 do 1, kde hodnota 1 znamená úplnú zhodu a hodnota 0

znamená nulovú zhodu. **Prah podobnosti** určuje, aká vysoká musí byť miera podobnosti, aby boli dva záznamy považované za zhodné. Napríklad, ak je prah podobnosti nastavený na hodnotu 0.8, záznamy s mierou podobnosti nižšou ako 0.8 budú považované za nezhodné, zatiaľ čo záznamy s mierou podobnosti vyššou ako 0.8 budú považované za zhodné. Výsledkom fuzzy párovania je množina zhukov, ktoré sú označené ako čiastočne zhodné. Fuzzy párovanie sa často používa v oblasti spracovania textov, ako napríklad pri porovnávaní názvov firiem, adries, mien zákazníkov atď. V praxi sa využíva na čistenie a zlúčenie dát z rôznych zdrojov, aby boli informácie správne a úplné. (11)

2 Analýza súčasného stavu

V analytickej časti diplomovej práce bude z úvodu predstavená univerzita VUT a Fakulta podnikateľská, ktorej dáta budú analyzované v priebehu tejto diplomovej práce. Po predstavení bude vysvetlený priebeh prijímacieho konania a priebeh štúdia študentov, aby bolo zrejmé s akými dátami sa pracuje. Na záver analytickej časti budú vytvorené analýzy, ktoré by mali byť užitočné a zaujímavé pre vedenie a napomáhať manažérskemu rozhodovaniu.

2.1 Predstavenie univerzity VUT a fakulty podnikateľskej

2.1.1 Univerzita VUT

Vysoké učení technické v Brně je verejná univerzita, ktorá ponúka vzdelanie v širokej škále odborov od ekonomických, technických, prírodovedných až po umelecké a to konkrétne na ôsmich fakultách a troch vysokoškolských ústavoch. Je prvou vysokou školou na Morave a jej história siaha až do roku 1899. Významnou činnosťou univerzity VUT je aj jej zameranie na výskum a vývoj, ktoré prebieha na jej piatich výskumných centrách. (12)

2.1.2 Fakulta podnikateľská

Fakulta podnikateľská je jednou z ôsmich fakúlt na univerzite VUT. Ponúka široké spektrum bakalárskych, magisterských a doktorandských študijných programov, ktoré sú zamerané na podnikanie a manažment. Formy štúdia na fakulte podnikateľskej sú prevažne denné, ale ponúka aj externé štúdium a kombinované štúdium pre tých, ktorí potrebujú flexibilné možnosti štúdia. Fakulta tiež ponúka študentské mobility ako Erasmus, vďaka ktorým si môžu študenti rozšíriť svoje vedomosti a skúsenosti v zahraničí. Fakulta je aj významným výskumným centrom, ktoré sa aktívne zúčastňuje medzinárodných výskumných projektov a spolupracuje s viacerými univerzitami a inštitúciami po celom svete. Fakulta má moderné zariadenia a výskumné laboratóriá, ktoré umožňujú študentom a výskumníkom pracovať na aktuálnych projektoch. (13)

Fakulta v minulosti bola rozdelená na programy, ktoré sa ďalej delili na obory, aby študentom umožnila si zvoliť špecializáciu. V súčasnosti už sú však len programy rozdelené podľa typu štúdia.

Bakalárske programy:

- **Ekonomika podniku**
- **Manažérska informatika**
- **Procesní management**
- **Účetnictví a daně**
- **Enterpreneurship and Small Business Developement** – vyučovaný v angličtine

Navazující magisterské:

- **Informačný management** – možnosť kombinovaného štúdia
- **Strategický rozvoj podniku** - možnosť kombinovaného štúdia
- **Účetnictví a finanční řízení podniku** - možnosť kombinovaného štúdia
- **Mezinárodní ekonomika a obchod**
- **International Business and Management** – vyučovaný v angličtine

Doktorské programy:

- **Ekonomika a management**
- **Economics and Management** – vyučovaný v angličtine
- **Řízení a ekonomika podniku** – možnosť kombinovaného štúdia
- **Company Management and Economics** – možnosť kombinovaného štúdia, vyučovaný v angličtine (13)

2.2 Priebeh štúdia študentov

V tejto kapitole bude popísané, ako prebieha štúdium študentov od rozhodnutia, že chcú na fakulte študovať až po prípadné doštudovanie. Kapitola slúži na lepšie pochopenie analýz, ktoré budú neskôr v práci tvorené.

2.2.1 Prijímacie riadenie

Keď sa študent rozhodne, že chce na fakulte študovať, najprv si musí byť osobne alebo elektronicky podať prihlášku. Do prihlášky študent zadáva:

- Osobné údaje: Meno, priezvisko, titul, dátum narodenia, pohlavie, rodné číslo, pohlavie, štátna príslušnosť, miesto narodenia
- Kontaktné údaje: korešpondenčná adresa, e-mail a telefón
- Údaje o predchádzajúcom vzdelaní: názov a adresa strednej či vysokej školy a dátum ukončenia a začatia štúdia. Pri vysokej škole sa vyplní aj fakulta a odbor, pri strednej škole sa vypisuje aj rok maturitnej skúšky a typ strednej školy.
- Výber študijného programu a formy štúdia: študijný obor a či kombinované alebo prezenčné štúdium
- Jazyk štúdia: anglický alebo český
- Prípadné poznámky a žiadosti, ktoré by mali byť zohľadnené pri prijímacom procese a štúdiu
- Tému práce v prípade doktorandského štúdia

Po vyplnení prihlášky ju študent odošle a uhradí za ňu poplatok. Po prijatí poplatku je študent pozvaný na prijímaciu skúšku. Študent sa jej však môže vyhnúť pri prihláške na magisterské štúdium ak študoval bakalárske štúdium na FP a mal priemer do 1.9, alebo na bakalárske štúdium ak:

- Pri SCIO teste všeobecných študijných predpokladov bol percentil 60 a vyšší
- Z matematiky a informatiky mal najhoršie známku 2
- Zúčastnil sa súťaže Business Point na prvých troch priečkach

V prípade, že študent spĺňa jednu z týchto kritérií, môže požiadať o prepáčenie prijímacej skúšky a doložiť dokumenty potvrdzujúce túto skutočnosť.

Ak však študent nespĺňa podmienky na prepáčenie prijímacej skúšky, neostáva mu nič iné ako sa na nich dostaviť. Fakulta na prijímaciu skúšku ponúka aj prípravný kurz. Prijímacia skúška pozostáva z 2 testov:

- Test všeobecných študijných predpokladov: 45 minút a 25 úloh. Za správnu odpoveď 4 body a za nesprávnu sa 1 odčíta, študent teda môže získať body v intervale $\langle -25, 100 \rangle$. Študent musí získať aspoň 30 bodov.
 - Test z anglického jazyka – už nie je, v minulosti však bol. Naposledy bol v roku 2021. Maximálny počet bodov bol 25, minimálny počet bodov na prijatie bol 8.
- (14)

Ak študent dosiahol dostatok bodov z prijímacej skúšky alebo ju nemusel ani podstupovať, fakulta zverejní dátum zápisu na ktorý sa študent príde zapísať a stáva sa študentom.

V roku 2020 kvôli pandémie prijímacie skúšky neboli.

2.2.2 Pribeh štúdia študentov od nástupu po ukončenie

Študent po nastúpení na štúdium musí splniť všetky predmety, ktoré má zapísané a získať tak za semester 60 kreditov. Aby splnil predmet, musí splniť väčšinou zápočet a skúšku, ktoré sú špecifické pre každý jeden predmet. Po úspešnom zakončení predmetu skúškou je študent ohodnotený stupnicou A,B,C,D,E,F. A je najlepšie hodnotenie a F neprospel, čo znamená, že študent predmet nespĺnil a musí si ho preniesť do ďalšieho roka. Ak ho však nedá ani v ňom, štúdium bude ukončené z dôvodu nesplnených podmienok Po splnení všetkých predmetov v záverečnom ročníku napíše záverečnú prácu na vybranú tému a následne si prácu musí obhájiť pri štátnej záverečnej skúške, kde budú overené aj jeho znalosti získané za štúdium na univerzite. Ak študent nezvládne štátnu záverečnú skúšku, taktiež má ešte jeden pokus a ak nezvládne ani ten, môže znovu skúsiť o rok. Študenti, čo sa rozhodnú, že toto štúdium nie je pre nich ho zanechajú písomným ohlásením.

2.3 Analýzy, čo by vedenie univerzity mohli zaujímať a ich prínosy

V tejto časti semestrálnej práce budú vytvorené analýzy dát, ktoré by priniesli výsledky zaujímavé pre univerzitu. Analýzy budú neskôr vizualizované v praktickej časti a ich výsledky analyzované.

2.3.1 Analýza regionálneho pôvodu študentov a ich výber oborov

Táto analýza sa bude zaoberať identifikáciou oblastí pôvodu študentov a o aké obory na univerzite majú študenti záujem podľa toho odkiaľ pochádzajú. Analýza umožní univerzite identifikovať oblasti s potenciálnom na zvýšenie záujmu o štúdium a zlepšiť marketingové aktivity v týchto oblastiach. Analýza bude obsahovať:

- Vizualizácie geografickej distribúcie študentov podľa miest a štátov
- Identifikácia oblastí s najväčším počtom prihlášok
- Porovnanie preferencií oborov medzi študentmi z regiónov s najväčším záujmom

2.3.2 Analýza vývoja počtu prihlášok a úspešnosti študentov v čase

Táto analýza poskytne univerzite prehľad o vývoji počtu prihlášok a úspešnosti študentov v priebehu času. Identifikácia zmien môže pomôcť univerzite pri plánovaní a prispôbovaní stratégií. Napríklad, ak sa zistí pokles počtu prihlášok, univerzita by mohla zvážiť zlepšenie svojho marketingového prístupu alebo nastaviť inak prijímacie konanie. To však ale dosť záleží aj na príčinách poklesu, na ktoré sa analýza taktiež bude snažiť prísť. Zároveň analýza sleduje úspešnosť študentov v priebehu času a identifikuje oblasti, kde sú študenti úspešnejší, čo by mohlo naznačovať silné stránky univerzity alebo oblasti, kde je potrebné zlepšenie. Analýza bude obsahovať:

- Sledovanie trendu v počte prihlášok, prijatých a zapísaných študentov
- Sledovanie úspešnosti študentov – počet bodov na prijímacích skúškach, výberový priemer celkovo a podľa oborov
- Identifikácia oborov s najväčším poklesom a nárastom záujmu

2.3.3 Analýza úspešnosti študentov v rôznych fázach štúdia v rôznych oboroch

Táto analýza určí, ktoré ročníky v ktorom obore sú pre študentov najnáročnejšie či najľahšie a na základe toho môže určiť lepšie učebné osnovy. Analýza výsledky bude porovnávať v rôznych formách a typoch štúdia:

- Porovnanie úspešnosti študentov na začiatku, v polovici a na konci štúdia.
- Identifikácia oborov, kde študenti zaznamenávajú výrazný pokles alebo zlepšenie úspešnosti v priebehu štúdia.

2.3.4 Analýza vzťahu medzi úspechom na prijímacích skúškach a vysokoškolským úspechom

Analýza bude skúmať či existuje nejaký vzťah medzi úspechom na prijímacích skúškach a úspechom počas štúdia. Univerzita získa lepší prehľad o efektívnosti svojich prijímacích skúšok a môže ich prispôbiť tak, aby lepšie predikovali úspech študentov.

- Porovnanie výsledkov prijímacích skúšok študentov s ich priemerom známok počas štúdia
- Zistenie či majú výsledky prijímacích skúšok vplyv na to, či študent dokončí štúdium
- Určenie či test všeobecného prehľadu alebo anglického jazyka má lepšie prediktívne vlastnosti
- Analýza úspešnosti študentov podľa typu prijatia

2.3.5 Analýza dokončenia štúdia

Táto analýza bude sledovať či začaté štúdium sa rovná ukončené štúdium. Analýza bude spočívať v hľadaní prečo a akí študenti štúdium neukončia a zároveň v hľadaní možnosti odstránenia tohto problému ak tam nejaký je. Bude obsahovať:

- Počet zápisov a úspešne ukončených štúdií podľa typu a formy štúdia
- Počet zápisov a úspešne ukončených štúdií podľa pohlavia, roku a či už predtým mali vysokú školu

2.3.6 Analýza stredných škôl podľa typu

Analýza bude zisťovať, aký je počet prijatých a zapísaných študentov podľa typu strednej školy, ako sa im darí na škole a o aké obory majú záujem. Podľa toho bude možné upraviť zameranie oborov a správne nasmerovať marketing. Bude obsahovať:

- Analýzu počtu prihlášok, zápisov, priebehu štúdia a prijímacích skúšok podľa typu strednej školy
- Porovnanie oborov podľa typu strednej školy študentov

2.3.7 Analýza vysokých škôl a fakúlt

Analýza sa bude zameriavať na hľadanie konkrétnych vysokých škôl a fakúlt, z ktorých študenti prichádzajú na fakultu, aby mohla lepšie smerovať svoj marketing. Bude obsahovať:

- Počet prihlášok podľa vysokých škôl
- Počet prihlášok podľa fakúlt vysokých škôl, z ktorých sa hlási najviac študentov

3 Vlastné návrhy riešenia

V návrhovej časti diplomovej práce budú z úvodu popísané dáta, ktoré budú slúžiť k analýzám. Následne budú vyčistené a modelované, aby sa dali vizualizovať a v záverečnej časti budú vytvorené vizualizácie a vyhodnotené výsledky analýz na základe týchto vizualizácií.

3.1 Popis získaných dát

Od fakulty podnikateľskej boli získané dva excelovské súbory. Súbory boli importované do PowerBI, kde sa s nimi bude pracovať. Pri importe sa prvý riadok tabuliek nastavil ako hlavička stĺpcov v PowerQuery, kde bude vykonaná aj analýza a čistenie dát.

3.1.1 Popis prvého súboru

Prvý súbor s názvom „Export_finall.xlsx“ obsahuje všetky údaje o prihláške na fakultu podnikateľskú, vrátane toho aká bola jej forma, na aký obor je, výsledky prijímacieho riadenia a či bude študent zapísaný. Či bol študent prijatý, určuje rozhodnutie dekana v tabuľke. Dáta taktiež obsahujú všetky údaje, ktoré tam študent o sebe napísal pri podaní prihlášky, ako adresa či stredná a vysoká škola, ktorú navštevoval v minulosti. Osobné informácie o študentoch budú v tejto práci anonymizované. Excelovská tabuľka je rozdelená na sedem listov, pričom každý z nich zodpovedá jednému roku od roku 2016 do roku 2022. Každý z týchto listov obsahuje rovnakých 137 stĺpcov, čo samozrejme znamená, že veľa z nich je nepotrebných či prázdnych a je potreba dáta vyčistiť. Pre lepšiu predstavu dát slúži snímka dát v nasledujúcom obrázku.

	A ^B _C narodnost	A ^B _C VS	1 ² ₃ kolo	A ^B _C pohlavi	1 ² ₃ rozhodnuti dekana	A ^B _C 123 prumer 1. rocnik
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 79% Error 0% Empty 21%	Valid 0% Error 0% Empty 100%
1	Česká republika	NE		1 F		null
2	Ghanská republika	ANO		1 F		10
3	Česká republika	ANO		1 M		null
4	Česká republika	ANO		1 F		20
5	Slovenská republika	ANO		1 M		null
6	Česká republika	NE		1 F		20
7	Česká republika	NE		1 F		20
8	Česká republika	ANO		1 M		10
9	Polská republika	NE		1 M		20
10	Polská republika	NE		1 M		20
11	Ghanská republika	ANO		1 M		20
12	Ruská federace	ANO		1 F		null
13	Turecká republika	NE		1 M		11
14	Syrská arabská republika	ANO		1 F		null
15	Česká republika	ANO		1 F		10

Obrázok č. 7: Prvý súbor

(Zdroj: Vlastné spracovanie)

3.1.2 Popis druhého súboru

Súbor „Export_studium.xlsx“ taktiež obsahuje sedem listov, pričom každý zodpovedá školskému roku, a teda sa jedná o prelomy rokov od 2016/2017 až po 2022/2023. Každý list obsahuje 10 stĺpcov a to konkrétne id študenta, ročník a stupeň, obor, formu štúdia, priemer a stĺpec popis zmeny štúdia, ktorý určuje čo sa deje so štúdiom študenta, či sa zapísal alebo štúdium ukončil a podobne. Súbor je teda vlastne nejaký zoznam zmien v štúdiu študentov. Pre lepšiu predstavu prikladám v nasledujúcom obrázku snímku dát.

1 ² ₃ ročník	1 ² ₃ stupeň	A ^B _C o./s. zkratka	A ^B _C p. zkratka	A ^B _C forma studia	1.2 v. prům.	1 ² ₃ stav studia	datum zm. studia	A ^B _C popis zmeny studia
Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 70% Error 0% Empty 30%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
1	2	MGR-EBF	MGR-EBF	P		null	7/1/2016	převedení z přijímacího řízení
3	1	BAK-UAD	BAK	P	1.53	61	6/23/2017	ukončení studia včetně SZZ
1	1	BAK-EPM	BAK	P		null	5/3/2016	převedení z přijímacího řízení
1	2	MGR-Z	MGR-Z	Z	1.85	99	2/3/2017	Ukončení krátkodobého studia (např. Socrates, Erasmus)
2	1	BAK-EP	BAK	P	2.48	11	9/1/2016	řádný zápis
1	2	MGR-ŘEP	MGR	P	2.1	11	8/25/2016	řádný zápis
1	2	MGR-ŘEP-KS	MGR-KS	K		null	6/30/2016	převedení z přijímacího řízení
1	1	BAK-EP	BAK	P		null	4/7/2017	nesplnění požadavků stud. programu
1	2	MGR-Z	MGR-Z	Z	1.09	99	6/20/2017	Ukončení krátkodobého studia (např. Socrates, Erasmus)
1	2	MGR-Z	MGR-Z	Z	1.16	99	7/4/2017	Ukončení krátkodobého studia (např. Socrates, Erasmus)
3	1	BAK-EPM	BAK	P	2.19	61	6/16/2016	Ukončení studia včetně SZZ
2	1	BAK-UAD	BAK	P	2.22	11	9/1/2016	řádný zápis
1	2	MGR-Z	MGR-Z	Z		null	2/2/2017	Ukončení krátkodobého studia (např. Socrates, Erasmus)
2	1	BAK-Min	BAK-Min	P	1.91	11	9/1/2016	řádný zápis

Obrázok č. 8: Druhý súbor

(Zdroj: Vlastné spracovanie)

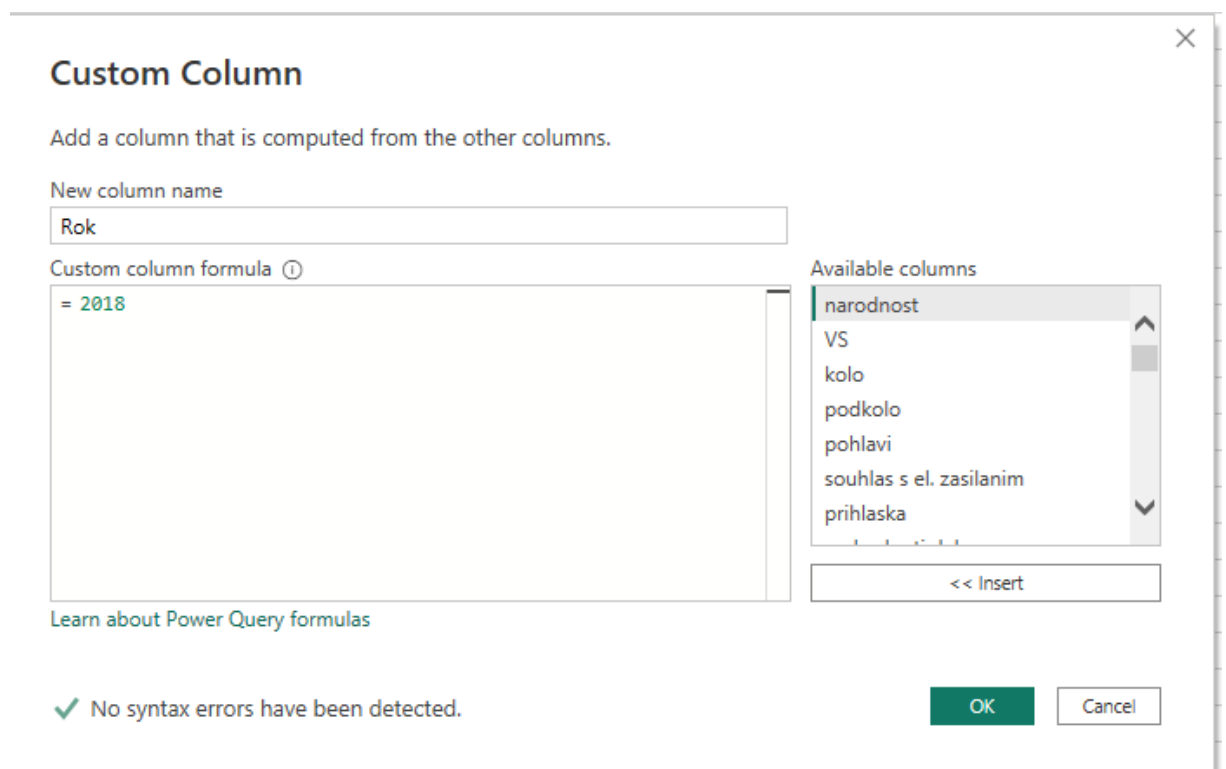
3.2 Čistenie dát

V tejto časti práce budú identifikované potrebné dáta na analýzu a vizualizácie, zmenia sa na požadovaný formát a eliminujú sa tie nepotrebné.

3.2.1 Čistenie prvého súboru

Ako bolo spomínané, súbor obsahuje 137 stĺpcov v siedmich listoch. Keďže každý list obsahuje rovnaké stĺpce, je vhodné ich spojiť do jedného a následne identifikovať potrebné a užitočné stĺpce a zvyšné eliminovať.

Aby sme ich spojili, najprv treba v každom liste pridať nový stĺpec, kde bude rok podania prihlášky, pretože v tabuľke nie je stĺpec s časovým údajom, ktorý by bolo možné využiť na rozlíšenie roku podania po spojení. Toto je možné dosiahnuť vytvorením nového vlastného stĺpca, kde sa jednoducho do vzorca stĺpca dopíše manuálne rok podania prihlášky ako v nasledujúcom obrázku.



Obrázok č. 9: Funkcia na získanie roku

(Zdroj: Vlastné spracovanie)

Keď už je umožnené rozlíšenie listov po spojení, ostáva už len spojiť listy dokopy pomocou funkcie v PBI „Append Queries“, kde po zvolení všetkých listov a potvrdení bola vytvorená nová tabuľka so všetkými údajmi pomenovaná názvom „Prihlášky“. Korektnosť spojenia je vhodné overiť pomocou funkcie transformácie „Count Rows“ v každom liste, ktorá zráta počet riadkov v tabuľke. Súčet riadkov v jednotlivých listoch by sa mal rovnať počtu riadkov v novej tabuľke. Po kontrole bolo zistené, že počet riadkov a teda počet prihlášok za všetky riadky je 22173.

Po spojení dát je potrebné získať relevantné, ktoré budú využité na vizualizáciu pri analýzach popísaných v analytickej časti. Dáta v niektorých stĺpcoch postačia aj v originálnom formáte, no niektoré bude treba zmeniť na požadovaný formát. Z tabuľky „Prihlášky“ bude potrebné získať tieto informácie:

1. **Rok podania prihlášky** – Stĺpec „Rok“ vytvorený v predošlom kroku
2. **Pohlavie** – Stĺpec „pohlavi“
3. **Cudzí kľúč na prepojenie s tabuľkou „Studium“** – stĺpec „student id“, ktorý bude v práci anonymizovaný
4. **Či sa študent zapísal** – Stĺpec „zapsán“, ktorého hodnoty sú 0 a 1. Pomocou funkcie „Replace Values“ boli hodnoty nahradené za „Zapísaný a nezapísaný“. V dátach je aj stĺpec „bude zapsán“, ktorý však nemá správne hodnoty.
5. **Či bol študent prijatý/neprijatý a ako** – táto informácia sa skrýva v stĺpci „rozhodnuti dekana“. Sú to však rôzne čísla od 10 do 38 a zo zvyšných dát nebolo možné dedukovať ich význam, preto bola nutná komunikácia s univerzitou, ktorá poskytla k číslam legendu. Aby sa dalo s údajmi pracovať, je potrebné vytvoriť nový stĺpec s názvom „Forma prijatia“, kde číslam priradený zodpovedajúci textový reťazec s informáciou či bol študent prijatý a akou formou pomocou funkcie „if“. Dosiahneme to vytvorením vlastného stĺpcu, ako je znázornené v nasledujúcom obrázku, ktorý obsahuje kód v jazyku M. Hodnoty, kde rozhodnutie dekana nebolo, patria študentom, čo neboli prijatí pre nesplnenie podmienok.

```

Table.AddColumn(
  "Forma prijatia",
  each
    if [rozhodnuti dekana] = 38 then
      "prijatý po obnovení príjmacieho konania"
    else if [rozhodnuti dekana] = 11 then
      "prijatý bez prijímacej skúšky"
    else if [rozhodnuti dekana] = 10 then
      "prijatý na základe prijímacej skúšky"
    else if [rozhodnuti dekana] = 14 then
      "prijatý na základe SCIO testu"
    else if [rozhodnuti dekana] = 16 then
      "prijatý prednostne"
    else if [rozhodnuti dekana] = 20 then
      "neprijatý z dôvodu neprospechu"
    else if [rozhodnuti dekana] = 22 then
      "neprijatý pre nedostačujúcu kapacitu"
    else if [rozhodnuti dekana] = 27 then
      "doktorandi neprijatí pre obsadenosť tém"
    else if List.Contains({35, 36, 37}, [rozhodnuti dekana]) then
      "doktorandi neprijatí z formálnych záležitostí"
    else
      "neprijatý pre nesplnenie podmienok"
)

```

Obrázok č. 10: Získanie prijatých študentov pomocou kódu v M

(Zdroj: Vlastné spracovanie)

Taktiež bol vytvorený ešte jeden stĺpec s názvom “bude prijatý”, kde sa číslom rozhodnutia 11,10,14 a 38 priradilo „ANO“ a zvyšným číslam „NIE“. Je to preto aby bol formát zhodný s tabuľkou „bude zapsán“ a tak sa umožnila rýchla filtrácia dát pri analýzach, kde toto nebude hlavným predmetom.

6. **Na aké štúdium sa študenti hlásia** – stĺpce „typ studia“ a „forma studia“, poskytujú informácie o stupni a prezenčnosti štúdia. Ostatné stĺpce popisujú obor či program názvami a skratkami a hlavne obsahujú zastaralé názvy, ktoré sa menili, ako bolo spomenuté v analytickej časti. Je vhodné názvy upraviť na aktuálne, aby bolo umožnená dobrá analýza z hľadiska starších údajov a z hľadiska prehľadnosti. Keďže obory boli špecializáciou programu, zmenili sa na programy bez špecializácie. Programy ostali také aké boli. Nasledujúcim kódom na obrázku budú získané aktuálne programy/obory tak, že ak obor je bez

špecializácie, využije sa názov programu a ak špecializáciu má, využije sa jej názov.

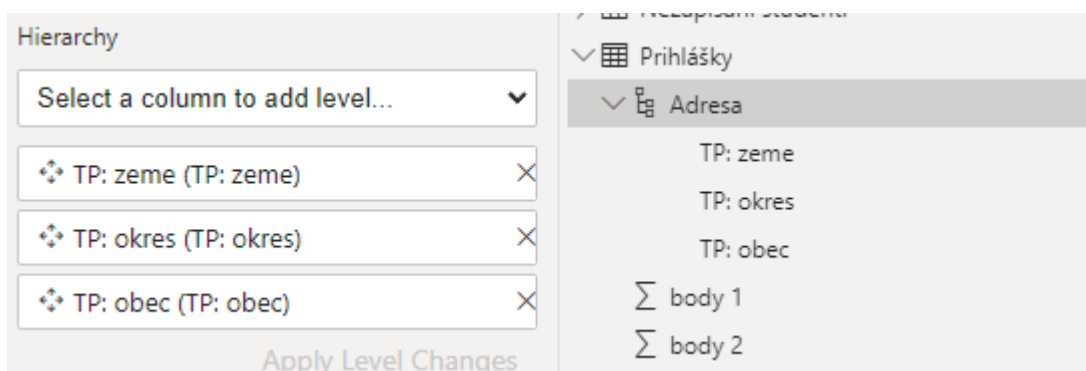
```
Table.AddColumn(  
    "Študovaný obor",  
    each  
        if [#"obor/specializace nazev"] = "bez specializace" then  
            [program nazev]  
        else  
            [#"obor/specializace nazev"]  
    )
```

Obrázok č. 11: Získanie aktuálnych oborov pomocou kódu v M

(Zdroj: Vlastné spracovanie)

Zostali ešte 3 obory pre zahraničných študentov, ktorých názvy sa pomocou funkcie „Replace Values“ zmenia na „Neaktuálny“. Po tejto zmene je v novovytvorenom stĺpci „Študovaný obor“ 12 aktuálnych oborov a jeden obor s neaktuálnymi a dá sa tak s nimi pekne pracovať.

7. **Adresa študenta** - stĺpce „TP:zeme“, „TP:okres“ a „TP:obec“ sú časti adresy trvalého pobytu vhodné na analýzu. Pre prípad vizualizácie na mape bude uchovaný aj stĺpec „TP:PSČ“. Keďže okres spadá pod štát a mesto pod obec, v tomto prípade bude vytvorená logická hierarchia v relačnom modeli.



Obrázok č. 12: Hierarchia adresy

(Zdroj: Vlastné spracovanie)

8. **Body za kolá prijímacieho riadenia** – stĺpec „body 1“ zodpovedá výsledkom testu všeobecného prehľadu a stĺpec „body 2“ testu z anglického jazyka. Celkovému počtu bodov za všetky kolá zodpovedá stĺpec „celkový počet bodu za vsechna kola“.
9. **Akú strednú školu vyštudovali** – Názov strednej školy sa nachádza v stĺpci „SS: nazev“ a na analýzu bude ponechaný aj stĺpec s obcou strednej školy „SS:obec“. Je však mnoho rôznych stredných škôl. Na lepšiu predstavu slúži nasledujúci snímok dát.

SS: nazev
SUPS a VOS, příspěvková organizace
Švehlova střední škola polytechnická
Svob.čeb.škola, ZŠ a gymnázium s.r.o.
SZdŠ a VOŠ zdr. E. Pöttinga a JŠ s PSJZ
SZeŠ a SOŠ
SZŠ a VOŠZ Žďár nad Sázavou
Táborské soukromé gymnázium a Základní škola, s.r.o.
Taneční konzervatoř
Taneční konzervatoř, Brno, Nejedlého 3
Tauferova SOŠ veterinární Kroměříž
Tříinecká obchodní akademie, spol. s r.o.
TRIVIS - Střední škola veřejnoprávní a Vyšší odborná škola bezpečnosti silniční dopravy Jihlava, s.r.o.
TRIVIS - Střední škola veřejnoprávní Prostějov, s.r.o.
TRIVIS - Střední škola veterinární Emila Holuba Brno, s.r.o.
TRIVIS Střední škola veřejnoprávní
TRIVIS-Střední škola veřejnopráv.,s.r.o
VÍTKOVICKÁ STŘEDNÍ PRŮMYSLOVÁ ŠKOLA
Vojenská SŠ a VOŠ Ministerstva obrany
Vojenská střední škola
Vojenská střední škola a Vyšší odborná škola Ministerstva obrany v Moravské Třebové
VOŠ a ISŠ strojnická
VOŠ a Obchodní akademie Chotěboř
VOŠ a SOŠ Bystřice nad Pernštejnem
VOŠ a SOŠ Gustava Habrmana
VOŠ a SOŠ technická
VOŠ a SPŠ
VOŠ a SPŠ elektrotechnická
VOŠ a SPŠ elektrotechnická Fr. Křížíka
VOŠ a SŠ technická
VOŠ a SŠ zdravotnická a sociální
VOŠ a SŠ, s.r.o.
VOŠ a Střední zemědělská škola
VOŠ a SUPŠ Brno
VOŠ ekon. a zdravotnická a Střední škola

Obrázok č. 13: Názvy stredných škôl pred čistením

(Zdroj: Vlastné spracovanie)

Ako je možné vidieť na obrázku, názvy sú pestré. Pomocou funkcie „count distinct“ bolo zistené, že v dátach je 1162 jedinečných názvov škôl. Je to preto lebo to tam zadávali študenti manuálne, ale aj širokou škálou stredných škôl. Unifikovať rôzne názvy jednej strednej školy je dlhý proces, ktorý by nemal veľký význam, pretože je mnoho rôznych stredných škôl v rôznych mestách a preto by bolo najrozumnejšie školy rozdeliť na skupiny, ktoré sa ďalej budú dať analyzovať podľa geografickej polohy. Po rýchlej analýze dát bolo zistené, že školy je najvhodnejšie rozdeliť podľa počtu prihlášok na:

- Integrované stredné školy – v názve majú „integ“ alebo „ISS/ISS“
- Stredné odborné učilište - v názve „učili/ucili“ alebo „SOU“
- Gymnázium - v názve majú „gymn“ alebo v prípade chyby „gimn“
- Obchodná akadémia – v názve „obch“ alebo „OA“
- Zahraničné škola – v názve „zahan“ alebo bez názvu – po analýze bolo zistené že školy bez názvu sú školy študentov zo zahraničia
- Stredné odborné školy – ostatné školy, keďže tu spadá veľa rôznych škôl a zvyšné kategórie už boli vymedzené

Aby boli tieto skupiny získané z dát, je potrebné vytvoriť nový vlastný stĺpec „SS:typ“, ktorý pomocou funkcie „Text.Contains“ zistí, či kontrolovaný reťazec obsahuje reťazec špecifický pre skupinu ako bolo popísané vyššie. Keďže jazyk M rozlišuje malé a veľké písmená, je taktiež potrebné využiť funkciu „Text.Lower“, ktorá pred kontrolou reťazca zmení písmena na malé a preto také musia byť aj kontrolované reťazce. Pomocou funkcie „if“ sa následne každej strednej škole priradí správny typ podľa definovanej podmienky – obsahu reťazca. Kód je zobrazený v nasledujúcom obrázku.


```

Table.AddColumn(
  "SS:typ",
  each
    if [{"SS: nazev"} = null or Text.Contains(Text.Lower(["SS: nazev"]), "zahran") then
      "Zahranická škola"
    else if Text.Contains(Text.Lower(["SS: nazev"]), "integ")
      or Text.Contains(Text.Lower(["SS: nazev"]), "isš")
      or Text.Contains(Text.Lower(["SS: nazev"]), "iss")
    then
      "Integrovaná stredná škola"
    else if Text.Contains(Text.Lower(["SS: nazev"]), "učili")
      or Text.Contains(Text.Lower(["SS: nazev"]), "ucili")
      or Text.Contains(Text.Lower(["SS: nazev"]), "sou")
    then
      "Stredné odborné učilište"
    else if Text.Contains(Text.Lower(["SS: nazev"]), "gymn")
      or Text.Contains(Text.Lower(["SS: nazev"]), "gimn")
    then
      "Gymnázium"
    else if Text.Contains(Text.Lower(["SS: nazev"]), "obch")
      or Text.Contains(Text.Lower(["SS: nazev"]), "oa")
    then
      "Obchodní akadémia"
    else
      "Stredná odborná škola"
)

```

Obrázok č. 14: Získanie typu stredných škôl pomocou kódu v M

(Zdroj: Vlastné spracovanie)

Na záver je ešte potrebné vytvoriť hierarchiu, aby stredné školy spadali pod ich typ.

SS:typ	Počet prihlášok
⊕ Gymnázium	7864
⊕ Zahranická škola	4727
⊕ Stredná odborná škola	4559
⊕ Obchodní akadémia	4438
⊕ Stredné odborné učilište	457
⊕ Integrovaná stredná škola	128
Total	22173

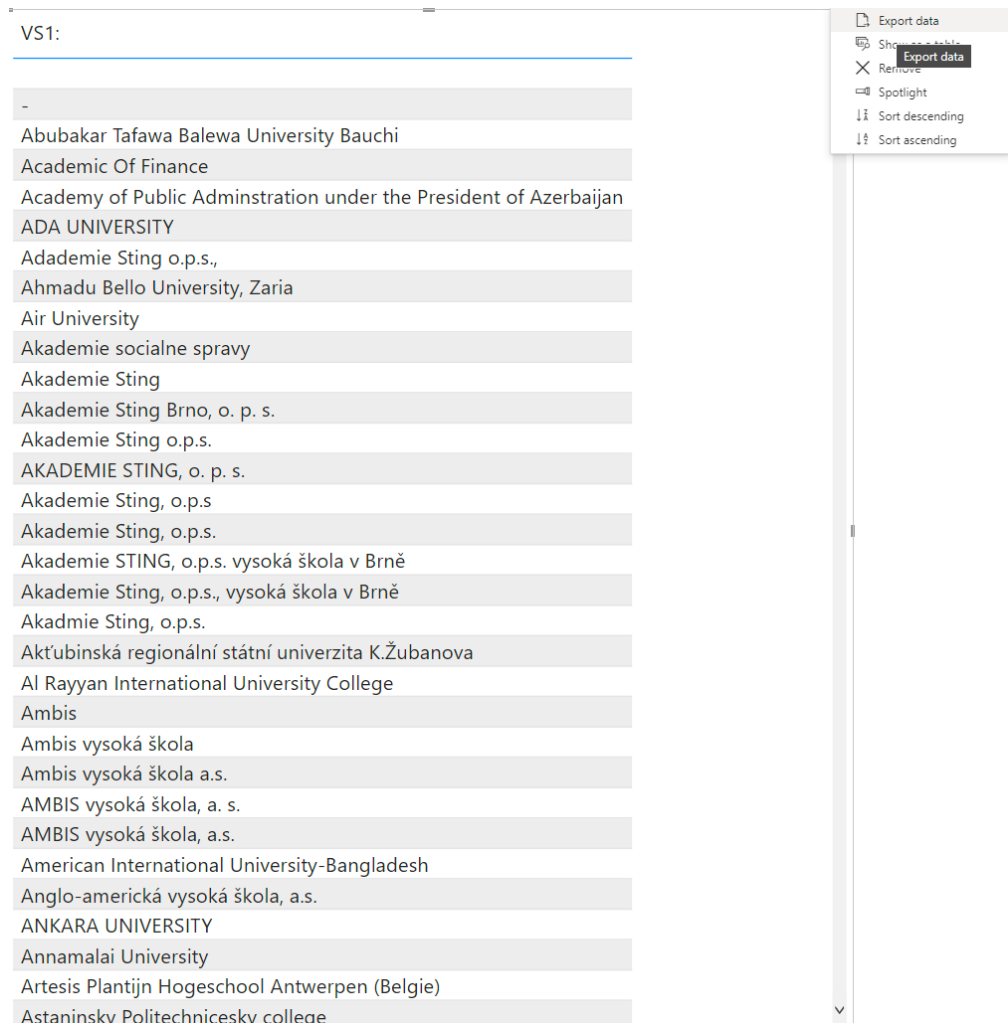
Obrázok č. 15: Hierarchia stredných škôl

(Zdroj: Vlastné spracovanie)

10. **Akú vysokú školu majú vyštudovanú**– V dátach je veľa stĺpcov súvisiacich s vysokou školou. V prvom rade sú tam 3 stĺpce s názvom vysokej školy, v prípade ak by študent študoval viac vysokých škôl. Stĺpec pre druhú a tretiu vysokú školu je však skoro celý prázdny, jedná sa skôr o doktorandské štúdiá a väčšinou sú to študenti z fakulty podnikateľskej, takže pre analýzu bude najrozumnejšie využívať stĺpec s prvou vysokou školou „VS1:“. K nej však prislúchajú aj stĺpce fakulty, študijného programu, študijného oboru a dátum zahájenia a ukončenia. Z týchto stĺpcov pre analýzu bude najzaujímavejší stĺpec fakulty „VS: fakulta“.

Pomocou funkcie „Count distinct“ bolo zistené, že je 905 jedinečných mien vysokých škôl a 981 jedinečných mien fakúlt. Je to kvôli tomu, že študenti manuálne dopisujú vyštudované vysoké školy. Ak teda chodili napríklad na VUT mohli tam napísať "VUT", "Vysoké učení", "VUT v Brně" etc. Dáta v tejto forme nie sú vhodné na analýzu, preto je potrebné spojiť súvisiace názvy univerzít do jedného spoločného. Nie je možné využiť postup ako pri stredných školách a rozdeliť vysoké školy do skupín, pretože všeobecné univerzity môžu mať rôzne fakulty bez špecifického zamerania. Je potrebné upraviť mená univerzít, no ručný prepis by bol časovo veľmi náročný, preto sa to bude robiť nasledovne:

1. **Export dát do Excelu** – ako prvý krok je potrebné vytvoriť kópiu tabuľky „Prihlášky“ pomocou „Reference“ v PowerQuery. V novej tabuľke sa tlačidlom „Choose columns“ zvolí stĺpec „VS1:“ a zvyšné stĺpce sa vymažú. Ďalej je potrebné vymazať duplikáty pomocou „Remove Duplicates“, čím budú získané jedinečné názvy. Následne sa v reportovacom zobrazení vytvorí tabuľková vizualizácia novej tabuľky. Pomocou „Export Data“ sa dáta vyexportujú do Excelu.



Obrázok č. 16: Export názvov vysokých škôl

(Zdroj: Vlastné spracovanie)

- Zhlukovanie dát a vytvorenie mapovacej tabuľky** – v ďalšom kroku je potrebné k 905 jedinečným názvom uložených v Exceli vytvoriť nový stĺpec, ktorý nesprávnemu názvu priradí správny a vytvoriť tak mapovaciu tabuľku. Manuálna úprava by trvala veľmi dlho, preto je nutné si pomôcť a to konkrétne využitím fuzzy párovania a skriptovacieho jazyku Python.

V nasledujúcom obrázku je zobrazený kód v jazyku Python, ktorý po spustení príkazovým riadkom zaradom prejde všetky názvy univerzít v exportovanom Exceli a na základe fuzzy párovania bude zhlukovať podobné názvy dokopy do spoločných zhlukov, ktoré uloží do vedľajšieho stĺpca. Ak názov univerzity

nepasuje do doteraz vytvoreného zhluku, vytvorí nový zhluk. V nasledujúcom obrázku je detailne popísané, čo presne kód vykonáva.

```
# Nahrание potrebných knižníc
import pandas as pd
from rapidfuzz import process, fuzz

# Nahrание jedniečnych univerzít z excelovského súboru
cesta = "C:/Users/hrint/Desktop/2.rocnik/diplomka/data/VS_list.xlsx"
mena_univerzít = pd.read_excel(cesta)

# Nastavenie prahu podobnosti - bolo potrebné testovať, aby zhluky univerzít neboli málo, ale ani veľmi
unikátne (0-100)
prah_podobnosti = 85

# Vytvorenie prázdnej mapovacej tabuľky, kde sa budú mapovať originálne mená na zhlukované mená
mapovacia_tabuľka = pd.DataFrame(columns=["OriginalneMeno", "ZhlukovaneMeno"])

# Iterovaný prechod menami univerzít
for i, row in mena_univerzít.iterrows():
    originalne_meno = row["Univerzity"]

    # Ak je mapovacia tabuľka prázdna, pridá sa prvá univerzita ako nový zhluk
    if mapovacia_tabuľka.empty:
        new_row = pd.DataFrame({"OriginalneMeno": [originalne_meno], "ZhlukovaneMeno":
[originalne_meno]})
        mapovacia_tabuľka = pd.concat([mapovacia_tabuľka, new_row], ignore_index=True)
        continue

    # Nájdienie najlepšej zhody v zhluku pomocou fuzzy prirovnávania
    nalepsia_zhoda = process.extractOne(originalne_meno, mapovacia_tabuľka["ZhlukovaneMeno"],
scorer=fuzz.token_set_ratio)

    # Ak je skóre najlepšej zhody nad prahom podobnosti, originálne meno sa priradí zhlukovanému menu
    if nalepsia_zhoda[1] > prah_podobnosti:
        new_row = pd.DataFrame({"OriginalneMeno": [originalne_meno], "ZhlukovaneMeno":
[nalepsia_zhoda[0]])
        mapovacia_tabuľka = pd.concat([mapovacia_tabuľka, new_row], ignore_index=True)
    else:
        # Ak sa nenájde žiadna zhoda nad prahom podobnosti, vytvorí sa nový zhluk s názvom univerzity
        new_row = pd.DataFrame({"OriginalneMeno": [originalne_meno], "ZhlukovaneMeno":
[originalne_meno]})
        mapovacia_tabuľka = pd.concat([mapovacia_tabuľka, new_row], ignore_index=True)

# Uloženie mapovacej tabuľky s originálnymi a zhlukovanými názvami univerzít do nového excelovského
súboru
mapovacia_tabuľka.to_excel("zhlukovany_VS_list.xlsx", index=False)
```

Obrázok č. 17: Vytvorenie zhlukov vysokých škôl pomocou fuzzy porovnávania

(Zdroj: Vlastné spracovanie)

	A	B
1	OriginalneMeno	ZhlukovaneMeno
19	Aktubinská regionální státní univerzita K.Žubanova	Aktubinská regionální státní univerzita K.Žubanova
20	Al Rayyan International University College	Al Rayyan International University College
21	Ambis	Ambis
22	Ambis škola	Ambis
23	Ambis vysoká	Ambis
24	AMBIS vysoká škola, a. s.	AMBIS vysoká škola, a. s.
25	AMBIS vysoká škola, a.s.	AMBIS vysoká škola, a. s.
26	American International University-Bangladesh	American International University-Bangladesh
27	Anglo-americká vysoká škola, a.s.	Anglo-americká vysoká škola, a.s.
28	ANKARA UNIVERSITY	ANKARA UNIVERSITY
29	Annamalai University	Annamalai University
30	Artesis Plantijn Hogeschool Antwerpen (Belgie)	Artesis Plantijn Hogeschool Antwerpen (Belgie)
31	Astaninsky Politechnicesky college	Astaninsky Politechnicesky college
32	Az?rbaycan Tibb Universiteti	Az?rbaycan Tibb Universiteti
33	Azad	Azad
34	Azerbaijan State University of Economics	Azerbaijan State University of Economics
35	Azerbaijan Cooperation University	Azerbaijan Cooperation University
36	Azerbaijan State Economic University	Azerbaijan State University of Economics
37	Azerbaijan State Oil and Industrial University	Azerbaijan State Oil and Industrial University
38	Azerbaijan State Oil and Industry University	Azerbaijan State Oil and Industrial University
39	Azerbaijan State University of Economics	Azerbaijan State University of Economics
40	B.I.B.S.	B.I.B.S.
41	B.I.B.S. a.s.	B.I.B.S.
42	B.I.B.S. Brno	B.I.B.S.

Obrázok č. 18: Vytvorené zhluky po spustení kódu v Pythone

(Zdroj: Vlastné spracovanie)

V obrázku č.18 sa nachádza výsledok po prevedení kódu. Ako je možné vidieť, nie všetko sa zhlukovalo správne, čo je dôsledkom nastavenia prahu podobnosti na 85. Po testovaní to bola však najvhodnejšia hodnota prahu, pretože nižší už zhlukoval aj čo nemal a vyšší zas nezhlukoval čo mal zhlukovať. Väčšine dát však bol priradený správny názov a ušetrilo to mnoho času. Zvyšok dát je potrebné upraviť ručne, aby každému názvu prislúchala správna univerzita. Pri úprave dát je nutné opraviť aj nevhodné údaje ako „nestudoval som“ alebo názov strednej školy na prázdnu hodnotu, aby sa študenti, ktorí tieto údaje zadávali, nerátali ako študenti vysokých škôl.

3. **Namapovanie správnych názvov**– ďalším krokom je importovať mapovaciu tabuľku do PowerBI pomocou „Get Data“ a nastaviť prvý riadok ako hlavičku stĺpca. Teraz je už možné tabuľky spojiť. V tabuľke „Prihlášky“ tlačidlo „Merge queries“ umožní spojenie tabuliek pomocou originálneho mena ako primárneho kľúča. Typ spojenia „Left join“ spojí všetky dáta z „Prihlášky“ a zhodné dáta z mapovacej tabuľky. Alternatívne by to v tomto prípade mohol byť aj „Inner join“, pretože by sa mali všetky názvy zhodovať.

Merge

Select a table and matching columns to create a merged table.

Prihlášky

SS: nazev	SS: obec	VS	VS1:	VS1: fakulta
...ní pedagogická škola, Obchodní akad...	Znojmo	NE	<i>null</i>	<i>null</i>
...ihana	<i>null</i>	ANO	University for Development Studies	Mathematical Science
	Brno	ANO	Mendelova univerzita v Brně	Provozně ekonomická fakulta
DŠ knih.,inf.a soc.sl.	Brno	ANO	Vysoké učení technické	Fakulta podnikatelská

ZhlukovaneVS

OriginalneMeno	ZhlukovaneMeno
-	<i>null</i>
Abubakar Tafawa Balewa University Bauchi	Abubakar Tafawa Balewa University Bauchi
Academic Of Finance	Academic Of Finance
Academy of Public Adminstration under the President o...	Academy of Public Adminstration under the President o...
ADA UNIVERSITY	ADA UNIVERSITY

Join Kind

Left Outer (all from first, matching from second)

Use fuzzy matching to perform the merge

▷ Fuzzy matching options

✓ The selection matches 10905 of 22172 rows from the first table.

OK

Cancel

Obrázok č. 19: Spojenie mapovacej tabuľky s prihláškami

(Zdroj: Vlastné spracovanie)

Spojením sa vytvoril nový stĺpec s referenciou na mapovaciú tabuľku. V pravom rohu názvu nového stĺpca je potrebné zaškrtnúť, nech sa zobrazujú len zhlukované vysoké školy. Pomocou funkcie „Count Distinct“ sa na záver overí úspešnosť. Počet jedinečných názvov univerzít sa redukoval z 905 na 417 – zhlukovanie bolo úspešné.

A ^B _C VS1:	A ^B _C ZhlukovaneMeno
<ul style="list-style-type: none"> ● Valid 51% ● Error 0% ● Empty 49% 	<ul style="list-style-type: none"> ● Valid 51% ● Error 0% ● Empty 49%
Ostravská univerzita	Ostravská univerzita v Ostravě
Rašínova vysoká škola v Brně	Rašínova vysoká škola
university of ghana	university of ghana
Slezská Univerzita	Slezská univerzita Opava
Vysoká škola Karla Engliše	Vysoká škola Karla Engliše v Brně
Slovenská Technická Univerzi...	Slovenská technická univerzita v Bratisl...
STU	Slovenská technická univerzita v Bratisl...
STU	Slovenská technická univerzita v Bratisl...
STU v Bratislave zo sídlom v T...	Slovenská technická univerzita v Bratisl...
Vysoké učení technické v brně	Vysoké učení technické v Brně
Vysoké Učení Technické	Vysoké učení technické v Brně
Žilinská univerzita v Žiline	Žilinská univerzita v Žilině
VYSOKÉ UČENÍ TECHNICKÉ V ...	Vysoké učení technické v Brně
VYSOKÉ UČENÍ TECHNICKÉ V ...	Vysoké učení technické v Brně

Obrázok č. 20: Zhlukované názvy škôl po namapovaní

(Zdroj: Vlastné spracovanie)

Na záver je vhodné pre prehľadnosť upraviť meno stĺpca na „VŠ“. Po úprave názvov vysokých škôl ostáva upraviť názvy fakúlt. Počet jedinečných mien fakúlt je 981. Predošlý spôsob úpravy mien sa dá aplikovať aj v tomto prípade, no predsa len je to celkom časovo náročné a ideálny bolo dáta analyzovať a hľadať zjednodušenie.

V nasledujúcom obrázku sú zhlukované názvy vysokých škôl a fakulty s počtom prihlášok za posledné 2 roky.

VŠ	Počet prihlášok	
+	2871	
+	Vysoké učení technické v Brně	1441
+	Masarykova univerzita	248
+	Mendelova univerzita v Brně	240
	Provozně ekonomická fakulta	85
	Provozně ekonomická	78
	Fakulta regionálního rozvoje a mezinárodních studií	13
	PEF	10
	Agronomická fakulta	7
	Agronomická	5
	Provozně-ekonomická	5
	Provozně-ekonomická fakulta	5
	Ekonomická	3
	FRRMS	3
	LDF MENDELU	2
	Lesnická a dřevařská	2
	Provozne ekonomická	2
	Provozovně ekonomická	2
	Regionálního rozvoje a mezinárodních studií	2
	agronomicka	1
	agronomická fakulta	1
	Ekonomie	1
	Faculty of Business and Economics	1
	Fakulta provozně ekonomická	1
	Fakulta regionálního rozvoje a mezinárodních studií	1
	ICV	1
	Institut celoživotního vzdělávání	1
	Lesnická a dřevařská fakulta	1
	Provozne ekonomická fakulta	1
	Provozně ekonomická fakulta Mendelovy univerzity v Brně	1
	Provozně ekonomická fakulta MENDELU	1
	Provozní Ekonomická Fakulta	1
	Regional Development and International Studies	1
	Regionalního Rozvoje a Mezinárodních Studií	1
	Zahradnická	1
+	Univerzita Palackého	33
+	Univerzita Tomáše Bati	31
Total	5342	

Obrázok č. 21: Názvy fakúlt pred čistením

(Zdroj: Vlastné spracovanie)

Najviac študentov je z VUT, MUNI a Mendelovej univerzity. Zvyšný počet prihlášok u iných univerzít za posledné 2 roky nie je najvyšší a podobne je to aj ostatné roky. Zdlhové zhlukovanie by v tomto prípade nemalo veľký význam, pretože ide hlavne o fakulty týchto troch univerzít. Fakulty navyše môžu mať rôzne obory a vyučovať rôzne veci, čo nevyplývajú z názvy fakulty, preto je pre

analýzu najvhodnejšie brať do úvahy len tieto tri univerzity a smerovať na nich marketing alebo si na základe nich vytvoriť predstavu o fakultách iných univerzít. Najprv sa vyfiltrujú názvy zhlukovaných VŠ na spomínané tri univerzity. Následne sa opakuje postup ako pri vysokých školách. Vyexportujú sa názvy fakúlt, vytvorí sa mapovacia tabuľka a spojí sa s tabuľkou „Prihlášky“ pomocou „Left join“. V tomto prípade nie je možné využiť „Inner Join“, preto nie všetky názvy fakúlt boli zhlukované. Z tohto dôvodu je však ešte potrebný jeden krok navyše, a to vytvoriť nový stĺpec s názvom „Fakulta“, ktorý je tvorený názvom z novej tabuľky ak existuje a ak nie tak sa využije názov z „VŠ: Fakulta“. V nasledujúcom obrázku je možné vidieť nové názvy fakúlt.

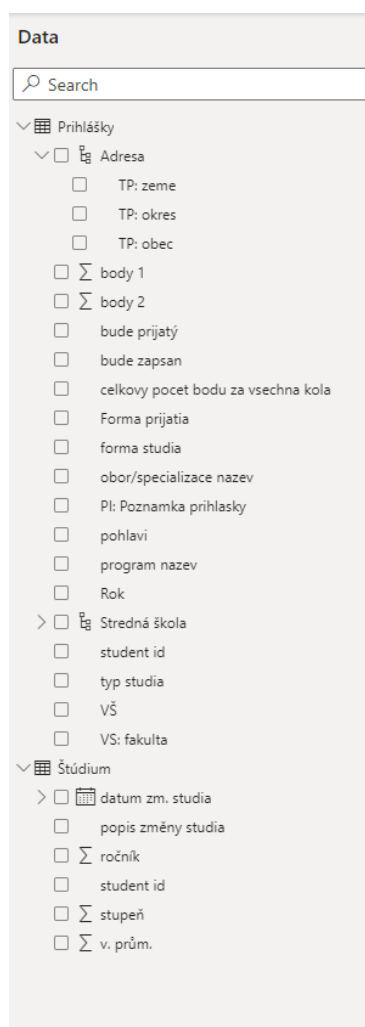
VŠ	Počet prihlášok
Vysoké učení technické v Brně	1441
Podnikatelská	967
Strojnícka	160
Informatická	150
Elektrotechnická	90
Stavebná	44
Chemická	25
Centrum sportovních aktivit	4
Architektúra	1
Masarykova univerzita	248
Ekonomicko – správná	79
Informatická	55
Prírodovedecká	38
Filozofická	22
Právnická	15
Športové štúdie	11
Pedagogická	10
Sociálnych štúdií	9
Lekárska	7
Fakulta filozofická	1
Farmaceutická	1
Mendelova univerzita v Brně	240
Provozně ekonomická fakulta	196
Regionálneho rozvoja a medzinárodných štúdií	21
Agronomická	14
Lesnícká a drevárska	5
Informatická	1
Institut celoživotního vzdělávání	1
Podnikatelská	1
Zahradnícka	1
Univerzita Palackého	33
Univerzita Tomáše Bati	31
Vysoká škola polytechnická v Jihlavě	26
AMBIS vysoká škola, a. s.	20
České vysoké učení technické v Praze	18
Akademie Sting	15
NEWTON College a.s.	14
Total	5342

Obrázok č. 22: Názvy fakúlt po čistení

(Zdroj: Vlastné spracovanie)

3.2.2 Čistenie druhého súboru

Z druhého súboru „Studium“ budú na analýzu potrebné všetky stĺpce, okrem stĺpca „o./s. zkratka“, ktorý má chýbajúce údaje a je len dodatkom stĺpca skratky programu a stĺpca „stav studia“, ktorý je len číselným vyjadrením „popis zmeny studia“, ktorý zaznamenáva akúkoľvek zmenu štúdia študentov ako zápis alebo ukončenie štúdia. Rok štúdia sa v tomto prípade bude brať zo stĺpca „dátum zm.studia“. Na spojenie tabuliek bude využitý „student id“. Po identifikácii a úprave potrebných stĺpcov ostáva už len zmazať zvyšné stĺpce. Na záver sa ešte upraví dátový typ stĺpcov pomocou „Detect data type“ na správny. V nasledujúcom obrázku sú zobrazené finálne dáta po vyčistení v reportovom zobrazení. Čistením sa redukoval celkový počet stĺpcov zo 147 na 25.



Obrázok č. 23: Finálne názvy stĺpcov po vyčistení

(Zdroj: Vlastné spracovanie)

3.3 Spojenie dát

Ako bolo spomenuté pri obsahu dát oboch súborov, na spojenie dát sa využije stĺpec „student id“, čo je vlastne id študenta. V tabuľke „Studium“ nemá nenulové hodnoty, pretože tabuľka je register zmien štúdia študentov, ktorí začínajú zápisom pri ktorom sa id študenta vytvorí. Nie je to však jedinečný identifikátor, pretože zmien v štúdiu môže byť viac a tým pádom sa budú vyskytovať duplicitné „student id“. V tabuľke „Prihlasky“ v tomto stĺpci sú aj prázdne riadky. Sú to študenti, ktorí neboli zapísaní. Taktiež je tam 253 študentov, ktorí majú id študenta ale neboli zapísaní. To však chybné dáta nie sú, sú to študenti ktorí študovali na fakulte podnikateľskej a chceli začať nové štúdium, no nakoniec sa nezapísali. Z tohto dôvodu sa aj v tejto tabuľke môžu vyskytovať duplikáty id študenta.

Na spojenie je nutné najprv vytvoriť prepojenie medzi tabuľkami. To sa dá uskutočniť v modelovom zobrazení v PowerBI. Spojenie je možné vytvoriť jednoducho kliknutím a podržaním „student id“ v jednej tabuľke a pretiahnutím na „student id“ v druhej tabuľke. Typ spojenia bude „many to many“, pretože obe tabuľky obsahujú duplikáty a „Cross filter direction“ bude nastavené na „Both“, aby bolo umožnené filtrovanie na základe stĺpcov oboch tabuliek.

Edit relationship

Select tables and columns that are related.

Štúdium

student id	ročník	stupeň	o./s. zkratka	forma studia	v. prům.	stav studia	datum zm. studia
225372	1	1	---	prezenční studium	null	1	Thursday, June 4, 2020
209995	1	1	---	prezenční studium	null	1	Thursday, June 4, 2020
227652	1	1	---	prezenční studium	null	1	Thursday, June 4, 2020

Prihlášky

udía	typ studia	obor/specializace nazev	Rok	SS:typ	body 1	body 2	student id	cel
studium	bakalářský	bez specializace	2021	Gymnázium	null	null	null	
studium	bakalářský	bez specializace	2021	Gymnázium	null	null	null	
studium	bakalářský	bez specializace	2021	Středná odborná škola	null	null	null	

Cardinality: Many to many (*:*)
Cross filter direction: Both

Make this relationship active
 Assume referential integrity

Apply security filter in both directions

Obrázok č. 24: Tvorenie prepojenia medzi tabuľkami

(Zdroj: Vlastné spracovanie)

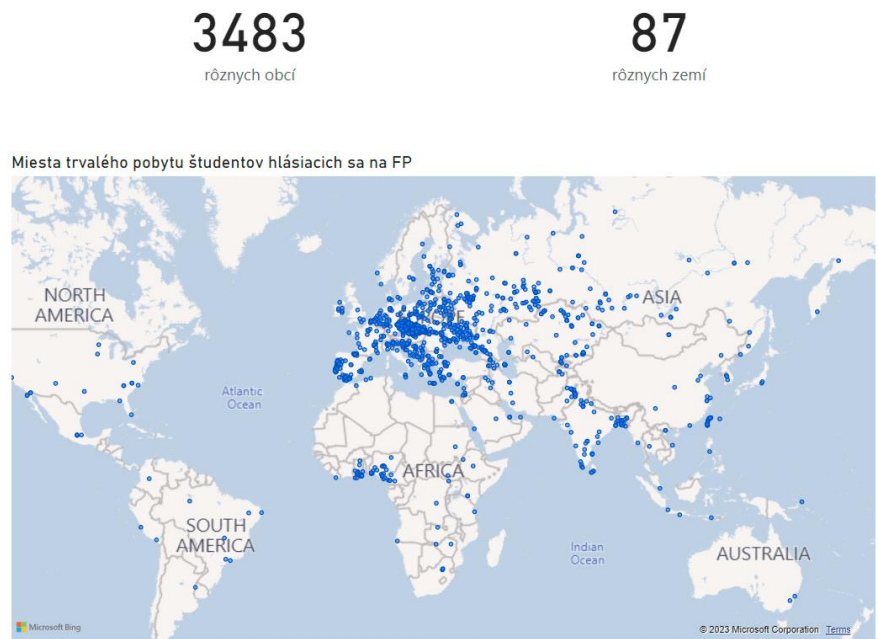
Následne sa v tabuľke prihlášok pomocou „Merge Queries“ tabuľky spoja pomocou „Left Outer Join“ a do tabuľky sa pridá odkaz na tabuľku Štúdium tak, aby obsahoval všetky študentské id z tabuľky prihlášky a zhodné z tabuľky „Štúdium“.

3.4 Vizualizácia analýz

V tejto časti diplomovej práce budú vizualizované analýzy z analytickej časti. K vizualizáciám z Power BI bude priložený komentár k výsledkom. Zhrnutie výsledkov analýz a návrh opatrení sa bude nachádzať v ďalšej kapitole.

3.4.1 Analýza regionálneho pôvodu študentov a ich výber oborov

TP: zeme	Počet prihlášok
<input type="checkbox"/> Czech Republic	17666
<input type="checkbox"/> Slovak Republic	2539
<input type="checkbox"/> Russian Federation	476
<input type="checkbox"/> Ukraine	333
<input type="checkbox"/> Republic of Kazakhstan	217
<input type="checkbox"/> Republic of Belarus	122
<input type="checkbox"/> Kingdom of Spain	83
<input type="checkbox"/> Portuguese Republic	62
<input type="checkbox"/> Republic of Ghana	54
<input type="checkbox"/> Republic of Azerbaijan	42
<input type="checkbox"/> Republic of Turkey	38
<input type="checkbox"/> Hellenic Republic	36
<input type="checkbox"/> Federal Republic of Germany	34
<input type="checkbox"/> Republic of India	29
<input type="checkbox"/> Federal Republic of Nigeria	28
<input type="checkbox"/> People's Republic of Bangladesh	25
<input type="checkbox"/> Republic of Serbia	24
<input type="checkbox"/> French Republic	23
<input type="checkbox"/> Republic of Italy	22
<input type="checkbox"/> Islamic Republic of Pakistan	21
<input type="checkbox"/> Taiwan (Province of China)	21
<input type="checkbox"/> Republic of Latvia	19
<input type="checkbox"/> Georgia	17
<input type="checkbox"/> Republic of Poland	17
<input type="checkbox"/> Republic of Moldova	13
<input type="checkbox"/> Republic of Uzbekistan	13
<input type="checkbox"/> Kingdom of Belgium	11
<input type="checkbox"/> Republic of Lithuania	11
<input type="checkbox"/> People's Republic of China	10
<input type="checkbox"/> Republic of Bulgaria	9
<input type="checkbox"/> Kingdom of the Netherlands	8
<input type="checkbox"/> Kyrgyz Republic	8
<input type="checkbox"/> Republic of Tajikistan	8
<input type="checkbox"/> United Mexican States	8
<input type="checkbox"/> Kingdom of Sweden	7
Total	22173



Obrázok č. 25: Rozdelenie študentov podľa zemi

(Zdroj: Vlastné spracovanie)

Ako je možné vidieť na obrázku č.25, študenti FP pochádzajú z rôznych zemí. Z Česka a Slovenska je však až 90 percent študentov, preto sa bude pracovať hlavne s tými štátmi. Na obrázku č.26 a 27 sú zobrazené obce v Česku a Slovensku zoradené podľa počtu prihlášok.

TP: obec	Počet prihlášok
Brno	4034
Břeclav	236
Blansko	216
Olomouc	214
Třebíč	182
Znojmo	179
Zlín	171
Prostějov	166
Uherské Hradiště	152
Bratislava	148
Jihlava	143
Kuřim	137
Vyškov	133
Hodonín	126
Přerov	122
Žďár nad Sázavou	108
Šlapanice	106
Velké Meziříčí	89
Kroměříž	86
Rosice	86
Boskovice	83

Obrázok č. 26: Obce s najväčším počtom študentov

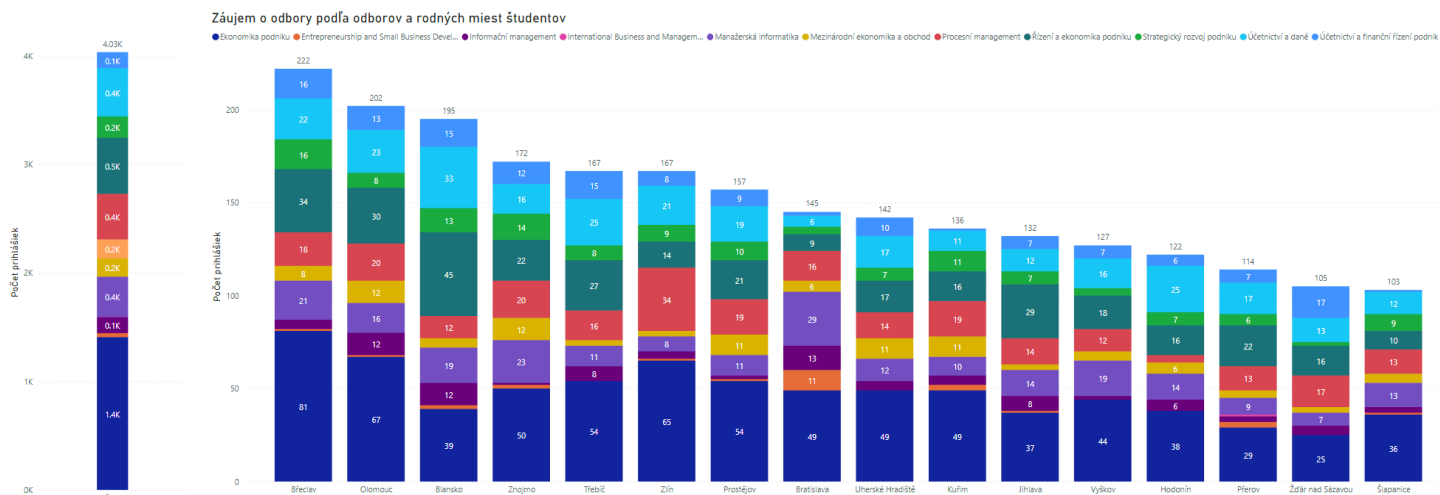
(Zdroj: Vlastné spracovanie)



Obrázok č. 27: Sústredenie najväčšieho počtu študentov

(Zdroj: Vlastné spracovanie)

Významná časť študentov je z Brna. Zvyšok je rozprchnutý po celom Česku a Slovensku. Cieľom tejto analýzy je však nie len zistiť, kde smerovať marketing, ale aj záujem o obory. Záujem sa bude sledovať u regiónov, z ktorých bolo aspoň 100 študentov od roku 2016. Záujem sa však bude sledovať len za posledné 2 roky, pretože je nutné sledovať aktuálne odbory a trendy.

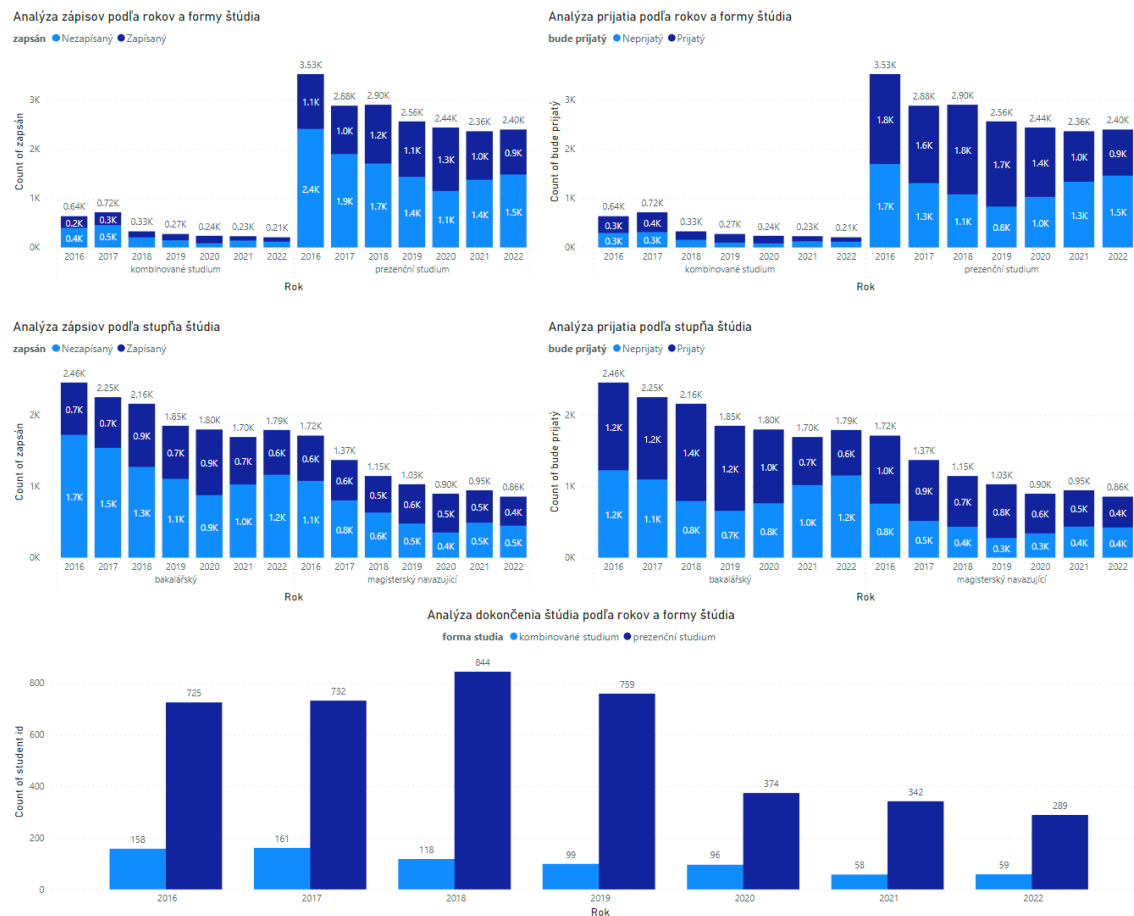


Obrázok č. 28: Záujem o obory podľa obcí

(Zdroj: Vlastné spracovanie)

Najobľúbenejším odborom vo všetkých mestách je Ekonomika podniku a druhým najobľúbenejším odborom je naväzujúci magisterský obor Riadenie a ekonomika podniku. V Bratislave Slovákom imponuje aj manažerska informatika, zatiaľ čo v Blansku imponuje študentom Účetnictví a daně a v Zlíne zas Procesní management.

3.4.2 Analýza vývoja počtu prihlášok a úspešnosti študentov v čase



Obrázok č. 29: Prihlásení a zapísaní študenti podľa rokov a štúdia

(Zdroj: Vlastné spracovanie)

V obrázku č.29 je vidieť, že počet prihlásených a počet zapísaných študentov od roku 2016 podstatne klesol. Oproti roku 2021 sa zaznamenal nárast, ale stále je to nižší počet prihlášok ako v roku 2020. Rovnaký trend je v oboch formách a typoch štúdia.

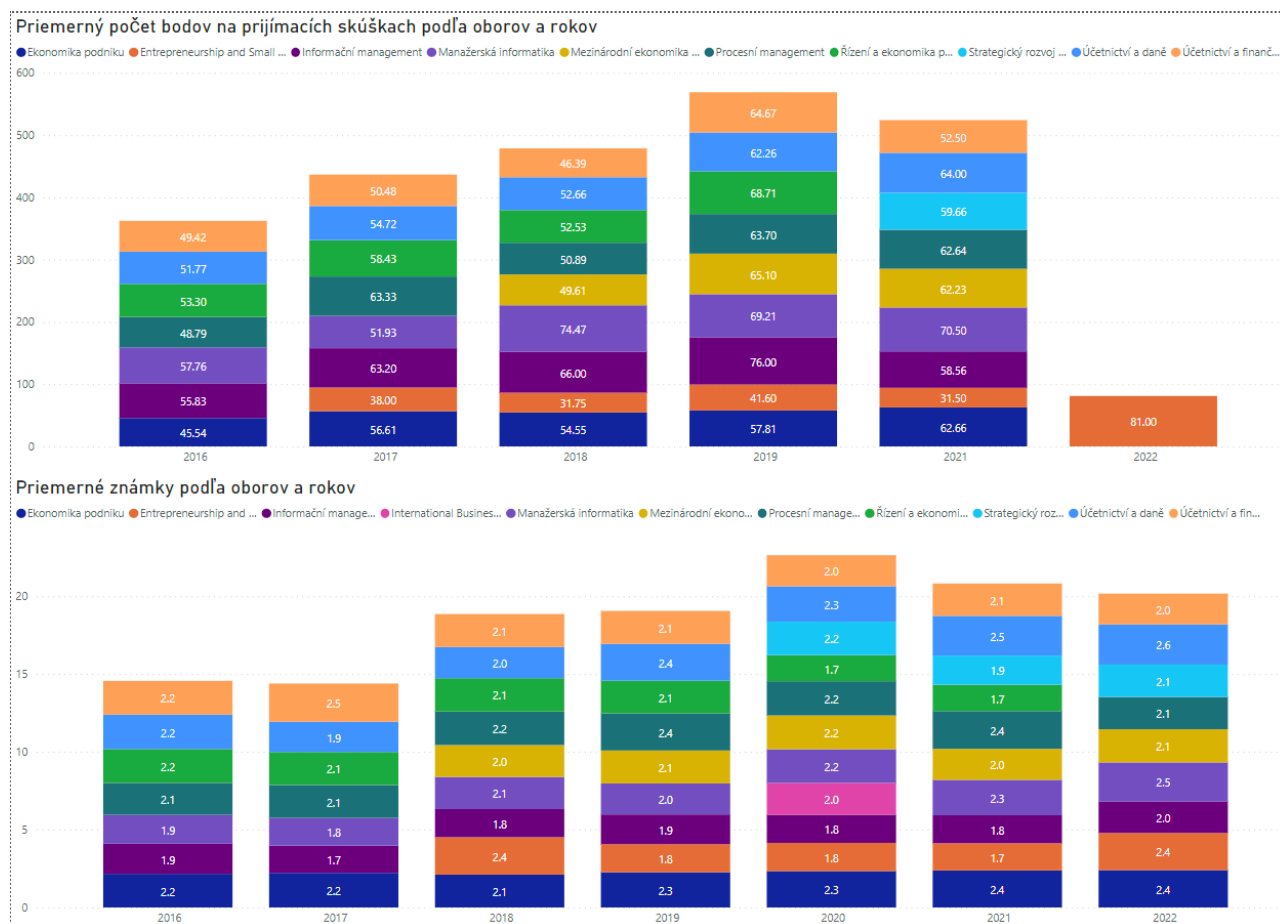
Na obrázku č.30 je možno vidieť detailné počty prihlášok a zápisov podľa oborov. Jediné o informačný management a manažersku informatiku je stále rovnaký záujem. U zvyšku oborov je zaznamenaný pokles.

Študovaný obor	2016	2017	2018	2019	2020	2021	2022	Total	Študovaný obor	2016	2017	2018	2019	2020	2021	2022	Total
Company Management and Economics	46	6	4	5	6	3	4	74	Company Management and Economics	46	6	4	5	6	3	4	74
Neprijatý		5	3	5	5	3	1	22	Nezapísaný		5	4	5	5	3	3	25
Prijatý	46	1	1		1			52	Zapísaný	46	1			1			49
Ekonomika podniku	1095	959	872	904	908	843	886	6467	Ekonomika podniku	1095	959	872	904	908	843	886	6467
Neprijatý	579	498	361	378	404	580	637	3437	Nezapísaný	772	691	522	563	468	581	637	4234
Prijatý	516	461	511	526	504	263	249	3030	Zapísaný	323	268	350	341	440	262	249	2233
Entrepreneurship and Small Business Development	39	43	51	38	26	48	245		Entrepreneurship and Small Business Development	39	43	51	38	26	48	245	
Neprijatý	35	13	12	11	11	22	104		Nezapísaný	39	19	23	13	11	22	127	
Prijatý	4	30	39	27	15	26	141		Zapísaný		24	28	25	15	26	118	
Informační management	158	154	156	159	152	187	150	1116	Informační management	158	154	156	159	152	187	150	1116
Neprijatý	62	38	45	32	52	69	67	365	Nezapísaný	74	43	62	46	54	74	67	420
Prijatý	96	116	111	127	100	118	83	751	Zapísaný	84	111	94	113	98	113	83	696
International Business and Management					36	67	108	211	International Business and Management					36	67	108	211
Neprijatý					36	48	74	158	Nezapísaný					36	56	93	185
Prijatý						19	34	53	Zapísaný						11	15	26
Manažerská informatika	350	349	344	254	338	304	343	2282	Manažerská informatika	350	349	344	254	338	304	343	2282
Neprijatý	95	112	74	53	152	171	185	842	Nezapísaný	187	191	171	147	171	171	185	1223
Prijatý	255	237	270	201	186	133	158	1440	Zapísaný	163	158	173	107	167	133	158	1059
Mezinárodní ekonomika a obchod			153	190	160	169	143	815	Mezinárodní ekonomika a obchod			153	190	160	169	143	815
Neprijatý			67	41	59	76	78	321	Nezapísaný			85	83	63	85	78	394
Prijatý			86	149	101	93	65	494	Zapísaný			68	107	97	84	65	421
Procesní management	539	459	451	215	189	201	221	2275	Procesní management	539	459	451	215	189	201	221	2275
Neprijatý	305	243	194	71	89	119	136	1157	Nezapísaný	423	334	318	137	101	119	136	1568
Prijatý	234	216	257	144	100	82	85	1118	Zapísaný	116	125	133	78	88	82	85	707
Řízení a ekonomika podniku	854	577	606	554	20	21	17	2649	Řízení a ekonomika podniku	854	577	606	554	20	21	17	2649
Neprijatý	406	234	233	159	4	6	3	1045	Nezapísaný	535	334	348	275	6	6	4	1508
Prijatý	448	343	373	395	16	15	14	1604	Zapísaný	319	243	258	279	14	15	13	1141
Strategický rozvoj podniku					374	367	340	1081	Strategický rozvoj podniku					374	367	340	1081
Neprijatý					123	163	156	442	Nezapísaný					130	190	156	476
Prijatý					251	204	184	639	Zapísaný					244	177	184	605
Účetnictví a daně	476	409	364	373	285	250	240	2397	Účetnictví a daně	476	409	364	373	285	250	240	2397
Neprijatý	249	210	152	135	105	136	171	1158	Nezapísaný	343	288	240	228	117	137	171	1524
Prijatý	227	199	212	238	180	114	69	1239	Zapísaný	133	121	124	145	168	113	69	873
Účetnictví a finanční řízení podniku	185	231	229	133	171	154	107	1210	Účetnictví a finanční řízení podniku	185	231	229	133	171	154	107	1210
Neprijatý	85	89	92	47	71	80	50	514	Nezapísaný	118	137	143	78	72	84	50	682
Prijatý	100	142	137	86	100	74	57	696	Zapísaný	67	94	86	55	99	70	57	528
Total	3703	3183	3222	2838	2677	2592	2607	20822	Total	3703	3183	3222	2838	2677	2592	2607	20822

Obrázok č. 30: Prihlášky a zápisy na jednotlivé obory

(Zdroj: Vlastné spracovanie)

Na obrázku č.31 sú zobrazené priemerné počty bodov z prijímacích skúšok a priemerné známky podľa jednotlivých oborov. Grafy ukazujú, že študenti síce majú lepší úspech na prijímacích skúškach, ale majú horší študijný priemer. Na celkový priemer sa detailnejšie pozrie ďalšia analýza.



Obrázok č. 31: Priemerné počty bodov a známok podľa oborov

(Zdroj: Vlastné spracovanie)

3.4.3 Analýza úspešnosti študentov v rôznych fázach štúdia v rôznych oboroch

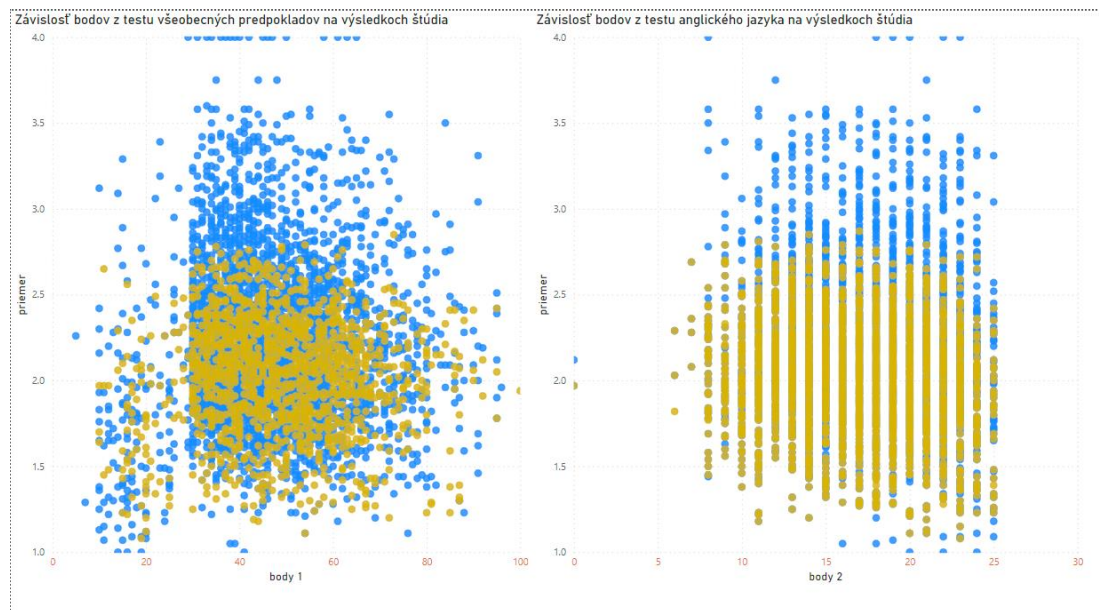
Bakalárske štúdium				
Študovaný obor	1	2	3	Total
Procesní management	2.28	2.11	2.07	2.18
Účetníctví a daně	2.26	2.13	2.11	2.18
Ekonomika podniku	2.26	2.13	2.07	2.18
Manažerská informatika	2.16	2.02	1.95	2.07
Entrepreneurship and Small Business Development	1.80	1.82	1.82	1.81
Total	2.24	2.10	2.05	2.15
Magisterské štúdium				
forma studia	1	2	Total	
☒ kombinované štúdium	2.33	2.13	2.24	
Účetníctví a finanční řízení podniku	2.51	2.23	2.38	
Informační management	2.23	2.04	2.15	
Strategický rozvoj podniku	2.22	2.04	2.14	
☒ prezenční štúdium	2.02	1.99	2.00	
Účetníctví a finanční řízení podniku	2.12	2.05	2.09	
Mezinárodní ekonomika a obchod	2.03	2.01	2.02	
Strategický rozvoj podniku	2.02	1.98	2.00	
Informační management	1.98	1.93	1.95	
International Business and Management	1.59	1.64	1.61	
Total	2.07	2.00	2.04	
Doktorské štúdium				
forma studia	1	2	Total	
☒ kombinované štúdium	2.35	2.20	2.27	
Účetníctví a finanční řízení podniku	2.51	2.23	2.38	
Řízení a ekonomika podniku	2.29	2.20	2.24	
☒ prezenční štúdium	2.11	2.05	2.08	
Účetníctví a finanční řízení podniku	2.12	2.05	2.09	
Řízení a ekonomika podniku	2.11	2.05	2.08	
Total	2.19	2.09	2.14	

Obrázok č. 32: Priemery podľa typu štúdia a oborov

(Zdroj: Vlastné spracovanie)

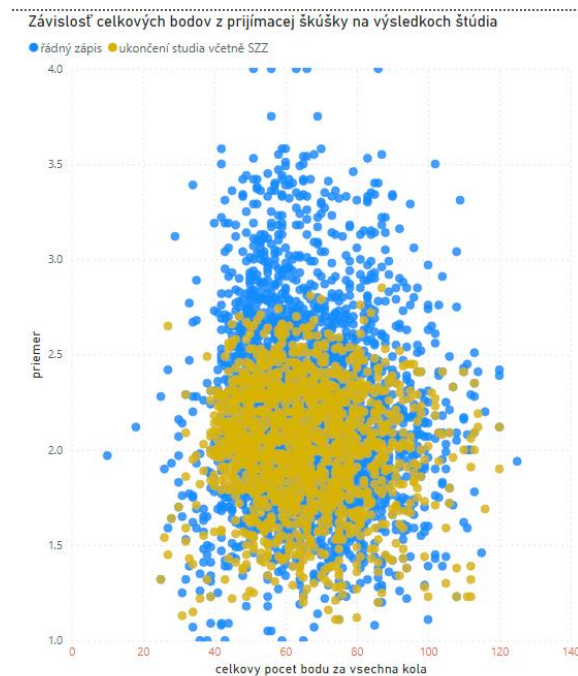
Na obrázku č.32 je možné vidieť, že prváci majú najhorší priemer. Je to však preto, že veľa študentov to v prvom ročníku vzdá a tak táto hodnota je skreslená. Prezenčné naväzujúce magisterské obory majú lepšie výsledky ako prezenčné bakalárske programy. Čo sa týka kombinovaných štúdií, tie sú na tom podstatne horšie ako prezenčné. V prípade oboru Účetníctví a finanční řízení podniku je táto hodnota až 2.38, čo znamená, že teoreticky každý študent má z predmetov len známku D a sem tam niečo lepšie. Priemer prezenčných štúdií je až o 0.24 lepší ako u kombinovaných.

3.4.4 Analýza vzťahu medzi úspechom na prijímacích skúškach a vysokoškolským úspechom



Obrázok č. 33: Vzťah bodov z prijímacích skúšok a priemerom a ukončením štúdia

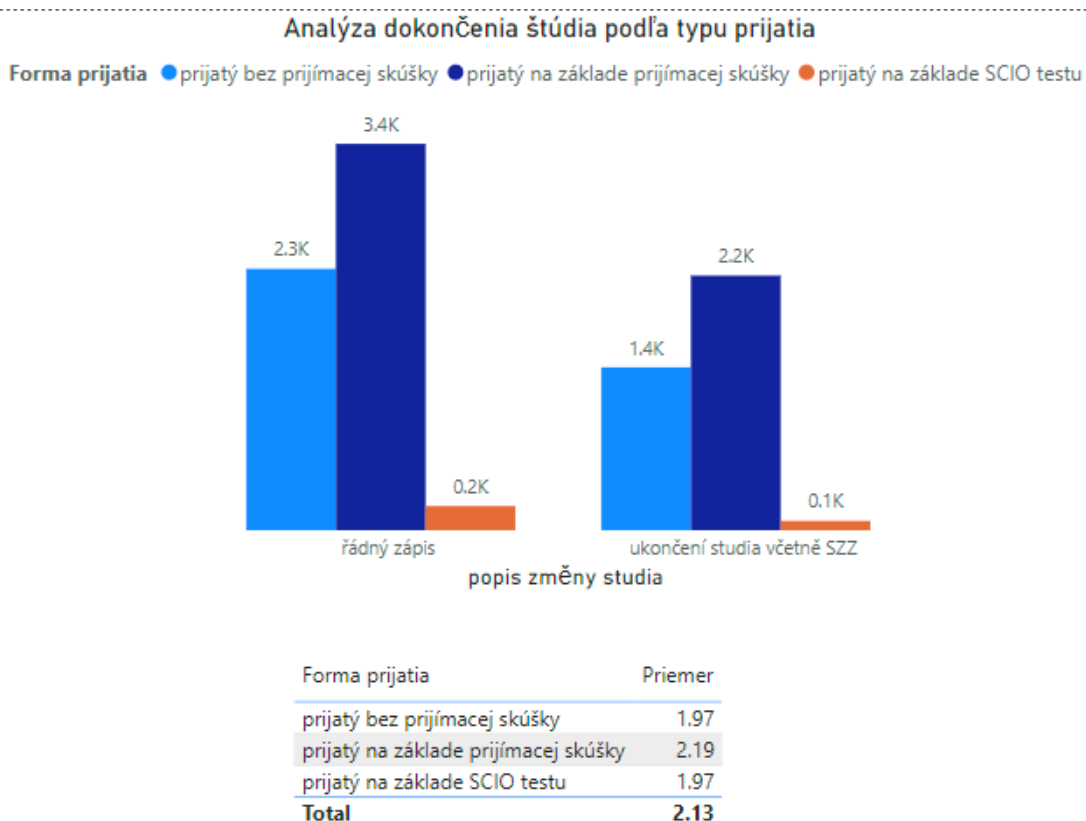
(Zdroj: Vlastné spracovanie)



Obrázok č. 34: Celkové body a ich vzťah

(Zdroj: Vlastné spracovanie)

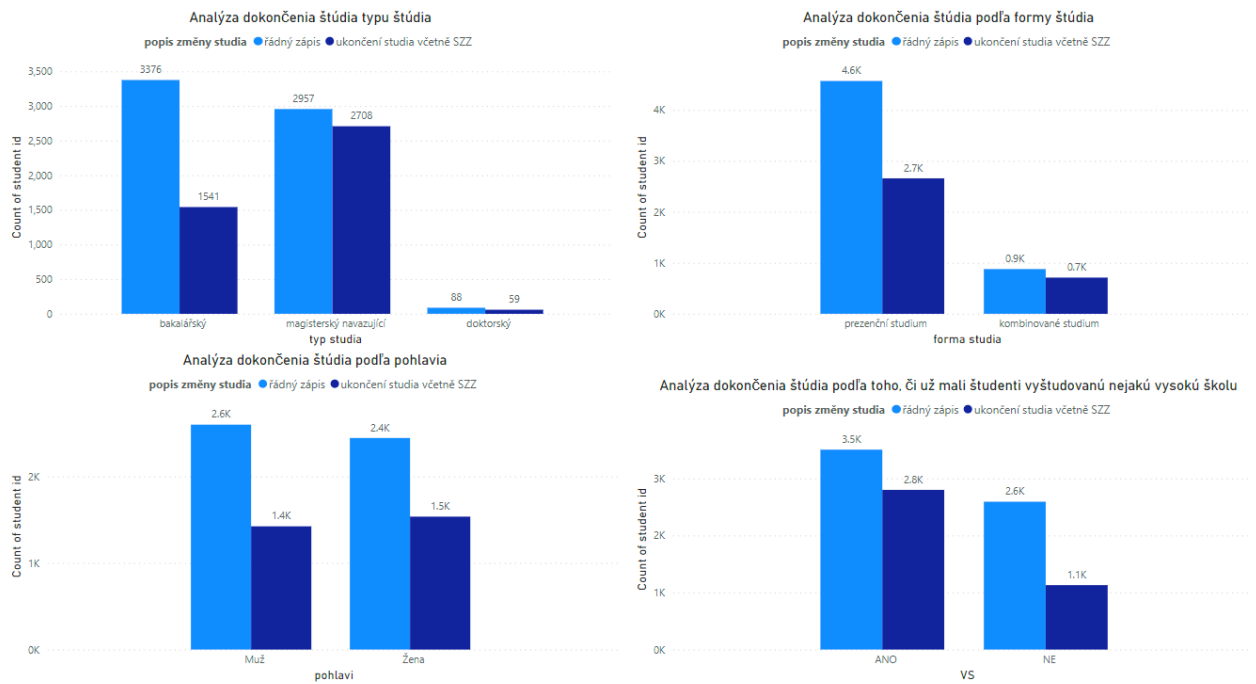
Na obrázkoch č.33 a č.34 je možné vidieť bodové grafy, kde sú porovnávané počty bodov z prijímacích skúšok s priemerom študentov a zároveň sú študenti rozdelení na úspešne doštudovaných a nedoštudovaných. Výsledky sú rovnomerné. Taktiež bolo zistené, ako je možné vidieť na obrázku č.35, že počas štúdia sú úspešnejší študenti, ktorí na prijímacej skúške neboli, takže výsledky prijímacích skúšok nie sú najlepším predikátorom úspechu na školu.



Obrázok č. 35: Dokonečnie štúdia podľa typu prijatia

(Zdroj: Vlastné spracovanie)

3.4.5 Analýza dokončenia štúdia



Obrázok č. 36: Analýza dokončenia štúdia

(Zdroj: Vlastné spracovanie)

Na obrázku č.36 je analýza, ktorá sledovala úspešnosti dokončenia štúdia. Zistilo sa, že ženy sú mierne úspešnejšie ako muži, že magisterské programy sú študenti náchylnejší dokončiť a že ak už má už študent vysokú školu, je pravdepodobné že dokončí aj túto.

3.4.6 Analýza stredných škôl podľa typu

Výsledky prijímacích testov a štúdia podľa typu strednej školy

Bude študent prijatý?	Počet prihlášiek	Priemer bodov z 1.testu	Priemer bodov z 2.testu	Priemer celkových bodov z prijímacej skúšky
Neprijatý	9965	25.93	14.90	40.59
Gymnázium	3364	27.21	15.67	42.20
Integrovaná stredná škola	73	24.53	14.48	38.83
Obchodní akadémia	2042	25.44	14.08	39.40
Stredná odborná škola	2200	25.48	13.14	38.63
Stredné odborné učilište	241	25.25	13.79	39.53
Zahraničná škola	2045	25.72	17.69	43.18
Prijatý	12208	46.60	17.82	64.66
Gymnázium	4500	48.07	18.31	66.70
Integrovaná stredná škola	55	46.77	16.03	63.00
Obchodní akadémia	2396	45.38	16.38	61.96
Stredná odborná škola	2359	46.26	16.57	63.14
Stredné odborné učilište	216	45.45	16.62	62.13
Zahraničná škola	2682	45.60	20.28	65.91
Total	22173	38.84	16.79	56.18

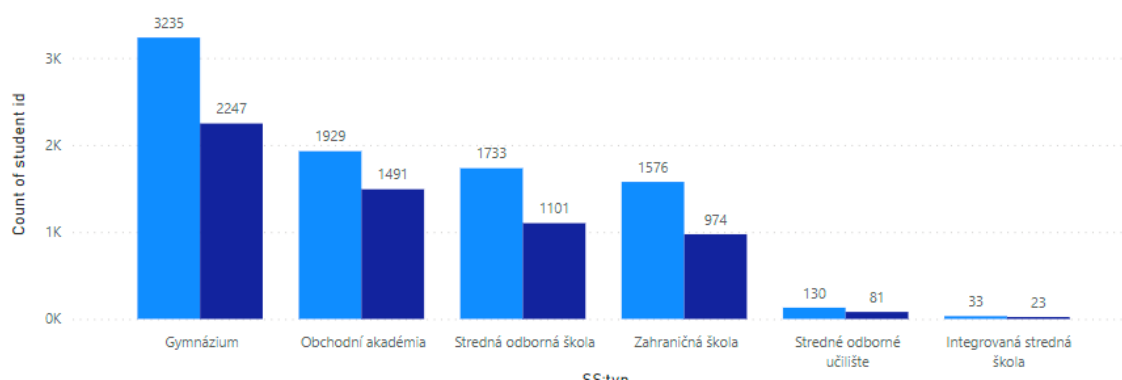
Obrázok č. 37: Stredné školy podľa typu – prihlášky

(Zdroj: Vlastné spracovanie)

Na obrázku č.37 je tabuľka, ktorá ukazuje počet prijatých študentov podľa typu strednej školy a ich výsledky na prijímacích skúškach. Najviac študentov je z gymnázia, obchodnej akadémie a strednej odbornej školy. Prijímacie skúšky píše žiaci rovnako, študenti zo zahraničných škôl sú lepší v teste z anglického jazyka.

Koľko zapísaných študentov doštudovalo podľa typu školy

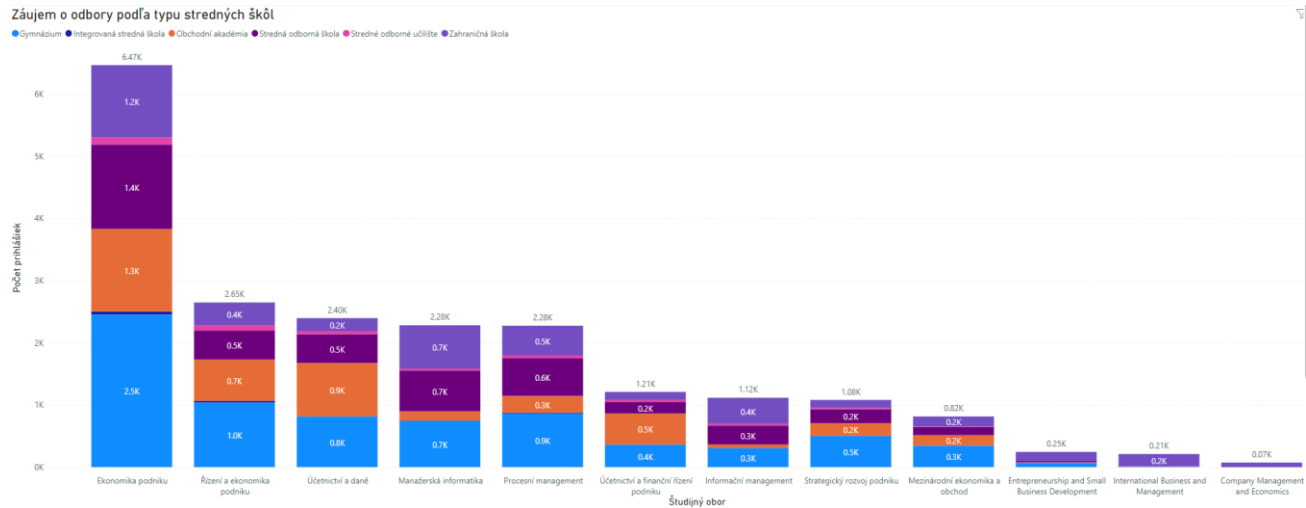
Počet študentov čo doštudovali vrátane SZZ ● rádny zápis ● ukončení studia včetně SZZ



Obrázok č. 38: Dokonečnie podľa typu strednej školy

(Zdroj: Vlastné spracovanie)

Študenti z obchodnej akadémie majú lepšie predpoklady dokončiť školu, zatiaľ čo študenti gymnázií často školu nedokončia. Niekde medzi sú zvyšné školy.



Obrázok č. 39: Stredné školy podľa oborov

(Zdroj: Vlastné spracovanie)

Záujem o obory podľa stredných škôl je celkom rovnomerne rozprestretý, s tým že o ekonomiku podniku je najväčší záujem. Študenti zo stredných odborných škôl a zahraničných škôl preferujú manažersku informatiku a procesný management, zatiaľ čo študenti z obchodných akadémií preferujú ekonomické odbory.

3.4.7 Analýza vysokých škôl podľa názvov a fakúlt

VŠ	Počet prihlášok	Fakulty najčastejších VŠ	Počet prihlášok
Vysoké učení technické v Brně	5588	<input type="checkbox"/> Vysoké učení technické v Brně	5586
Mendelova univerzita v Brně	1661	<input type="checkbox"/> Podnikatelská	3860
Masarykova univerzita	1055	<input type="checkbox"/> Informatická	605
Vysoká škola polytechnická v Jihlavě	252	<input type="checkbox"/> Strojnícka	456
Univerzita Tomáše Bati	176	<input type="checkbox"/> Elektrotechnická	385
NEWTON College a.s.	165	<input type="checkbox"/> Stavebná	178
Univerzita Palackého	122	<input type="checkbox"/> Chemická	89
Technická Univerzita Ostrava - Vysoká škola báňská	92	<input type="checkbox"/> Architektúra	7
Vysoká škola ekonomická v Praze	81	<input type="checkbox"/> Centrum sportovních aktivit	4
Akademie Sting	72	<input type="checkbox"/> Fakulta	1
Vysoká škola obchodní a hotelová Brno	66	<input type="checkbox"/> Výtvarná	1
České vysoké učení technické v Praze	65	<input type="checkbox"/> Mendelova univerzita v Brně	1660
Vysoká škola Karla Engliše v Brně	65	<input type="checkbox"/> Provozně ekonomická fakulta	1394
Soukromá vysoká škola ekonomická Znojmo	62	<input type="checkbox"/> Regionálního rozvoje a mezinárodních studií	167
Univerzita obrany v Brně	61	<input type="checkbox"/> Agronomická	50
Univerzita Pardubice	56	<input type="checkbox"/> Lesnická a dřevářská	25
B.I.B.S.	51	<input type="checkbox"/> Zahradnická	11
Slovenská technická univerzita v Bratislave	51	<input type="checkbox"/> Podnikatelská	10
Slezská univerzita Opava	43	<input type="checkbox"/> Informatická	1
Karlova univerzita	38	<input type="checkbox"/> Institut celoživotního vzdělávání	1
Moravská vysoká škola Olomouc	38	<input type="checkbox"/> Pedagogická	1
Vysoká škola technická a ekonomická v Českých Budějovicích	37	<input type="checkbox"/> Masarykova univerzita	1054
Univerzita Komenského v Bratislave	36	<input type="checkbox"/> Ekonomicko – správní	343
AMBIS vysoká škola, a. s.	30	<input type="checkbox"/> Informatická	180
Ekonomická univerzita v Bratislave	30	<input type="checkbox"/> Přírodovědecká	143
Univerzita Hradec Králové	27	<input type="checkbox"/> Filozofická	130
Jihočeská univerzita v Českých Budějovicích	25	<input type="checkbox"/> Právnícká	86
Česká zemědělská univerzita	22	<input type="checkbox"/> Pedagogická	49
Evropský polytechnický institut	20	<input type="checkbox"/> Športové štúdie	46
Žilinská univerzita v Žilině	18	<input type="checkbox"/> Lekárska	33
Veterinární a farmaceutická univerzita Brno	17	<input type="checkbox"/> Sociálních studií	33
SVŠE Znojmo	14	<input type="checkbox"/> Fakulta filozofická	9
Kwame Nkrumah University of Science & Technology	13	<input type="checkbox"/> Farmaceutická	1
Ostravská univerzita v Ostravě	12	<input type="checkbox"/> Lesnická a dřevářská	1
Technická univerzita v Liberci	12	<input type="checkbox"/> Vysoká škola polytechnická v Jihlavě	246
Total	10894	<input type="checkbox"/> Univerzita Tomáše Bati	176

Obrázok č. 40: Analýza vysokých škôl

(Zdroj: Vlastné spracovanie)

Na obrázku č.40 sú na ľavej strane zobrazené počty prihlášok podľa vysokých škôl. Najviac študentov prichádza z najväčších brnenských vysokých škôl a preto sú na pravej strane zobrazené fakulty týchto škôl, aby sa zistilo, z ktorých chodí najviac študentov. Veľa študentov pokračuje na podnikateľskej škole na magisterské štúdium, ale veľa zase príde aj z FEKTu, FSI alebo FITu. Veľký záujem študentov prichádza taktiež z ekonomickej fakulty MUNI a hlavne z Provozně ekonomickej fakulty Mendelovej univerzity.

3.5 Zhrnutie výsledkov analýz a návrh opatrení

3.5.1 Pokles prihlášok

Bol zistený trend výrazného poklesu prihlášok na Fakultu, na základe ktorých sa aj zmiernili podmienky prijatia, ale zaznamenalo sa len mierne zlepšenie. Na porozumenie dôvodu poklesu je potrebné zistiť, či klesol aj počet vysokoškolákov v ČR.

	2018/2019	2019/2020	2020/2021	2021/2022	2022/2023
Děti v mateřských školách	41 618	41 796	41 058	41 612	42 422
Žáci v základních školách	103 570	105 272	106 890	107 848	112 842
Žáci ve středních školách	45 611	45 755	46 762	47 978	49 863
Studenti vysokých škol ¹⁾	27 884	27 415	28 169	28 222	27 868

Obrázok č. 41: Počet študentov v Brně

(Zdroj: (15))

Podľa obrázku č.41 , ktorá zobrazuje údaje z Českého štatistického úradu, počet študentov v Brně neklesol. Preto treba hľadať príčinu inde. Analýza počtu prihlášok podľa oborov ukázala, že jediný obor kde neklesol počet študentov a záujem je manažérska informatika a informačný management, čo je obor viac smerujúci k IT. Z tohto sa dá usúdiť, že študenti začali čo sa týka ekonomie voliť iné univerzity, ako Masarykovu univerziu alebo PEF Mendelu z dôvodu, že majú lepšiu reputáciu a pravdepodobne lepší marketing. Moje odporúčanie Fakulte by bolo:

- zlepšiť a aktualizovať vzdelávacie programy, aby boli čo najrelevantnejšie pre dnešný trh a odlišovali sa od ekonomických oborov MUNI a PEF Mendelu.
- Aktívne propagovanie úspechov absolventov a zamestnatel'nosti, aby študenti videli hodnotu štúdia na VUT.
- Zlepšenie marketingovej komunikácie a propagácie univerzity, aby bola atraktívnejšia pre potenciálnych študentov.

Na základe analýz bolo zistené aj najlepšie smerovanie marketingu podľa historických dát. Fakulta by tak hrala na svoje silné stránky. Bolo zistené že:

- Najviac študentov je z Jihomoravského kraja a to konkrétne v Brně, Břeclavi, Blansku, Olomouci, Třebíči a podobne
- Ekonomické odbory preferujú hlavne študenti obchodných akadémií a gymnázií
- Na naväzujúce magisterské programy prichádza veľa ľudí z ekonomickej fakulty MUNI a Mendelovej univerzity

Na základe týchto informácií by svoj marketing Fakulta mohla smerovať na gymnázia a obchodné akadémie v Brně a Jihomoravskom kraji. Taktiež by sa mohla zamerať na študentov PEF Mendelu a MUNI, pretože z ich bakalárskych programov prechádza veľa študentov. Ideálne by bolo študentov nalákať na ekonomický program relevantný pre dnešný trh odlišujúci sa od bakalárskych programov MUNI a Mendelovej univerzity.

3.5.2 Začaté a nedokončené štúdiá

Bolo zistené, že veľa študentov sa zapíše a nedokončí štúdium. Sú to hlavne študenti gymnázií a stredných odborných škôl na bakalárskych prezenčných programoch. Môže to mať rôzne príčiny, ako napríklad že programy sú špecializované a oni vedia skôr všeobecné veci, čo by mohlo byť pravdou keďže študenti obchodných akadémií sú úspešnejší. Pre týchto študentov je však aj typické, že ešte nemajú dobrú predstavu o budúcej kariére.

Vhodným opatrením by bolo viac investovať do marketingu dňa otvorených dverí a upraviť ho tak, aby študenti mali dobrú predstavu o tom čo ich čaká, poprípade upraviť študijné osnovy tak, aby vysvetľovali učivo viac od základov, aby s tým študenti zo všeobecných škôl nemali problém alebo zaviesť už od prvého ročníka prax, aby si učivo vedeli previesť do reálneho života.

3.5.3 Zlé študijné výsledky bakalárskych oborov a kombinovaných štúdií

Bolo zistené, že magisterské prezenčné odbory sú pre študentov najideálnejšie. Študenti dosahujú dobrých známok ale nie v zmysle, že by tie študijné odbory boli ľahké. Bakalárske obory majú najhorší priemer hlavne v prvých ročníkoch, na čo už reagovalo predošlé zhrnutie a návrh. Zlé výsledky kombinovaných štúdií pravdepodobne znamenajú, že študenti sa venujú viac práci ako škole a snažia sa len jednoducho prejsť. Bolo by vhodné upraviť učebné osnovy tak, aby študentov vedeli zapojiť viac na víkendových hodinách, keď je na to vyhradený čas, aby sa štúdiu viacej venovali.

Záver

V diplomovej práci bol z úvodu spracovaný teoretický základ, nutný pre pochopenie informácií v diplomovej práci.

V analytickej časti bola predstavená univerzita VUT a Fakulta podnikateľská vrátane možných smerov na Fakulte podnikateľskej, ktorými sa môžu študenti vydať. Po predstavení bol vysvetlený priebeh štúdia študentov na Fakulte a v závere boli vytvorené analýzy, ktorých cieľom bolo zistiť cenné informácie, ktoré by mohli pomôcť vedeniu Fakulty podnikateľskej pri strategickom plánovaní.

V praktickej časti boli popísané dáta o prihláškach a štúdiu študentov, ktoré boli získané od vedenia fakulty. Tieto dáta boli vyčistené a modelované na prehľadné dáta vhodné na analýzu. Keď dáta mali vhodnú formu, boli zostrojené vizualizácie analýz a popísané ich výsledky, na základe ktorých bolo vytvorené celkové zhrnutie a návrhy opatrení.

Ciele diplomovej práce boli úspešne splnené, práca je okrem hľadania cenných informácií aj dobrým návodom na čistenie dát.

Zoznam použitej literatúry

- (1) POŽÁR, Josef. *Manažerská informatika*. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk, 2010. ISBN 978-80-7380-276-9.
- (2) SKLENÁK, Vilém. *Data, informace, znalosti a Internet*. Praha: C.H. Beck, 2001. C.H. Beck pro praxi. ISBN 80-717-9409-0.
- (3) LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*. Brno: Computer Press, 2012. ISBN 978-80-251-3729-1.
- (4) NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. *Business intelligence: jak využít bohatství ve vašich datech*. Praha: Grada, 2005. Management v informační společnosti. ISBN 80-247-1094-3.
- (5) LACKO, Luboslav. *Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle*. Brno: Computer Press, 2003. ISBN 80-722-6969-0.
- (6) HARINATH, Sivakumar. *Professional Microsoft SQL Server 2012 Analysis Services with MDX and DAX*. Indianapolis: Wiley, 2012. ISBN 978-1-118-10110-0.
- (7) What Is Data Mining? How It Works, Benefits, Techniques, and Examples. In: *Investopedia* [online]. [cit. 2023-05-13]. Dostupné z: <https://www.investopedia.com/terms/d/datamining.asp>
- (8) The Rise and Fall of the OLAP Cube. In: *The Hollistics Blog* [online]. [cit. 2023-05-13]. Dostupné z: <https://www.hollistics.io/blog/the-rise-and-fall-of-the-olap-cube/>

- (9) *Vizualizace dat | Microsoft Power BI* [online]. [cit. 2023-05-13]. Dostupné z: <https://powerbi.microsoft.com/cs-cz/>
- (10) Welcome to Python.org [online]. [cit. 2023-05-13]. Dostupné z: <https://www.python.org/>
- (11) *What is Fuzzy Matching?* | Redis. In: *Redis | The real time data platform* [online]. [cit. 2023-05-14]. Dostupné z: <https://redis.com/blog/what-is-fuzzy-matching/>
- (12) Profil univerzity - VUT. In: *Vysoké učení technické v Brně* [online]. [cit. 2023-05-10]. Dostupné z: <https://www.vut.cz/vut/profil>
- (13) Studuj Fakultu podnikatelskou [online]. [cit. 2023-05-10]. Dostupné z: <https://www.studujfp.cz/cs/?studium=all>
- (14) Přijímačky Fakulta podnikatelská VUT. In: *Fakulta podnikatelská, VUT v Brně* [online]. [cit. 2023-05-10]. Dostupné z: <https://www.fp.vut.cz/cs/pro-uchazece/prijimacky#bakalarske-studium>
- (15) Vzdělávání - Kraj | ČSU v Brně. In: *Český statistický úřad | ČSU v Brně* [online]. [cit. 2023-05-15]. Dostupné z: <https://www.czso.cz/csu/xb/vzdelavani-xb>

Zoznam obrázkov

Obrázok č. 1: Dáta a informácie	14
Obrázok č. 2: Vrstvy BI.....	16
Obrázok č. 3: Data mining.....	19
Obrázok č. 4: OLAP datová kocka	20
Obrázok č. 5: Prepojenie jednotlivých Power BI aplikácií.....	22
Obrázok č. 6: Logo jazyku Python	24
Obrázok č. 7: Prvý súbor	34
Obrázok č. 8: Druhý súbor.....	34
Obrázok č. 9: Funkcia na získanie roku.....	35
Obrázok č. 10: Získanie prijatých študentov pomocou kódu v M.....	37
Obrázok č. 11: Získanie aktuálnych oborov pomocou kódu v M.....	38
Obrázok č. 12: Hierarchia adresy	38
Obrázok č. 13: Názvy stredných škôl pred čistením	39
Obrázok č. 14: Získanie typu stredných škôl pomocou kódu v M	41
Obrázok č. 15: Hierarchia stredných škôl.....	41
Obrázok č. 16: Export názvov vysokých škôl	43
Obrázok č. 17: Vytvorenie zhlukov vysokých škôl pomocou fuzzy porovnávania	44
Obrázok č. 18: Vytvorené zhluky po spustení kódu v Pythone.....	45
Obrázok č. 19: Spojenie mapovacej tabuľky s prihláškami.....	46

Obrázok č. 20: Zhlukované názvy škol po namapovaní	47
Obrázok č. 21: Názvy fakúlt pred čistením	48
Obrázok č. 22: Názvy fakúlt po čistení.....	49
Obrázok č. 23: Finálne názvy stĺpcov po vyčistení	50
Obrázok č. 24: Tvorenie prepojenia medzi tabuľkami	52
Obrázok č. 25: Rozdelenie študentov podľa zemí	53
Obrázok č. 26: Obce s najväčším počtom študentov	54
Obrázok č. 27: Sústreďenie najväčšieho počtu študentov	54
Obrázok č. 28: Záujem o obory podľa obcí	55
Obrázok č. 29: Prihlásení a zapísaní študenti podľa rokov a štúdiá	56
Obrázok č. 30: Prihlášky a zápisy na jednotlivé obory.....	57
Obrázok č. 31: Priemerné počty bodov a známok podľa oborov	58
Obrázok č. 32: Priemery podľa typu štúdiá a oborov	59
Obrázok č. 33: Vzťah bodov z prijímacích skúšok a priemerom a ukončením štúdiá ...	60
Obrázok č. 34: Celkové body a ich vzťah.....	60
Obrázok č. 35: Dokonečnie štúdiá podľa typu prijatia	61
Obrázok č. 36: Analýza dokončenia štúdiá	62
Obrázok č. 37: Stredné školy podľa typu – prihlášky	63
Obrázok č. 38: Dokonečnie podľa typu strednej školy.....	63
Obrázok č. 39: Stredné školy podľa oborov	64
Obrázok č. 40: Analýza vysokých škôl	65

Obrázok č. 41: Počet študentov v Brně..... 66