

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačního inženýrství



Diplomová práce

**Využití dashboardů v záležitosti vizualizace a analýzy
dat**

Bc. Lukáš Tůma

© 2022 ČZU v Praze

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Provozně ekonomická fakulta

ZADÁNÍ DIPLOMOVÉ PRÁCE

Lukáš Tůma

Systemové inženýrství a informatika
Informatika

Název práce

Využití dashboardů v záležitosti vizualizace a analýzy dat

Název anglicky

Usage of Dashboards in the matter of Data Visualization and Analysis

Cíle práce

Diplomová práce je zaměřena na problematiku využití dashboardů pro účely analýzy a vizualizace dat jakožto nástroje pro zpracování a efektivní využití informací pro potřeby rozhodování a řízení podniku. Náplní a účelem této práce je:

- vymezit historii a principy analýzy a vizualizace dat, objasnit s tím související základní pojmy,
- popsat typy grafů a jejich vhodné uplatnění během vizualizace dat včetně použití dashboardů pro potřeby rozhodování a řízení podniku,
- popsat využití metody storytelling jakožto efektivní metody v oblasti vizualizace dat,
- zmapovat používané nástroje používané pro vizualizaci a analýzu dat a specifikovat jejich výhody a nevýhody,
- demonstrovat možnosti těchto nástrojů prostřednictvím realizace několika dashboardů pro vybraný podnik,
- zobecnit přínosy takového využití pro další možná použití v podnikatelské praxi.

Metodika

Použitá metodika zadané diplomové práce bude založena na studiu a analýze dostupných informačních zdrojů a existujících řešení v dané oblasti. Navrhované řešení bude zohledňovat identifikované požadavky a očekávání spojená s řešenou záležitostí. Na podkladě syntézy teoretických poznatků a dosažených výsledků budou formulovány závěry této diplomové práce a následně zobecněny pro další možná použití.

Závazný harmonogram:

Teoretické principy řešené problematiky, literární rešerše – předmět 1. zápočtu z DP: do 5.9.2021

Zmapování momentální situace řešené problematiky, identifikace požadavků s tím spojených: do 30.11.2021

Navržení možného řešení a jejich následné ověření: předmět 2. zápočtu z DP: do 31.1.2022

Zobecnění navržených záležitostí pro další možná použití – předmět 3. zápočtu z DP: do 25.3.2022

Doporučený rozsah práce

50-60 stran

Klíčová slova

datová analýza, vizualizace dat, graf, dashboard, report, storytelling

Doporučené zdroje informací

CUESTA, Hector. Analýza dat v praxi. Brno: Computer Press, 2015. ISBN 978-80-251-4361-2.

GEMIGNANI, Chris, Richard GALENTINO a Patrick Jude SCHUERMANN. Efektivní analýza a využití dat.

Brno: Computer Press, 2015. ISBN 978-80-251-4571-5.

KNAFLIC, Cole Nussbaumer. Storytelling with Data: A Data Visualization Guide for Business Professionals.

Wiley, 2015. ISBN 9781119002253.

SLEEPER, Ryan. Practical Tableau: 100 Tips, Tutorials, and Strategies from a Tableau Zen Master. O'Reilly

Media, Inc, USA, 2018. ISBN 1491977310.

SPENCE, Robert. Information Visualization. 3rd Edition. Switzerland: Springer International Publishing,

2014. ISBN 978-3-319-07341-5.

Předběžný termín obhajoby

2021/22 LS – PEF

Vedoucí práce

doc. Dr. Ing. Václav Vostrovský

Garantující pracoviště

Katedra informačního inženýrství

Elektronicky schváleno dne 23. 11. 2021

Ing. Martin Pelikán, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 25. 11. 2021

Ing. Martin Pelikán, Ph.D.

Děkan

V Praze dne 26. 01. 2022

Čestné prohlášení

Prohlašuji, že svou diplomovou práci "Vyžití dashboardů v záležitosti vizualizace a analýzy dat" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 29.03.2022

Poděkování

Rád bych touto cestou poděkoval vedoucímu diplomové práce doc. Ing. Václavovi Vostrovskému, Ph.D. za vstřícný přístup a užitečné rady, které mi pomohly při zpracování diplomové práce.

Využití dashboardů v záležitosti vizualizace a analýzy dat

Abstrakt

Tato diplomová práce se zabývá datovou analýzou a vizualizací dat. Diplomová práce se skládá ze dvou částí. První část je teoretická část, která začíná popisem historie vizualizace dat, nejvýznamnějších historických milníků a dále je popsán vliv vizualizace dat na schopnost člověka zpracovat a zapamatovat si důležité informace. Po tomto úvodu následuje popis hlavních metody vizualizace dat, samotného procesu tvorby vizualizace a nejvýznamnějších chyb vizualizace, které mohou vést k nepochopení prezentovaných informací. Dále je popsána metoda data storytelling, jakožto moderní a efektivní metoda používaná při předávání informací obsažených v datech publiku. Následující kapitola teoretické části popisuje dva typy nástrojů pro vizualizaci dat. Prvním typem je hotový software připravený k použití pro každého, druhým typem nástrojů jsou programovací jazyky (například Python, JavaScript či R), které nabízí široké možnosti pro tvorbu vizualizace dat. Poslední část pokrývá principy Business Intelligence a rozdíl mezi reportem a dashboardem. Druhá část diplomové práce je praktická část, který je orientována na vytvoření několika dashboardů pro vybraný podnik. Tyto dashboardy budou sloužit vedení podniku k efektivnějšímu rozhodování a řízení společnosti.

Klíčová slova: datová analýza, vizualizace dat, graf, dashboard, report, data storytelling

Usage of Dashboards in the matter of Data Visualization and Analysis

Abstract

This diploma thesis focuses on data analysis and data visualization. The diploma consists of two parts. The first is the Theoretical part, which begins with describing the history of data visualization, the most important historical milestones, and the influence of data visualization on the human ability to process and remember important information. Following this introduction, the main methods of data visualization, the visualization process itself, and the most important visualization mistakes that can lead to misunderstanding of the presented information are described. After this, the method of data storytelling is described as a modern and effective method used to convey the information, which is contained in the data, to the audience. The next section of the theoretical part describes the two types of tools for data visualization. The first type is ready-to-use software for everyone; the second type is programming languages (such as Python, JavaScript or R), which offer a wide range of features for creating data visualization. The final section covers the principles of Business Intelligence and the difference between a report and a dashboard are described. The second part of this thesis is the Practical part, which focuses on the creation of several dashboards for the selected company. These dashboards will be used by the leaders of the company for more effective decision-making and company management.

Keywords: data analysis, data visualization, graph, dashboard, report, data storytelling

Obsah

1. Úvod.....	13
2. Cíl práce a metodika	16
2.1. Cíl práce	16
2.2. Metodika	16
2.2.1. Zmapování současného stavu	17
2.2.2. Sběr a analýza požadavků.....	18
2.2.3. Realizace centrálního úložiště dat.....	18
2.2.4. Realizace ETL procedur	18
2.2.5. Realizace dashboardů ve vybraném nástroji.....	19
2.2.6. Ověření splnění požadavků.....	19
3. Teoretická východiska	20
3.1. Historie vizualizace dat	20
3.2. Vliv vizualizace na paměť.....	24
3.3. Metody vizualizace dat	25
3.3.1. Tabulky	26
3.3.2. Grafy	26
3.3.3. Mapy	31
3.3.4. Proces tvorby vizualizace	34
3.4. Skupiny dat.....	38
3.4.1. Kvalitativní členění.....	39
3.4.2. Kvantitativní členění.....	39
3.5. Typy médií	39
3.5.1. Tištěné noviny.....	40
3.5.2. Marketingové tiskoviny	41
3.5.3. Televizní přenos.....	41
3.5.4. Internet.....	42
3.6. Chyby vizualizace	42
3.6.1. Faktor lživosti	42
3.6.2. Nevhodné prvky.....	43
3.7. Data Storytelling	46
3.8. Datové zdroje, získání a zpracování dat.....	48
3.8.1. Datové zdroje.....	48
3.8.2. Databáze.....	49
3.8.3. Jazyk SQL.....	50
3.9. Nástroje používané k vizualizaci dat.....	51
3.9.1. Programátorské nástroje	51

3.9.2. Hotový software	54
3.10. Business Intelligence	60
4. Vlastní práce	63
4.1. Představení společnosti	63
4.1.1. Reference	63
4.1.2. Partneři	64
4.2. Popis současného stavu	64
4.2.1. DMS ELO	65
4.2.2. Docházkový systém	66
4.2.3. CRM systém.....	68
4.2.4. Získávání informací pro potřeby řízení společnosti.....	69
4.3. Sběr a analýza požadavků	70
4.4. Návrh řešení	72
4.4.1. Součásti řešení.....	72
4.4.2. Architektura systému	73
4.4.3. Použité technologie	73
4.5. Realizace centrálního úložiště dat	74
4.5.1. Návrh E-R diagramu	75
4.5.2. Realizace úložiště pomocí SQL skriptu	75
4.6. Realizace ETL procedur	76
4.6.1. Dokument mapování zdroje na cíl	76
4.6.2. Aplikace ETL procedury.....	77
4.6.3. Načítání kurzů měn	78
4.6.4. Řešené problémy	79
4.7. Realizace dashboardů	80
4.7.1. Návrh drátěných modelů dashboardů	80
4.7.2. Realizace SQL dotazů	84
4.7.3. Výběr vhodných typů grafů	87
4.7.4. Realizace dashboardů.....	88
4.8. Testování	91
5. Výsledky a diskuse	95
5.1. Zobecnění přínosů navrženého řešení	95
5.2. Aspekty znemožňující zúročení navrženého řešení	96
5.3. Možnosti dalšího rozvoje	97
6. Závěr.....	100
7. Seznam použitých zdrojů.....	104
8. Přílohy	107

Seznam obrázků

Obrázek 1 - Vývoj velikosti trhu s rozšířenou a virtuální realitou	15
Obrázek 2 - První známý statistický graf.....	20
Obrázek 3 - Graf Williama Playfaira	21
Obrázek 4 - Mapa Johna Snowa o vypuknutí epidemie cholery	21
Obrázek 5 - Graf Charlese Minarda "Napoleon's 1812 Russian Campaign"	22
Obrázek 6 - Graf Florence Nightingale "Coxcomb"	23
Obrázek 7 - Lidská paměť	25
Obrázek 8 - Sloupcový graf	27
Obrázek 9 - Koláčový graf.....	28
Obrázek 10 - 100% skládaný sloupcový graf	28
Obrázek 11 - Spojnicový graf.....	29
Obrázek 12 - Bodový graf	29
Obrázek 13- Krabicový diagram.....	30
Obrázek 14 – Histogram	31
Obrázek 15 - Barevná mapa.....	32
Obrázek 16 - Proporční symbolická mapa.....	33
Obrázek 17 - Bodová mapa	33
Obrázek 18 - Spojnicová mapa	34
Obrázek 19 - Rozhodovací strom pro volbu typu grafu	37
Obrázek 20 – Vizualizace v tištěných novinách	40
Obrázek 21 - Vizualizace v rámci televizního přenosu	41
Obrázek 22 - Zavádějící osy	43
Obrázek 23 - Izolované pozorování	44
Obrázek 24 - 3D efekty a koláčové grafy	45
Obrázek 25 - Graf se dvěma osami.....	45
Obrázek 26 - Metoda Martini glass	48
Obrázek 27 - Open Data	49
Obrázek 28 - Jazyk SQL	51
Obrázek 29 - Python Seaborn	53
Obrázek 30 - Popularita jazyka R	54
Obrázek 31 - Microsoft Excel.....	55
Obrázek 32 - Google Data Studio.....	56
Obrázek 33 - Power BI Designer	57
Obrázek 34 - Tableau Desktop	59
Obrázek 35 - ArcGIS Online	60
Obrázek 36 - Report vs dashboard.....	61
Obrázek 37 - Reference společnosti Exxon s.r.o.	64
Obrázek 38 - DMS ELO	66
Obrázek 39 - Docházkový systém	67
Obrázek 40 - Export dat z Docházkového systému	68
Obrázek 41 - Systém CRM.....	69
Obrázek 42 - Architektura řešení	73
Obrázek 43 - E-R diagram centrálního úložiště dat.....	75
Obrázek 44 - Drátěný model dashboardu obchodních příležitostí	81
Obrázek 45 - Drátěný model dashboardu ziskovosti projektů.....	81
Obrázek 46 - Drátěný model dashboardu činností zaměstnanců	82

Obrázek 47 - Drátěný model dashboardu nákladů středisek.....	82
Obrázek 48 - Drátěný model dashboardu neuhrazených pohledávek.....	83
Obrázek 49 - Drátěný model dashboardu tržeb	83
Obrázek 50 - Dashboard Obchodní příležitosti.....	88
Obrázek 51 - Dashboard Ziskovost projektů	89
Obrázek 52 - Dashboard Činnosti zaměstnanců	89
Obrázek 53 - Dashboard Náklady středisek.....	90
Obrázek 54 - Dashboard Neuhrazené pohledávky.....	90
Obrázek 55 - Dashboard Tržby	91

Seznam tabulek

Tabulka 1 - Základní informace o společnosti Exxon s.r.o.	63
Tabulka 2 - Katalog požadavků	71
Tabulka 3 - Požadavek na získání historického kurzu měny	79
Tabulka 4 - Ověření požadavků dle katalogu požadavků	94

Seznam použitých zkratk

API	Application Programming Interface
BI	Business Intelligence
CRM	Customer Relationship Management
CSV	Comma-separated values
DB	Database
DBMS	Database Management Systems
DLL	Dynamic-link library
DMS	Document Management System
ECM	Enterprise Content Management
ER	Entity-Relationship Model
ETL	Extract, Transform, Load
EVA	Economic Value Added
GDPR	General Data Protection Regulation
IS	Information System
IT	Information Technology
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
PDF	Portable Document Format
RDBMS	Relational Database Management Systems
ROA	Return on Assets
ROE	Return on Equity
SQL	Structured Query Language
SŘBD	Systém řízení báze dat
URL	Uniform Resource Locator
XML	Extensible Markup Language

1. Úvod

Každý podnik může být dlouhodobě úspěšný a prosperující pouze v případě, že management podniku na všech úrovních řízení provádí správná rozhodnutí ve správný čas. S rostoucím množstvím informací, které mohou být pro rozhodování relevantní a které jsme schopni získat, sehrávají počítače v procesu rozhodování stále důležitější roli. Kvalita rozhodovacích procesů ovlivňuje zásadním způsobem fungování každého podniku. V rozhodovacím procesu je důležité, aby relevantní, bezchybné a kompletní informace měla k dispozici správná osoba v čase, kdy jsou pro rozhodnutí potřeba. Pokud jedna z uvedených skutečností selže, pravděpodobnost učinění nesprávného rozhodnutí je velmi vysoká. Pokud by například manažer měl jako podklad pro své rozhodnutí k dispozici nepravdivé, nekompletní či neaktuální informace, bude tím jeho rozhodnutí velmi negativně ovlivněno. Stejně tak je velký problém, pokud manažer dostane k dispozici správné informace, ale o několik dní později, než jsou požadovány.

Cesta k informacím potřebným k rozhodování však nemusí být vždy jednoduchá, protože tyto informace mohou být uloženy na různých místech, v různých informačních systémech podniku a schopnost manažera na ně pohlížet v kontextu a odhalovat skryté souvislosti může být omezená. Manažer navíc často nepotřebuje detailní informace uložené v transakčních systémech využívaných pro běžnou operativní činnost podniku, ale využije spíše souhrnné informace, ideálně ve formě vizualizací, ze kterých by na první pohled mohl vyčíst například vývoj prodeje daného produktu v čase, sezónní výkyvy prodeje, či měl k dispozici jednoduchý ukazatel výkonnosti podniku. Pokud na data manažer pohlíží pouze v tabulkách či databázích, je velmi složité v nich najít potřebné informace a vyvodit smysluplné závěry. Vizualizace dat je tak velmi účinným nástrojem pro podporu rozhodování manažerů a má pozitivní vliv na množství informací, které je schopen člověk zpracovat a zapamatovat si.

Pokud mají manažeři k dispozici vizualizace podnikových dat v podobě dashboardů, vytvořené v nástrojích vhodných pro daný účel a s použitím správných vizualizačních metod, jedná se bezesporu o konkurenční výhodu takového podniku oproti podnikům, které takovou podporu pro své rozhodování a řízení k dispozici nemají.

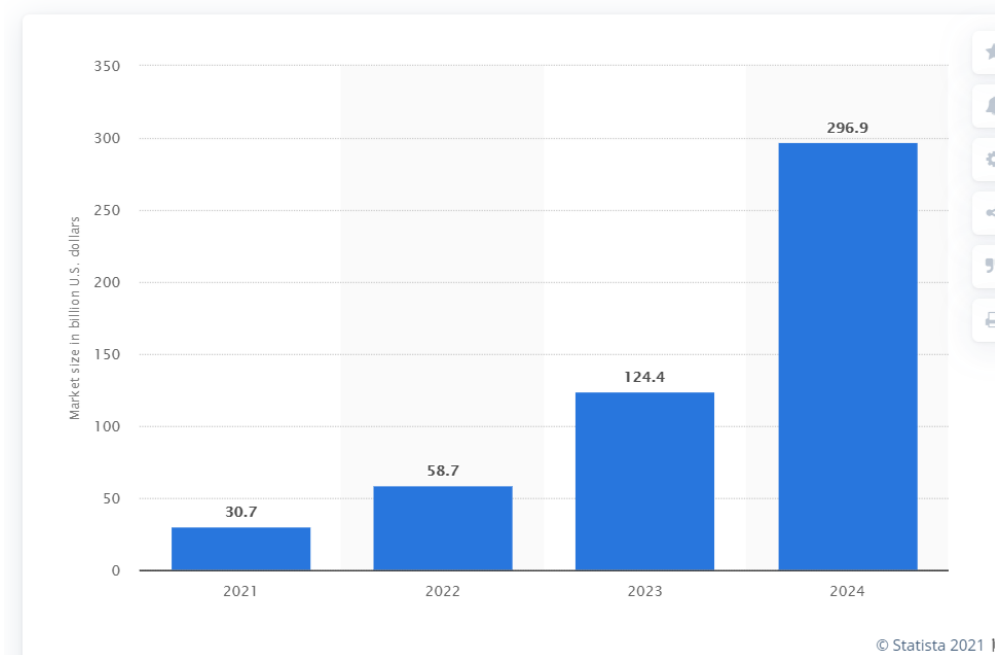
Vizualizace dat je vhodnou metodou nejen jako podpora řízení podniku bez ohledu na sektor podnikání či velikost podniku, ale jedná se i o účinnou metodu zprostředkování informací publiku například ve školství, zpravodajství a mnoha dalších oblastech.

Vizualizace dat je aktuální a stále se rozvíjející oblast, o čemž svědčí několik hlavních trendů, kterými se vizualizace dat aktuálně ubírá. Jedním takovým trendem je vizualizace velkých objemů dat, tzv. big data, v reálném čase, což v rychle se měnícím podnikatelském prostředí může být způsob, jak zůstat v obraze a předstihnout konkurenci. Vizualizace dat v reálném čase tak pomáhá urychlit rozhodovací proces a umožňuje včasné a lepší rozhodování. Příkladem takovéto vizualizace dat může být nástroj MonkeyLearn Dashboard, který v reálném čase monitoruje na sociálních sítích zpětnou vazbu zákazníků na daný produkt či službu a vizualizuje ji do přehledných a srozumitelných grafů. Pokud by se na sociálních sítích objevila negativní zpětná vazba od zákazníka, společnost na ni může ihned adekvátně reagovat.

Dalším zajímavým trendem je vizualizace dat prostřednictvím videa. Takovýto způsob vizualizace může být ještě zajímavější a poutavější než statická vizualizace, protože mysl člověka je uzpůsobena k tomu, aby videu věnovala pozornost. Například účastníci studie (Nizam, 2021) odpovídali na otázky o 323 % lépe, pokud měli k dispozici video pomůcku oproti účastníkům, kteří měli k dispozici pouze textové informace. Dle stejné studie dává 67 % uživatelů přednost seznámení s novým produktem nebo službou prostřednictvím videa před obrázkem nebo textem.

Oblastí, kde vizualizace také zažívá rychlý vývoj, je rozšířená a virtuální realita. Očekává se, že trh s rozšířenou a virtuální realitou bude mít v roce 2022 hodnotu přibližně 58,7 miliardy amerických dolarů a v dalších letech bude růst tohoto trhu ještě rychlejší. Například aplikace Ikea Place od společnosti Ikea, která umožňuje potenciálním zákazníkům prostřednictvím rozšířené reality vidět vybraný sortiment přímo v jejich domácnosti. Prostřednictvím katalogu produktů si uživatel přímo v aplikaci vybere produkt a poté jej jednoduchým přetažením umístí do virtuální místnosti, která vznikne vyfotografováním domácnosti zákazníka. Virtuální a rozšířená realita umožňuje uživateli interagovat s informacemi, jako by byly hmatatelné.

Augmented reality (AR) and virtual reality (VR) market size (in billion U.S. dollars)



Obrázek 1 - Vývoj velikosti trhu s rozšířenou a virtuální realitou

(Zdroj: <https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/>)

Vizualizace dat usnadňuje lepší osvojení klíčových obchodních informací, rychlejší přístup k cenným obchodním poznatkům, lepší pochopení obchodních operací a přesnou analýzu nálad zákazníků. Umožňuje také identifikaci vzorů a trendů a efektivnější rozhodování v reálném čase. Vzhledem k tomu, že objem dat rychle roste, je vizualizace dat jedním z nejlepších způsobů, jak co nejlépe využít všechna podniková data a udržet si náskok na konkurenčním trhu.

2. Cíl práce a metodika

2.1.Cíl práce

Diplomová práce je zaměřena na problematiku využití dashboardů pro účely analýzy a vizualizace dat jakožto nástroje pro zpracování a efektivní využití informací pro potřeby rozhodování a řízení podniku. Hlavním cílem předkládané diplomové práce je realizace vzorových dashboardů pro vybraný podnik umožňující managementu tohoto podniku optimalizovat svoje strategická a taktická rozhodování na podkladě výrazného zpřehlednění k tomu potřebných informací. V rámci procesu rozhodování a řízení společnosti bude mít management společnosti k dispozici požadované a aktuální informace v přehledné podobě umožňující jejich komplexní využití. Dílčími cíli v rámci takto specifikovaného hlavního cíle jsou pak tyto následující záležitosti:

- a) vymezit historii a principy analýzy a vizualizace dat, objasnit s tím související základní pojmy,
- b) popsat typy grafů a jejich vhodné uplatnění během vizualizace dat včetně použití dashboardů pro potřeby rozhodování a řízení podniku,
- c) popsat využití metody data storytelling jakožto efektivní metody v oblasti vizualizace dat,
- d) zmapovat používané nástroje používané pro vizualizaci a analýzu dat a specifikovat jejich výhody a nevýhody,
- e) demonstrovat možnosti těchto nástrojů prostřednictvím realizace několika dashboardů pro vybraný podnik,
- f) zobecnit přínosy takového využití pro další možná použití v podnikatelské praxi.

2.2.Metodika

Řešení předložené diplomové práce se bude mimo jiné opírat o následující metody a techniky.

- Interview – metodicky vedený rozhovor založený na přímém dotazování, tedy na verbální komunikaci výzkumného pracovníka s respondentem. Interview bude využito v rámci rozhovoru s vedením podniku, na základě kterého budou sesbírány požadavky na budoucí dashboardy.

- E-R model – technika, pomocí které lze vytvořit datový model. Tato technika bude využita pro návrh datového modelu datového úložiště, ze kterého budou dashboardy načítat potřebná data.
- Mapování toků dat – pro zmapování toku dat bude vytvořen dokument mapující zdroj dat na cíl dat, pomocí kterého budou popsány datové toky mezi zdrojovými informačními systémy a datovým úložištěm.
- Dekompozice – rozklad problému na dílčí podproblémy, které se pak řeší samostatně. Rozdělení programu do modulů a metod. Metoda dekompozice bude využita při realizaci procedury načítající data do úložiště dat pro dashboardy.
- Metody vizualizace dat – pro realizaci dashboardů budou využity různé metody vizualizace dat (grafy), například sloupcový graf, koláčový graf a další, které budou vybrány během návrhu a realizace dashboardů.
- Akceptační testování – metodika ověření shody vytvořeného řešení s požadavky sesbíranými během interview.

Pro splnění předsevzatého cíle je zvolena následující metodika (posloupnost kroků) řešení.

2.2.1. Zmapování současného stavu

V první fázi praktické části diplomové práce bude zmapován a popsán současný stav ve vybraném podniku, tedy jakým způsobem získává vedení podniku informace potřebné pro rozhodování a řízení podniku. Dále budou popsány používané informační systémy, které bude možné použít jako potenciální datové zdroje pro tvorbu dashboardů a budou popsána data, která jsou v rámci daných informačních systémů uchovávána. Dále budou popsány možnosti pro tvorbu reportů, které jsou aktuálními informačními systémy poskytovány a podnikem využívány.

2.2.2. Sběr a analýza požadavků

Cílem sběru a analýzy požadavků na řešení bude vymezit hranice budoucího řešení a vyjasnit si zadání s vedením podniku. Zdrojem informací zde budou především rozhovory s vedením podniku a také poznatky o informačních zdrojích získané předcházejícím popisem současného stavu. Výsledkem by měla být dokumentace požadavků v podobě strukturovaného textu, který bude obsahovat především popis požadavku, kategorizaci a prioritu požadavku. Každý požadavek by měl být jednoznačný a také ověřitelný, aby bylo možné po realizaci řešení říci, zda byl požadavek splněn, a to v dostatečné kvalitě. Na závěr sběru požadavků budou tyto požadavky odsouhlaseny vedením podniku.

2.2.3. Realizace centrálního úložiště dat

Po fázi vytvoření katalogu požadavků proběhne návrh centrálního úložiště dat pomocí datového E-R modelu a poté realizace úložiště pomocí skriptu v jazyce SQL. V rámci návrhu E-R modelu budou navrženy entity a jejich atributy, identifikátory a vztahy mezi entitami. Pro vytvoření E-R diagramu bude použita notace vyvinutá Jamesem Martinem, tzv. notace „ptačího pařátu“ (Crow's foot notation). Po vytvoření E-R modelu bude model transformován na návrh relační databáze. Při procesu transformace bude dbáno na to, aby databázové tabulky byly správně normalizovány, aby obsahovaly vhodná integritní omezení a aby atributy měly správné datové typy. Následně bude vytvořen SQL skript pro vytvoření relační databáze dle navrženého modelu.

2.2.4. Realizace ETL procedur

V rámci řešení praktické části budou procedury ETL (extract, transform, load) zajišťovat načtení dat ze zdrojových informačních systémů do centrálního úložiště dat a dále jejich případnou transformaci. Pojem transformace může zahrnovat různé dílčí procesy a kroky, které závisí na požadavcích transformace. Může se jednat o vyčištění dat, slučování, třídění, ověřování dat, zajištění referenční integrity či agregování. Dále bude stanoven vhodný interval a časové okno pro spouštění ETL procedur, které bude záviset na objemu zpracovávaných dat a času potřebném na jejich zpracování. V rámci správné specifikace ETL procedur bude vytvořen dokument označovaný jako dokument mapování zdroje na cíl, kde budou znázorněny datové toky ze zdrojových systémů, přes centrální datové úložiště až po konkrétní dashboardy. Procedury ETL budou realizovány

v programovacím jazyce, který bude zvolen jako nejvhodnější pro načítání dat ze zdrojových systémů.

2.2.5. Realizace dashboardů ve vybraném nástroji

Po implementaci centrálního datového úložiště a jeho naplněním daty pomocí ETL procedur bude přistoupeno k tvorbě samotných dashboardů. Dle stanovených požadavků budou nejdříve vytvořeny tzv. drátěné modely dashboardů (wireframe), které budou představovat grafické znázornění hlavních prvků dashboardu. Jejich přínosem je přehlednost a možnost včasného a snazšího provedení úprav než v případě úprav již vytvořeného řešení. Po vytvoření drátěných modelů a jejich vzájemném odsouhlasení vedením podniku budou dashboardy realizovány ve vybraném nástroji pro tvorbu vizualizací.

2.2.6. Ověření splnění požadavků

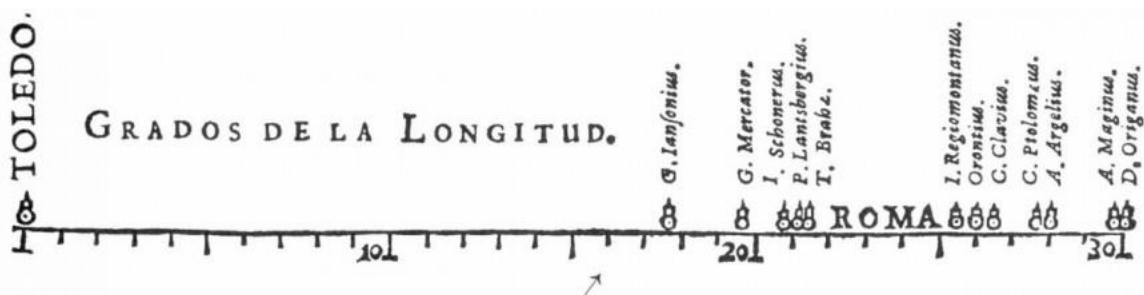
Zda realizované řešení odpovídá požadavkům specifikovaným v katalogu požadavků, a tedy i očekávání uživatelů, bude ověřeno v rámci uživatelského akceptačního testování, které bude provedeno uživateli řešení, tedy managementem podniku. Pro účastníky akceptačního testování proběhne před testováním základní zaškolení. Bude se jednat o demonstrační test, který hodnotí připravenost systému k použití. Případné nedostatky či závady nalezené během akceptačního testování budou zdokumentovány a následně bude navrženo řešení těchto nalezených problémů. Po ukončení akceptačního testování bude řešení připraveno pro zahájení rutinního provozu.

3. Teoretická východiska

3.1. Historie vizualizace dat

Před 17. stoletím existovala vizualizace dat především v oblasti map, které zobrazovaly města, silnice, přírodní zdroje a podobně. S rostoucími požadavky na přesnější mapování a měření rostla i potřeba lepší vizualizace.

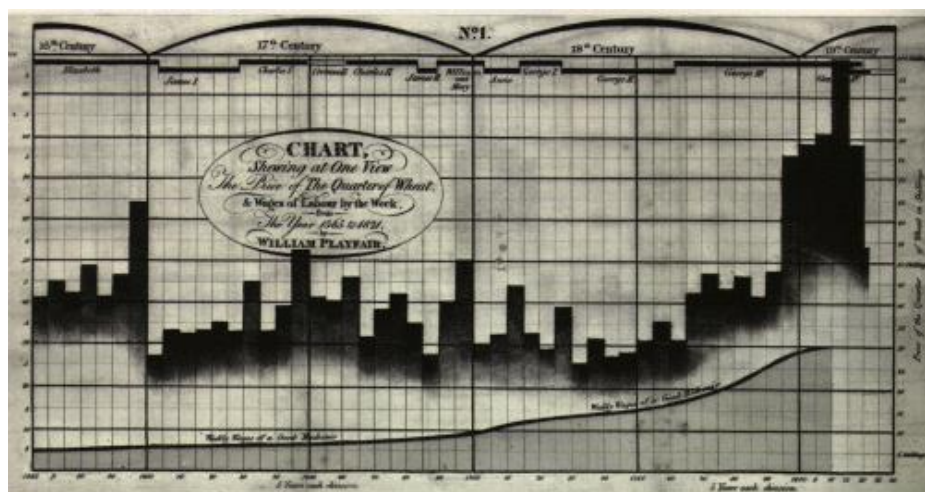
První vizualizace statistických údajů byla publikována v roce 1644 nizozemským astronomem a kartografem Michaelem Florentem van Langrenem. Součástí pamfletu s názvem „La Verdadera Longitud por Mar y Tierra“ byl jednorozměrný čárový graf znázorňující v té době dvanáct známých odhadů vzdálenosti mezi Toledem a Římem a také jméno každého astronoma, který odhad vytvořil. (DUNDAS BI, 2019)



Obrázek 2 - První známý statistický graf

(Friendly, 2003)

Během 18. století probíhaly pokusy o vytváření tematických map, které zobrazovaly geologické, ekonomické a lékařské údaje. V tomto období působí také skotský inženýr, a ekonom William Playfair, který je považován za vynálezce dnes nejpoužívanějších grafů (spojnicový, sloupcový, koláčový). Ve své knize s názvem „*The Commercial and Political Atlas; Representing, by Means of Stained Copper-Plate Charts, the Exports, Imports, and General Trade of England, at a Single View*“ představil spojnicový graf a grafy časových řad, které jsou v knize uvedeny ve 43 variantách. Tato kniha byla vydána v roce 1785 v Londýně. Následující graf Williama Playfaira znázorňuje vývoj ceny pšenice, týdenních mezd a vládnoucího panovníka v rozmezí od roku 1565 do roku 1820. (Dundas BI, 2019)



Obrázek 3 - Graf Williama Playfaira

(Dundas BI, 2019)

V rámci druhé poloviny 19. století lze zmínit hned několik slavných příkladů vizualizace dat. Prvním je mapa Johna Snowa znázorňující vypuknutí epidemie cholery v Londýně v roce 1854. John Snow zmapoval případy úmrtí na cholera a každý případ znázornil na mapě jako pruh, jak je vidět na obrázku níže. Nejednalo se pouze o mapu, ale byla to jedna část podrobné statistické analýzy, která dopomohla k objasnění vzniku epidemie a k pochopení, jak se toto infekční onemocnění přenáší. (Rogers, 2013)



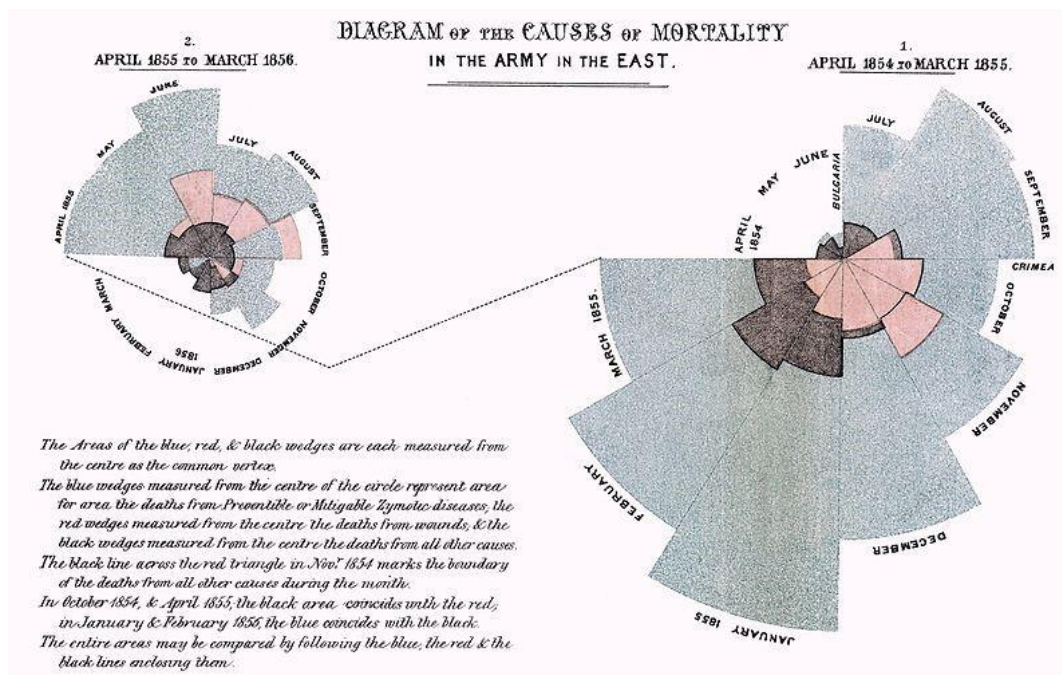
Obrázek 4 - Mapa Johna Snowa o vypuknutí epidemie cholery

(Rogers, 2013)

Druhým příkladem je graf Charlese Minarda z roku 1869 znázorňující počet mužů v Napoleonově armádě během jeho tažení do Ruského impéria. Na začátku června 1812 Napoleonova armáda na hranici s Ruskem čítala na 422 000 mužů z celé Evropy. Půl roku

po vpádu do Ruska byl Napoleon donucen se vrátit od Moskvy zpět k řece Němen a jeho armáda tehdy činila pouhých 10 000 vojáků. Graf Charlese Minarda znázorňuje postup Napoleonovi armády v čase, klesající počet mužů a vývoj teploty ve stupních Celsia. Trasa na východ k Moskvě je znázorněna béžovou čarou, zatímco návrat armády je vyznačen černě. Šířka čáry znázorňuje počet mužů v Napoleonově armádě. (Brinch, 2019)





Obrázek 6 - Graf Florence Nightingale "Coxcomb"

(Rogers, 2010)

Toto období bývá označováno jako zlatý věk statistické grafiky, k čemuž přispěla průmyslová revoluce, která umožnila vznik moderního podnikání, vznikaly státní statistické úřady. Důvodem byl rostoucí význam číselných informací v oblasti průmyslu, sociálního plánování nebo dopravy.

Rostoucí obliba statistických vizualizací zaznamenala na počátku 20. století malou překážku. Tato éra je označována jako doba temna pro vizualizaci dat. Statistickí se stále více zabývali přesnými čísly a považovali vizualizace za příliš nepřesné. Přestože se vývoj v oblasti statistické analýzy mohl odklonit od vizualizací dat, v tomto období dochází k popularizaci a aplikaci statistické grafiky v mnoha vědních oborech. Grafy různých druhů se začaly rychle používat v učebnicích, obchodní sféře a podobně. (Dundas BI, 2019)

Druhá polovina 20. století bývá označována jako znovuzrození vizualizace dat, které bylo důsledkem rozvoje počítačové techniky a počítačového zpracování dat. Počítače umožnily statistikům shromažďovat a zpracovávat data ve stále větších objemech a také rychle a snadno vizualizovat informace. V šedesátých a sedmdesátých letech 20. století působili vědci John W. Tukey a Jacques Bertin, kteří vytvořili základy vědy týkající se

vizualizace informací v oblasti statistiky a kartografie. Na počátku devadesátých let byla profesorem Edwardem Tuftem publikována významná práce nesoucí název „The Visual Display of Quantitative Information“, která se dodnes používá na univerzitách v kurzech věnujících se vizualizaci dat a statistické analýze. (Sridharan, 2017)

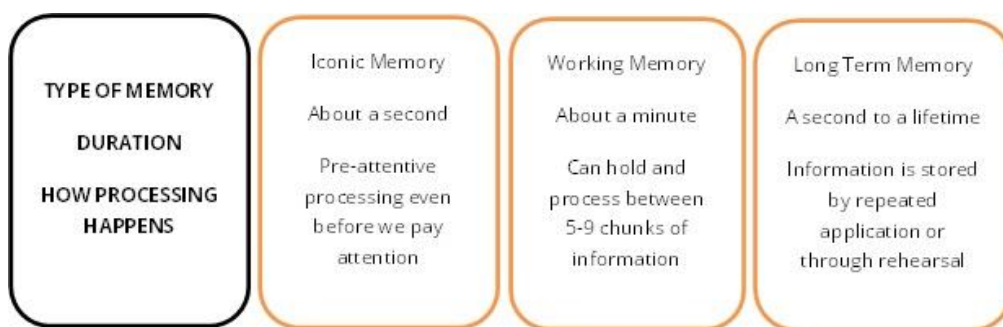
3.2. Vliv vizualizace na paměť

Říká se, že jeden obrázek vydá za tisíc slov. Proč je pro člověka snazší porozumět složitým informacím ve formě vizualizace než informacím obsažených například v tabulkách? Hlavní účel vizualizace dat je pomoci při správném rozhodování prostřednictvím porozumění trendů, vzorců a vztahů obsažených ve vizualizaci. Vizualizace dat je v mnoha ohledech pomůckou pro podporu lidské paměti. Vizualizaci ale člověk nevnímá očima, ale především svým mozkiem. (Kesavan, 2016)

- Zrakové vnímání je selektivní. Pokud by lidské vědomí vnímalo vše, byl by člověk velmi rychle přetížený. Proto zrakové vnímání selektivně vnímá především věci, které upoutají pozornost.
- Oči jsou přitahovány známými vzorci. Člověk často vidí to, co očekává, že uvidí. Autor vizualizace by tedy měl pracovat také s tím co lidé vědí a co očekávají.
- Lidská paměť je omezená, proto při pohledu na vizualizaci je paměť schopna udržet jen omezené množství informací.

Zrakové vnímání je akt vidění obrazu, což je zpracováno zrakovou oblastí v zadní části mozku. Toto zpracování je velmi rychlé a efektivní. Oproti tomu poznání je akt myšlení, zpracování informací, srovnávání a zkoumání vztahů je zpracováno v přední části mozku a je mnohem pomalejší. V rámci vizualizace dat dochází k větší míře využití zrakového vnímání a menšímu využití kognitivních schopností. Lidskou paměť lze rozdělit na tři typy, a to sensorickou paměť, krátkodobou paměť a dlouhodobou paměť. Sensorická a krátkodobá paměť jsou paměti, ve kterých je vizualizace zpracována. Když člověk vidí vizualizaci, informace zůstávají v sensorické paměti po velmi krátkou dobu, pouze zlomek sekundy. V tomto krátkém okamžiku jsou informace zpracovávány a ukládány automaticky ještě předtím, než informaci člověk věnuje pozornost. V tuto chvíli je detekováno několik vizuálních atributů. Pochopení toho, jakým způsobem lze nechat

vyniknout konkrétní atribut, může pomoci vytvořit vizualizace, které zdůrazňují ty správné důležité informace. (Kesavan, 2016)



Obrázek 7 - Lidská paměť

(Kesavan, 2016)

Krátkodobá paměť dále zpracovává informace ze sensorické paměti, které člověka zajímají. Informace v krátkodobé paměti zůstávají kolem jedné minuty a kapacita krátkodobé paměti se pohybuje mezi 5 až 9 položkami (Miller's Law, Millerovo magické číslo), přičemž kapacitu krátkodobé paměti lze zvýšit procesem nazývaným Chunking, neboli sdružováním údajů do smyslových jednotek, které následně tvoří v krátkodobé paměti jediný prvek. Vizualizace dat používají výhody tohoto procesu, protože lze více informací spojit dohromady, a tedy zpracovat mnohem více informací, než když člověk zrakově vnímá data ve formě tabulky. V tomto tkví síla vizualizace dat, tedy ve schopnosti člověka pomocí vizualizace získat a zpracovat složitější informace a více informací si také zapamatovat. (Kesavan, 2016)

3.3. Metody vizualizace dat

K práci s daty, a především k jejich prezentaci, potřebujeme tabulky a grafy, pomocí kterých lze ukázat příjemci informací, co data říkají. Je nutné správně zvolit vhodný typ tabulky a způsob zobrazení dat. Tabulky jsou vhodné, pokud je potřeba data uvést v přesném tvaru nebo je očekáváno jejich další zpracování. Tabulkou je vhodné zobrazit hrubá data a používá se pro demonstraci výsledků výpočtů a závěrů. Obrazovou informaci je možné zprostředkovat pomocí grafů a map. V následující kapitoly detailněji popisují vybrané metody vizualizace dat.

3.3.1. Tabulky

Tabulky jsou užitečné pro prezentaci a srovnání určitého množství dat, přičemž umožňují prezentovat data s přesností, které nelze dosáhnout prezentací dat pomocí grafu.

Tabulky je možné rozdělit na dva typy: (Hendl, 2014)

a) **Demonstrační tabulky**

Demonstrační tabulky ilustrují určitý názor a neobsahují zbytečné podrobnosti. Je možné dále členit na tabulky textové, statistické a prezentační.

- Textové tabulky obsahují textová data, což bývá obvykle tehdy, jestliže tabulka obsahuje kvalitativní údaje.
- Statistické tabulky prezentují popisné statistiky nebo výsledky statistické analýzy. Obsahují průměry, směrodatné odchylky, variační koeficienty, percentily atd. Specifickým případem statistické tabulky jsou tzv. kontingenční tabulky, které obsahují popis závislosti kvalitativních proměnných pomocí četností, relativních četností, sloupcových a řádkových relativních četností.
- Prezentační tabulky zobrazují numerická data, kvalitativní i kvantitativní údaje. Pomocí prezentační tabulky je možné upozornit na vztahy mezi primárními daty.

b) **Referenční tabulky**

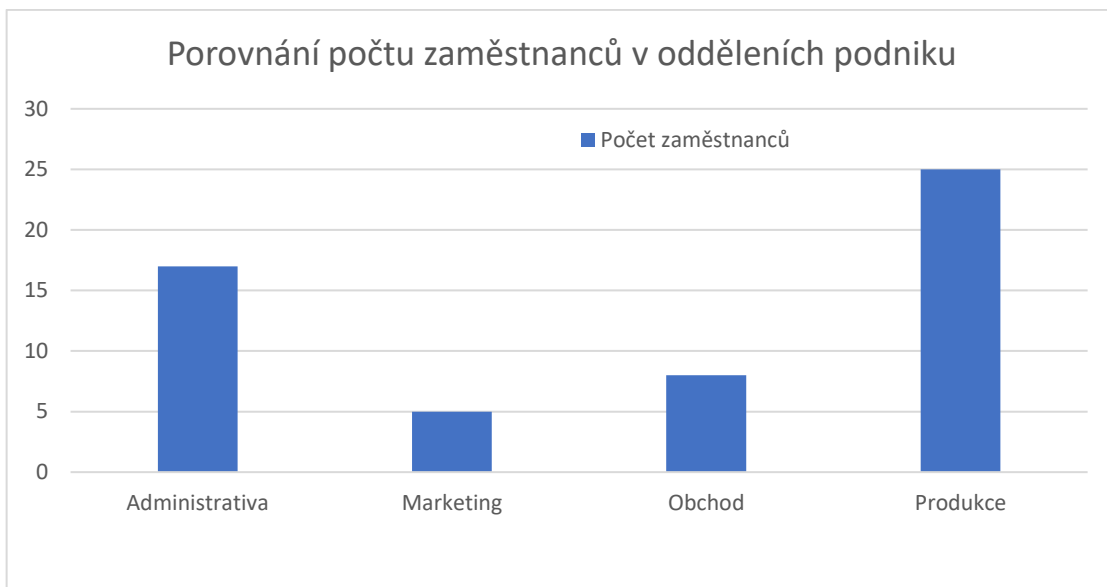
Obsahují přesná a podrobná data. Tento typ tabulky se využívají pro další potřebné výpočty a objevují se například v přílohách vědeckých či akademických prací.

3.3.2. Grafy

Grafy jsou vhodné pro efektivní prezentaci dat příjemci informací, jelikož poskytují vizuální vjem rozdělení dat a jejich trendů. Neslouží jako zdroj číselného materiálu pro další zpracování. Jako základní a nejpoužívanější statistické grafy lze uvést grafy sloupcové, koláčové, spojnicové, krabicové, bodové grafy, histogramy a polygony. Jejich cílem je především zobrazit rozdělení dat a znázornit závislost mezi daty nebo na čase.

a) **Sloupcový graf**

Znázorňuje množinu dat tak, že délka příslušného sloupce je funkcí znázorněného údaje. Délka sloupce tedy proporcionálně odpovídá velikosti hodnoty. V grafu lze vedle sebe zobrazovat několik množin dat. Sloupce mohou být zobrazeny svisle i vodorovně.



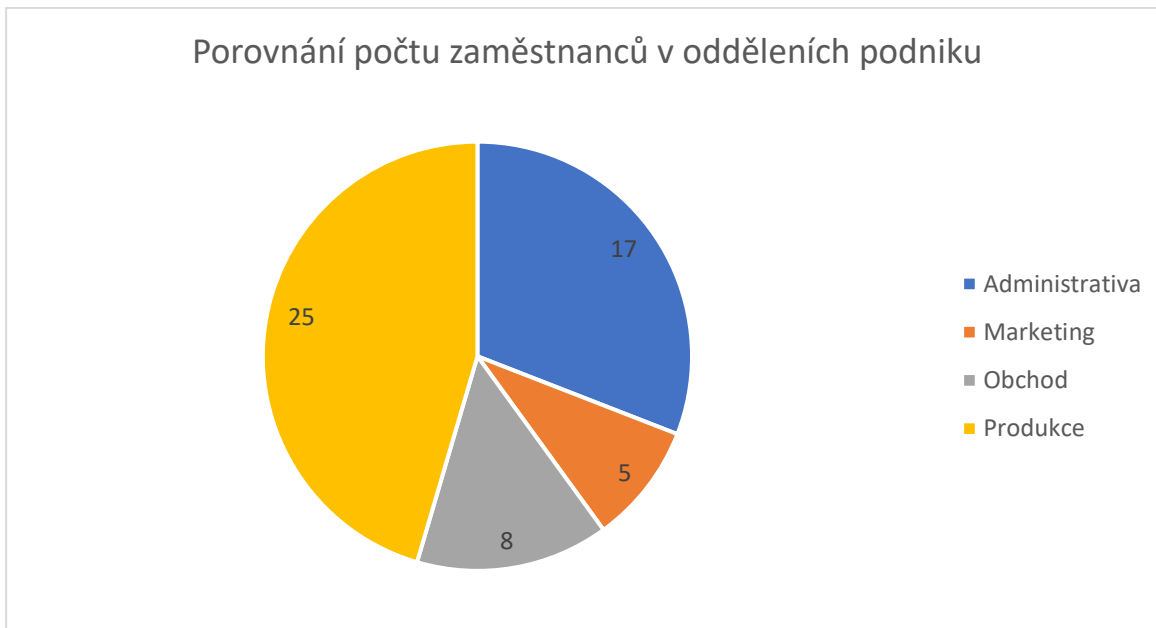
Obrázek 8 - Sloupcový graf

(Zdroj: vlastní zpracování)

Chybou při použití sloupcového grafu může být absence měřítka. Pokud graf nezačíná v nule, je nutné na to příjemce informací upozornit, jelikož takový graf by danou skutečnost zkreslil.

b) Koláčový graf

Znázorňuje hodnoty množiny údajů, které odpovídají velikosti části z celku, pomocí kruhových výsečí. Všechny výseče dohromady tvoří celý kruh. Jednotlivé výseče jsou obvykle znázorněny odlišnou barvou. Používanou modifikací koláčového grafu je prstencový graf. (Hendl, 2014)

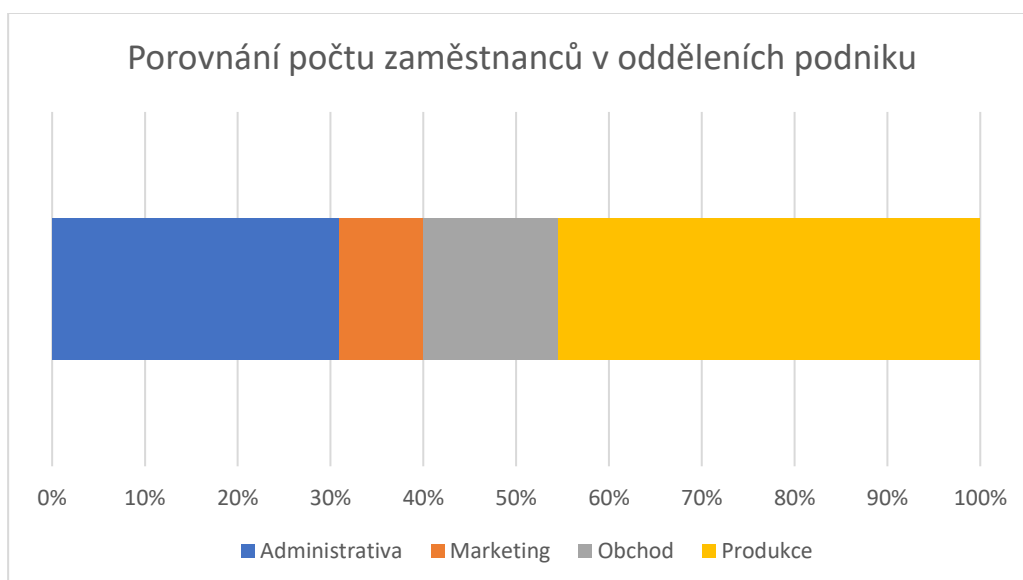


Obrázek 9 - Koláčový graf

(Zdroj: vlastní zpracování)

c) 100% skládaný sloupcový graf

V případě 100% skládaného sloupcového grafu se také jedná o modifikaci koláčového grafu. Tento typ grafu je možné použít v případě, kdy je cílem zobrazit vztah části k celku, přičemž celkový součet není důležitý. (Hendl, 2014)

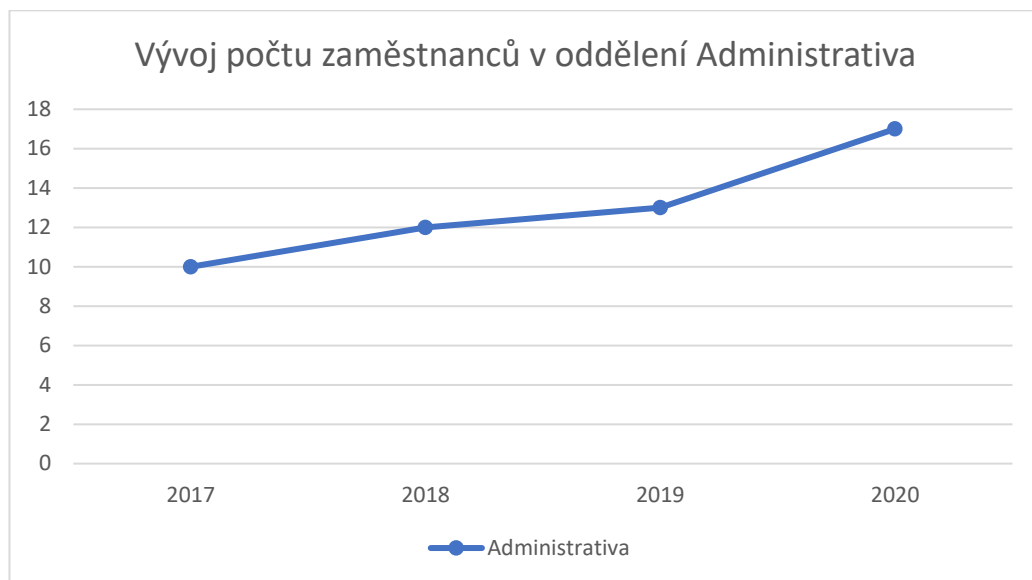


Obrázek 10 - 100% skládaný sloupcový graf

(Zdroj: vlastní zpracování)

d) Spojnicový graf

Spojnicový graf je vhodné použít pro znázornění trendu nebo vývoje proměnné v závislosti na čase.

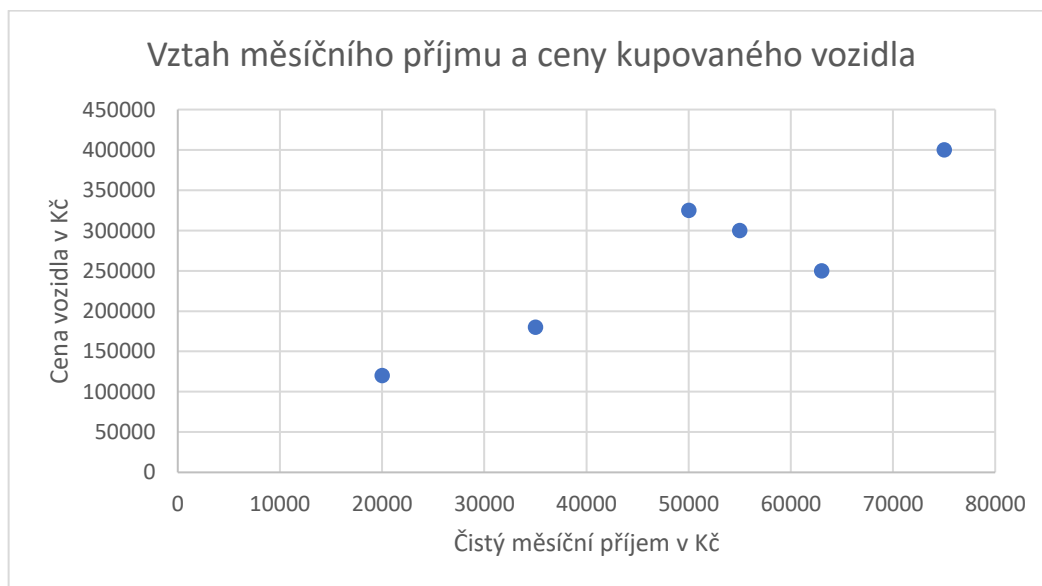


Obrázek 11 - Spojnicový graf

(Zdroj: vlastní zpracování)

e) Bodový graf

Spojnicové a bodové grafy vypadají podobně, nicméně rozdíl je v tom, že na obou osách bodového grafu jsou znázorněny spojitě proměnné a průsečíky v bodovém grafu jsou kombinací daných hodnot do jednoho bodu. Bodový graf zobrazuje na dvou osách jednu sadu dat.

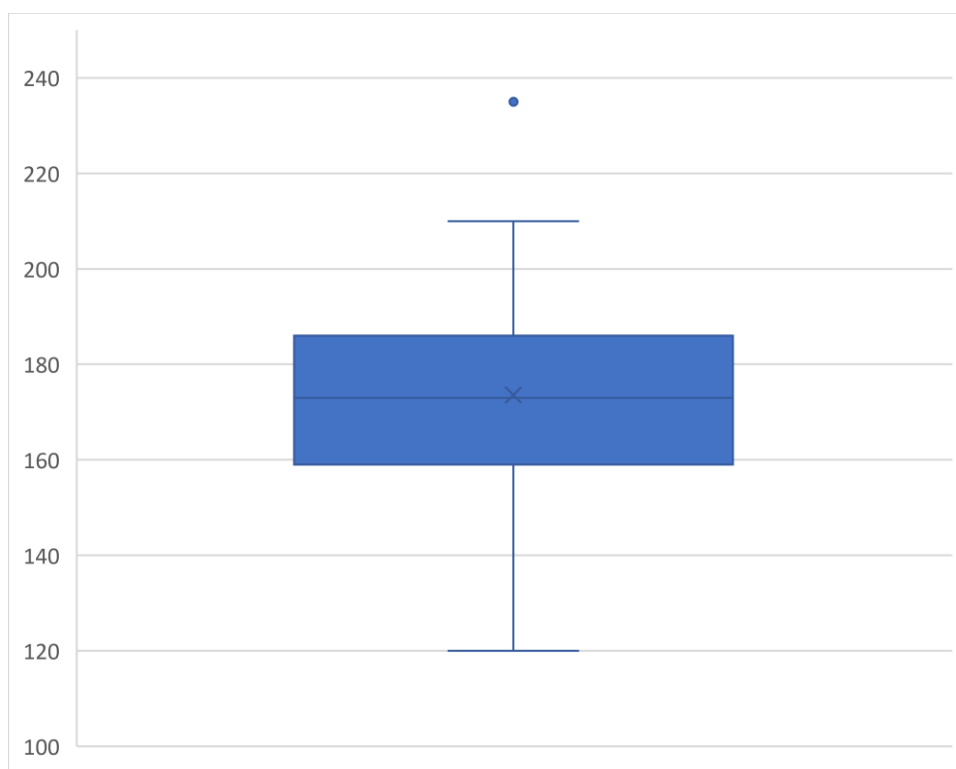


Obrázek 12 - Bodový graf

(Zdroj: vlastní zpracování)

f) Krabicový graf

Krabicový graf slouží pro popis rozložení souboru dat a je vhodný pro porovnání několika souborů dat. Lze pomocí něho identifikovat vychýlené hodnoty. Krabice je ohraničena dolním a horním kvartilem dat, což znamená, že krabice graficky znázorňuje 50% datového souboru. Pokud se jedná o nesymetrické rozložení dat, medián neleží v prostředku krabice. Horní a spodní linka nad a pod krabicí znázorňuje maximum a minimum v datovém souboru, které nejsou vychýlené hodnoty, ty jsou identifikovány pomocí teček. (Hendl, 2014)

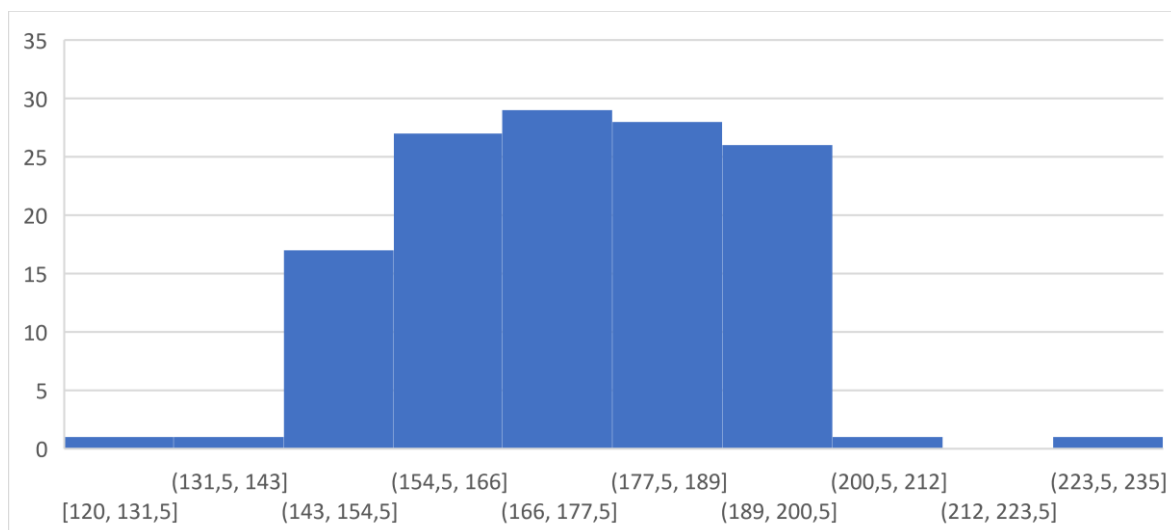


Obrázek 13- Krabicový diagram

(Zdroj: vlastní zpracování)

g) Histogram a polygon

Histogram je graf podobný sloupcovému grafu, protože oba tyto typy grafů zobrazují informace pomocí sloupců. Hlavní rozdíl je v tom, že osa X histogramu je rozdělena na intervaly a na ose Y jsou znázorněny četnosti v jednotlivých intervalech. Délka sloupce stejně jako u sloupcového grafu vyjadřuje četnost v daném intervalu. Sloupce na sebe přiléhají a tím je zdůrazněno, že jde o znázornění kvantitativních, a ne kategoriálních proměnných.



Obrázek 14 – Histogram

(Zdroj: vlastní zpracování)

Polygon je spojnicový graf, ale stejně jako u histogramu tvoří jeho souřadnice hodnota statistického znaku a její četnost v daném datovém souboru.

Histogramy a polygony četností se používají pro znázornění rozložení nebo distribuci numerických údajů. Tvar grafu určuje charakter rozložení dat. Pokud je histogram symetrický, jedná se o symetrické rozdělení. Pokud histogram nebo polygon vykazují zešikmení, jedná se o zešikmené rozložení. Histogram a polygon mohou mít i více vrcholů (maxim), pak se jedná o vícemodální rozložení. Zkoumáním těchto grafů lze také identifikovat vychýlené hodnoty. (Hendl, 2014)

3.3.3. Mapy

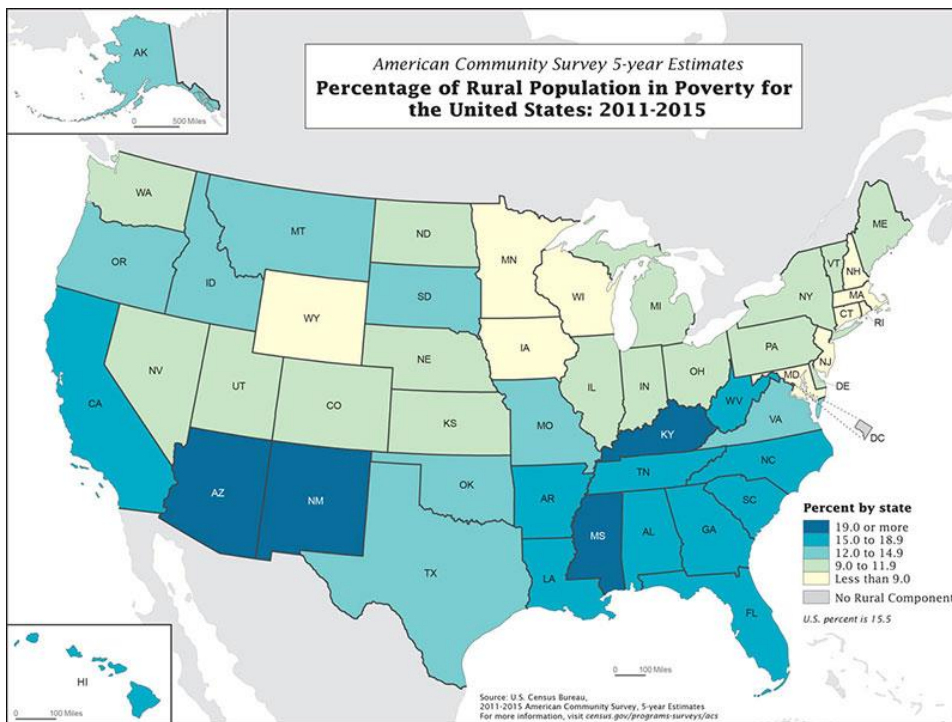
Mapy jsou výborným prostředkem pro prezentaci statistické informace v geografických souvislostech. Jsou vizuálně přitažlivé, zjednodušují rozpoznávání vztahu dat k danému místu a pomáhají uživatelům identifikovat geografické trendy v datech lépe, než se to může podařit pomocí běžného statistického grafu nebo tabulky. Známým příkladem datových map jsou geografická zobrazení výskytu nemocí. Jiným příkladem takového zobrazení dat jsou klimatologické mapy nebo mapy o demografické situaci v různých oblastech. (Hendl, 2014)

Mapy lze rozdělit na tři typy:

a) Barevné mapy a mapy s různou výplní

Jsou vhodné pro znázornění intenzity určité veličiny v dané oblasti pomocí různých barech či odstínů. Uživatel pak jednoduchým způsobem identifikuje vysokou, střední či

nízkou intenzitu. Příkladem může být například mapa míry nezaměstnanosti v různých okresech daného státu.

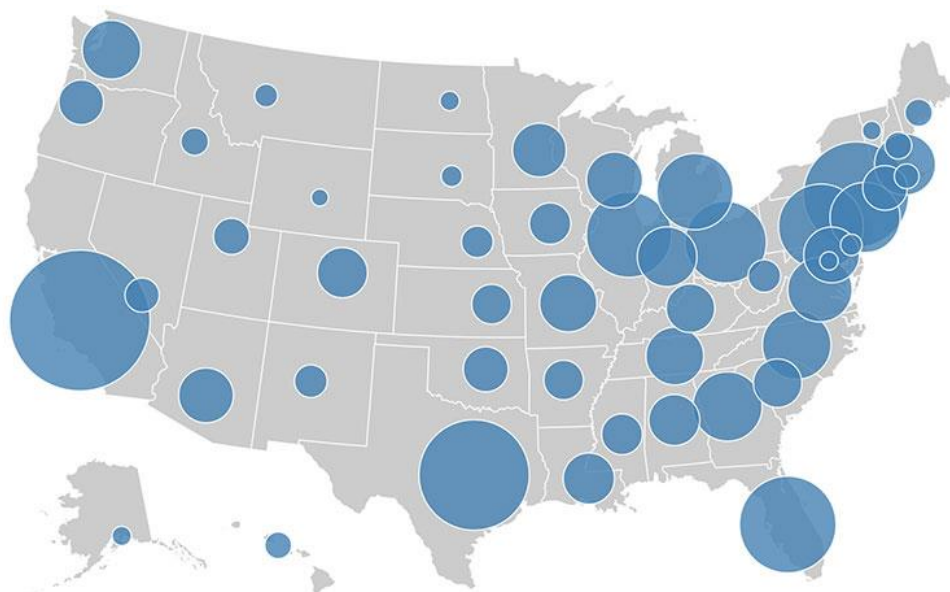


Obrázek 15 - Barevná mapa

(Stepanov, 2021)

b) Proporční symbolické mapy

Proporční mapa používá k reprezentaci hodnoty v konkrétní oblasti symbol, obvykle kruh, o určité velikosti. Čím větší symbol je, tím větší je hodnota zobrazované proměnné v dané oblasti.

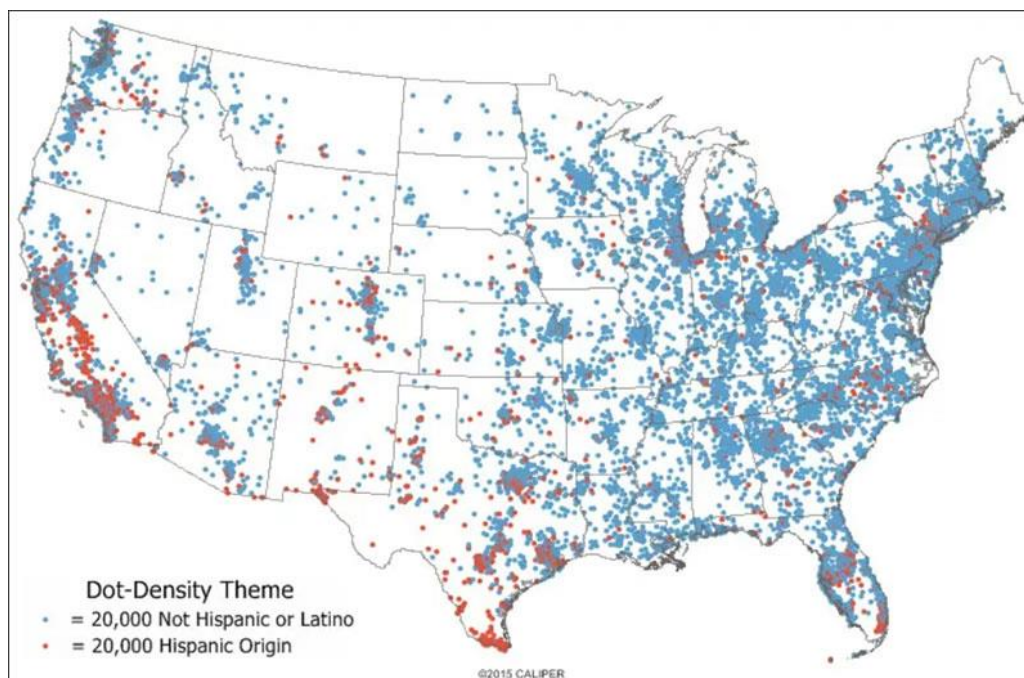


Obrázek 16 - Proporční symbolická mapa

(Stepanov, 2021)

c) Bodové mapy

Jednotlivé výskyty událostí se vyznačují bodem v dané oblasti. Bodové mapy jsou efektivním způsobem, jak zobrazit geografické rozložení dané proměnné. Mohou zobrazovat například hustotu obyvatelstva, národnostní složení obyvatelstva v dané oblasti a podobně.

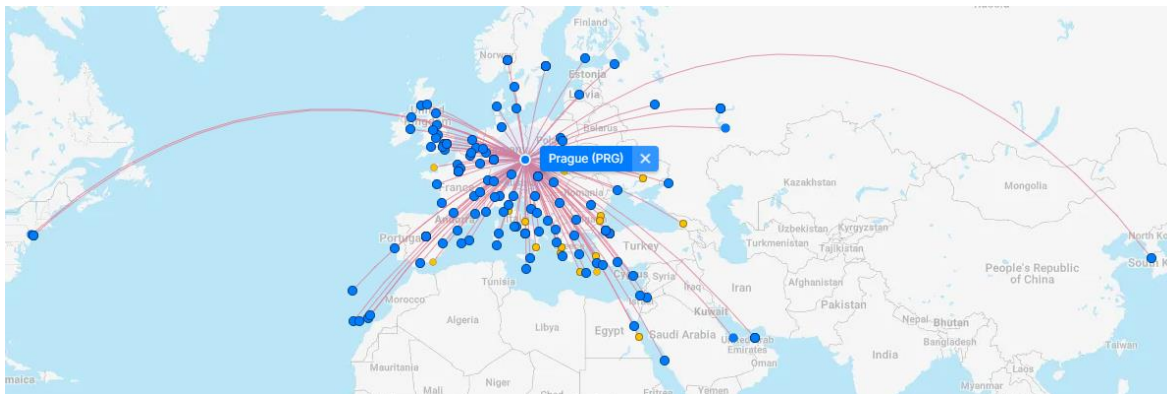


Obrázek 17 - Bodová mapa

(Stepanov, 2021)

d) Spojnicová mapa

Mapa, která zobrazuje také jednotlivé body, ale navíc mezi nimi ukazuje spojnice. Takovéto výstupy jsou často využívány leteckými společnostmi pro znázornění jejich letů, nebo cestovními kanceláři, které svým klientům ukazují, kam všude s nimi mohou vyjet na dovolenou.



Obrázek 18 - Spojnicová mapa

(Zdroj: <https://www.flightconnections.com/flights-from-prague-prg>)

Obecné zásady pro tvorbu map (Hendl, 2014):

- Jasně formulovaný název s případným podnázvem.
- Pojmenování geografické oblasti.
- Indikace okamžiku nebo časového intervalu zachycených dat.
- Klíč vysvětlující, co myslíme jednotlivými barvami nebo typem výplně oblasti, definice intervalů se nesmí překrývat.
- Indikovat zdroj dat.
- Textové shrnutí v dokumentu.
- Dostatečná velikost pro zobrazení všeho, co je potřeba.
- V případě potřeby použijeme zvláštní mapu s přesným označením míst a hranic.

3.3.4. Proces tvorby vizualizace

Proces tvorby vizualizace lze rozdělit na několik následujících bodů.

a) Analýza publika

Před samotnou tvorbou grafu je potřeba přemýšlet o tom, kdo je publikum, kterému bude graf prezentován. Jedná se o jeden z nejdůležitějších aspektů v procesu tvorby grafu,

který ale bývá opomíjen. Dále je třeba myslet například na to, kolik času bude k dispozici pro prezentaci grafu. Jaké rozhodování bude na základě grafu publikum dělat, jaká přesnost dat je pro rozhodování publika nezbytná a podobně.

Pokud lze definovat primárního diváka s určitou osobností, který je v publiku procentuálně hojně zastoupen, jedná se o homogenní publikum. Tedy takové publikum, které je sjednoceno dle určitého parametru. Zacílit na takové publikum je jednodušší. Homogenním publikem může být například tým zaměstnanců v jedné firmě, třída dětí na základní škole, účastníci jednoho kurzu atd. Tyto skupiny mají sdílené vlastnosti, kterých je možné využít během přípravy grafu.

Nestejnorodé skupiny uživatelů přinášejí plno problémů. Nelze se zafixovat pouze na jednu z nich. Je potřeba se ptát, čím oslovíme každou z nich, a hlavně čím neznevýhodníme ani jednu z nich. Problematické bývá například různá úroveň vzdělání, různé zájmy, vzhled do dané problematiky nebo různé názory. Je nutné tedy počítat s tím, že uživatel, který přichází k naší vizualizaci, nemusí být znalý problematiky a nemusí ji tedy pochopit vhodným způsobem bez správného okomentování. Zároveň ale není dobré znalého uživatele zahlcovat banalitami. Je potřeba najít rovnováhu.

b) Určíme, co chceme sdělit

Dalším důležitým krokem je stanovení, co chceme grafem říci. Jaký účel má mít vytvořený graf. Pokud známe důvod, proč graf tvoříme, je snadnější zvolit i správný typ grafu. Je třeba tedy vědět, čeho má být vizualizací dosaženo.

Prosté předání faktů je jeden z nejběžnějších účelů vytvořené vizualizace. Cílem je tedy předat realitu tak, jak je, a dále do ní nevnášet další manipulační nebo emoční artefakty. Tento přístup je často ražen v odborné sféře. Akademické subjekty nebo společnosti zabývající se výzkumem prezentují své výstupy na odborných fórech, na každoročních konferencích či v periodicky vydávaných časopisech. Důležité je v případě předávání faktických informací uvést zdroj datových podkladů. Díky tomu se zvyšuje transparentnost uvedených údajů a každý si může dohledat, zda je interpretace správná.

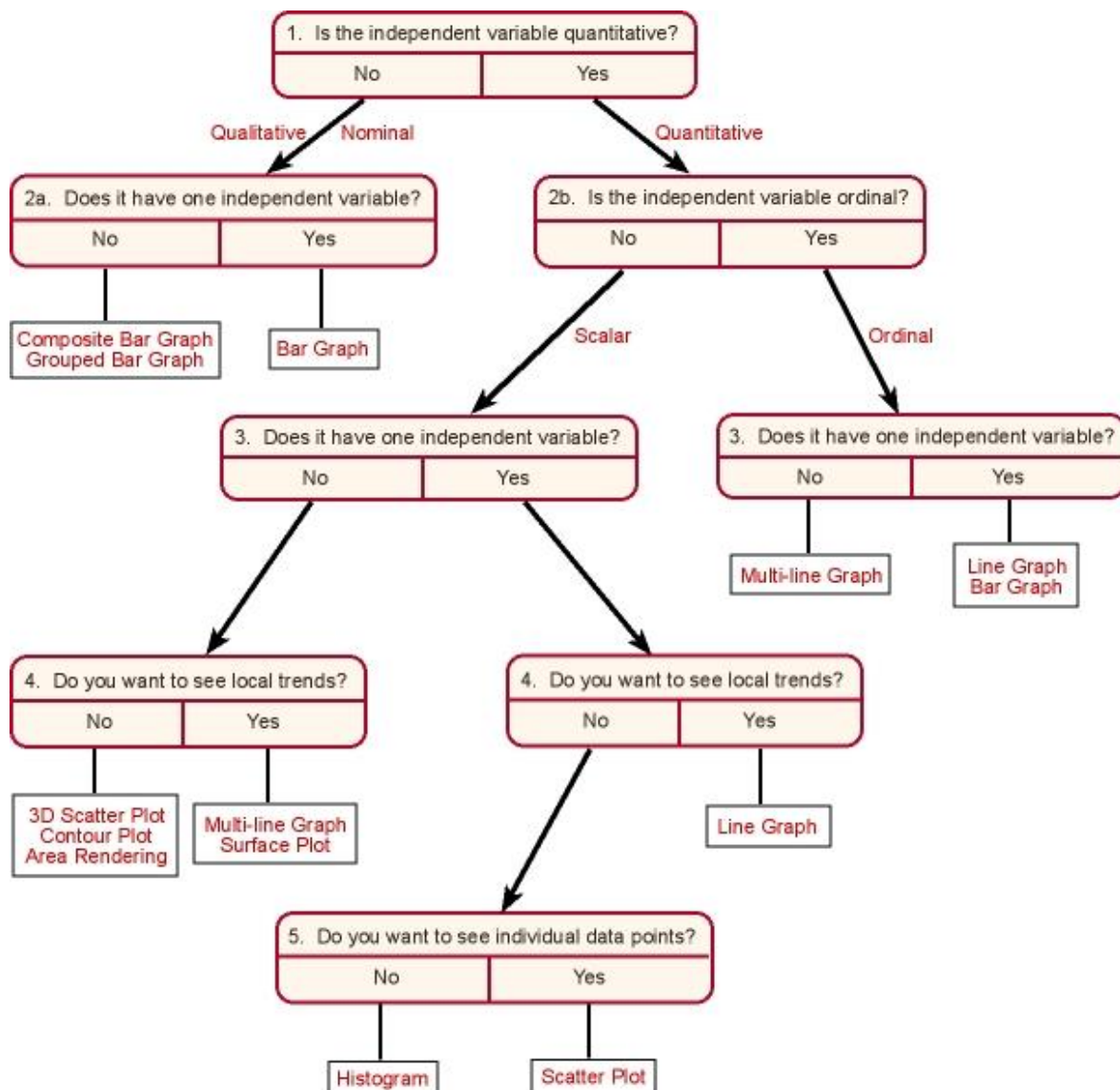
Cílem vizualizace může být také útok na city publika. Často se jedná například o vizualizace z oblasti ochrany životního prostředí. Autor vizualizace se může snažit o zdůraznění problémů, které nastávají, nebo se snaží upozornit na ty, které by mohly v dohledné době přijít. Dalším příkladem může být graf znázorňující volební preference

v následujících volbách. Takový graf může do určité míry ovlivnit rozhodnutí skupiny voličů.

Velmi častým cílem vizualizace je i podpora strategických firemních rozhodnutí. Může se jednat o volbu směru, kterým se bude vyvíjet obchodní strategie nebo například zhodnocení stavu projektu. Při práci na různých produktech je také vhodné si v pravidelných cyklech zhodnocovat jejich stav, kvalitu a případný prodej. Se všemi těmito rozhodnutími by správně měla pomáhat podkladová data, která má firma k dispozici. Pokud se rozhoduje na jejich základě, hovoříme o tzv. datově řízeném rozhodování. Aby bylo možné efektivně data analyzovat, je nutné je nejdříve předzpracovat a následně i vizualizovat. Tím se odhalí v datech skryté informace související s různými aspekty firmy/projektu/produktu. Je poté jednodušší prezentované jevy prezentovat a diskutovat v širším kolektivu.

c) Vybereme typ grafu

Pokud pro prezentaci informací zvolíme nesprávný typ grafu, může to mít neblahý dopad na zprostředkování informací cílovému publiku. Pro správné rozhodnutí je třeba se seznámit s jednotlivými typy grafů, znát jejich specifika a vhodné použití. Pro správnou volbu grafu nám může pomoci následující rozhodovací strom:



Obrázek 19 - Rozhodovací strom pro volbu typu grafu

(NC State University, 2005)

Kromě tohoto rozhodovacího stromu mohou s výběrem vhodného grafu pomoci další interaktivní nástroje, například na adrese <https://depictdatastudio.com/charts/>, nebo nástroj From Data to Viz na adrese <https://www.data-to-viz.com/>.

d) Příprava dat

Dalším krokem je příprava a uspořádání dat, které jsou k dispozici. Často nejsou data ve vhodné podobě a je potřeba je uspořádat, přepočítat nebo transformovat. K těmto účelům je možné použít tabulkové procesory a jiné statistické nástroje, které poskytují mnoho prostředků. Teprve poté, co jsou data připravená ve správné formě, je možné přistoupit k tvorbě vizualizace.

e) **Tvorba vizualizace**

V rámci tvorby vizualizace je nutné zvolit správný nástroj. Nedostatek znalostí práce v určitém nástroji nebo limitace zvoleného nástroje jsou hlavní důvody, proč se v mnoha případech nepodaří vytvořit přesně to, co bylo navrženo. Popisu nástrojů používaných pro vizualizaci dat se detailněji věnuje kapitola 3.9.

Mezi obecné zásady pro tvorbu grafů lze zahrnout, že každý graf musí mít jasně popsán název a účel grafu, popsané jednotlivé osy, třídy, sloupky, škály každé osy s vyznačením začátku. Dále je třeba uvést zdroj dat a v grafu by se nemělo vyskytovat nadbytečné množství čar a jiných grafických prvků.

f) **Ověření vizualizace**

Po samotné tvorbě vizualizace je vhodné ověřit, zda plní zamýšlený účel. Je možné k tomu použít tzv. „The Grandmother Test“, který spočívá v ověření vizualizace skrze jiného spolupracovníka, který se na návrhu či tvorbě vizualizace nepodílel. Pokud takový člověk váhá nad účelem vizualizace nebo se jeho interpretace neshoduje se zamýšleným účelem vizualizace, je potřeba vizualizaci upravit. Další možností ověření vizualizace je například použití „The Data Visualization Checklist“ vytvořený Stephanie Evergreen a Ann K. Emery v únoru 2018, vložený v příloze A.

g) **Sdílení vizualizace**

Posledním krokem v procesu návrhu a tvorby vizualizace je přizpůsobení vizualizace vybranému médiu, kde bude vizualizace publikována nebo sdílena. Nosič vizualizace z velké míry určuje, co vše si autor vizualizace může dovolit a jakým způsobem by se měl k tvorbě postavit. Typy médií se historicky mění, stejně jako se mění jejich procentuální zastoupení napříč společnostmi. Nosičem vizualizace mohou být například tištěné noviny, marketingové letáky, televizní přenos, internet a další.

3.4. Skupiny dat

Vlastnosti datového typu, jež mají jednotlivé záznamy v datové sadě, ovlivňují volbu grafu nejvýrazněji. Skupiny záznamů v datech určují, jakým způsobem bude možné data zpracovat a jaké operace lze nad nimi provádět.

Dle (Jarkovský et al. 2017) lze data rozdělit na následující typy.

3.4.1. Kvalitativní členění

a) Binární dat

Nejjednodušší skupinou jsou data nabývající dvou možných hodnot. Často se používají jako příznaky, zda je něco přítomno nebo ne, případně, že objekt má, nebo nemá danou vlastnost. Hodnoty jsou pak často true/false, 1/0, ano/ne.

b) Nominální data

Záznamy této skupiny obsahují více skupin, nicméně nemají mezi sebou nějaké zvláštní pravidlo, jak by měly být seřazeny. Často se jedná o různé výčty hodnot, případně číselníky.

c) Ordinální data

Důležitou vlastností této skupiny je, že se záznamy této skupiny dají seřadit.

d) Intervalová data

Kromě možnosti data řadit, je možné sledovat a ohodnocovat vzdálenosti mezi jednotlivými kategoriemi dat. Dává zde i smysl hodnoty mezi sebou sčítat a odečítat.

e) Poměrová data

U dat v této skupině, kromě předchozích vlastností, je možné sledovat i poměry mezi jednotlivými záznamy.

3.4.2. Kvantitativní členění

a) Spojitá data

Data této skupiny mohou nabývat libovolných hodnot v daném intervalu. Hodnot v daném intervalu je potom nekonečně mnoho.

b) Diskrétní data

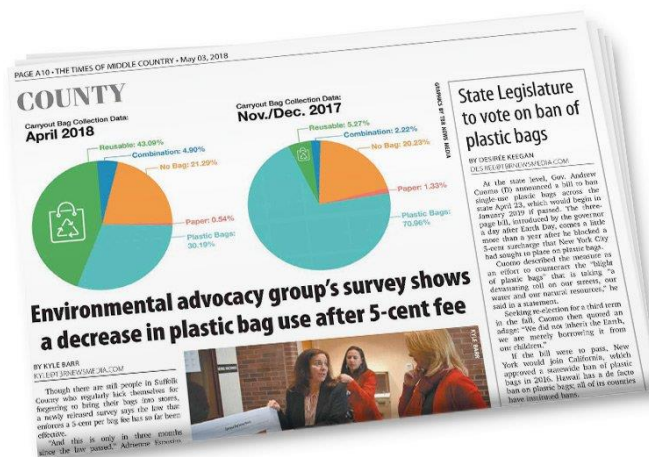
Narozdíl od spojitých souborů dat jsou diskrétní množiny tvořeny konečným počtem prvků v určitém intervalu.

3.5. Typy médií

Nosič vizualizace z velké míry určuje, co vše je možné si dovolit a jakým způsobem je potřeba se k tvorbě vizualizace postavit. Typy médií se historicky mění, stejně jako se mění jejich procentuální zastoupení napříč společnostmi. Dynamické reagování na vyvíjející se trend může znamenat konkurenční výhodu. (Engeto, 2020)

3.5.1. Tištěné noviny

Tento typ média je celosvětově hodně rozšířený a rozhodně není záhodno se jím nezabývat. Alespoň pro dokreslení celkového obrázku. Zde se totiž vizualizace dat vyskytují denně a musí dosahovat velice vysoké kvality, protože jakmile se vytisknou, nelze je již změnit. V novinách si také nelze dovolit žádnou míru interaktivity, proto je nutné mít vše už hotové tak, jak je pro čtenáře zamýšleno. Na druhou stranu není správné vizualizaci zahltit zbytečnými detaily, které ji učiní nepřehlednou. Většina výtvorů v tomto typu média se zaměřuje na faktické informace – ekonomické trendy, výsledky voleb, sportovní statistiky. Z tohoto hlediska by měly být výstupy dobře srozumitelné i pro širokou veřejnost z různých sociálních vrstev. Zároveň by vizualizace neměly být zavádějící nebo dokonce lživé. Toto ale platí nehledě na zvolené médium. Problematická může být i barevná paleta, kterou některé novinové redakce nabízejí – někdy je možno vybírat pouze z odstínů šedi. (Engeto, 2020)



Obrázek 20 – Vizualizace v tištěných novinách

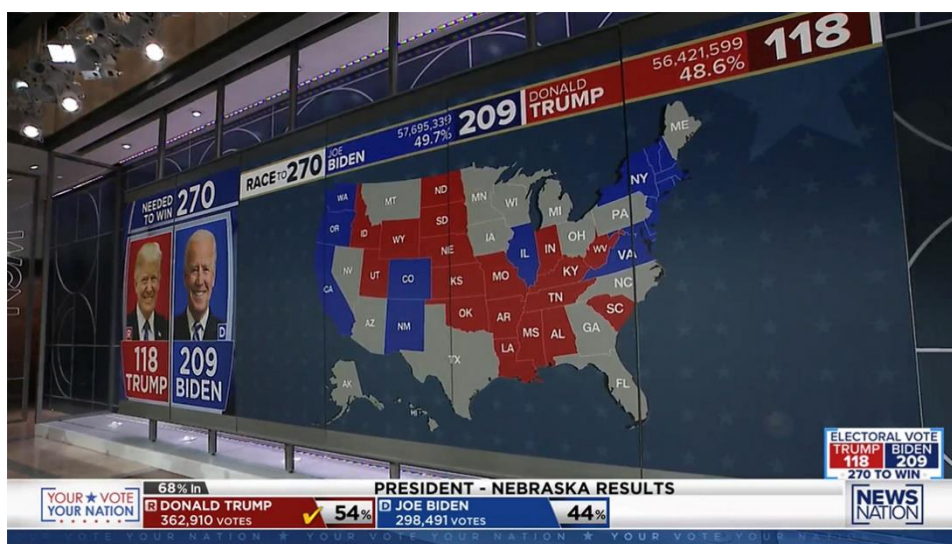
(Zdroj: <https://dvacreative.com/project/data-visualization-tbr-news-media/>)

3.5.2. Marketingové tiskoviny

Může se jednat o různé letáky, brožury, ale třeba i tiskoviny, které prezentují nějakou instituci např. reklamní materiály škol, náborové letáky a mnoho dalších. V mnohém se neliší od tištěných novinových příspěvků, ale mají pár odlišností. Jednou z největších je často potřeba předávat kromě faktické informace i emoci a budovat potřebu. Často je tak možné dosáhnout nějakým výrazným efektem ve vizualizaci – zvýrazněním, dodaným textem nebo barevnou kombinací. Tento typ nosiče je velmi často určen pro konkrétní cílovou skupinu a lze tedy mířit rovnou na ni. Je možné počítat i s větším rozpočtem na takovou datovou vizualizaci. (Engeto, 2020)

3.5.3. Televizní přenos

Během běžného vysílání není běžné, že by se datové vizualizace vůbec objevovaly. Velkou výjimkou je ovšem zpravodajství. Stále více se setkáváme s obrazovou podporou slovního a textového předávání informací. Lidé se totiž raději dívají na již hotové vizuální výstupy, než aby pouze poslouchali a museli si slova vizualizovat v hlavě. Hlavním atributem je zde velikost doprovodných informací a rychlost, ve které jsou výstupy předávány publiku. Další důležitou věcí je minimum detailů. Toto prostředí to dovoluje, jelikož je grafický výstup vždy doprovázen dostatečným slovním komentářem moderátora. (Engeto, 2020)



Obrázek 21 - Vizualizace v rámci televizního přenosu

(Zdroj: <https://www.newscaststudio.com/2020/12/17/ross-s-rocket-surgery-helps-power-election-night-graphics-including-newsnation/>)

3.5.4. Internet

V dnešní době je internet stěžejním médiem. Dokáže dokonce emulovat všechny předchozí typy médií a je tedy velmi složité zvolit tu správnou formu komunikace v konkrétní situaci. Nikdy totiž nedosáhneme všech atributů na vysoké úrovni. Na internetu je zároveň obrovská konkurence a důležitá je tedy kvalita a originalita publikovaného obsahu. Jedině tak je možné zaujmout čtenáře, kteří vytvořené vizualizaci často věnují pouze sekundy svého času. Udržet si je, a dokonce motivovat k tomu, aby se vrátili, je o to složitější. (Engeto, 2020)

3.6. Chyby vizualizace

Někdy se stává, že vizualizace neprezentují přesně to, co by měly nebo co bylo zamýšleno. Může to být zapříčiněno jak špatně zvoleným způsobem prezentace, tak zpracováním dat. K desinterpretaci dopomáhá i špatné vysvětlení, co graf ukazuje, nebo kontext, ve kterém se výstup nachází. Tyto problémy jsou spjaté i s možnou nezkušeností autora výstupu.

3.6.1. Faktor lživosti

To, jak je vizualizace lživá, lze dokonce vypočítat. Vzorec pro vypočítání představuje a vysvětluje ve své knize *The Visual Display of Quantitative Information* profesor Edward Tufte.

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

Vzorec pro výpočet faktoru lživosti grafiky

(Tufte, 2007)

Aby byla zajištěna integrita vizualizace, její faktor lživosti by měl mít hodnotu mezi 0,95 a 1,05. Pokud je hodnota menší nebo větší, znamená to podstatné (a často zamýšlené) zkreslení, daleko za drobné nepřesnosti (např. způsobené vykreslením na různých zařízeních). Například kdyby měl faktor lživosti hodnotu 0,98, pak vizualizaci nelze označit za lživou, protože odchylka vykreslení v grafice nebo jiná drobná chyba má za příčinu dané zkreslení. Na druhé straně, pokud by měl faktor hodnotu 1,6 (respektive 0,7), můžeme obecně předpokládat, že autor záměrně grafiku zkreslil. Je-li faktor lživosti větší

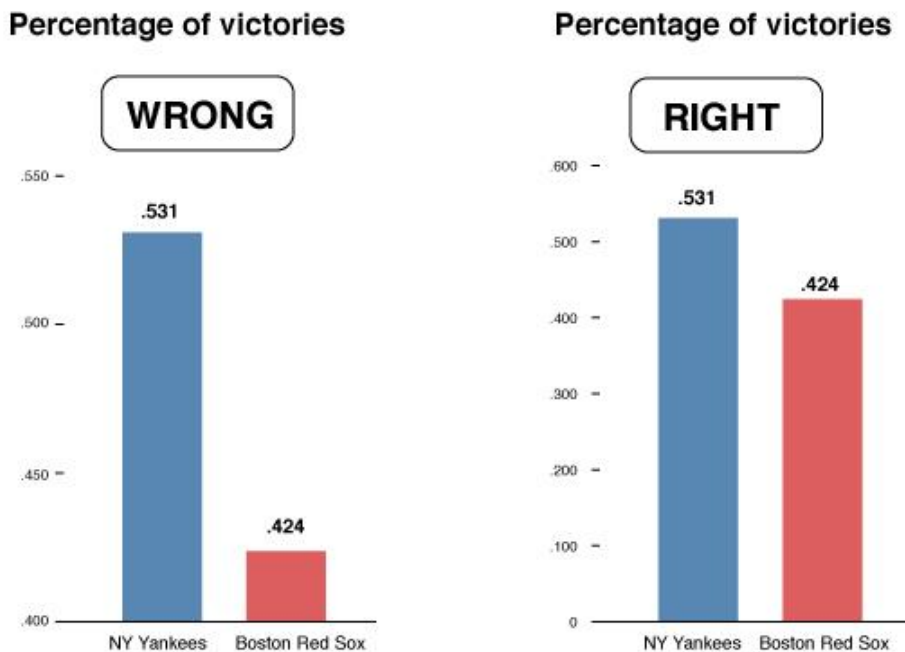
než 1, grafiku se lze označit za „přehnanou“ nebo „přemrštěnou“. Je-li menší než 1, grafika je „zmírněná“. (Tufte, 2007)

3.6.2. Nevhodné prvky

V této kapitole budou představeny čtyři problematické techniky, které zkreslují datové vizualizace.

a) Zavádějící osy

Osy by měly začínat od nuly. Někdy je snaha přiblížit publiku pohled na zajímavou část grafu, ale není to vhodným způsobem interpretace. Rozdíly mezi dvěma výsledky se pak stávají více dramatickými, než ve skutečnosti jsou. První graf má osu Y, jejíž počátek je v čísle 400. Osa Y v druhém začíná standardně od nuly. Vizuelní rozdíl v prvním případě pak na publikum působí jako mnohem větší.

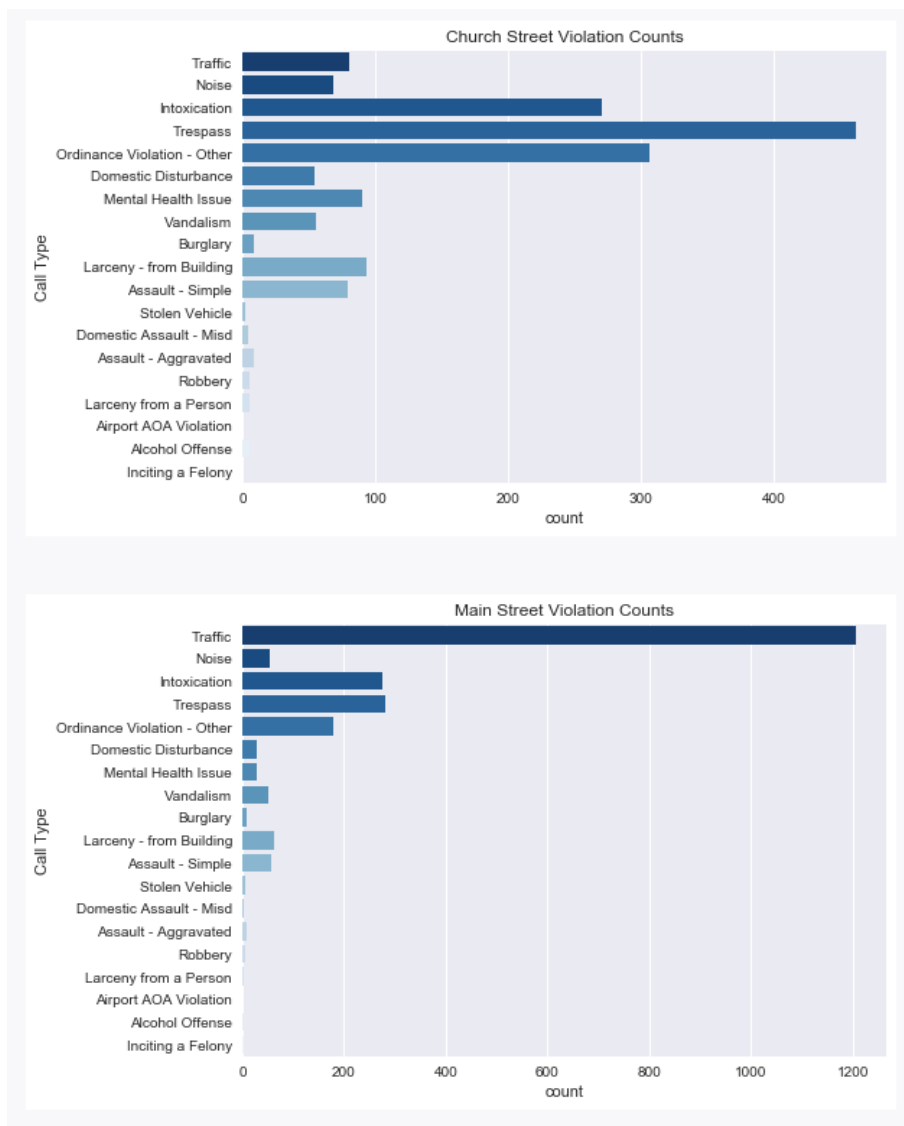


Obrázek 22 - Zavádějící osy

(Chiqui, 2015)

b) Izolované porovnávání

Pokud porovnání dvou nebo více datových vstupů je provedeno více grafy umístěnými vedle sebe, může se stát, že jejich osy nejsou shodné. A to i markantně. Publikum si toho může všimnout až po delším zkoumání vizualizace, ale první pohled na data je ovlivněn tímto úkazem. Jako korekce postačí sjednotit maxima obou (všech) grafů.



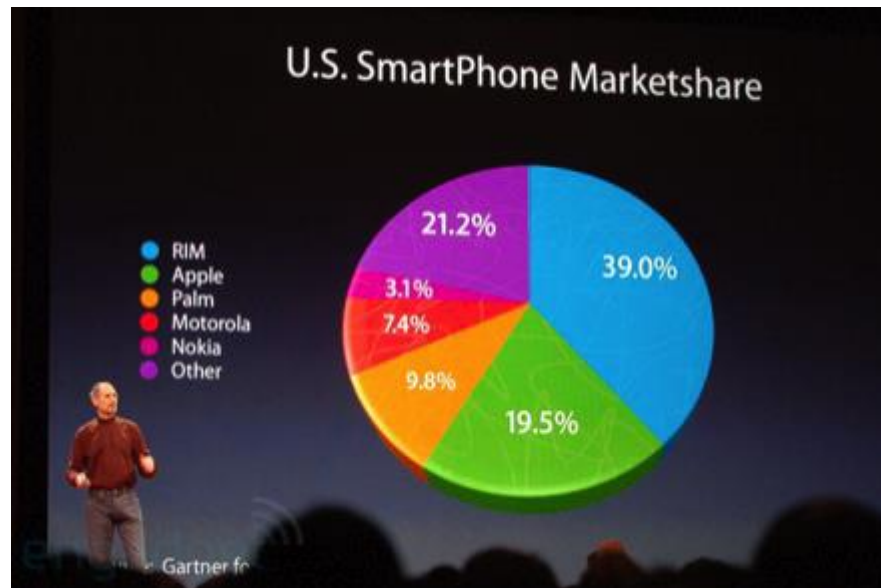
Obrázek 23 - Izolované pozorování

(Fortney, 2017)

c) 3D efekty a koláčové grafy

Koláčové grafy bývají častým problémem. Zejména kvůli tomu, že lidské oko nemá tak dobrou schopnost porovnat jednotlivé díly mezi sebou. Sloupcový graf je v tomto ohledu uživatelsky mnohem přívětivější. Koláčový graf je nevhodný v kombinaci s 3D efektem. V těchto případech jsou poměry těžko detekovatelné a plocha navíc v mnoha případech nesouhlasí s uváděným procentuálním číslem.

Například na koláčovém grafu níže by se mohlo zdát, že dílek s hodnotou 19,5% je větší než dílek s hodnotou 21,2%.

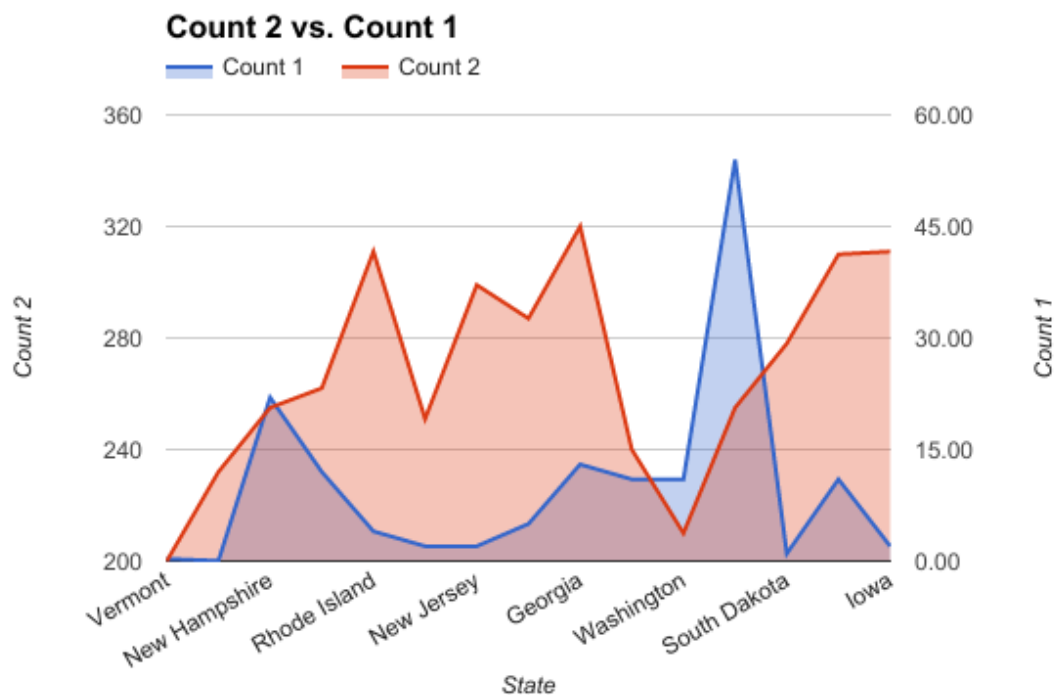


Obrázek 24 - 3D efekty a koláčové grafy

(Paragraph, 2008)

d) Grafy se dvěma osami

V případě nevhodného vysvětlení grafu se dvěma nezávislými svislými osami se může stát, že si publikum neuvědomí rozdílnost dvou měřítek a může dojít ke zmatení publika.



Obrázek 25 - Graf se dvěma osami

(Fortney, 2017)

3.7.Data Storytelling

Ve školách bývají děti vzdělávány v češtině a matematice, kde se dozvědí, jak skládat slova do vět a věty do příběhů a v hodinách matematiky pochopí význam čísel, ale zřídka kdy jsou tyto dvě oblasti spojovány. Nikdo neučí, jakým způsobem vyprávět příběh pomocí čísel, nicméně tento úkol nabývá stále více na důležitosti. Technologie nám umožňují shromažďovat stále větší množství dat a s tím souvisí rostoucí potřeba dát všem nashromážděným datům smysl. Schopnost vizualizovat data a vyprávět pomocí nich příběhy je klíčem k jejich přeměně na informace, které lze použít k lepšímu rozhodování.

Příběhy fungují, protože lidé jsou připraveni na vnímání příběhů, nikoliv na vnímání nesouvisících faktů nebo statistických dat. Ve skutečnosti, 63% posluchačů si zapamatuje prezentovaný příběh a pouze 5% posluchačů si zapamatuje statistické údaje. Je velký rozdíl v množství zapamatovaných informací po prezentaci s efektivní vizualizací dat oproti prezentaci tabulky s čísly. Paralely mezi vizualizací dat a vyprávěním příběhů jsou nepopiratelné. Stejně jako v příbězích, Data Storytelling zahrnuje tři hlavní elementy a to postavy, dějovou linku a vyprávění. V Data Storytelling jsou postavy klíčové ukazatele výkonnosti (KPI), vyprávění je vizualizace dat a dějová linka je porozumění vyprávěnému. (Sleeper, 2018)

Data Storytelling je technika, při které se autor vizualizace snaží pomocí vytvořené vizualizace vyprávět příběh, který u publika probudí zájem a díky tomu je publiku snazší předat myšlenku, kterou má vizualizace obsahovat. Zároveň by mělo dojít i k povzbuzení uživatele zacházet u datového souboru do větší hloubky a hledat zajímavé závěry.

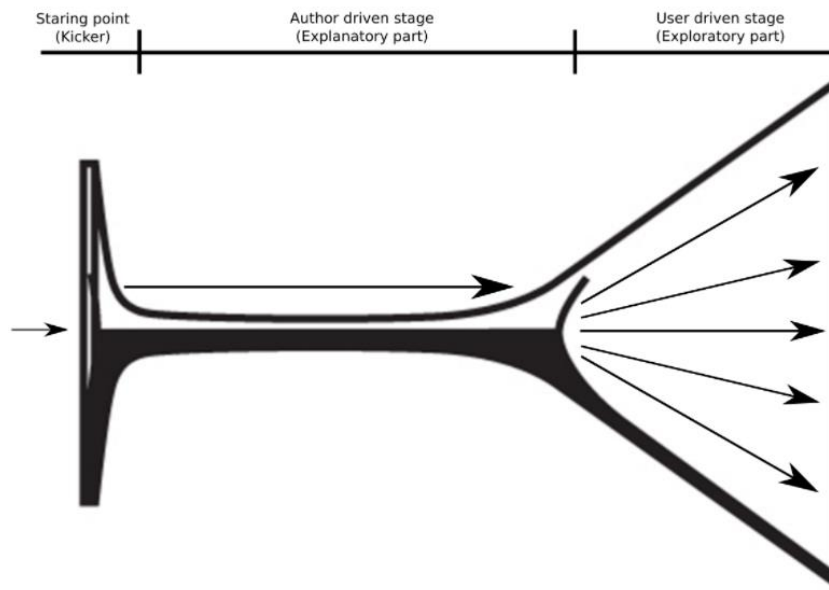
Existují dva hlavní typy analýzy dat, které mají stejný cíl, tedy ukázat data grafickou formou, odhalit informace v datech uvnitř a převzt je na znalosti. Liší se pouze způsob, jakým se snaží těchto cílů dosáhnout. Prvním typem je průzkumová analýza, která hledá v souboru dat zajímavá data a odhaluje skryté vzorce v nich. Zpravidla jsou vizualizace postavené na průzkumové analýze vytvořeny pro odborníky v daném oboru, kteří jsou schopni interpretovat zobrazené výstupy a najít si svou vlastní cestu k nalezení závěrů. Druhým typem je vysvětlující analýza, která pracuje s dostupnými údaji a vypráví příběh, který publiku pomůže vzniknout do problematiky snadnější a srozumitelnější formou. Publikum je v tomto případě pasivnější než v případě průzkumové analýzy.

Zajímavým problémem je výběr typu vizualizace. Existuje mnoho způsobů, jak prezentovat informace přenášené daty. Forma vizualizace a vybrané vyprávění má velký dopad na interpretaci dat. Vyprávěči příběhů integrují vizualizace do vyprávění stále častěji, a dokonce nechají písemnou část plně nahradit vizualizací. Existuje více přístupů, jak vytvořit vizualizaci, a každý z nich je charakterizován jinými vlastnostmi.

První typ vizualizací je zcela veden autorem. Uživateli je předložena sada statických obrazů, které lze číst pouze pasivně. Velkou výhodou tohoto přístupu je, že vizualizace mohou být velmi propracované do nejmenších detailů a autor tak může vyprávět dobrý příběh, který je v mnoha momentech velice sofistikovaný. Dalšími vlastnostmi vizualizací řízenými autorem je lineární řazení scén, kdy uživatel nemá možnost volby alternativního způsobu dosažení cíle a žádná nebo malá interaktivita.

Na opačné straně, než je autorský řízený příběh, je příběh řízený uživateli. Zde mohou uživatelé manipulovat s grafy a provádět vlastní průzkumovou analýzu. Interaktivitu vizualizací lze obvykle přizpůsobit přidáním možností filtrování napříč několika dimenzemi. Různí uživatelé mohou vyvodit různé závěry v závislosti na jejich průchodu datovým souborem a jejich předchozích zkušenostech. Vizualizace založené na řízení uživateli jsou charakterizovány nepředvídatelným sledem kroků uživatele a volnou interaktivitou uživatele s grafy nebo částmi grafů.

Poslední typ tvorby příběhu má nejlepší části z předchozích dvou. Autor postaví část příběhu, který se věnuje dané problematice, například poukáže na důležité momenty, odchylky nebo anomálie. Zároveň uživatel může podrobněji upravovat a filtrovat prezentované grafy a prohlížet data do hloubky. Typickým představitelem této skupiny je přístup s názvem „Martini glass“ inspirovaný tvarem sklenice na Martini. Tento narativní přístup má tři fáze: v první fázi je celá doména představena čtenáři. Ve druhé fázi je čtenář prostřednictvím silného scénáře motivovaného autorem, kde se k vysvětlení domény nebo konkrétního problému používají texty, popisy, anotace nebo animace. Po skončení autorova vyprávění může čtenář aktivně prozkoumat vizualizaci pomocí interaktivních prvků.



Obrázek 26 - Metoda Martini glass

(Komenda, 2016)

Ani jedna z cest – autorem řízená, uživatelem řízená nebo hybridní – není tou preferovanou. Vždy záleží na celkovém kontextu a také na tom, jak autor vizualizace chce, aby uživatel na data pohlížel.

3.8. Datové zdroje, získání a zpracování dat

3.8.1. Datové zdroje

Data pro analýzu a vizualizaci lze získat z mnoha zdrojů. Tyto zdroje spadají do jedné ze dvou kategorií: interní a externí. Interní zdroje jsou ty, které vznikají prostřednictvím aktivit firmy jako záznamy o zákaznících, firemní web, záznamy z poštovních či telefonních kampaní nebo databáze či datové sklady, které jsou přímo určeny k uchování firemních dat. Externí zdroje mívají firmy jako úvěrové kanceláře, obchodníci a kompilátoři seznamů a společnosti s rozsáhlými databázemi zákazníků jako vydavatelé a katalogoví prodejci. (Rud, 2001)

Interní zdroje dat jsou ty, které jsou udržovány uvnitř firmy nebo organizace. Jde často o data s nejvyšší prediktivní schopností pro modelování, protože reprezentují informace, které jsou specifické pro výrobky a služby dané firmy. Typickými zdroji interních dat jsou databáze zákazníků, databáze provedených transakcí, databáze historie nabídek a datové sklady. Datový sklad je struktura, která spojuje informace ze dvou nebo

více databází. Dává dohromady data do jednoho centrálního úložiště, provádí jejich určitou integraci, vyčištění a sumarizaci a informace dále distribuuje. (Rud, 2001)

Mezi externí zdroje se mohou řadit například prodejci a kompilátoři různých seznamů či katalogů, kteří obvykle shromažďují a prodávají jména, adresy a telefonní čísla společně s demografickými, behaviorálními či psychografickými údaji. (Rud, 2001)

Externích datových zdrojů a datových sad je stále více. Jedná se o datové sady ve formátu Open Data. Iniciativa poskytovat data v jednoduše strojově zpracovatelných formátech stále narůstá, a to zejména ve veřejném sektoru. Tím se rapidně zvětšuje i okruh lidí, kteří s daty pracují. Licence otevřených dat totiž umožňuje libovolné použití dat, a to komerční. Česká republika v trendu otevírání dat ve standardizované formě nijak nezaostává a data z veřejného sektoru jsou dostupná skrze Národní katalog otevřených dat. Datové sady jsou kategorizovány zejména dle jejich poskytovatele. Lze zde nalézt data převážně ve formátech JSON a CSV společně s jejich metadatovými soubory. Hlavními poskytovateli dat jsou Český úřad zeměměřičský a katastrální, jednotlivá ministerstva ČR, Český statistický úřad, statutární města, Český telekomunikační úřad a další.



Obrázek 27 - Open Data

(Zdroj: <https://blog.scielo.org/en/2014/07/14/the-open-data-movement-international-consolidation/>)

3.8.2. Databáze

Databáze, často zkracována jako DB, představuje souhrn informací. Je uspořádána tak, že programy přistupující k databázi mohou požadovaná data rychle najít a pracovat s nimi. Ukládání dat zároveň probíhá kontrolovaně, a pokud jsou správně nadefinované integritní omezení, tak by data měla zůstat konzistentní.

Pro vytvoření databáze a spravování dat v ní je potřeba systém pro správu databázi (DBMS – DataBase Management System). To je sbírka programů, které nám umožní přistupovat, organizovat a filtrovat data v databázi. Znamé databázové systémy jsou MySQL, PostgreSQL, Microsoft SQL Server nebo Oracle DB. Databáze obecně nebývají přenosné napříč různými databázovými systémy. Různé DBMS ale mohou spolupracovat pomocí standardů, jako je SQL. (Engeto, 2019)

Databáze neslouží jenom jako uložistiše dat. Uložit data můžeme i v klasickém souboru se základní strukturou (CSV, JSON). Databáze představují komplexní systém, který se jednoduše používá a ovládá se jazykem SQL. Nicméně je nutné si uvědomit, že označení „databáze“ není úplně správné, přesnější je pojem RDBMS (Relation DataBase Management System). Tento pojem je možné přeložit jako „systém řízení báze dat“. (Engeto, 2019)

Komplexita databázových systémů se vyznačuje zabezpečením, výkonem a konzistencí dat, kterou zajišťuje zpracování SQL dotazů v transakcích. Takto není konzistence dat ohrožena ani v případě, že více uživatelů edituje jednu položku najednou. Pokud je transakce neplatná nebo může data ohrozit, databázový systém ji zamítne a transakce se neuskuteční.

3.8.3. Jazyk SQL

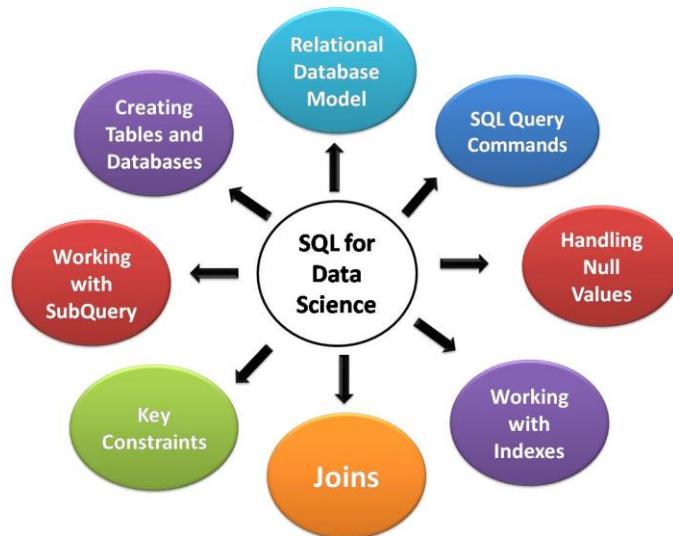
Jazyk SQL, anglicky Structured Query Language, lze přeložit jako strukturovaný dotazovací jazyk, který lze použít pro ukládání, přístup, aktualizaci a extrahování velkého množství dat za účelem následující datové analýzy. Tento jazyk je základním prostředkem pro práci s daty uloženými v relačních databázích, jakožto nejčastějším zdrojem dat.

Jazyk SQL je textově orientovaný. Vznikl o hodně let dříve než grafická uživatelská rozhraní a k psaní jeho příkazů stačí pouhý textový editor. Současné systémy Microsoft Access, Microsoft SQL Server, Oracle Database, MySQL a jiné databázové produkty obsahují grafické nástroje, které plní mnohé funkce jazyka SQL. Příkazy SQL můžeme rozdělit do několika hlavních kategorií: (Kroenke, 2015)

- DDL (Data Definition Language) – příkazy, které umožňují definovat databázové struktury,
- DML (Data Manipulation Language) – příkazy, které slouží k dotazování a úpravám databázových dat,

- DCL (Data Control Language) – příkazy pro správu uživatelských rolí a práv,
- TCL (Transactional Control Language) – příkazy pro správu databázových transakcí.

Jelikož datová analýza je věda pracující s daty, které jsou často uloženy v databázích a jazyk SQL je prostředek pro dotazování, manipulaci a extrahování dat z databáze, je jazyk SQL v rámci datové analýzy jedním z nejpoužívanějších jazyků.



Obrázek 28 - Jazyk SQL

(Zdroj: <https://www.analyticsvidhya.com/blog/2021/06/sql-for-data-science-a-beginners-guide>)

3.9. Nástroje používané k vizualizaci dat

V následující kapitole je uveden seznam nejpoužívanějších nástrojů pro vizualizaci dat rozdělen na dvě kategorie. První kategorií jsou programátorské nástroje a druhou kategorií nástrojů je hotový software, který je pro vizualizaci dat možno použít bez programátorských dovedností.

3.9.1. Programátorské nástroje

Programátorské nástroje poskytují větší svobodu při zpracování dat a může být pomocí nich zpracováno automatizovaně velké množství dat. Programovací jazyky JavaScript, Python a R jsou oblíbené jazyky pro vizualizaci dat a mají velké množství knihoven, které se dají pro tyto účely použít. (FineReport, 2018)

a) Python

Python je vysokoúrovňový programovací jazyk s univerzálním použitím a jednou z oblastí, kde jej lze použít, je právě datová analýza a vizualizace dat. Pro tyto účely nabízí řadu různých knihoven. Datoví analytici používají pro práci s Pythonem jeho distribuci Anaconda, která obsahuje nástroje pro datovou analýzu, data science a strojové učení a také IDE. Součástí Anacondy je i Jupyter notebook, nejpoužívanější prostředí pro datovou analýzu v Pythonu. V Jupyter notebooku je možné navíc pouštět i další jazyky, například jazyk R.

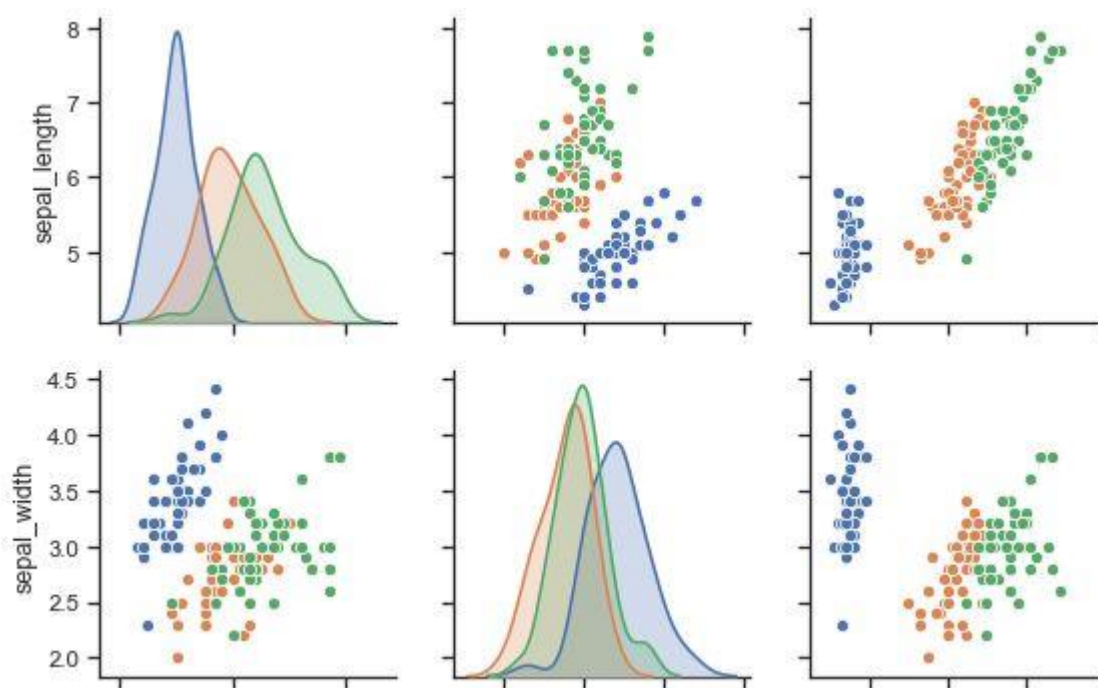
NumPy (Numerical Python) je knihovna vytvořená za účelem usnadnění operací a výpočtů s poli o mnoha dimenzích. Základní objekt této knihovny, NumPy pole, umožňuje reprezentovat data tak, aby s nimi bylo možné dále efektivně pracovat. Například ve strojovém učení, se s reprezentací dat pomocí NumPy polí pracuje zcela běžně. Kromě toho NumPy obsahuje celou řadu funkcí užitečných pro transformaci dat. Na NumPy pole lze aplikovat mnoho dalších funkcí jako logaritmus, exponent, směrodatná odchylka, maticové operace a podobně. Pokud je funkce aplikována na vícerozměrné pole, je možné zvolit dimenzi. Operace s polem NumPy je efektivnější než s běžným polem s daty. Je to dáno tím, že NumPy provádí takzvané vektorizované operace, kdy operace je aplikována na celé pole najednou místo jednotlivých hodnot postupně. Vektorizací se NumPy vyhýbá cyklům, které jsou výpočetně náročné. Vektorizované operace jsou mnohem rychlejší a při práci s velkým množstvím dat jsou jedinou možností, jak operace provést.

Další knihovnou určenou pro manipulaci s daty je knihovna Pandas. Jedná se o nadstavbu knihovny NumPy a umožňuje intuitivní a snadnou práci s daty uloženými v NumPy polích. Základním objektem je DataFrame, což je skupina jednorozměrných NumPy polí. DataFrame má řádkový a sloupcový index, pomocí kterých je možné se v datech snadno orientovat.

Matplotlib je komplexní knihovna pro tvorbu statických, animovaných a interaktivních vizualizací v Pythonu. Zobrazení dat v grafu je jedním ze způsobů, jakým se datový analytik seznamuje s daty, než je bude dále používat, například v modelech. Grafy mohou být i konečným výstupem analýzy. V Matplotlibu jsou dva způsoby konstrukce grafů, jsou jím pyplot a objektově orientovaný přístup. Graf v Matplotlibu se skládá z různých objektů. Pokud je použit objektově orientovaný přístup, je nutné explicitně vytvořit tyto objekty a poté je editovat zvlášť. Při použití pyplotu to Matplotlib udělá automaticky, nicméně zde jsou možnosti značně omezené. Pro tvorbu efektivních grafů, je

nutné používat objektově orientovaný přístup. Pyplot většinou slouží pouze k rychlému vykreslení kontrolních grafů.

Knihovna Seaborn je nadstavbou knihovny Matplotlib určená zejména pro statistiky. Pro pokročilejší práci a pochopení některých funkcí v Seaborn je tedy zapotřebí pokročilejší znalost statistiky a pravděpodobnosti. Stejně jako v Matplotlib lze i v Seaborn pracovat s datovými strukturami DataFrame z knihovny Pandas, vytvářet různé typy grafů a snadno je customizovat.



Obrázek 29 - Python Seaborn

(Zdroj: <https://www.fromthegenesis.com/seaborn-python-data-visualization-tool/>)

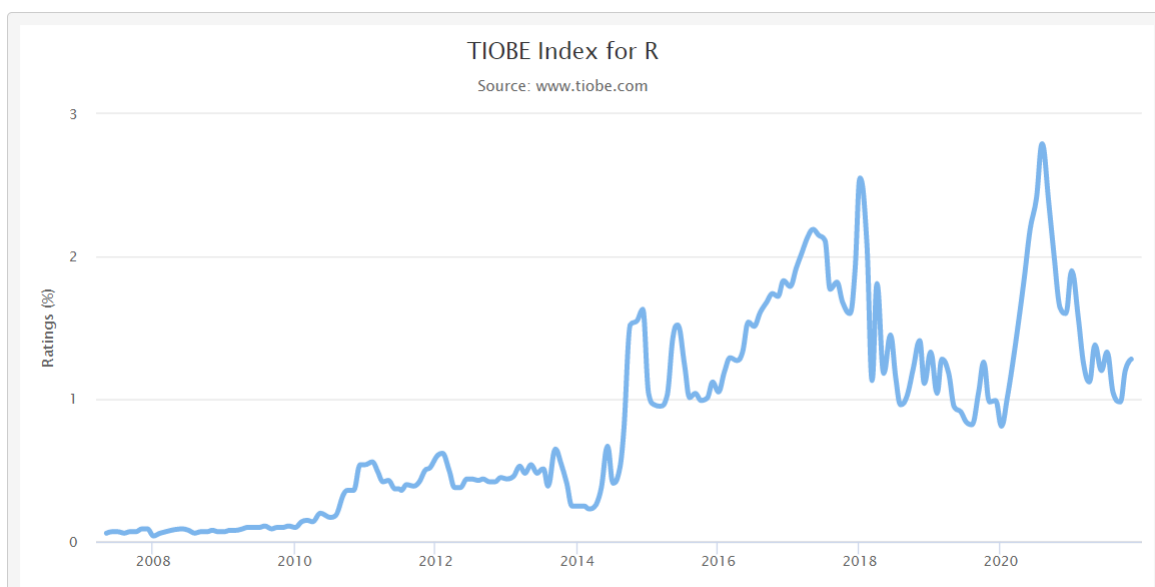
b) JavaScript

Programovací jazyk JavaScript je pro tvorbu vizualizací velmi vhodným, protože je spouštěn přímo v klientském prohlížeči a poskytuje tak mnoho možností pro interaktivitu uživatele. Existuje mnoho hotových knihoven, které lze v tomto jazyku pro tvorbu vizualizací použít. Jedním z nejznámějších a nejpoužívanějších zástupců je knihovna d3.js, která obsahuje funkce pro práci s různými typy grafů a dále ji lze použít například pro animace, datovou analýzu a interakci uživatele. Princip knihovny d3.js spočívá v tom, že při vizualizaci jsou hlavním stavebním kamenem data a knihovna má za cíl co nejvíce usnadnit práci s nimi, aniž by autora vizualizace omezovala ve vzhledu výsledné vizualizace. Pro uspořádání dat a jejich vizualizaci knihovna obsahuje mnoho algoritmů,

které lze použít. Jako další zástupce podobných knihoven lze jmenovat knihovny ChartJS, Highcharts či Echarts.

c) R

R je doménově specifickým jazykem pro zpracování a statistickou analýzu dat, včetně jejich grafického zobrazení, která je typicky následovaná zveřejněním na webu ve formě statické nebo interaktivní prezentace. Na R je ovšem nejdůležitější nikoli vlastní jazyk, ale ekosystém, který je okolo něho postaven. Jazyk R byl představen již před 27 lety a i přesto, že se jednalo a stále jedná o velmi rozšířený jazyk v oblasti zpracování dat, statistických analýz a vizualizace výsledků, stále jde o doménově specifický jazyk, který se příliš nehodí pro vývoj obecných aplikací vhodných pro produkční nasazení. Proto by se podle všech předpokladů neměl tento jazyk umísťovat na předních místech žebříčku oblíbenosti, nicméně jeho popularita s rostoucí mírou důležitosti zpracování a analýzy dat roste. Velmi silnou stránkou programovacího jazyka R je tvorba grafů, která je většinou velmi intuitivní. (Tišnovský, 2020)



Obrázek 30 - Popularita jazyka R

(Zdroj: <https://www.tiobe.com/tiobe-index/r/>)

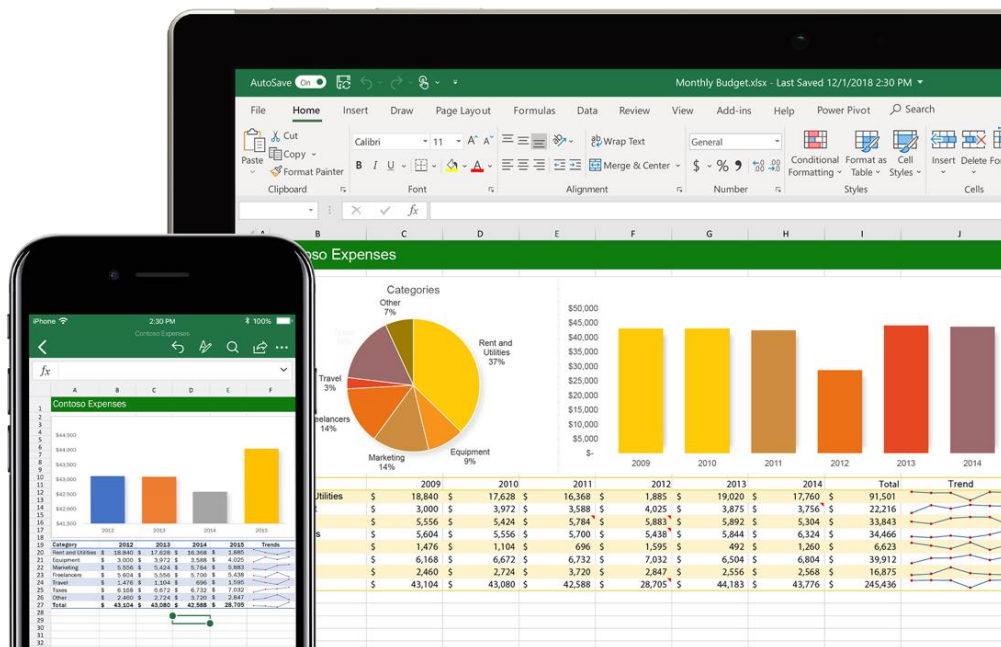
3.9.2. Hotový software

Následující software je možné pro vizualizaci dat použít bez jakýchkoli programátorských znalostí. Kromě vizualizace dat tyto nástroje poskytují funkce pro

napojení na zdroj dat, například databázi, analýzu dat a jejich další zpracování. (FineReport, 2018)

a) Microsoft Excel

MS Excel je součást kancelářského balíku MS Office a zároveň nejpoužívanější aplikace typu tabulkového procesoru. Lze v něm tvořit mnoho různých typů grafů, od základního sloupcového či spojnicového až po různé typy mapových grafů. Přestože se jedná o základní nástroj v oblasti analýzy a vizualizace dat, nachází uplatnění v mnoha společnostech, kde je tou první volbou pro vytvoření grafu.



Obrázek 31 - Microsoft Excel

(Zdroj: <https://www.microsoft.com/en-ww/microsoft-365/excel>)

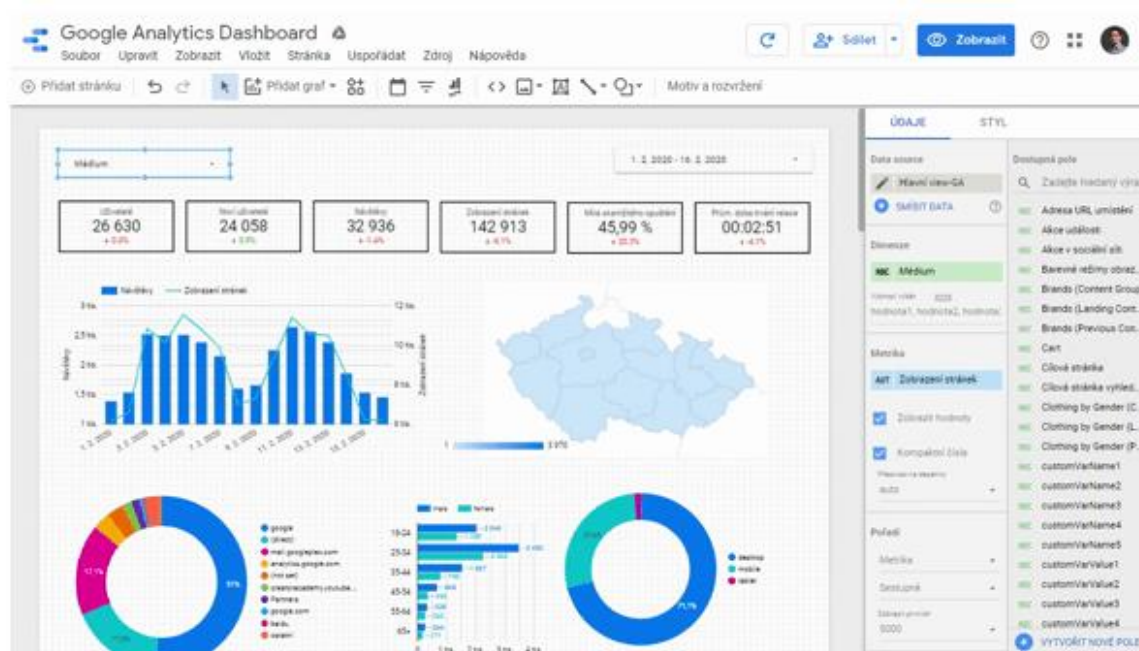
b) Google Data Studio

Nástroj Google Data Studio je produktem společnosti Google a je dostupný volně všem uživatelům, kteří mají u společnosti Google účet.

Google Data Studio je bezplatný online nástroj, díky kterému jsou data přehledně na jednom místě. Nabízí jednoduché ovládání, nastavení vlastního vzhledu, interaktivní možnosti filtrování, dynamické ovládací prvky, porovnání časového období, matematické nebo řetězcové funkce pro vytváření vlastních vypočítaných polí, sdílení i export reportu

v PDF, správu práv uživatelů, automatické odesílání reportu na e-mail nebo sledování dashboardu pomocí ID měření služby Google Analytics. (Herel, 2020)

Data v již vytvořeném dashboardu se automaticky aktualizují při každém otevření dashboardu. Nástroj Google Data Studio není nutné využívat jen pro zpracování dat služby Google Analytics, ale lze jej připojit i na databáze Microsoft SQL Server, MySQL, PostgreSQL, načíst datové soubory ve formátu CSV, JSON, XML a mnoho dalších.



Obrázek 32 - Google Data Studio

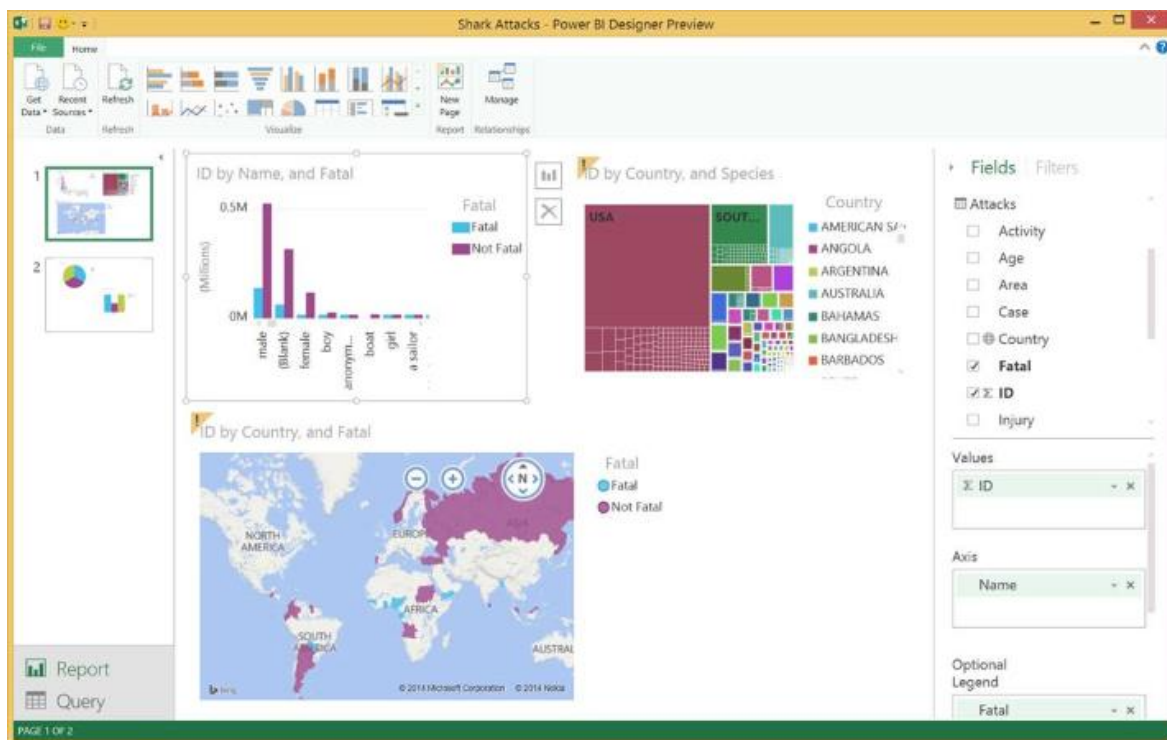
(Herel, 2020)

c) Microsoft Power BI

Microsoft Power BI obsahuje analytické a reportingové nástroje pro zpracování většího objemu dat. 3D vyobrazení analytických výstupů nad mapovými podklady, automaticky vytvářené reporty na základě dotazů kladených v běžném jazyce (angličtině) či interaktivní animované grafy zobrazující vývoj v čase. BI je zkratka pojmu Business Intelligence, který označuje sběr, integraci, analýzu, interpretaci a prezentaci informací pro získání lepšího pochopení chování na trhu a obchodních souvislostech. (Plíva, 2018)

Pod názvem Power BI se skrývá několik různých variant aplikace. První variantou jsou různé doplňky do aplikace Microsoft Excel, které umožňují napojení na datové zdroje, načtení dat (Power Query), vytváření analytických sestav prostřednictvím tabulek a grafů (Power View) a sestavení datových modelů vytvářením relací, vlastních měřítek, hierarchií a klíčových ukazatelů výkonu (Power Pivot). Druhou variantou jsou serverové aplikace

podporující sdílení vytvořených reportů a sestav prostřednictvím cloudových služeb. Jedná se o cloudovou službu Power BI nebo Power BI pro Office 365. Poslední variantou nástroje Power BI je klientská aplikace Power BI Designer nainstalovaná na pracovní stanici uživatele, která sjednocuje veškeré výše zmiňované funkce do jedné aplikace.



Obrázek 33 - Power BI Designer

(Zdroj: <https://www.microsoft.com/en-ww/microsoft-365/excel>)

d) Tableau Software

Tableau je jeden z nejoblíbenějších analytických produktů a lídr ve vizualizaci dat. Motto nástroje Tableau zní: „Tableau helps people see and understand their data“. Tableau má několik různých variant (Sleeper, 2018):

- Tableau Desktop: Personal
Jedná se o základní placenou variantu určenou pro vývoj vizualizací. Umožňuje uchovat vytvořené vizualizace v soukromí, ale možnosti pro napojení na datové zdroje a distribuci vytvořených vizualizací jsou omezené, jelikož se jedná pouze o variantu instalovanou lokálně na pracovní stanici.
- Tableau Desktop: Professional
Poskytuje stejné funkce pro vývoj vizualizací, jako předchozí varianta, ale umožňuje napojení na všechny typy datových zdrojů, které Tableau

podporuje. Navíc umožňuje ukládat vytvořené vizualizace na Tableau Server nebo Tableau Online.

- Tableau Reader

Tableau Reader je nástroj zdarma, který umožňuje otevírat vizualizace vytvořené v nástroji Tableau Desktop a další interaktivní práci s nimi, ale neumožňuje jejich úpravu a vývoj.

- Tableau Public

Tato varianta produktu Tableau je opět zdarma a poskytuje funkce potřebné pro vytváření vizualizací. Nicméně výsledné vizualizace musí být uloženy v cloudu, kde jsou veřejně dostupné. Z tohoto důvodu tato varianta není vhodná například pro obchodní či jinak citlivá data.

- Tableau Server

Tableau Server poskytuje centrální úložiště pro všechny vizualizace vytvořené v Tableau, kde jsou dostupná pro uživatele pomocí webového prohlížeče. Jedná se o „on premise“ software, tedy software je instalován na server dané společnosti nebo cloud společností používaný.

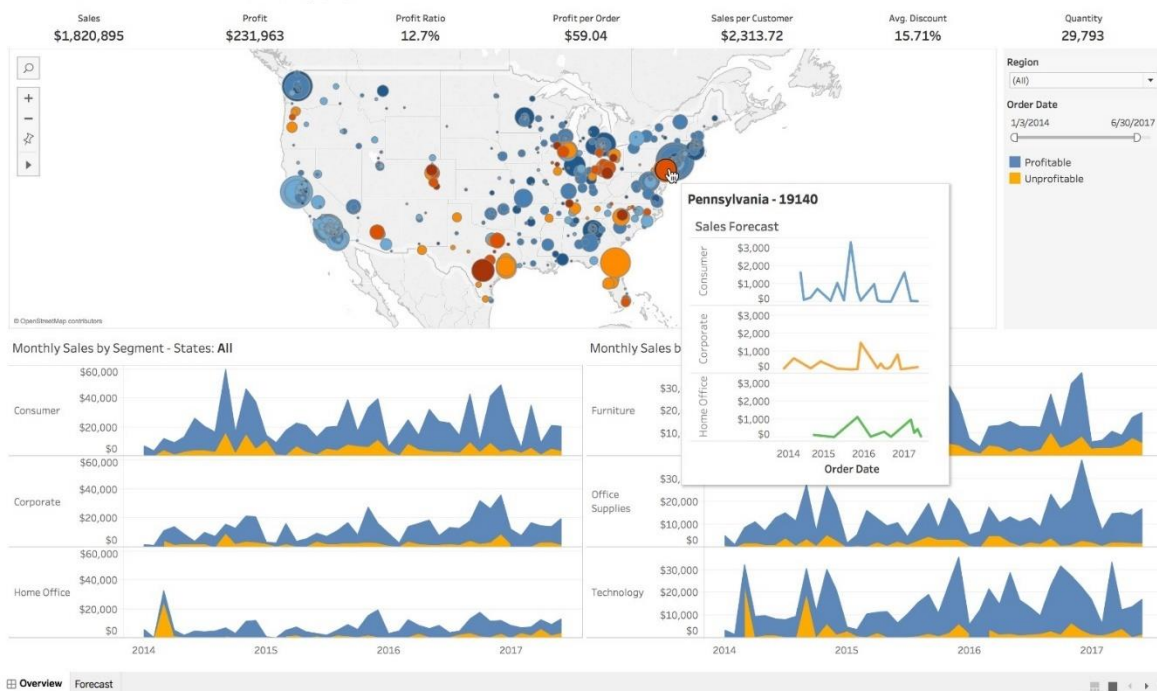
- Tableau Online

Varianta Tableau Online je podobná jako Tableau Server, nicméně centrální úložiště není server/cloud společnosti, ale cloud partnerské společnosti Tableau.

Rozhodnutí, kterou variantu produktu Tableau používat, je potřeba zvážit podle čtyř následujících kritérií:

- Z jakých datových zdrojů budou použita data pro vizualizaci?
- Kdo bude vizualizaci používat a jakým způsobem bude sdílena?
- Je potřeba, aby se vizualizace automaticky aktualizovala na základě aktualizace dat?
- Je vyžadována bezpečnost na úrovni on-premise řešení nebo mohou být vizualizace ukládány v cloudu?

Executive Overview - Profitability (All)



Obrázek 34 - Tableau Desktop

(Zdroj: <https://www.tableau.com/products>)

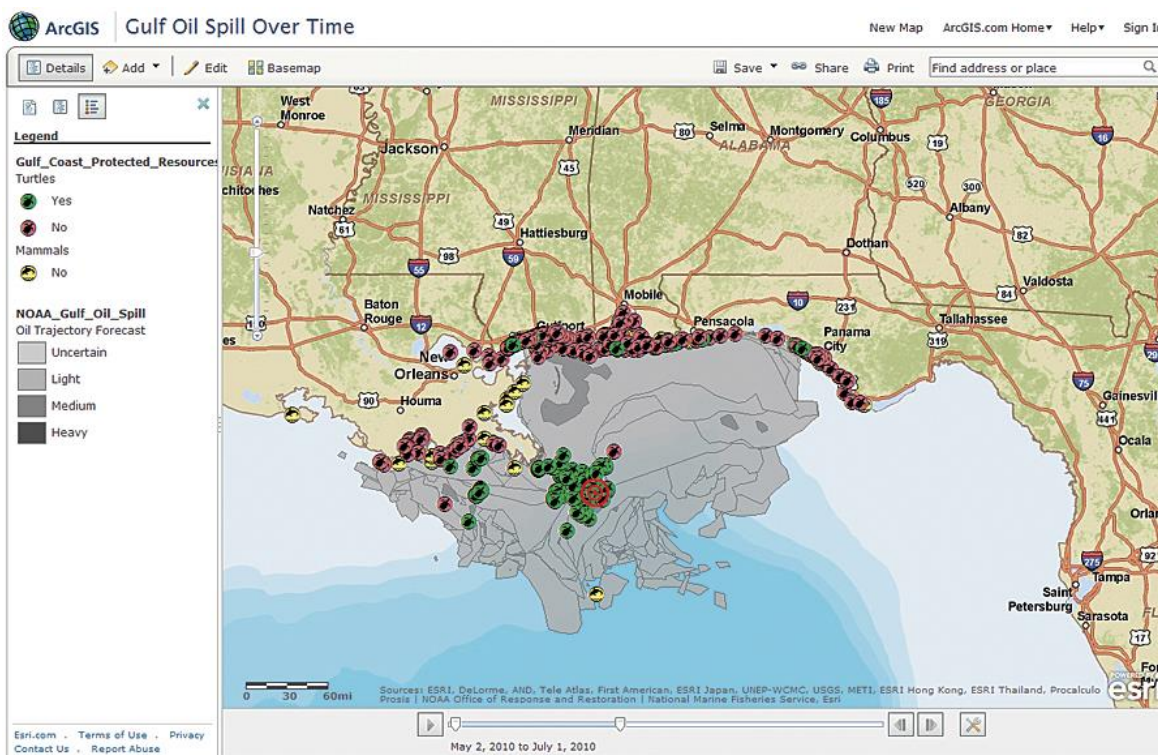
Výhodou nástroje Tableau oproti dvěma předchozím nástrojům je přímá podpora tvorby datově řízeného příběhu. Lze vytvořit interaktivní slideshow z již vytvořených stránek reportu a vhodně je provázat. Čtenář má pak mnohem lepší prožitek z prozkoumávání výstupu a je možné mu obsah snadně rozdělit do tematických celků.

e) Nástroje pro mapové vizualizace

Pro vytvoření vizualizace na reálném mapovém podkladu je možné využít již hotové knihovny a nástroje. V případě knihoven se jedná většinou o řešení v jazyce JavaScript a základní použití bývá dobře popsáno v dokumentaci. Jako zástupce lze zmínit open-source knihovnu Leaflet, která je dobře kombinovatelná i s dalšími knihovnami, jako je již zmiňovaná knihovna d3.js. Jako druhého zástupce lze jmenovat Mapbox. Jedná se o placenou službu, kterou je možné používat i v mobilních aplikacích.

Volně dostupnou aplikací, ve které je možné efektivně pracovat s geografickými daty a tvořit mapové výstupy, je Esri ArcGIS Online. Jedná se o produkt společnosti Esri, který lze využívat ihned po vytvoření účtu a přihlášení. Existují různé úrovně licencování, které jsou více popsány na webových stránkách produktu, nicméně lze vytvořit i bezplatný účet pro nekomerční účely, ale s tímto účtem jsou vytvořené vizualizace vždy

veřejné. Nástroj ArcGIS Online umožňuje tvorbu vizualizace pomocí mapového podkladu nebo i 3D vizualizace vybraného místa na planetě, do které lze zahrnout externě dodaná data. Do zveřejněné vizualizace lze přidávat také interaktivní prvky, kterými může uživatel s vizualizací dále manipulovat.



Obrázek 35 - ArcGIS Online

(Zdroj: <https://www.esri.com/news/arcnews/summer11/articles/arcgis-online-whats-new.html>)

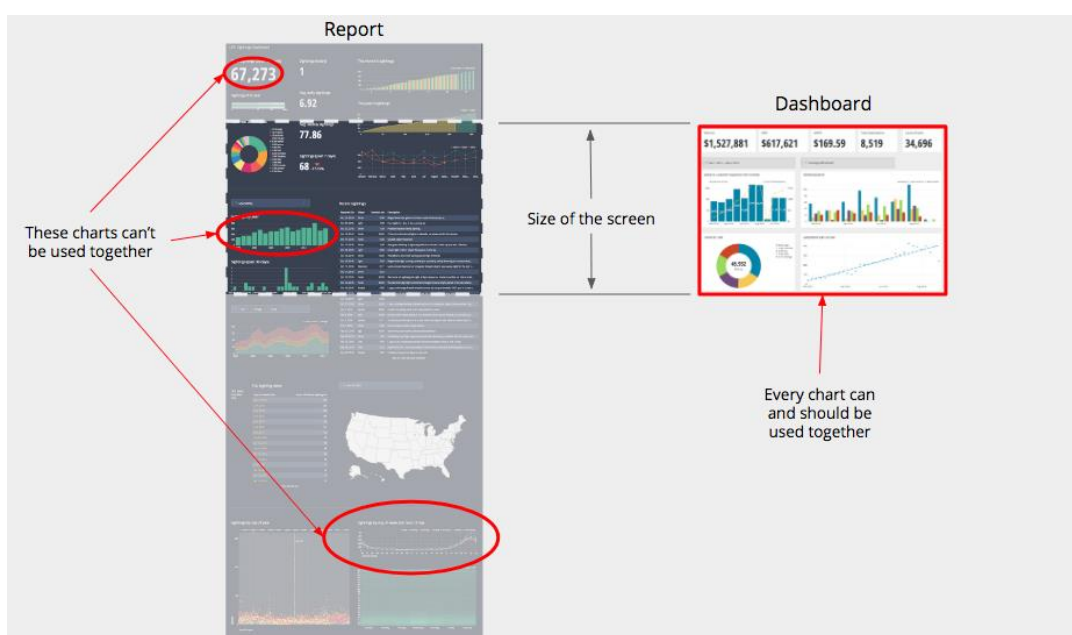
3.10. Business Intelligence

Systémy Business Intelligence (BI) jsou informační systém, které manažerům a dalším profesionálům pomáhají při analýze aktuálních a minulých aktivit a předpovídání budoucích událostí. Na rozdíl od ostatních systémů transakčního zpracování nepodporují provozní aktivity nebo primární obchodní aktivity, jako je například záznam či zpracování objednávek, ale místo toho jsou určeny k podpoře analytických a rozhodovacích činností managementu. Manažerům usnadňují hodnocení, analýzu, plánování, předpovídání budoucích událostí a v důsledku i rozhodování a řízení. (Kroenke a Auer, 2015)

Dashboards jsou nedílnou součástí BI podniku. Měly by být navrženy a vytvořeny takovým způsobem, aby se s jejich pomocí dala analyzovat klíčová data podniku a sloužily k podpoře rozhodování a řízení podniku. Místo toho, aby datoví analytici ručně sestavovali

tabulky, mohou moderní platformy BI automaticky k datům přistupovat, analyzovat a zobrazovat prostřednictvím webových dashboardů.

Rozdíl, mezi dashboardem a reportem spočívá v tom, že report obsahuje sestavu tabulek a grafů s větším detailem, obsahuje tedy podrobnější informace. Report je vhodnější pro nižší management, který pro řízení na dané úrovni potřebuje detailnější informace o stavu. Dashboard oproti tomu obsahuje zjednodušený pohled, dává odpověď na konkrétní otázku a je tedy ideální pro vyšší management, který potřebuje rychlý přehled o výkonu daného ukazatele. Dashboards bývají vizuální a interaktivní a umožňují uživateli vytvářet vlastní analýzy. Reporty bývají spíše statické.



Obrázek 36 - Report vs dashboard

(Zdroj: <https://chartio.com/blog/dashboards-vs-reports-how-theyre-the-same-how-theyre-different>)

Klíčové vlastnosti moderních BI dashboardů se dají shrnout do následujících několika bodů:

- Přizpůsobitelné rozhraní
- Interaktivita
- Schopnost načítat data z datových zdrojů téměř v reálném čase
- Přístupnost z webového prohlížeče
- Sdílení mezi uživateli a podpora jejich spolupráce
- Standardní šablony

Hlavní význam dashboardů a BI je, že umožňují organizacím porozumět komplexním datům, které uchovávají ve svých či externích informačních systémech a činí informace přístupnými i pro netechnické uživatele.

4. Vlastní práce

V rámci praktické části této diplomové práce bylo realizováno několik dashboardů pro společnost Exon s.r.o..

4.1. Představení společnosti

Společnost Exon s.r.o. je předním dodavatelem efektivních řešení v oblasti digitální kanceláře. Dodává renomované systémy správy dokumentů (DMS/ECM) a komplexní služby digitalizace dokumentů zahrnující vytěžování dat z formulářů či faktur, zabezpečené archivace dokumentů, dodávky profesionálních digitalizačních pracovišť, zakázkového vývoje software, instalace velkoformátových a knižních skenerů a s tím souvisejícího poradenství.

Zároveň je společnost Exon s.r.o. konzultační společnost v oblasti digitalizace firemních procesů a bezpapírové organizace včetně dodání a implementace vhodné technologie. Vytváří analýzy současného stavu interních procesů (workflow) a analýzy oběhu dat a dokumentů, na základě které vytvoří návrh optimálního řešení a vhodných postupů dle mnohaletých zkušeností a know-how v oboru digitalizace a elektronické správy dokumentů a jejich obsahu.

Obchodní firma	Exon s.r.o.
Vedení společnosti:	Ing. Radek Chramosta, jednatel
Datum vzniku:	24.8.2004
Počet zaměstnanců:	cca 50
IČO:	26376326
DIČ:	CZ26376326
Sídlo:	Vrážská 73/10, 153 00 Praha
Webové stránky:	www.exon.cz
Email:	exon@exon.cz

Tabulka 1 - Základní informace o společnosti Exon s.r.o.

(Zdroj: justice.cz, exon.cz)

4.1.1. Reference

Společnost Exon s.r.o. spolupracuje se zákazníky ze širokého spektra oborů:

- Bankovní a finanční sektor
- Výroba, služby a obchod
- Městské úřady a magistráty
- Akademické instituce

- Státní správa
- Doprava a logistika



Obrázek 37 - Reference společnosti Exon s.r.o.

(Zdroj: <https://www.exon.cz/cs/reference>)

4.1.2. Partneri

ELO Digital Office GmbH sídlící v německém Stuttgartu je distributor a výrobce špičkového systému pro správu dokumentů a obsahu – DMS/ECM ELO. Společnost Exon s r.o. technologickou platformu ELO využívá pro digitalizační projekty desítek zákazníků nejen v České republice.

Společnost i2S sídlící ve francouzském Bordeaux je světová špička v digitalizaci kulturního dědictví, technologii pro digitalizaci (knižní a velkoformátové skenery) a zpracování obrazu. Společnost Exon s.r.o. je autorizovaný partner pro Českou a Slovenskou republiku.

4.2. Popis současného stavu

Společnost Exon s.r.o. používá několik informačních systémů. Jedná se především o DMS ELO, Docházkový systém a CRM systém. V těchto systémech je uložena většina dat společnosti. Dále společnost využívá služby externí účetní firmy, v jejímž účetním systému jsou uložena data týkající se účetnictví, mezd atd.

4.2.1. DMS ELO

Do systému pro správu dokumentů DMS ELO jsou vkládány veškeré elektronické dokumenty, které jsou do společnosti přijaty nebo jsou ve společnosti vytvořeny. Jedná se tedy o systém, ve kterém jsou spravovány a archivovány dokumenty společnosti a související metadata. Systém byl implementován v druhé polovině roku 2011 a od té doby je dále soustavně rozvíjen. Vzhledem k tomu, že společnost byla založena v roce 2004, systém obsahuje data z velké části existence společnosti, přičemž některá data od vzniku společnosti byla do systému také dodatečně naimportována.

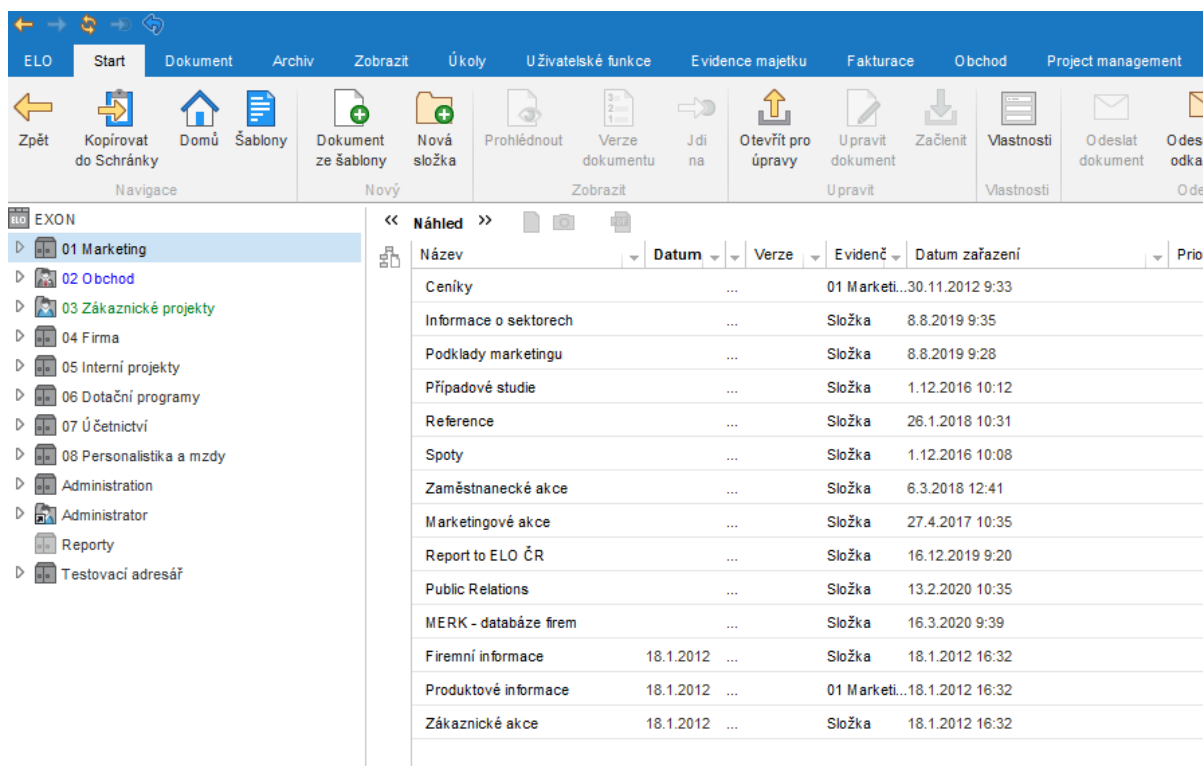
Systém obsahuje dokumenty a související data z následujících oblastí:

- Marketing
- Obchod
- Zákaznické a interní projekty
- Dotační programy
- Účetnictví
- Personalistika
- další firemní dokumenty (autopark, pošta, nákup, směrnice, GDPR, IT, ...).

Systém DMS ELO ve společnosti používají pro svou práci všichni zaměstnanci, přičemž každý může mít nastavená různá uživatelská práva, dle své pozice v organizační struktuře společnosti.

Možnosti systému DMS ELO pro tvorbu reportů jsou značně omezené. Přestože se jedná o systém obsahující mnoho funkcí pro správu elektronických dokumentů, správu workflow, uživatelů a podobně, v základu neobsahuje téměř žádné funkce pro vytvoření reportu, pouze jednoduchou funkci pro export metadat. Při požadavku na report je nutné, aby export požadovaných informací provedl administrátor například z databáze či jiným způsobem. Dále systém nabízí možnost naprogramovat vlastní funkci, která by například provedla export požadovaných informací do aplikace MS Excel. V tomto případě je však potřeba práce programátora.

Databázový systém, ve kterém jsou veškerá data ukládána, je v tomto případě Microsoft SQL Server 2019.



Obrázek 38 - DMS ELO

(Zdroj: vlastní zpracování)

4.2.2. Docházkový systém

Dalším ve společnosti široce využívaným systémem je Docházkový systém. Jedná se o systém, který používají všichni zaměstnanci i brigádníci na každodenní bázi. Každý zaměstnanec či brigádník v systému eviduje svou docházku, tedy příchod do práce, odchod z práce a přestávky. Stejně tak jsou každý den vyplňovány vykonané činnosti, které jsou děleny do třech kategorií, a to interní činnosti, projektové činnosti a třetí kategorií je marketing a obchod. Každá kategorie je dále členěna na další podkategorie. Do interních činností patří například marketing, administrativa, údržba, vzdělávání a podobně. Do projektových činností patří činnosti, které jsou zaměstnanci vykonávány v rámci jednotlivých projektů, ať už zákaznických či interních. Do třetí kategorie jsou vkládány činnosti týkající se například tvorby marketingových materiálů a kampaní, přípravy zákaznických akcí a činnosti související s obchodními příležitostmi, tedy vyhledávání potenciálních zákazníků, příprava nabídek, jednání s potenciálními zákazníky a podobně. Dále je v systému evidována dovolená zaměstnanců, pracovní cesty, nemoc či další typy dní. Jednotlivé činnosti jsou taktéž v rámci systému schvalovány vedoucím daného zaměstnance či projektovým manažerem, pokud se jedná o projektovou činnost. Systém

také umožňuje vedoucímu zaměstnanci nařizovat svým podřízeným mimořádnou práci mimo běžný režim, tedy například práci v den pracovního klidu.

práce na pracovišti

Denní výkaz docházky ?

Práce na pracovišti	Příchod	Přestávky		Celkem	Odchod
		Od	Do		
Zaznamenáno:	08:59	11:11 16:52	12:00 00:00	00:48	00:00
Požadováno:	08:59	11:11 16:52	12:00 00:00	00:48	00:00

Zadání činnosti ?

Od:* : (hh:mm)
 Do:* : (hh:mm)
 Projekt/Servis:*
 Fáze Projektu/Servisu:*
 Oblast:*
 Podoblast:*
 Činnost:*
 Homeworking:*

Obrázek 39 - Docházkový systém

(Zdroj: vlastní zpracování)

Oproti systému DMS ELO nabízí Docházkový systém mnoho možností pro export dat. Systém nabízí předdefinované exporty například pro export pracovních cest zaměstnanců, export podkladů pro mzdy, přehled čerpání dovolených, přehled práce mimo běžný režim (o víkendy) a export evidence pracovní doby. Kromě předdefinovaných exportů systém umožňuje i tvorbu vlastních uživatelských exportů. Docházkový systém ale umožňuje export dat pouze do formátu CSV nebo XLSX, nepodporuje žádnou formu vizualizace dat.

Export docházky ?

Exportovat stránku:	<input type="button" value="CSV"/>
	<input type="button" value="XLSX"/>
Exportovat přehled pracovních cest:	<input type="button" value="XLSX"/>
Exportovat všechny zaměstnance:	<input type="button" value="XLSX"/>
Exportovat podklady pro mzdy: <input type="text" value="EXON s.r.o."/> ▼	<input type="button" value="XLSX"/>
Exportovat přehled čerpání dovolených:	<input type="button" value="XLSX"/>
Exportovat přehled práce mimo běžný režim:	<input type="button" value="XLSX"/>
Exportovat evidenci pracovní doby:	<input type="button" value="XLSX"/>

Obrázek 40 - Export dat z Docházkového systému

(Zdroj: vlastní zpracování)

Pro práci s daty používá Docházkový systém databázový systém MySQL.

4.2.3. CRM systém

Systém CRM je ve společnosti Exon s.r.o. používán především pro správu potenciálních zákazníků, obchodních příležitostí a obchodních kontaktů. Ke každému potenciálnímu zákazníkovi je možné v systému evidovat například název společnosti, odvětví, e-mailovou adresu, adresu webových stránek, IČ, DIČ a další údaje. Dále je možné pro každého potenciálního zákazníka vytvořit libovolné množství obchodních kontaktů, tedy osob, které je možné u potenciálního zákazníka kontaktovat. V neposlední řadě je možné pro potenciální zákazníky vytvářet obchodní příležitosti a dále s nimi v systému pracovat, vyhodnocovat a podobně.

Systém CRM používají především zaměstnanci z oddělení marketingu a oddělení obchodu. Systém pracuje s daty uloženými v databázovém systému Firebird.

Nový Klient

Obecné	Dokumenty	Aktivity	Obchod	Organizace	Finance	Sklad	Personalistika	Marketing	Ostatní
* Jméno	<input type="text"/>								
* Typ vztahu	Zákazník	Odpovědná osoba		Administrator					
Odvětví	Zkratka		<input type="text"/>						
IČ	<input type="text"/>	ARES	DIČ	<input type="text"/>	CZ+IČ				
E-mail	<input type="text"/>	info@.cz	Napsat e-mail						
Web	<input type="text"/>								
Telefon	<input type="text"/>	Volat	Mobilní telefon	<input type="text"/>	Volat				
Typ	<input type="text"/>								
Poznámky	<div style="border: 1px solid #ccc; padding: 5px;"><p>Rich text editor toolbar: Bold, Italic, Underline, Text color, Background color, Bulleted list, Numbered list, Indent, Outdent, Link, Unlink, Undo, Redo, Print, Edit.</p><div style="border: 1px solid #ccc; height: 100px; width: 100%;"></div></div>								
<input type="checkbox"/> Režim HTML									

Obrázek 41 - Systém CRM

(Zdroj: vlastní zpracování)

4.2.4. Získávání informací pro potřeby řízení společnosti

Výše zmiňované systémy obsahují většinu provozních dat společnosti, nicméně poskytují pouze omezené možnosti pro tvorbu reportů. Kromě toho, data mezi systémy spolu do jisté míry souvisí, ale v rámci jednoho systému je možné tvorba reportu pouze s daty, které daný systém obsahuje. Systémy také podporují reporty do různých formátů. Největší slabinou systémů v rámci tvorby reportů je, že reporty obsahují data pouze ve formě tabulek. V okamžiku, kdy vedení společnosti vyžaduje komplexnější report obsahující data z více systémů nebo data zpracovaná do grafů, je pro tyto účely vyžadována práce specialisty z IT oddělení, který data z jednotlivých systémů vyexportuje, sloučí do jednoho reportu a dále zpracuje dle požadavku. Tento proces však vyžaduje určitou práci zaměstnance IT a také čas. Management společnosti tedy nemá informace potřebné pro řízení společnosti k dispozici v daném okamžiku, kdy jsou požadovány, každý požadavek na report s sebou přináší určité náklady v podobě práce zaměstnance IT a

jelikož je report zpracován manuálně, existuje také riziko v podobě nepřesného zpracování dat či nepochopení zadání.

4.3.Sběr a analýza požadavků

Na základě rozhovorů s vedením společnosti byl sestaven katalog požadavků na vytvářené dashboardy v následující tabulce.

Číslo	Název	Popis	Stakeholder	Priorita	Návaznost
1	Přehled obchodních příležitostí	Přehled obchodních příležitostí se základními informacemi (název zákazníka a obchodní příležitosti, odpovědná osoba, zdroj obchodní příležitosti a odhad ceny/objemu obchodní příležitosti).	Ředitel, Vedoucí obchodu	1	-
2	Filtr obchodních příležitostí	Obchodní příležitosti bude možné filtrovat dle odpovědné osoby (obchodník), zdroje obchodní příležitosti a data vytvoření obchodní příležitosti.	Ředitel, Vedoucí obchodu	3	1
3	Vizualizace obchodních příležitostí	Vizualizace množství obchodních příležitostí dle zdroje pomocí vhodného grafu.	Ředitel, Vedoucí obchodu	3	1
4	Klíčový ukazatel obchodních příležitostí	Součet odhadované ceny vyfiltrovaných obchodních příležitostí.	Ředitel, Vedoucí obchodu	4	1
5	Přehled projektů	Přehled projektů se základními informacemi (název zákazníka a projektu, stav projektu, náklady projektu v rámci faktur přijatých a odpracovaných MD, výnosy projektu na základě vydaných faktur, zisk).	Ředitel, Projektová kancelář	1	-
6	Filtr projektů	Možnost filtrovat projekty dle stavu projektu a roku vytvoření projektu.	Ředitel, Projektová kancelář	3	5
7	Klíčový ukazatel projektů	Součet částek faktur přijatých, částek faktur vydaných, nákladů MD a zisku dle vyfiltrovaných projektů.	Ředitel, Projektová kancelář	3	5

8	Vizualizace činností zaměstnanců	Vizualizace množství vykazovaných činností zaměstnanců dle typu a podtypu činností.	Ředitel, Vedoucí všech středisek	1	-
9	Filtr činností zaměstnanců	Možnost filtrovat činnosti zaměstnanců podle střediska a dle projektu.	Ředitel, Vedoucí všech středisek	3	8
10	Přehled nákladů středisek	Přehled nákladů středisek se základními informacemi (středisko, číslo faktury, částka faktury, dodavatel, předmět dodávky).	Ředitel, Vedoucí všech středisek	1	-
11	Filtr nákladů středisek	Možnost filtrovat náklady středisek dle střediska a období.	Ředitel, Vedoucí všech středisek	3	10
12	Vizualizace nákladů středisek	Vizualizace velikosti nákladů jednotlivých středisek vhodným grafem.	Ředitel, Vedoucí všech středisek	3	10
13	Klíčový ukazatel nákladů středisek	Součet nákladů jednotlivých středisek.	Ředitel, Vedoucí všech středisek	2	10
14	Přehled neuhrazených pohledávek	Přehled pohledávek se základními informacemi (číslo vydané faktury, částka pohledávky, datum splatnosti, zákazník).	Ředitel, Vedoucí střediska ekonomiky	1	-
15	Vizualizace neuhrazených pohledávek	Vizualizace podílu neuhrazených pohledávek po splatnosti a před splatností. Znázornění pěti zákazníků s aktuálně největším objemem neuhrazenými pohledávkami.	Ředitel, Vedoucí střediska ekonomiky	2	14
16	Vizualizace faktur vydaných	Vizualizace vývoje čtvrtletních/ročních tržeb.	Ředitel, Vedoucí střediska ekonomiky	1	-
17	Filtr faktur vydaných	Možnost filtrovat faktury vydané dle předmětu projektu a dle období vydání faktury.	Ředitel, Vedoucí střediska ekonomiky	3	16

Tabulka 2 - Katalog požadavků

(Zdroj: interview s vedením společnosti Exxon s.r.o.)

4.4. Návrh řešení

Navrhované řešení spočívá ve vytvoření několika dashboardů ve vybraném nástroji pro vizualizaci dat, které by sloužily managementu jako podpora pro rozhodování a řízení výkonnosti společnosti. Dalším cílem je vybudování centrálního datového úložiště s jednoznačnou interpretací dat a terminologií, které by sloužilo jako jednotný datový zdroj pro vytvořené dashboardy a dále vytvoření automatizovaných procesů (ETL procedur) pro přenos dat ze zdrojových informačních systémů do centrálního datového úložiště, které by každý den do centrálního úložiště přenesly aktuální data. Řešení ve svém důsledku přispěje ke zvýšení konkurenceschopnosti společnosti.

Výhody takového řešení spočívají především v:

- silné podpoře rozhodování managementu,
- centralizaci a vyčištění podnikových dat,
- aktuálnosti údajů v dashboardech.

4.4.1. Součásti řešení

a) Centrální datové úložiště

Centrální datové úložiště bude relační databáze, která bude umožňovat řešit úlohy zaměřené především na analytické dotazování. Vzhledem k tomu, že editace dat již uložených v datovém skladu bude minimální, data budou v datovém skladu ukládána a organizována s ohledem na rychlost čtení, tedy na co nejlepší a nejrychlejší provádění složitých dotazů.

b) ETL procedury

ETL procedury jsou procedury, které zajišťují přenos dat ze zdrojových informačních systémů do centrálního datového skladu. Zdrojové systémy obsahují mnoho dat, které nemusí být nutně zajímavé pro analýzu a reporting, proto ETL procedury budou přenášet do datového skladu pouze ta data, která budou následně v systému zpracovávána a zobrazena pomocí dashboardů.

Během přenosu dat ze zdrojových systémů do datového skladu bude docházet také k vyčištění dat, jejich opravě a transformaci do požadované podoby a na požadovanou úroveň detailu. Data z primárních zdrojových systémů mohou mít a velmi často mají různorodou kvalitu. Během procesu čištění dat bude jejich kvalita zvýšena tak, aby při pozdější analýze nebyla data zatížena velkými chybami. Příčinou špatné kvality dat jsou

primárně uživatelé zdrojových systémů, kteří vědomě či nevědomě zkreslují informace, které předávají ústně a chybují v písemném záznamu informací do zdrojových systémů.

c) **Dashboardy**

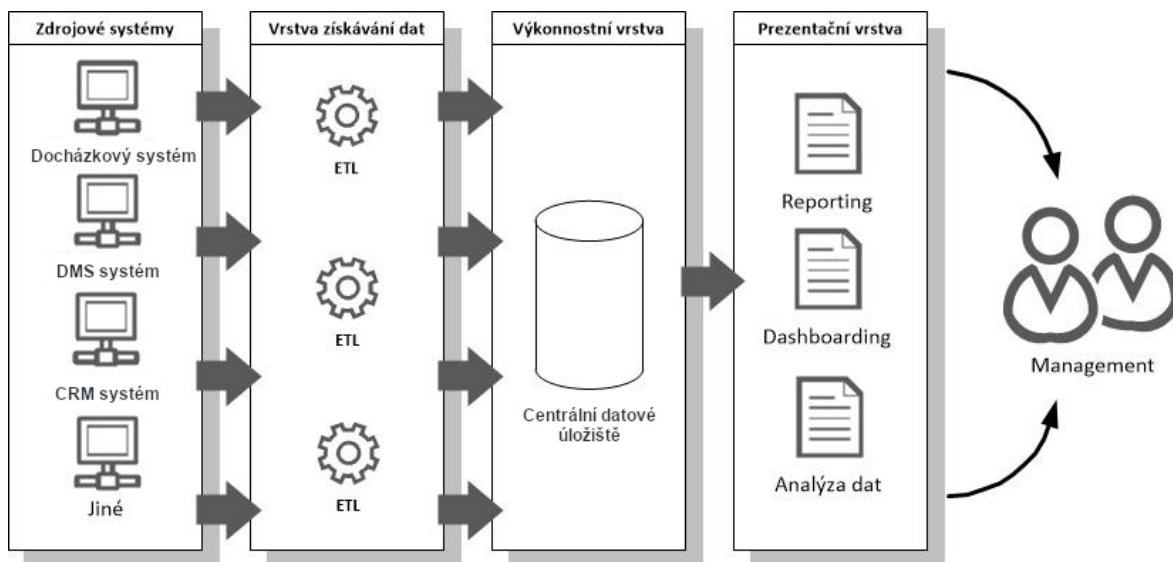
Hlavní součástí řešení budou samotné dashboardy určené managementu pro řízení společnosti. Cílem je vytvoření následujících dashboardů:

- Obchodní příležitosti
- Ziskovost projektů
- Činnosti zaměstnanců
- Náklady středisek
- Neuhrazené pohledávky
- Tržby

4.4.2. **Architektura systému**

Architektura systému bude rozdělena do tří vrstev:

- Vrstva získávání dat (ETL procedury)
- Výkonnostní vrstva (centrální datové úložiště)
- Vrstva uživatelské prezentace (dashboardy)



Obrázek 42 - Architektura řešení

(Zdroj: vlastní zpracování)

4.4.3. **Použité technologie**

Společnost Exxon s.r.o. do značné míry využívá produkty a technologie společnosti Microsoft, proto pro realizaci centrálního datového úložiště byl vybrán databázový systém

Microsoft SQL Server. Jedná se o vysoce výkonný, stabilní a technologicky vyspělý produkt a společnost Exxon s.r.o. již tímto databázovým systémem disponuje a provozuje v rámci něj několik databází. Dále umožňuje snadné zálohování, případnou obnovu dat a rozsáhlé možnosti řízení přístupu k datům pomocí přístupových práv.

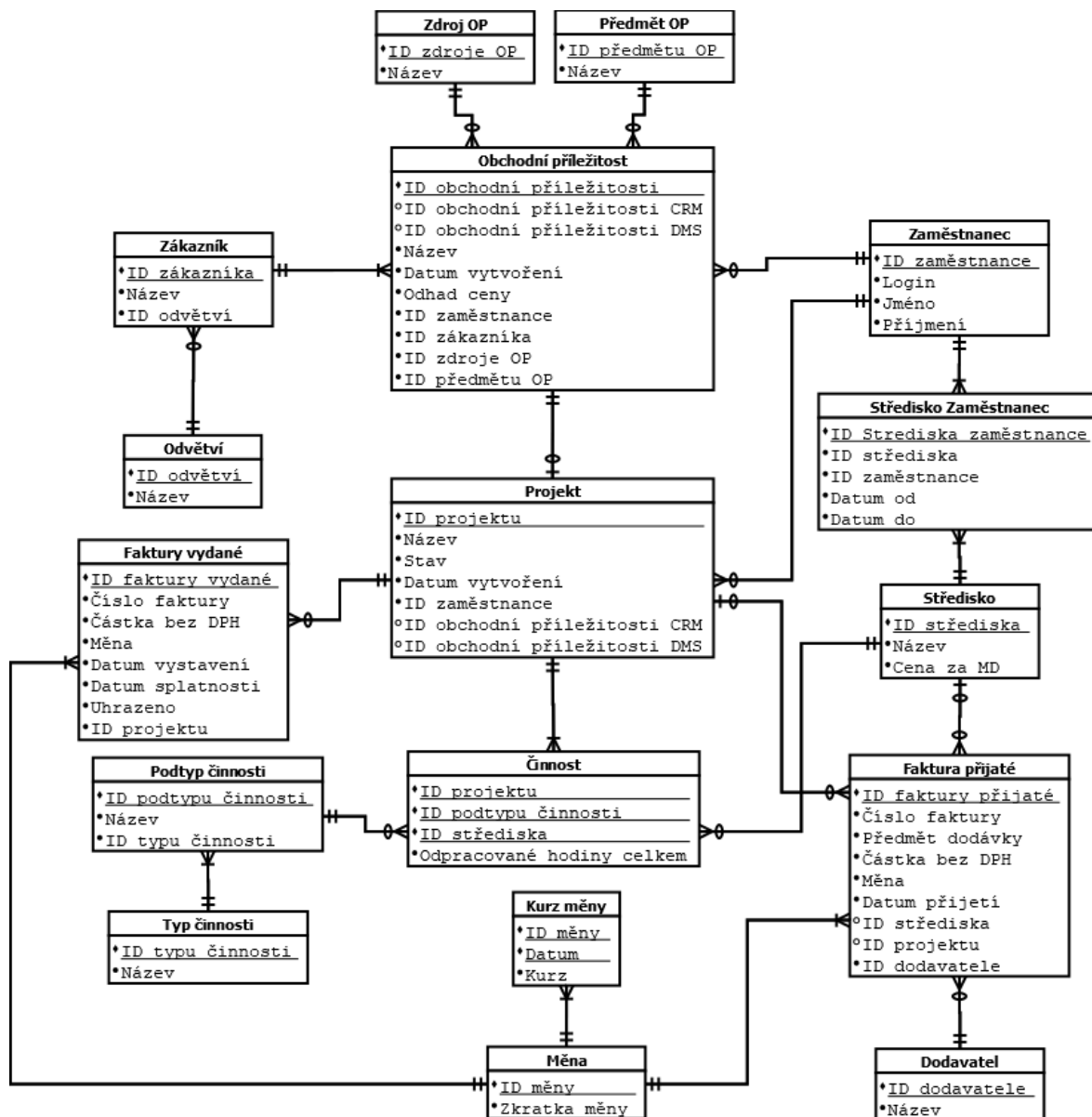
Pro vytvoření ETL procedur byl vybrán programovací jazyk C# a .NET Framework. Jedná se o rozsáhlou softwarovou platformu, která je určena pro vývoj mnoha různých druhů aplikací. Jelikož DMS ELO poskytuje knihovnu připravenou pro platformu .NET, která umožňuje snadné načítání dat přímo ze systému ELO a práci s aplikačním rozhraním tohoto systému. Další možností načítání dat ze systému ELO je přímé načtení dat ze zdrojové databáze daného systému, což .NET a jazyk C# taktéž umožňuje. ETL procedury budou koncipovány jako konzolové aplikace pravidelně spouštěné jednou denně v nočních hodinách pomocí nástroje Task Scheduler, který je součástí systému Microsoft Windows, aby druhý den byly v centrálním úložišti připravena aktuální data.

Vytvoření samotných dashboardů proběhne v nástroji Tableau, jakožto v jednom z nejoblíbenějších nástrojů pro vizualizaci dat používaným v mnoha různých odvětvích. Členů v Tableau komunitě je více než 50 000, tedy lze s touto komunitou řešit případné technické problémy. Dashboardy je možné v nástroji Tableau vytvořit rychle a jednoduše bez programování.

4.5. Realizace centrálního úložiště dat

Dle katalogu požadavků byl vytvořen návrh logické struktury centrálního úložiště dat pomocí E-R modelu. Byly navrženy entity s atributy a vztahy mezi entitami. Při převodu E-R modelu do relačního modelu bylo dbáno na správnou normalizaci databáze, aby byla omezena redundance dat. Následně byl realizován SQL skript pro vytvoření databáze v databázovém systému Microsoft SQL Server.

4.5.1. Návrh E-R diagramu



Obrázek 43 - E-R diagram centrálního úložiště dat

(Zdroj: vlastní zpracování)

4.5.2. Realizace úložiště pomocí SQL skriptu

Kompletní skript v jazyce SQL je uveden v příloze B této práce. V této kapitole je vložena pouze část skriptu pro vytvoření centrální databáze a jedné vybrané tabulky.

Vytvoření centrální databáze:

```
IF NOT EXISTS(SELECT * FROM sys.databases WHERE name = 'CentralDataStorage')
BEGIN
    CREATE DATABASE CentralDataStorage
END
```

Vytvoření tabulky *ObchodniPrilezitost*:

```

IF NOT EXISTS (SELECT * FROM sys.objects WHERE object_id =
OBJECT_ID(N'[dbo].[ObchodniPrilezitost]') AND type IN (N'U'))
BEGIN
    CREATE TABLE ObchodniPrilezitost (
        Id_obchodni_prilezitosti INT PRIMARY KEY IDENTITY(1,1),
        Id_obchodni_prilezitosti_crm INT,
        Id_obchodni_prilezitosti_dms INT,
        Nazev VARCHAR(200) NOT NULL,
        DatumVytvoreni DATE NOT NULL,
        OdhadCeny INT NOT NULL CHECK (OdhadCeny >= 0),
        Id_zamestnanec INT NOT NULL FOREIGN KEY REFERENCES
Zamestnanec(Id_zamestnanec),
        Id_zakaznika INT NOT NULL FOREIGN KEY REFERENCES Zakaznik(Id_zakaznika),
        Id_zdroje_OP INT NOT NULL FOREIGN KEY REFERENCES ZdrojOP(Id_zdroje_OP),
        Id_predmetu_OP INT NOT NULL FOREIGN KEY REFERENCES PredmetOP(Id_predmetu_OP),
        CHECK ((Id_obchodni_prilezitosti_crm IS NOT NULL AND
Id_obchodni_prilezitosti_dms IS NULL) OR (Id_obchodni_prilezitosti_crm IS NULL AND
Id_obchodni_prilezitosti_dms IS NOT NULL))
    )
END

```

4.6. Realizace ETL procedur

Realizace ETL procedur proběhla na platformě Microsoft .NET Framework pomocí programovacího jazyka C#. Jako vývojové prostředí bylo použito Microsoft Visual Studio 2019.

4.6.1. Dokument mapování zdroje na cíl

Vytvořený dokument mapování zdroje na cíl má za úkol popsat, ze kterých zdrojových systémů a která data budou načítána pomocí ETL procedur a do jaké části centrálního datového úložiště budou tato data ukládána. Kompletní dokument mapování zdroje na cíl je uveden v příloze C této práce a obsahuje následující informace:

- Zdroj
 - Systém – zdrojový informační systém, ze kterého jsou daná data načtena
 - Entita – databázová tabulka či jiná zdrojová entita obsahující načítaná data
 - Atribut – zdrojový databázový sloupec či atribut zdrojové entity
 - Datový typ – datový typ načítaných dat
 - NULL – informace, zda načítaná data mohou nabývat neznámé hodnoty ve zdrojovém systému
- Cíl

- Název tabulky – název cílové tabulky v centrálním datovém úložišti, do které budou data uložena
- Název sloupce – název sloupce v dané databázové tabulce centrálního datového úložiště
- Datový typ – datový typ cílového databázového sloupce
- NULL – informace, zda načítaná data mohou v daném sloupci nabývat neznámé hodnoty

4.6.2. Aplikace ETL procedury

Jak již bylo řečeno dříve, pro vytvoření ETL procedury byl zvolen programovací jazyk C# a softwarová platforma Microsoft .NET Framework. Aplikace byla vytvořena jako konzolová aplikace. Zdrojové kódy aplikace jsou přílohou D této práce.

Aplikace je rozdělena do několika následujících částí:

- FirebirdClient – Slouží pro načítání dat ze systému CRM, který používá pro ukládání dat SŘBD Firebird. Z tohoto systému jsou načteny data především o zákaznících a obchodních příležitostech.
- MySqlConnection – Slouží pro načítání dat z Docházkového systému, který používá pro ukládání dat SŘBD MySQL. Z tohoto systému jsou přeneseny informace o zaměstnancích, střediscích a činnostech zaměstnanců, které jsou rozděleny na několik typů a podtypů.
- DMSClient – Slouží pro načítání dat ze systému DMS ELO. V tomto případě nejsou data načítána přímo z databáze daného systému, ale je použito aplikační rozhraní systému DMS ELO a DLL knihovna určená pro účely komunikace s tímto aplikačním rozhraním. Tato DLL knihovna je poskytována výrobcem systému DMS ELO.
- MsSqlClient – Slouží pro ukládání zpracovaných dat do centrálního datového úložiště, které bylo realizováno pomocí SŘBD Microsoft SQL Server.
- ExchangeRatesClient – Slouží pro načítání kurzů měn z API služby XE.com pro převody částek faktur přijatých a vydaných, které jsou v jiné měně, než v Českých korunách.
- App.config – Konfigurační soubor aplikace obsahuje údaje pro připojení do zdrojových informačních systémů a také do centrálního datového úložiště.

Hesla jsou uložena v hashované podobě algoritmem Advanced Encryption Standard (AES).

- CryptClient – Slouží pro načtení hashovaných hesel ze souboru App.config a jejich převedení do dešifrované podoby.
- Program – Hlavní soubor aplikace, který volá jednotlivé funkce pro postupné načítání dat ze zdrojových systémů.

V rámci testování aplikace byl celkový čas běhu aplikace změřen na 30-35 minut. Pro každodenní spouštění aplikace a přenos dat do centrálního datového úložiště bylo zvoleno časové okno ve 2:00 v noci, kdy je malá pravděpodobnost používání systémů zaměstnanci a hromadné načítání a přenos dat negativně neovlivní běh a používání zdrojových systémů. Pravidelné spouštění aplikace bylo nastaveno pomocí nástroje Plánovač úloh (Task Scheduler), který je standardní součástí operačního systému Microsoft Windows.

4.6.3. Načítání kurzů měn

Kurzy měn načítá ETL procedura přes webové API rozhraní služby XE (The World's Trusted Currency Authority), která umožňuje získávání různých druhů informací o kurzech měn. Jedná se především o konverzi z nebo do požadované měny a získávání historických kurzů měn. K API službě je možno přistupovat jednoduše přes URL adresu s příslušnými parametry s uvedením typu požadavku a předání vstupních dat potřebných pro zpracování požadavku.

Převod měny do jedné nebo více cílových měn dle historického kurzu v daném dni lze uskutečnit následujícím požadavkem.

Příklad URL adresy:

https://xecdapi.xe.com/v1/historic_rate.json/?from=USD&date=2011-03-05&to=CAD,JPY&amount=150

Komponenta	Popis
xecdapi.xe.com	Doména, na které je API rozhraní dostupné.
v1	Verze API rozhraní.
historic_rate	Typ požadavku, který se vykoná.
.json/	Formát výsledných dat. Podporovaný je formát .json, .csv nebo .xml.
from=USD	Měna, ze které má být učiněn převod do požadovaných měn, ve formátu ISO 4217.

date=2011-03-05	Datum kurzu měny.
to=CAD,JPY	Seznam měn, do kterých má být převod proveden, ve formátu ISO 4217.
amount=150	Částka, která má být převedena do cílových měn.

Tabulka 3 - Požadavek na získání historického kurzu měny

(Zdroj: https://www.xe.com/xecurrencydata/XE_Currency_Data_API_Specifications.pdf)

Příklad odpovědi na požadavek:

```
Status: 200 OK
X-RateLimit-Limit: 900
X-RateLimit-Remaining: 861
X-RateLimit-Reset: 1350085394
{
  "terms": "http://www.xe.com/legal/dfs.php",
  "privacy": "http://www.xe.com/privacy.php",
  "from": "USD",
  "amount": 1.0,
  "timestamp": "2011-03-05T17:00:00Z",
  "to": [
    {
      "quotecurrency": "CAD",
      "mid": 0.9727929102
    },
    {
      "quotecurrency": "JPY",
      "mid": 82.3000430273
    }
  ]
}
```

4.6.4. Řešené problémy

Během realizace ETL procedur bylo nutné se vypořádat s několika problémy. V první řadě s tím, že zdrojové systémy jsou ve společnosti Exon s.r.o. používány různě dlouhou dobu a data v nich tedy pokrývají různě dlouhá období. Dále realizaci ETL procedur komplikovala skutečnost, že dané informační systémy nejsou po celou dobu používání stejné, průběžně dochází k jejich větším či menším úpravám a mění se tedy i způsob práce s uchovávanými daty a rozsah uchovávaných dat. Dochází pak například k situaci, že v současné podobě informačního systému je vyžadováno, aby vytvářený projekt měl vazbu na dříve vytvořenou obchodní příležitost. V dřívější podobě informačního systému toto vyžadováno nebylo a bylo možné vytvořit projekt bez existence obchodní příležitosti a u dříve vytvořených projektů tedy vazba na obchodní příležitost nemusí existovat. Stejná situace je například u vazby mezi vydanými fakturami a projekty. Třetím problémem je nejednoznačná identifikace stejných informací v různých informačních systémech. Například obchodní příležitost může být založena pouze v CRM

systemu (současný stav používání IS, pokud se obchodní příležitost nerozvine do tvorby nabídky), pouze v DMS systému (historický stav před nasazením CRM systému) nebo v obou systémech (současný stav používání IS, pokud obchodní příležitost dospěje do stádia tvorby nabídky). Pokud neexistuje jednoznačný identifikátor mezi různými informačními systémy, je obtížné tuto vazbu identifikovat.

Zmiňované informační systémy jsou aktuálně ve společnosti Exxon s.r.o. nastaveny tak, aby existovaly jednoznačné identifikátory a vazby mezi uchovávanými daty v různých informačních systémech. Data jsou mezi informačními systémy také v některých případech automaticky přenášena a existuje tak i integrace těchto informačních systémů, ale ne vždy tomu tak bylo. V rámci získání co možná nejobjektivnějších a kompletních informací ze zdrojových systémů, společnost Exxon s.r.o. postupně doplní chybějící vazby všude, kde to bude možné.

4.7. Realizace dashboardů

Následující kapitola popisuje realizaci dashboardů, kdy byly nejdříve vytvořeny drátěné modely dashboardů, následně byly vytvořeny SQL dotazy, které načítají data z centrálního datového úložiště a pomocí nástroje From Data to Viz byly vybrány vhodné grafy použité v dashboardech. Všechny podklady potřebné pro realizaci dashboardů byly v tomto okamžiku připraveny a bylo možné poté přistoupit k realizaci samotných dashboardů.

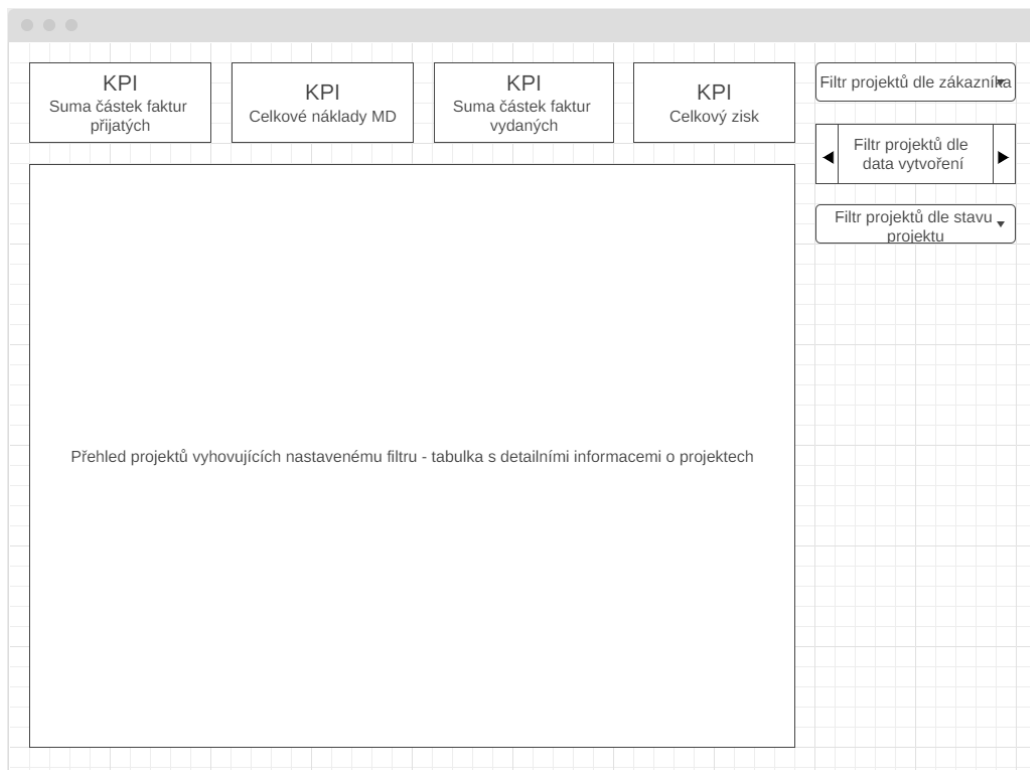
4.7.1. Návrh drátěných modelů dashboardů

Na základě katalogu požadavků byly pro znázornění podoby budoucích dashboardů vytvořeny drátěné modely (wireframe), pomocí kterých si lze snáze představit, jak dashboardy budou vypadat, jakou budou mít strukturu a rozložení prvků, které budou dashboardy obsahovat. Pro vytvoření drátěných modelů byl použit nástroj wireframe.cc.



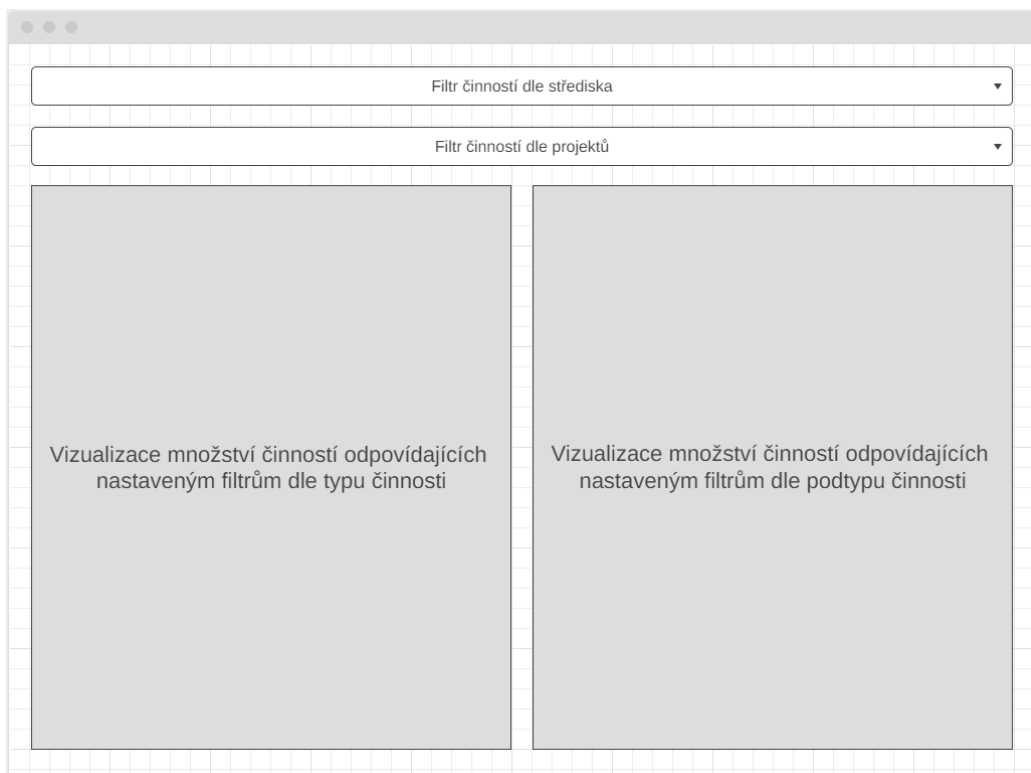
Obrázek 44 - Drátěný model dashboardu obchodních příležitostí

(Zdroj: vlastní zpracování)



Obrázek 45 - Drátěný model dashboardu ziskovosti projektů

(Zdroj: vlastní zpracování)



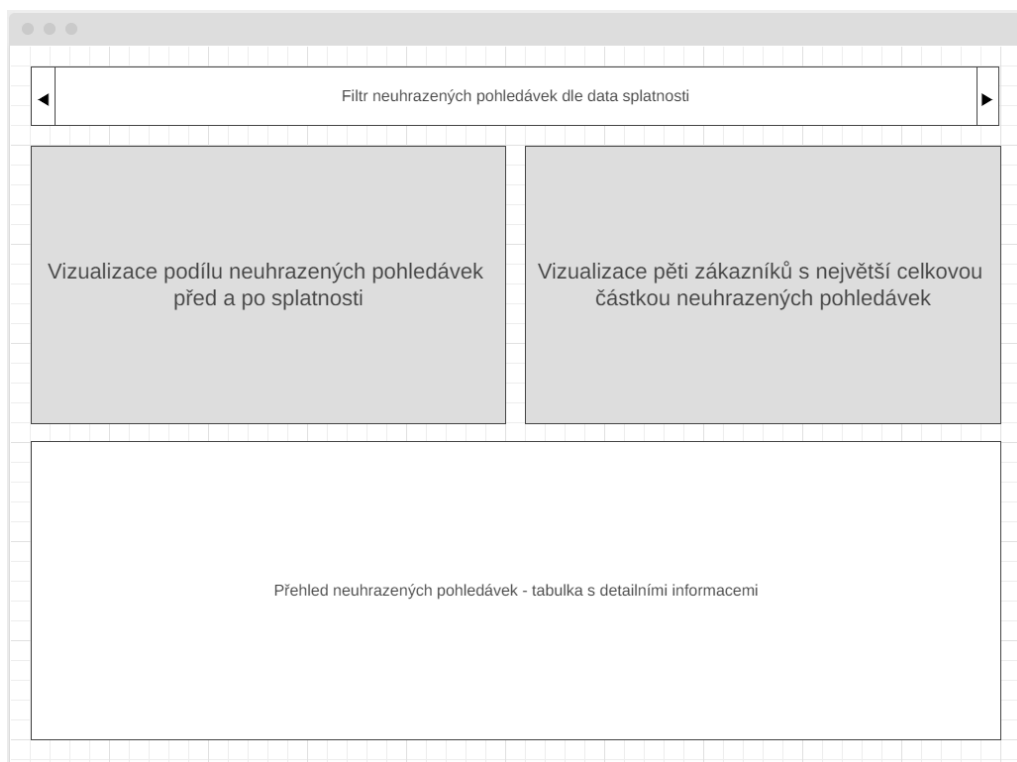
Obrázek 46 - Drátěný model dashboardu činností zaměstnanců

(Zdroj: vlastní zpracování)



Obrázek 47 - Drátěný model dashboardu nákladů středisek

(Zdroj: vlastní zpracování)



Obrázek 48 - Drátěný model dashboardu neuhrazených pohledávek

(Zdroj: vlastní zpracování)



Obrázek 49 - Drátěný model dashboardu tržeb

(Zdroj: vlastní zpracování)

4.7.2. Realizace SQL dotazů

Před samotnou realizací dashboardů bylo nutné vytvořit SQL dotazy, pomocí kterých vizualizační nástroj získá data z centrálního datového úložiště a následně tato data lze dále zpracovat do podoby finálních dashboardů. Pro každý dashboard byl vytvořen jeden SQL dotaz.

- SQL dotaz pro dashboard Obchodní příležitosti

```
SELECT zak.Nazev AS 'Zakaznik'
      ,op.Nazev AS 'NazevOP'
      ,op.DatumVytvoreni
      ,op.OdhadCeny
      ,zop.Nazev AS 'ZdrojOP'
      ,CONCAT(zam.Jmeno, ' ', zam.prijmeni) AS 'Zamestnanec'
      ,pop.Nazev AS 'PredmetOP'
FROM [CentralDataStorage].[dbo].[ObchodniPrilezitest] op
INNER JOIN [CentralDataStorage].[dbo].[ZdrojOP] zop ON op.Id_zdroje_OP =
zop.Id_zdroje_OP
INNER JOIN [CentralDataStorage].[dbo].[Zamestnanec] zam ON op.Id_zamestnance =
zam.Id_zamestnance
INNER JOIN [CentralDataStorage].[dbo].[PredmetOP] pop ON op.Id_predmetu_OP =
pop.Id_predmetu_OP
INNER JOIN [CentralDataStorage].[dbo].[Zakaznik] zak ON op.Id_zakaznika =
zak.Id_zakaznika
```

- SQL dotaz pro dashboard Ziskovost projektů

```
SELECT proj.Id_projektu, proj.Nazev
      ,proj.Stav
      ,proj.DatumVytvoreni
      ,CONCAT(zam.Jmeno, ' ', zam.prijmeni) AS 'Zamestnanec'
      ,(SELECT zak.Nazev FROM [CentralDataStorage].[dbo].[Zakaznik] zak,
[CentralDataStorage].[dbo].[ObchodniPrilezitest] op
      WHERE zak.[Id_zakaznika] = op.Id_zakaznika
      AND COALESCE(op.Id_obchodni_prilezitosti_crm,
op.Id_obchodni_prilezitosti_dms) = COALESCE(proj.Id_obchodni_prilezitosti_crm,
proj.Id_obchodni_prilezitosti_dms)) AS 'Zakaznik'
      ,(SELECT COALESCE(SUM(CASE
      WHEN (SELECT m.Zkratka_meny FROM [CentralDataStorage].[dbo].[Mena] m
WHERE fp.Id_meny = m.Id_meny) = 'CZK' THEN fp.CastkaBezDPH
      ELSE fp.CastkaBezDPH * (SELECT km.Kurz FROM
[CentralDataStorage].[dbo].[Kurz_meny] km WHERE km.Id_meny = fp.Id_meny AND km.Datum
= fp.DatumPrijeti)
      END),0) FROM [CentralDataStorage].[dbo].[FakturaPrijata] fp WHERE
fp.Id_projektu = proj.Id_projektu) AS 'SumaNakladyFP'
      ,(SELECT COALESCE(SUM(CASE
      WHEN (SELECT m.Zkratka_meny FROM [CentralDataStorage].[dbo].[Mena] m
WHERE fv.Id_meny = m.Id_meny) = 'CZK' THEN fv.CastkaBezDPH
      ELSE fv.CastkaBezDPH * (SELECT km.Kurz FROM
[CentralDataStorage].[dbo].[Kurz_meny] km WHERE km.Id_meny = fv.Id_meny AND km.Datum
= fv.DatumVystaveni)
      END),0) FROM [CentralDataStorage].[dbo].[FakturaVydana] fv WHERE
fv.Id_projektu = proj.Id_projektu) AS 'SumaPrijmyFV'
```

```

        ,(SELECT COALESCE(SUM((cin.OdpracovaneHodinyCelkem / 8.0) *
st.CenaZaMD),0)
        FROM [CentralDataStorage].[dbo].[Cinnost] cin,
[CentralDataStorage].[dbo].[Stredisko] st
        WHERE cin.Id_projektu = proj.Id_projektu AND cin.Id_strediska =
st.Id_strediska) AS 'NakladyMD'
        FROM [CentralDataStorage].[dbo].[Projekt] proj
        INNER JOIN [CentralDataStorage].[dbo].[Zamestnanec] zam ON proj.Id_zamestnanec =
zam.Id_zamestnanec
        WHERE COALESCE(proj.Id_obchodni_prilezitosti_crm,
proj.Id_obchodni_prilezitosti_dms) IS NOT NULL

```

- SQL dotaz pro dashboard Činnosti zaměstnanců

```

SELECT cin.OdpracovaneHodinyCelkem
        ,CONCAT((SELECT zak.Nazev FROM [CentralDataStorage].[dbo].[Zakaznik] zak,
[CentralDataStorage].[dbo].[ObchodniPrilezitost] op
        WHERE zak.[Id_zakaznika] = op.Id_zakaznika
        AND COALESCE(op.Id_obchodni_prilezitosti_crm,
op.Id_obchodni_prilezitosti_dms) = COALESCE(proj.Id_obchodni_prilezitosti_crm,
proj.Id_obchodni_prilezitosti_dms)), ' : ', proj.Nazev) AS 'Nazev projektu'
        ,st.Nazev AS 'Stredisko'
        ,(SELECT tc.Nazev FROM [CentralDataStorage].[dbo].[TypCinnosti] tc WHERE
cin.Id_podtypu_cinnosti = pc.Id_podtypu_cinnosti AND pc.Id_typu_cinnosti =
tc.Id_typu_cinnosti) AS 'TypCinnosti'
        ,pc.Nazev AS 'PodtypCinnosti'
FROM [CentralDataStorage].[dbo].[Cinnost] cin
INNER JOIN [CentralDataStorage].[dbo].[Stredisko] st ON cin.Id_strediska =
st.Id_strediska
INNER JOIN [CentralDataStorage].[dbo].[PodtypCinnosti] pc ON cin.Id_podtypu_cinnosti
= pc.Id_podtypu_cinnosti
INNER JOIN [CentralDataStorage].[dbo].[Projekt] proj ON cin.Id_projektu =
proj.Id_projektu
WHERE COALESCE(proj.Id_obchodni_prilezitosti_crm, proj.Id_obchodni_prilezitosti_dms)
IS NOT NULL

```

- SQL dotaz pro dashboard Náklady středisek

```

SELECT fp.CisloFaktury
        ,fp.PredmetDodavky
        ,fp.OznaceniDodavky
        ,CASE
        WHEN (SELECT m.Zkratka_meny FROM [CentralDataStorage].[dbo].[Mena] m WHERE
fp.Id_meny = m.Id_meny) = 'CZK' THEN fp.CastkaBezDPH
        ELSE fp.CastkaBezDPH * (SELECT km.Kurz FROM
[CentralDataStorage].[dbo].[Kurz_meny] km WHERE km.Id_meny = fp.Id_meny AND km.Datum
= fp.DatumPrijeti)
        END AS 'CastkaBezDPHvCZK'
        ,fp.DatumPrijeti
        ,st.Nazev AS 'Stredisko'
        ,dod.Nazev AS 'Dodavatel'
FROM [CentralDataStorage].[dbo].[FakturaPrijata] fp
INNER JOIN [CentralDataStorage].[dbo].[Stredisko] st ON fp.Id_strediska =
st.Id_strediska
INNER JOIN [CentralDataStorage].[dbo].[Dodavatel] dod ON fp.Id_dodavatele =
dod.Id_dodavatele

```

- SQL dotaz pro dashboard Neuhrazené pohledávky

```

SELECT fv.CisloFaktury
      ,CASE
          WHEN (SELECT m.Zkratka_meny FROM [CentralDataStorage].[dbo].[Mena] m
WHERE fv.Id_meny = m.Id_meny) = 'CZK' THEN fv.CastkaBezDPH
          ELSE fv.CastkaBezDPH * (SELECT km.Kurz FROM
[CentralDataStorage].[dbo].[Kurz_meny] km WHERE km.Id_meny = fv.Id_meny AND km.Datum
= fv.DatumVystaveni)
          END AS 'CastkaBezDPHvCZK'
      ,CASE
          WHEN DATEDIFF(day, GETDATE(), fv.DatumSplatnosti) < 0 THEN
              (CASE
                  WHEN (SELECT m.Zkratka_meny FROM [CentralDataStorage].[dbo].[Mena]
m WHERE fv.Id_meny = m.Id_meny) = 'CZK' THEN fv.CastkaBezDPH
                  ELSE fv.CastkaBezDPH * (SELECT km.Kurz FROM
[CentralDataStorage].[dbo].[Kurz_meny] km WHERE km.Id_meny = fv.Id_meny AND km.Datum
= fv.DatumVystaveni)
                  END)
              ELSE 0
          END AS 'PoSplatnosti'
      ,fv.DatumSplatnosti
      ,(SELECT zak.Nazev FROM [CentralDataStorage].[dbo].[Zakaznik] zak,
[CentralDataStorage].[dbo].[ObchodniPrilezitost] op
WHERE zak.[Id_zakaznika] = op.Id_zakaznika
AND COALESCE(op.Id_obchodni_prilezitosti_crm,
op.Id_obchodni_prilezitosti_dms) = COALESCE(proj.Id_obchodni_prilezitosti_crm,
proj.Id_obchodni_prilezitosti_dms)) AS 'Zakaznik'
FROM [CentralDataStorage].[dbo].[FakturaVydana] fv
INNER JOIN [CentralDataStorage].[dbo].[Projekt] proj ON fv.Id_projektu =
proj.Id_projektu
WHERE fv.Uhrazeno = 0
AND COALESCE(proj.Id_obchodni_prilezitosti_crm, proj.Id_obchodni_prilezitosti_dms)
IS NOT NULL

```

- SQL dotaz pro dashboard Tržby

```

SELECT fv.CisloFaktury
      ,CASE
          WHEN (SELECT m.Zkratka_meny FROM [CentralDataStorage].[dbo].[Mena] m
WHERE fv.Id_meny = m.Id_meny) = 'CZK' THEN fv.CastkaBezDPH
          ELSE fv.CastkaBezDPH * (SELECT km.Kurz FROM
[CentralDataStorage].[dbo].[Kurz_meny] km WHERE km.Id_meny = fv.Id_meny AND km.Datum
= fv.DatumVystaveni)
          END AS 'CastkaBezDPHvCZK'
      ,fv.DatumVystaveni
      ,CASE
          WHEN (SELECT COALESCE(proj.Id_obchodni_prilezitosti_crm,
proj.Id_obchodni_prilezitosti_dms) FROM [CentralDataStorage].[dbo].[Projekt] proj
WHERE proj.Id_projektu = fv.Id_projektu) IS NULL THEN 'Nespecifikováno'
          ELSE (SELECT pop.Nazev FROM [CentralDataStorage].[dbo].[PredmetOP]
pop, [CentralDataStorage].[dbo].[ObchodniPrilezitost] op WHERE pop.Id_predmetu_OP =
op.Id_predmetu_OP
AND COALESCE(op.Id_obchodni_prilezitosti_crm,
op.Id_obchodni_prilezitosti_dms) = (SELECT
COALESCE(proj.Id_obchodni_prilezitosti_crm, proj.Id_obchodni_prilezitosti_dms) FROM
[CentralDataStorage].[dbo].[Projekt] proj WHERE proj.Id_projektu = fv.Id_projektu))
          END AS 'PredmetProjektu'
FROM [CentralDataStorage].[dbo].[FakturaVydana] fv

```

4.7.3. Výběr vhodných typů grafů

Výběr vhodných typů grafů, které byly použity při realizaci dashboardů, byl proveden pomocí nástroje From Data to Viz, který je dostupný na adrese www.data-to-viz.com. Grafy byly vybírány tak, aby v rámci realizace dashboardů bylo použito více různých typů grafů.

- Dashboard Obchodní příležitosti

V rámci tohoto dashboardu je realizován jeden graf, který znázorňuje zastoupení jednotlivých obchodních příležitostí dle zdroje. Jedná se tedy o kategorická data jedné proměnné a jako vhodný graf byl vybrán „Treemap“, ve vizualizačním nástroji Tableau nazývaný „Square“, který znázorňuje množinu kategorií pomocí sady sousedících obdélníků o různé velikosti, která je úměrná hodnotě dané kategorie.

- Dashboard Činnosti zaměstnanců

V dashboardu Činnosti zaměstnanců byly realizovány dva grafy, které znázorňují množství odpracovaných činností dle typu a podtypu činnosti. Jedná se tedy o kategorická data o dvou hierarchických úrovních a pro realizaci byly vybrány dva koláčové grafy, v nástroji From Data to Viz i Tableau objevující se pod názvem „Pie“.

- Dashboard Náklady středisek

Dashboard Náklady středisek obsahuje také jeden graf, který znázorňuje velikost nákladů na jednotlivá střediska zaměstnanců. Jedná se tedy taktéž o kategorická data jedné proměnné a jako vhodný graf byl vybrán „Circular Packing“, v nástroji Tableau je tento graf pod názvem „Circle“. Tento graf je obdobou grafu „Treemap“, nicméně kategorie znázorňuje v podobě kruhů o velikosti úměrné hodnotě dané kategorie.

- Dashboard Neuhrazené pohledávky

Dashboard Neuhrazené pohledávky obsahuje dva grafy. První graf má znázorňovat velikost pohledávek, které jsou již po splatnosti a velikost pohledávek zatím před splatností. V tomto případě se tedy jedná o kategorická data jedné kategorie a jako vhodný graf byl vybrán koláčový graf. Druhý graf v tomto dashboardu znázorňuje pět zákazníků s největším množstvím neuhrazených pohledávek. Jako vhodný graf byl vybrán sloupcový graf, v nástroji From Data to Viz zvaný „Barplot“, v nástroji Tableau je dostupný pod názvem „Bar“.

- Dashboard Tržby

Dashboard Tržby obsahuje jeden graf, který znázorňuje vývoj tržeb v čase. Jedná se tedy o data časové řady a jako vhodný graf byl taktéž zvolen sloupcový graf. Hodnota tržeb v daném čase je znázorněna ve formě sloupce o velikosti úměrné velikosti tržeb.

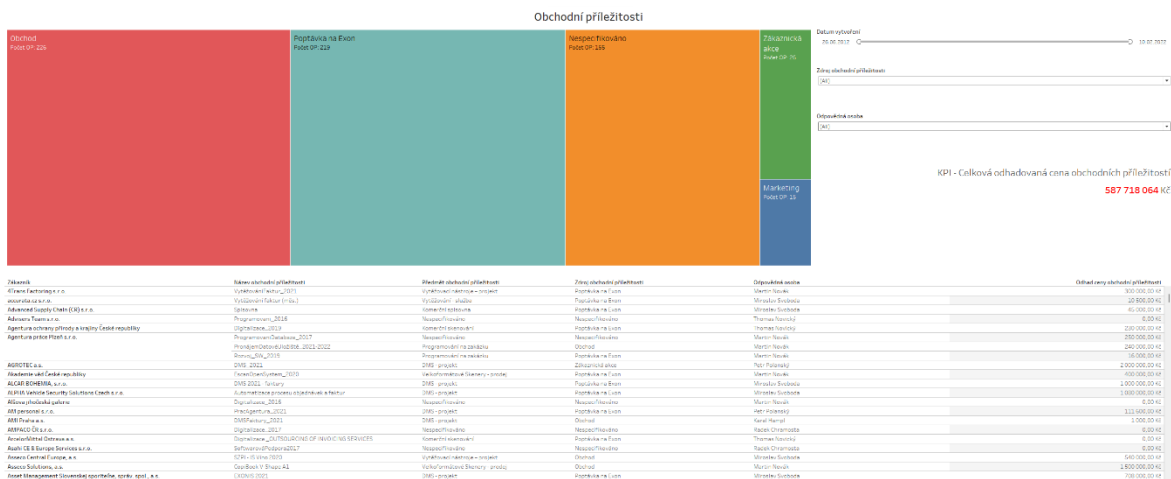
- Dashboard Ziskovost projektů

Tento dashboard neobsahuje žádný graf, pouze data ve formě tabulky a další prvky ve formě KPI ukazatelů a filtrů.

4.7.4. Realizace dashboardů

Po realizaci SQL dotazů již bylo možné přistoupit k tvorbě samotných dashboardů. Pro tvorbu dashboardů byl vybrán vizualizační nástroj Tableau. Dashboardy byly vytvořeny na základě dříve uvedených drátěných modelů.

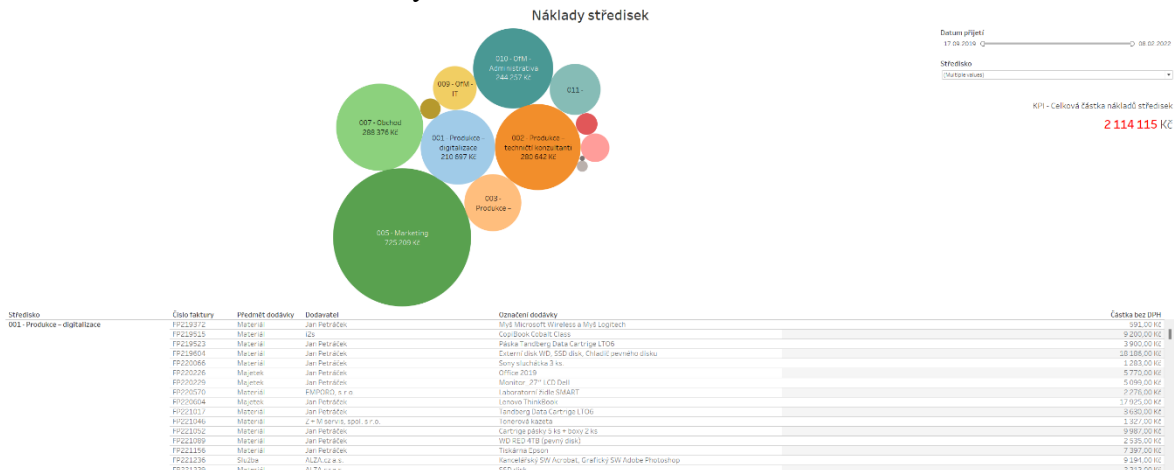
- Dashboard Obchodní příležitosti



Obrázek 50 - Dashboard Obchodní příležitosti

(Zdroj: vlastní zpracování)

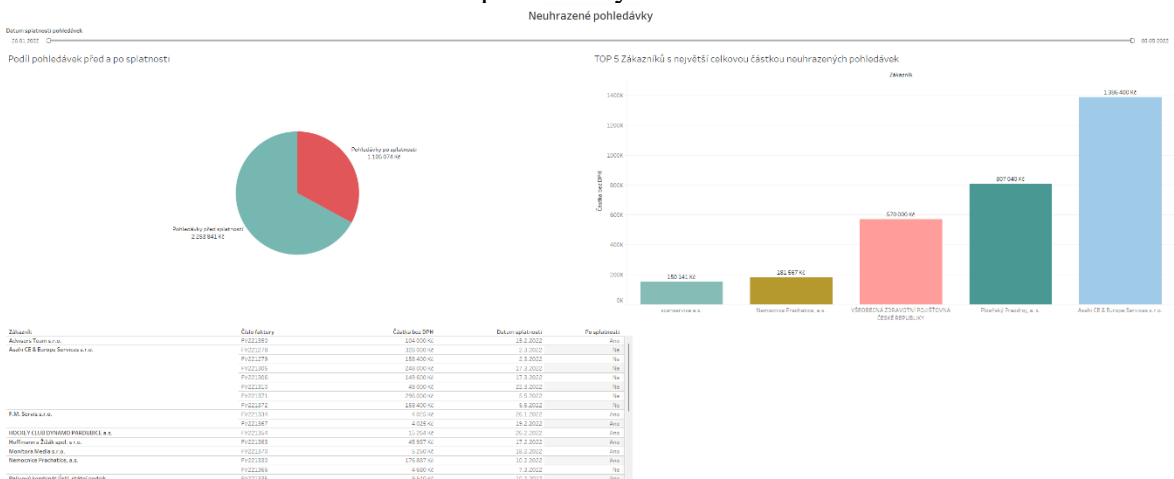
- Dashboard Náklady středisek



Obrázek 53 - Dashboard Náklady středisek

(Zdroj: vlastní zpracování)

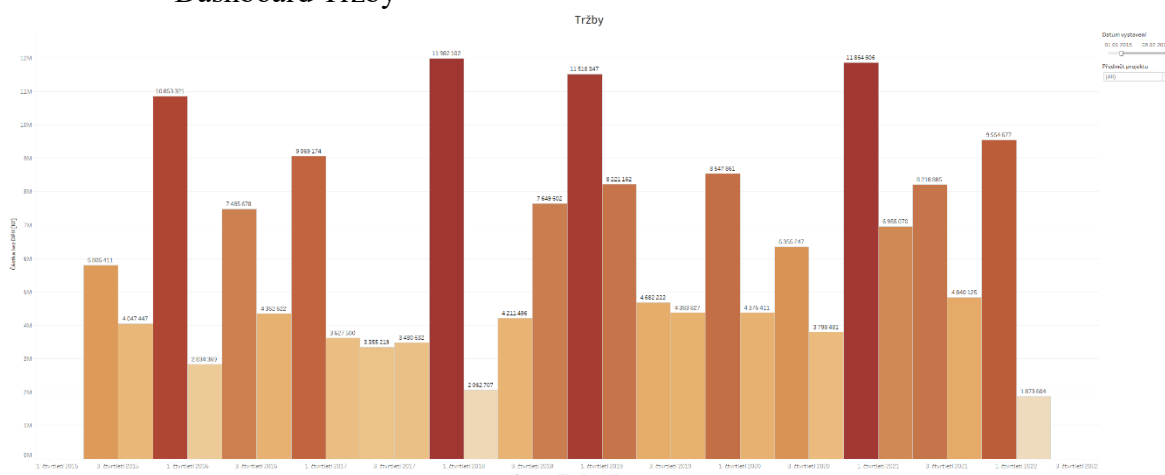
- Dashboard Neuhrazené pohledávky



Obrázek 54 - Dashboard Neuhrazené pohledávky

(Zdroj: vlastní zpracování)

- Dashboard Tržby



Obrázek 55 - Dashboard Tržby

(Zdroj: vlastní zpracování)

4.8. Testování

Testování vytvořených dashboardů mělo za cíl vyhodnotit, zda dashboardy splňují požadavky, které byly stanoveny v katalogu požadavků dle tabulky č. 2. Zainteresané osoby, tzn. vedení a střední management, ze společnosti Exxon s.r.o. se seznámily s vytvořeným řešením a měly vyhrazený prostor o délce jednoho týdne k testování. Toto testování lze označit za akceptační testování, které se provádí pro ověření funkčnosti vytvořeného systému.

Číslo	Název	Popis	Stakeholder	Výsledek
1	Přehled obchodních příležitostí	Přehled obchodních příležitostí se základními informacemi (název zákazníka a obchodní příležitosti, odpovědná osoba, zdroj obchodní příležitosti a odhad ceny/objemu obchodní příležitosti).	Ředitel, Vedoucí obchodu	Splněno – dashboard obsahuje tabulku s přehledem obchodních příležitostí. Obchodní příležitosti v tabulce se aktualizují na základě zvolených filtrů viz. následující požadavek.
2	Filtr obchodních příležitostí	Obchodní příležitosti bude možné filtrovat dle odpovědné osoby (obchodník), zdroje obchodní příležitosti a data vytvoření obchodní příležitosti.	Ředitel, Vedoucí obchodu	Splněno – dashboard obsahuje filtry pro filtrování obchodních příležitostí dle odpovědné osoby, zdroje a data vytvoření obchodní příležitosti.
3	Vizualizace	Vizualizace množství	Ředitel,	Splněno – dashboard

	obchodních příležitostí	obchodních příležitostí dle zdroje pomocí vhodného grafu.	Vedoucí obchodu	obsahuje graf typu „Treemap“. Při změně filtrů dochází také k aktualizaci hodnot grafu.
4	Klíčový ukazatel obchodních příležitostí	Součet odhadované ceny vyfiltrovaných obchodních příležitostí.	Ředitel, Vedoucí obchodu	Splněno – dashboard obsahuje klíčový ukazatel v podobě součtu odhadované ceny obchodních příležitostí. Součet se aktualizuje na základě nastavených filtrů obchodních příležitostí.
5	Přehled projektů	Přehled projektů se základními informacemi (název zákazníka a projektu, stav projektu, náklady projektu v rámci faktur přijatých a odpracovaných MD, výnosy projektu na základě vydaných faktur, zisk).	Ředitel, Projektová kancelář	Splněno – dashboard obsahuje tabulku s přehledem projektů a s požadovanými informacemi. Projekty v tabulce se aktualizují na základě zvolených filtrů viz. následující požadavek.
6	Filtr projektů	Možnost filtrovat projekty dle stavu projektu a roku vytvoření projektu.	Ředitel, Projektová kancelář	Splněno – dashboard obsahuje filtr projektů dle data vytvoření a stavu projektu. Navíc lze projekty filtrovat dle zákazníka.
7	Klíčový ukazatel projektů	Součet částek faktur přijatých, částek faktur vydaných, nákladů MD a zisku dle vyfiltrovaných projektů.	Ředitel, Projektová kancelář	Splněno – dashboard obsahuje čtyři klíčové ukazatele, a to součet částek faktur přijatých a faktur vydaných, náklady MD a výsledný zisk. Tyto ukazatele jsou aktualizovány na základě nastavených filtrů projektů.
8	Vizualizace činností zaměstnanců	Vizualizace množství vykazovaných činností zaměstnanců dle typu a podtypu činností.	Ředitel, Vedoucí všech středisek	Splněno – dashboard obsahuje dva koláčové grafy, které znázorňují množství činností zaměstnanců dle typu a podtypu činností.
9	Filtr činností zaměstnanců	Možnost filtrovat činnosti zaměstnanců podle střediska a dle projektu.	Ředitel, Vedoucí všech středisek	Splněno – dashboard obsahuje filtr činností zaměstnanců dle střediska a projektu. Jako filtr lze použít navíc první koláčový graf, tedy když je vybrána výšeč typu činnosti, druhý

				graf je dle výběru aktualizován a jsou zobrazovány jen podtypy činností odpovídající vybranému typu činností.
10	Přehled nákladů středisek	Přehled nákladů středisek se základními informacemi (středisko, číslo faktury, částka faktury, dodavatel, předmět dodávky).	Ředitel, Vedoucí všech středisek	Splněno – dashboard obsahuje tabulku s náklady jednotlivých středisek. Tabulka je aktualizována dle nastavených filtrů.
11	Filtr nákladů středisek	Možnost filtrovat náklady středisek dle střediska a období.	Ředitel, Vedoucí všech středisek	Splněno – dashboard obsahuje filtr nákladů dle střediska a data přijetí faktury.
12	Vizualizace nákladů středisek	Vizualizace velikosti nákladů jednotlivých středisek vhodným grafem.	Ředitel, Vedoucí všech středisek	Splněno – dashboard obsahuje graf typu „Circular Packing“. Při změně filtrů dochází také k aktualizaci hodnot grafu.
13	Klíčový ukazatel nákladů středisek	Součet nákladů jednotlivých středisek.	Ředitel, Vedoucí všech středisek	Splněno – dashboard obsahuje klíčový ukazatel ve formě součtu nákladů středisek, který se aktualizuje na základě nastavených filtrů.
14	Přehled neuhrazených pohledávek	Přehled pohledávek se základními informacemi (číslo vydané faktury, částka pohledávky, datum splatnosti, zákazník).	Ředitel, Vedoucí střediska ekonomiky	Splněno – dashboard obsahuje tabulku s přehledem neuhrazených pohledávek s požadovanými informacemi. Přehled je aktualizován na základě nastavených filtrů.
15	Vizualizace neuhrazených pohledávek	Vizualizace podílu neuhrazených pohledávek po splatnosti a před splatností. Znárodnění pěti zákazníků s aktuálně největším objemem neuhrazenými pohledávkami.	Ředitel, Vedoucí střediska ekonomiky	Splněno – dashboard obsahuje dva grafy. První graf je koláčový graf znázorňující podíl neuhrazených pohledávek před splatností a po splatností. Druhý graf je sloupcový graf znázorňující pět zákazníků s největším součtem částek neuhrazených pohledávek.
16	Vizualizace faktur vydaných	Vizualizace vývoje čtvrtletních/ročních tržeb.	Ředitel, Vedoucí střediska ekonomiky	Splněno – dashboard obsahuje graf znázorňující vývoj tržeb. Jako graf byl použit sloupcový graf. Graf

				je aktualizován na základě nastavených filtrů.
17	Filtr faktur vydaných	Možnost filtrovat faktury vydané dle předmětu projektu a dle období vydání faktury.	Ředitel, Vedoucí střediska ekonomiky	Splněno – dashboard obsahuje filtry pro filtrování tržeb dle období vystavení faktury a dle předmětu projektu.

Tabulka 4 - Ověření požadavků dle katalogu požadavků

(Zdroj: akceptační testování)

Dle uvedeného vyplývá, že požadavky stanovené v rámci katalogu požadavků byly splněny.

5. Výsledky a diskuse

Tato kapitola je věnována v první řadě zobecněním přínosů navrženého řešení, aspektům znemožňující zúročení navrženého řešení a také možnostem potenciálního rozvoje řešení.

5.1. Zobecnění přínosů navrženého řešení

Navržené řešení není pouze software, ale jedná se o ucelenou koncepci, jejímž cílem je využití dat v každodenních procesech podniku prostřednictvím vizualizace dat a dashboardů, které jsou nedílnou součástí Business Intelligence. Hlavní přínosy takového řešení lze zobecnit do několika následujících bodů.

- Centralizace dat do jednoho úložiště

Pro učinění správných rozhodnutí jsou většinou potřeba data z různých zdrojových systémů. Díky centralizaci dat do jednoho úložiště dochází ke zvýšení integrity těchto dat, vyčištění dat a redukci duplicit a data mohou být analyzována jako jeden celek. Centrální datové úložiště má jednoznačnou interpretaci dat a terminologii. Pokud je správně nastavená entitní a referenční integrita, mohou být během načítání dat identifikovány problémy s daty přicházejícími do centrálního úložiště. Tyto problémy mohou být dále opraveny přímo ve zdrojových systémech, což v budoucnu napomůže k lepším výsledkům při analýze dat.

- Úspora času

Pomocí automatické aktualizace dat v centrálním datovém úložišti prostřednictvím ETL procedur a dashboardů napojených na toto centrální úložiště se lze vyhnout manuální tvorbě reportů z transakčních systémů a jejich dalšímu zpracování do požadované podoby. Díky tomu dochází k úspoře času, jelikož uživatelé mají k dispozici každý den aktuální informace bez dalších manuálních zásahů či čekání na zpracování reportu jinou osobou. Ušetřený čas pak lze věnovat analýze dat nebo jiným činnostem či projektům.

- Zvýšení efektivity rozhodovacího procesu

Účelem BI dashboardu je odpovědět na dvě hlavní otázky a to „Jak dobře si podnik vede?“ a „Co má podnik dělat dál?“. Tyto otázky se mohou týkat podniku jako celku nebo jeho jednotlivých oddělení. Po druhé zmiňované otázce musí přijít rozhodnutí, ke kterému by měl pomoci dashboard. Jeho účelem je právě pomoci k rozhodnutí na základě

poskytnutých informací. Díky možnosti snadno určit, co data skutečně znamenají, lze přijímat lepší rozhodnutí relevantní pro podnik a k dosahování jeho cílů.

- **Predikce trendů**

Dashboards by měly poskytovat pohled na minulý i současný vývoj vybraných veličin, aby vedení podniku mohlo předvídat, jak bude vývoj vypadat v budoucnu a případně se mohl podnik pokusit změnit svůj směr. Znalost trendů znamená, že podnik je schopen efektivněji plánovat na další období, provádět změny svých procesů a daný trend pozitivně ovlivnit, pracovat na zlepšení svých výsledků nebo je možno například provést preventivní kroky proti vzniku různých problémů.

- **Zaměření na klíčové ukazatele výkonnosti**

Vytvořené dashboards poskytují rychlý přehled o nejdůležitějších podnikových procesech a obsažené informace mají být výstižné a okamžitě viditelné. Umožňují rychlý přístup ke klíčovým ukazatelům výkonnosti (KPI), což jsou klíčové metriky pro hodnocení podnikového procesu, které jsou v rámci dashboardu zasazeny do vhodného kontextu. Pomocí těchto ukazatelů si podnik také může stanovit cíle a pak hodnotit jejich dosažení ve stanoveném období. Bývá obtížné udržet všechna oddělení podniku v souladu a v práci na společných cílech. Klíčové ukazatele výkonnosti mohou být prostředek, jak složitě popsané cíle převést do srozumitelných metrik poskytujících zpětnou vazbu o vývoji podniku.

5.2.Aspekty znemožňující zúročení navrženého řešení

Hlavním aspektem, na kterém závisí úspěch či neúspěch každého BI řešení, jsou data. Dostupnost dat v požadovaném rozsahu a dostatečné kvalitě je klíčová, nicméně kvalita dat představuje významný problém a zároveň bývá velkou výzvou pro všechny firmy. Špatná kvalita dat, může v samotném důsledku vést až ke ztrátě zákazníka, protože má vliv na chybné nebo pomalé rozhodovací procesy a zhoršení image podniku. Pokud podnik nemá k dispozici data pro vytvoření dashboardu, který by vedení podniku potřebovalo pro efektivnější rozhodování a řízení, nelze takový dashboard vytvořit. Jediným řešením pak je, že podnik začne daná data zaznamenávat. Dashboard lze poté vytvořit, nicméně až s odstupem několika měsíců či dokonce roků, kdy budou dostupná data za dostatečně dlouhé období, bude mít dashboard vypovídající hodnotu o vývoji dané veličiny.

V kapitole 4.6.4. byly popsány některé problémy, které bylo nutné řešit v rámci realizace ETL procedury, které taktéž souvisí s kvalitou dat ve zdrojových systémech. Dalšími problémy mohou spočívat například v nejasných a nejednotných číselnících používaných v transakčních systémech, kterých může být v každém takovém systému celá řada. Tyto číselníky navíc nemusí být statické, ale často bývají dynamicky upravovány, pokud v nich nějaká položka chybí. ETL proces pak musí počítat například se situací, kdy jednomu typu zboží odpovídá více položek v daném číselníku a měl by danou situaci nějak řešit. Dalšími problémy v datech mohou být chybějící hodnoty, překlepy uživatelů, duplicitní záznamy či nesprávné konverze datových typů.

Kvůli špatné kvalitě dat může dojít k tomu, že dashboardy v navrženém řešení nebudou mít vypovídající hodnotu, nebudou poskytovat dostatečné podklady pro rozhodování a uživatelé systému ztratí v navržené řešení důvěru a přestanou jej používat. Způsob, jak tomuto zamezit, spočívá především v neustálém procesu zlepšování kvality dat ve zdrojových systémech, aby kvalita dat byla na dostatečné úrovni pro podporu informačních potřeb podniku a dále spočívá také v silné podpoře řešení u vrcholového managementu podniku.

Další aspekt znemožňující zúročení navrženého řešení může spočívat v nedostatečné technické podpoře. Pokud se v rámci navrženého řešení vyskytne technický problém nebo požadavek na úpravu, ať už v ETL procesu nebo v samotných dashboardech, je nutné mít k dispozici odpovědnou osobu se znalostmi technologií a nástrojů, které byly pro řešení použity. Stejný problém pak může vzniknout v případě odchodu této odpovědné osoby z daného podniku.

5.3. Možnosti dalšího rozvoje

Navržené řešení má velký potenciál pro další rozvoj, a to především v rozšiřování možností vytvořených dashboardů a realizace nových, zatím nevytvořených dashboardů. Jediné omezení je kladeno na existenci dostatečně kvalitních dat ve zdrojových systémech. Na základě datových zdrojů, se kterými navržené řešení pracuje, by bylo možné dále vytvořit například následující dashboardy.

- Přehled provizí obchodníků

Na základě vydaných faktur získávají obchodníci určitou část provize, pokud daný projekt vzešel z obchodní příležitosti, kterou měl daný obchodník na starost. Přehled

provizí obchodníků je aktuálně tvořen každý měsíc manuálně ekonomickým oddělením společnosti, nicméně dashboard vytvořený pro tyto účely by byl vhodným způsobem, jak tento proces automatizovat a zefektivnit.

- Přehled čerpání dovolených

Vývoj čerpání dovolené zaměstnanců v daném roce sleduje osoba odpovědná za personalistiku, nicméně čerpání dovolené je nutné pro každého zaměstnance zjistit jednotlivě. V případě několika desítek zaměstnanců se tak jedná o zbytečně zdlouhavý proces. Podkladová data jsou uložena v Docházkovém systému, ze kterého jsou již data pomocí ETL procedury načítána, proto je tento přehled dalším vhodným námětem pro nový dashboard.

- Vývoj počtu zaměstnanců a brigádníků

Jedním ze zajímavých ukazatelů, které také mohou vypovídat o vývoji společnosti, je vývoj počtu zaměstnanců a brigádníků, kteří jsou v případě potřeby přijímáni. Podkladová data by bylo možné získat opět v Docházkovém systému, který používá každý zaměstnanec a brigádník pro vykazování své docházky a činností a má tedy v tomto systému uživatelský účet.

- Dashboardy finanční analýzy

V rámci řešení by bylo možné také vytvořit několik různých dashboardů pro potřeby finanční analýzy společnosti, které by mohly zobrazovat klíčové ukazatele, jako jsou ROA, ROE, EVA a podobně. Pro potřeby takových dashboardů by bylo pravděpodobně nutné načítat data z účetního systému, nicméně účetní služby jsou aktuálně ve společnosti Exxon s.r.o. outsourcovány a společnost nemá vlastní systém pro data potřebná pro výpočet těchto ukazatelů.

Jako další potenciální rozvoj řešení lze zmínit možnou úpravu aplikace, která zajišťuje ETL proces. V současné chvíli aplikace automaticky neposkytuje žádnou zpětnou vazbu o tom, kolik dat bylo ze zdrojových systémů načteno a přeneseno do centrálního úložiště a kolik dat a jakých naopak přeneseno nebylo kvůli různým problémům s chybějícími údaji či chybné referenční integritě a podobně. Aplikace by bylo možné upravit tak, aby při každém spuštění vytvořila report o nepřenesených datech a tento report odeslala například e-mailem odpovědnému zaměstnanci, který by o případných problémech byl ihned informován a mohl je řešit. Nejlepším způsobem, jak poté chybu opravit, je přímá oprava dat v systému, kde se chyba nachází.

V teoretické části práce byla popsána mimo jiné technika Data Storytelling. V praktické části práce tato technika nebyla využita, nicméně může se jednat taktéž o způsob, jakým navržené řešení lze dále rozvíjet. Vizualizační nástroj Tableau, ve kterém byly dashboardy vytvořeny, tuto techniku podporuje prostřednictvím spojení několika souvisejících dashboardů do sekvence vizualizací, pomocí které lze příběh uživateli prezentovat.

6. Závěr

S rostoucím množstvím dat, která jsou moderní informační společností vytvářena a ukládána, roste i potřeba tato data dále efektivně zpracovat, vizualizovat a vytěžit z nich co nejvíce informací, která lze dále využít pro potřeby jednotlivce, podniku či společnosti obecně. Dá se říci, že tato potřeba bude s rostoucím množstvím dat dále stoupat a růst na významu.

Aby mohl být podnik správně a efektivně řízen a mohl dosahovat svých cílů, musí mít vedení podniku, na všech úrovních řízení, k dispozici správné informace ve správný čas. Tato skutečnost je klíčová pro realizaci správných rozhodnutí, zvláště v turbulentním tržním prostředí, kde se podmínky mohou rychle měnit. Jak bylo již dříve v této práci řečeno, vizualizace dat je metodou, která napomáhá člověku ke zpracování většího množství informací a lepšímu pochopení souvislostí, než pokud jsou data například ve formě tabulky.

Vizualizace dat je vhodnou metodou nejen jako podpora řízení podniku bez ohledu na sektor podnikání či velikost podniku, ale jedná se i o účinnou metodu zprostředkování informací publiku například ve školství, zpravodajství a mnoha dalších oblastech.

Hlavním cílem diplomové práce byla realizace několika dashboardů pro konkrétní podnik, které vedení podniku využije pro svou potřebu rozhodování a řízení podniku. Tohoto cíle bylo dosaženo v rámci praktické části této práce, kde bylo realizováno centrální datové úložiště pro data získaná ze zdrojových informačních systémů prostřednictvím vytvořené aplikace ETL procedury. Tato data pak poskytla datovou základnu pro vytvoření dashboardů. Dalšími cíli práce bylo v její teoretické části vymezit historii a principy vizualizace dat, popsat typy grafů a dalších metod vizualizace, jejich vhodné uplatnění během vizualizace dat a popsat použití dashboardů pro potřeby rozhodování a řízení podniku. Dalším cílem bylo popsat metodu data storytelling jakožto efektivní metodu v oblasti vizualizace dat a její výhody a zmapovat používané nástroje pro vizualizaci dat, které byly v rámci práce rozděleny na dvě kategorie, a to programovací jazyky poskytující prostředky pro vizualizaci dat a druhou kategorií byl hotový software, který je možné použít pro vizualizaci dat bez znalosti programování.

Metodika dosažení těchto cílů byla v první řadě ve studiu informačních zdrojů uvedených v kapitole 7. V praktické části práce bylo postupováno dle následujících kroků:

1. zmapování současného stavu
 - popis používaných informačních systémů společnosti
 - popis možností pro vytvoření reportu, které používané informační systémy poskytují
 - popis aktuálního způsobu vytváření reportů pro potřeby vedení podniku
2. návrh řešení
 - návrh architektury řešení
 - výběr technologií a nástrojů pro realizaci řešení
3. vytvoření katalogu požadavků
 - interview s vedením podniku
 - vytvoření katalogu požadavků na dashboardy
4. centrální úložiště dat
 - návrh struktury úložiště pomocí ER modelu, který byl vytvořen pomocí Crow's foot notace
 - převod ER modelu na návrh relační databáze, normalizace databáze, návrh integritních omezení
 - realizace skriptu pro vytvoření centrálního úložiště dat pomocí jazyka SQL pro SŘBD Microsoft SQL Server
5. aplikace ETL procedury
 - vytvoření dokumentu mapování zdroje na cíl
 - dekompozice problému na jednotlivé dílčí části/moduly
 - realizace aplikace ET procedury na platformě .NET
 - výběr časového okna pro spouštění ETL procedury a nastavení jejího spouštění
6. realizace dashboardů
 - vytvoření drátěných modelů dashboardů
 - vytvoření SQL dotazů pro načítání dat z centrálního datového úložiště
 - realizace dashboard v nástroji Tableau
7. ověření požadavků
 - seznámení vedení s dashboardy
 - akceptační testování a ověření splnění požadavků

Praktická část práce tedy spočívala v návrhu a realizaci řešení pro podporu rozhodování a řízení podniku Exxon s.r.o. Řešení bylo rozděleno do tří samostatných vrstev/částí:

1. Vrstva získávání dat, která zajišťuje přenos dat ze zdrojových informačních systémů do centrálního datového úložiště.
2. Výkonnostní vrstva, jejíž součástí je centrální datové úložiště.
3. Prezentační vrstva, která obsahuje dashboardy určené managementu podniku jako podpora pro rozhodování.

Pro realizaci centrálního datového úložiště byl vybrán databázový systém Microsoft SQL Server, jelikož společnost Exxon s.r.o. tento systém již vlastní a využívá i pro jiné systémy. Veškeré skripty v jazyce SQL, které byly v rámci této práce realizovány, byly tedy vytvořeny pro tuto databázovou platformu. Pro vytvoření aplikace ETL procedury byla vybrána softwarová platforma Microsoft .NET Framework a programovací jazyk C#, jelikož pro načítání dat ze zdrojového systému DMS ELO byla použita knihovna vytvořená výrobcem tohoto systému, která je připravena právě pro použití v tomto programovacím jazyce. Kvůli zjednodušení problému ETL aplikace byla aplikace dekomponována na několik samostatných částí. Dashboardy byly realizovány ve vizualizačním nástroji Tableau, jakožto v jednom z nejpoužívanějších nástrojů pro vizualizaci dat s širokou členskou základnou. Samotné dashboardy bylo možné v tomto nástroji vytvořit bez nutnosti programování. Případná změna technologií použitých v jednotlivých vrstvách by znamenala především změnu jen v dané vrstvě. Například, pokud by se společnost Exxon s.r.o. rozhodla, že dashboardy vytvoří znova v jiném vizualizačním nástroji, centrálního datového úložiště ani aplikace ETL procedury by se tato změna nijak nedotkla.

Způsob, jakým společnost Exxon s.r.o. získávala informace ze svých informačních systémů, spočíval v manuální tvorbě reportů ve formě tabulek ve formátu MS Excel nebo CSV. Tyto reporty si uživatel informačního systému buď musel vytvořit sám, nebo musel být report zpracován prostřednictvím odpovědné osoby z oddělení IT, která vytvořila report dle požadavků vedení společnosti. Tento způsob přinášel mnoho nevýhod v podobě časové náročnosti pro tvorbu reportu, nepochopení požadavků ze strany IT oddělení, report byl vytvořen jen z dat v rámci jednoho informačního systému a hlavní nevýhoda spočívala v samotné podobě reportu, kde data byla zpracována ve formě tabulek, ze kterých je složité vyčíst ty podstatné informace a souvislosti nebo vyzorovat trend zkoumané veličiny.

Realizované řešení tak přináší společnosti Exxon s.r.o. oproti původnímu stavu značný posun v oblasti informační podpory pro rozhodování a řízení. Dashboardy může management společnosti použít kdykoli pro své potřeby bez nutnosti dalšího manuálního zásahu do dashboardů. Data jsou automaticky každý den v centrálním datovém úložišti aktualizována, dashboardy tedy obsahují aktuální informace nutné pro realizaci efektivních rozhodovacích procesů a v konečném důsledku to společnosti Exxon s.r.o. může přinést zvýšení konkurenceschopnosti a zlepšení pozice na trhu.

7. Seznam použitých zdrojů

BRINCH, Sara. Charles-Joseph Minard's map of Napoleon's flawed Russian campaign: An ever-current classic. *Science Norway* [online]. 2019 [cit. 2021-9-1]. Dostupné z: <https://sciencenorway.no/blog-blog-from-numbers-to-graphics-statistics/charles-joseph-minards-map-of-napoleons-flawed-russian-campaign-an-ever-current-classic/1618695>

DUNDAS BI. *A Brief History of Data Visualization* [online]. 2019 [cit. 2021-9-1]. Dostupné z: <https://www.dundas.com/resources/blogs/introduction-to-business-intelligence/brief-history-data-visualization>

ENGETO - DATOVÁ AKADEMIE. *Databáze: Úvod do databází*. 2019. Dostupné také z: <https://engeto.com/cs/kurz/databaze/studium/lDNA1XZkQ1633zrUelqlgQ/uvod-do-databazi>

ENGETO - DATOVÁ AKADEMIE. *Základy vizualizace dat: Síla vizualizací*. 2020. Dostupné také z: https://engeto.com/cs/kurz/zaklady-vizualizace-dat/studium/6RKej_bWTGKWAZinwbztog/sila-vizualizaci

FINERREPORT. Compare 6 Types and 14 Data Visualization Tools. *FineReport* [online]. 2018 [cit. 2021-11-18]. Dostupné z: https://www.finereport.com/en/data-visualization/compare-6-types-and-14-data-visualization-tools.html#2_Visual_Reporting_or_Blnbsp

FORTNEY, Kendall. *5 Ways Data Visualizations can Lie* [online]. 2017 [cit. 2021-9-1]. Dostupné z: <https://towardsdatascience.com/5-ways-data-visualizations-can-lie-46e54f41de37>

FRIENDLY, Michael. Effect ordering for data displays. *Comput Stat Data Anal. ResearchGate* [online]. 2003 [cit. 2021-09-01]. Dostupné z: https://www.researchgate.net/figure/Langrens-1644-graph-of-determinations-of-the-distance-in-longitude-from-Toledo-to_fig14_4898061

HENDL, Jan. *Statistika v aplikacích*. Praha: Portál, 2014. ISBN 978-80-262-0700-9.

HEREL, Marek. *Komplexní návod: Jak na Google Data Studio* [online]. 2020 [cit. 2021-11-18]. Dostupné z: <https://blog.h1.cz/aktualne/jak-na-google-data-studio>

CHIQUI, Esteban. *A Quick Guide to Spotting Graphics That Lie* [online]. 2015 [cit. 2021-9-1]. Dostupné z: <https://www.nationalgeographic.com/science/article/150619-data-points-five-ways-to-lie-with-charts>

JARKOVSKÝ, Jiří, Ladislav DUŠEK a Jiří KALINA. *Základní typy dat* [online]. 2017 [cit. 2021-10-08]. Dostupné z: https://is.muni.cz/el/sci/jaro2017/Bi7541/um/02_zakladni_typy_dat.pdf. Vytvořil Institut biostatistiky a analýz, Masarykova univerzita.

KESAVAN, Vidya. The Role of Visual Perception in Data Visualization. *Daydreaming Numbers* [online]. 2016 [cit. 2022-01-20]. Dostupné z: <http://daydreamingnumbers.com/blog/visual-perception-data-visualization/>

KNAFLIC, Cole Nussbaumer. *Storytelling With Data: A Data Visualization Guide for Business Professionals*. New Jersey: John Wiley & Sons, 2015. ISBN 9781119002253.

KOMENDA, Martin a Matěj KAROLYI. *A Survey on Data-driven Decision Support Systems Using Effective Narrative Pattern* [online]. Brno, 2016 [cit. 2021-11-14]. Dostupné z: https://www.researchgate.net/publication/305618042_A_Survey_on_Data-driven_Decision_Support_Systems_Using_Effective_Narrative_Pattern
KROENKE, David a David AUER. *Databáze*. Brno: Computer Press, 2015. ISBN 978-802-5143-520.

NC STATE UNIVERSITY. *Graph Types* [online]. 2005 [cit. 2021-9-1]. Dostupné z: <https://projects.ncsu.edu/labwrite//res/gh/gh-graphtype.html>

NIZAM, Alexa. 67 Video Marketing Stats You Need to Know for 2022 (Including Original Research!). *Lemonlight* [online]. 2021 [cit. 2022-02-16]. Dostupné z: <https://www.lemonlight.com/blog/67-video-marketing-stats-you-need-to-know-for-2022/>

PARAGRAPH. *The chart-junk of Steve Jobs* [online]. 2008 [cit. 2021-9-1]. Dostupné z: <https://paragraft.wordpress.com/2008/06/03/the-chart-junk-of-steve-jobs/>

PLÍVA, Michal. *Power BI* [online]. 2018 [cit. 2021-11-20]. Dostupné z: <https://ci.vse.cz/office365/dalsi-webove-aplikace/power-bi/>

ROGERS, Simon. Florence Nightingale, datajournalist: information has always been beautiful. *The Guardian* [online]. 2010 [cit. 2021-9-1]. Dostupné z: <https://www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics>

ROGERS, Simon. John Snow's data journalism: the cholera map that changed the world. *The Guardian* [online]. 2013 [cit. 2021-9-1]. Dostupné z: <https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map>

RUD, Olivia Parr. *Data Mining: praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníků (CRM)*. Praha: Computer Press, 2001. ISBN 80-722-6577-6.

SLEEPER, Rayan. *Practical Tableau: 100 Tips, Tutorials, and Strategies from a Tableau Zen Master*. Sevastopol: O'Reilly Media, 2018. ISBN 978-1-491-97731-6.

SRIDHARAN, Mithun. Data Visualization: History and origins. *Think Insights* [online]. 2017 [cit. 2021-09-02]. Dostupné z: <https://thinkinsights.net/digital/data-visualization-history/>

STEPANOV, Roman. The Types of Statistical Maps and Great Examples to See. *MAPSVG* [online]. 2021 [cit. 2021-9-1]. Dostupné z: <https://mapsvg.com/blog/statistical-maps>

TIŠNOVSKÝ, PAVEL. Programovací jazyk R: úvodní informace. *Root.cz* [online]. 2020 [cit. 2021-11-22]. Dostupné z: <https://www.root.cz/clanky/programovaci-jazyk-r-uvodni-informace/>

TUFTE, Edward Rolf. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press, 2007.