

**EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY**

Evidenčné číslo: 103006/B/2022/36124048425515268

**VYUŽITIE POWER BI DESKTOP NA ANALÝZU
A VIZUALIZÁCIU ÚDAJOV**

Bakalárska práca

2022

Veronika Hesková

**EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY**

**VYUŽITIE POWER BI DESKTOP NA ANALÝZU
A VIZUALIZÁCIU ÚDAJOV**

Bakalárska práca

Študijný program: hospodárska informatika

Študijný odbor: informatika

Školiace pracovisko: Katedra matematiky a aktuárstva

Vedúci záverečnej práce: doc. Mgr. Vladimír Mucha, PhD.

Bratislava 2022

Veronika Hesková

Čestné vyhlásenie

Čestne vyhlasujem, že záverečnú prácu som vypracovala samostatne a že som uviedla všetku použitú literatúru.

Dátum:

Podpis študenta:

ABSTRAKT

HESKOVÁ, Veronika: *Využitie Power BI Desktop na analýzu a vizualizáciu údajov*. - Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky; Katedra matematiky a aktuárstva. - Vedúci záverečnej práce: doc. Mgr. Vladimír Mucha, PhD. - Bratislava: FHI EU, 2022, 63 s.

Cieľom záverečnej práce bolo popísať možnosti využitia aplikácie Power BI Desktop, ako súčasť nástroja Power BI, na analýzu a vizualizáciu údajov prostredníctvom interaktívneho dashboardu, ktorý bol zameraný na momentálnu pandemickú situáciu vo svete. Práca je rozdelená do 4 kapitol a obsahuje 47 obrázkov a 1 tabuľku. Prvá kapitola je venovaná definovaniu pojmov Business Intelligence a Big Data, ako zastrešujúcim pojmom práce, spolu s charakteristikou procesu analýzy a vizualizácie údajov. V rámci tejto kapitoly je popísaná taktiež aplikácia Power BI Desktop z teoretického hľadiska, ako jedna zo služieb Power BI, ktorej neoddeliteľnou súčasťou je jazyk DAX. V ďalšej časti sa charakterizuje hlavný cieľ spolu s čiastkovými cieľmi a taktiež je venovaná pozornosť metodike a metódam skúmania, využitých pri spracovávaní bakalárskej práce. Záverečná kapitola sa zaoberá vytvorením interaktívnej zostavy v aplikácii Power BI Desktop, ktorej predchádzala celá postupnosť krokov s ich podrobnou charakteristikou, od pripojenia sa k údajom až po pridávanie vizualizácií do stránok zostavy.

Kľúčové slová: Business Intelligence, Power BI Desktop, údaje, dashboard

ABSTRACT

HESKOVÁ, Veronika: *Using Power BI Desktop for data analysis and visualization*. - University of Economics in Bratislava. Faculty of Economic Informatics; Department of Mathematics and Actuarial Science. - Supervisor: doc. Mgr. Vladimír Mucha, PhD. - Bratislava: FHI EU, 2022, 63 p.

The aim of the thesis was to describe the possibilities of using Power BI Desktop, as part of the Power BI tool, to analyse and visualise data through an interactive dashboard that focused on the current pandemic situation in the world. The thesis is divided into 4 chapters and contains 47 figures and 1 table. The first chapter is devoted to defining the concepts of Business Intelligence and Big Data, as the overarching concepts of the thesis, along with a characterization of the data analysis and visualization process. Within this chapter, the Power BI Desktop application is also described from a theoretical point of view, as one of the Power BI services, of which the DAX language is an integral part. In the next part, the main objective is characterized together with the sub-objectives and attention is paid to the methodology and research methods used in the elaboration of the bachelor thesis. The final chapter deals with the creation of an interactive report in the Power BI Desktop application, preceded by a sequence of steps with their detailed characteristics, from connecting to the data to adding visualizations to the report pages.

Keywords: Business Intelligence, Power BI Desktop, data, dashboard

Obsah

ÚVOD.....	7
1 SÚČASNÝ STAV RIEŠENEJ PROBLEMATIKY DOMA I V ZAHRANIČÍ	9
1.1 BUSINESS INTELLIGENCE	9
1.1.1 <i>Big Data</i>	11
1.1.2 <i>Analýza a vizualizácia údajov</i>	13
1.1.3 <i>Komparácia nástrojov BI</i>	17
1.2 POWER BI.....	19
1.2.1 <i>Power BI Desktop</i>	22
1.2.2 <i>Jazyk DAX</i>	23
2 CIEĽ PRÁCE	25
3 METODIKA PRÁCE A METÓDY SKÚMANIA.....	26
3.1 NAČÍTANIE ÚDAJOV	26
3.2 TRANSFORMÁCIA ÚDAJOV, VYTVÁRANIE MODELU	27
3.3 APLIKOVANIE JAZYKA DAX.....	30
3.4 TVORBA VIZUÁLOV	31
3.5 VYTVÁRANIE DASHBOARDOV	38
4 VÝSLEDKY PRÁCE A DISKUSIA	40
4.1 NAČÍTANIE ÚDAJOV A ICH ÚPRAVY, TVORBA MODELU.....	40
4.2 VYPOČÍTANIE DOPLŇUJÚCICH HODNÔT	50
4.3 ZOBRAZENIE FINÁLNYCH STRÁNOK ZOSTAVY A ICH ANALÝZA	54
ZÁVER	59
ZOZNAM POUŽITEJ LITERATÚRY	61

Úvod

V súčasnej dobe popularita systémov Business Intelligence neustále narastá. Spoločnosti sa naprieč všetkými odbormi usilujú o prosperitu a udržanie konkurencieschopnosti, k čomu nepochybne prispieva získavanie znalostí z vlastných dát za pomoci podnikových informačných systémov. Spolu s tým nadobúdajú na význame aplikácie Business Intelligence a reportingové nástroje na prácu s dátami. Nástroje Business Intelligence môžu organizáciám poskytnúť rýchly a efektívny rozhodovací proces založený na zdrojoch údajov, ktorý môže ovplyvniť prežitie organizácie na trhu. S rastúcim významom BI dochádza k vývoju nových nástrojov, vďaka ktorým sa zjednodušuje práca s dátami.

Jedným z takýchto nástrojov, pre jednoduchšiu a rýchlejšiu analýzu aj následnú vizualizáciu dát, je práve Power BI Desktop, na ktorého možnosti využitia je zameraná táto bakalárska práca. Jedná sa o Self Service Business Intelligence nástroj od spoločnosti Microsoft, pomocou ktorého je možné analyzovať dáta a následne vytvoriť vizuálne atraktívne a prehľadné reporty. Power BI získava údaje z viacerých zdrojov, či už v cloude alebo na serveri, a kombinuje ich na jednej výkonnej platforme. Rovnako umožňuje sledovanie zásadných ukazovateľov, a to aj v reálnom čase, čím je užitočný v mnohých oddeleniach firmy - od HR a manažmentu až po výrobu. Power BI má totižto schopnosť zobrazovať rôzne úrovne pohľadov na to, ako organizácia funguje, aké má zisky a problémy, a živé štatistiky na aktualizáciu interných alebo externých zainteresovaných strán organizácie. To všetko je prezentované vizuálne príťažlivým spôsobom.

Cieľom práce je prezentovať možnosti aplikácie Power BI Desktop v oblasti analýzy a vizualizácie údajov, so zameraním na popísanie jej schopností preberať údaje z rôznych dátových zdrojov, na ich transformáciu potrebnú pre vytvorenie dátového modelu a na ich vizualizáciu v podobe zaujímavých grafických výstupov.

Úvodná kapitola je rozdelená na dve hlavné časti. Prvú časť tvorí charakteristika pojmu Business Intelligence, nasledujúc pojmom Big Data, doplnená o podobnosti a rozdiely medzi nimi. Rovnako má za úlohu vysvetliť proces analýzy a vizualizácie údajov a porovnať tri rôzne, často používané nástroje Business Intelligence. Jedným z nich je práve nástroj Power BI Desktop, ktorý je neoddeliteľnou a veľmi podstatnou súčasťou našej záverečnej práce. Druhá časť prvej kapitoly je venovaná už samotnému nástroju Power BI a službe Power BI Desktop, ktoré najskôr predstavíme z teoretického hľadiska, popíšeme

hlavné komponenty tohto prostredia a zameriame sa taktiež na jazyk DAX, ktorý umožňuje tvorbu nových mierok a výpočtov v aplikácii.

V druhej kapitole si určíme hlavný cieľ záverečnej práce, ktorý sme sa rozhodli dosiahnuť vďaka stanoveným čiastkovým cieľom. Na úspešné naplnenie čiastkových cieľov a tým aj hlavného cieľa, využijeme metodiku práce a metódy skúmania uvedené v tretej kapitole bakalárskej práce.

Štvrtou a zároveň záverečnou kapitolou demonštrujeme možnosti analýzy a vizualizácie údajov v aplikácii Power BI Desktop na získaných údajoch, zaoberajúcich sa problematikou COVID-19. Pôjde o databázu, v ktorej sú zaznamenávané počty prípadov po celom svete. V rámci tejto kapitoly aplikujeme celý postup - od získania surových dát z externého zdroja, cez vytváranie dátového modelu a prácu s jazykom DAX až po samotnú analýzu a vizualizáciu údajov - na konkrétnych získaných údajoch. Tým umožníme aj neskúsenému čitateľovi zostrojiť si svoj prvý dashboard, ktorým by mohol odprezentovať akékoľvek údaje a analyzovať ich.

1 Súčasný stav riešenej problematiky doma i v zahraničí

1.1 Business Intelligence

Na úvod je potrebné vysvetliť, čo vlastne pojem Business Intelligence znamená. Podľa Novotného (2005) Business Intelligence označuje celý komplex činností, úloh a technológií, ktoré dnes stále častejšie tvoria bežnú súčasť riadenia podnikov a ich informačných systémov. Predstavuje súbor procesov, aplikácií a technológií IS/ICT, ktoré vo väčšine prípadov podporujú analytické a plánovacie činnosti podnikov a organizácií a sú postavené na princípe multidimenzionality, pod ktorým si môžeme predstaviť možnosť pozerat' sa na realitu z niekoľkých možných uhlov pohľadu.

Základom myšlienky získavania informácií a tvorby reportov z rozsiahlych dátových súborov je tieto dáta určitým spôsobom zbierať, uchovávať a následne po spracovaní a vizualizácii aj extrahovať (Chaudhuri, 2011). Prvá inštancia, ktorá sa týmto odvetvím zaoberala bola označovaná ako BI&A 1.0 a bola založená na analýze predovšetkým štruktúrovaných dát uložených v relačných systémových databázach. Databázový dotaz, online analytické spracovanie (OLAP) a reportovacie nástroje založené na intuitívnej, ale jednoduchšej grafike sa používajú na skúmanie dôležitých údajových charakteristík. Riadenie podnikovej výkonnosti (BPM) pomocou výsledkových kariet a panelov pomáha analyzovať a vizualizovať rôzne metriky výkonnosti. Väčšina technológií na spracovanie údajov a analytické technológie už bola začlenená do popredných komerčných platforiem BI, ktoré ponúkajú hlavnú predajcovia IT vrátane Microsoft, IBM, Oracle a SAP (Chen, 2012).

Aplikácie business intelligence sa zameriavajú na analytické a plánovacie funkcie hlavne v oblasti podnikovej informatiky. Zároveň sú však na nich kladené nasledujúce nároky:

- zaistenie hodnotenia sledovaných podnikových ukazovateľov na definovanom rozsahu podnikových dát - môže ísť napríklad o objem tržieb, počet reklamácií, počet pracovníkov a pod.,
- možnosť analýzy týchto ukazovateľov na základe rozličných hľadísk, tiež označovaných ako dimenzie alebo na základe ich kombinácií - objem tržieb môžeme napríklad analyzovať podľa zákazníkov, tovaru, územia či predajcov,
- vysoká flexibilita riadiacich a rozhodovacích aktivít pracovníkov podniku,

- analýza vývoja podnikových ukazovateľov a ich fluktuácia v čase na rôznych úrovniach detailu sledovaných hodnôt (Pour, Maryška a Novotný, 2012).

Iniciatíva BI však sama osebe nie je kľúčovým aspektom, ktorý odlišuje jednu organizáciu od druhej, pretože jej hodnota veľmi závisí od ľudí, ktorí interpretujú informácie a prijímajú rozhodnutia. Skôr zhmotňuje potenciál organizácie stať sa lepšou – rýchlejšou, inteligentnejšou, uvedomelejšou, kreatívnejšou pri riešení problémov - za predpokladu, že zainteresované strany efektívne využívajú informácie. Navyše, BI riešenie nemožno implementovať tradičným spôsobom, ako by sa to stalo pri implementácii iných systémov. BI je skôr prieskumná: zahŕňa proces objavovania, ktorý prebieha nepretržite, často pri implementácii riešenia. Implementácia BI si vyžaduje nielen dynamickejší prístup, aký umožňujú agilné metodiky, ale aj zahrnutie oveľa rôznorodejších špecialistov, medzi ktorých môžu patriť programátori, databázoví experti, dátoví vedci, dátoví inžinieri, štatistici alebo matematici (Melo and Machado, 2020).

Pour (2006) však poukázal aj na určité problémy, ktoré sa spájajú s nasadzovaním BI v praxi:

- nedostatočná kvalifikačná pripravenosť a s tým často aj súvisiaci nezáujem užívateľskej sféry, čo v konečnom dôsledku znamená, že iniciatívu v tejto oblasti podnikovej informatiky preberajú útvary ICT,
- problematická kvalita zdrojových dát, ktorá je najčastejšou príčinou oneskorenia alebo dokonca aj znehodnotenia BI projektov,
- uplatnenie Business Intelligence v praxi predstavuje novú kvalitu a štýl riadenia firmy, avšak za predpokladov, že firma a jej útvary, vrátane ICT, budú takéto zmeny akceptovať. Tento predpoklad však nie je vo väčšine prípadov splnený.

Aj napriek tomu môžeme tvrdiť, že BI sa stále rýchlejšie a rýchlejšie rozširuje. V súčasnosti, keďže údaje predstavujú novú triedu ekonomických aktív, podobnú mene alebo zlatu, BI sa stala výzvou pre informačné technológie - Priemysel 4.0 - ako aj veľmi dôležitou otázkou riadenia. Jeho dôležitosť bola uznaná najmä pre rozvoj rozhodovacích schopností založených na analýze, ktoré sa odrážajú v softvéri a počítačových systémoch (Romero et al., 2021).

1.1.1 Big Data

V súčasnosti sa odhaduje, že asi 98% všetkých informácií je uložených v digitálnej forme; kým v roku 2000 bol podiel digitálnych dát na globálnych informačných zdrojoch len 25 %. Podľa štatistických domnienok sa predpokladá, že v súčasnosti údaje pribúdajú 40-krát rýchlejšie ako populácia. Na ilustráciu rozsahu nárastu dát si stačí vziať do úvahy skutočnosť, že samotná sociálna sieť Twitter generuje 12 terabajtov dát na dennej báze. Ľudia využívajúci internet prostredníctvom rôznych typov služieb, portálov, aplikácií, stacionárnych a mobilných zariadení sú istým spôsobom generátorom nekonečného množstva dát.

Big Data, pochádzajúce najmä z internetu, sociálnych médií, rôznych vládnych portálov, distribuovaných databáz, ako aj z mobilných zariadení, sú považované za novú formu kapitálu (Olszak, 2021). Ak by sme chceli ísť do histórie tohto pojmu, nebolo by potrebné ísť veľmi ďaleko do minulosti. Termín „Big Data“ totižto prvýkrát použil počítačový vedec zo spoločnosti Silicon Graphics až v polovici 90. rokov 20. storočia. V roku 2008 začalo tento termín používať niekoľko časopisov z technologického priemyslu na označenie väčších objemov dát, vo všeobecnosti v rozsahu petabajtov ale až v roku 2012 prišiel väčší zlom, kedy sa tento pojem dostal do mainstreamu (Howson, 2014).

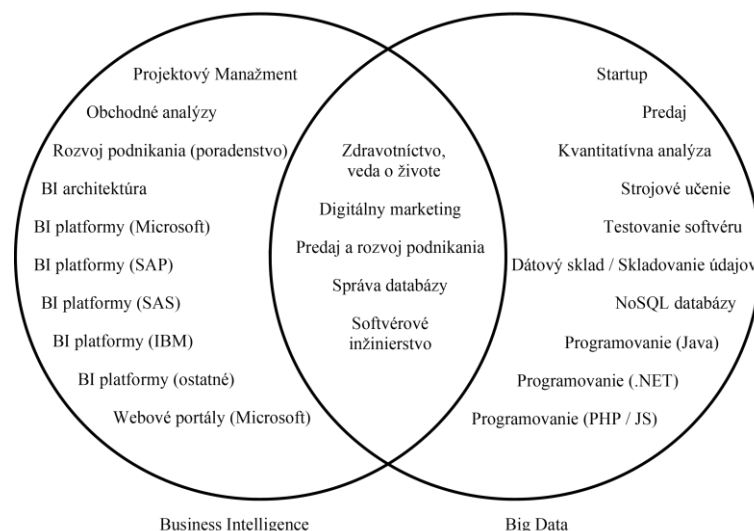
Podľa Olszak (2021) Big Data môžu predstavovať jednu z piatich „kritických technických oblastí“, ktoré prispievajú k Business Intelligence a analytike, pričom ďalšími sú textová analytika, webová analytika, sieťová analytika a mobilná analytika. Big Data považujú za podmnožinu Business Intelligence. Uviedli, že Business Intelligence možno vnímať ako zastrešujúci pojem, ktorý zahŕňa aplikácie a nástroje, ktoré umožňujú prístup a analýzu veľkých dát a informácií s cieľom zlepšiť výkonnosť podniku, konkurenčnú výhodu a vytvoriť obchodnú hodnotu. Mnohí autori tvrdia, že veľké dáta znamenajú nové spôsoby vedenia podniku, rozhodovania a porozumenia zákazníkom, dodávateľom a ďalším zainteresovaným stranám organizácií. Uvádza sa, že Big Data umožňujú vytvárať inovatívne obchodné modely, produkty a služby. Poskytuje organizáciám spôsob, ako prekonať konkurenciu. Tento druh údajov možno použiť na dosiahnutie lepšieho pochopenia zákazníkov, zamestnancov, partnerov a operácií organizácie.

Už z doposiaľ uvedených informácií je zrejmé, že medzi oblasťami Business Intelligence a Big Data existuje istá súvislosť. Najmä pokiaľ ide o všeobecné koncepcie a metódy IT a obchodné zručnosti, pozorovali sme značné prekrývanie medzi BI a BD, ktoré

môžeme vidieť na Obrázku 1. Napríklad práca v oboch oblastiach si vyžaduje určité množstvo softvérového inžinierstva a databázovej kompetencie. Zručnosti v oblasti predaja a rozvoja podnikania pre správu riešení BI a BD sa tiež prekrývajú. Napokon, znalosti z oblasti sa prekrývajú v zdravotníctve a digitálnom marketingu, doménach, o ktorých je známe, že sú založené najmä na údajoch.

V nadväznosti na túto časť si preto uvedieme aj opačnú stránku, teda skutočnosti, v ktorých sa tieto dve oblasti odlišujú. Zatiaľ čo úlohy BD vyžadujú kvantitatívnu analýzu a zručnosti strojového učenia, v úlohách BI sa o takýchto pojmoch výslovne nehovorí. Prečo boli nájdené také rozdiely, hoci BI aj BD sa zameriavajú na podporu rozhodovania prostredníctvom kvantitatívnej analýzy údajov? Existujú dve možné vysvetlenia tohto zistenia: Po prvé, vznikajúca literatúra o veľkých údajoch neustále zdôrazňuje ich rôznorodosť, čo naznačuje, že veľké údaje sa nevzťahujú na relačné údaje spravované v podnikových systémoch alebo dátových skladoch, ale na prúdy údajov v rôznych formátoch a z rôznych zdrojov, väčšinou z internetu. Kvôli tejto rôznorodosť údajov sa riešenia na analýzu veľkých údajov spoliehajú menej na štandardné softvérové produkty ako na riešenia na mieru. Po druhé, súčasné projekty veľkých dát hľadajú odpovede na vysoko špecializované otázky a sú často porovnateľnejšie s výskumnými projektmi ako s tradičnými IT projektmi. Kvôli tejto rôznorodosť otázok vyžadujú riešenia na analýzu veľkých dát viac prispôbené softvérové nástroje a lepšie metodologické zručnosti ako tradičné BI (Debortoli, Müller and Brocke, 2014).

Obrázok 1: Podobnosti a rozdiely v oblastiach kompetencií BI a BD



Zdroj: Vlastné spracovanie podľa Debortoli, Müller and Brocke, 2014

1.1.2 Analýza a vizualizácia údajov

V nasledujúcej kapitole sa zameriame na dve hlavné činnosti tejto práce z všeobecného hľadiska, ktorými sú analýza a vizualizácia dát. Následne tieto skutočnosti budeme môcť odprezentovať aj na reálnych dátach.

Slovo analytika vo veľkej miere nahradilo predchádzajúce jednotlivé komponenty počítačových technológií na podporu rozhodovania. V skutočnosti mnohí odborníci a akademici teraz používajú slovo analytika namiesto BI. Aj keď to mnohí autori a konzultanti definovali trochu inak, analytiku možno vnímať ako proces vývoja rozhodnutí alebo odporúčaní pre akcie na základe poznatkov generovaných z historických údajov. Podľa Inštitútu pre operačný výskum a manažment (INFORMS) predstavuje analytika kombináciu výpočtovej techniky, techník manažérskej vedy a štatistiky na riešenie skutočných problémov. Samozrejme, mnohé iné organizácie navrhli svoje vlastné interpretácie a motivácie pre analýzu.

Technológie, ktoré sú dnes k dispozícii, môžu tiež automaticky vydávať upozornenia pre osoby s rozhodovacou právomocou, keď si výkon takéto upozornenia vyžaduje. Na úrovni spotrebiteľov vidíme takéto upozornenia na počasie alebo iné problémy. Ale podobné upozornenia môžu byť generované aj v špecifických nastaveniach, keď predaje klesnú nad alebo pod určitú úroveň v určitom časovom období alebo keď sa zásoby konkrétneho produktu míňajú. Všetky tieto aplikácie sú možné vďaka analýze a dotazom na údaje, ktoré organizácia zhromažďuje.

Ďalšia úroveň analýzy môže zahŕňať štatistickú analýzu na lepšie pochopenie vzorcov. Tie sa potom môžu posunúť o krok ďalej na vývoj prognóz alebo modelov na predpovedanie toho, ako by zákazníci mohli reagovať na konkrétnu marketingovú kampaň alebo prebiehajúce ponuky služieb/produktov. Keď má organizácia dobrý prehľad o tom, čo sa deje a čo sa pravdepodobne stane, môže použiť aj iné techniky na prijímanie najlepších rozhodnutí za daných okolností.

Spomínané úrovne možno charakterizovať ako deskriptívnu, prediktívnu a normatívnu úroveň. Do istej miery sú to nezávislé kroky a jeden typ analytických aplikácií vedie k druhému.

- **Deskriptívna** (alebo reportingová) **analytika** sa vzťahuje na poznanie toho, čo sa deje v organizácii a pochopenie niektorých základných trendov a príčin takýchto udalostí.

- Cieľom **prediktívnej analýzy** je určiť, čo sa pravdepodobne stane v budúcnosti. Táto analýza je založená na štatistických technikách, ako aj na iných, nedávno vyvinutých technikách, ktoré spadajú do všeobecnej kategórie dolovania údajov.
- Cieľom **normatívnej analýzy** je rozpoznať, čo sa deje, ako aj pravdepodobnú predpoveď a urobiť rozhodnutia na dosiahnutie najlepšieho možného výkonu (Sharda, Delen and Turban, 2018).

Samotný proces analýzy je podľa EMC - Education Services (2015) v oblasti analytiky, zameranej na údaje dôležitým faktom, ktorý je potrebné pochopiť ešte pred začatím s analýzou a s prácou s údajmi. Životný cyklus analýzy údajov je navrhnutý špeciálne pre problémy s veľkými údajmi a projekty vedy o údajoch. Životný cyklus má šesť fáz a projektová práca môže prebiehať v niekoľkých fázach naraz. Životný cyklus definuje najlepšie postupy analytického procesu od objavovania až po dokončenie projektu a čerpá zo zavedených metód v oblasti analýzy údajov a rozhodovacej vedy. Táto syntéza bola vyvinutá po zhromaždení vstupov od vedcov v oblasti údajov a po konzultácii so zavedenými prístupmi, ktoré poskytli informácie o častiach procesu.

Životný cyklus zahŕňa šesť fáz. Tímy sa bežne učia nové veci vo fáze, ktorá ich prinúti vrátiť sa späť a vylepšiť prácu vykonanú v predchádzajúcich fázach na základe nových poznatkov a informácií, ktoré boli odhalené. Týmito fázami sú:

1. **Objav.** V prvej fáze sa tím naučí oblasť podnikania vrátane relevantnej histórie. Poukazujú sa dostupné zdroje na podporu projektu z hľadiska ľudí, technológie, času a údajov. Dôležité činnosti v tejto fáze zahŕňajú rámcovanie obchodného problému ako analytickú výzvu a formulovanie počiatkových hypotéz na testovanie.
2. **Príprava údajov** - vyžaduje si prítomnosť analytického testovacieho prostredia, v ktorom môže tím pracovať s údajmi a vykonávať analýzy. Tím musí vykonať extrakciu, načítanie a transformáciu údajov. Rovnako sa tiež podnikajú kroky na úpravu týchto údajov.
3. **Plánovanie modelu.** Počas tejto fázy určuje tím metódy, techniky a pracovný tok, ktorý má v úmysle použiť v nasledujúcej fáze. Skúmajú sa údaje pre zistenie vzťahov medzi premennými, z ktorých budú následne vybraté kľúčové premenné a najvhodnejšie modely.
4. **Budovanie modelu.** Tím zostavuje a vykonáva modely na základe práce vykonanej vo fáze plánovania modelu. Zvažuje sa, či existujúce nástroje budú

stačiť na spustenie modelov, alebo či bude potrebné iné, výkonnejšie prostredie.

5. **Komunikácia výsledkov.** V tejto fáze tím v spolupráci s hlavnými zainteresovanými stranami určí, či sú výsledky projektu úspešné alebo neúspešné na základe kritérií vyvinutých v prvej fáze. Tím by mal identifikovať kľúčové zistenia, kvantifikovať obchodnú hodnotu a vypracovať zhrnutie zistení zainteresovaným stranám.
6. **Uvedenie do prevádzky.** V poslednej fáze tím dodáva záverečné správy, brífingy, kódy a technické dokumenty. Rovnako tiež môže byť spustený pilotný projekt na implementáciu modelov v produkčnom prostredí.

Ak by sme sa zamerali na vizualizáciu, tak bez ohľadu na typ dát by mali vizualizačné techniky spĺňať požiadavky koncového používateľa a tým pádom aj poskytovať jednoduché, časovo úsporné a efektívne riešenie prezentácie dát (Yipeng, 2017). Správna vizualizácia dát je základom efektívnosti reportingu. Práca s vizualizačnými nástrojmi väčšinou vyžaduje ručnú voľbu vizuálov. Daný výber zobrazovaných dát je teda na samotnom človeku, rovnako ako aj voľby ich zobrazenia. To vytvára požiadavku na znalosť dát a vizualizačných techník (Wongsuphasawat et al., 2016). Správne použitie vizualizačných technológií dokáže zjednodušiť komplexné dáta a tiež pomôcť pochopiť pravidlá, ktoré sú ukryté v dátach (Yipeng, 2017).

Správne dodržanie pravidiel vizualizácie dát by malo vo výsledku znamenať zjednodušenie pri čítaní dát z informačného zdroja. Zjednodušiť čitateľnosť komplexných dát umožňujú nasledujúce tri charakteristiky spracovania pomocou vizualizačných technológií:

- interaktivita zaisťujúca jednoduchý výber dát,
- multidimenzionalita - možnosť vybrať si informácie z rôznych dimenzií (atribútov a charakteristík) dát a zobraziť ich v zoskupenej jednodimenzionalnej forme,
- dáta sú prezentované formou obrázkov, grafov a kriviek a vytvárajú tak viditeľné vzorce a vnútorné vzťahy medzi údajmi (Yipeng, 2017).

Výber grafu sa môže z prvého pohľadu javiť ako jednoduchá záležitosť ale v skutočnosti ide o naozaj premyslenú činnosť, ktorej by sme mali venovať rovnako veľa pozornosti ako získavaniu dát, či transformácii. Každý typ grafu je vhodný pre zobrazenie určitého vzťahu medzi dátami. Powell (2018) pre jednotlivé vzťahy odporučil určité grafy a toto rozdelenie môžeme vidieť v nasledujúcej Tabuľke 1:

Tabuľka 1: Kategórie grafov

Kategórie	Graf	Využitie
Porovnanie	Čiarový Stĺpcový a pruhový	Porovnanie v čase Porovnanie skupín
Vzťah	XY bodový a bublinový	Relatívna pozícia bodu v závislosti na dvoch premenných
Rozloženie	Histogram	Frekvencia výskytu v závislosti od diskkrétnej premennej (alebo skupiny)
Zloženie	Vodopádový Lievikový	Zmeny hodnoty v čase v závislosti od kategórie Rozdiel medzi štádiami procesu

Zdroj: Vlastné spracovanie podľa Powel, 2018

Powell (2018) ďalej uvádza, čomu je dobré sa vyhnúť pri vizualizácii údajov:

- snaha zobrazit' nadmerné množstvo informácií na jednu stranu reportu, vzhľadom na to, že v konečnom dôsledku bude tento report veľmi neprehľadný,
- nekonzistencia zobrazenia úrovne detailu alebo aj nevyrovnané reporty,
- snaha zaujať farbami namiesto informáciami, a v neposlednom rade nesprávna voľba grafov.

Výberom vhodných grafov následne vytvárame tzv. dashboardy. Podľa Fewa (2006) je dashboard vizuálne zobrazenie najdôležitejších informácií potrebných na dosiahnutie jedného alebo viacerých cieľov; konsolidované a usporiadané na jednej obrazovke, takže informácie možno sledovať na prvý pohľad. Vo svojej knihe *Information Dashboard Design* analyzoval mnohé, v tom čase dostupné vizualizačné dashboardy za účelom kategorizácie a vytvorenia odporúčaní pre dobré princípy dizajnu pre dashboardy. Následne poukázal na najkritickejšie aspekty dashboardov:

- zobrazujú informácie potrebné na dosiahnutie konkrétnych cieľov,
- zmestia sa na jednu obrazovku počítača,
- v súčasnosti je najlepším médiom pre dashboard webový prehliadač,
- informácie by malo byť možné sledovať na prvý pohľad,
- majú malé, stručné, jasné a intuitívne zobrazovacie mechanizmy,
- sú prispôsobiteľné.

1.1.3 Komparácia nástrojov BI

SelectHub (2022) uvádza rebríček najobľúbenejších a najúspešnejších nástrojov pre platformy Business Intelligence. V príslušnej podkapitole si uvedieme a zároveň aj porovnáme 3 nástroje, ktoré patria medzi 10 najlepších či už v rámci celkových parametrov na danú platformu alebo aj pri zameraní sa na konkrétny parameter, ktorý je v našom prípade vizualizácia a analýza údajov. Ide o nástroje nasledujúcich spoločností:

- Tableau - Tableau Desktop,
- QLIK - QLIK Sense (Desktop),
- Microsoft - Microsoft Power BI Desktop.

Tableau Desktop

Tableau ponúka bohatú škálu riešení business intelligence. Konkrétne Tableau Desktop je výkonný nástroj na vytváranie interaktívnych dashboardov, definujúcich príbehy pre vizuálny prieskum dát a prezentáciu pripravených vizualizácií. Táto aplikácia podporuje veľké množstvo typov súborov akými sú napríklad Access, Excel alebo textové súbory, rôzne typy databáz, medzi ktoré patria PostgreSQL, MySQL alebo Oracle, ale tiež rozmanitú škálu serverov a služieb - Amazon Redshift, Microsoft SQL Server, Google Analytics, atď., z ktorých dokáže extrahovať údaje. Po extrakcii údajov zo zdroja údajov je možné ich transformovať do špecifického formátu, ktorý je možné následne použiť vo vizualizáciách. V dashboarde je možné jednotlivé atribúty údajov filtrovať, triediť, spájať, rozdeľovať alebo dokonca úplne mazať. Tableau Desktop privíta používateľa priateľským rozhraním, ktoré je typické intuitívnym spôsobom štruktúrovania údajov, významnou súčasťou je aj takzvané „drag and drop“ rozhranie. Ide o pôsobivý nástroj so širokým sortimentom možností úprav prezentovaných údajov.

Jedna z najväčších predností Tableau spočíva v jeho funkcii Show Me. Podporuje užívateľa pri výbere vizualizácií, ktoré sú kompatibilné s vybranými atribútmi. Umožňuje tiež zmenu vizualizácie na inú, ktorá má rovnaké atribúty. Ďalšou užitočnou funkciou pre aspekt analýzy dát Tableau je možnosť obohatiť vizualizácie o trendové čiary, prognózy alebo sumarizovať dáta cez boxplot, čo uľahčuje detekciu určitých vzorov v dátach.

Má to však aj určité nevýhody. Napríklad, keď sa vizualizácie stanú komplikovanejšími, pre používateľa bude ťažké sledovať atribúty, ktoré boli vybraté predtým. Okrem toho je Tableau k dispozícii iba pre Windows a Mac. Najväčšou nevýhodou

Tableau však je, že mu chýba rozšíriteľnosť, čo znamená, že nedokáže zahrnúť nové vizualizácie jednoduchým a intuitívnym spôsobom (Tableau, 2022).

QLIK Sense (Desktop)

Qlik Sense Desktop je v porovnaní s cloudovou verziou výkonnejšia, čo sa týka manipulácie s dátami a navyše je dostupný zadarmo. Aby sme mohli Qlik Sense používať, musíme si najprv vytvoriť účet na stránke Qlik a pri spustení desktopovej aplikácie je potrebné sa prihlásiť. Napriek tomu, že Qlik ponúka väčšinu bežných pripojení k zdrojom údajov - ODBC, Microsoft SQL Server, PostgreSQL, Oracle, nepodporuje však niektoré online služby, ktoré sú tiež široko používané, konkrétne Amazon Redshift alebo Google Analytics. Pri úprave načítaných dát ponúka Qlik veľmi slabú detekciu dátového typu. Rozmery a miery je potrebné pri vytváraní vizualizácií manuálne rozdeliť.

Veľmi užitočnou funkciou pri úprave načítaných údajov je stĺpcová štatistika. Pri výbere stĺpca v tabuľke, Qlik zobrazí stĺpcový graf s počtom výskytov každej hodnoty. Pre stĺpce s číselnými hodnotami je navyše možné zobrazit' os s vyznačenými hodnotami pre minimum, maximum, priemer a medián. Možnosti úpravy tabuľky sú bohužiaľ v tomto prípade veľmi obmedzené. Je možné nahradiť hodnoty a rozdeliť stĺpce, ale zlúčenie 2 stĺpcov je ťažkopádne.

Na vytvorenie vizualizácie Qlik Desktop používa prístup, ktorý má samostatné režimy na kontrolu vizualizácií a ich vytváranie. Vytvorenie vizualizácie je veľmi jednoduché; požadovanú vizualizáciu stačí presunúť na dashboard a pridať potrebné atribúty. Ak však atribúty nie sú správne vybraté, vizualizácia sa aj tak vytvorí, ale nesprávne. Významnou nevýhodou nástroja je, že nie je možné vybrať niektoré atribúty a získať návrhy, s ktorými vizualizáciami ich možno použiť. Preto nie je vhodný pre laických používateľov. Vytvorené vizualizácie vyzerajú príťažlivo a sú plne prispôsobiteľné čo sa týka farieb, štítku, orientácie atď., sú navyše navzájom prepojené, čo znamená, že výber v jednej vizualizácii je zvýraznený aj v iných. Dizajn Qlik sense je možné prispôbiť preferenciám užívateľa prostredníctvom tém a možnosť písať si vlastné rozšírenia dáva vývojárom možnosť zahrnúť vizualizácie šité na mieru ich potrebám (Qlik, 2022).

Microsoft Power BI Desktop

Posledným nástrojom je Power BI Desktop, ktorý si však v tejto časti popíšeme len z informatívneho, nie moc praktického hľadiska, aby bolo možné poukázať na výhody či nevýhody tejto služby pri porovnaní zvyšných dvoch. V ďalších kapitolách následne

odprezentujeme prácu s týmto produktom vo všeobecnosti ale tiež aj na konkrétnych a reálnych údajoch.

Power BI je nástroj spoločnosti Microsoft na vizualizáciu a prieskum údajov. Ponúka známy dizajn s hornými prvkami navigácie na páse pre používateľov, ktorí sú už zvyknutí na prácu s Microsoft Windows a balíkom Microsoft Office. Konkrétne Power BI Desktop je jednou z troch rôznych verzií tejto služby. Hlavným rozdielom a dôvodom, prečo je práve táto verzia obľúbenou, je jej voľná a bezplatná dostupnosť.

Power BI pokrýva všetky najpoužívanejšie súbory na import údajov a okrem toho podporuje aj obrovské množstvo databáz, online služby a ďalšie dátové zdroje. Po načítaní údajov je možné ich transformovať do požadovaného tvaru ešte pred vytvorením danej vizualizácie vďaka veľmi kvalitnému editoru. V zásade existujú dva spôsoby vytvorenia vizualizácie: Môžete vybrať atribúty, ktoré vás zaujímajú, a potom kliknúť na tlačidlo vizualizácie, alebo môžete kliknutím na tlačidlo vytvoriť prázdnu vizualizáciu a potom presunúť atribúty do požadovaných kanálov.

Samotné vizualizačné tlačidlá nemajú popis svojich požiadaviek, takže sú potrebné existujúce znalosti o tom, ako je každá vizualizácia štruktúrovaná a ako funguje, a preto nie je Power BI vhodná pre neodborníkov. Celkovo je Power BI výkonným nástrojom pre analytikov údajov, ktorý ponúka mnoho aspektov, ktoré by si expertní používatelia mohli želať, ako napríklad širokú škálu podporovaných zdrojov údajov, spúšťanie R-skriptov, prispôsobenia a rozširiteľnosť. Na druhej strane jeho používateľské rozhranie s hrubými vizualizačnými oknami sťažuje hľadanie vlastností a nedostatok používateľskej podpory sťažuje používanie nástroja neodborníkom. Okrem toho je k dispozícii iba pre systém Microsoft Windows (Microsoft, 2022a).

1.2 Power BI

Ako už bolo v predošlej kapitole spomenuté, Power BI je nástroj self-service BI od spoločnosti Microsoft, na podporu samoobslužných a podnikových riešení BI, v ktorom môžeme z rôznych dátových zdrojov vytvárať rôznorodé analýzy. Môžeme napríklad vytvoriť rýchly pohľad na dáta z Excelu alebo z lokálnej databázy. V Power BI môžeme ale tiež vytvárať analýzy vo veľkých rozmeroch a v reálnom čase. Rovnako môžeme tento nástroj využiť ako analytický nástroj na podporu rozhodovania, v rôznych úrovniach

spoločnosti. Tieto dostupné možnosti dosiahol Microsoft prepojením funkcionalít už existujúcich nástrojov, ktorými sú napríklad:

- *Power Query* - využíva sa na pripravovanie dát, napríklad ich určitou transformáciou, ktorú je nevyhnutné vykonať pred vstupom do dátového modelu Power Pivot,
- *Power Pivot* - doplnujúci komponent k MS Excel dostupný od verzie 2010, ktorý je určený pre podporu samoobslužných riešení BI,
- *Power View* - doplnok, ktorý môžeme použiť na tvorbu interaktívnych reportov, ak chceme klásť dôraz najmä na vizuálnu stránku (Pour et al., 2018).

Microsoft ich však doplnil o ďalšie, zaujímavejšie konektory a služby. Z tých môžeme za najzaujímavejšie považovať napríklad *DirectQuery*, vďaka čomu sa dáta nenahrávajú do dátových modelov Power BI, ale sú uložené v samostatných dátových zdrojoch. Služba *Power BI Gateway* umožňuje rýchly a zabezpečený prenos dát alebo cloudová služba *Power BI Service* a *Power BI mobile apps* na prístup k dashboardom a reportom z akéhokoľvek mobilného zariadenia so systémom Windows, iOS či Android (Pour et al., 2018).

Ak by sme sa zamerali na konkrétne služby, ktoré sú dostupné v rámci Power BI, tak okrem Power BI Desktop nám ponúka nasledujúce produkty:

- **Power BI Pro** - Power BI Pro je jednou z licenčných možností služby Power BI (ďalšie dve sú Power BI Free a Power BI Premium). Služba Power BI je cloudová služba podnikovej analýzy (powerbi.com), ktorá umožňuje hostiť údaje, reporty a dashboardy online a zdieľať ich so svojimi spolupracovníkmi. Keďže Power BI je hostovaná v cloude a spravuje ju spoločnosť Microsoft, to znamená, že používateľ v rámci svojej organizácie nemusí kupovať, inštalovať a udržiavať lokálnu infraštruktúru.
- **Power BI Premium** - Power BI Premium, zamerané na veľké organizácie, ponúka prostredie s vyhradenou kapacitou, čím poskytuje organizácii konzistentnejší výkon bez toho, aby bolo nevyhnutné kupovať licencie pre jednotlivých používateľov. Ak by sme predpokladali, že chceme zdieľať reporty s viac ako 500 používateľmi v rámci našej organizácie a väčšina z týchto používateľov vyžaduje prístup iba na čítanie, tak tým pádom namiesto licencovania každého používateľa môžeme znížiť náklady zakúpením plánu Power BI Premium, ktorý nevyžaduje licencie pre ľudí, ktorí si dané reporty

chcú len pozrieť. Power BI Premium pridáva aj funkcie, ktoré nie sú dostupné v Power BI Pro, medzi ktoré patria napríklad väčšie veľkosti množín údajov alebo prírastkové obnovovanie údajov.

- **Power BI Mobile** - predstavuje set voľne dostupných mobilných aplikácií pre iOS, Android a Windows, ktoré používateľom umožňujú používať mobilné zariadenia, ako sú tablety a smartfóny, na získavanie prehľadov údajov na cestách. Mobilný používateľ môže napríklad zobrazovať reporty a dashboardy nasadené v službe Power BI a pracovať s nimi.
- **Power BI Embedded** - Power BI Embedded je súhrnný názov pre podmnožinu rozhraní API služby Power BI na vkládanie obsahu. Power BI Embedded, integrovaná so službou Power BI, umožňuje vývojárom vkladať interaktívne zostavy Power BI do vlastných aplikácií pre interných alebo externých používateľov.
- **Power BI Report Server** - Power BI Report Server, ktorý sa vyvinul z Microsoft SQL Server Reporting Services (SSRS), umožňuje nasadiť dátové modely a zostavy Power BI na lokálny server. To nám dáva na výber, akým spôsobom chceme reporty nasadiť a zdieľať: cloud a/alebo lokálne. Treba však podotknúť, že výber nemusí byť exkluzívny. Môžeme sa napríklad rozhodnúť nasadiť niektoré zostavy do služby Power BI, aby sme využili všetky funkcie, ktoré ponúka, zatiaľ čo ostatné zostavy nasadíme na interný portál Power BI Report Server, ktorý zachováva všetky funkcie SQL Server Reporting Services (Lachev, 2022).

Na záver tejto časti si ešte uvedieme výhody a nevýhody Power BI. Podľa Řehořa (2015) sú hlavnou výhodou omnoho nižšie náklady, oproti štandardným variantom reportingových nástrojov. Používatelia môžu k tvorbe využiť oveľa viac zdrojov dát a vďaka tomu nie sú viazaní na dátový sklad. Transformácia zdrojových dát je výrazne jednoduchšia, čím umožňuje transformovať dáta aj používateľom, ktorí sa v IT oblasti veľmi dobre neorientujú. Výraznou výhodou je tiež relatívne jednoduché ovládanie, ktoré je podobné MS Excel, čo nám môže evokovať, že každý, kto vie ako sa pracuje s MS Excelom bude intuitívne vedieť ako používať Power BI. Celý proces zaúčania sa s programom je vďaka tomu kratší a jednoduchší. Výhodou je taktiež to, že používateľ si môže podľa svojich požiadaviek kedykoľvek svoje dáta analyzovať a sám vyhodnotiť, čo znamená, že nemusí čakať na výstupy, ktoré by mu inak pripravil niekto iný, napríklad v rámci organizácie by to bolo IT oddelenie.

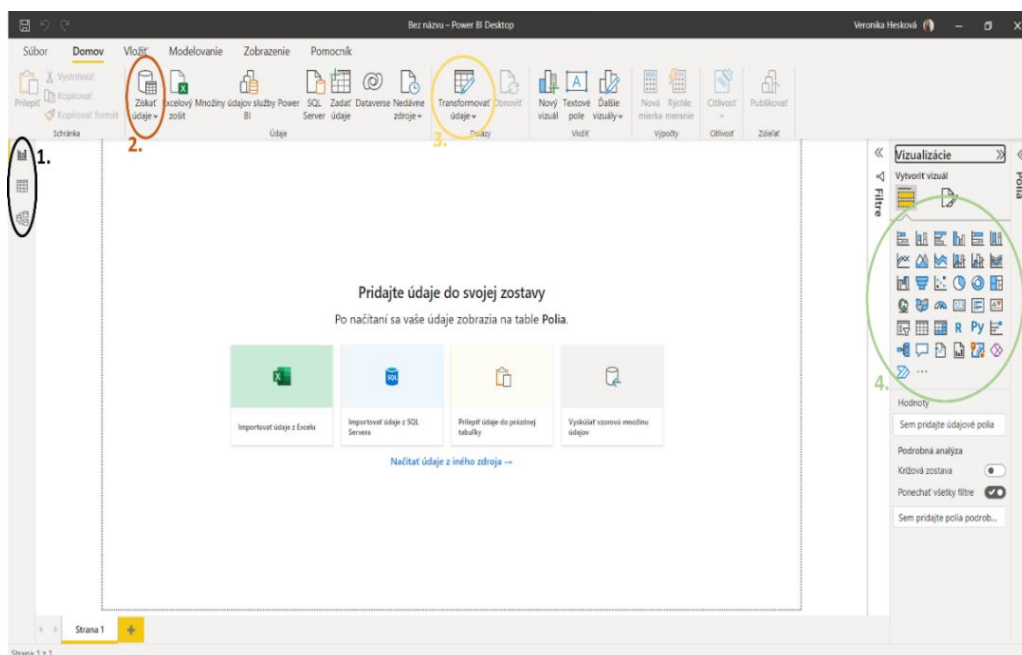
Nevýhodou je nemožnosť realizácie celopodnikovej úlohy. Následkom môže byť nezhodnosť rôznych pohľadov nad jednou vymedzenou oblasťou, čo je spôsobené buď tým, že každý užívateľ použil iný zdroj alebo nedostatočnou odbornosťou pracovníkov, vďaka čomu vznikajú následne nesprávne reporty. V nadväznosti na to s tým súvisí aj riziko neadekvátnych rozhodnutí nad nekorektnými údajmi.

1.2.1 Power BI Desktop

Power BI Desktop, ako už bolo spomenuté je voľne dostupná počítačová aplikácia pre Windows, ktorá umožňuje analytikom navrhovať samoobslužné dátové modely a vytvárať interaktívne zostavy prepojené s týmito modelmi alebo s externými zdrojmi dát (Lachev, 2022). Jensen (2020) zhrnul funkcionality Power BI Desktop do nasledujúcich štyroch bodov:

- extrahovať, transformovať a načítať údaje zo širokej škály zdrojov údajov (s Power Query),
- vytvoriť dátový model pozostávajúci z tabuliek a vzťahov,
- robiť všetky druhy výpočtov,
- vytvárať prehľady, ktoré môžu pozostávať z tabuliek, grafov a filtrov.

Obrázok 2: Prvé zobrazenie aplikácie Power BI Desktop



Zdroj: Vlastné spracovanie

Vďaka prechádzajúcej snímke obrazovky na Obrázku 2 môžeme vidieť základné zobrazenie aplikácie Power BI Desktop. Predstavíme si názvy a účely niektorých najdôležitejších funkcií, ktoré nám táto aplikácia ponúka:

1. **Zobrazenie zostavy** - Plátno zostavy používané na navrhovanie vizualizácií údajov. Toto je predvolené zobrazenie, ktoré sa otvorí pri spustení aplikácie Power BI Desktop. **Zobrazenie údajov** - Poskytuje zobrazenie údajov v našom modeli. Vyzerá to podobne ako typická tabuľka Excel, ale je len na čítanie. **Zobrazenie modelu** - Používa sa predovšetkým vtedy, keď má dátový model viacero tabuliek a vzťahov, ktoré je potrebné medzi nimi definovať.
2. **Získať údaje** - Používa sa na výber údajových konektorov a konfiguráciu podrobností o zdroji údajov.
3. **Transformovať údaje** - Spustí Editor Power Query, ktorý sa používa na aplikovanie transformácií údajov na prichádzajúce údaje.
4. **Vizualizácie** - Predstavuje ponuku všetkých základných dostupných vizualizácií, pomocou ktorých môžeme naše dáta zobraziť na ploche v rámci dashboardu. Power BI Desktop nám umožňuje vybrať si spomedzi 37 typov rôznych grafov, tabuliek či teritoriálnych zobrazení avšak ak by to používateľovi nestačilo, vie si načítať aj iné, viac vyhovujúce jeho podmienkam vizuály (Knight et al., 2020).

1.2.2 Jazyk DAX

Z pohľadu Power BI je rovnako nevyhnutné spomenúť aj jazyk DAX. DAX je skratka pre jazyk Data Analysis Expression. Tento jazyk sa používa v produktoch spoločnosti Microsoft na analýzu údajov: Power BI, Excel Power Pivot, SQL Server Analysis Services Tabulator Edition a Azure Analysis Services. Jazyk je kombinovanou verziou trocha T-SQL, vzorca Excel a C# (Rad, 2021).

Tento jazyk sa tiež nazýva funkcionálny jazyk, kde sa celý kód uchováva v rámci funkcie. Tento programovací jazyk má dva dátové typy – číselné a iné. Prvý typ bude zahŕňať celé čísla, desatinné čísla a menu, zatiaľ čo druhý bude zahŕňať binárne a reťazcové objekty. Rovnako však je možné používať zmiešané typy údajov ako vstupy v akomkoľvek vzorci DAX a konverzia prebehne automaticky. Výstupné hodnoty sa vždy skonvertujú na typ údajov, ktorý sme zadali vo vzorcoch a funkciách.

Funkcie jazyka DAX sú preddefinované vzorce uložené v službe Power BI. Tieto funkcie sa používajú na vykonávanie niektorých výpočtov hodnôt v tabuľke v argumente. Tieto argumenty bude potrebné pridať v určitom poradí a môžu to byť čísla, text, odkazy na stĺpce, logické hodnoty ako TRUE alebo FALSE, iná funkcia alebo vzorec alebo konštanty. Funkcie tiež vykonajú špecifickú operáciu na tých hodnotách, ktoré sú zahrnuté v argumente. Funkcie, ktoré v rámci jazyka DAX môžeme využiť sú tieto:

- **Agregačné funkcie** (SUM, SUMX, MAX, MIN, AVERAGE).
- **Funkcie počítania** (COUNT, COUNTA, COUNTROWS,...).
- **Logické funkcie** (AND, OR, NOT, IF, IFERROR).
- **Textové funkcie** (REPLACE, SEARCH, FIXED,...).
- **Dátumové funkcie** (DATE, HOUR, NOW,...).
- **Informačné funkcie** (ISBLANK, ISTEXT,...).

Dôležitosť jazyku DAX by sa nemala podceňovať. Ak chceme vytvoriť zostavu, ktorú ľudia ľahko pochopia so všetkými dostupnými údajmi, musíme mať základné znalosti o aplikácii Power BI Desktop. Ak chceme zlepšiť spôsob, akým premietame svoje prehľady, mali by sme sa naučiť jazyk DAX. Predpokladajme, že chceme vytvoriť prehľad, ktorý nám pomôže analyzovať percento rastu predaja v rôznych mestách, štátoch alebo okresoch v krajine. Túto analýzu môžeme použiť aj na porovnanie rastu s predajom každý rok. Ak to chceme urobiť, budeme musieť vytvoriť novú mieru v službe Power BI pomocou jazyka DAX. Týmto spôsobom môžeme vždy vytvoriť nové miery a použiť ich na vytvorenie niektorých exkluzívnych vizualizácií. Môžeme tiež získať niekoľko jedinečných náhľadov na údaje. Iba vtedy, keď máme tieto jedinečné poznatky o údajoch, môžeme vyvinúť požadované riešenia. DAX teda uľahčuje vykonávanie analýzy údajov v službe Power BI (Morris, 2020).

2 Cieľ práce

Hlavným cieľom bakalárskej práce je popísať a prezentovať možnosti aplikácie Power BI Desktop pri analýze a vizualizácii údajov.

Pre dosiahnutie hlavného cieľa našej záverečnej práce je potrebné dosiahnuť nasledujúce **čiasťkové ciele**:

1. Vysvetliť pojmy Business Intelligence a Big Data - ich prepojenie ale taktiež poukázať na rozdiely.
2. Popísať proces analýzy a vizualizácie údajov a ich možnosti v rámci oblasti Business Intelligence.
3. Charakterizovať nástroj Power BI so zameraním na aplikáciu Microsoft Power BI Desktop zo všeobecného hľadiska - popísať základné atribúty rozhrania a predstaviť jazyk DAX, dôležitý pre prácu s touto aplikáciou pri spracovávaní údajov.
4. Priblížiť prácu s aplikáciou Microsoft Power BI Desktop aj na konkrétnych dátach zostrojením interaktívneho dashboardu, čoho súčasťou je popísanie funkcionality aplikácie pre potreby získania údajov z rôznych dátových zdrojov, ich transformovania na vytvorenie dátového modelu a ich vizualizácie v podobe zaujímavých výstupov (grafy, filtrovanie, mapy alebo napríklad dekompozičný strom). Jednotlivé výstupy, vďaka zostrojeniu vzťahov v dátovom modeli, budú v rámci stránok zostavy preukazovať spomínanú interaktivitu.

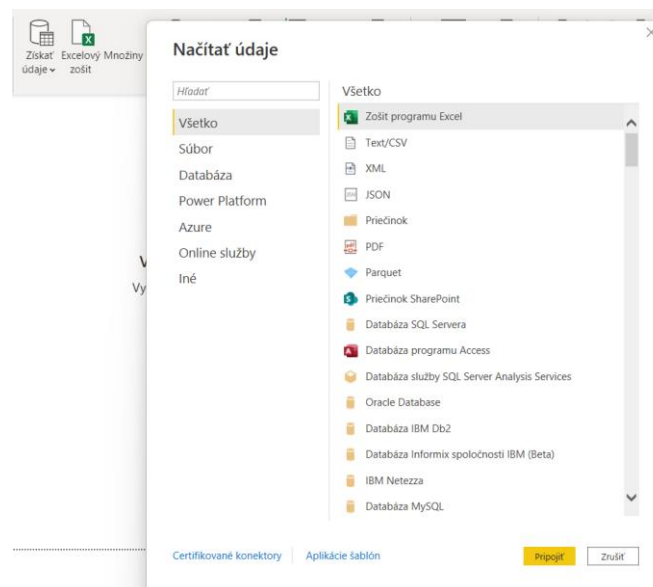
3 Metodika práce a metody skúmania

3.1 Načítanie údajov

Pri práci s Power BI Desktop je potrebné sa najprv pripojiť k určitým údajom. Táto aplikácia nám ponúka možnosť pripojiť sa k širokému spektru zdrojov dát. Pripojenie k údajom prebieha tak, že na domovskej stránke vyberieme možnosť **Získať údaje** a následne možnosť **Viac**. Vďaka tomu sa nám zobrazí okno **Načítať údaje**, kde vidíme ponuku všetkých možností zdrojov, ku ktorým je možné sa pripojiť pomocou Power BI Desktop, Obrázok 3. Medzi najčastejšie využívané zdroje údajov patria napríklad:

- Excel,
- PDF,
- Text/CSV,
- databáza aplikácie Access,
- databáza SQL serveru,
- web,
- skript jazyka Python,
- Google Analytics, atď. (Maslyuk, 2018).

Obrázok 3: Načítanie dát



Zdroj: Vlastné spracovanie

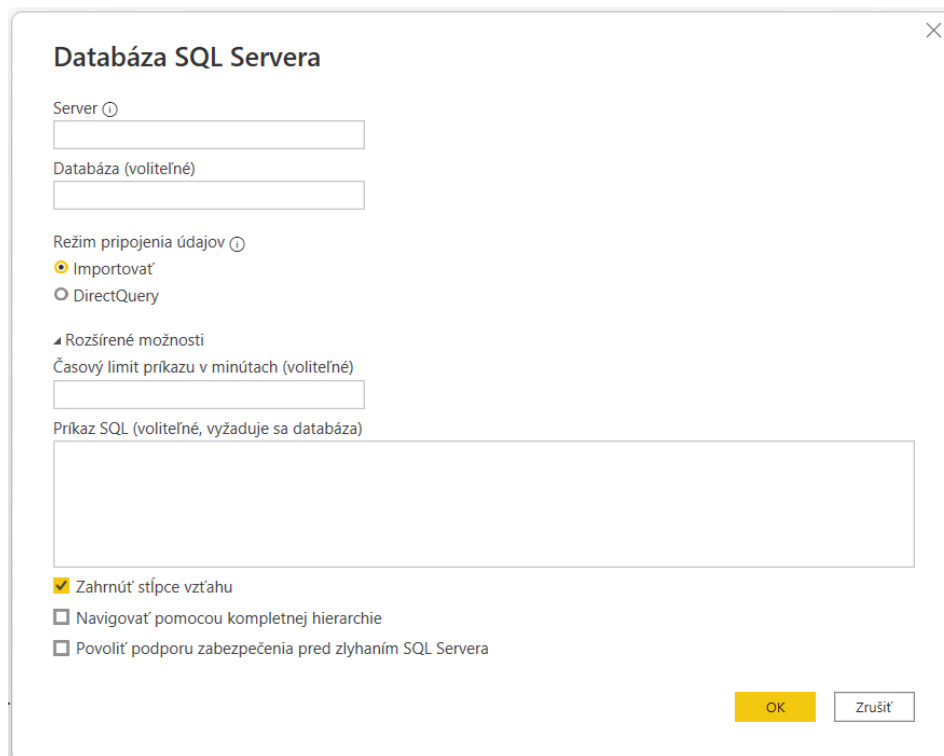
Po vybraní nášho zdroja údajov budeme vyzvaní k zadaniu doplňujúcich informácií, ktoré sú potrebné k tomu, aby sa mohol Power BI Desktop pripojiť k zdroju údajov. Môže ísť napríklad o zadanie URL adresy alebo prihlasovacích údajov, Obrázok 4 a Obrázok 5.

Obrázok 4: Výzva k zadaniu URL adresy



Zdroj: Vlastné spracovanie

Obrázok 5: Výzva k zadaniu prihlasovacích údajov do databázy



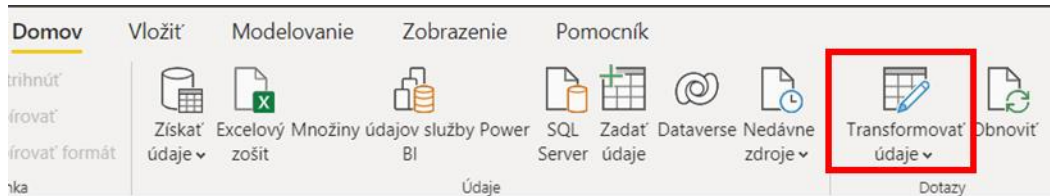
Zdroj: Vlastné spracovanie

3.2 Transformácia údajov, vytváranie modelu

Po úspešnom pripojení k zdroju alebo aj viacerým zdrojom údajov je používateľovi umožnené tieto načítané údaje transformovať do takej podoby, v akej mu budú najviac

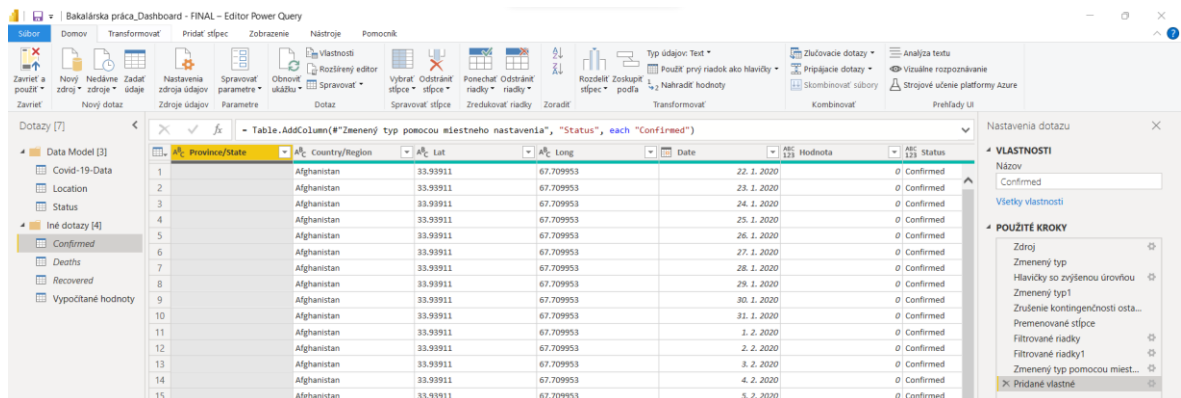
vyhovovať. V Power BI Desktop je transformácia údajov možná vďaka integrovanému Editoru dotazov - Editor Power Query, Obrázok 7, ku ktorému sa dostaneme pomocou tlačidla **Transformovať údaje**, Obrázok 6.

Obrázok 6: Transformácia údajov



Zdroj: Vlastné spracovanie

Obrázok 7: Zobrazenie prostredia v editore Power Query



Zdroj: Vlastné spracovanie

V editore Power Query je následne možné vykonávať rôzne zmeny, aby údaje boli vo finále pre nás v požadovanom tvare. Medzi možné úpravy patrí:

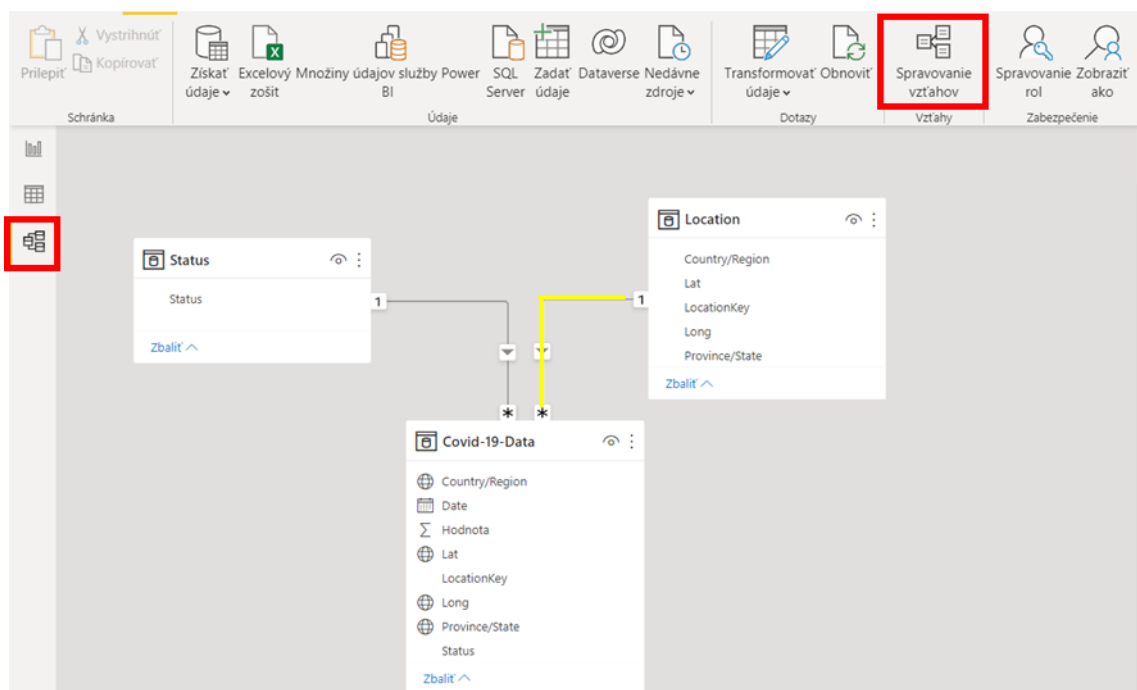
- zmena dátového typu,
- vymazanie alebo pridanie stĺpca,
- kombinovanie súborov - táto transformácia uskutoční analýzu všetkých vstupných súborov, ktoré by sa mali použiť. Je možné vybrať konkrétny objekt z daného súboru, ktorý má byť extrahovaný a následne sú automaticky vykonané potrebné dotazy. Takto je možné kombinovať všetky súbory v rámci vybranej zložky za podmienky, že sú rovnakého typu a majú rovnakú štruktúru (napríklad rovnaké stĺpce),
- zmena stĺpcov na riadky,
- použitie prvého riadku ako hlavičky.

Každá vykonaná zmena sa zaznamenáva a jednotlivé kroky je možné nájsť v použítom postupe - pravý okraj na Obrázku 7.

Po dokončení všetkých zmien je potrebné úpravy uložiť a v danom tvare údaje transformovať do prostredia Power BI Desktop - tlačidlo **Zavrieť a použiť**. Ešte pred samotným vytváraním potrebných mierok a grafických vizualizácií je potrebné zaoberať sa dátovým modelom, ku ktorému sa dostaneme zakliknutím tlačidla **Model**, v ľavej zvislej lište, spomenutej už v teoretickej časti.

V prípade, že používame pri tvorbe dashboardu viac tabuliek, ktoré majú medzi sebou nejaké väzby, je potrebné sa zaoberať práve dátovým modelom a ich vzťahy definovať. V niektorých prípadoch sa vzťahy vytvoria automaticky pri importe, čo sa stane jedine v prípade, že stĺpce majú rovnaký názov a rovnaké dátové atribúty. Ak však k tomu automaticky nedôjde, používateľ vytvorí vzťah manuálne v časti **Vzťahy - Spravovanie vzťahov**, Obrázok 8. Relácie môžeme vytvárať priamo v grafickom modeli - presunutím vzťahu od jedného príslušného atribútu z jednej tabuľky, k druhému atribútu. Power BI nám odporučí kardinalitu aj spôsob filtrovania. Sú možné vzťahy „Many to One (*:1)“, „One to One (1:1)“ a „One to Many (1:*)“ a filtrovanie jednosmerné alebo obojsmerné. Vzťahy je možné dodatočne upraviť dvojklikom na stred vzťahu alebo v rámci okna **Spravovanie vzťahov** (Maslyuk, 2018).

Obrázok 8: Práca s dátovým modelom



Zdroj: Vlastné spracovanie

3.3 Aplikovanie jazyka DAX

Ako bolo už v rámci teoretickej časti spomenuté, jazyk DAX je programovacím jazykom, pomocou ktorého pracujeme s údajmi v prostredí Power BI. Tento jazyk má vlastnú syntax a funkcie, aj keď niektoré operácie sú veľmi podobné príkazom v Exceli. Na druhej strane je však v mnohých prípadoch celkom odlišný - napr. v Exceli sa môžeme odkazovať na individuálne bunky poľa, v DAXe sa môžeme odkazovať na tabuľky alebo hodnoty v stĺpcoch.

K písaniu vzorcov v jazyku DAX sa môžeme vo všeobecnosti dostať cez hlavnú kartu **Domov**, časť **Výpočty** - tlačidlo **Nová mierka**. Každý výpočet sa začína znakom „=“, za ktorým nasledujú prvky, identifikujúce názvy stĺpcov tabuliek. Mierka napísaná v jazyku DAX môže byť použitá pri tvorbe kalkulovaného stĺpca tabuľky v editore Power Query, ktorý je vypočítavaný z hodnôt alebo prípadne konštánt v rámci jedného riadku (Maslyuk, 2018).

Rovnako už boli spomenuté rôzne typy funkcií, ktoré sa pri písaní v tomto jazyku dajú použiť. V rámci našej práce sme využili najmä agregáčnej funkcie v spojení s logickými funkciami. Proces práce s metódou jazyk DAX si môžeme demonštrovať na nasledujúcom Obrázku 9, ktorý vyjadruje výpočet celkových potvrdených prípadov s ochorením COVID-19.

Po zakliknutí **Nová mierka** nám prostredie automaticky vytvorí nový riadok, určený na zápis v jazyku DAX, s názvom Mierka. Názov je však samozrejme možné ľubovoľným spôsobom zmeniť, v našom prípade sme názov mierky upravili na „Potvrdené prípady“. Za názov bolo potrebné vložiť znamienko rovnosti, ktoré je nasledované samotným vzorcom. Využili sme funkciu CALCULATE, s parametrami Všetky prípady a Status = „Confirmed“. Tým sme funkcii určili, aby spočítavala len tie prípady, ktorých status je definovaný ako potvrdený. Po stlačení klávesy **ENTER** sa mierka pridala k poliam a bola dostupná k využitiu pre ďalšie iné účely.

Obrázok 9: Výpočet potvrdených prípadov

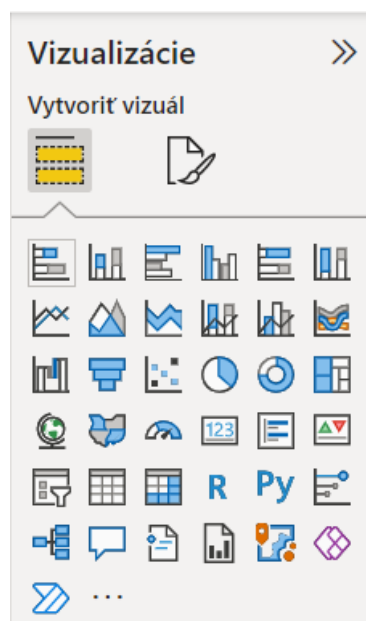
```
1 Potvrdené prípady =  
2 CALCULATE([Všetky prípady] , 'Status'[Status]="Confirmed")
```

Zdroj: Vlastné spracovanie

3.4 Tvorba vizuálov

Následne po tom, čo údaje dostaneme do požadovaného stavu a máme taktiež vypočítané všetky potrebné hodnoty, je možné začať vytvárať vizuály v časti **Vizualizácie**, Obrázok 10, ktorú môžeme nájsť v pravej časti obrazovky aplikácie Power BI Desktop alebo môžeme jednoducho presunúť vybrané pole na plátno zostavy a aplikácia automaticky zvolí graf, ktorý sa podľa nej ku danému poľu hodí. Používateľ si tento graf môže ponechať alebo následne vybrať taký, ktorý mu bude najviac vyhovovať.

Obrázok 10: Časť Vizualizácie

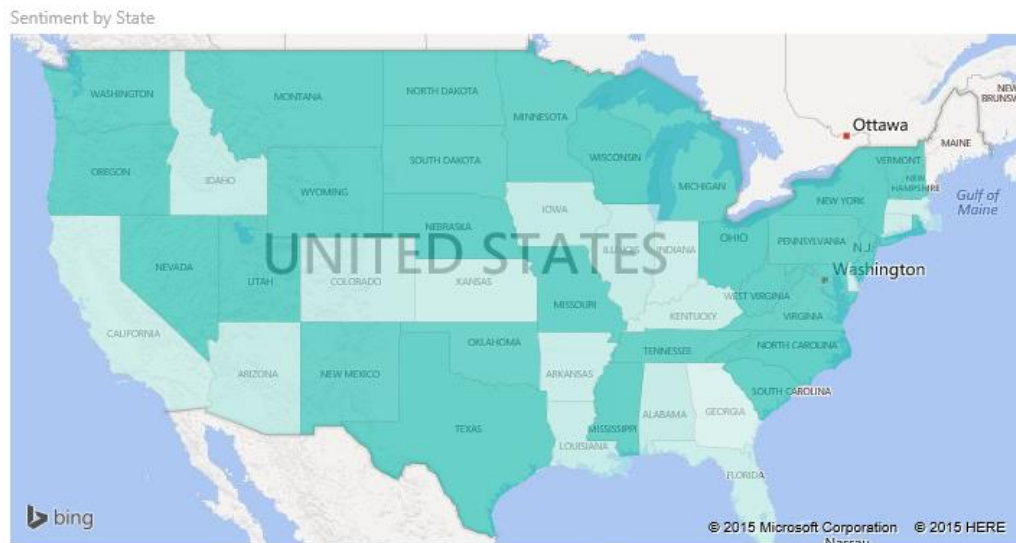


Zdroj: Vlastné spracovanie

Power BI Desktop ponúka ako už bolo spomínané široký výber základných vizuálov, konkrétne 37, napríklad:

- mapa, Obrázok 11,
- stĺpcový graf,
- čiarový graf,
- tabuľka,
- vizuál R skriptu,
- dekompozičný strom, atď.

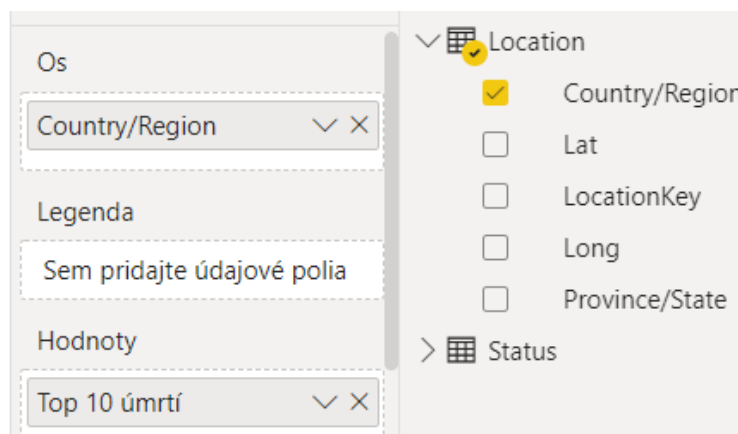
Obrázok 11: Ukážka vizuálu - Mapa



Zdroj: Microsoft, 2022(b)

Aby sa vo vizuáloch zobrazovali požadované informácie, stačí jednoducho presunúť vybrané údajové polia do daných kategórií, príklad je zobrazený na Obrázku 12, kde Krajina/Región bol presunutý do kategórie s názvom Os, čo znamená, že krajiny alebo regióny budú zobrazené na osi. Nie každá vizualizácia však ponúkala zobrazenie hodnôt pomocou osi, ako to bolo v tomto prípade. Napríklad pri vizualizácii Mapa bolo možné vložiť polia určujúce miesto, zemepisnú šírku, zemepisnú dĺžku, veľkosť bubliny a popisy alebo pri zobrazení dát cez Kľúčové vplyvy bolo dôležité vložiť pole, ktoré chceme analyzovať a podľa akého aspektu, ale taktiež môžeme vložiť pole, určujúce spôsob rozbalenia. Vkladanie polí do daných oblastí je dôležitým faktorom pri správnom grafickom zobrazení údajov.

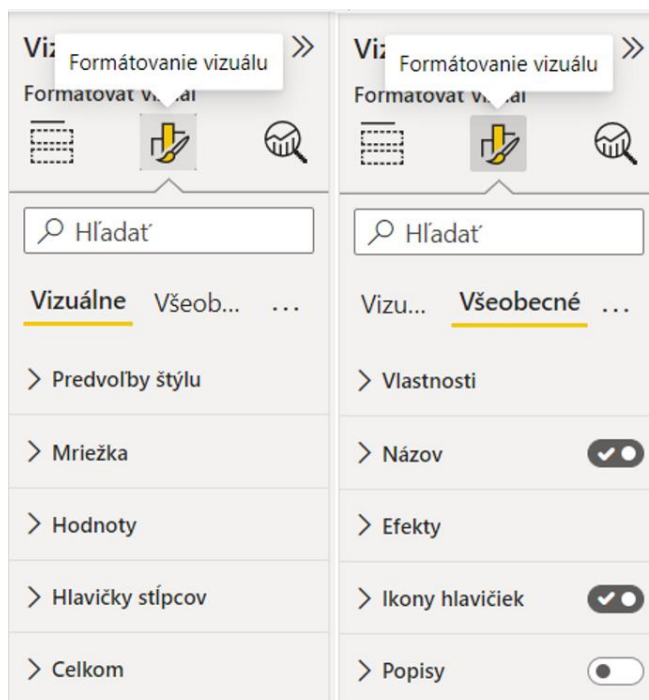
Obrázok 12: Výber údajových polí



Zdroj: Vlastné spracovanie

Ďalej je taktiež možné upravovať formát vizuálov po stlačení možnosti **Formátovanie vizuálu**. Na Obrázku 13 vidíme, že aplikácia rozdelila formátovanie vizuálu na **Vizuálne** (napríklad zmena farby hodnôt) a **Všeobecné** (napríklad zmena názvu grafu, či jeho odstránenie z vizuálu).

Obrázok 13: Úprava formátu



Zdroj: Vlastné spracovanie

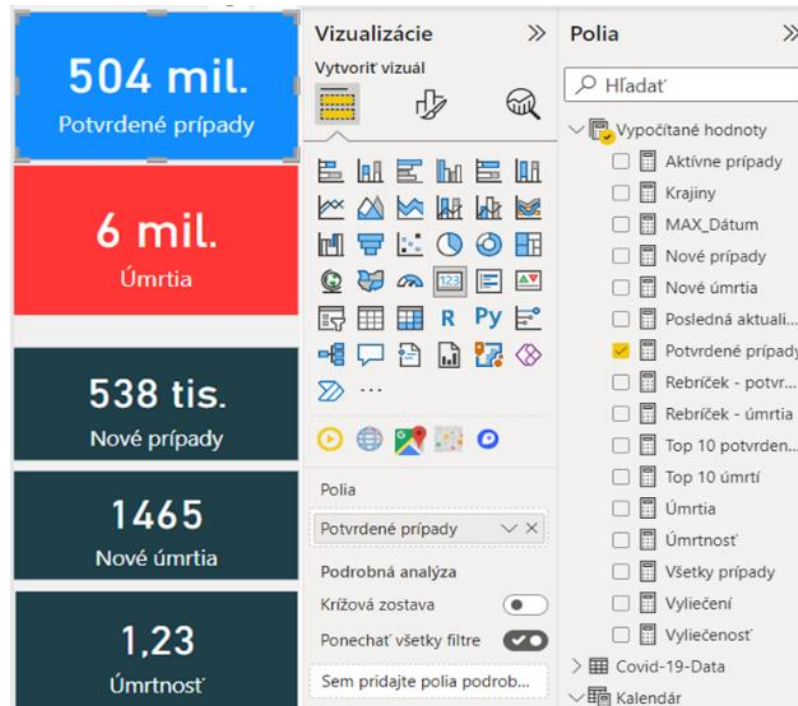
Vzhľadom na oboznámenie sa so všeobecným postupom vytvárania vizuálov je možné zamerať sa na jednotlivé konkrétne typy vizualizácií - využité metódy a ich špecifikácie. Pre lepšiu predstavivosť budú tieto vizualizácie popísané na našich získaných údajoch a následne budú aj zakomponované vo finálnych stránkach zostavy.

Vizuál - Karta

Prvý vizuál, ktorý sme vytvorili sa nazýva Karta. Vo všeobecnosti je úlohou tohto typu vizualizácie vyjadriť nejakú hodnotu či text alebo aj oboje naraz ako to je zobrazené na snímku obrazovky, Obrázok 14. Z časti **Vizualizácie** sme si zaklikli vybraný vizuál - Karta, čím sa nám automaticky pridal do zostavy. Následne stačilo len zakliknúť vybrané pole v časti **Polia** alebo ho presunúť do ľavej časti a hodnota sa zobrazila v našom okne. K formátovaniu vizuálu sme sa dostali kliknutím na druhú ikonu - **Formátovanie vizuálu** v hornej časti **Vizualizácie**. V rámci formátovania sme si zmenili farbu písma aj farbu

pozadia, ale je možné zmeniť aj typ či veľkosť písma, pridať alebo odstrániť názov vizuálu, zobraziť alebo odstrániť ikony hlavičiek a mnohé iné, v závislosti aj od typu vizualizácie. Rovnako je možné upravovať veľkosť vizuálu v zostave.

Obrázok 14: Tvorba vizuálu Karta



Zdroj: Vlastné spracovanie

Vizuál - Rýchly filter

Ďalej sme využili vizuál Rýchly filter, do ktorého sme vložili pole s dátumami. Tento vizuál však nemusí vyzerat' rovnako, ako to vidíme na Obrázku 15. Aplikácia nám totižto poskytuje rôzne možnosti, ako by naše vložené pole mohla filtrovať: Pred, Po, Zoznam, Rozbaľovací zoznam, Relatívny dátum, Relatívny čas a Medzi, ktoré sme aplikovali v našom zobrazení.

Obrázok 15: Vizuál Rýchly filter



Zdroj: Vlastné spracovanie

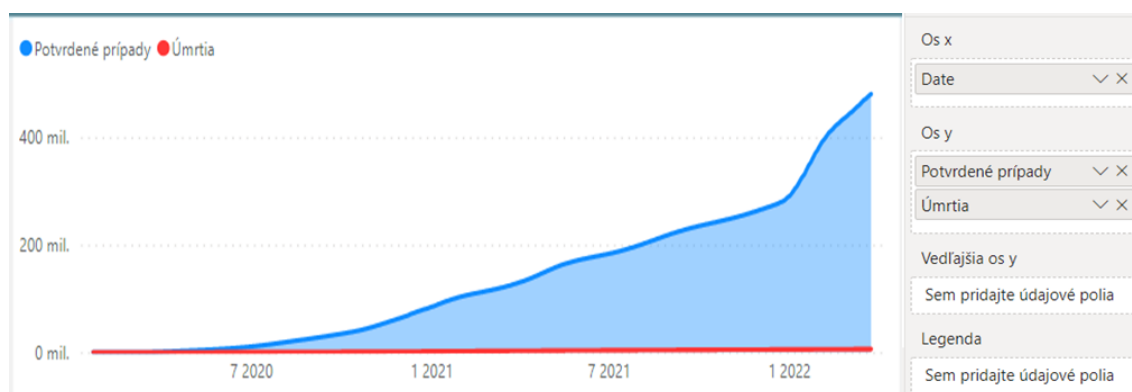
Vizuál - Graf

Neoddeliteľnou súčasťou vizualizácií údajov sú grafy, ktoré v početnom množstve ponúka aj prostredie aplikácie Power BI Desktop. Tvorbu grafov sme realizovali v nasledujúcej postupnosti:

1. Výber grafu.
2. Výber polí, zobrazených na grafe.
3. Formátovanie grafu.

Jedinú odlišnosť s doposiaľ znázornenými grafmi môžeme vidieť na snímku obrazovky, Obrázok 16, kde spolu s výsledným grafom je zobrazená časť karty **Vizualizácie**. Na rozdiel od predošlých, kde bol riadok na vkladanie poľa len jeden, pri grafoch je ich viac a bolo na našej voľbe, ktoré pole chceme na akej osi zobraziť.

Obrázok 16: Tvorba vizuálu Plošný graf



Zdroj: Vlastné spracovanie

Vizuál - Tabuľka

Medzi časté vizualizačné prostriedky by sme mohli zaradiť aj klasickú tabuľku, ako sme ju zostrojili na Obrázku 17. Vytvoriť samotnú tabuľku bolo jednoduché - presunutím polí do daného vizuálu. Dôležité pri tabuľke bolo ju používateľsky naformátovať, čo znamená, aby pre používateľa nebola nezaujímavá. Na to sme využili podmienené formátovanie čísel, kde sfarbenie malo vyjadrovať situáciu s ochorením COVID-19 v danej krajine. Červená farba mala evokovať „zlú“ situáciu a zelená farba evokovala „dobrú“ situáciu. Doplnili sme taktiež nadpis tabuľky, striedavé zafarbenie riadkov a krajiny v tabuľke sme zoradili podľa počtu potvrdených prípadov.

Obrázok 17: Vizuál Tabuľka

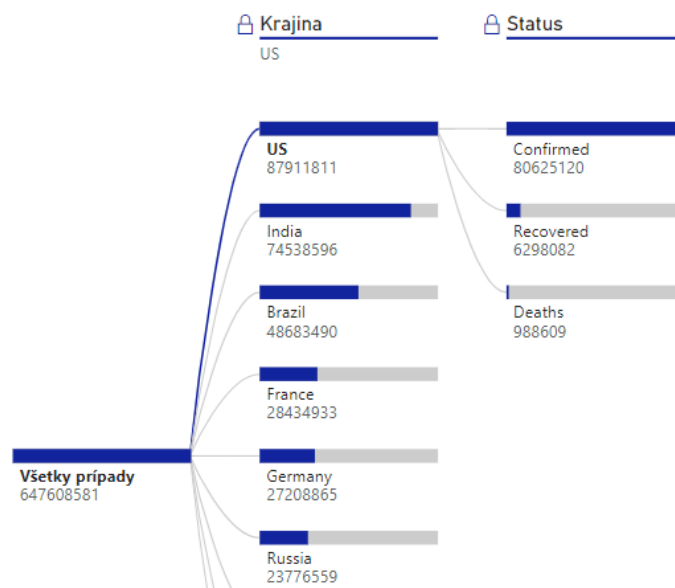
Všetky potvrdené prípady/úmrтия na krajinu		
Country/Region	Úmrтия	Potvrdené prípady
US	976652	79946097
India	521004	43019453
Brazil	659046	29839103
France	142664	25106739
United Kingdom	165049	20848913
Germany	127599	19492672
Russia	359693	17481799
Turkey	97736	14789483
Italy	158700	14304111
Korea, South	14899	11815841
Spain	102392	11451676
Argentina	127904	9025257
Vietnam	42258	8919557
Netherlands	22508	7904943

Zdroj: Vlastné spracovanie

Vizuál - Dekompozičný strom

Zaujímavým spôsobom vyjadruje údaje v grafickej podobe aj nasledujúci dekompozičný strom, Obrázok 18. Tento grafický vizuál sa však nedá použiť vždy, pretože musí vo výsledku dávať určitý zmysel. Je dôležité si najprv premyslieť, ktorý údaj by sme chceli analyzovať a takpovediac si ho rozobrať na drobné. V našom prípade to bol atribút Všetky prípady. Tie sme postupne mali rozdelené a vysvetlené podľa krajiny a po kliknutí na určitú krajinu sme mohli vidieť štatistiky danej krajiny v jednom grafe podľa statusu.

Obrázok 18: Vizuál Dekompozičný strom

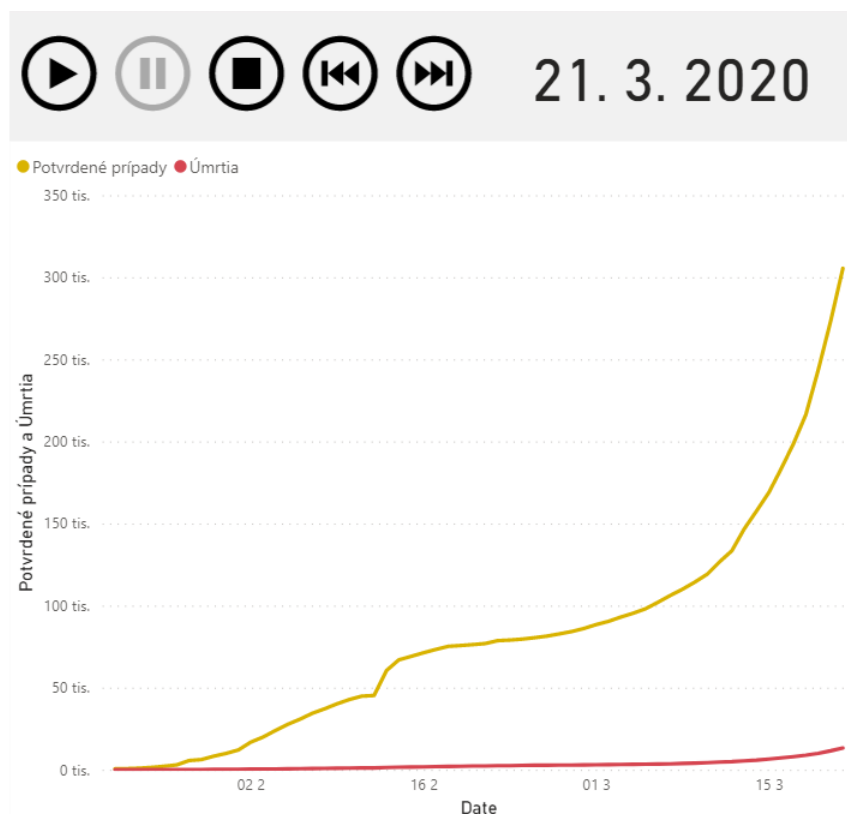


Zdroj: Vlastné spracovanie

Vizuál - Play Axis

Posledný vizuál, vyjadrujúci ďalšiu metódu využítú v záverečnej práci a ktorého tvorbu si bližšie popíšeme môžeme vidieť na Obrázku 19. Ide o vizuál, ktorý je animačný a nie je zaradený medzi základné, a preto sme si ho doplnili cez tlačidlo **Načítať ďalšie vizuály**. Vzhľadom na jeho charakteristiku animácie sme celý priebeh nemohli zachytiť na snímku obrazovky ale môžeme vidieť, že vedľa „prehrávača“ je dátum, ktorý má vyjadrovať, do akého času sa práve v tom momente zobrazuje priebeh pandémie na grafe. Ešte treba dodať, že os x má začiatkový bod fixný, ktorým je prvý dátum v rámci našej tabuľky údajov a takpovediac posledný bod sa mení stále počas toho ako je animácia spustená. Na os y sme pridali Potvrdené prípady a Úmrtia a zmenili im aj farby. Spomenúť musíme taktiež vytvorený kalendár, ktorý sme špeciálne pre tento graf pridávali do dátového modelu. Kalendár, ktorý sme neprepájali so žiadnou inou tabuľkou sme vložili do vizuálu „prehrávača“ a do vizuálu zobrazujúceho dátum v hornej časti, pretože ak v týchto dvoch vizuáloch by bol pridaný dátum z tabuľky „Covid-19-Data“, potom by graf nezobrazoval údaje ako animáciu ale vždy by zobrazil iba dva samostatné body, vyjadrujúce počet v daný deň, ktoré by sa síce menili v čase ale nie ako priamka.

Obrázok 19: Animačný vizuál



Zdroj: Vlastné spracovanie

3.5 Vytváranie dashboardov

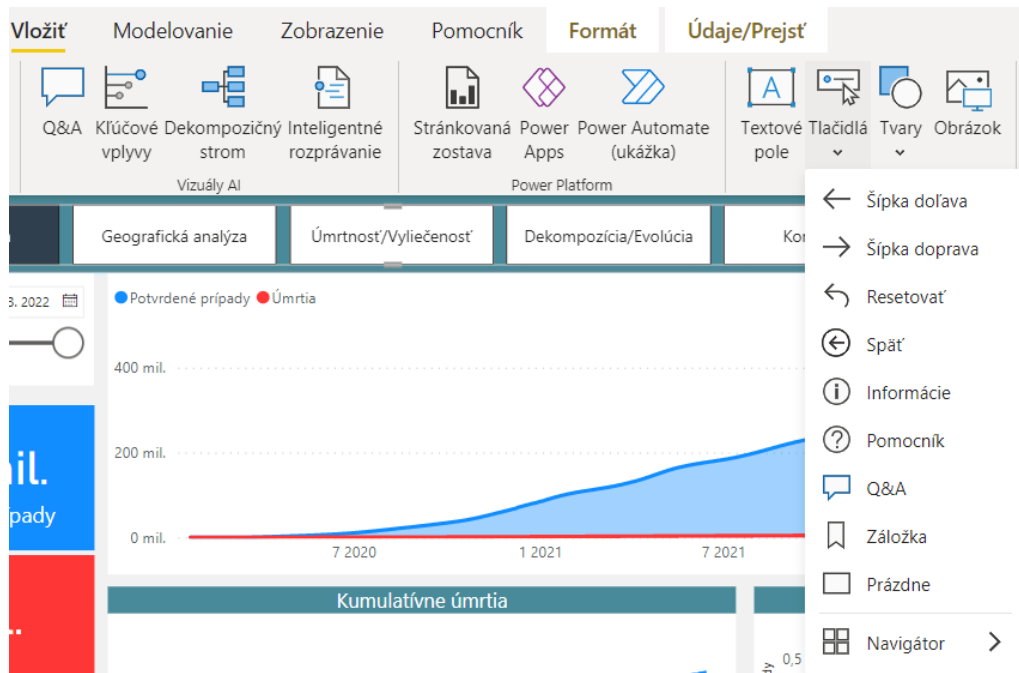
Zostavy, či inak povedané dashboardy predstavujú kolekciu vizuálov zobrazujúcich rôzne aspekty údajov v jednom súbore Power BI Desktop. Zostavy môžu byť tvorené jednou alebo viacerými stránkami (podobne ako aj Excel súbory môžu mať viac hárkov). Jednotlivé vizuály sú medzi sebou prepojené, čo umožňuje ich interaktivitu - teda pri zakliknutí určitej možnosti v jednom vizuále to ovplyvní všetky ostatné vizuály na danej strane (Maslyuk, 2018).

Medzi dôležité súčasti dashboardov, ktoré však nemožno charakterizovať ako vizuály, patria rôzne tlačidlá či navigačné panely. Ide o metódy, ktoré taktiež majú vlastnosť interaktivity, ale nenachádzajú sa v rámci zoznamu vizualizácií.

Tlačidlo - Navigátor

Je možné ho pridať do zostavy cez kartu **Vložiť**, časť **Prvky** a následne zvoliť **Tlačidlá**, vďaka čomu nám aplikácia ponúka voľbu rôznych tlačidiel spolu s tlačidlom **Navigátor**. Následne je možné vybrať, či chceme vytvoriť navigátor všetkých záložiek alebo ako to bolo v našom prípade - navigátor všetkých stránok zostavy, Obrázok 20.

Obrázok 20: Tvorba navigátora

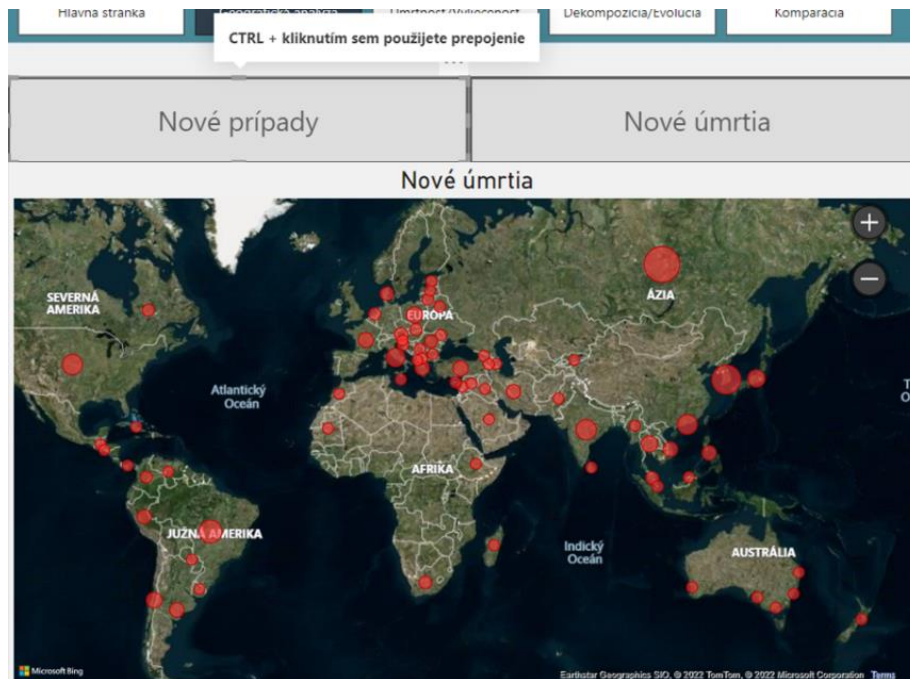


Zdroj: Vlastné spracovanie

Tlačidlo - Prázdne tlačidlo

Prázdne tlačidlo je možné modifikovať úplne podľa svojich predstáv - zvoliť si vlastný text, formátovanie, štýl a akciu tlačidla. Pri zostrojovaní našich konkrétnych zostáv sme vytvorili dve Prázdne tlačidlá, do ktorých sme vložili text, ako to možno vidieť na Obrázku 21. Pod tlačidlá sme umiestnili dve grafické zobrazenia - Mapy, s tým že jedna mapa vyjadrovala počet nových prípadov a druhá mapa počet nových úmrtí. Ak by sme však tieto dve mapy umiestnili vedľa seba, zabralo by to všetok dostupný priestor na stránke. Preto sme mapy umiestnili tak, aby prekrývala jedna druhú a tlačidlám pridali funkciu akcie s prepínaním na vybraný vizuál. To znamená, že používateľ si sám zvolí, ktorú mapu chce vidieť pomocou dostupných tlačidiel.

Obrázok 21: Tvorba prázdneho tlačidla



Zdroj: Vlastné spracovanie

Výsledne dashboardy, ktorých súčasťou sú už predstavené vizuály, budú zobrazené a charakterizované ako celok v závere nasledujúcej kapitoly.

4 Výsledky práce a diskusia

Jedným z cieľov bakalárskej práce bola praktická ukážka a implementácia aplikácie Power BI Desktop, ktorá je využívaná najmä na analýzu a vizualizáciu určitých údajov. Na demonštráciu tohto cieľa sme sa v rámci praktickej časti zamerali na tému dnešných dní, ktorou je pandémia a s ňou súvisiace údaje. Pracovali sme s databázou údajov, ktoré nám poskytujú informácie o počte prípadov s ochorením COVID-19 naprieč všetkými krajinami. Údaje sú extrahované a zbierané z databáz štátov či štátnych zdravotníckych organizácií po celom svete a zverejnené prostredníctvom platformy GitHub. Databáza je denne aktualizovaná a je spravovaná univerzitou v meste Baltimore, konkrétne Centrom pre systémové vedy a inžinierstvo, ktoré je pod záštitou tejto univerzity.

4.1 Načítanie údajov a ich úpravy, tvorba modelu

Naším zdrojom údajov bola webová stránka, z ktorej sme potrebné uchovávané údaje importovali. Ešte treba pripomenúť, že informácie o počte prípadov boli rozdelené do troch súborov - vyliečené prípady, prípady s úmrtím a potvrdené prípady. Pripojenie sa k údajom bolo uskutočnené prostredníctvom URL adresy spomínanej webovej stránky (<https://github.com/CSSEGISandData/COVID-19>).

Načítanie údajov sme začali rovnako, ako to bolo popísané v kapitole 3.1. Spomedzi skupiny zdrojov s názvom **Iné**, sme si vybrali možnosť **Web** a následne sme boli vyzvaní k zadaniu URL adresy, ako sme to už videli na Obrázku 4.

Po potvrdení tlačidlom **OK** nám bola poskytnutá ukážka údajov, Obrázok 22, ako budú údaje vyzerat' v tabuľkovom formáte. V hornej časti okna sme mohli vidieť URL adresu, na ktorú sme sa pripojili, spolu s typom kódovania súboru, oddeľovačom jednotlivých informácií a počtom riadkov, na základe ktorých aplikácia zistila typ údajov. Vzhľadom na to, že údaje boli uložené vo formáte CSV a v riadkoch boli oddeľované čiarkami, sme zvolili ako oddeľovač práve interpunkčné znamienko - čiarka.

Obrázok 22: Vzorka údajov pri načítaní z webovej stránky

https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_d...

Pôvod súboru: 65001: Unicode (UTF-8) | Oddeľovač: Čiarka | Zisťovanie typu údajov: Na základe prvých 200 riadkov

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12
Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20
	Afghanistan	33.93911	67.709953	0	0	0	0	0	0	0	0
	Albania	41.1533	20.1683	0	0	0	0	0	0	0	0
	Algeria	28.0339	1.6596	0	0	0	0	0	0	0	0

Ukážka stiahnutá dňa utorok 30. novembra 2021

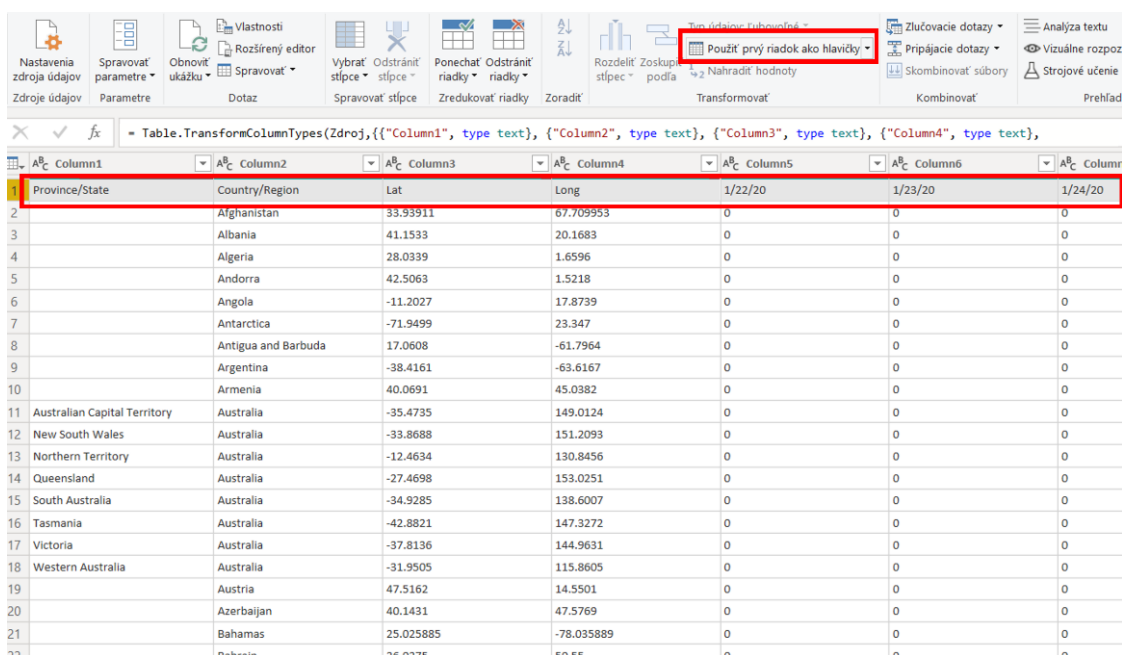
Extrahovať tabuľky pomocou príkladov | Načítať | Transformovať údaje | Zrušiť

Zdroj: Vlastné spracovanie

V dolnej časti nám už len boli ponúknuté tri možnosti - **Načítať**, **Transformovať údaje** alebo **Zrušiť**, čím by sme zatvorili toto okno a žiadne údaje by neboli do aplikácie extrahované. Ak by sme si vybrali, že údaje priamo načítame do prostredia Power BI Desktop, tak by boli načítané presne v takom stave v akom sú zaznamenávané do CSV súboru. My sme si však zvolili, že ešte pred samotným načítaním si tieto údaje chceme transformovať na takú podobu, v ktorej nám budú najviac vyhovovať, aby sa nám s nimi čo najlepšie pracovalo.

Výberom transformácie údajov nám aplikácia sprístupnila Editor Power Query, ktorého rozhranie môžeme vidieť na Obrázku 23. Editor nám umožnil vykonať rôzne typy zmien, jednou z nich bolo použitie prvého riadku ako hlavičky jednotlivých stĺpcov, ktorú vidíme taktiež na tomto obrázku.

Obrázok 23: Úprava údajov v editore Power Query



Zdroj: Vlastné spracovanie

Ďalšou dôležitou úpravou bolo zrušenie kontingenčnosti stĺpcov, okrem prvých štyroch - Provincia/Štát, Krajina/Región, Zemepisná šírka a Zemepisná dĺžka. Zrušenie teda nastalo v stĺpcoch, ktorých názov sa odvíjal od dátumu, v ktorom bol zaznamenaný určitý počet prípadov. Týmto krokom sa zmenila štruktúra tabuľky údajov takým spôsobom, že na rozdiel od Obrázku 23, na ktorom bola v každom riadku iná krajina, ku ktorej prislúchali hodnoty v stĺpcoch podľa dátumov, teda stĺpcov bolo podstatne viac ako riadkov, po tejto zmene bol celkový počet stĺpcov znížený na šesť. K prvým štyrom stĺpcom, u ktorých sme túto zmenu nevykonali sa pridali dva - jeden bol naplnený dátumami, ktoré predtým predstavovali názvy stĺpcov a druhý obsahoval číselné hodnoty vyjadrujúce počty prípadov. Nadväzujúc na tento krok sme pozmenili názov stĺpca, do ktorých sa transformovali dátumy - „Date“, Obrázok 24.

Obrázok 24: Úprava údajov v editore Power Query(2)

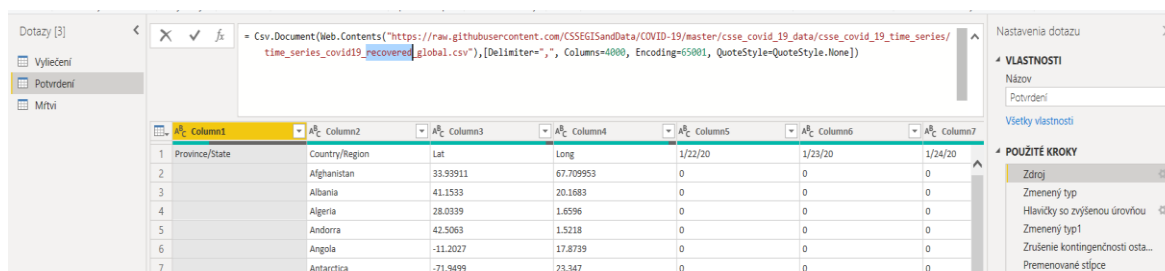
Province/State	Country/Region	Lat	Long	Date	Hodnota
	Afghanistan	33.93911	67.709953	1/22/20	0
	Afghanistan	33.93911	67.709953	1/23/20	0
	Afghanistan	33.93911	67.709953	1/24/20	0
	Afghanistan	33.93911	67.709953	1/25/20	0
	Afghanistan	33.93911	67.709953	1/26/20	0
	Afghanistan	33.93911	67.709953	1/27/20	0

Zdroj: Vlastné spracovanie

Ako sme v úvode tejto kapitoly spomínali, pracovali sme s tromi rôznymi súbormi. Aby sme však nemuseli predošlé zmeny aplikovať postupne na všetky tri súbory, vykonali sme zmeny na súbore, ktorý uchovával informácie o počte vyliečených ľudí z ochorenia COVID-19. Keď už sme mali tento súbor v požadovanom stave, tak po pravom kliknutí myšou na dotaz s názvom Vyliečení sme si vybrali, že chceme tento dotaz duplikovať a urobili sme to dvakrát. Duplikované súbory sme pomenovali Potvrdení a Mŕtvi. Aby však aj hodnoty správne vyjadrovali počet prípadov, teda počet potvrdených a počet mŕtvych, tak bolo potrebné urobiť nasledujúcu zmenu, ktorú môžeme vidieť na Obrázku 25.

Webová adresa, ktorú sme uviedli ako zdroj, bola totižto špecifická pre prípady, ktoré boli vyliečené z tohto ochorenia. Ak sme teda chceli vidieť čísla, reprezentujúce celkový počet potvrdených prípadov alebo čísla, reprezentujúce úmrtia, museli sme zmeniť zvýraznenú časť webovej adresy - „recovered“ na „confirmed“ v prvom prípade a „deaths“ v druhom prípade.

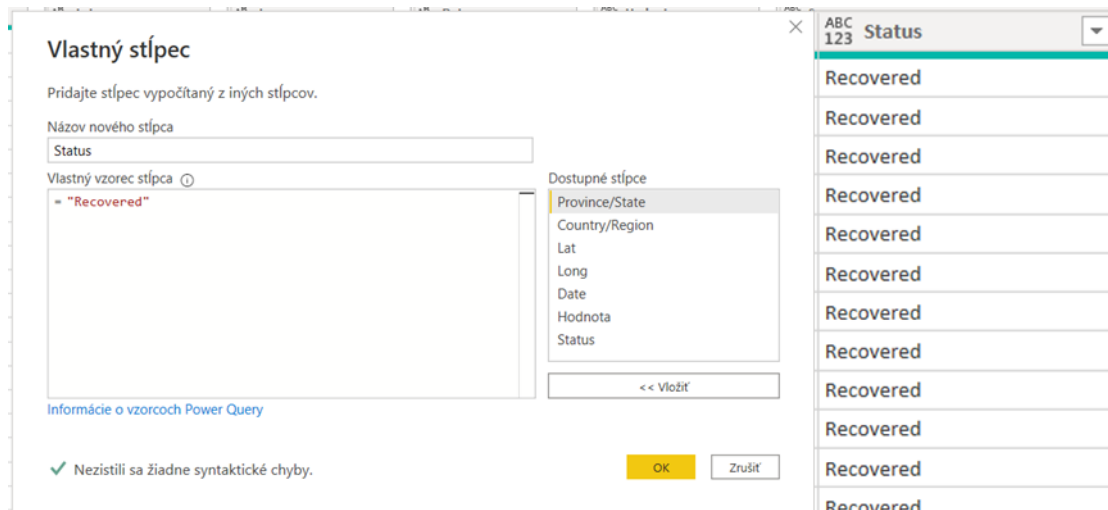
Obrázok 25: Úprava údajov v editore Power Query(3)



Zdroj: Vlastné spracovanie

Dodatočnú úpravu, ktorú by nebolo efektívne vykonať ešte pred duplikáciou dotazu, sme uplatnili na tieto tri dotazy zvlášť. Chceli sme totižto doplniť nový stĺpec pre každý dotaz, ktorý by bol naplnený hodnotou, v tomto prípade textom, vyjadrujúcim typ prípadu. V karte **Pridať stĺpec**, časť **Všeobecné**, sme vybrali **Vlastný stĺpec** a ďalej sme pracovali s nasledujúcim oknom, Obrázok 26, ktorý nám umožnil vytvorenie nového stĺpca v danej tabuľke na základe vlastného vzorca. Vo všetkých troch dotazoch sme novovytvorený stĺpec pomenovali Status a podľa toho, v ktorej tabuľke sme práve stĺpec vytvárali, sme do časti Vlastný vzorec stĺpca napísali „Recovered“, „Confirmed“ alebo „Deaths“. Vzorec sme potvrdili stlačením **OK**. V pravej časti Obrázku 26 rovnako môžeme vidieť výsledok použitia takéhoto vzorca.

Obrázok 26: Úprava údajov v editore Power Query(4)



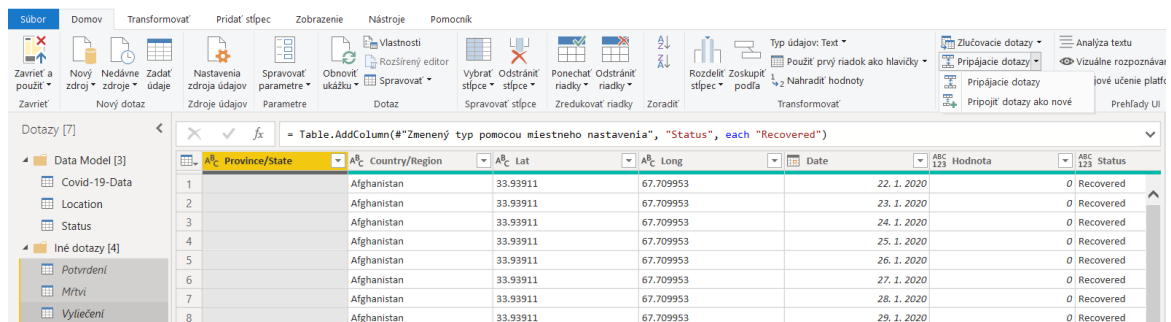
Zdroj: Vlastné spracovanie

Aby sme už ďalej nemuseli pracovať s týmito tromi dotazmi zvlášť, urobili sme si z nich jednu komplexnú tabuľku. Avšak predtým sme si najprv vytvorili novú skupinu, do ktorej sme plánovali novú tabuľku vložiť. Docielili sme to nasledujúcimi krokmi:

1. Právý klik myšou v ľavej časti - Dotazy.
2. Výber možnosti - **Nová skupina...**
3. Vyplnenie názvu novej skupiny a popisu, ktorý bol v našom prípade rovnaký - „DataModel“

Po vytvorení skupiny sme mohli prísť k zlúčeniu našich tabuliek. Označili sme si všetky tri tabuľky, ako to vidíme na Obrázku 27 a v karte **Domov**, časť **Kombinovať**, sme vybrali **Pripájacie dotazy - Pripojiť dotazy ako nové**.

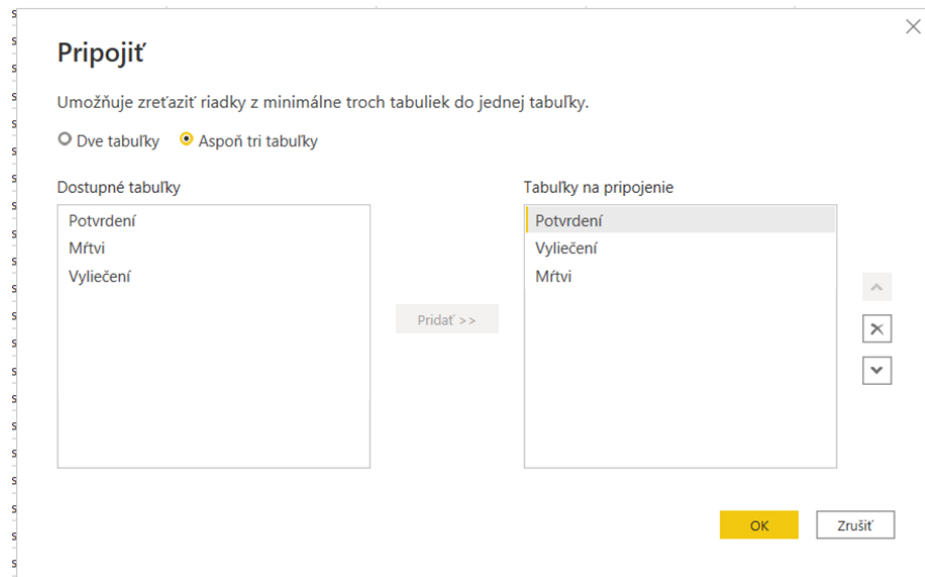
Obrázok 27: Úprava údajov v editore Power Query(5)



Zdroj: Vlastné spracovanie

V nasledujúcom okne, Obrázok 28, ktorý nám bol zobrazený po predošlých krokoch, sme si zvolili že chceme zreťaziť riadky z minimálne troch tabuliek, v našom prípade išlo konkrétne o tri. Tabuľky z ľavej časti - **Dostupné tabuľky** sme do pravej časti - **Tabuľky na pripojenie** dostali tým, že sme na každú z nich klikli a následne tlačidlom **Pridať** sme ich presunuli. Zmenu sme aj v tomto prípade potvrdili tlačidlom **OK**.

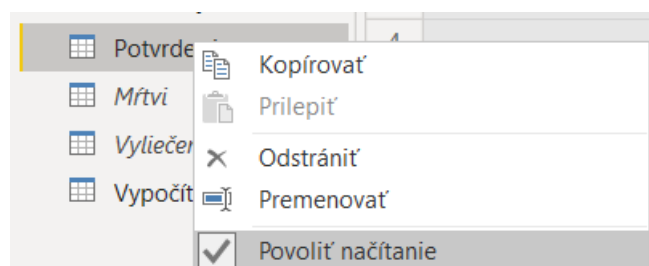
Obrázok 28: Úprava údajov v editore Power Query(6)



Zdroj: Vlastné spracovanie

Po tom ako sme umožnili prepojenie troch tabuliek do jednej, sme si túto novovytvorenú tabuľku premenovali na „Covid-19-Data“ a presunuli tabuľku do skupiny „DataModel“. Keďže v ďalších častiach práce sme už nepotrebovali pracovať s pôvodnými tabuľkami, pretože všetky údaje z nich budú načítavané do spoločnej tabuľky, tak sme u pôvodných mohli znepřístupniť povolenie načítavania, Obrázok 29.

Obrázok 29: Úprava údajov v editore Power Query(7)

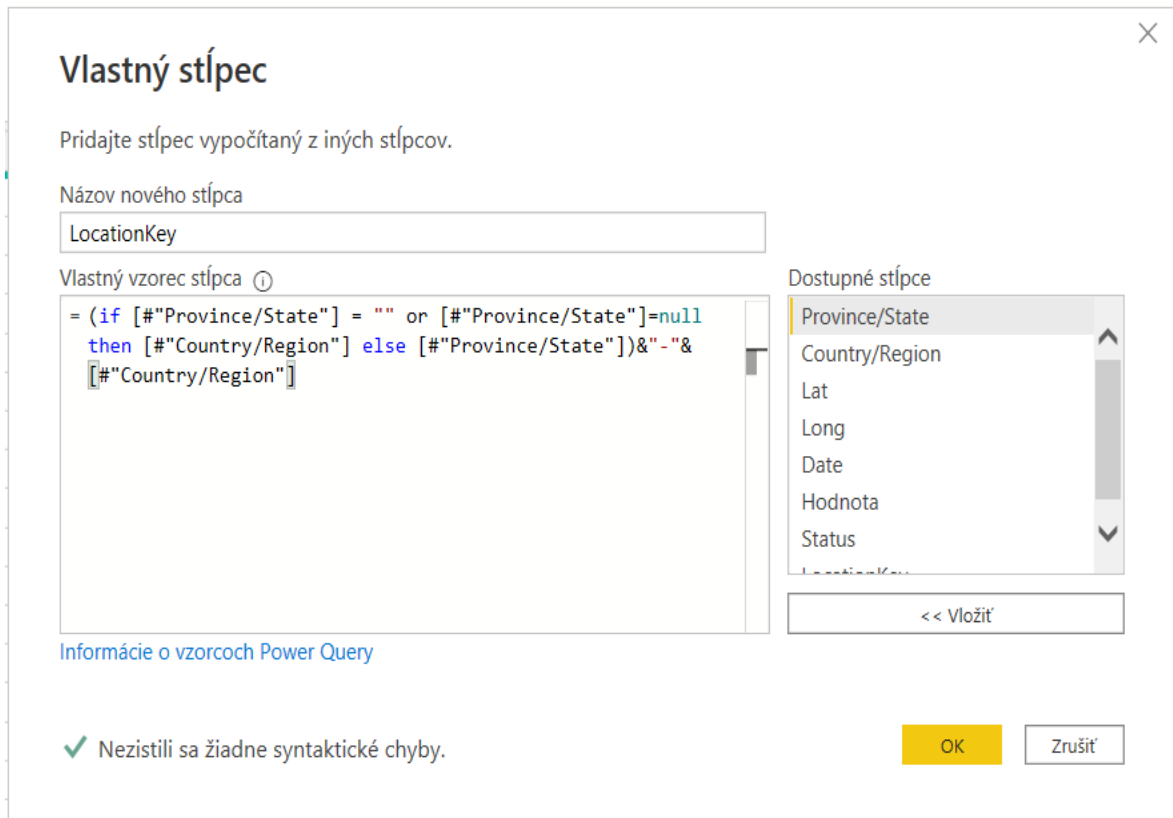


Zdroj: Vlastné spracovanie

Nasledujúce zmeny a úpravy súviseli s vytvorením novej tabuľky, ktorá by uchovávala informácie výhradne len o lokácii. Ešte predtým ako sme však začali so samotným vytváraním novej tabuľky, urobili sme nasledujúce zmeny. Prvou zmenou bolo pridanie nového stĺpca do tabuľky „Covid-19-Data“ na jednoznačnú a unikátnu identifikáciu miesta či krajiny, ku ktorej prislúcha daný počet prípadov. Stĺpec sme pridali rovnakým spôsobom, ako sme to urobili pri stĺpci Status avšak v tomto prípade bol vzorec o niečo zložitejší.

Na Obrázku 30 môžeme vidieť, že názov nového stĺpca je „LocationKey“ a ak by sme analyzovali vzorec, tak riadky tohto stĺpca budú naplnené v závislosti od stĺpcov „Province/State“ a „Country/Region“, oddelené pomlčkou. Viaceré krajiny však v stĺpci provincia/ štát nemajú uvedenú žiadnu hodnotu, teda riadky pre tieto krajiny sú prázdne a v takom prípade bude nový stĺpec obsahovať dvakrát názov krajiny, oddelený pomlčkou, napríklad „Slovakia-Slovakia“. Funkciu unikátnosti tento stĺpec plnil najmä pri krajinách, ktoré boli v rámci databázy rozdelené na väčšie územné celky - provincie, ako v prípade USA či Austrálie, kde riadok mohol vyzerat' napríklad takto „Victoria-Australia“.

Obrázok 30: Úprava údajov v editore Power Query(8)



Zdroj: Vlastné spracovanie

V tomto stave sme mali všetko pripravené na to, aby sme mohli vytvoriť novú, už spomínanú tabuľku. Vytvorili sme ju duplikovaním tabuľky „Covid-19-Data“ a následne premenovali na názov „Location“. Ak by sme však novú tabuľku ponechali s takýmito údajmi, jej vytvorenie by bolo zbytočné, vzhľadom na to, že by sme mali dve identické tabuľky. Aby spĺňala svoj účel, teda zaznamenávať iba názvy krajín, sme označili stĺpce „Date“, „Hodnota“ a „Status“ a odstránili ich z novej tabuľky.

Doplňujúci krok môžeme vidieť na nasledujúcom Obrázku 31. Keďže v tabuľke, ktorú sme duplikovali sa každá krajina nachádzala toľkokrát, koľko bolo zaznamenaných dní doposiaľ, odkedy začala pandémia vo svete, v novej tabuľke „Location“ by evidovanie v takomto počte nemalo zmysel. Preto sme označili stĺpec „LocationKey“ a po pravom kliknutí myšou sme vybrali možnosť **Odstrániť duplikáty**. Na snímke môžeme vidieť už tabuľku po vykonanej zmene a rovnako aj to, že každá krajina sa vyskytuje v tabuľke iba raz, s výnimkou krajín, ktoré sú rozdelené na ďalšie územné celky a v takom prípade sa názov krajiny opakuje v závislosti od počtu provincií, čo bolo našim cieľom.

Obrázok 31: Úprava údajov v editore Power Query(9)

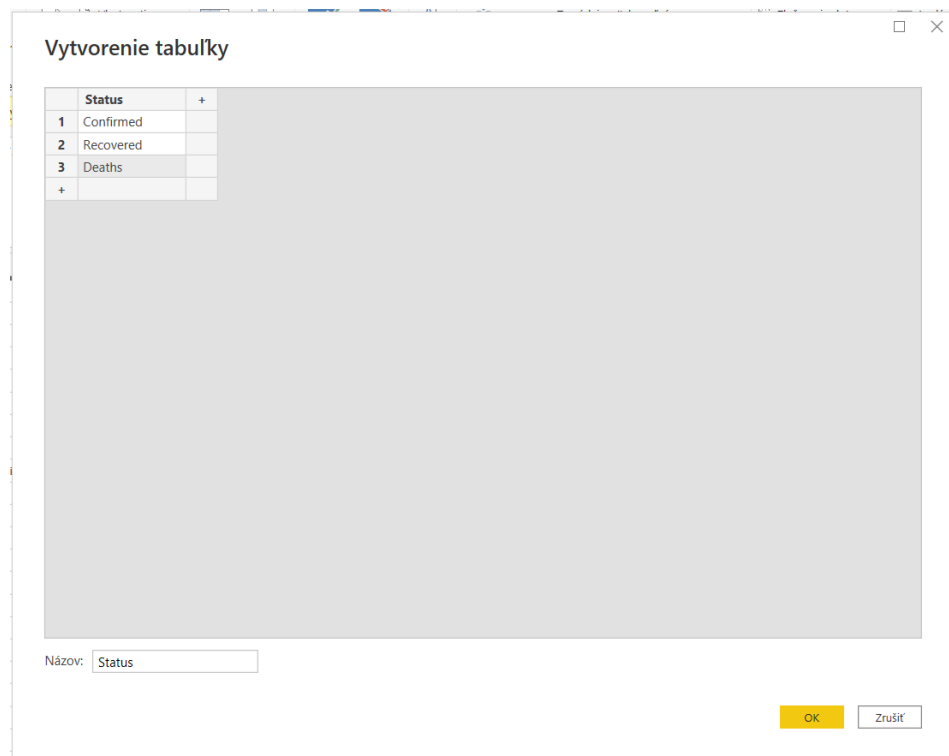
	A _C Province/State	A _C Country/Region	A _C Lat	A _C Long	A _C LocationKey
1		Afghanistan	33.93911	67.709953	Afghanistan-Afghanistan
2		Albania	41.1533	20.1683	Albania-Albania
3		Algeria	28.0339	1.6596	Algeria-Algeria
4		Andorra	42.5063	1.5218	Andorra-Andorra
5		Angola	-11.2027	17.8739	Angola-Angola
6		Antigua and Barbuda	17.0608	-61.7964	Antigua and Barbuda-Antigua and Barbuda
7		Argentina	-38.4161	-63.6167	Argentina-Argentina
8		Armenia	40.0691	45.0382	Armenia-Armenia
9	Australian Capital Territory	Australia	-35.4735	149.0124	Australian Capital Territory-Australia
10	New South Wales	Australia	-33.8688	151.2093	New South Wales-Australia
11	Northern Territory	Australia	-12.4634	130.8456	Northern Territory-Australia
12	Queensland	Australia	-27.4698	153.0251	Queensland-Australia
13	South Australia	Australia	-34.9285	138.6007	South Australia-Australia
14	Tasmania	Australia	-42.8821	147.3272	Tasmania-Australia
15	Victoria	Australia	-37.8136	144.9631	Victoria-Australia
16	Western Australia	Australia	-31.9505	115.8605	Western Australia-Australia
17		Austria	47.5162	14.5501	Austria-Austria
18		Azerbaijan	40.1431	47.5769	Azerbaijan-Azerbaijan
19		Bahamas	25.025885	-78.035889	Bahamas-Bahamas
20		Bahrain	26.0275	50.55	Bahrain-Bahrain
21		Bangladesh	23.685	90.3563	Bangladesh-Bangladesh
22		Barbados	13.1939	-59.5432	Barbados-Barbados

Zdroj: Vlastné spracovanie

Poslednú zmenu, ktorú sme ešte pred samotným načítaním údajov do aplikácie urobili bolo vytvorenie novej tabuľky. Tak ako sme v predošlých krokoch vytvorili špecifickú tabuľku len pre krajiny a štáty, rovnako sme chceli mať unikátnu tabuľku pre statusy, ktorých počet bol v tomto prípade len tri. Vzhľadom na jednoduchosť tejto tabuľky nebolo potrebné vytvárať vzorec, či duplikovať iné tabuľky. V karte **Domov**, časť **Nový**

dotaz sme zvolili **Zadat' údaje** a ďalej sme pokračovali v okne, zobrazenom na snímku obrazovky, Obrázok 32. Identicky sme pomenovali stĺpec aj názov tejto tabuľky. Zostalo nám už len naplniť tabuľku hodnotami - „Confirmed“, „Recovered“ a „Deaths“. Následne sme zmenu potvrdili stlačením **OK**, čím sa nám nová tabuľka pridala do dátového modelu.

Obrázok 32: Úprava údajov v editore Power Query(10)

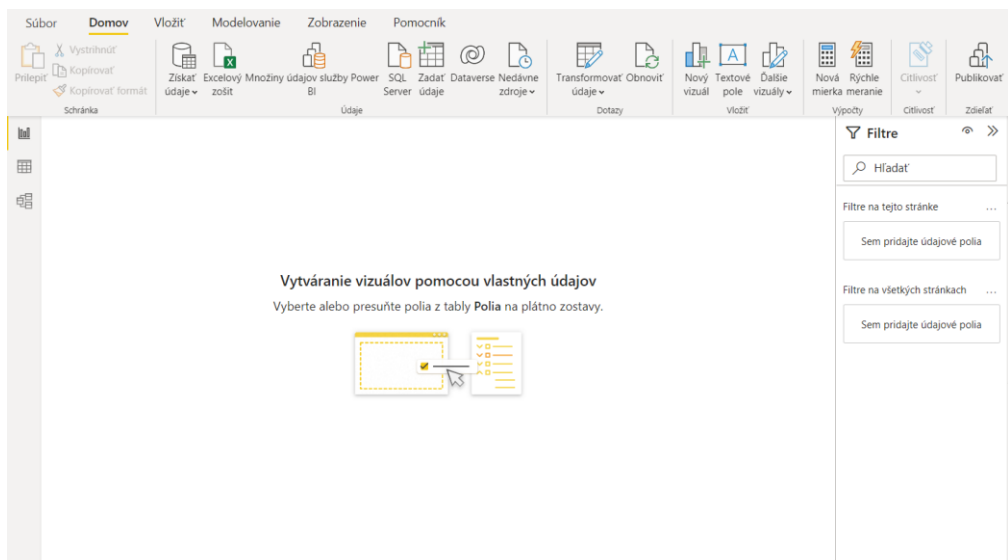


Zdroj: Vlastné spracovanie

Doplnením predošlej tabuľky sme dostali údaje do požadovaného stavu a mohli sme preto prejsť k ich načítaniu do prostredia aplikácie Power BI Desktop, aby sme následne mohli s nimi pracovať. V karte **Domov** sme zaklikli **Zavrieť a použiť**, čím sme zavreli okno editora dotazov, použili všetky vykonané zmeny a dostali sa späť do hlavného prostredia aplikácie, Obrázok 33.

V ľavej časti tejto snímky možno vidieť tri ikony zvisle pod sebou, o ktorých sme už uviedli základné informácie v rámci teoretickej časti. Aj keď úpravy údajov sme už dokončili, zostalo nám ešte upraviť dátový model a vzťahy medzi jednotlivými tabuľkami, aby sme mohli začať s tvorením zostavy a prevádzaním údajov do grafickej podoby. A práve k dátovému modelu sme sa dostali pomocou tretej ikony, na ktorej možno vidieť tri prepojené tabuľky.

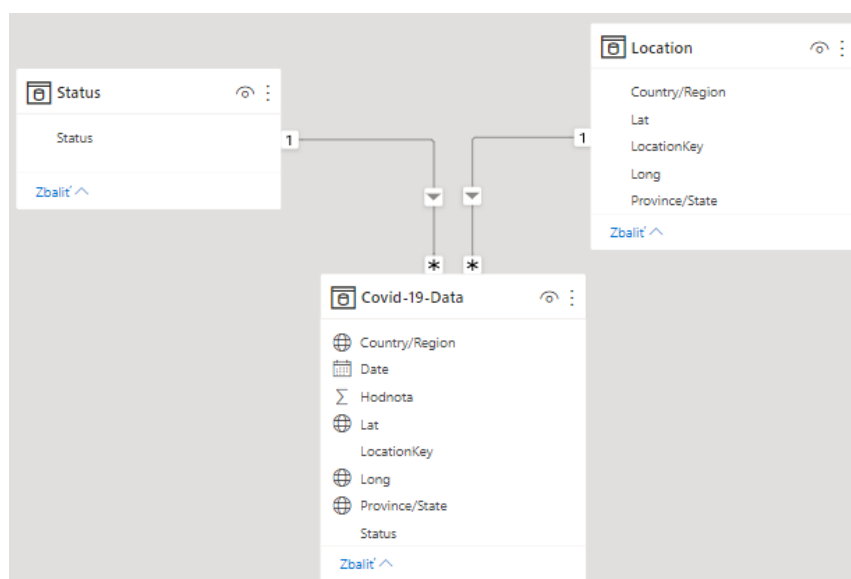
Obrázok 33: Domovská obrazovka aplikácie



Zdroj: Vlastné spracovanie

V rámci Modelu sme potrebovali prepojiť tri tabuľky „Status“, „Covid-19-Data“ a „Location“. Vzhľadom na to, že tabuľka so všetkými údajmi o prípadoch s ochorením Covid-19 bola hlavná, potrebovali sme to vyjadriť aj vzťahmi. Na nasledujúcej snímke obrazovky, Obrázok 34, už môžeme vidieť finálnu verziu dátového modelu aj s vyjadrením vzťahov medzi tabuľkami. Cez atribút Status sme prepojili hlavnú tabuľku s tabuľkou opisujúcou len statusy prípadov a cez atribút LocationKey sme prepojili hlavnú tabuľku s tabuľkou „Location“.

Obrázok 34: Dátový model



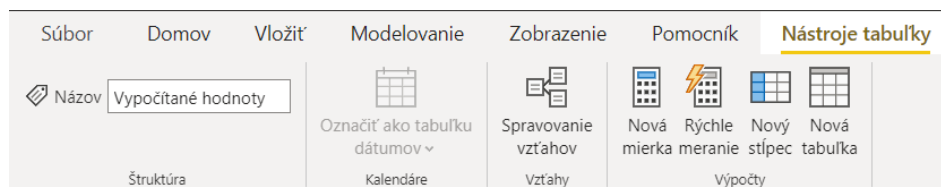
Zdroj: Vlastné spracovanie

4.2 Vypočítanie doplňujúcich hodnôt

V tejto časti bakalárskej práce sme sa zamerali na využitie jazyka DAX pri počítaní hodnôt, ktoré sme mali v záujme neskôr preniesť do grafickej podoby. Vzhľadom na to, že údaje, tak ako sme ich načítali do prostredia by nám vo všeobecnosti neprezradili veľa, bolo potrebné viaceré hodnoty dodatočne zistiť.

Pred samotným písaním vzorcov sme sa rozhodli vytvoriť novú ale prázdnu tabuľku s názvom „Vypočítané hodnoty“, znovu pomocou tlačidla **Zadať údaje**. Síce sme pri vytváraní tabuľky nevložili žiadne údaje, tabuľka sa vytvorila s prázdny stĺpcom „Stĺpec 1“. Ak by sme hneď nepotrebný prázdny stĺpec vymazali, vymazali by sme s ním aj novú tabuľku, čo nebolo v našom pláne. Preto sme sa najprv začali venovať počítaniu potrebných hodnôt pre našu zostavu. Aby sme našu prvú vypočítanú hodnotu priradili do tabuľky „Vypočítané hodnoty“, potrebovali sme na ňu kliknúť v pravej časti **Polia**. Vďaka tomu sme mohli vidieť novú kartu, týkajúcu sa tejto tabuľky, ako to vidíme na Obrázku 35. V rámci tejto karty **Nástroje tabuľky**, v časti **Výpočty** sme klikli na **Nová mierka** a dostali sa tak k možnosti zapísať výraz DAX, ktorý vypočítal hodnoty z našich údajov. Takýmto postupom sme sa riadili pri každej jednej vytvorenej mierke.

Obrázok 35: Nová mierka



Zdroj: Vlastné spracovanie

Prvú mierku, ktorú sme sa rozhodli vytvoriť bola na výpočet všetkých prípadov. V jednoduchosti povedané sme chceli spočítať počet potvrdených prípadov, počet úmrtí a počet ľudí, ktorí sa z ochorenia vyliečili. Avšak jeden človek, ktorému ochorenie potvrdili sa následne mohol vyliečiť alebo ochoreniu mohol podľahnúť a tým pádom bol započítaný do dvoch skupín naraz, a z toho dôvodu tento výpočet nebude predstavovať pre nás kľúčovú výpovednú hodnotu. Pre nás najdôležitejšími číslami, ktoré sme chceli zistiť bolo počet potvrdených prípadov, počet vyliečených prípadov a počet úmrtí, ale aby sme kvôli tomu nemuseli písať tri komplikované vzorce, vytvorili sme taký jeden - Všetky prípady, Obrázok 36, z ktorého sme pre nás dôležité čísla už dopočítali jednoducho.

Tento vzorec obsahoval rôzne funkcie. Jednou z nich bola SUMX, ktorú sme uplatnili práve trikrát a jej podstatou vo všeobecnosti je ísť riadok po riadku - každý riadok vyhodnocovať a spočítavať zvlášť. V tomto prípade spočítaval prípady pre každý status, každú krajinu a každú provinciu. Aby ale spočítaval len správne čísla, využili sme funkciu CALCULATE spolu s funkciou MAX. Tým sme zabezpečili, že pre každý z troch uvedených atribútov spočíta maximálne, teda najväčšie číslo, a to z dôvodu, že v stĺpci „Hodnota“ boli uvádzané čísla vyjadrujúce počet prípadov a nie počet nových prípadov. Následne boli tieto hodnoty ukladané do premennej SUM_CASES.

Obrázok 36: Výpočet všetkých prípadov

```

1 Všetky prípady =
2 VAR SUM_CASES=
3 SUMX ( VALUES('Status'[Status]),
4       SUMX(
5         VALUES(Location[Country/Region] ),
6         SUMX( VALUES(Location[Province/State]), CALCULATE(MAX('Covid-19-Data'[Hodnota]) ) )
7       )
8 )
9 RETURN
10 IF (SUM_CASES = 0 , 0 , SUM_CASES)

```

Zdroj: Vlastné spracovanie

Počet všetkých ľudí, ktorí na ochorenie doposiaľ zomreli sme ďalej počítali v nadväznosti na predošlý vzorec. Využili sme jednoduchú funkciu CALCULATE, ktorá pracovala s výslednou hodnotou mierky „Všetky prípady“ a doplnili sme ju o podmienku, aby spočítavala len tie riadky, kde status bol „Deaths“, ako to vidíme na snímku obrazovky, Obrázok 37. Ide o obdobný spôsob, akým sme počítali „Potvrdené prípady“ s uvedeným vzorcom v rámci metodiky práce. Pri tretej skupine prípadov - vyliečené prípady - sme status nahradili slovom „Recovered“. Treba však poukázať na skutočnosť, že približne od leta v roku 2021 nie všetky krajiny uchovávajú informácie o počte vyliečených ľudí, čo znamená, že aj keď výpočet je správny tak výsledná hodnota neodzrkadľuje realitu. V nadväznosti na tento fakt, ani počet aktívnych prípadov, ktorý sme zistili odčítaním úmrtí a vyliečených ľudí od potvrdených prípadov, nebude ukazovať správnu hodnotu. Vzhľadom na to sme tieto dve čísla vo veľkom množstve nezpracovávali do grafických vizualizácií.

Obrázok 37: Výpočet prípadov s úmrtím

```

1 Úmrtia =
2 CALCULATE([Všetky prípady] , 'Status'[Status]="Deaths")

```

Zdroj: Vlastné spracovanie

Ďalšiu hodnotu, ktorá nás zaujímala, bolo zistiť počet nových prípadov a nových úmrtí. Keďže údaje sú aktualizované na dennej báze, rozhodli sme sa aj tieto hodnoty zisťovať v závislosti od jedného dňa. Z toho vyplýva, že počet nových prípadov je možné vypočítať, keď od počtu prípadov za dnešný deň odpočítame počet prípadov za predošlý deň. Preto sme si aj v rámci vzorca, Obrázok 38, vytvorili dve premenné: CURRENT_DAY, do ktorej sme vložili počet prípadov z dnešného dňa a PREV_DAY, do ktorej sme vložili počet prípadov z predošlého dňa. Toto číslo sme zistili využitím funkcie CALCULATE, ktorej funkcia DATEADD určila, že má spočítať len potvrdené prípady s dátumom o jeden deň menším. So zámenou polí [Potvrdené prípady] na [Úmrtia] sme rovnakým spôsobom zistili počet Nových úmrtí.

Obrázok 38: Výpočet nových prípadov

```
1 Nové prípady =  
2 VAR CURRENT_DAY = [Potvrdené prípady]  
3 VAR PREV_DAY =  
4     CALCULATE(  
5         [Potvrdené prípady] , DATEADD('Covid-19-Data'[Date], -1, DAY)  
6     )  
7 RETURN  
8 IF (ISBLANK(PREV_DAY) , 0 , CURRENT_DAY-PREV_DAY)
```

Zdroj: Vlastné spracovanie

Vzhľadom na to, že čísla ako také, môžu byť niekedy klamlivé, sme sa zamerali na výpočet percentuálneho vyjadrenia úmrtnosti a vyliečenosti. Znovu pôjde takmer o identické vzorce, takže si vzorec vysvetlíme len pri jednej mierke - Úmrtnosť, Obrázok 39. Riadok napísaný v jazyku DAX týmto predstavuje výpočet pomeru počtu mŕtvych ľudí kvôli ochoreniu COVID-19 k počtu celkových potvrdených prípadov. S využitím funkcie DIVIDE sme tieto dve čísla podelili, vynásobili ich číslom 100 a zistili percentuálne vyjadrenie mierky.

Obrázok 39: Výpočet úmrtnosti

```
1 Úmrtnosť = DIVIDE( [Úmrtia], [Potvrdené prípady])*100
```

Zdroj: Vlastné spracovanie

Jednou z posledných mierok, ktorú sme vytvorili, bola „Rebríček - potvrdené prípady“, ktorej podstatou bolo zoradiť krajiny podľa počtu potvrdených prípadov zostupne.

Jej vytvorenie bolo dôležité pre ďalšiu mierku „Top 10 potvrdených prípadov“, Obrázok 40, ktorá mala za úlohu určiť prvých 10 krajín sveta s najväčším denným prírastkom potvrdených prípadov. Rovnakým spôsobom sme vytvorili dané mierky aj v nadväznosti na Úmrtia.

Obrázok 40: Zistenie krajín s najväčším počtom prípadov

```
1 Rebríček - potvrdené prípady = RANKX( ALL(Location[Country/Region]),[Potvrdené prípady],,DESC)
1 Top 10 potvrdených prípadov = IF ([Rebríček - potvrdené prípady] <=10, [Potvrdené prípady], BLANK())
```

Zdroj: Vlastné spracovanie

Mierka s názvom „Posledná aktualizácia“, ktorú môžeme vidieť na nasledujúcom snímku obrazovky, Obrázok 41, mala zobrazovať maximálny, teda posledný dátum, v rámci ktorého boli do databázy zaznamenané údaje o prípadoch s ochorením COVID-19. V tomto prípade išlo len o využitie jednoduchšej funkcie MAX.

Obrázok 41: Posledná aktualizácia

```
1 Posledná aktualizácia = MAX('Covid-19-Data'[Date])
```

Zdroj: Vlastné spracovanie

Záverečná mierka bola vytvorená len za účelom správneho fungovania animačného vizuálu, ktorý sme pridali do našej zostavy. Bližšie vysvetlenie jej podstaty v správnom fungovaní vizuálu už bolo uvedené v kapitole 3.4 Tvorba vizuálov. Ale aby sme vysvetlili vzorec ako taký, tak išlo o zapojenie funkcie CALENDAR, ktorá nám vytvorila tabuľku „Kalendár“ s rovnakým začiatčným a koncovým dátumom, aké sú v tabuľke „Covid-19-Data“ a zvyšné dni v tomto intervale doplnila automaticky, Obrázok 42. Tabuľku Kalendár sme však zámerne neprepojili so žiadnou inou tabuľkou v dátovom modeli.

Obrázok 42: Vytvorenie kalendára

```
1 Kalendár =
2 CALENDAR(MIN('Covid-19-Data'[Date]),MAX('Covid-19-Data'[Date]))
```

Zdroj: Vlastné spracovanie

4.3 Zobrazenie finálnych stránok zostavy a ich analýza

Prvá stránka zostavy zobrazuje základné informácie a mala by predstavovať hlavnú stránku celej zostavy. Navigátor spolu s dátumom poslednej aktualizácie je dostupný na každej stránke v rovnakej podobe aby sa používateľ nemusel vždy vracat' na tú hlavnú, ak by medzičasom zabudol na dátum a taktiež aby sa mohol z každej stránky dostať na akúkoľvek inú. Môžeme tu ďalej vidieť v jednoduchom číselnom a zaokrúhlenom vyjadrení počet doposiaľ potvrdených prípadov, počet úmrtí či počet nových prípadov za posledný deň na celom svete. Zapojili sme tu základné grafy - plošný, ktorý v čase porovnáva množstvo potvrdených prípadov s množstvom obetí; prstencový graf; skupinový stĺpcový graf, na zobrazenie úmrtí a potvrdených prípadov zvlášť a posledný čiarový graf, z ktorého by sme vedeli určiť krajiny s najhoršou pandemickou situáciou na svete, vzhľadom na počet ľudí s týmto ochorením a počet ľudí, ktorí mu už podľahli.

Samozrejmosťou je, že grafy sú navzájom interaktívne, a preto zmena v časovej osi v ľavom hornom rohu by mala vplyv na všetky grafy na stránke, Obrázok 43. Každý jeden graf či vizuál by popisoval a vyjadroval situáciu vo svete v novom, zmenenom časovom intervale.

Obrázok 43: Hlavná stránka zostavy

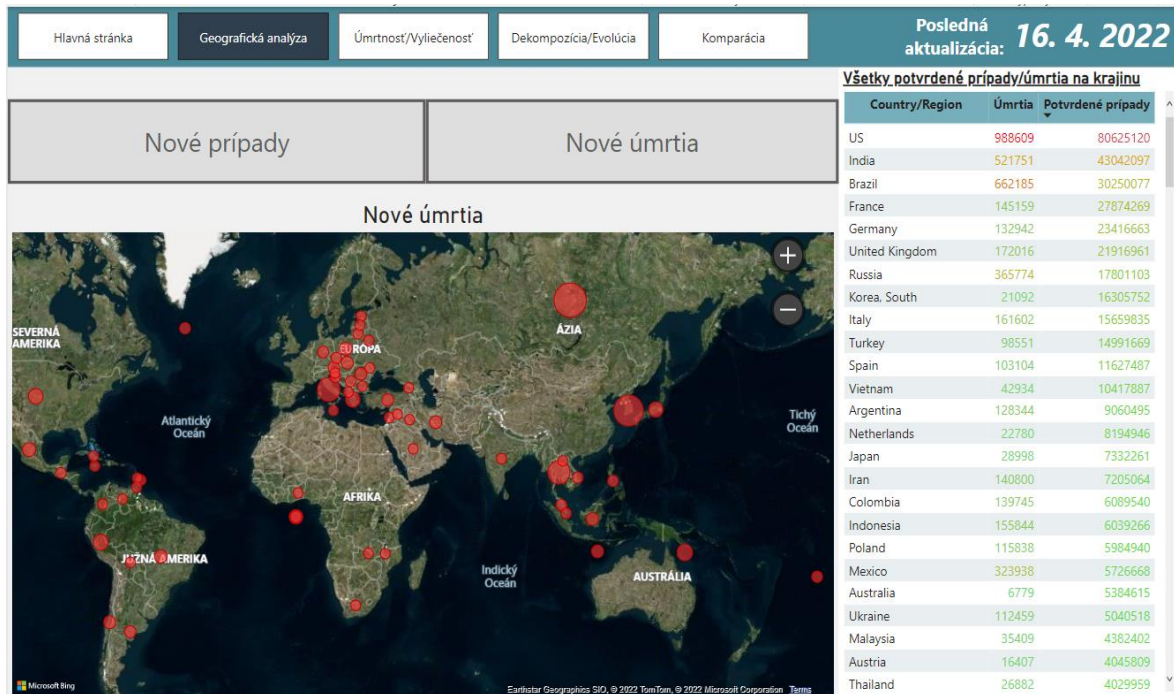


Zdroj: Vlastné spracovanie

Druhá stránka zostavy - Geografická analýza - bola vytvorená pre podrobnejšie charakterizovanie pandémie z geografického hľadiska, Obrázok 44. Využili sme na to vypočítané hodnoty Nové prípady a Nové úmrtia a pomocou bublín na mape sme mohli určovať, v ktorej krajine sa ochorenie za posledný deň veľmi rozšírilo alebo kde za posledný deň najviac ľudí ochoreniu podľahlo. Aktuálne na snímke obrazovky vidíme mapu s novými úmrtiami vo svete ale po prekliknutí cez tlačidlo Nové prípady by bolo možné zobraziť mapu s bublinami, ktoré by vyjadrovali množstvo nových prípadov pre dané krajiny. Mapa samozrejme nie je statická a bolo možné ju priblížiť či oddialiť a pohybovať sa po nej posúvaním podľa potreby.

Aby ale pozorovateľ tejto zostavy mohol mať aj konkrétnejšiu predstavu o číslach, doplnili sme na stránku prehľadnú tabuľku, ktorá farebne vyjadruje nielen situáciu za posledný deň ale situáciu počas celého trvania pandémie. Tabuľka bola s mapou interaktívne prepojená, čo znamená, že pri zakliknutí určitej bubliny na mape sa v tabuľke zobrazila len táto krajina, spolu so zodpovedajúcimi číslami a naopak pri zakliknutí jednej či viacerých krajín v tabuľke boli na mape zobrazené bubliny len týchto vybraných krajín.

Obrázok 44: Geografická analýza zostavy



Zdroj: Vlastné spracovanie

V poradí tretia stránka - Úmrtnosť/Vyliečenosť - bola vytvorená len z dvoch grafických vizuálov. Ako nám už názov napovedá, vďaka zobrazeniam sme mohli analyzovať podiel mŕtvych a podiel vyliečených ľudí naprieč všetkými krajinami zvlášť. 100-percentný skladaný pruhový graf sme doplnili ešte o vizuál Rýchly filter, vďaka čomu sme si mohli vybrať napríklad 2 či 3 krajiny, čím by sa graf stal prehľadnejším a bol by vhodný na porovnávanie percentuálneho podielu mŕtvych a vyliečených ľudí vo vybraných krajinách zeme, Obrázok 45.

Obrázok 45: Úmrtnosť/ Vyliečenosť



Zdroj: Vlastné spracovanie

Stránka Dekompozícia/Evolúcia bola predposlednou v našej zostave a mala by umožniť používateľovi bližšie sa pozrieť na čísla naprieč jednotlivými krajinami podľa statusov a zároveň graficky znázorniť priebeh pandémie naprieč celým svetom z časového hľadiska pomocou animačných prvkov daných vizuálov, Obrázok 46.

Ak by sme však chceli vidieť evolúciu pandémie len v určitej krajine, bolo by to možné vďaka interaktivite medzi týmito dvoma vizuálmi. Momentálne na snímke obrazovky vidíme začínajúcu pozíciu čiar na grafe pre celý svet, čo je možné ovplyvniť dvojklikom na danú krajinu v dekompozičnom strome. Následne po spustení „prehrávača“ animácie by sme mohli vidieť celý evolučný proces už len pre vybranú krajinu.

Obrázok 46: Dekompozícia/Evolúcia

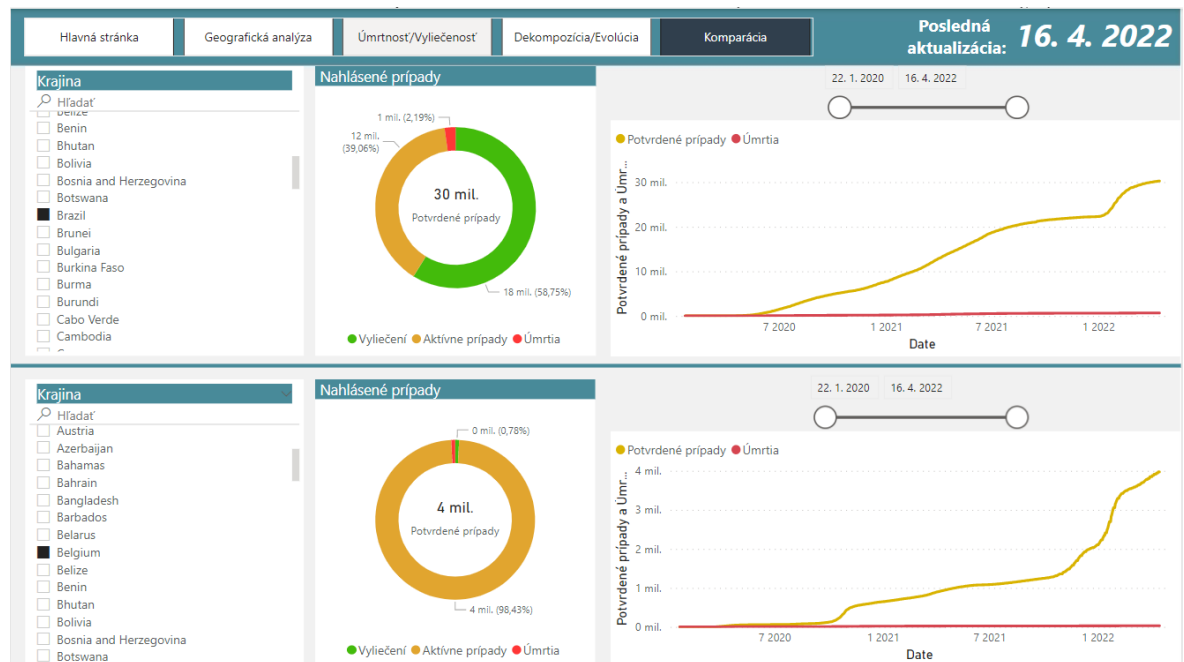


Zdroj: Vlastné spracovanie

Posledná stránka - Komparácia - bola rozdelená na dve polovice, kde každá z nich obsahuje totožné grafické vizualizácie. Dôležité pri tejto stránke bolo, aby vizualizácie v jednej časti neovplyvňovali a nemali žiadne interakcie s vizualizáciami v druhej časti a naopak, pretože inak by účel tejto stránky nebol naplnený. Účelom mala byť práve možnosť komparácie, teda porovnania situácie v dvoch krajinách alebo v dvoch skupinách rôznych krajín. Toto sme zabezpečili cez kartu **Formát** po kliknutí na určitú vizualizáciu a následne cez tlačidlo **Upraviť interakcie**. Následne sme pri vizuáloch, ktoré sme nechceli aby ovplyvňoval náš vybraný vizuál, zrušili interaktivitu kliknutím na tlačidlo **Žiadne**. Rovnaký postup sme aplikovali na všetky vizuály, nachádzajúce sa na tejto stránke.

Výber jednej i druhej krajiny používateľ zvolí zakliknutím v Rýchlom filtri v ľavej časti obrazovky. Ako už bolo naznačené, nemusí však ísť vždy o porovnanie len dvoch krajín - je možné v obidvoch rýchlych filtroch zakliknúť aj viac ako jednu krajinu. Následne s výberom sa zmenia aj grafy - prstencový graf a čiarový graf. Rovnako je umožnené používateľovi stanoviť si časový interval, ktorému majú vizualizácie zodpovedať. Pre lepšiu predstavivosť fungovania komparácie na tejto stránke môžeme vidieť na snímke obrazovky, Obrázok 47, porovnanie situácie v Brazílii a v Belgicku.

Obrázok 47: Komparácia



Zdroj: Vlastné spracovanie

Záver

Efektívne využívanie Business Intelligence a analytiky je zásadným rozdielom medzi spoločnosťami, ktoré v súčasnom modernom prostredí uspejú, a spoločnosťami, ktoré zlyhávajú. Softvéry BI, medzi ktoré patrí aj Power BI Desktop, využívajú algoritmy na extrahovanie praktických poznatkov z údajov spoločnosti a usmernenie jej strategických rozhodnutí. Využitie poskytnutých údajov z takéhoto softvéru na zlepšenie spoločnosti by malo byť neoddeliteľnou súčasťou základných hodnôt podnikania.

Cieľom práce bolo popísať a odprezentovať možnosti analýzy a vizualizácie údajov v prostredí aplikácie Power BI Desktop, ako jedného z nástrojov z rodiny Microsoft Power BI. K naplneniu cieľa bola v prvom rade vysvetlená základná teória z oblasti Business Intelligence a Big Data, taktiež proces analýzy a vizualizácie údajov. Následne bol čitateľ oboznámený s produktom Power BI a jeho architektúrou, aj v porovnaní s inými populárnymi nástrojmi - Tableau Desktop a QLIK Sense (Desktop). Dôležitou a neoddeliteľnou súčasťou bolo teoretické zhrnutie jazyka DAX, ktorý je potrebné ovládať pri analýze údajov v tomto prostredí. V rámci tretej kapitoly boli predstavené využité metódy práce a skúmania, bez ktorých by naplnenie čiastkových cieľov a tým pádom aj dosiahnutie hlavného cieľa nebolo možné. V našom prípade išlo o charakteristiku postupu vytvárania dashboardu v aplikácii zo všeobecného hľadiska - od možnosti importu údajov do prostredia, cez prácu s dátovým modelom a upravovanie vzťahov v modeli, až po spôsob prevádzania obyčajných dát do zaujímavých grafických zobrazení a vizualizácií.

V druhom rade bola vypracovaná interaktívna zostava, zameraná na aktuálnu tému COVID-19, ktorá demonštruje použitie nástroja Power BI Desktop pri spracovaní úloh Business Intelligence. Konkrétne bolo vytvorených 5 stránok zostavy, ktoré pomocou rôznych aspektov zobrazovali pandemickú situáciu vo svete. Bolo možné vidieť napríklad rebríček krajín, v ktorých doposiaľ zomrelo najviac ľudí na toto ochorenie, ale taktiež bolo možné určiť krajiny, v ktorých za predošlý deň pribudlo najviac ochorení či najviac mŕtvych.

Prínosom práce je predovšetkým prípadová štúdia - spracovanie údajov o prípadoch COVID-19 po celom svete - ktorú možno chápať aj ako názorný návod k práci s aplikáciou Power BI Desktop. Používatelia uvažujúci nad implementáciou nástroja Power BI môžu po prečítaní práce získať ucelenú predstavu o tom, ako je práca s týmto nástrojom náročná alebo ako vyzerá užívateľské rozhranie. Požívateľom, ktorí sa už pre Power BI rozhodli, môže

práca poslúžiť ako návod, ktorý im pomôže so spracovaním ich prvých dashboardov, či ako prehľad možností využitia rôznych prvkov vizualizácie analýz.

Práca odkazuje na veľa zaujímavých zdrojov, ktoré môžu čitatelia následne využiť k prehĺbeniu svojich znalostí ako z oblasti Business Intelligence či Big Data, tak aj z oblasti používania nástrojov Power BI, keďže popularita tejto problematiky stále narastá a je preto dôležité jej venovať dostatok pozornosti.

Zoznam použitej literatúry

1. DEBORTOLI, Stefan – MÜLLER, Oliver – BROCKE, Jan. Comparing Business Intelligence and Big Data Skills [online]. In *Bus Inf Syst Eng*. Springer, 2014, roč. 6, č. 6, s. 289-300 [cit. 2022-03-02]. Dostupné na: <https://link.springer.com/article/10.1007/s12599-014-0344-2#citeas>.
2. EHRENMUELLER-JENSEN, Markus. *Self-Service AI with Power BI Desktop : Machine Learning Insights for Business*. New York : Apress, 2020. 358 p. ISBN 978-1-4842-6230-6.
3. FEW, Stephen. *Information Dashboard Design : The effective Visual Communication of Data*. Sebastopol : O'Reilly Media, 2006. 223 p. ISBN 0-596-10016-7.
4. HOWSON, Cindi. *Successful Business Intelligence*. 2 ed. McGraw-Hill Education, 2014. 337 p. ISBN 978-0-07-180919-1.
5. CHAUDHURI, Surajit - DAYAL Umeshwar - NARASAYYA Vivek. An Overview of Business Intelligence Technology [online]. In *Communications of the ACM*. New York : ACM, 2011, roč. 54, č. 8, s. 88–98 [cit. 2022-03-02]. Dostupné na: <https://cacm.acm.org/magazines/2011/8>.
6. CHEN, Hsinchun – CHIANG, Roger – STOREY, Veda. Business Intelligence and Analytics: From Big Data to Big Impact [online]. In *MIS Quarterly*. Minnesota: Management Information Systems Research Center, 2012, roč. 36, č. 4, s. 1165-1188 [cit. 2022-03-02]. ISSN 0276-7783. Dostupné na: <http://www.jstor.org/stable/41703503>.
7. KNIGHT, Devin et al. *Microsoft Power BI Quick Start Guide*. 2 ed. Birmingham : Packt, 2020. 375 p. ISBN 978-1-80056-157-1.
8. LACHEV, Teo. *Applied Microsoft Power BI : Bring Your Data to Life*. 7 ed. USA : Prologika Press, 2022. 592 p. ISBN 978-1-7330461-3-8.
9. MASLYUK, Daniil. *Analyzing and Visualizing Data with Microsoft Power BI*. Pearson Education, Inc., 2018. 670 p. ISBN 978-1-5093-0702-9.
10. MELO, Pedro Novo – MACHADO, Carolina. *Business Intelligence and Analytics in Small and Medium Enterprises*. Boca Raton : CRC Press, 2020. 166 p. ISBN 978-0-367-17388-3.
11. MICROSOFT. Čo je Power BI? [online]. In *Začíname so službou Power BI : dokumentácia*. 2022(a) [cit. 2022-03-02]. Dostupné na: <https://docs.microsoft.com/sk-sk/power-bi/fundamentals/power-bi-overview>.

12. MICROSOFT. Vytváranie a používanie kartogramov v službe Power BI [online]. In *Vytváranie zostáv a tabúl' v službe Power BI*. 2022(b) [cit. 2022-03-20]. Dostupné na: <https://docs.microsoft.com/sk-sk/power-bi/visuals/power-bi-visualization-filled-maps-choropleths>.
13. MORRIS, Mike. *Power BI : Moving Beyond the Basics of Power BI and Learning More about DAX Language*. 2020. 126 p. ISBN 979-8609057471.
14. NOVOTNÝ, Ota – POUR, Jan – Slánský, David. *Business Intelligence : jak využit bohatství ve vašich datech*. Praha : GRADA, 2005. 192 s. ISBN 80-247-1094-3.
15. OLSZAK, Celina M. *Business Intelligence and Big Data : Drivers of Organizational Success*. Boca Raton : CRC Press, 2021. 195 p. ISBN 978-0-367-37394-8.
16. POUR, Jan – MARYŠKA, Miloš – NOVOTNÝ. *Business intelligence v podnikové praxi*. Praha : Professional Publishing, 2012. 276 s. ISBN 978-80-7341-065-2.
17. POUR, Jan et al. *Self Service Business Intelligence : Jak si vytvorit vlastní analytické, plánovací a reportingové aplikace*. Praha : GRADA, 2018. 352 s. ISBN 978-80-271-0616-5.
18. POUR, Jan. Co lze očekávat od Business Intelligence? [online]. Praha : Vysoká škola ekonomická, 2006. s. 1-9 [cit. 2022-03-02]. Dostupné na: https://www.cesvsem.cz/data/data/ces-soubory/konference-seminare/gf_Praha0906_ICT.pdf.
19. POWELL, Brett. *Mastering Microsoft Power BI*. Birmingham. Packt Publishing, 2018. 638 p. ISBN 978-1-78829-723-3.
20. QLIK. Qlik Sense: Go from passive to active analytics [online]. 2022 [cit. 2022-03-02]. Dostupné na: <https://www.qlik.com/us/products/qlik-sense>.
21. RAD, Reza. *Power BI DAX Simplified: DAX and calculation language of Power BI demystified by practical examples*. New Zealand : RADACAD, 2021. 496 p. ISBN 9798538933952.
22. ROMERO, Carlos Andres Tavera et al. Business Intelligence: Business Evolution after Industry 4.0 [online]. In *Sustainability*. MDPI Open Access Journals, 2021, roč. 13, č. 18, s. 1-12 [cit. 2022-03-02]. Dostupné na: https://www.mdpi.com/2071-1050/13/18/10026#framed_div_cited_count.
23. ŘEHOŘ, František. Podstatné charakteristiky faktorů [online]. In *F455 : Self Service BI*. MBI, 2014, s. 86 [cit. 2022-03-02]. Dostupné na: <https://mbi.vse.cz/public/cs/obj/FACTOR-110>.

24. SELECTHUB. Best Business Intelligence Software Tools [online]. 2022 [cit. 2022-03-02]. Dostupné na: <https://www.selecthub.com/business-intelligence-tools/>.
25. SHARDA, Ramesh – DELEN, Dursun – TURBAN, Efraim. *Business Intelligence, Analytics, and Data Science : A Managerial Perspective*. 4 ed. London : Pearson Education, 2018. 514 p. ISBN 978-1-292-22054-3.
26. TABLEAU. Get Tableau Desktop as part of Tableau Creator [online]. 2022 [cit. 2022-03-02]. Dostupné na: <https://www.tableau.com/resource/business-intelligence>.
27. WONGSUPHASAWAT, Kanit et al. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations [online]. In *IEEE Transactions on Visualization and Computer Graphics*. IEEE, 2016, roč. 22, č. 1, p. 649-658 [cit. 2022-03-02]. ISSN 1941-0506. Dostupné na: <https://ieeexplore.ieee.org/document/7192728>.
28. YIPENG, Li – SIYUAN, Hou. Methods and Techniques in Data Visualization Model [online]. In *International Conference on Computer Technology, Electronics and Communication*. Dalian: IEEE, 2017, p. 71-74 [cit. 2022-03-02]. ISBN 978-1-5386-5784-3. Dostupné na: <https://ieeexplore.ieee.org/document/8789106>.