

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky



**Analýza trhu s mobilními aplikacemi
s využitím Competitive Intelligence a jeho
nástrojů**

DIPLOMOVÁ PRÁCE

Studijní program: Informační systémy a technologie

Specializace: Business Intelligence

Autor: Bc. Hoang Nam Dao

Vedoucí diplomové práce: PhDr. Jan Černý, Ph.D.

Praha, duben 2024

Prohlášení

Prohlašuji, že jsem diplomovou práci Analýza trhu s mobilními aplikacemi s využitím Competitive Intelligence a jeho nástrojů vypracoval samostatně za použití v práci uvedených pramenů a literatury.

V Praze dne 29. dubna 2024

.....

Hoang Nam Dao

Poděkování

Chtěl bych poděkovat PhDr. Janu Černému, Ph.D., za cenné rady a nápady, které mi značně pomohly při tvorbě této diplomové práce a také za trpělivost, která byla klíčová pro její úspěšné dokončení.

Abstrakt

Tato diplomová práce se zaměřuje na analýzu trhu s mobilními aplikacemi s hlavním důrazem na integraci principů Competitive Intelligence (CI) a technik Business Intelligence (BI). Trh s mobilními aplikacemi se dynamicky vyvíjí, proto je klíčové pro vývojové firmy rozumět konkurenčnímu prostředí, identifikovat trendy a adaptovat se na měnící se potřeby uživatelů. Distribuční platformy pro aplikace třetích stran, jako jsou Google Play nebo App Store, jsou zdrojem cenných dat, která mohou pomoci s výzkumem a identifikací potenciálně lukrativních odvětví. Správná analýza a interpretace těchto dat může poskytnout vývojovým firmám strategickou výhodu prostřednictvím hlubšího porozumění trhu a zlepšení jejich konkurenční pozice.

V této práci byly stanoveny klíčové zpravodajské otázky (KIQ), přičemž odpovědi na tyto otázky by mohly pomoci firmě formulovat strategie pro získání konkurenční výhody. Pro účel hledání odpovědi na tyto otázky je pro firmu Robology vyvinut analytický nástroj, který automatizuje procesy sběru, čištění, transformace a analýzy dat z distribučních platforem, jako jsou Google Play a App Store. Tato data jsou vizualizována pomocí interaktivního dashboardu v nástroji Google Looker, což firmě umožňuje snadno interpretovat informace o tržním prostředí. Cílem diplomové práce je vyvinout interaktivní nástroj, který poskytne vývojovým firmám komplexní přehled o konkurenčním prostředí v sektoru mobilních aplikací a umožní jim formulovat odpovědi na jejich klíčové zpravodajské otázky.

Klíčová slova

web scraping, mobilní aplikace, Competitive Intelligence, datová analýza, Business Intelligence

Abstract

This thesis focuses on the analysis of the mobile application market, with a primary emphasis on integrating Competitive Intelligence (CI) principles and Business Intelligence (BI) techniques. The mobile application market is evolving dynamically, making it crucial for development companies to understand the competitive environment, identify trends, and adapt to changing user needs. Third-party application distribution platforms such as Google Play and the App Store are sources of valuable data that can aid in research and identification of potentially lucrative sectors. Proper analysis and interpretation of this data can provide development agencies with a strategic advantage through a deeper understanding of the market and improving their competitive position.

In this thesis, Key Intelligence Questions (KIQs) were established, and the answers to these questions could help the company formulate strategies for gaining a competitive advantage. To find answers to these questions, an analytical tool was developed for Robology that automates the processes of collecting, cleaning, transforming, and analyzing data from distribution platforms such as Google Play and the App Store. This data is visualized using an interactive dashboard in the Google Looker tool, which allows the company to interpret information about the market environment easily. The thesis aims to develop an interactive tool that provides development companies with a comprehensive overview of the competitive environment in the mobile application sector and enables them to formulate responses to their key intelligence questions.

Keywords

mobile application, data analysis, Competitive Intelligence, web scraping, Business Intelligence

Obsah

1 Mobilní aplikace	12
1.1 Popis trhu s mobilními aplikacemi.....	13
1.2 Distribuce mobilních aplikací	14
1.3 Modely monetizace aplikací	15
1.3.1 Paid apps	15
1.3.2 Freemium model a In-App Purchases.....	15
1.3.3 Předplatné.....	15
1.3.4 Reklama	15
1.4 Klíčoví hráči	16
1.5 Klíčové metriky	18
2 Competitive Intelligence.....	19
2.1 Plánování a řízení	21
2.1.1 Plánování a definice KIT pro analýzu trhu mobilních aplikací.....	21
2.2 Sběr informací.....	23
2.2.1 Web scraping	23
2.2.2 Externí poskytovatelé dat.....	26
2.3 Analýza	26
2.3.1 Vybrané metody pro analýzu trhu s mobilními aplikacemi.....	27
2.4 Distribuce	27
2.4.1 Zpravodajský produkt pro analýzu trhu s mobilními aplikacemi	28
3 Business Intelligence	29
3.1 Rozdíl mezi Business Intelligence a Competitive Intelligence	29
3.2 Architektura BI	30
3.3 Datové zdroje v BI	32
3.4 Databázové komponenty	33
3.5 Datová transformace.....	34
3.6 Analytické nástroje.....	34
4 Návrh architektury analytického nástroje.....	36
4.1 Frontend	36
4.2 Backend	37
4.3 Vrstva přípravy a transformace dat	37
4.4 Datové úložiště.....	38
4.5 Vizualizace dat.....	38

4.6 Diagram architektury	39
5 Vývoj nástroje pro analýzu trhu s mobilními aplikacemi	40
5.1 Nástroje a služby použité při vývoji	40
5.2 Vývoj frontendu.....	40
5.2.1 Uživatelské rozhraní	40
5.2.2 Integrace uživatelského rozhraní s backendem.....	42
5.3 Vývoj backendu	43
5.3.1 API endpointy	43
5.3.2 Extrakce dat.....	44
5.3.3 Spuštění datové transformace.....	44
5.4 Implementace datové transformace	45
5.4.1 Popis extrahovaných dat.....	46
5.4.2 Příprava a čištění dat.....	52
5.4.3 Tvorba odvozených atributů	54
5.4.4 Analýza sentimentu	55
5.4.5 Textová analýza	56
5.4.6 Kategorizace dat	57
5.4.7 Odebrání nepotřebných atributů	58
5.4.8 Výsledné datové soubory	59
5.5 Integrace datového úložiště	60
5.6 Vizualizace dat.....	61
5.6.1 Napojení na datový zdroj a propojení datových souborů	61
5.6.2 Design dashboardu.....	62
5.7 Nasazení aplikace na webový server	65
6 Analýza trhu s mobilními aplikacemi.....	68
6.1 Představení kontextu analýzy	68
6.2 Plánování a řízení.....	69
6.3 Sběr informací.....	70
6.4 Analýza	71
6.4.1 Identifikace klíčových hráčů v segmentu	71
6.4.2 Sledování tržních trendů a průzkum funkcionalit.....	75
6.4.3 Identifikace monetizačních strategií.....	78
6.4.4 Identifikace slabých stránek konkurenčních aplikací	81
6.4.5 Formulace strategie pro vstup na zahraniční trh	85
6.5 Vyhodnocení.....	88

6.5.1 Identifikace klíčových hráčů v segmentu	88
6.5.2 Sledování tržních trendů a průzkum funkcionalit.....	88
6.5.3 Identifikace monetizačních strategií v rámci segmentu.....	89
6.5.4 Identifikace slabých stránek konkurenčních aplikací	89
7 Možnosti pro budoucí vývoj nástroje	90
7.1 Rozšíření o funkcionalitu App Store Optimization.....	90
7.2 Integrace dat od externích poskytovatelů.....	90
7.3 Reporting a automatizovaná extrakce dat	91
7.4 Ukládání dat pro historický kontext	91
Závěr.....	92
Použitá literatura	94
Přílohy.....	I
Příloha A: Odkaz na webovou aplikaci	I
Příloha B: Odkaz na repozitář se zdrojovým kódem aplikace.....	I
Příloha C: Odkaz na dashboard	II

Seznam obrázků

Obrázek 1: Vývoj počtu uživatelů smartphonů. [3].....	13
Obrázek 2: Vývoj obratu segmentů mobilních aplikací [7].....	14
Obrázek 3: Nejpopulárnější frameworky pro tvorbu mobilních aplikací [18]	16
Obrázek 4: Nejstahovanější aplikace pro období březen 2024 [20]	17
Obrázek 5: Vizualizace zpravodajského cyklu a jeho fází [24]	20
Obrázek 6: Fáze web scrapingu [25]	24
Obrázek 7: Obecná koncepce architektury BI	30
Obrázek 8: Konceptuální návrh architektury nástroje pro analýzu dat, Zdroj: autor	31
Obrázek 9: Zjednodušený diagram webové aplikace, Zdroj: autor	39
Obrázek 10: Snímek obrazovky pracovního prostředí platformy Webflow, Zdroj: autor....	41
Obrázek 11: Snímek obrazovky prostředí GCS po nahrání datových souborů, Zdroj: autor	61
Obrázek 12: Konfigurace propojení datových souborů v rámci Google Looker Studio, Zdroj: autor	62
Obrázek 13: Ukázka stránky dashboardu s informacemi o mobilních aplikacích na platformě App Store, Zdroj: autor	63
Obrázek 14: Ukázka stránky dashboardu sloužící k textové analýze recenzí mobilních aplikacích na platformě App Store, Zdroj: autor	64
Obrázek 15: Snímek obrazovky administračního rozhraní DigitalOcean App platform, Zdroj: autor	66
Obrázek 16: Grafický návrh mobilní aplikace pro kolaborativní plánování dovolené, Zdroj: autor	68
Obrázek 17: Zadávání vyhledávacích kritérií pro extrakci dat ve webové aplikaci, Zdroj: autor	71
Obrázek 18: Získaný seznam deseti nejlépe hodnocených aplikací v segmentu aplikací pro plánování cest na platformě Google Play, Zdroj: autor	71
Obrázek 19: Přehled metrik deseti nejstahovanějších aplikací platformě Google Play, Zdroj: autor	72
Obrázek 20: Získaný seznam deseti nejlépe hodnocených aplikací v segmentu aplikací pro plánování cest na platformě App Store, Zdroj: autor	72
Obrázek 21: Přehled metrik deseti nejlépe hodnocených aplikací na platformě App Store, Zdroj: autor	72
Obrázek 22: Zralost aplikací v rámci segmentu aplikací pro plánování cest z platformy Google Play, Zdroj: autor	73
Obrázek 23: Zralost aplikací v rámci segmentu aplikací pro plánování cest z platformy App Store, Zdroj: autor	73
Obrázek 24: Detail mobilní aplikace na platformě AppMagic, Zdroj: autor	74
Obrázek 25: Tabulka s popisem aplikací a informacemi o poslední aktualizaci, Zdroj: autor	75
Obrázek 26: Vizualizace atributů pro identifikaci monetizačních strategií z platformy Google Play, zdroj: Autor	78
Obrázek 27: Tabulka s přehledem spodních a horních hodnot cen pro nákupy v aplikaci, Zdroj: autor	79
Obrázek 28: Funkce jednotlivých typů předplatného aplikace Wanderlog [43]	79

Obrázek 29: Výstřížek z platformy AppMagic, zobrazující odhad tržeb z mikrotransakcí, Zdroj: autor	80
Obrázek 30: Seznam aplikací s negativním sentimentem recenzí z platformy Google Play, zdroj: Autor	81
Obrázek 31: Ukázka funkcionality dashboardu při výběru aplikace, Zdroj: Autor	82
Obrázek 32: Graf zobrazující průměrný počet dní od poslední aktualizace podle kategorií sentimentu z platformy Google Play, Zdroj: autor	84
Obrázek 33: Tabulka se seznamem podporovaných jazyků a počtem aplikací, které dané jazyky podporují, Zdroj: autor	85
Obrázek 34: Geografická heatmapa na základě podpory jazyků, Zdroj autor	86
Obrázek 35: Výstřížek z platformy AppMagic, zobrazující zeměpisné rozložení uživatelské báze, Zdroj: autor	86

Úvod

Mobilní aplikace jsou v dnešní době součástí každodenního života mnoha lidí. Z toho, co se se ještě před dvaceti lety považovalo pouze za dočasný trend, se stalo v současnosti něčím, co proniká do téměř každého aspektu našeho života. S rostoucím počtem uživatelů chytrých telefonů po celém světě se trh s mobilními aplikacemi neustále rozšiřuje a stává se konkurenčnějším. Je velmi obtížné se na tomto trhu orientovat, jelikož se neustále objevují nové aplikace, což dělá trh s mobilními aplikacemi mimořádně dynamickým a proměnlivým. Pro firmy působící v tomto sektoru je tedy nezbytné nejen sledovat nejnovější trendy a inovace, ale také důkladně rozumět konkurenčnímu prostředí.

Jednou z takovou firem je Robology, softwarová agentura, kterou jsem s mým společníkem založil v roce 2021. Naše firma se specializuje na tvorbu webových a mobilních aplikací na míru. Nejde nám však jen o technické aspekty vývoje, ale klademe také důraz na pochopení toho, co uživatelé chtějí a co by rádi používali. Proto prozkoumáváme konkurenční aplikace, které se v příslušné kategorii již osvědčily, a to již v počátečních fázích koncepce a designu našich řešení. Rozhodl jsem se proto využít principů Competitive Intelligence (CI) a Business Intelligence (BI) k vytvoření nástroje, který by nám ulehčil orientaci v tomto prostředí a umožnil nám lépe pochopit konkurenci.

Cílem této diplomové práce je vývoj interaktivního nástroje, který poskytne detailní přehled o konkurenčním prostředí v oblasti mobilních aplikací. Tento nástroj získává aktuální data o aplikacích z distribučních platforem Google Play a App Store a umožňuje firmě analyzovat klíčové ukazatele, sledovat konkurenci a identifikovat trendy na trhu s mobilními aplikacemi. Získaná data jsou vizualizována v interaktivním dashboardu pomocí nástroje Google Looker Studio, což usnadňuje jejich interpretaci a pomáhá firmě formulovat efektivní strategie pro získání konkurenční výhody.

Tato práce je rozdělena na teoretickou a praktickou část. V teoretické části se věnuji trhu s mobilními aplikacemi, významu a metodám Competitive Intelligence a Business Intelligence. Také definuji klíčové zpravodajské otázky, které jsou formulovány tak, aby pomohly firmě se strategickým rozhodování v rámci tohoto konkurenčního prostředí.

Praktická část práce popisuje kompletní popis návrhu, vývoje a nasazení interaktivního nástroje. Nástroj je koncipovaný jako webová aplikace, která zajistí extrakci dat z distribučních platforem pomocí web scrapingu, čištění a transformaci dat pomocí skriptu vytvořeného v Pythonu, načtení dat do cloudového úložiště a jejich následná vizualizace v dashboardu za použití nástroje Google Looker Studio. V rámci praktické ukázky použití nástroje je provedena analýza trhu s mobilními aplikacemi na základě položených klíčových zpravodajských otázek. Dále jsou v této práci popsány možnosti dalšího vývoje vytvořeného analytického nástroje.

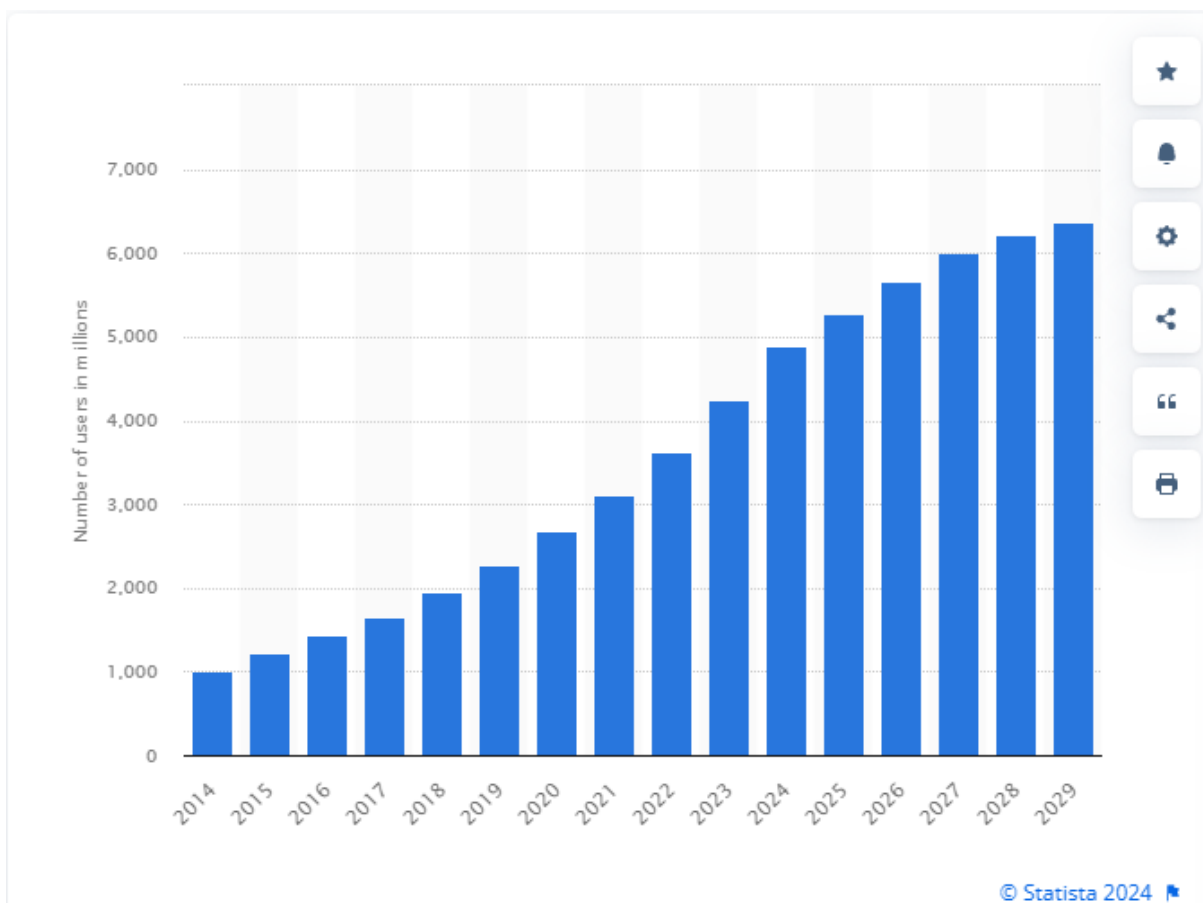
1 Mobilní aplikace

Mobilní aplikace jsou software, který je vyvíjen specificky pro použití na mobilních zařízeních, jako jsou telefony, tablety nebo hodinky. Tyto aplikace rozšiřují funkce mobilních zařízení a usnadňují uživatelům plnění široké škály úkolů, od zábavy a komunikace až po produktivitu a správu financí. [1]

Historie mobilních aplikací sahá do 90. let 20. století, kdy se začaly objevovat první hry a jednoduché programy pro tehdejší mobilní telefony. Příkladem může být ikonická hra Snake, vydaná v roce 1997, která byla dostupná na mobilních telefonech Nokia 6110. V té době bylo standardem, že mobilní aplikace přicházely předinstalované výrobcem společně s telefonem. Průlom však nastal v roce 2007, kdy společnost Apple vydala chytrý telefon iPhone a s ním také distribuční kanál App Store. Ten obsahoval při spuštění pouze 500 aplikací. Od té doby se však trh s mobilními aplikacemi rozrostl do obrovských rozměrů a dnes existují miliony aplikací pro různé účely. [2]

S rozvojem trhu s mobilními aplikacemi také úzce souvisí zvyšující se dostupnost chytrých telefonů. Počet uživatelů smartphonů každoročně prudce stoupá – v roce 2014 jich bylo 1,01 miliard, v roce 2024 už 4,88 miliard a podle prognóz by jich v roce 2029 mělo být 6,37 miliard. [3] Díky této rostoucí bázi uživatelů chytrých telefonů se mobilní aplikace staly pro firmy integrálním kanálem pro přístup k zákazníkům. Aplikace se mohou stát základem obchodní strategie firmy a představovat jedním z hlavních zdrojů jejich příjmů (např. Facebook, Airbnb nebo X). Existují dokonce společnosti, které začaly svoje podnikání přímo s mobilními aplikacemi a své služby poskytují výhradně jejich prostřednictvím (např. Uber, Bolt, Liftago). I pro tradičně netechnologické firmy však aplikace představují důležité médium, které jim pomáhá získávat zákazníky, udržovat si jejich loajalitu a nabízet různé promo akce (např. Lidl Plus, Kaufland, IKEA). [4]

Mobilní aplikace také nejsou nutně určeny jenom pro interakci firem se zákazníky. Mobilní aplikace využívají i státy, které se snaží držet krok s digitálními trendy a zpřístupnit občanům důležité služby přímo skrze jejich mobilní zařízení. Např. v rámci České republiky byla v roce 2024 spuštěna aplikace eDoklady, která umožňuje občanům prokazovat svoji totožnost bez nutnosti nosit občanský průkaz. [5] Dalším příkladem může být aplikace Tečka (nyní přejmenovaná na EZKarta), která se hojně využívala jako digitální očkovací průkaz v době pandemie COVID-19. [6] Tyto iniciativy naznačují rostoucí trend využívání mobilních aplikací ve veřejné správě, což občanům umožňuje snadnější a rychlejší přístup k různým veřejným službám.



Obrázek 1: Vývoj počtu uživatelů smartphonů. [3]

1.1 Popis trhu s mobilními aplikacemi

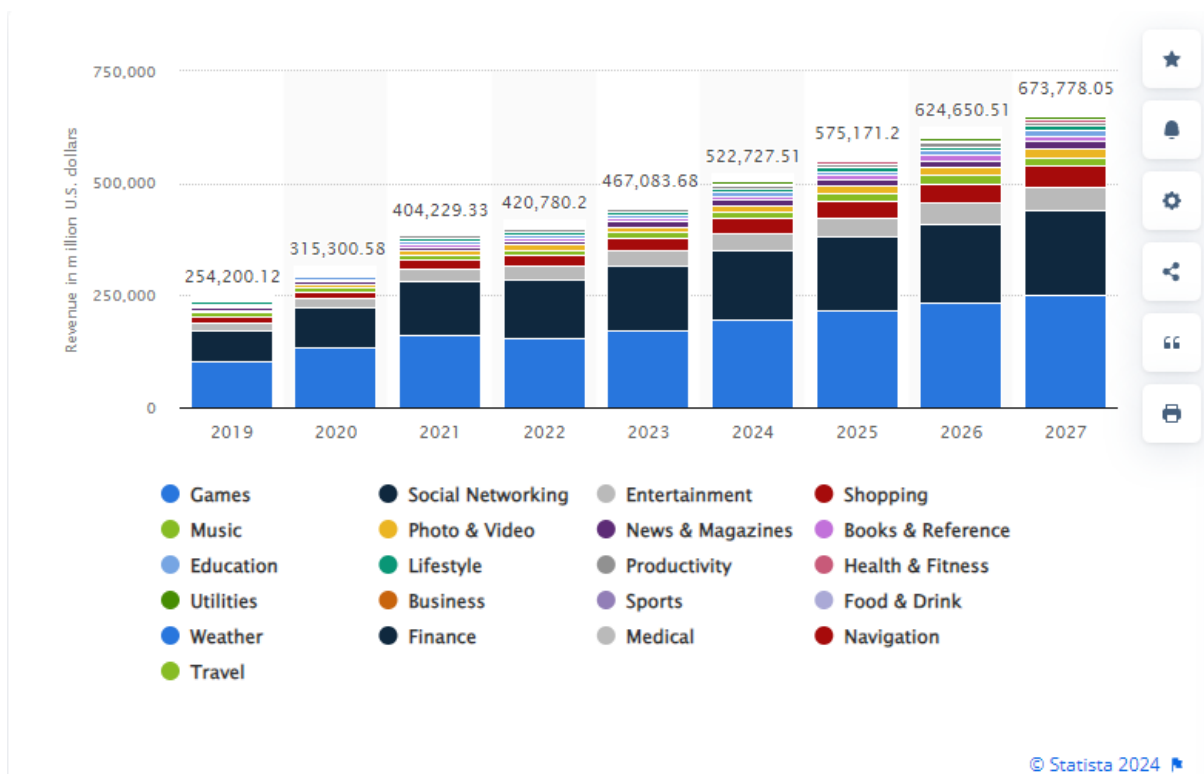
Trh s mobilními aplikacemi je jedním z nejrychleji rostoucích odvětví na světě. V roce 2023 dosáhl globální trh s mobilními aplikacemi obrátu 467,1 miliard USD a očekává se, že do roku 2026 tento trh dosáhne celkového obrátu 624,7 miliard USD. [7] Jak již bylo zmíněno, tento růst je důsledkem rostoucího počtem chytrých telefonů a zvyšující se potřebě používání mobilních aplikací v každodenním životě.

Počet dostupných aplikací a vývoj této hodnoty se liší dle distribuční platformy. Zatímco na Google Play bylo v roce 2018 dostupných 3,6 milionu aplikací, v roce 2023 se počet dostupných aplikací snížil na 2,4 milionu. Konkurenční App Store zažívá opačný trend, kdy počet aplikací každoročně roste. V roce 2016 byly na App Store necelé 2 miliony aplikací, v roce 2023 již bylo zaznamenáno 3,8 milionu dostupných aplikací. [8] [9]

Počet stažení aplikací se stejně jako obrát každým rokem zvyšuje. Zatímco v roce 2016 bylo zaznamenáno 140,7 miliard stažení, v roce 2023 byl počet stažení 257 miliard.[10]

Trh s mobilními aplikacemi lze rozdělit do segmentů. Nejpopulárnějším a zároveň nejvýdělečnějším segmentem jsou hry. V roce 2023 byl celkový obrat tohoto segmentu 172,3 miliardy USD a kumulativní počet stažení činil 134,9 miliard. [7] [11]

Z hlediska objemu používání jsou aplikace v segmentu sociální média uživateli nejvíce frekventované. [12]. V roce 2023 měli celkový obrat 143 miliard USD a 12,8 miliard kumulativních stažení. [7] [11] Mezi další populární segmenty patří segmenty jako zábava, nákupy a životní styl.



Obrázek 2: Vývoj obratu segmentů mobilních aplikací [7]

Nejvýznamnějšími zeměmi pro trh s mobilními aplikacemi jsou Čína, Indie, Spojené státy americké a Brazílie. Lze pozorovat, že počet stažení je přímo úměrný populaci daných zemí. [13]

1.2 Distribuce mobilních aplikací

Mobilní aplikace jsou distribuovány hlavně skrze digitální platformy, které slouží jako centrální místo pro jejich stahování a instalaci. Mezi nejpoužívanější distribuční platformy patří App Store a Google Play. Kromě těchto platform existuje také možnost distribuce aplikací prostřednictvím obchodů třetích stran či obchodů přímo od výrobců telefonů. Příkladem těchto obchodů jsou platformy jako Amazon Appstore, Samsung Galaxy Apps, Firefox Marketplace nebo Microsoft Store. [14]

Mimo oficiální obchody existuje také možnost distribuce aplikací jinými přímými kanály. To může být provedeno přímým sdílením instalačních souborů aplikace. Nicméně tato praxe je méně běžná a může s sebou nést určitá bezpečnostní rizika, jako je možnost stažení škodlivého softwaru, protože tyto aplikace nejsou podrobeny stejným kontrolám, které nabízí oficiální distribuční platformy.

1.3 Modely monetizace aplikací

1.3.1 Paid apps

Placené aplikace vydělávají prostřednictvím samotného nákupu aplikace. Uživatelé jsou přímo účtováni za stažení aplikace z distribuční platformy. Tyto aplikace poté většinou nenabízejí žádné další nákupy v aplikaci. [15]

1.3.2 Freemium model a In-App Purchases

Freemium model nabízí základní verzi aplikace zdarma, zatímco prémiové funkce nebo obsah jsou placené. Tento model může aplikaci získat širokou uživatelskou základnu. Výhoda také spočívá v možnosti uživatelů vyzkoušet si aplikaci před zakoupením plné verze nebo předplatného. Příkladem může být aplikace MyFitnessPal, která nabízí pokročilé funkce, jako jsou personalizované jídelníčky a přístup k složení potravin pouze platícím uživatelům.

Freemium modely také mohou využívat in-app purchases, neboli nákupů v aplikaci, které umožní uživatelům zakoupit virtuální zboží nebo služby přímo v aplikaci. Tento přístup je obzvláště populární ve hrách, kde si uživatelé mohou koupit virtuální měnu, speciální vybavení nebo bonusové levely. Příkladem může být populární hra Candy Crush Saga, kde si můžou uživatelé za reálné peníze koupit virtuální měnu, se kterou si odemykají nové úrovně. [16]

1.3.3 Předplatné

Model předplatného poskytuje uživatelům přístup k prémiovému obsahu nebo službám za pravidelný poplatek. Tento model je vhodný pro služby, které poskytují neustálou hodnotu, jako jsou streamovací služby, novinové aplikace nebo cloudové služby. Příkladem aplikace, která využívá tento typ monetizace je např. Spotify, kde si uživatel platí předplatné, aby mohl streamovat hudbu bez omezení. Tato metoda může být zkombinována s modelem freemium. [15]

1.3.4 Reklama

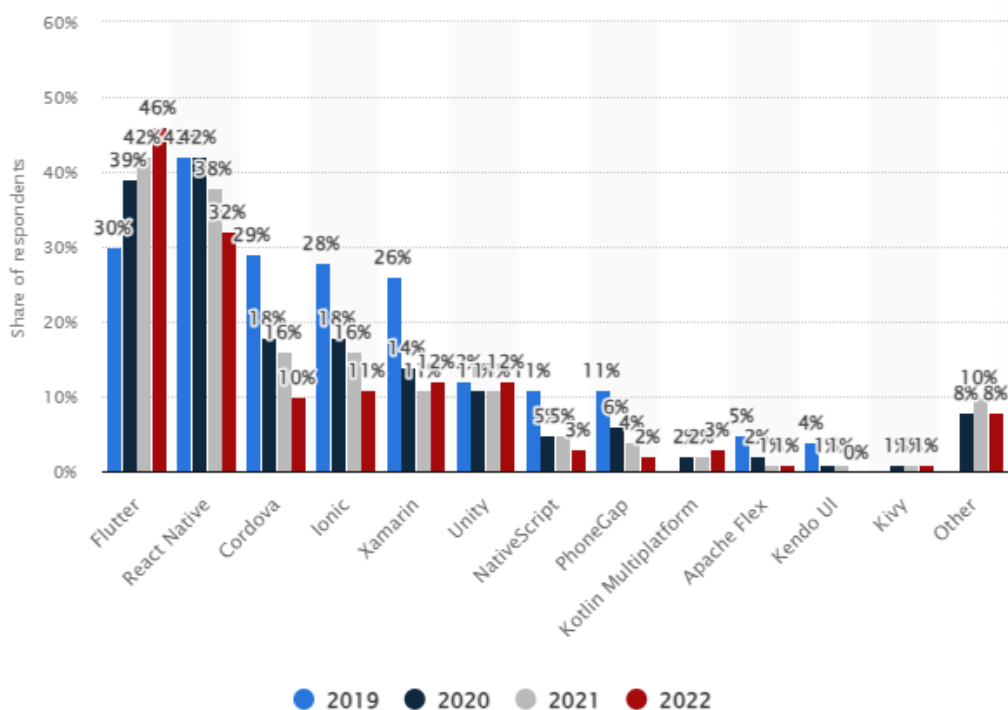
Zobrazování reklam je jednou z nejjednodušších a nejrozšířenějších forem monetizace. Aplikace zobrazují bannery, videa nebo interaktivní reklamy a generují příjmy na základě zobrazení nebo kliknutí uživatelů. Pro uživatele je to však jeden z nejagresivnějších způsobů monetizace. Reklamy u nich mohou vyvolat negativní emoce, zvláště když je reklam

v aplikaci velké množství. Důležité je také jejich umístění – pokud je uživateli zabráněno v průchodu aplikací, je možné že se uživatel do aplikace již nevrátí. Tento model může být uplatněn samostatně nebo v kombinaci s ostatními modely monetizace. [16]

1.4 Klíčoví hráči

Trhu s mobilními aplikacemi dominují převážně velké technologické společnosti. Nejvýznamnějšími hráči na tomto trhu jsou firmy Google a Apple, které provozují nejpopulárnější distribuční platformy Google Play a App Store. Tyto platformy nejsou pouze distribučním kanálem pro miliony aplikací, ale poskytují také vývojářům nástroje a směrnice pro tvorbu a publikaci aplikací. Tyto firmy si také berou provize ze všech transakcí, které se uskuteční na těchto platformách, čímž generují značné zisky. Společnost Apple je v této souvislosti známá pro své kontroverzní praktiky, které porušují antimonopolní zákony. Všem vývojářům totiž zakazuje jakékoliv transakce mimo jejich platformu, přičemž si z každé proběhlé transakce bere provizi třicetiprocentní provizi. Za toto porušení zákonů dostal Apple od Evropské unie v březnu roku 2024 pokutu ve výši 2 miliardy USD. [17]

Google a Apple stojí také za vývojem dvou nejpoužívanějších mobilních operačních systémů na trhu – Android a iOS. Společnost Google také vyvinula framework Flutter, který je v současné době nejpopulárnějším nástrojem pro vývoj multiplatformních mobilních aplikací. [18] Pozice těchto firem na trhu je nezpochybnitelná a v budoucnosti je velmi nepravděpodobné, že by mohl někdo ohrozit jejich dominanci.

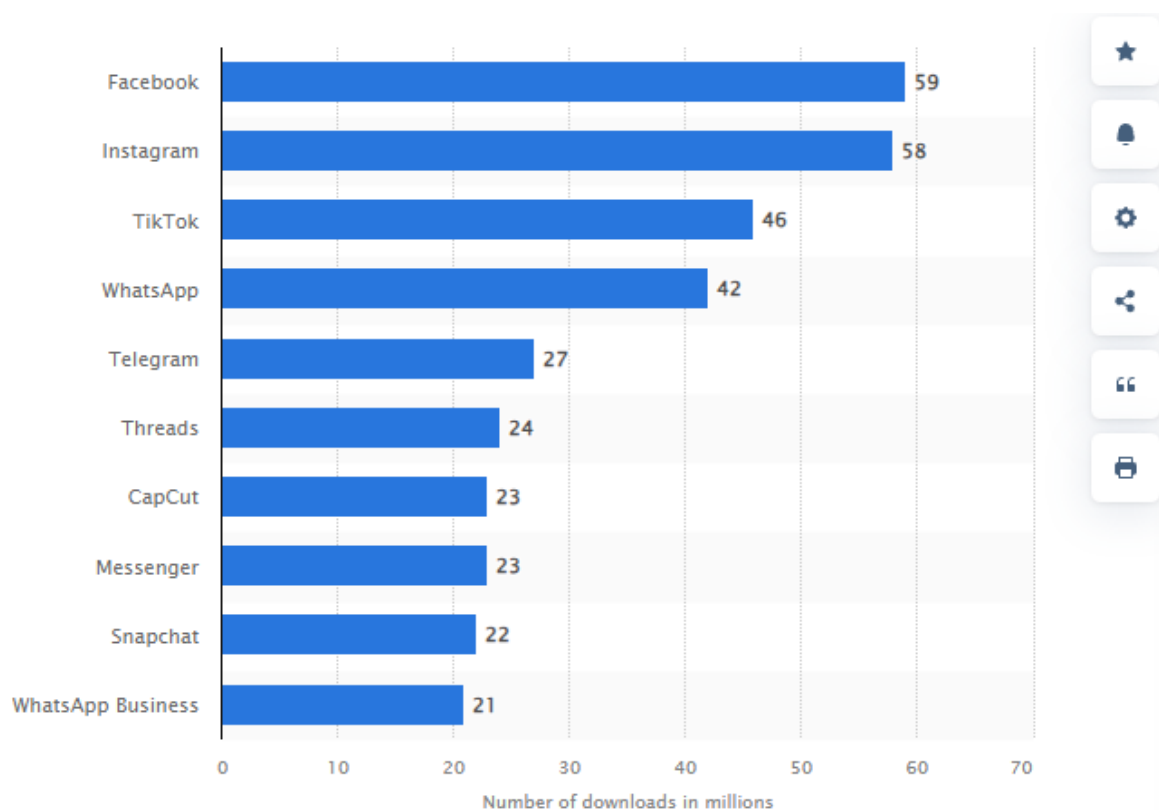


Obrázek 3: Nejpopulárnější frameworky pro tvorbu mobilních aplikací [18]

Za dalšího klíčového hráče se dá považovat firma Meta. V jejím portfoliu jsou globálně nejpoužívanější aplikace jako např. Facebook, Instagram, WhatsApp a Threads. Meta také stojí za vývojem React Native, druhým nejpopulárnějším frameworkem pro vývoj mobilních aplikací. [18]

Kromě těchto klíčových hráčů existuje mnoho dalších společností, které se významně podílí na trhu s mobilními aplikacemi. Jsou to společnosti, které mají vlastní platformy s velkou uživatelskou základnou, jako jsou např. Amazon (e-commerce a streaming), Netflix (streaming) nebo Spotify (hudba).

Poslední firmou, která stojí za zmínku je Tencent, čínský technologický gigant, který má téměř monopolní postavení na čínském trhu. Tencent stojí za aplikací WeChat, která je považována za tzv. „super-app“, díky jejímu nepřehlednému množství funkcionalit. Mezi tyto funkcionality můžeme řadit např. instant-messaging, platební služby, streamování obsahu, hraní her, sdílení fotek, e-commerce a mnoho dalšího. WeChat je velmi unikátní v tom, že je jednou z nejpoužívanějších aplikací na světě, přestože je používán primárně pouze v Číně a zároveň je nezávislý na distribučních platformách Google Play a App Store. [19]



Obrázek 4: Nejstahovanější aplikace pro období březen 2024 [20]

1.5 Klíčové metriky

Hlavní ukazatele umožňují hodnotit výkonnost mobilních aplikací a jejich úspěšnost na trhu. Tyto metriky poskytují firmám cenné informace pro strategické rozhodování a optimalizaci produktů.

Jednou z klíčových metrik v tomto odvětví je míra odlivu uživatelů (churn rate), která vyjadřuje procento uživatelů, kteří v určitém časovém období přestanou aplikaci používat. Vysoká míra odlivu může signalizovat problémy ve funkčnosti aplikace, uživatelské zkušenosti nebo její celkové hodnotě. Další zásadní metrikou je míra udržení uživatelů (retention rate), ukazující, kolik uživatelů se pravidelně vrací k aplikaci. Vysoká míra udržení naznačuje schopnost aplikace efektivně zapojit uživatele a poskytnout jim hodnotu, která je motivuje k opakovanému používání. Tento ukazatel je úzce spojen s dlouhodobým úspěchem aplikace na trhu. [21] Tyto metriky jsou obvykle dostupné pouze vydavatelům aplikací. Existují však také služby, které pomocí statistických algoritmů a výzkumných panelů dokáží tyto metriky odhadnout.

I z veřejně dostupných metrik lze však získat vhled do úspěšnosti a popularity aplikací na trhu. Mezi nejvýznamnější veřejně dostupné metriky patří počet stažení a číselné hodnocení aplikace, které lze snadno nalézt na distribučních platformách, jako jsou Google Play nebo App Store. Počet stažení ukazuje celkový zájem o aplikaci a může sloužit jako předběžný ukazatel její popularity. Číselné hodnocení, obvykle na škále od jedné do pěti hvězd, poskytuje přehled o spokojenosti uživatelů s aplikací a její kvalitě. Tyto metriky jsou často prvními, které potenciální uživatelé i vývojáři hodnotí, aby získali rychlý přehled o tom, jak je aplikace přijímána na trhu. Dále jsou veřejně dostupné i informace o frekvenci aktualizací aplikace, které mohou naznačovat aktivní péči vývojářů o aplikaci a její neustálé zlepšování.

Každá metrika má také odlišnou váhu pro to, jestli je aplikace úspěšná. Podle výzkumu [22], kde autoři sledovali detailně aplikace, které se dostali čerstvě do žebříčku nejlepších 300 aplikací v rámci svých distribučních platforem, si vysoké umístění často udrží vývojáři, kteří mají diverzifikované portfolio napříč různými žánry. To může znamenat, že mají úspěch zavedené společnosti, které již mají rozmanité zkušenosti s vývojem různých aplikací. Autoři také zjistili, zjistili, že nabídka bezplatných aplikací, investice do méně populárních kategorií, průběžné aktualizace funkcí a ceny aplikací, stejně jako vyšší počet uživatelských zpětných vazeb na aplikace, jsou pozitivně spjaté s jejich výkonem.

2 Competitive Intelligence

Competitive Intelligence (CI), je systematický, legální a etický proces sběru, analýzy a interpretace informací o externím prostředí firmy. Hlavním cílem CI je zvýšit konkurenceschopnost společnosti prostřednictvím procesu řízení, sběru, analýzy a distribuce informací a znalostí, které podporují rozhodování na taktické a strategické úrovni. Hlavní výhodou využívání CI je to, že poskytuje velké množství relevantních a včasných informací, které umožňují firmě činit správná rozhodnutí. [23]

Hlavní úkoly CI jsou následující [23]:

- **Získávat a analyzovat dostupné informace** s cílem předpovídat tržní změny a možné kroky konkurence.
- **Vyhledávání nových zákazníků a identifikace nových příležitostí** pro expanzi a růst firmy.
- **Poskytování aktuálních informací** o nových technologiích, standardech, legislativních změnách a sociálně-politických faktorech, které mohou ovlivnit fungování společnosti.
- **Snižovat rizika** spojená s rozhodovacími procesy ve společnosti.
- **Analyzovat interní procesy a strategie**, poskytovat zpětnou vazbu a upozorňovat na potenciální slabiny firmy.
- **Identifikovat a upozorňovat na zastaralé postupy**, které mohou oslabovat konkurenční pozici společnosti.
- **Identifikace a pochopení chyb konkurence**, aby se z nich společnost mohla poučit a získat konkurenční výhodu.

CI může nabývat různých forem v závislosti na specifickém účelu a cílech, které si klade. Rozlišujeme mezi aktivním (ofenzivním) a pasivním (defenzivním) CI. V této práci je primárně využíváno aktivní CI, jehož hlavním cílem je shromažďování a analýza informací o tržních podmínkách, konkurenci, aktuálních trendech a budoucích příležitostech či hrozbách. Pasivní CI je opakem aktivního CI a jeho cílem je ochrana důležitých firemních informací a znalostí před konkurencí. [23]

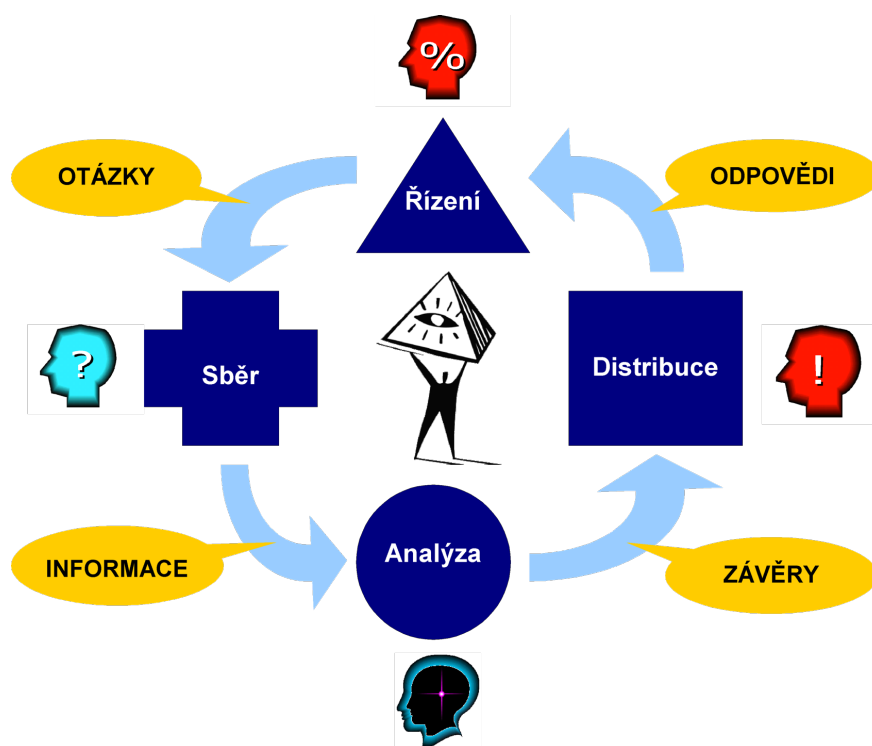
Competitive Intelligence můžeme dělit podle zaměření [23]:

- **Strategické CI** je zaměřeno především na tvorbu dlouhodobých strategií. Tato forma CI se soustředí na pohled do budoucnosti a predikce budoucího vývoje s cílem shromáždit dostatek relevantních informací, které poskytnou pevný základ pro strategické plánování, zejména v kontextu plánovaného růstu podniku.
- **Taktické CI** se na rozdíl od strategického CI zaměřuje na krátkodobé plány, které by měly být přesnější, jelikož faktor nejistoty je nižší. Podporuje především procesy

prodeje, marketingu a cenotvorby. Proces taktického CI je většinou spuštěn jako reakce na externí faktor.

- **Technické/Technologické CI** se specializuje na výzkum a vývoj, přičemž je zaměřeno na výsledky inovací a technického vývoje, s ohledem na pokrok konkurence. Jeho hlavním cílem je udržet krok s konkurencí, nebo být v oblasti inovací o krok před ní.
- **Competitor CI** se specificky zaměřuje na sběr informací o nejbližších konkurentech, nikoli o celkovém tržním prostředí.
- **Customer CI** se zaměřuje na současné i budoucí zákazníky a zahrnuje informace o cenových hladinách výrobků, promo akcích a jejich efektivitě.
- **Finanční CI** se soustředí na finanční sféru a zahrnuje analýzu dat a informací o finanční situaci zkoumaného subjektu.
- **Market CI** se zaměřuje na analýzu současného trhu, jeho vývoje, charakteristik, velikosti, nových zákaznických trendů a směrů marketingu.
- **Partner CI** je oblastí, která se zaměřuje na obchodní partnery a udržení si kladných vztahů s nimi i do budoucna, nejen s dodavateli, ale také s komplementárními partnery a veřejným sektorem nebo finančními institucemi.

Základním prvkem Competitive Intelligence je zpravodajský cyklus, který je strukturován do těchto základních fází: plánování a řízení, sběr informací, analýza a distribuce. Celý cyklus je kontinuální, kde každý následující cyklus CI je upravován na základě zpětné vazby od iniciátora procesu, což umožňuje neustálé zlepšování a přizpůsobování procesu. [23]



Obrázek 5: Vizualizace zpravodajského cyklu a jeho fází [24]

Níže představím jednotlivé fáze zpravodajského cyklu CI a popíšu, jakým způsobem budou souviset s mým cílem analýzy trhu s mobilními aplikacemi.

2.1 Plánování a řízení

Plánování a řízení je klíčovou fází procesu CI, která určuje směřování celého procesu a zajišťuje, že shromážděné informace budou pro firmu relevantní a užitečné. V rámci této fáze se definují klíčová zpravodajská témata (KIT), která vyznačují konkrétní oblasti zájmů a účel, proč se zpravodajský cyklus provádí.

Tato témata se dělí do tří základních kategorií [23]:

- **Rozhodovací témata**, která jsou přímo spojena s plánovanými rozhodnutími managementu a jsou charakteristická svým jasně definovaným obsahem a termínem. Rozhodovací témata obvykle zahrnují specifické informace potřebné k podpoření konkrétního rozhodnutí, jako je například expanze na nový trh, investice do nové technologie, nebo strategické akvizice.
- **Předmětná témata**, která se věnují konkrétním subjektům, jako jsou konkurenti, obchodní partneři, státní orgány, banky a další relevantní aktéři. Cílem předmětných témat je předvídat chování těchto subjektů a pochopit, jak jejich akce mohou ovlivnit organizaci.
- **Varovná témata**, která jsou zaměřena na průběžné sledování specifických indikátorů, které mohou naznačovat budoucí hrozby nebo příležitosti. Varovná témata umožňují organizaci rychle reagovat na potenciální problémy nebo využít nečekané možnosti, ještě předtím, než se plně projeví. Tímto způsobem může firma minimalizovat rizika nebo maximalizovat užitek ze vzniklých situací.

V rámci těchto témat dochází ke stanovení klíčových zpravodajských otázek (KIQ), které určují specifické informační požadavky a vedou k identifikaci zdrojů dat a jejich následnému sběru.

2.1.1 Plánování a definice KIT pro analýzu trhu mobilních aplikací

V naší softwarové agentuře Robology¹ se specializujeme na vývoj aplikací na míru pro naše klienty. V rámci návrhu aplikace, předtím, než začínáme s jejím vývojem, prozkoumáváme konkurenční prostředí a bereme si inspiraci od zavedených a úspěšných vydavatelů aplikací. Tato analýza však probíhá manuálně a nestrukturovaně a výstupy těchto analýz jsou nekonzistentní.

Chtěli bychom tento proces zformalizovat, usnadnit a vylepšit, a proto si definujeme klíčová zpravodajská témata, která nás jako firmu zajímají. Výstup z níže zmíněných klíčových

¹ <https://robology.cz/>

zpravodajských témat je relevantní pro téměř každý začátek procesu návrhu mobilní aplikace.

Identifikace klíčových hráčů v konkrétním segmentu mobilních aplikací

Když vyvíjíme aplikaci, je pro nás zásadní rozpoznat, kdo jsou hlavní konkurenti v rámci konkrétního segmentu a jaké strategie tyto subjekty implementují. Toto poznání je klíčové pro porozumění strukturám tržních sil a konkurenční dynamiky, což umožňuje firmě efektivně navigovat tržním prostředím a přizpůsobit své strategické plánování.

Sledování tržních trendů a průzkum funkcionalit v rámci segmentu

Proces sledování tržních trendů by nám poskytl důležité informace o trendech, které mohou ovlivnit budoucí poptávku, technologické inovace a tržní potřeby. Tímto způsobem může naše firma proaktivně reagovat na trh a přizpůsobovat své produkty.

Identifikace monetizačních metod mobilních aplikací

Identifikace efektivních monetizačních strategií by firmě pomohla při tvorbě udržitelných obchodních modelů, což je jedním z nejdůležitějších faktorů pro dlouhodobý úspěch na trhu s mobilními aplikacemi.

Identifikace slabých stránek konkurenčních aplikací

Analyzování slabých stránek konkurenčních aplikací může firmě nabídnout cenné vhledy o tom, čeho se při vývoji aplikací vyvarovat. Zároveň tato analýza umožňuje formulovat strategie pro poskytování lepších služeb a pro vývoj inovativních řešení.

Formulace strategie pro vstup na zahraniční trhy

Odpovědi na klíčové informační otázky v rámci tohoto tématu mohou firmě pomoci při návrhu a implementaci marketingových a cenových strategií, které jsou přizpůsobeny lokálním potřebám a preferencím.

2.2 Sběr informací

Sběr informací představuje fázi zpravodajského cyklu CI, během které se shromažďují data z různých informačních zdrojů. Tento proces by měl být prováděn systematicky a eticky, přičemž je nutné, aby odpovídal aktuálním potřebám a cílům organizace. Ačkoliv je tato fáze zpravodajského cyklu považována za jednu z nejdůležitějších, je nezbytné zdůraznit, že bez následného zpracování a analýzy informací sběr sám o sobě nepřináší organizaci žádný užitek a vede k plýtvání lidskými a finančními zdroji. [23]

Z hlediska metodologie může sběr informací probíhat dvěma základními způsoby. Prvním je primární sběr informací, kde organizace sama generuje data pomocí průzkumů, anket, rozhovorů a dalších výzkumných technik. Druhým způsobem je sekundární sběr dat, který zahrnuje shromažďování veřejných informací z různých databází, archivů a publikací.

Existuje mnoho zdrojů, ze kterých může firma čerpat. Mezi validní zdroje informací můžeme zařadit např. veřejné databáze, tiskové zprávy, sociální média, marketingové materiály, webové stránky či účetní uzávěrky. [24]

Výsledkem efektivně provedeného sběru dat by měla být strukturovaná databáze informací, která poskytne solidní základ pro další kroky zpravodajského cyklu CI – od analýzy, přes interpretaci, až po prezentaci zjištění.

Níže představím informační zdroje, metody a nástroje, které jsou relevantní pro sběr dat z trhu s mobilními aplikacemi.

2.2.1 Web scraping

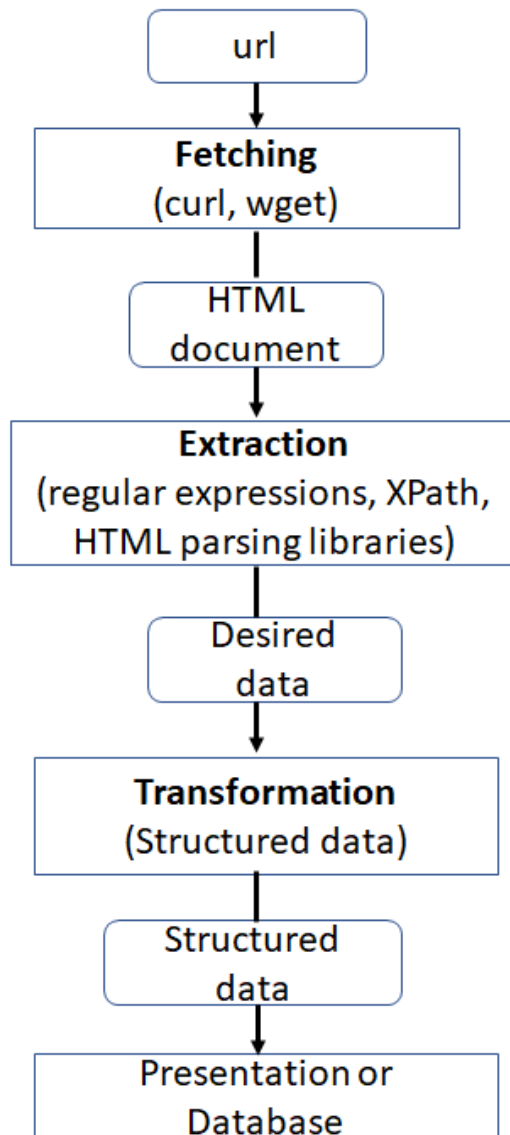
V rámci trhu s mobilními aplikacemi je v kontextu CI velkou výhodou centralizace dat. Distribuční platformy Google Play a App Store slouží nejen jako prodejní kanály, ale rovněž jako zdroje cenných strukturovaných dat o konkurenčním prostředí. K tomu, abychom se dostali k datům z těchto platforem je použit metodu sběru dat zvanou web scraping.

Termín web scraping označuje techniku automatického sběru dat z webových stránek. Tento proces zahrnuje extrakci informací z internetu, obvykle z HTML stránek, a jejich převod do strukturovanějšího formátu, jako je CSV či JSON. Web scraping se používá v různých oblastech, včetně tržního výzkumu, analýzy cen, shromažďování konkurenčních dat, monitorování veřejných názorů, agregace obsahu a dalších. Hlavní výhodou web scrapingu je možnost rychlého získání velkého objemu dat. [25]

Proces web scrapingu má tři fáze [26]:

- **Získání/načtení dat**, při kterém se odesílá HTTP požadavek na server, na kterém je cílová webová stránka hostována. Tento požadavek simuluje návštěvu stránky prostřednictvím webového prohlížeče, je však realizován automatickým skriptem. Po obdržení odpovědi od serveru, která obvykle obsahuje HTML kód stránky, je možné tento kód stáhnout a uložit pro další zpracování.

- **Extrakce a parsování dat**, při kterém se využívají nástroje a knihovny pro parsování HTML, aby bylo možné z kódu stránky extrahovat specifické informace, jako jsou texty, obrázky, odkazy a další relevantní obsah.
- **Transformace dat**, při které se data transformují do požadované struktury pro další využití.



Obrázek 6: Fáze web scrapingu [25]

Metody pro extrakci dat

Existuje mnoho různých způsobů, jak získávat data z webových stránek. Mezi tyto metody můžeme řadit:

- **Online služby pro web scraping**, jako jsou Octoparse² a ParseHub³, které nabízejí web scraping jako službu. Tyto služby umožňují uživatelům specifikovat data k extrakci a poskytují možnosti pro automatizovaný sběr dat. Výhodou je uživatelsky přívětivé rozhraní, které zpřístupňuje web scraping i méně technickým uživatelům. [25]
- **Rozhraní API**, přes které se dá dostat ke strukturovaným datům přímým způsobem. API (Application Programming Interface) je soubor definic a protokolů pro integraci softwaru. Umožňuje aplikacím vzájemnou komunikaci a výměnu dat. Některé platformy poskytují API veřejně, aby podpořily vývojáře třetích stran ve vytváření nových služeb, které by mohly rozšířit funkčnost dané platformy. Veřejná API mohou být také využívána komerčně, kde platformy nabízejí svá data za finanční kompenzaci. [27] I když je API přístup obecně efektivnější a spolehlivější než web scraping, nemusí být dostupné pro všechny služby nebo mohou mít omezené možnosti, které nesplňují všechny uživatelské požadavky. [28]
- **Vlastní skripty pro web scraping**, které si výzkumníci a vývojáři vyvíjí svépomocí. Python a JavaScript jsou dvěma nejpopulárnějšími jazyky pro psaní těchto skriptů díky svým rozsáhlým knihovnám a frameworkům, které web scraping značně usnadňují. [28] Jelikož platformy Google Play a App Store nemají veřejná API, budu muset pro sběr dat využít této metody. V rámci této práce budu pro web scraping používat jazyk JavaScript s knihovnami google-play-scraper a app-store-scraper. Tyto knihovny jsou přímo vytvořené pro extrakci dat z distribučních platforem Google Play a App Store.

Etické aspekty

Při provádění web scrapingu je důležité zvážit etické a právní aspekty, jako jsou autorská práva, soukromí a smluvní podmínky webových stránek. Mnoho lidí považuje tuto metodu sběru dat za neetickou a v některých případech může být také nelegální.

Při web scrapingu bychom měli respektovat jistá etická i právní omezení [29]:

- Při web scrapingu záměrně nepřetěžujeme infrastrukturu cílové webové stránky.
- Data, která se snažíme extrahovat jsou veřejně dostupná a nejsou chráněna heslem.
- Získaná data by měla být primárně faktické povahy a jejich extrakce neporušuje práva jiné osoby (včetně autorských práv).

² <https://www.octoparse.com/>

³ <https://www.parsehub.com/>

- Informace získané procesem web scrapingu by měly sloužit k vytváření produktů, které transformují původní obsah a nepředstavují jen kopii. Neměly by být využity k odvedení uživatelů od původního zdroje nebo k vytváření produktů, které by byly ve značné míře totožné s produkty cílového webu.

2.2.2 Externí poskytovatelé dat

Vedle primárních zdrojů dat distribučních platforem Google Play a App Store můžeme pro sběr dat využít služby externích poskytovatelů dat. Platformy jako SensorTower⁴, AppTopia⁵ nebo AppMagic⁶ poskytují sofistikovanější sady údajů, které zahrnují demografické rozdělení uživatelů, analýzy uživatelského chování, benchmarking konkurentů, údaje o monetizaci a retenci uživatelů, angažovanosti a efektivitě reklamních kampaní. Tyto informace jsou získávány z různých zdrojů – veřejně dostupných dat, přímo od partnerských aplikací či agregací uživatelských dat. Např. platforma SensorTower má statistický panel s miliony uživatelů, od kterých kontinuálně sbírají data o jejich používání mobilních aplikací. [30] Externí poskytovatelé také využívají pokročilé algoritmy a umělou inteligenci pro provádění prediktivních analýz a prognózu trendů, což překračuje možnosti běžných metod web scrapingu.

I přes širokou škálu výhod, které tyto externí zdroje dat přinášejí, existují i potenciální nevýhody. Data, která tyto platformy poskytují nemusí být přesná, jelikož jsou některé metriky vypočítané pomocí aproximace. Reálné konkrétní informace o mobilních aplikacích znají pouze vydavatelé daných aplikací. Další nevýhodou je extrémně vysoká cena těchto služeb. Cena základní měsíční licence pro firemní přístup na tyto platformy se pohybuje kolem 400USD. [31]

Vlastní scraping dat, i když je časově náročnější a vyžaduje specifické technické dovednosti, umožňuje firmám větší kontrolu nad procesem získávání a správy dat. Vlastní extrakce je ekonomicky výhodnější, zvláště pro startupy nebo firmy s omezenými zdroji, což je důvod, proč jsem se pro tento typ sběr dat rozhodl.

2.3 Analýza

Jednou z nejdůležitějších fází cyklu CI je analýza, při které dochází k přeměně získaných informací na znalosti. V této fázi jsou nejdříve data transformována do struktury vhodné pro analýzu. To může zahrnovat agregaci dat, kategorizaci, výpočet nových proměnných a jejich převod do formátů, které jsou kompatibilní s analytickými nástroji. Tato fáze také zahrnuje tvoření anotací a metadat, která pomáhají v další analýze a interpretaci dat.

⁴ <https://sensortower.com/>

⁵ <https://apptopia.com/>

⁶ <https://appmagic.rocks/>

Po zpracování dat následuje identifikace vzorců, trendů a anomálií, což umožňuje analytikům rozpoznat potenciální hrozby nebo příležitosti. Tento krok vyžaduje schopnost syntézy a integrace rozličných datových zdrojů, což je často podporováno použitím specializovaných analytických modelů a metodologií, jako jsou SWOT analýzy, Porterova analýza pěti sil nebo PESTLE modely. V této fázi analytici také formulují doporučení pro akce, které by měly být podniknuty v reakci na nově získané informace. [24]

Výsledky analýzy jsou prezentovány ve formě zpravodajských zpráv či jiných prostředků pro distribuci, které jsou přizpůsobeny potřebám a očekáváním koncových uživatelů. Fáze analýzy je zásadní pro zpracování a interpretaci informací tak, aby byly přínosné pro rozhodovací procesy. Je založena na kritickém myšlení, odborných znalostech a systematickém přístupu k hodnocení informací

2.3.1 Vybrané metody pro analýzu trhu s mobilními aplikacemi

Pro to, abychom mohli analyzovat získané informace z distribučních platform, bude nutné je očistit, zkatégorizovat a celkově transformovat tak, abychom z nich dokázali vytěžit co nejvíce poznatků. K tomu využijí principů Business Intelligence a jeho nástrojů, abych tohoto cíle dosáhl. Pro přípravu dat, které budou vhodné pro analýzy použijí programovací jazyk Python, který umožní očištění, kategorizaci mapování a kalkulaci odvozených hodnot, které dodají datům větší výpovědní hodnotu. Po tomto zpracování budou data poslána do cloudového úložiště, odkud budou dostupná pro analýzu prostřednictvím nástroje Google Looker Studio, který umožní provádět pokročilé datové analýzy a vizualizace.

2.4 Distribuce

Distribuce informací představuje závěrečnou fázi zpravodajského cyklu Competitive Intelligence (CI). V této fázi jsou výsledky analýzy převedeny do zpravodajské formy, která je přizpůsobena pro bezprostřední využití cílovými uživateli. Zpravodajský produkt, by měl splňovat následující parametry [23]:

- **Obahuje využitelné informace**, přičemž by měly být zpracovány v relevantním kontextu a s prakticky využitelnými doporučeními, včetně identifikace potenciálních rizik.
- **Srozumitelná forma**, která umožní rychlé a jednoznačné pochopení situace.
- **Včasná dostupnost**, aby bylo možné přijmout nezbytná opatření dříve, než dojde ke krizové situaci.

V souvislosti s těmito požadavky hovoříme také o tzv. 4R – right time, right quality, right place a right product. Koncept 4R zdůrazňuje, že zpravodajské informace musí být doručeny ve správný čas, aby maximálně podpořily rozhodovací procesy a předešly krizím. Dále musí být zpravodajské informace správné kvality, což znamená přesné, spolehlivé a relevantní, aby byly základem pro informovaná rozhodnutí. Správné místo znamená, že informace jsou přístupné tam, kde jsou potřeba a ve formátu, který je kompatibilní s používanými systémy.

Správný produkt zajišťuje, že informace jsou prezentovány v užitečné formě, přizpůsobené specifickým potřebám uživatelů.

Důležitým aspektem distribuce informací je překonávání komunikačních bariér uvnitř organizace. Ideální je vytvoření otevřeného dialogu mezi analytikem CI a kompetentním manažerem, což podporuje efektivní přenos informací a zajišťuje, že jsou informace správně interpretovány a využity. Je třeba také nastavit správné procesy pro práci lidí se systémem, aby každý pracovník věděl, jaké má kompetence a povinnosti.

Formy distribuce informací mohou být různé a závisí na závažnosti a důvěrnosti informace. Tyto formy zahrnují osobní rozhovory nebo briefingy kompetentních manažerů, automatickou distribuci upozornění na identifikované hrozby či příležitosti prostřednictvím e-mailu nebo zobrazení na podnikovém intranetu. [23]

V rámci distribuce by se také nemělo zapomenout na získání zpětné vazby od iniciátora zpravodajského cyklu, přičemž by mělo být zhodnocena hodnota zpravodajského produktu. Tato zpětná vazba pak slouží jako podklad pro případné úpravy v rámci další iterace zpravodajského cyklu. [32]

2.4.1 Zpravodajský produkt pro analýzu trhu s mobilními aplikacemi

V kontextu této práce jsem si definoval, že bude pro distribuci informací využíván dashboard. Dashboard bude automaticky nahrávat data z cloudového úložiště, čímž se zajistí aktuálnosti informací. Dashboard mi umožní prezentovat informace v přehledné podobě pomocí vizualizací, což značně usnadní jejich interpretaci pro ostatní uživatele. Další výhodou je snadná dostupnost, jelikož bude dashboard hostován na cloudové službě, čímž se zajistí přístup pomocí jakéhokoliv zařízení s přístupem na internet.

3 Business Intelligence

V rámci této práce je také důležité vysvětlit termín Business Intelligence, jehož koncepty budou využity při analýze trhu s mobilními aplikacemi.

V současném vysoce konkurenčním prostředí je zásadní, aby analytici a manažeři mohli činit rychlá a přesná rozhodnutí. Pro tyto účely potřebují mít k dispozici objektivní informace, které jsou snadno dostupné, nevyžadují složitou manipulaci a umožňují rychle formulovat požadavky na nové informace odpovídající aktuálním potřebám firmy. Zajištěním těchto požadavků se v organizaci zabývá oblast zvaná Business Intelligence (BI).

Dle definice: „*Business Intelligence je sada procesů, aplikací a technologií, jejichž cílem je účinně a účelně podporovat rozhodovací procesy ve firmě. Podporují analytické a plánovací činnosti podniků a organizací a jsou postaveny na principech multidimenzionálních pohledů na podniková data.*“ [33]

Technologie BI jsou schopny manipulovat s velkým množstvím strukturovaných a nestruturovaných dat. Primárním úkolem BI je usnadnit interpretaci těchto dat – poskytovat k nim interaktivní přístup, umožňovat jejich manipulaci a tím poskytovat podnikovým analytikům nástroje pro provádění analýz, které jsou klíčové pro strategická rozhodnutí vedení. Zjednodušeně řečeno, Business Intelligence se zakládá na procesu přeměny dat na informace, které následně vedou k rozhodnutím a poté k akcím. [34]

3.1 Rozdíl mezi Business Intelligence a Competitive Intelligence

Přestože oblasti Competitive Intelligence (CI) a Business Intelligence (BI) byly původně vnímány jako odlišné a specializované disciplíny, v současné době dochází k jejich stále častějšímu propojování a vzájemnému doplňování. Tradičně bylo rozlišováno, že CI pracuje primárně s externími, nestruturovanými daty, zatímco BI se zaměřuje na interní, strukturovaná data. V současnosti se však hranice mezi těmito pojmy stírají, a to díky technologickému pokroku, který umožňuje BI systémům využívat i externí a nestruturovaná data.

Pro efektivní rozhodování je nezbytná integrace jak interních, tak externích dat, ať již strukturovaných, tak nestruturovaných. Tento přístup vede k lepšímu porozumění situací, které organizaci ovlivňují.

Výhody externích dat spočívají v poskytování dalšího stupně informační podpory pro manažerské rozhodování, zatímco interní data nabízejí přehled o interních zdrojích a procesech. Nicméně, obrovský objem dostupných externích dat představuje výzvu, kterou je třeba řešit pomocí specializovaných nástrojů pro zpracování nestruturovaných dat.

V architektuře podnikového informačního systému lze tedy CI a BI vidět jako vzájemně se doplňující komponenty, které společně formují tzv. Strategic Intelligence. Tato integrace poskytuje firmám strategickou výhodu v rychle se měnícím a vysoce konkurenčním prostředí, což je klíčové pro udržení a posílení jejich tržní pozice. [23]

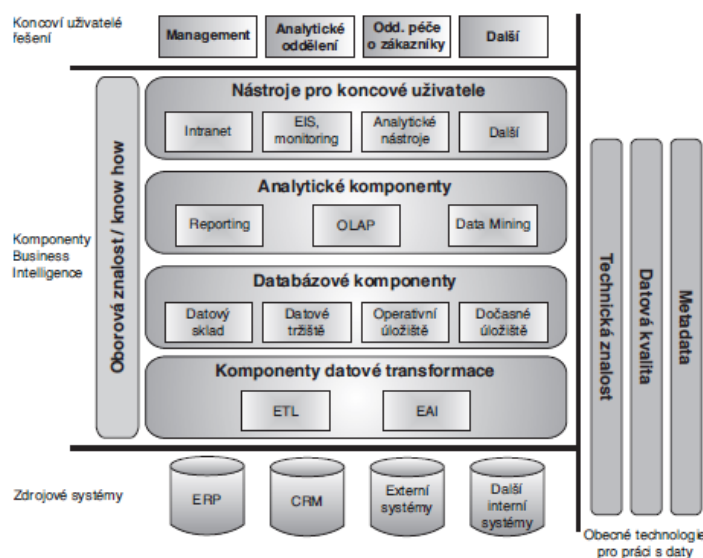
3.2 Architektura BI

Řešení BI spočívá v kombinaci různých komponent, jejichž konfigurace se odvíjí od specifických potřeb jednotlivých firem a uživatelů. Rozsah těchto řešení může být od jednoduchých a nenákladných až po velmi složité systémy, které vyžadují značné technologické, finanční a personální investice. Ačkoliv neexistuje standardizované nebo dominantní řešení architektury BI, lze alespoň identifikovat obecnou koncepci architektury těchto systémů.

V rámci obecné koncepce BI systému lze vymezit tyto vrstvy [33]:

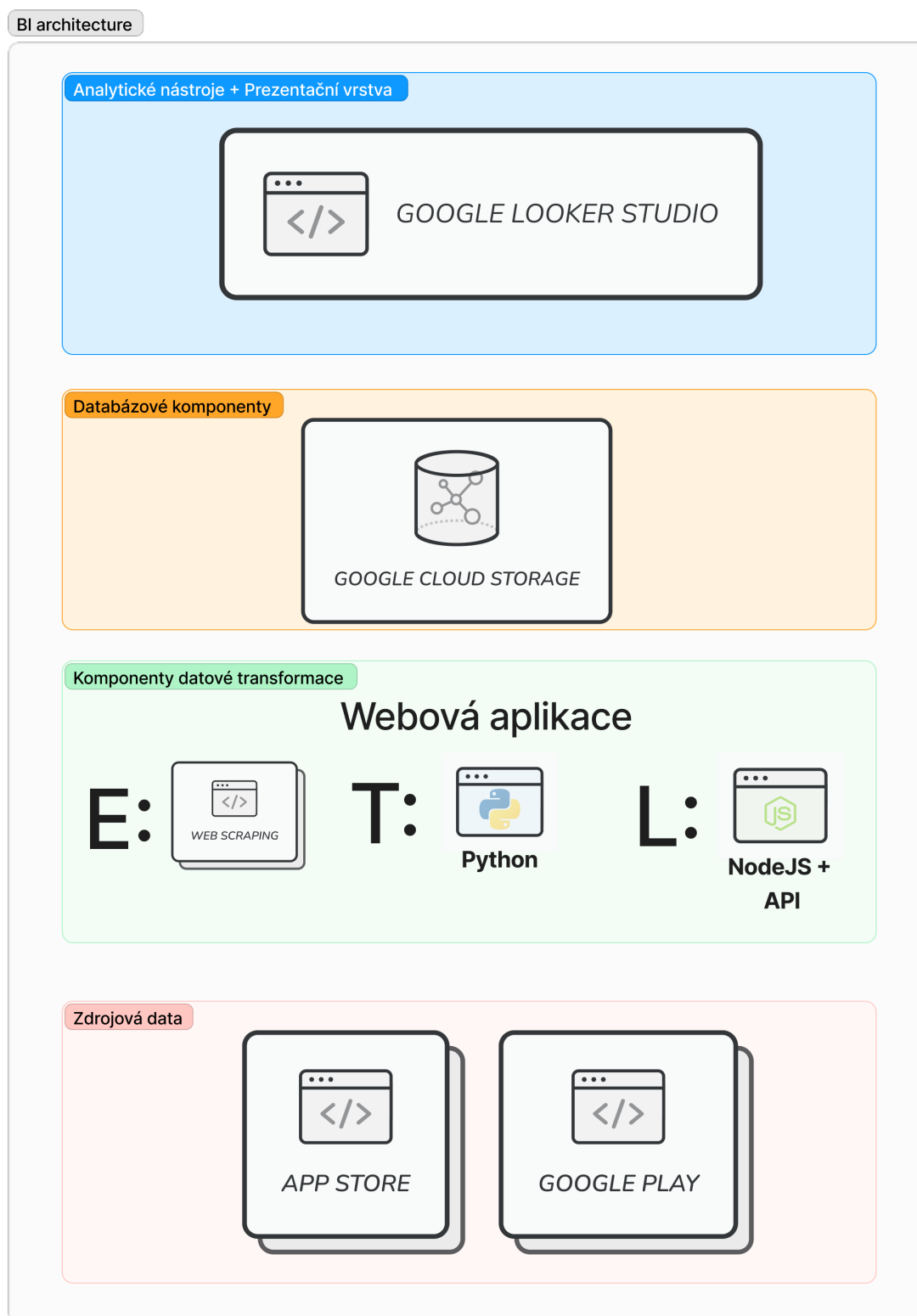
- **Vrstva pro extrakci, transformaci, čištění a nahrávání dat** (komponenty datové transformace).
- **Vrstva pro ukládání dat** (databázové komponenty).
- **Vrstva pro analýzu dat** (analytické komponenty).
- **Prezentační vrstva** (nástroje pro koncové uživatele).
- **Vrstva oborové znalosti** (oborová znalost/know how).

Všechny tyto vrstvy hrají klíčovou roli v procesu Business Intelligence. Tento proces spočívá v získání, transformaci a čištění dat ze zdrojových systémů za pomoci komponentů datové transformace, následném uložení těchto dat do databází za použití databázových komponentů, provedení analýzy za využití analytických komponentů, a nakonec publikování výsledků analýzy za pomoci nástrojů pro koncové uživatele. [33] Následující obrázek znázorňuje obecnou koncepci architektury řešení BI v rámci informačního systému.



Obrázek 7: Obecná koncepce architektury BI

V rámci této práce však nebudeme počítat s architekturou v rámci interního informačního systému. Nástroj, který byl vytvořen, simuluje roli BI komponentů v rámci architektury BI. Pro upřesnění jsem vytvořil tento model, který tuto skutečnost znázorňuje.



Obrázek 8: Konceptuální návrh architektury nástroje pro analýzu dat, Zdroj: autor

3.3 Datové zdroje v BI

Datové zdroje jsou základem každého systému Business Intelligence (BI). Kvalita a dostupnost dat určují efektivitu analýz a přesnost rozhodování v podniku. Můžeme rozlišovat tři hlavní typy datových zdrojů – interní, externí a osobní. [35]

Interní datové zdroje

Interní data pochází přímo z různých interních systémů organizace a zahrnují [35]:

- **Transakční data a informace z bodů prodeje (POS)**, což jsou systémy, které obsahují důležité údaje o finančních transakcích.
- **Systémy pro správu vztahů se zákazníky (CRM)**, které poskytují informace o vztazích se zákazníky, které jsou klíčové pro tvorbu marketingové strategie.
- **Interní záznamy** ve formě firemních dokumentů, které obsahují důležité informace o firemních operacích a procedurách.
- **Archivy a historická data v digitální podobě**, jako jsou interní dokumenty a záznamy, poskytují vhled do firemních operací a dlouhodobých trendů.
- **Data z IoT zařízení**, jako jsou např. senzory ve vozidlech, mohou nabídnout podrobné informace o využití vozidla či spotřebě paliva.
- **Ostatní interní systémy**, jako jsou např. systémy pro projektové, finanční a HR řízení.

Externí datové zdroje

Pro doplnění interních dat mohou organizace využívat i externí zdroje dat. Mezi ně můžeme řadit [35]:

- **Sociální média**, jako jsou Facebook, X či Reddit, ze kterých lze čerpat data o zákaznických preferencích a tržních trendech.
- **Vládní a veřejně dostupná data** nabízí širokou škálu informací, včetně statistik obyvatelstva, zdravotnických dat a ekonomických indikátorů, které mohou poskytnout cenné informace o tržních podmínkách.
- **Geografické informační systémy (GIS)** poskytují data spojená s konkrétními geografickými lokacemi, což umožňuje provádět detailní analýzy založené na lokaci.
- **Webové stránky, blogy a online fóra** jsou bohatým zdrojem nestructurovaných dat různého charakteru.

Osobní datové zdroje

Ve zvláštních případech mohou být zdrojem dat osobní databáze a excelové tabulky klíčových pracovníků, které mohou poskytnout cenná data, která jsou vhodná pro integraci s ostatními datovými zdroji. [35]

3.4 Databázové komponenty

Databázové komponenty jsou klíčové pro ukládání, aktualizaci a správu dat v systémech Business Intelligence (BI). Tyto komponenty zahrnují datové sklady, datová tržiště, operativní datová úložiště a dočasná úložiště dat, které společně podporují různé potřeby analytických a operativních procesů v organizaci.

Datový sklad (Data Warehouse)

Datový sklad je klíčovým komponentem pro BI, který slouží jako centrální úložiště historických a aktuálních dat. Tato data jsou strukturovaná tak, aby podporovala činnosti jako online analytické zpracování (OLAP), data mining a reporting. Na rozdíl od produkčních databází, které jsou zaměřeny na zpracování dat a podnikové operace, datový sklad je určen primárně k podpoře rozhodovacích procesů. Data v datovém skladu jsou integrována z různých zdrojů a uspořádána podle klíčových obchodních dimenzí, jako jsou prodeje, produkty nebo zákazníci, a jsou časově rozlišená pro analýzu trendů a predikce. [34]

Datové tržiště (Data Mart)

Datové tržiště je obvykle menší a zaměřené na specifickou oblast nebo oddělení firmy a fungují jako decentralizovaný datový sklad. Tato uspořádání mohou být buď závislá, přímo odvozená z datového skladu, což zajišťuje vysokou kvalitu a konzistenci dat, nebo nezávislá, kdy jsou navržena speciálně pro určité obchodní potřeby a nejsou přímo spojena s hlavním datovým skladem. [34]

Operativní datové úložiště (ODS)

Operativní datové úložiště poskytuje aktuální pohled na provozní data, která jsou integrována z různých firemních systémů. Na rozdíl od datového skladu, ODS obsahuje minimální historická data a je zaměřeno na podporu krátkodobých, operativních rozhodnutí. [34]

Dočasné úložiště dat (Staging Area)

Dočasné úložiště dat slouží k prvotnímu shromáždění dat z různých produkčních systémů před jejich transformací. Tato data jsou zpravidla nekonzistentní a neagregovaná, a slouží k zefektivnění a zrychlení procesů extrakce a transformace dat. Po zpracování jsou data z tohoto úložiště přenesena do dalších komponent a z dočasného úložiště jsou odstraněna. [34]

3.5 Datová transformace

Pro tuto diplomovou práci je relevantní komponenta datové transformace zvaná ETL (Extract, Transform, Load). ETL představuje způsob, jakým firmy kombinují data z více systémů do jediné databáze, datového úložiště, datového skladu nebo datového jezera. ETL lze využít k ukládání starších dat nebo, jak je dnes běžnější, k agregaci dat pro analýzu a podporu obchodních rozhodnutí. [36] Výše zmíněná webová aplikace bude v rámci nástroje pro analýzu trhu s mobilními aplikacemi sloužit jako komponenta pro ETL.

ETL systémy jsou založeny na procesu ETL. Extrakcí dat se rozumí získání dat z datových zdrojů, které mohou být homogenní či heterogenní. Tato část procesu je nejdůležitějším aspektem ETL, jelikož správná extrakce dat přímo ovlivňuje úspěšnost všech dalších procesů. Hlavní překážka spočívá v různorodosti datových zdrojů. Firmy mají často mnoho různých datových zdrojů, přičemž každý z nich může využívat odlišnou technologii pro ukládání dat s odlišnými datovými typy a strukturami. Při extrakci proto probíhá i ověřování dat, aby se potvrdilo, zda mají data ze zdrojových systémů správnou hodnotu. Transformace spočívá ve vyčištění dat a jejich následné převedení do vhodné struktury či formátu. Na extrahovaná data se aplikují různá pravidla a funkce, které je připravují pro načtení do databázových systémů. Tyto funkce zahrnují výpočty nových hodnot z existujících sloupců, kombinování hodnot z různých sloupců, standardizaci formátů sloupců a odstranění duplicitních záznamů. Poslední částí procesu ETL je nahrání, při kterém se data nahrají do specifických datových struktur, resp. do datových schémat a datových skladů. [37]

3.6 Analytické nástroje

Jak už z názvu vypovídá, tyto komponenty pokrývají činnosti spojené s analýzou dat. Cílem je převést data, která jsou uložena v databázových komponentech, na znalosti, které pomohou zaměstnancům při rozhodování. Zmíním se pouze o relevantních komponentech souvisejících s touto diplomovou prací.

Report

Reporting je činnost spojená s dotazováním se do databází pomocí standardních rozhraní těchto databází. Lze ho dělit na standardní a ad hoc reporting. Standardní reporting je založen na principu spouštění předpřipravených dotazů v určitých časových periodách. Při ad hoc reportingu jsou na databáze formulovány jednorázové specifické dotazy. Výstupem jsou grafické reporty či dashboardy, které se snaží o prezentaci klíčových informací pracovníkům. [33]

Dashboarding

Dashboard je sada nejdůležitějších informací na jednom místě. Na rozdíl od reportingu se tyto informace typicky zobrazují na jedné obrazovce či stránce, kdežto reporty bývají často na více stránkách. Dashboardy umožňují koncovým uživatelům flexibilně ovládat a měnit

obsah. Jsou typické svojí stručností, přehledností a poskytují identifikaci důležitých informací pro uživatele. [38]

Reporting

Reporting zahrnuje činnosti dotazování na databáze pomocí standardního rozhraní. Rozlišujeme mezi standardním reportingem, který zahrnuje pravidelné spouštění předdefinovaných dotazů, a ad hoc reportingem, který umožňuje formulování jednorázových specifických dotazů. Výsledkem reportingu jsou často grafické reporty nebo dashboardy, které prezentují klíčové informace de forma přehledná a srozumitelná. [38]

Dashboard

Dashboard představuje agregaci nejdůležitějších informací na jednom místě, typicky na jedné obrazovce nebo stránce, na rozdíl od vícestránkového reportingu. Dashboardy umožňují uživatelům flexibilní kontrolu a úpravu obsahu, přičemž jsou charakteristické svou stručností, přehledností a schopností rychle identifikovat klíčové informace. Tento analytický nástroj je využit v této práci. [38]

Umělá inteligence

Umělá inteligence, která je v současnosti často diskutovaným tématem, zahrnuje schopnosti počítačů vykonávat činnosti podobné těm, které provádějí lidé. To zahrnuje vnímání prostředí, budování znalostí, uvažování, rozhodování a reakce na situace. Umělá inteligence se spoléhá na techniky jako datová integrace, data mining a strojové učení k vytváření komplexních samostatných systémů schopných adresovat specifické problémy. V budoucnosti se očekává, že umělá inteligence bude hrát klíčovou roli ve vývoji podnikových BI systémů. [38]

4 Návrh architektury analytického nástroje

Cílem této diplomové práce je vyvinout pro firmu Robology komplexní analytický nástroj, jenž jí umožní hlouběji porozumět konkurenčnímu prostředí a efektivně formulovat obchodní strategie. Jelikož je trh s mobilními aplikacemi velmi proměnlivý, byl jedním z hlavních požadavků na tento nástroj, aby obsahoval co nejaktuálnější data. Důraz je kladen také na flexibilitu, dostupnost a snáze používání.

Podle těchto požadavků byla navržena webová aplikace, která extrahuje aktuální data o aplikacích z distribučních platforem Google Play a App Store na základě kritérií specifikovaných uživatelem prostřednictvím technik web scrapingu. Data jsou po extrakci transformována tak, aby byla připravena pro podrobnou analýzu a vizualizaci. V procesu transformace dochází k očištění dat, jejich kategorizaci a k výpočtu nových metrik, které poskytují hlubší vhledy do trhu s mobilními aplikacemi. Tato data jsou následně vizualizována prostřednictvím interaktivního dashboardu, který nabízí intuitivní uživatelské rozhraní pro prohlížení různých metrik a umožňuje uživatelům snadně interpretovat data.

Aplikace je navržena, aby byla modulární a škálovatelná pro budoucí rozšíření a vývoj nových funkcionalit. Webová aplikace⁷, dashboard⁸ a celý zdrojový kód⁹ je veřejně přístupný.

Webová aplikace je rozdělena do několika klíčových komponentů. V této podkapitole tyto komponenty představím a stručně popíšu technologie, které jsem při vývoji této aplikace použil.

4.1 Frontend

Frontend (uživatelské rozhraní) je postaven na technologiích HTML, CSS a JavaScript. HTML (HyperText Markup Language) je standardní značkovací jazyk používaný pro tvorbu webových stránek a určuje strukturu obsahu na internetových stránkách. CSS (Cascading Style Sheets) definuje vizuální styl těchto stránek a JavaScript se používá pro přidání interaktivity komponentů a zajišťuje logiku pro posílání informací z vyplněných formulářů na server. Uživatelé prostřednictvím tohoto uživatelského rozhraní mohou specifikovat kritéria pro extrakci dat a posílat požadavky na zpracování.

⁷ <https://robology-app-analytics-dy2n9.ondigitalocean.app/>

⁸ <https://lookerstudio.google.com/reporting/7f569003-3a74-419c-b2f2-ecoff39f5761>

⁹ <https://github.com/Robology-Github/app-scraper-dp>

4.2 Backend

Backend (server) je klíčovou složkou celé aplikace, která je zodpovědná za logiku toku dat, integraci s externími službami a komunikaci s frontendem. Tato část aplikace je vybudována na platformě Node.js, oblíbeném nástroji pro vývoj serverových aplikací. Node.js bylo zvoleno z důvodu jeho schopnosti efektivně zvládat více požadavků najednou bez toho, aby docházelo k zablokování hlavního procesu. To je možné díky jeho asynchronní povaze, kdy může provádět několik úloh současně v pozadí. Node.js také umožňuje používat JavaScript nejen na frontendu, ale i na backendu, což usnadňuje jednotný vývoj celé aplikace pomocí jednoho programovacího jazyka. [39]

Další technologií, kterou jsem při vývoji využil, je framework Express.js. Tento framework umožňuje jednoduše definovat, jak aplikace reaguje na různé uživatelské interakce pomocí API endpointů. API endpointy přijímají požadavky z webového rozhraní, v této aplikaci přijímají kritéria pro extrakci dat z distribučních platform Google Play a App Store.

Extrakce dat je realizována pomocí Node.js modulů *google-play-scraper*¹⁰ a *app-store-scraper*¹¹. Pomocí těchto modulů můžeme zasílat požadavky na servery Google Play a App Store prostřednictvím rozhraní s předpřipravenými metodami, které nám umožní definovat kritéria pro sběr dat. Moduly dokážou simulovat chování uživatele procházejícího stránku dané distribuční platformy a programově extrahovat HTML soubor.

Jakmile je HTML stránka z distribuční platformy získána, moduly používají různé mechanismy pro extrakci a parsování dat z HTML tagů. Tyto data následně převedou do strukturovaného JSON formátu pomocí mapovacích skriptů. Modul při procesu extrakce iteruje skrze stránky, extrahuje a akumuluje data o aplikacích, dokud není dosaženo nastaveného počtu požadovaných aplikací.

Moduly obsahují mechanismy pro řízení toku požadavků, jako např. udržení konzistence cookies mezi požadavky, stránkování a držení stejné relace prohlížeče. Zajišťují také regulaci frekvence odesílaných požadavků, čímž se minimalizuje riziko blokování serverem pro podezřelou aktivitu. V rámci testování jsem narazil na to, že bezpečný limit pro počet extrahovaných aplikací je 200. Během vývoje se stalo, že mi distribuční platformy zablokovaly přístup na několik hodin, kvůli nadměrnému počtu požadavků.

4.3 Vrstva přípravy a transformace dat

Po získání dat pomocí web scrapingu je spuštěn skript napsaný v jazyce Python, který zajišťuje komplexní transformaci a zpracování dat. Python je programovací jazyk oblíbený mezi datovými analytiky především díky jeho bohaté sadě knihoven pro manipulaci s daty. Pro čištění, kategorizaci a manipulaci s daty používám knihovnu *pandas*, která poskytuje

¹⁰ <https://github.com/facundoolano/google-play-scraper>

¹¹ <https://github.com/facundoolano/app-store-scraper>

nástroje pro rychlé a snadné zpracování velkých objemů dat. V rámci analýzy textu recenzí využívám knihovny *NLTK* pro zpracování přirozeného jazyka, *LangDetect* pro detekci jazyka textu a knihovnu *Transformers* spolu s modelem *BERT* pro pokročilou analýzu sentimentu uživatelských recenzí. Dále používám knihovnu *TextBlob* pro další textové analýzy a *ISO 639* pro identifikaci jazykových kódů. Tato kombinace knihoven mi umožnila komplexně transformovat data, aby byla relevantní pro další analýzu a vizualizaci. Po transformaci vznikají výsledné soubory ve formátu csv, které jsou připraveny k nahrání do datového úložiště.

4.4 Datové úložiště

Pro ukládání dat byla zvolena služba Google Cloud Storage¹² (GCS). Díky jejímu API je možné snadno integrovat GCS s aplikacemi, což umožňuje aplikacím bezpečně manipulovat daty. Data jsou organizována do tzv. bucketů, které umožňují efektivní správu a jsou konfigurovatelné pro různé geografické oblasti, což optimalizuje rychlost přístupu a redukuje náklady. Ve společnosti Robology používáme Google Cloud Platform jako infrastrukturu pro hostování databází a souborů, což byl také z jedním faktorů pro výběr této služby.

4.5 Vizualizace dat

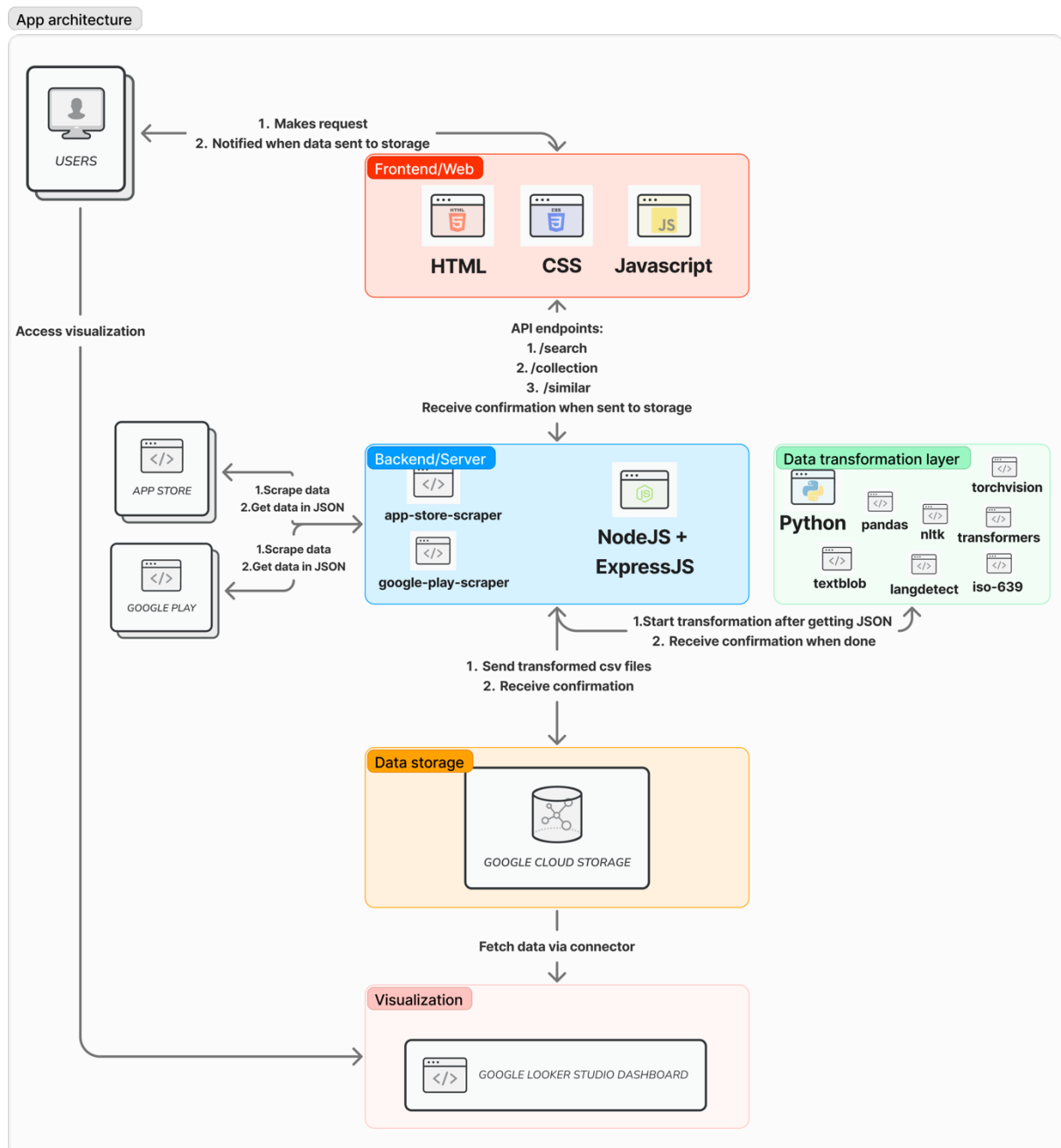
Jak již bylo zmíněno, pro vizualizaci dat je využit nástroj Google Looker Studio¹³. S jeho pomocí jsem vytvořil dashboardy, ve kterých může uživatel interaktivně prozkoumávat data, provádět hloubkové analýzy a získat detailní přehled o vybraných aplikacích, které si sám nadeřinoval v rámci webové aplikace. Data jsou z datového úložiště automaticky načítána do dashboardu pomocí poskytnutého nativního konektoru poskytovaného společností Google. Google Looker Studio tak umožňuje dynamicky pracovat s daty a vytvářet vizualizace v reálném čase.

¹² <https://cloud.google.com/storage>

¹³ <https://cloud.google.com/looker-studio>

4.6 Diagram architektury

Pro lepší představu jsem vytvořil zjednodušený diagram architektury webové aplikace, který popisuje hlavní komponenty a ukazuje, jak jsou mezi sebou propojené.



Obrázek 9: Zjednodušený diagram webové aplikace, Zdroj: autor

5 Vývoj nástroje pro analýzu trhu s mobilními aplikacemi

V rámci této kapitoly objasním, jak jsem postupoval při vývoji webové aplikace.

5.1 Nástroje a služby použité při vývoji

Při vývoji aplikace jsem využil tyto nástroje a služby:

- **Visual Studio Code** jako vývojové prostředí.
- **Github** pro verzování aplikace.
- **Webflow** pro návrh a tvorbu uživatelského rozhraní.
- **Github Copilot** pro vysvětlování kódu pomocí umělé inteligence.
- **Google Cloud Storage** jako úložiště transformovaných dat
- **Google Looker Studio** pro vytvoření vizualizací.
- **Docker** pro automatizovanou orchestraci nasazení na webový server.
- **DigitalOcean** pro nasazení a hosting webové aplikace.
- **Jupyter Notebook** pro tvorbu skriptů pro transformaci dat v pythonu.

Při vývoji a dokumentaci funkcí aplikace byl použit anglický jazyk.

5.2 Vývoj frontendu

Jak již bylo zmíněno, frontend v této aplikaci má za účel poskytnout rozhraní, kam mohou uživatelé specifikovat svá kritéria pro extrakci dat. V této kapitole popíšu, jak v této aplikaci frontend funguje a stručně vysvětlím, jak jsem při vývoji postupoval.

5.2.1 Uživatelské rozhraní

Při tvorbě uživatelského rozhraní jsem musel vymyslet, jak bude uživatel ovládat aplikaci, a jakým způsobem bude definovat kritéria, podle kterých chce získat data o aplikacích. Nejlepší variantou se ukázaly být oddělené formuláře pro každý typ extrakce:

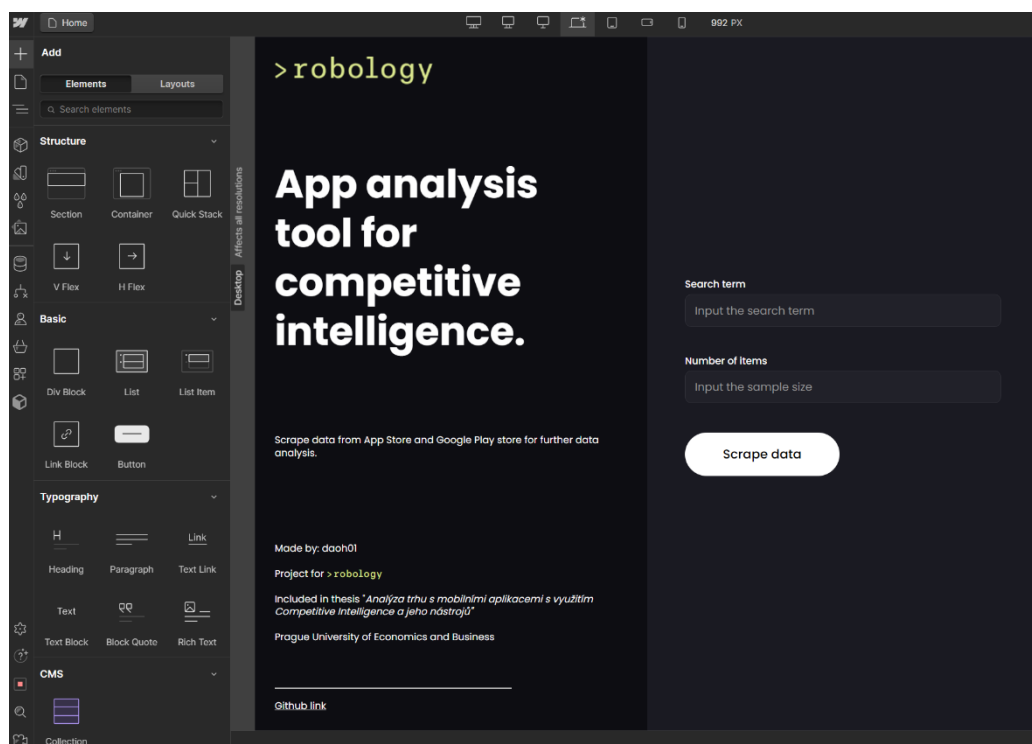
- **Pro extrakci dat aplikací podle klíčového slova** musí uživatelé zadat klíčové slovo, zemi a počet aplikací, které chtějí extrahovat.
- **Pro extrakci dat aplikací podle kolekcí daných distribučních platforem** musí vybrat předdefinovanou kolekci, zemi a počet aplikací k extrakci.
- **Pro web scraping podobných aplikací** musí uživatelé poskytnout název specifické aplikace a zemi, což web scraperu umožní hledat podobné aplikace.

Tyto typy extrakce a jejich kritéria korespondují s metodami, které můžou být volány v rámci knihoven pro extrakci dat z distribučních platform.

Pro tvorbu tohoto uživatelského rozhraní jsem využil no-code platformu Webflow¹⁴. Tato platforma umožňuje tvorbu webových stránek bez nutnosti psaní kódu, což se děje prostřednictvím grafického rozhraní, kde je možné vizuálně konfigurovat komponenty webu. Webflow generuje sémanticky čistý kód z graficky navrženého rozhraní, který je připraven k dalším úpravám, což značně zrychluje proces vývoje.

V tomto nástroji jsem j vizuální strukturu komponentů, jako jsou vstupní pole a tlačítka. Po dokončení návrhu jsem projekt exportoval, čímž jsem získal HTML, CSS a JavaScriptový kód, který jsem následně mohl upravit podle potřeby. V dalších fázích vývoje jsem do webového rozhraní přidával další funkcionality, jako jsou dialogová okna pro zobrazení chyb nebo úspěchu, načítací animace při odesílání dat a navigace mezi jednotlivými formuláři. Tyto výsledné soubory tvoří můj frontend:

- *index.html* – hlavní soubor s HTML kódem.
- *app.js* – JavaScriptový soubor obsahující logiku aplikace.
- *webflow.css* – CSS soubor generovaný platformou Webflow.
- *styling.css* – vlastní styly definované pro další úpravy designu.
- *normalize.css* – CSS soubor použitý pro zajištění konzistentního zobrazení ve všech prohlížečích.



Obrázek 10: Snímek obrazovky pracovního prostředí platformy Webflow, Zdroj: autor

¹⁴ <https://webflow.com/>

5.2.2 Integrace uživatelského rozhraní s backendem

Po vytvoření frontendové části aplikace bylo nezbytné zajistit, že data z formulářů budou správně zpracována serverovou částí aplikace. Data jsou z frontendu posílána na server pomocí AJAX (Asynchronous JavaScript And XML), což umožňuje asynchronní komunikaci mezi klientem a serverem. Tento přístup umožňuje aplikaci posílat a přijímat data a manipulovat s komponenty webu bez nutnosti znovu načítat celou stránku. [40]

Pro všechny formuláře jsem definoval funkce, které na webu zachycují uživatelské kliknutí na tlačítko pro odeslání. Když se taková událost stane, funkce získá všechna zadaná data z webových komponentů a odešle je na definované endpointy prostřednictvím HTTP požadavků. Zobrazí se animovaný indikátor, který uživateli říká, že jeho požadavek je zpracováván. Backend při tomto stavu extrahuje, transformuje a nahraje data do datového úložiště. Po obdržení odpovědi od serveru o tom, zda proces proběhl úspěšně, webová aplikace zobrazí okno se zprávou a odkazem na dashboard, viz Výpis 1.

Výpis 5: Ukázka kódu pro odeslání dat na endpoint `/search` v rámci souboru `app.js`, Zdroj: autor

```
1 // searchForm logic
2 document.addEventListener("DOMContentLoaded", (event) => {
3   document
4     .getElementById("searchForm")
5     .addEventListener("submit", async function (e) {
6       e.preventDefault();
7       const backdrop = document.getElementById("loading-backdrop"); /
8       backdrop.style.display = "block"; // Show the backdrop and spinner
9       const searchTerm = document.getElementById("searchTerm").value;
10      const numResults = document.getElementById("numResults").value;
11      const country = document.getElementById("searchCountryList").value;
12      try {
13        const response = await fetch(
14          // Fetch the data
15          `/search?term=${encodeURIComponent(
16            searchTerm
17          )}&num=${encodeURIComponent(numResults)}
18            &country=${encodeURIComponent(country)}`
19        );
20        if (!response.ok) throw new Error("Failed to fetch data");
21        const data = await response.json();
22        showModal("Data successfully scraped and transformed.
23          It may take up to 3 minutes to upload the data");
24      } catch (error) {
25        showModal("Error during data fetching: " + error.message);
26      } finally {
27        backdrop.style.display = "none"; // Hide the backdrop and spinner
28      }
29    });
30 });
```

5.3 Vývoj backendu

Vývoj backendu byl klíčovým aspektem diplomové práce, jelikož je to nejdůležitější část webové aplikace, která je zodpovědná za veškerý tok dat. V této kapitole objasním, jak jsem postupoval při vývoji a popíšu principy a technologie, na kterých je backend postaven. Celá logika této části aplikace je obsažena v souboru *server.js*.

5.3.1 API endpointy

Pro zajištění komunikace s frontendem byly navrženy a implementovány API endpointy. Tyto endpointy slouží jako přístupové body pro specifické požadavky z frontendu a byly definovány s využitím frameworku Express.js. Definoval jsem definoval tři hlavní API endpointy, s nimiž frontend komunikuje: */search*, */collection* a */similar*. Pomocí endpointu */search* může uživatel extrahovat data z distribučních platform na základě klíčových slov. Endpoint */collection* umožňuje uživatelům extrahovat data na základě kolekcí, které jsou definovány distribučními platformami – pro tento projekt jsem vybral kolekce *Top Free* (nejlepší zdarma), *Top Paid* (nejlepší placené) a *Top Grossing* (nejvíce vydělávající), které jsou shodné pro obě distribuční platformy. Endpoint */similar* umožňuje uživatelům extrahovat data o aplikacích, které jsou podobné uživatelsky zadané aplikaci. V rámci těchto endpointů je rovněž důležité, aby uživatel specifikoval zemi, protože platformy Google Play a App Store poskytují pro každou zemi odlišná data.

Jakmile uživatel odešle požadavek z uživatelského rozhraní, backend extrahuje parametry vyhledávání z URL dotazu. Tyto parametry jsou zásadní pro definování kritérií pro data, která mají být získána. Pokud uživatel nezadá některý z těchto parametrů, server vrátí chybový stav 400 (Bad Request) s vysvětlením, že požadavek není kompletní. Toto opatření slouží k tomu, aby server zpracovával pouze kompletní a správně formulované dotazy. V případě, že jsou všechny parametry správné, server pokračuje tím, že spustí funkci pro extrakci a transformaci dat. Výsledky této funkce jsou poté vráceny ve formátu JSON. Pokud během zpracování požadavku dojde k jakýmkoli chybám, server vrátí chybový stav 500 (Internal Server Error). Celý tento proces je patrný z výpisu kódu, viz Výpis 2.

Výpis 2: Ukázka kódu pro definici endpointu */search* v rámci souboru *search.js*, Zdroj: autor

```
1 app.get("/search", async (req, res) => {
2   const { term, country, num } = req.query;
3
4   if (!term || !country || !num) {
5     return res.status(400).send("Missing term or num parameter");
6   }
7
8   try {
9     const results = await searchAndFetchAppDetails(term, country, num);
10    res.json(results);
11  } catch (error) {
12    console.error(error);
13    res.status(500).send("Error fetching app details");
14  }}
```

5.3.2 Extrakce dat

Pro proces extrakce dat jsem vytvořil tři klíčové funkce: *searchFetchAppDetails*, *collectionFetchAppDetails* a *similarFetchAppDetails*. Každá z těchto funkcí je asociována s konkrétním API endpointem:

- Funkce *searchFetchAppDetails* se spouští z endpointu */search*.
- Funkce *collectionFetchAppDetails* se spouští z endpointu */collection*.
- Funkce *similarFetchAppDetails* se spouští z endpointu */similar*.

Tyto funkce jsou aktivovány vždy, když je z webového rozhraní odeslán dotaz na jejich odpovídající API endpoint s korektně definovanými kritérii pro extrakci dat. Pro extrakci dat jsou využívány následující metody z modulů *google-play-scraper* a *app-store-scraper*:

- Metoda *search* vrací seznam aplikací, které odpovídají zadanému vyhledávacímu termínu.
- Metoda *list* vrací seznam aplikací z předdefinovaných kolekcí na dané distribuční platformy.
- Metoda *similar* vrací seznam aplikací podobných zadané aplikaci.
- Metoda *app* umožňuje získání kompletních detailů o aplikaci.
- Metoda *review* získává recenze pro specifickou aplikaci.

Funkce nejprve pomocí metod *list*, *search* a *similar* získávají seznamy aplikací dle zadaných kritérií. Po získání těchto seznamů následně volají metody *app* a *review* k získání detailních dat o jednotlivých aplikacích a jejich recenzích. Výstupy těchto metod jsou poté sloučeny do jednoho JSON souboru. Data jsou následně transformována do formátu CSV, což mi umožňuje s nimi lépe manipulovat během procesu čištění a transformace. Z tohoto procesu vznikají dva soubory, *GooglePlayOutput.csv* a *AppStoreOutput.csv*. Jakmile jsou tyto soubory s extrahovanými daty vytvořeny, spouští se skript *transform.py*, který zahajuje proces čištění a transformace dat.

5.3.3 Spuštění datové transformace

Pro spuštění datové transformace jsem implementoval funkci *executePythonScript*, která zajišťuje spuštění Python skriptu přímo z běhu aplikace. Tato funkce využívá metodu *spawn* pro vytvoření nového procesu, ve kterém se Python skript spustí s předem definovanými parametry. Je nutné specifikovat, v jakém prostředí bude tento proces spuštěn – v mém případě jsem musel nastavit cestu k pythonovému interpreteru. Tento přístup umožňuje integraci různých programovacích jazyků v jedné aplikaci. Při volání funkce je také potřeba definovat název souboru se skriptem, název funkce pro transformaci v rámci skriptu, cestu k souboru s extrahovanými daty a cestu k transformovanému souboru s daty.

Po dokončení běhu skriptu se vyhodnocuje návratový kód, podle kterého aplikace rozhodne, zda byl skript ukončen úspěšně. V případě, že byl ukončen neúspěšně, je chybová hláška

odeslána na uživatelské rozhraní. Pokud je skript úspěšně dokončen, transformovaná data jsou nahrána do datového úložiště.

Výpis 3: Definice funkce `executePythonScript` v rámci souboru `search.js`, Zdroj: autor

```
1 function executePythonScript(functionName, inputFilePath, outputFilePath)
2 {
3   return new Promise((resolve, reject) => {
4     const pythonProcess = spawn(process.env.PYTHON_PATH || "python3", [
5       "transform.py",
6       functionName,
7       inputFilePath,
8       outputFilePath,
9     ]);
10
11    pythonProcess.stdout.on("data", (data) => {
12      console.log(`stdout: ${data}`);
13      console.log(process.env.PYTHON_PATH);
14    });
15
16    pythonProcess.stderr.on("data", (data) => {
17      console.error(`stderr: ${data.toString()}`);
18      console.log(process.env.PYTHON_PATH);
19    });
20
21    pythonProcess.on("close", (code) => {
22      if (code === 0) {
23        console.log("Python script completed successfully");
24        resolve(); // Resolve the promise upon successful completion
25      } else {
26        console.error("Python script failed with code " + code);
27        reject(new Error("Python script failed with code " + code));
28      }
29    });
30  });
```

5.4 Implementace datové transformace

Cílem této kapitoly je poskytnout podrobný popis dat, která byla extrahována pro účely této analýzy a ukázat, jak jsem postupoval při implementaci datové transformace. Soubor, který

obsahuje celý kód pro transformaci se jmenuje *transform.py* a je dostupný v repozitáři aplikace na Githubu¹⁵.

5.4.1 Popis extrahovaných dat

Po extrakci dat nám vznikly soubory *GooglePlayOutput.csv* a *AppStoreOutput.csv*. Každý soubor obsahuje velké množství atributů, které poskytují detailní informace o jednotlivých aplikacích. Níže dané soubory představím.

GooglePlayOutput.csv

Data v tomto textovém souboru mají 53 atributů. Některé příklady hodnot atributů jsem musel zkrátit, jelikož obsahovaly tisíce znaků.

Tabulka 1: Popis atributů ze souboru GooglePlayOutput.csv

Název atributu	Popis atributu	Příklad hodnoty
title	Název aplikace	Google Chrome
description	Popis aplikace bez HTML tagů	Google Chrome is a fast, easy to use, and secure web browser...
descriptionHTML	Popis aplikace s HTML tagy	Google Chrome is a fast, easy to use, and secure web browser...
summary	Krátký popis aplikace	Chrome is the fast and secure browser built by Google...
installs	Počet instalací	10,000,000,000+
minInstalls	Minimální počet instalací	10000000000
maxInstalls	Maximální odhadovaný počet instalací	15416916054
score	Průměrné hodnocení	3.919535

¹⁵ [Sxoohttps://github.com/Robology-Github/app-scraper-dp/blob/1ccddf087a54e1352b2236457106917df56a2b5d/transform.py](https://github.com/Robology-Github/app-scraper-dp/blob/1ccddf087a54e1352b2236457106917df56a2b5d/transform.py)

scoreText	Textová reprezentace průměrného hodnocení	3.9
ratings	Počet hodnocení	44887136
reviews	Počet recenzí	[Good Nice app This update of new chrome is bad I can't even download any songs the old version is good...]
histogram	Distribuce hodnocení	{5: 24976585, 4: 6854340, 3: 3926701, 2: 2342692, 1: 7564518}
price	Cena	0.0
originalPrice	Původní cena	0.0
discountEndDate	Datum konce slevy	[]
free	Jestli je aplikace zdarma	True
currency	Měna ceny	USD
priceText	Textová reprezentace ceny	Free
available	Značí, zda je aplikace dostupná	True
offersIAP	Značí, zda aplikace nabízí nákupy v aplikaci	False
IAPRange	Rozsah cen nákupů v aplikaci	[]
androidVersion	Minimální potřebná verze Androidu	8.1
androidVersionText	Textová reprezentace minimální verze Androidu	Varies with device
developer	Vývojář aplikace	Google LLC
developerId	ID vývojáře	5700313618786177705
developerEmail	Email vývojáře	apps-help@google.com
developerWebsite	Webová stránka vývojáře	http://www.google.com/chrome/

developerAddress	Adresa vývojáře	1600 Amphitheatre Parkway, Mountain View 94043
privacyPolicy	Odkaz na zásady ochrany osobních údajů	http://www.google.com/policies/privacy
developerInternalID	Interní ID vývojáře	[]
genre	Žánr aplikace	Communication
genreId	ID žánru	COMMUNICATION
categories	Kategorie, do kterých aplikace patří	[Communication]
icon	Odkaz na ikonu aplikace	https://play-lh.googleusercontent.com/Or72iU...
headerImage	Odkaz na hlavní obrázek aplikace	https://play-lh.googleusercontent.com/fc6C8Z...
screenshots	Odkazy na screenshoty aplikace	[https://play-lh.googleusercontent.com/a-/AC...]
video	Odkaz na video	[]
videoImage	Odkaz na obrázek z videa	[]
previewVideo	Náhledové video	[]
contentRating	Věkové hodnocení	Everyone
contentRatingDescription	Popis věkového hodnocení	[]
adSupported	Značí, zda aplikace podporuje reklamy	True
released	Datum vydání aplikace	Feb 7, 2012
updated	Datum poslední aktualizace	1713213219000
version	Verze aplikace	VARY
recentChanges	Nedávné změny v aplikaci	Thanks for choosing Chrome! This release

		contains the following features...
comments	Recenze uživatelů, stejný jako atribut reviews, obsahuje méně recenzí	["I use Chrome on every device, and I honestly love it. But with the last update on my phone (Samsung Android), it tries to refresh the page every one or to seconds without touching anything. It's really annoying. Had to remove it and use an earlier version of it. Please try to fix it ASAP. It's unusable with this bug in it. Once fixed, it's a 5 Star product. Thanks,..."]
preregister	Značí, zda je dostupná registrace před vydáním aplikace	False
earlyAccessEnabled	Značí, zda je aplikace dostupná pro předběžný přístup	False
isAvailableInPlayPass	Značí, zda je aplikace dostupná v Play Pass	False
appId	Identifikační kód aplikace v Google Play	com.android.chrome
url	URL adresa na Google Play	https://play.google.com/store/apps/details?id=com.android.chrome

AppStoreOutput.csv

Tento dataset obsahuje 33 atributů.

Tabulka 2: Popis atributů ze souboru AppStoreOutput.csv

Název atributu	Popis atributu	Příklad hodnoty
----------------	----------------	-----------------

id	ID aplikace	535886823
appId	Identifikační kód aplikace	com.google.chrome.ios
title	Název aplikace	Google Chrome
url	URL adresa na App Store	https://apps.apple.com/cz/app/google-chrome/id535886823
description	Popis aplikace	Download the new Google Chrome for your iPhone and iPad...
icon	Odkazy na ikony aplikace	[https://play-lh.googleusercontent.com/QRRGW2tMZ4-FNwoXWk6WWiXHaQCGxuwM-92HrBhlA4Wod_AGmjVmQkiHyAqQjW2yByc, ...]
genres	Žánry, do kterých aplikace patří	["Utilities", "Productivity"]
genreIds	ID žánrů	["6002", "6007"]
primaryGenre	Primární žánr	Utilities
primaryGenreId	ID primárního žánru	6002
contentRating	Věkové hodnocení	17+
languages	Podporované jazyky	["FI", "FR", "DE", "EL", "HU", "ID", "IT", "JA", "KO", "MS", "NB", "PL", "PT", "RU", "ZH", "SK", "ES", "SV", "TH", "ZH", "TR", "UK", "VI"]
size	Velikost aplikace (v bytech)	93798400
requiredOsVersion	Minimální verze OS potřebná	12.0
released	Datum vydání aplikace	2012-06-28
updated	Datum poslední aktualizace	2022-08-22

releaseNotes	Poznámky k nové verzi	Thanks for choosing Chrome! This version includes...
version	Verze aplikace	103.0.5060.63
price	Cena aplikace	0.0
currency	Měna	USD
free	Jestli je aplikace zdarma	True
developerId	ID vývojáře	281956209
developer	Název vývojáře	Google LLC
developerUrl	URL stránka vývojáře na App Store	https://apps.apple.com/cz/developer/google-llc/id281956209
developerWebsite	Webová stránka vývojáře	https://www.google.com
score	Průměrné hodnocení zákazníků	3.69103
reviews	Text uživatelských recenzí	[Cursor in search input if the input is already filled on position one? Really? Whose idea was that? If itsy a bug please please fix it. All thing are ok ...]
currentVersionScore	Hodnocení aktuální verze	4.69393
currentVersionReviews	Recenze aktuální verze	10453
screenshots	Odkazy na screenshoty	["https://is1-ssl.mzstatic.com/image/thumb/Pur..."]
ipadScreenshots	Odkazy na screenshoty pro iPad	["https://is1-ssl.mzstatic.com/image/thumb/Pur..."]
appletvScreenshots	Odkazy na screenshoty pro Apple TV	[]

supportedDevices	Podporované zařízení	["iPhone5s-iPhone5s", "iPadAir-iPadAir", ...]
------------------	----------------------	---

5.4.2 Příprava a čištění dat

V procesu přípravy a čištění dat jsem se zaměřil na úpravy, které mi pomohou při pozdějších krocích transformace dat. V rámci přípravy dat jsem pracoval v Jupyter Notebook. Jupyter Notebook je interaktivní vývojové prostředí, které mi umožnilo implementovat a spouštět Python skripty pro jednotlivé transformace. Pracovní sešit i s jeho historií úprav je dostupný veřejně ve formě souboru *workspace_transformation.ipynb*, který je dostupný v rámci repozitáře projektu na Githubu¹⁶. Tato předpříprava se týká obou datasetů.

Konverze datových typů

Přípravu jsem začal konverzí datových typů, aby bylo zajištěno, že jsou data ve správném formátu pro analýzu a zpracování. Níže jsou atributy a konverze, které jsem provedl:

- **Datum a čas** – u atributů *released* a *updated* jsem provedl konverzi na datový typ *datetime* pomocí funkce *pd.to_datetime()*, abych bylo možné provádět výpočty pro stáří aplikace nebo dobu od poslední aktualizace aplikace.
- **Numerické hodnoty** – atribut *score* byl převeden z textového formátu na numerický datový typ pomocí funkce *pd.to_numeric()*, aby bylo možné provádět výpočty spojené s hodnocením.
- **Celočíselné hodnoty** – atribut *free* byl převeden z boolean na celočíselný datový typ *int* pomocí metody *astype(int)*, aby bylo možné hodnoty číst

Příprava dat pro textové analýzy

V atributu *reviews*, který obsahuje textové řetězce uživatelských recenzí, jsem provedl odstranění nealfanumerických znaků a nadbytečných mezer, aby byly připravené pro textové analýzy. Tento krok byl proveden pomocí regulárních výrazů s použitím pythonové knihovny *re*.

Pro analýzu frekvencí slov a bigramů jsem text nejprve očistil od tzv. stopwords. Stopwords jsou slova, která se v textu vyskytují velmi často, ale nesou malou sémantickou hodnotu – jsou to např. předložky a spojky. Postup je následovný – pomocí knihovny *langdetect* se zjistí primární jazyk obsažený v atributu recenzí. Následně se načte soubor se seznamem stopwords pro daný jazyk. Seznamy stopwords jsou získávány z knihovny *nlTK*, která

¹⁶ https://github.com/Robology-Github/app-scrapers/blob/1ccddf087a54e1352b2236457106917df56a2b5d/workspace_transformation.ipynb

podporuje 23 jazyků. Nevýhodou tohoto přístupu je, že pokud pro daný jazyk neexistuje soubor se stopwords, nedojde k správnému předzpracování textu. V takovém případě je alespoň využít anglický seznam, jelikož se anglická slova v recenzích objevují často. Všechny tyto činnosti jsem definoval v rámci funkce *preprocess_and_split_reviews*. Výsledný atribut se nazývá *processed_reviews*.

Normalizace dat

Hodnoty atributů *genres* a *languages* z datasetu pro App Store a atribut *categories* z datasetu Google Play jsou ve formě listů. Např. hodnota atributu *languages*, která označuje, jaké jazyky jsou v rámci aplikace dostupné, mohou vypadat takto: `["Utilities", "Productivity"]`.

V procesu normalizace dat jsem využil metodu *explode* z knihovny *pandas*, která mi umožnila rozdělit seznamové hodnoty na jednotlivé řádky, čímž jsem dosáhl lepší struktury dat pro následnou analýzu. Tyto hodnoty byly společně s atributem *appId*, uloženy do separátních datových souborů. Tímto vznikly následující soubory:

- **AppStore_Genres.csv** - obsahuje data o žánrech jednotlivých mobilních aplikací z platformy App Store.
- **AppStore_Languages.csv** - obsahuje data o podpoře jazyků a státech, jednotlivých aplikací z platformy App Store
- **GooglePlay_Categories.csv** – obsahuje data o kategoriích jednotlivých mobilních aplikací z platformy Google Play

Analogicky postupuji při normalizaci atributů *bigrams* a *word_freq*, které vzniknou v pozdější fázi transformace.

Mapování hodnot

Pro určení trhů, na které jsou aplikace cíleny, jsem prováděl mapování jazykových kódů z atributu *languages* na seznam zemí. Proto jsem si vytvořil slovník *language_to_countries*, který obsahuje jednotlivé jazykové kódy, které jsem získal z App Store a přiřadil jsem je k zemím podle veřejného datasetu¹⁷, který obsahuje informace o státech a jejich jazycích.

Výpis 4: Ukázka kódu pro tvorbu slovníku pro mapování jazykových kódů ke státům, Zdroj: autor

```
1 language_to_countries = {
2     'AF': ['Afghanistan'],
3     'AM': ['Armenia'],
4     'AN': ['Netherlands Antilles'],
5     # další záznamy...
6 }
```

¹⁷ <https://developers.expediagroup.com/docs/products/rapid/lodging/content/country-to-destination-language-mapping>

Tyto informace o zemích jsou uloženy spolu s atributem `language` do souboru `AppStore_Languages.csv`.

5.4.3 Tvorba odvozených atributů

Odvozené atributy poskytují nové perspektivy a hlubší porozumění zkoumaným jevům. Tento proces nejenže zvyšuje hodnotu a využitelnost původních dat, ale také umožňuje lépe interpretovat vzorce a trendy relevantní pro trh mobilních aplikací.

Odvozené atributy pro data z Google Play

- `days_since_last_update` – počet dní od poslední aktualizace aplikace, který vypovídá o údržbě aplikace.
- `app_age` – počet dní od doby, kdy byla aplikace vydána.
- `1*`, `2*`, `3*`, `4*`, `5*` – z atributu `histogram`, který obsahuje rozpad počtu hodnocení na jednotlivé skóre, jsem vyextrahoval tyto hodnoty a vytvořil z nich separátní atributy.
- `rating_ratio` – poměr pozitivních hodnocení (4 a 5 hvězd) k negativním (1 a 2 hvězdy).
- `engagement_score` – skóre zapojenosti uživatelů, kalkulované jako součin hodnocení a počtu hodnocení dělený minimálním počtem instalací.
- `install_to_rating` – poměr mezi počtem instalací a hodnocením. Poměr pomáhá posoudit, jak aktivně uživatelé hodnotí aplikaci ve vztahu k jejímu počtu instalací.

Výpis 5: Ukázka kódu tvorby odvozených atributů v rámci souboru `transform.py`, Zdroj: autor

```
#Calculations
df['app_age'] = (df['updated'] - df['released']).dt.days
df['rating_ratio'] = (df['4*'] + df['5*']) / (df['1*'] + df['2*'])
df['engagement_score'] = (df['score'] * df['ratings']) / df['minInstalls']
df['install_to_rating'] = df['minInstalls'] / (df['ratings'] + 1e-10)
df['days_since_last_update'] = (datetime.now(timezone.utc) -
df['updated']).dt.days
```

Odvozené atributy pro data z App Store

- `supports_iPhone`, `supports_iPad`, `supports_Mac` – jsou indikátory, které označují podporu aplikace pro různá zařízení Apple. Tyto atributy jsem získal z pole `supportedDevices`, kde jsem seskupil hodnoty pro jednotlivé modely Apple zařízení.
- `days_since_last_update` – počet dní od poslední aktualizace aplikace, který vypovídá o údržbě aplikace.
- `app_age` – počet dní od doby, kdy byla aplikace vydána.
- `language_name` – celý název jazyka. Pomocí knihovny `iso639` jsem konvertoval kódy jazyků z atributu `language` na plný název pro snadnější interpretaci.
- `IAPMin`, `IAPMax` – značí nejnižší a nejvyšší částku pro nákupy v dané aplikaci. Jsou extrahovány z původního atributu `IAPRange`.

- *CurrencySymbolMin*, *CurrencySymbolMax* – značí měnu, ve které se provádí nákupy v aplikaci, jsou extrahovány z atributu *IAPRange* .

5.4.4 Analýza sentimentu

Analýza sentimentu recenzí z obchodů Google Play a App Store představuje klíčový prvek mé diplomové práce, zaměřující se na porozumění uživatelských názorů spojených s mobilními aplikacemi. Tento proces vyžaduje sofistikované nástroje a metody, aby bylo možné efektivně zpracovat a klasifikovat velký objem textových dat. Níže krátce představím jazykový model, který jsem pro analýzu sentimentu použil.

Pro analýzu sentimentu jsem zvolil multijazyčný model BERT (Bidirectional Encoder Representations from Transformers), který je schopen rozpoznat sentiment v textech napsaných v různých jazycích. BERT je pokročilý model zpracování přirozeného jazyka vyvinutý společností Google v roce 2018. Je trénován na datech z Wikipedie. [41] Historicky mohly jazykové modely číst vstupní text pouze sekvenčně – buď zleva doprava nebo zprava doleva – ale nikoli oběma směry současně. BERT je odlišný, protože je navržen tak, aby četl oběma směry najednou. Těmto typům jazykových modelů se říká „bidirectional transformers“. BERT tímto položil základy pro mnoho dnešních modelů, které se v současnosti využívají, např. lehčí verze BERTu byly implementovány v modelu GPT-2 od společnosti OpenAI. [42]

Předtím, než se analýza uskuteční, je provedeno předzpracování atributu *reviews*. Tento proces je již popsán v dřívější podkapitole. Po předzpracování textu je každý upravený text recenze poskytnut modelu BERT, který vrátí sentiment jako jednu z pěti kategorií (1 hvězda až 5 hvězd). Je také nutné zmínit, že v rámci extrakce dat byly specifikovány kritéria, která zajišťují, že bude extrahováno 200 posledních recenzí, abychom mohli zjistit co nejaktuálnější sentiment.

Analýza sentimentu je provedena pomocí metody *sentiment-analysis*, která je součástí knihovny *transformers*. Metoda automatizovaně zpracovává text, provádí tokenizaci (rozklad textu na jednotlivé složky, tzv. tokeny), a následně automaticky klasifikuje sentiment textu.

Výstupy jsem následně rozkategorizoval na pět předdefinovaných kategorií sentimentu: *Negative*, *Slightly negative*, *Neutral*, *Slightly positive* a *Positive* a uložil do atributu *sentiment_category*, viz Výpis 5.

Výpis 5: Ukázka kódu analýzy sentimentu v rámci souboru *transform.py*, Zdroj: autor

```
1 # Load the sentiment analysis pipeline with the multilingual BERT model
2 sentiment_analyzer = pipeline("sentiment-analysis", model="nlptown/bert-
  base-multilingual-uncased-sentiment")
3 tokenizer = AutoTokenizer.from_pretrained("nlptown/bert-base-multilingual-
  uncased-sentiment")

def compute_sentiment_category_mbert(text):
4     try:
5         # Directly pass the text to the pipeline
6         # The pipeline handles tokenization and truncation internally
7         result = sentiment_analyzer(text, truncation=True,
max_length=512)[0]
8         label = result['label']
9
10        # Mapping the model output to custom categories
11        if label == '1 star':
12            return 'Negative'
13        elif label == '2 stars':
14            return 'Slightly negative'
15        elif label == '3 stars':
16            return 'Neutral'
17        elif label == '4 stars':
18            return 'Slightly positive'
19        else: # '5 stars'
20            return 'Positive'
21    except Exception as e:
22        print(f"Error processing text: {e}")
23        return 'Missing' # Default to 'Missing' in case of an error
24
25 df['sentiment_category']=
  df['processed_reviews'].apply(compute_sentiment_category_mbert)
26
27
```

5.4.5 Textová analýza

Tato analýza má za úkol extrahovat nejčastější slova a slovní spojení z textových řetězců v atributu *processed_reviews*. To poslouží k detailnější analýze recenzí, jelikož můžeme zjistit důležité informace o tom, jaké mají uživatelé problémy, nebo naopak co se jim na aplikaci líbí.

Začal jsem extrakcí bigramů, což jsou páry sousedních slov, které mohou odhalit zajímavé vzory v textu. Funkce *get_and_flatten_bigrams* bere jako vstup zpracovaný text a generuje seznam bigramů s jejich frekvencemi. Využívá k tomu n-gram model z knihovny *TextBlob*.

Pro analýzu frekvence slov jsem vytvořil funkci `flatten_word_frequencies`, která rozdělí text na jednotlivá slova a pak spočítá jejich četnost pomocí objektu `Counter` z pythonového modulu `collections`.

Vzniklé atributy jsou poté normalizovány a převedeny do tabulkové formy pomocí `explode` stejným způsobem, jako to bylo provedeno pro atributy s žánry a jazyky. Ty jsou následně uloženy do těchto samostatných souborů:

- **AppStore_Bigrams.csv** - obsahuje nalezené bigramy a jejich frekvenci pro jednotlivé aplikace z platformy App Store.
- **AppStore_Word_Frequencies.csv** - obsahuje data o frekvenci slov získaných z uživatelských recenzí mobilních aplikací z platformy App Store.
- **GooglePlay_Bigrams.csv** - obsahuje nalezené bigramy a jejich frekvenci pro jednotlivé aplikace z platformy Google Play.
- **GooglePlay_Word_Frequencies.csv** - obsahuje data o frekvenci slov získaných z recenzí mobilních aplikací z platformy Google Play.

5.4.6 Kategorizace dat

Kategorizace dat výrazně usnadní uživatelům tohoto nástroje interpretaci dat. Níže popíšu, jak jsem kategorizoval jednotlivé hodnoty vybraných atributů.

Kategorizace ceny

Cena aplikace je jedním z hlavních faktorů, který ovlivňuje rozhodování uživatelů o instalaci. Ceny jsem kategorizoval na základě percentilů namísto pevně stanovených hodnot. Pevně stanovené hodnoty by nereflektovali odlišné měny a nevypovídalo by to o tom, zda je aplikace skutečně levná či drahá. Vytvořil jsem atribut `price_category` s těmito hodnotami:

- **Free** – aplikace, které jsou nabízeny zdarma. Tato kategorie zahrnuje všechny aplikace s cenou 0.
- **Low** – aplikace s cenou nižší než 25. percentil.
- **Medium** – aplikace s cenou mezi 25. a 50. percentilem.
- **High** – aplikace s cenou mezi 50. a 75. percentilem.
- **Very High** – aplikace s cenou vyšší než 75. percentil.

Kategorizace věku aplikace

Věk aplikace jsem kategorizoval na základě doby od jejího prvního vydání, což poskytuje informace o tom, v jakém stádiu produktového cyklu se aplikace nachází. Vytvořil jsem atribut `app_age`, s těmito hodnotami:

- **Brand New** – aplikace vydány do 30 dnů.
- **Recently Launched** – aplikace vydány do 90 dnů.
- **Established** – aplikace starší než 90 dní, ale mladší než jeden rok.
- **Mature** – aplikace, které jsou na trhu 1 až 3 roky.

- **Very Mature** – aplikace na trhu déle než 3 roky.

Kategorizace hodnocení sentimentu

V rámci kategorizace pro hodnoty z analýzy sentimentu je vytvořen atribut *sentiment_category* s těmito hodnotami:

- **Negative** – text v atributu *reviews* je klasifikačním modelem hodnocen jako "1 star".
- **Slightly negative** – text v atributu *reviews* je klasifikačním modelem hodnocen jako 2 stars".
- **Neutral** – text v atributu *reviews* je klasifikačním modelem hodnocen "3 stars".
- **Slightly positive** – text v atributu *reviews* je klasifikačním modelem hodnocen "4 stars".
- **Positive** – text v atributu *reviews* je klasifikačním modelem hodnocen "5 stars".

Kategorizace frekvence aktualizací

Frekvence aktualizací aplikace může ukázat, zdali probíhá aktivní vývoj a údržba mobilní aplikace. V nově vytvořeném atributu *update_frequency* jsou následující hodnoty:

- **Very Recent Updates** – aplikace aktualizované do 30 dnů.
- **Recently Updated** – aplikace aktualizované do 90 dnů.
- **Moderately Updated** – aplikace aktualizované do 180 dnů.
- **Rarely Updated** – aplikace aktualizované do 365 dnů.
- **Stále** – aplikace bez aktualizací déle než rok.

Kategorizace velikosti souboru aplikace

Hodnoty pro velikosti souboru aplikací jsou v atributu *size* jsou uvedeny v bytech, což není přehledné. Vytvořil jsem atribut *file_size_category*, kde je velikost souboru aplikace kategorizována následovně:

- **Small** – aplikace s velikostí souboru do 50 MB.
- **Medium** – aplikace s velikostí souboru mezi 50 MB a 200 MB.
- **Large** – aplikace s velikostí souboru nad 200 MB.

5.4.7 Odebrání nepotřebných atributů

Některé atributy již nejsou nadále potřebné pro vizualizaci dat, a proto jsem je odebral před vytvoření transformovaného datového souboru. Odebral jsem následující atributy:

- **App Store** – *id*, *icon*, *genreIds*, *primaryGenreId*, *requiredOsVersion*, *version*, *developerid*, *developerUrl*, *screenshots*, *ipadScreenshots*, *appletvScreenshots*,

languages, genres, supportedDevices, currency, developerId, reviews, processed_reviews, bigrams, word_freq

- **Google Play** – *descriptionHTML, summary, installs, maxInstalls, scoreText, reviews, histogram, currency, androidVersion, androidVersionText, androidMaxVersion, previewVideo, developerId, developerEmail, developerAddress, privacyPolicy, developerInternalID, genreId, icon, headerImage, screenshots, video, videoImage, contentRatingDescription, version, recentChanges, comments, originalPrice, discountEndDate, categories, priceText, 1*, 2*, 3*, 4*, 5*, processed_reviews, bigrams_df, word_freq_df*

5.4.8 Výsledné datové soubory

Po transformaci vznikají následující soubory:

App Store

- **AppStoreOutput_cleaned.csv** – hlavní soubor s transformovanými daty z platformy App Store.
- **AppStore_Genres.csv** – soubor obsahující normalizované atributy s žánry aplikací z platformy App Store.
- **AppStore_Languages.csv** – soubor obsahující normalizované atributy s jazyky a státy z platformy App Store.
- **AppStore_Bigrams.csv** – soubor obsahující extrahované bigramy z recenzí a jejich frekvenci z platformy App Store.
- **AppStore_Word_Frequencies.csv** – soubor obsahující extrahovaná slova a jejich frekvenci z platformy App Store.

Google Play

- **GooglePlayOutput_cleaned.csv** – hlavní soubor s transformovanými daty z platformy Google Play.
- **GooglePlay_Categories.csv** – soubor obsahující normalizované atributy s kategoriemi aplikací z platformy Google Play.
- **GooglePlay_Bigrams.csv** – soubor obsahující extrahované bigramy z recenzí a jejich frekvenci z platformy Google Play.
- **GooglePlay_Word_Frequencies.csv** – soubor obsahující extrahovaná slova a jejich frekvenci z platformy Google Play.

5.5 Integrace datového úložiště

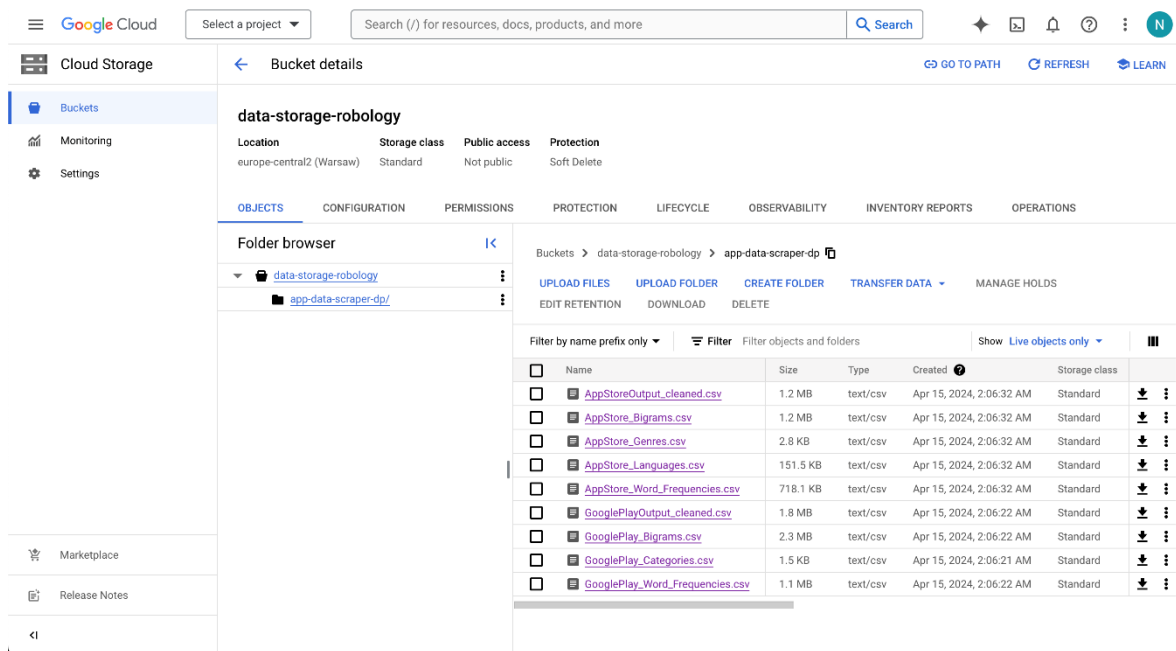
Při integraci Google Cloud Storage (GCS) do mého projektu na analýzu trhu s mobilními aplikacemi jsem prošel několika klíčovými kroky, aby bylo možné bezpečně a efektivně ukládat a spravovat data.

Prvním krokem bylo vytvoření nového projektu v Google Cloud Console, což je služba, kde se spravují všechny cloudové zdroje, které nabízí společnost Google. Projekt jsem pojmenoval *app-scraper-dp*. Po vytvoření projektu jsem musel aktivovat službu Google Cloud Storage, což zahrnovalo přidání platebních údajů. Jakmile byla služba aktivována, vytvořil jsem bucket s názvem *data-storage-robology*, což je základní kontejner pro uchovávání datových objektů. Při vytváření bucketu jsem zvolil název, který musí být globálně jedinečný, a specifikoval jsem region *europa-central2* (Warsaw). Nastavení regionu bucketu je velmi důležité, jelikož ovlivňuje rychlost a cenu za přenos dat.

K tomu, aby byla komunikace mezi GCS a webovou aplikací bezpečná, je nutné, aby se používali autentizační údaje. Pro tento účel jsem vytvořil služební účet, který aplikaci umožňuje komunikovat s GCS. Pro služební účet jsem nastavil příslušné role, které určují jeho oprávnění. Pro tento účet jsem poté vygeneroval privátní klíče ve formátu JSON, které slouží k autentizaci při interakci aplikace s GCS.

Informace z JSON klíče jsem uložil do environmentálních proměnných aplikace. Tento přístup zajišťuje, že citlivé informace nejsou pevně zakódované v zdrojovém kódu a jsou bezpečně uloženy. Pokud by se autentizační údaje kódovali natvrdo, mohlo by se stát, že by je mohl někdo využívat k přístupu na úložiště, popřípadě by mohl využívat služby v rámci projektu Google Cloudu za naše peníze. V kódu aplikace jsem použil knihovnu Google Cloud Storage poskytovanou Googlem pro interakci s úložištěm. Při inicializaci klienta GCS v aplikaci se načtou autentizační údaje z environmentálních proměnných.

Pro nahrávání dat do GCS jsem vytvořil funkci *uploadFileToGCS*, která asynchronně nahrává soubory s transformovanými daty do cloudového úložiště v rámci mého definovaného bucketu. Přitom se autentizuje údaji služebního účtu, které si načítají z environmentálních proměnných aplikace. Funkci se musí poskytnout název souboru, název bucketu a cílová cesta do bucketu.



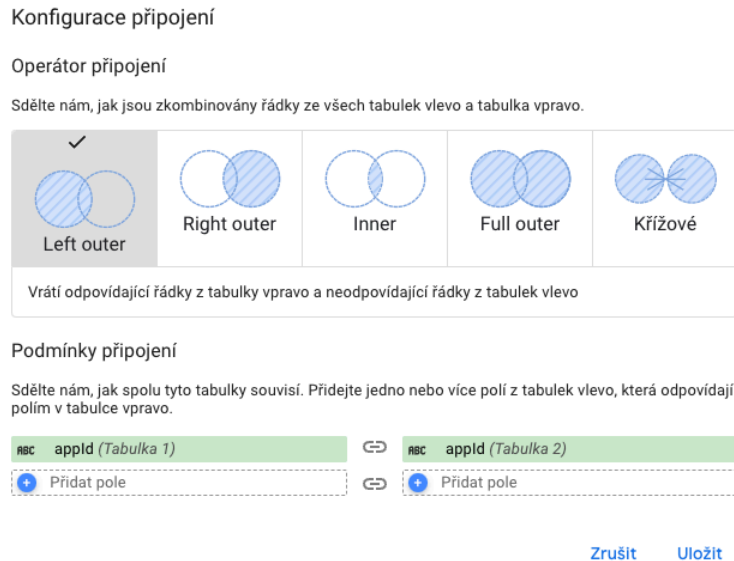
Obrázek 11: Snímek obrazovky prostředí GCS po nahrání datových souborů, Zdroj: autor

5.6 Vizualizace dat

V této kapitole se zaměřím na to, jak jsem propojil data z datového úložiště Google Cloud Storage s analytickou platformou Google Looker Studio. Dále popíšu vytvořený dashboard, který bude sloužit jako zpravodajský produkt v rámci zpravodajského cyklu CI.

5.6.1 Napojení na datový zdroj a propojení datových souborů

Při použití Google Looker Studio je nezbytné definovat zdroj dat, ze kterého budou vizualizace čerpat. Data jsou připojena pomocí předdefinovaného konektoru pro Google Cloud Storage, kde specifikuji cestu k souborům umístěným v příslušném bucketu. Google Looker Studio následně automaticky načte strukturu dat a umožní začít s vytvářením vizualizací. Vzhledem k tomu, že pracuji s více datovými soubory, je nutné je propojit. Toto propojení realizuji prostřednictvím uživatelského rozhraní, kde mohu definovat, na základě jakých atributů mohu data propojit. K dispozici jsou různé typy spojení, jako je left outer, right outer, inner nebo full outer. Dále je možné specifikovat, která data z každé tabulky budou do finálního datasetu zahrnuta. Všechny datové soubory mimo hlavní dataset, které obsahují doplňující data jako jsou kategorie, jazyky, textové analýzy jsem propojil pomocí atributu *appId* s hlavním transformovaným datovým souborem pomocí left outer join. Jakmile je tento proces dokončen, mohu začít s vytvářením vizualizací.



Obrázek 12: Konfigurace propojení datových souborů v rámci Google Looker Studio,
Zdroj: autor

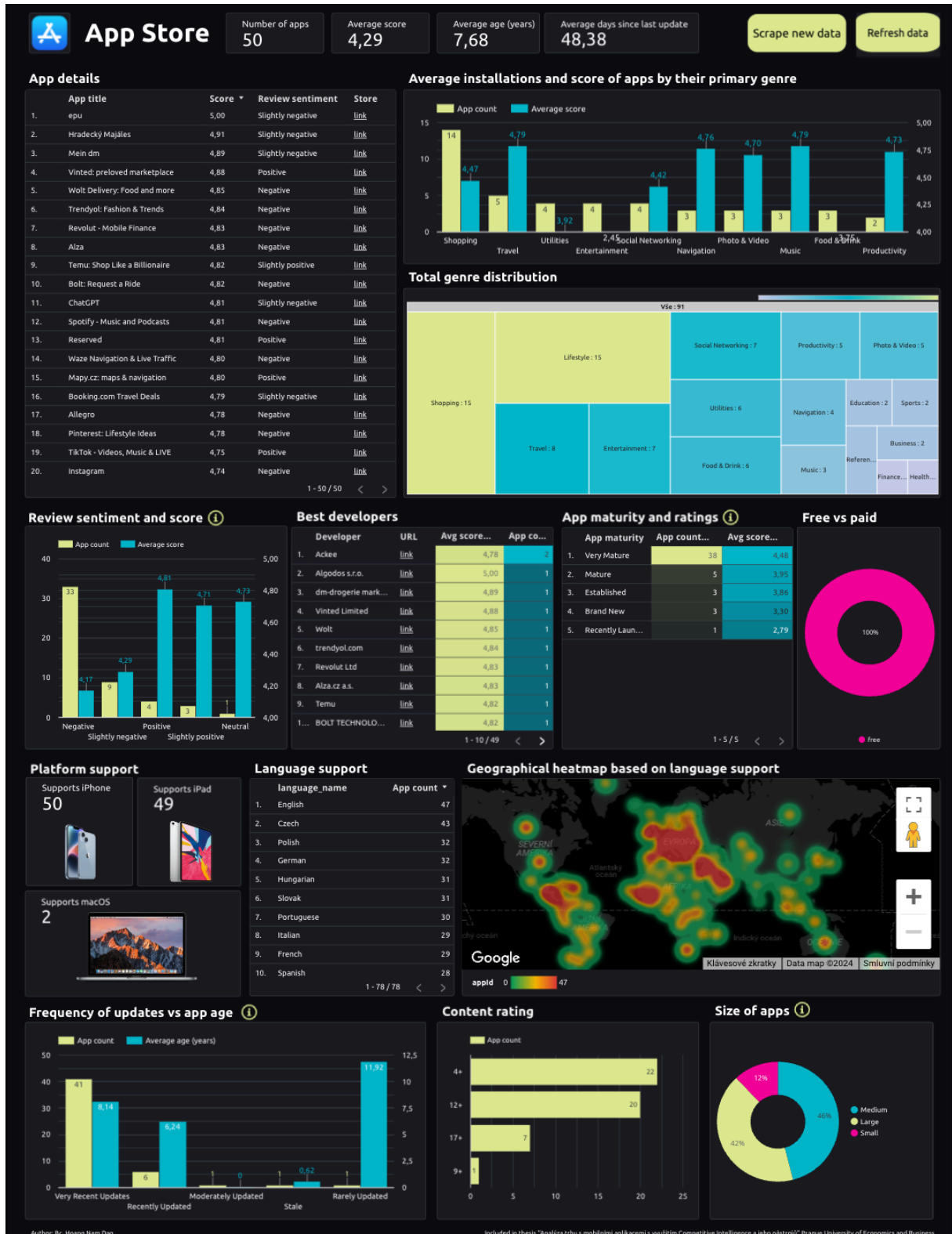
5.6.2 Design dashboardu

Celkově byly v dashboardu vytvořeny 4 stránky. Níže popíšu jednotlivé stránky a jejich vizualizace.

App Store

- **App details** – tabulka se seznamem aplikací obsahující informace jako název aplikace, hodnocení, sentiment recenzí a odkaz na obchod.
- **Average installations and score of apps by their primary genre** – sloupcový graf zobrazující průměrný počet instalací a hodnocení aplikací podle hlavního žánru.
- **Total genre distribution** – treemap zobrazující distribuci žánrů aplikací, kde větší bloky představují žánry s větším počtem aplikací.
- **Review sentiment and score** – kombinovaný sloupcový a pruhový graf zobrazující hodnocení a sentiment recenzí aplikací.
- **Best developers** – tabulka nejlepších vývojářů s informacemi jako URL, průměrné hodnocení a počet aplikací.
- **App maturity and ratings** – bodový graf s rozložením aplikací podle zralosti obsahu a průměrného hodnocení.
- **Free vs paid** – kruhový graf ukazující podíl bezplatných a placených aplikací.
- **Platform support** – počet aplikací, které podporují iPhone, iPad a macOS.
- **Language support** – seznam podporovaných jazyků s počtem aplikací, které daný jazyk podporují.
- **Geographical heatmap based on language support** – heatmapa zobrazující geografickou podporu různých jazyků.
- **Frequency of updates vs app age** – sloupcový graf porovnávající frekvenci aktualizací s průměrným stářím aplikací.

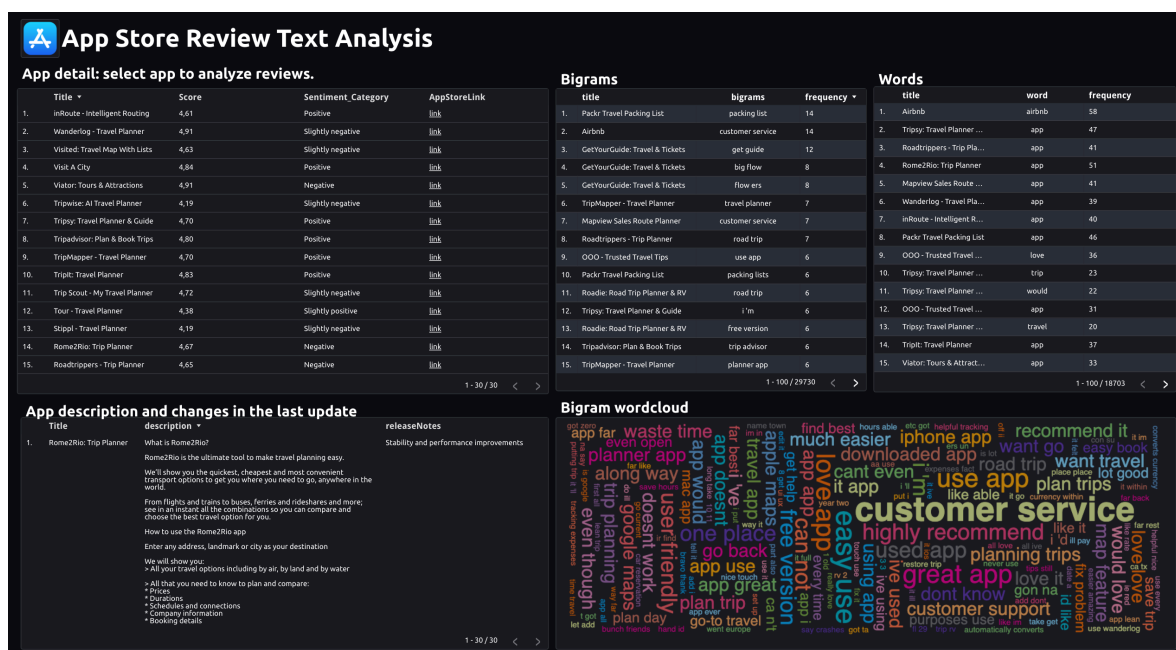
- **Content rating** – horizontální sloupcový graf zobrazující počet aplikací podle věkového hodnocení obsahu.
- **Size of apps** – výšečový graf ukazující rozložení velikostí aplikací.



Obrázek 13: Ukázka stránky dashboardu s informacemi o mobilních aplikacích na platformě App Store, Zdroj: autor

App Store Review Text Analysis

- **App details** – tabulka, která obsahuje seznam aplikací s informacemi jako název aplikace, skóre, sentiment recenzí, a odkaz na App Store.
- **Bigrams** – tabulka zobrazující bigramy (dvojslovné fráze) s jejich frekvencí výskytu v recenzích.
- **Words** – tabulka zobrazující získaná slova a jejich frekvenci v recenzích.
- **App description and changes in the last update** – seznam aplikací s jejich popisem a textem od vývojářů s informacemi o poslední aktualizaci aplikace.
- **Bigram wordcloud** – vizualizace nejčastěji vyskytujících se bigramů, přičemž fráze, které mají větší zastoupení mají větší velikost



Obrázek 14: Ukázka stránky dashboardu sloužící k textové analýze recenzí mobilních aplikací na platformě App Store, Zdroj: autor

Google Play

- **App details** – tabulka s názvy aplikací, počtem instalací, hodnocením, sentimentem recenzí, odkazy na obchod a vývojářem.
- **Average installations and score of apps by their primary genre** – sloupcový graf ukazující průměrný počet instalací a hodnocení aplikací rozdělených podle jejich primárního žánru.
- **Total genre distribution** – treemap graf znázorňující celkové rozdělení žánrů aplikací.
- **Best developers** – tabulka nejlepších vývojářů s odkazy, počtem aplikací, průměrným hodnocením a průměrným počtem instalací.
- **Review sentiment and score** – kombinovaný sloupcový a pruhový graf, který zobrazuje sentiment recenzí a průměrné skóre aplikací.

- **App maturity and ratings** – bodový graf zobrazující zralost aplikací a jejich hodnocení.
- **Rating ratio** – pruhový graf zobrazující poměr hodnocení.
- **Engagement score** – sloupcový graf znázorňující rozložení kategorií pro metriku *engagement_score*.
- **Install to rating ratio** – sloupcový graf zobrazující rozložení kategorií pro metriku *install_to_rating*.
- **Ad support** – výsečový graf ukazující podíl aplikací s podporou reklam.
- **Free vs paid** – výsečový graf znázorňující poměr bezplatných a placených aplikací.
- **Offers in-app purchases** – výsečový graf zobrazující podíl aplikací nabízejících nákupy v aplikaci.
- **In-app purchases pricing overview** – tabulka s přehledem cenových rozpětí nákupů v aplikaci.
- **Frequency of updates vs app age** – sloupcový graf zobrazující frekvenci aktualizací ve vztahu k průměrnému věku aplikace.
- **Frequency of update by genre** – sloupcový graf ukazující frekvenci aktualizací aplikací podle žánru.
- **Content rating** – sloupcový graf zobrazující počet aplikací podle věkového hodnocení (např. PEGI 13).

Google Play Review Text Analysis

- **App details** – tabulka, která obsahuje seznam aplikací s informacemi jako název aplikace, skóre, sentiment recenzí, a odkaz na Google Play.
- **Bigrams** – tabulka zobrazující bigramy (dvojslovné fráze) s jejich frekvencí výskytu v recenzích.
- **Words** – tabulka zobrazující získaná slova a jejich frekvenci v recenzích.
- **App description and changes in the last update** – seznam aplikací s jejich popisem a textem od vývojářů s informacemi o poslední aktualizaci aplikace.
- **Bigram wordcloud** – vizualizace nejčastěji vyskytujících se bigramů, přičemž fráze, které mají větší zastoupení mají větší velikost

5.7 Nasazení aplikace na webový server

Nasazení webové aplikace je proces, jehož cílem je zpřístupnit webovou aplikaci koncovým uživatelům na internetu. Pro orchestraci nasazení webové aplikace jsem zvolil technologii službu Digital Ocean App Platform pro hostování a technologii Docker pro implementaci nasazení.

Pro nasazení a správu aplikace jsem zvolil Digital Ocean App Platform¹⁸, což mi poskytlo přístup k automatizovaným procesům pro nasazení, škálování a monitorování aplikace. Tato platforma nabízí implementaci Continuous Integration/Continuous Delivery (CI/CD) pipeline, což mi umožňuje automatické nasazení aplikace do produkčního prostředí pokaždé, když promítnu změny v kódu do repozitáře na Githubu. Díky integraci s cloudovými službami a automatickému škálování se aplikace dokáže přizpůsobit zvýšenému zatížení bez nutnosti manuálního zásahu.

The screenshot shows the DigitalOcean App Platform interface for an application named 'robology-app-analytics'. At the top, there is a 'Back to Apps' link, the application name, and 'Create' and 'Actions' buttons. Below this is a navigation bar with tabs for 'Overview', 'Insights', 'Activity', 'Runtime Logs', 'Console', and 'Settings'. The 'Overview' tab is selected, showing a 'Learn' link. The main content area is divided into several sections: 'App Health Check' (Available), 'INSIGHTS' (Compute table for 'app-scrapers-dp' showing 3% RAM and 2% CPU usage), 'HTTP Routes', 'Estimated Monthly App Cost' (\$25.00), 'App Resources' (Docs, API Docs, Tutorials), 'Recent Activity' (listing deployment events with triggers and timestamps), and 'Troubleshooting Topics' (links to error code reference and scaling guides).

← Back to Apps

robology-app-analytics Create Actions

in [app-scrapers-dp](#) / [robology-app-analytics-dy2n9.ondigitalocean.app](#)

Overview Insights Activity Runtime Logs Console Settings

Overview Learn

App Health Check

✓ Available

INSIGHTS

Compute Last 5m

Component	RAM	CPU
app-scrapers-dp	3%	2%

Insights >

HTTP Routes

Estimated Monthly App Cost

\$25.00

App Resources

- Docs
- API Docs
- Tutorials

Recent Activity

Apr 15 2024

LIVE robogentlenam's deployment went live

- Trigger: robogentlenam pushed [1ccddf0](#) to [Robology-Github/app-scrapers-dp/main](#)

01:56:05 AM • 12m 51s build

! nam.dao@robology.cz's deployment failed during build phase

- Trigger: nam.dao@robology.cz configured app-scrapers-dp settings

01:39:29 AM • 6m 20s build

! nam.dao@robology.cz's deployment failed during build phase

- Trigger: nam.dao@robology.cz created a new app

01:30:12 AM • 6m 48s build

Activity >

Troubleshooting Topics

- App Platform Error Code Reference
- How to Troubleshoot Apps in App Platform
- How to Scale Apps in App Platform

Obrázek 15: Snímek obrazovky administračního rozhraní DigitalOcean App platform, Zdroj: autor

¹⁸ <https://www.digitalocean.com/products/app-platform>

Docker jsem použil pro kontejnerizaci celé aplikace, což zajistilo její konzistenci a izolaci běhového prostředí. Tato technologie mi umožnila balit všechny potřebné závislosti (např. potřebné knihovny, specifické verze technologií) přímo s aplikací, což značně zjednodušuje nasazení na různých platformách a eliminuje tak problémy s nekompatibilitou. Docker také podporuje rychlé nasazování a snadné vrácení k předchozím verzím aplikace, což je zásadní pro efektivní údržbu produkčního prostředí.

Pro webovou aplikaci jsem vytvořil dockerfile. Ten je také k nalezení na v repozitáři projektu na Githubu¹⁹. Dockerfile je textový soubor, který obsahuje sadu příkazů a argumentů používaných pro vytvoření Docker image, což je v podstatě šablona obsahující operační systém, aplikaci a všechny její závislosti. V rámci tohoto souboru je definováno, co bude v image obsaženo a jak má být aplikace konfigurována a spuštěna. Platforma Digital Ocean z tohoto souboru získává instrukce pro to, jak má být prostředí webové aplikace připraveno.

Výpis 6: Výpis části kódu s instrukcemi pro přípravu prostředí ze souboru *Dockerfile*, Zdroj: autor

```
# Use an official Node.js runtime as a parent image
FROM node:21.7.1

# Set the working directory in the container to /app
WORKDIR /app

# Copy the current directory contents into the container at /app
COPY . .

# Install dependencies required to compile OpenSSL
RUN apt-get update && \
    apt-get install -y build-essential checkinstall zlib1g-dev wget

# Install Python dependencies from requirements.txt
# Ensure you have a requirements.txt file in your project directory
COPY requirements.txt /app/
RUN pip3 install --no-cache-dir -r requirements.txt

# Run TextBlob's download script
RUN python -m textblob.download_corpora lite
RUN pip3 install torch torchvision torchaudio --index-url
https://download.pytorch.org/whl/cpu

# Install any needed packages specified in package.json for Node.js
RUN npm install

# Run your application
CMD ["npm", "start"]
```

¹⁹ <https://github.com/Robology-Github/app-scraper-dp/blob/1ccddf087a54e1352b2236457106917df56a2b5d/Dockerfile>

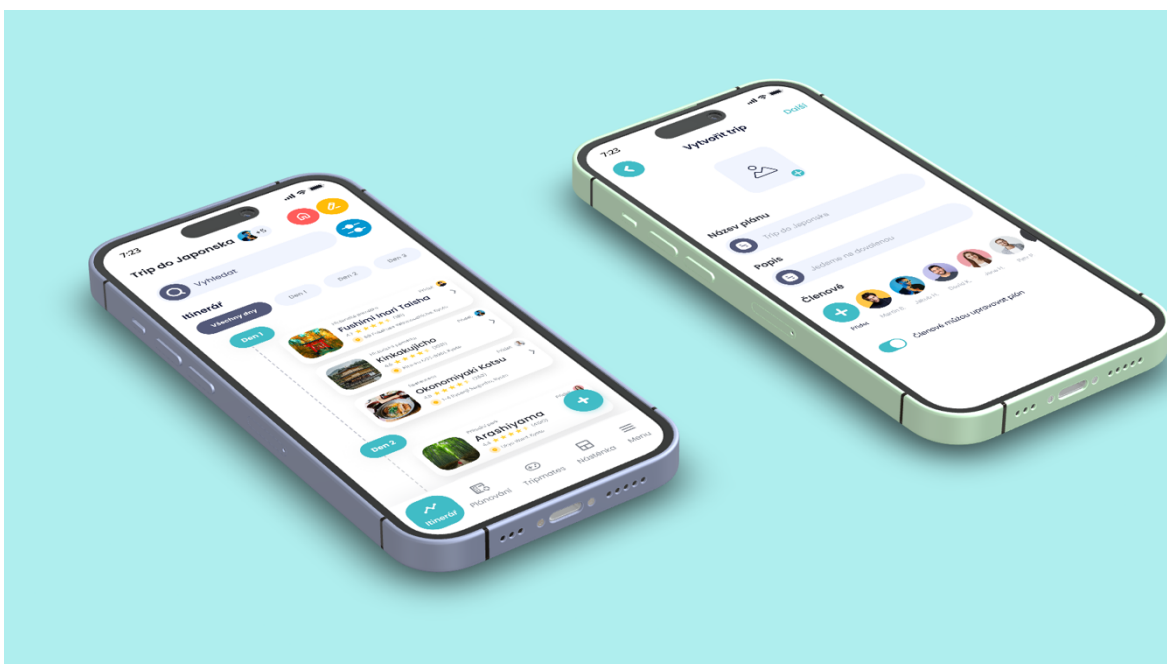
6 Analýza trhu s mobilními aplikacemi

V této kapitole se budu zabývat analýzou trhu s mobilními aplikacemi. K tomu využiji vytvořený nástroj a principy CI. Představím také kontext pro analýzu, díky kterému jsem si definoval klíčové zpravodajské otázky.

6.1 Představení kontextu analýzy

Nejnovějším projektem, kterým se ve firmě zabýváme, je koncept mobilní aplikace pro kolaborativní plánování dovolené. Aplikace by umožnila kolaboraci mezi uživateli prostřednictvím intuitivních funkcí, jako je real-time úprava itineráře, hlasování o aktivitách a vyhledávání míst na základě zadaných preferencí uživatelů. Významným prvkem je také integrace umělé inteligence, která by uživatelům umožnila kompletně vygenerovat cestovní itinerář podle jejich kritérií.

Projekt je v současné době ve fázi výzkumu a ověřování, přičemž jsou již navrženy základní funkce a obrazovky.



Obrázek 16: Grafický návrh mobilní aplikace pro kolaborativní plánování dovolené, Zdroj: autor

Cílem této analýzy je prozkoumat, jak bychom mohli tento koncept proměnit v úspěšnou mobilní aplikaci. Důkladná analýza tržního prostředí segmentu cestovních aplikací a výzkum tržních trendů v rámci tohoto segmentu nám poskytnou nezbytný základ pro

určení důležitých funkcionalit a pomůžou nám definovat strategie, které by firmě umožnily nejen aplikaci úspěšně uvést na trh, ale také zajistí její dlouhodobou udržitelnost a růst.

V rámci této práce jsem si v dřívějších kapitolách stanovil několik zpravodajských témat, které by mohly být zásadní pro pochopení dynamiky a trendů definovaného tržního prostředí. Pro každé z těchto témat jsem formuloval řadu klíčových zpravodajských otázek, jejichž zodpovězení by mohlo firmě pomoci s prozkoumáním trhu a formulací strategie pro získání konkurenční výhody.

Pro hledání odpovědí na tyto otázky využiji nástroj, který byl v rámci této práce vytvořen. Pro získání doplňujících informací, použiji také demo verzi platformy AppMagic²⁰, která obsahuje detailnější data o jednotlivých aplikacích, jako např. odhady tržeb, segmentace uživatelů, data pro jednotlivé trhy, rozpad stažení dle zemí a mnoho dalšího.

6.2 Plánování a řízení

V rámci fáze plánování a řízení budu pro klíčová zpravodajská témata, která byla zmíněna na začátku práce, definovat klíčové zpravodajské otázky, které by firmě pomohly poskytnout vhled do externího prostředí tohoto konkrétního segmentu s mobilními aplikacemi.

Identifikace klíčových hráčů v segmentu

- Jaké aplikace můžeme na tomto segmentu považovat za klíčové hráče?
- Jsou mezi klíčovými hráči aplikace, které vznikly v posledním roce?
- V jakých zemích sídlí vývojáři daných aplikací?

Sledování tržních trendů a průzkum funkcionalit

- Jaké funkce se nejčastěji objevují v aplikacích pro plánování cest?
- Na jaké funkce se vývojáři soustředí v posledních aktualizacích aplikací?
- Jak je v těchto mobilních aplikacích využívána umělá inteligence?

Identifikace monetizačních strategií v rámci segmentu

- Jaké monetizační modely jsou použité v aplikacích v rámci tohoto segmentu?
- Jaká je cenová hladina pro in-app purchases v rámci tohoto segmentu?
- Jaké aplikace jsou v rámci tohoto segmentu nejvýdělečnější?

Identifikace slabých stránek konkurenčních aplikací

- Jaké jsou nejčastější stížnosti uživatelů v tomto segmentu?
- Na jaké kritické chyby uživatelé naráží ve špatně hodnocených aplikacích?

²⁰ <https://appmagic.rocks/top-charts/apps>

- Existuje spojitost mezi sentimentem recenzí a frekvencí aktualizace aplikace?

Formulace strategie pro vstup na zahraniční trhy

- Jaké jazyky jsou v rámci tohoto segmentu nejčastěji podporované?
- Na jaké trhy mobilní aplikace v tomto segmentu cílí?
- Jaký jsou rozdíly mezi českým a americkým trhem v rámci tohoto segmentu?

6.3 Sběr informací

Fáze sběru informací bude provedena pomocí webové aplikace, která byla vyvinuta v rámci této diplomové práce. Do formuláře pro extrakci dat podle klíčových slov jsem zadal následující kritéria.

Použitá vyhledávací kritéria

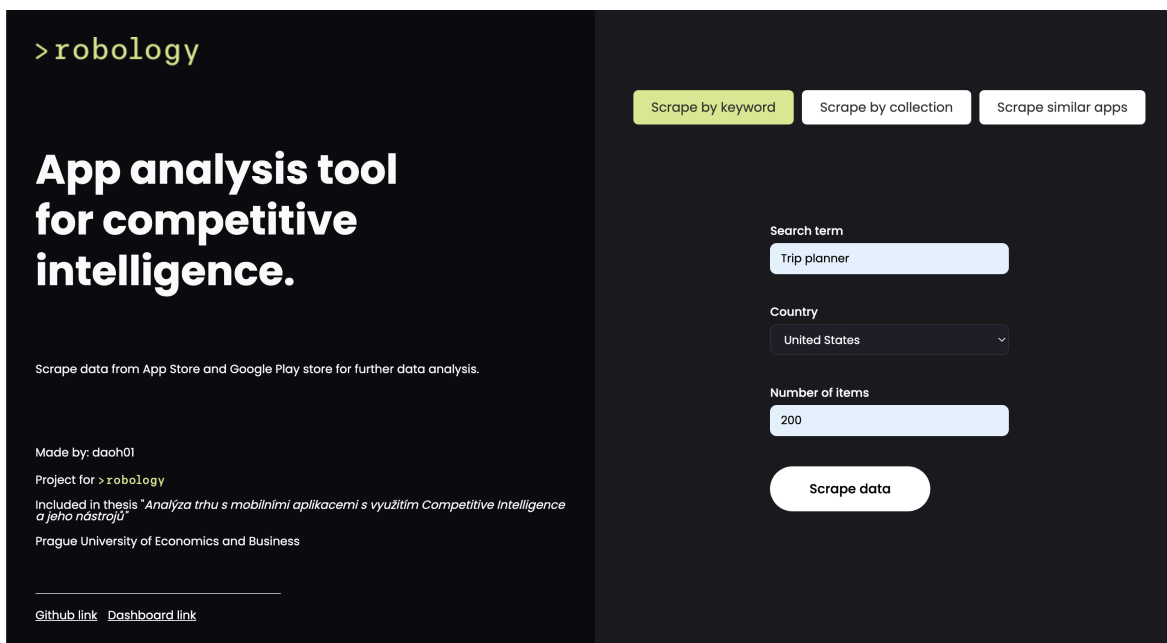
- Vyhledávací řetězec (Search term) = „*Trip planner*“.
- Země (Country) = „*United States*“
- Počet položek (Number of items) = „*200*“.

Výraz „*Trip planner*“ (Plánovač cest) je nejužitečnější pro typ aplikace, kterou plánujeme vyvíjet. Kritérium vyhledávání pro Spojené státy americké je vybráno kvůli tomu, že se jedná o největší a nejkonkurenčnější trh. Pokud jsou aplikace na tomto trhu považovány za klíčové hráče, je velmi pravděpodobné, že budou také globálními hráči.

Počet položek byl definován na hodnotu 200. Tato hodnota neznámá to, že z distribučních platform vyextrahujeme přesně tolik položek, slouží však jako maximální horní limit. Obvykle se spíše stane to, že na platformě nenajde tolik aplikací, které by splňovaly uživatelská kritéria, aby se horního limitu při extrakci dosáhlo.

Po zadání těchto kritérií a spuštění extrakce byla data transformována a nahrána automaticky do úložiště, odkud čerpá nástroj Google Looker Studio, ve kterém budeme provádět analýzu. Web scraper podle daných kritérií získal data o 30 aplikacích z platformy Google Play a 199 aplikací z platformy App Store.

Jako suplement k získaným informacím také využiji platformu AppMagic, která poskytuje detailnější data, která nejsou na distribučních platformách dostupná. Jedná se např. o odhady tržeb, popularita v různých zemích atp. Přestože je tato platforma placená, v demo verzi umožňuje uživatelům vyhledávat detaily o aplikacích ručně podle jejich názvu či kategorie.



Obrázek 17: Zadávání vyhledávacích kritérií pro extrakci dat ve webové aplikaci, Zdroj: autor

6.4 Analýza

6.4.1 Identifikace klíčových hráčů v segmentu

Po určení klíčových hráčů jsem se zaměřil na aplikace s nejvyšším počtem stažení a nejvyšším hodnocením. V dashboardu pro platformu Google Play jsem vyfiltroval aplikace podle počtu stažení a seřadil je sestupně. Bylo nalezeno těchto 10 nejstahovanějších aplikací viz Obrázek 18.

	Title	Installs ▼	Score	Sentiment	Store	Developer
1.	Tripadvisor: Plan & Book Trips	100 000 000	4,40	Positive	Link	Tripadvisor
2.	Citymapper	10 000 000	4,80	Slightly negati...	Link	Citymapper Li...
3.	KAYAK: Flights, Hotels & Cars	10 000 000	4,78	Slightly negati...	Link	KAYAK.com
4.	Triplt: Travel Planner	5 000 000	4,71	Positive	Link	Triplt, Inc.
5.	Circuit Route Planner	5 000 000	4,64	Slightly negati...	Link	Circuit Routin...
6.	Wanderlog - Trip Planner App	1 000 000	4,69	Positive	Link	Wanderlog
7.	Roadtrippers - Trip Planner	1 000 000	3,54	Negative	Link	Roadtrippers
8.	Visit A City	1 000 000	4,76	Negative	Link	Visit A City
9.	Polarsteps - Travel Tracker	1 000 000	4,68	Positive	Link	Polarsteps
10.	Zeo Fast Multi Stop Route Plan	1 000 000	4,53	Slightly negati...	Link	Zeo Business

Obrázek 18: Získaný seznam deseti nejlépe hodnocených aplikací v segmentu aplikací pro plánování cest na platformě Google Play, Zdroj: autor

V rámci dashboardu jsou definovány následující rychlé přehledy, které značí, že jsou aplikace velmi dobře hodnocené, s průměrným počtem stažení přes 13 milionů. Jsou také aktivně udržované, jelikož jsou aktualizované průměrně jednou za 16 dní.



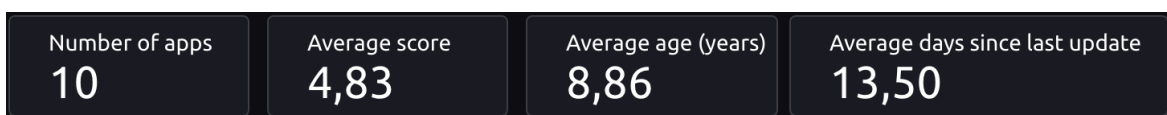
Obrázek 19: Přehled metrik deseti nejstahovanějších aplikací platformě Google Play, Zdroj: autor

Pro platformu App Store, jsem použil jako klíčovou metriku uživatelské skóre, jelikož není počet stažení na této platformě dostupný. To, že má aplikace vysoké hodnocení nutně nemusí znamenat, že se jedná o klíčového hráče na trhu. I přesto je tato informace užitečná, jelikož tato informace signalizuje, že je aplikace uživateli považována za kvalitní. Zde je 10 nejlépe hodnocených aplikací v tomto segmentu na platformě App Store viz Obrázek 20.

	App title	Score ▾	Sentiment	Store	Developer
1.	Viator: Tours & Attractions	4,91	Negative	link	Viator
2.	Wanderlog - Travel Planner	4,91	Slightly negative	link	Travel and Vacation Pl...
3.	Nowy: AI Travel Social Planner	4,89	Positive	link	Nowy Inc.
4.	Visit A City	4,84	Positive	link	Visit A City
5.	Triplt: Travel Planner	4,83	Positive	link	Triplt
6.	Hopper: Flights, Hotels & Cars	4,81	Positive	link	Hopper
7.	Tripadvisor: Plan & Book Trips	4,80	Slightly positive	link	Tripadvisor
8.	Expedia: Hotels, Flights & Car	4,79	Negative	link	Expedia, Inc.
9.	Pogo - Travel Planner	4,78	Slightly negative	link	Pogo Travel Ltd
10.	Airbnb	4,78	Negative	link	Airbnb, Inc.

Obrázek 20: Získaný seznam deseti nejlépe hodnocených aplikací v segmentu aplikací pro plánování cest na platformě App Store, Zdroj: autor

Z rychlého přehledu metrik opět můžeme pozorovat, že jsou aplikace dobře hodnocené a udržované.



Obrázek 21: Přehled metrik deseti nejlépe hodnocených aplikací na platformě App Store, Zdroj: autor

Dále jsme si definovali otázku, zda jsou mezi klíčovými hráči na trhu aplikace, které jsou mladší jednoho roku. Tato otázka umožňuje identifikovat nové aplikace, které rychle získávají popularitu a tržní podíl. Informace o věku aplikace také může identifikovat, zda je tento segment aplikací zralý.

V dashboardu můžeme nalézt informace o zralosti aplikací. Níže jsou data pro aplikace z platformy Google Play, na kterých můžeme pozorovat, že je tento segment aplikací velmi zralý. Všichni klíčoví hráči spadají do kategorie *Very Mature*, což znamená že byly aplikaci vydány před více než třemi lety. Z rychlého přehledu lze vyzorovat, že průměrný věk těchto aplikací je 8,48 let, což znamená, že je tento segment aplikací velmi zralý.

	App maturity	App count	Avg score
1.	Very Mature	10	4,55

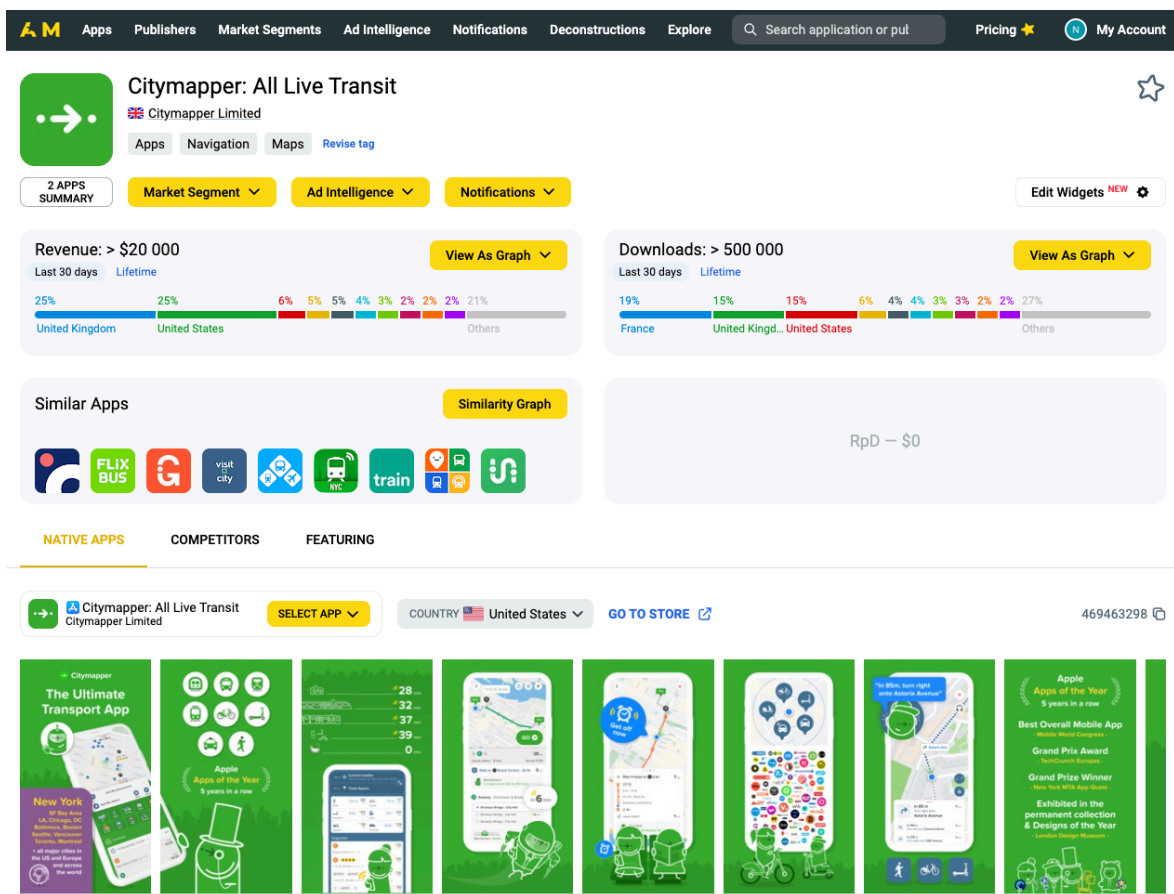
Obrázek 22: Zralost aplikací v rámci segmentu aplikací pro plánování cest z platformy Google Play, Zdroj: autor

Na platformě App Store lze pozorovat, že opět převažují aplikace s hodnotou *Very Mature*, které jsou starší tří let. Oproti platformě Google Play je zde přítomna i jedna mladší aplikace (*Novy:AI Travel Social Planner*) s hodnotou *Mature*, které byly vydána před jedním až dvěma roky.

	App maturity	App count...	Avg score...
1.	Very Mature	9	4,83
2.	Mature	1	4,89

Obrázek 23: Zralost aplikací v rámci segmentu aplikací pro plánování cest z platformy App Store, Zdroj: autor

Pro zodpovězení poslední KIT, je nutné identifikovat země, ve kterých sídlí tvůrci daných aplikací. Vědomí o tom, kde konkurence sídlí, může napomoci v rozpoznání jejich možných výhod, jako je přístup k technologickým zdrojům, talentu, nebo nižším mzdovým nákladům. Tato informace může být klíčová při formulaci efektivních konkurenčních strategií. K tomu, abych našel tyto informace, využiji platformu AppMagic, kde jsou informace o zemích původů aplikací dostupné. Informace hledám pro jednotlivé aplikace, přičemž používám jejich název jako vyhledávací termín.



Obrázek 24: Detail mobilní aplikace na platformě AppMagic, Zdroj: autor

Po analýze jsem vytvořil následující tabulku s informacemi, z jakých zemí vydavatelé těchto aplikací pocházejí, viz Tabulka 3. Celkově je ve vzorku 16 unikátních aplikací, jelikož jsou čtyři aplikace, které se řadí mezi klíčové hráče na obou distribučních platformách.

Tabulka 3: Seznam zemí a jejich zastoupení ve vzorku, Zdroj: autor

Stát	Počet aplikací	Procentuální zastoupení
Spojené státy	10	62,50 %
Velká Británie	3	18,75 %
Izrael	1	6,25 %
Nizozemsko	1	6,25 %
Kanada	1	6,25 %

Bylo zjištěno, že drtivě největší zastoupení mezi klíčovými hráči jsou firmy ze Spojených států amerických. Toto zjištění není překvapivé, jelikož má tato země nejvyspělejší technologický sektor na světě.

Za zmínku také ale stojí i aplikace od vývojářů z České republiky. Jedná se o aplikace *Wordlee Trip Planner & Journal* a *Sygyic:Travel Maps and Trip planner*. Ty sice nebyly zařazeny do pozice klíčových hráčů podle hodnocení či počtu stažení, ale představují lokální konkurenci, na kterou bychom dávat pozor.

6.4.2 Sledování tržních trendů a průzkum funkcionalit

Pro analýzu funkcionalit využiji pole s popisem funkcionality aplikací, které se nachází na stránkách dashboardu App Store Review Text Analysis a Google Play Review Text Analysis. Následně prozkoumám popisy aplikací, kde vývojáři specifikují jejich funkce a také informace o poslední aktualizaci uživatelů, kde je popsáno, jaké funkce byly přidány v nejnovější verzi. Tento postup jsem aplikoval na všechny aplikace zmíněné v předchozí analýze. Celkem jsem provedl detailní analýzu funkcí šestnácti unikátních aplikací, z nichž čtyři byly zastoupeny na obou platformách.

App description and changes in the last update		
Title	description	releaseNotes
1. Tripadvisor: Plan & Book Trips	<p>** Be one of the first to try the beta version of our newly-launched AI itinerary builder, offering personalized recs powered by traveler tips and reviews. **</p> <p>As a travel guidance company, Tripadvisor brings people, passions and places together. We aim to help make you a better traveler, from travel planning, to booking, to taking a trip. The app lets you get the most out of your trip—whether you're planning or on-the-go. Discover where to stay, what to do and where to eat based on guidance from millions of travelers who have been there before. Find deals on hotels, book experiences, reserve tables at great restaurants and discover great places nearby. No matter what type of trip you're looking to take, the Tripadvisor app makes planning it easy and also lets you guide others on their way.</p> <p>GET TRAVEL GUIDANCE</p> <ul style="list-style-type: none"> • Find millions of reviews about hotels, restaurants, tours, attractions and other experiences from travelers like you • Read articles and guides to discover new ways to experience a destination with tips from travel experts • Discover great deals on top-rated hotels, tours, activities, experiences and other things to do • Find great places to eat nearby and make restaurant reservations <p>ORGANIZE YOUR TRIP IN ONE PLACE</p> <ul style="list-style-type: none"> • Use the updated Trips planning tool to save hotels, restaurants, and 	<p>Improvements to the Trip planning product which allow travelers to:</p> <ul style="list-style-type: none"> - Receive a custom AI itinerary backed by traveler reviews - View saves in an organized list within the saves tab - Build an itinerary by adding saves and organizing them by day - View saves plotted on a map

Obrázek 25: Tabulka s popisem aplikací a informacemi o poslední aktualizaci, Zdroj: autor

Z analýzy mi vzešla následující tabulka, která znázorňuje procentuální zastoupení funkcionalit a počet aplikací, které tyto funkce nabízejí, viz Tabulka 4.

Tabulka 4: Seznam funkcionalit a jejich zastoupení ve vzorku, Zdroj: autor

Funkce	Počet aplikací	Procentuální zastoupení
Plánování cest	16	100,00 %
Itinerář	9	56,25 %
Objevování atrakcí	9	56,25 %
Tipy od cestovatelů	8	50,00 %
Kolaborativní plánování	5	31,25 %
Sdílení itineráře	7	43,75 %
AI itinerář	3	18,75 %
Import vstupenek/letenek/ubytování	6	37,50 %
Rezervace ubytování	6	37,50 %
Rezervace letenek	3	18,75 %
Rezervace aut	3	18,75 %
Rezervace zážitků	5	31,25 %
Hledání spojů MHD	2	12,50 %
Navigace	7	43,75 %
Aktuální informace o letech	5	31,25 %
Informace o bezpečnostní situaci	1	6,25 %
Úložiště dokumentů	3	18,75 %
Plánování rozpočtu	2	12,50 %
AI asistent	2	12,50 %

Z analýzy dat je zřejmé, že funkce plánování cest je univerzálně nabízena ve všech zkoumaných aplikacích, což odráží její základní význam pro segment aplikací pro plánování cest. Tato funkce je nezbytná pro základní operace související s organizací a plánováním cest, což potvrzuje její 100% zastoupení mezi analyzovanými aplikacemi. Zajímavým zjištěním je, že funkcionalita itineráře a objevování atrakcí není tak běžná, jak bychom si na první pohled mohli u aplikací pro plánování cest myslet. Tyto funkce má jen přes polovina aplikací z vybraného vzorku.

Funkce pro sdílení itineráře a kolaborativní plánování, jsou pro nás důležité, jelikož jsou stěžejním principem naší mobilní aplikace. Tyto funkce jsou zastoupeny ve více než 30 % aplikací, což naznačuje, že trend sdílení a spolupráce mezi uživateli při plánování cest není tak rozšířený.

Rezervace ubytování, letů a zážitků integrované přímo do aplikací, jsou nabízeny téměř v 40 % případů, což zjednodušuje proces rezervace a zvyšuje pohodlí uživatelů. Tyto funkce jsou vyhrazené pro skutečně nejvýznamnější aplikace, jelikož jsou nejnáročnější na implementaci – jak technickou, tak obchodní.

Aktuální informace o bezpečnostní situaci, navigace a plánování rozpočtu, jsou funkce, které mohou uživatelům výrazně vylepšit zážitek. Jsou však dostupné v nižším procentu aplikací, ale představují důležitý zdroj pro hledání inspirace pro implementaci nových funkcí.

Použití umělé inteligence zatím není v segmentu plánování cest příliš rozšířené. Pouze 12,50 % aplikací nabízí nějakou formu AI asistenta nebo AI generování itineráře. Toto zjištění ukazuje, že přestože AI má potenciál transformovat způsob, jakým uživatelé plánují cesty, stále je v tomto segmentu prostor pro inovace a další rozvoj.

Po prozkoumání funkcí, které se objevují v posledních aktualizacích jsem zpracoval následující tabulku, viz Tabulka 5. Některé aplikace adresovaly v poslední aktualizaci více funkcí, proto nemusí být sloupec počet aplikací roven 16.

Tabulka 5: Seznam funkcionalit zastoupených v posledních aktualizacích aplikací, Zdroj: autor

Předmět aktualizace	Počet aplikací	Procentuální zastoupení
Itinerář	1	6.25 %
Objevování atrakcí	1	6.25 %
AI itinerář	2	12.50 %
Rezervace ubytování	1	6.25 %
Rezervace zážitků	1	6.25 %
Koupě lístků pro MHD	1	6.25 %
Porovnání cen letenek	1	6.25 %
Informace o bezpečnostní situaci	1	6.25 %
Úložiště dokumentů a fotek	1	6.25 %
Vylepšení UI/UX	4	25.00 %
Oprava chyb	5	31.25 %

Offline mode	1	6.25 %
Přidání podpory pro jazyky	1	6.25 %

Z těchto informací můžeme pozorovat, že v rámci poslední aktualizace aplikace vývojáři řešili nejčastěji opravy chyb a vylepšení UI/UX. Co se týče přidání konkrétních funkcí, můžeme pozorovat, že jsou rozprostřeny poměrně rovnoměrně. Lze však pozorovat mírný nárůst aplikací, jejichž poslední aktualizace přidala funkce s AI.

Třetí otázka se týkala využití již zmiňovaného AI v rámci těchto aplikací. To může být tedy dvěma způsoby:

- **AI Itinerář**, který dokáže vygenerovat celý program cesty podle kritérií definovaných uživatelem.
- **AI Asistent**, který může uživateli pomoci s navigací, informacemi o lokalitě, tipy a triky při cestování apod.

6.4.3 Identifikace monetizačních strategií

Na dashboardu pro Google Play máme k dispozici několik klíčových atributů týkajících se způsobů, jakými aplikace generují příjmy. Tyto atributy zahrnují podporu reklamy, informace o tom, zda je aplikace placená nebo zdarma, a indikaci, zda aplikace umožňuje nákupy uvnitř samotné aplikace.



Obrázek 26: Vizualizace atributů pro identifikaci monetizačních strategií z platformy Google Play, zdroj: Autor

Z těchto vizualizací lze vyčíst, že většina aplikací (76,7 %) v tomto segmentu nepoužívá reklamy jako monetizační model. Současně je 100 % zkoumaných aplikací nabízeno uživatelům zdarma, což potvrzuje převládající model freemium na celkovém trhu s mobilními aplikacemi. Nakonec 70 % aplikací poskytuje možnost nákupů v aplikaci, což odráží běžnou praxi monetizace prostřednictvím mikrotransakcí. Tyto poznatky poukazují na to, že většina aplikací se spoléhá na in-app purchases jako na hlavní zdroj příjmů.

Pro zjištění cenové hladiny těchto mikrotransakcí, máme v dashboardu k dispozici tabulku, ve které jsou spodní a horní hodnoty cen, společně s měnou, viz Obrázek 21.

In-app purchases pricing overview				
	Title ▾	Lowest price	Highest price	Currency
1.	iplan.ai - Travel Planner	4	10	\$
2.	Zeo Fast Multi Stop Route Plan	2	200	\$
3.	Wanderlog - Trip Planner App	2	80	\$
4.	Visit A City	5	50	\$
5.	Tripmaker: AI Travel Assistant	4	60	\$

1 - 5 / 21 < >

Obrázek 27: Tabulka s přehledem spodních a horních hodnot cen pro nákupy v aplikaci, Zdroj: autor

Ceny za mikrotransakce obvykle začínají na jednom americkém dolaru, můžou se však vyšplhat i do stovek dolarů. Jednotlivé aplikace jsem prozkoumal manuálně přes proklik na jejich web a zjistil jsem, že většina aplikací funguje na bázi měsíční či ročního předplatného. Uživatel, který si toto předplatné pořídí, dostane přístup k dalším funkcionalitám aplikace, které jsou v základu uzamčené. Příkladem je třeba aplikace Wanderlog, která nabízí v rámci předplatného funkce jako offline přístup, neomezený prostor pro dokumenty a mnoho dalšího.

Price comparison

Freemium

\$0

Sign up

- Add unlimited places
- Live collaboration with friends
- Import flights/hotels/cars
- Smart recommendations
- Itinerary and map in 1 view

Pro Annual

\$49.99/year

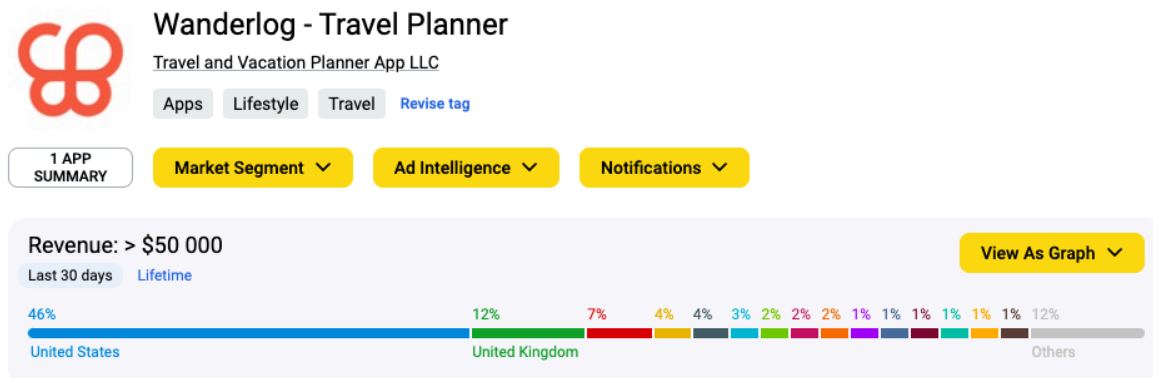
Start free trial

Everything in the free version **plus**:

- Offline access
- Flight and car rental deals
- Unlimited attachments
- Export to Google Maps
- Optimize route
- Auto Gmail scanning
- Dark mode in app
- Unlimited AI assistance
- Add places from anywhere on the web

Obrázek 28: Funkce jednotlivých typů předplatného aplikace Wanderlog [43]

Dále je důležité identifikovat, jaké aplikace jsou v daném vzorku nejvýdělečnější. K tomu jsem opět využil platformu AppMagic, která má u některých aplikací dostupnou metriku obratu. Je však nutno podotknout, že se jedná o odhad.



Obrázek 29: Výstřižek z platformy AppMagic, zobrazující odhad tržeb z mikrotransakcí, Zdroj: autor

Pro mobilní aplikace, kde byla metrika dostupná, jsem vytvořil následující tabulku:

Tabulka 6: Odhad tržeb z nákupů v aplikacích za posledních 30 dní, Zdroj: autor

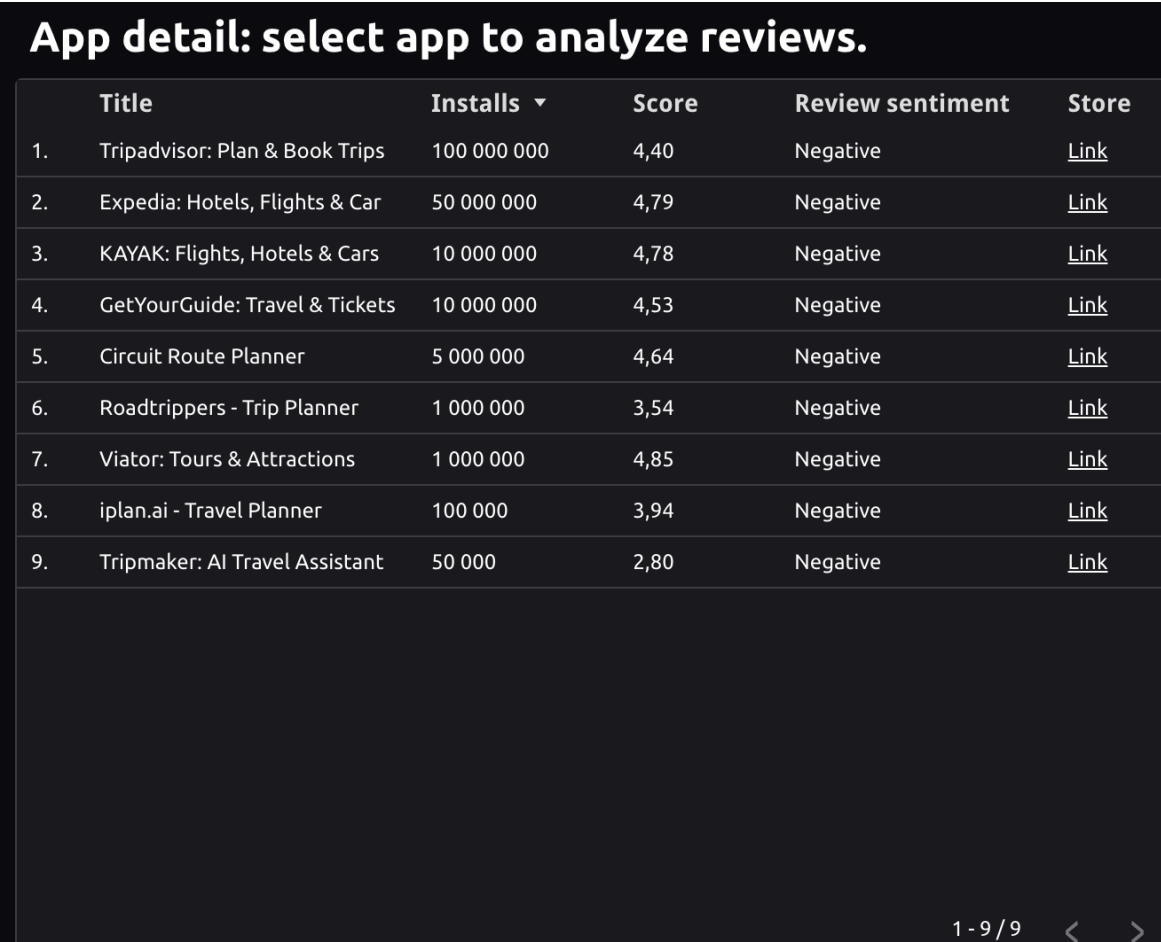
Aplikace	Odhad tržeb z nákupů v aplikaci za posledních 30 dní
Citymapper: All Live Transit	>20 000 USD
TripIt: Travel Planner	>20 000 USD
Wanderlog – Travel Planner	>50 000 USD
Roadtrippers – Trip Planner	>20 000 USD
Visit A City	<5 000 USD
Zeo Fast Multi Stop Route Plan	> 20 000 USD

Z dat, která jsou dostupná, lze vyčíst, že se obraty z nákupů v aplikaci v tomto segmentu pohybují kolem desítek tisíců dolarů měsíčně. Největší odhadovaný obrat má aplikace Wanderlog – Travel Planner, pro niž se odhaduje tržba nad 50000 USD měsíčně.

6.4.4 Identifikace slabých stránek konkurenčních aplikací

Pro tuto analýzu použiji stránky dashboardu App Store Review Text Analysis a Google Play Review Text Analysis. Tyto stránky obsahují informace o frekvenci nejčastějších bigramů a jednotlivých slov, ze kterých mohu vyčíst klíčová slova, která charakterizují sentiment uživatelů.

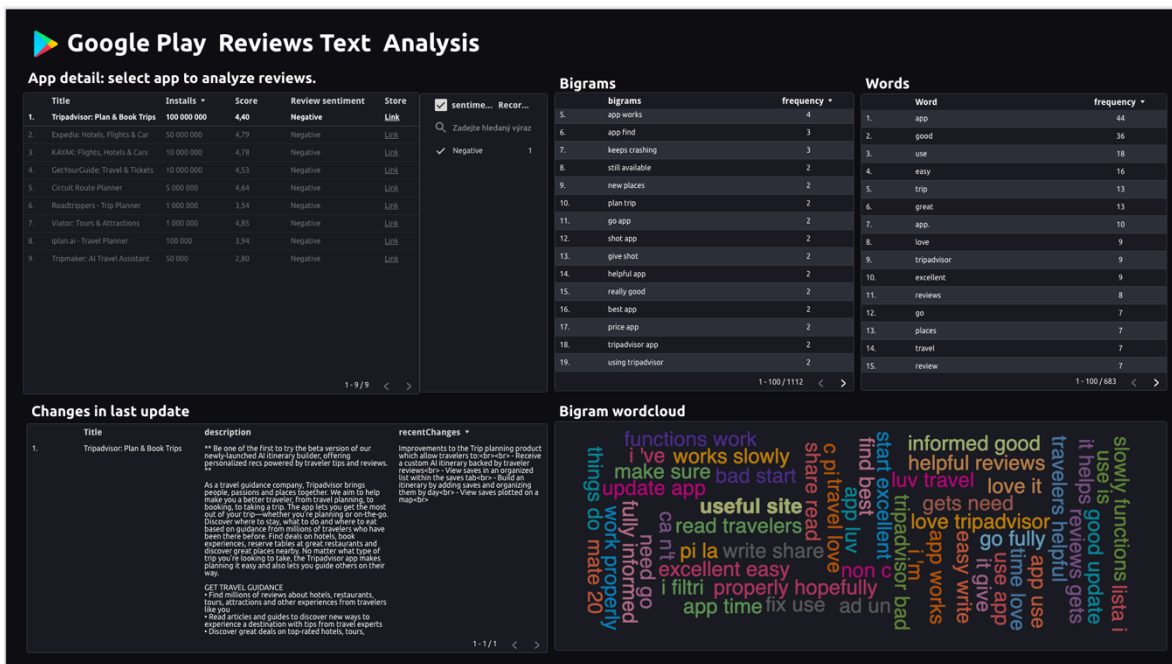
Pokud se zaměřím na aplikace s negativním sentimentem uživatelských recenzí, bude poměrně snadné najít slabé stránky daných aplikací. Je ještě nutné podotknout, že vzorek pro klasifikaci byl 200 nejnovějších recenzí – celkový sentiment recenzí aplikace nemusí být negativní. I přesto je relevantní sledovat nejaktuálnější sentiment uživatelských aplikací v téměř reálném čase. Pomocí filtru jsem si zobrazil aplikace s negativním sentimentem v rámci platformy Google Play.



	Title	Installs ▾	Score	Review sentiment	Store
1.	Tripadvisor: Plan & Book Trips	100 000 000	4,40	Negative	Link
2.	Expedia: Hotels, Flights & Car	50 000 000	4,79	Negative	Link
3.	KAYAK: Flights, Hotels & Cars	10 000 000	4,78	Negative	Link
4.	GetYourGuide: Travel & Tickets	10 000 000	4,53	Negative	Link
5.	Circuit Route Planner	5 000 000	4,64	Negative	Link
6.	Roadtrippers - Trip Planner	1 000 000	3,54	Negative	Link
7.	Viator: Tours & Attractions	1 000 000	4,85	Negative	Link
8.	iplan.ai - Travel Planner	100 000	3,94	Negative	Link
9.	Tripmaker: AI Travel Assistant	50 000	2,80	Negative	Link

Obrázek 30: Seznam aplikací s negativním sentimentem recenzí z platformy Google Play, zdroj: Autor

V rámci této tabulky mohu kliknout na jednotlivou aplikaci a tím se spustí křížové filtrování, které mi ukáže data o nejčastějších slovech a bigramech v recenzích, popisu aplikace a informace o nejnovější aktualizaci, společně s vizualizací frekvence bigramů pomocí wordcloudu.



Obrázek 31: Ukázka funkcionality dashboardu při výběru aplikace, Zdroj: Autor

Detailně jsem si dané bigramy a slova manuálně prošel a vytvořil jsem tabulku s negativními bigramy, které jsem zaznamenal, viz Tabulka 6. Recenze od uživatelů někdy nemusí mít výpovědní hodnotu, např. v rámci průzkumu jsem našel bigramy jako „stuck philippines“ nebo „bad weather“, které nejsou věcné či nesouvisí přímo s mobilní aplikací. Proto bylo důležité, abych při této analýze použil svůj lidský úsudek.

Tabulka 7: Příklady nalezených bigramů z uživatelských recenzí, Zdroj: autor

Title	Příklady nalezených negativních bigramů
Tripadvisor: Plan & Book Trips	<i>keeps crashing, missing features, closed twice, 2023 unstable, trips ruined, crashing huawei,</i>
Expedia: Hotels, Flights & Car	<i>problem bookings, high fees, sign mistakenly, wrong page, reinstalling app, glitch always, poor customer, app unusable, no refunds, payment failed</i>
KAYAK: Flights, Hotels & Cars	<i>app closes, late processing, stop spam, spam terrible, error ui, uninstalling use, unusable now, completely unusable, incorrect hotel,</i>

	<i>site error, worst app, closes always, please fix</i>
GetYourGuide: Travel & Tickets	<i>scam ever, bombard policies, hidden policies, deceptive descriptions, refused refund, app terrible, troubleshoot promo, problem order, description scam, service useless, scam beware</i>
Circuit Route Planner	<i>iphone update, please fix, latest version, waste time, samsung uninstalled, uninstalled tried, freezing phone, duplicates update, app buggy, freezes always</i>
Roadtrippers – Trip Planner	<i>impossible cancel, stops monthly, crash crash, app worthless, behind paywall, money blocked, app useless, app sucks, screw customers, keeps crashing, unusable since, cancellation denied, ridiculous expensive, expensive fee, planning impossible, gps fail</i>
Viator: Tours & Attractions	<i>caution booking, spam folder, reservation scam, scam expect, small font, fraud company, refund hearing, viator fraud, app useless, operator cancelled, service useless</i>
iplan.ai - Travel Planner	<i>unimpressed app, problems purchased, frustrated inaccurate, inaccurate locations, still dissapointed, error status, error 500, generated error, crashed times, doesnt work, 502 error, gives 502, screen killed, buggy rushed</i>
Tripmaker: AI Travel Assistant	<i>buggy impossible, impossible delete, limited</i>

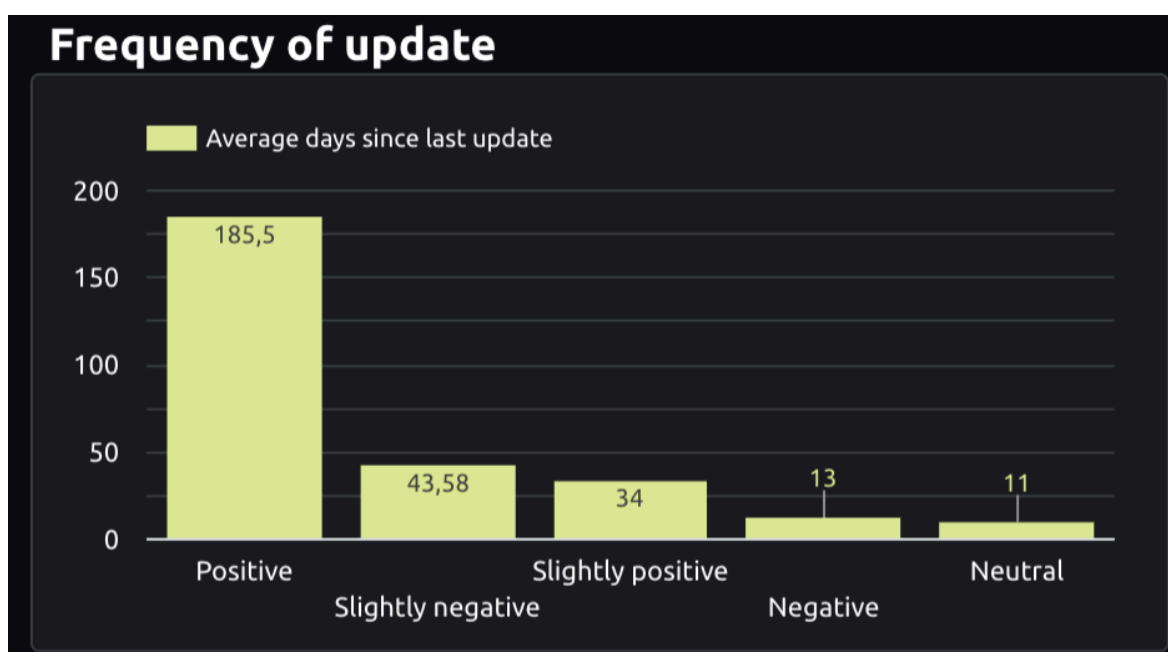
	<i>unfortunate, much useless, super glitchy, deleted items, extremely limited, scam app, pretty limited, limited knowledge,</i>
--	---

Z této analýzy jsem našel odpovědi na položené otázky. Nejčastější stížnosti rozdělím do těchto kategorií:

- **Stížnosti na zákaznický servis** – uživatelé si často stěžují na zákaznickou podporu. Časté jsou problémy s vrácením peněz, zrušením cesty, ve skrytých podmínkách. Tyto stížnosti jsou zastoupeny hlavně v aplikacích, které poskytují rezervaci letenek, ubytování či zážitků.
- **Stížnosti na chyby v aplikaci** – uživatelé hlásí technické problémy. Nejčastěji se jedná o pády aplikací, nefunkčnosti na specifických typech zařízení, bugy v aplikaci, chybě ve zpracování platby. Tyto stížnosti lze nalézt v recenzích pro každou aplikaci.
- **Stížnosti na cenu** – uživatelé si často stěžují na ceny za předplatné. V bigramech se často objevovali číslovky, které korespondovaly s cenou předplatných.

Z analýzy také vyplynulo, že aplikace, které v sobě mají integrované AI, mají často negativní recenze, specifické pro tuto technologii. Uživatelé si stěžují na nepřesné nebo neúplné informace nebo technické chyby, které jim aplikaci zabraňují využívat.

Dále jsem se zaměřil na klíčovou otázku, zda existuje korelace frekvence aktualizací se sentimentem recenzí. Pro zjištění odpovědi na tuto otázku jsem využil dostupnou vizualizaci na dashboardu, která tyto atributy porovnává.

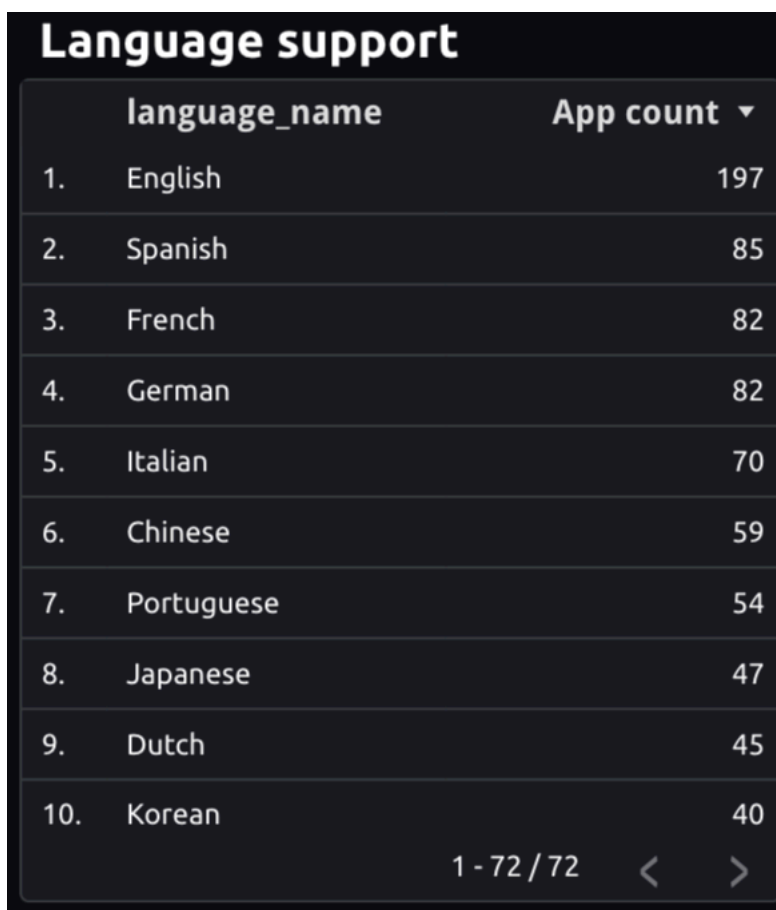


Obrázek 32: Graf zobrazující průměrný počet dní od poslední aktualizace podle kategorií sentimentu z platformy Google Play, Zdroj: autor

Bylo zjištěno, že mezi těmito atributy existuje korelace, je však opačná, než jsem předpokládal – aplikace, které mají pozitivní sentiment uživatelských recenzí, jsou zároveň nejméně aktualizované (průměrně jednou za 185,5 dní). Naproti tomu, pokud má aplikace negativní sentiment, je aktualizovaná průměrně jednou za 13 dní, což je druhá nejvyšší průměrná frekvence aktualizací. Z tohoto zjištění jsem formuloval hypotézu, že se zvýšená frekvence aktualizace aplikací odráží od špatného sentimentu recenzí. Vývojáři tak musí reagovat na uživatelskou zpětnou vazbu. Naopak, pokud má aplikace pozitivní sentiment, vývojáři ji nemají důvod aktualizovat tak často.

6.4.5 Formulace strategie pro vstup na zahraniční trh

Pro tuto analýzu se odrazím z dat o jazykové podpoře, které jsou dostupné z platformy App Store. K dispozici mám data o jednotlivých jazycích a také o jednotlivých státech, které jsem si namapoval podle dané jazykové podpory. Na dashboardu o datech z App Store je definována níže vyobrazená tabulka.



	language_name	App count ▾
1.	English	197
2.	Spanish	85
3.	French	82
4.	German	82
5.	Italian	70
6.	Chinese	59
7.	Portuguese	54
8.	Japanese	47
9.	Dutch	45
10.	Korean	40

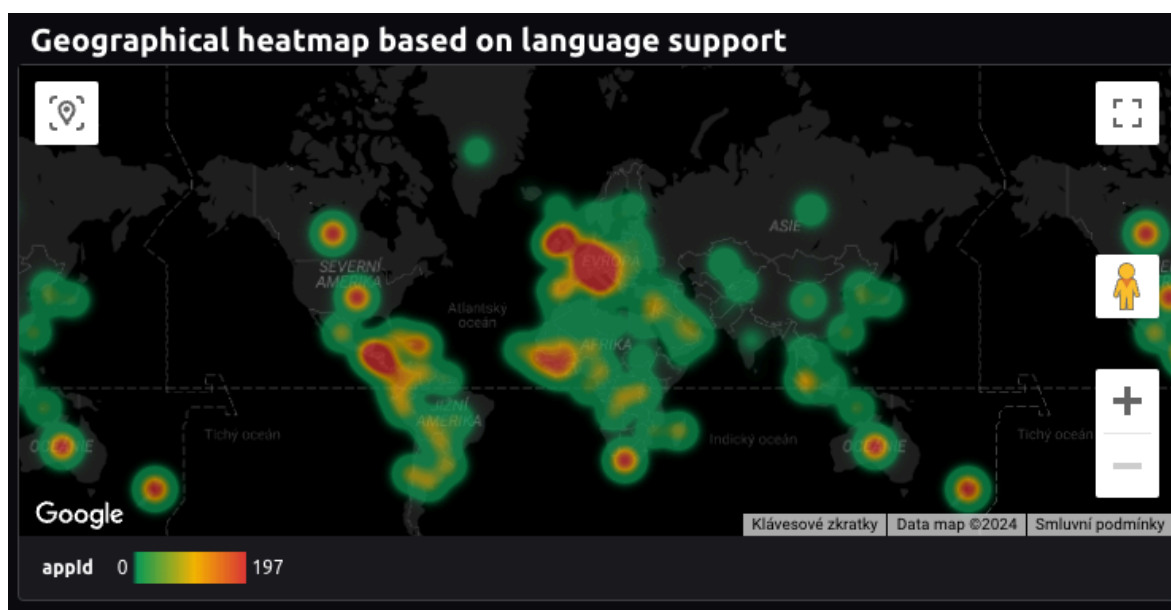
1 - 72 / 72 < >

Obrázek 33: Tabulka se seznamem podporovaných jazyků a počtem aplikací, které dané jazyky podporují, Zdroj: autor

V rámci zkoumaného vzorku aplikací v daném segmentu lze identifikovat, že angličtina je nejvíce podporovaným jazykem, s podporou ve 197 aplikacích, což odráží její celosvětovou důležitost. Pro aplikace je podpora anglického jazyka nutnost.

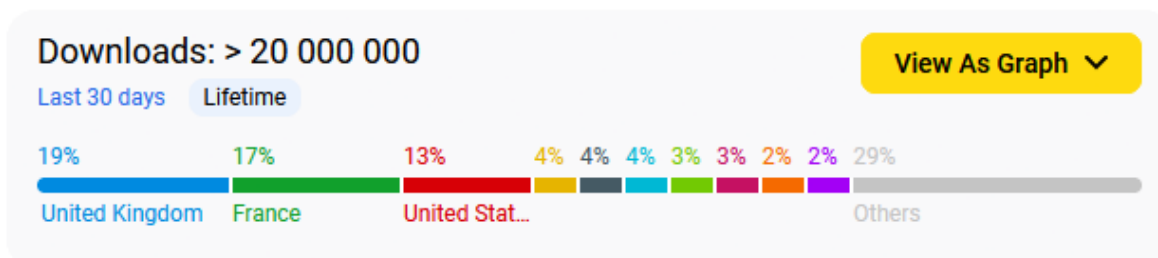
Následují španělština a francouzština, obě podporované v 85 resp. 82 aplikacích, což signalizuje strategický záměr oslovit trhy v Evropě, Severní a Jižní Americe, kde jsou tyto jazyky široce rozšířené. Němčina a italština, každá s podporou ve 82 resp. 70 aplikacích, dále ukazují na významnost Evropské unie jako cílového trhu pro tento segment.

V rámci asijského trhu je významná jazyková podpora čínštiny, japonštiny a korejštiny, což jsou také velmi konkurenční a technologicky pokročilé trhy. Tato zjištění podporuje i geografická heatmapa, která je založená právě na podpoře jazyků.



Obrázek 34: Geografická heatmapa na základě podpory jazyků, Zdroj autor

Poslední definovanou otázkou v rámci tohoto zpravodajského tématu je zjistit, z jakých zemí je uživatelská báze klíčových hráčů na trhu. K tomu jsem opět využil platformu AppMagic, ze které lze získat informace o zeměpisném rozložení uživatelské báze.



Obrázek 35: Výstřížek z platformy AppMagic, zobrazující zeměpisné rozložení uživatelské báze, Zdroj: autor

Do analýzy jsem zahrnul aplikace, které jsem identifikoval jako klíčové hráče.

Tabulka 8: Seznam aplikací a jejich Top 5 zemí podle počtu stažení, Zdroj: autor

Aplikace	Top 5 zemí podle počtu stažení (v pořadí)
Tripadvisor: Plan & Book Trips	USA, Itálie, Velká Británie, Francie, Indie
Citymapper: All Live Transit	Velká Británie, Francie, USA, Brazílie, Španělsko
KAYAK: Flights, Hotels & Cars	USA, Brazílie, Francie, Mexiko, Španělsko
TripIt: Travel Planner	USA, Austrálie, Velká Británie, Indie, Kanada
Circuit Route Planner	Brazílie, USA, Mexiko, Velká Británie, Argentina
Wanderlog - Travel Planner App	USA, Austrálie, Velká Británie, Kanada, Singapur
Roadtrippers - Trip Planner	USA, Kanada, Austrálie, Francie, Velká Británie
Visit A City	USA, Velká Británie, Indie, Indonésie, Španělsko
Polarsteps - Travel Tracker	Nizozemsko, Francie, Německo, Belgie, Švýcarsko
Zeo Fast Multi Stop Route Plan	Brazílie, Chile, Mexiko, Velká Británie, Malajsie
Viator: Tours & Attractions	USA, Velká Británie, Austrálie, Kanada, Indie
Nowy: AI Travel Social Planner	USA, Kanada, Jihoafrická republika, Austrálie, Španělsko
Hopper: Flights, Hotels & Cars	USA, Kanada, Mexiko, Kolumbie, Velká Británie
Pogo - Travel Planner	Velká Británie, USA, Kanada, Nový Zéland, Austrálie
Airbnb	USA, Brazílie, Čína, Francie, Mexiko
Expedia: Hotels, Flights & Car	USA, Mexiko, Velká Británie, Kanada, Japonsko

Z těchto dat vyplývá, že hlavním trhem pro tento segment jsou Spojené státy americké, následované Velkou Británií. Další významné trhy jsou země Evropské unie jako např. Francie, Itálie a Španělsko. Všechna tato zjištění korelují s analýzou jazykové podpory. Tyto země by tedy představovaly hlavní cíl pro expanzi naší aplikace.

6.5 Vyhodnocení

Na základě provedené analýzy trhu s mobilními aplikacemi jsem zformuloval odpovědi na klíčové zpravodajské otázky, které by měly pomoci v strategickém rozhodování a k dalšímu vývoji zmiňované mobilní aplikace pro plánování cest.

6.5.1 Identifikace klíčových hráčů v segmentu

Jaké aplikace můžeme na tomto segmentu považovat za klíčové hráče?

Nejvýznamnějšími hráči jsou aplikace s vysokým počtem stažení a hodnocení jsou TripAdvisor, Expedia, KAYAK, TripIt, Wanderlog a Roadtrippers, které nabízejí rozsáhlé funkce pro plánování cest.

Jsou mezi klíčovými hráči aplikace, které vznikly v posledním roce?

Všichni klíčoví hráči jsou na trhu déle, což naznačuje, že je tento tržní segment zralý. Nové aplikace, které rychle získávají popularitu a tržní podíl, nebyly identifikovány.

V jakých zemích sídlí vývojáři daných aplikací?

Vývojáři klíčových aplikací jsou převážně ze Spojených států amerických, což ukazuje na silné technologické a inovační zázemí v této zemi.

6.5.2 Sledování tržních trendů a průzkum funkcionalit

Jaké funkce se nejčastěji objevují v aplikacích pro plánování cest?

Základní funkce zahrnují plánování cest, itineráře a objevování atrakcí, což je konzistentní napříč většinou aplikací v tomto segmentu.

Na jaké funkce se vývojáři soustředí v posledních aktualizacích aplikací?

Poslední aktualizace nejčastěji přinášejí vylepšení UI/UX a opravy chyb, lze však pozorovat rostoucí zaměření na integraci umělé inteligence.

Jak je v těchto mobilních aplikacích využívána umělá inteligence?

Umělá inteligence je používána především pro generování itinerářů a personalizované návrhy cest, ale je zastoupena jen v menším počtu aplikací, což ukazuje na prostor pro inovace.

6.5.3 Identifikace monetizačních strategií v rámci segmentu

Jaké monetizační modely jsou použité v aplikacích v rámci tohoto segmentu?

Převažuje model freemium s nákupy v aplikacích, což umožňuje uživatelům základní funkcionality zdarma s možností rozšíření při pořízení předplatného. Dále se v menší míře vyskytuje monetizace pomocí reklamy.

Jaká je cenová hladina pro in-app purchases v rámci tohoto segmentu?

Nákupy v aplikacích se pohybují od několika dolarů za malé funkce až po stovky dolarů, záleží na typu předplatného a také jeho délce.

Jaké aplikace jsou v rámci tohoto segmentu nejvýdělečnější?

Aplikace Wanderlog vykazuje vysoké tržby, což naznačuje silnou uživatelskou základnu a úspěšný monetizační model.

6.5.4 Identifikace slabých stránek konkurenčních aplikací

Jaké jsou nejčastější stížnosti uživatelů v tomto segmentu?

Uživatelé často uvádějí problémy s chybami v aplikaci, nedostatečnou zákaznickou podporou a vysokými cenami za prémiové funkce.

Na jaké kritické chyby uživatelé naráží ve špatně hodnocených aplikacích?

Kritické chyby zahrnují pády aplikací, chyby při placení či, neúplné funkcionality.

Existuje spojitost mezi sentimentem recenzí a frekvencí aktualizace aplikace?

Výsledky ukazují, že aplikace s negativními recenzemi jsou aktualizovány častěji, což naznačuje snahu vývojářů rychle reagovat na uživatelskou zpětnou vazbu.

7 Možnosti pro budoucí vývoj nástroje

V této kapitole popíšu možnosti pro budoucí vývoj nástroje pro analýzu trhu s mobilními aplikacemi. Představím několik funkcionalit, které by mohly nástroj výrazně vylepšit a umožnit tak ještě lépe pochopit tržní prostředí.

7.1 Rozšíření o funkcionalitu App Store Optimization

App Store Optimization (ASO) je proces zaměřený na zlepšení viditelnosti aplikací v rámci distribučních platforem Google Play nebo App Store, prostřednictvím optimalizace klíčových slov a dalších metadat. Cílem ASO je optimalizovat klíčová slova a další metadatové elementy aplikace, aby se zvýšila šance, že uživatelé aplikaci naleznou při vyhledávání. ASO zahrnuje analýzu trendů vyhledávání, optimalizaci popisků aplikací, obrázků a klíčových slov tak, aby byly co nejvíce relevantní pro hledané termíny.

Pro implementaci bychom mohli využít modulu *aso*²¹, který je postaven na existujících modulech a *app-store-scrapers*. Knihovna *aso* poskytuje robustní funkcionalitu pro analýzu a optimalizaci klíčových slov na základě reálných dat shromážděných z obchodů s aplikacemi. Modul nabízí nástroje pro efektivní optimalizaci vyhledávání aplikací pomocí několika klíčových funkcí: poskytuje hodnocení klíčových slov na základě obtížnosti hledání a provozu, což umožňuje identifikovat nejrelevantnější klíčová slova. Dále generuje návrhy klíčových slov podle kategorie, podobnosti a konkurenčních aplikací. Dokáže také vypočítat také index viditelnosti, který hodnotí, jak dobře je aplikace vyhledatelná.

Díky tomuto rozšíření bychom mohli provádět analýzy klíčových slov a aktualizovat strategie ASO pro naše vyvíjené aplikace na základě získaných dat.

7.2 Integrace dat od externích poskytovatelů

Integrace externích datových zdrojů, jako jsou Sensor Tower, Apptopia a AppMagic by nám umožnila analyzovat trh s mobilními aplikacemi do většího detailu. Tyto platformy poskytují cenné informace, které nejsou dostupné přímo z platforem Google Play nebo App Store. Tyto platformy mají odhady tržeb pro jednotlivé aplikace, detailní segment Integrace takových dat může významně zvýšit hodnotu našich analýz tím, že poskytne širší spektrum informací, které můžeme o mobilní aplikaci zjistit. Bohužel je toto poměrně nákladná záležitost, kvůli vysokým měsíčním cenám, které se pohybují v desítkách tisíc korun.

²¹ <https://github.com/facundoalano/aso>

7.3 Reporting a automatizovaná extrakce dat

V rámci rozvoje by mohl být také implementována funkce pro automatizaci extrakce dat s následným reportingem. Tato funkce by naší firmě umožnila průběžně monitorovat vývoj trhu a rychle reagovat na změny. Automatizace by mohla zahrnovat naplánování web scrapingových úloh v pravidelných intervalech, automatické generování reportů a jejich odesílání uživatelům.

Pro zavedení automatizace by bylo třeba vyvinout funkci pro plánování extrakce dat, která by byla integrována do webové aplikace. Plánovač by nám umožňoval definovat kritéria pro sběr dat, časové intervaly pro sběr dat a příjemce reportů. Po automatizované extrakci by proběhla transformace a vygenerování reportu, který by se následně odeslal definovaným příjemcům.

7.4 Ukládání dat pro historický kontext

Schopnost uchovávat a analyzovat historická data by poskytlo firmě hlubší vhled do dlouhodobých trendů a změn v tržních podmínkách. Integrace historických dat do našeho nástroje pro analýzu trhu s mobilními aplikacemi by nám umožnilo lepší pochopení dlouhodobých trendů, což je klíčové pro efektivní plánování vývoje produktů a marketingových strategií. Tato data by také pomohla při predikci budoucího tržního vývoje a umožnila by hodnocení výkonnosti aplikací v čase.

Ukládání historických dat by vyžadovalo rozšířenou databázovou infrastrukturu schopnou efektivně zpracovávat a ukládat velké objemy dat. Implementace takového systému by mohla zahrnovat vývoj a integraci s cloudovými datovými službami jako např. Google Big Query²², které poskytují škálovatelné řešení pro tvorbu datových skladů. Kromě toho by bylo možné použít nástroje pro data mining a strojové učení pro analýzu shromážděných historických dat, což by naší firmě mohlo umožnit předpovídat budoucí trendy.

²² <https://cloud.google.com/bigquery>

Závěr

Cílem této práce bylo vytvořit interaktivní analytický nástroj pro softwarovou agenturu Robology, který nabídne komplexní pohled na konkurenční prostředí v oblasti mobilních aplikací, přičemž pro tento účel integruje principy Competitive Intelligence a Business Intelligence. Analýza tohoto tržního prostředí probíhá využitím aktuálních dat získaných z distribučních platforem mobilních aplikací Google Play a App Store.

V teoretické části této diplomové práce se zabývám významem a metodami Competitive Intelligence a Business Intelligence, představením trhu s mobilními aplikacemi, jeho klíčových hráčů a klíčových metrik. Dále definuji klíčová zpravodajská témata, jejichž zodpovězení by mohlo firmě Robology pomoci ve strategickém rozhodování.

Praktická část podrobně popisuje návrh, vývoj a implementaci analytického nástroje. Tento nástroj je koncipován jako webová aplikace, která zajišťuje extrakci dat z distribučních platforem prostřednictvím web scrapingu, čištění a transformaci dat s využitím Pythonu a jeho knihoven, ukládání dat do cloudového úložiště a jejich následnou vizualizaci v interaktivním dashboardu na platformě Google Looker Studio.

Webová aplikace se skládá z frontendu (uživatelské rozhraní) vyvinutého v programovacích jazycích HTML, CSS a JavaScript a backendu (server), který je postaven na technologii Node.js. Na této technologii jsou založeny také open source moduly google-play-scraper a app-store-scraper, umožňující extrakci aktuálních dat z distribučních platforem pro mobilní aplikace Google Play a App Store. Data jsou v backendu transformována pomocí skriptu v jazyce Python. Tento skript zahrnuje čištění dat, kalkulaci odvozených atributů, kategorizaci atributů, textovou analýzu, a hlavně analýzu sentimentu uživatelských recenzí s využitím jazykového modelu BERT. Po transformaci jsou data odesílána do cloudového úložiště Google Cloud Storage, odkud jsou automaticky posílána do předdefinovaného dashboardu v Google Looker Studio.

Výsledkem je analytický nástroj s interaktivním rozhraním, do kterého může uživatel zadat svá kritéria pro extrakci dat z distribučních platforem Google Play a App Store. Nástroj data automaticky transformuje do struktury vhodné pro analýzu a nahrává je do cloudového úložiště, odkud se data automaticky nahrávají do analytického nástroje Google Looker Studio. Tato data jsou vizualizována v dashboardu, který umožňuje uživateli provádět vlastní analýzy a lépe pochopit tržní prostředí v oblasti mobilních aplikací.

Následně je demonstrováno využití tohoto nástroje při analýze trhu mobilních aplikací na základě definovaných klíčových zpravodajských témat, jako jsou identifikace klíčových hráčů v konkrétním segmentu, sledování tržních trendů, identifikace monetizačních strategií, identifikace slabých stránek konkurence a formulace strategie pro vstup na zahraniční trhy. Byl vysvětlen kontext analýzy, přičemž byl popsán produkt, který je nyní firmou Robology vyvíjen. Na základě tohoto kontextu byly definovány klíčové zpravodajské otázky. Sběr dat a následná analýza probíhala za použití vyvinutého analytického nástroje

a dalších externích zdrojů dat. Byly nalezeny odpovědi na klíčové zpravodajské otázky, které mohou firmě Robology pomoci při formulaci strategie pro vývoj konkrétního produktu.

Dále práce nastiňuje možnosti pro budoucí rozvoj a vylepšení analytického nástroje. Jsou představeny možnosti vývoje, jako je integrace funkcionality pro optimalizaci klíčových slov, podle kterých se aplikace hledají, vytvoření automatického reportingu, integrace placených externích dat či vývoj datového skladu pro ukládání historických dat.

Vytvořený analytický nástroj výrazně pomůže firmě Robology s analýzou a průzkumem tržního prostředí trhu s mobilními aplikacemi a stane se nedílnou součástí každého procesu analýzy konkurenčního prostředí. Webová aplikace, dashboard a celý zdrojový kód jsou veřejně přístupné a dostupné k vyzkoušení či pro vlastní implementaci řešení.

Použitá literatura

- [1] JAIN, Varsha a Vijay VISWANATHAN. The Usage and Application of Mobile Apps. In: [online]. 2015, s. 1242–1255. Dostupné z: doi:10.4018/978-1-4666-8239-9.ch100
- [2] SUNILA, Goray. The History of Mobile Apps and Evolution of Mobile Platforms. *The History of Mobile Apps and Evolution of Mobile Platforms* [online]. 28. listopad 2023 [vid. 2024-04-10]. Dostupné z: <https://webandcrafts.com/blog/history-of-mobile-apps>
- [3] DEGENHARD, J. Global: number of smartphone users 2014-2029. *Statista* [online]. [vid. 2024-04-11]. Dostupné z: <https://www.statista.com/forecasts/1143723/smartphone-users-in-the-world>
- [4] TAO, Kungpo a Paulette EDMUNDS. Mobile APPs and Global Markets. *Theoretical Economics Letters* [online]. 2018, **08**, 1510–1524. Dostupné z: doi:10.4236/tel.2018.88097
- [5] DIGITÁLNÍ A INFORMAČNÍ AGENTURA. Nová éra prokazování totožnosti. *eDoklady* [online]. [vid. 2024-04-11]. Dostupné z: <https://edoklady.gov.cz>
- [6] ČR, ÚZIS. EZKarta | NZIP. *NZIP.cz* [online]. [vid. 2024-04-11]. Dostupné z: <https://www.nzip.cz/ezkarta>
- [7] STATISTA RESEARCH DEPARTMENT. Global: mobile app revenue by segment 2019-2027. *Statista* [online]. 16. listopad 2023 [vid. 2024-04-10]. Dostupné z: <https://www.statista.com/forecasts/1262892/mobile-app-revenue-worldwide-by-segment>
- [8] LAURA CECI. Google Play Store: number of apps 2023. *Statista* [online]. 19. leden 2024 [vid. 2024-04-11]. Dostupné z: <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/>
- [9] LAURA CECI. Number of apps from the Apple App Store 2023. *Statista* [online]. 18. leden 2024 [vid. 2024-04-11]. Dostupné z: <https://www.statista.com/statistics/268251/number-of-apps-in-the-itunes-app-store-since-2008/>
- [10] LAURA CECI. Annual number of mobile app downloads worldwide 2023. *Statista* [online]. 5. duben 2024 [vid. 2024-04-10]. Dostupné z: <https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/>
- [11] STATISTA RESEARCH DEPARTMENT. Global: mobile app downloads by segment 2019-2027. *Statista* [online]. 16. listopad 2023 [vid. 2024-04-10]. Dostupné z: <https://www.statista.com/forecasts/1262881/mobile-app-download-worldwide-by-segment>
- [12] LI, Tong, Tong XIA, Huandong WANG, Zhen TU, Sasu TARKOMA, Zhu HAN a Pan HUI. Smartphone App Usage Analysis: Datasets, Methods, and Applications. *IEEE*

- Communications Surveys & Tutorials* [online]. 2022. Dostupné z: doi:10.1109/COMST.2022.3163176
- [13] LAURA CECI. App downloads by country 2022. *Statista* [online]. 29. srpen 2023 [vid. 2024-04-11]. Dostupné z: <https://www.statista.com/statistics/1287159/app-downloads-by-country/>
- [14] PRAKASH, Anand. 7 Most Popular Mobile App Distribution Platforms [online]. 29. září 2020 [vid. 2024-04-11]. Dostupné z: <https://www.appventurez.com/blog/mobile-app-distribution-platforms>
- [15] TANG, Ailie K. Y. Mobile App Monetization: App Business Models in the Digital Era. *International Journal of Innovation, Management and Technology* [online]. 2016, 224–227. ISSN 20100248. Dostupné z: doi:10.18178/ijimt.2016.7.5.677
- [16] APPEL, Gil, Barak LIBAI, Eitan MULLER a Ron SHACHAR. On the monetization of mobile apps. *International Journal of Research in Marketing* [online]. 2019, 37. Dostupné z: doi:10.1016/j.ijresmar.2019.07.007
- [17] COOBAN, Anna. Apple hit with landmark \$2 billion EU antitrust fine | CNN Business. *CNN* [online]. 4. březen 2024 [vid. 2024-04-11]. Dostupné z: <https://www.cnn.com/2024/03/04/tech/apple-europe-antitrust-fine-music-streaming/index.html>
- [18] LIONEL SUJAY VAILSHERY. Cross-platform mobile frameworks used by global developers 2022. *Statista* [online]. 1. červen 2023 [vid. 2024-04-11]. Dostupné z: <https://www.statista.com/statistics/869224/worldwide-software-developer-working-hours/>
- [19] MONTAG, Christian, Benjamin BECKER a Chunmei GAN. The Multipurpose Application WeChat: A Review on Recent Research. *Frontiers in Psychology* [online]. 2018, 9 [vid. 2024-04-11]. ISSN 1664-1078. Dostupné z: doi:10.3389/fpsyg.2018.02247
- [20] LAURA CECI. Most downloaded apps worldwide 2024. *Statista* [online]. 8. duben 2024 [vid. 2024-04-20]. Dostupné z: <https://www.statista.com/statistics/1448008/top-downloaded-mobile-apps-worldwide/>
- [21] UXCAM. Top 51 Important Mobile App KPIs to Measure Performance 2024. *Bluespace* [online]. [vid. 2024-04-15]. Dostupné z: <https://uxcam.com/blog/top-50-mobile-app-kpis>
- [22] LEE, Gunwoong a Raghu SANTANAM. Determinants of Mobile Apps' Success: Evidence from the App Store Market. *Journal of Management Information Systems* [online]. 2014, 31, 133–170. Dostupné z: doi:10.2753/MIS0742-1222310206
- [23] MOLNÁR, Zdeněk. *Competitive intelligence, aneb, Jak získat konkurenční výhodu*. 1. vydání (tištěné). B.m.: Oeconomica, 2012. ISBN 978-80-245-1908-1.
- [24] SAFEGRAPH. *The Ultimate Guide to Competitive Intelligence Research* | SafeGraph [online]. [vid. 2024-04-11]. Dostupné z: <https://safegraph.com/guides/competitive-intelligence>

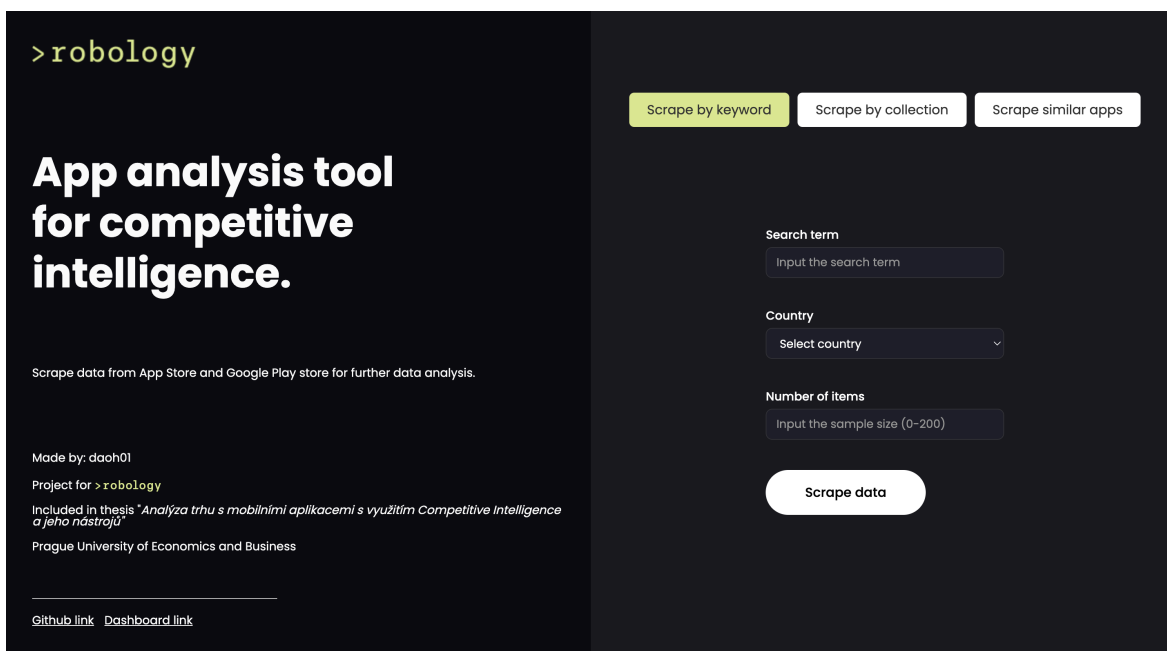
- [25] DAVID BARTON. What is web scraping? Complete guide to how it works. *Apify Blog* [online]. 24. srpen 2023 [vid. 2024-04-17]. Dostupné z: <https://blog.apify.com/what-is-web-scraping/>
- [26] PERSSON, Emil. *Evaluating tools and techniques for web scraping* [online]. 2019 [vid. 2024-04-17]. Dostupné z: <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-271206>
- [27] AMAZON. What is an API? - Application Programming Interface Explained - AWS. *Amazon Web Services, Inc.* [online]. [vid. 2024-04-21]. Dostupné z: <https://aws.amazon.com/what-is/api/>
- [28] HARKIRAN. What is Web Scraping and How to Use It? *GeeksforGeeks* [online]. 22. červen 2020 [vid. 2024-04-17]. Dostupné z: <https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/>
- [29] ZAMORA, Amber. Making Room For Big Data: Web Scraping and an Affirmative Right to Access Publicly Available Information Online. nedatováno.
- [30] SENSORTOWER. Sensor Tower | Responsibly Sourced Data. *SensorTower* [online]. [vid. 2024-04-17]. Dostupné z: <https://sensortower.com/responsibly-sourced-data>
- [31] ANDREA KNEZOVIC. Top 12 Mobile Intelligence Providers [Pricing Included]. *Udonis Mobile Marketing Agency* [online]. 20. únor 2024 [vid. 2024-04-17]. Dostupné z: <https://www.blog.udonis.co/mobile-marketing/top-mobile-intelligence-providers>
- [32] SATISH THANDRA. *The Intelligence Cycle and its Importance in the Competitive Intelligence Practices - Thandra Consulting* [online]. [vid. 2024-04-16]. Dostupné z: <https://www.thandraconsulting.com/blogs/post/the-intelligence-cycle-and-its-importance-in-the-competitive-intelligence-practices>
- [33] OTA NOVOTNÝ, JAN POUR, a DAVID SLÁNSKÝ. *Business intelligence : jak využít bohatství ve vašich datech*. 1. vyd. B.m.: Grada, 2005. Management v informační společnosti. ISBN 978-80-247-1094-5.
- [34] SHARDA, Ramesh, Dursun DELEN, Efraim TURBAN, Janine E ARONSON, Ting-Peng LIANG a David KING. *Business intelligence, analytics, and data science: a managerial perspective*. 2019. ISBN 978-93-5306-702-1.
- [35] HARIHARAN, Kunnathuvalappil a NAVEEN. *Data Sources for Business Intelligence* [online]. SSRN Scholarly Paper. 30. listopad 2018 [vid. 2024-04-24]. Dostupné z: <https://papers.ssrn.com/abstract=3914444>
- [36] GOOGLE CLOUD. What is ETL? *Google Cloud* [online]. [vid. 2024-04-24]. Dostupné z: <https://cloud.google.com/learn/what-is-etl>
- [37] BENTLEY, Drew, ed. *Business Intelligence and Analytics*. B.m.: Larsen and Keller Education, 2017. ISBN 978-1-63549-056-5.
- [38] SLÁNSKÝ, David. *Data a analytika pro 21. století*. 2018. ISBN 978-80-88260-25-7.
- [39] NODE.JS. *Node.js — Introduction to Node.js* [online]. [vid. 2024-04-19]. Dostupné z: <https://nodejs.org/en/learn/getting-started/introduction-to-nodejs>

- [40] W3SCHOOLS. *What is AJAX* [online]. [vid. 2024-04-21]. Dostupné z: https://www.w3schools.com/whatis/whatis_ajax.asp
- [41] DEVLIN, Jacob, Ming-Wei CHANG, Kenton LEE a Kristina TOUTANOVA. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* [online]. B.m.: arXiv. 24. květen 2019 [vid. 2024-04-25]. Dostupné z: doi:10.48550/arXiv.1810.04805. arXiv:1810.04805 [cs]
- [42] CAMERON HASHEMI-POUR. What is the BERT language model? | Definition from TechTarget.com. *Enterprise AI* [online]. [vid. 2024-04-25]. Dostupné z: <https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model>

Přílohy

Příloha A: Odkaz na webovou aplikaci

Příloha A obsahuje odkaz na webovou aplikaci, která vznikla v praktické části této diplomové práce.



[Odkaz na webovou aplikaci zde](#)

Příloha B: Odkaz na repozitář se zdrojovým kódem aplikace

Příloha B obsahuje odkaz na repozitář na platformě Github, který obsahuje zdrojový kód vytvořené webové aplikace.

[Odkaz na repozitář zde](#)

Příloha C: Odkaz na dashboard

Příloha C obsahuje odkaz na dashboard na platformě Google Looker Studio, který v rámci diplomové práce sloužil pro analýzu trhu s mobilními aplikacemi.



[Odkaz na dashboard zde](#)