

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET ORGANIZACIJE I INFORMATIKE**  
**V A R A Ž D I N**

**Luka Miroić**

**Analiza android aplikacija u Google Play**  
**Store trgovini**

**PROJEKT IZ KOLEGIJA SKLADIŠTA PODATAKA I POSLOVNA**  
**INTELIGENCIJA**

**Varaždin, 2021.**

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET ORGANIZACIJE I INFORMATIKE**  
**V A R A Ž D I N**

**Luka Miroić**

**Matični broj: 44847/16–R**

**Studij: *Organizacija poslovnih sustava***

**Analiza android aplikacija u Google Play Store trgovini**  
**PROJEKT IZ KOLEGIJA SKLADIŠTA PODATAKA I POSLOVNA INTELIGENCIJA**

**Mentor:**

Prof. dr. sc. Kornelije Rabuzin

**Varaždin, lipanj 2021**

## **Sažetak**

Cilj ovog rada je prikaz izrade skladišta podataka te analiza podataka u tom skladištu koji su vezani za android aplikacije, opis same aplikacije (veličina, naziv, verzija androida potrebna za instalaciju i sl.), te podaci vezani za iste kao što su broj preuzimanja, ocjena korisnika i dr. Analiza podataka se provodi nakon što su provedeni koraci izrade skladišta podataka, ETL procesi i čišćenje podataka te na kraju priprema samih podataka za izradu izvještaja. Za provedbu tih koraka sam koristio sljedeće alate: MySQL Server, MySQL Workbench te Microsoft PowerBI Desktop.

**Ključne riječi:** PowerBI, MySQL, ETL, Izvještaji, Google Play Store, aplikacije, analiza

# Sadržaj

Sadržaj .....	iv
1. Uvod .....	1
2. Opis projekta.....	2
2.1. Korištene tehnologije.....	2
2.1.1. MySQL Server .....	2
2.1.2. MySQL Workbench .....	3
2.1.3. Microsoft Power BI .....	3
3. Kreiranje skladišta podataka.....	3
3.1. Dataset android aplikacija .....	3
3.2. Princip ETL-a nad podacima dataseta .....	5
4. Model zvijezde i kreiranje tablica u skladištu .....	7
4.1. Dimenzijska tablica android_version .....	8
4.2. Dimenzijska tablica category .....	8
4.3. Dimenzijska tablica rating .....	9
4.4. Dimenzijska tablica size .....	9
4.5. Dimenzijska tablica content_rating .....	10
4.6. Dimenzijska tablica app .....	11
4.7. UPDATE upiti nad tablicom dataset .....	12
4.8. ALTER upiti nad tablicom dataset .....	12
4.9. Činjenična tablica apps_facts .....	12
5. Analiza pomoću Power BI Desktop alata.....	14
5.1. Izvještaji u Power BI alatu .....	15
6. Zaključak .....	20
Popis literature.....	21
Popis slika .....	22

# 1. Uvod

Mnoga poduzeća u današnje vrijeme vode bitku sa velikom količinom podataka koja, barem za većinu poduzeća u RH, nisu informatizirani nego su spremljeni u papirnatom obliku u arhivima. U današnje vrijeme je raspolaganje bitnim i pravovremenim informacijama, vrlo bitna prednost na tržištu zbog toga što je konkurencija na tržištu sve veća, upravo iz tog razloga poduzeća koja analiziraju svoje podatke i na njima temelje vlastite strateške odluke se nalaze u bitnoj prednosti u usporedbi sa poduzećima koja ne analiziraju podatke koje generira njihovo poslovanje.

Nadalje čak ni transakcijski sustavi, u velikim poduzećima sa velikom količinom različitih tipova podataka spremljenim na različitim mjestima, nisu u stanju isporučiti tražene podatke u kratkom vremenu. Iz tog razloga su razvijena skladišta podataka, kod kojih se svi podaci nalaze na jednom mjestu te tako unapređuje konkurentnost poduzeća na tržištu. Danas postoji mnogo poslužitelja skladišta podataka i programskih alata za analizu podataka, a takva integrirana softverska rješenja koja sadrže sve bitne komponente sustava za analizu podataka se nazivaju platforme poslovne inteligencije (eng. *Business Intelligence platforms*). Takav softver je dostupan na tržištu po različitim cijenama, ovisno o tome koje mogućnosti nude, te poduzeće mora odabrati softversko rješenje koje zadovoljava njihove potrebe uz troškove koji se uklapaju u njihov budžet. Igor Mekterović i Lj. Brkić govore kako su skladišta podataka subjektivno orijentiran, integriran, postojan i vremenski različit skup podataka koji služi kao potpora odlučivanju.[3]

Budući da smo definirali neke osnovne pojmove vezane za projekt, pa sada možemo reći kako smo u svrhu ovog projekta radili upravo skladište podataka iz skupa podataka preuzetog sa Kaggle.com web stranice te da smo na temelju tog skladišta podataka izradili izvještaje pomoću kojih se radi analiza podataka tog skladišta. Najprije je prikazan postupak izrade samog skladišta podataka, zajedno sa potrebnim alatima za provedbu istog. Nakon toga je prikazan postupak učitavanja podataka u Microsoft Power BI alat te prikaz izvještaja koji su kreirani za podatke unutar skladišta.

## 2. Opis projekta

U svrhu izrade ovog projekta fokusirali smo se na izradu skladišta podataka sa podacima vezanim za android aplikacije na Google Play Store-u. Google Play Store je aplikacija tvrtke Google na kojoj developeri mobilnih aplikacija mogu prodati svoj proizvod ili dati besplatno na korištenje svim korisnicima. Neki osnovni cilj bio je dobiti neki uvid u to koje aplikacije su najpopularnije, koliko ima aplikacija za određenu kategoriju i sl.

U preuzetom skupu podataka (eng. *Dataset*) imamo sljedeće podatke koje ćemo promatrati:

- App: naziv aplikacije
- Category: naziv kategorije u koju spada određena aplikacija
- Rating: prosječna ocjena korisnika za određenu aplikaciju
- Reviews: broj ljudi koji su dali ocjenu za određenu aplikaciju
- Size: veličina određene aplikacije
- Installs: broj preuzimanja određene aplikacije
- Type: tip aplikacije (dali se plaća ili je besplatna)
- Price: cijena aplikacije ukoliko se određena aplikacija plaća
- Content\_Rating: govori nam o tome za koga je aplikacija namijenjena
- Genres: žanr aplikacije, gotovo identično kao i kategorija aplikacije
- Last\_Updated: govori nam kada je aplikacija posljednji put ažurirana
- Current\_Ver: trenutna verzija određene aplikacije
- Android\_Ver: minimalna verzija android sustava na uređaju za rad određene aplikacije

### 2.1. Korištene tehnologije

Za izradu ovog projekta korištene su sljedeće tehnologije: MySQL Server, MySQL Workbench te Microsoft Power BI Desktop. Baza podataka je kreirana na lokalnom serveru pomoću MySQL Server aplikacije, samo skladište podataka izrađeno pomoću MySQL Workbench alata dok su izvještaji kreirani u alatu Microsoft Power BI Desktop.

#### 2.1.1. MySQL Server

MySQL je jedan od najrasprostranjenijih sustava za upravljanje bazom podataka otvorenog koda na tržištu. MySQL su relacijske baze podataka koje su idealne za skladištenje i pretraživanje velikih količina podataka. MySQL Server je alat koji je u vlasništvu tvrtke Oracle kao i cijeli MySQL sustav, njegovom instalacijom kod pokretanja računala se kreira lokalni server na adresi 127.0.0.1 sa portom 3306 koji je defaultni port za MySQL protokol.

### 2.1.2.MySQL Workbench

Za izradu skladišta podataka korišten je alat MySQL Workbench koji je zapravo alat za dizajn baze podataka koji integrira razvoj SQL-a, administraciju, dizajn, stvaranje i održavanje u jedinstveno integrirano sučelje za MySQL baze podataka. Kod MySQL Workbencha se sve radi u grafičkom sučelju, odnosno mogu se, ali i ne moraju, pisati SQL naredbe putem upita za upravljanje bazom podataka. Ovaj alat dolazi u komercijalnom i besplatnom izdanju te sa sigurnošću možemo reći da je jedan od najpopularnijih MySQL alata.

### 2.1.3.Microsoft Power BI

Da bi kreirali izvještaje iz podataka unutar skladišta podataka korišten je Microsoft Power BI Desktop alat koji omogućuje korisniku unos podataka iz raznih izvora od kojih je jedan i MySQL baza podataka, koja je i korištena. Power BI omogućuju transformaciju podataka ukoliko za to postoji potreba prije samo učitavanja, omogućuje pregled podataka, organiziranje, kreiranje izvještaja iz učitanih podataka i još mnogo različitih funkcionalnosti koje u kraćem vremenu u kojem sam radio sa njim nije bilo moguće primijeniti. Microsoft Power BI alat dolazi u više inačica: Power BI Premium, Power BI Pro, Power BI Desktop i Power BI Mobile od koje je besplatan jedino Power BI Desktop kojeg sam i koristio za izradu ovog projekta.

## 3. Kreiranje skladišta podataka

U ovom dijelu rada ćemo se fokusirati na skup podataka (eng. *Dataset*) kojeg sam pronašao na web stranici Kaggle.com te kako ćemo primijeniti ETL princip (eng. *Extract Transform Load*) da bi podatke mogli učitati u skladište podataka te da bi isti bili dostupni za analizu i kreiranje izvještaja u Power BI Desktop alatu.

### 3.1. Dataset android aplikacija

Skup podataka koji je izabran za izradu ovog projekta je, kao što je već navedeno iznad u radu, popis android aplikacija i njihovih atributa koje se mogu naći na Google Play Store aplikaciji. Nadalje taj skup podataka ćemo koristiti kako bi izvršili analizu nad istim i dobili odgovore na pitanja kao što su koje aplikacije su najpopularnije, koje su najbolje ocijenjene i sl.

Preuzeti skup podataka dolazi u .csv formatu te sadrži 10842 zapisa, odnosno 10842 različite aplikacije, međutim iz nekog razloga od 9 ili 10 testiranih skupova podataka vezanih za android aplikacije, pomoću MySQL Workbencha sam uspio učitati u bazu samo oko 300 zapisa što je ujedno i najviše od svih testiranih.

A1	App,Category,Rating,Reviews,Size,Installs,Type,Price,Content Rating,Genres,Last Updated,Current Ver,Android Ver																				
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	App,Category,Rating,Reviews,Size,Installs,Type,Price,Content Rating,Genres,Last Updated,Current Ver,Android Ver																				
2	Photo Editor & Candy Camera & Grid & ScrapBook,ART_AND_DESIGN,4.1,159,19M,"10,000+",Free,0,Everyone,Art & Design,"January 7, 2018",1.0.0,4.0.3 and up																				
3	Coloring t Pretend Play,"January 15, 2018",2.0.0,4.0.3 and up																				
4	U Launcher Lite æ" FREE Live Cool Themes, Hide Apps,ART_AND_DESIGN,4.7,87510,8.7M,"5,000,000+",Free,0,Everyone,Art & Design,"August 1, 2018",1.2.4,4.0.3 and up																				
5	Sketch - Draw & Paint,ART_AND_DESIGN,4.5,215644,25M,"50,000,000+",Free,0,Teen,Art & Design,"June 8, 2018",Varies with device,4.2 and up																				
6	Pixel Draw Creativity,"June 20, 2018",1.1,4.4 and up																				
7	Paper flowers instructions,ART_AND_DESIGN,4.4,167,5.6M,"50,000+",Free,0,Everyone,Art & Design,"March 26, 2017",1.0,2.3 and up																				
8	Smoke Effect Photo Maker - Smoke Editor,ART_AND_DESIGN,3.8,178,19M,"50,000+",Free,0,Everyone,Art & Design,"April 26, 2018",1.1,4.0.3 and up																				
9	Infinite Painter,ART_AND_DESIGN,4.1,36815,29M,"1,000,000+",Free,0,Everyone,Art & Design,"June 14, 2018",6.1.61.1,4.2 and up																				
10	Garden Coloring Book,ART_AND_DESIGN,4.4,13791,33M,"1,000,000+",Free,0,Everyone,Art & Design,"September 20, 2017",2.9.2,3.0 and up																				
11	Kids Paint Creativity,"July 3, 2018",2.8,4.0.3 and up																				
12	Text on Photo - Fontee,ART_AND_DESIGN,4.4,13880,28M,"1,000,000+",Free,0,Everyone,Art & Design,"October 27, 2017",1.0,4.4.1 and up																				
13	Name Art Photo Editor - Focus n Filters,ART_AND_DESIGN,4.4,8788,12M,"1,000,000+",Free,0,Everyone,Art & Design,"July 31, 2018",1.0.15,4.0 and up																				
14	Tattoo Name On My Photo Editor,ART_AND_DESIGN,4.2,44829,20M,"10,000,000+",Free,0,Teen,Art & Design,"April 2, 2018",3.8,4.1 and up																				
15	Mandala Coloring Book,ART_AND_DESIGN,4.6,4326,21M,"100,000+",Free,0,Everyone,Art & Design,"June 26, 2018",1.0,4.4.4 and up																				
16	3D Color Pixel by Number - Sandbox Art Coloring,ART_AND_DESIGN,4.4,1518,37M,"100,000+",Free,0,Everyone,Art & Design,"August 3, 2018",1.2.3,2.3 and up																				
17	Learn To Draw Kawaii Characters,ART_AND_DESIGN,3.2,55,2.7M,"5,000+",Free,0,Everyone,Art & Design,"June 6, 2018",NaN,4.2 and up																				
18	Photo Designer - Write your name with shapes,ART_AND_DESIGN,4.7,3632,5.5M,"500,000+",Free,0,Everyone,Art & Design,"July 31, 2018",3.1,4.1 and up																				
19	350 Diy Room Decor Ideas,ART_AND_DESIGN,4.5,27,17M,"10,000+",Free,0,Everyone,Art & Design,"November 7, 2017",1.0,2.3 and up																				
20	FlipaClip - Cartoon animation,ART_AND_DESIGN,4.3,194216,39M,"5,000,000+",Free,0,Everyone,Art & Design,"August 3, 2018",2.2.5,4.0.3 and up																				
21	Ibis Paint X,ART_AND_DESIGN,4.6,224399,31M,"10,000,000+",Free,0,Everyone,Art & Design,"July 30, 2018",5.5,4.4.1 and up																				
22	Logo Maker - Small Business,ART_AND_DESIGN,4.0,450,14M,"100,000+",Free,0,Everyone,Art & Design,"April 20, 2018",4.0,4.1 and up																				
23	Boys Photo Editor - Six Pack & Men's Suit,ART_AND_DESIGN,4.1,654,12M,"100,000+",Free,0,Everyone,Art & Design,"March 20, 2018",1.1,4.0.3 and up																				
24	Superheroes Wallpapers   4K Backgrounds,ART_AND_DESIGN,4.7,7699,4.2M,"500,000+",Free,0,Everyone 10+,Art & Design,"July 12, 2018",2.2.6.2,4.0.3 and up																				
25	Mcqueen Action & Adventure,"March 7, 2018",1.0.0,4.1 and up																				
26	HD Mickey Minnie Wallpapers,ART_AND_DESIGN,4.7,118,23M,"50,000+",Free,0,Everyone,Art & Design,"July 7, 2018",1.1.3,4.1 and up																				
27	Harley Quinn wallpapers HD,ART_AND_DESIGN,4.8,192,6.0M,"10,000+",Free,0,Everyone,Art & Design,"April 25, 2018",1.5,3.0 and up																				
28	Colorfit - I Creativity,"October 11, 2017",1.0.8,4.0.3 and up																				
29	Animated Photo Editor,ART_AND_DESIGN,4.1,203,6.1M,"100,000+",Free,0,Everyone,Art & Design,"March 21, 2018",1.03,4.0.3 and up																				
30	Pencil Sketch Drawing,ART_AND_DESIGN,3.9,136,4.6M,"10,000+",Free,0,Everyone,Art & Design,"July 12, 2018",6.0,2.3 and up																				
31	Easy Realistic Drawing Tutorial,ART_AND_DESIGN,4.1,223,4.2M,"100,000+",Free,0,Everyone,Art & Design,"August 22, 2017",1.0,2.3 and up																				
32	Pink Silver Bow Keyboard Theme,ART_AND_DESIGN,4.2,1120,9.2M,"100,000+",Free,0,Everyone,Art & Design,"July 12, 2018",6.7.12.2018,4.0.3 and up																				
33	Art Drawing Ideas,ART_AND_DESIGN,4.1,227,5.2M,"50,000+",Free,0,Everyone,Art & Design,"May 31, 2018",1.2,2.3 and up																				
34	Anime Manga Coloring Book,ART_AND_DESIGN,4.5,5035,11M,"100,000+",Free,0,Everyone,Art & Design,"July 19, 2018",2.20,4.0 and up																				
35	Easy Origami Ideas,ART_AND_DESIGN,4.2,1015,11M,"100,000+",Free,0,Everyone,Art & Design,"January 6, 2018",1.1.0,4.1 and up																				
36	I creative Idea,ART_AND_DESIGN,4.7,353,4.2M,"10,000+",Free,0,Teen,Art & Design,"April 27, 2018",1.6,4.1 and up																				
37	How to draw I advice and Cat Noir,ART_AND_DESIGN,3.8,564,9.2M,"100,000+",Free,0,Everyone,Art & Design,"July 11, 2018",2.1.4.1 and up																				

Slika 1. Prikaz skupa podataka u .csv formatu

Na slici 1 možemo vidjeti popis prvih 40-ak aplikacija te attribute za svaku aplikaciju, pa tako možemo vidjeti da se svaki zapis odnosno redak sastoji od sljedećih stupaca ili atributa: App, Category, Rating, Reviews, Size, Installs, Type, Price, Content\_Rating, Genres, Last\_Updated, Current\_Ver i Android\_Ver. U MySQL Workbench alat smo učitali sve podatke međutim u daljnjoj analizi dataseta uvidjeli smo da postoje irelevantni atributi za koje su sve vrijednosti jednake pa su tako oni nisu stavljani niti u činjeničnu niti u dimenzijske tablice npr. Type i Price.



## 3.2. Princip ETL-a nad podacima dataseta

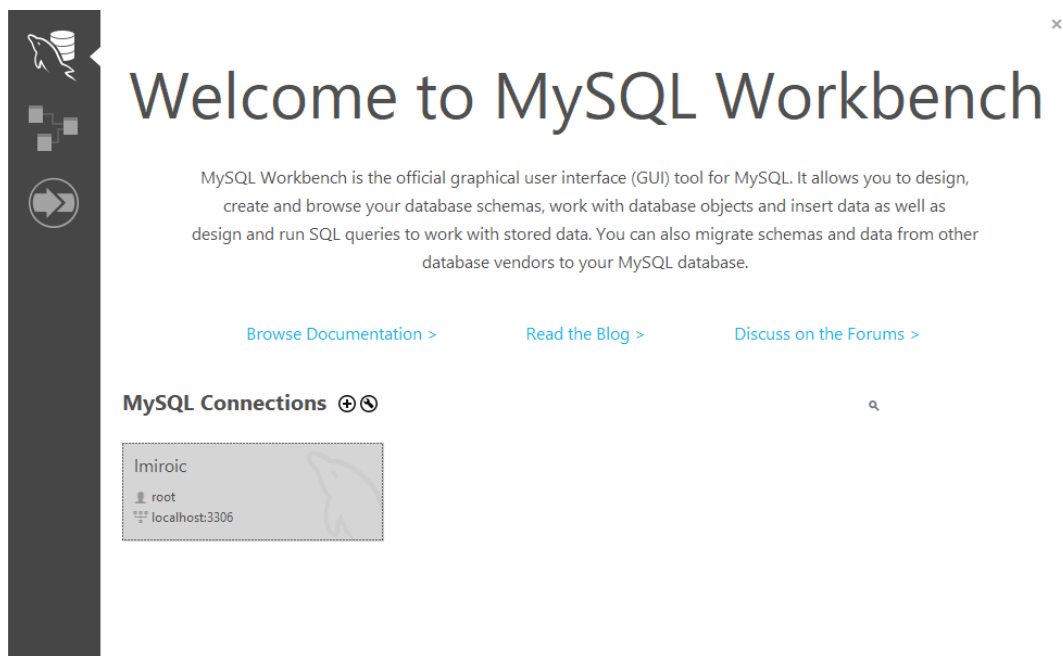
ETL ili Extract, Transform, Load je termin koji se koristi za proces koji prenosi podatke iz nekog izvora do skladišta podataka odnosno, kao što i sam naziv govori, proces kod kojeg se podaci ekstrahiraju iz nekog izvora podataka, transformiraju te učitavaju u skladište podataka.[1]

Ekstrakcija je prvi korak ETL-a, a definiramo ga kao postupak vađenja podataka iz izvora podataka za naše skladište podataka, taj izvor može biti strukturiran ili nestrukturiran, iako je u većini slučajeva radi o strukturiranim podacima u digitalnom obliku kao što je bilo u našem slučaju (.csv format). [3]

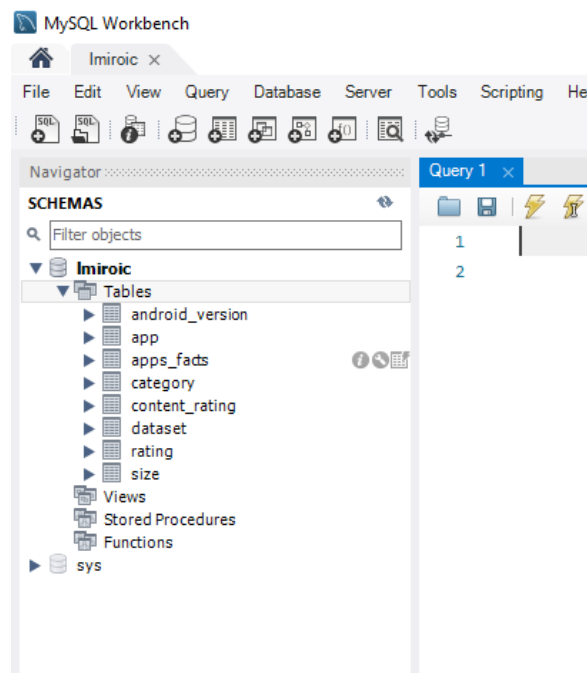
Transformacija podataka je postupak kojim ekstrahirane podatke prilagođavamo za unos u skladište podataka, to podrazumijeva zamjene vrijednosti, deriviranje novih vrijednosti, transformacije slova, generiranje ključeva i sl.[4]

Učitavanje podataka je posljednji korak ETL-a te se u njemu transformirani podaci unose u finalnu tablicu u skladištu podataka.[4]

U svrhu ovog projekta je bilo potrebno osposobiti MySQL Workbench alat te ga spojiti na lokalni server koji sam ja nazvao Imiroic, što se može vidjeti na slici 2. i zatim kreirati sve potrebne tablice za provedbu ETL-a. Na slici 3 možemo vidjeti popis kreiranih tablica unutar lokalnog servera.



Slika 2. Prikaz servera na koji se spajamo



Slika 3. Prikaz baze podataka

Nakon što smo napravili bazu podataka, desni klik na nju zatim odaberemo opciju Table Dana Import Wizard. Zatim se otvara prozor u kojem odaberemo csv datoteku koju želimo učitati, odaberemo da se kreira nova tablica te sam ju ja nazvao dataset kao što se i može vidjeti na slici iznad. Kod učitavanja podataka odabrao sam latin 1(iso8859-1), svakom stupcu je odmah određen i tip vrijednosti te su zatim učitani podaci iz csv datoteke u tablicu dataset (međutim samo njih 300-ak što je dovoljno za potrebu provedbe ove analize). Na sljedećim slikama 4. i 5. možemo vidjeti strukturu tablice i učitane podatke u dataset tablicu.




Table Name:

dataset

Charset/Collation:

utf8mb4

utf8mb4\_0900\_ai\_ci

Engine:

InnoDB

Schema:

Imiroic

Comments:

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
App	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Category	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Rating	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Reviews	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Size	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Installs	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Type	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Price	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Content_Rating	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Genres	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
LastUpdated	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Current_Ver	TEXT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
Android_Ver	INT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL

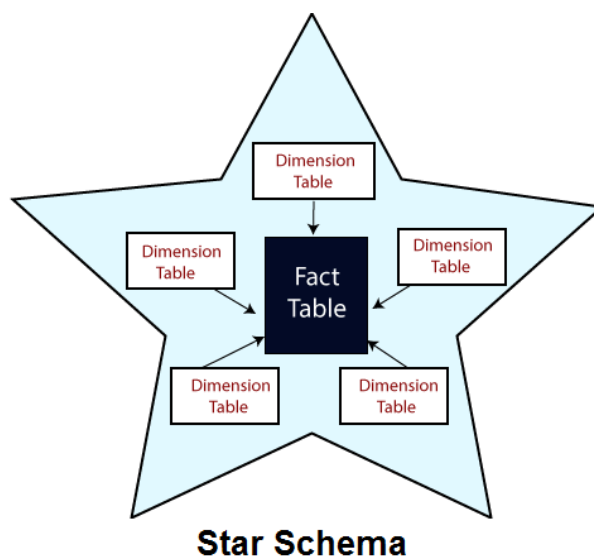
Slika 4. Prikaz atributa tablice dataset

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content_Rating	Genres	Last Updated	Current_Ver	Android_Ver
1	1	1	159	1	10,000+	Free	0	1	Art & Design	January 7, 2018	1.0.0	1
2	1	2	967	2	500,000+	Free	0	1	Art & Design;Pretend Play	January 15, 2018	2.0.0	1
3	1	3	87510	3	5,000,000+	Free	0	1	Art & Design	August 1, 2018	1.2.4	1
4	1	4	215644	4	50,000,000+	Free	0	2	Art & Design	June 8, 2018	Varies with device	2
5	1	5	967	5	100,000+	Free	0	1	Art & Design;Creativity	June 20, 2018	1.1	3
6	1	6	167	6	50,000+	Free	0	1	Art & Design	March 26, 2017	1.0	4
7	1	7	178	1	50,000+	Free	0	1	Art & Design	April 26, 2018	1.1	1
8	1	1	36815	7	1,000,000+	Free	0	1	Art & Design	June 14, 2018	6.1.61.1	2
9	1	6	13791	8	1,000,000+	Free	0	1	Art & Design	September 20, 2017	2.9.2	5
10	1	3	121	9	10,000+	Free	0	1	Art & Design;Creativity	July 3, 2018	2.8	1
11	1	6	13880	10	1,000,000+	Free	0	1	Art & Design	October 27, 2017	1.0.4	6
12	1	6	8788	11	1,000,000+	Free	0	1	Art & Design	July 31, 2018	1.0.15	7
13	1	8	44829	12	10,000,000+	Free	0	2	Art & Design	April 2, 2018	3.8	6
14	1	9	4326	13	100,000+	Free	0	1	Art & Design	June 26, 2018	1.0.4	3
15	1	6	1518	14	100,000+	Free	0	1	Art & Design	August 3, 2018	1.2.3	4
16	1	10	55	15	5,000+	Free	0	1	Art & Design	June 6, 2018	NaN	2
17	1	3	3632	16	500,000+	Free	0	1	Art & Design	July 31, 2018	3.1	6
18	1	4	27	17	10,000+	Free	0	1	Art & Design	November 7, 2017	1.0	4
19	1	5	194216	18	5,000,000+	Free	0	1	Art & Design	August 3, 2018	2.2.5	1
20	1	9	224399	19	10,000,000+	Free	0	1	Art & Design	July 30, 2018	5.5.4	6
21	1	11	450	2	100,000+	Free	0	1	Art & Design	April 20, 2018	4.0	6
22	1	1	654	11	100,000+	Free	0	1	Art & Design	March 20, 2018	1.1	1
23	1	3	7699	20	500,000+	Free	0	3	Art & Design	July 12, 2018	2.2.6.2	1
24	1	3	118	21	50,000+	Free	0	1	Art & Design	July 7, 2018	1.1.3	6
25	1	12	192	22	10,000+	Free	0	1	Art & Design	April 25, 2018	1.5	5
26	1	3	20260	4	500,000+	Free	0	1	Art & Design;Creativity	October 11, 2017	1.0.8	1

Slika 5. Prikaz tablice dataset

## 4. Model zvijezde i kreiranje tablica u skladištu

U ovom poglavlju ćemo definirati dimenzijske tablice i činjeničnu tablicu nad kojima ćemo vršiti analizu podataka i prikazivati izvještaje. Model zvijezda je klasična karakteristika skladišta podataka te je u tu svrhu potrebno definirati iznad navedene dimenzijske i činjenične tablice upravo zbog toga što svaki model sadrži jednu tablicu, s kompozitnim primarnim ključem ili činjenična tablica (eng. *Fact table*) i niza manjih tablica koje se nazivaju dimenzijske tablice (eng. *Dimension table*). Pa tako možemo reći da dimenzijske tablice daju odgovor na pitanje prema čemu mjerimo, a činjenične tablice daju odgovor na pitanje što mjerimo.[2]



Slika 6. Model zvijezde [5]

## 4.1. Dimenzijska tablica android\_version

Dimenzijska tablica android\_version se sastoji od dva atributa *idandroid\_version* i *android\_version*. Primarni ključ tablice je stavljen *Auto increment*, a *android\_version* je tipa TEXT, a da bi ju napunili podacima trebalo je napisati sljedeći upit:

```
INSERT INTO android_version(android_version) SELECT DISTINCT Android_Ver FROM dataset;
```

	idandroid_version	android_version
▶	1	4.0.3 and up
	2	4.2 and up
	3	4.4 and up
	4	2.3 and up
	5	3.0 and up
	6	4.1 and up
	7	4.0 and up
	8	2.3.3 and up
	9	Varies with device
	10	2.2 and up
	11	5.0 and up
	12	6.0 and up
	13	1.6 and up
	14	1.5 and up
	15	2.1 and up
	16	7.0 and up
	17	4.3 and up
*	NULL	NULL

Slika 7. Dimenzijska tablica android\_version

## 4.2. Dimenzijska tablica category

Dimenzijska tablica category se sastoji od dva atributa *idcategory* i *category*. Primarni ključ je i u ovoj tablici stavljen *auto increment*, a *category* kao TEXT, a da bi se tablica napunila podacima potrebno je napisati sljedeći upit:

```
INSERT INTO category(category) SELECT DISTINCT Category FROM dataset;
```

	idcategory	category
▶	1	ART_AND_DESIGN
	2	AUTO_AND_VEHICLES
	3	BEAUTY
	4	BOOKS_AND_REFERENCE
	5	BUSINESS
*	NULL	NULL

Slika 8. Dimenzijska tablica category

### 4.3. Dimenzijska tablica rating

Dimenzijska tablica rating se također sastoji od dva atributa: *idrating* i *rating*. I u ovoj tablici je primarni ključ stavljen kao *auto increment*, a *rating* kao FLOAT, a da bi se napunila tablica sa podacima potrebno je napisati sljedeći upit:

**INSERT INTO** rating(rating) **SELECT DISTINCT** Rating **FROM** dataset;

	idrating	rating
▶	1	4.1
	2	3.9
	3	4.7
	4	4.5
	5	4.3
	6	4.4
	7	3.8
	8	4.2
	9	4.6
	10	3.2
	11	4
	12	4.8
	13	4.9
	14	3.6
	15	3.7
	16	3.3
	17	3.4
	18	3.5
	19	3.1
*	NULL	NULL

Slika 9. Dimenzijska tablica rating

### 4.4. Dimenzijska tablica size

Dimenzijska tablica size se kao i prijašnje tablice sastoji od dva atributa: *idsize* i *size* te je i u ovoj tablici primarni ključ stavljen kao *auto increment*, a *size* kao TEXT, a da bi se tablica napunila sa podacima potrebno je napisati sljedeći upit:

**INSERT INTO** size(size) **SELECT DISTINCT** Size **FROM** dataset;

	idsize	size
▶	1	19M
	2	14M
	3	8.7M
	4	25M
	5	2.8M
	6	5.6M
	7	29M
	8	33M
	9	3.1M
	10	28M
	11	12M
	12	20M
	13	21M
	14	37M
	15	2.7M
	16	5.5M
	17	17M
	18	39M
	19	31M
	20	4.2M
	21	23M
	22	6.0M
	23	6.1M
	24	4.6M
	25	9.2M

Slika 10. Dimenzijska tablica size

## 4.5. Dimenzijska tablica content\_rating

Dimenzijska tablica content\_rating se sastoji od dva atributa: *idcontent\_rating* i *content\_rating*. Primarni ključ je kao i u prethodnim tablicama stavljen kao *auto increment*, a *content\_rating* kao FLOAT, a da bi napunili tablicu podacima potrebno je napisati sljedeći upit: **INSERT INTO** content\_rating(content\_rating) **SELECT DISTINCT** Content\_Rating **FROM** dataset;

Result Grid     Filter Rows: <input type="text"/>		
	idcontent_rating	content_rating
▶	1	Everyone
	2	Teen
	3	Everyone 10+
	4	Mature 17+
*	NULL	NULL

Slika 11. Dimenzijska tablica content\_rating

## 4.6. Dimenzijska tablica app

Dimenzijska tablica app se sastoji od četiri atributa: *idapp*, *app*, *installs*, *reviews*. I u ovoj tablici je primarni ključ stavljen *auto increment*, *app* je TEXT, kao i *installs* dok je *reviews* INT, a da bi popunili ovu tablicu moramo napisati sljedeći upit:

**INSERT INTO** app(app,installs,reviews) **SELECT** App,Installs,Reviews **FROM** dataset;

	idapp	app	installs	reviews
▶	1	Photo Editor & Candy Camera & Grid & ScrapBook	10,000+	159
	2	Coloring book moana	500,000+	967
	3	U Launcher Lite â€” FREE Live Cool Themes, Hid...	5,000,000+	87510
	4	Sketch - Draw & Paint	50,000,000+	215644
	5	Pixel Draw - Number Art Coloring Book	100,000+	967
	6	Paper flowers instructions	50,000+	167
	7	Smoke Effect Photo Maker - Smoke Effect Photo Maker - Smoke Editor		
	8	Infinite Painter	1,000,000+	36815
	9	Garden Coloring Book	1,000,000+	13791
	10	Kids Paint Free - Drawing Fun	10,000+	121
	11	Text on Photo - Fontee	1,000,000+	13880
	12	Name Art Photo Editor - Focus n Filters	1,000,000+	8788
	13	Tattoo Name On My Photo Editor	10,000,000+	44829
	14	Mandala Coloring Book	100,000+	4326
	15	3D Color Pixel by Number - Sandbox Art Coloring	100,000+	1518
	16	Learn To Draw Kawaii Characters	5,000+	55
	17	Photo Designer - Write your name with shapes	500,000+	3632
	18	350 Diy Room Decor Ideas	10,000+	27
	19	FlipaClip - Cartoon animation	5,000,000+	194216
	20	ibis Paint X	10,000,000+	224399
	21	Logo Maker - Small Business	100,000+	450
	22	Boys Photo Editor - Six Pack & Men's Suit	100,000+	654
	23	Superheroes Wallpapers   4K Backgrounds	500,000+	7699
	24	HD Mickey Minnie Wallpapers	50,000+	118
	25	Harley Quinn wallpapers HD	10,000+	192

Slika 12. Dimenzijska tablica app

Nakon što smo kreirali sve dimenzijske tablice, moramo još definirati i činjeničnu tablicu i popuniti ju vrijednostima, ali prije toga je potrebno ažurirati dataset. Najprije će biti navedeni upiti za update nad datasetom, a nakon toga alter tablice dataseta u tip podataka INT zbog vanjskih ključeva kako bi sve uspješno povezali sa dimenzijskim tablicama. Nadalje nakon update i alter upita će se popuniti i činjenična tablica.

## 4.7. UPDATE upiti nad tablicom dataset

**UPDATE** dataset d **INNER JOIN** app a on d.App = a.app

**SET** d.App=a.idapp;

**UPDATE** dataset d **INNER JOIN** android\_version av on d.Android\_Ver = av.android\_version

**SET** d.Android\_Ver=av.idandroid\_version;

**UPDATE** dataset d **INNER JOIN** category c on d.Category = c.category

**SET** d.Category=c.idcategory;

**UPDATE** dataset d **INNER JOIN** rating r on d.Rating = r.rating

**SET** d.Rating=r.idrating;

**UPDATE** dataset d **INNER JOIN** size s on d.Size = s.size

**SET** d.Size=s.idsize;

**UPDATE** dataset d **INNER JOIN** content\_rating cr on d.Content\_Rating = cr.content\_rating

**SET** d.Content\_Rating=cr.idcontent\_rating;

## 4.8. ALTER upiti nad tablicom dataset

**ALTER TABLE** dataset **MODIFY** App int;

**ALTER TABLE** dataset **MODIFY** Category int;

**ALTER TABLE** dataset **MODIFY** Rating int;

**ALTER TABLE** dataset **MODIFY** Size int;

**ALTER TABLE** dataset **MODIFY** Android\_Ver int;

**ALTER TABLE** dataset **MODIFY** Content\_Rating int;

## 4.9. Činjenična tablica apps\_facts

Činjenična tablica se sastoji od 8 atributa, a to su:

- idapps\_facts
- idapp
- idcategory



- idrating
- idsize
- idcontent\_rating
- idandroid\_version
- current\_ver

Svi atributi koji počinju sa id zapravo predstavljaju vanjske ključeve na pripadajuće dimenzijske tablice povezane sa činjeničnom tablicom. Nazivi vanjskih ključeva su isti kao i primarni ključ u dimenzijskim tablicama da bi ih kod kasnije analize, alatu Power BI Desktop lakše prepoznao da su povezane da ne bi morali ručno povezivati tablice. Da bi naposljetku napunili i činjeničnu tablicu potrebno je napisati sljedeći upit:

```
INSERT INTO apps_facts(idapp, idcategory, idrating, idsize, idcontent_rating, idandroid_version, current_ver) SELECT App, Category, Rating, Size, Content_Rating, Android_Ver,Current_Ver FROM dataset;
```

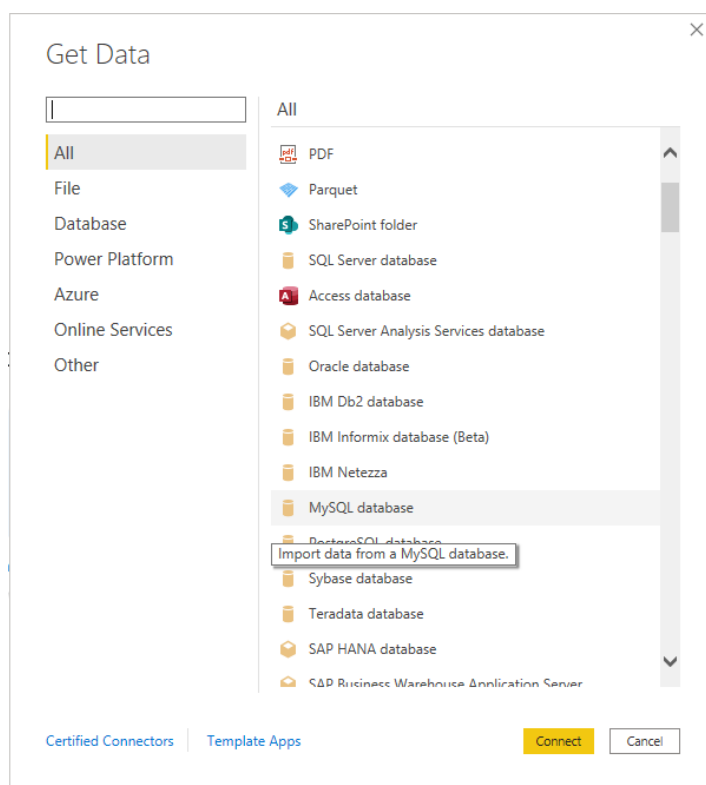
Nakon što smo unijeli podatke u činjeničnu tablicu, tablica apps\_facts izgleda kao na slici ispod.

	idapps_facts	idapp	idcategory	idrating	idsize	idcontent_rating	idandroid_version	current_ver
512	1	1	1	1	1	1	1	1.0.0
513	2	1	2	2	1	1	1	2.0.0
514	3	1	3	3	1	1	1	1.2.4
515	4	1	4	4	2	2		Varies with device
516	5	1	5	5	1	3		1.1
517	6	1	6	6	1	4		1.0
518	7	1	7	1	1	1		1.1
519	8	1	1	7	1	2		6.1.61.1
520	9	1	6	8	1	5		2.9.2
521	10	1	3	9	1	1		2.8
522	11	1	6	10	1	6		1.0.4
523	12	1	6	11	1	7		1.0.15
524	13	1	8	12	2	6		3.8
525	14	1	9	13	1	3		1.0.4
526	15	1	6	14	1	4		1.2.3
527	16	1	10	15	1	2		NaN
528	17	1	3	16	1	6		3.1
529	18	1	4	17	1	4		1.0
530	19	1	5	18	1	1		2.2.5
531	20	1	9	19	1	6		5.5.4
532	21	1	11	2	1	6		4.0
533	22	1	1	11	1	1		1.1
534	23	1	3	20	3	1		2.2.6.2
535	24	1	3	21	1	6		1.1.3
536	25	1	12	22	1	5		1.5
537	26	1	3	4	1	1		1.0.8

Slika 13. Činjenična tablica apps\_facts

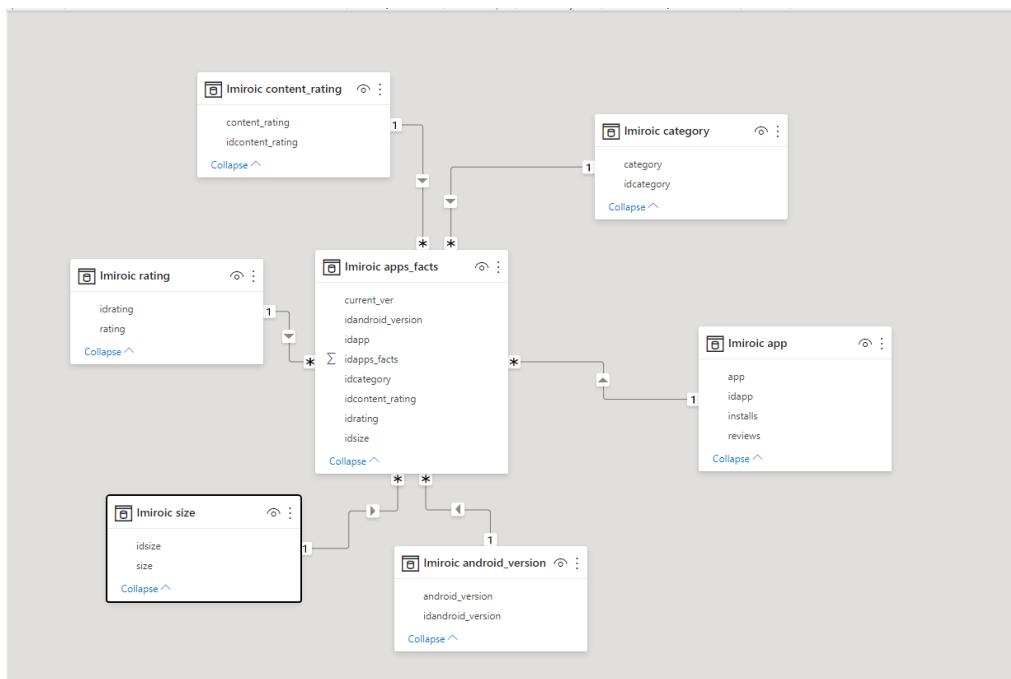
## 5. Analiza pomoću Power BI Desktop alata

Za analizu ovog skladišta podataka kojeg smo izradili prethodno u radu, korišten je Microsoft Power BI Desktop alat koji omogućuje analizu podataka iz raznih izvora, pa tako i MySQL baze podataka, što je slučaj i kod nas. Pa tako će u ovom poglavlju biti prikazani koraci učitavanja podataka iz skladišta podataka u alat te će biti kreirani razni izvještaji sa tih podataka. Pa tako prilikom pokretanja alata se na ekranu pojavi opcija *Get Data* odaberemo tu opciju, te se nakon toga otvara prozor u kojem možemo odabrati izvor od kuda dohvaćamo podatke ( u našem slučaju to je bila MySQL baza podataka) na adresi 127.0.0.1 na defaultnom portu za MySQL protokol 3306, a baza podataka se zove Imiroic.



Slika 14. Odabir vrste izvora podataka

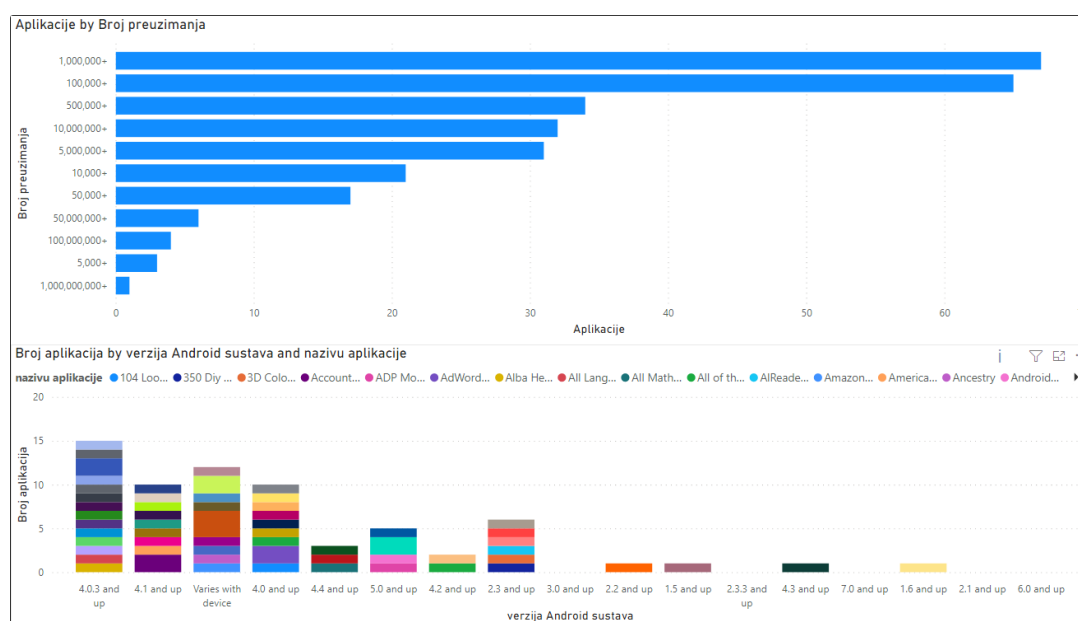
Nakon što smo se povezali na bazu podataka se otvara prozor sa popisom svih tablica koje se nalaze u bazi podataka, zatim odaberemo željene tablice te pritisnemo *Load*. Nakon što su sve tablice učitane alat generira, ukoliko je sve uredno napravljeno u MySQL Workbench alatu, model zvijezde koji izgleda kao na slici ispod.



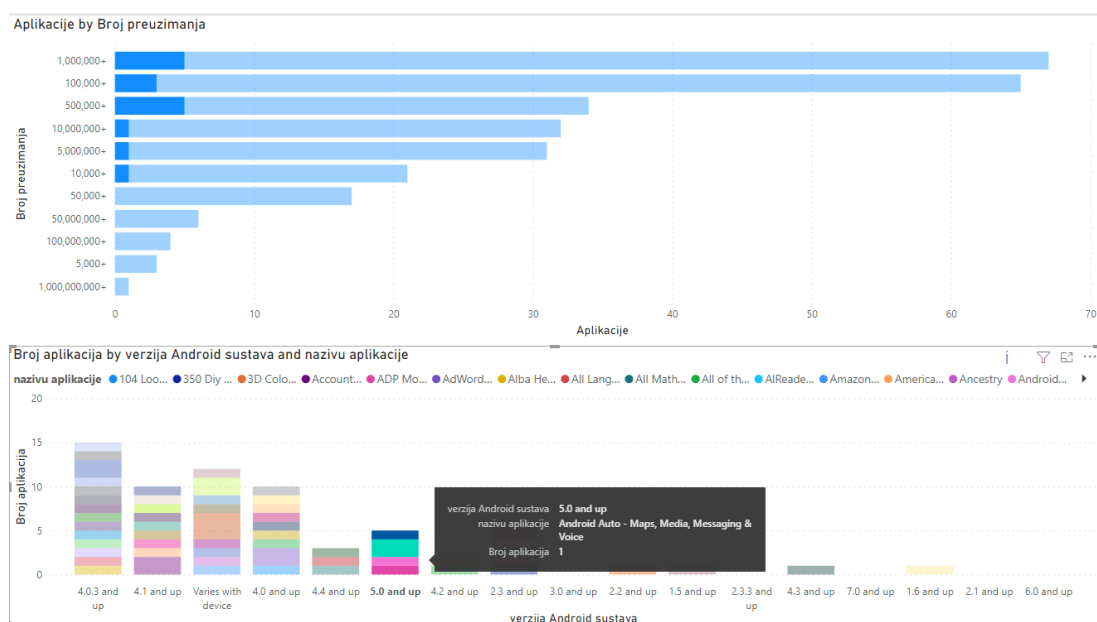
Slika 15. Generirani model zvijezde u Power BI za odabrano skladište podataka

## 5.1. Izvještaji u Power BI alatu

Za analizu podataka pokušao sam napraviti četiri vrste izvještaja na četiri zasebne stranice Power BI Desktop alata. Prvo možemo vidjeti dva grafa koja se odnose na prikaz broja aplikacija po broju preuzimanja, a drugi popis aplikacija ovisno o android verziji potrebnoj za pokretanje aplikacije. Na slici 16 vidimo prikaz grafova za android verzije aplikacije i broj preuzimanja općenito, a na slici 17 možemo vidjeti koliko preuzimanja imaju aplikacije sa android verzijom 5.0 i više i koliko takvih aplikacija ima.

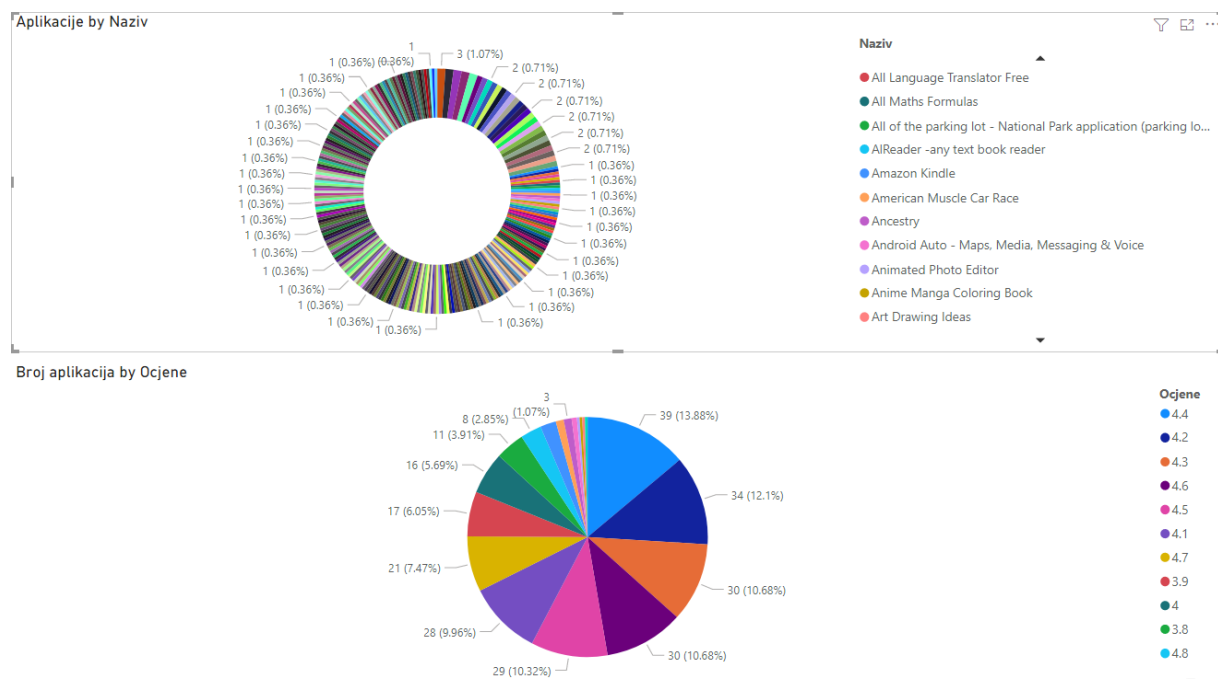


Slika 16. Prikaz preuzimanja aplikacija i njihove android verzije općenito



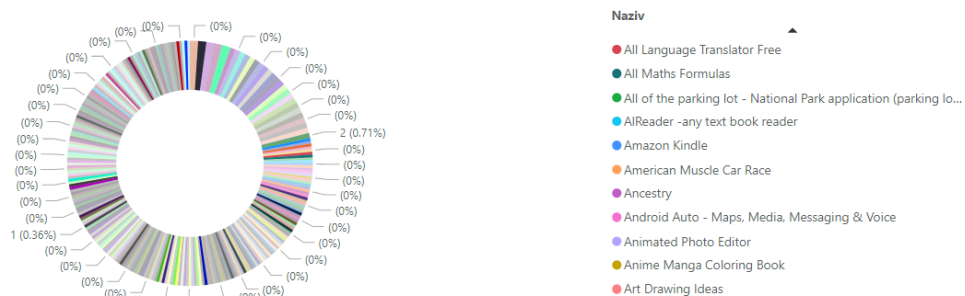
Slika 17. Prikaz preuzimanja aplikacija za android verziju 5.0+

Zatim možemo vidjeti dva grafa gdje na prvom vidimo popis svih aplikacija, a na drugom broj aplikacija po ocjenama koje su im korisnici dali. Možemo vidjeti da je najveća ocjena koju je aplikacija ostvarila 4.9 od 5, ali da najviše aplikacija ima ocjenu 4.4. Na slici 18 vidimo općeniti prikaz popisa aplikacija te broj aplikacija po određenoj ocjeni, dok na slici 19 možemo vidjeti da su označene na prvom grafu sve aplikacije sa ocjenom 4.4 koja je odabrana na drugom grafu.

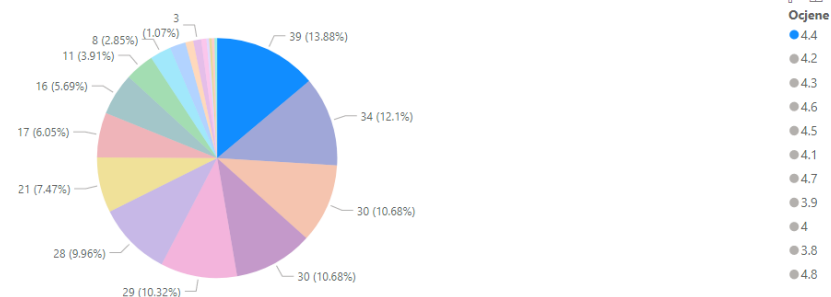


Slika 18. Prikaz aplikacija i njihovih ocjena općenito

Aplikacije by Naziv

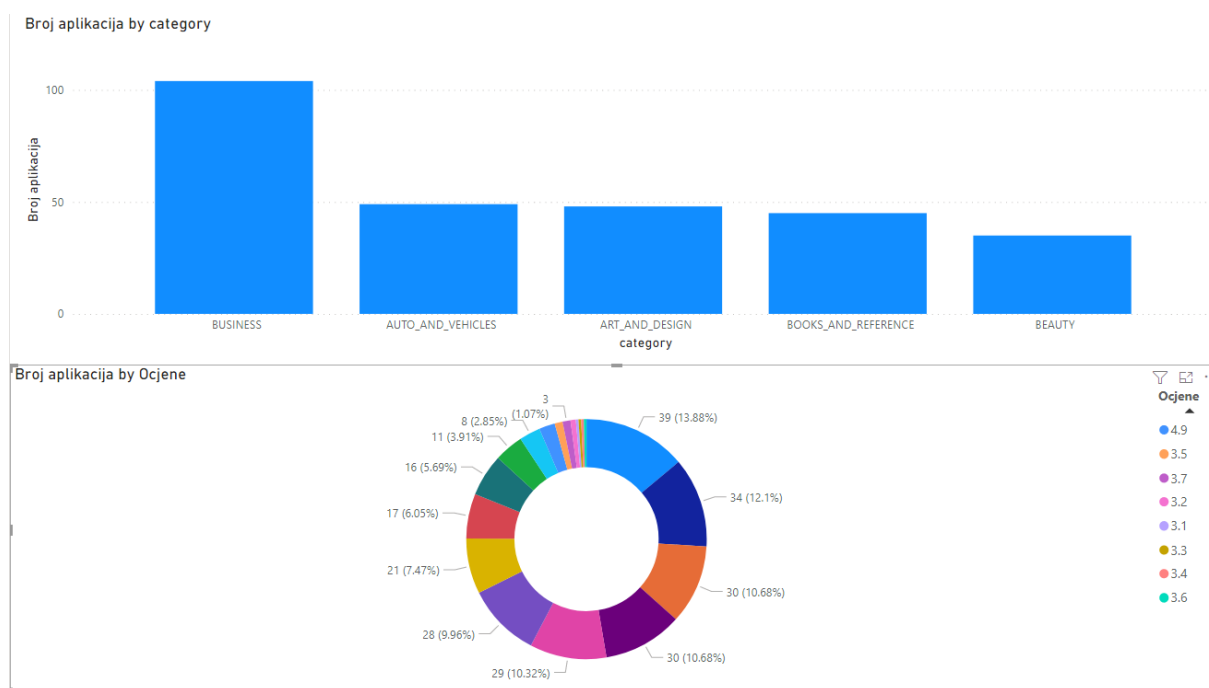


Broj aplikacija by Ocjene

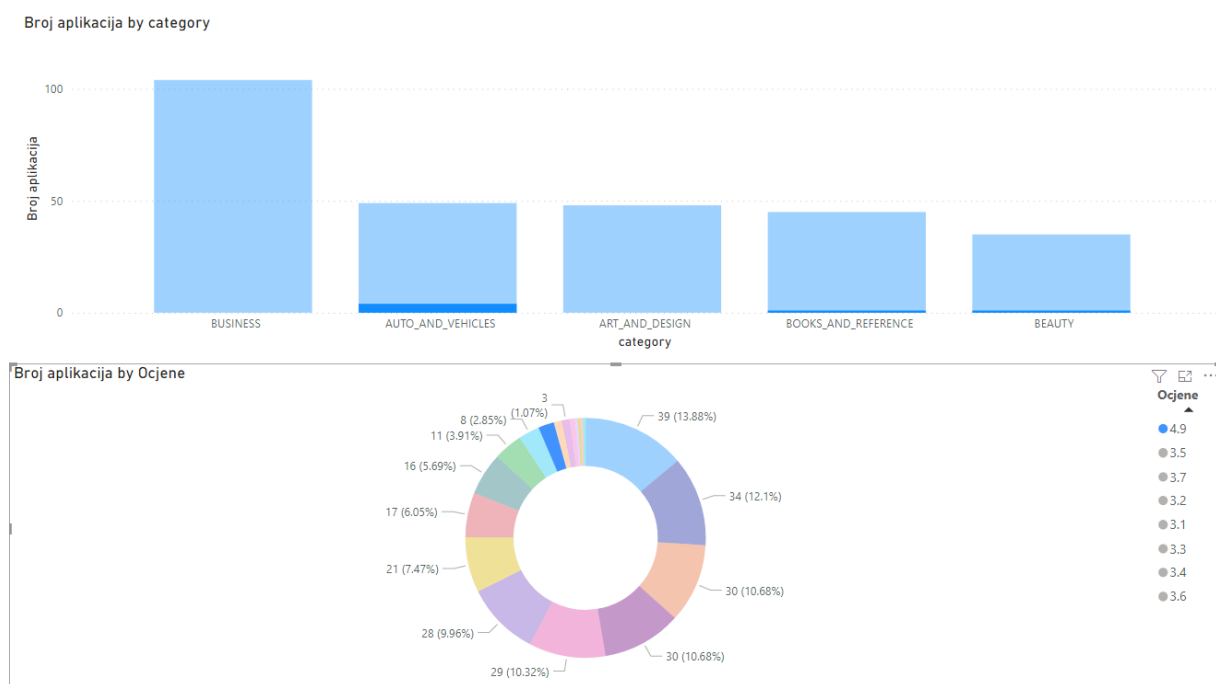


Slika 19. Prikaz aplikacija za ocjenu 4.4

Nadalje možemo vidjeti dva grafa od kojih jedan prikazuje broj aplikacija za određenu kategoriju aplikacije, a na drugom vidimo broj aplikacija po ocjeni koju su korisnici dali za određenu aplikaciju. Pa tako na slici 20. je prikazan općeniti prikaz svih kategorija i svih ocjena iz kojih možemo vidjeti da je najviše aplikacija iz kategorije *Business*, a na slici 21. možemo vidjeti broj aplikacija i u koju kategoriju spadaju najbolje ocjenjene aplikacije sa ocjenom 4.9. Pa tako vidimo da najbolje ocjenjene aplikacije spadaju u kategoriju *Auto and vehicles*.

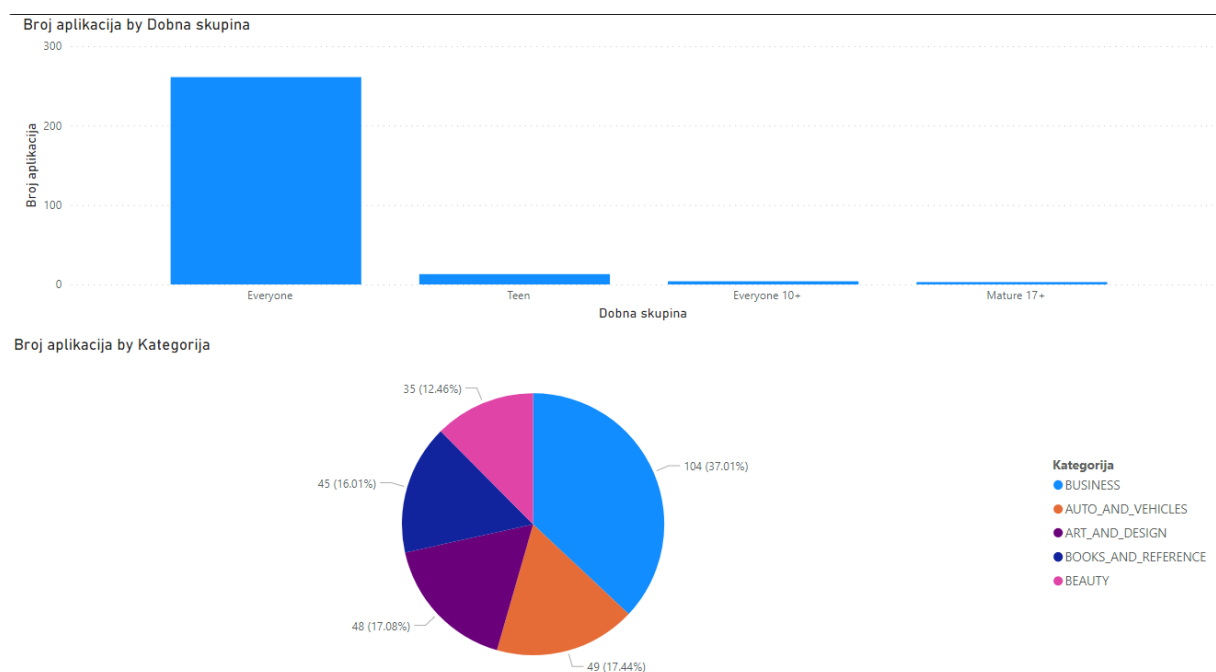


Slika 20. Prikaz kategorija i ocjena za aplikacije općenito

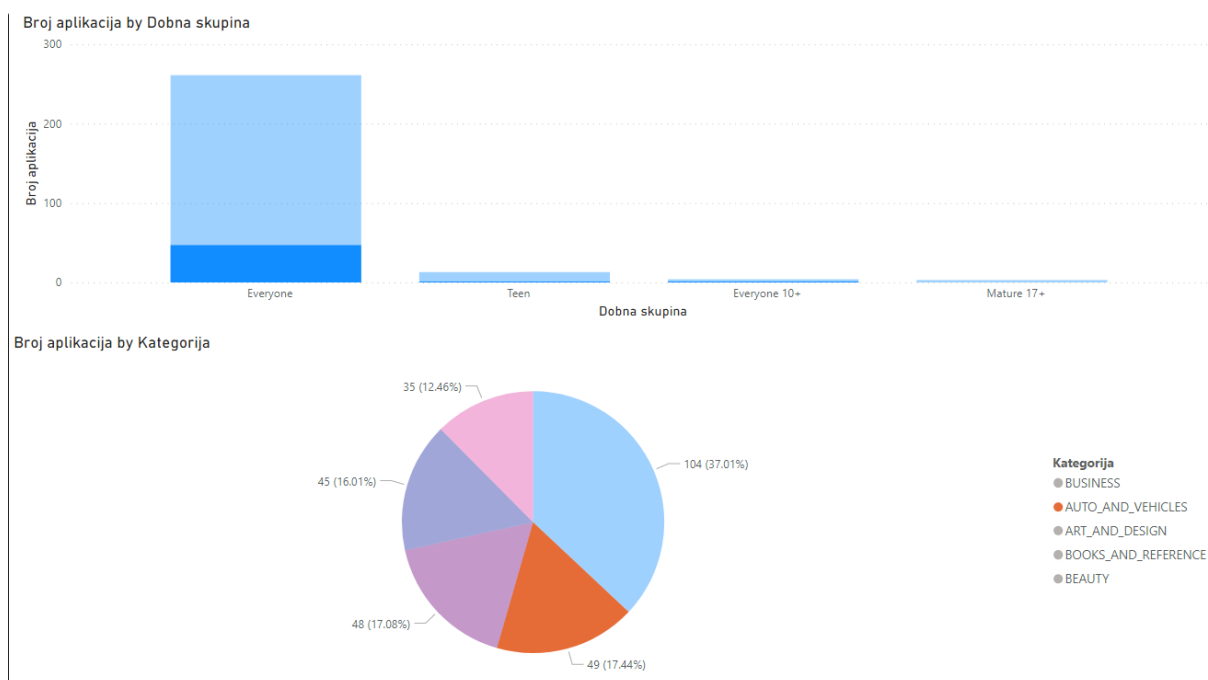


Slika 21. Prikaz u koju kategoriju spadaju aplikacije sa ocjenom 4.9

I za zadnju skupnu grafova imamo također dva grafa, gdje na prvom grafu vidimo popis broja aplikacija po dobnim skupinama za koje su aplikacije namijenjene, a na drugom grafu vidimo popis kategorija u koje aplikacije spadaju. Pa tako na slici 22. vidimo općeniti prikaz iznad objašnjenih grafova, a na slici 23. vidimo broj aplikacija u dobnoj skupini za kategoriju *Auto and vehicles*.



Slika 22. Prikaz dobnih skupina i kategorija aplikacija općenito



Slika 23. Prikaz broja aplikacija u dobnim skupinama za odabranu kategoriju

## 6. Zaključak

Ovaj projekt je izrađen da bi se analizirale android aplikacije dostupne na Google Play Store trgovini, pronaći koje su najpopularnije, i u koje kategorije spadaju, koje su aplikacije najbolje ocijenjene od strane korisnika te također u koju kategoriju spadaju te na kraju za koje dobne skupine ima najviše aplikacija i u koje kategorije spadaju određene aplikacije dobni skupina. Isto tako provođenjem ovog projekta mogli smo se upoznati sa novim tehnologijama koje su jedne od najpopularnijih u svijetu upravljanja bazama podataka i analizom skladišta podataka pa smo tako radili u alatima kao što su MySQL Workbench i Microsoft Power BI Desktop. Za izradu ovog rada koristili smo besplatan skup podataka koji je bio preuzet sa web stranice Kaggle.com, koji je bio jedan od rijetkih iz kojih se moglo uvesti barem dovoljno podataka u svrhu uspješne provedbe ovog rada, dok iz ostalih se moglo učitati samo par desetaka zapisa u MySQL Workbench alat. Nadalje možemo zaključiti kako je ovaj skup podataka ipak bio dovoljno kvalitetno popunjen te nije stvarao problem kod analize podataka u Power BI alatu u kojem su uspješno kreirane četiri stranice, svaka sa svojom skupinom grafova.



## Popis literature

- [1] B. Vangie (2021.): *ETL – Extract, Transform, Load*. Dostupno 30.5.2021. na <https://www.webopedia.com/definitions/etl/>
- [2] K. Rabuzin (2021.): „*Predavanja 5 i 6*“, nastavni materijali sa kolegija Skladišta podataka i poslovna inteligencija, Varaždin. Dostupno na Moodle sustavu.
- [3] I. Mekterović, Lj. Brkić (2017.): *Skladišta podataka i poslovna inteligencija*, skripta iz kolegija Poslovna inteligencija, Zagreb. Dostupno 31.5.2021. na [https://www.fer.unizg.hr/download/repository/SKRIPTA\\_Skladista\\_podataka\\_i\\_poslovna\\_inteligencija.pdf](https://www.fer.unizg.hr/download/repository/SKRIPTA_Skladista_podataka_i_poslovna_inteligencija.pdf)
- [4] Guru99.com (2021.), „*ETL (Extract, Transform and Load) Process in Data Warehouse*“. Dostupno 31.5.2021. na <https://www.guru99.com/etl-extract-load-process.html>
- [5] JavaTPoint (2021.), „*What is Star Schema?*“. Dostupno 31.5.2021. na <https://www.javatpoint.com/data-warehouse-what-is-star-schema>

## Popis slika

Slika 1. Prikaz skupa podataka u .csv formatu.....	4
Slika 2. Prikaz servera na koji se spajamo.....	5
Slika 3. Prikaz baze podataka .....	6
Slika 4. Prikaz atributa tablice dataset .....	6
Slika 5. Prikaz tablice dataset.....	7
Slika 6. Model zvijezde [5] .....	7
Slika 7. Dimenzijska tablica android_version .....	8
Slika 8. Dimenzijska tablica category.....	8
Slika 9. Dimenzijska tablica rating .....	9
Slika 10. Dimenzijska tablica size .....	10
Slika 11. Dimenzijska tablica content_rating.....	10
Slika 12. Dimenzijska tablica app .....	11
Slika 13. Činjenična tablica apps_facts.....	13
Slika 14. Odabir vrste izvora podataka .....	14
Slika 15. Generirani model zvijezde u Power BI za odabrano skladište podataka.....	15
Slika 16. Prikaz preuzimanja aplikacija i njihove android verzije općenito.....	15
Slika 17. Prikaz preuzimanja aplikacija za android verziju 5.0+ .....	16
Slika 18. Prikaz aplikacija i njihovih ocjena općenito.....	16
Slika 19. Prikaz aplikacija za ocjenu 4.4 .....	17
Slika 20. Prikaz kategorija i ocjena za aplikacije općenito.....	17
Slika 21. Prikaz u koju kategoriju spadaju aplikacije sa ocjenom 4.9 .....	18
Slika 22. Prikaz dobnih skupina i kategorija aplikacija općenito .....	18
Slika 23. Prikaz broja aplikacija u dobnim skupinama za odabranu kategoriju.....	19