

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

**Ivica Strmečki**

**Izrada skladišta podataka i analiza  
istih o prodaji rabljenih automobila**

**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**

**Ivica Strmečki**

**Matični broj: 0016123439**

**Studij: Organizacija poslovnih sustava**

**Izrada skladišta podataka i analiza istih o prodaji rabljenih  
automobila**

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

**Mentor:**

Prof. dr. sc. Kornelije Rabuzin

**Varaždin, lipanj 2021**

# Sadržaj

1. Uvod .....	1
2. Karakteristike projekta .....	2
2.1. Opis skupa podataka .....	2
2.2. Primijenjene tehnologije .....	3
2.2.1. MySQL .....	3
2.2.2. MySQL Workbench .....	3
2.2.3. Power BI .....	3
3. ETL i skladište podataka .....	4
4. Dimenzijsko modeliranje .....	8
4.1. Dimenzijska tablica fuel .....	9
4.2. Dimenzijska tablice owner .....	10
4.3. Dimenzijska tablica seller .....	11
4.4. Dimenzijska tablica transmission .....	12
4.5. Dimenzijska tablica car .....	13
4.6. Činjenična tablica car_facts .....	15
5. Power BI .....	17
6. Zaključak .....	24
7. Popis literature .....	25
8. Popis slika .....	26

# 1. Uvod

U današnje, veoma tehnološki napredno vrijeme i veliku količinu raznih podataka, dostupnost i raspoloživost istih je od vrlo velike važnosti. Poslovanje svakog poduzeća, koje želi zauzeti ciljani dio tržišta, steći konkurentsku prednost, odnosno u konačnici uspješno poslovati, temelji se na raznim podacima od kojih svaki ima određenu važnost. Samim time potrebna je integracija i dobra organizacija istih jer velika količina raznih, neadekvatnih podataka nam nije od velike važnosti i koristi. Samim time do izražaja dolaze razna skladišta podataka i poslovna inteligencija te se stavlja vrlo veliki naglasak na njihovu ulogu u uspješnom poslovanju gledajući organizaciju, jednostavnost, kvalitetu i efikasnost pronalaska potrebnih podataka za donošenje vrlo bitnih odluka u poslovanju.

Tako se u središtu razrade ovog projektnog zadatka nalazi izrada skladišta podataka te kreiranje izvještaja na temelju istih. Samim time će biti prikazan i *dataset*, odnosno skup podataka koji predstavljaju temelj za izradu skladišta podataka i same analize. Isto tako će biti opisan proces implementacije skladišta i ostalih povezanih, bitnih tehnologija, a prije samog zaključka biti će dani izvještaji zajedno sa njihovim značenjem.

## 2. Karakteristike projekta

Prije provođenja same analize, unutar ovog rada, bit će izrađeno skladište podataka. Da bismo uopće mogli izraditi skladište podataka, potrebni su nam podaci koji će ispunjavati isto. Kao što sam već spomenuo, dobro organizirani podaci predstavljaju temelj uspješnog funkcioniranja određenog sustava, što se opet temelji na provođenju raznih analiza, odnosno rudarenju podataka. Samim time ćemo prije konačne analize i ispisivanja raznih izvještaja urediti bazu podataka da bi podaci bili bolje organizirani i pregledniji te da bismo mogli u konačnici efikasnije iščitati one koji su u našoj interesnoj domeni.

### 2.1. Opis skupa podataka

U središtu razrade ovog projekta nalaze se podaci koji su vezani uz prodaju rabljenih automobila, a dolaze sa indijske web stranice CarDekho.com. Istoimena stranica je zapravo web stranica čija je glavna svrha pomoć kupcu prilikom kupnje željenog automobila. Samim time, na navedenoj web stranici isti može pronaći različite podatke vezane uz željeni automobil poput recenzija, detaljnijih specifikacija, cijene, kao i razne usporedbe te videozapise i slike vezane uz automobil. Sama tvrtka surađuje s više od 4000 auto-kuća pa je samim time prisutna velika količina podataka o raznim automobilima, a ujedno surađuje i sa raznim financijskim institucijama pa ju stvarno možemo smatrati dobrom podrškom prilikom kupnje limenog ljubimca [1].

Samim time, rad se temelji na analizi podataka koji su vezani uz rabljene automobile sa tržišta, a odnosi se na prikaz raznih atributa i njihovu međusobnu povezanost. Prilikom preuzimanja *dataseta* dobivamo .csv datoteku unutar koje imamo ukupno 7907 podataka i 13 atributa koji su zajedno sa svojim tipovima prikazani u nastavku.

#	Field	Schema	Table	Type
1	name	sppi	car_details	TEXT
2	year	sppi	car_details	INT
3	selling_price	sppi	car_details	INT
4	km_driven	sppi	car_details	INT
5	fuel	sppi	car_details	TEXT
6	seller_type	sppi	car_details	TEXT
7	transmission	sppi	car_details	TEXT
8	owner	sppi	car_details	TEXT
9	mileage	sppi	car_details	TEXT
10	engine	sppi	car_details	TEXT
11	max_power	sppi	car_details	TEXT
12	torque	sppi	car_details	TEXT
13	seats	sppi	car_details	INT

*Slika 1. Prikaz atributa i njihovih tipova*

## **2.2. Primijenjene tehnologije**

Prilikom izrade cjelokupnog projekta korišteno je nekoliko različitih tehnologija, ovisno o zadacima koje je bilo potrebno izvršiti. Generalni popis korištenih tehnologija je MySQL, MySQL Workbench i PowerBI, a njihovi kratki opisi slijede u nastavku rada.

### **2.2.1. MySQL**

Prije svega, MySQL je softver koji se danas nalazi u vlasništvu Oracle korporacije. To je jedan od najraširenijih sustava na tržištu koji upravlja relacijskim bazama podataka otvorenog koda. Relacijske baze podataka su namijenjene organizaciji velike količine podataka u različiti broj tablica koje se sastoje od stupaca (atributa) i redova (slogova), a tablice mogu biti međusobno povezane i upravo ta povezanost pomaže u strukturiranju samih podataka. Raširenost sustava je vrlo velika pa tako danas postoje inačice za veliki broj različitih operacijskih sustava kao što su Linux, Windows, Mac OS i drugi [2].

### **2.2.2. MySQL Workbench**

Prije svega, MySQL Workbench predstavlja jednog od najpoznatijih i najkorištenijih MySQL alata. To je vizualni alat koji se koristi za dizajniranje, modeliranje, generiranje i upravljanje bazom podataka. Kako integrira administraciju, razvoj, dizajn te stvaranje i održavanje SQL-a, možemo reći da sadrži sve potrebno da bi pojedinac uspješno upravljao raznim podacima i bazama podataka. Alat je dostupan na sustavima Windows, Linux i Mac OS X [3].

### **2.2.3. Power BI**

Posljednji korak našeg projekta odnosi se na kreiranje izvještaja na temelju analiziranih podataka. U tu svrhu koristi se besplatan user-friendly alat Power BI. Navedeni alat nam nudi velike mogućnosti gledajući povezivanje velikog broja različitih podataka, pripremu istih, odnosno njihovu organizaciju, analiziranje te ispis različitih vrsta izvještaja koji su svojim izgledom veoma jasni i razumljivi gotovo svakom korisniku. Kao i kod većine alata koji se koriste u različite svrhe postoji nekoliko verzija od kojih su neke Power BI Pro, Power BI Mobile, Power BI premium i verzija koja je korištena unutar ovog rada, a to je Power BI Desktop [4].

### 3. ETL i skladište podataka

Koraci koji su primijenjeni prilikom izrade ovog projekta mogu se poistovjetiti zapravo s ETL principom. Točnije, s procesima ekstrakcije (engl. Extract), transformacije (engl. Transform) te učitavanja (engl. Load) željenih podataka. Značenje prethodno spomenutih termina je sljedeće:

- Ekstrakcija – proces koji se odnosi na proces istraživanja i prikupljanja željenih podataka iz različitih izvora, te taj proces ujedno predstavlja temeljnu komponentu ETL procesa.
- Transformacija – proces koji se odnosi na kritične zadatke u ETL procesi. Točnije u kojem se prethodno prikupljeni podaci mijenjaju, odnosno transformiraju u onaj oblik koji je pogodan za učitavanje u određeno skladište podataka. Ovdje treba spomenuti i termin „pročišćavanja“ podataka koji se zapravo odnosi na uklanjanje neadekvatnih podataka čije postojanje nam nije od velike važnosti, primjerice irelevantni atributi, prazna, nepopunjena polja i slično.
- Učitavanje podataka – proces koji slijedi nakon što su svi podaci transformirani i prilagođeni u onom smjeru u kojem je potrebno da bi isti mogli biti učitani u skladište podataka, te da bi isti donijeli najveću moguću korist bilo u pogledu poslovanja ili za neke druge svrhe

Dakle, prvi korak cjelokupnog rada odnosi se na pretraživanje i preuzimanje, točnije prikupljanje željenih podataka. Prethodno sam već spomenuo kako temelj ovog rada predstavljaju podaci koji su vezani uz prodaju rabljenih automobila. Isti su preuzeti sa stranice *Kaggle* u .csv formatu dokumenta, a samo porijeklo podataka je na početku objašnjeno [5].

Nakon preuzimanja samog .csv dokumenta isti je otvoren putem alata MS Excel što možemo vidjeti na sljedećoj slici.

	A
1	name,year,selling_price,km_driven,fuel,seller_type,transmission,owner,mileage,engine,max_power,torque,seats
2	Maruti Swift Dzire VDI,2014,450000,145500,Diesel,Individual,Manual,First Owner,23.4 kmpl,1248 CC,74 bhp,190Nm@ 2000rpm,5
3	Skoda Rapid 1.5 TDI Ambition,2014,370000,120000,Diesel,Individual,Manual,Second Owner,21.14 kmpl,1498 CC,103.52 bhp,250Nm@ 1500-2500rpm,5
4	Honda City 2017-2020 EXi,2006,158000,140000,Petrol,Individual,Manual,Third Owner,17.7 kmpl,1497 CC,78 bhp,"12.7@ 2,700(kgm@ rpm)",5
5	Hyundai i20 Sportz Diesel,2010,225000,127000,Diesel,Individual,Manual,First Owner,23.0 kmpl,1396 CC,90 bhp,22.4 kgm at 1750-2750rpm,5
6	Maruti Swift VXI BSIII,2007,130000,120000,Petrol,Individual,Manual,First Owner,16.1 kmpl,1298 CC,88.2 bhp,"11.5@ 4,500(kgm@ rpm)",5
7	Hyundai Xcent 1.2 VTVT E Plus,2017,440000,45000,Petrol,Individual,Manual,First Owner,20.14 kmpl,1197 CC,81.86 bhp,113.75Nm@ 4000rpm,5
8	Maruti Wagon R LXI DUO BSIII,2007,96000,175000,LPG,Individual,Manual,First Owner,17.3 km/kg,1061 CC,57.5 bhp,"7.8@ 4,500(kgm@ rpm)",5
9	Maruti 800 DX BSII,2001,45000,5000,Petrol,Individual,Manual,Second Owner,16.1 kmpl,796 CC,37 bhp,59Nm@ 2500rpm,4
10	Toyota Etios VXD,2011,350000,90000,Diesel,Individual,Manual,First Owner,23.59 kmpl,1364 CC,67.1 bhp,170Nm@ 1800-2400rpm,5
11	Ford Figo Diesel Celebration Edition,2013,200000,169000,Diesel,Individual,Manual,First Owner,20.0 kmpl,1399 CC,68.1 bhp,160Nm@ 2000rpm,5
12	Renault Duster 110PS Diesel RxL,2014,500000,68000,Diesel,Individual,Manual,Second Owner,19.01 kmpl,1461 CC,108.45 bhp,248Nm@ 2250rpm,5
13	Maruti Zen LX,2005,92000,100000,Petrol,Individual,Manual,Second Owner,17.3 kmpl,993 CC,60 bhp,78Nm@ 4500rpm,5
14	Maruti Swift Dzire VDI,2009,280000,140000,Diesel,Individual,Manual,Second Owner,19.3 kmpl,1248 CC,73.9 bhp,190Nm@ 2000rpm,5
15	Maruti Swift 1.3 VXI,2007,200000,80000,Petrol,Individual,Manual,Second Owner,,,,,
16	Maruti Wagon R LXI Minor,2009,180000,90000,Petrol,Individual,Manual,Second Owner,18.9 kmpl,1061 CC,67 bhp,84Nm@ 3500rpm,5
17	Mahindra KUV 100 mFALCON G80 K8 5str,2016,400000,40000,Petrol,Individual,Manual,First Owner,18.15 kmpl,1198 CC,82 bhp,115Nm@ 3500-3600rpm,5
18	Maruti Ertiga SHVS VDI,2016,778000,70000,Diesel,Individual,Manual,Second Owner,24.52 kmpl,1248 CC,88.5 bhp,200Nm@ 1750rpm,7
19	Hyundai i20 1.4 CRDI Asta,2012,500000,53000,Diesel,Individual,Manual,Second Owner,23.0 kmpl,1396 CC,90 bhp,22.4 kgm at 1750-2750rpm,5
20	Maruti Alto LX,2002,150000,80000,Petrol,Individual,Manual,Second Owner,19.7 kmpl,796 CC,46.3 bhp,62Nm@ 3000rpm,5
21	Hyundai i20 2015-2017 Asta 1.4 CRDI,2016,680000,100000,Diesel,Individual,Manual,First Owner,22.54 kmpl,1396 CC,88.73 bhp,219.7Nm@ 1500-2750rpm,5
22	Mahindra Verito 1.5 D4 BSIII,2011,174000,100000,Diesel,Individual,Manual,Second Owner,21.0 kmpl,1461 CC,64.1 bhp,160Nm@ 2000rpm,5
23	Honda WR-V i-DTEC VX,2017,950000,50000,Diesel,Individual,Manual,First Owner,25.5 kmpl,1498 CC,98.6 bhp,200Nm@ 1750rpm,5
24	Maruti Swift Dzire ZDI,2015,525000,40000,Diesel,Individual,Manual,First Owner,26.59 kmpl,1248 CC,74 bhp,190Nm@ 2000rpm,5
25	Maruti SX4 ZDI,2012,600000,72000,Diesel,Individual,Manual,First Owner,21.5 kmpl,1248 CC,88.8 bhp,200Nm@ 1750rpm,5
26	Tata Tigor 2017-2020 XZ,2018,500000,35000,Petrol,Individual,Manual,First Owner,20.3 kmpl,1199 CC,83.81 bhp,114Nm@ 3500rpm,5
27	Maruti Baleno Delta 1.2,2016,575000,45000,Petrol,Individual,Manual,First Owner,21.4 kmpl,1197 CC,83.1 bhp,115Nm@ 4000rpm,5
28	Maruti Alto 800 LXI,2017,275000,28000,Petrol,Individual,Manual,First Owner,24.7 kmpl,796 CC,47.3 bhp,69Nm@ 3500rpm,5
29	Chevrolet Enjoy TCDI LTZ 7 Seater,2013,300000,70000,Diesel,Individual,Manual,First Owner,18.2 kmpl,1248 CC,73.8 bhp,172.5Nm@ 1750rpm,7
30	Maruti Wagon R LXI Minor,2009,220000,120000,Petrol,Individual,Manual,First Owner,18.9 kmpl,1061 CC,67 bhp,84Nm@ 3500rpm,5
31	Maruti Omni E MPI STD BS IV,2018,254999,25000,Petrol,Individual,Manual,First Owner,16.8 kmpl,796 CC,34.2 bhp,59Nm@ 2500rpm,8
32	Maruti Vitara Brezza LDI Option,2017,670000,70000,Diesel,Individual,Manual,First Owner,24.3 kmpl,1248 CC,88.5 bhp,200Nm@ 1750rpm,5
33	Fiat Palio 1.2 ELX,2003,70000,50000,Petrol,Individual,Manual,Second Owner,,,,,
34	Maruti Omni 8 Seater BSIV,2012,150000,35000,Petrol,Individual,Manual,Second Owner,14.0 kmpl,796 CC,35 bhp,6.1kgm@ 3000rpm,5
35	Hyundai i20 1.2 Asta Dual Tone,2018,730000,2388,Petrol,Individual,Manual,First Owner,18.6 kmpl,1197 CC,81.83 bhp,114.7Nm@ 4000rpm,5
36	Maruti Vitara Brezza LDI,2017,650000,16200,Diesel,Individual,Manual,First Owner,24.3 kmpl,1248 CC,88.5 bhp,200Nm@ 1750rpm,5

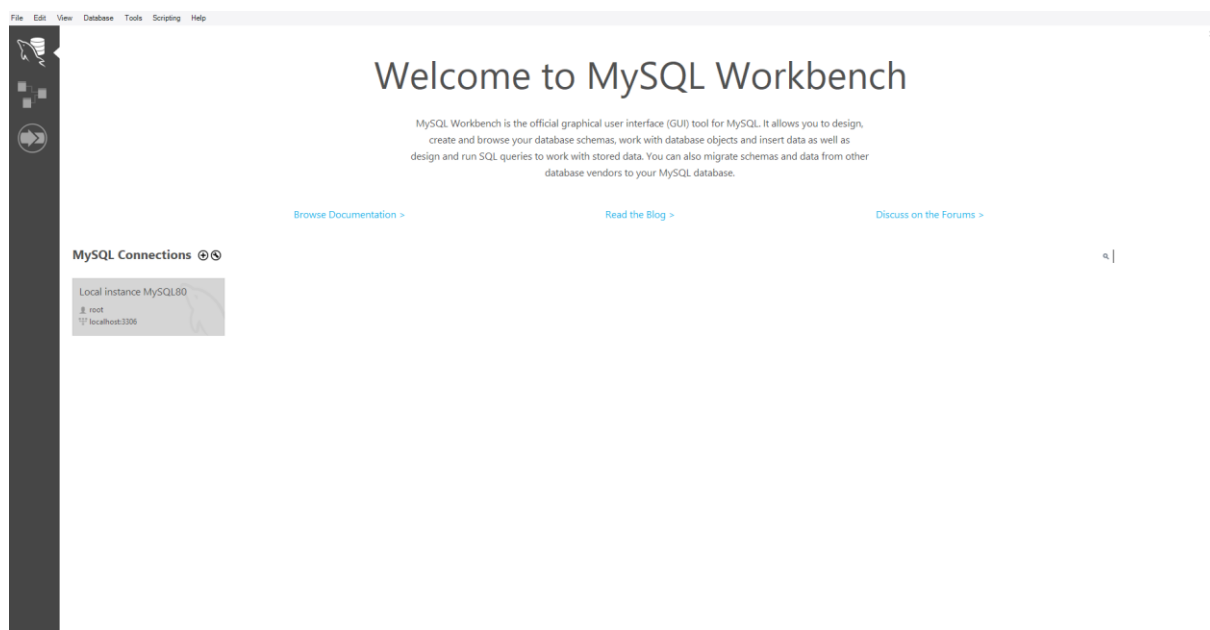
Slika 2. Pregled dijela izvornih podataka .csv datoteke

U preuzetoj datoteci nalazi se vrlo velika količina raznih podataka o samim automobilima. Točnije, 7907 redaka od kojih je svaki redak povezan s određenom markom automobila te su ostali atributi retka, koje sadrži naš *dataset*, također vezani uz isti.

Prethodno sam u radu dao pregled atributa i njihovih tipova tako da ću sada samo nabrojati od kojih se atributa sastoji svaki zapis, odnosno redak tablice. To su: name (naziv), year ( godina proizvodnje), selling\_price (prodajna cijena), km\_driven (prevoženi kilometri), fuel (vrsta goriva), seller\_type (vrsta prodavača), transmission (prijenos), owner (koji po redu vlasnik), mileage (kilometraža), engine (zapremnina motora), max\_power (maksimalna snaga), torque (vučna sila), seats (broj sjedala). Pregledavajući preuzeto utvrdio sam kako nema irelevantnih atributa, nema praznih polja već je svaki redak popunjen, tako da sam nakon početnog pregleda preuzetu .csv datoteku učitao u MySQL Workbench.



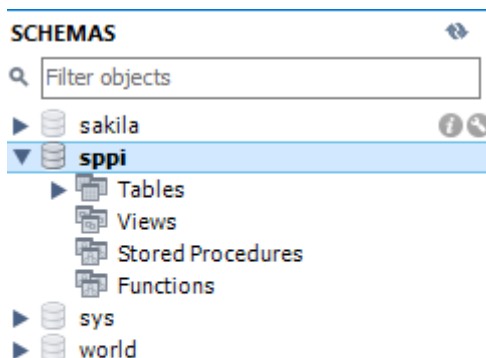
Prije samog učitavanja .csv datoteke u bazu bilo je potrebno postaviti istu unutar MySQL Workbencha, a početno sučelje MySQL Workbencha s dostupnim MySQL konekcijama izgleda kao na sljedećoj slici.



*Slika 3. Prikaz početnog sučelja MySQL Workbencha*

Nakon otvaranja samog programa, potrebno je odabrati željeni server na koji ćemo se spojiti, a u našem slučaju to je server pod nazivom „Local instance MySQL80“. Taj naziv je automatski generiran prilikom početnog postavljanja MySQL Workbencha, pa je tako i ostalo.

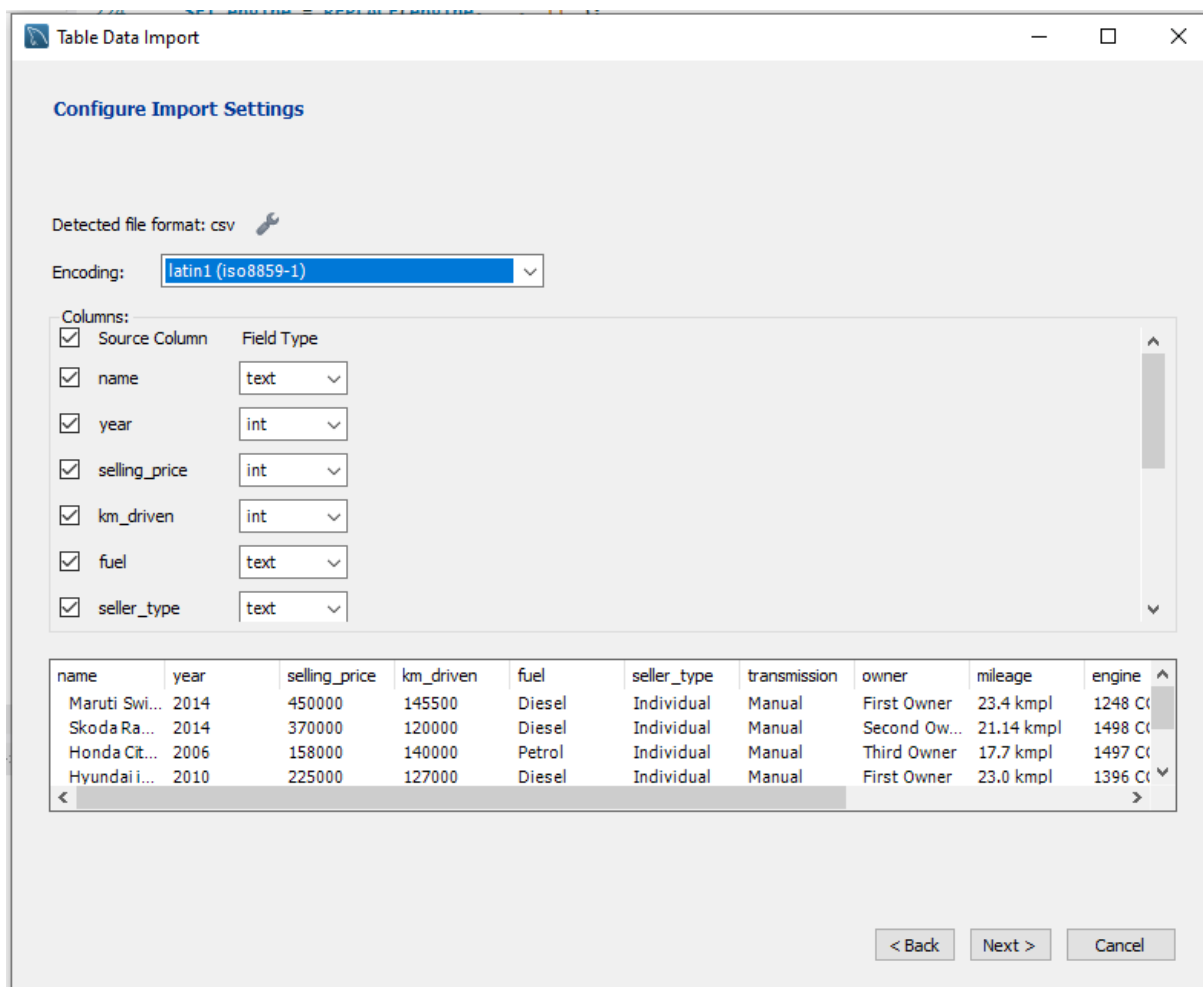
Nakon otvaranja servera, kreirao sam bazu podataka pod nazivom „sppi“ kao i određene dimenzijske tablice i činjeničnu, koje će detaljnije biti objašnjene kasnije.



*Slika 4. Prikaz baze podataka "sppi"*

Na slici 4. možemo vidjeti kreiranu bazu podataka i određene kategorije unutar iste, od kojih je prva po redu *Tables*, gdje će ujedno biti dostupne sve naše kreirane tablice.

Nakon što je kreirana *Schema*, u našem slučaju „*sppi*“, potrebno je ubaciti preuzetu datoteku u .csv formatu. Da bismo je ubacili u kreiranu bazu podataka potrebno je učiniti sljedeće: *desni klik na Schema „sppi“ → Table Data Import Wizard → Browse* i pronađemo željenu, preuzetu .csv datoteku te odemo *Next*. Nakon toga je potrebno podesiti još neke postavke kako bi ubacivanje (eng. Import) podataka prošlo uspješno, bez nekih poteškoća ili pogrešaka. Isto tako moguće je odabrati koje attribute iz preuzete .csv datoteke želimo ubaciti u kreiranu bazu podataka i moguće je iste od značiti ukoliko ih ne želimo.

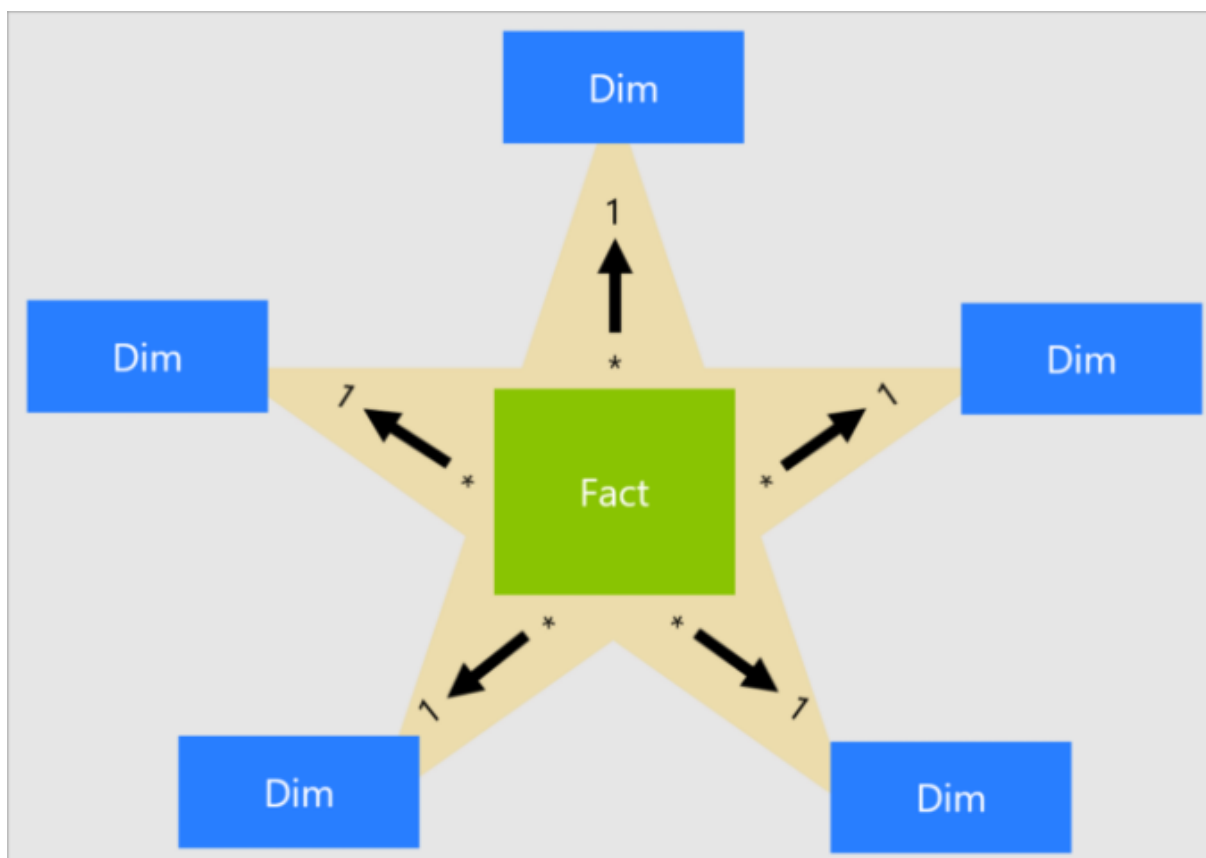


Slika 5. Pregled "Table Data Import" sučelja

Na prethodnoj slici možemo vidjeti kako izgleda naš *Table Data Import Wizard*. Kako sam prethodno utvrdio da nema irelevantnih ili nepopunjenih atributa, te su svi atributi, osim jednoga čija nam informacija nije važna, označeni za ubacivanje u bazu podataka. Isto tako, možemo vidjeti kojeg su tipa dostupni atributi te kratak pregled kako će naša konačna tablica izgledati. Nakon što su uređene sve ponuđene mogućnosti odemo na *Next* i pričekamo par trenutaka te nakon toga odaberemo *Finish*. Samim time, naša .csv datoteka je ubačena u bazu podataka, *Load* proces ETL principa je završen te smo spremni za daljnji rad.

## 4. Dimenzijsko modeliranje

Nakon provedenih koraka ETL principa, podaci su spremljeni u tablicu. Sada je potrebno definirati dimenzije i činjenice koje će doći do izražaja kasnije prilikom analize i kreiranja izvještaja. Prilikom spomenutog modeliranja možemo, uz ostale, susresti poznati model zvijezde, odnosno logički model podataka koji nam pomaže ubrzati često pokretane upite i pomoći prilikom organizacije samih podataka radi lakšeg i efikasnijeg snalaženja, kreiranjem činjenične tablice i pripadnih dimenzijskih tablica [6].



*Slika 6. Prikaz generalne strukture zvjezdanog modela*

Kao što možemo vidjeti na prethodnoj slici prikaza zvjezdanog modela, u središtu se nalazi *fact*, odnosno činjenica, što zapravo označava činjenični tablicu koja nam govori što mjerimo. Na „krakovima“ zvijezde možemo vidjeti dimenzije, odnosno dimenzijske tablice koje nam govore prema čemu mjerimo te su iste povezane sa središnjom činjeničnom tablicom. U nastavku rada ćemo spomenute tablice kreirati na temelju naših, odabranih podataka [6].

## 4.1. Dimenzijska tablica fuel

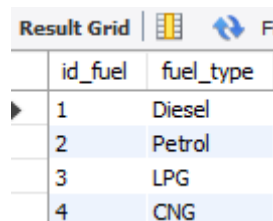
Prva opisana dimenzijska tablica je tablica *fuel*. Ona se sastoji od dva atributa, a to su: *id\_fuel* i *fuel\_type*. Prvi atribut (*id\_fuel*) tipa INT je prije svega definiran kako ne smije biti bez vrijednosti te je stavljeno automatsko dodjeljivanje iste. Isto tako je prvi atribut označen kao primarni ključ istoimene tablice. Za drugi atribut (*fuel\_type*) je definirani tip podataka TEXT. Naredbu koja je korištena prilikom kreiranja tablice dana je u nastavku.

```
CREATE TABLE fuel
(
  id_fuel int NOT NULL AUTO_INCREMENT,
  fuel_type text,
  primary key (id_fuel)
);
```

Nakon kreiranja tablice, istu smo popunili pripadajućim podacima sljedećim upitom.

```
INSERT INTO fuel (fuel_type) SELECT DISTINCT fuel FROM car_details;
```

Samim time smo dobili novu tablicu koja u sebi sadrži spomenute podatke, a izgled same tablice dan je na sljedećoj slici.



	id_fuel	fuel_type
▶	1	Diesel
	2	Petrol
	3	LPG
	4	CNG

Slika 7. Prikaz kreirane tablice fuel

Nakon što je tablica popunjena, što možemo vidjeti na prethodnoj slici, ista je ažurirana u početnoj tablici na sljedeći način:

```
UPDATE car_details c INNER JOIN fuel f ON c.fuel = f.fuel_type
SET c.fuel = f.id_fuel;
```

## 4.2. Dimenzijska tablice owner

Sljedeća po redu je tablica *owner*. Unutar iste imamo atribut *id\_owner*, koji predstavlja primarni ključ te je isto definiran kao i prethodni. Također unutar iste tablice imamo i atribut *owner\_type* tipa TEXT. Način kreiranja dan je u nastavku.

```
CREATE TABLE owner
(
  id_owner int NOT NULL AUTO_INCREMENT,
  owner_type text,
  primary key (id_owner)
);
```

Popunjavanje kreirane tablice *owner* izvršeno je sljedećim upitom.

```
INSERT INTO owner(owner_type) SELECT DISTINCT owner FROM car_details;
```

Nakon popunjavanja tablice, njezin prikaz možemo vidjeti na slici 8.

	id_owner	owner_type
►	1	First Owner
	2	Second Owner
	3	Third Owner
	4	Fourth & Above Owner
	5	Test Drive Car

Slika 8. Prikaz kreirane tablice *owner*

Kreirana tablica *owner* ažurirana je na sljedeći način:

```
UPDATE car_details c INNER JOIN owner o ON c.owner = o.owner_type
SET c.owner = o.id_owner;
```

### 4.3. Dimenzijska tablica seller

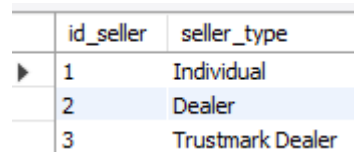
Dimenzijska tablica *seller* sastoji se također od dva atributa. To su *id\_seller* i *seller\_type* (tip prodavača). Unutar ove kreirane tablice *id\_seller* je stavljen kao primarni ključ i *auto increment*, a *seller\_type* tipa TEXT. Tablica je kreirana na sljedeći način:

```
CREATE TABLE seller
(
  id_seller int NOT NULL AUTO_INCREMENT,
  seller_type text,
  primary key (id_seller)
);
```

Nakon kreiranja slijedi popunjavanje i to se izvršava upitom koji slijedi.

```
INSERT INTO seller(seller_type) SELECT DISTINCT seller_type FROM car_details;
```

Nakon što smo kreirali tablicu i popunili željenim podacima, njezin izgled možemo vidjeti na sljedećoj slici.



	id_seller	seller_type
▶	1	Individual
	2	Dealer
	3	Trustmark Dealer

Slika 9. Prikaz kreirane tablice owner

Kao i prilikom stvaranja prethodnih dimenzijskih tablica, zadnji korak odnosi se na ažuriranje podataka početno učitane tablice, a to sam u ovom slučaju izvršio na sljedeći način:

```
UPDATE car_details c INNER JOIN seller s ON c.seller_type = s.seller_type
SET c.seller_type = s.id_seller;
```

## 4.4. Dimenzijska tablica transmission

Dimenzijska tablica *transmission* sadži također dva atributa. Prvi atribut je *id\_transmission* i on ujedno predstavlja primarni ključ tablice. Također je *auto increment* kao i u prethodnim slučajevima. Drugi atribut odnosi se na *transmission\_type* i taj atribut je tipa TEXT. Tablica je kreirana na način koji slijedi.

```
CREATE TABLE transmission
(
  id_transmission int NOT NULL AUTO_INCREMENT,
  transmission_type text,
  primary key (id_transmission)
);
```

Nakon kreiranja je popunjena podacima upitom:

```
INSERT INTO transmission(transmission_type) SELECT DISTINCT transmission
FROM car_details;
```

te je izgled same tablice prikazan na slici 10.

	id_transmission	transmission_type
►	1	Manual
	2	Automatic

Slika 10. Prikaz kreirane tablice transmission

U posljednjem koraku je kreirana tablica ažurirana unutar početne na sljedeći način:

```
UPDATE car_details c INNER JOIN transmission t ON c.transmission =
t.transmission_type
SET c.transmission = t.id_transmission;
```

## 4.5. Dimenzijska tablica car

Posljednje kreirana dimenzijska tablica je tablica *car*. Za razliku od prethodno kreiranih ona sadrži veći broj atributa, a to su sljedeći: *id\_car* koji ujedno predstavlja primarni ključ te je *auto increment*, *name\_car* tipa TEXT, *engine* tipa TEXT, *max\_power* tipa TEXT, *torque* tipa TEXT i *seats* tipa INT. Proces kreiranja nije ništa kompleksniji već samo malo duljeg zapisa kao što možemo vidjeti u nastavku.

```
CREATE TABLE car
(
  id_car int NOT NULL AUTO_INCREMENT,
  name_car text,
  engine text,
  max_power text,
  seats int,
  primary key(id_car)
);
```

Popunjavanje podataka izvršava se upitom koji slijedi, a na slici 11. je prikazan dio izgleda kreirane i popunjene tablice.

```
INSERT INTO car(name_car,engine,max_power,seats) SELECT DISTINCT
name,engine,max_power,seats FROM car_details;
```



	id_car	name_car	engine	max_power	seats
►	1	Maruti Swift Dzire VDI	1248	74	5
	2	Skoda Rapid 1.5 TDI Ambition	1498	103.52	5
	3	Honda City 2017-2020 EXi	1497	78	5
	4	Hyundai i20 Sportz Diesel	1396	90	5
	5	Maruti Swift VXI BSIII	1298	88.2	5
	6	Hyundai Xcent 1.2 VTVT E Plus	1197	81.86	5
	7	Maruti Wagon R LXI DUO BSIII	1061	57.5	5
	8	Maruti 800 DX BSII	796	37	4
	9	Toyota Etios VXD	1364	67.1	5
	10	Ford Figo Diesel Celebration Edition	1399	68.1	5
	11	Renault Duster 110PS Diesel RxL	1461	108.45	5
	12	Maruti Zen LX	993	60	5
	13	Maruti Swift Dzire VDi	1248	73.9	5
	14	Maruti Wagon R LXI Minor	1061	67	5
	15	Mahindra KUV 100 mFALCON G80 ...	1198	82	5
	16	Maruti Ertiga SHVS VDI	1248	88.5	7
	17	Hyundai i20 1.4 CRDi Asta	1396	90	5
	18	Maruti Alto LX	796	46.3	5
	19	Hyundai i20 2015-2017 Asta 1.4 C...	1396	88.73	5
	20	Mahindra Verito 1.5 D4 BSIII	1461	64.1	5
	21	Honda WR-V i-DTEC VX	1498	98.6	5
	22	Maruti Swift Dzire ZDI	1248	74	5
	23	Maruti SX4 ZDI	1248	88.8	5
	24	Tata Tigor 2017-2020 XZ	1199	83.81	5
	25	Maruti Baleno Delta 1.2	1197	83.1	5
	26	Maruti Alto 800 LXI	796	47.3	5
	27	Chevrolet Enjoy TCDi LTZ 7 Seater	1248	73.8	7
	28	Maruti Omni E MPI STD BS IV	796	34.2	8
	29	Maruti Vitara Brezza LDi Option	1248	88.5	5
	30	Maruti Omni 8 Seater BSIV	796	35	5

*Slika 11. Prikaz kreirane tablice car*

Kao i u svim prethodno kreiranim tablicama, posljednji korak odnosi se na ažuriranje podataka unutar početno kreirane tablice, odnosno tablice koju smo na početku ubacili u bazu podataka. Kako imamo malo više atributa unutar ove dimenzijske tablice, proces ažuriranja je malo složeniji, što možemo vidjeti u nastavku.

```

UPDATE car_details c, car s
SET c.name = s.id_car
WHERE c.name = s.name_car
AND c.engine = s.engine
AND c.max_power = s.max_power
AND c.seats = s.seats;

```

## 4.6. Činjenična tablica car\_facts

Nakon što su kreirane sve dimenzijske tablice, kreirana je i jedna činjenična na koju će iste biti povezane vanjskim ključevima. Prije nego je sama tablica kreirana, zbog vanjskih ključeva smo morali učiniti sljedeći:

```
ALTER TABLE car_details MODIFY name int;
```

```
ALTER TABLE car_details MODIFY fuel int;
```

```
ALTER TABLE car_details MODIFY owner int;
```

```
ALTER TABLE car_details MODIFY seller_type int;
```

```
ALTER TABLE car_details MODIFY transmission int;
```

Sama činjenična tablica se sastoji od sljedećih atributa:

- *Idcar\_facts*
- *Id\_car*
- *Year*
- *Selling\_price*
- *Km\_driven*
- *Id\_fuel*
- *Id\_seller*
- *Id\_transmission*
- *Id\_owner*
- *mileage*

Svi atributi koji imaju prefiks *Id*, osim prvog atributa (*Idcar\_facts*), koji predstavlja primarni ključ same činjenične tablice, predstavljaju vanjske ključeve na prethodno kreirane dimenzijske tablice koje su s činjeničnom povezane. Ostali atributi su isti kao u početno datom skupu podataka.

Nakon što je činjenična tablica kreirana, u istu su ubačeni podaci *insert* naredbom, a to je realizirano na sljedeći način:

```
INSERT INTO car_facts (id_car, year, selling_price, km_driven, id_fuel, id_seller,  
id_transmission, id_owner, mileage) SELECT name, year, selling_price, km_driven, fuel,  
seller_type, transmission, owner, mileage FROM car_details;
```

Konačan izgled jednog dijela naše kreirane činjenične tablice možemo vidjeti na slici koja slijedi.

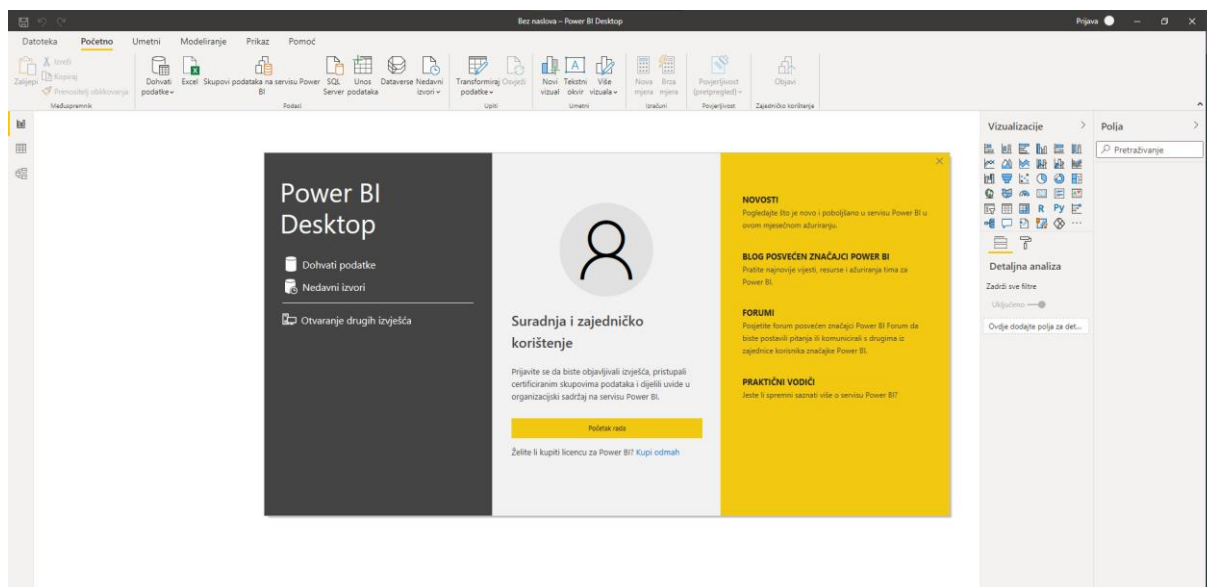
	idcar_facts	id_car	year	selling_price	km_driven	id_fuel	id_seller	id_transmission	id_owner	mileage
►	1	1	2014	450000	145500	1	1	1	1	23.4
	2	2	2014	370000	120000	1	1	1	2	21.14
	3	3	2006	158000	140000	2	1	1	3	17.7
	4	4	2010	225000	127000	1	1	1	1	23.0
	5	5	2007	130000	120000	2	1	1	1	16.1
	6	6	2017	440000	45000	2	1	1	1	20.14
	7	7	2007	96000	175000	3	1	1	1	17.3
	8	8	2001	45000	5000	2	1	1	2	16.1
	9	9	2011	350000	90000	1	1	1	1	23.59
	10	10	2013	200000	169000	1	1	1	1	20.0
	11	11	2014	500000	68000	1	1	1	2	19.01
	12	12	2005	92000	100000	2	1	1	2	17.3
	13	13	2009	280000	140000	1	1	1	2	19.3
	14	14	2009	180000	90000	2	1	1	2	18.9
	15	15	2016	400000	40000	2	1	1	1	18.15
	16	16	2016	778000	70000	1	1	1	2	24.52
	17	17	2012	500000	53000	1	1	1	2	23.0
	18	18	2002	150000	80000	2	1	1	2	19.7
	19	19	2016	680000	100000	1	1	1	1	22.54

*Slika 12. Prikaz kreirane činjenične tablice car\_facts*

## 5. Power BI

Već prethodno u radu, Power BI alat je opisan kao alat za analiziranje podataka i izradu izvještaja, odnosno vizualizaciju rezultata. Samim time, nakon što smo kreirali skladište podataka i sve popratne tablice, spremni smo na zadnji korak koji se temelji na izradi željenih izvještaja, a to će biti prikazano u nastavku rada.

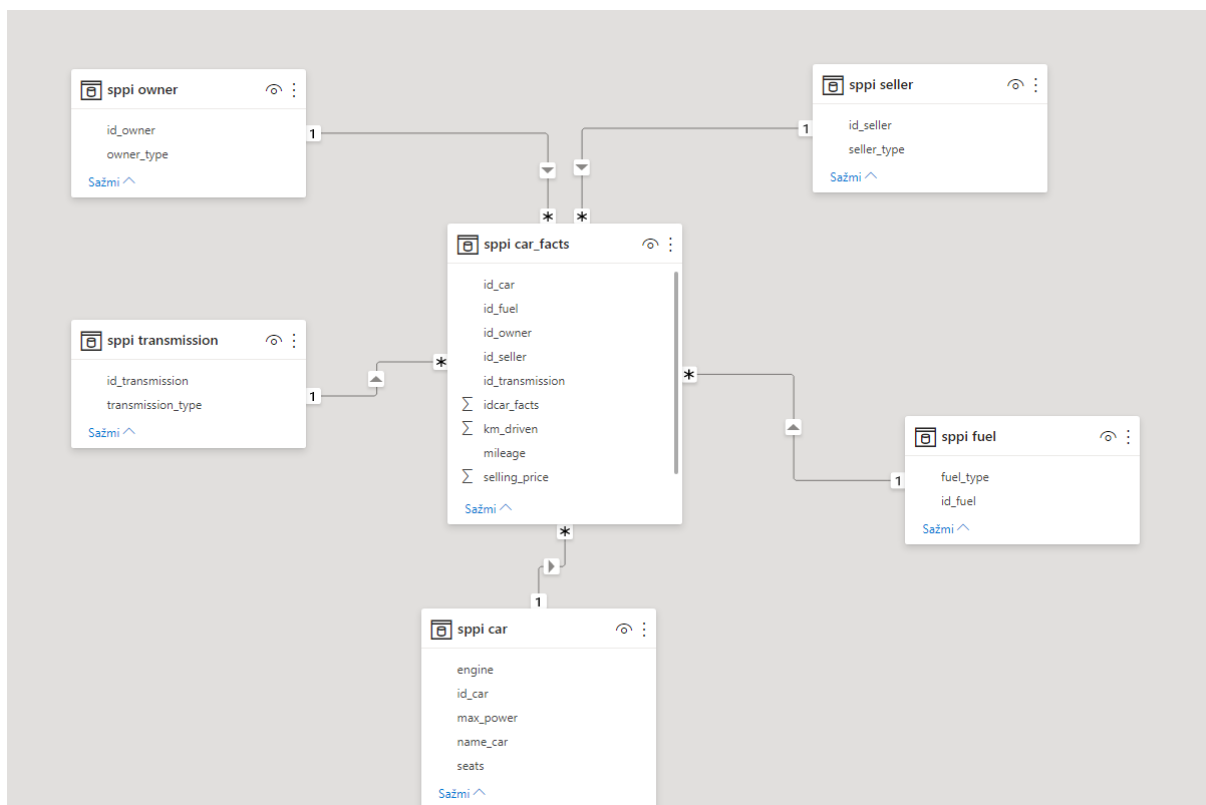
Nakon instalacije i otvaranja samog alata, otvara nam se sučelje koje možemo vidjeti na sljedećoj slici.



Slika 13. Prikaz početnog sučelja Power BI alata

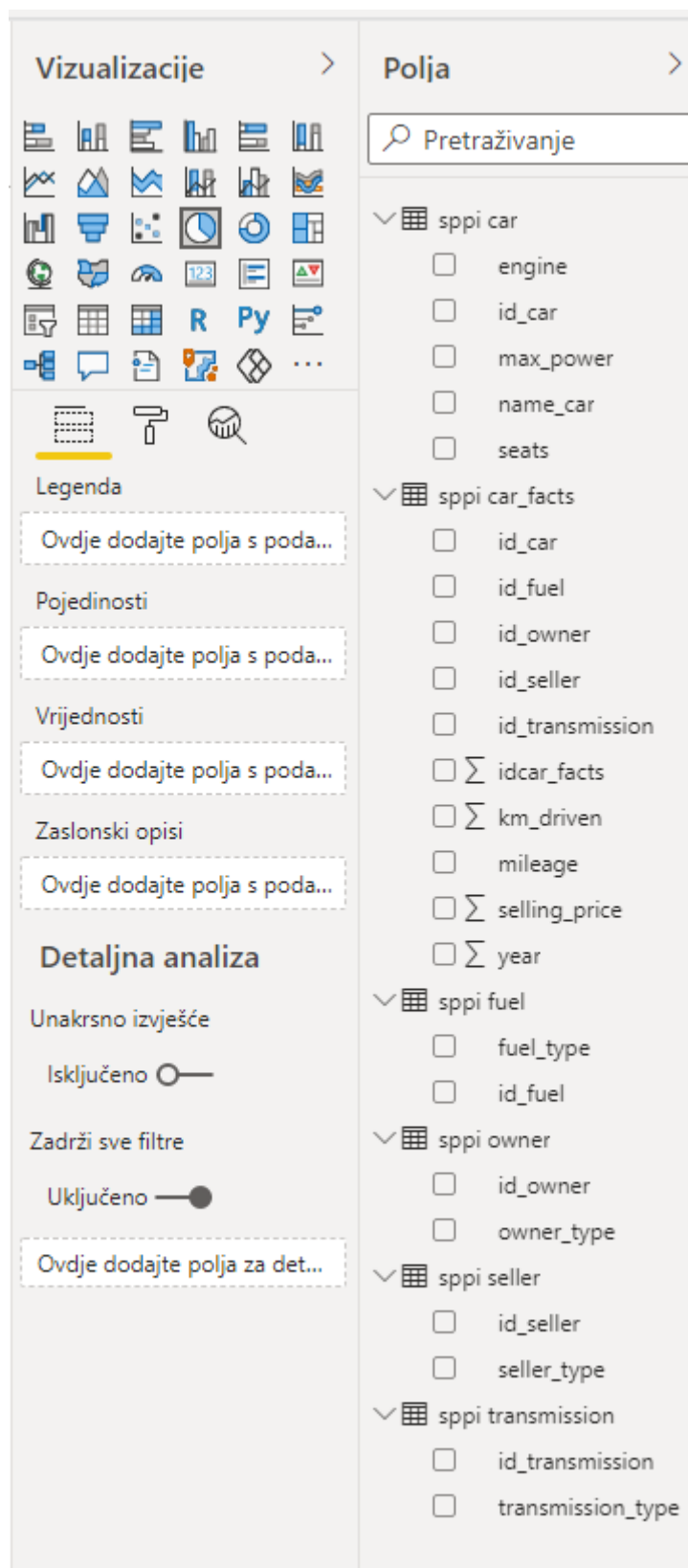
Prvi korak odnosi se na dohvaćanje željenih podataka odabirom opcije *Dohvati podatke*. Nakon toga odaberemo izvor iz kojeg želimo uzeti podatke i učitati ih u Power BI. U našem slučaju je adresa izvora 127.0.0.1, a port 3306 dok je željena baza podataka „spri“.

Nakon što smo uspješno definirali i otvorili željeno, prilikom otvaranja samog programa automatski nam se kreira, uz ostale mogućnosti, model zvijezde sa pripadajućim dimenzijskih i činjeničnom tablicom te atributima, a možemo ga vidjeti na sljedećoj slici.



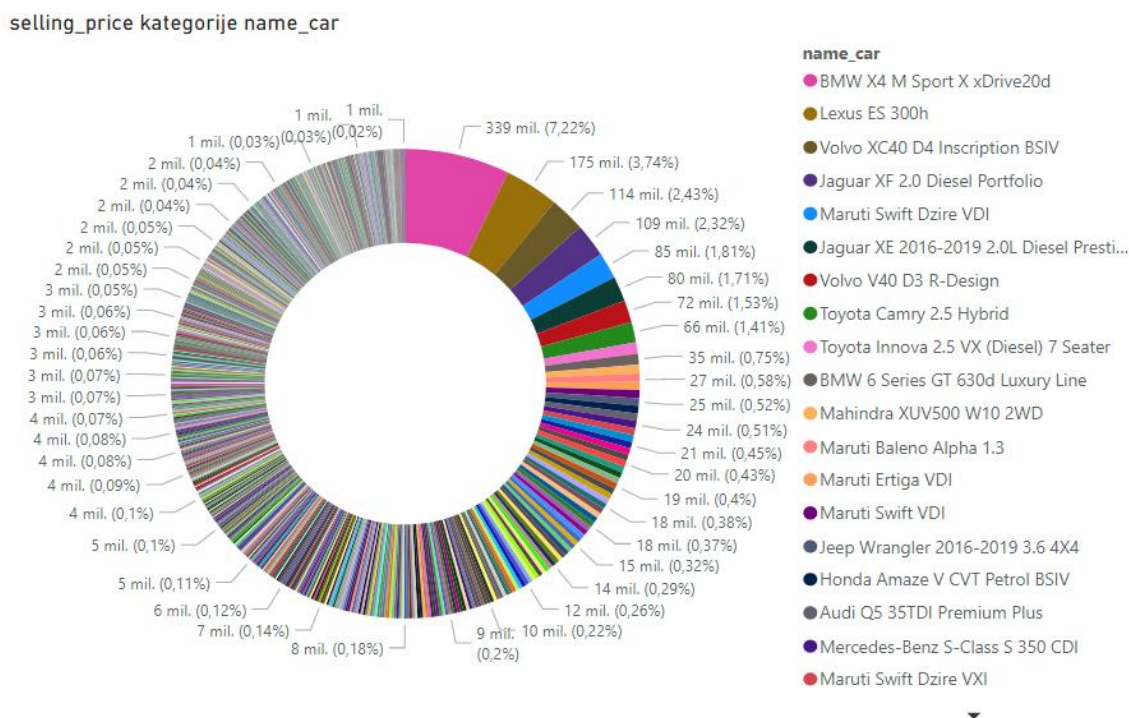
*Slika 14. Model zvijezde odabranog skladišta podataka*

Nakon što smo uspješno učitali sve željene tablice i podatke u program Power BI spremni smo napraviti nekoliko izvještaja. Isti se kreiraju vrlo jednostavnim odabirom željenih podataka unutar grafa i odabirom željenog oblika izvještaja npr. pita dijagram, stupičasti, linijski, prstenasti i slično. Pregled dostupnih dijagrama i podataka koje možemo odabrati prilikom kreiranja izvještaja dan je na sljedećoj slici.

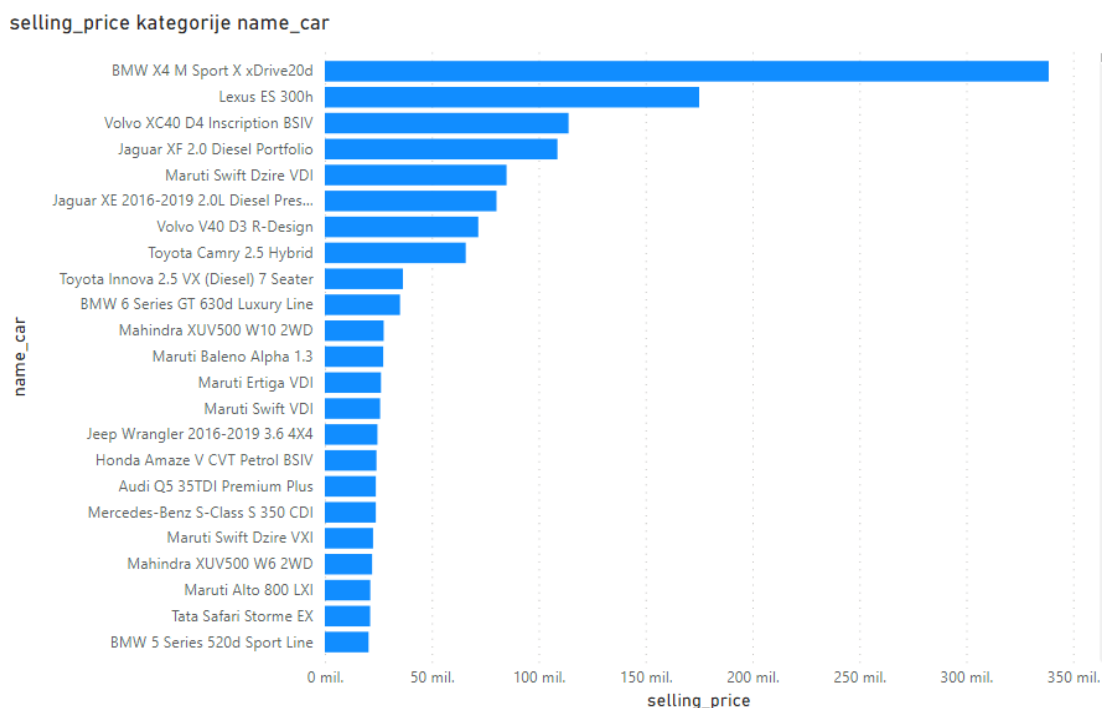


Slika 15. Prikaz dostupnih podataka i raznih oblika izvještaja

Prvi kreirani izvještaj odnosi se na popis naziva automobila i ukupne sume njihove prodajne cijene. Dakle, možemo vidjeti popis različitih marki automobila koje su dostupne unutar naših podataka, a marke automobila se ponavljaju veći broj puta, pa je u konačnici suma cijene vrlo visoka. Prvi izvještaj dan je u obliku prstenastog grafikona i grupiranog trakastog grafikona što možemo vidjeti na slikama koje slijede.



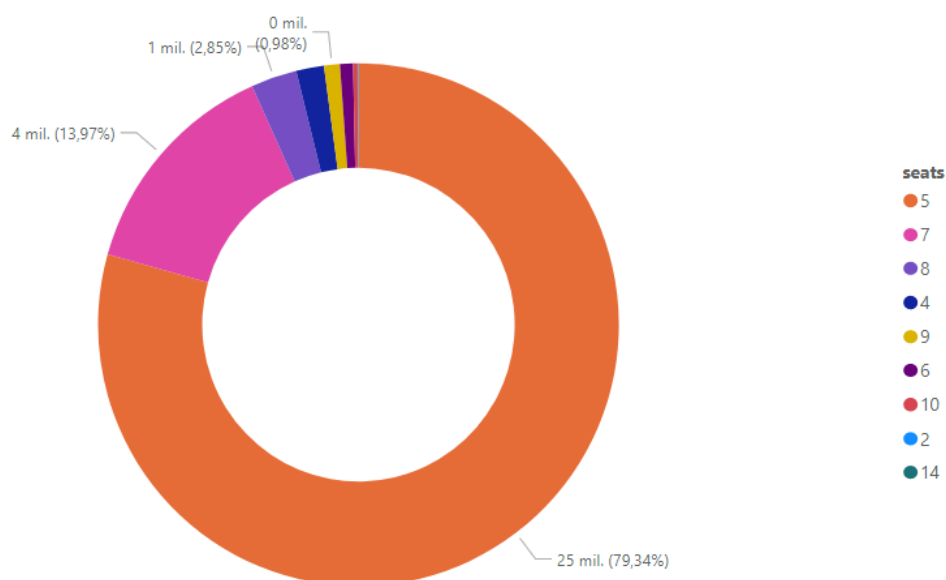
Slika 16. Prstenasti grafikon marke automobila i sume njihove prodajne cijene



Slika 17. Grupirani trakasti grafikon marke automobila i sume njihove prodajne cijene

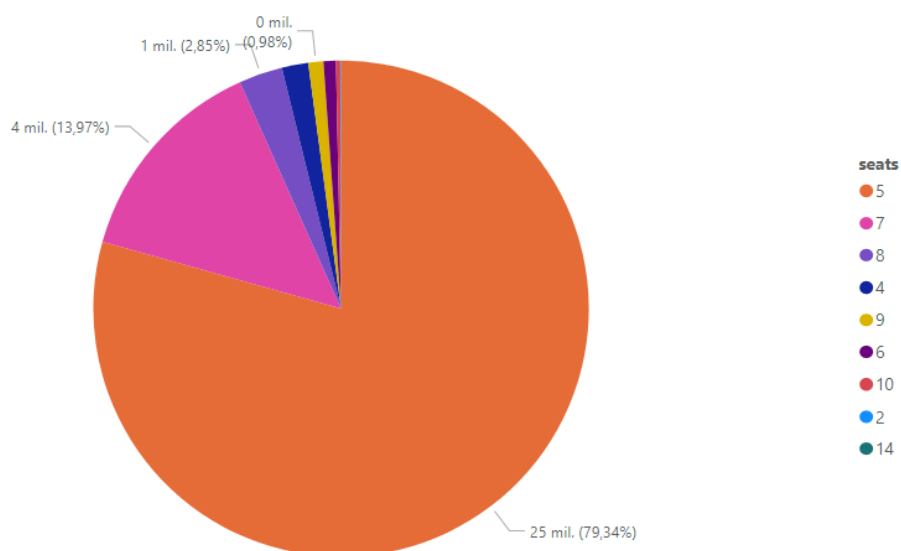
Drugi izvještaj odnosi se na broj sjedala. Točnije, prikazuje nam koliko imamo automobila ukupno iz našeg skladišta podataka koji imaju određeni broj sjedala, odnosno u našem slučaju 2, 4, 5, 6, 7, 8, 9, 10, 14, te je taj podatak prikazan prstenastim i tortnim grafikonom.

idcar\_facts kategorije seats



Slika 18. Prstenasti grafikon broja automobila s obzirom na broj sjedala

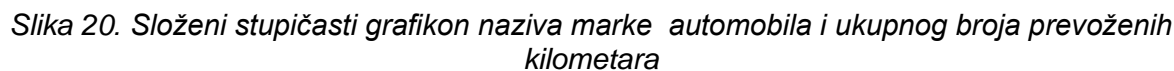
idcar\_facts kategorije seats



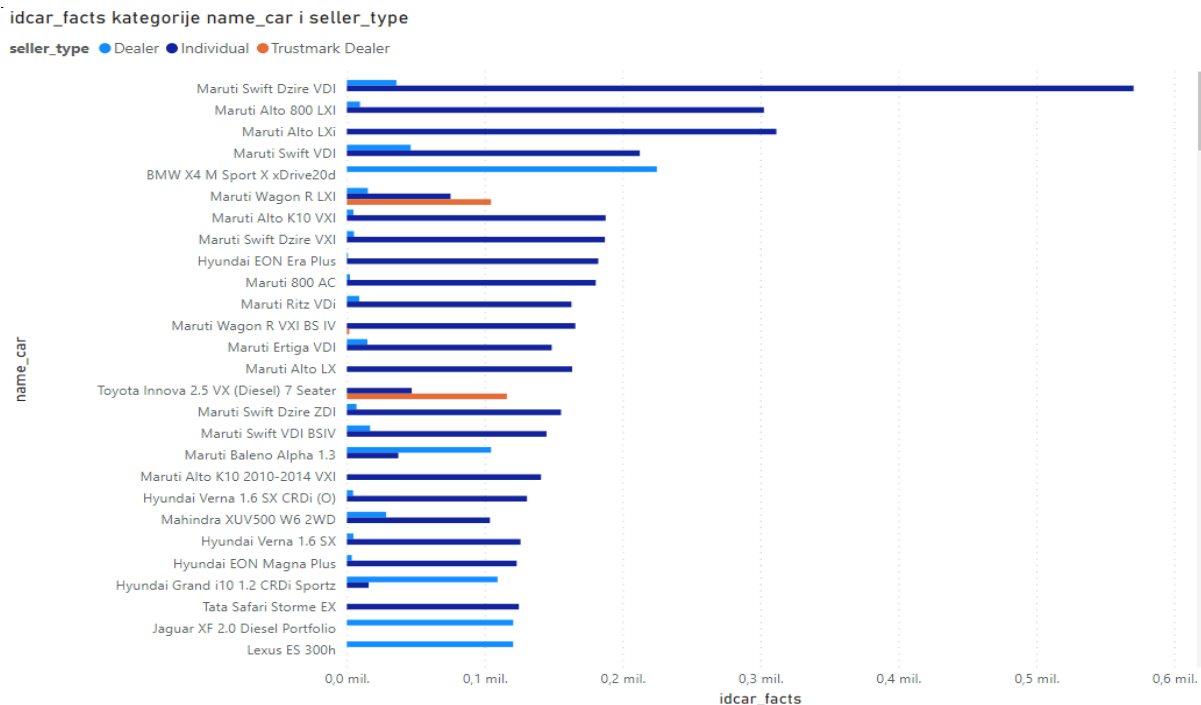
Slika 19. Tortni grafikon broja automobila s obzirom na broj sjedala



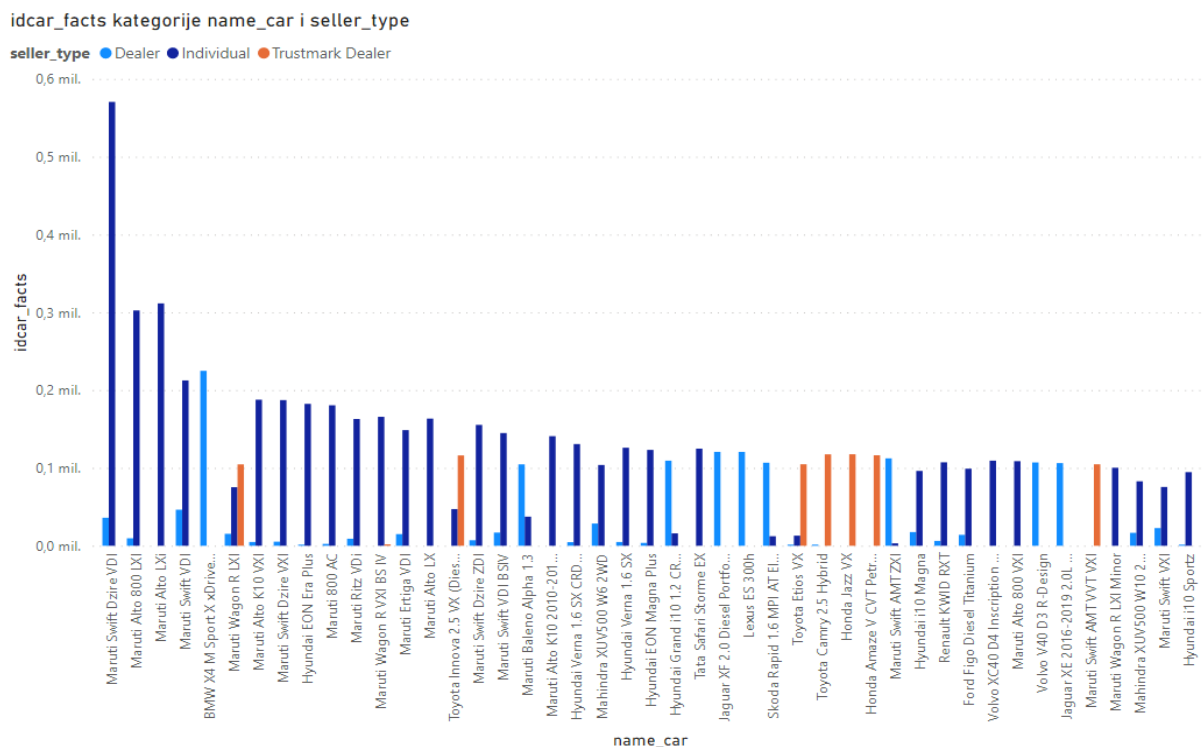
km\_driven kategorije name\_car



Četvrti izvještaj vezan je uz nazive marke automobila, njihov ukupni broj, i vrstu prodavača. Dakle, unutar grafikona možemo vidjeti kojeg automobila ima najviše, točnije koja se marka automobila najviše ponavlja unutar našeg skladišta podataka te koji tip prodavača (trgovac, pojedinac ili provjereni trgovac) najviše prodaje isti.



Slika 22. Grupirani trakasti grafikon ukupnog broja određene marke automobila i vrste prodavača istog



Slika 23. Linijski i grupirani grafikon ukupnog broja određene marke automobila i vrste prodavača istog

## 6. Zaključak

Kako podaci danas imaju vrlo veliku ulogu gotovo svakog aspekta života, organizacija istih od velike je važnosti. U tu svrhu postoje različiti načini na koje možemo podatke grupirati, razvrstavati i skladištiti, kako bi pristup njima bio što jednostavniji i efikasniji. Isto tako je veoma bitna komponenta transformiranje podataka kako bi isti bili kvalitetni u donošenju raznih analiza, i u našem slučaju, kreiranju raznih vrsta izvještaja. Tu vrlo veliku ulogu igraju skladišta podataka i poslovna inteligencija. Zahvaljujući skladištima podataka mi možemo sve podatke dobro analizirati, transformirati i razvrstati tako da je pretraživanje istih na kvalitetnoj razini. Proces pronalaska, transformacije i učitavanja podataka u kreirano skladište podataka te kasnija analiza istih, uz kreiranje izvještaja, je bio i cilj ovog projektnog zadatka: U tu svrhu smo krenuli sa sustavom *Kaggle* gdje smo pronašli željene podatke, korištene u daljnjoj obradi. Uz spomenuti sustav koristili smo i alate MySQL Workbench i Power BI. Koristeći MySQL Workbench kreirali smo skladište podataka i odabrane podatke, nakon pronalaska i transformacije, učitali u isto. Zatim smo ih organizirali u pripadajuće dimenzijske i činjenične tablice koje su nam u konačnici bile od koristi prilikom konačne analize i ispisa izvještaja. Naš primjer pokazuje koliko je velika prednost korištenja alata namijenjenih obradi samih podataka, na način opisan kroz cijeli projektni rad, te bi svakome bio plus uložiti trud u shvaćanje cjelokupnog funkcioniranja skladišta podataka jer će istom biti sigurno od koristi prilikom donošenja kvalitetnih i preciznijih poslovnih odluka, ali i drugih.

## 7. Popis literature

- [1] CarDekho, »CarDekho,« 2021.. [Mrežno]. Dostupno na: <https://www.cardekho.com/>. [Pokušaj pristupa 2. 6. 2021.].
- [2] R. B, »Hostinger,« 18. 5. 2018.. [Mrežno]. Dostupno na: <https://www.hostinger.com/tutorials/what-is-mysql>. [Pokušaj pristupa 2. 6. 2021.].
- [3] K. Lee, »Chron,« [Mrežno]. Dostupno na: <https://smallbusiness.chron.com/mysql-workbench-36016.html>. [Pokušaj pristupa 2. 6. 2021.].
- [4] Guru, »Guru99,« [Mrežno]. Dostupno na: <https://www.guru99.com/power-bi-tutorial.html>. [Pokušaj pristupa 2. 6. 2021.].
- [5] Guru, »Guru,« [Mrežno]. Dostupno na: <https://www.guru99.com/etl-extract-load-process.html>. [Pokušaj pristupa 2. 6. 2021.].
- [6] M. Sanchez-Ayala, »Data Modeling: The Star Schema,« 6. 4. 2020.. [Mrežno]. Dostupno na: <https://medium.com/@marcosanchezayala/data-modeling-the-star-schema-c37e7652e206>. [Pokušaj pristupa 2. 6. 2021.].

## 8. Popis slika

Slika 1. Prikaz atributa i njihovih tipova .....	2
Slika 2. Pregled dijela izvornih podataka .csv datoteke .....	5
Slika 3. Prikaz početnog sučelja MySQL Workbencha .....	6
Slika 4. Prikaz baze podataka "sppi" .....	6
Slika 5. Pregled "Table Data Import" sučelja .....	7
Slika 6. Prikaz generalne strukture zvjezdanog modela .....	8
Slika 7. Prikaz kreirane tablice fuel .....	9
Slika 8. Prikaz kreirane tablice owner .....	10
Slika 9. Prikaz kreirane tablice owner .....	11
Slika 10. Prikaz kreirane tablice transmission .....	12
Slika 11. Prikaz kreirane tablice car .....	14
Slika 12. Prikaz kreirane činjenične tablice car_facts .....	16
Slika 13. Prikaz početnog sučelja Power BI alata .....	17
Slika 14. Model zvijezde odabranog skladišta podataka .....	18
Slika 15. Prikaz dostupnih podataka i raznih oblika izvještaja .....	19
Slika 16. Prstenasti grafikon marke automobila i sume njihove prodajne cijene .....	20
Slika 17. Grupirani trakasti grafikon marke automobila i sume njihove prodajne cijene .....	20
Slika 18. Prstenasti grafikon broja automobila s obzirom na broj sjedala .....	21
Slika 19. Tortni grafikon broja automobila s obzirom na broj sjedala .....	21
Slika 20. Složeni stupičasti grafikon naziva marke automobila i ukupnog broja prevoženih kilometara .....	22
Slika 21. Tortni grafikon naziva marki automobila i ukupnog broja prevoženih kilometara .....	22
Slika 22. Grupirani trakasti grafikon ukupnog broja određene marke automobila i vrste prodavača istog .....	23
Slika 23. Linijski i grupirani grafikon ukupnog broja određene marke automobila i vrste prodavača istog .....	23