

Sveučilište Jurja Dobrile u Puli  
Fakultet Informatike u Puli

Mateo Zlatunić

Skup podataka prodaje proizvoda na globalnoj razini  
Seminarski rad

Pula, 21. svibnja, 2024.

Sveučilište Jurja Dobrile u Puli  
Fakultet Informatike u Puli

Mateo Zlatunić

Skup podataka prodaje proizvoda na globalnoj razini  
Seminarski rad

JMBAG: 030101458

Status: redoviti student

Studijski smjer: Informatika

Kolegij: Skladišta i rudarenje podataka

Mentor: doc.dr.sc Goran Oreški

Pula, 21. svibnja, 2024.



## IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Mateo Zlatunić, ovime izjavljujem da je ovaj seminarski rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio seminarskog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

 Student

Pula, 21. svibnja, 2024.

## Sadržaj

Uvod .....	1
Skup podataka.....	2
Pravilno korištenje <i>dataset-a</i> .....	2
Odabir dataset-a.....	2
Analiza dataset-a .....	3
Relacijski model podataka .....	4
Pred procesiranje.....	4
Kreiranje <i>ER</i> modela.....	4
Stvaranje i punjenje baze.....	6
Kreiranje <i>EER</i> modela .....	6
Dimenzijski modeli .....	8
Izrada modela .....	8
ETL proces: Pentaho Data Integration .....	12
Kreiranje dimenzijkog modela <i>location</i> .....	13
Unos skupa i baze podata.....	13
Transformacija podataka.....	14
Učitavanje i kreiranje podataka u dimenzijsku bazu podataka .....	15
Kreiranje dimenzijkog modela <i>market</i> .....	17
Kreiranje dimenzijskog modela <i>category</i> .....	18
Kreiranje dimenzijkog modela <i>sub-category</i> .....	19
Kreiranje dimenzijkog modela <i>orders</i> .....	20
Kreiranje dimenzijkog modela <i>segment</i> .....	21
Kreiranje dimenzijkog modela <i>customer</i> .....	22
Kreiranje dimenzijkog modela <i>product</i> .....	23
Kreiranje dimenzijkog modela <i>sold_goods</i> .....	24

Kreiranje tablice činjenica .....	25
Grafički prikaz podataka .....	27
Zaključak.....	36
Literatura .....	38

# Uvod

Poslovna inteligencija ključna je komponenta za donošenje informacijskih odluka i ostvarivanje prednosti nad konkurencijom. Korištenjem alata i tehnika za analizu podataka, ima ulogu u otkrivanju uzoraka, trendova i ključnih informacija koje mogu utjecati na poslovanje organizacije. U kontekstu samog seminarskog rada, prikupljeni podaci omogućit će tvrtkama da prate performanse proizvoda na različitim tržištima, rastuće trendove potražnje, te razvijanje marketinške strategije za potrebe potrošača.

Cilj projekta jest praktično i teorijski opisati, problemski proces obrađivanja skupa podataka, odrediti ključne faktore koji utječu na uspješnost prodaje proizvoda, te krajnji ishod prikazivanja i korištenja istih podataka u svrhe donošenja poslovnih odluka, na temelju prošlih aktivnosti. Proces će se sastojati od provjere točnosti količine podataka iz pripadajućeg skupa, pred procesiranja nepotrebnih atributa, te spremanja istih u bazu podataka i prikaza EER modela, zajedno sa dimenzijskim modelom i završnom fazom grafičkog prikaza podataka.

# Skup podataka

## Pravilno korištenje *dataset-a*

Određivanje skupova podataka bitan je segment u samom uvodu poslovnog procesa, iz razloga što će svi ostali procesi u konačnici ovisiti o samom skupu koji se od početka i koristio. Važno je koristiti skup dovoljnih količina podataka, uzimaju u obzir već odvijene događaje, te da jasno i kvalitetno opisuju okolnosti procesa. Obzirom da analiza generalno sadrži vremenske okolnosti, sastoji se od kvantitativnih i kvalitativnih podataka, točnije numeričkih izraza i opisnih podataka. Po mogućnosti, potrebno je da prikazuje različite i smislenije tipove podataka, te da ne sadrži nepostojeće vrijednosti, prilikom spremanja istih u bazu podataka.

## Odabir dataset-a

*Dataset* odabran za analizu podataka opisuje prodaju proizvoda na globalnoj razini, bez indikatora u ime koje organizacije su isti namijenjeni, što ne predstavlja toliki problem, jer sama količina podataka zadovoljava kriterije od kojih se pravilan skup u osnovi mora i sastojati, te podaci koji se grafički budu prikazivali, bili bi validni i za organizaciju koja ih bude koristila za donošenje odluka. *Dataset* za koji će se detaljno razmatrati poslovni proces, dostupan je preko sljedeće poveznice:

<https://www.kaggle.com/datasets/ronysoliman/global-superstore-dataset>

## Analiza dataset-a

Skup podataka prodaje proizvoda sadrži preko 50 tisuća redova i 20 atributa, što je idealna količina, jer će procedura transformacije podataka biti brža. Zastupljeni su podaci o proizvodu, lokaciji, kupcima, segmentima, troškovima i još mnogo toga. Takva veličina omogućuje provođenje detaljnijih analiza, pri čemu skupovi podataka pružaju bolji pregled tržišta, trendova i obrazaca o prodaji. Naravno, ovisno je od situacije do situacije kakav omjer podataka je prihvatljiv za daljnje analiziranje, te kroz koje procedure će sirovi podaci i prolaziti. Prema podacima, skup ne sadrži nedostajuće vrijednosti. Ovo je dobro jer eliminira potrebu za dodatnom obradom podataka prije analize. Nedostajuće vrijednosti mogu uzrokovati probleme u analizi i dovesti do netočnih zaključaka. Činjenica da su svi podaci kompletni olakšava proces analize i omogućuje fokusiranje na daljnju razradu, umjesto čišćenja podataka. Skup sadrži različite tipova podataka, što će kasnije biti korisno za buduće procese i za krajnji grafički prikaz rezultata, jer što je više raznolikosti, bolja će kombinacija biti između kvalitativnih i kvantitativnih podataka.

Također, postoji vremenska dimenzija, prikazujući datume naručivanja i dostavljanja proizvoda krajnjem korisniku. Moguće je uspoređivati prodaju u različitim vremenskim razdobljima i omogućuje predviđanje budućih prodajnih trendova na temelju povijesnih podataka. U suštini, baza podataka predstavlja 8 dimenzijskih modela, povezano sa tablicom činjenica, kvantitativnih tipova podataka. Kvantitativni podaci, od ostalih stvari uključuju brojke poput cijena, troškova i količina, dok s druge strane kvalitativni podaci tipa naziva proizvoda, lokacija i opisa kupaca pružaju kontekst i dublje razumijevanje prodajnih obrazaca. Kombinacijom ovih dviju vrsta dobiva se cjelovita slika donošenja informativnih poslovnih odluka.



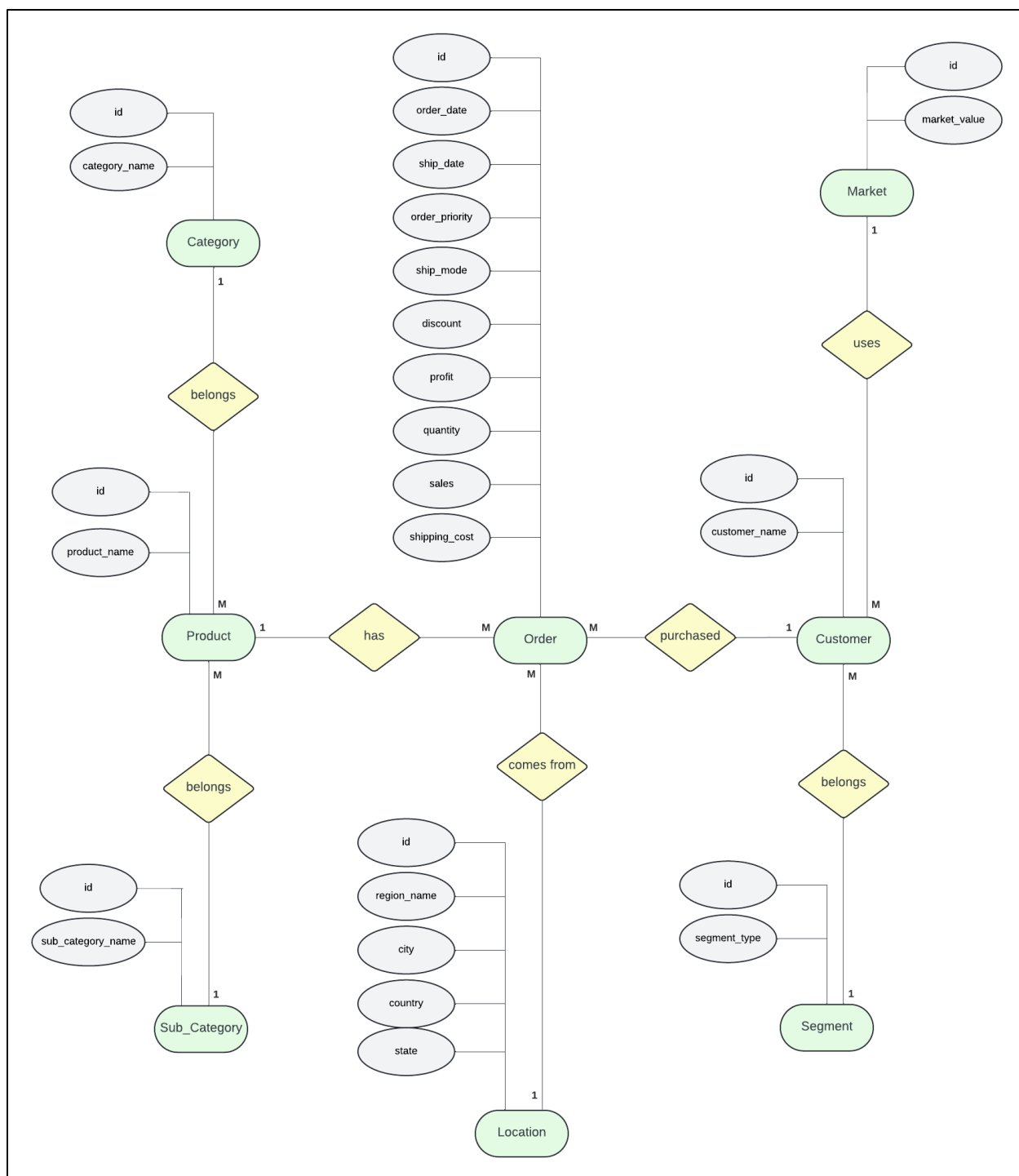
# Relacijski model podataka

## Pred procesiranje

Nakon definiranja skupa podataka i analiziranja njegovog sadržaja, dolazi pred procesiranje podataka, točnije uklanjanje nepotrebnih vrijednosti ili uređivanje onih vrijednosti koji baš nisu odgovarali pripadajućem atributu. U slučaju prodaje proizvoda, iz skupa je uklonjeno nekoliko atributa jedinstvenih ključeva, iz razloga što nisu jedinstveno predstavljali svaki red posebno. Datoteka se u zadnjoj fazi pred procesiranja dijeli na dvije datoteke. Prvi dio sadrži 80% podataka, zadužen za punjenje baze u *MySQL Workbench* programu, a drugi će se koristiti tokom ETL procesa.

## Kreiranje *ER* modela

Sljedeći korak jest logički odrediti entitete pripadajućih atributa, sa odgovarajućim kardinalnim vezama. Kao entiteti, definirani su: *category*, *product*, *sub\_category*, *order*, *location*, *market*, *customer* i *segment*. Prevladavajuća veza je *many to one* ili *one to many*, ovisno o promatranom entitetu preko kojeg se uspoređuje odnos sa drugim entitetom. Grafičkim prikazom, *ER* model praktično opisuje buduće stanje relacija unutar baze podataka.



Slika 1: ER model prodaje proizvoda

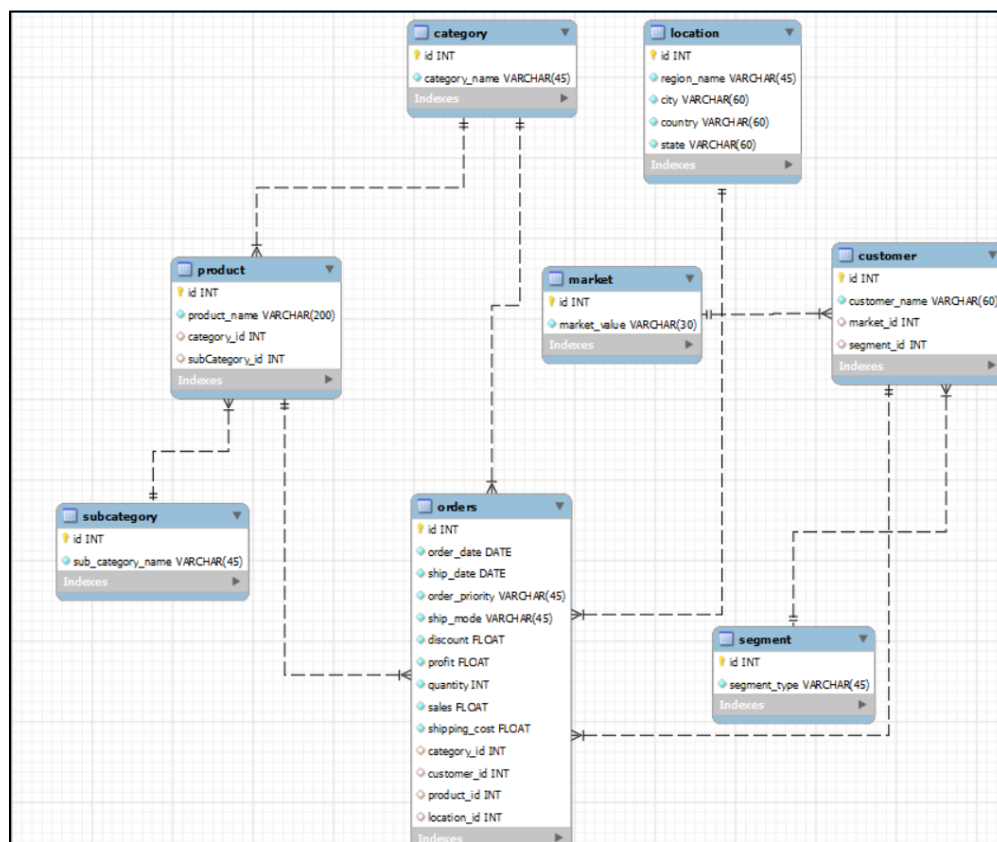
## Stvaranje i punjenje baze

Sljedeća faza jest pomoću *python* skripte stvoriti shemu baze podataka koristeći *SQLAlchemy ORM* i popuniti tu bazu podacima, pred procesiranim skupom podataka iz CSV formata. Pošto već postoje dvije kreirane datoteke, koristit će se ona koja sadrži do 80% podataka i učitati putem *Pandas-a*. Korištenjem *SQLAlchemy ORM-a*, definiraju se tablice *Order*, *Customer*, *Location*, *Product*, *Category*, *SubCategory*, *Market*, i *Segment*. Svaka klasa predstavlja jednu tablicu, a atributi tih klasa definiraju stupce tablica u našoj bazi. Konekcija sa *MySQL* bazom ostvaruje se definiranjem naziva *host-a*, korisničkog imena, lozinke i naziva sheme. Postojeće tablice se brišu, a zatim se kreiraju nove prema definiranom modelu. Prolazi se kroz svaki redak i stvara se odgovarajući zapis za svaku tablicu. Obraćena je pažnja na tablicu *Location*, gdje se provjerava postoji li zapis s istim atributima, kako bi se izbjeglo dupliciranje. Tablice bez stranih ključeva, prvenstveno se ubacuju u bazu, kako bi se njihov *ID* kasnije mogao iskoristiti za one tablice koje sadrže strane ključeve, uspostavljajući odnos između tablica. Na kraju procesa, sve promjene su spremljene u bazu i sesija se zatvara.

## Kreiranje *EER* modela

Nakon unošenja podataka u postojeću bazu, slijedi kreiranje *EER* modela. Proces se gotovo automatski izvršava, korištenjem *Reverse Engineer* opcije, samo je bitno definirati shemu. Ovakav tip sličan je *ER* modelu, samo detaljnije obuhvaća zahtjeve korisnika, pružajući jasniji i konzistentniji dizajn baze podataka. Sadrži drugačiji zapis kardinalnih veza entiteta, ali prikazuje istu količinu entiteta kao i *ER* model.

Slika 2: Uspješno kreiran DB *superstore*



Slika 3: EER model

## Dimenzijski modeli

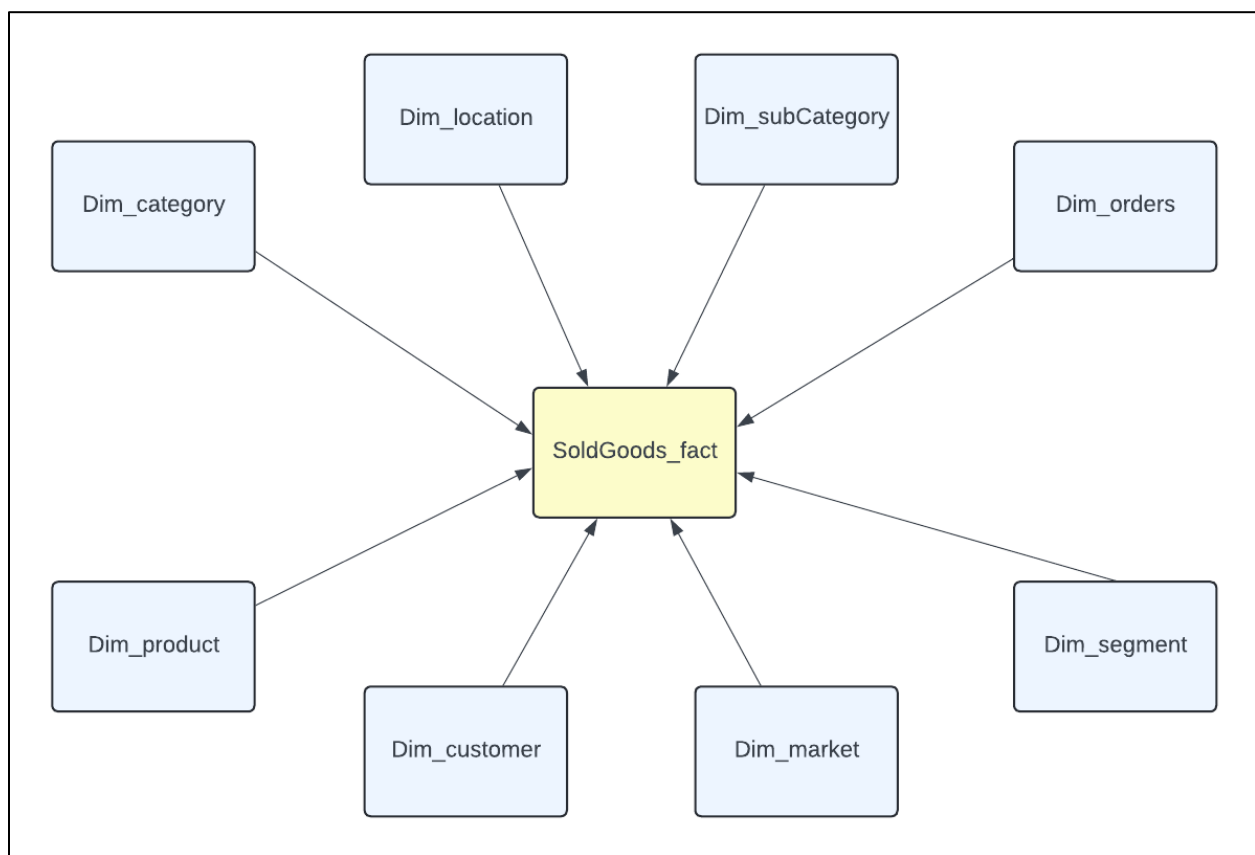
Dimenzijsko modeliranje odnosi se na prikaz analitičkih podataka krajnjim korisnicima, omogućavajući brže performanse upita, na temelju svoje jednostavno definirane strukture u odnosu na ostale modele. Struktura je jednostavna iz razloga što se sastoji od dimenzijskih tablica i tablice činjenica. Dimenzije sadrže opisne podatke, dok tablica činjenica se sastoji od brojčanih vrijednosti i definiranih vremenskih okolnosti. Svaka tablica ima svoj surogat ključ, ali samo tablica činjenica sadrži strane ključeve ostalih dimenzija. Na taj način lakše se prikupljaju i prikazuju podatci, te će se moći grafički prikazati.

### Izrada modela

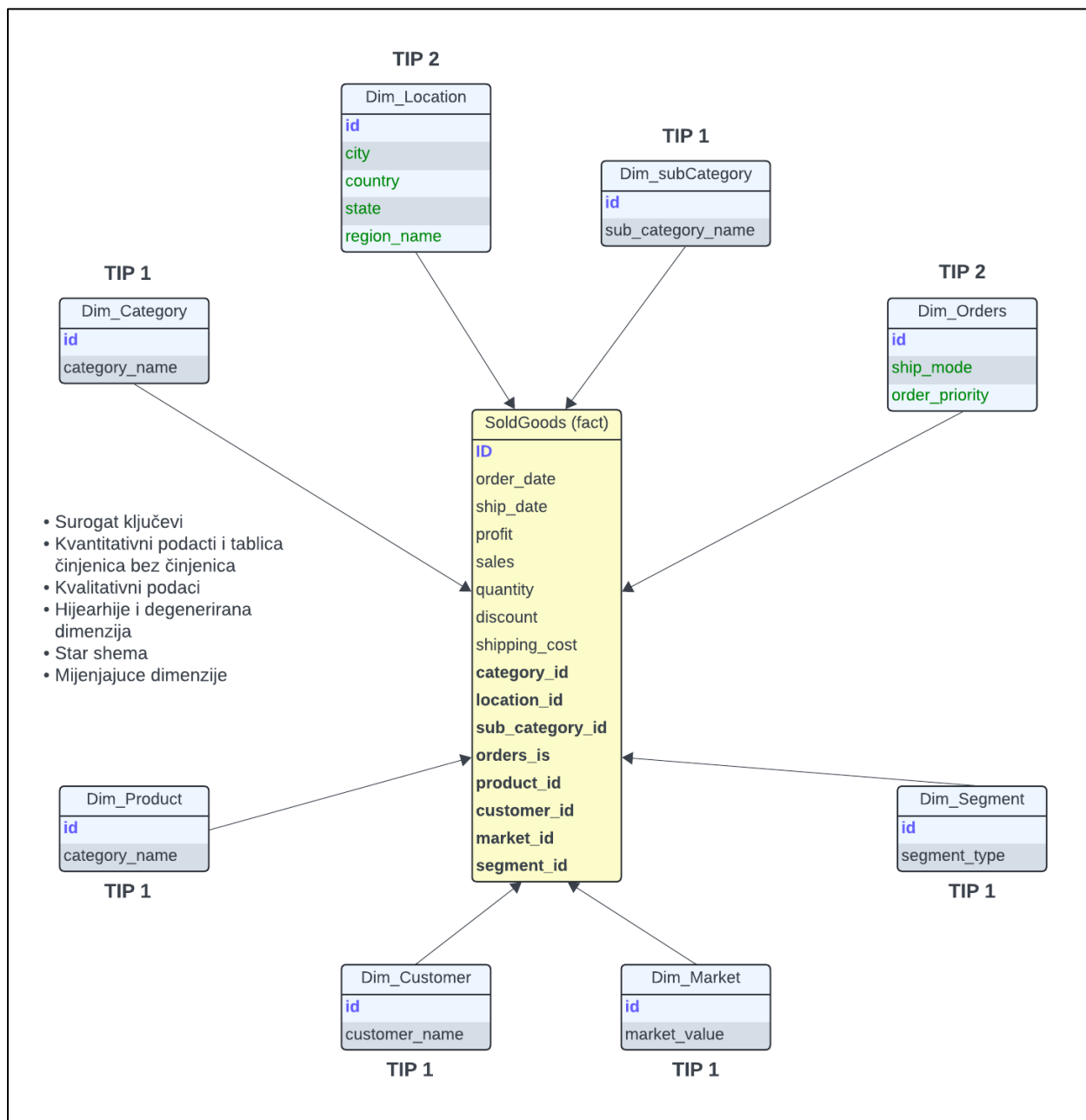
Prodaja proizvoda predstavljena je kao *star* shema, jer susjedne dimenzije se ne sastoji od vlastitih pod dimenzija. Model prikazuje 8 dimenzija i jednu tablicu činjenica, kao i hijerarhijsko sortiranje atributa po određenoj važnosti, kao što je slučaj sa tablicom *lokacija*. Takva dimenzijska tablica sadrži nazive gradova, država, zemalja i regija, lokacije u kojima naši kupci prebivaju i iz kojih naručuju različite proizvode. Vremenska dimenzija je definirana po datumu naručivanja i dostavljanja, te su podaci smješteni pod tablicu činjenica. Količina, prodaja, popusti, cijena dostave i profit proizvoda također pripadaju tablici činjenica, jer takvi podaci čine kvantitativne tipove podataka. Za dimenzijske tablice definirane su mogućnosti izvedbe prvog i drugog tipa. Tablice koje su prvog tipa, izmjenjuju stare vrijednosti atributa novijom, dok tablice drugog tipa pridodaju nove vrijednosti u dimenzije. Pomoću *python* skripte, kreirana je baza podataka *pentaho*, čija svrha jest pohraniti podatke u dimenzijske modele, u svrhe praktičnog prikaza ETL procesa, kao što je prikazano na navedenim slikama.

Opis dimenzijskih tablica:

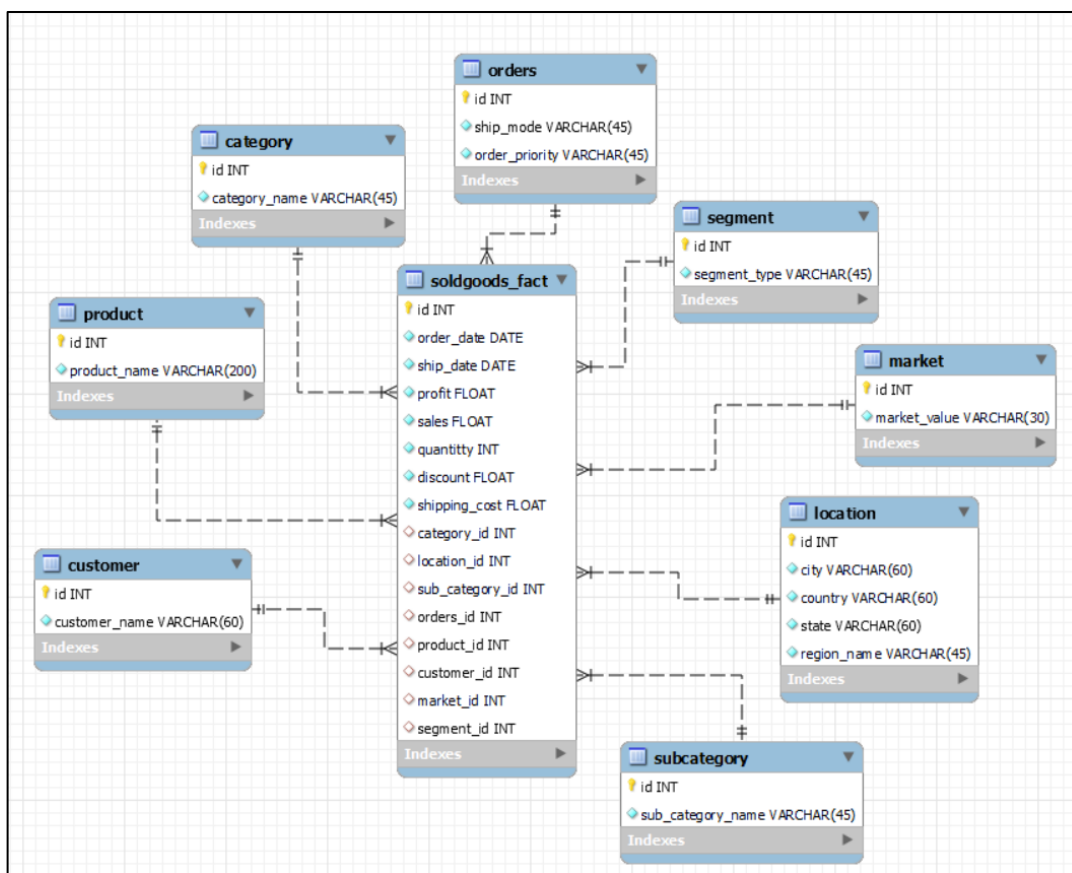
- Dim\_Category – opisuje kategoriju proizvoda
- Dim\_subCategory – opisuje pod kategoriju proizvoda
- Dim\_Product – naziv samog proizvoda
- Dim\_Location – lokacije u kojima kupci prebivaju
- Dim\_Orders – vezano za prioritetsnu klasu dostave proizvoda
- Dim\_Segment – tip namjene proizvoda
- Dim\_Market – opisuje pripadnu geografsku regiju kupaca
- Dim\_Customer – naziv kupaca koji obavljaju narudžbe
- Dim\_Sold\_Goods – tablica činjenica



Slika 4: Skica dimenzijskih modela



Slika 5: Prikaz dimenzijskog modeliranja



Slika 6: Dimenzijsko modeliranje u MySQL programu

```

24 # <=====
25 # DROP DATABASE pentaho;
26 # CREATE DATABASE pentaho;
27 • USE pentaho;
28 # <=====
29
30 • SELECT * FROM sold_goods;
31
32 • SELECT COUNT(*) FROM category;
33 • SELECT COUNT(*) FROM subcategory;
34 • SELECT COUNT(*) FROM customer;
35 • SELECT COUNT(*) FROM location;
36 • SELECT COUNT(*) FROM market;
37 • SELECT COUNT(*) FROM orders;
38 • SELECT COUNT(*) FROM product;
39 • SELECT COUNT(*) FROM segment;
40 • SELECT COUNT(*) FROM sold_goods;

```

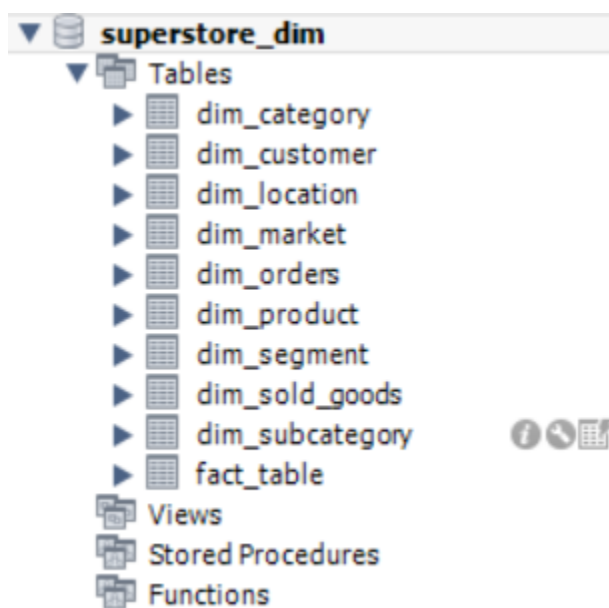
	id	order_date	ship_date	profit	quantity	sales	shipping_cost	discount	category_id	customer_id	product_id	location_id	sub_category
1	1	2021-06-23	2021-06-27	-2.044	7	121.716	7.756	0.4	1	1	1	1	1
2	2	2020-11-10	2020-11-16	13.02	2	81.66	6.78	0	1	2	2	2	1
3	3	2020-11-19	2020-11-23	3.9032	2	27.88	5.48	0	3	3	3	3	3
4	4	2018-12-26	2018-12-28	-280.92	3	312.12	31.529	0.6	4	4	4	4	4
5	5	2019-10-16	2019-10-21	-202.98	3	434.88	31.63	0.4	4	5	5	5	4
6	6	2020-10-14	2020-10-16	-39.6	6	158.22	28.92	0.4	4	6	6	6	6
7	7	2020-02-14	2020-02-21	3.5752	4	8.72	1.03	0	1	7	7	7	7
8	8	2020-02-06	2020-02-10	208.98	3	950.22	131.43	0	3	8	8	8	8
9	9	2021-10-21	2021-10-23	-28.9568	4	579.136	103.4	0.2	4	9	9	9	9
10	10	2020-09-04	2020-09-11	0	3	27.34	0.76	0	1	10	10	10	10

Slika 7: Uspješno kreirana baza podataka *pentaho*, za ETL proces



## ETL proces: Pentaho Data Integration

Do sada se vršila analiza podataka, izrada ER modela, kreiranje i spremanje podataka u određenu bazu, te izrada dimenzijskih modela. Sljedeći korak jest kreirati postupak spremanja podataka u bazu, točnije u bazu dimenzijskih modela. Proces se sastoji ekstrakcije, čija je svrha prikupiti podatke iz različitih izvora, pohranjujući ih za daljnju obradu. Transformacije, vršeći čišćenje, formatiranje i manipulaciju podataka, kako bi bili konzistentni i korisni za analizu, te posljednje faze što jest učitavanje, opisujući postupak pohrane transformiranih podataka u određeni sustav. Korišten alat pod ovakve svrhe biti će *pentaho data integration*, softver namijenjen za kreiranje ETL procesa. U uvodnom dijelu kreirana su dva izvora podataka, jedan predstavlja skup, a drugi pohranu podataka u bazu *pentaho*. Koristeći navedena dva izvora, prikazati će se kako zapravo to izgleda u praksi. Prvo, u prošlim kreiranim prikazima dimenzijskih modela, može se primijetiti količina dimenzija, uključujući i tablicu činjenica. Obrada će se prvenstveno vršiti na kvalitativnim podacima, te za kraj na kvantitativnim tipovima podataka. Također, za ETL proces kreirana je baza podataka, isključivo vezana za skladištenje dobivenih dimenzijskih modela.

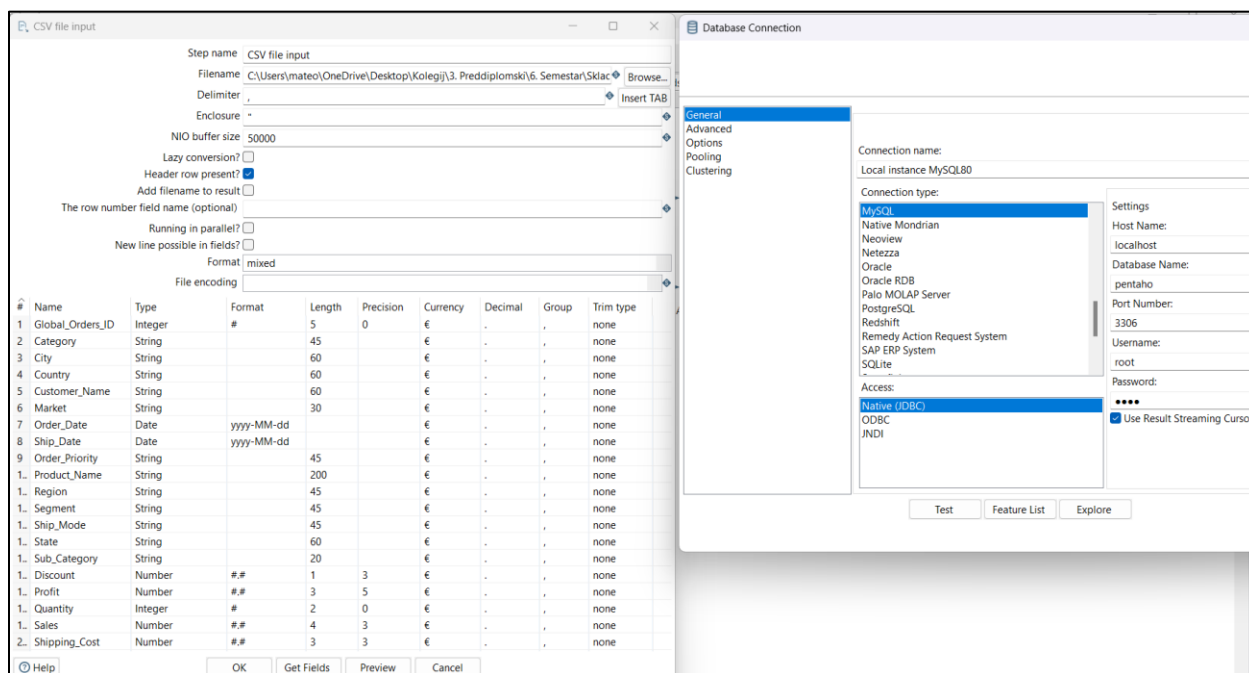


Slika 8: Dimenzijska baza podataka *superstore\_dim*

## Kreiranje dimenzijskog modela *location*

### Unos skupa i baze podata

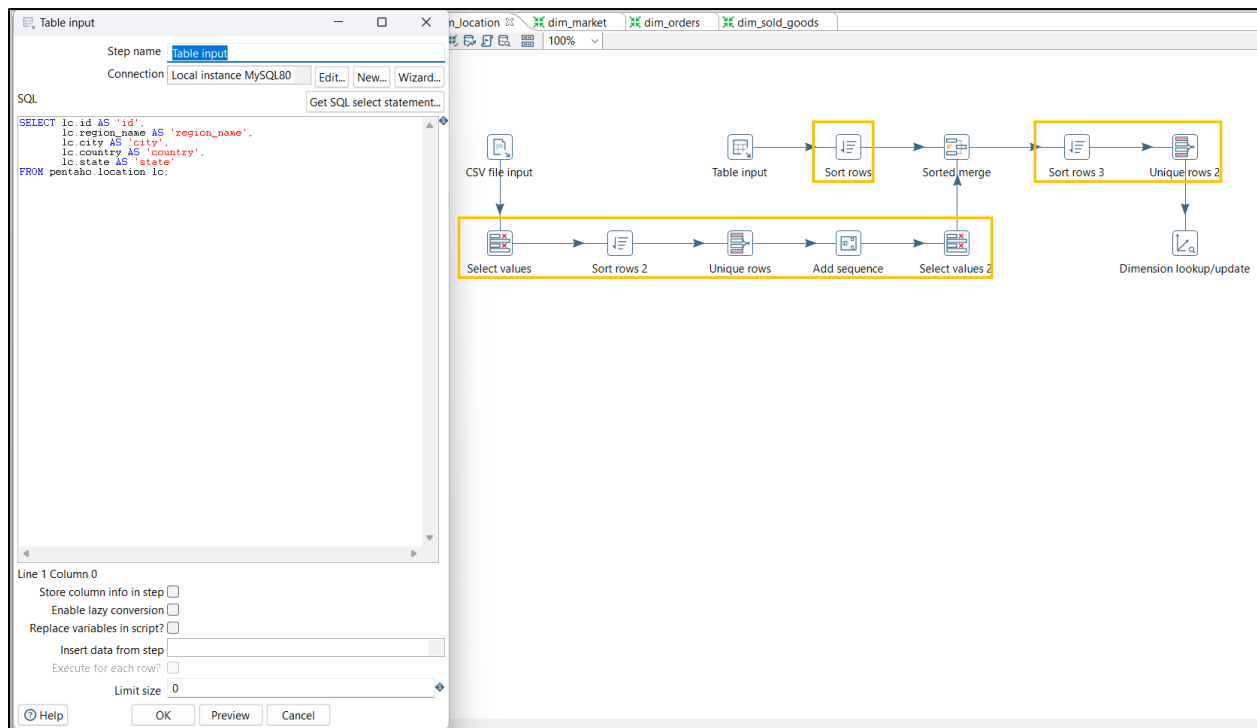
Prodaja proizvoda sadrži neiskorišten skup od 20% podataka, dok ostalih 80% je pohranjeno u *pentaho* bazu podataka. Prvi korak bi bio učitati bazu i skup podataka, koji će se koristiti kroz ostatak procesa. Iz baze se dohvaća tablica *location*, koja od ostalih stvari opisuje regije i države prebivališta kupaca, što će kasnije činiti naš dimenzijski model. Jednom kada se učitava skup podataka, važno je definirati njegove tipova podataka.



Slika 9: Kreirana baza i skup podataka

## Transformacija podataka

Slijedi proces transformacije, u kojem se za bazu, korištenjem *SQL* programskog jezika stvara upit, namijenjen za dohvaćanje svih atributa tablice lokacija. Podaci iz definiranog skupa, prolaze kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti i odabir sekvence. Relacije baze podataka mogu se samo sortirati, jer je baza podataka već obrađena. Sortirana baza i odabrane vrijednosti skupa podataka spajaju se na temelju zajedničkih vrijednosti, što bi bio njihov *ID*. Zatim se vrši proces sortiranja kao i proces odabira jedinstvenih vrijednosti.



Slika 10: MySQL sintaksa i označena transformacija podataka

## Učitavanje i kreiranje podataka u dimenzijsku bazu podataka

Za završni dio definira se naziv modela i pripadna shema unutar baze podataka. Obzirom da se kreira dimenzijska tablica, morat će unijeti surogat ključ, kako bi se kasnije mogla povezati sa tablicom činjenica. Navode se atributi *date\_from*, *date\_to* i *version* za definiranje sporo mijenjajuće dimenzije. Prilikom definiranja dimenzije, obzirom da smo spojili podatke sa različitih izvora, pod opciju *ključa* navodimo samo ID za lakše prepoznavanje podataka, dok u opciju *polja* normalno navodimo ostale vrijednosti koje smo bili kreirali.

Step name: Dimension lookup/update

Update the dimension? ☒

Connection: Local instance MySQL80

Target schema: superstore\_dim

Target table: dim\_location

Commit size: 100

Enable the cache? ☒

Pre-load the cache? ☐

Cache size in rows (0 = cache all): 5000

Keys Fields

#	Dimension field	Stream field to compare with	Type of dimension update
1	region_name	region_name	Insert
2	city	city	Insert
3	country	country	Insert
4	state	state	Insert

Technical key field: location\_tk

New name:

Creation of technical key

☒ Use table maximum + 1

☐ Use sequence

☐ Use auto increment field

Version field: version

Stream Datefield:

Date range start field: date\_from

Min. year: 2010

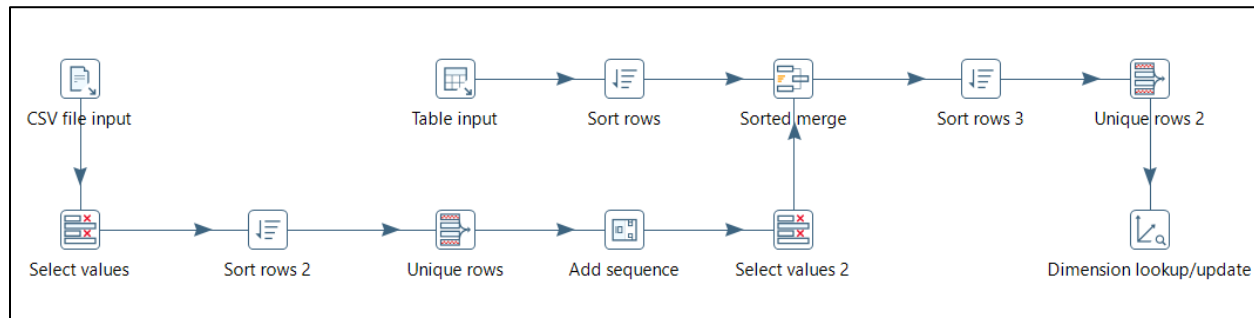
Use an alternative start date? ☐ <Select Options>

Table date range end: date\_to

Max. year: 2030

OK Cancel Get Fields SQL

Slika 11: Dimenzijski proces kreiranja tablice



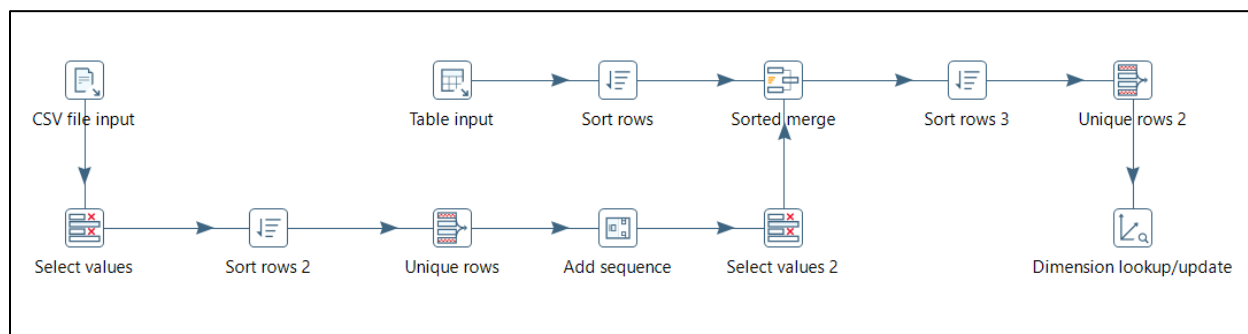
Slika 12: ETL proces dimenzije *location*

version	location_tk	date_from	date_to	id	region_name	country	city	state
1	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL
1	1	2010-01-01	2031-01-01	1	South	Peru	Lima	Lima (city)
1	2	2010-01-01	2031-01-01	2	North	United Kingdom	Oxford	England
1	3	2010-01-01	2031-01-01	3	West	United States	Montebello	California
1	4	2010-01-01	2031-01-01	4	South	Brazil	Salvador	Bahia
1	5	2010-01-01	2031-01-01	5	Caribbean	Dominican Republic	Santo Domingo	Santo Domingo
1	6	2010-01-01	2031-01-01	6	Oceania	Australia	Canberra	Australian Capital Territory
1	7	2010-01-01	2031-01-01	7	East	United States	New York City	New York
1	8	2010-01-01	2031-01-01	8	South	Spain	Madrid	Madrid
1	9	2010-01-01	2031-01-01	9	West	United States	Fort Collins	Colorado
1	10	2010-01-01	2031-01-01	10	Central Asia	India	Avadi	Tamil Nadu
1	11	2010-01-01	2031-01-01	11	West	United States	Seattle	Washington
1	12	2010-01-01	2031-01-01	12	Central	France	Montereau-Fa...	Île-de-France
1	13	2010-01-01	2031-01-01	13	North Asia	South Korea	Seoul	Seoul
1	14	2010-01-01	2031-01-01	14	North Asia	China	Zhuhai	Guangdong
1	15	2010-01-01	2031-01-01	15	South	Italy	Brindisi	Apulia
1	16	2010-01-01	2031-01-01	16	South	Italy	Palermo	Sicily
1	17	2010-01-01	2031-01-01	17	West	United States	Los Angeles	California

Slika 13: Dimenzijski prikaz podataka tablice *location*

## Kreiranje dimenzijskog modela *market*

Učitavajući bazu i skup podataka, na isti način koristimo izvore podataka. Kada uspostavimo konekciju sa bazom, putem *MySQL* upita dohvaćamo sve atribute vezane za tablicu *market*, koja opisuje tržište, odnosno geo lokaciju regija. Postepeno se prolazi kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti, te odabir sekvence. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*. Detaljniji postupak prikazan je na poglavlju yyyy.



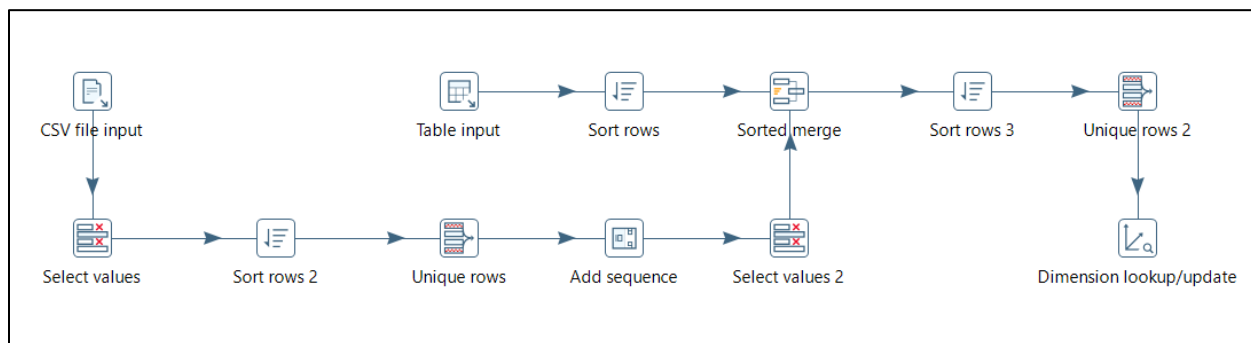
Slika 14: ETL proces dimenzije *market*

	version	market_tk	date_from	date_to	id	market_value
▶	1	0	NULL	NULL	NULL	NULL
	1	1	2010-01-01	2031-01-01	1	LATAM
	1	2	2010-01-01	2031-01-01	2	EMEA
	1	3	2010-01-01	2031-01-01	3	USCA
	1	4	2010-01-01	2031-01-01	4	LATAM
	1	5	2010-01-01	2031-01-01	6	APAC
	1	6	2010-01-01	2031-01-01	7	USCA
	1	7	2010-01-01	2031-01-01	8	EMEA
	1	8	2010-01-01	2031-01-01	9	USCA
	1	9	2010-01-01	2031-01-01	10	APAC
	1	10	2010-01-01	2031-01-01	11	USCA
	1	11	2010-01-01	2031-01-01	12	EMEA
	1	12	2010-01-01	2031-01-01	13	APAC
	1	13	2010-01-01	2031-01-01	15	EMEA
	1	14	2010-01-01	2031-01-01	17	USCA
	1	15	2010-01-01	2031-01-01	19	EMEA
	1	16	2010-01-01	2031-01-01	21	USCA
	1	17	2010-01-01	2031-01-01	22	EMEA
	1	18	2010-01-01	2031-01-01	24	USCA

Slika 15: Dimenzijski prikaz podataka tablice *market*

## Kreiranje dimenzijskog modela *category*

Učitavajući bazu i skup podataka, na isti način koristimo izvore podataka. Kada uspostavimo konekciju sa bazom, putem *MySQL* upita dohvaćamo sve atribute vezane za tablicu *category*, koja opisuje kategorije proizvoda. Postepeno se prolazi kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti, te odabir sekvence. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*. Detaljniji postupak prikazan je na poglavlju yyyy.



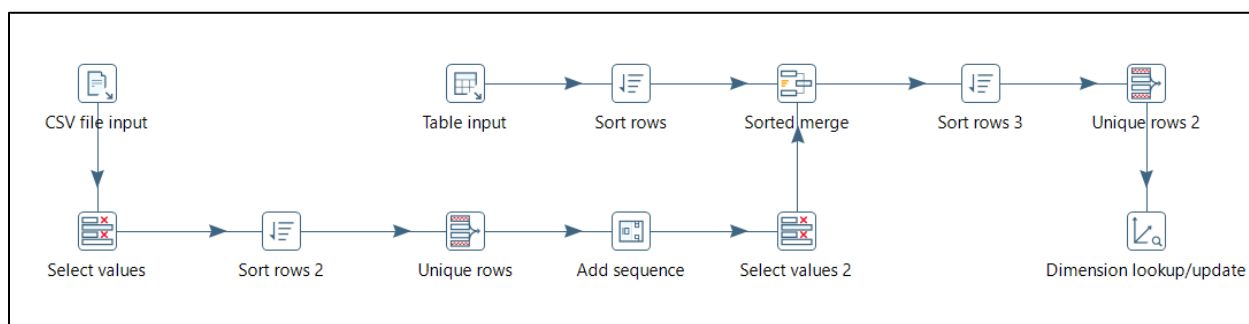
Slika 16: ETL proces dimenzije *category*

	version	category_tk	date_from	date_to	id	category_name
▶	1	0	NULL	NULL	NULL	NULL
	1	1	2010-01-01	2031-01-01	1	Office Supplies
	1	2	2010-01-01	2031-01-01	3	Technology
	1	3	2010-01-01	2031-01-01	4	Furniture
	1	4	2010-01-01	2031-01-01	7	Office Supplies
	1	5	2010-01-01	2031-01-01	8	Technology
	1	6	2010-01-01	2031-01-01	9	Furniture
	1	7	2010-01-01	2031-01-01	10	Office Supplies
	1	8	2010-01-01	2031-01-01	17	Furniture
	1	9	2010-01-01	2031-01-01	18	Office Supplies
	1	10	2010-01-01	2031-01-01	19	Furniture
	1	11	2010-01-01	2031-01-01	20	Office Supplies
	1	12	2010-01-01	2031-01-01	21	Furniture
	1	13	2010-01-01	2031-01-01	23	Office Supplies
	1	14	2010-01-01	2031-01-01	25	Furniture
	1	15	2010-01-01	2031-01-01	26	Office Supplies
	1	16	2010-01-01	2031-01-01	27	Furniture
	1	17	2010-01-01	2031-01-01	28	Office Supplies

Slika 17: Dimenzijski prikaz podataka tablice *category*

## Kreiranje dimenzijskog modela *sub-category*

Učitavajući bazu i skup podataka, na isti način koristimo izvore podataka. Kada uspostavimo konekciju sa bazom, putem *MySQL* upita dohvaćamo sve atribute vezane za tablicu *sub-category*, koja opisuje pod kategorije proizvoda. Postepeno se prolazi kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti, te odabir sekvence. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*. Detaljniji postupak prikazan je na poglavlju yyyy.



Slika 18: ETL proces dimenzije *sub-category*

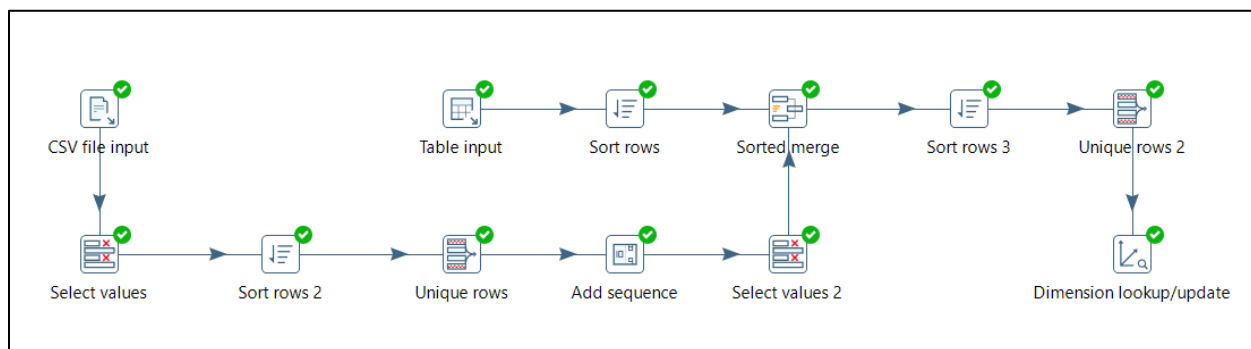
version	subCategory_tk	date_from	date_to	id	sub_category_name
1	0	NULL	NULL	NULL	NULL
1	1	2010-01-01	2031-01-01	1	Supplies
1	2	2010-01-01	2031-01-01	3	Accessories
1	3	2010-01-01	2031-01-01	4	Bookcases
1	4	2010-01-01	2031-01-01	6	Furnishings
1	5	2010-01-01	2031-01-01	7	Fasteners
1	6	2010-01-01	2031-01-01	8	Copiers
1	7	2010-01-01	2031-01-01	9	Chairs
1	8	2010-01-01	2031-01-01	10	Storage
1	9	2010-01-01	2031-01-01	11	Binders
1	10	2010-01-01	2031-01-01	12	Art
1	11	2010-01-01	2031-01-01	13	Storage
1	12	2010-01-01	2031-01-01	14	Binders
1	13	2010-01-01	2031-01-01	15	Paper
1	14	2010-01-01	2031-01-01	16	Storage
1	15	2010-01-01	2031-01-01	17	Chairs
1	16	2010-01-01	2031-01-01	18	Storage
1	17	2010-01-01	2031-01-01	19	Bookcases

Slika 19: Dimenzijski prikaz podataka tablice *sub-category*



## Kreiranje dimenzijkog modela *orders*

Učitavajući bazu i skup podataka, na isti način koristimo izvore podataka. Kada uspostavimo konekciju sa bazom, putem *MySQL* upita dohvaćamo sve atribute vezane za tablicu *orders*, koja opisuje uslugu dostavljanja proizvoda krajnjim kupcima. Postepeno se prolazi kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti, te odabir sekvence. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*. Detaljniji postupak prikazan je na poglavlju yyyy



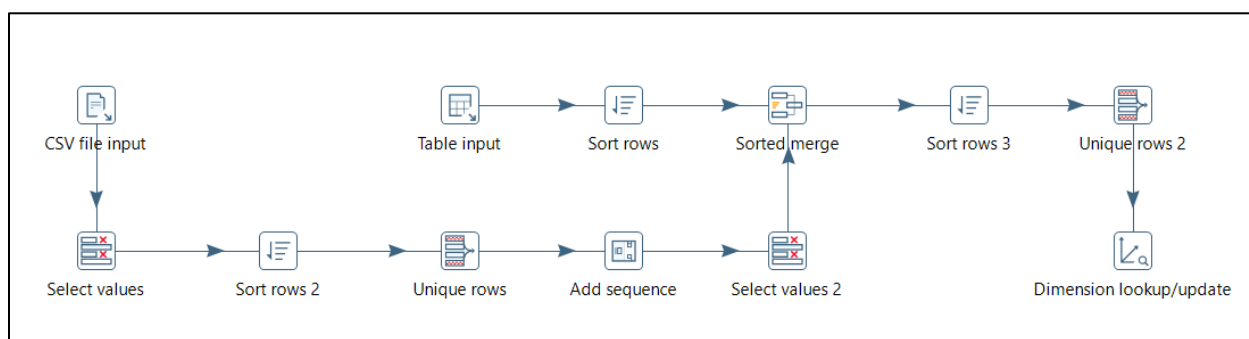
Slika 20: ETL proces dimenzije *orders*

Result Grid							
Filter Rows:							
Exports: Wrap Cell Contents:							
	version	order_tk	date_from	date_to	id	ship_mode	order_priority
▶	1	0	NULL	NULL	NULL	NULL	NULL
	1	1	2010-01-01	2031-01-01	1	Standard Class	Medium
	1	2	2010-01-01	2031-01-01	3	Standard Class	High
	1	3	2010-01-01	2031-01-01	5	Standard Class	Medium
	1	4	2010-01-01	2031-01-01	6	First Class	High
	1	5	2010-01-01	2031-01-01	7	Standard Class	Low
	1	6	2010-01-01	2031-01-01	8	Standard Class	High
	1	7	2010-01-01	2031-01-01	9	Second Class	Critical
	1	8	2010-01-01	2031-01-01	10	Standard Class	Medium
	1	9	2010-01-01	2031-01-01	11	First Class	High
	1	10	2010-01-01	2031-01-01	13	Standard Class	Low
	1	11	2010-01-01	2031-01-01	14	Standard Class	Medium
	1	12	2010-01-01	2031-01-01	15	Second Class	Critical
	1	13	2010-01-01	2031-01-01	16	Standard Class	Medium
	1	14	2010-01-01	2031-01-01	19	Second Class	High
	1	15	2010-01-01	2031-01-01	20	Second Class	Medium
	1	16	2010-01-01	2031-01-01	21	Standard Class	Low
	1	17	2010-01-01	2031-01-01	22	Standard Class	Medium

Slika 21: Dimenzijski prikaz podataka tablice *orders*

## Kreiranje dimenzijskog modela *segment*

Učitavajući bazu i skup podataka, na isti način koristimo izvore podataka. Kada uspostavimo konekciju sa bazom, putem *MySQL* upita dohvaćamo sve atribute vezane za tablicu *segment*, koja opisuje pod koje namjene će se proizvodi koristiti. Postepeno se prolazi kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti, te odabir sekvence. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*. Detaljniji postupak prikazan je na poglavlju yyyy.



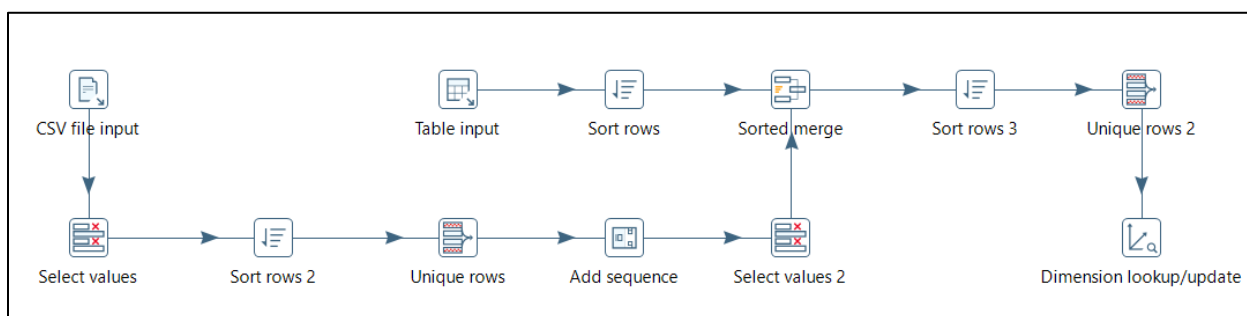
Slika 22: ETL proces dimenzije *segment*

	version	segment_tk	date_from	date_to	id	segment_type
▶	1	0	NULL	NULL	NULL	NULL
	1	1	2010-01-01	2031-01-01	1	Consumer
	1	2	2010-01-01	2031-01-01	2	Corporate
	1	3	2010-01-01	2031-01-01	4	Consumer
	1	4	2010-01-01	2031-01-01	6	Home Office
	1	5	2010-01-01	2031-01-01	7	Consumer
	1	6	2010-01-01	2031-01-01	10	Corporate
	1	7	2010-01-01	2031-01-01	12	Home Office
	1	8	2010-01-01	2031-01-01	13	Consumer
	1	9	2010-01-01	2031-01-01	15	Corporate
	1	10	2010-01-01	2031-01-01	16	Consumer
	1	11	2010-01-01	2031-01-01	18	Corporate
	1	12	2010-01-01	2031-01-01	19	Consumer
	1	13	2010-01-01	2031-01-01	22	Home Office
	1	14	2010-01-01	2031-01-01	23	Consumer
	1	15	2010-01-01	2031-01-01	24	Corporate
	1	16	2010-01-01	2031-01-01	25	Consumer
	1	17	2010-01-01	2031-01-01	27	Home Office

Slika 23: Dimenzijski prikaz podataka tablice *segment*

## Kreiranje dimenzijkog modela *customer*

Učitavajući bazu i skup podataka, na isti način koristimo izvore podataka. Kada uspostavimo konekciju sa bazom, putem *MySQL* upita dohvaćamo sve attribute vezane za tablicu *customer*, koja opisuje pod koje namjene će se proizvodi koristiti. Postepeno se prolazi kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti, te odabir sekvence. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*. Detaljniji postupak prikazan je na poglavlju yyyy.



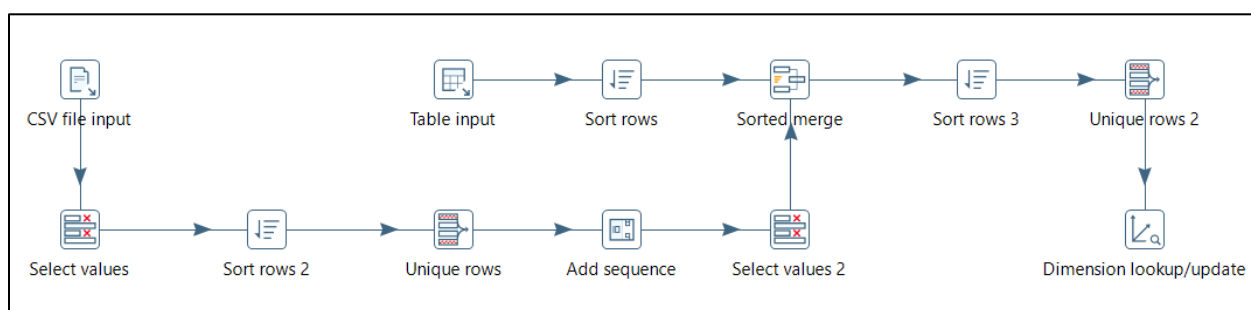
Slika 24: ETL proces dimenzije *customer*

Result Grid						
Filter Rows:						
	version	customer_tk	date_from	date_to	id	customer_name
▶	1	0	NULL	NULL	NULL	NULL
	1	1	2010-01-01	2031-01-01	1	Lena Radford
	1	2	2010-01-01	2031-01-01	2	Gene Hale
	1	3	2010-01-01	2031-01-01	3	Brendan Murry
	1	4	2010-01-01	2031-01-01	4	Mick Brown
	1	5	2010-01-01	2031-01-01	5	Clay Ludtke
	1	6	2010-01-01	2031-01-01	6	Corey Roper
	1	7	2010-01-01	2031-01-01	7	Patrick Gardner
	1	8	2010-01-01	2031-01-01	8	Mike Vittorini
	1	9	2010-01-01	2031-01-01	9	Helen Abelman
	1	10	2010-01-01	2031-01-01	10	David Wiener
	1	11	2010-01-01	2031-01-01	11	Rob Williams
	1	12	2010-01-01	2031-01-01	12	Mike Pelletier
	1	13	2010-01-01	2031-01-01	13	Carol Triggs
	1	14	2010-01-01	2031-01-01	14	Angele Hood
	1	15	2010-01-01	2031-01-01	15	Tom Stivers
	1	16	2010-01-01	2031-01-01	16	Anne McFarland
	1	17	2010-01-01	2031-01-01	17	Erica Smith

Slika 25: Dimenzijski prikaz podataka tablice *customer*

## Kreiranje dimenzijskog modela *product*

Učitavajući bazu i skup podataka, na isti način koristimo izvore podataka. Kada uspostavimo konekciju sa bazom, putem *MySQL* upita dohvaćamo sve atribute vezane za tablicu *product*, koja opisuje nazive proizvoda. Postepeno se prolazi kroz selekciju, sortiranje, odabir jedinstvenih vrijednosti, te odabir sekvence. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*. Detaljniji postupak prikazan je na poglavlju yyyy.



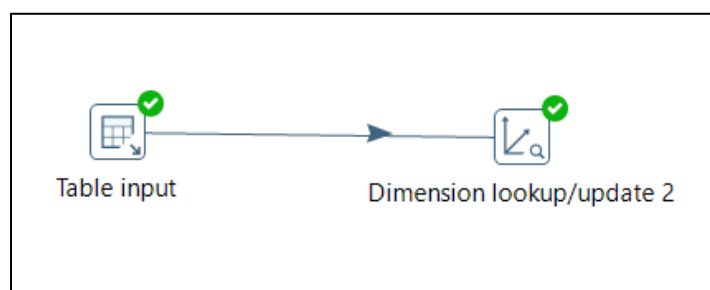
Slika 26: ETL proces dimenzije *product*

	version	product_tk	date_from	date_to	id	product_name
▶	1	0	NULL	NULL	NULL	NULL
	1	1	2010-01-01	2031-01-01	1	Kleencut Shears, Serrated
	1	2	2010-01-01	2031-01-01	2	Kleencut Trimmer, Steel
	1	3	2010-01-01	2031-01-01	3	Sony Micro Vault Click 4 GB USB 2.0 Flash Drive
	1	4	2010-01-01	2031-01-01	4	Sauder Library with Doors, Mobile
	1	5	2010-01-01	2031-01-01	5	Dania Library with Doors, Traditional
	1	6	2010-01-01	2031-01-01	6	Deflect-O Door Stop, Black
	1	7	2010-01-01	2031-01-01	7	Advantus Push Pins
	1	8	2010-01-01	2031-01-01	8	Brother Fax Machine, High-Speed
	1	9	2010-01-01	2031-01-01	9	Office Star - Ergonomic Mid Back Chair with 2-W...
	1	10	2010-01-01	2031-01-01	10	Smead Box, Blue
	1	11	2010-01-01	2031-01-01	11	GBC DocuBind TL200 Manual Binding Machine
	1	12	2010-01-01	2031-01-01	12	BIC Canvas, Fluorescent
	1	13	2010-01-01	2031-01-01	13	Smead File Cart, Single Width
	1	14	2010-01-01	2031-01-01	14	Avery Hole Reinforcements, Economy
	1	15	2010-01-01	2031-01-01	15	SanDisk Cards & Envelopes, Recycled
	1	16	2010-01-01	2031-01-01	16	Smead Box, Blue
	1	17	2010-01-01	2031-01-01	17	Global Deluxe Steno Chair

Slika 27: Dimenzijski prikaz podataka tablice *product*

## Kreiranje dimenzijkog modela *sold\_goods*

Po kreiranoj, dimenzijskoj *python* skripti, tablica *sold\_goods* predstavlja tablicu činjenica. U ETL procesu zanemareni su strani ključevi iz razloga što će se tablica činjenica gledati kao jedinstvena tablica i tek kasnije povezati sa ostalim dimenzijama, Učitavajući bazu podataka, putem *MySQL* upita dohvaćamo sve atribute vezane za tablicu *sold\_goods*, koja sadrži kvantitativne tipove podataka. Završnom fazom definira se dimenzijska tablica, spremajući model u bazu podataka *superstore\_dim*.



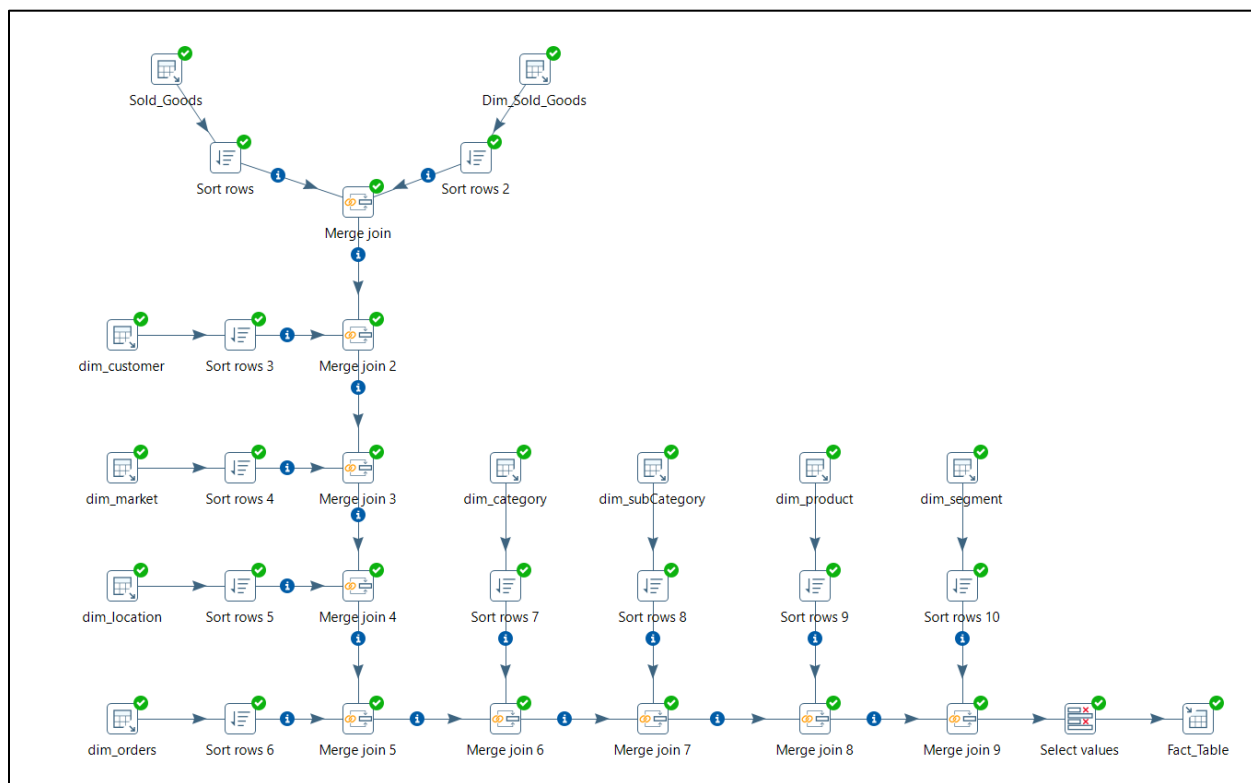
Slika 28: ETL proces dimenzije *sold\_goods*

Result Grid												
Filter Rows:												
Export: Wrap Cell Content: Fetch rows:												
	version	sold_goods_tk	date_from	date_to	id	order_date	ship_date	profit	quantity	sales	discount	shipping_cost
▶	1	0	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
	1	1	2010-01-01	2031-01-01	1	2021-06-23	2021-06-27	-2.044	7	121.716	0.4	7.756
	1	2	2010-01-01	2031-01-01	2	2020-11-10	2020-11-16	13.02	2	81.66	0	6.78
	1	3	2010-01-01	2031-01-01	3	2020-11-19	2020-11-23	3.9032	2	27.88	0	5.48
	1	4	2010-01-01	2031-01-01	4	2018-12-26	2018-12-28	-280.92	3	312.12	0.6	31.529
	1	5	2010-01-01	2031-01-01	5	2019-10-16	2019-10-21	-202.98	3	434.88	0.4	31.63
	1	6	2010-01-01	2031-01-01	6	2020-10-14	2020-10-16	-39.6	6	158.22	0.4	28.92
	1	7	2010-01-01	2031-01-01	7	2020-02-14	2020-02-21	3.5752	4	8.72	0	1.03
	1	8	2010-01-01	2031-01-01	8	2020-02-06	2020-02-10	208.98	3	950.22	0	131.43
	1	9	2010-01-01	2031-01-01	9	2021-10-21	2021-10-23	-28.9568	4	579.136	0.2	103.4
	1	10	2010-01-01	2031-01-01	10	2020-08-04	2020-08-11	9	3	32.31	0	0.76
	1	11	2010-01-01	2031-01-01	11	2021-04-12	2021-04-14	302.373	5	895.92	0.2	88.77
	1	12	2010-01-01	2031-01-01	12	2021-04-26	2021-05-01	38.46	2	109.92	0	17.03
	1	13	2010-01-01	2031-01-01	13	2020-06-18	2020-06-24	-149.52	4	257.76	0.5	38.13
	1	14	2010-01-01	2031-01-01	14	2021-01-21	2021-01-26	2.64	2	8.82	0	0.48
	1	15	2010-01-01	2031-01-01	15	2018-04-17	2018-04-19	9.6	5	241.65	0	19.59
	1	16	2010-01-01	2031-01-01	16	2021-10-18	2021-10-25	-21.54	5	32.31	0.4	2.13
	1	17	2010-01-01	2031-01-01	17	2021-09-14	2021-09-20	-20.7846	3	184.752	0.2	10.89

Slika 29: Dimenzijski prikaz podataka tablice *sold\_goods*

## Kreiranje tablice činjenica

Jednom kada se definiraju sve dimenzije, može se kreirati tablica činjenica. Definira se baza podataka, te povezuju dva tipa *table input-a* koja čine tablicu činjenica. Oba dva tipa sadrže kvantitativne podatke, dok jedna sintaksa predstavlja dimenzijske attribute. Sortiranje će se vršiti po ID-u dimenzijskih modela, a *merge* sortiranih redova pomoću *inner join* metode. Na taj način se dobiva skup dimenzija, tako da je dodatno potrebno selektirati one vrijednosti koje karakteriziraju tablicu činjenica, što obuhvaća strane surogate ključeve drugih dimenzija, te brojčane vrijednosti samih činjenica. Zadnja faza prikazuje *table output* selektiranih vrijednosti, te se može pohraniti u definiranu bazu.



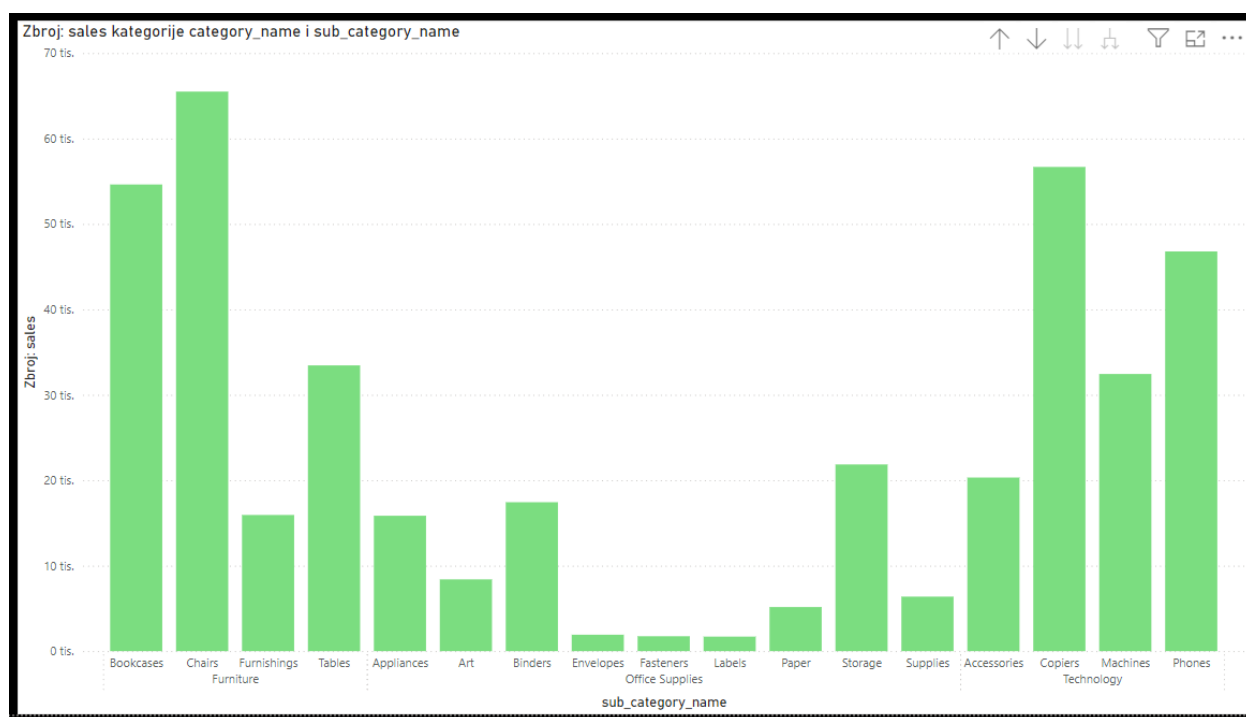
Slika 30: Tablica činjenica

Result Grid											
Filter Rows:											
Export:											
Wrap Cell Content:											
	segment_tk	order_date	ship_date	profit	quantity	sales	discount	shipping_cost	sold_goods_tk	customer_tk	market_tk
1		2021-06-23 00:00:00	2021-06-27 00:00:00	-2.044	7	121.716	0.4	7.756	1	1	1
5		2020-02-14 00:00:00	2020-02-21 00:00:00	3.5752	4	8.72	0	1.03	7	7	6
6		2020-08-04 00:00:00	2020-08-11 00:00:00	9	3	32.31	0	0.76	10	10	9
12		2021-08-28 00:00:00	2021-08-30 00:00:00	320.94	6	729.54	0	135.88	19	19	15
18		2018-09-25 00:00:00	2018-09-29 00:00:00	-3.42	2	12.78	0.4	1.03	28	28	20
24		2021-07-04 00:00:00	2021-07-07 00:00:00	-25.902	2	27.738	0.7	6.07	39	39	27
25		2021-11-10 00:00:00	2021-11-14 00:00:00	270.12	2	628.2	0	49.92	41	41	29
26		2020-06-09 00:00:00	2020-06-16 00:00:00	0.2	2	22.76	0	0.863	42	42	30
27		2021-12-02 00:00:00	2021-12-05 00:00:00	4.6552	1	16.192	0.2	3.39	43	43	31
28		2018-12-07 00:00:00	2018-12-11 00:00:00	13.74	2	275.88	0	24.91	48	48	35
36		2020-02-14 00:00:00	2020-02-18 00:00:00	0	5	251.2	0.2	52.047	60	60	45
37		2019-06-12 00:00:00	2019-06-12 00:00:00	1124.9	9	3068.36	0.1	555.77	63	63	47
39		2019-08-19 00:00:00	2019-08-23 00:00:00	86.25	1	172.53	0	16.25	67	67	50
46		2020-09-01 00:00:00	2020-09-03 00:00:00	59.7	3	145.68	0	51.091	78	78	57
47		2019-06-06 00:00:00	2019-06-09 00:00:00	12.12	4	67.44	0	6.66	79	79	58
50		2019-08-19 00:00:00	2019-08-21 00:00:00	6.24	2	39.36	0	13.42	84	84	62
55		2019-09-02 00:00:00	2019-09-04 00:00:00	25.1225	1	100.49	0	16.83	89	89	65
68		2020-02-10 00:00:00	2020-02-14 00:00:00	34.0300	2	22.5301	0.47	2.66	02	02	67

Slika 31: Prikaz tablice činjenica

## Grafički prikaz podataka

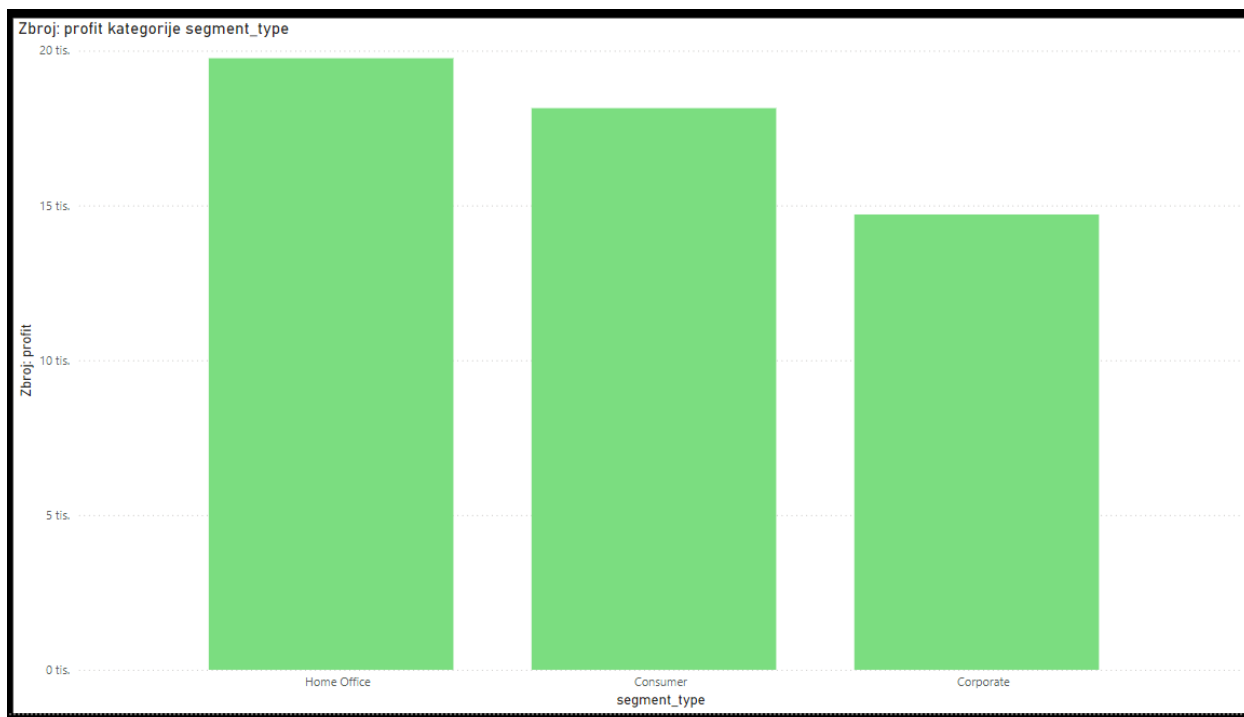
Nakon definiranja tablice činjenica, te dimenzijskih tablica, sljedeći korak jest kreirati grafičko sučelje za prikaz podataka. Postoje brojni alati koji su namijenjeni za ovakav tip zadatka. U svrhu ovog projekta, korišten je *power BI* softver, razvijen od strane *Microsoft-a*, preko kojeg će se prikazati *dashboard-ovi* kvalitativnih i kvantitativnih podataka.



Slika 32: Dashboard prodaje po kategorijama i podkategorijama

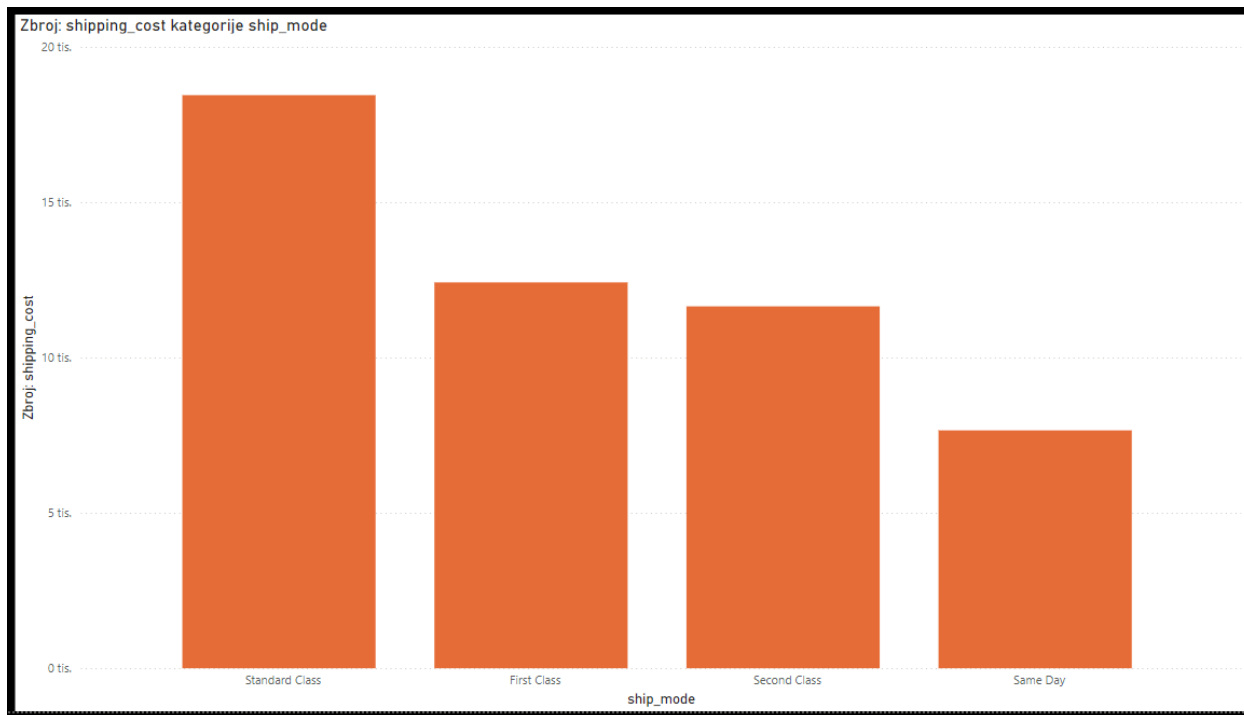
Za kreiranje grafa koristili su se podaci iz tablice činjenica, točnije podaci vezani za prodaju proizvoda. Uz činjenice, pridruženi su nazivi kategorija i pod kategorija. Prodaja je grafički prikazana po y-osi, dok su ostale sekcije proizvoda na x-osi. Korišten je stupčasti grafikon jer jasno prikazuje usporedbu prodaje između različitih kategorija i pod kategorija. Dobiveni rezultat prikazuje koja sekcija je najviše prodala proizvoda.





Slika 33: Dashboard prodaje po profitu i segmentu

Kombinacijom kvantitativnih i kvalitativnih podataka, graf se sastoji od profita i segmenta. Prati se profit dobiti prodaje proizvoda, te se uspoređuju namjene kod kojih je dobit najveći. Korišten je grupirano stupčasti graf, jer se može uočiti koji segment kupaca donosi najviše profita. Profit je postavljen po y-osi, a opće namjene na x-osi, što predstavlja ured, potrošača i korporaciju. Iz navedenog, zaključno je da prodaja proizvoda najviše prisutna kod uredskog sektora.



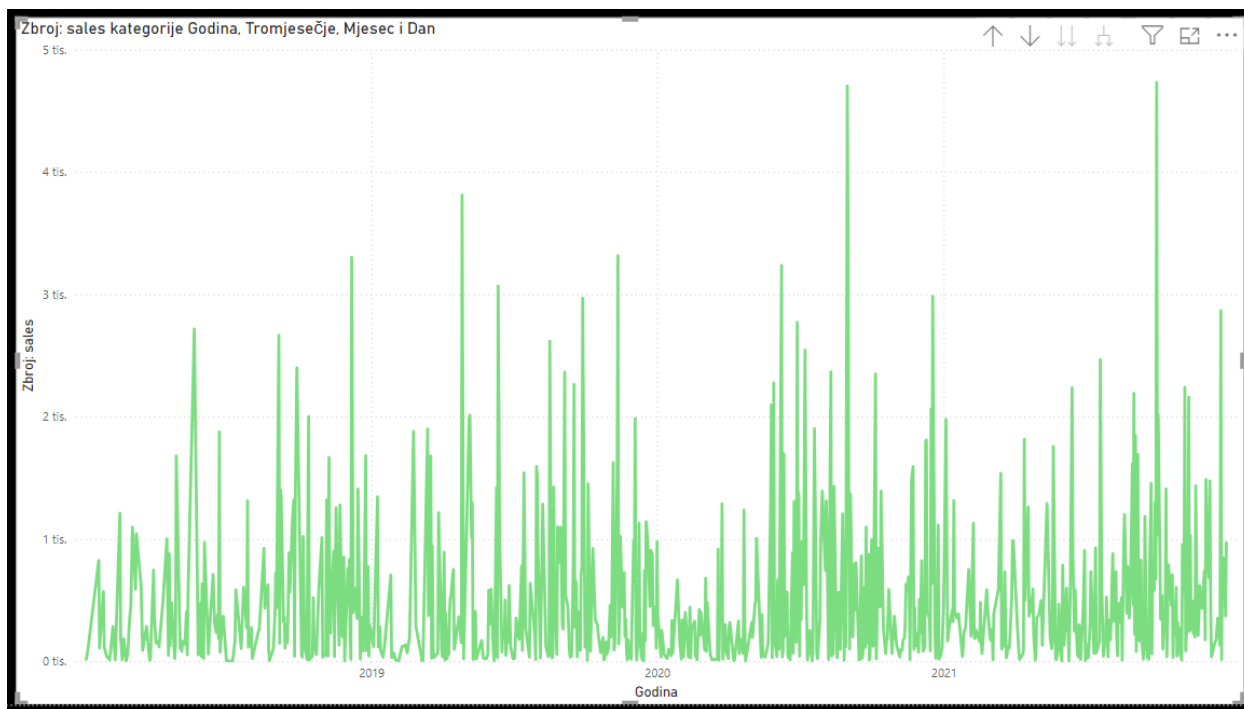
Slika 34: Dashboard prodaje po trošku i vrsti dostave

Graf prati koliki je trošak dostave u odnosu na vrstu dostave. Kao kvalitativni podaci, prisutne su klase pružanja usluga. Na x-osi nalaze se opcije dostavljanja, a y-os prikazuje ukupne troškove dostave. Najveći gubitak prihoda se pojavljuje u standardnoj klasi dostavljачkih usluga.



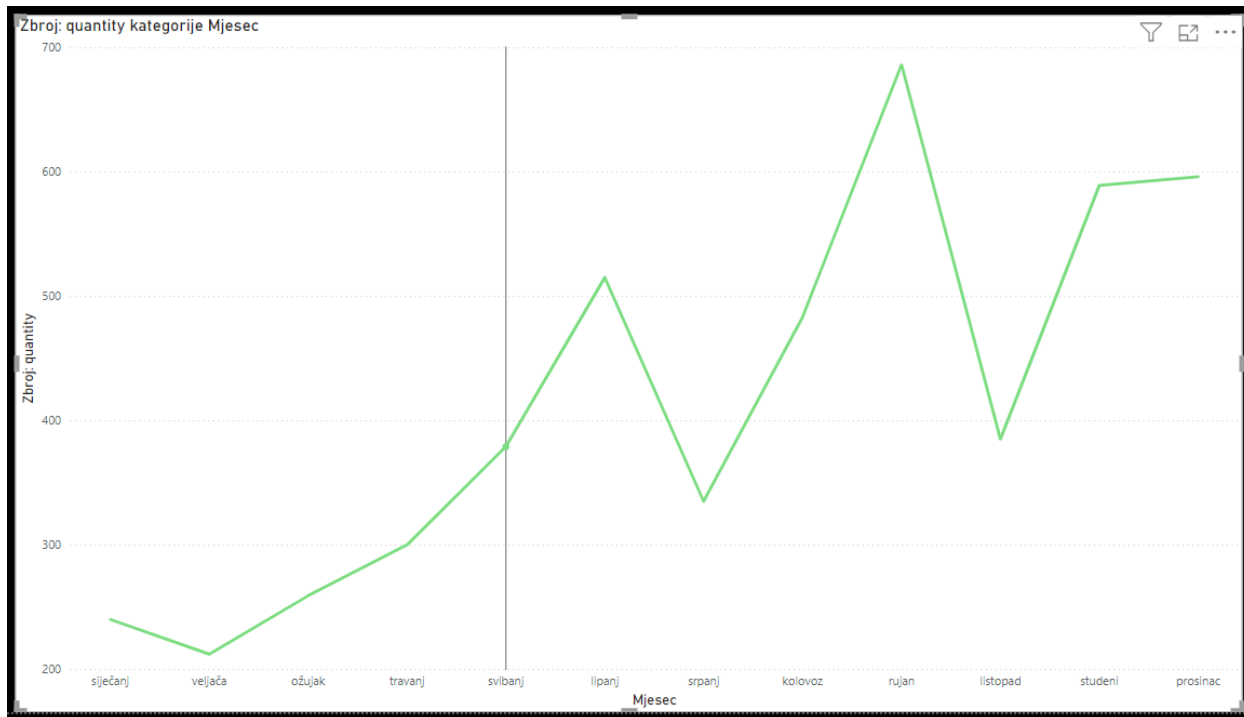
Slika 35: Dashboard prodaje po profitu i lokaciji

Korišteni su podaci profita iz tablice činjenica, u kombinaciji s državom i zemljom iz *Dim\_Location* tablice. Dobiveni profit nam je najznačajniji u Afričkim i Europskim zemljama. Pošto se radi o lokacijama, za kreiranja grafa korišten je oblik karte, kako bi se geografski dao osvrt na globalno stanje ostvarenog profita, te je također korisno za analizu profitabilnosti po različitim regijama svijeta.



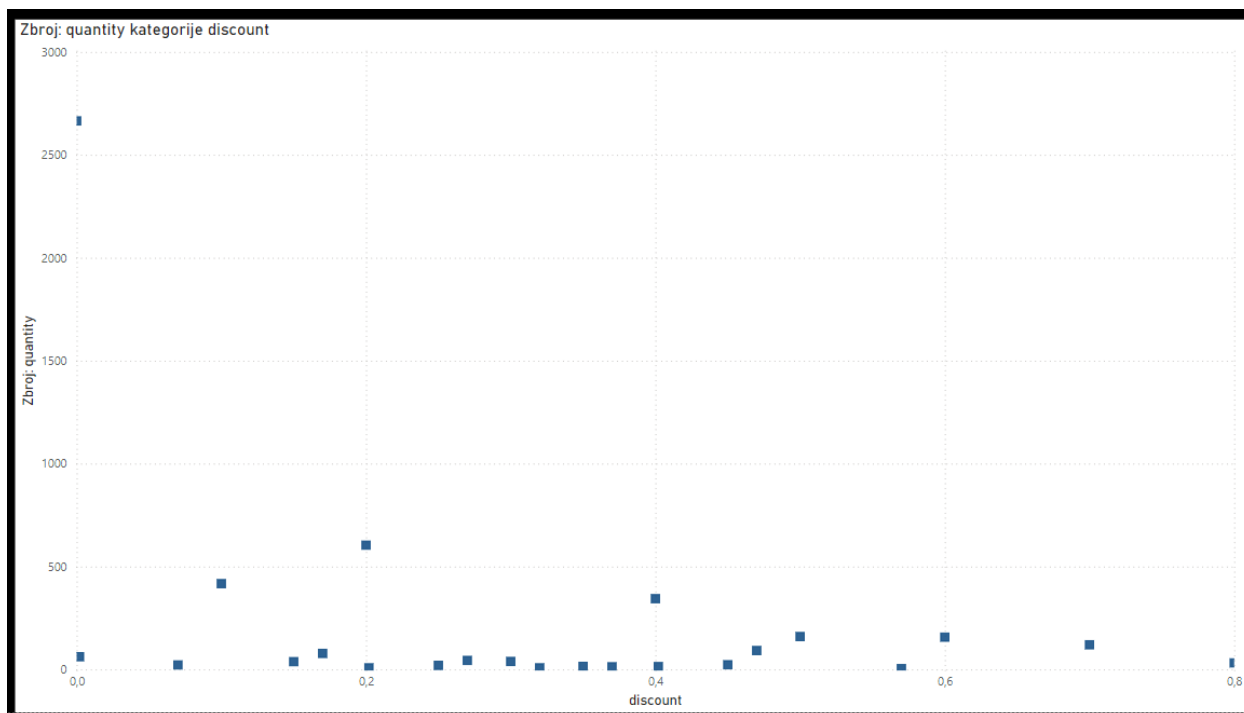
Slika 36: Dashboard prodaje po datumu narudžbe

Linijski graf prikazuje prodaju i datum naručivanja proizvoda, te je idealan za prikaz vremenskih promjena općenito. Na x-osi postavljeni su datumi, koji sadrže već odvijene događaje, a y-os prati prodaju u tom rasponu. Podaci su skladišteni najranije do 2019.-te godine, sve do 2021. U tom rasponu, prodaja naših proizvoda dostiže svoj vrhunac u kasnijem dijelu godina.



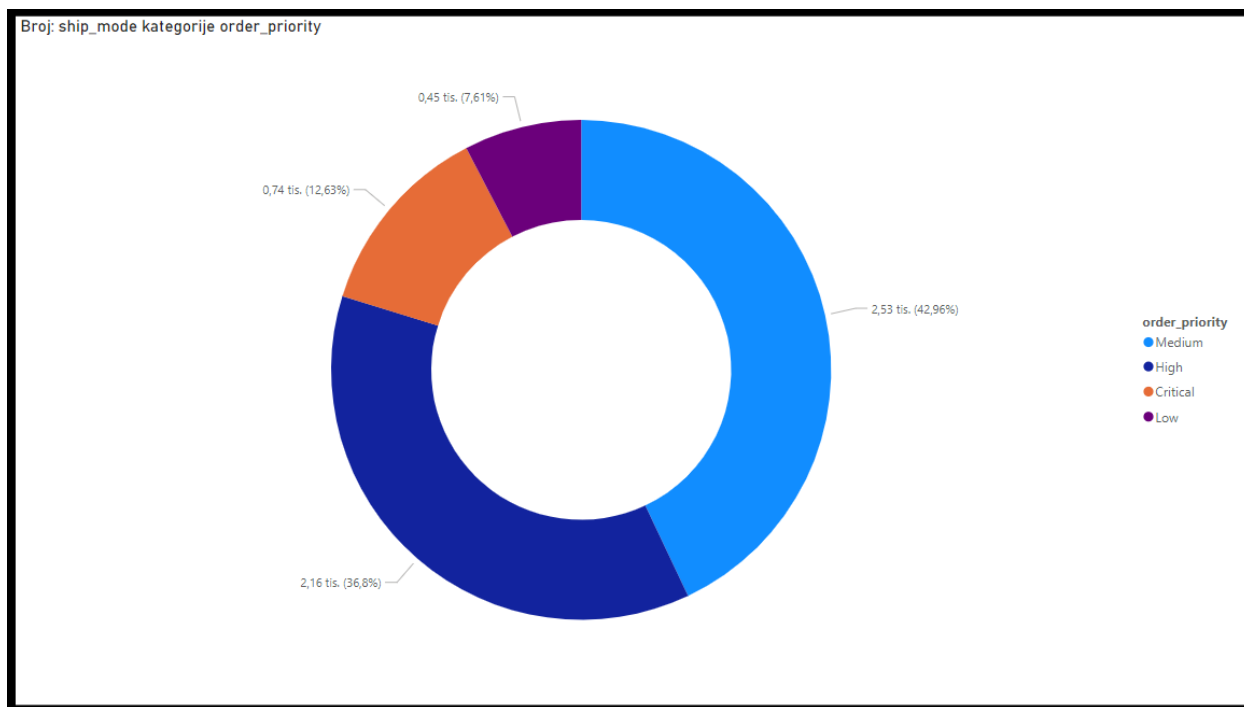
Slika 37: Dashboard prodaje po količini u vremenskom intervalu

Podaci količine i datuma naručivanja, uzeti su iz tablice činjenica. Vizualno je prikazan linijski grafikon, koji po x-osi prikazuje interval mjeseci jedne godine, dok y-os prati količine samih proizvoda. Svrha u ovome jest pratiti sezonske varijacije u količini prodaje tijekom godine, omogućujući uvid u obrasce i trendove prodaje na mjesečnoj razini, te promatrati sezonske vrhunce i padove prodaja, što može biti korisno za planiranje zaliha i marketinške strategije.



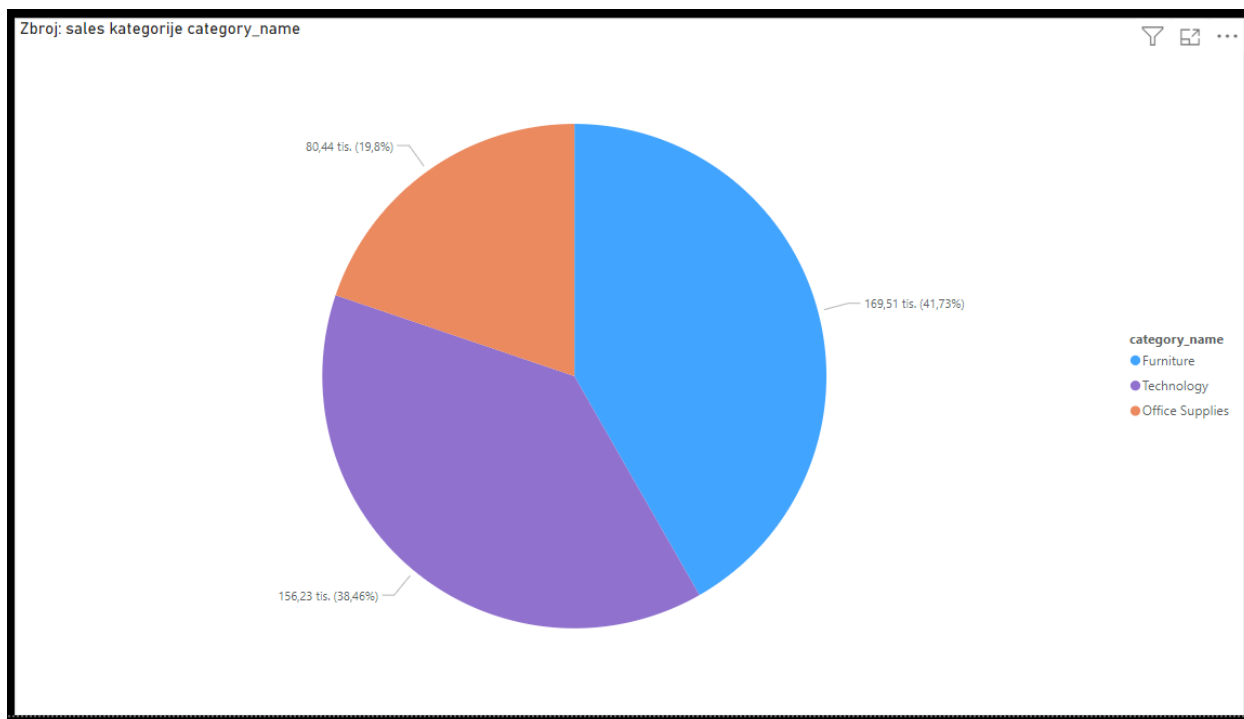
Slika 38: Slika 37: Dashboard prodaje po popustu i količini

Graf se sastoji od popusta i količine proizvoda. Korišten je raspršeni tip grafa, jer prikazuje kako varijacije u popustima utječu na količinu prodaje, što može pomoći u razumijevanju učinkovitosti popusnih strategija. Rezultat govori o tome da 2665 proizvoda nije sadržavalo popuste kupljenih proizvoda.



Slika 39: Dashboard prodaje po vrsti dostave i tipu prioriteta

Prstenasti graf prikazuje prioritete dodijeljene svakoj narudžbi i načine na koje će narudžba biti dostavljena. Koristi se za prikaz udjela različitih kategorija u odnosu na cijelu skupinu. Prioritet naručivanja nalazi se na srednjoj razini



Slika 40: Dashboard prodaje po nazivu proizvoda

Za kreiranje grafa koristili su se podaci za prodaju kao i podatci kategorije. Rezultat je da se nad klasom namještaja najviše ostvarila prodaja proizvoda.



## Zaključak

Izrada skladišta podataka je složen proces koji zahtijeva znanje i stručnost sa odgovarajućim alatima i programskim jezicima za stvaranje sustava za odlučivanje. Cijeli proces obrade podataka svodio se na kreiranje korisničkog sučelja, za donošenje poslovnih odluka, pri čemu sučelje iz tog razloga mora biti jednostavno za upotrebu, čak i za korisnike bez tehničkog predznanja. Da bi skladište podataka funkcioniralo kako je i zamišljeno, programer opsežno mora voditi analizu skupa podataka, pravilno osmisliti dimenzijske tablice i uspješno provesti ETL procese za punjenje dimenzijskih skladišta podataka. U ovom seminarskom radu prikazan je primjer takvog procesa. Kreiran je sustav koji grafičkim sučeljem krajnji korisnik može bolje razumjeti prodaju proizvoda na globalnoj razini, u nadi donošenja boljih i informativnih poslovnih odluka.

Slika 1: ER model prodaje proizvoda .....	5
Slika 2: Uspješno kreiran DB superstore.....	7
Slika 3: EER model .....	7
Slika 4: Skica dimenzijskih modela .....	9
Slika 5: Prikaz dimenzijskog modeliranja .....	10
Slika 6: Dimenzijsko modeliranje u MySQL programu .....	11
Slika 7: Uspješno kreirana baza podataka pentaho, za ETL proces.....	11
Slika 8: Dimenzijska baza podataka superstore_dim.....	12
Slika 9: Kreirana baza i skup podataka .....	13
Slika 10: MySQL sintaksa i označena transformacija podataka .....	14
Slika 11: Dimenzijski proces kreiranja tablice.....	15
Slika 12: ETL proces dimenzije location.....	16
Slika 13: Dimenzijski prikaz podataka tablice location .....	16
Slika 14: ETL proces dimenzije market.....	17
Slika 15: Dimenzijski prikaz podataka tablice market .....	17
Slika 16: ETL proces dimenzije category .....	18
Slika 17: Dimenzijski prikaz podataka tablice category.....	18
Slika 18: ETL proces dimenzije sub-category.....	19
Slika 19: Dimenzijski prikaz podataka tablice sub-category .....	19
Slika 20: ETL proces dimenzije orders.....	20
Slika 21: Dimenzijski prikaz podataka tablice orders.....	20
Slika 22: ETL proces dimenzije segment .....	21
Slika 23: Dimenzijski prikaz podataka tablice segment.....	21
Slika 24: ETL proces dimenzije customer .....	22
Slika 25: Dimenzijski prikaz podataka tablice customer.....	22
Slika 26: ETL proces dimenzije product.....	23
Slika 27: Dimenzijski prikaz podataka tablice product.....	23
Slika 28: ETL proces dimenzije sold_goods.....	24
Slika 29: Dimenzijski prikaz podataka tablice sold_goods .....	24
Slika 30: Tablica činjenica .....	25
Slika 31: Prikaz tablice činjenica.....	26
Slika 32: Dashboard prodaje po kategorijama i podkategorijama.....	27
Slika 33: Dashboard prodaje po profitu i segmentu .....	28
Slika 34: Dashboard prodaje po trošku i vrsti dostave .....	29
Slika 35: Dashboard prodaje po profitu i lokaciji .....	30
Slika 36: Dashboard prodaje po datumu narudžbe.....	31
Slika 37: Dashboard prodaje po količini u vremenskom intervalu.....	32
Slika 38: Slika 37: Dashboard prodaje po popustu i količini .....	33
Slika 39: Dashboard prodaje po vrsti dostave i tipu prioriteta .....	34
Slika 40: Dashboard prodaje po nazivu proizvoda .....	35

## Literatura

- Kimball, R., Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling. John Wiley and Sons.
- Oreški, G., (2024). Predavanja: Skladišta i rudarenje podataka.