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

Valentina Požar

## ANALIZA PROMETNIH PREKRŠAJA U MARYLAND COUNTYU

PROJEKT IZ KOLEGIJA SKLADIŠTA PODATAKA I POSLOVNA INTELIGENCIJA

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

Valentina Požar	
Studij: Baze podataka i baze znanja	
ANALIZA PROMETNIH PREKR	ŠAJA U MARYLAND COUNTYU
PROJEKT IZ KOLEGIJA SKLADIŠTA PO	DDATAKA I POSLOVNA INTELIGENCIJA
	Mentor/Mentorica:
	Prof. dr. sc. Kornelije Rabuzin
	Martina Šestak, mag. inf.

### Sadržaj

Sadržaj	i
1. Uvod	1
2. Skladište podataka	2
3. Izrada skladišta podataka	2
3.1. Pronalazak odgovarajućih podataka	2
3.2. Modeliranje skladišta podataka	4
3.2.1. Dimenzijska tablica "agencija"	7
3.2.2. Dimenzijska tablica "datum"	8
3.2.3. Dimenzijska tablica "karakteristike"	9
3.2.4. Dimenzijska tablica "mjesto"	10
3.2.5. Dimenzijska tablica "vozac"	11
3.2.6. Dimenzijska tablica "vozilo"	12
3.2.7. Dimenzijska tablica "vrijeme"	13
3.2.8. Činjenična tablica "evidencija_prekrsaja"	14
3.3. Model zvijezde	16
4. Qlik Sense	17
4.1. Izrada izvještaja	18
4.1.1. Prvi izvještaj	21
4.1.2. Drugi izvještaj	22
4.1.3. Treći izvještaj	22
4.1.4. Četvrti izvještaj	23
4.1.5. Peti izvještaj	23
5. Zaključak	24
Popis literature	25
Popis slika	26

#### 1. Uvod

Cilj ovog projektnog zadatka je izrada skladišta podataka za odabranu domenu podataka te vizualizacija istih pomoću odabranog alata. U nastavku će biti prikazano skladište podataka izrađeno u Microsoft SQL Server Management Studiu te određeni izvještaji kreirani u alatu Qlik Sense.

U današnje vrijeme je jako bitno da svi menadžeri imaju na raspolaganju lako dostupne i konzistentne podatke. Na taj način bi uvijek istovremeno, sažeto i precizno mogli dobiti prikaz organizacije i njenog okruženja u cjelini. Podaci i informacije omogućuju poduzećima da ostvare veći profit tako što će ostvarivati konkurentsku prednost na tržištu. Upravo zbog toga je bitno da menadžeri dobivaju informacije što prije i u ispravnom obliku. Podaci koje poduzeća dobivaju su najčešće pohranjeni u relacijskim bazama podataka. Problem je što zbog veličine tih baza često nije moguće pretraživati u realnom vremenu. Kako bi se mogle donositi pravilne poslovne odluke potrebno je imati uvid i u vremenski tijek donošenja raznih poslovnih odluka, a baze podataka u tom slučaju ne predstavljaju zadovoljavajuće rješenje.

Skladišta podataka omogućuju da se velike količine različitih podataka iz različitih izvora spremaju na jedno mjesto. Na taj način menadžeri mogu dobivati konkretne podatke i izvještaje za cijelo poduzeće neovisno o različitim izvorima podataka i o dislociranosti. Također, dobro dizajnirana skladišta podataka omogućuju brzo izvršavanje upita nad velikom količinom podataka, dok velik broj alata za vizualizaciju podataka omogućuje kreiranje različitih izvještaja, grafova i ostalih tipova vizualizacije podataka.

#### 2. Skladište podataka

Postoji veliki broj različitih definicija za skladište podataka koje dolaze od različitih autora. Međutim, najpopularnija definicija dolazi od Billa Inmona, oca skladišta podataka, koji je rekao sljedeće: "Skladište podataka je predmetno orijentiran, integriran, postojan i vremenski različit skup podataka koji služi kao potpora procesu donošenja odluka od strane menadžmenta."

Predmetno orijentiran – skladište podataka se može koristiti za analizu određenog područja, na primjer "prodaja" može biti određeni predmet.

Integriran – skladište podataka integrira podatke iz različitih izvora podataka, na primjer mogu postojati dva izvora podataka koji imaju različite načine identificiranja proizvoda, ali u skladištu podataka postoji samo jedan način identificiranja proizvoda.

Postojan – kada se podaci pohrane u podatkovnom skladištu, oni se neće promijeniti što znači da se povijesni podaci u skladištu podataka nikada ne smiju mijenjati.

Vremenski različit – povijesni podaci se čuvaju u skladištu podataka što znači da se uvijek mogu dohvatiti podaci stari 3 mjeseca, 6 mjeseci, 12 mjeseci ili čak stariji podaci. To je suprotno od transakcijskog sustava, gdje se često čuvaju samo najnoviji podaci. Na primjer, sustav transakcija može sadržavati najnoviju adresu klijenta, a skladište podataka može sadržavati sve adrese povezane s klijentom.

S druge strane, Ralph Kimball je dao sažetu definiciju skladišta podataka koja govori da je skladište podataka kopija transakcijskih podataka posebno strukturiranih za upite i analizu. Tom definicijom je prikazana funkcija skladišta podataka, za razliku od Inmonove definicije koja više opisuje kako je skladište podataka izgrađeno (1Keydata, 2019).

#### 3. Izrada skladišta podataka

#### 3.1. Pronalazak odgovarajućih podataka

Za izradu skladišta podataka odabrana je tema analize prometnih prekršaja u Maryland Countyu u razdoblju od 2012. do 2018. godine. Kao izvor podataka preuzet je skup podataka s web stranice <a href="www.kaggle.com">www.kaggle.com</a> gdje se mogu besplatno preuzeti i koristiti podaci u akademske svrhe. Odabrani skup podataka sadrži sve događaje vezane za

prometne prekršaje u razdoblju od 2012. do 2018. godine te ima oko 1,04 milijuna zapisa. Podaci uključuju sljedeće stavke:

- nesreća je li prometni prekršaj uključuje nesreću
- agencija naziv agencije koja izdaje prometni prekršaj
- alkohol je li vozač tijekom prometnog prekršaja bio pod utjecajem alkohola
- vrsta uhićenja
- članak članak državnog zakona
- pojasevi je li vozač tijekom prometnog prekršaja koristio pojas
- naplata numerički kod za određenu naknadu
- boja boja vozila
- komercijalna licenca je li vozač koji je skrivio prometni prekršaj ima licencu za komercijalne upravljačke programe
- komercijalno vozilo je li vozilo kojim je počinjen prometni prekršaj komercijalno vozilo
- doprinos nesreći je li prometni prekršaj doprinio nesreći
- datum zaustavljanja datum prometnog prekršaja
- opis tekstualni opis određene naknade
- DL država naziv države koja izdaje vozačku dozvolu
- grad vozača grad kojem pripada adresa vozača
- država vozača država kojoj pripada adresa vozača
- fatalno je li tijekom prometnog prekršaja došlo do smrtnog slučaja
- spol spol vozača (F = žensko, M = muško)
- geolokacija geografske koordinate
- HAZMAT je li prometni prekršaj uključuje opasne materijale
- geografska širina lokacija prometnog prekršaja
- geografska dužina lokacija prometnog prekršaja
- lokacija mjesto prometnog prekršaja, može biti adresa ili raskrižje
- proizvođač naziv proizvođača vozila kojim je skrivljen prometni prekršaj
- model model vozila kojim je skrivljen prometni prekršaj
- osobna ozljeda je li prometni prekršaj uključuje osobnu ozljedu
- oštećenje imovine je li prometni prekršaj uključuje oštećenje imovine
- rasa rasa kojoj pripada vozač
- država naziv države koja izdaje registraciju vozila
- podagencija sudski kod koji predstavlja distrikt zaduženja časnika
- vrijeme zaustavljanja vrijeme prometnog prekršaja
- vrsta vozila vrsta vozila kojim je počinjen prometni prekršaj

- vrsta prekršaja vrsta prometnog prekršaja
- radna zona je li prometni prekršaj počinjen u radnoj zoni
- godina godina proizvodnje vozila kojim je počinjen prometni prekršaj

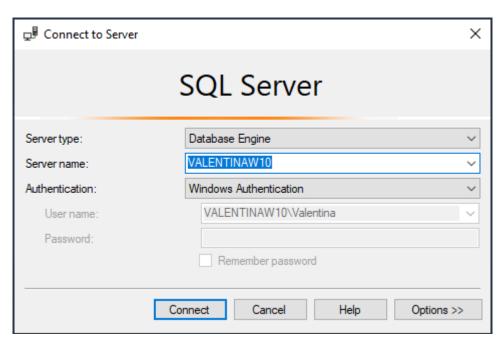
Izvorni podaci su dani u .csv (*eng. comma-separated value*) formatu. Može se vidjeti kako podaci u izvornom formatu nisu baš dobro posloženi, pa ih je zbog toga potrebno očistiti i transformirati kako bi se dobili što kvalitetniji podaci. Na slici 1 je prikazan originalni prikaz podataka i izgled u Excelu. Može se vidjeti da su neki podaci nerazumljivi te da je teško iščitati kojem stupcu pripada koji redak.

Date Of Stop	Time Of St	Agency	SubAgeno	Description	Location	Latitude	Longitude	Accident	Belts	Personal I	Property I	Fatal	Commerc	HAZMAT	Commerc	Alcohol	Work Zon	State	VehicleTy Y€
9/24/2013	17:11:00	MCP	3rd distric	DRIVING '	\ 8804 FLOV	VER AVE		No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
8/29/2017	10:19:00	MCP	2nd distric	DRIVER FA	WISCONS	38.98173	-77.0928	No	No	No	No	No	No	No	No	No	No	VA	02 - Auton
12/1/2014	12:52:00	MCP	6th distric	FAILURE S	CHRISTOP	39.16289	-77.2291	No	No	No	Yes	No	No	No	No	No	No	MD	02 - Auton
8/29/2017	9:22:00	MCP	3rd distric	FAILURE Y	CHERRY H	39.05698	-76.9546	No	No	No	Yes	No	No	No	No	No	No	MD	02 - Auton
8/28/2017	23:41:00	MCP	6th distric	FAILURE O	355 @ SOL	JTH WESTL	AND DRIVE	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
8/27/2013	0:55:00	MCP	2nd distric	NEGLIGEN	CONNECT	ICUT/CHEV	Y CHASE L	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
10/8/2013	13:23:00	MCP	4th distric	DRIVING '	GEORGIA	39.09338	-77.0796	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
4/24/2015	0:38:00	MCP	1st district	DRIVER FA	EB MONTE	ROSE PKW	//EAST JEFF	No	No	No	No	No	No	No	No	No	No	VA	02 - Auton
8/28/2017	23:41:00	MCP	6th distric	FAILURE O	355 @ SOL	JTH WESTL	AND DRIVE	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
8/28/2017	23:41:00	MCP	6th distric	DRIVING '	\ 355 @ SOU	JTH WESTL	AND DRIVE	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
2/14/2014	20:10:00	MCP	1st district	FAILURE T	GATEWAY	39.23484	-77.2815	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
3/19/2012	16:10:00	MCP	2nd distric	DRIVER U	CLARENDO	38.98273	-77.1008	No	No	No	No	No	No	No	No	No	No	VA	02 - Auton
6/9/2012	21:19:00	MCP	3rd distric	OCCUPAN	2068 HARL	39.06914	-76.9697	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
12/20/2012	0:41:00	MCP	2nd distric	DRIVING	NORFOLK	38.98358	-77.0931	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
7/20/2012	23:12:00	MCP	5th distric	FAILURE T	WISTERIA	39.16181	-77.2536	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
9/11/2012	21:47:00	MCP	4th distric	PERSON E	TWIG RD A	39.0962	-76.987	No	Yes	No	No	No	No	No	No	No	No	MD	02 - Auton
7/23/2013	23:15:00	MCP	4th distric	FAILURE T	108 / OLD	39.15167	-77.0706	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
10/28/2013	17:41:00	MCP	6th distric	DRIVER FA	EMORY GF	39.12993	-77.1671	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
8/28/2017	23:41:00	MCP	6th distric	PERSON E	355 @ SOL	JTH WESTL	AND DRIVE	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
12/17/2015	13:15:00	MCP	5th distric	FAILURE T	19916 CRY	39.22279	-77.2537	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
8/28/2017	23:41:00	MCP	6th distric	PERSON E	355 @ SOL	JTH WESTL	AND DRIVE	No	No	No	No	No	No	No	No	No	No	MD	02 - Auton
5/12/2015	8:26:00	MCP	1st distric	PERSON E	woottor	39.06754	-77.1494	No	No	No	No	No	No	No	No	No	No	VA	28 - Other

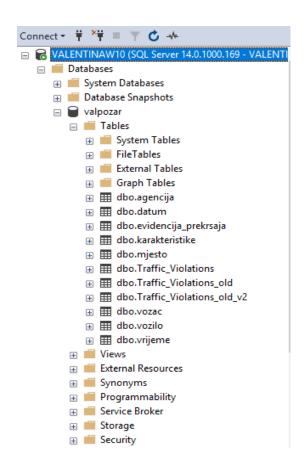
Slika 1: Izgled izvornih podataka u .csv datoteci

#### 3.2. Modeliranje skladišta podataka

Za potrebe ovog projektnog rada skladište podataka je izrađeno u Microsoft SQL Server Management Studiu na Microsoft SQL Serveru. Kako bi se počelo s radom, potrebno se povezati na server putem Windows ovjere (*eng. Windows Authentication*). Nakon toga se može izraditi nova baza podataka koja se u ovom slučaju zove *valpozar*. Unutar baze podataka izrađene su tablice koje su korištene za izradu skladišta podataka. Izrađeno je sedam dimenzijskih tablica: dbo.agencija, dbo.datum, dbo.karakteristike, dbo.mjesto, dbo.vozac, dbo.vozilo i dbo.vrijeme, te jedna činjenična tablica dbo.evidencija\_prekrsaja.

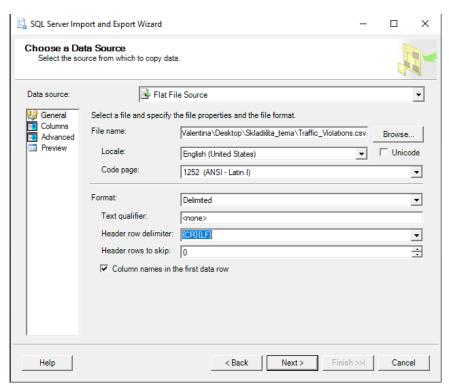


Slika 2: Povezivanje na server

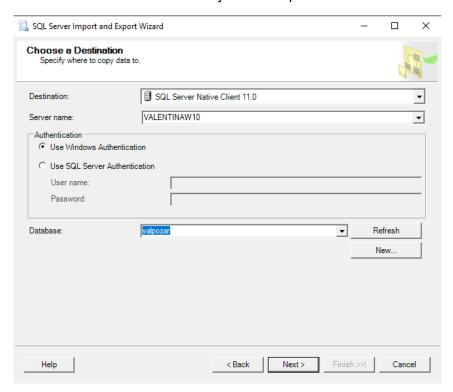


Slika 3: SQL server i skladište podataka

Kako bi se kreirane tablice mogle napuniti podacima, potrebno je učitati izvorne podatke. To se može napraviti uz pomoć SQL Server Import and Export Wizarda. Prvo se mora odabrati otkud će se učitati podaci, a u ovom slučaju to je .csv datoteka. Nakon toga se treba odabrati gdje će se ti podaci učitati, a to je SQL Server Native Client 11.0.



Slika 4: Učitavanje izvornih podataka



Slika 5: Odabir odredišne lokacije

	Commercial Vehicle	State	Year	Make	Model	Color	Location	
190	No	MD	2004	HONDA	CRV	SILVER	MD 28 / SHADY GROVE RD	
191	No	MD	2004	HONDA	PILOT	WHITE	15700 BLK SHADY GROVE RD.	
192	No	MD	2004	INFINITI	G35	WHITE	SELFRIDGE RD AT GRIDLEY RD	
193	No	MD	2004	INFINITY	4S	SILVER	CONNECTICUIT AVE / GEORGIA AVE	
194	No	MD	2004	MAZDA	3	BLUE	R28/FOREST AVE	
195	No	MD	2004	SUZUKI	4DR SUV	BLACK	E/B NORBECK RD. AT AVERY RD.	
196	No	MD	2004	TOYOTA	COROLLA	SILVER	RANDOLPH RD/ HUNTERS LA	
197	No	MD	2005	ACURA	MDX	SILVER	ROCKVILLE PIKE / HALPINE RD	
198	No	MD	2005	FORD	EXPLORER	RED	12100 BLOCK PARKLAWN	
199	No	MD	2005	HOND	TK	BLACK	E. GUDE DR/SOUTHLAWN LA	
200	No	MD	2005	HONDA	ACCORD	RED	WB RANDOLPH ROAD/TIVOLI LAKE BL	
201	No	MD	2005	JEEP	TK LIMITED S	SILVER	W/B E GUDE AT SOUTHLAWN LN	
202	No	MD	2005	MAZD	VN	GOLD	CHAPMAN AVE AND BOU	
< ous	No	MD	2005	MISSAM	ΔΙ ΤΙΜΔ	RITIE	VEIRS MILL RD / RORINDALE DR	>

Slika 6: Učitani izvorni podaci

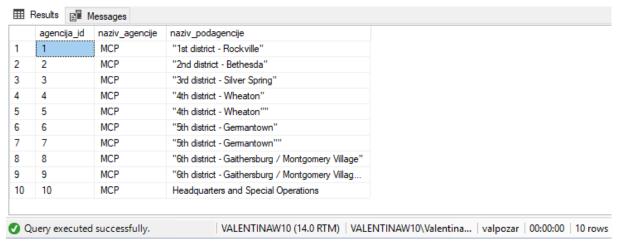
#### 3.2.1. Dimenzijska tablica "agencija"

Na slici 7 prikazana je struktura dimenzijske tablice "agencija". Može se vidjeti da je primarni ključ tablice agencija\_id. Tablica sadrži još dva atributa, a to su naziv\_agencije i naziv\_podagencije. U ovoj tablici nalaze se nazivi svih agencija i podagencija koje izdaju određeni prometni prekršaj. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.

VAI	ENTINAW10.valpozar - dbo.a	gencija → X	
	Column Name	Allow Nulls	
P	agencija_id	int	
	naziv_agencije	varchar(400)	
	naziv_podagencije	varchar(400)	

Slika 7: Struktura tablice "agencija"

Slika 8: Upit za spremanje podataka u tablicu "agencija"



Slika 9: Prikaz podataka u tablici "agencija"

#### 3.2.2. Dimenzijska tablica "datum"

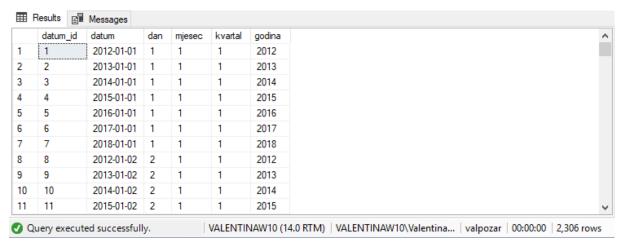
Na slici 10 prikazana je struktura dimenzijske tablice "datum". Može se vidjeti da je primarni ključ tablice datum\_id. Tablica sadrži još pet atributa, a to su datum, dan, mjesec, kvartal i godina. U ovoj tablici nalaze se svi datumi kada se određeni prometni prekršaj dogodio. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.

VAL	.ENTINAW10.valpozar - dbo.d	atum ⊅ ×	
	Column Name	Data Type	Allow Nulls
₽Ÿ	datum_id	int	
	datum	date	
	dan	int	
	mjesec	int	
	kvartal	int	
	godina	int	

Slika 10: Struktura tablice "datum"

```
INSERT INTO datum
SELECT ROW_NUMBER() OVER (ORDER BY subq.date_of_stop)
,          subq.date_of_stop datum
,          DATEPART(DAY, subq.date_of_stop)
,          DATEPART(MONTH, subq.date_of_stop)
,          DATEPART(QUARTER, subq.date_of_stop)
,          DATEPART(YEAR, subq.date_of_stop)
FROM (SELECT DISTINCT "Date Of Stop" date_of_stop
          FROM Traffic_Violations) AS subq;
```

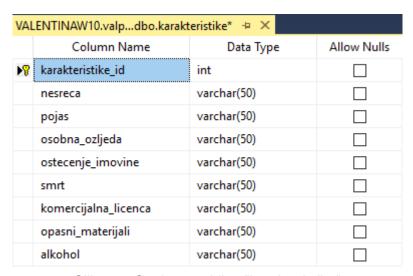
Slika 11: Upit za spremanje podataka u tablicu "datum"



Slika 12: Prikaz dijela podataka u tablici "datum"

#### 3.2.3. Dimenzijska tablica "karakteristike"

Na slici 13 prikazana je struktura dimenzijske tablice "karakteristike". Može se vidjeti da je primarni ključ tablice karakteristike\_id. Tablica sadrži još osam atributa, a to su nesreca, pojas, osobna\_ozljeda, ostecenje\_imovine, smrt, komercijalna\_licenca, opasni\_materijali, alkohol. U ovoj tablici se nalaze moguće karakteristike koje opisuju prometni prekršaj. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.



Slika 13: Struktura tablice "karakteristike"

Slika 14: Upit za spremanje podataka u tablicu "karakteristike"

	karakteristike_id	nesreca	pojas	osobna_ozljeda	ostecenje_imovine	smrt	komercijalna_licenca	opasni_materijali	alkohol	-
1	1	No	No	No	No	No	No	No	No	
2	2	No	No	No	No	No	No	No	Yes	
3	3	No	No	No	No	No	No	Yes	No	
4	4	No	No	No	No	No	Yes	No	No	
5	5	No	No	No	No	No	Yes	No	Yes	
6	6	No	No	No	No	No	Yes	Yes	No	
7	7	No	No	No	No	Yes	No	No	No	
8	8	No	No	No	No	Yes	Yes	No	No	
9	9	No	No	No	Yes	No	No	No	No	
10	10	No	No	No	Yes	No	No	No	Yes	
11	11	No	No	No	Yes	No	Yes	No	No	

Slika 15: Prikaz dijela podataka u tablici "karakteristike"

#### 3.2.4. Dimenzijska tablica "mjesto"

Na slici 16 prikazana je struktura dimenzijske tablice "mjesto". Može se vidjeti da je primarni ključ tablice mjesto\_id. Tablica sadrži još tri atributa, a to su naziv, geografska\_sirina i geografska\_duzina. U ovoj tablici se nalaze podaci o točnoj lokaciji gdje se dogodio određeni prometni prekršaj. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.

VAL	.ENTINAW10.valpozar - dbo.n	njesto ⊅ X	
	Column Name	Data Type	Allow Nulls
₽₽	mjesto_id	int	
	naziv	varchar(50)	
	geografska_sirina	decimal(18, 14)	
	geografska_duzina	decimal(18, 14)	

Slika 16: Struktura tablice "datum"

Slika 17: Upit za spremanje podataka u tablicu "mjesto"

	mjesto_id	naziv	geografska_sirina	geografska_duzina	^
1	1	"10005 FIELDS Rt - GA"	39.11665166666670	-77.20651166666670	
2	2	"10008 HEMSWELL LANt - POTOMAt - MD"	39.38690000000000	-77.41396000000000	
3	3	"10011 STEDWICK Rt - MONTGOMERY VILLAGE MD"	39.17070333333333	-77.20716333333333	
4	4	"10100 BLK CEDAR LANt - KENSINGTOt - MD"	39.02142500000000	-77.08510500000000	
5	5	"10100 BLK SEVEN LOCKS ROAt - S/B"	39.01581886666670	-77.16039921666670	
6	6	"10217 SUNDANCE CT "	39.07048431666670	-77.16321353333333	
7	7	"10313 DICKENS AVt - BETHESDt - MD 20814"	38.98061500000000	-77.08972500000000	
8	8	"10400 AMHERST AVt - WHEATOt - MD 20902"	39.02570666666670	-77.04476833333333	
9	9	"10410 HUNTLEY AVE. SILVER SPRINt - MD"	39.02758833333333	-77.03743166666670	
10	10	"10440 ROCKVILLE PIKt - BETHESDt - MD"	38.99753500000000	-77.10961500000000	
11	11	"10500 CONNECTICUT AVE. KENSINGTOt - MD"	38.99788666666670	-77.10996500000000	

Slika 18: Prikaz dijela podataka iz tablice "mjesto"

#### 3.2.5. Dimenzijska tablica "vozac"

Na slici 19 prikazana je struktura dimenzijske tablice "vozac". Može se vidjeti da je primarni ključ tablice vozac\_id. Tablica sadrži još četiri atributa, a to su spol, drzava, grad, rasa. U ovoj tablici se nalaze podaci o svim vozačima koji su skrivili neki prometni prekršaj. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.

VAL	VALENTINAW10.valpozar - dbo.vozac → ×								
	Column Name	Data Type	Allow Nulls						
₽₽	vozac_id	int							
	spol	varchar(50)							
	drzava	varchar(50)							
	grad	varchar(50)							
	rasa	varchar(50)							

Slika 19: Struktura tablice "vozac"

Slika 20: Upit za spremanje podataka u tablicu "vozac"

	vozac_id	spol	drzava	grad	rasa
	1	F	AB	ST. ALBERT	WHITE
	2	М	AB	CALGARY	BLACK
	3	M	AB	CALGARY	WHITE
4	4	F	AB	SPRUCE GROVE	WHITE
5	5	М	AB	EDMONTON	ASIAN
6	6	М	AB	EDMONTON	OTHER
7	7	М	AB	CALGARY	HISPANIC
	8	М	AB	CALGARY	OTHER
9	9	F	AK	ANCHORAGE	WHITE
10	10	F	AK	FAIRBANKS	WHITE
11	11	М	AK	ANCHORAGE	OTHER

Slika 21: Prikaz dijela podataka iz tablice "vozac"

#### 3.2.6. Dimenzijska tablica "vozilo"

Na slici 22 prikazana je struktura dimenzijske tablice "vozilo". Može se vidjeti da je primarni ključ tablice vozilo\_id. Tablica sadrži još sedam atributa, a to su tip\_vozila, komercijalno\_vozilo, drzava\_registracije, godina\_proizvodnje, marka\_vozila, model\_vozila i boja\_vozila. U ovoj tablici se nalaze podaci o svim vozilima kojima su počinjeni neki prometni prekršaji. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.

VAI	.ENTINAW10.valpozar - dbo.v	ozilo ⊅ ×	
	Column Name	Data Type	Allow Nulls
<b>▶</b> 8	vozilo_id	int	
	tip_vozila	varchar(50)	
	komercijalno_vozilo	varchar(50)	
	drzava_registracije	varchar(50)	
	godina_proizvodnje	varchar(50)	
	marka_vozila	varchar(50)	
	model_vozila	varchar(50)	
	boja_vozila	varchar(50)	

Slika 22: Struktura tablice "vozilo"

Slika 23: Upit za spremanje podataka u tablicu "vozilo"

	vozilo_id	tip_vozila	komercijalno_vozilo	drzava_registracije	godina_proizvodnje	marka_vozila	model_vozila	boja_vozila	^
112	112	02 - Automobile	No	MD	2002	ACUARA	3.2	GRAY	
113	113	03 - Station	No	MD	2015	ACUARA	4D	BLACK	
114	114	02 - Automobile	No	MD	2013	ACUEA	4S	WHITE	
115	115	02 - Automobile	No	NC	1998	ACUFRA	XL	GREEN	
116	116	02 - Automobile	No	MD	2008	ACUFRA	MDX	GRAY	
117	117	02 - Automobile	No	MD	2005	ACUIRA	TL	SILVER	
118	118	01 - Motorcycle	No	MD	2000	ACUR	4S 3.2 TL	RED	
119	119	01 - Motorcycle	No	MD	2001	ACUR	4S	GOLD	
120	120	01 - Motorcycle	No	MD	2003	ACUR	4S	GOLD	
121	121	01 - Motorcycle	No	MD	2004	ACUR	4S	BLACK	V
<								>	

Slika 24: Prikaz dijela podataka iz tablice "vozilo"

#### 3.2.7. Dimenzijska tablica "vrijeme"

Na slici 25 prikazana je struktura dimenzijske tablice "vrijeme". Može se vidjeti da je primarni ključ tablice vrijeme\_id. Tablica sadrži još četiri atributa, a to su vrijeme, sekunda, minuta i sat. U ovoj tablici se nalaze sva vremena kada su počinjeni neki prometni prekršaji. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.

VALENTINAW10.valpozar - dbo.vrijeme → ×					
	Column Name	Allow Nulls			
₽₽	vrijeme_id	int			
	vrijeme	time(7)			
	sekunda	int			
	minuta	int			
	sat	int			

Slika 25: Struktura tablice "vrijeme"

Slika 26: Upit za spremanje podataka u tablicu "vrijeme"

	vrijeme_id	vrijeme	sekunda	minuta	sat	
1	1	00:00:00.0000000	0	0	0	
2	2	00:01:00.0000000	0	1	0	
3	3	00:02:00.0000000	0	2	0	
4	4	00:03:00.0000000	0	3	0	
5	5	00:04:00.0000000	0	4	0	
6	6	00:05:00.0000000	0	5	0	
7	7	00:06:00.0000000	0	6	0	
8	8	00:07:00.0000000	0	7	0	
9	9	00:08:00.0000000	0	8	0	
10	10	00:09:00.0000000	0	9	0	
11	11	00:10:00.0000000	0	10	0	

Slika 27: Prikaz dijela podataka iz tablice "vrijeme"

#### 3.2.8. Činjenična tablica "evidencija\_prekrsaja"

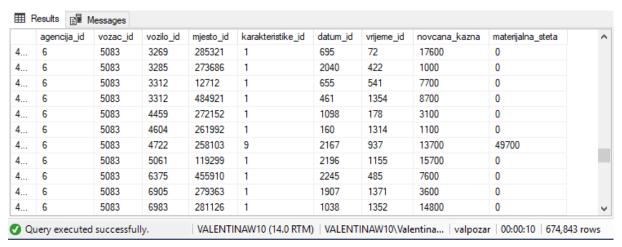
Na slici 28 prikazana je struktura činjenične tablice "evidencija\_prekrsaja". Činjenična tablica sadrži vanjske ključeve od dimenzijskih tablica s kojima je povezana: agencija\_id, vozac\_id, vozilo\_id, mjesto\_id, karakteristike\_id, datum\_id i vrijeme\_id. Osim toga, navedena tablica sadrži dva atributa koja predstavljaju činjenice, a to su novcana\_kazna i materijalna\_steta. U nastavku će biti prikazan upit kojim je navedena tablica napunjena podacima.

VALENTINAW10.valpidencija_prekrsaja → ×					
	Column Name	Data Type	Allow Nulls		
₽Ÿ	agencija_id	int			
8	vozac_id	int			
8	vozilo_id	int			
8	mjesto_id	int			
8	karakteristike_id	int			
8	datum_id	int			
8	vrijeme_id	int			
	novcana_kazna	int			
	materijalna_steta	int			

Slika 28: Struktura tablice "evidencija\_prekrsaja"

```
INSERT INTO evidencija prekrsaja
SELECT ag.agencija id
      vo.vozac id
      vz.vozilo id
      mj.mjesto id
      ka.karakteristike id
      dt.datum id
      vr.vrijeme id
      ROUND (CAST (ABS (CHECKSUM (NEWID ())) AS bigint) %20000+500, -2)
      ROUND (CAST (ABS (CHECKSUM (NEWID ())) AS bigint) %80000+3000, -2)
FROM agencija ag
      vozac vo
      vozilo vz
      mjesto mj
      karakteristike ka
      datum dt
       vrijeme vr
      Traffic Violations tv
WHERE tv. "Agency"
                               = ag.naziv agencije
AND tv. "SubAgency"
                               = ag.naziv podagencije
      tv. "Gender"
                               = vo.spol
AND
      tv."Driver City"
tv."Driver State"
AND
                               = vo.grad
AND
                                = vo.drzava
AND
      tv."Race"
                                = vo.rasa
      tv."VehicleType"
AND
                                = vz.tip_vozila
      tv.[Commercial Vehicle] = vz.komercijalno_vozilo
AND
AND
      tv."State"
                               = vz.drzava registracije
      tv. "Year"
AND
                               = vz.godina proizvodnje
      tv."Make"
AND
                               = vz.marka vozila
      tv."Model"
                               = vz.model_vozila
AND
                               = vz.boja_vozila
AND
      tv."Color"
AND
      tv."Location"
                               = mj.naziv
AND
      tv."Latitude"
                               = mj.geografska_sirina
AND
      tv."Longitude"
                              = mj.geografska_duzina
AND
      tv. "Accident"
                               = ka.nesreca
AND
      tv."Belts"
                               = ka.pojas
      AND
AND
      tv."Fatal"
AND
                                = ka.smrt
      tv. "Commercial License" = ka.komercijalna licenca
AND
      tv."HAZMAT"
AND
                               = ka.opasni materijali
AND
      tv. "Alcohol"
                               = ka.alkohol
      tv."Date Of Stop"
tv."Time Of Stop"
AND
                               = dt.datum
                               = vr.vrijeme
AND
AND
       tv.[Property Damage]
                                = 'Yes'; -- 674 843
```

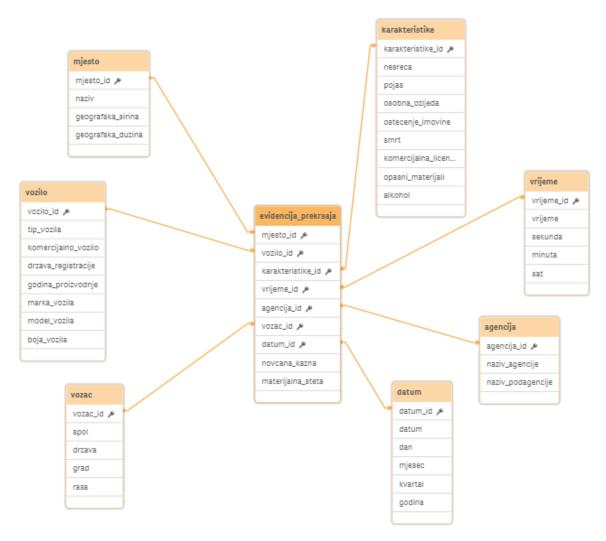
Slika 29: Upit za spremanje podataka u tablicu "evidencija\_prekrsaja"



Slika 30: Prikaz dijela podataka iz tablice "evidencija\_prekrsaja"

#### 3.3. Model zvijezde

Skladište podataka dizajnirano je prema modelu zvijezde. Model zvijezde sastoji se od jedne ili više činjeničnih tablica. Na činjenične tablice se veže više dimenzijskih tablica. Činjenične tablice nam govore što mjerimo, dok nam dimenzijske govore prema čemu mjerimo odnosno daju kontekst brojevima i mjerama iz činjenične tablice (Rabuzin, 2018). Nakon analiziranja podataka koji se nalaze u tablici "Traffic\_Violations" definirano je sedam dimenzijskih tablica te jedna činjenična tablica. U nastavku će biti prikazan model zvijezde koji se sastoji od ranije definiranih i objašnjenih dimenzijskih tablica te činjenične tablice.



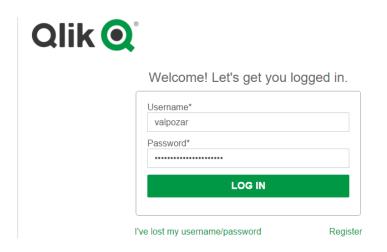
Slika 31: ERA model (model zvijezde)

#### 4. Qlik Sense

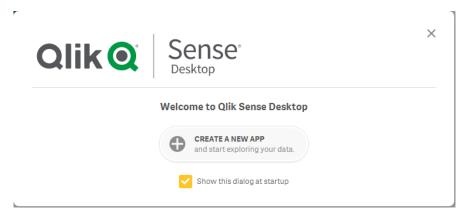
Qlik Sense je jedan od najboljih alata za analizu i vizualizaciju podataka zahvaljujući snažnom motoru, visokoj brzini obrade te odličnom izgledu i dojmu. Također, alat je intuitivan i "user-friendly" te omogućuje bilo kome u organizaciji da kreira vizualizacije, dublje istražuje podatke, otkriva veze te identificira nove mogućnosti. Godine 2018., Qlik je lansirao najnoviju verziju Qlik Sensea na tržište, koju krase nevjerojatne značajke. Neke prednosti Qlik Sensea su: vizualizacija povlačenjem i ispuštanjem (*eng. drag and drop*) koja otkriva skrivene obrasce, pametno pretraživanje za maksimalnu fleksibilnost i ostalo (Professional Advantage, bez dat.).

#### 4.1. Izrada izvještaja

Kako bi se mogao početi koristiti Qlik Sense Desktop potrebno se prvo registrirati, a zatim prijaviti. Nakon prijave može se kreirati nova aplikacija.



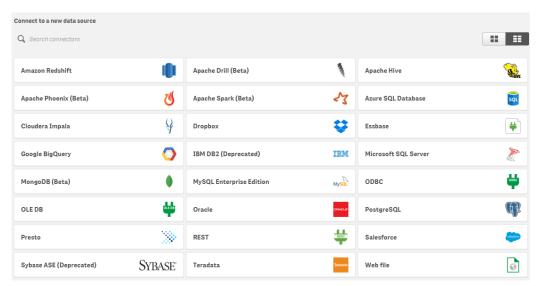
Slika 32: Prijava u Qlik Sense Desktop



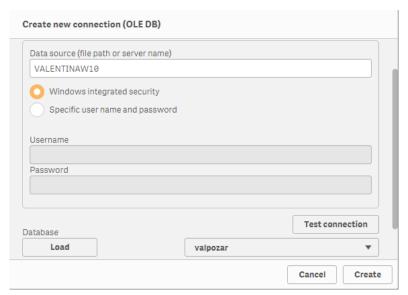
Slika 33: Kreiranje nove aplikacije

Kada se kreira nova aplikacija i unese naziv aplikacije, može se odabrati način povezivanja s izvorom podataka. U ovom slučaju odabran je OLE DB. OLE DB je skup sučelja koja izlažu podatke iz različitih izvora, kao što su proračunske tablice, projektni planovi, SQL, naslijeđene baze podataka i drugo. OLE DB sučelja pružaju aplikacije s jednakim pristupom podacima pohranjenim u različitim izvorima informacija ili pohranama podataka. Ta sučelja podržavaju količinu funkcionalnosti sustava za upravljanje bazom podataka (DBMS) prikladnu za pohranu podataka, omogućujući joj da dijeli svoje podatke (Webopedia, 2019).

Nakon odabira načina povezivanja s izvorom podataka, potrebno je unijeti podatke o bazi kako bi se mogli uspješno povezati.

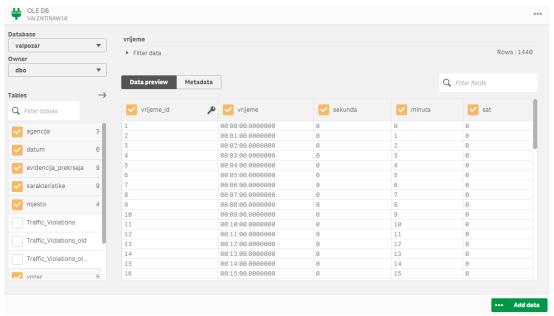


Slika 34: Odabir načina povezivanja s izvorom podataka

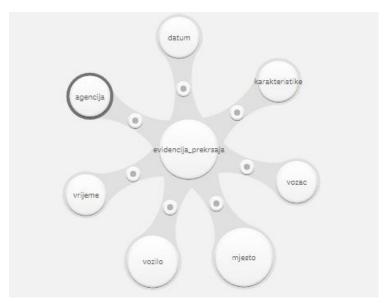


Slika 35: Kreiranje nove konekcije

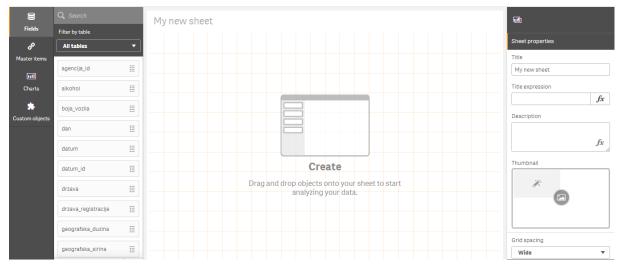
Nakon ovog koraka prikazuju se tablice iz kreiranog skladišta podataka. Moraju se označiti one tablice koje su potrebne za kreiranje raznih izvještaja te se potom mogu kreirati izvještaji sa svim podacima iz označenih tablica.



Slika 36: Odabir potrebnih tablica



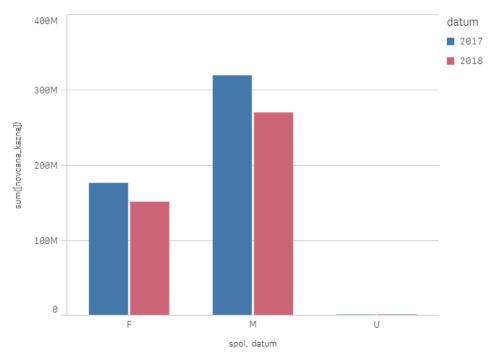
Slika 37: Prikaz povezanih tablica u alatu Qlik Sense



Slika 38: Kreiranje izvještaja

#### 4.1.1. Prvi izvještaj

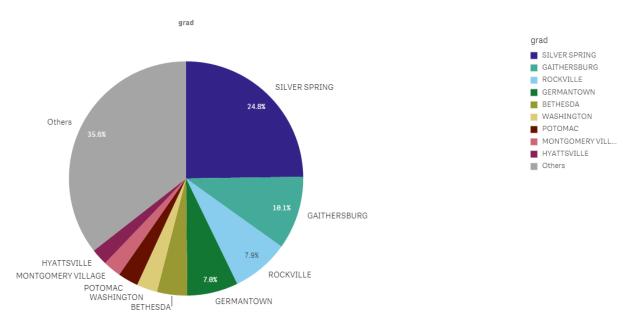
Prvi izvještaj prikazuje ukupnu sumu novčane kazne prema spolu vozača koji su počinili prometni prekršaj te prema godini kada je prekršaj počinjen. Može se vidjeti da su i 2017. godine i 2018. godine puno više novčano kažnjavane osobe muškog spola nego osobe ženskog spola. Također, ukupna suma novčane kazne je za osobe muškog spola veća 2017. godine nego 2018. godine. Isto tako, ukupna suma novčane kazne za osobe ženskog spola je veća 2017. godine nego 2018. godine.



Slika 39: Prvi izvještaj

#### 4.1.2. Drugi izvještaj

Drugi izvještaj prikazuje postotke ukupne sume materijalne štete prema nazivima gradova vozača koji su počinili prometni prekršaj.



Slika 40: Drugi izvještaj

#### 4.1.3. Treći izvještaj

Treći izvještaj prikazuje ukupnu sumu novčane kazne prema datumu te prema informaciji je li vozač koji je počinio prometni prekršaj koristio pojas ili ne. Može se vidjeti da su u navedenom razdoblju veću novčanu kaznu dobivali vozači koji nisu koristili pojas.



Slika 41: Treći izvještaj

#### 4.1.4. Četvrti izvještaj

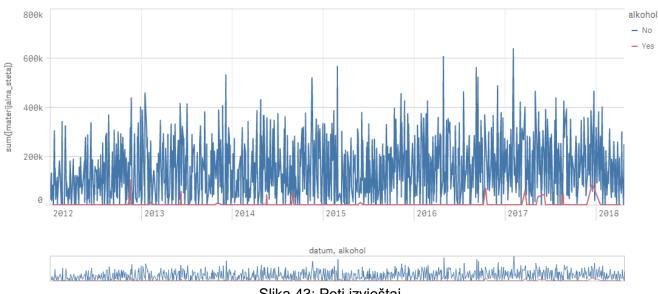
Četvrti izvještaj prikazuje sumu novčane kazne prema nazivu podagencije koja izdaje prometni prekršaj te prema nazivu lokacije gdje je prometni prekršaj počinjen.

naziv_podagencije	Q	naziv Q	sum([novcana_kazna])
Totals			8920500
"2nd district - Bethesda"		WISCONSIN AVE/STANFORD ST	27900
"2nd district - Bethesda"		WISCONSIN/ FAIRMONT AVE	26700
"2nd district - Bethesda"		WISCONSIN/NORFOLK	13600
"2nd district - Bethesda"		WISONSIN AVE/ WOODMONT AVE	10600
"2nd district - Bethesda"		WOODMONT AVE @ OLD GEORGETOWN RD	3400
"2nd district - Bethesda"		WOODMONT AVE/CORDELL AVE	8500
"3rd district - Silver Spring"		"8801 ARLISS STREEt - ON WALDEN ROAD"	30400
"3rd district - Silver Spring"		"9408 CURRAN RD SILVER SPRINt - MD"	13100
"3rd district - Silver Spring"		2ND AVE / RILEY RD	5500
"3rd district - Silver Spring"		16TH STREET AT SECOND AVENUE	8200
"3rd district - Silver Spring"		198/ OLD COLUMBIA PIKE	10800
"3rd district - Silver Spring"		198/ RIDING STABLE ROAD	4800
"3rd district - Silver Spring"		601 SLIGO AVE	19700
"3rd district - Silver Spring"		700 UNIVERSITY BLVD	14100
"3rd district - Silver Spring"		1220 EAST WEST HIGHWAY	13100
"3rd district - Silver Spring"		2101 FAIRLAND ROAD	7800
"3rd district - Silver Spring"		8482 PINEY BRANCH RD	9300
"3rd district - Silver Spring"		8701 ARLISS STREET	16400

Slika 42: Četvrti izvještaj

#### 4.1.5. Peti izvještaj

Peti izvještaj prikazuje ukupnu sumu materijalne štete prema godini te prema informaciji je li vozač koji je počinio prometni prekršaj bio u alkoholiziranom stanju.



Slika 43: Peti izvještaj

#### 5. Zaključak

U ovom projektnom zadatku prikazani su koraci po kojima je izrađeno skladište podataka. Prvi korak je bio pronalaženje izvora podataka te preuzimanje podataka u .csv formatu. Nakon toga je slijedilo čišćenje podataka, te prepravljanje formata nekih stupaca. Sljedeći korak je bio učitavanje .csv datoteke u alat Microsoft SQL Server Management Studio u kojem se izrađivalo skladište podataka. U tom alatu su se kreirale dimenzijske tablice i činjenična tablica na temelju učitanih podataka. Potom su se kreirane tablice popunile podacima primjenjujući upite. Vizualizacija podataka je rađena u alatu Qlik Sense. Podaci koji su analizirani i vizualizirani u ovom projektnom zadatku odnose se na analizu prometnih prekršaja na području Maryland Countya tijekom određenog vremenskog razdoblja. Qlik Sense je prilično jednostavan alat za korištenje, omogućuje različite prikaze podataka koji se mogu podesiti na različite načine sukladno potrebama korisnika. Općenito, uspostava skladišta podataka i sama poslovna inteligencija omogućuju brže, kvalitetnije i preciznije donošenje poslovnih odluka.

#### **Popis literature**

Kaggle (2018). *Traffic Violations in Maryland County*. Preuzeto 17.05.2019. s https://www.kaggle.com/rounak041993/traffic-violations-in-maryland-county

1Keydata (2019). *Data Warehouse Definition*. Preuzeto 10.06.2019. s https://www.1keydata.com/datawarehousing/dana-warehouse-definition.html

Professional Advantage (bez dat.). *What is Qlik Sense*. Preuzeto 13.06.2019. s <a href="https://www.pa.com.au/products/qlik-sense/">https://www.pa.com.au/products/qlik-sense/</a>

Rabuzin, K. (2018). *Skladišta podataka i poslovna inteligencija*. Materijali s predavanja. Dostupno 17.05.2019. na <a href="https://elf.foi.hr/course/view.php?id=142">https://elf.foi.hr/course/view.php?id=142</a>

Webopedia (2019). *OLE DB*. Preuzeto 13.06.2019 s <a href="https://www.webopedia.com/TERM/O/OLE\_DB.html">https://www.webopedia.com/TERM/O/OLE\_DB.html</a>

### Popis slika

Slika	1: Izgled izvornih podataka u .csv datoteci	4
Slika	2: Povezivanje na server	5
Slika	3: SQL server i skladište podataka	5
Slika	4: Učitavanje izvornih podataka	6
Slika	5: Odabir odredišne lokacije	6
Slika	6: Učitani izvorni podaci	7
Slika	7: Struktura tablice "agencija"	7
Slika	8: Upit za spremanje podataka u tablicu "agencija"	7
Slika	9: Prikaz podataka u tablici "agencija"	8
Slika	10: Struktura tablice "datum"	8
Slika	11: Upit za spremanje podataka u tablicu "datum"	8
Slika	12: Prikaz dijela podataka u tablici "datum"	9
Slika	13: Struktura tablice "karakteristike"	9
Slika	14: Upit za spremanje podataka u tablicu "karakteristike"	10
	15: Prikaz dijela podataka u tablici "karakteristike"	
	16: Struktura tablice "datum"	
Slika	17: Upit za spremanje podataka u tablicu "mjesto"	11
Slika	18: Prikaz dijela podataka iz tablice "mjesto"	11
Slika	19: Struktura tablice "vozac"	11
	20: Upit za spremanje podataka u tablicu "vozac"	
Slika	21: Prikaz dijela podataka iz tablice "vozac"	12
Slika	22: Struktura tablice "vozilo"	12
Slika	23: Upit za spremanje podataka u tablicu "vozilo"	13
Slika	24: Prikaz dijela podataka iz tablice "vozilo"	13
Slika	25: Struktura tablice "vrijeme"	13
Slika	26: Upit za spremanje podataka u tablicu "vrijeme"	14
Slika	27: Prikaz dijela podataka iz tablice "vrijeme"	14
	28: Struktura tablice "evidencija_prekrsaja"	
Slika	29: Upit za spremanje podataka u tablicu "evidencija_prekrsaja"	15
Slika	30: Prikaz dijela podataka iz tablice "evidencija_prekrsaja"	16
Slika	31: ERA model (model zvijezde)	17
Slika	32: Prijava u Qlik Sense Desktop	18
Slika	33: Kreiranje nove aplikacije	18
Slika	34: Odabir načina povezivanja s izvorom podataka	19
Slika	35: Kreiranje nove konekcije	19
Slika	36: Odabir potrebnih tablica	20
Slika	37: Prikaz povezanih tablica u alatu Qlik Sense	20

Slika 38: Kreiranje izvještaja	21
Slika 39: Prvi izvještaj	21
Slika 40: Drugi izvještaj	22
Slika 41: Treći izvještaj	22
Slika 42: Četvrti izvještaj	23
Slika 43: Peti izvještaj	23