



KRIMINAL U BOSTONU

AUTOR: MATEJ VIŠNJIĆ, INFORMATIKA 2

KOLEGIJ: SUSTAVI POSLOVNE INTELIGENCIJE

OPIS

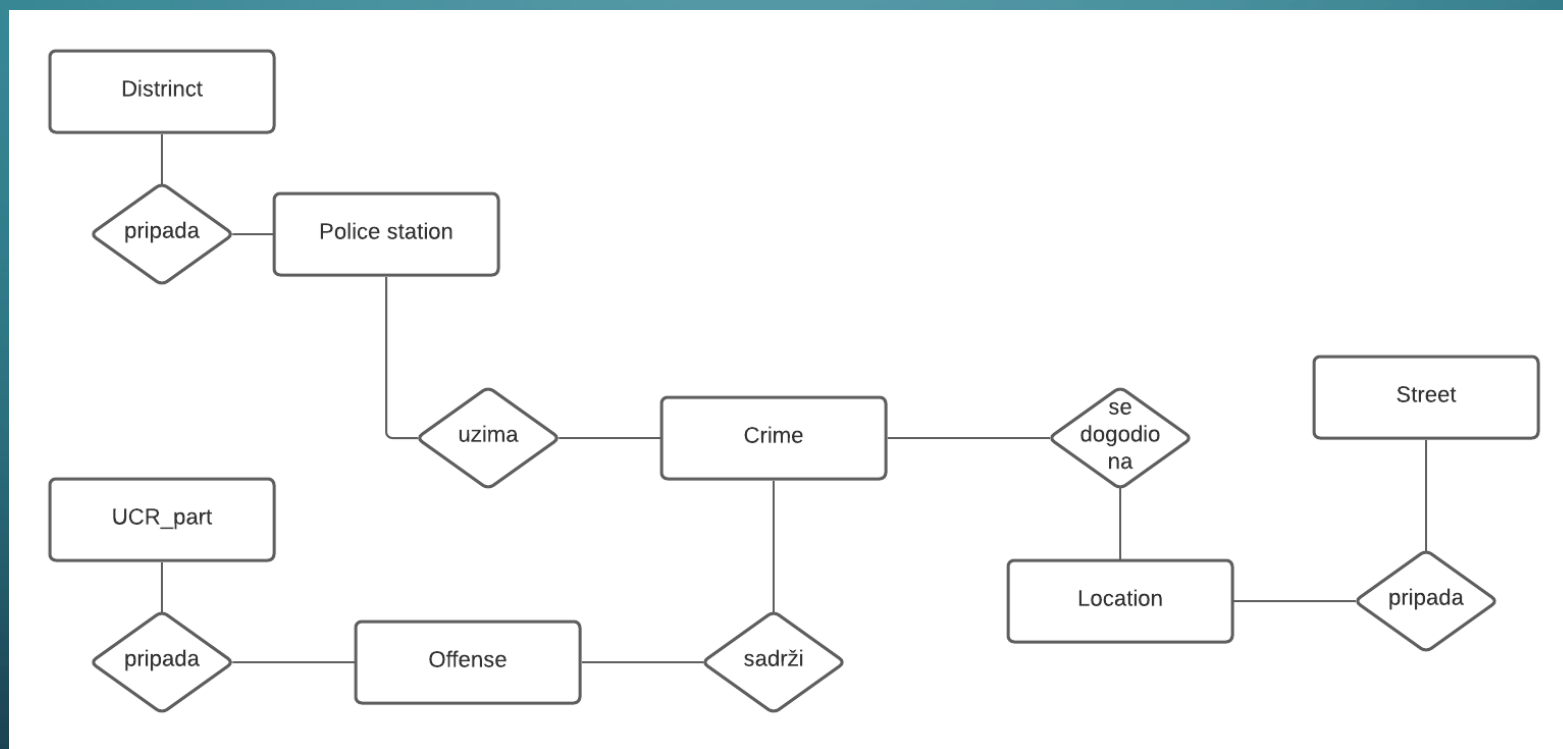
- U ovome projektu nalaze se podaci koji su sakupljeni od 2015 – 2018 godine.
- Podaci sadrže zapise zločina u Bostonu koji su se desili u tom periodu.

Svaki zločin ima svoj identifikacijski broj, vrijeme i lokaciju kada je počinjen, opis zločina, vrsta, da li je pucano. Imamo i popis policijskih postaji koje su vodile slučaj, i njihov okrug gdje se nalaze.

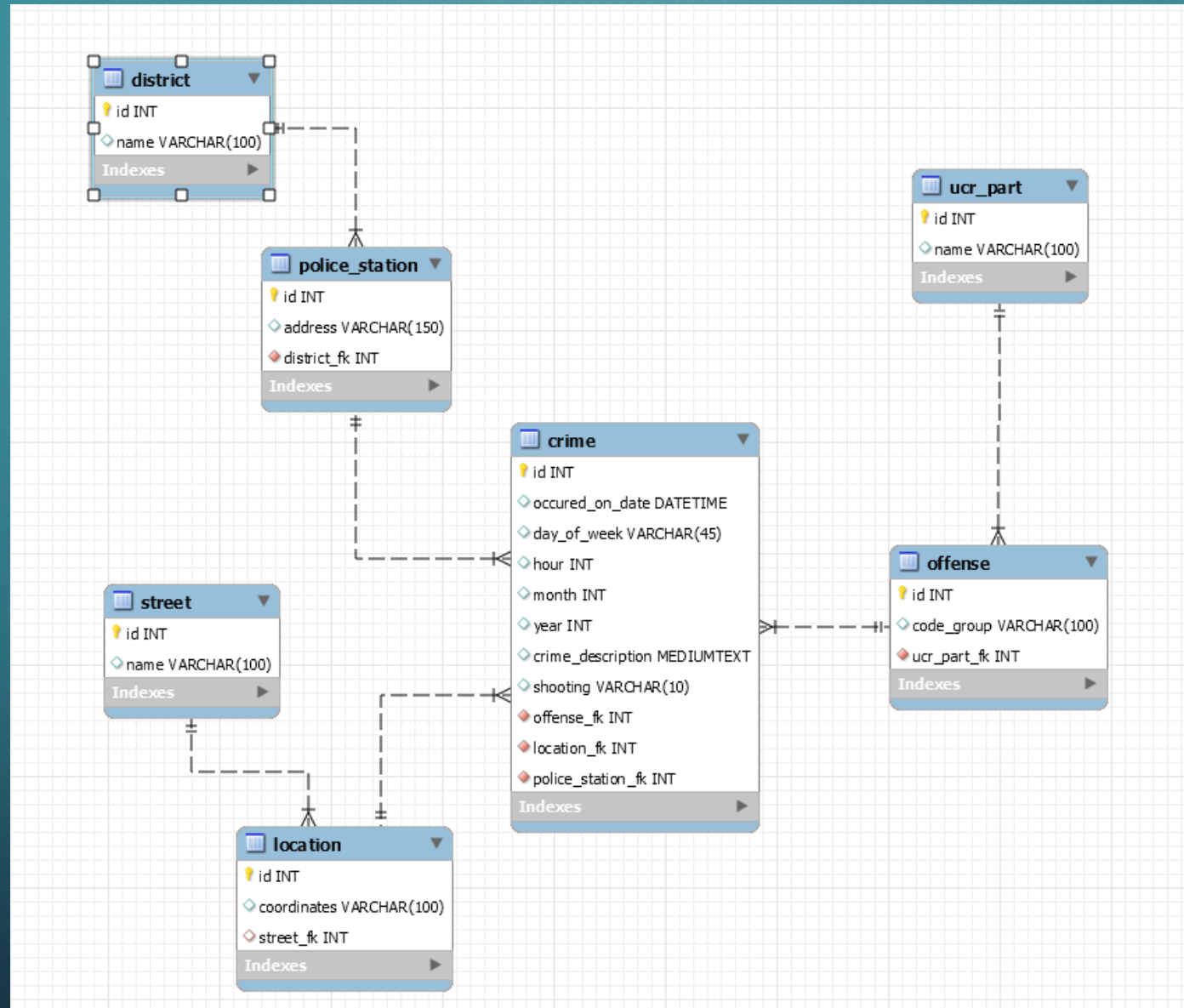
PODACI

- Link: <https://www.kaggle.com/AnalyzeBoston/crimes-in-boston>
- Analiza podataka:
 - Oko 319 tisuća zapisa
 - Dimenzija vremena i dimenzija lokacije
 - Različiti tipovi podataka
 - Postoji vremenska dimenzija

ER DIAGRAM



ER MODEL

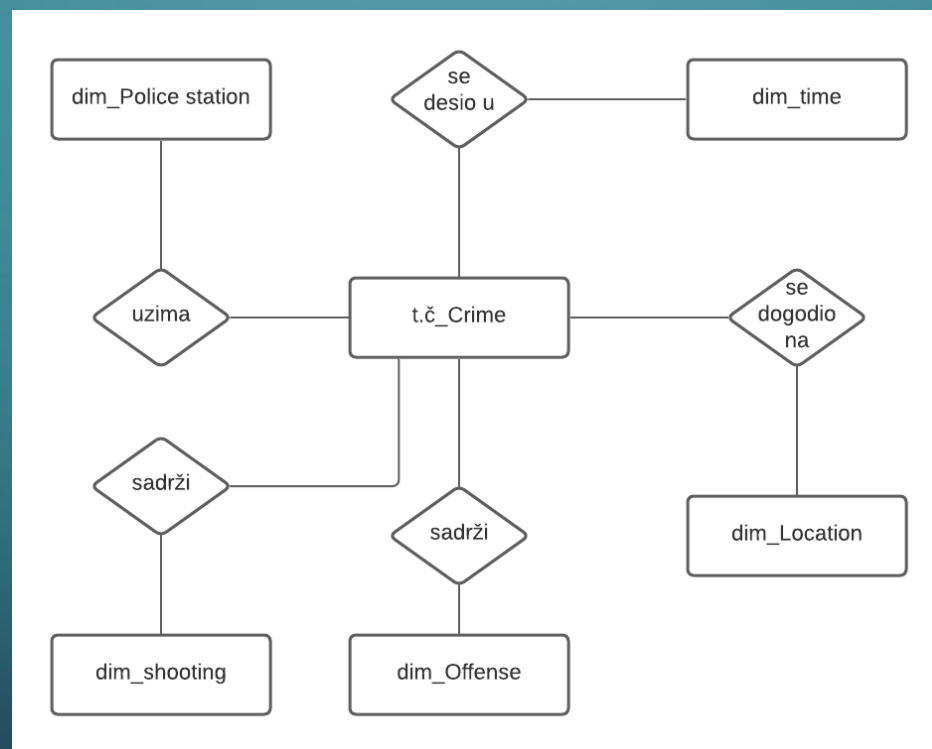


PUNJENJE RELACIJSKE BAZE PODATAKA

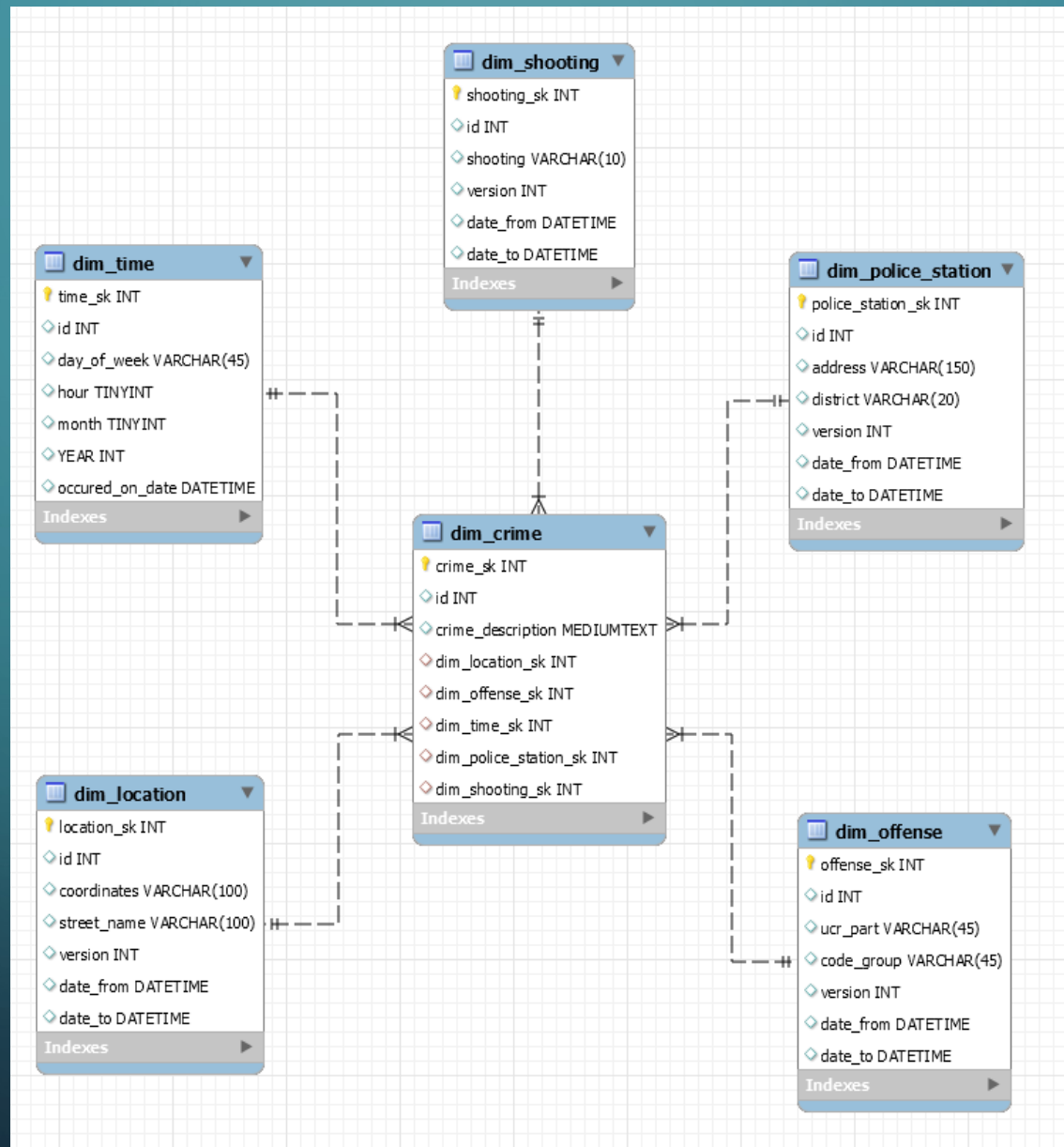
```
#DDL
district_ddl = "CREATE TABLE crimes.district (id INT NOT NULL PRIMARY KEY, name VARCHAR(100), UNIQUE INDEX id_UNIQUE (id ASC));"
connection.execute(district_ddl)
street_ddl = "CREATE TABLE crimes.street (id INT NOT NULL PRIMARY KEY, name VARCHAR(100), UNIQUE INDEX id_UNIQUE (id ASC));"
connection.execute(street_ddl)
ucr_part_ddl = "CREATE TABLE crimes.ucr_part (id INT NOT NULL PRIMARY KEY, name VARCHAR(100), UNIQUE INDEX id_UNIQUE(id ASC));"
connection.execute(ucr_part_ddl)
location_ddl = "CREATE TABLE crimes.location (id INT NOT NULL PRIMARY KEY, coordinates VARCHAR(100), street_fk INT, UNIQUE INDEX id_UNIQUE(id ASC), CONSTRAINT :
connection.execute(location_ddl)
police_station_ddl = "CREATE TABLE crimes.police_station (id INT NOT NULL PRIMARY KEY, address VARCHAR(150), district_fk INT NOT NULL, UNIQUE INDEX id_UNIQUE(id
connection.execute(police_station_ddl)
offense_ddl = "CREATE TABLE crimes.offense (id INT NOT NULL PRIMARY KEY, code_group VARCHAR(100), ucr_part_fk INT NOT NULL, UNIQUE INDEX id_UNIQUE(id ASC), COI
connection.execute(offense_ddl)
crime_ddl = "CREATE TABLE crimes.crime (id INT NOT NULL PRIMARY KEY, occurred_on_date DATETIME, day_of_week VARCHAR(45), hour INT, month INT, year INT, crime_de
connection.execute(crime_ddl)
```

```
#POLICE_STATION
# id,address,district_fk
district_id=[]
for i,row in df.iterrows():
    district_id.append(int(district_data['id'].iloc[i]))
policestation_data = pd.DataFrame({'id':list(range(1,len(district_id)+1)), 'district_fk': district_id})
policestation_data.to_sql(con=mydb, name='police_station', if_exists='append', index=False)
```

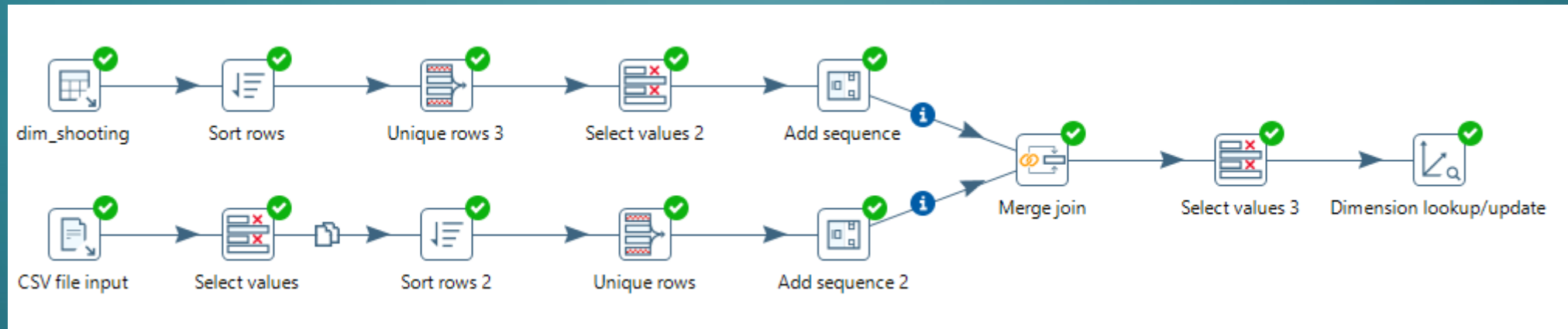

DIMENZIJSKI DIAGRAM



DIMENZIJSKI MODEL

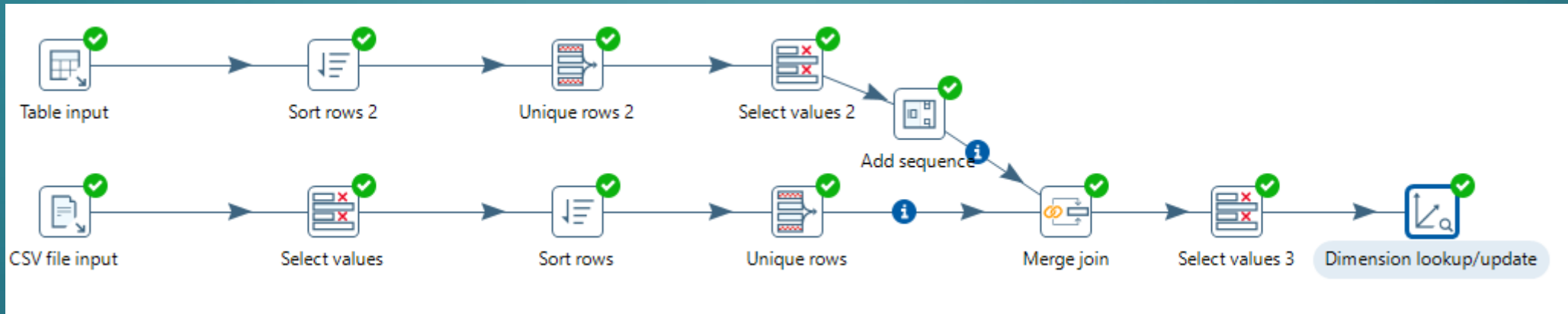


DIMENZIJA DIM_SHOOTING

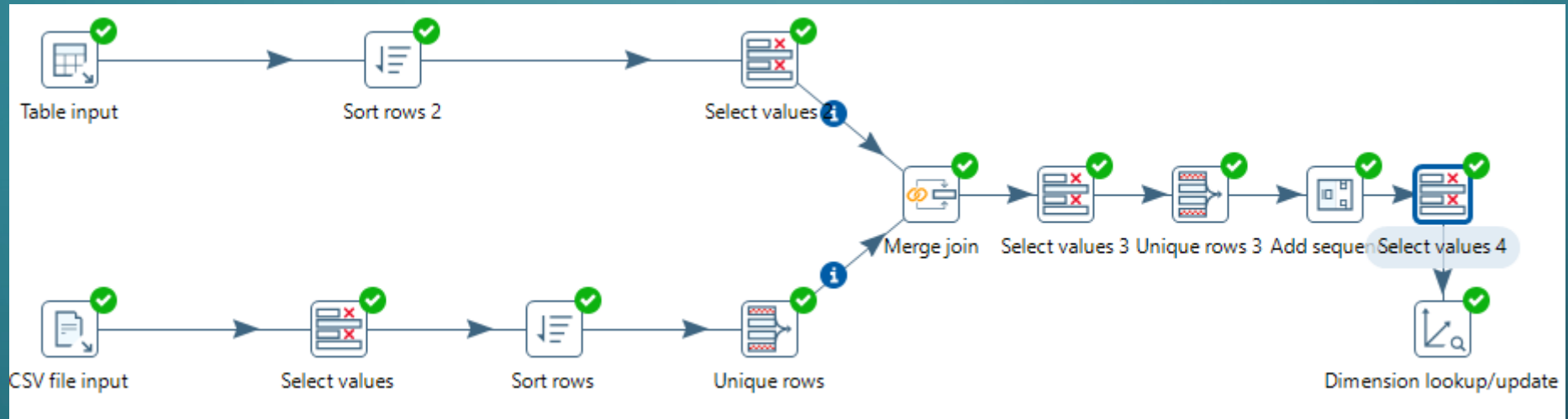


	shooting_sk	id	shooting	version	date_from	date_to
▶	0	NULL	NULL	1	NULL	NULL
	1	1	N	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	2	2	Y	1	1900-01-01 00:00:00	2200-01-01 00:00:00
*	NULL	NULL	NULL	NULL	NULL	NULL

DIMENZIJA DIM_POLICE_STATION

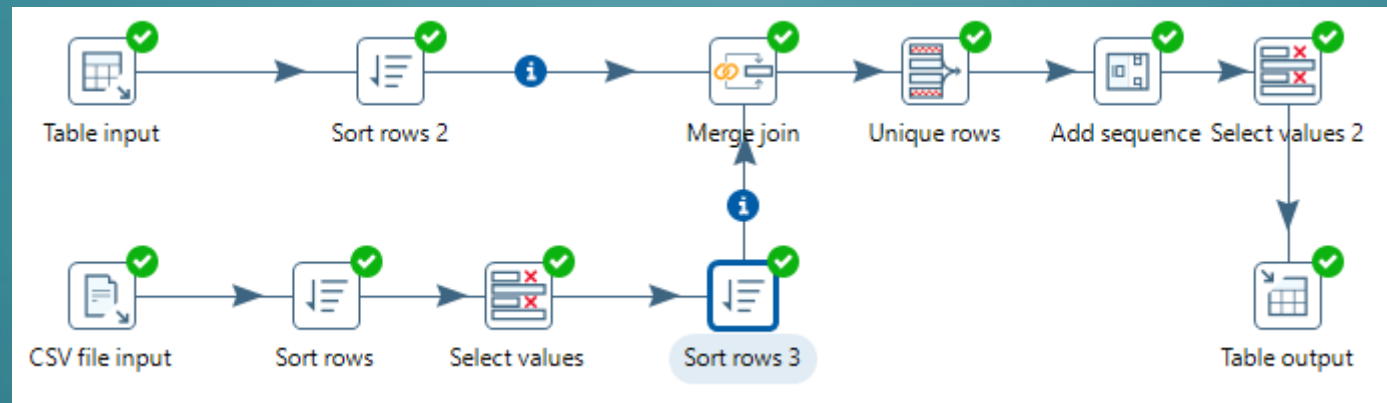
[illegible]

DIMENZIJA DIM_OFFENSE



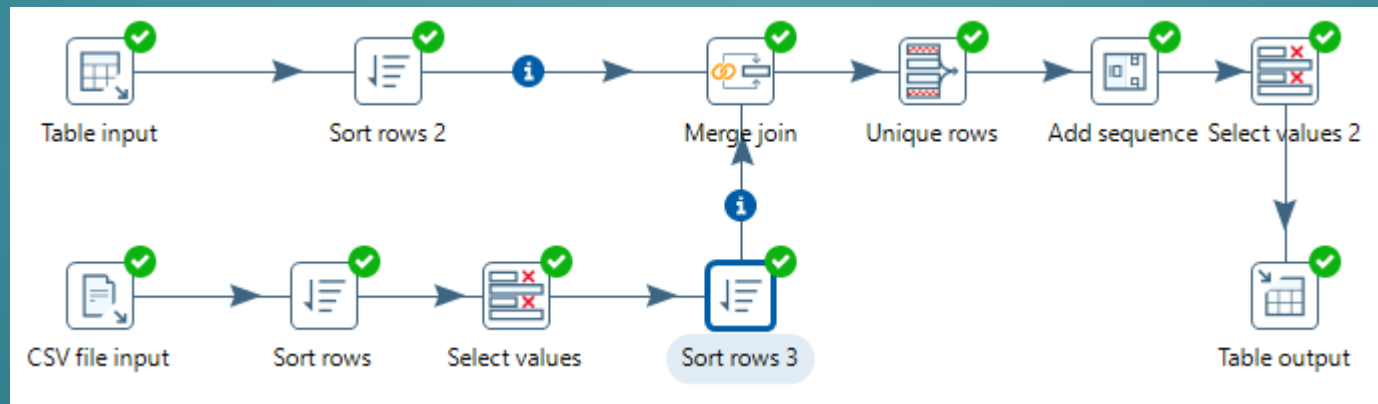
	offense_sk	id	ucr_part	code_group	version	date_from	date_to
	0	NULL	NULL	NULL	1	NULL	NULL
	1	1	Part One	Aggravated Assault	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	2	2	Part Three	Aircraft	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	3	3	Other	Arson	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	4	4	Part Three	Assembly or Gathering Violations	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	5	5	Part One	Auto Theft	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	6	6	Other	Auto Theft Recovery	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	7	7	Part Two	Ballistics	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	8	8	Part Two	Biological Threat	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	9	9	Part Two	Bomb Hoax	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	10	10	Other	Burglary - No Property Taken	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	11	11	Part One	Commercial Burglary	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	12	12	Part Two	Confidence Games	1	1900-01-01 00:00:00	2200-01-01 00:00:00
	13	13	Part Two	Counterfeiting	1	1900-01-01 00:00:00	2200-01-01 00:00:00
▶	14	14	Part Two	Criminal Harassment	1	1900-01-01 00:00:00	2200-01-01 00:00:00

DIMENZIJA DIM_LOCATION



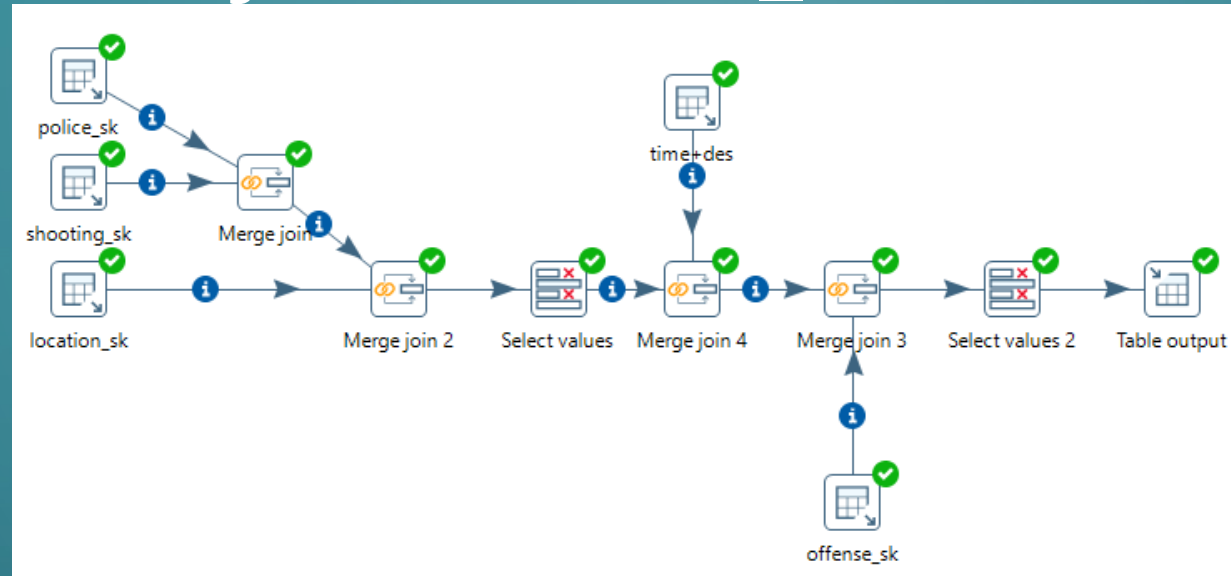
	location_sk	id	coordinates	street_name	version	date_from	date_to
▶ 0		NULL	NULL	NULL	1	NULL	NULL
1		1	(42.33361000, -71.07337000)	ALBANY ST	1	1900-01-01 00:00:00	2200-01-01 00:00:00
2		2	(42.31730370, -71.07799648)	BLUE HILL AVE	1	1900-01-01 00:00:00	2200-01-01 00:00:00
3		3	(42.34060371, -71.08174185)	COLUMBUS AVE	1	1900-01-01 00:00:00	2200-01-01 00:00:00
4		4	(42.35053956, -71.13106722)	COMMONWEALTH AVE	1	1900-01-01 00:00:00	2200-01-01 00:00:00
5		5	(42.33253100, -71.07213000)	MASSACHUSETTS AVE	1	1900-01-01 00:00:00	2200-01-01 00:00:00
6		6	(0.00000000, 0.00000000)	0 BURRELL	1	1900-01-01 00:00:00	2200-01-01 00:00:00
7		7	(0.00000000, 0.00000000)	0 MASS AVE	1	1900-01-01 00:00:00	2200-01-01 00:00:00
8		8	(0.00000000, 0.00000000)	00 CENTRE ST	1	1900-01-01 00:00:00	2200-01-01 00:00:00
9		9	(0.00000000, 0.00000000)	00 MASS AVE	1	1900-01-01 00:00:00	2200-01-01 00:00:00
10		10	(0.00000000, 0.00000000)	00 OLD COLONY AVE	1	1900-01-01 00:00:00	2200-01-01 00:00:00
11		11	(0.00000000, 0.00000000)	12 BREED ST	1	1900-01-01 00:00:00	2200-01-01 00:00:00
12		12	(0.00000000, 0.00000000)	12 BROOKVIEW ST	1	1900-01-01 00:00:00	2200-01-01 00:00:00
13		13	(0.00000000, 0.00000000)	12 WOODWARD PARK ST	1	1900-01-01 00:00:00	2200-01-01 00:00:00
14		14	(0.00000000, 0.00000000)	13 DIXWELL ST	1	1900-01-01 00:00:00	2200-01-01 00:00:00

DIMENZIJA DIM_TIME



	time_sk	id	day_of_week	hour	month	YEAR	occured_on_date
▶	1	1	Friday	0	1	2016	2016-01-01 00:00:00
	2	2	Friday	0	1	2016	2016-01-01 00:01:00
	3	3	Friday	0	1	2016	2016-01-01 00:05:00
	4	4	Friday	0	1	2016	2016-01-01 00:09:00
	5	5	Friday	0	1	2016	2016-01-01 00:13:00
	6	6	Friday	0	1	2016	2016-01-01 00:14:00
	7	7	Friday	0	1	2016	2016-01-01 00:15:00
	8	8	Friday	0	1	2016	2016-01-01 00:16:00
	9	9	Friday	0	1	2016	2016-01-01 00:28:00
	10	10	Friday	0	1	2016	2016-01-01 00:30:00
	11	11	Friday	0	1	2016	2016-01-01 00:35:00
	12	12	Friday	0	1	2016	2016-01-01 00:37:00
	13	13	Friday	0	1	2016	2016-01-01 00:38:00
	14	14	Friday	0	1	2016	2016-01-01 00:39:00
	15	15	Friday	0	1	2016	2016-01-01 00:43:00

TABLICA ČINJENICA DIM_CRIME



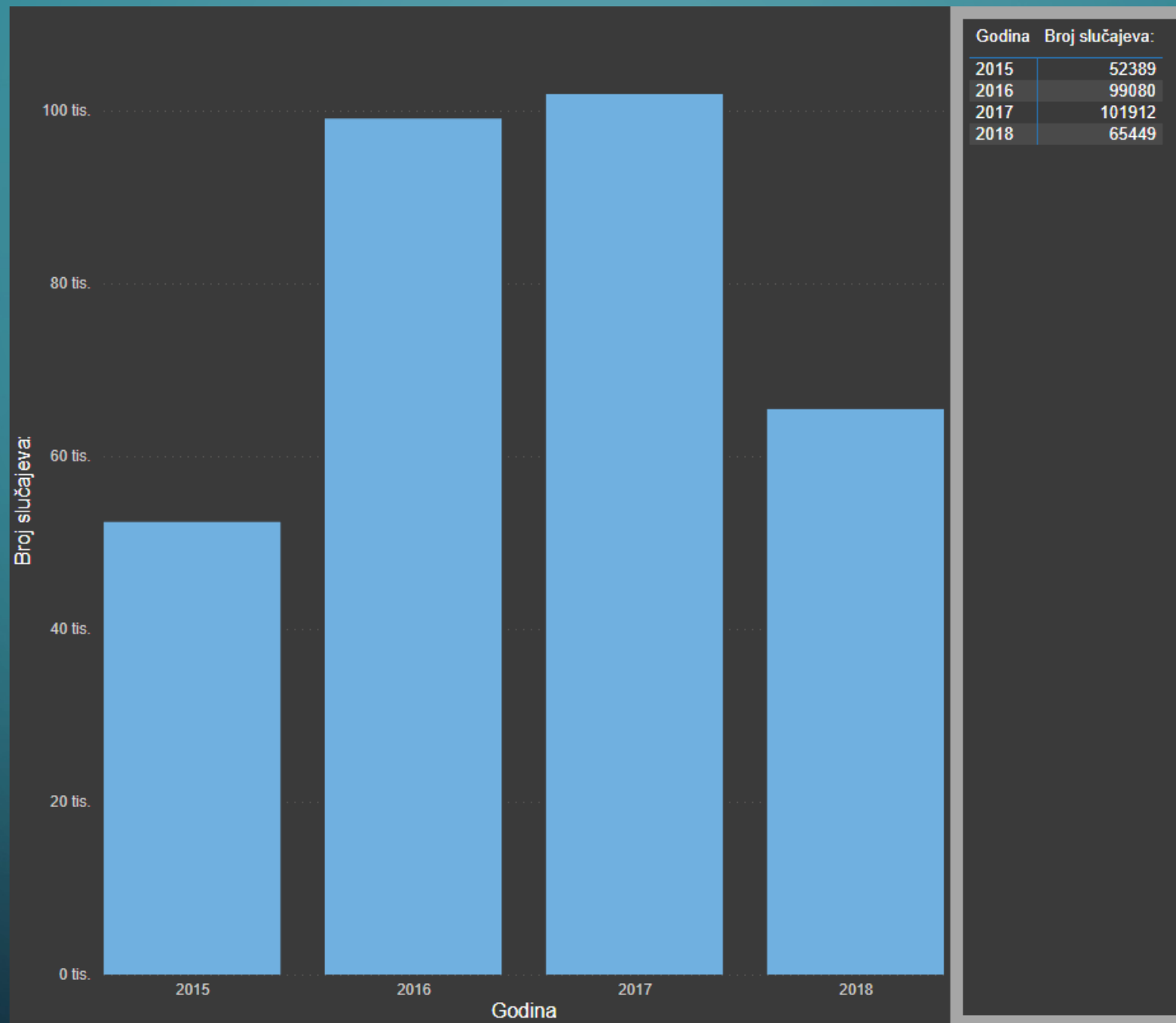
```
SELECT dp.police_station_sk, c.id
FROM crimes.crime c,
     crimes.offense o,
     crimes.location l,
     crimes.police_station p,
     crimes.district d
INNER JOIN crimes_dimenzije.dim_police_station dp
ON dp.district = d.name
WHERE c.offense_fk = o.id
AND c.location_fk = l.id
AND c.police_station_fk = p.id
AND d.id = p.district_fk;
```

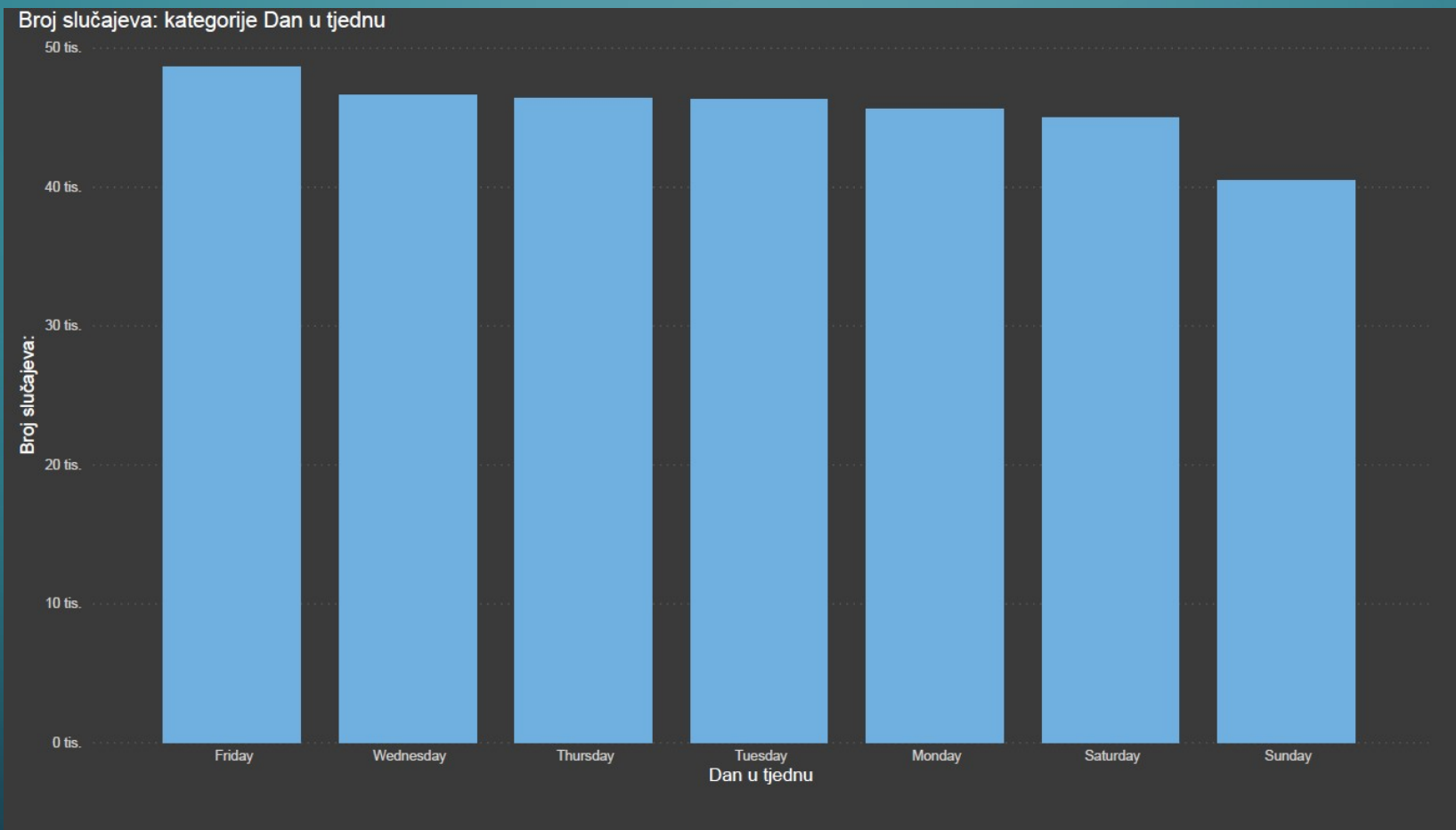

TABLICA ČINJENICA DIM_CRIME

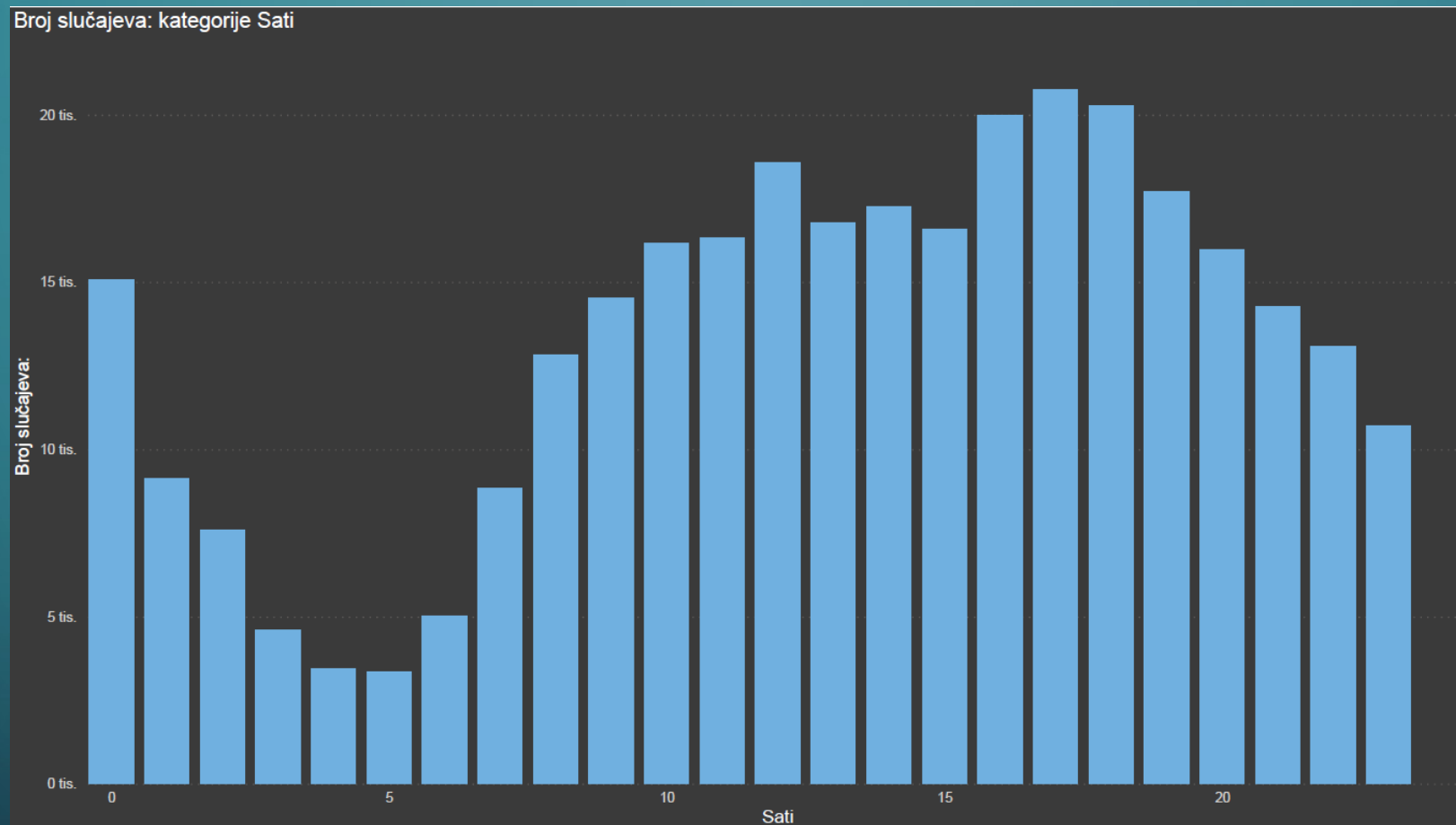
	crime_sk	id	crime_description	dim_location_sk	dim_offense_sk	dim_time_sk	dim_police_station_sk	dim_shooting_sk
▶	1	1	LARCENY ALL OTHERS	2536	35	110302	8	1
	2	2	VANDALISM	2074	64	5917	6	1
	3	3	TOWED MOTOR VEHICLE	785	63	190628	9	1
	4	4	INVESTIGATE PROPERTY	3067	33	213269	9	1
	5	5	INVESTIGATE PROPERTY	1241	33	213267	5	1
	6	6	M/V ACCIDENT INVOLVING PEDESTRIAN - INJURY	4075	44	213268	6	1
	7	7	AUTO THEFT	3100	5	213270	4	1
	8	8	VERBAL DISPUTE	2460	65	202386	4	1
	9	9	ROBBERY - STREET	2741	59	202387	7	1
	10	10	VERBAL DISPUTE	2504	65	202384	6	1
	11	11	VERBAL DISPUTE	3160	65	190631	7	1
	12	12	INVESTIGATE PROPERTY	1174	33	202383	7	1
	13	13	FIRE REPORT - HOUSE, BUILDING, ETC.	2709	20	190632	9	1
	14	15	PROPERTY - LOST	2968	53	122881	5	1
	15	16	SICK/INJURED/MEDICAL - PERSON	2987	41	190629	13	1

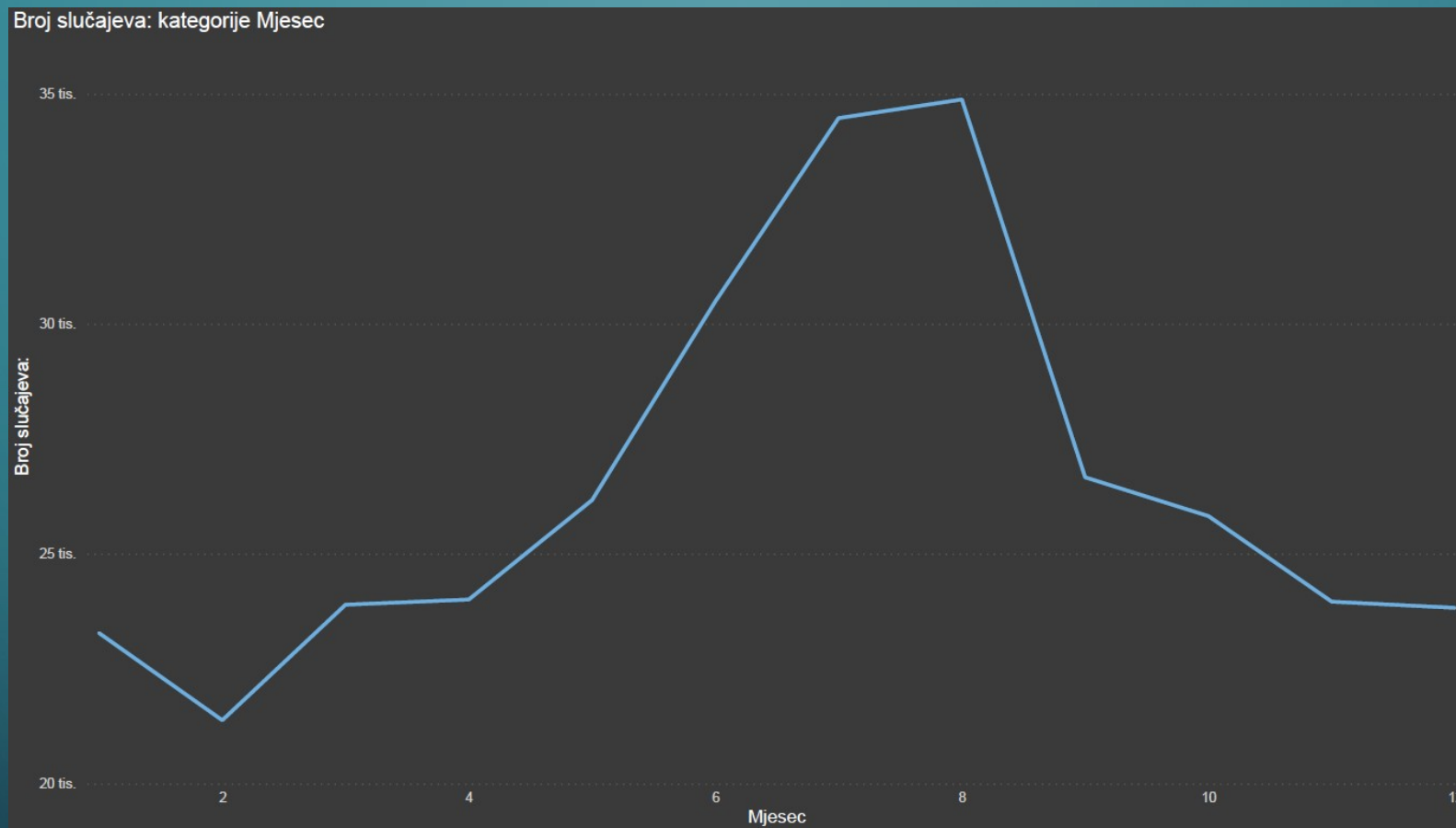
VIZUALIZACIJA PODATAKA

- Za vizualizaciju koristio sam Microsoft Power BI
- Jednostavna usluga tvrtke Microsoft
- Mogućnost poslovne inteligencije s jednostavnim sučeljem
- Krajnji korisnici mogu stvoriti vlastita izvješća i nadzorne ploče

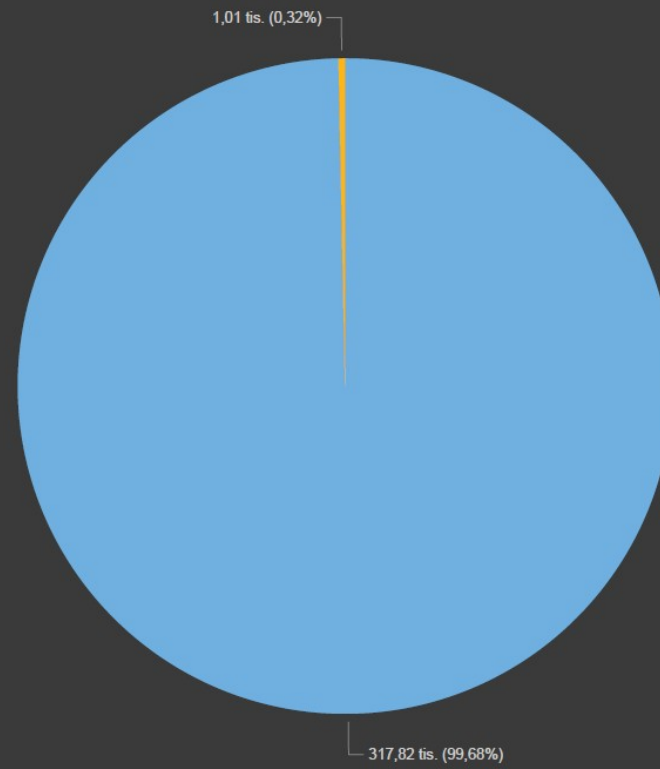




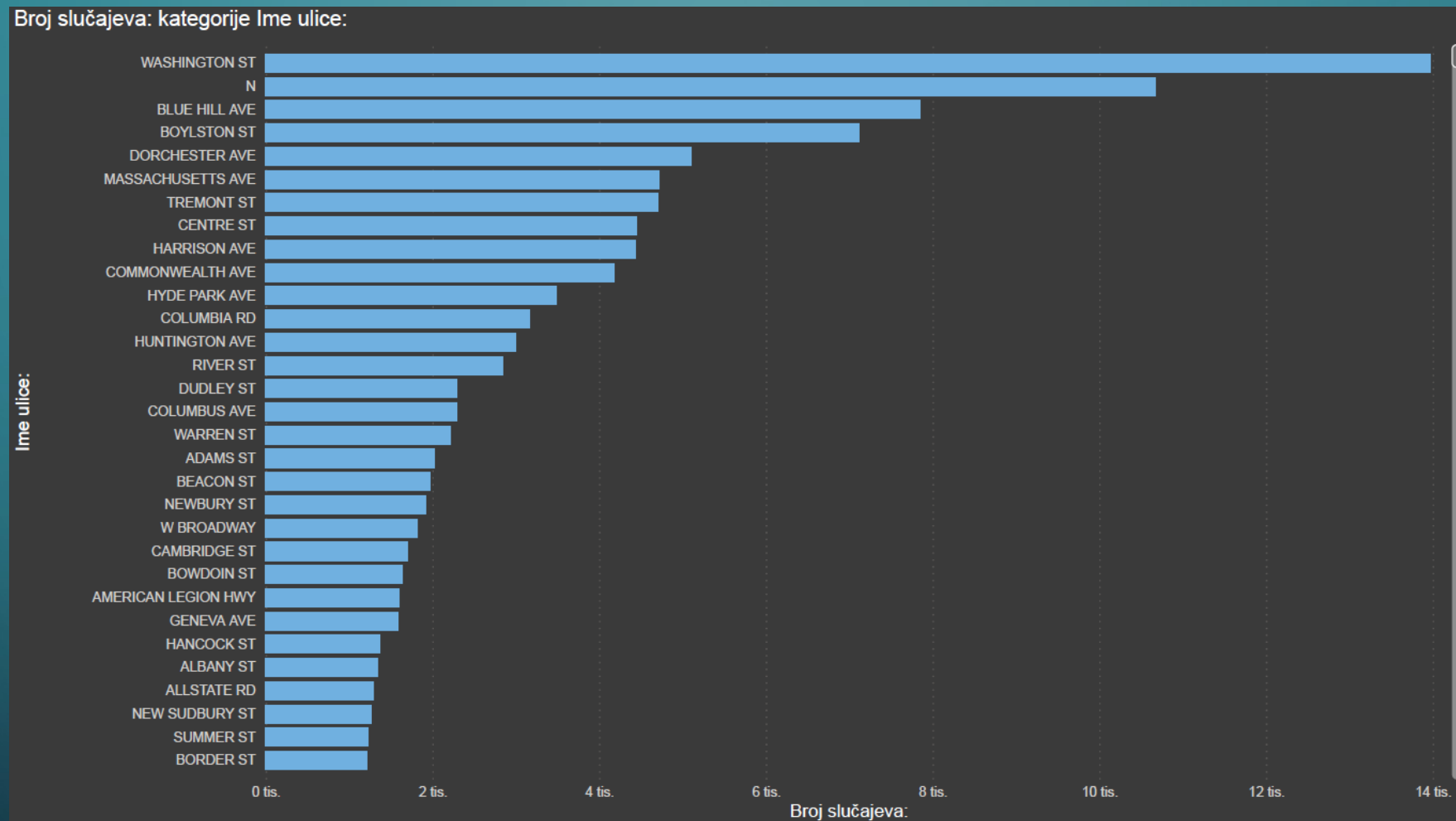




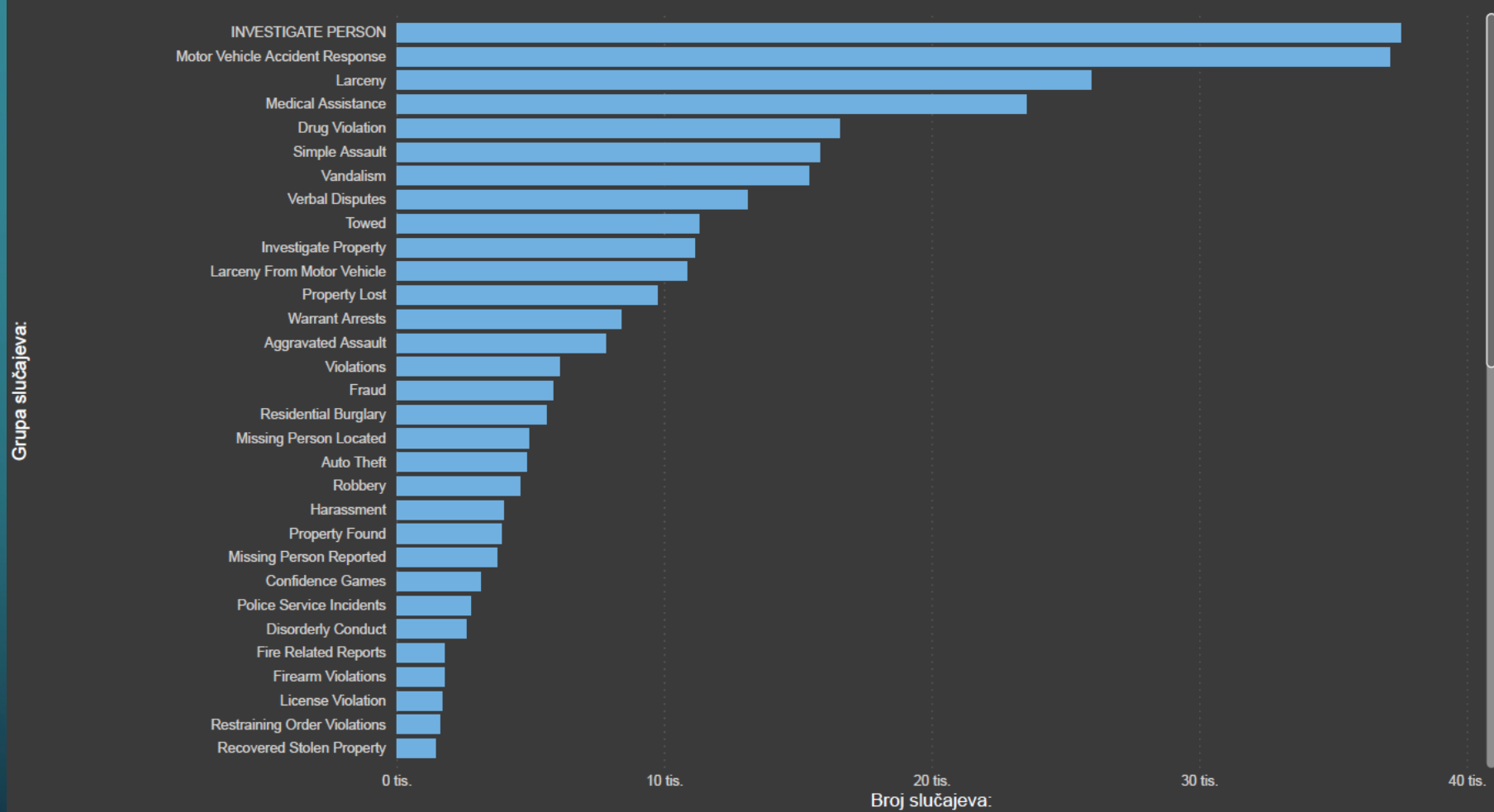
Broj slučajeva: kategorije Pucanj:



Pucanj:
● N
● Y



Broj slučajeva: kategorije Grupa slučajeva:



Broj slučajeva: kategorije Okrug

