

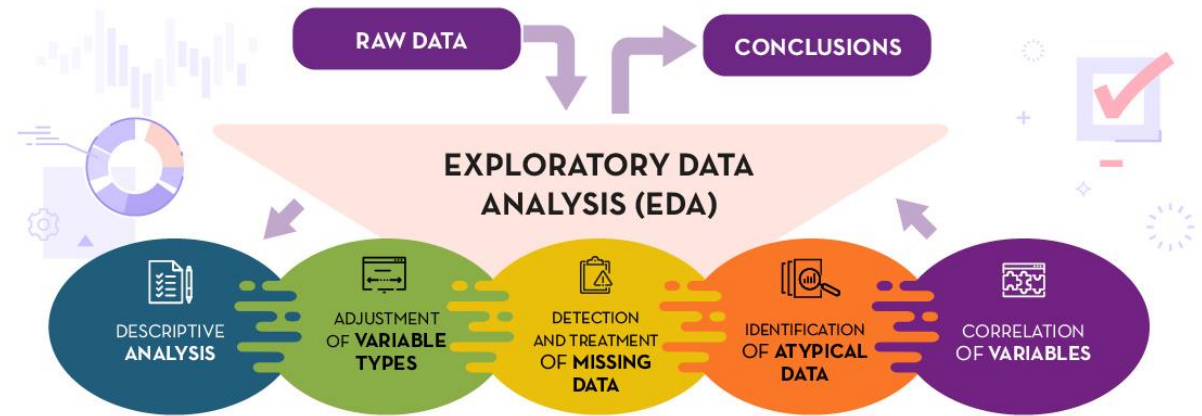
Eksplorativna analiza podataka o vremenskim prilikama i kvalitetu vazduha u Šangaju u periodu 2014-2021

Autor: Miloš Sirar IN 3/2020

Predmet: Metode i tehnike nauke o podacima

Sadržaj istraživanja

- Definisanje ideje i istraživačkog cilja
- Prikupljanje podataka
- Priprema podataka
- Istraživanje podataka
- Prezentovanje rezultata



Definisanje ideje i istraživačkog cilja

- Tema istraživanja
- Koja je korist ovog istraživanja
- Podaci
- Šta je cilj istraživanja



Šangaj

- Najveći grad Kine, uz obalu Pacifika
- Broj stanovnika: oko 27 miliona (oko 40 miliona)
- Vlažna suptropska klima
- Četiri godišnja doba
- Proleće – često promenljivo i kišovito
- Jesen – suva i topla
- Zime – umerene
- Leta – topla i vlažna sa povremenim pljuskovima



Prikupljanje podataka

- Dva skupa podataka sa Kaggle-a
 1. Podaci o vremenskim uslovima – podaci_vreme_sangaj
 2. Podaci o štetnim česticama i gasovima – podaci_kvalitet_vazduha_sangaj



Programski jezik i biblioteke



Priprema podataka – shanghai.csv

- Učitavanje podataka
- Provera da li postoje nepostojeće vrednosti
- Promena naziva kolona
- Uklanjanje kolona
- Promena tipova kolona
- Dodavanje novih kolona

Priprema podataka – shanghai.csv

```
Tipovi kolona:  
date_time          object  
maxtempC           int64  
mintempC           int64  
totalSnow_cm       float64  
sunHour            float64  
uvIndex            int64  
moon_illumination  int64  
moonrise           object  
moonset            object  
sunrise            object  
sunset             object  
DewPointC          int64  
FeelsLikeC         int64  
HeatIndexC         int64  
WindChillC         int64  
WindGustKmph       int64  
cloudcover         int64  
humidity           int64  
precipMM           float64  
pressure           int64  
tempC              int64  
visibility          int64  
winddirDegree      int64  
windspeedKmph      int64  
location           object  
dtype: object
```



```
Prikaz nakon uklanjanja kolona koje nisu potrebne za ovo istraživanje:  
datum_merenja      object  
max_dnevna_temperatura  int64  
min_dnevna_temperatura  int64  
oblacnost           int64  
padavine_mm         float64  
vidljivost          int64  
pravac_vetra        int64  
brzina_vetra_kmh    int64  
dtype: object
```



```
datum_merenja      datetime64[ns]  
godina_merenja     int64  
mesec_merenja      object  
dan_merenja        int64  
max_dnevna_temperatura  int64  
min_dnevna_temperatura  int64  
oblacnost           int64  
padavine_mm         float64  
vidljivost          int64  
pravac_vetra        int64  
brzina_vetra_kmh    int64  
dtype: object
```


Priprema podataka – shanghai-air-quality.csv

- Učitavanje podataka
- Promena naziva kolona
- Promena tipova kolona
- Zamena nepostojećih vrednosti sa -1.0

```
Tipovi kolona:  
date      object  
pm25      object  
pm10      object  
o3        object  
no2       object  
so2       object  
co        object  
dtype: object
```



```
Prikaz nakon promene naziva kolona na srpski jezik:  
datum_merenja  object  
PM25           object  
PM10           object  
O3             object  
NO2            object  
SO2            object  
CO             object  
dtype: object
```



```
datum_merenja  datetime64[ns]  
PM25           float64  
PM10           float64  
O3             float64  
NO2            float64  
SO2            float64  
CO             float64  
dtype: object
```

Priprema podataka – spajanje podataka

- Spajanje dva skupa podataka pomoću „ključa“, odnosno datuma merenja

```
podaci = pd.merge(podaci_vreme_sangaj, podaci_kvalitet_vazduha_sangaj, on="datum_merenja", how="left")
```

- Objedinjen skup podataka
- Sada može da počne rad nad tim podacima

Podaci o vremenskim uslovima



```
datum_merenja      datetime64[ns]
godina_merenja      int64
mesec_merenja       object
dan_merenja         int64
max_dnevna_temperatura int64
min_dnevna_temperatura int64
oblacnost           int64
padavine_mm         float64
vidljivost          int64
pravac_vetra        int64
brzina_vetra_kmh    int64
PM25                float64
PM10                float64
O3                  float64
NO2                  float64
SO2                  float64
CO                  float64
dtype: object
(2587, 17)
```



Podaci o štetnim česticama i gasovima

Početna istraživanja podataka

- Da li svih 8 godina imaju dovoljan broj uzoraka?

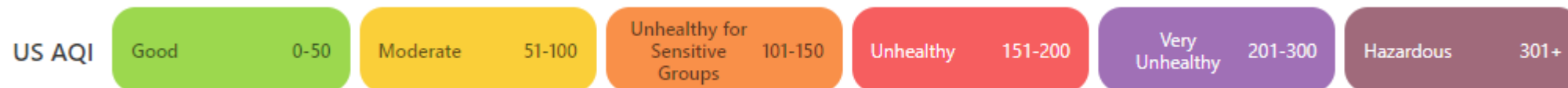
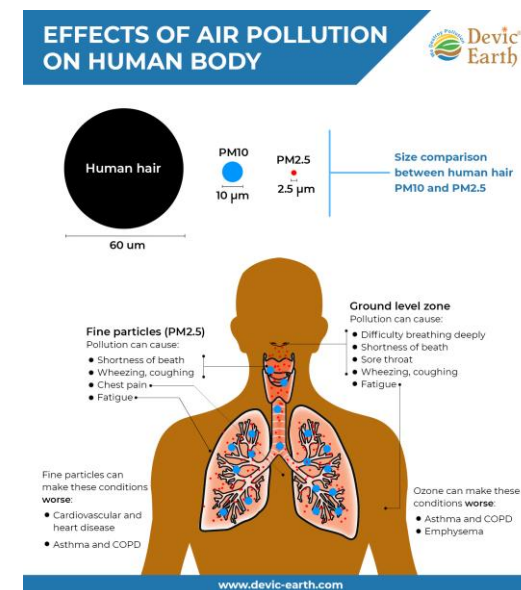
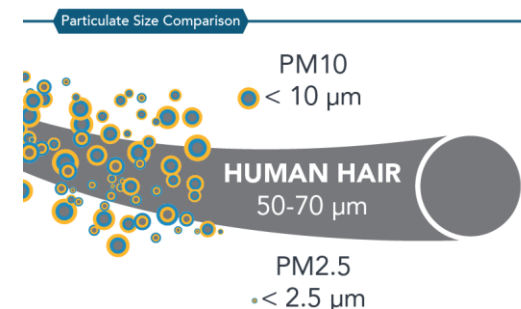
```
godina_merenja
2014    365
2015    365
2016    366
2017    365
2018    365
2019    365
2020    366
2021     30
Name: datum_merenja, dtype: int64

Podaci za 2021. godinu ce biti izbaceni iz skupa podataka, jer nemaju dovoljan broj podataka za analizu.

godina_merenja
2014    365
2015    365
2016    366
2017    365
2018    365
2019    365
2020    366
Name: datum_merenja, dtype: int64
```

Izračunavanje AQI (Air Quality Index)

- Vazduh je pun sitnih štetnih čestica i gasova
- Oni u manjoj ili većoj meri utiču na kvalitet vazduha
- Čestice:
 1. pm 2.5
 2. pm 10
- Gasovi:
 1. O₃ – ozon
 2. NO₂ – azot dioksid
 3. SO₂ – sumpor dioksid
 4. CO – ugljen monoksid



Izračunavanje AQI (Air Quality Index)

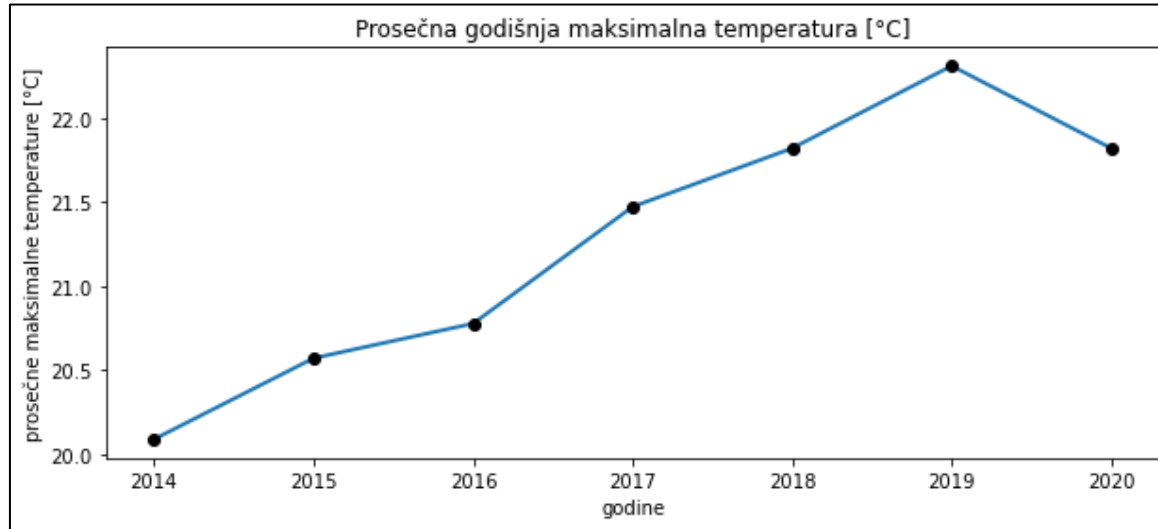
- Potrebni sub indeksi da bi se izračunao AQI
- Svaki sub indeks ima svoju formulu
- AQI dobija vrednost maksimalne vrednost nekog od sub indeksa

```
datum_merenja      datetime64[ns]
godina_merenja      int64
mesec_merenja       object
dan_merenja         int64
max_dnevna_temperatura  int64
min_dnevna_temperatura  int64
oblacnost           int64
padavine_mm         float64
vidljivost          int64
pravac_vetra        int64
brzina_vetra_kmh    int64
PM25                float64
PM10                float64
O3                  float64
NO2                 float64
SO2                 float64
CO                  float64
PM25_sub_index      float64
PM10_sub_index      float64
O3_sub_index        float64
NO2_sub_index       float64
SO2_sub_index       float64
CO_sub_index        float64
dtype: object
```

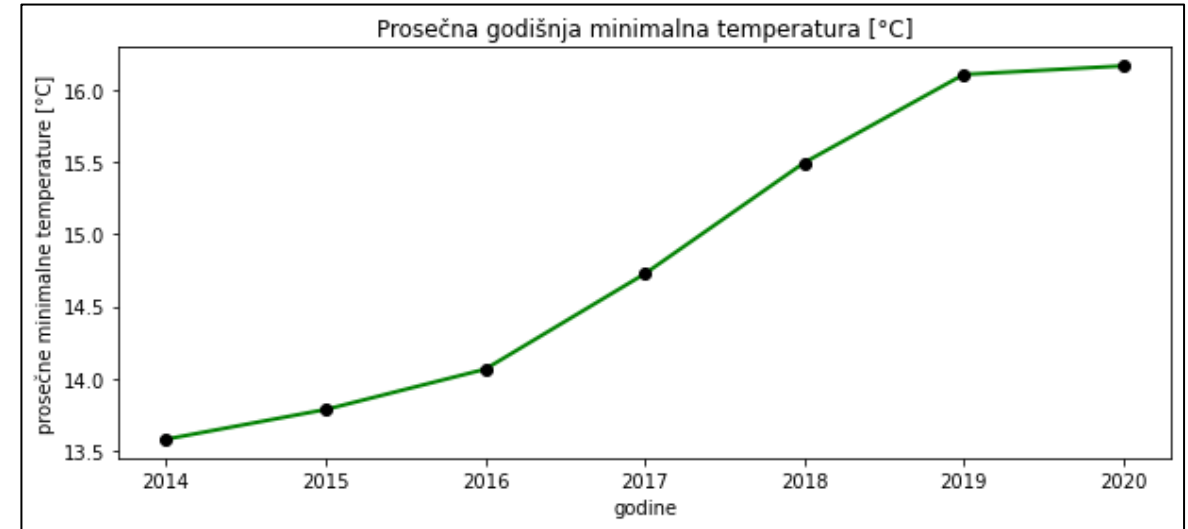


```
datum_merenja      datetime64[ns]
godina_merenja      int64
mesec_merenja       object
dan_merenja         int64
max_dnevna_temperatura  int64
min_dnevna_temperatura  int64
oblacnost           int64
padavine_mm         float64
vidljivost          int64
pravac_vetra        int64
brzina_vetra_kmh    int64
PM25                float64
PM10                float64
O3                  float64
NO2                 float64
SO2                 float64
CO                  float64
PM25_sub_index      float64
PM10_sub_index      float64
O3_sub_index        float64
NO2_sub_index       float64
SO2_sub_index       float64
CO_sub_index        float64
AQI                 float64
AQI_objasnjenje     object
dtype: object
```

Analize 1 i 2 – prosečne godišnje maksimalne i minimalne temperature °C

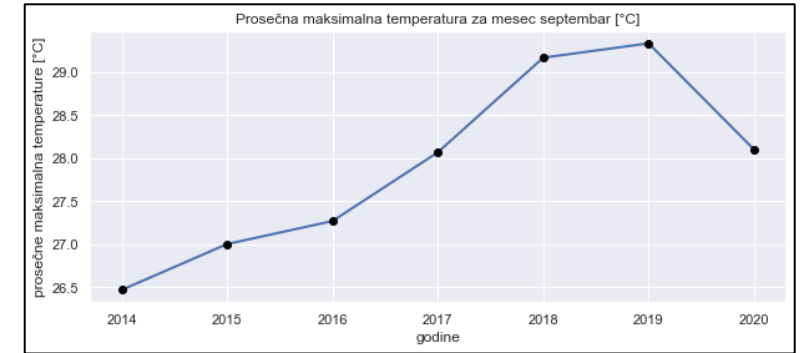
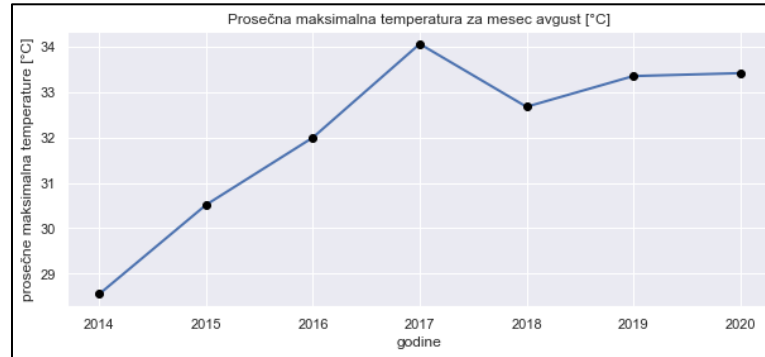
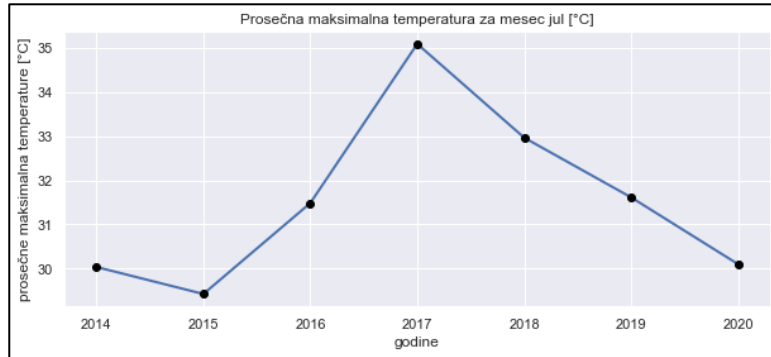


Godina	Prosečna max. temperatura °C	Povećanje u odnosu na 2014. °C
2014.	20.09	
2015.	20.57	+0.48
2016.	20.78	+0.69
2017.	21.48	+1.39
2018.	21.82	+1.73
2019.	22.31	+2.22
2020.	21.82	+1.73



Godina	Prosečna min. temperatura °C	Povećanje u odnosu na 2014. °C
2014.	13.58	
2015.	13.79	+0.21
2016.	14.07	+0.49
2017.	14.73	+1.15
2018.	15.50	+1.92
2019.	16.11	+2.53
2020.	16.17	+2.59

Analize 3, 4, 5 i 6

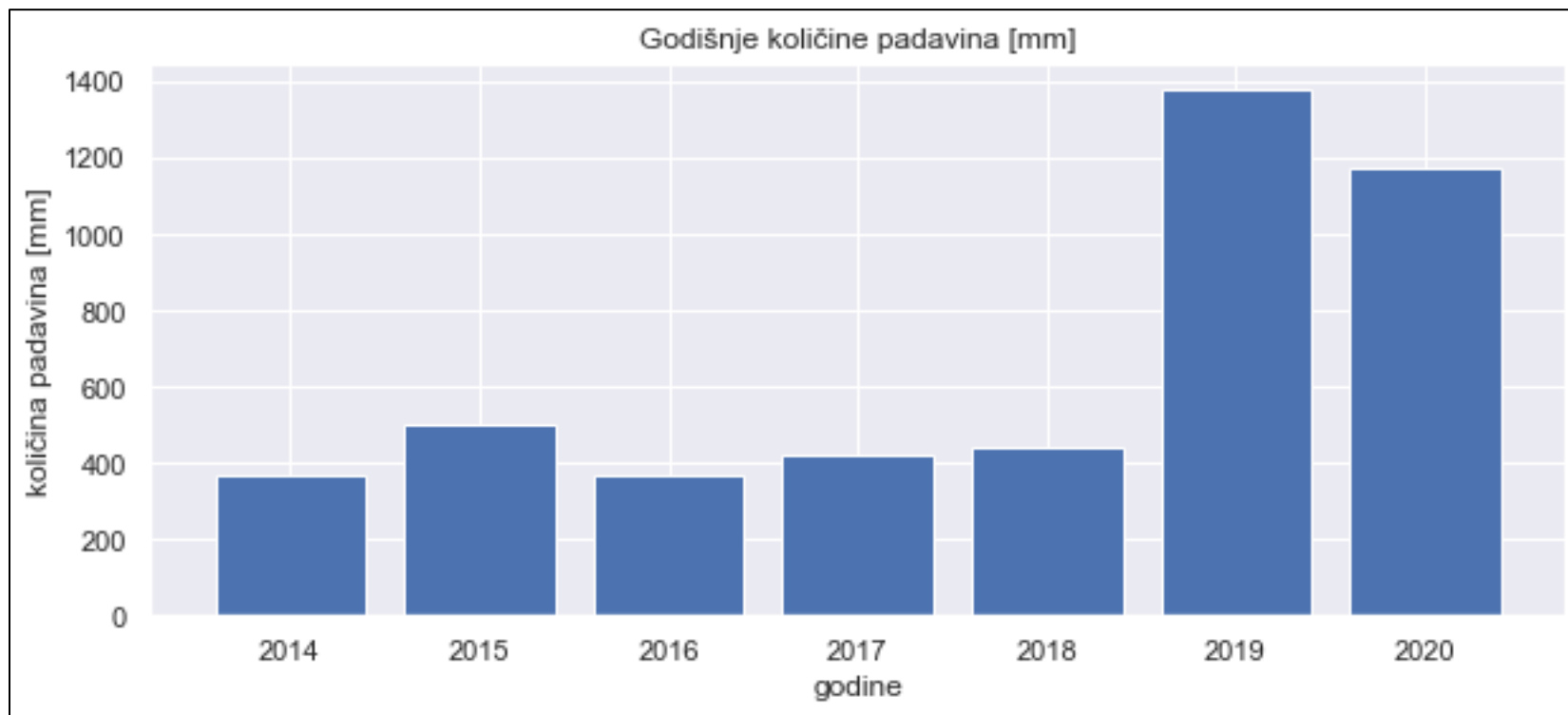


Prikaz rezultata na slike_analiza_5_i_6

Analize 7 i 8 - maksimalne temperature avgust i septembar

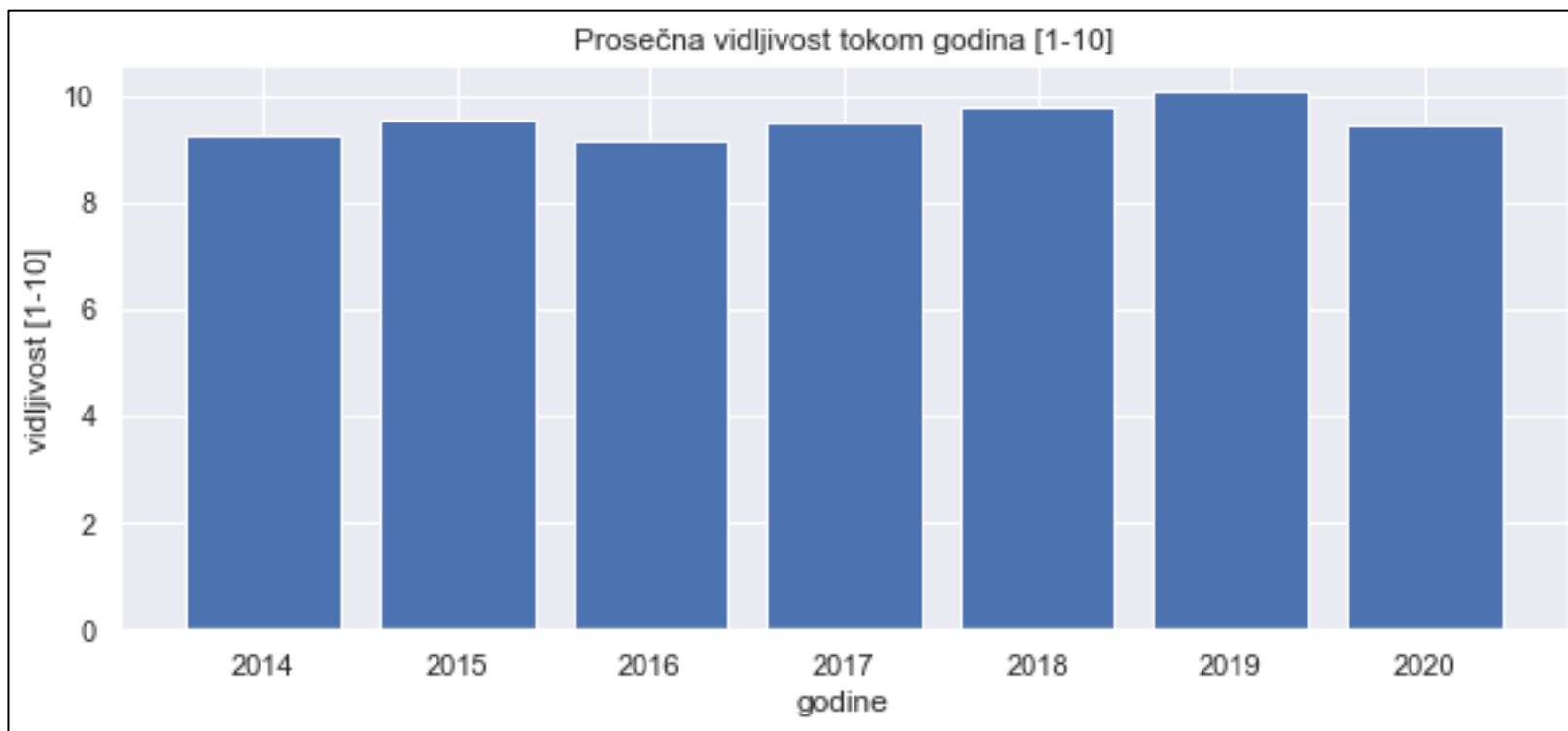
Prikaz rezultata na slike_analiza_7_i_8

Analiza 9 – ukupne godišnje količine padavina



Godina	Padavine mm
2014.	365.1
2015.	498.6
2016.	365.6
2017.	420.2
2018.	443.7
2019.	1379.3
2020.	1175.8

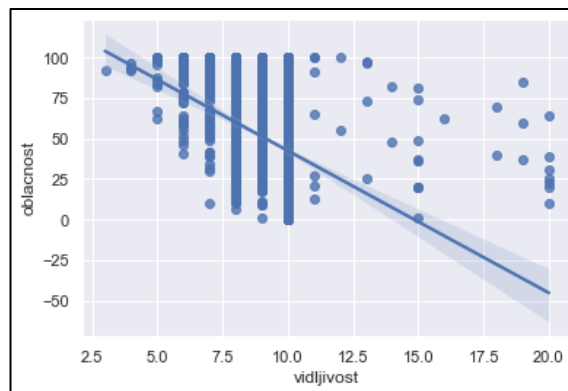
Analiza 10 – prosečna vidljivost tokom godina



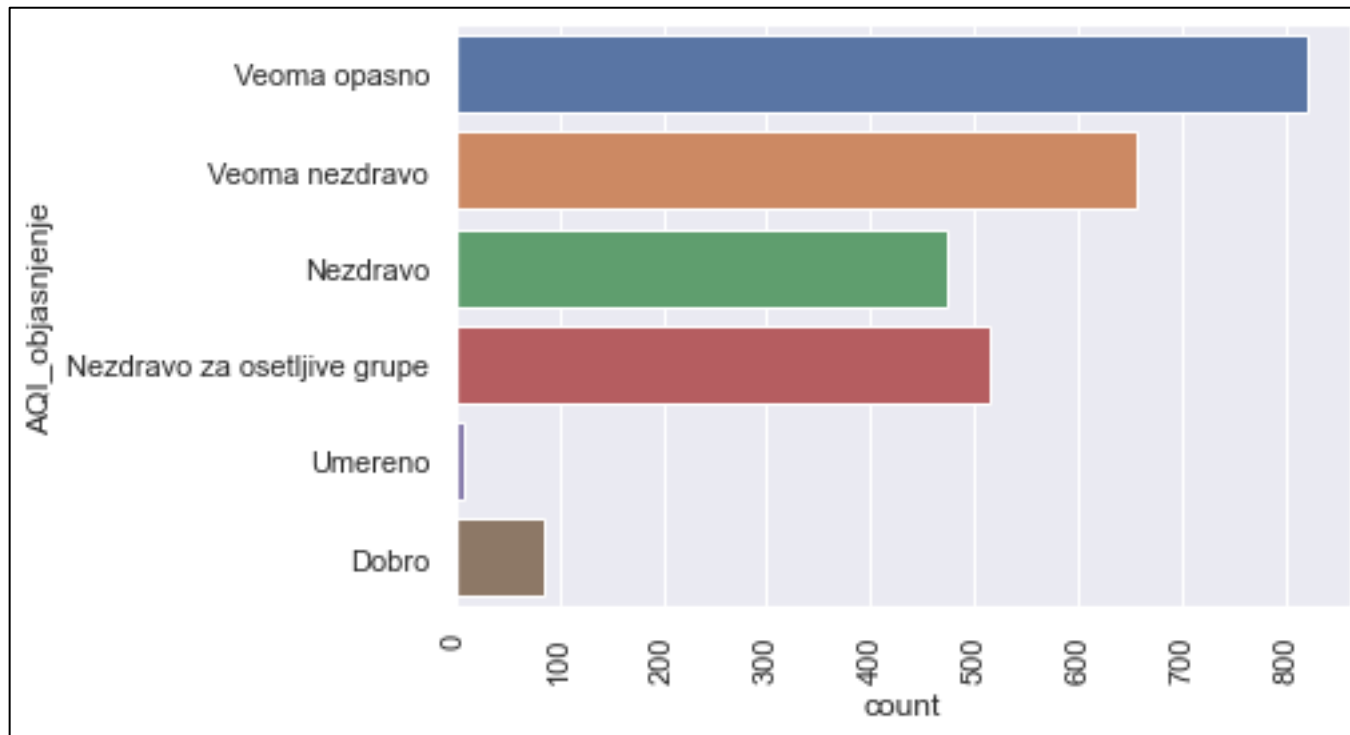
Analiza 11 - svi dani kada je vidljivost <= 5

datum_merenja	godina_merenja	mesec_merenja	dan_merenja	max_dnevna_temperatura	min_dnevna_temperatura	oblačnost	padavine_mm	vidljivost	pravac_vetra	brzina_vetra_kmh	PM25	PM10	O3	NO2	SO2	CO	PM25_sub_index	PM10_sub_index	O3_sub_index	NO2_sub_index	SO2_sub_index	CO_sub_index	AQI	AQI_objasnenje	
6/26/2014	2014	jun	26	25	2	100	0.5	5	73	8	56	27	29	21	2	6	93.33333333	27	29	26.25	2.5	150	10	Nezdravo za osetljive grupe	
7/15/2014	2014	jul	15	24	2	94	18.6	4	107	6	85	59	44	32	5	7	183.33333333	59	44	40	6.25	162.5	13	Nezdravo	
8/18/2014	2014	avgust	18	26	2	93	15.2	4	228	8	87	21	29	22	2	6	190	21	29	27.5	2.5	150	10	Nezdravo	
8/19/2014	2014	avgust	19	22	2	92	27	3	30	9	42	22	27	25	3	6	70	22	27	31.25	3.75	150	10	Nezdravo za osetljive grupe	
9/19/2014	2014	septembar	19	19	1	82	7	5	298	10	95	51	40	25	8	8	216.66666667	51	40	31.25	10	175	27	Veoma nezdravo	
1/28/2016	2016	januar	28	11	0	99	0	5	281	9							0	0	0	0	0	0	0	0	Dobro
7/3/2016	2016	jul	3	27	2	88	0.1	5	163	9	103	54	71	16	4	6	243.33333333	54	71	20	5	150	23	Veoma nezdravo	
9/7/2016	2016	septembar	7	22	1	86	19.6	5	101	9	109	62	84	33	5	9	263.33333333	62	84	41.25	6.25	187.5	23	Veoma nezdravo	
9/26/2016	2016	septembar	26	29	2	62	0.5	5	84	10	62	31	13	12	4	3	106.66666667	31	13	15	5	112.5	12	Nezdravo za osetljive grupe	
10/22/2016	2016	oktobar	22	23	2	100	15.8	5	100	20	34	30	39	10	4	4	56.66666667	30	39	12.5	5	125	15	Nezdravo za osetljive grupe	
4/7/2017	2017	april	7	15	1	92	4	4	46	10	129	61	37	31	5	7	306.9230769	61	37	38.75	6.25	162.5	37	Veoma opasno	
6/13/2017	2017	jun	13	20	1	97	21.7	4	101	10	59	30	45	15	2	4	98.33333333	30	45	18.75	2.5	125	15	Nezdravo za osetljive grupe	
6/23/2017	2017	jun	23	27	2	96	19.4	5	99	5	103	36	31	22	2	6	243.33333333	36	31	27.5	2.5	150	23	Veoma nezdravo	
6/24/2017	2017	jun	24	25	2	83	5.6	5	187	4	70	47	20	27	3	8	133.33333333	47	20	33.75	3.75	175	15	Nezdravo	
12/22/2019	2019	decembar	22	12	1	98	20.7	5	181	6	93	18	27	21	3	8	210	18	27	26.25	3.75	175	20	Veoma nezdravo	
1/7/2020	2020	januar	7	21	1	67	1	5	207	21	78	33	11	29	3	11	160	33	11	36.25	3.75	214.2857143	24	Veoma nezdravo	
2/7/2020	2020	februar	7	8	1	100	10.7	5	61	13	57	21	32	5	3	4	95	21	32	6.25	3.75	125	15	Nezdravo za osetljive grupe	
9/18/2020	2020	septembar	18	21	1	98	18.6	5	139	11	78	63	29	34	2	11	160	63	29	42.5	2.5	214.2857143	24	Veoma nezdravo	

Korelacija sa oblačnošću = -0.37 => slaba negativna linearna korelacija



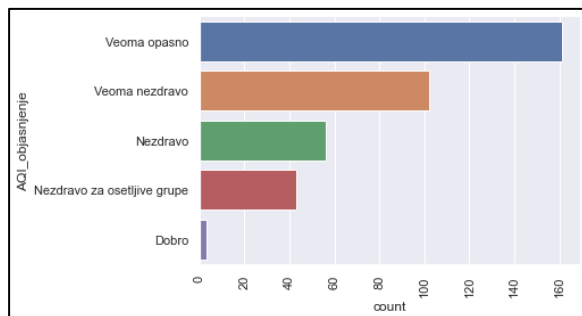
Analiza 12 – AQI objašnjenje za sve godine



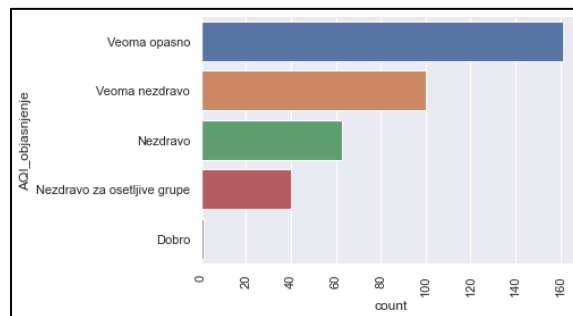
AQI objašnjenje	Broj dana tog tipa
Veoma opasno	819
Veoma nezdravo	657
Nezdravo za osetljive grupe	514
Nezdravo	473
Umereno	9
Dobro	85

Analiza 13 – AQI objašnjenje po godinama

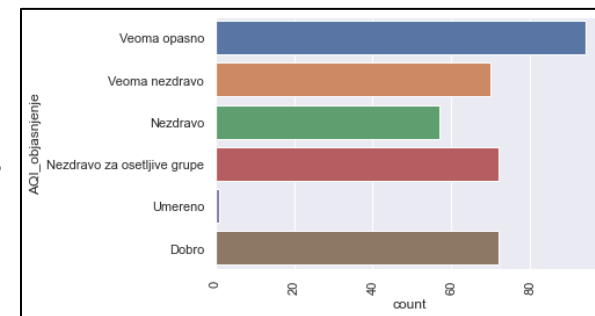
2014.



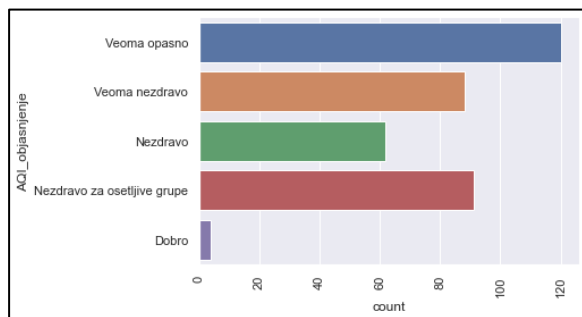
2015.



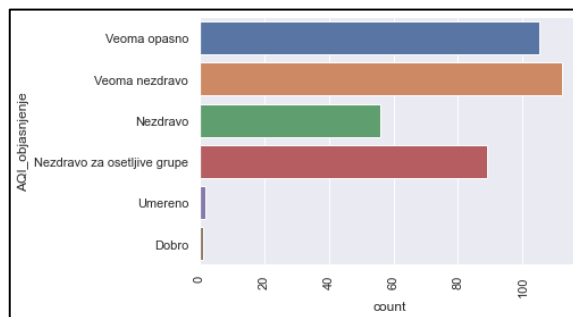
2016.



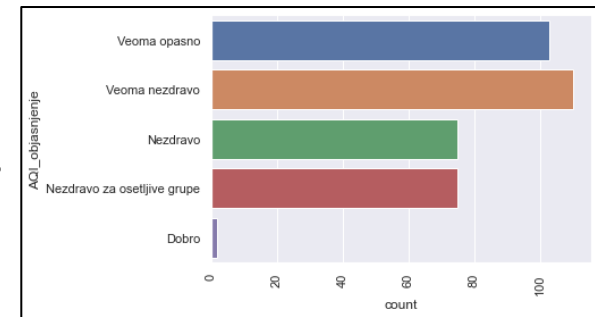
2017.



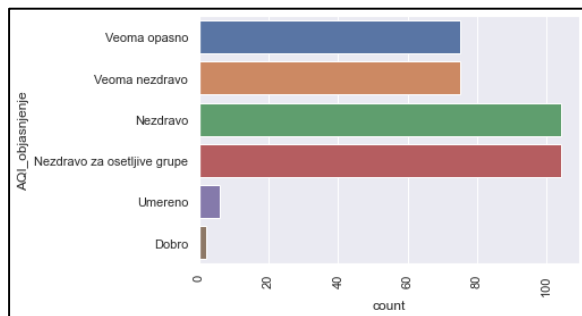
2018.



2019.



2020.

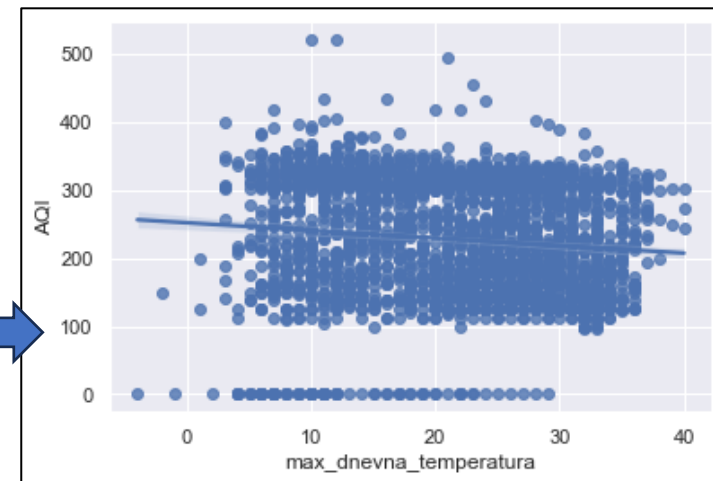
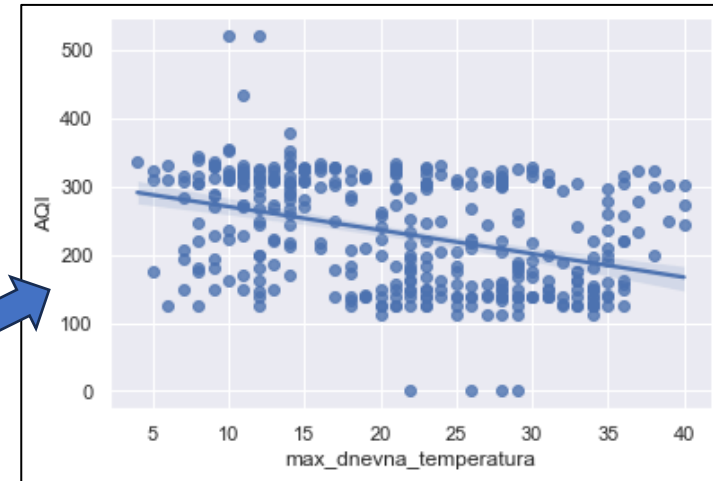


Analiza 14 – korelacija

- Provera korelacije između vremenskih parametara i AQI
- Provereni su sledeći parametri:
 1. Maksimalna dnevna temperatura °C
 2. Minimalna dnevna temperatura °C
 3. Padavine mm
 4. Vidljivost
 5. Brzina vetra km/h
 6. Oblačnost
 7. Pravac vetra

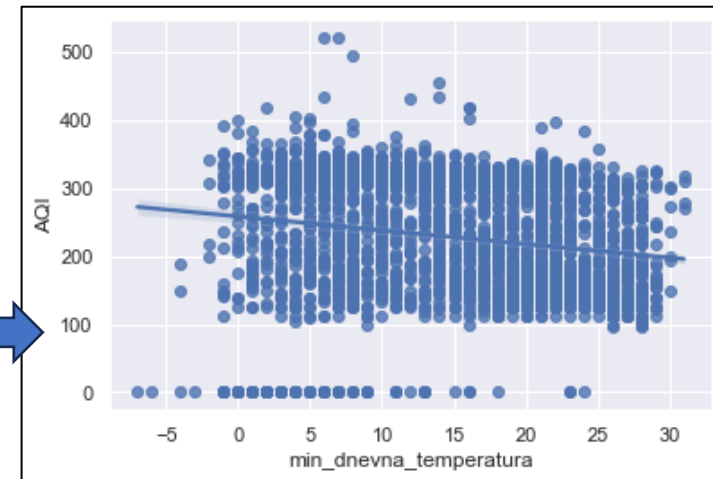
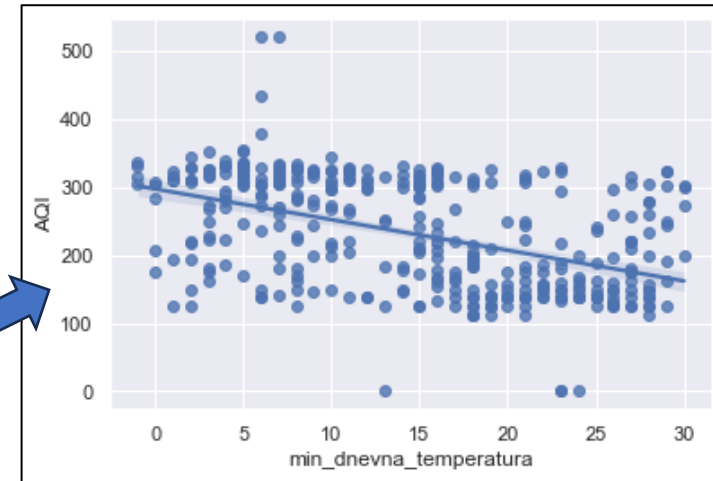
Analiza 14 – maksimalna dnevna temperatura °C

Godina	Koeficijent korelacije r
2014.	-0.2
2015.	-0.26
2016.	0.34
2017.	-0.38
2018.	-0.2
2019.	-0.14
2020.	-0.12
sve godine	-0.11



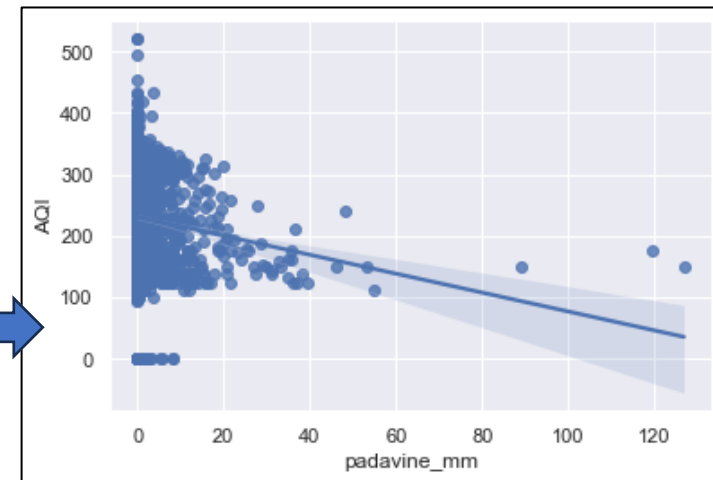
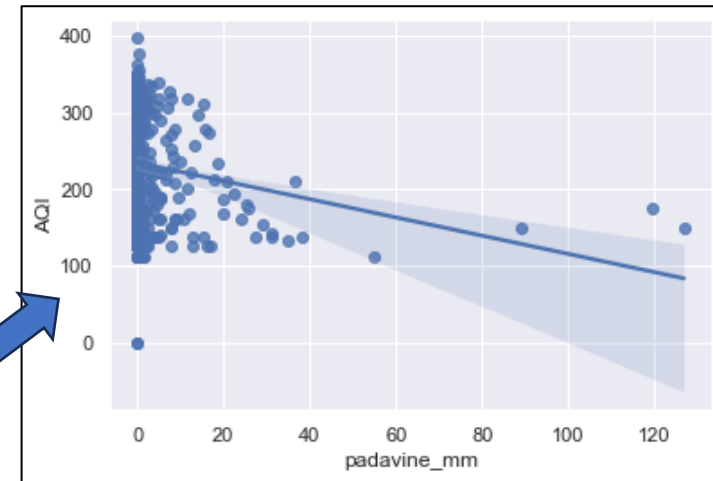
Analiza 14 – minimalna dnevna temperatura °C

Godina	Koeficijent korelacije r
2014.	-0.32
2015.	-0.36
2016.	0.32
2017.	-0.47
2018.	-0.31
2019.	-0.21
2020.	-0.19
sve godine	-0.19



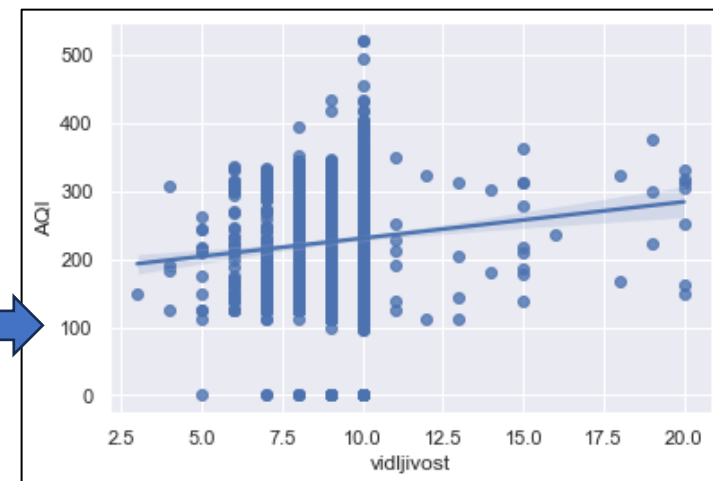
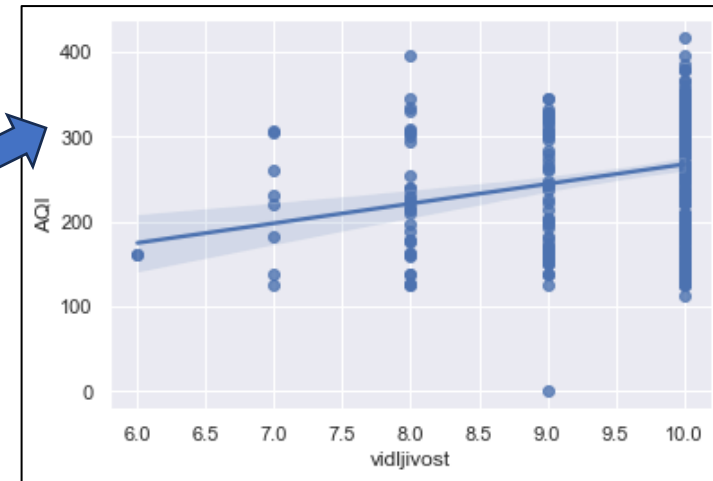
Analiza 14 – padavine mm

Godina	Koeficijent korelacije r
2014.	-0.16
2015.	-0.19
2016.	-0.01
2017.	-0.18
2018.	-0.11
2019.	-0.19
2020.	-0.06
sve godine	-0.11



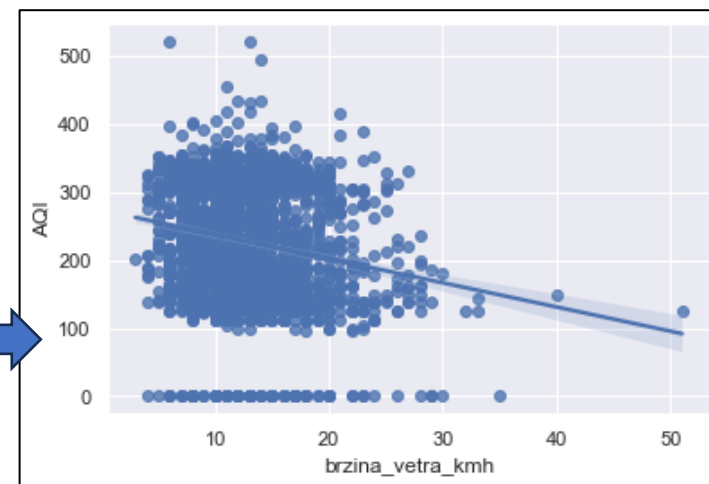
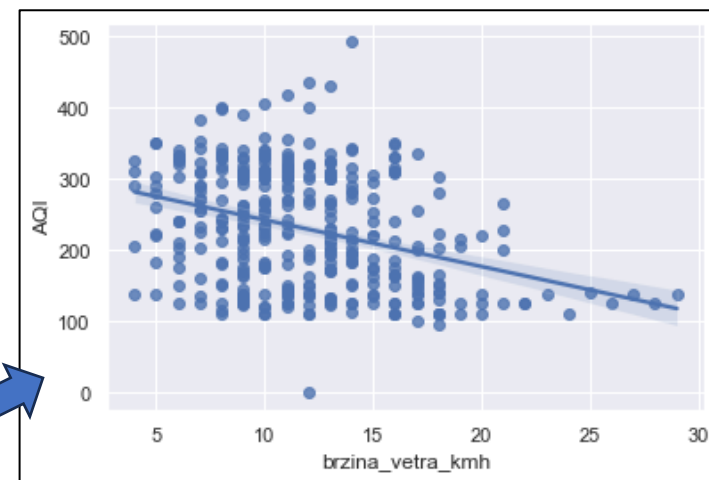
Analiza 14 – vidljivost 1-10

Godina	Koeficijent korelacije r
2014.	0.22
2015.	0.25
2016.	-0.04
2017.	0.08
2018.	-0.02
2019.	0.11
2020.	-0.05
sve godine	0.08



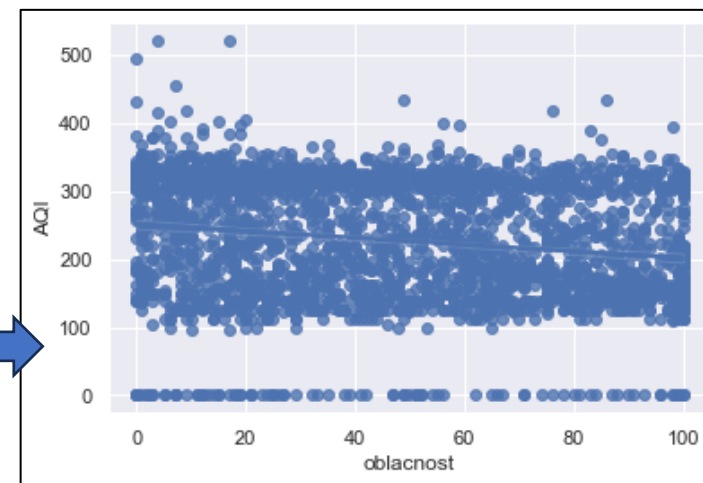
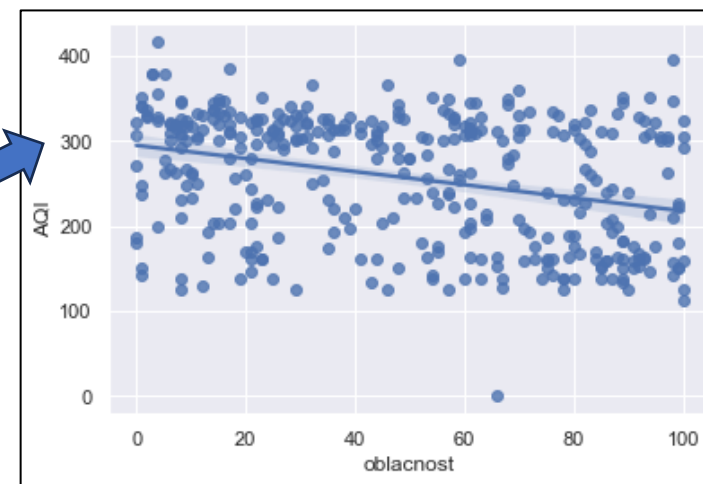
Analiza 14 – brzina vetra kh/m

Godina	Koeficijent korelacije r
2014.	-0.05
2015.	0.21
2016.	-0.25
2017.	-0.17
2018.	-0.34
2019.	-0.30
2020.	-0.16
sve godine	-0.19



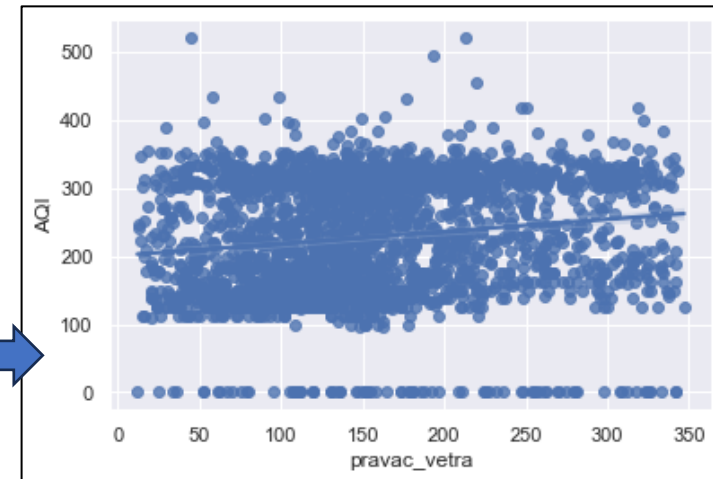
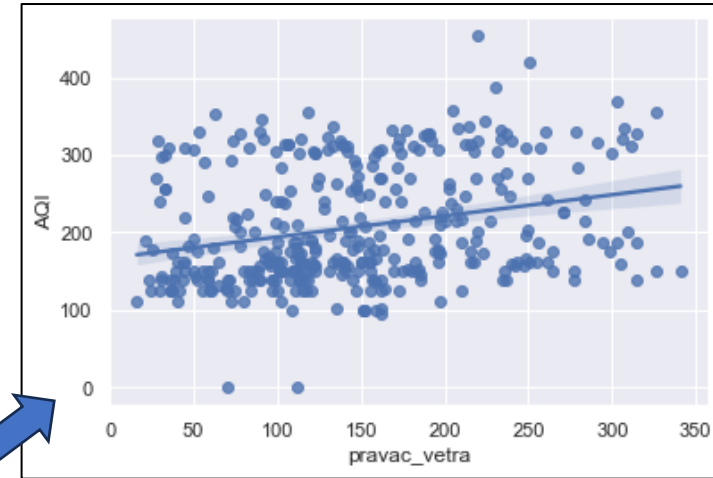
Analiza 14 – oblačnost

Godina	Koeficijent korelacije r
2014.	-0.26
2015.	-0.32
2016.	0.02
2017.	-0.27
2018.	-0.20
2019.	-0.21
2020.	-0.05
sve godine	-0.17



Analiza 14 – pravac vetra

Godina	Koeficijent korelacije r
2014.	0.21
2015.	0.27
2016.	-0.08
2017.	0.24
2018.	0.19
2019.	0.14
2020.	0.27
sve godine	0.16



Deskriptivna statistika

	max_dnevna_temperatura	min_dnevna_temperatura	oblacnost	padavine_mm	\
count	2557.00	2557.00	2557.00	2557.00	
mean	21.26	14.85	46.60	1.82	
std	8.58	8.13	30.68	6.02	
min	-4.00	-7.00	0.00	0.00	
25%	14.00	8.00	19.00	0.00	
50%	22.00	15.00	44.00	0.00	
75%	28.00	22.00	73.00	1.10	
max	40.00	31.00	100.00	126.80	

	vidljivost	pravac_vetra	brzina_vetra_kmh	PM25	PM10	O3	\
count	2557.00	2557.00	2557.00	2472.00	2472.00	2472.00	
mean	9.54	154.63	12.71	105.76	47.49	42.06	
std	1.29	76.34	4.58	41.10	20.83	23.65	
min	3.00	12.00	3.00	-1.00	-1.00	-1.00	
25%	9.00	99.00	10.00	75.00	32.00	28.00	
50%	10.00	144.00	12.00	100.00	44.00	38.00	
75%	10.00	202.00	15.00	133.00	59.00	52.00	
max	20.00	347.00	51.00	409.00	164.00	143.00	

	N02	S02	CO	AQI
count	2472.00	2472.00	2472.00	2557.00
mean	19.88	4.79	5.98	228.15
std	9.62	3.78	2.34	87.24
min	-1.00	-1.00	-1.00	0.00
25%	13.00	2.00	4.00	157.00
50%	18.00	4.00	6.00	227.00
75%	25.00	6.00	7.00	309.00
max	69.00	37.00	19.00	522.00

Zaključci

- Povećava se prosečna maksimalna i minimalna temperatura kroz period (i meseci i godina)
- Najveće temperature se pomeraju sa jula na avgust
- Takođe i septembar ima značajno povećanje temperature
- Postoji slaba negativna korelacija između vidljivosti i oblačnosti
- Prikazani su AQI indeksi ukupno i kroz godine
- Pokazano je da ne postoji značajne korelacije između parametara vremenskih prilika i AQI indeksa

Šta bi moglo onda da bude faktor zagađenja vazduha

- Po podacima sa interneta to su:
 1. sagorevanje ulja
 2. motorna vozila
 3. industrijska prašina
 4. hemijski otpadi koji odlaze u atmosferu



Šta bi moglo onda da bude faktor zagađenja vazduha

- Svi ovi faktori su povezani sa brzim društveno-ekonomskim razvojem
- U Šangaju je 2020. godine zabeleženo oko 3 475 000 vozila, dok je u 2014. godini taj broj bio oko 1 830 000
- Šangaj ima najveći broj fabrika u Kini



Šta se radi na tome da se ovo poboljša

- Šangaj postaje zeleniji grad
- Do 2035. godine će biti završeno 2000 parkova
- Smanjenje emisije ugljenika
- Donošenja novih zakona o ispuštanju štetnih materija



HVALA NA PAŽNJI

