

# LV1: Obdelava in vizualizacija podatkov

## 1 Uvod

---

Namen vaje:

- Obnovitev sintaksa Python, Jupyter notebooks
- Knjižnica pandas za uvoz in vizualizacijo podatkov

### 1.1 Instalacija Python, Jupyter in Visual Studio Code

---

Python: na šolskih računalnikih je nameščen Python 3.8

Če nimate Python:

- <https://www.python.org/downloads/>
- Naložite instaler in namestite Python na C disk npr. v direktorij C:\Python312

Visual Studio Code: namestite zadnjo verzijo (Help , Install Update)

Namestite vtičnike za Visual Studio Code : zavihek Extensions

- V iskanje vpišite: Python Extension Pack, in namestite paket vtičnikov avtorja Leo Jhon Song

## 2 Pandas knjižnjica

---

Pandas je open-source knjižnjica zgrajena na NumPy, in namenjena hitri analizi podatkov, pa tudi vizualizaciji. Delamo lahko s podatkih iz različnih virov.

DataFrames je temeljni objekt za zapis podatkov. Lahko si ga predstavljamo kot tabelo v excelu.

### 2.1 Delo s podatki - DataFrames

---

Pandas dataframe je podatkovna struktura z dvema dimenzijama, vsebuje stolpce in vrstice. Nad dataframi lahko izvajamo razne operacije kot so seštevanje, deljenje, množenje, ...

### 2.2 Izbira podatkov, dodajanje novih, odstranjevanje

---

#Podatke zberemo (dataframe/serija):

```
df['W']/df.loc['A']
```

#Nove stolpce dodamo v tabelo

```
df['new'] = df['W'] + df['Y']
```

Uporabniku prilagojene komunikacije

```
#Nove vrstico dodamo v tabelo
```

```
df.append([1,2,3,4,5], ignore_index=True)
```

```
#Odstranimo stolpec
```

```
df.drop('new', axis=1, inplace=True)
```

```
#Odstranimo vrstico
```

```
df.drop('E',axis=0)
```

## 2.3 Uvoz podatkov

---

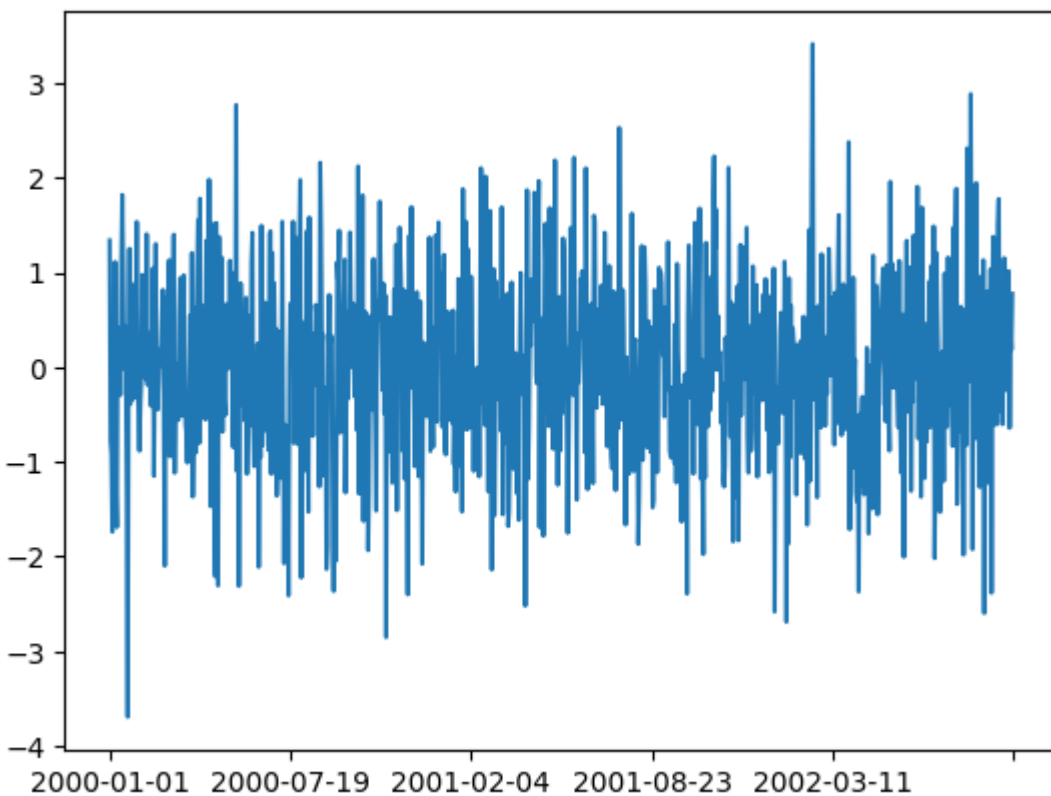
```
#Podatke preberemo in shranimo v
```

```
df1 = pd.read_csv('df1',index_col=0)
```

## 2.4 Vizualizacija podatkov vgrajena v Pandas

---

Z pandas lahko vizualiziramo podatke z različnimi grafi (primer linechart).



### 3 Samostojna naloga: Vizualizacija covid podatkov

Naloga:

Uvozi podatke iz covid dataseta (Excel) v dataframe. Preveri obseg podatkov in stolpce (tipe podatkov).  
Preveri, katere države imamo (katere iso\_code).

Iz covid-data.xls preberi podatke v spremenljivko cdata, in si oglej podatke

Prebermo z

```
cdata = pd.read_excel('covid-data.xlsx')
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	female_smokers	male_smokers	hand
0	ALB	Europe	Albania	2020-02-25	NaN	NaN	NaN	NaN	NaN	NaN	...	7.1	51.2	
1	ALB	Europe	Albania	2020-02-26	NaN	NaN	NaN	NaN	NaN	NaN	...	7.1	51.2	
2	ALB	Europe	Albania	2020-02-27	NaN	NaN	NaN	NaN	NaN	NaN	...	7.1	51.2	
3	ALB	Europe	Albania	2020-02-28	NaN	NaN	NaN	NaN	NaN	NaN	...	7.1	51.2	
4	ALB	Europe	Albania	2020-02-29	NaN	NaN	NaN	NaN	NaN	NaN	...	7.1	51.2	

5 rows x 67 columns

Izpiši vse kode držav, ki so prisotne v setu

```
cdata['iso_code'].unique()
```

```
['ALB' 'AND' 'AUT' 'BLR' 'BEL' 'BIH' 'BGR' 'HRV' 'CYP' 'CZE' 'DNK' 'EST'  
'FRO' 'FIN' 'FRA' 'DEU' 'GIB' 'GRC' 'GGY' 'HUN' 'ISL' 'IRL' 'IMN' 'ITA'  
'JEY' 'OWID_KOS' 'LVA' 'LIE' 'LTU' 'LUX' 'MLT' 'MDA' 'MCO' 'MNE' 'NLD'  
'MKD' 'NOR' 'POL' 'PRT' 'ROU' 'RUS' 'SMR' 'SRB' 'SVK' 'SVN' 'ESP' 'SWE'  
'CHE' 'UKR' 'GBR' 'VAT']
```

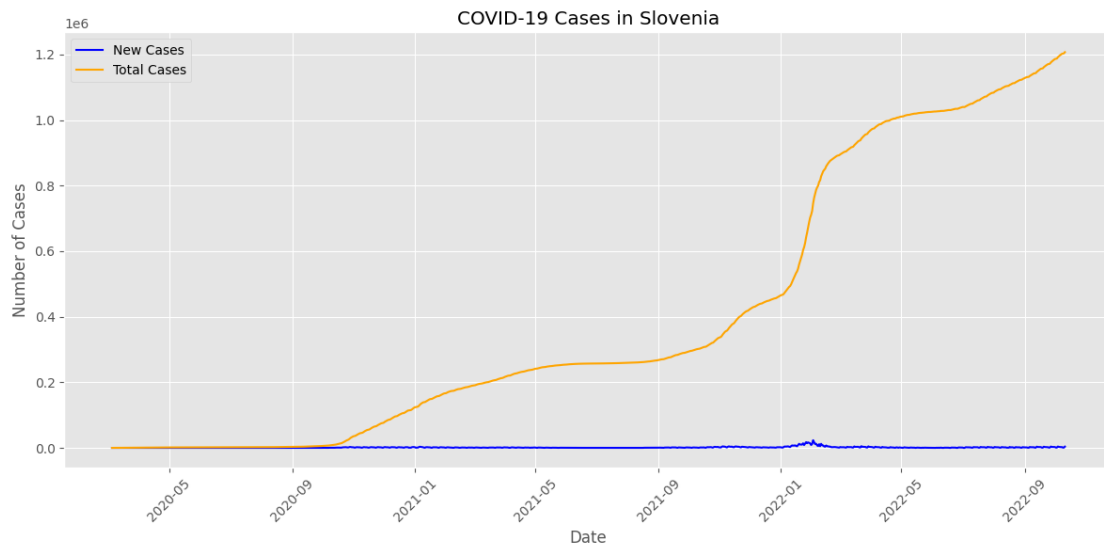
Izloči podatke (vrstice) pripadajoče izbrani državi, npr. koda SLO, in zapiši v novo spremenljivko (npr. slodata)

```
slodata = cdata[cdata['iso_code'] == 'SVN']
```

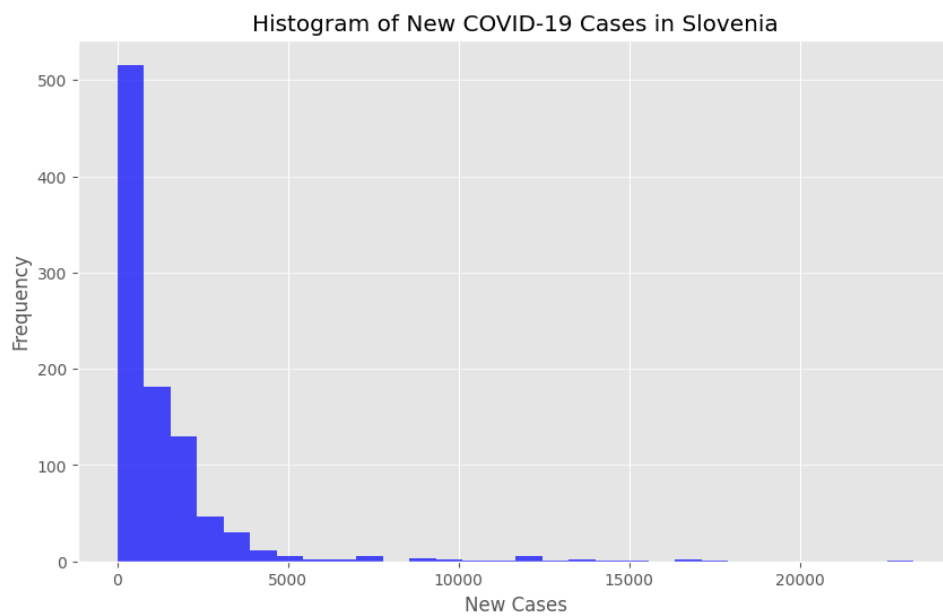
Izriši časovni graf izbranega stolpca podatkov (ali več stolpcev)

Graf starih in novih primerov glede na čas

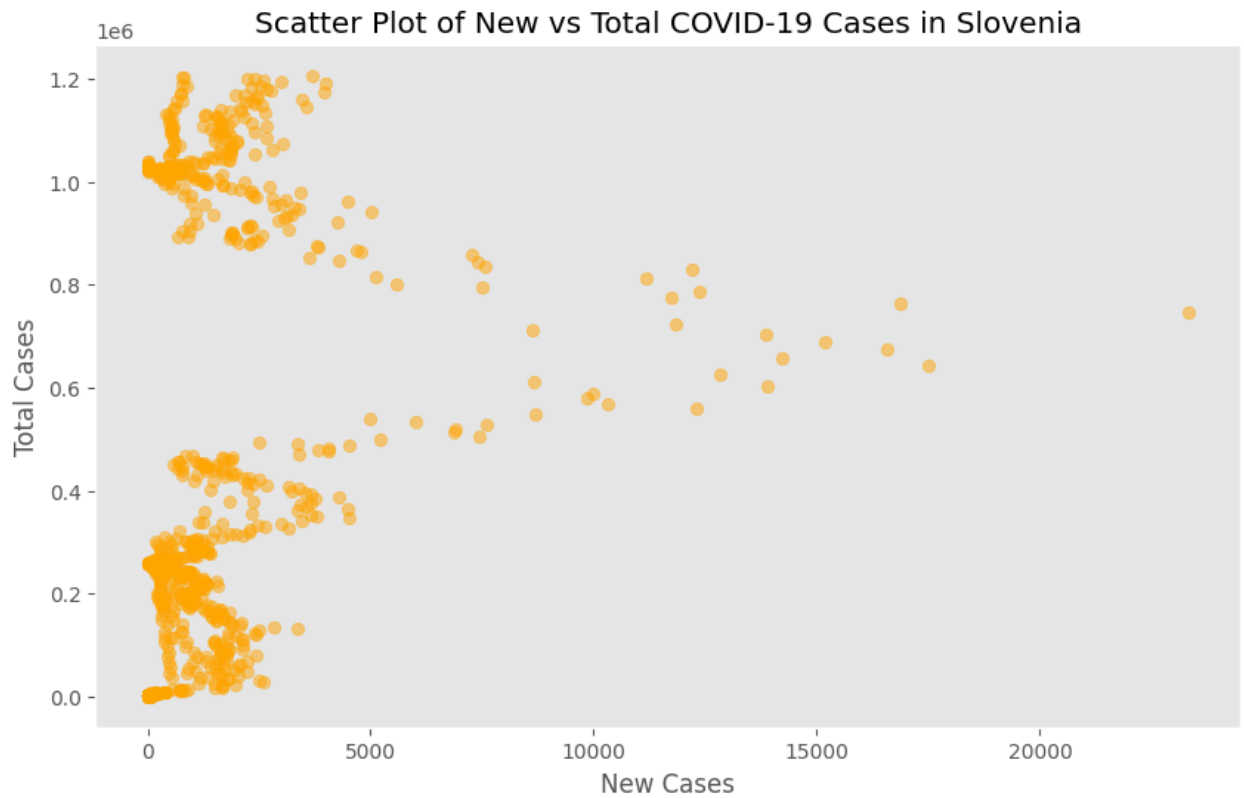
## Uporabniku prilagojene komunikacije



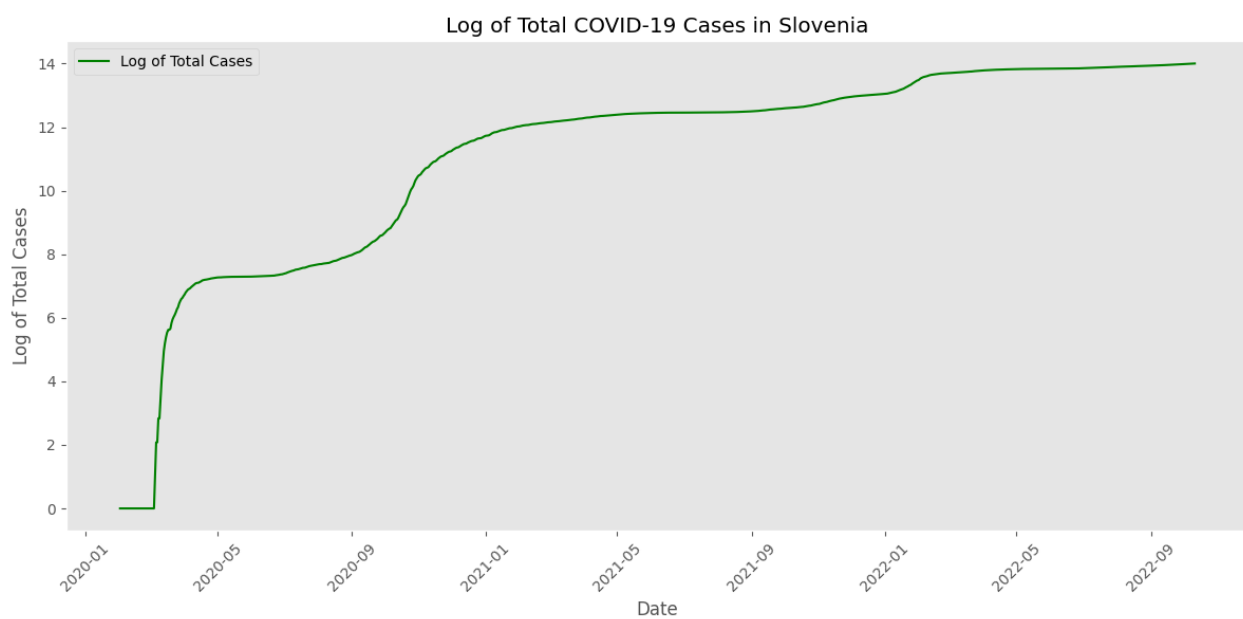
1. Izriši histogram vrednosti ene ali več izbrane spremenljivke (stolpca)  
Histogram pojavitve novih primerov



2. Izriši scatter diagram izbranih dveh ali treh (barva) spremenljivk



3. Dodatne ideje: izračunaj kakšen nov parameter, in ga vizualiziraj (npr. logaritem kumulativnega števila primerov Covid, ali kakšno razmerje, delež, primerjave med državami.. )



Uporabniku prilagojene komunikacije