

PANDAS

Csv: `pandas.read_csv`

Separator: `sep=';`

Excel: `pandas.read_excel`

Tidy data:

1. Wyświetlanie kilku pierwszych wierszy danych:
 - `df.head()`
2. Sprawdzanie wymiarów ramki danych:
 - `df.shape`
3. Sprawdzanie typów danych w ramce danych:
 - `df.dtypes`
4. Filtracja danych na podstawie warunków:
 - `df_filtered = df[df['kolumna'] > warunek]`
5. Sortowanie danych względem określonej kolumny:
 - `df_sorted = df.sort_values('kolumna')`
6. Grupowanie danych na podstawie określonej kolumny:
 - `df_grouped = df.groupby('kolumna').mean()`
7. Transponowanie danych:
 - `df_transposed = df.transpose()`
8. Usuwanie duplikatów:
 - `df_duplicates_removed = df.drop_duplicates()`
9. Uzupełnianie brakujących wartości:
 - `df_filled = df.fillna(value)`
10. Zapisywanie danych do pliku:
 - `df.to_csv('nazwa_pliku.csv', index=False)`

WYKRESY:

Wykres liniowy:

`plt.plot()`

Wykres punktowy:

`plt.scatter()`

Wykres słupkowy:

`plt.bar()`

Wykres słupkowy skumulowany:

`plt.barh()`

Wykres kołowy:

`plt.pie()`

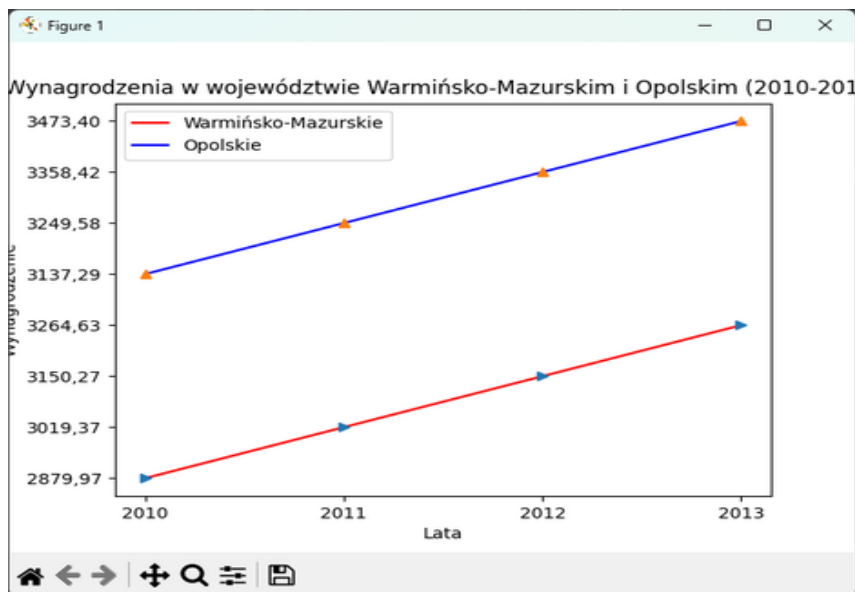
Wykres histogramu:

`plt.hist()`

```

import matplotlib.pyplot as plt
import pandas as pd
data = pd.read_csv("wynagrodzenia21.csv", sep=';')
data_sorted = data.sort_values('2010')
cleaned = data.dropna()
warminsko_mazurskie = data.loc[data['Województwo'] == 'WARMIŃSKO-MAZURSKIE',
'2010':'2013'].values.flatten()
opolskie = data.loc[data['Województwo'] == 'OPOLSKIE', '2010':'2013'].values.flatten()
lata = ['2010', '2011', '2012', '2013']
plt.plot(lata, warminsko_mazurskie, 'r', label='Warmińsko-Mazurskie')
plt.plot(lata, warminsko_mazurskie, '>')
plt.plot(lata, opolskie, 'b', label='Opolskie')
plt.plot(lata, opolskie, '^')
plt.title('Wynagrodzenia w województwie Warmińsko-Mazurskim i Opolskim (2010-2013)')
plt.legend()
plt.xlabel('Lata')
plt.ylabel('Wynagrodzenie')
plt.savefig("zad3.webp")
plt.show()

```



```

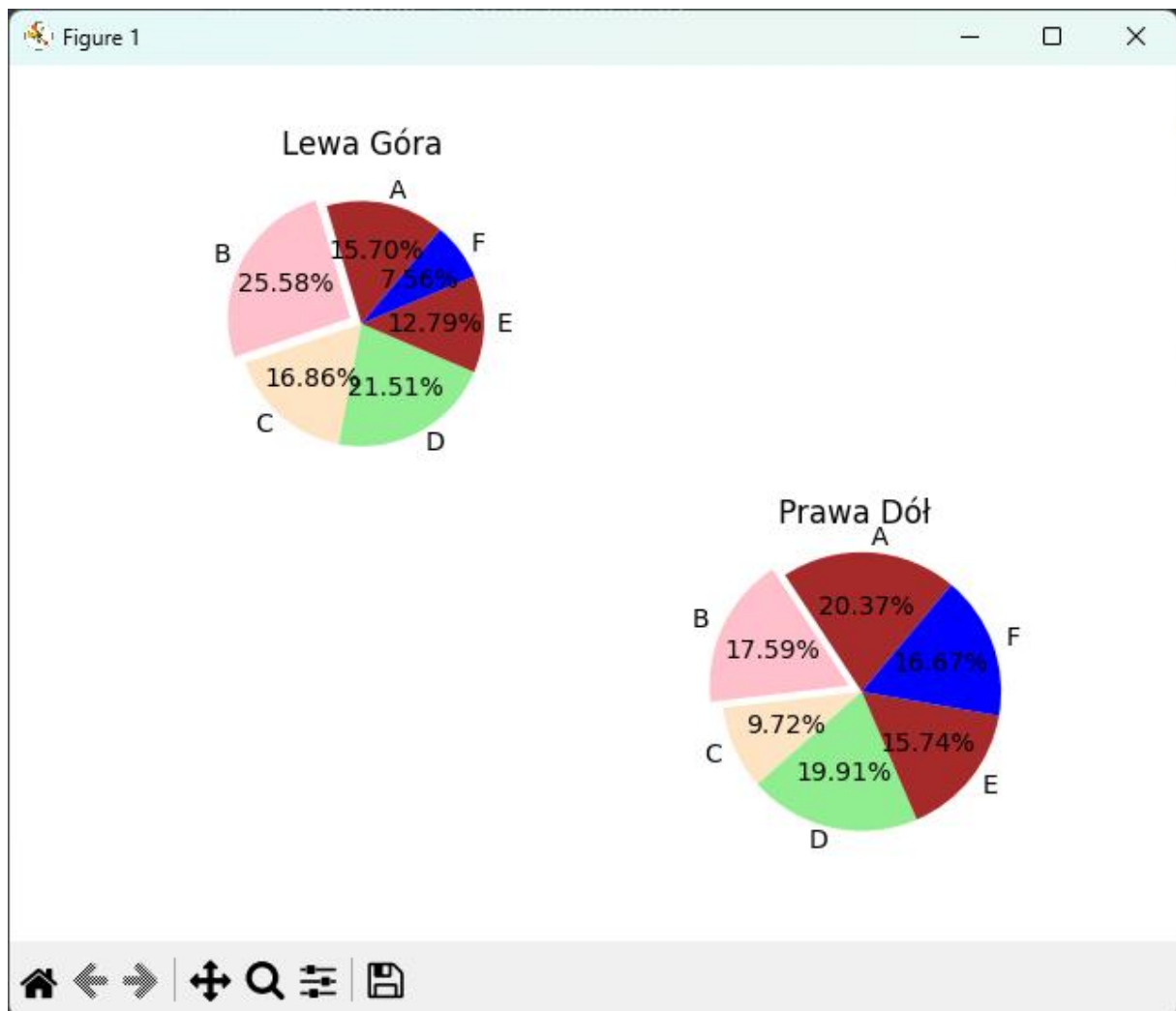
import pandas as pd
import matplotlib.pyplot as plt
dane = pd.read_csv("wynagrodzenia21.csv", sep=';', decimal=',')
wojewodztwa = dane[dane['Województwo'].isin(['PODLASKIE', 'POMORSKIE'])]
roczne_wynagrodzenia = dane.loc[dane['Województwo'].isin(['PODLASKIE',
'POMORSKIE']), '2010':'2013']
plt.figure(figsize=(10, 6))
plt.plot(roczne_wynagrodzenia.T)
plt.legend(wojewodztwa['Województwo'], loc='upper left')
plt.xlabel('Rok')
plt.ylabel('Wynagrodzenie')
plt.title('Średnie roczne wynagrodzenia w poszczególnych województwach (do 2013)')
plt.grid()
plt.show()

```

```

import matplotlib.pyplot as plt
sizes = [15.70,25.58,16.86,21.51,12.79,7.56]
sizes2 = [20.37,17.59,9.72,19.91,15.74,16.67]
labes = ["A","B","C","D","E","F"]
colors = ['brown','pink','bisque','lightgreen','brown','blue']
explode = [0,0.1,0,0,0,0]
plt.subplot(2,2,1)
plt.title("Lewa Góra")
plt.pie(sizes, explode=explode,labels=labes,colors=colors,autopct='% 1.2f%% ',startangle=50)
plt.subplot(2,2,4)
plt.title("Prawa Dół")
plt.pie(sizes2,
explode=explode,labels=labes,colors=colors,autopct='% 1.2f%% ',startangle=50)
plt.axis('equal')
plt.savefig("zad1.jpg")
plt.show()

```



```

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
dane = pd.read_excel("ceny21.xlsx")
maka = dane[dane["Rodzaje towarów"]=="mąka pszenna - za 1kg"]
kasza = dane[dane["Rodzaje towarów"]=="kasza jęczmienna - za 0,5kg"]
x = np.arange(0, len(maka["Wartość"]))
plt.scatter(x, maka["Wartość"], color="red", marker='*', label="mąka pszenna - za 1kg")
plt.scatter(x, kasza["Wartość"], color="blue", marker='^', label="kasza jęczmienna - za 0,5kg")
plt.title("Dane zboża")
plt.xticks(x, maka["Rok"])
plt.xlabel('Rok')
plt.ylabel('Cena w zł')
plt.annotate('Borys', xy=(4.5, 2.2), color='black', fontsize=14)
plt.legend()
plt.savefig("zad2.png")
plt.show()

```

