

# Analiza i bazy danych

## Ćwiczenie 6 - Exploratory Data Analysis

Natalia Książek 406780 Grupa 5 - piątek 11:00

```
In [117... import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

wielkopolskie = pd.read_csv('G:\V_semestr\AiBD_Natalia_Ksiazek\Lab6-Exploratory-Data-Analysis\original-data\14_WIELKOPOLSKIE.csv', encoding_errors='ignore')

In [4]: wielkopolskie.columns

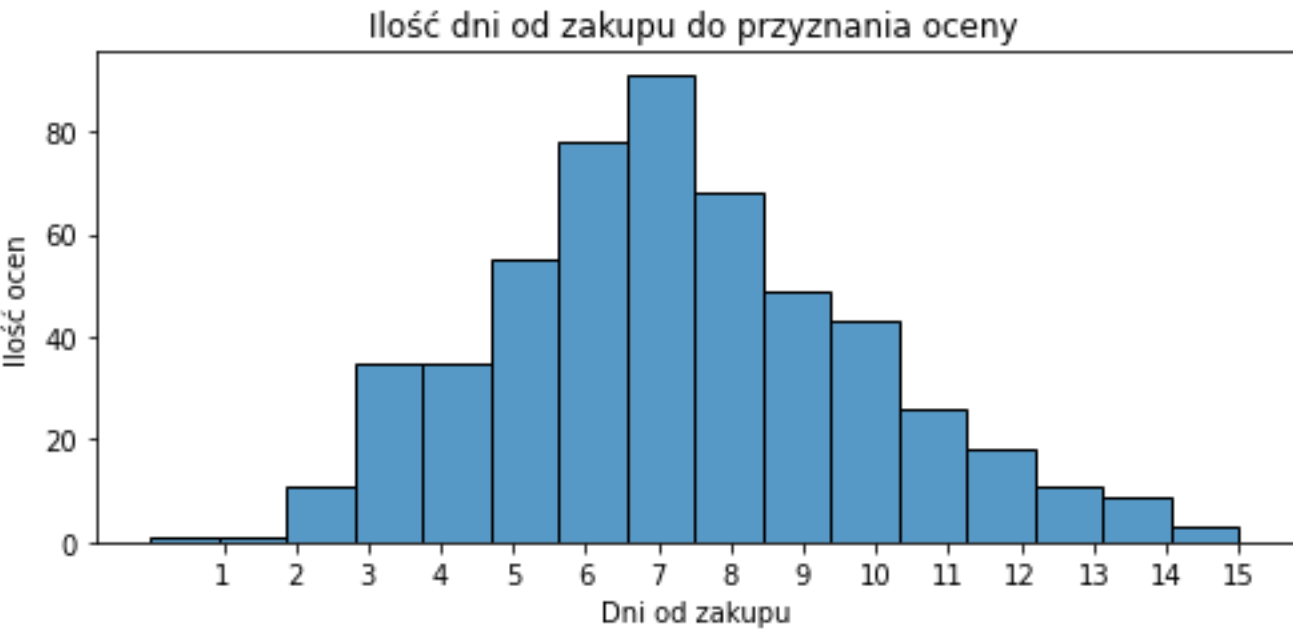
Out[4]: Index(['Unnamed: 0', 'Dni od zakupu', 'Marka', 'Wiek kupującego',
        'Płeć kupującego', 'Ocena'],
        dtype='object')
```

### 1. Ilość dni od zakupu do przyznania oceny

```
In [152... x = wielkopolskie['Dni od zakupu'].value_counts()
x = [x.index[i] for i in range(len(x))]
len(x)

Out[152]: 16

In [154... fig1 = sns.histplot(data=wielkopolskie, x='Dni od zakupu', bins=16)
fig1.set(ylabel='Ilość ocen', title='Ilość dni od zakupu do przyznania oceny')
plt.xticks([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15])
plt.show()
```



Klienci najczęściej wystawiali opinie o odkurzaczu 7-9 dni po zakupie.

### 2. Najczęściej wybierane marki odkurzaczy

```
In [106... fig2 = sns.histplot(data=wielkopolskie, x='Marka')
fig2.set(ylabel='Ilość zakupionych odkurzaczy', title='Najczęściej wybierane marki odkurzaczy')
plt.show()
```



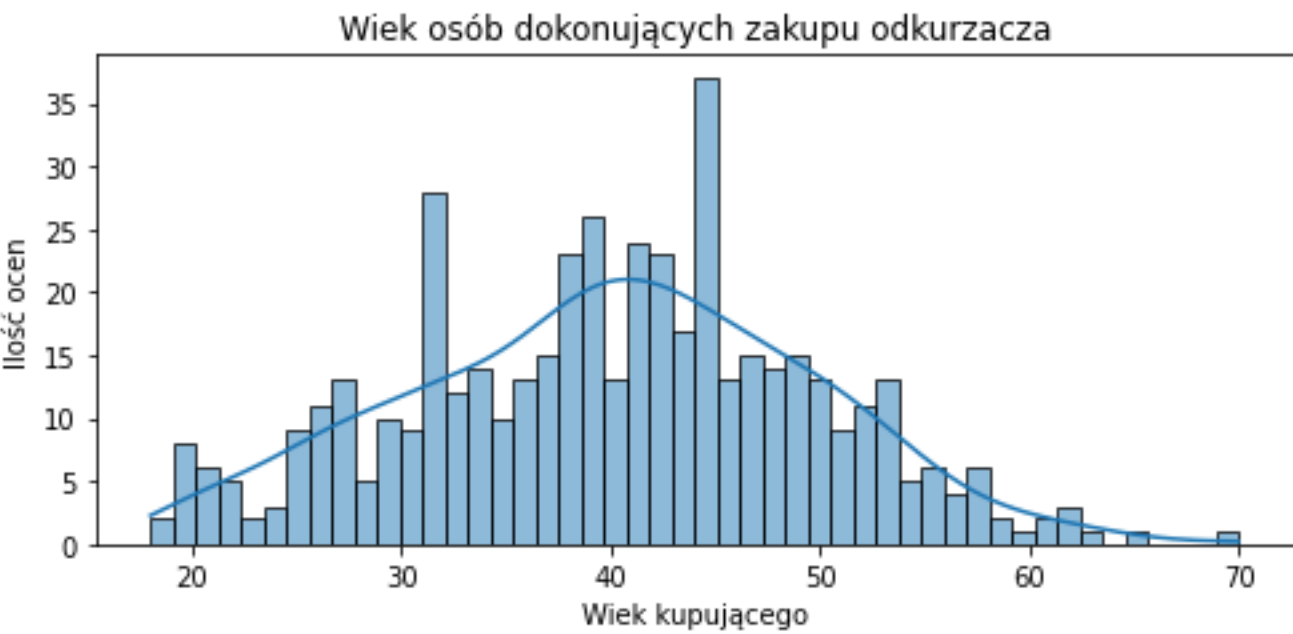
Najczęściej wybraną marką odkurzaczy był *Electrolux*.

### 3. Wiek osób dokonujących zakupu

```
In [111... x = wielkopolskie['Wiek kupującego'].value_counts()
x = [x.index[i] for i in range(len(x))]
len(x)

Out[111]: 48

In [162... fig3 = sns.histplot(data=wielkopolskie, x='Wiek kupującego', bins=48, kde=True)
fig3.set(ylabel='Ilość ocen', title='Wiek osób dokonujących zakupu odkurzacza')
plt.show()
```



Ilość zakupionych odkurzaczy rośnie do 40 roku życia, wtedy to zaczyna ona sukcesywnie spadać.

### 4. Płeć kupujących

```
In [158... fig4 = sns.histplot(data=wielkopolskie, x='Płeć kupującego')
fig4.set(ylabel='Ilość ocen', title='Płeć kupujących')
plt.show()
```



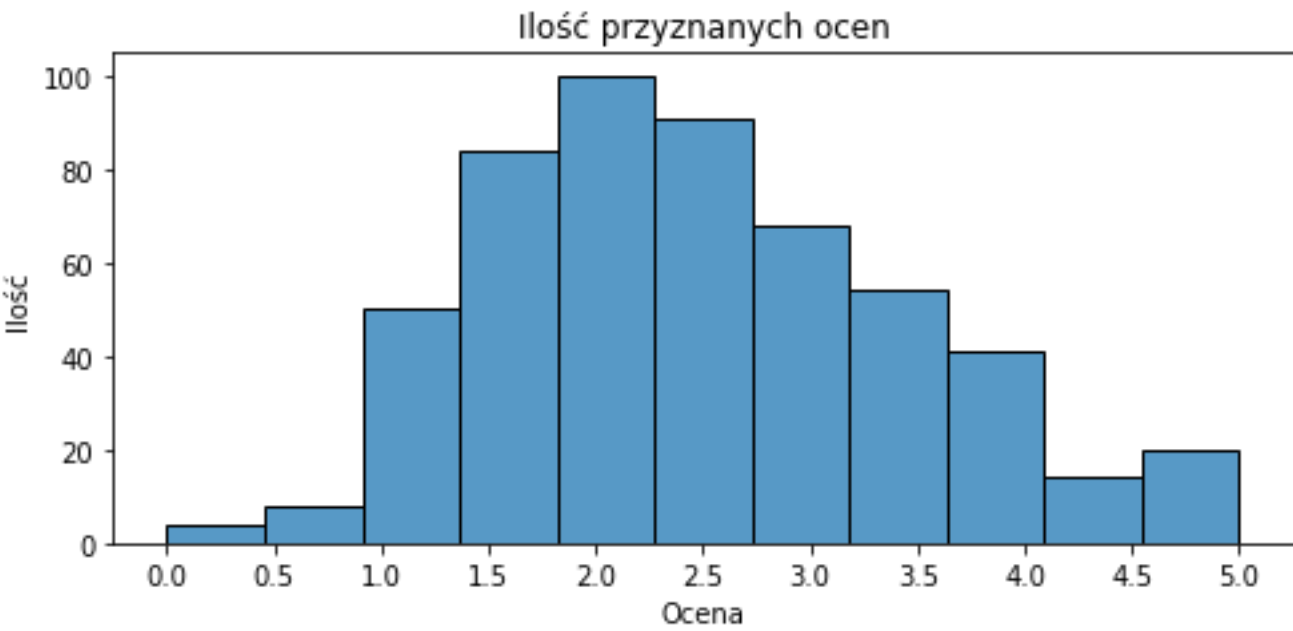
Zakupem odkurzaczy w województwie Wielkopolskim zainteresowani są w większości mężczyźni.

### 5. Oceny

```
In [150... x = wielkopolskie['Ocena'].value_counts()
x = [x.index[i] for i in range(len(x))]
print(len(x))

11

In [165... fig5 = sns.histplot(data=wielkopolskie, x='Ocena', bins=11)
fig5.set(ylabel='Ilość', title='Ilość przyznanych ocen')
x = np.arange(0, 5.5, 0.5)
plt.xticks(x)
plt.show()
```



Sprzedawane przez firmę odkurzacze nie są oceniane dobrze przez konsumentów. Najczęściej przyznawanymi ocenami są wartości z zakresu 1,5 - 2,5.