

The Final Paper

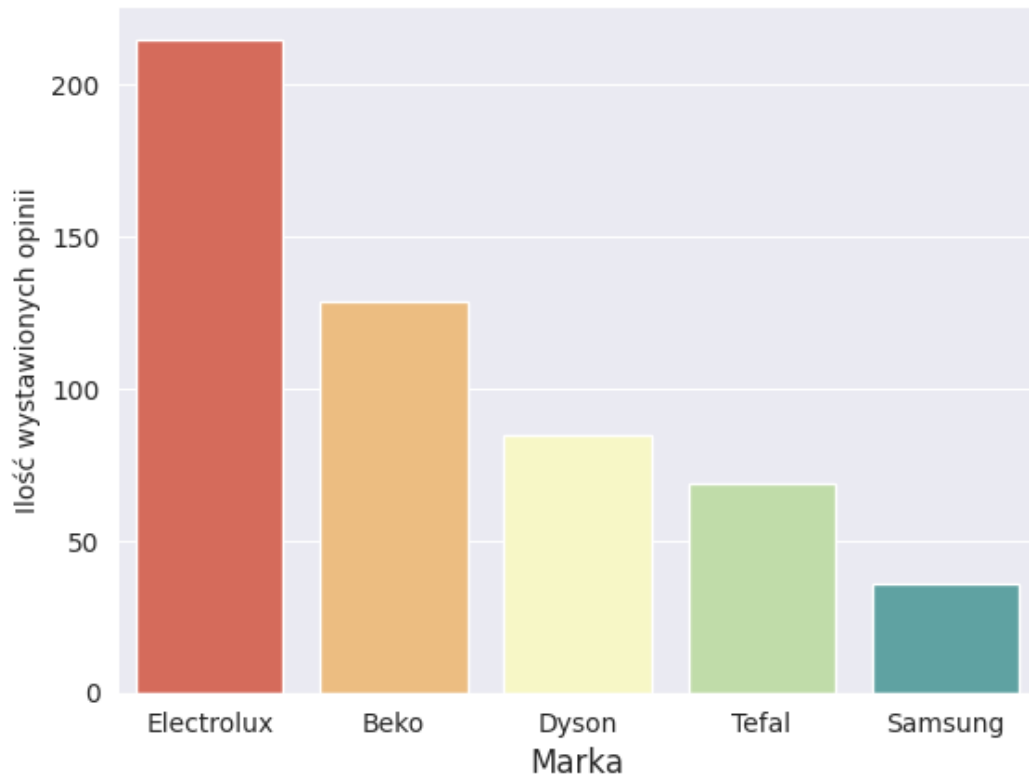
November 14, 2022

```
[58]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os
```

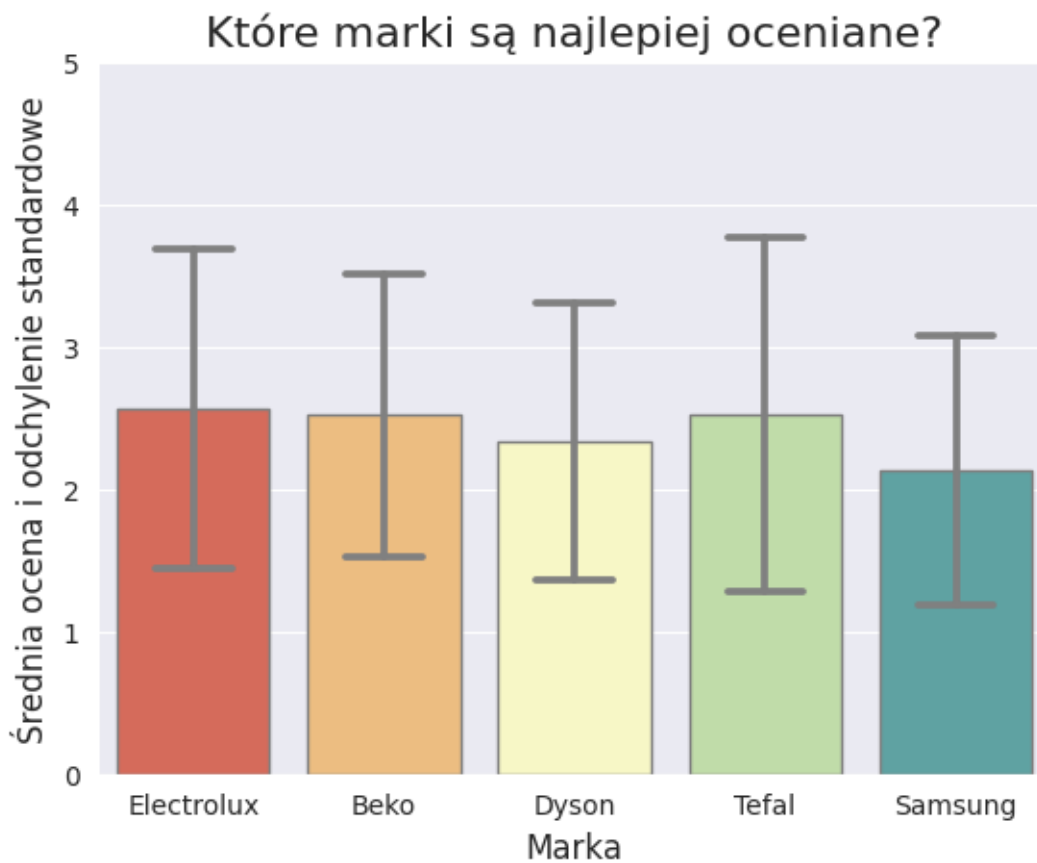
```
[59]: df = pd.read_csv('../analysis_data/Prepared_Data.csv')
clean_df = pd.read_csv('../analysis_data/Clean_Data.csv')
group_age_df = pd.read_csv('../analysis_data/Data_Group_Age.csv')
group_days_delay_df = pd.read_csv('../analysis_data/
↳Data_Group_Days_Delay_Rating.csv')
```

```
[60]: sns.countplot(df, x='Brand', palette="Spectral")
plt.title('Jaka marka odkurzaczy jest najbardziej popularna?', fontsize=16)
plt.xlabel('Marka', fontsize=12)
plt.ylabel('Ilość wystawionych opinii')
plt.show()
```

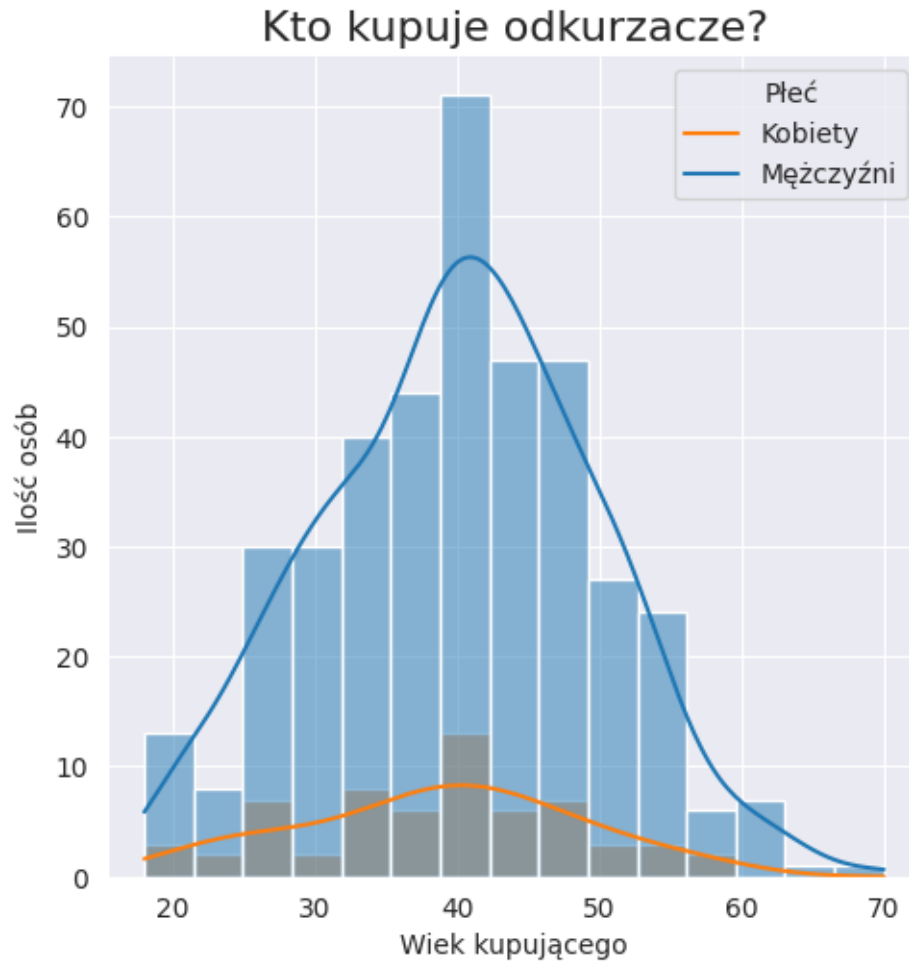
Jaka marka odkurzaczy jest najbardziej popularna?



```
[61]: sns.barplot(df, x='Brand', y='Rating', edgecolor=".5", errorbar='sd',  
               errcolor=".5", capsize=.4, palette="Spectral")  
plt.title('Które marki są najlepiej oceniane?', fontsize=16)  
plt.xlabel('Marka', fontsize=12)  
plt.ylabel('Średnia ocena i odchylenie standardowe', fontsize=12)  
plt.ylim([0, 5])  
plt.show()
```



```
[62]: sns.displot(clean_df, x='Buyer_Age', hue='Buyer_Sex', kde=True, legend=False)
plt.title('Kto kupuje odkurzacze?', fontsize=16)
plt.legend(['Kobiety', 'Mężczyźni'], title='Płeć')
plt.xlabel('Wiek kupującego')
plt.ylabel('Ilość osób')
plt.show()
```



- 1 Przy czym 11% ankietowanych nie podało swojego wieku, a prawie 3% swojej płci

```
[63]: plt.figure(figsize=(12, 4))

plt.subplot(1, 2, 1)
sns.barplot(group_age_df.where(group_age_df.Buyer_Sex == 'K').
    ↪sort_values(by='Buyer_Age_Group'), x='Buyer_Age_Group', y='Rating',
    ↪palette="flare", errorbar=None)
plt.title('Jak oceniali kobiety?', fontsize=16)
plt.xlabel('Wiek ankietujących kobiet', fontsize=12)
plt.ylabel('Średnia ocena', fontsize=12)
plt.ylim([0, 5])

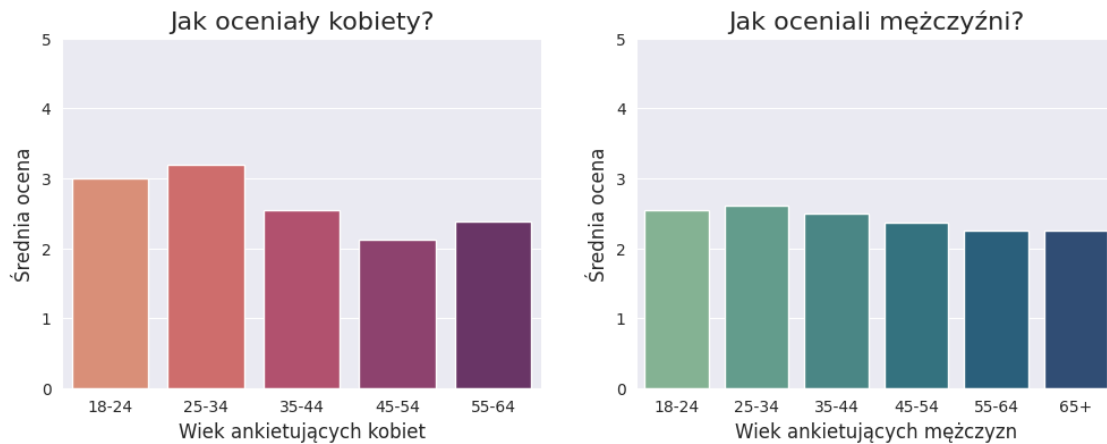
plt.subplot(1, 2, 2)
```

```

sns.barplot(group_age_df.where(group_age_df.Buyer_Sex == 'M').
    ↪sort_values(by='Buyer_Age_Group'), x='Buyer_Age_Group', y='Rating',
    ↪palette="crest", errorbar=None)
plt.title('Jak oceniali mężczyźni?', fontsize=16)
plt.xlabel('Wiek ankietujących mężczyzn', fontsize=12)
plt.ylabel('Średnia ocena', fontsize=12)
plt.ylim([0, 5])

plt.show()

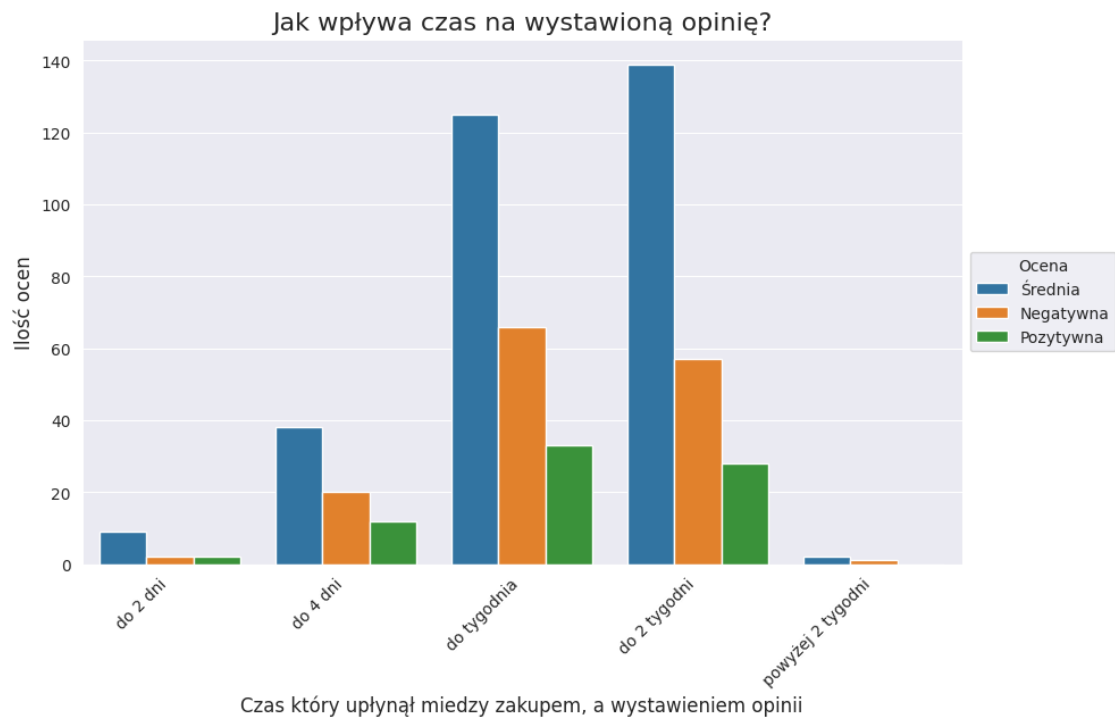
```



```

[64]: plt.figure(figsize=(10, 6))
sns.countplot(data=group_days_delay_df,
    ↪x='Days_From_Purchase_Range', hue='Rating_Group')
plt.title('Jak wpływa czas na wystawioną opinię?', fontsize=16)
plt.xlabel('Czas który upłynął między zakupem, a wystawieniem opinii',
    ↪fontsize=12)
plt.xticks(rotation=45, ha='right')
plt.legend(title='Ocena', loc='center left', bbox_to_anchor=(1, 0.5))
plt.ylabel('Ilość ocen', fontsize=12)
plt.show()

```



[64] :