

Mariusz Kuchta gr. 2a (czw. 10.00 - 11.30)

Laboratorium 6 - 18.11.21

Exploratory Data Analysis

In [55]:

```
len("MariuszKuchta") % 16
```

Out[55]: 13

Analizowany będzie zestaw danych nr 13

In [56]:

```
import pandas as pd

# czytaj pierwszą kolumnę jako index
d = pd.read_csv("../OriginalData/13_WARMINSKO-MAZURSKIE.csv", index_col=0)

d.head()
```

Out[56]:

	Dni od zakupu	Marka	Wiek kupującego	Płeć kupującego	Ocena
0	5	Dyson	45.0	K	2.0
1	9	Dyson	NaN	K	1.0
2	11	Beko	44.0	K	2.0
3	10	Electrolux	40.0	bd.	2.0
4	5	Electrolux	62.0	K	2.5

In [57]:

```
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt

np.min(d['Wiek kupującego'])
```

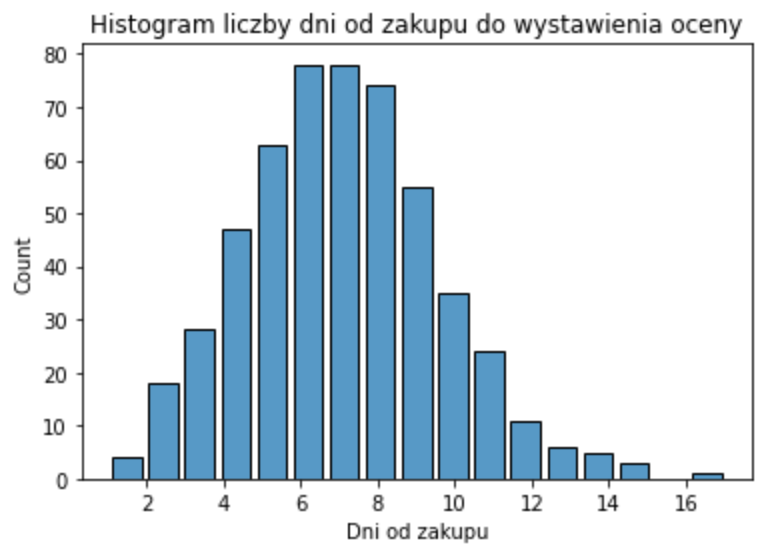
Out[57]: 18.0

Histogramy poszczególnych zmiennych.

In [58]:

```
sb.histplot(d['Dni od zakupu'], bins=17, shrink=0.8)
plt.title("Histogram liczby dni od zakupu do wystawienia oceny")
```

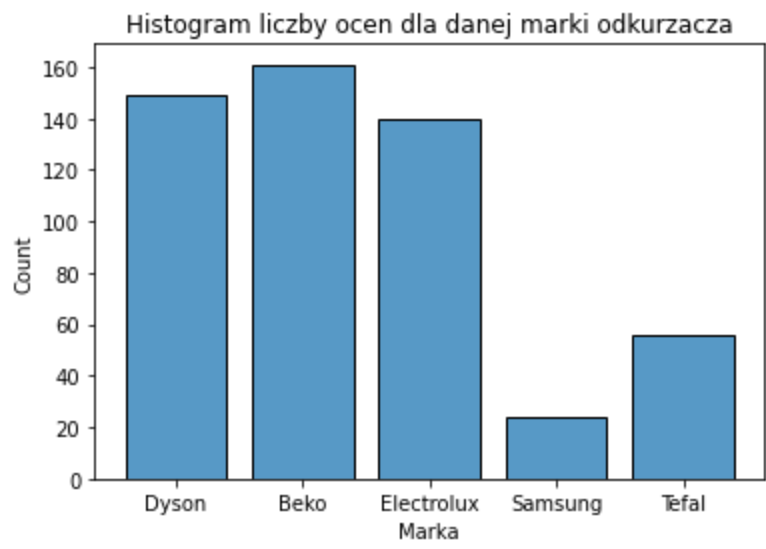
Out[58]: Text(0.5, 1.0, 'Histogram liczby dni od zakupu do wystawienia oceny')



In [59]:

```
sb.histplot(d['Marka'], bins=5, shrink=0.8)
plt.title("Histogram liczby ocen dla danej marki odkurzacza")
```

Out[59]: Text(0.5, 1.0, 'Histogram liczby ocen dla danej marki odkurzacza')

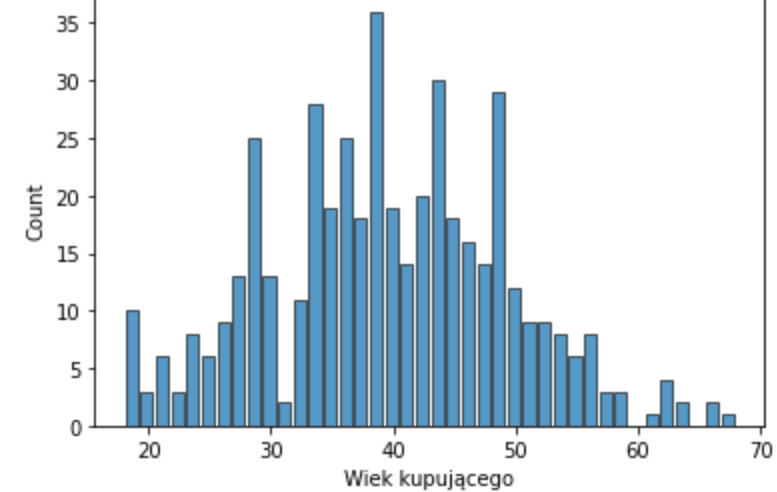


In [73]:

```
sb.histplot(d['Wiek kupującego'], bins=40, shrink=0.8)
plt.title("Histogram liczby ocen w zależności od wieku wystawiającego klienta")
np.sum(np.isnan(d['Wiek kupującego']))
```

Out[73]: 67

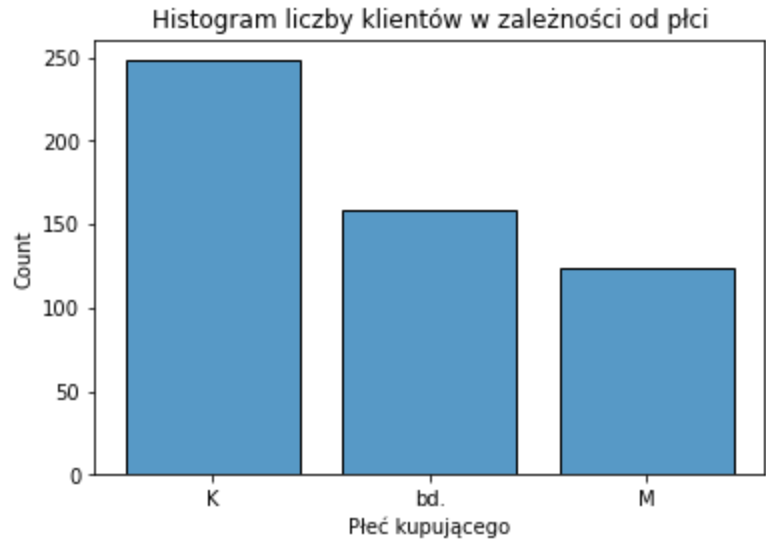
Histogram liczby ocen w zależności od wieku wystawiającego klienta



In [61]:

```
sb.histplot(d['Płeć kupującego'], bins=3, shrink=0.8)
plt.title("Histogram liczby klientów w zależności od płci")
```

Out[61]: Text(0.5, 1.0, 'Histogram liczby klientów w zależności od płci')



In [62]:

```
sb.histplot(d['Ocena'], bins=10, shrink=0.8)
plt.title("Histogram liczby ocen w zależności od ich wartości")
```

Out[62]: Text(0.5, 1.0, 'Histogram liczby ocen w zależności od ich wartości')

