

In [97]:

```
import pandas as pd
```

In [98]:

```
df = pd.read_csv(r'../OriginalData/1_KUJAWSKO-POMORSKIE.csv')
```

## Ogólne informacje dotyczące zbioru danych

In [99]:

```
print('Rozmiar zbioru danych: ', df.shape)
```

Rozmiar zbioru danych: (458, 6)

In [100]:

```
df.dtypes
```

Out[100]:

```
Unnamed: 0          int64
Dni od zakupu       int64
Marka              object
Wiek kupującego    float64
Płeć kupującego    object
Ocena              float64
dtype: object
```

## Opis poszczególnych zmiennych

### Dni od zakupu

**liczba dni, która minęła od zakupu odkurzacza[dni]**

**Liczba brakujących wartości :**

In [101]:

```
df['Dni od zakupu'].isnull().sum()
```

Out[101]:

0

In [102]:

```
df['Dni od zakupu'].describe()
```

Out[102]:

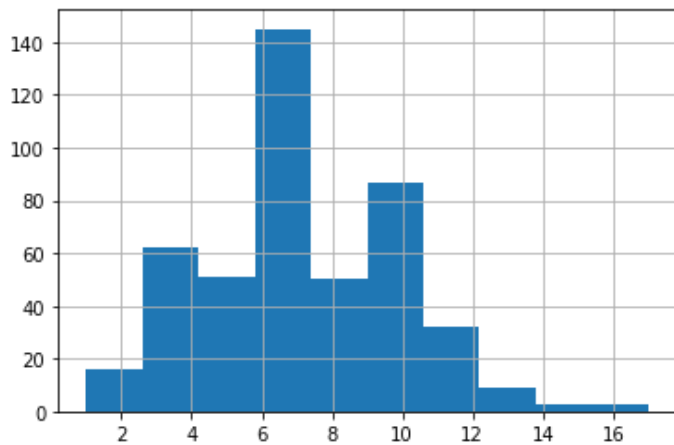
```
count    458.000000
mean      7.074236
std       2.708070
min       1.000000
25%       5.000000
50%       7.000000
75%       9.000000
max      17.000000
Name: Dni od zakupu, dtype: float64
```

In [103]:

```
df['Dni od zakupu'].hist()
```

Out[103]:

<AxesSubplot:>



## Marka

### nazwa marki odkurzacza

Liczba brakujących wartości :

In [104]:

```
df['Marka'].isnull().sum()
```

Out[104]:

0

In [105]:

```
df['Marka'].value_counts()
```

Out[105]:

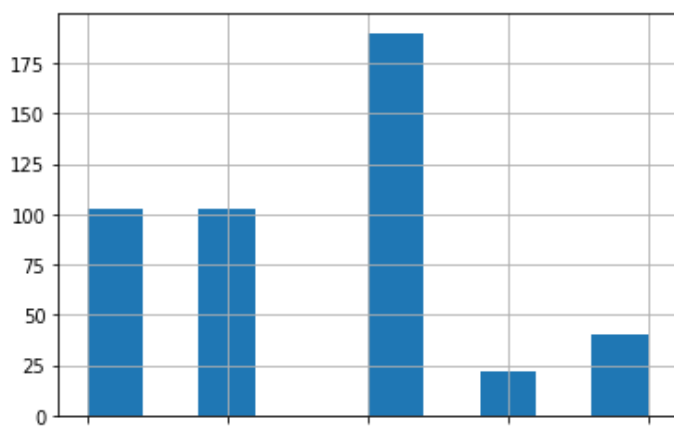
```
Dyson          190
Electrolux     103
Beko           103
Tefal          40
Samsung        22
Name: Marka, dtype: int64
```

In [106]:

```
df['Marka'].hist()
```

Out[106]:

<AxesSubplot:>



## Wiek kupującego

**Liczba brakujących wartości :**

In [107]:

```
df['Wiek kupującego'].isnull().sum()
```

Out[107]:

45

In [108]:

```
df['Wiek kupującego'].describe()
```

Out[108]:

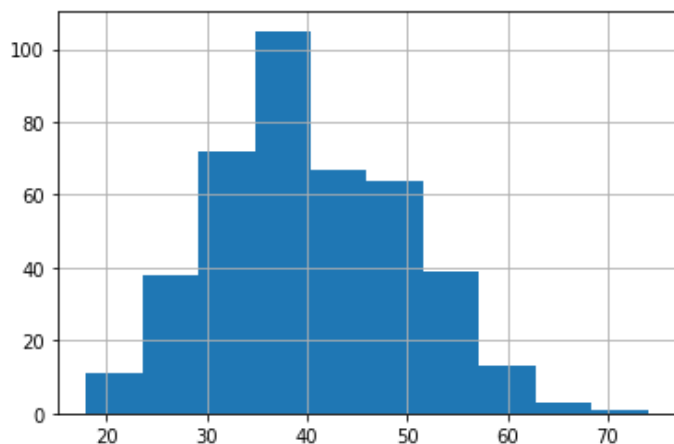
```
count    413.000000
mean      40.125908
std        9.558640
min       18.000000
25%       33.000000
50%       39.000000
75%       46.000000
max       74.000000
Name: Wiek kupującego, dtype: float64
```

In [109]:

```
df['Wiek kupującego'].hist()
```

Out[109]:

<AxesSubplot:>



## Płeć kupującego

**określa płeć osoby kupującej**

**M - mężczyzna, K - kobieta, bd - brak.danych**

**Liczba brakujących wartości :**

In [110]:

```
df['Wiek kupującego'].isnull().sum()
```

Out[110]:

In [111]:

```
df['Płeć kupującego'].value_counts()
```

Out[111]:

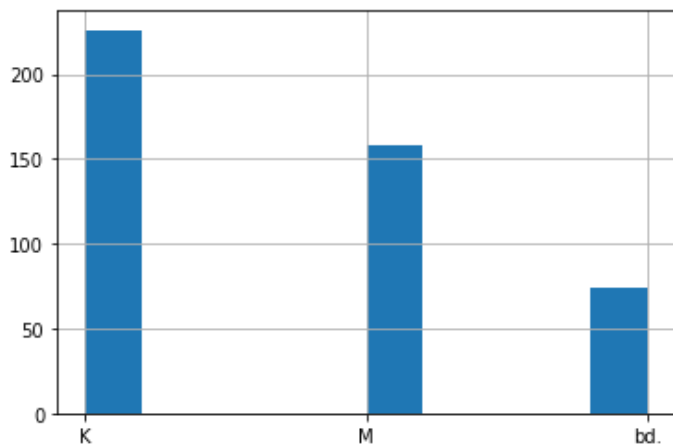
```
K      226
M      158
bd.     74
Name: Płeć kupującego, dtype: int64
```

In [112]:

```
df['Płeć kupującego'].hist()
```

Out[112]:

<AxesSubplot:>



## Ocena

### ocena urządzenia wystawiona przez użytkownika

#### zakres 0-5

#### Liczba brakujących wartości :

In [113]:

```
print(df['Wiek kupującego'].isnull().sum())
```

45

In [114]:

```
df['Wiek kupującego'].describe()
```

Out[114]:

```
count    413.000000
mean     40.125908
std       9.558640
min       18.000000
25%      33.000000
50%      39.000000
75%      46.000000
max       74.000000
Name: Wiek kupującego, dtype: float64
```

In [115]:

```
df['Ocena'].hist()
```

Out[115]:

<AxesSubplot:>

