Week -5

Movies

In [1]:
```
import pandas as pd
pd.set_option('display.precision', 3)
```

In [2]:
```
file_df = pd.read_csv("Movies.csv")
```

In [3]:
```
print("Movie with the maximum rating is:", file_df.loc[file_df["Rating"].idxmax()][
```

```
Movie with the maximum rating is: Movie9
```

In [4]:
```
hindi_df = file_df[file_df["Language"] == "Hindi"]
hindi_df.to_csv("HindiMovies.csv", index=False)
```

Cereals

In [5]:
```
cereal_df = pd.read_csv("Cereal.csv")

print("======================")
print("5 number summary:")
print("======================\n")
print(cereal_df.describe().loc[['min', '25%', '50%', '75%', 'max']])
```

```
======================
5 number summary:
======================

     Calories  Protein (g)  Fat  Sodium  Dietary Fiber  Carbs  Sugars  \
min      50.0          1.0  0.0    0.00           0.00   -1.0   -1.00
25%     100.0          2.0  0.0  131.25           0.25   12.0    3.00
50%     110.0          2.0  1.0  180.00           1.50   14.0    6.00
75%     110.0          3.0  1.0  217.50           3.00   17.0   10.75
max     160.0          6.0  5.0  320.00          14.00   23.0   15.00

     Display Shelf  Potassium  Vitamins and Minerals  Serving Size Weight  \
min            1.0      -1.00                    0.0                  0.5
25%            1.0      40.00                   25.0                  1.0
50%            2.0      90.00                   25.0                  1.0
75%            3.0     113.75                   25.0                  1.0
max            3.0     330.00                  100.0                  1.5

     Cups per Serving
min             -1.00
25%              0.67
50%              0.75
75%              1.00
max              1.50
```

In [6]:
```
# To replace -1 values
protein_to_vitamin = cereal_df.loc[:, "Calories":]
for col in protein_to_vitamin:
    mean_val = cereal_df[col][cereal_df[col] != -1].mean()
    cereal_df[col] = cereal_df[col].replace(-1, mean_val)
```

In [7]:
```
print("\n====================")
print("After replacing -1:")
print("====================\n")
print(cereal_df.describe().loc[['min', '25%', '50%', '75%', 'max']])
```

```
====================
After replacing -1:
====================

     Calories  Protein (g)  Fat  Sodium  Dietary Fiber   Carbs  Sugars  \
min      50.0          1.0  0.0    0.00           0.00   5.000   0.000
25%     100.0          2.0  0.0  131.25           0.25  12.000   3.000
50%     110.0          2.0  1.0  180.00           1.50  14.404   6.438
75%     110.0          3.0  1.0  217.50           3.00  17.000  10.750
max     160.0          6.0  5.0  320.00          14.00  23.000  15.000

     Display Shelf  Potassium  Vitamins and Minerals  Serving Size Weight  \
min            1.0      15.00                    0.0                  0.5
25%            1.0      41.25                   25.0                  1.0
50%            2.0      90.00                   25.0                  1.0
75%            3.0     113.75                   25.0                  1.0
max            3.0     330.00                  100.0                  1.5

     Cups per Serving
min              0.25
25%              0.67
50%              0.75
75%              1.00
max              1.50
```

In [8]:
```
# To treat noisy values
for col in protein_to_vitamin:
    median_val = cereal_df[col].median()
    cereal_df[col] = cereal_df[col].apply(
        lambda v: median_val
        if (v < cereal_df[col].quantile(0.05) or v > cereal_df[col].quantile(0.95))
        else v)
```

In [9]:
```
print("\n====================")
print("After treating noise:")
print("====================\n")
print(cereal_df.describe().loc[['min', '25%', '50%', '75%', 'max']])
```

```
====================
After treating noise:
====================

     Calories  Protein (g)  Fat  Sodium  Dietary Fiber   Carbs  Sugars  \
min      70.0          1.0  0.0    0.00           0.00   9.000   0.000
25%     100.0          2.0  0.0  131.25           0.25  13.000   3.000
50%     110.0          2.0  1.0  180.00           1.50  14.404   6.219
75%     110.0          3.0  1.0  207.50           3.00  17.000  10.000
max     130.0          4.0  2.0  280.00           5.00  21.000  14.000

     Display Shelf  Potassium  Vitamins and Minerals  Serving Size Weight  \
min            1.0       25.0                    0.0                 1.00
25%            1.0       45.0                   25.0                 1.00
50%            2.0       90.0                   25.0                 1.00
75%            3.0      110.0                   25.0                 1.00
max            3.0      240.0                  100.0                 1.33

     Cups per Serving
min              0.50
25%              0.69
50%              0.75
75%              1.00
max              1.00
```