Problem 1: Overall Average Rating (2 points)

We would like to compute the average rating for all the apps in the dataset. This will give us an idea of which apps are performing above average and which are performing below average, in terms of rating.

Write a python script that does that. Use Pandas Data Frames with the mean() function applied to the rating column.

Here are the steps that you should follow in your code:

1. Read the csv file into a dataframe using the pandas.read_csv() function
2. Extract the 'Rating' column and compute its mean
3. Print the mean

In [1]:
```
import pandas as pd

df = pd.read_csv('sample_googleplaystore.csv', header=0)
ratingCol = df['Rating']
ratingMean = ratingCol.mean()
print(f'Average Rating: {ratingMean:.2f}')
```

```
Average Rating: 4.32
```

Problem 2: Average Rating per Category (3 points)

Now we would like to compute the average rating within each category. This can tell us which categories tend to have higher ratings (by higher, we mean above average), and which ones have lower ratings.

There are multiple ways to do that in pandas. Try using the groupby() function, or using a for loop as discussed below.

To do that, you should perform the following:

1. Read the csv file into a dataframe using the pandas.read_csv() function
2. Extract the different categories. This can be done by applying the unique() function to get the count of each category: c = data['Category'].unique().tolist()
3. Now you can compute the rating for each of the categories in the list c. Use a for loop to scan c and , extract the ratings for that category, then apply the mean() function to find the average rating. For each category cat in c:
   A. Extract the ratings for all apps that fall into this category: x = data[data['Category'] == cat]
   B. Compute the average over the 'Rating' column: avg = x['Rating'].mean()
   C. Print the average for this category cat: print(cat, "=", "%.2f" % avg)

In [2]:
```
df = pd.read_csv('sample_googleplaystore.csv', header=0)

categoryList = df['Category'].unique().tolist()

for category in categoryList:
    categoryRating = df[df['Category'] == category]
```

```
    categoryRatingAvg = categoryRating['Rating'].mean()
    print(f'{category} Average Rating = {categoryRatingAvg:.2f}')
```

```
ART_AND_DESIGN Average Rating = 4.32
AUTO_AND_VEHICLES Average Rating = 4.30
BEAUTY Average Rating = 4.33
BOOKS_AND_REFERENCE Average Rating = 4.37
BUSINESS Average Rating = 4.29
```

Problem 3: Full Data Summarization (5 points)

To summarize the whole dataset, find the following:

1. For each non-numeric column, find the number of unique labels and the frequency of each. For example, for the type column, find the number of free versus paid apps. (Check the value_counts() function)
2. For each numeric column, find the min, max and average values. (Check the min(), max() and mean() functions). Attributes such as size, in mega bytes, are considered non numeric because they have alpha-numeric characters. They can be treated as numeric values after data cleaning, which we will cover soon.

In [4]:
```
# Read the dataset into a DataFrame
df = pd.read_csv('sample_googleplaystore.csv', header=0)

# Task 1: For each non-numeric column, find the number of unique labels and
non_numeric_columns = df.select_dtypes(exclude='number').columns
for column in non_numeric_columns:
    print(df[column].value_counts(), '\n')

# Task 2: For each numeric column, find the min, max, and average values
numeric_columns = df.select_dtypes(include='number').columns
for column in numeric_columns:
    min_value = df[column].min()
    max_value = df[column].max()
    avg_value = df[column].mean()
    print(column)
    print(f"Minimum value: {min_value}")
    print(f"Maximum value: {max_value}")
    print(f"Average value: {avg_value:.2f}\n")
```

```
App
Quick PDF Scanner + OCR FREE                          2
Box                                                   2
FBReader: Favorite Book Reader                        1
Free Books - Spirit Fanfiction and Stories            1
Google Play Books                                     1
                                                     ..
Speed Camera Detector - Traffic & Speed Alert         1
Used car search Goo net whole car Go to net           1
CarMax – Cars for Sale: Search Used Car Inventory     1
BEST CAR SOUNDS                                       1
Google Ads                                            1
Name: count, Length: 237, dtype: int64

Category
BUSINESS              52
ART_AND_DESIGN        49
AUTO_AND_VEHICLES     49
BOOKS_AND_REFERENCE   48
BEAUTY                41
Name: count, dtype: int64

Size
Varies with device    49
14M                    8
17M                    7
25M                    7
2.9M                   6
                      ..
4.5M                   1
9.8M                   1
52M                    1
9.0M                   1
7.5M                   1
Name: count, Length: 94, dtype: int64

Installs
100,000+          60
1,000,000+        50
500,000+          27
10,000,000+       26
10,000+           24
5,000,000+        23
50,000+           17
50,000,000+        5
5,000+             3
100,000,000+       3
1,000,000,000+     1
Name: count, dtype: int64

Type
Free    237
Paid      2
Name: count, dtype: int64

Price
0         237
$4.99       2
Name: count, dtype: int64

Content Rating
Everyone        214
Teen             14
Everyone 10+      6
```

```
Mature 17+        5
Name: count, dtype: int64

Genres
Business                          52
Auto & Vehicles                   49
Books & Reference                 48
Art & Design                      42
Beauty                            41
Art & Design;Creativity            5
Art & Design;Pretend Play          1
Art & Design;Action & Adventure    1
Name: count, dtype: int64

Last Updated
30-Jul-18    12
2-Aug-18     11
31-Jul-18     9
1-Aug-18      7
26-Jul-18     6
             ..
15-Jul-18     1
16-Jul-18     1
25-May-18     1
14-Oct-16     1
11-Apr-17     1
Name: count, Length: 121, dtype: int64

Current Ver
Varies with device    45
1                     15
1.1                    6
3                      5
1.6                    5
                      ..
2.3.5.1                1
1.79                   1
2.9                    1
2.2.21                 1
1.12.0                 1
Name: count, Length: 132, dtype: int64

Android Ver
4.0.3 and up          54
4.1 and up            48
Varies with device    37
4.0 and up            32
2.3 and up            13
4.2 and up            11
4.4 and up            10
3.0 and up            10
5.0 and up             7
2.2 and up             4
6.0 and up             3
2.3.3 and up           2
1.5 and up             2
7.0 and up             2
1.6 and up             1
2.1 and up             1
5.1 and up             1
4.3 and up             1
Name: count, dtype: int64

Rating
```

```
Minimum value: 3.1
Maximum value: 4.9
Average value: 4.32

Reviews
Minimum value: 2
Maximum value: 2914724
Average value: 77748.46
```

In [ ]: