

Faculty of Engineering & Technology
Department of Computer Science

COMP4381 - Data Science
Instructor: Ahmed Sabbah

Individual Assignment

Spring 2024/2025
Date: 5/4/2025

Analyzing Google Play Store Apps Dataset

Dataset: [Google Play Store Apps](#)

Download Instructions:

1. Download **googleplaystore.csv** from Kaggle.
2. Upload it to your Jupyter Notebook environment.

Tasks

Answer the following questions using pandas. Include code, outputs, and explanations in a Jupyter Notebook.

1. Loading and Initial Exploration

- Load the dataset into a pandas DataFrame.
- Display the first **6 rows** and the last **6 rows**.
- Show the **shape** of the dataset (number of rows and columns).
- Generate **descriptive statistics** for numerical columns (e.g., Rating, Reviews, Installs) with interpretation.

2. Data Cleaning:

- Print the **data types** of all columns and check the ability to have **null** values.
- Drop rows where Type (**Free/Paid**) is missing.
- For the Rating column, fill missing values with the **median rating** of the same app category. Explain why you chose the median.
- Convert the Installs column to a numeric type (remove commas and "+", e.g., "1,000,000+" → 1000000).
- Convert the Price column to a numeric type (\$0.99 → 0.99).

3. Data Filtering

- Create a new DataFrame containing only **free apps** (Type = Free).
- Filter apps with a **rating greater than 4.5** and **more than 1 million installs**.



Faculty of Engineering & Technology
Department of Computer Science

COMP4381 - Data Science

Instructor: Ahmed Sabbah

Individual Assignment

Spring 2024/2025

Date: 5/4/2025

- List the top 3 most installed apps in the "**Education**" category. **Hint** (case sensitive)

4. Grouping and Aggregation

- Group the data by Category and calculate the **average rating** for each category.
- Which category has the lowest and highest average rating?
- Group the data by **Content Rating** (e.g., Everyone, Teen) and calculate each group's **median and sum number of installs**.

5. New Column Creation

- Create a new column, **Revenue Estimate** for paid apps, by multiplying Price (convert to numeric) by Installs.

6. Visualization

- Use matplotlib to create a **histogram** of app ratings.
- Create a **bar plot** showing the top 10 most common app categories.

Submission Instructions

1. Submit a Jupyter Notebook (.ipynb) with code, outputs, and explanations.
2. **Jupyter Notebook name must include your name and ID: (AhmedSabbah123456)**
3. Ensure all code cells run without errors.
4. Use markdown cells to explain your work.
5. You may use any AI generative code model such as ChatGPT if you want. But you will be doing yourself a terrible disservice if you use this to fake Python programming rather than learn it, and you will get zero marks (**Be ready any time for a discussion of your code**)—now is your chance to learn!