

Analysis and visualization of massive dataset using Pandas

In this homework, you will use Pandas to analyze the Google Play Store dataset (retrieved from Kaggle). You will be manipulating and visualizing using various NumPy and Pandas functions to extract the essence of the given data. The specific instructions are as follows:

1. Load the `googleplaystore_user_reviews.csv` (reviews) and `googleplaystore.csv` (apps) data sets as pandas DataFrames. You should use the `pd.read_csv()` function for this purpose.
2. Delete any review that does not contain either a **Translated_Review** or a **Sentiment**. The `pd.dropna` function will be helpful for this.
3. Remove any apps with invalid **Rating** (>5).
4. Produce a pie chart with the **Android Ver** requirements for the different apps. Group together all versions that make up less than 5% of the total apps into a single 'Others' category.
5. Create a similar pie chart for app **Category**. In this case, group together categories that make up less than 3% of the apps.
6. Show histograms of the Rating and Reviews across all apps, with 20 bins each.
7. Plot a bar chart for different **Sentiment** categories. Each bar should represent the frequency of each Sentiment within the total dataset.
8. Combine the two DataFrames into a single one, based on the **App** names. You should make sure that all apps from the apps DataFrame are kept, and no app beyond those is added. The `pd.merge` function will be useful for this.
9. Group the **Sentiment** by rounded **Rating**, and produce a bar chart where you display the different sentiments grouped by rating.

You are free to use any NumPy, Pandas, and Matplotlib functions.

Please attach a code file (or jupyter notebook) you have developed for the assignment, and a report containing the produced figures and your reasoning on the general implications of the data based on the figures.