Play Store apps data cleaning and preprocessing

The Play Store hosts a wide range of applications, and the data associated with these apps offers significant potential for analysis. Developers and businesses can leverage this data to understand market trends, optimize their apps, and maximize user engagement.

In this project, we will perform a data cleaning and preprocessing task on the Play Store apps dataset, addressing missing values, outliers, and other issues while extracting valuable insights.

Column Name	Description
Application name	Name of the application
Sort ascending	Sort in ascending order
Sort descending	Sort in descending order
Category	Category the app belongs to
Rating	Overall user rating of the app (as when scraped)
Reviews	Number of user reviews for the app (as when scraped)
Size	Size of the app (as when scraped)
Installs	Number of user downloads/installs for the app (as when scraped)
Type	Whether the app is paid or free
Price	Price of the app (as when scraped)
Content Rating	Age group the app is targeted at (e.g., Children / Mature 21+ / Adult)
Genres	An app can belong to multiple genres (apart from its main category). For example, a musical family game will belong to multiple genres.

Questions

- **Work a Data Cleaning**

  Tasks:

  - Fix Rating
  - Fix Size
  - Fix Price
  - Fix Category
  - Fix Android Version
  - Handle Missing Values
  - Outlier Detection and Handling
  - Fix anything need fixing

Analytical Questions:

1. **What is the most expensive app on the Play Store?**
2. **Which genre has the highest number of apps?**
3. **What is the average size of free vs. paid apps?**
4. **What are the top 5 most expensive apps with a perfect rating (5)?**
5. **How many apps have received more than 50K reviews?**
6. **What is the average price of apps, grouped by genre and number of installs?**
7. **How many apps have a rating higher than 4.7, and what is their average price?**
8. **What is Google's estimated revenue from apps with 5,000,000+ installs?**
   (Assuming Google takes a 30% cut from app sales)
9. **What are the maximum and minimum sizes of free vs. paid apps?**
10. **Is there a correlation between an app's rating, number of reviews, size, and its price?**
11. **How many apps exist for each type (free/paid) across different content ratings?**
12. **How many apps are compatible with Android version 4.x?**

Deliverables:

1. **Preprocessing Notebook:** Submit a Jupyter notebook detailing the data cleaning steps taken for each column (with comments).
2. **Analysis Results:** Present your answers to the analytical questions in a clear and organized format.
3. **Visualizations:** Include visualizations to support the insights drawn from the dataset for all analytical questions and in cleaning.
4. Upload your project to GitHub with more comments and markdowns and a good show at Notebook.
5. The submission deadline, which will be December 17, 2024.
6. After evaluation you will post project and notebook in linkedIn

Good luck with your data preprocessing project! Make sure to clean the data thoroughly before answering the analytical questions.