**Google Play Store Data Analytics Internship Report**

**Name:** Reva Saxena  
 **Email:** revasaxena90@gmail.com  
 **Internship Company:** NullClass

**1. Introduction**

This report outlines the work I did during my internship on **Google Play Store Data Analytics** using Python. The purpose of the internship was to gain hands-on experience in analyzing data from the Google Play Store, with a focus on understanding app performance and trends through data visualizations. I worked with Python libraries like Pandas, Plotly, and Matplotlib to clean the data, perform analysis, and create interactive visualizations. In this report, I’ll walk you through the tasks I completed, the skills I learned, the challenges I faced, and the final outcomes of the project.

**2. Background**

The **Google Play Store** is one of the largest digital marketplaces for apps, with millions of applications available for download. This dataset provides key insights into app performance metrics, such as ratings, installs, revenue, and categories. For this project, I analyzed this data to uncover trends and create visualizations that help in understanding the factors influencing app success. By using Python’s data analysis libraries, I was able to filter, manipulate, and visualize the data, ultimately making it easier to derive actionable insights from it.

**3. Learning Objectives**

The main objectives I set for this internship were:

* To sharpen my skills in **data analysis** and **data visualization** using Python.
* To create various types of interactive visualizations that show insights about app performance on the Play Store.
* To explore how data can be used to make decisions about app development and marketing strategies.
* To improve my knowledge of Python libraries such as **Pandas**, **Plotly**, and **Matplotlib**.

**4. Activities and Tasks**

**Task 1: Scatter Plot for Paid Apps**

I created a scatter plot to explore the relationship between the number of installs and the revenue generated by paid apps. Each app was color-coded by its category, making it easy to spot trends and outliers.

**Task 2: Interactive Choropleth Map**

This task involved creating an interactive map to show the global distribution of installs across different app categories. I applied filters to show specific categories and focused on apps that had a significant number of installs.

**Task 3: Grouped Bar Chart Comparison**

I compared the top 10 app categories by the number of installs, analyzing how their average ratings and review counts compared to each other. This chart provided a good overview of which categories performed best.

**Task 4: Dual-Axis Chart for Free vs Paid Apps**

Here, I created a dual-axis chart comparing installs and revenue for free and paid apps across the top three categories. This allowed me to visualize the performance difference between paid and free apps in a clear way.

**Task 5: Correlation Heatmap**

To understand the relationships between installs, ratings, and review counts, I generated a heatmap. This helped highlight any significant correlations between these variables, especially for apps that were updated in the past year.

**Task 6: Time Series Line Chart**

I visualized the trend of total installs over time, segmented by app category. This allowed me to identify patterns or growth trends in specific app categories.

**Task 7: Bubble Chart**

This task involved creating a bubble chart to analyze the relationship between app size, ratings, and installs. It allowed for filtering apps that had ratings higher than 3.5, providing an in-depth view of successful apps.

**5. Skills and Competencies**

Throughout this internship, I developed several important skills:

* **Data Cleaning and Preprocessing**: I became proficient in cleaning data, handling missing values, and preparing datasets for analysis.
* **Data Visualization**: I created various interactive visualizations to convey the insights I gained from the data. These included scatter plots, bar charts, heatmaps, and more.
* **Python Libraries**: I used **Pandas** for data manipulation, **Plotly** for creating interactive charts, and **Matplotlib** for static visualizations.
* **Real-Time Data Handling**: I learned how to work with real-time data and integrate filters for dynamic analysis.
* **Problem-Solving**: I tackled various challenges related to large datasets, performance optimization, and custom visualizations.

**6. Challenges and Solutions**

**Challenges**

* **Handling Missing Data**: The dataset contained some missing values, which could affect the accuracy of analysis.
* **Complex Visualizations**: Creating interactive visualizations with custom filters took time and required a deep understanding of the tools I was using.
* **Performance Issues**: The dataset was large, and rendering interactive charts took longer than expected, especially when displaying real-time data.

**Solutions**

* I handled missing data using methods like fillna() and dropna() in Pandas.
* For complex visualizations, I carefully customized the charts using Plotly’s advanced features like filtering and dynamic updates.
* To improve performance, I optimized how the data was loaded and processed, making the visualization process faster.

**7. Outcomes and Impact**

The final outcome of the internship was a series of meaningful visualizations that shed light on app performance trends. Some key insights include:

* A strong correlation between installs and revenue for paid apps.
* The most popular app categories and their respective ratings and review counts.
* An analysis of app performance over time, identifying growth periods for certain categories.

These visualizations offer valuable insights for developers and marketers looking to improve their apps based on user feedback and performance metrics.

**8. Conclusion**

This internship was a fantastic opportunity for me to enhance my skills in data analytics and visualization. I not only learned how to analyze large datasets but also how to present complex data in a way that’s easily understandable. The project helped me understand the real-world applications of data analytics, especially in the context of app development and marketing. I now feel more confident in my ability to tackle data-related challenges and create insightful visualizations that can guide decision-making.

**9. References**

* [Pandas Documentation](https://pandas.pydata.org/pandas-docs/stable/)
* [Plotly Documentation](https://plotly.com/python/)
* [Matplotlib Documentation](https://matplotlib.org/stable/contents.html)