

INTERNSHIP REPORT



INTERNSHIP TITLE: DATA ANALYST INTERN

COMPANY/ORGANIZATION: NULLCLASS

INTERNSHIP PERIOD: 10-01-2025 TO 10-03-2025

SUBMITTED BY: KRISHNA G HEGDE

SUBMISSION DATE: 18-02-2025

INTRODUCTION

The **Data Analyst Internship** at **NullClass** provided me with hands-on experience in **data analysis, preprocessing, visualization and dashboard creation**. This internship was designed to enhance my skills in working with real-world datasets and extracting meaningful insights.

As a **Data Analyst Intern**, I focused on analyzing **Google Play Store data**, where I explored various app categories, user ratings and reviews patterns. My tasks involved **data preprocessing, filtering, statistical analysis, and visualization using Python**. Additionally, I developed an **interactive dashboard** to present key insights effectively.

This report outlines my experience, the tasks I completed, the challenges I faced, and the skills I developed during the internship.

BACKGROUND

The Data Analyst Internship at NullClass was designed to provide **practical exposure** to real-world datasets and analytical methodologies. The primary objective was to gain hands-on experience in **data preprocessing, analysis, and visualization**, along with understanding the key aspects of **data-driven decision-making**.

Throughout the internship, I worked with **Google Play Store data**, focusing on extracting meaningful insights from app categories, user ratings, and review patterns. I learned how to **clean and preprocess raw datasets, handle missing values, and apply statistical techniques** to enhance data quality. Additionally, I developed skills in **Plotly for interactive data visualization** and used **HTML and JavaScript to create dynamic dashboards** for presenting insights effectively.

Beyond the assigned internship tasks, my training also included **sentiment analysis on user reviews**, helping me understand how text data can be analyzed to extract user sentiments and trends. While this was not implemented in the internship project, it expanded my knowledge of **natural language processing (NLP) techniques**.

This internship served as a **bridge between theoretical learning and practical implementation**, refining my **analytical, technical, and problem-solving skills** to tackle real-world data challenges.

LEARNING OBJECTIVES

The primary objective of this internship was to gain **practical experience in data analysis** and enhance my ability to work with real-world datasets. The specific learning goals included:

- **Data Preprocessing & Cleaning:** Understanding how to handle missing values, filter datasets, and preprocess raw data for analysis.
- **Exploratory Data Analysis (EDA):** Applying statistical methods to extract insights from the data.
- **Data Visualization:** Creating interactive and meaningful visualizations using **Plotly** to present key findings.
- **Dashboard Development:** Learning to design and implement dynamic dashboards using **HTML and JavaScript** for data representation.
- **Statistical Analysis:** Applying mathematical and statistical techniques to uncover trends and patterns in the dataset.
- **Real-World Problem Solving:** Enhancing my ability to work with large datasets, interpret results, and derive actionable insights.
- **Project Documentation:** Improving skills in documenting findings, structuring reports, and presenting insights in a professional format.

This internship allowed me to bridge the gap between **theoretical knowledge and practical application**, strengthening my **data analysis, visualization, and problem-solving** skills.

ACTIVITIES AND TASKS

During my internship at NullClass, I worked on various tasks related to data analysis, visualization, and dashboard development using **Python, Plotly, HTML, and JavaScript**. Below is a breakdown of the key activities I completed:

1. DATA COLLECTION & PREPROCESSING

- Filtered the dataset to include only relevant data, focusing on **Google Play Store apps**.
- **Converted categorical columns to numeric** (e.g., removed symbols like, + and \$ from "installs" and "price" columns).
- Handled missing values, standardized data formats, performed necessary **data cleaning** and ensured data consistency for analysis.

2. EXPLORATORY DATA ANALYSIS (EDA)

- Analyzed **app categories, user ratings, review patterns, and revenue distribution**.
- Applied **statistical methods** to compute key insights such as average rating, total installs, and revenue trends.

3. DATA VISUALIZATION

Developed interactive visualizations to gain deeper insights into the data:

- **Scatter Plot:**
 - Visualized the relationship between **revenue and number of installs** for **paid apps only**.
 - Added a **trendline** to highlight correlation.
 - Color-coded points based on **app categories**.
- **Grouped Bar Chart:**

- Compared **average ratings** and **total review count** for the **top 10 app categories**.
 - Applied filters: Excluded categories with an **average rating below 4.0**, app size **below 10MB**, and last update **before January**.
 - Set a **time restriction (3 PM – 5 PM IST)** to display the chart only during the allowed period.
- **Dual-Axis Chart:**
 - Compared **average installs** and **revenue** for **free vs. paid apps** within the **top 3 app categories**.
 - Applied filters: Excluded apps with **fewer than 10,000 installs**, revenue **below \$10,000**, and ensured **Android version > 4.0**, size **> 15MB**, content rating "**Everyone**", and app name **≤ 30 characters**.
 - Implemented a **time-based display restriction (1 PM – 2 PM IST)**.
- **Bubble Chart:**
 - Analyzed the relationship between **app size (in MB)** and **average rating**, with **bubble size representing the number of installs**.
 - Included a filter to display only apps with **rating > 3.5**, **installs > 50K**, and belonging to the "**Games**" category.
 - Restricted the chart to be **visible only between 5 PM – 7 PM IST**.
- **Time Series Line Chart:**
 - Plotted **total installs over time**, segmented by **app category**.
 - Highlighted **significant growth periods** where installs increased **by more than 20% month-over-month**.
 - Applied filters: Included only apps with **content rating "Teen"**, app names **starting with 'E'**, and installs **> 10K**.
 - Restricted display time to **6 PM – 9 PM IST**.

4. DASHBOARD DEVELOPMENT

- Integrated **all visualizations** into an **interactive dashboard** using **HTML and JavaScript**.

- Ensured proper **time-based filtering** to show specific charts only during their allowed periods.
- Created a **user-friendly interface** for better accessibility and insight extraction.

5. INSIGHTS & REPORTING

- Extracted meaningful **business insights** from the analysis.
- Summarized trends in **app categories, installs, ratings, and revenue distribution**.
- Documented key observations to assist in **data-driven decision-making**.

SKILLS AND COMPETENCIES

During my internship at NullClass, I developed and strengthened several technical and analytical skills, including:

1. Data Analysis & Preprocessing

- Data cleaning, handling missing values, and data transformation.
- Filtering datasets based on conditions for meaningful insights.

2. Programming & Scripting

- Writing Python scripts for data analysis and visualization.
- Using Pandas and NumPy for efficient data manipulation.

3. Data Visualization

- Creating interactive plots using Plotly.
- Implementing time-based filtering in visualizations.
- Designing effective charts such as scatter plots, grouped bar charts, dual-axis charts, bubble charts, and time series plots.

4. Dashboard Development

- Integrating charts into an interactive dashboard using HTML and JavaScript.
- Ensuring user-friendly and responsive design.

5. Problem-Solving & Logical Thinking

- Applying statistical methods to extract insights.
- Debugging and troubleshooting errors in data processing and visualization.

6. Time Management & Attention to Detail

- Effectively handling multiple tasks with specific constraints while ensuring accuracy in data representation and report writing.

FEEDBACK AND EVIDENCE

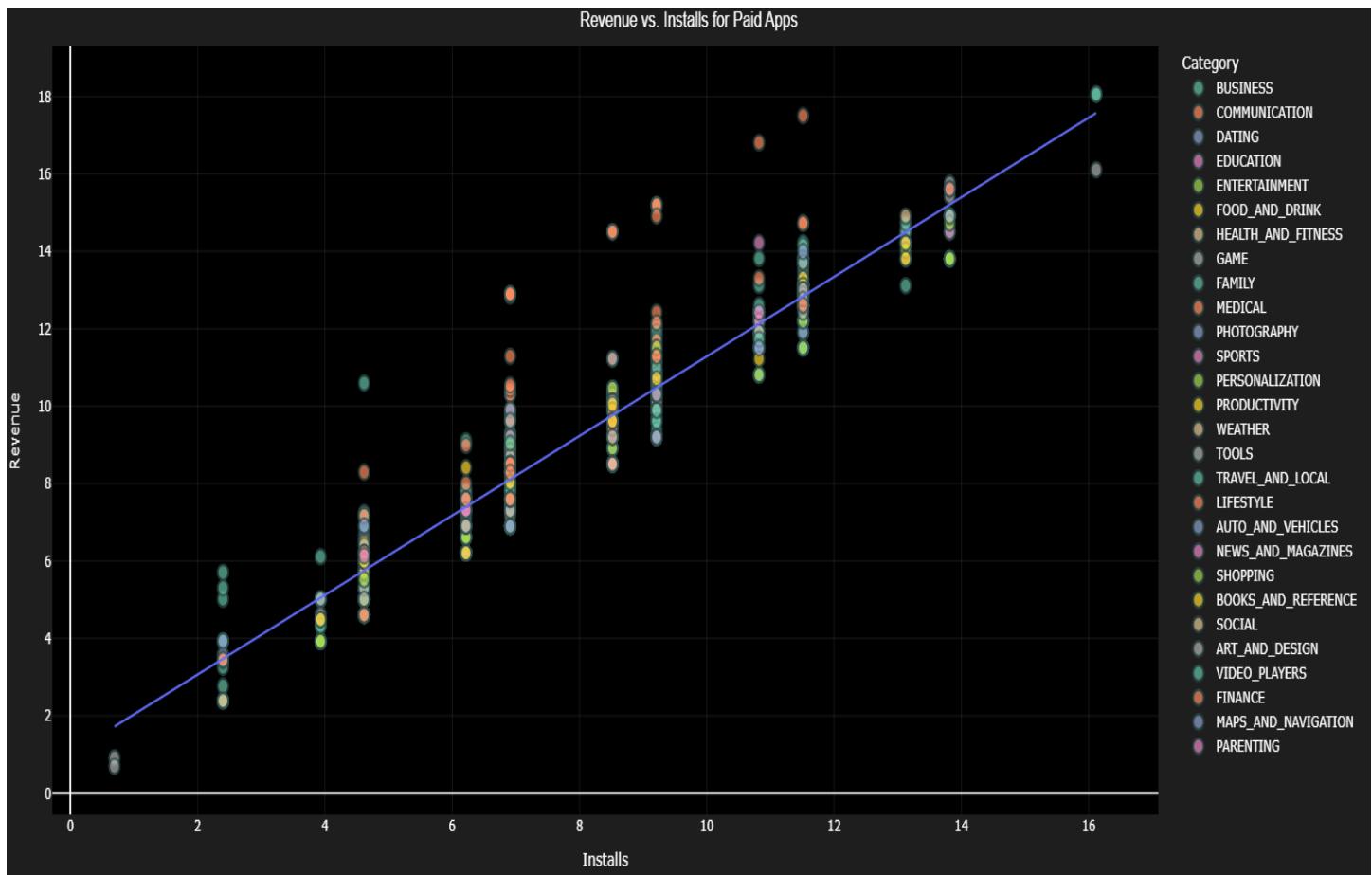
Since no direct mentor feedback was provided, this section presents **evidence of my work** through key outputs, including visualizations, and a dashboard screenshot.

1. Data Visualizations

Below are the interactive visualizations I created during the internship, each designed to provide meaningful insights into the Google Play Store dataset.

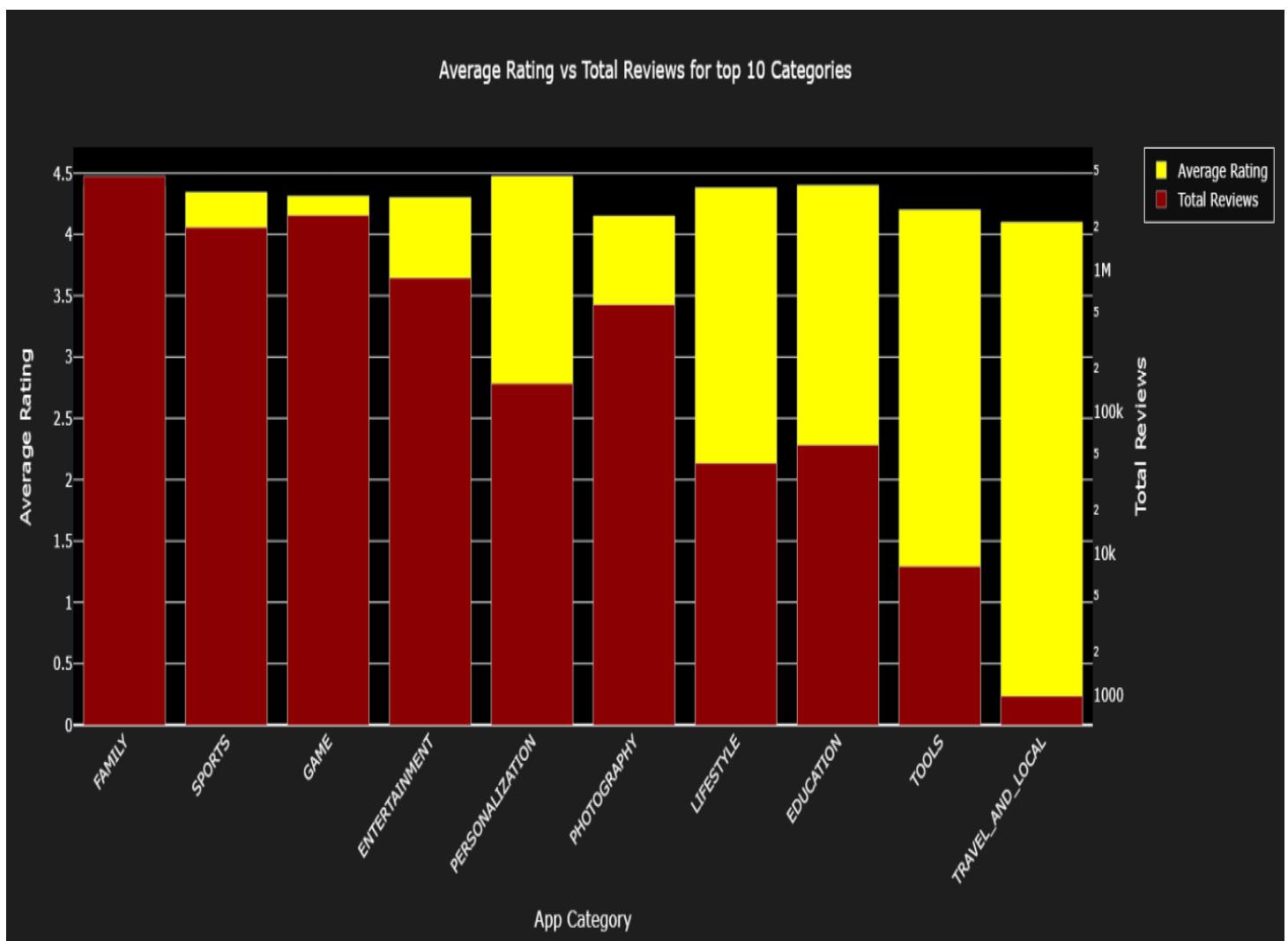
📌 Scatter Plot:

- Shows the relationship between **Revenue and Installs** for **paid apps only**.
- A **trendline** highlights the correlation, and data points are **color-coded by app category**.



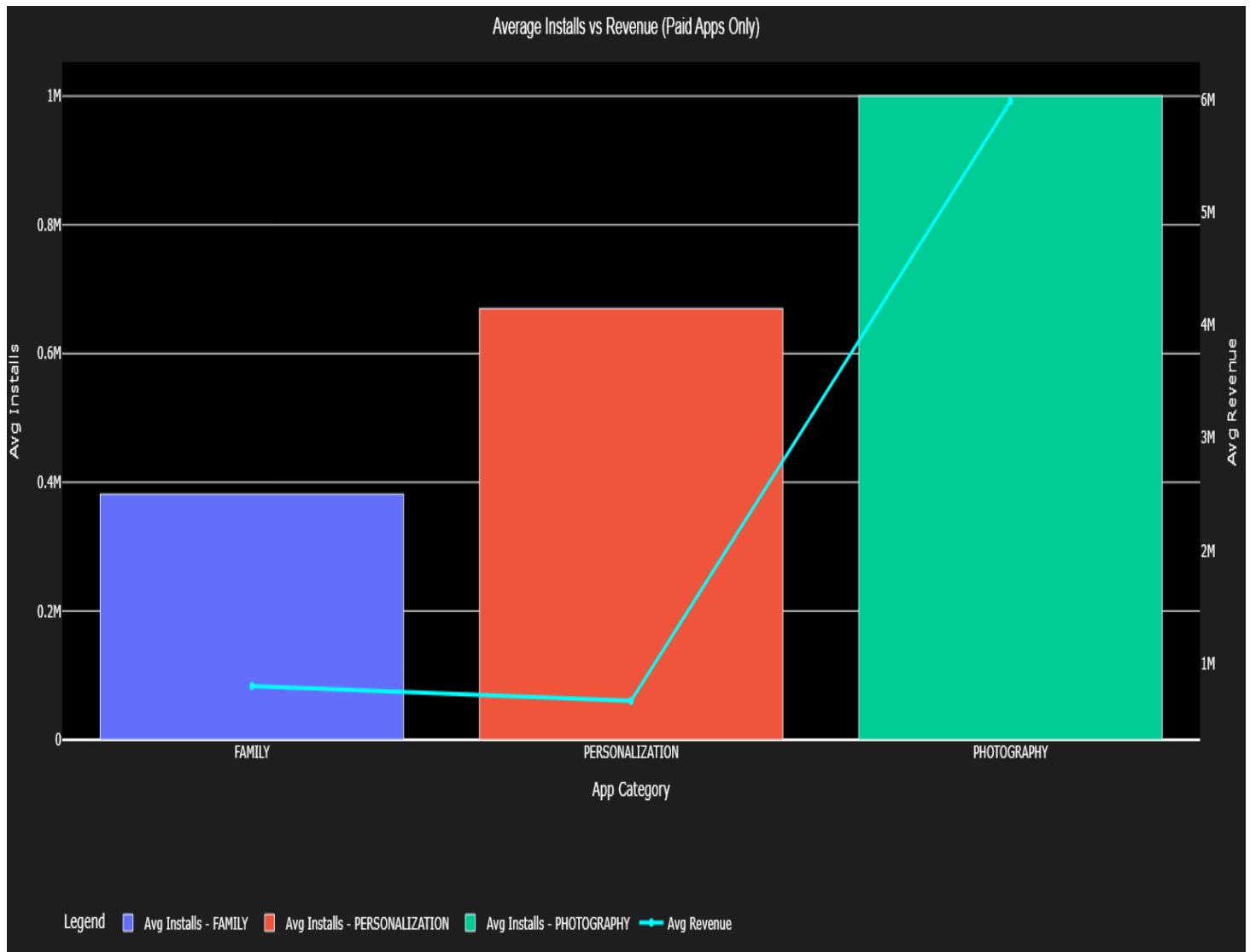
❖ Grouped Bar Chart:

- Compares the **average rating** and **total review** count for the **top 10 app categories**.
- Includes filters to exclude categories with an average rating below 4.0, app size below 10MB, and last update before January.
- Time-restricted display (visible only between **3 PM – 5 PM IST**).



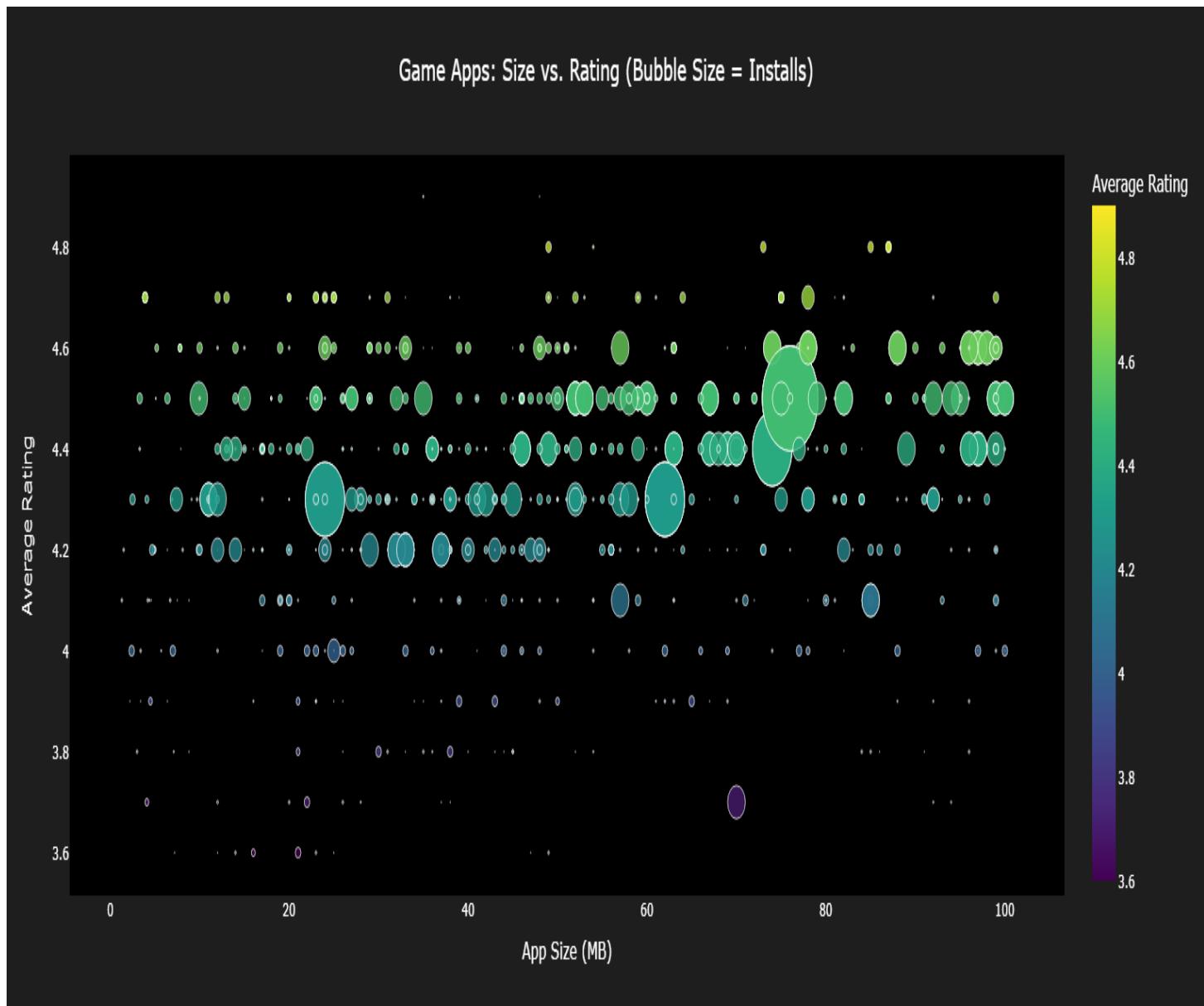
❖ Dual-Axis Chart:

- Compares **average installs and revenue** for **free vs. paid apps** in the **top 3 categories**.
- Applied filters: Installs > 10,000, Revenue > \$10,000, Android version > 4.0, size > 15MB, content rating “Everyone.”
- Display restricted to **1 PM – 2 PM IST**.



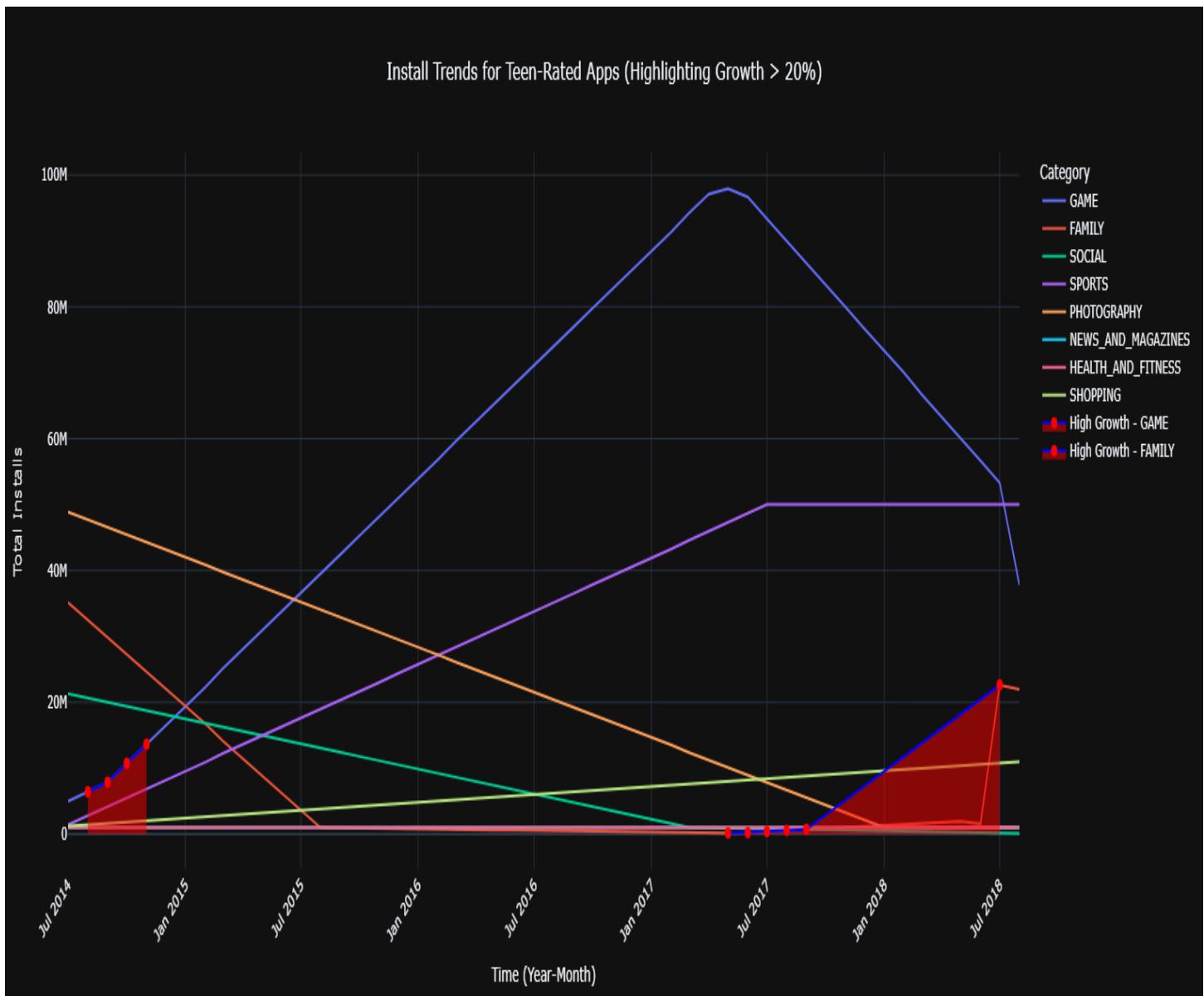
📌 Bubble Chart:

- Analyses the relationship between **app size and average rating**, with bubble size representing the **number of installs**.
- Displays only apps with a rating **above 3.5**, installs **more than 50K**, and belonging to the **Games** category.
- Time-based restriction: **5 PM – 7 PM IST**.



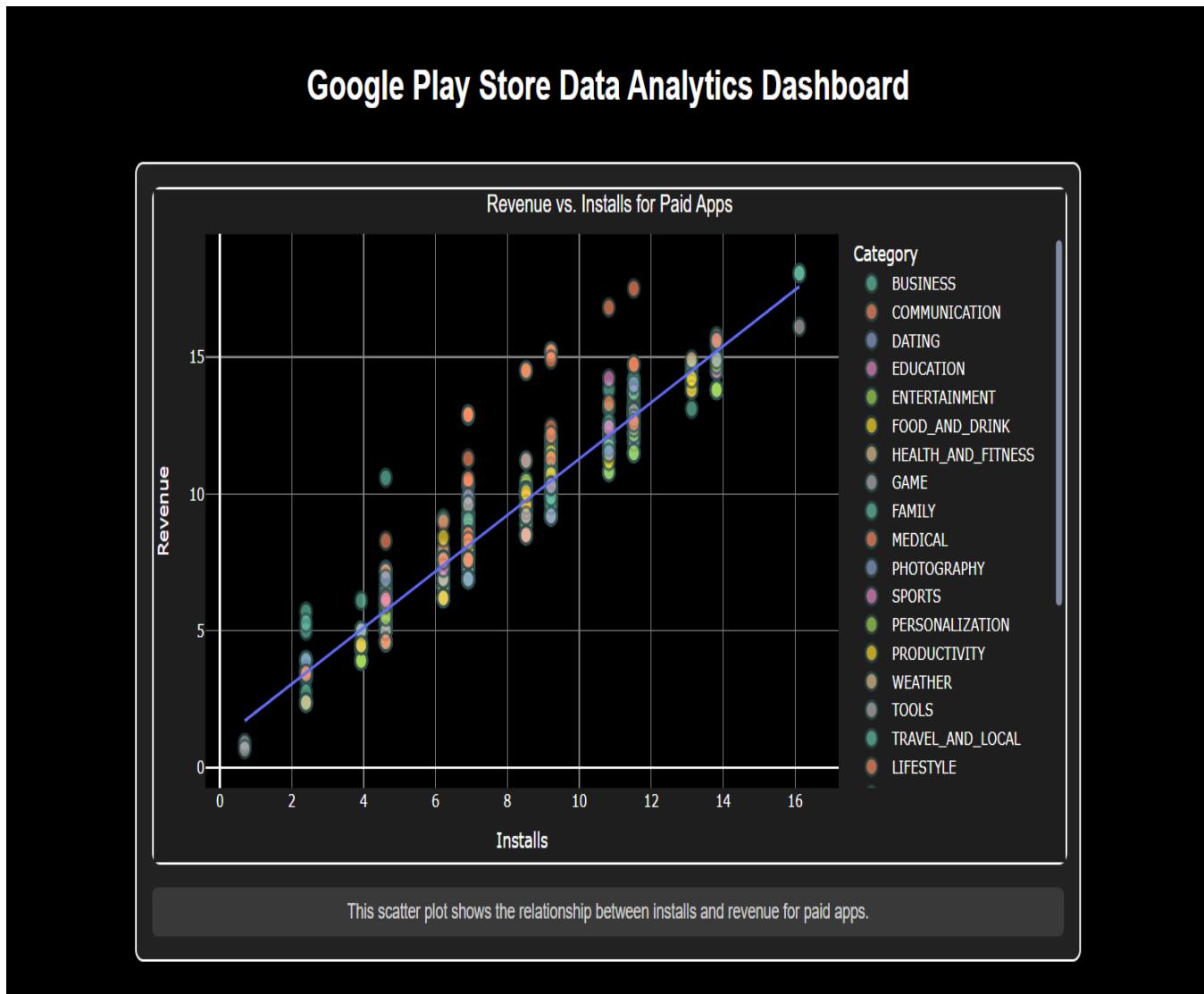
📌 Time Series Line Chart:

- Shows **install trends over time**, segmented by app category.
- Highlights significant growth periods (**install increase > 20% month-over-month**).
- Filters: Apps with **Teen content rating**, name starting with ‘E’, and installs **above 10K**.
- Visible only from **6 PM – 9 PM IST**.



2. Interactive Dashboard Interface

An interactive dashboard was created using **HTML & JavaScript** to present insights dynamically. However, due to **time-based filtering**, all plots do not appear at the same time. The following screenshot represents the **dashboard layout** and structure:



Note: Certain visualizations only appear within their designated time windows (as implemented in the logic).

CHALLENGES AND SOLUTIONS

During my internship at NullClass, I encountered several challenges while working on data analysis, visualization, and dashboard development. Overcoming these hurdles significantly contributed to my learning and problem-solving skills. Below are the key challenges I faced and the solutions I implemented:

1. LEARNING PLOTLY FOR DATA VISUALIZATION

- **Challenge:** Initially, I struggled with using Plotly to create interactive visualizations as I had limited prior experience with it. Understanding its syntax, layout configurations, and customization options was challenging.
- **Solution:** I referred to the official Plotly documentation and tutorials, experimented with different parameters, and gradually improved my understanding. By implementing multiple charts, I became more comfortable with its functionalities.

2. IMPLEMENTING TIME-BASED FILTERING IN PLOTLY

- **Challenge:** One of the major requirements was to display visualizations only within specific time slots. Managing this condition dynamically in Plotly was complex.
- **Solution:** I explored JavaScript-based time filtering techniques and implemented a logic that checked the current system time before rendering the graphs, ensuring that they appeared only during the designated periods.

3. ADDING A TRENDLINE IN SCATTER PLOT (TASK 1)

- **Challenge:** While creating the scatter plot to analyze revenue vs. installs, I faced issues in correctly adding a trendline to visualize the correlation.

- **Solution:** I researched different approaches and used the "**trendline**" parameter in **Plotly Express**, which helped me correctly fit and display the trendline for better interpretation of the data.

4. HANDLING DATA FILTERING IN DUAL-AXIS CHART (TASK 3)

- **Challenge:** Creating a **dual-axis chart** comparing installs and revenue for free vs. paid apps was one of the most difficult tasks. The challenge arose because free apps naturally have **\$0 revenue**, and since the filter required **revenue > \$10,000**, all free apps were automatically excluded.
- **Solution:** I analyzed the filtering conditions and discovered that the exclusion of free apps was an inherent result of the data preprocessing step. Instead of manually removing free apps, I simply ensured that the final dataset aligned with the expected filtering criteria. This helped in maintaining the integrity of the analysis.

5. SHADING THE GROWTH TREND IN TIME SERIES CHART (TASK 5)

- **Challenge:** In the time series visualization, I had to **highlight periods where installs increased by more than 20% month-over-month** by shading those regions. Implementing this logic dynamically in Plotly was difficult.
- **Solution:** I computed the percentage increase in installs for each month and used conditional formatting to shade the regions where the growth exceeded 20%. After multiple iterations and debugging, I successfully applied the shaded regions to highlight significant growth trends.

6. DASHBOARD DEVELOPMENT WITH HTML AND JAVASCRIPT

- **Challenge:** I had no prior experience with HTML and JavaScript, making dashboard development a major hurdle. Integrating the visualizations and ensuring the proper layout, responsiveness, and functionality was difficult.
- **Solution:** I learned the basics of **HTML, CSS, and JavaScript** through online resources and applied my knowledge step by step. I used **Plotly's JavaScript**

library to embed the charts and ensured that all filtering conditions and time-based visibility worked correctly.

OUTCOME:

Overcoming these challenges helped me gain a deeper understanding of **data visualization, dashboard development, and time-based filtering techniques**. I also strengthened my problem-solving abilities, debugging skills, and ability to work with new technologies efficiently.

OUTCOME AND IMPACT

Successfully completing this internship allowed me to achieve significant learning and skill development in various areas, including:

1. Enhanced Technical Proficiency:

- Gained in-depth knowledge of **Plotly** for interactive visualizations.
- Learned to implement **time-based filtering** and advanced chart customizations.
- Developed skills in **HTML and JavaScript** for dashboard integration.

2. Strengthened Problem-Solving Abilities:

- Overcame challenges in **data filtering, trendline implementation, and dual-axis visualization**.
- Applied **logical thinking** to resolve errors in data representation and dashboard functionality.

3. Improved Debugging and Analytical Thinking:

- Debugged complex **data preprocessing and visualization** errors.
- Applied **statistical insights** to make data-driven decisions effectively.

4. Practical Experience in Real-World Data Analysis:

- Worked with real-world Google Play Store data, handling **large datasets, missing values, and category-based filtering**.
- Created impactful **dashboards and visual reports** to extract meaningful business insights.

5. Increased Confidence in New Technologies:

- Adapted quickly to **new tools and programming techniques** with minimal prior experience.
- Strengthened my ability to learn independently and apply **innovative solutions** to complex tasks.

This experience has prepared me to confidently tackle real-world **data analytics, visualization, and dashboard development** challenges in future roles.

CONCLUSION

My internship at **NullClass** was a transformative learning experience that significantly enhanced my technical and analytical skills. Throughout this journey, I worked with real-world datasets, tackled complex data visualization challenges, and developed an interactive dashboard using **Plotly, Python, HTML, and JavaScript**. The hands-on exposure to **data preprocessing, filtering, and time-based visualization techniques** has strengthened my ability to derive meaningful insights from raw data.

Despite initial challenges, such as **implementing dual-axis charts, trendline calculations, and time-restricted visualizations**, I successfully navigated them through persistent problem-solving and continuous learning. This experience has not only improved my **technical expertise** but also boosted my **confidence in handling new technologies and working independently on data-driven projects**.

Overall, this internship has been a valuable stepping stone in my journey toward becoming a **skilled data analyst and visualization expert**, equipping me with the practical knowledge required to thrive in the field of **data science and business intelligence**. I look forward to applying these learnings to future projects and continuously expanding my expertise in the domain.