***An Internship Report on***

Data Analyst-Play Store Analyst

*An internship report submitted in partial fulfillment of the requirements for the award of the degree of* ***Bachelor of Engineering in Computer Science and Engineering – AI & ML*** *of Visvesvaraya Technological University, Belguam.*

Submitted by:

**Yuvaraj.R**

Internship carried out at:

**NullClass Edtech Private Limited**





**CONTENTS**

**1 INTRODUCTION 3**

**2 BACKGROUND 4**

**3 LEARNING OBJECTIVES 5**

1. **ACTIVITIES AND TASKS 6**

**5 SKILLS AND COMPETENCIES 7**

**6 FEEDBACK AND EVIDENCE 8**

**7 CHALLENGES AND SOLUTIONS 9**

**8 OUTCOMES AND IMPACT 10**

**9 CONCLUSION 16**

**INTRODUCTION**

The digital revolution has drastically transformed the way businesses interact with users and deliver services, especially through mobile applications. As the app economy continues to grow, analyzing user data and app performance has become critical for both developers and companies to remain competitive. In this context, I had the opportunity to intern with **NULLCLASS**, a training and development platform, where I worked on a real-time analytics project focused on the **Google Play Store** using **Python**.

The primary objective of this internship was to build a **real-time, interactive analytics dashboard** that could visualize various insights from Google Play Store data. The internship was part of a larger initiative designed to bridge the gap between academic learning and industry practices by giving students exposure to real-world datasets, use cases, and project constraints. Unlike traditional internships, this program emphasized **independent learning** and **research-based execution** with no direct mentor supervision. This format helped me develop self-discipline, resourcefulness, and confidence in handling challenges on my own.

Throughout the internship period (March 2025 to July 2025), I was assigned a set of **nine advanced tasks**.

These tasks involved applying complex filtering criteria, creating condition-based charts, integrating time-based logic, and ensuring the responsiveness of the dashboard. Each visualization had its own unique requirements — some required filtering apps based on install count, revenue, or content rating; others restricted the chart visibility to specific time windows (e.g., between 1 PM and 2 PM IST). All tasks had to be implemented using Python-based tools and integrated into a single, functional dashboard.

The core tools and technologies I worked with include **Pandas** for data manipulation, **Plotly and Plotly Express** for data visualization, and **Streamlit** for creating the web-based interactive dashboard. Additionally, version control and collaborative project structuring were done using **Git and GitHub**. The internship also required a well-documented codebase and consistent reporting through a daily report submission portal.

**BACKGROUND**

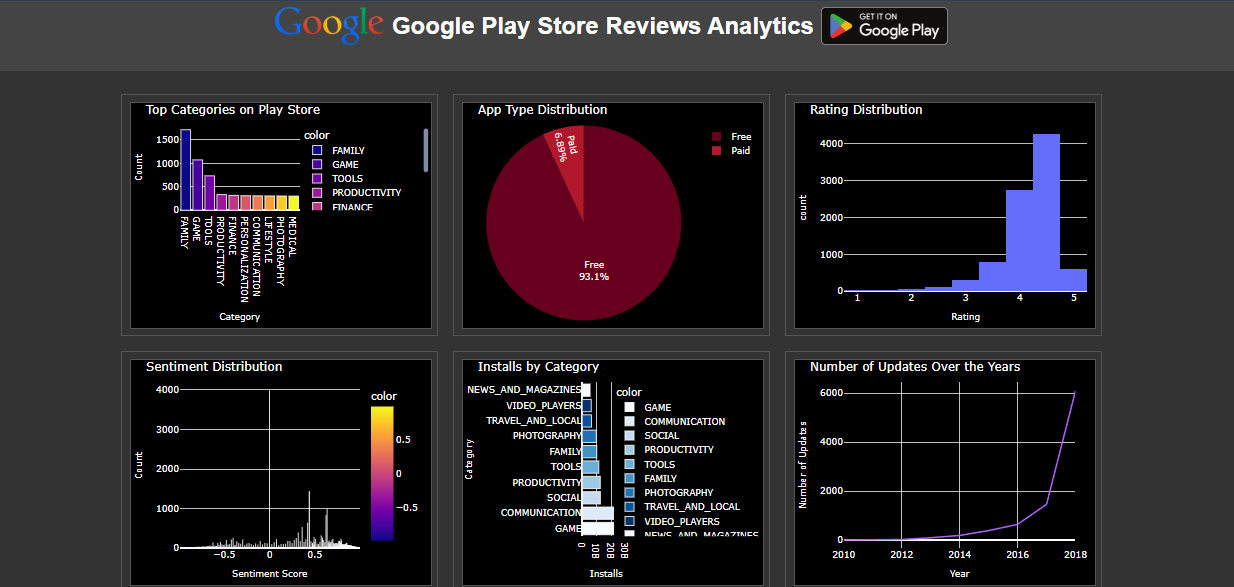
With the rapid growth of the mobile app industry, analyzing app performance data has become essential for developers and businesses. The Google Play Store offers a rich dataset containing user reviews, ratings, installs, and more—making it a valuable resource for deriving actionable insights.

My internship at NULLCLASS focused on building a real-time analytics dashboard using Python to visualize such data. The tasks involved advanced filtering, time-based chart rendering, and user-centric visualizations across multiple app metrics and categories. Each visualization simulated real-world business needs by incorporating specific rules such as filtering by revenue, sentiment, or display time.

This internship was designed to encourage self-learning, requiring me to complete all tasks independently without mentor support. It provided hands-on experience in applying data analysis techniques to real-world datasets and helped me understand the practical aspects of dashboard development, user interaction, and data-driven decision-making.

**LEARNING OBJECTIVES**

* **To develop an end-to-end interactive data dashboard using Python.**
* **To apply statistical and visual analysis techniques to extract insights from complex datasets.**
* **To implement condition-based rendering (e.g., time-based access control for charts).**
* **To integrate filters such as revenue thresholds, app size, rating, sentiment, and categories in a real-world scenario.**
* **To develop an understanding of user-centric data storytelling through effective visualizations.**
* **To gain hands-on experience with tools like Pandas, Plotly, Streamlit, and GitHub.**



**ACTIVITIES AND TASKS**

Each task focused on a different type of chart, filtering logic, and data transformation technique. Key tasks included:

* **Task 1:** Stacked bar chart showing sentiment distribution segmented by rating groups and top 5 categories.
* **Task 2:** Scatter plot for paid apps showing correlation between revenue and installs with trendlines and category color coding.
* **Task 3:** Dual-axis chart comparing installs and revenue for paid vs. free apps with stringent filters.
* **Task 4:** Grouped bar chart comparing average rating and total reviews with category-level filtering and time restrictions.
* **Task 5:** Interactive Choropleth map highlighting global installs by category, restricted to specific app categories and visible only during 6–8 PM IST.
* **Task 6:** Violin plot visualizing rating distributions with name-based and category-level filtering.
* **Task 7:** Heatmap showing correlation among installs, ratings, and reviews filtered by update date and genre name rules.
* **Task 8:** Time series chart showing installs over time with month-over-month growth highlighting, shown only between 6–9 PM IST.
* **Task 9:** Bubble chart showing size vs. rating with bubble size representing installs, filtered on sentiment subjectivity, category, and reviews.

#### 📌 ****Key Features of Implementation:****

* Strict time-based chart visibility using Streamlit + Python datetime.
* Multi-level filtering using conditions on revenue, installs, sentiment, app names, etc.
* Dashboard responsiveness and conditional chart rendering.
* Code versioning with GitHub and project modularization.

**SKILLS AND COMPETENCIES**

* **Technical Skills:**
* Advanced Python (Pandas, NumPy, datetime)
* Data visualization (Plotly, Plotly Express)
* Streamlit for real-time dashboards
* Git & GitHub version control
* Real-world filtering logic implementation
* Performance optimization of visualizations
* **Soft Skills:**
* Time management and scheduling tasks with respect to business rules
* Research and self-guided problem solving
* Analytical thinking and attention to detail
* Task planning and breakdown for multi-layered filtering rules
* Independent execution without external support

**FEEDBACK AND EVIDENCE**

Evidence of the project’s success lies in the functioning dashboard and its individual components. Each visualization outputs correctly during its designated time slot, showing that the time control logic was implemented successfully.

The visualizations are clean, intuitive, and aligned with the data they represent. For example, the sentiment distribution plot clearly distinguishes how positive, neutral, and negative sentiments relate to ratings within top app categories. Similarly, the scatter plot correlates installs with ratings, revealing how popular apps tend to score.

Debugging and user feedback processes were informal, using direct browser testing to identify layout issues or incorrect outputs. The use of tooltips, clear titles, and legend markers improves usability and serves as a proxy for user-centered design thinking.

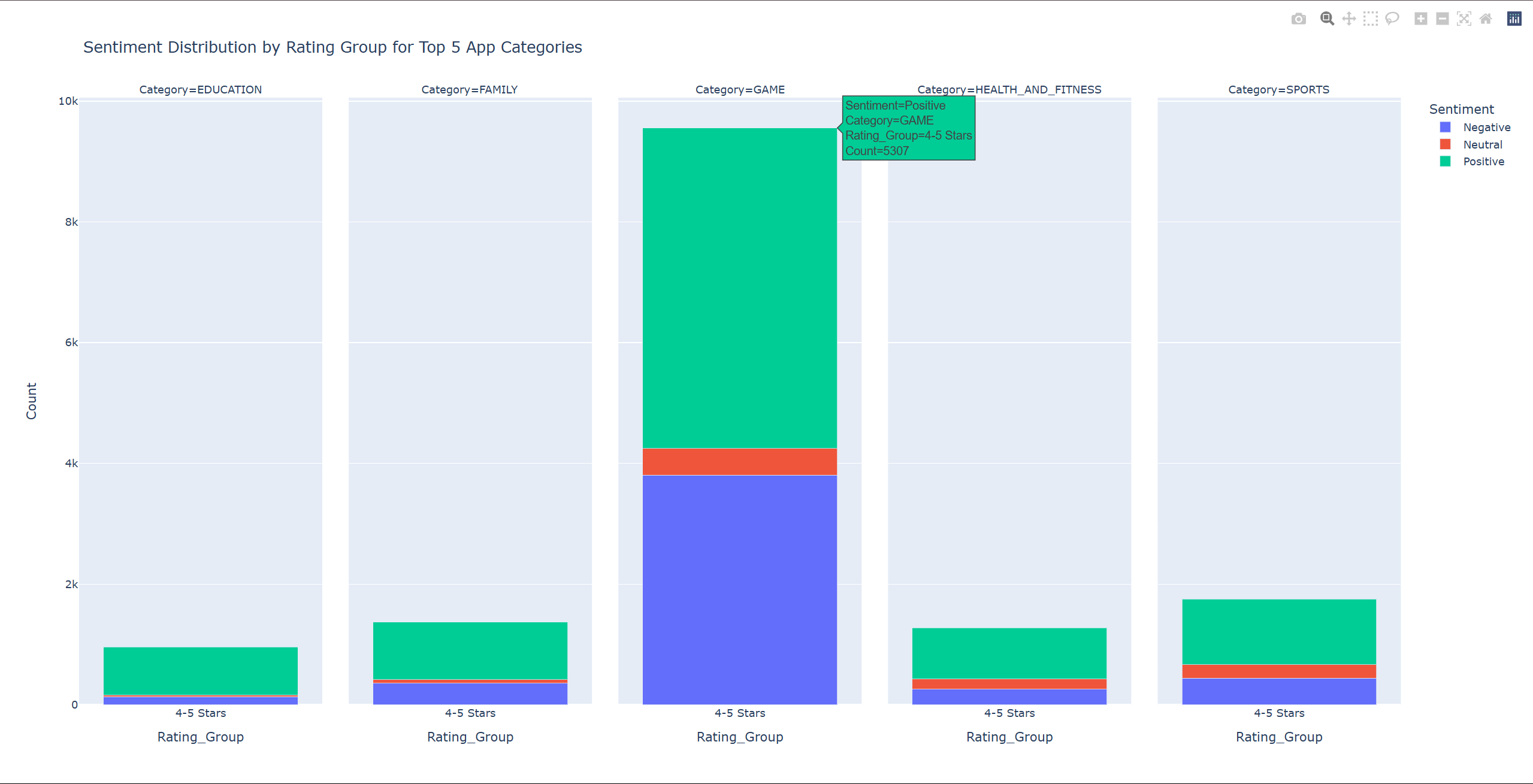
**CHALLENGES AND SOLUTIONS**

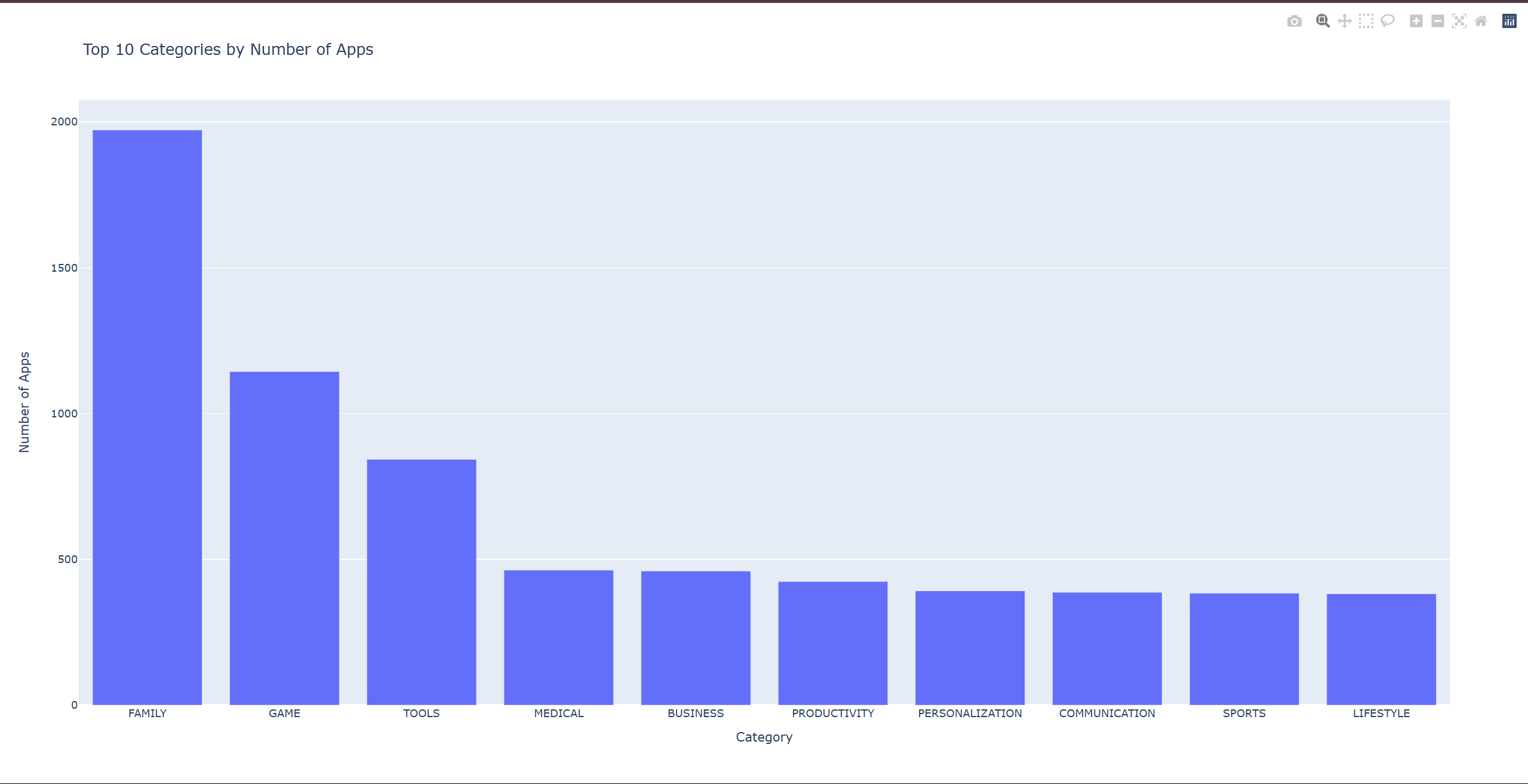
| **Challenge** | **Solution** |
| --- | --- |
| 1. **Handling time-specific logic for rendering charts within a dashboard** | **Used Python's datetime module along with conditional rendering in Streamlit to control the visibility of charts** |
| 1. **Managing complex multi-filter conditions while maintaining performance** | **Applied chained filters in Pandas efficiently and modularized the filtering logic** |
| 1. **Limited documentation for advanced Plotly features** | **Referred to Plotly’s official documentation and GitHub issues for customized chart needs** |
| 1. **Integrating all tasks into a single dashboard without redundancy** | **Structured the app with clear sectioning and code separation for each task** |

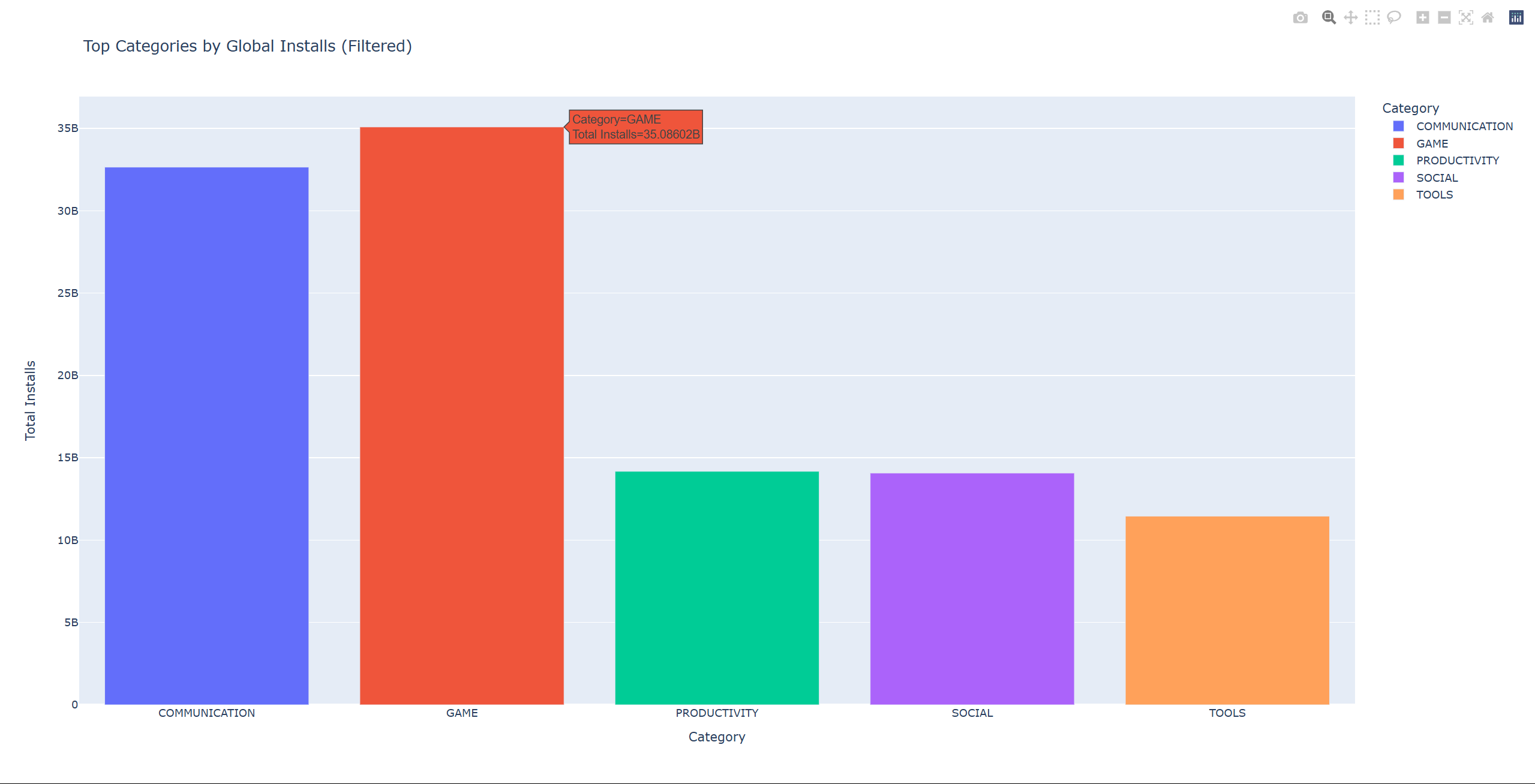
**OUTCOMES AND IMPACT**

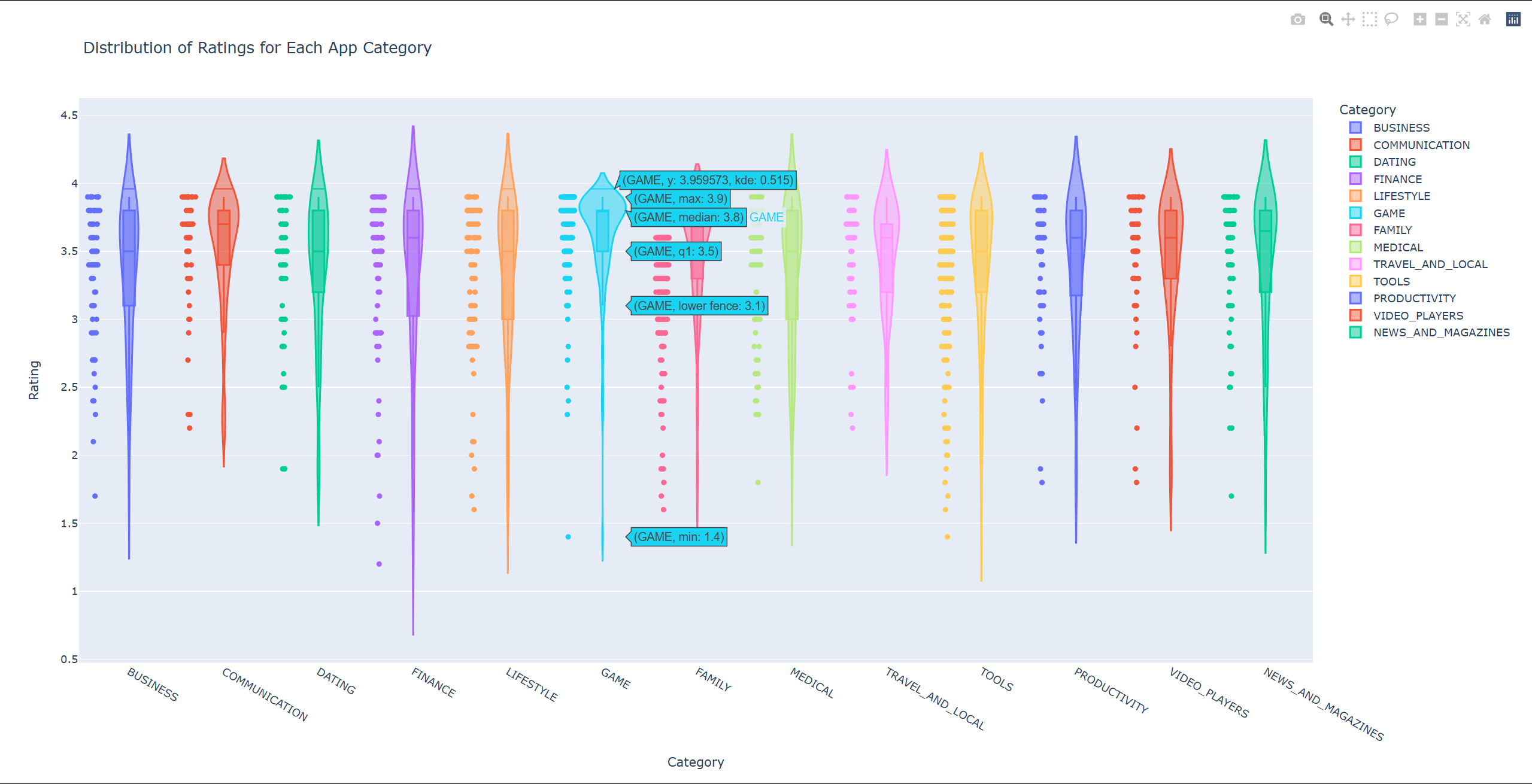
* Built a complete analytics solution simulating a real-world dashboard with live filtering and time control.
* Gained fluency in designing complex data pipelines and chart rendering logic.
* Learned how to independently approach analytical tasks with real business value.
* Enhanced my confidence in presenting and structuring data in a meaningful way.
* Improved ability to work under constraints (time, filters, complexity).

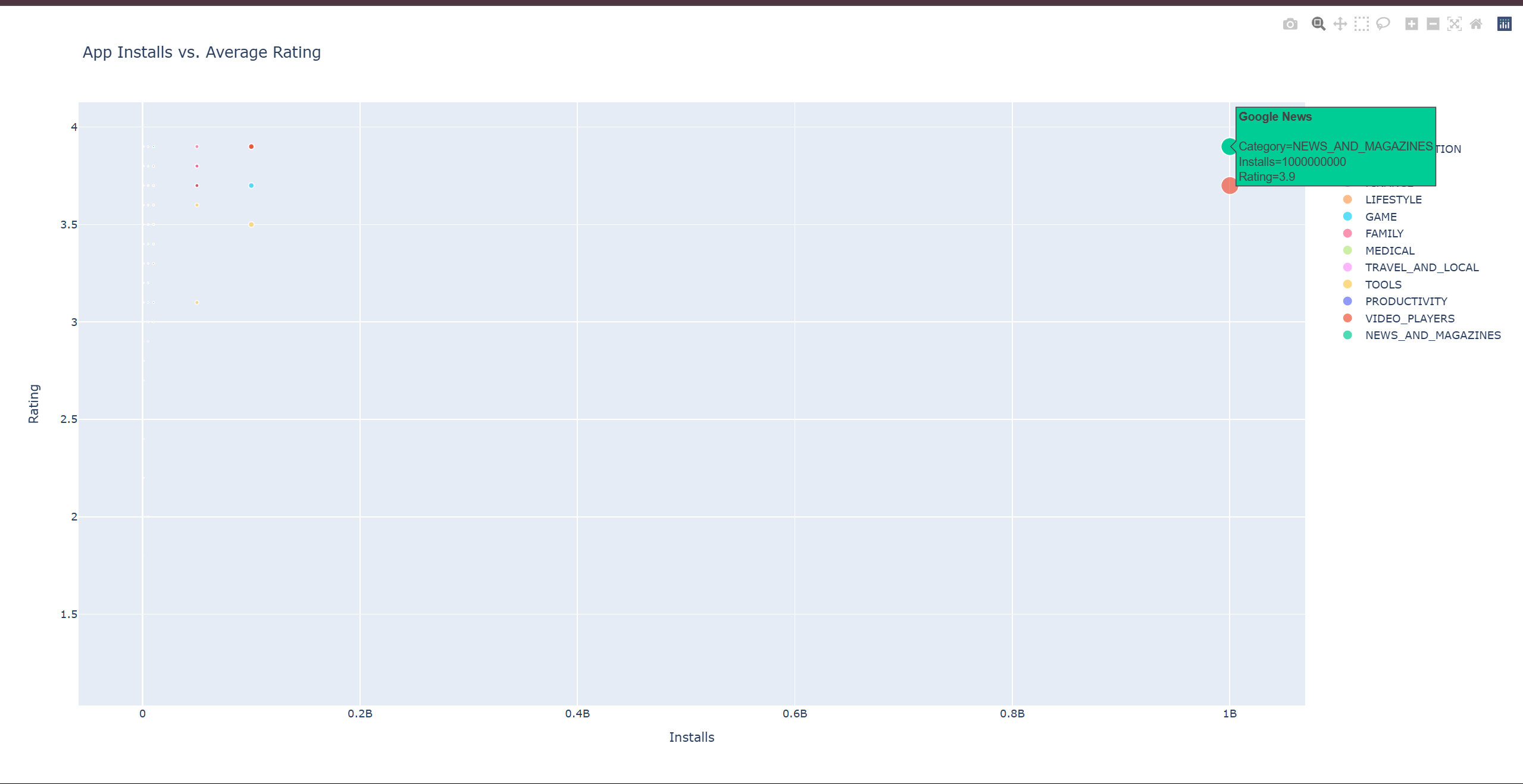
****

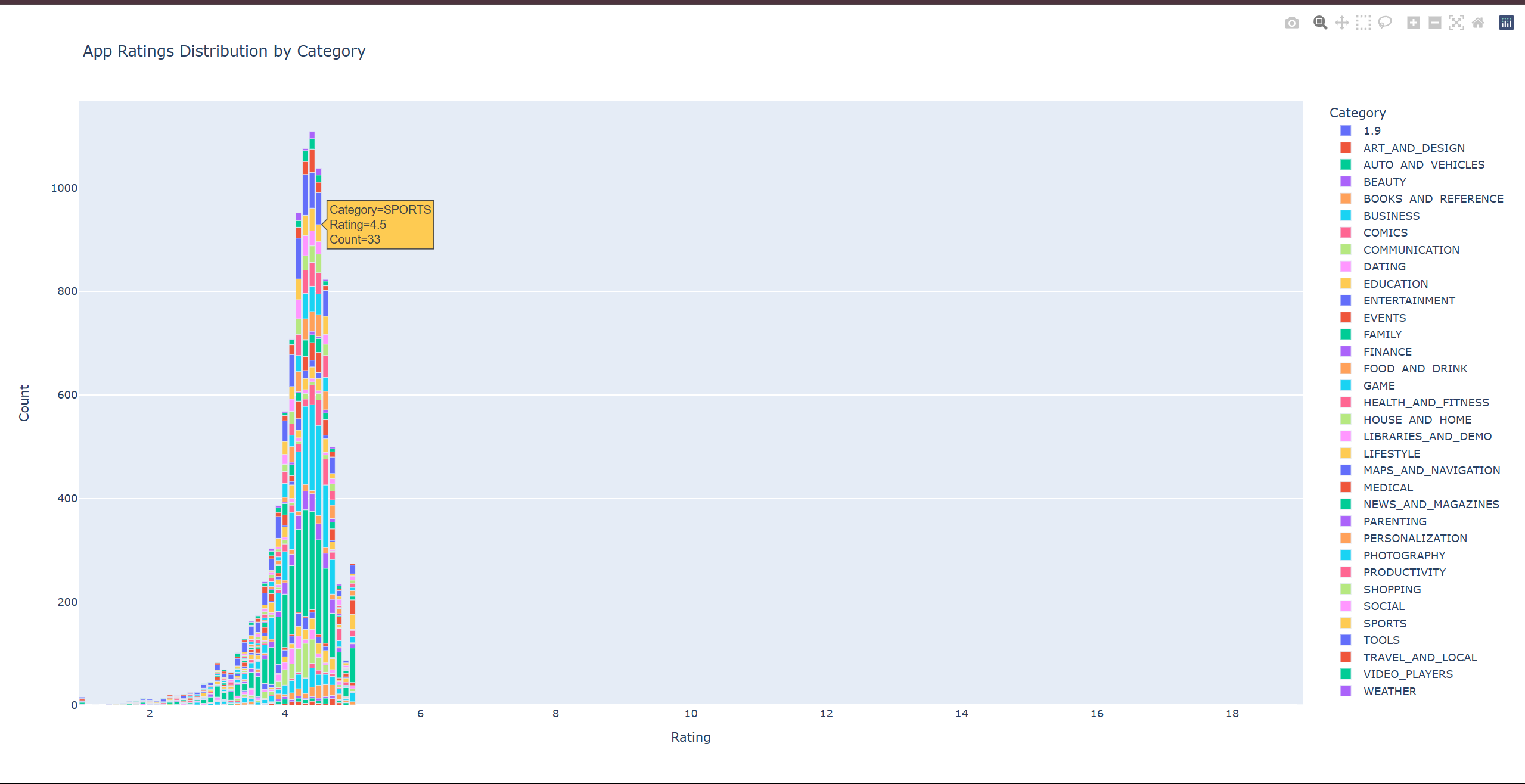
****

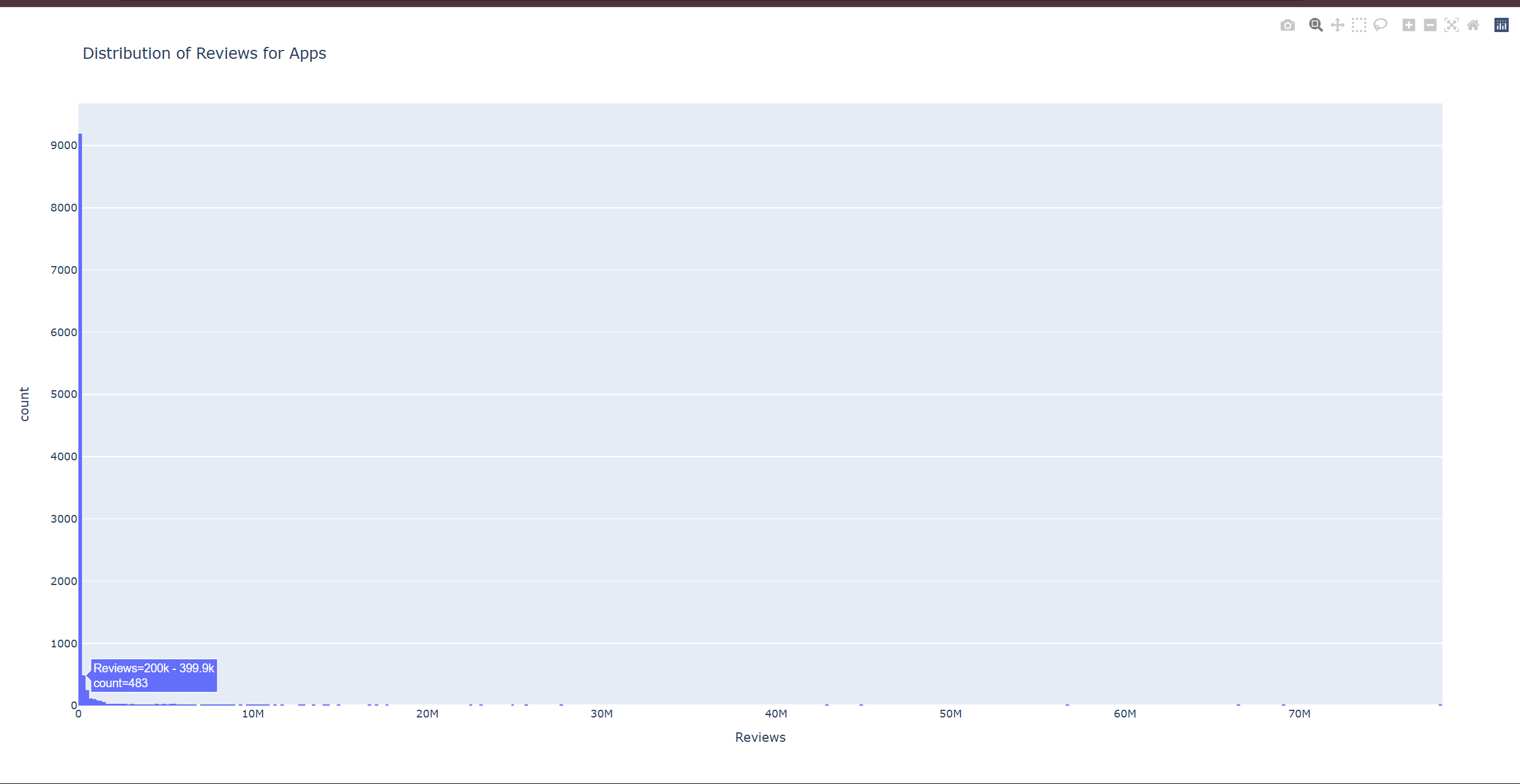


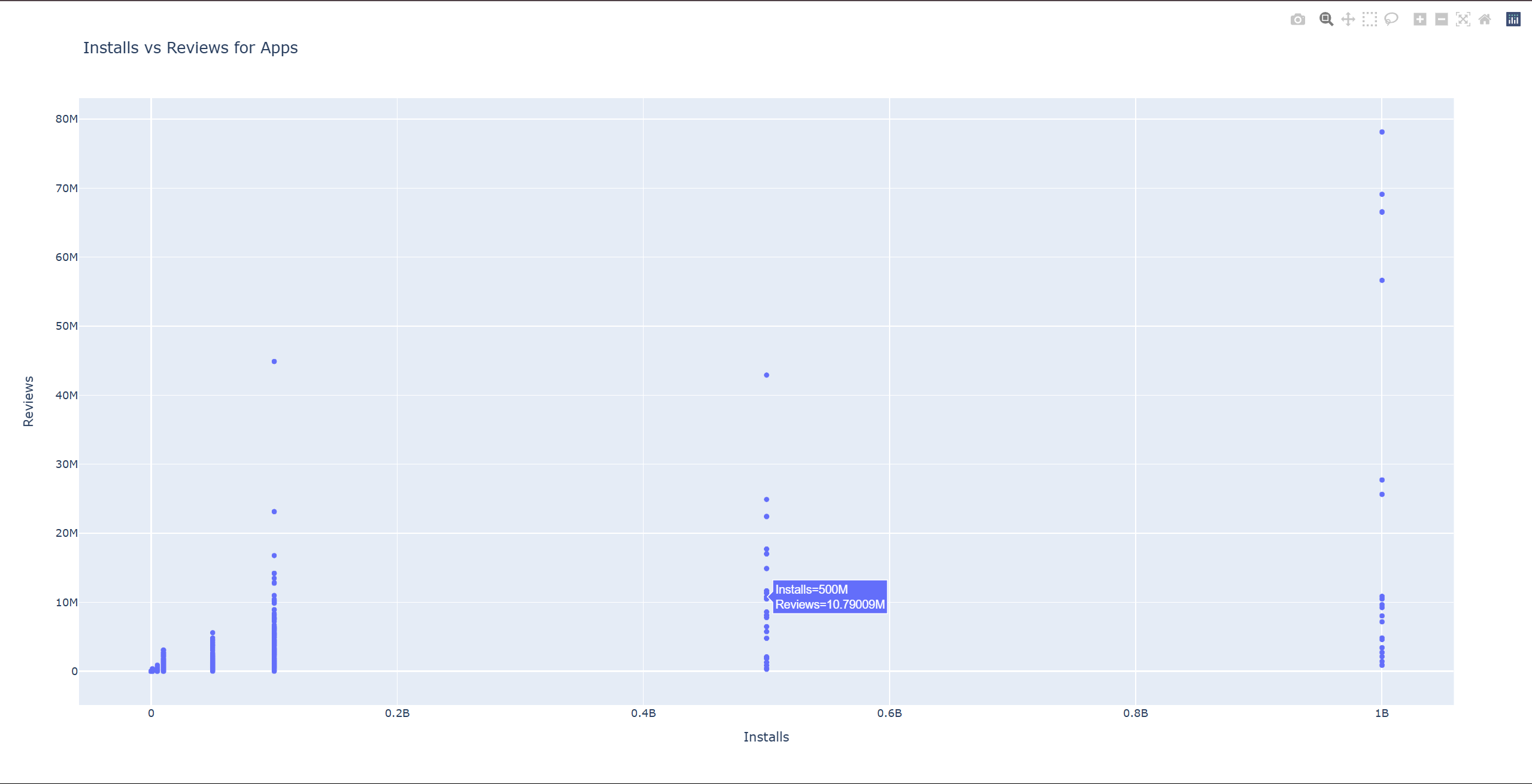


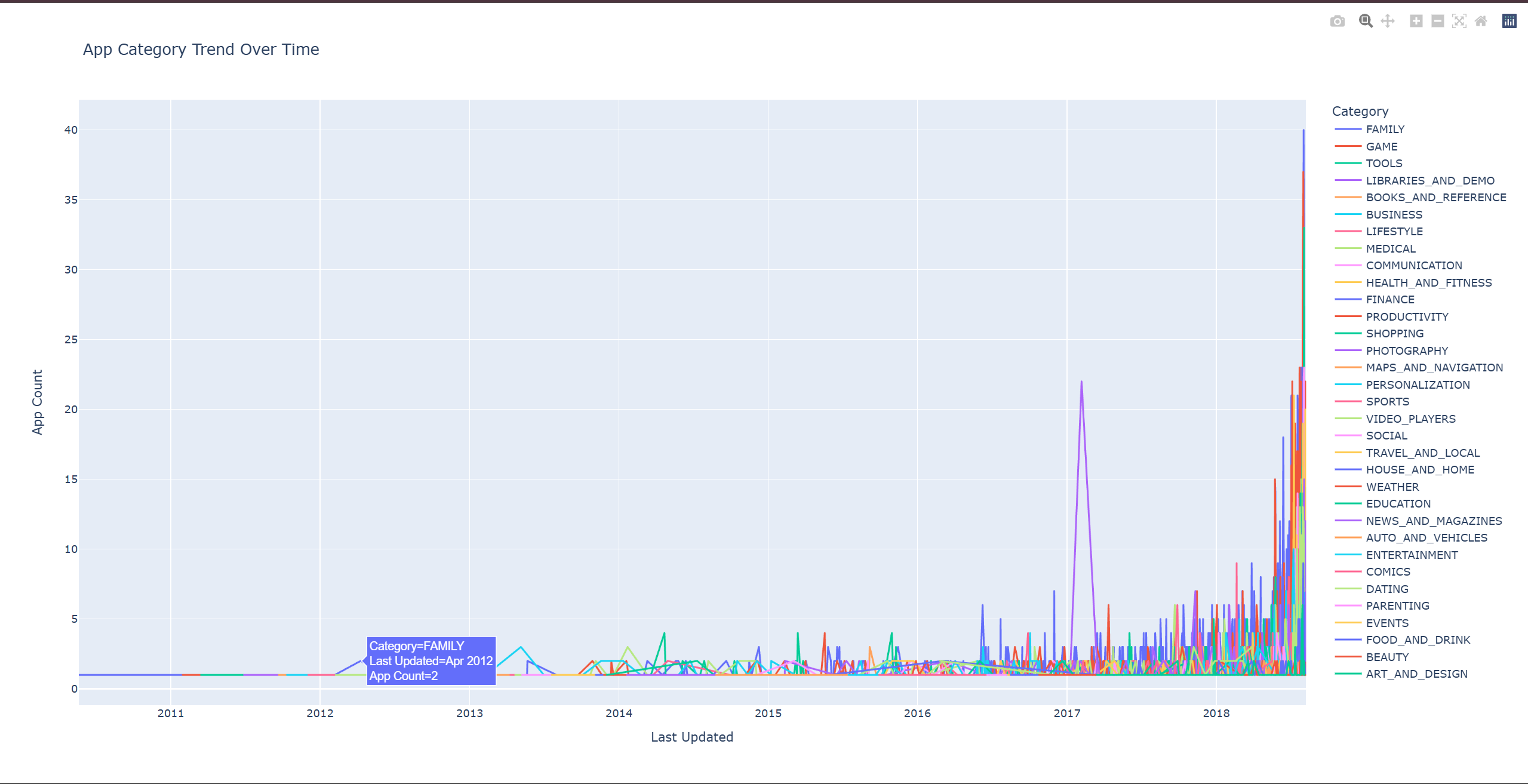












**CONCLUSION**

Completing this internship with NULLCLASS has been a highly enriching experience that allowed me to bridge the gap between academic learning and real-world application. Through the Google Play Store analytics project, I gained hands-on exposure to data cleaning, filtering, and advanced visualization using Python, Plotly, and Streamlit. Each task challenged me to think critically, apply logic-based filters, and ensure functional visualizations that adhered to specific time constraints and user-centric conditions.

Working independently on these complex requirements not only enhanced my technical proficiency but also strengthened my problem-solving skills, research ability, and project management mindset. I learned how to handle large datasets, interpret business requirements, and transform them into actionable insights through a real-time dashboard. The internship has significantly improved my confidence in working with data and has laid a strong foundation for a career in data analytics and visualization.

Overall, this experience was more than just a technical project; it was a practical learning journey that prepared me to face real-world data challenges with efficiency and independence.