**INTERNSHIP REPORT**

**ON**

**DATA ANALYTICS IN GOOGLE PLAYSTORE DATA**

**AT**

****

**NULL CLASSES PRIVATED LIMITED**

**INTERNSHIP DURATION: 1 MONTH**

**(1-07-2025 TO 30-07-2025)**

**SUBMITTED BY:**

**ANUJ YADAV**

**CONTACT DETAILS:**

**9098663187**

**Anujyadavitm178@gmail.com**

**Declaration**

I, Anuj Yadav, hereby declare that the internship report titled:

“data analytics in google playstore data” is the result of my own work, carried out during the internship period.

I further declare that this report has not been submitted to any other university or institution for the award of any degree , diploma, certificate.

All sources of information and data used in this report have been acknowledged, and any similarities with other work are purely coincidental .

This work is a true representation of my efforts, learning, and experiences gained during the internship period from 1-07-2025 to 30-07-2025.

**Acknowledgement**

I would like to express my deepest gratitude to Null Classes for giving me the opportunity to work as an intern and gain practical exposure in the field of data analytics and visualization. This internship has played a crucial role in enhancing my technical and professional skills.

I am especially thankful to my mentor and the entire team at Null Classes for their continuous support, guidance, and constructive feedback throughout the internship period. Their mentorship helped me gain a deeper understanding of real-world dashboard development, time-based access control, and interactive data representation using Python, JavaScript, and modern visualization libraries.

Working at Null Classes has been a valuable and enriching experience, and I am truly grateful for the opportunity.

**Table of contents**

|  |  |  |
| --- | --- | --- |
| **s.no** | **contents** | **Pg.no** |
| 1 | introduction | 1 |
| 2 | Background knowledge | 2 |
| 3 | Learning objectives | 3-4 |
| 4 | Activities and tasks | 5-7 |
| 5 | Skills and competencies | 8-9 |
| 6 | Feedback and evidences | 10-14 |
| 7 | Challenges and solutions | 15-16 |
| 8 | Outcomes and impact | 17-18 |
| 9 | conclusion | 19 |

**Introduction**

In today’s fast-paced digital economy, mobile applications have become an integral part of everyday life, powering services from entertainment to business management. As a result, understanding app performance metrics such as user ratings, installs, revenue generation, and user engagement has become essential for developers, marketers, and decision-makers.

This internship project aimed to design and implement the interactive visualization dashboards using Plotly in Python that helps analyze google playstore app data with precision and adaptability. Unlike static dashboards, these solutions integrates dynamic filtering based on category, user engagement metrics, time zones, and even sentiment analysis. Moreover, the project incorporates localized translations and time-sensitive chart visibility logic, enhancing both usability and relevance.

The visualizations created in this project include a scatter plot to understand the relationship between revenue and number of installs for paid apps, a grouped bar chart comparing ratings and reviews across the most downloaded categories with strict display-time filtering in dashboard itself only, and a complex bubble chart analysing app size and rating with multilingual support and conditional formatting.

A unique aspect of this work is its time-based dashboard logic: certain graphs are only displayed during specific time windows according to Indian Standard Time (IST), mimicking real-world dashboard access restrictions . Furthermore, the inclusion of natural language filters (such as excluding apps with certain letters, or including only apps with high sentiment subjectivity) adds depth and realism to the filtering logic.

Through this internship project, I have sought not only to build engaging visualizations but also to bridge the gap between raw data and informed decision-making. By integrating business logic, multilingual support, and intelligent filtering into the visualization layer, these dashboards serves as a scalable and extensible prototype for advanced analytics systems in the google playstore application ecosystem.

**Background**

In recent years, the mobile application industry has experienced tremendous growth. With millions of apps available across various categories like Games, Communication, Entertainment, Beauty, Business, and more, it has become increasingly important for developers and companies to understand how their apps are performing in the market. Key metrics such as installs, revenue, ratings, reviews, and user sentiment provide valuable insights—but only when they are analysed and presented in a meaningful way. This is where data visualization plays a critical role.

The main idea behind this project was to take real-world app data and transform it into interactive and informative charts that could help identify trends and patterns. Instead of simply displaying raw numbers, I wanted to create visual stories that could be explored and understood easily by stakeholders. For this, I used Python and Plotly ,html and javascript, powerful tools that allow the creation of highly interactive and customizable visualization dashboards.

The project involved creating dynamic charts. One of the key challenge was how to make the dashboard dynamic showing or hiding specific graphs based on certain conditions. For example, some charts were designed to appear only during specific time windows (e.g., between 3 PM to 5 PM or 5 PM to 7 PM IST). This simulates how dashboards in real-world businesses may show certain reports only during peak hours, team shifts, or business-critical periods.Another interesting aspect of the project was adding smart filters to the data. I applied several conditions such as showing apps only if their rating was above 3.5, reviews were greater than 500, installs were more than 50,000, and so on. I also excluded apps based on the presence of specific letters in their names and included sentiment-based filters to get more meaningful results. These filters helped narrow down the data to the most relevant and high-quality entries, improving the value of the visualizations.

Moreover, I added multilingual features to make the charts more user-friendly. Certain app categories were translated into local languages like "Beauty" shown in Hindi, "Business" in Tamil, and "Dating" in German depending on their context. This small but thoughtful feature added a personal and localized touch to the visualizations, making them more engaging.

Lastly, the visual design also included color-coding, such as highlighting the "Game" category in pink to make it stand out. These kinds of visual cues are important when trying to convey information quickly and effectively, especially in dashboards where users expect to make quick decisions.Overall, this project gave me the opportunity to explore how technical skills like data filtering, time-based logic, and multilingual customization can come together to build dashboards that are not only functional but also intelligent and user-centered. This background sets the foundation for the specific tasks I completed, the tools I used, and the outcomes I achieved in the next sections of this report.

**Learning objectives**

**1.Understand the Structure and Relevance of google playstore App Data**

One of the first learning goals was to explore and understand different features in a google play store app dataset. This included analysing columns such as app name, category, size, type (free or paid), number of installs, ratings, reviews, sentiment subjectivity, and last update date. Understanding how each of these metrics contributes to app performance was essential before moving into analysis and visualization.

**2. Strengthen Data Cleaning and Preprocessing Skills**

A major part of any data project is preparing the data for analysis. I aimed to improve my skills in cleaning messy data ,such as converting text values to numbers (e.g., size from "14M" to numeric), removing unwanted characters (like '+' and ',' from install counts), dealing with missing values, and creating new columns (e.g., converting size to MB). This helped ensure the accuracy and reliability of the visual output.

**3. Build Interactive and Customized Visualizations Using Plotly**

Another key objective was to learn how to create powerful and interactive charts using Plotly. This included building different types of plots—such as scatter plots, bar charts, and bubble charts—with advanced customization options. I also learned how to add trendlines, control color schemes, and size bubbles based on variables like number of installs. And also learned to visualize the charts as html files.

**4. Implement Time-Based Conditional Logic in Dashboards**

One unique learning goal was to simulate a real-time dashboard by making certain visualizations visible only during specific IST hours. For example, a grouped bar chart was shown only between 3 PM and 5 PM IST, and a bubble chart only between 5 PM and 7 PM IST. This helped me understand how time-based logic can be applied to dashboards to improve control, relevance, and user experience.

**5. Apply Advanced Filtering Techniques to Enhance Insights**

I aimed to sharpen my ability to use conditional filters effectively. These included logical conditions such as selecting only apps with a rating above 3.5, more than 500 reviews, sentiment subjectivity above 0.5, and excluding app names containing the letter 'S'. These filters made the visualizations more focused and actionable.

**6. Add Multilingual and Categorical Enhancements**

Another learning goal was to make the visualizations more user-friendly by translating certain app categories into regional languages—like Hindi, Tamil, and German—and using color highlights for specific categories like "Game". This introduced me to the idea of localization and personalization in data presentation.

**7. Improve Real-World Dashboard Thinking**

Overall, I wanted to develop a mindset that focuses on solving real-world data problems—such as when to show information, how to highlight important trends, and how to make insights understandable to a wider audience.

**Activities and Tasks**

**There were mainly three tasks performed by me in this internship project which are as follows:**

**Task1**:  Create a scatter plot to visualize the relationship between revenue and the number of installs for paid apps only. Add a trendline to show the correlation and color-code the points based on app categories.

**This was the first task given to me in this internship project ,in this task the main objective was to create a scatter plot between the revenue and the number of installs for paid apps only . a trendline was inserted to identify the trend how linearly the revenue increase with the number of installs and different categories were colour coded for easy identification.**

**The steps involved in performing this task are as follows:**

**Step 1:-** dataset was loaded using the pandas library and initial data information was checked using the pandas library such as checking for the null values, checking the number of rows and columns in the data, checking for the datatype of the dataset.

**Step 2:-** after the initial data checking now comes the part of data preprocessing , in data preprocessing the null values were replaced by mode of the particular column so as not to lose the important data, duplicate rows were removed because duplicated data is of no use and it can hinder the data analysis. Rows with rating more than 5 were removed because rating is always out of 5 means that rating more than 5 is a outlier data.

**Step 3:-** after the data preprocessing the data conversion was performed to convert the data into float and integer type from object type because the visualization cannot be performed on the object type data and various unwanted signs were removed to make the data normal such as ‘+’ sign from installs column and dollar sign from the price column.

**Step 4:-** the data was filtered according to the requirement of the task such filtering out the unpaid apps taking only the paid apps for analysis.

**Step 5:-** as there was need to generate a revenue column for particular apps , the revenue column was generated by multiplying ‘price’ and the ‘installs’ column.

**Step 6:-** a html file was defined to store and save the plot as a html file for displaying it directly on the browser.

**Step 7:-** finally the graph was created using plotly library of data visualization by taking installs on x axis and revenue on y axis and differentiating the colour of points on the basis of category.

**Task 2: Use a grouped bar chart to compare the average rating and total review count for the top 10 app categories by number of installs. Filter out any categories where the average rating is below 4.0 and size below 10 M and last update should be Jan month . this graph should work only between 3PM IST to 5 PM IST apart from that time we should not show this graph in dashboard itself.**

**this was the second task given to me in my internship project , in this task the main objective was to build a grouped bar chart to compare the average rating and and total review count of top 10 app categories by no. of installs.**

**The steps involved in performing the above task are as follows:**

**Step1:-** data loading with pandas library and initial data manipulations such as checking for null values ,checking for data info etc.

**Step 2:-** after the step 1 data preprocessing was performed such as replacing the null values by mode so as to maintain the data integrity and consistency, removing the duplicate data, filtering out the rows with rating more than 5.

**Step 3:-**after data preprocessing the data conversion was performed on the data such as converting the data type of installs column from object type to integer type and removing the”+” symbol , converting the size column to mb type from kb and and removing “M” and “k” symbol and converted the last updated column to datetime format from object type , reviews column was converted to integer type from object type.

**Step 4:-**after the data conversion now data filteration was done according to requirement of the task .

**Step 5:-** in the size column there were many apps with size mentioning that “varies with device ” so that replaced with a constant size of 11 as filter was to put for apps with size greater than 10.

**Step 6:-** now after the step 5 we group the data on the basis of category by taking the mean of rating and total sum of reviews and installs and then filter out the top 10 app categories on the basis of number if installs .

**Step7:-** after the step 6 we melt the data to long format just by taking the their review count and average rating.

**Step 8:-**create a html function to save and store the plot as html file and use javascript function to view the plot in dashboard between a particular timeline provided in the above task.

**Step 9:-**create the plot using the plotly library by taking category on x axis and count on y axis.

**Task 3:  Plot a bubble chart to analyze the relationship between app size (in MB) and average rating, with the bubble size representing the number of installs. Include a filter to show only apps with a rating higher than 3.5 and that belong to the Game, Beauty ,business , commics , commication , Dating , Entertainment , social and event categories. Reviews should be greater than 500 and the app name should not contain letter "S" and sentiment subjectivity should be more than 0.5 and highlight the Game Category chart in Pink color. We have to translate the Beauty category in Hindi and Business category in Tamil and Dating category in German while showing it on Graphs. Installs should be more than 50k as well as this graph should work only between 5 PM IST to 7 PM IST apart from that time we should not show this graph in dashboard itself.**

**This was the third task given to me in my internship project, the objective of this task was to build a bubble chart where the size of the bubble was decided by the number of installs ,the chart was to show a relationship between the average rating and app size and various filters were applied according to the requirement .**

**The various steps involved in performing this task are as follows :**

**Step1 :** various libraries were imported according to the requirement of the task ,dataset was loaded and the datasets were loaded as there two datasets required for the above task one apps dataset and one translated reviews dataset.

**Step 2:** initial data manipulations were performed on the datasets and data info was collected and the two datasets were merged on the basis of inner join.

**Step 3:** data preprocessing was performed on the data such as dropping the rows where there is no rating and filtering out the rows where the rating is greater than 5 and the duplicate rows were removed.

**Step 4:** data conversion was performed on various columns as per the requirement of the task such as converting them to integer format and removing unwanted signs from the values.

**Step 5:** filters were applied as per the requirement of the task and the categories of apps were translated to different languages as mentioned in the task.

**Step 6:** created a html function to save and store the plot as html file and use javascript function to view the plot in dashboard between a particular timeline provided in the above task.

**Step 7:** created the bubble chart using the plotly library by taking size on x axis and rating on y axis and setting the size of the bubble according to the number of installs.

**Skills and competencies**

**During this internship the various skills acquired by me are as follows:**

**Technical Skills**

**1:Python Programming**

* Gained practical experience in using Python for data analysis and visualization.
* Used libraries such as Pandas and NumPy for data cleaning, transformation, and filtering.

**2:Data Cleaning and Preprocessing**

* Learned to handle missing values, convert textual data (e.g., size in “M”) to numerical format, and remove special characters from columns like “Installs”.
* Created new columns such as revenue in task 1 to enhance analysis accuracy.

**3: Data Visualization with Plotly**

* Mastered the use of Plotly Express to build interactive charts such as scatter plots, bubble charts, and grouped bar charts.
* Added features like trendlines, color coding, and dynamic bubble sizes.
* Customized chart aesthetics for clarity and impact.

**4:Conditional Rendering Based on Time (IST) and creating dashboards using html**

* Implemented logic using javascript tile logic functionand to restrict chart visibility within specific IST time windows (e.g., 3–5 PM and 5–7 PM).
* Created the dynamic dashboards for easy view directly on browser using html.

**5: Advanced Filtering Logic**

* Applied multiple conditional filters such as:
  + Ratings greater than 3.5
  + Reviews above 500
  + Excluding app names containing certain letters (e.g., “S”)
  + Sentiment subjectivity above 0.5
  + Category-specific and install count filters

**6:Multilingual and Categorical Data Handling**

* Translated specific app categories into Hindi, Tamil, and German.
* Demonstrated the ability to localize charts for better user engagement.

**Professional and Analytical Competencies**

**1: Problem-Solving**

* Tackled challenges related to data inconsistencies, missing values, and complex filtering rules.
* Designed practical solutions to simulate real-world dashboard behavior.

**2: Analytical Thinking**

* Interpreted data patterns to identify meaningful insights.
* Designed visualizations to support data-driven decisions.

**3: Attention to Detail**

* Carefully implemented logic to avoid data misinterpretation.
* Ensured that only relevant and high-quality data was included in each chart.

**4: Creativity in Visualization**

* Used color coding, bubble sizing, and multilingual labels to make charts more intuitive and visually appealing.

**5: Time Management**

* Planned tasks efficiently to complete each visualization within scope and time constraints.
* Followed a logical workflow from data preprocessing to final chart export.

**6: Dashboard Thinking**

* Developed a mindset for creating professional dashboards that are not only functional but also user-friendly and adaptive to different user needs.

**Feedback**

Working on this project was a highly rewarding experience that allowed me to apply the technical concepts I have studied in a practical and meaningful way. When I started, I had a good understanding of basic data analysis and Python programming, but this internship challenged me to go beyond the fundamentals. I had to think critically, manage time effectively, and consider real-world business use cases while building visualizations.

One of the most enriching aspects of the project was dealing with complex filtering logic—especially when applying multiple conditions to narrow down large volumes of app data. It taught me the importance of precision and logic structuring in programming. The implementation of time-based visibility logic for the charts was completely new to me, and figuring out how to apply Indian Standard Time (IST) was a significant learning milestone. This feature added depth to the project and made the dashboard feel dynamic and smart, much like the professional dashboards used in data-driven organizations.

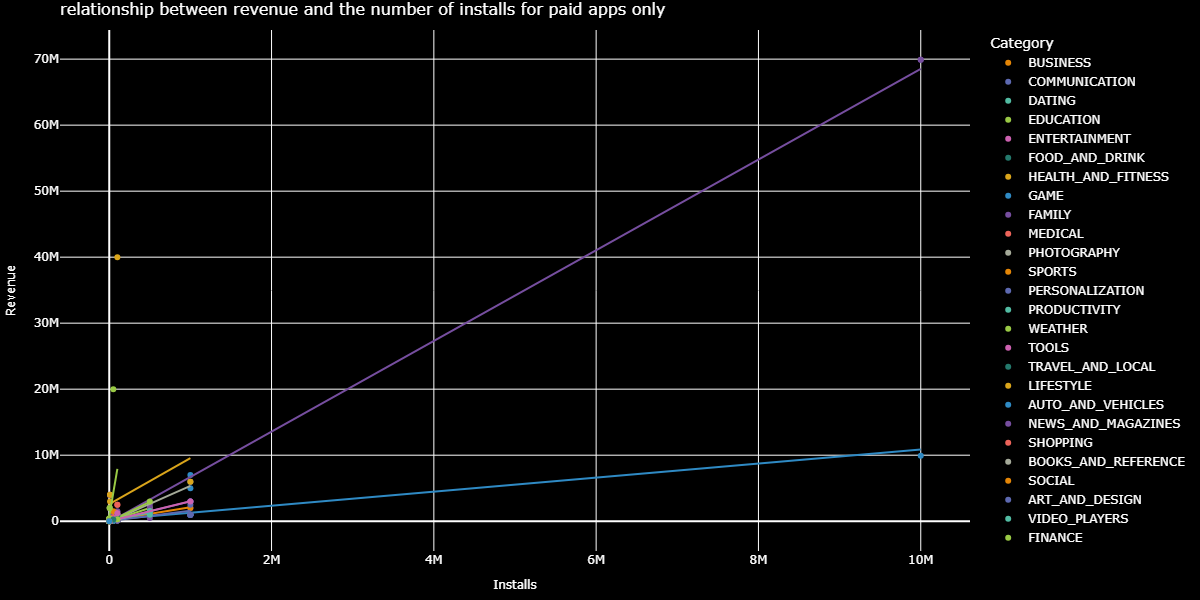
I also took pride in the attention given to user experience—ensuring charts were not just technically accurate but also clean, visually appealing, and easy to interpret. For instance, color-coding the "Game" category in pink and translating specific categories into Hindi, Tamil, and German added a personalized touch that made the visualizations more user-centric. These efforts reinforced the importance of data storytelling—turning raw data into something people can understand and use to make decisions.

This project also improved my ability to work independently, debug errors, and adapt my code structure based on evolving requirements. It was a valuable experience that strengthened both my technical and soft skills, preparing me for more complex roles in data analytics and dashboard development.

**Evidence**

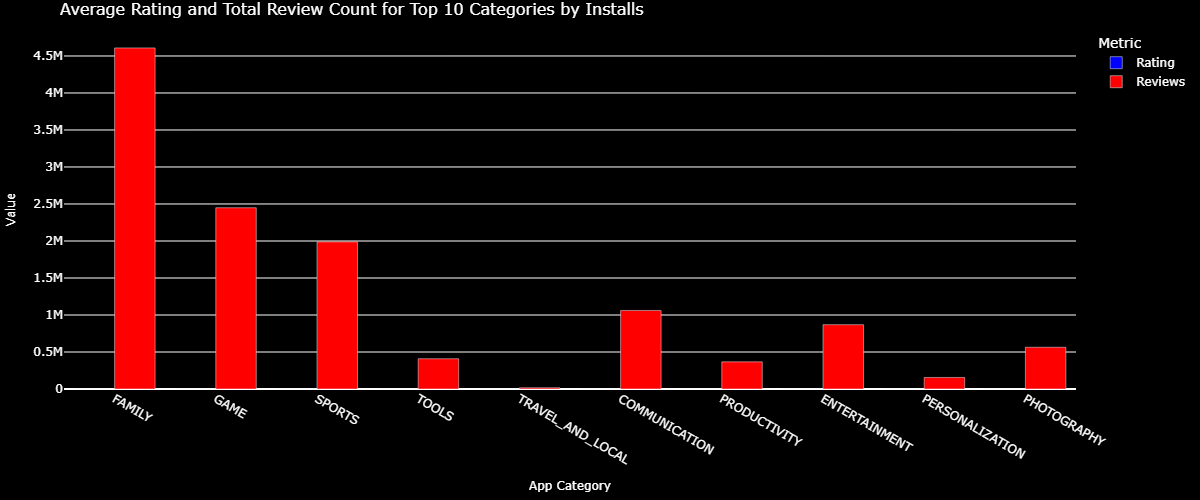
**Task 1:** refer the below html document for dashboard view



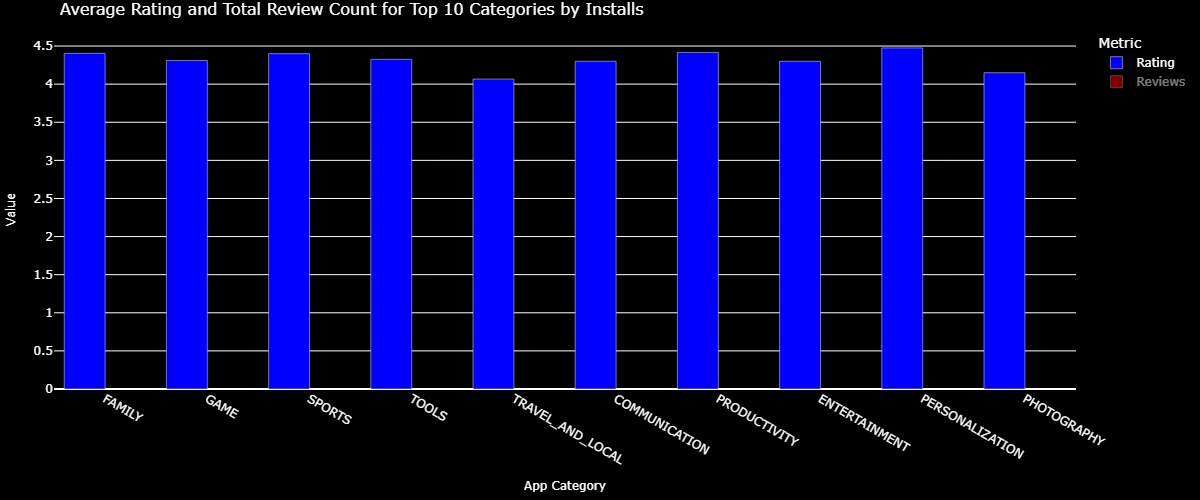
****

**Task 2:** refer the below html document for dashboard view



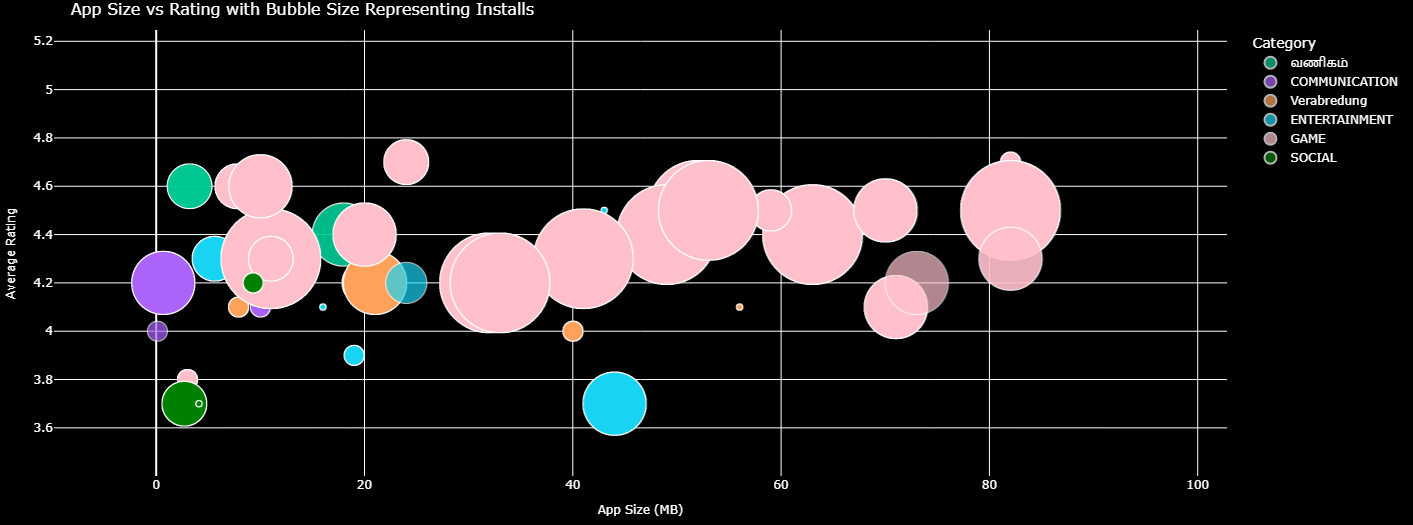
****

**Top of Form**

****

**Task 3:** refer the below html document for the dashboard view.**Bottom of Form**



****

**Challenges and Solutions**

**the various challenges faced by me in performing the given tasks are as follows:**

**1. Filtering with Multiple Complex Conditions**

**Challenge:**  
One of the most difficult tasks was applying multiple filters simultaneously ,such as filtering apps by category, ratings, reviews, app names not containing a particular letter, sentiment subjectivity, and install count. Combining all these conditions without affecting the integrity of the dataset was tricky.

**Solution:**  
I carefully structured the filtering logic step-by-step using Pandas. I ensured each condition was applied sequentially and used Boolean masking to combine them properly. Testing intermediate outputs helped confirm that no data was wrongly excluded or included.

**2. Time-Based Graph Display in IST (Using JavaScript)**

**Challenge:**  
Restricting specific graphs to appear only during certain hours (like 3–5 PM IST or 5–7 PM IST) was a unique requirement. Initially, I struggled with how to implement time-based visibility since most of the logic was handled in Python, and Python runs on the server side not in the browser.

**Solution:**  
I realized that this kind of dynamic visibility needed to be handled on the client side using JavaScript. I wrote JavaScript functions that detected the current time, converted it to IST using the browser’s local time offset, and then conditionally displayed or hid the graph containers. This allowed me to control dashboard visibility based on real-time clock logic without reloading the page or relying on server-side validation.  
It also taught me how to integrate front-end scripting with data visualization workflows, making the dashboard smarter and more interactive.

**3. Multilingual Category Translation in Graphs**

**Challenge:**  
Displaying specific app categories in different languages (e.g., Beauty in Hindi, Business in Tamil, and Dating in German) required me to find the correct translations and figure out how to reflect them dynamically in the chart labels.

**Solution:**  
I used a simple mapping dictionary to replace category names before plotting. I verified translations using online tools and ensured they were correctly rendered in the graphs. This task helped me think from a global user perspective.

**4. Data Cleaning and Missing Values**

**Challenge:**  
The raw dataset contained missing, inconsistent, or malformed values—especially in columns like size, installs, and reviews. These caused errors during plotting or filtering.

**Solution:**  
I used Pandas functions like fillna(), dropna(), and .replace() to clean the data. I also wrote custom functions to convert install counts from strings (e.g., “50,000+”) into numerical format. This made the dataset analysis-ready.

**5. Visual Customization and Highlighting**

**Challenge:**  
Highlighting the "Game" category in pink while keeping other categories neutral was difficult using standard colour options.

**Solution:**  
I wrote a conditional colour-mapping function using Plotly, which assigned pink to “Game” and default colours to others. This enhanced visual clarity and user focus.

**6. Maintaining Dashboard Clarity**

**Challenge:**  
With multiple filters, time-based logic, and customized designs, it became difficult to maintain a clean and easy-to-understand dashboard.

**Solution:**  
I used modular coding to separate chart logic, cleaned up graph titles and labels, and added tooltips where necessary. This helped ensure that the final output was both functional and visually attractive.

**Outcomes and Impact**

**1. Interactive and Intelligent Visualizations**

The project resulted in the creation of three advanced visualizations ,a scatter plot, grouped bar chart, and bubble chart—each driven by distinct business logic, conditions, and constraints. These visualizations were not only visually appealing but also deeply interactive, with filters and conditions that allowed users to gain precise insights from the data. They mimicked real-world dashboard behaviour, especially in the context of app store analytics.

**2. Time-Based Logic for Real-Time Insights**

By implementing JavaScript-based time logic, I was able to control when specific graphs appeared on the dashboard. This simulated real-world use cases where dashboards display or restrict content based on time-sensitive events or business hours. It added a layer of realism to the project and exposed me to the blend of front-end logic with back-end analytics.

**3. Mastery in Data Cleaning and Filtering**

Through hands-on practice, I became comfortable applying multiple filters simultaneously using conditions such as category type, ratings, app size, install count, sentiment metrics, and string operations. These operations deepened my understanding of how to cleanse and preprocess data to ensure accuracy in analysis.

**4. Cultural and Language Awareness in Visualization**

One unique aspect of this project was translating selected app category names into Hindi, Tamil, and German. This not only improved the inclusiveness and diversity of the visuals but also taught me how to incorporate multilingual support in data presentation—a valuable skill for global applications and user interfaces.

**5. Real-World Scenario Simulation**

The tasks mimicked real business challenges such as performance tracking of app categories, market segmentation based on sentiment, and product filtering using multiple conditions. By applying visualization logic to practical use cases (like paid apps vs free, install volumes, and sentiment-driven filtering), I gained experience in translating raw data into business insights.

**6. Growth in Technical Stack**

This internship enhanced my skills in several key technologies including:

* **Python** (Pandas, Plotly)
* **JavaScript** (for time-based dashboard view control)
* **HTML/CSS** (for setting the dashboard layout)

**7. Professional Readiness**

Most importantly, the internship gave me a taste of how data is handled in real companies—how dashboards are expected to perform, how visual clarity is prioritized, and how attention to small details (like translations or review counts) can make a big difference in understanding user behaviour. I now feel more confident in pursuing roles related to data analysis and dashboard development.

**Conclusion**

This internship has been an incredibly valuable experience that bridged the gap between theoretical knowledge and practical application in the field of data analytics and visualization. The tasks assigned challenged me to think critically, code efficiently, and design visualizations that are both informative and user-friendly.

By working on real-world inspired datasets and requirements, I gained hands-on experience in data preprocessing, dynamic filtering, multilingual labeling, and time-based content rendering. I explored and integrated technologies like Python (Pandas, Plotly), JavaScript, and HTML to create an intelligent dashboard that reflects actual business intelligence needs. The project taught me how to make data-driven decisions and present insights in a way that users can interact with and understand easily.

Moreover, this internship helped me improve both technical and soft skills. I learned how to tackle problems independently, troubleshoot unexpected results, and document my process effectively. These experiences have contributed to my growth as a budding data professional, and I feel more prepared to take on future roles in data analytics, dashboard design, and decision support systems.

In summary, the internship has not only enhanced my technical competencies but also reinforced my enthusiasm for pursuing a career in data science and analytics. It has given me a strong foundation to build upon, and I am excited to apply these skills in real-world scenarios ahead.

**Bottom of Form**