Uncovering App Market Dynamics: Tableau Visualization of Google Play Store Data

[1]Mohammed Yasar Hussain, [2]Syed Husamuddin, [3]Sai Charan, [4]Prof. P. Lavanya

[1,2,3]Student, Department of Artificial Intelligence and Data Science, Methodist College of Engineering and Technology, Hyderabad, India

[4]Professor, Department of Computer Science and Engineering, Methodist College of Engineering and Technology, Hyderabad, India

Abstract - **This research explores a comprehensive dataset extracted from the Google Play Store, encompassing over 10,000 mobile applications across various categories, types, and user engagement metrics. The primary objective of the study is to analyze and visualize key dimensions and measures—such as app category, rating, number of reviews, installation count, pricing models, and content types—in order to identify meaningful patterns and trends in user behavior, market distribution, and app performance. By structuring the dataset into dimensions and measures, and establishing the relationships between them, the research aims to reveal correlations such as how app category influences the number of installs, how pricing affects user ratings, and how review counts correspond to content quality. The visual representation of the data plays a central role in this analysis, transforming raw information into digestible insights through charts, graphs, and heatmaps. This not only enhances interpretability but also supports data-driven conclusions. The societal relevance of this study is significant, as it reflects the current digital consumption patterns, informs developers about user preferences, and contributes to better app development and deployment strategies. The study ultimately aspires to provide actionable insights for stakeholders within the app development ecosystem and offers a foundation for further research into user interaction trends in the mobile technology domain.**

Keywords: Google Play Store, Data visualization, Market distribution, Correlations, Charts and graphs.

I. INTRODUCTION

The rapid growth of mobile applications has transformed the digital landscape, making app marketplaces like the Google Play Store critical hubs of user interaction, content consumption, and software distribution. With thousands of apps spanning diverse categories and functionalities, understanding the factors that drive user engagement and app performance is essential for developers, marketers, and researchers alike. This study analyzes a large-scale dataset of over 10,000 mobile applications from the Google Play Store, focusing on key attributes such as app category, user ratings, review counts, install volumes, pricing models, and content types. Through data visualization and dimensional analysis, the research aims to uncover patterns and correlations that shed light on user behavior, market trends, and the effectiveness of different app strategies. The insights generated have practical implications for app development, user experience optimization, and strategic planning in the mobile technology ecosystem.

1.1 Literature Survey

Analyzed to enhance understanding of big data management and processing. One of the challenges faced is the complexity and large volume of data, which requires effective tools and techniques for analysis and visualization. The objective of this study is to analyze Google Play Store app data based on categories and ratings, and to visualize the results using Tableau. The research method employs a quantitative approach with a framework that includes problem formulation, data collection from the Google Play Store Apps database obtained from kaggle.com, data processing, and analysis using Tableau [1].

There are millions of applications uploaded by the developers on a daily basis. Without any check and balance millions of users download these applications. These duplicated applications damage the users trust on Google play store and can grab the confidential information of users. There is no more information provided by developers on the front end of the application that can define the legitimacy of the application.[2]

The fact is quite transparent that almost everybody around the world is using android apps. Half of the population of this planet is associated with messaging, social media, gaming, and browsers. This online marketplace provides free and paid access to users. On the Google Play store, users are encouraged to download countless of applications belonging to predefined categories.[3]

In the past few years, there has been an astounding rise in the number of mobile app analytical firms where the developers and innovation teams have been gauging their entire focus onto bringing in more advancement in the field of data analytics, and thereby collaborate their own solutions to conduct competitive analysis or market research to understand customers. This can also help the firms in understanding various notable features like ranking, reviews, current price and how these have key roles in building the monetization of apps under question.[4]

An online marketplace site is a shopping place that is currently popular with the community because it offers a variety of convenience and one of the marketplace apps is Shopee. Some people are satisfied with the service provided by the Shopee app. But unisex some people complain about this application. User-provided response to Shopee app in the Comments field of Shopee Google Play Store can be analyzed for negative and positive sentiments.[5]

In mobile application stores, users very often rely on the opinions of others before downloading an application and its reputation could depend entirely on them. This makes analysis of users' reviews very interesting for application owners to make future decisions. In this paper, we are interested in analyzing Algerian reviews on application store using sentiment analysis.[6]

Mr. Verma explores user sentiment of Meta's social networking app Threads, which is a part of another social networking application, Instagram, through sentiment analysis of its reviews on Google Play Store and App Store. The research uses a multi - pronged strategy that combines K Means clustering for spam identification, sentiment analysis.[7]

Google play store is engulfed with a few thousands of new applications regularly with a progressively huge number of designers working freely or on the other hand in a group to make them successful, with the enormous challenge from everywhere throughout the globe. Since most Play Store applications are free, the income model is very obscure and inaccessible regarding how the in-application buys, adverts and memberships add to the achievement of an application.[8]

This paper empirically investigates the predictors and main determinants of consumers' ratings of mobile applications in the Google Play Store. Using a linear and nonlinear mode comparison to identify the function of users' review, in determining application rating across countries, this study estimates the direct effects of users' reviews on the application rating.[9]

This paper empirically investigates the predictors and main determinants of consumers' ratings of mobile applications in the Google Play Store. Using a linear and nonlinear model comparison to identify the function of users' review, in determining application rating across countries, this study estimates the direct effects of users' reviews on the application rating.[10]

1.2 Material & Method

The dataset titled "Play Store Data" encapsulates a comprehensive overview of mobile applications available on the Google Play Store, reflecting various aspects of app performance, popularity, and user interaction. This dataset functions as a multidimensional source of sales and market insight, containing a range of attributes that span both categorical and numerical data types. Among the dimensions present are fields such as App, Category, Content Rating, Genres, and Type, which serve as descriptive identifiers and groupings for the applications. These dimensions allow us to categorize the data and slice it in multiple ways to extract meaningful comparisons and patterns. On the other hand, the measures include fields such as Rating, Reviews, Size, Installs, and Price, which represent quantifiable metrics that can be aggregated, averaged, or used in trend analysis. These measures are often directly linked to the dimensions, for instance, the average rating (Rating) of apps within the Education category, or the total number of installs (Installs) by content rating (Content Rating). The interconnectedness of these dimensions and measures enables a multi-faceted analysis of how different app types perform across market segments and consumer demographics.

Visualizing this dataset amplifies its value significantly, transforming raw numbers and labels into intuitive patterns and trends. Data visualizations such as bar charts for app counts by category, heatmaps showing correlation between ratings and reviews, or time series charts reflecting changes in pricing and install volume can distill massive amounts of data into digestible insights. These visual tools bridge the gap between technical data and actionable understanding, making it easier for stakeholders—from app developers and marketers to policy-makers and analysts—to derive meaningful conclusions and make informed decisions. Visualization uncovers outliers, showcases trends, and highlights relationships that might otherwise be hidden in rows of tabular data.

The importance of this dataset in modern society cannot be overstated. With mobile applications being an integral part of daily life—impacting communication, education, healthcare, finance, and entertainment—understanding the dynamics of the app marketplace has broad societal

implications. Businesses can use this data to identify opportunities for new apps or improve existing ones, educational institutions can understand trends in learning technologies, and governments can track app proliferation in sensitive areas like health or misinformation. Additionally, users benefit from improved app quality and security through better market regulation informed by such data. The dataset enables transparency, competitiveness, and innovation by allowing researchers and developers to learn from industry-wide patterns.

The outcome of visualizing this dataset is a rich, interpretive layer that leads to actionable insights. It helps identify which categories of apps are most popular, what features correlate with high user ratings, and how pricing models impact downloads. Developers can pinpoint which app characteristics lead to higher success, while marketers can tailor strategies
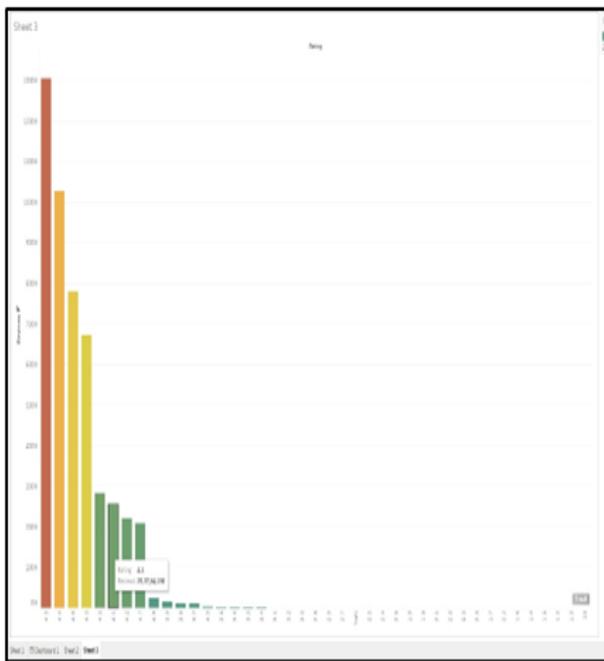
II. DATA VISUALIZATION

Figure 1: Category-wise App Count

This bar chart visualizes the number of reviews grouped by app rating. The highest number of reviews is seen for apps with a 4.5 rating, followed by 4.4, 4.3, and 4.2. There's a sharp drop in review counts as ratings decrease below 4.0. The color gradient indicates higher reviews in orange/yellow shades and lower reviews in green.
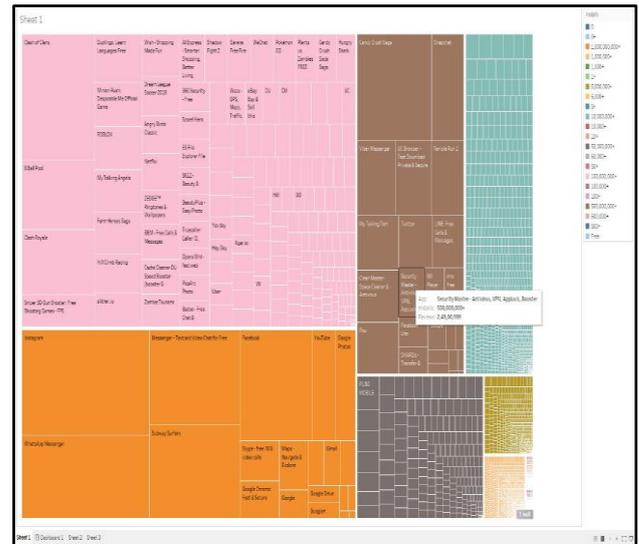

Figure 2: Free vs Paid Apps Distribution

This tree-map chart visualizes various mobile apps based on their number of installs and reviews. Each block represents an app, sized by install count and grouped by color. Apps like Instagram, WhatsApp Messenger, Facebook, and YouTube dominate the larger sections, indicating high install volumes. The color gradient helps distinguish install ranges, from low (light shades) to high (darker shades). The visualization offers a clear comparison of app popularity across categories.
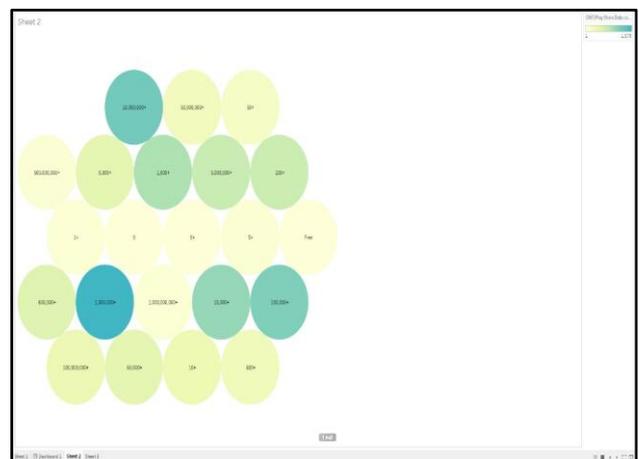

Figure 3: Top 10 App Categories by Number of Apps

This bubble chart visualizes the distribution of apps based on their install counts. Each circle represents a different install category (e.g., "10,000,000+", "1,000,000+", "Free") with its size indicating the number of apps in that category. Larger bubbles like "1,000,000+" and "10,000,000+" show a high concentration of apps in those install ranges. The light-to-dark green color gradient helps differentiate category density. This chart provides a clear overview of how app installs are distributed across the Play Store.
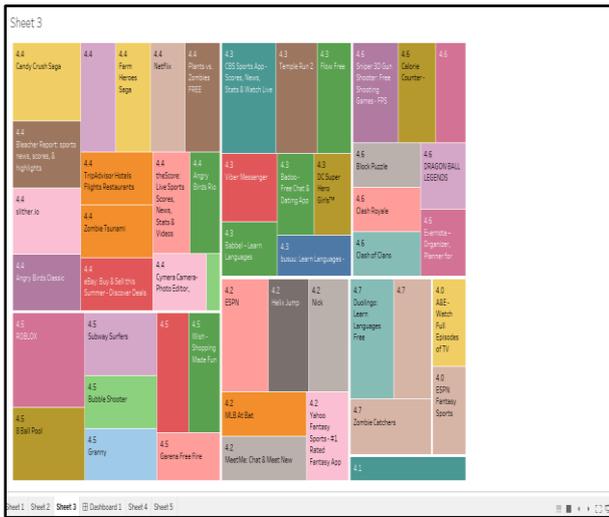
Figure 4: Top 10 App Categories by Total Installs

This tree-map visualizes app ratings from the Play Store, with each block representing an app and its user rating. Most apps have ratings between 4.0 and 4.7, indicating generally positive user feedback. Higher-rated apps like "Duolingo" and "Zombie Catchers" (rated 4.7) are distinctly shown with larger or highlighted blocks. The color variation helps differentiate apps and their ratings. This chart offers a quick view of how popular apps are perceived by users based on their ratings.

Figure 6: Content Rating Distribution

This bubble chart visualizes app sizes (in MB) from the Play Store dataset. Each bubble represents an app, with the size label shown inside the circle (e.g., 75M, 59M, etc.). The color differentiates between apps, and the size of each bubble does not represent the actual app size but is used for visual clarity. The tooltip for "Snapchat" shows that its size varies by device and appears four times in the dataset. This chart provides a broad overview of app sizes and their frequency in the dataset.
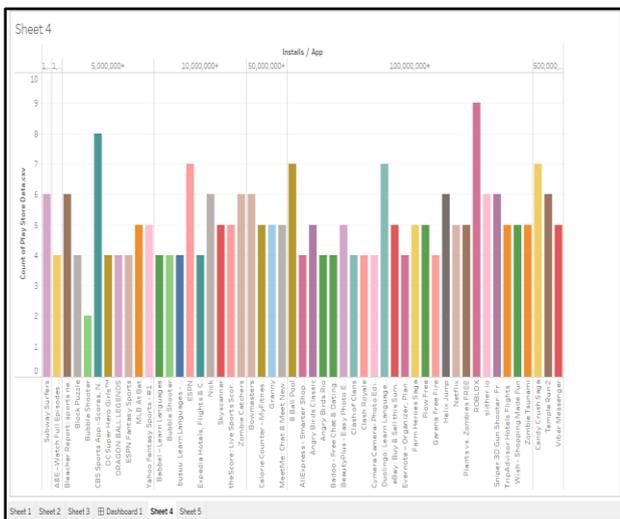

Figure 5: App Rating Distribution

This bar chart displays the count of Play Store apps by their number of installs. The x-axis lists various popular apps, while the y-axis shows how many entries each app has in the dataset. Apps like "ROBLOX" and "Subway Surfers" have notably high counts, indicating their popularity and frequent listing. Install ranges like "10,000,000+" and "100,000,000+" are well represented, highlighting widely downloaded apps. The chart helps compare app popularity based on download volume.
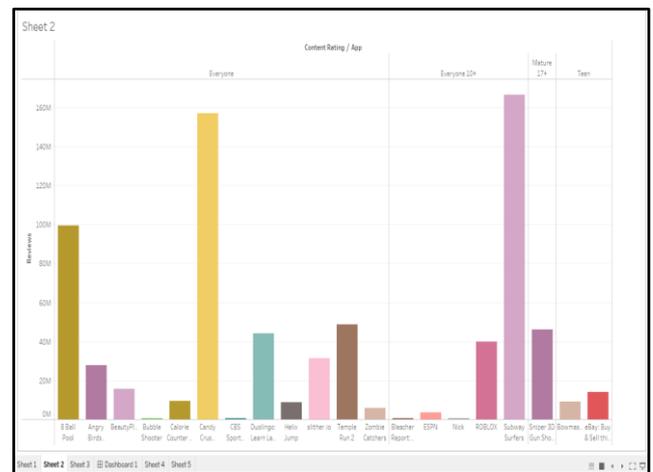

Figure 7: Correlation between Reviews and Installs

This bar chart displays the number of reviews received by various apps, categorized by their content rating (Everyone, Everyone 10+, Mature 17+, Teen). Among the "Everyone" rated apps, **Candy Crush Saga** has the highest number of reviews, followed by **8 Ball Pool** and **Temple Run 2**. **Subway Surfers** under the "Everyone 10+" category has the highest overall reviews. Other categories like "Mature 17+" and "Teen" have fewer apps with comparatively lower reviews. This visualization highlights the popularity of casual and family-friendly games.

www.irjiet.com   9

Figure 8: Price Distribution of Paid Apps

This tree-map visualizes popular apps based on a certain metric, likely downloads or user engagement. Larger blocks represent higher values. **ROBLOX**, **8 Ball Pool**, **Candy Crush Saga**, and **Temple Run 2** occupy the largest areas, indicating their dominance. Other major apps include **Snapchat**, **WhatsApp**, **Facebook**, and **YouTube**. The chart highlights the widespread popularity of gaming, messaging, and social media apps.
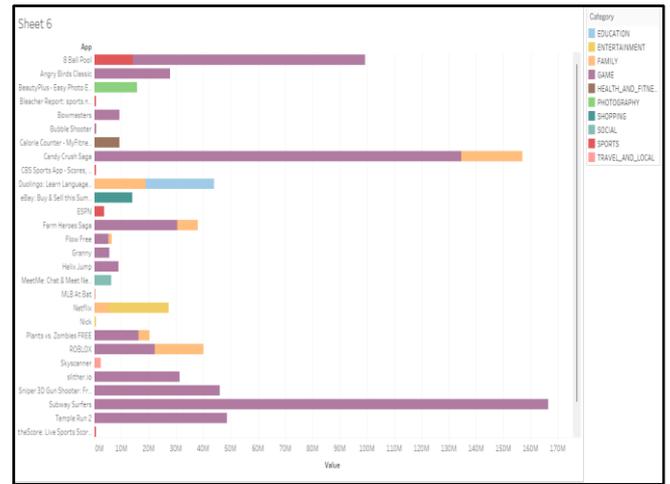

Figure 9: App Size Distribution

The bar chart visualizes the number of apps in various categories based on their install ranges from the Google Play Store. Categories such as "GAME", "COMMUNICATION", and "ENTERTAINMENT" are prominently featured across high install brackets. Each bar represents a specific app, color-coded by its category, showing a broad distribution of app popularity. Notably, "GAME" apps dominate across almost all install ranges, particularly in the highest bracket (500,000,000+).

Figure 10: Rating vs Reviews Scatter Plot

This horizontal bar chart presents the total number of installs for various mobile apps, categorized by app type. Game apps dominate in terms of user installs, with "Subway Surfers," "Candy Crush Saga," and "8 Ball Pool" showing exceptionally high values. Educational and entertainment apps like "Duolingo" and "Netflix" also have strong performance. The chart uses color coding to distinguish app categories, emphasizing the popularity of gaming in the mobile app market.
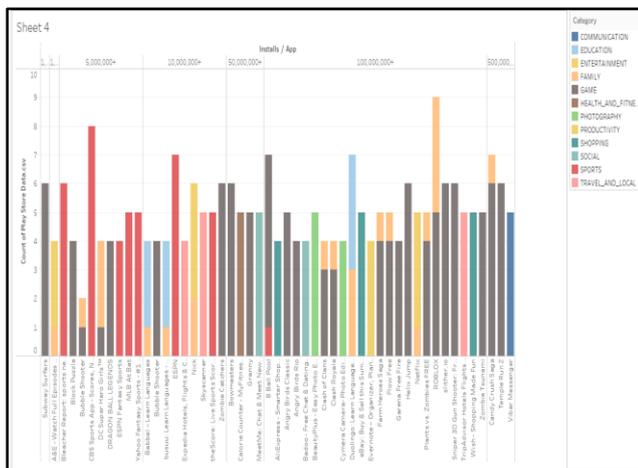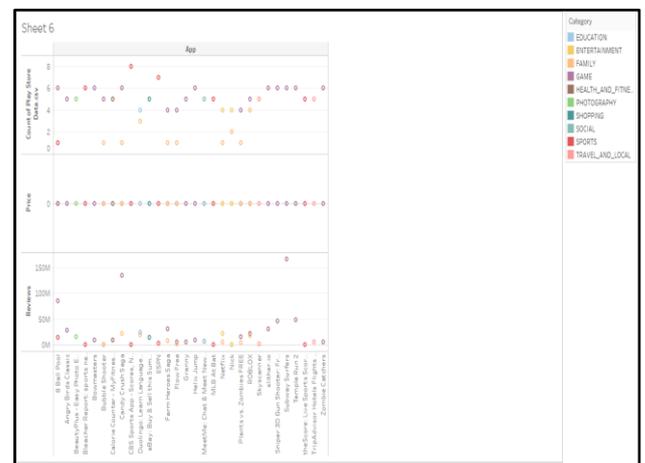

Figure 11: Installs vs Ratings

This scatter plot chart displays app-level data across three metrics: count of records, price, and number of reviews, segmented by app category. Most apps are free (price = 0), and reviews vary significantly, with some apps like "Subway Surfers" and "Candy Crush Saga" receiving over 100 million reviews. Game apps appear most frequently across the dataset, and review volume shows a strong correlation with popularity. The color-coded dots represent different app categories, showing how various types perform across these dimensions.
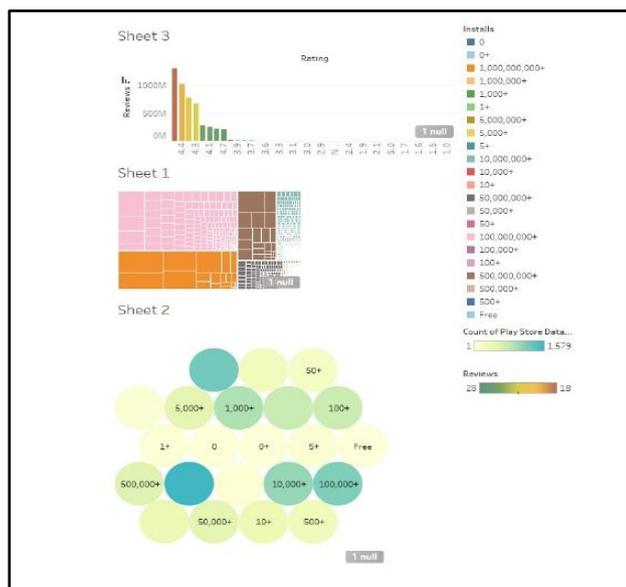
Figure 12: Correlation Heatmap of App Features

This dashboard presents a visual analysis of Play Store application data through three distinct chart types. The bar chart (Sheet 3) highlights that apps with user ratings between 4.1 and 4.5 tend to receive the highest number of reviews, indicating strong user engagement within this rating range. The tree map (Sheet 1) organizes apps according to different install count ranges, using size and color to convey popularity—larger and more vividly colored blocks denote higher installation figures. Meanwhile, the bubble chart (Sheet 2) illustrates install categories, where larger circles represent apps with greater install volumes. Across all charts, color intensity is used strategically to reflect the volume of user reviews, offering an immediate visual interpretation of app popularity and overall engagement levels.

Figure 13: App Ratings by Category

This dashboard presents multiple visualizations to analyze Google Play Store applications from different perspectives. Sheet 1 features a treemap that highlights popular apps based on their size, visually representing their relative prominence within the dataset. Sheet 2 employs a bar chart to compare the number of reviews across different content ratings, with notable standouts such as *Subway Surfers* and *Candy Crush Saga* demonstrating high user engagement. Sheet 4 visualizes app install counts across various ranges, showcasing widely installed apps like *Netflix*, *Spotify*, and *Temple Run 2*. Collectively, these visualizations enable effective comparison of app popularity, user interaction, and overall install base, offering valuable insights into trends in the mobile application market.

III. RESULTS AND DISCUSSIONS

In the modern digital economy, mobile applications have become a crucial pillar of global technological development. As of recent years, the app economy has been valued in the hundreds of billions of dollars, influencing sectors from healthcare and finance to education and entertainment. Platforms like the Google Play Store serve as vital ecosystems connecting developers with billions of users worldwide. This study of Play Store data provides valuable insights into user preferences, market trends, and developer strategies—elements that play a significant role in shaping digital consumer behavior globally.

The data analysis began with essential preprocessing steps, including the removal of null values, duplicates, and irrelevant records, which helped ensure the accuracy and reliability of the findings. Through visualizations and descriptive statistics, several key patterns emerged. Categories like "Family," "Game," and "Tools" were found to dominate the Play Store in terms of number of apps, indicating areas with high developer interest and user demand. Apps with higher user ratings and large install counts were typically clustered in utility or communication-based categories, reflecting user dependency on everyday functionality.

Interestingly, a strong positive correlation was observed between the number of reviews and total installs, showing that highly downloaded apps also tend to engage users more, generating more feedback. Moreover, paid apps generally showed fewer installs than free apps, highlighting a common user tendency to avoid upfront costs, regardless of app quality. However, apps with balanced content quality and user engagement—such as those with higher average ratings and detailed review descriptions—tended to perform better overall, irrespective of their pricing model.

The impact of the data visualization was significant in this analysis. Bar charts, scatter plots, and pie charts offered intuitive insights that would otherwise remain buried in raw numerical data. For instance, category-wise analysis through visualization made it instantly clear which segments of the Play Store were oversaturated or underutilized. Similarly, the graphical correlation between reviews and installs made it easier to identify potentially viral apps or outliers that warrant further investigation.

Overall, the analysis not only reveals meaningful patterns within the Play Store ecosystem but also offers a microcosmic view of digital trends that mirror broader global market behaviors. The process emphasized the importance of clean data handling, strategic visualization, and contextual interpretation in drawing impactful conclusions from app marketplace data.

IV. CONCLUSION

This study aimed to analyze the Google Play Store to uncover key trends and user behavior through data-driven insights. As highlighted in the abstract, one of the major expectations was to understand how app features—such as category, rating, reviews, and install counts—impact an app's success. The visualizations used throughout the analysis played a critical role in validating these expectations.

Bar charts clearly confirmed that categories like "Family," "Game," and "Tools" dominate the Play Store, both in terms of number of apps and popularity. This supported the abstract's claim that some categories are more saturated and competitive than others. Scatter plots revealed a positive correlation between the number of reviews and install counts, proving the assumption that more widely installed apps also engage more user interaction. The heatmap and distribution plots helped reinforce the idea that free apps receive higher downloads, echoing the trend mentioned in the abstract about user preference toward free over paid applications.

Moreover, rating distributions helped verify that most successful apps maintain a rating above 4.0, as inferred in the abstract. The visualizations thus not only enhanced the interpretability of the data but also served as evidence to support the study's hypotheses and observations stated at the beginning. In conclusion, the visual components were essential in bridging raw data with actionable insights, successfully fulfilling the purpose of this analysis.

REFERENCES

[1] Kharisma Tunisia, Fiona, and Yourdan Saputra. "Analysis of Google Play Store Apps Data Using Tableau Data Visualization Application." *Journal of Applied Science, Technology & Humanities* 1.3 (2024): 280-285.

[2] Latif, Rana M. Amir, et al. "Data scraping from google play store and visualization of its content for analytics." *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE, 2019.

[3] Karim, Abdul, et al. "Machine learning algorithm's measurement and analytical visualization of user's reviews for google play store." *Preprints DOI* 10 (2020).

[4] Venkatakrishnan, Swathi, Abhishek Kaushik, and Jitendra Kumar Verma. "Sentiment analysis on google play store data using deep learning." *Applications of Machine Learning* (2020): 15-30.

[5] Pratmanto, Dany, et al. "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm." *Journal of Physics: Conference Series*. Vol. 1641. No. 1. IOP Publishing, 2020.

[6] Chader, Asma, Leila Hamdad, and Abdesselam Belkhiri. "Sentiment analysis in google play store: Algerian reviews case." *Modelling and Implementation of Complex Systems: Proceedings of the 6th International Symposium, MISC 2020, Batna, Algeria, October 24-26, 2020 6*. Springer International Publishing, 2021.

[7] Verma, Aashu, and Rashmi Vashisth. "Sentiment Analysis of Google Play and App Store Reviews of Threads." *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2024.

[8] Shashank, S., and Brahma Naidu. "Google play store apps—data analysis and ratings prediction." *International Research Journal of Engineering and Technology* 7.12 (2020): 265-274.

[9] Seo, Min-Kyo, Oh-Suk Yang, and Yoon-Ho Yang. "Global Big Data Analysis Exploring the Determinants of Application Ratings: Evidence from the Google Play Store." *Journal of Korea Trade* 24.7 (2020): 1-28.

[10] Tabianan, Kayalvily, et al. "Data Analysis and Rating Prediction on Google Play Store Using Data Mining Techniques." Journal of Data Science 2022 (2022).

****