

# Understanding Categorical Mobile Applications in Google Marketspace: Google Play Store Data Analysis

Karthik Sivaraman Iyer  
School of Engineering and Applied Sciences  
The George Washington University  
Washington, DC  
[karthiksivaraman.iyer@gwu.edu](mailto:karthiksivaraman.iyer@gwu.edu)

Vidit Rakesh Sheth  
School of Engineering and Applied Sciences  
The George Washington University  
Washington, DC  
[viditrakesh.sheth@gwu.edu](mailto:viditrakesh.sheth@gwu.edu)

Avani Rao  
School of Engineering and Applied Sciences  
The George Washington University  
Washington, DC  
[avani.rao@gwu.edu](mailto:avani.rao@gwu.edu)

Akash Poddar  
School of Engineering and Applied Sciences  
The George Washington University  
Washington, DC  
[akash.poddar@gwu.edu](mailto:akash.poddar@gwu.edu)

***Abstract - The market for mobile applications has expanded at an unprecedented rate in recent years, with Google Play Store being one of the most well-known outlets for Android apps. This study examines trends in the Google Play Store app ecosystem in great detail, paying particular attention to important metrics like download numbers, categories, application sizes, user ratings, and features. The study attempts to identify patterns that support app success and user engagement by methodically analysing a wide sample of apps from various categories. Correlations between the download counts, user reviews, and the app's features are found using statistical techniques. Additionally, this study investigates how particular features—like in-app purchases, ad integrations, and regular updates—affect the functionality of apps. By examining the temporal evolution of trends, it aims to provide a thorough understanding of the mechanisms influencing the ecosystem of apps available on the Google Play Store. The results not only add to the body of knowledge regarding mobile app trends in academia, but they also provide developers, marketers, and app enthusiasts with useful advice on how to survive and prosper in the cutthroat app industry by streamlining their marketing strategies and, in the end, creating richer user experiences. This study provides an insightful look at the variables affecting an app's success in the ever-changing Google Play Store, particularly in light of the ongoing evolution of the mobile app industry.***

**Keywords:** Google Play Store, Mobile Applications, App Trends, Downloads, User Ratings, Features, App Success, Mobile App Ecosystem, App Marketing.

## I. INTRODUCTION:

In a world where digital experiences rule, mobile applications are essential to the way we interact and go about our daily lives. Google Play Store is one of the largest and most vibrant app stores out there, with millions of apps available. This study conducts a thorough trend analysis of the Google Play Store using a waterfall methodology to identify trends, correlations, and changes in user preferences.

The Google Play Store, a global centre for Android apps, captures a multifaceted environment where users can make a plethora of choices and developers can innovate. Comprehending the dynamics of trends is crucial for developers, marketers, and stakeholders who seek strategic insights in this dynamic ecosystem. Our examination takes place against the backdrop of earlier Play Store trends, which serves as a basis for dissecting the situation as it stands right now.

With an emphasis on installation statistics and user reviews, our methodology begins with the careful collection and classification of app metrics using a waterfall approach. After then, statistical techniques are used to identify trends and correlations and extract insightful information. In-depth analysis of category dynamics ensures a thorough grasp of the dominant patterns in the Google Play Store by adding contextual richness. This combination of approaches creates a strong framework that can offer stakeholders navigating the constantly changing app landscape strategic insights.

Some themes come to light when we consider previous trends. Utility apps, particularly those that deal with productivity, wellness, and health, have seen a noticeable increase in popularity. Apps that provide time management, fitness tracking, and meditation options are examples of how digital well-being has gained prominence. Furthermore, eco-friendly and sustainable lifestyle applications are becoming increasingly popular, reflecting the growing interest in sustainable living.

Examining the trends of today, we find a changing environment characterized by a fusion of functionality and user experience, with a stronger focus on entertainment. In the market, games, music & audio, tools, and education continue to be the top categories. Some noteworthy findings are that, even though they are smaller in size, arcade, puzzle, and casual games have a higher installation count. Even though they are less common, simulation and casino games are the most well-liked.

The study of recent patterns in the Google Play Store demonstrates the continued importance of genres such as entertainment, education, and games. Important factors become apparent for developers and companies working on application development. Games, Music & Audio, Tools, and Education make up the bulk of the market. In order to increase installation numbers and, consequently, ad revenue, developers might deliberately concentrate on developing smaller leisure-related applications. Developers have a strong incentive to investigate creating casino games in order to generate in-app revenue. The best-case scenario would entail a greater number of small arcade or casino games, which would provide an avenue for increased revenue generation.

We explore in the following sections the specific observations and statistical evaluations that support our comprehension of the subtle patterns found in the Google Play Store.

## II. APPROACH

We have implemented a rigorous, sequential strategy, starting with comprehensive data collection and cleaning followed by categorization of Google Play Store application data. The primary focus was to be on user reviews and app installation numbers since these are important measures of an app's popularity and success. We attempt to understand user preferences and market dynamics by identifying patterns and correlations in the data with the use of statistical tools and techniques.

Furthermore, our technique explores category dynamics in detail. This offers insightful background information to understand the changing landscape of the Mobile application marketplace.

The approach seeks to attain a thorough comprehension of the dominant markets and trends. Through a thorough analysis of around 2.5 million applications in various categories, we aim to identify the most common applications, particularly to understand the definitive characteristics as needed to develop a high demand low-cost application.

Our analysis goes beyond the previously listed techniques by incorporating analysis of user reviews and ratings, which allows us to gauge customer satisfaction and identify market demand and supply. Moreover, a sub-categorical study is

implemented to consider multiple demographics for a detailed understanding across different subdivisions of the market dominant category.

This paper touches upon another layer of insights by analysing application size as an indicator of production costs. The incorporation of these extra levels of research aims to provide stakeholders, marketers, and app developers with strategic insights that can help them make decisions as they navigate the ever-changing Google Play Store.

## III. WORKFLOW

### A. Data Collection

We started the process of data collection by gathering information about all the applications including names, categories, installation counts, user reviews, ratings, and surrounding metadata, from the Google Play Store. After collecting loads of data, we have segregated the most recent details about a variety of apps by using the Google Play Store API or web scraping methods. We take into account the size and feasibility of all quantifiable factors in our exercise of analysing the Android Mobile Application marketplace.

### B. Extract, Transform and Load

In the first step of ETL, it is very important to know about the details of the data. Therefore, we examine the collected data carefully and look for any anomalies, missing values or null values. Furthermore, we work at removing the missing values, anomalies and inconsistency of the data by performing various steps of data cleaning. In the step of Data cleaning, we have also dropped many columns such as Developers details which are of no use to analysis and create nuisance and could also reduce the efficiency. In addition, this step also resolved differences in app names or categories and dealt with duplicates, if any. This transformed data is further loaded into the data frame for performing analysis on the trends of applications on Google Play Store.

### C. Evaluation

The dataset is loaded into a python workspace, for statistical analysis and categorical research. To understand the popularity and performance of an app, we evaluate data regarding reviews, user ratings, and installation quantifiers. The paper also investigates the distribution of apps within various app categories, and spot prevailing trends to learn more about category dynamics. This evaluation is expressed in easy to interpret formats using graphs, using visualization tools like charts and graphs.

The above-mentioned processes can give an assurance that it is a methodical approach to trend analysis on the Google Play Store. Summarizing it, firstly we began with data collection, moving further to data transformation and cleansing, and in the end with an in-depth examination of user preferences, categorical dynamics of apps, and app trends.

## IV. DATA COLLECTION

The initial study phase involved a comprehensive data collection from the Google Play Store[11]. The dataset had different features that captured different aspects of each application. However, to ensure an objective analysis, some attributes were identified for exclusion during the data

cleansing process.

Sources: Detailing the methods and sources for collecting data on Google Play Store apps.

Criteria: Explaining the criteria for selecting particular attributes of apps for analysis.

Upon analyzing trends in Google Play Store apps, attribute selection is crucial. User performance and user preferences can be determined via each of these attributes as it renders different aspects to user's needs.

#### *A. App Name*

Essential for app identification. For any analysis it is fundamental, especially whenever tracking the popularity or perception of apps over time.

#### *B. App Id*

This unique identifier is vital for differentiating between apps, especially those with similar names. It's crucial for accurately aggregating data related to a single app.

#### *C. Category*

Comprehending the category of an app enables the analysis of trends within distinct market segments. It proves beneficial for contrasting the performance among various types of apps, such as games, productivity tools, and social media platforms.

#### *D. Rating*

This attribute is directly proportional to the user satisfaction and app quality. High ratings correlate with high app popularity and user retention.

#### *E. Rating Count*

Traces the level of user engagement and the feedback rate. A higher count can depict massive use and involvement, even if the rating is medium.

#### *F. Installs*

A key feature for app access and popularity. Higher install numbers often correlate with higher visibility and market penetration.

#### *G. Free vs. Paid (Price, Currency)*

This difference is critical for understanding market strategies. Willingness to pay for apps makes it easy to analyze trends in free vs paid applications.

#### *H. Size*

Decision to download applications is directly impacted by size of it, specifically in field with limited data capabilities or storage. Complexity and richness of app features might also be reflected by the size.

#### *I. Release Date and Last Updated*

Lifecycle and maintenance of an app can be strongly determined with these dates. Frequent updates might showcase active development and adaptation to user feedback.

#### *J. Content Rating*

Majorly affects building target audience and potential market size. It also helps in compliance and regulatory analysis.

#### *K. Ad Supported*

This gives an idea how advertising is increasing app revenue

model. Patterns and trends indicate user-tolerance towards these ads and in turn provide insights into marketing strategies.

#### *L. In-App Purchases*

Significantly diagnoses trend in monetization strategies. Unveils if an app depends on in-app purchases for revenue gains.

#### *M. Editors' Choice*

This attribute helps in categorizing high quality and trustworthiness, and analyzing these apps can reveal trends in what the platform considers exemplary.

## V. DATA CLEANING

Understanding the critical role of data cleaning in ensuring precise and reliable analysis of Google Play Store app trends.

#### *A. Redundant Entries*

The presence of multiple records necessitated the identification and removal of redundant entries. The importance of this stage cannot be overstated since it helps to maintain reliability and integrity in the data set thereby preventing skewed analyses. To maintain that we successfully managed to drop two columns "installation of device" and "minimum installs" and used a third column with aggregate installments of all applications. Redundancy issue was taken care by dropping unnecessary columns which added no value in our dataset.

#### *B. Rectifying Absent or Irregular Values*

Preservation of data integrity was through rectification of absent or irregular values; missing data points were addressed by filling with essential information, here we filled ratings attribute with average of category wherever missing, as rating needed data to analyze user satisfaction and app quality in turn determining app popularity and user retention. Also dropping null values where required while outliers were corrected upon flagging them for later investigation.

#### *C. Standardization of Data Formats*

To facilitate a seamless analysis, there was need to standardize data formats in terms of aligning units, date formats, or such relevant parameters for consistency across datasets. For the same we formatted ratings and converted it to type integers, introduced functions in code to convert essential size attribute from GB, KB to MB, therefore managing to format all sizes to MB as per requirement.

#### *D. Exclusion of Attribute*

Certain attributes such as Minimum Androids, Developer ID, Developer email, Developer website, Privacy Policy and Scrapped time were considered not necessary in looking into trends at hand. Eliminating these will therefore help to streamline all processes involved in subsequent steps.

#### *E. Resulting Dataset*

The final step has refined this dataset thereby making it more accurate and relevant as well. The resultant dataset is now poised for a more effective and insightful examination on trends at Google play store. In conclusion precise insights can only be gotten from meticulous collection as well as cleaning up of data sets thus non-essential attribute have to be excluded

during cleaning process so as to streamline a dataset making it more focused and efficient towards analysis. Insights precision and reliability, therefore, would give an all-rounded comprehension of Google play store applications trends for informed decision making in the dynamic digital space.

## VI. RESULTS AND DISCUSSIONS

There are millions of mobile applications uploaded by developers daily [3]. For developers, it is crucial to understand the marketplace before developing applications for the google play store. More specifically, with the development of numerous applications without check and balance, there is obfuscation of application code, leading to duplicative applications[7] and hence users' trust towards the Google play store is facing a major crisis, as user data can easily be fetched by application in the store and confidential information no longer seems to be confidential[4]. In this paper, we discuss the development of applications, based on market trends, for developers to be able to target the broader spectrum of users utilizing applications available on the Google play store[8][9]. To perform this analysis, we utilize python programming in cohesion with the wide variety of libraries available for data visualization such as matplotlib, seaborn, plotly, express and more.

The aim of the visualization will be to initially understand the categories and the number of installations for the various categories individually. Furthermore, we focus on the various number of applications for the distinct categories in the marketplace, to understand the focus on application development. Based on this analysis we understand which categories to further analyze to understand the popularity based on the number of installations. The paper also looks at the size and complication of the application to understand the work hours for development.

Information about the user ratings is also an important criterion to understand the popularity of categorical applications in the marketplace. This visualization is more helpful for game developers in the development phases, and for users of game applications for the selection of the game they want to play.

### A. Number of Installs by category

The applications in Google Play store are categorized by Google for developers to be able to understand what type of application they can develop. Based on our data cleaning and preliminary data processing we observe the following set of categories in the data(Fig 4.1).

	Category
0	Action
1	Arcade
2	Casual
3	Communication
4	Entertainment
5	Photography
6	Productivity
7	Social
8	Tools
9	Video Players & Editors

Fig 4.1: Sample Categories in Dataset

We further develop a bar plot to understand how many installations per category we have available. From the plot we identify that there are multiple sub-categories that can be merged into a single category called games.

By refining data to merge the installation metric for these sub-categories of games we are able to obtain the re-plot the same chart as follows.

This analysis helps us understand that the number of installs in Google Play Store is dominated by Gaming applications which is then followed by Tools and Communication applications. This helps us understand the huge presence and demand for gaming applications in the marketplace.

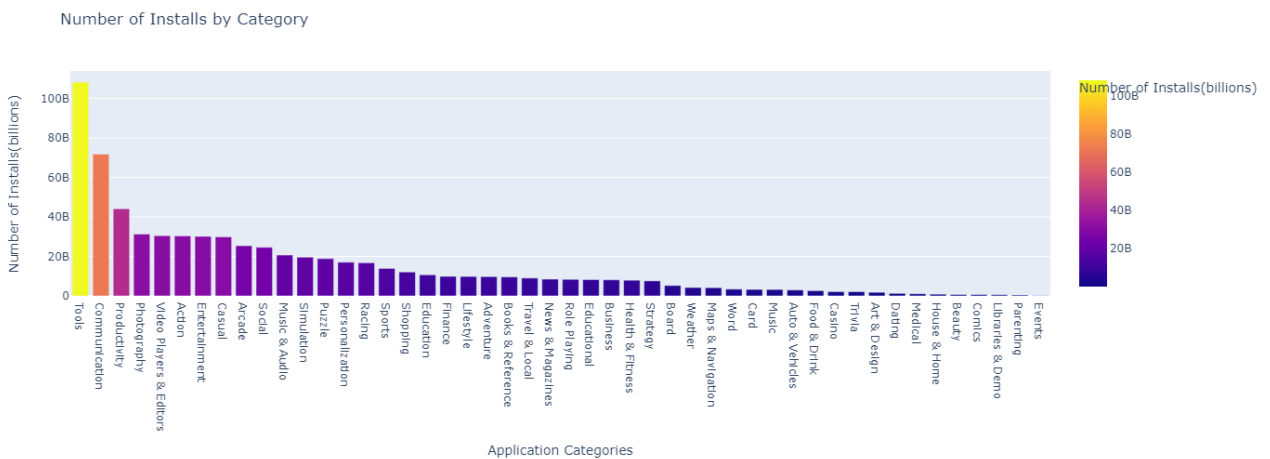


Fig 4.3 Bar plot of Number of installs by Category

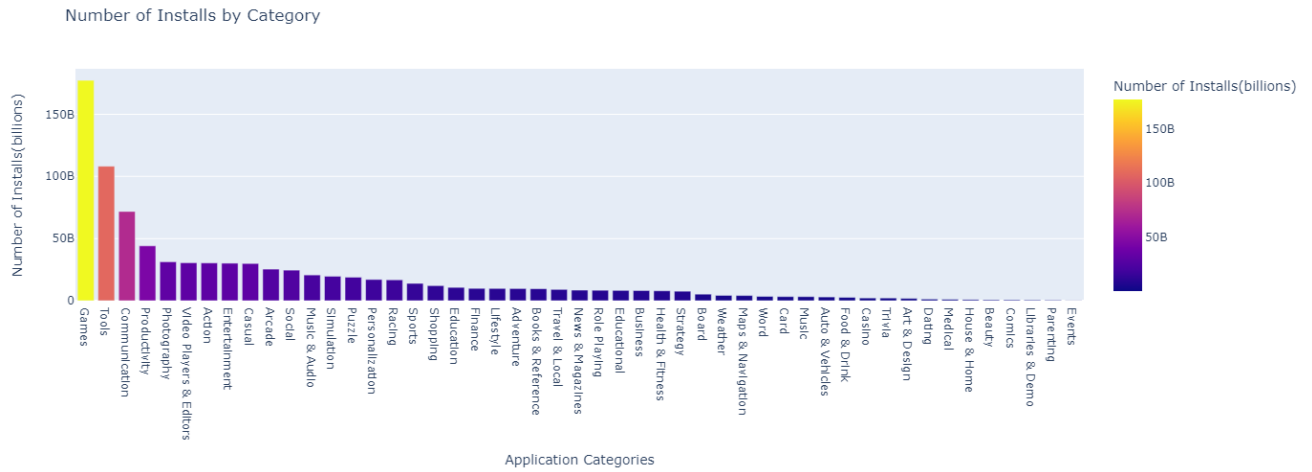


Fig 4.3 Bar plot of Number of installs by Category (after re-categorization of data)

### B. Number of Application by category

When looking at demand, as done above, the supply for applications must also be understood for developers to ensure that they enter a high revenue market[10]. The supply of

applications can also be understood by analyzing the number of applications present in Google Play Store for every category. A bar plot of the number of applications in all categories as in Fig 4.4 helps us do exactly that.

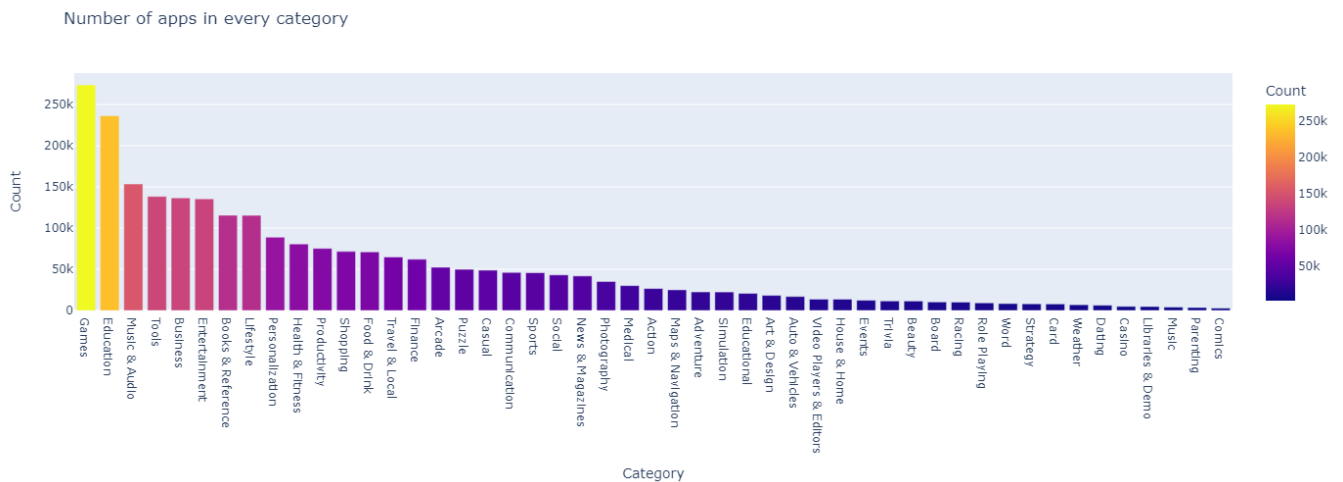


Fig 4.4: Number of mobile applications in every category

As observed, the demand for gaming applications has an equivalent high supply of Gaming applications in the Marketplace, followed by education, music & audio and Tools. Since we observe a high demand high supply gaming market, we further analyze the Games categories to drive better insights.

### C. Number of Application per Gaming category vs Total installs per category

The understanding of gaming applications leads us to believe that the different types of categories is an indicator for the development process. The bar vs line plot in Fig 4.5 helps us understand the different supply vs demand of the gaming sub-categories.

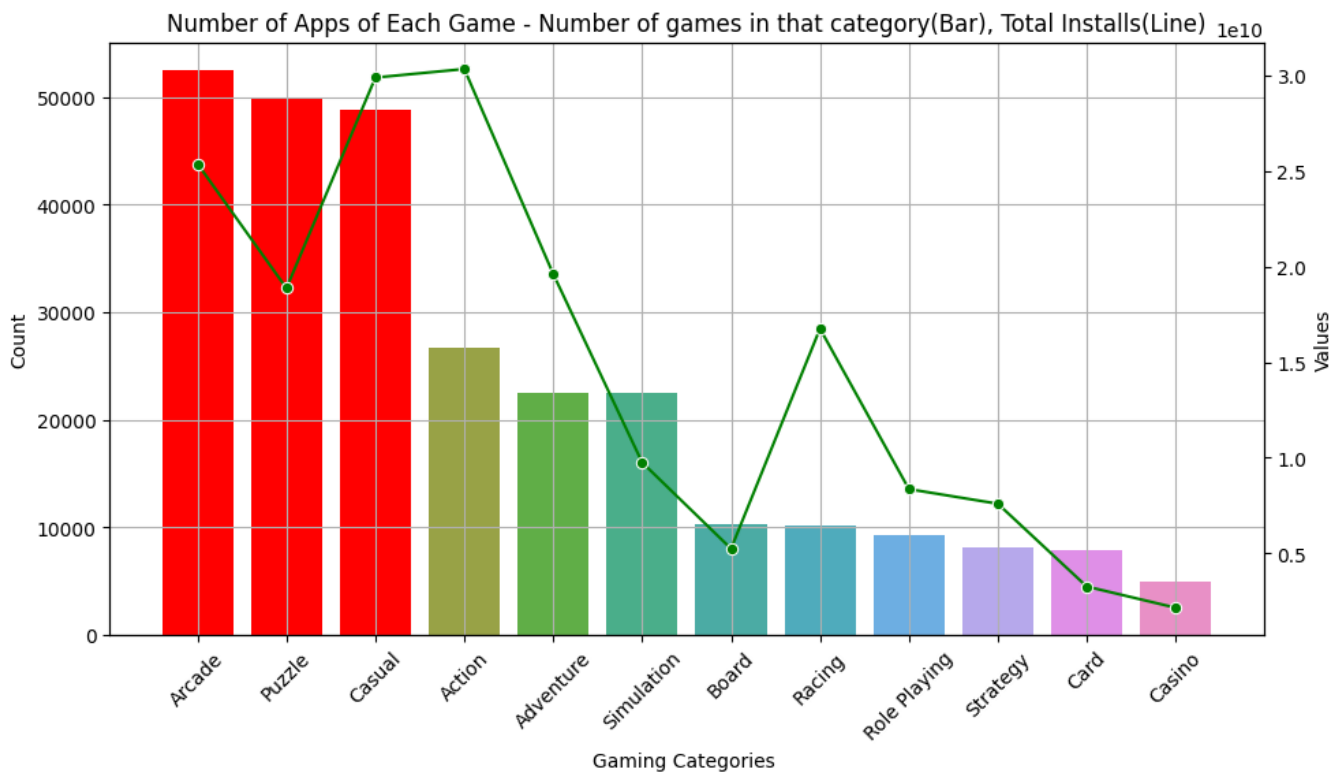


Fig 4.5 Demand (Number of Apps – Bar) vs Supply (Total Installs - Line)

The availability of arcade, puzzle and casual games is highest in Google Play Store. Action, Adventure and Simulation games have a moderate supply while Board, Racing, Role playing, Strategy Card and Casino games are among the least. The line plot indicates that users are most interested in Casual, Action, Arcade games and Racing, Puzzle, Adventure games lie among moderate demand range.

development of these games can be beneficial next to Adventure and Action games.

#### D. Size based analyses

Game development is a difficult task which requires high man hours which would translate into raised production costs. The size of games can be a decent indicator about how development intensive the category might be.

With this information we can state that since Racing games are among the least supplied and higher and demand,

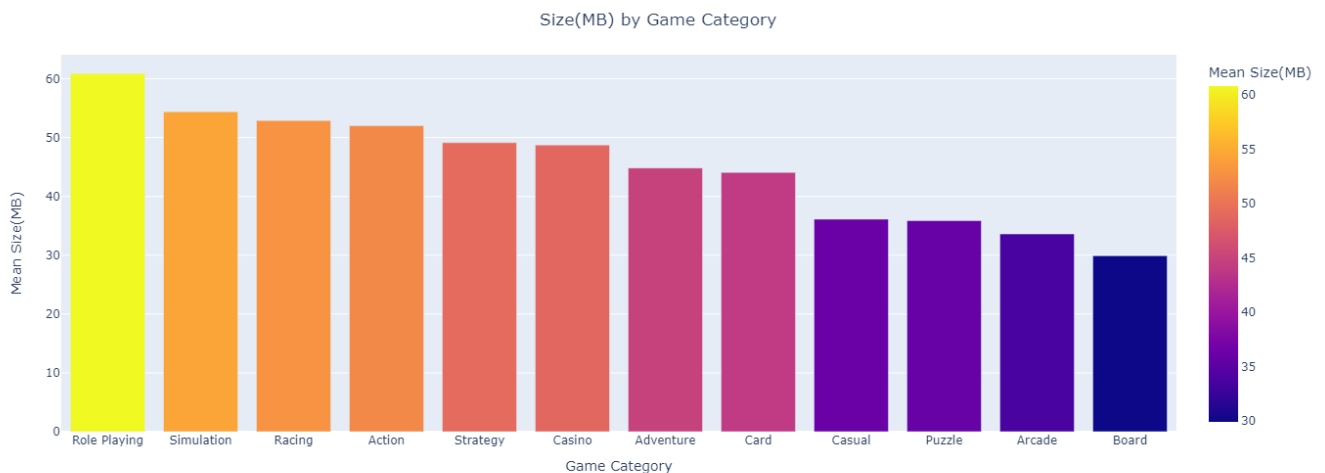


Fig 4.6 Average Size of applications in gaming categories (in Megabytes)

Fig 4.6 outlines the average size of application in gaming categories and shows that Board, Arcade, Puzzle and Casual games are amongst the least in Size(in megabytes). The Role playing, simulation, action and racing games are among the

largest in size with average sizes of 55 – 60 Megabytes. This is indicative of higher production costs induced by the development of games in these categories.

### E. App Rating

The indicator of user supply along with number of installs is user provided feedback i.e. application rating. Fig 4.7 illustrates the variation in average rating of all application of

Google Play Store. We observe that rating average between 2.4 and 3.4, which is not a vast deviation, indicating that all gaming applications are rated on a similar scale. The chart also points toward a higher rating for Role Playing and Casino and Simulation games while it has a lower rating for Casual and Arcade games.

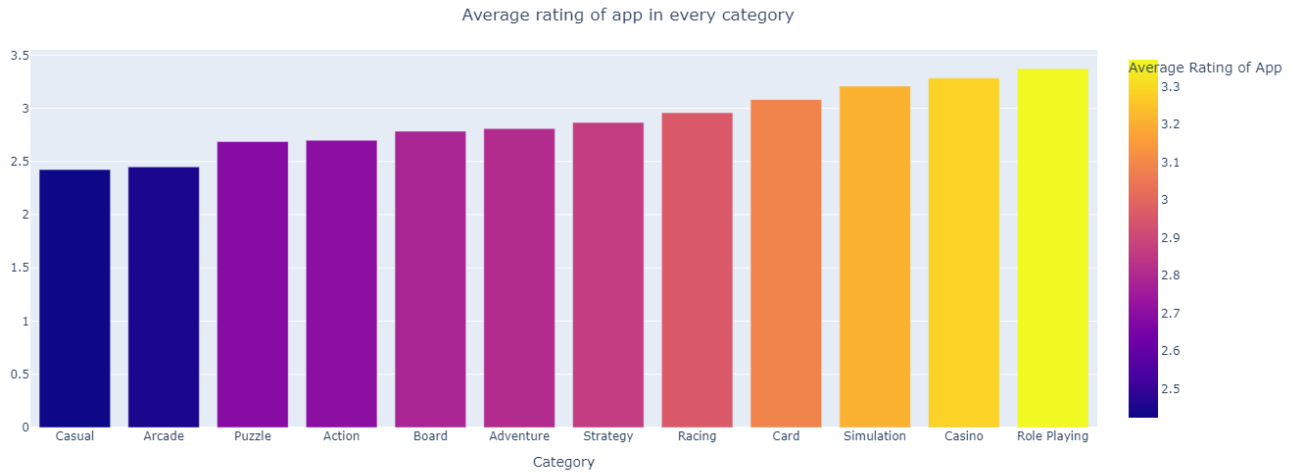


Fig 4.7 Average rating of Application in each category

Since this metric is user provided and subjective it can be considered as informative but skewed[6] and hence must be interpreted keeping various variables such as spam ratings, higher number of ratings causing a skew, improper application development in any particular category leading to higher number of bad reviews and many similar factors. While keeping these factors in focus, no strong insight can be derived. In addition to this information, we also consider the fact that the average rating is mediocre although it is the most demanded category in the marketplace, which is indicative of the lack of reviews.

## VII. CONCLUSIONS AND FUTURE WORK

In this paper, by using a Google Play Store dataset scraper built using python and libraries, we build a dataset of 2.5 million application available on the android mobile application marketplace [3]. The major applications available in the Google Play Store Marketplace are Gaming, Music and Audio, Education and Tools. These types of application also have various sub-categories present in the store. Further, the paper dives into the games category and its subcategories to understand the optimum application development which will lead to high revenue generation for developers and companies. The presence of demand and supply is almost equivalent except for categories of Action, Adventure, Racing and Casual games. When trying to understand development costs, we see that Casual, Arcade and Adventure games have a relatively low size, which we interpret as a low – moderate production cost. In addition to this we also consider user feedback for the games. This drives us to the conclusion that development of Casual games would be ideal. Furthermore, to enhance the user interaction, developers can focus on developing Casual games that have elements of Adventure and/or Arcade/ Action would be ideal as this

multi-categorization would bridge the supply-demand difference while ensuring a moderate production cost as well as good user interaction.

The dimension to look at in terms of revenue generation would be to understand the presence of Free and Paid applications in the marketplace. This is reviewed in [1] and shows us paid Arcade games have a higher number of installs which is cohesive with our analysis of developing Casual Arcade games. Additionally, development of free to install games should also be considered as it would be revenue generating through in-app purchases and advertisement revenue. In this scenario it is beneficial to develop games with higher installs, which would lead to higher ad-revenue generation. This is a pointer to develop casual or action games. The involvement of factors such as development duration, organizational vs individual development and length of in-app time [2] can be analyzed to understand better what application would be optimal.

## REFERENCES

- [1] R. M. Amir Latif, M. Talha Abdullah, S. U. Aslam Shah, M. Farhan, F. Ijaz and A. Karim, "Data Scraping from Google Play Store and Visualization of its Content for Analytics," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-8, doi: 10.1109/ICOMET.2019.8673523.
- [2] Wu Zhou, Yajin Zhou, Michael Grace, Xuxian Jiang, and Shihong Zou. 2013. Fast, scalable detection of "Piggybacked" mobile applications. In Proceedings of the third ACM conference on Data and application security and



privacy (CODASPY '13). Association for Computing Machinery, New York, NY, USA, 185–196.  
<https://doi.org/10.1145/2435349.2435377>

[3] AppBrain. Number of Available Android Applications. <http://www.appbrain.com/stats/number-of-android-apps>. Online;

[4] Haoyu Wang, Hao Li, Li Li, Yao Guo, and Guoai Xu. 2018. Why are Android apps removed from Google Play? a large-scale empirical study. In *Proceedings of the 15th International Conference on Mining Software Repositories (MSR '18)*. Association for Computing Machinery, New York, NY, USA, 231–242.  
<https://doi.org/10.1145/3196398.3196412>

[5] Spreitzenbarth, M., Freiling, F., Echtler, F., Schreck, T., & Hoffmann, J. (2013, March). Mobile-sandbox: having a deeper look into android applications. In *Proceedings of the 28th annual ACM symposium on applied computing* (pp. 1808-1815).

[6] M. -Y. Day and Y. -D. Lin, "Deep Learning for Sentiment Analysis on Google Play Consumer Review," 2017 IEEE International Conference on Information Reuse and Integration (IRI), San Diego, CA, USA, 2017, pp. 382-388, doi: 10.1109/IRI.2017.79.

[7] Dominik, W., Nicolas, H., Yasemin, A., Brad, R., Patrick, T., & Sascha, F. (2018). A large scale investigation of obfuscation use in Google Play. In *Annual Computer Security Applications Conference*.

[8] S. Satrio, E. K. Budiarjo and M. Rifki Shihab, "Minimum viable product analysis in hyperlocal marketplace applications using customer development method," 2022 10th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 2022, pp. 334-339, doi: 10.1109/ICoICT55009.2022.9914834.

[9] H. Pryshchepa and N. Kunanets, "Requirements for the application for replacement on the marketplace to promote IT product," 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2022, pp. 388-391, doi: 10.1109/CSIT56902.2022.10000576.

[10] Blank, S., & Dorf, B. (2020). *The startup owner's manual: The step-by-step guide for building a great company*. John Wiley & Sons.

[11] [https://www.google.com/aclk?sa=l&ai=DChcSEwj06CogYuDAXUQR0cBHfBIBGMYABAAGgJxdQ&ase=2&gclid=CjwKCAiApuCrBhAuEiwA8VJ6Jo866RGbCxfnO8R5cCxjr\\_N7B8GX9Ozm9MX-MmBL9ZaKTT5I9QkieRoCDxAQAvD\\_BwE&ei=Jed4ZYfxKPGx5NoP8pOCQA&sig=AOD64\\_3mSunU\\_Wjg6tj2VZFNBxJ7FYBRiA&q&sqi=2&nis=4&adurl&ved=2ahUKEwjHl5qogYuDAXXxGFkFHfKJAAgO0Qx6BAgJEAE](https://www.google.com/aclk?sa=l&ai=DChcSEwj06CogYuDAXUQR0cBHfBIBGMYABAAGgJxdQ&ase=2&gclid=CjwKCAiApuCrBhAuEiwA8VJ6Jo866RGbCxfnO8R5cCxjr_N7B8GX9Ozm9MX-MmBL9ZaKTT5I9QkieRoCDxAQAvD_BwE&ei=Jed4ZYfxKPGx5NoP8pOCQA&sig=AOD64_3mSunU_Wjg6tj2VZFNBxJ7FYBRiA&q&sqi=2&nis=4&adurl&ved=2ahUKEwjHl5qogYuDAXXxGFkFHfKJAAgO0Qx6BAgJEAE) – Google Play Store

[12] <https://www.kaggle.com/datasets/gauthamp10/google-playstore-apps> - Googler Play Store Dataset