In-Depth Analysis of Mobile Apps Statistics: A Study and Development of a Mobile App

[3]Maryam Abbasi,
[1]André Lopes, [1]Diogo Rodrigues,[1]João Saraiva,
[2]Pedro Martins, [2,3]Filipe Sá, [4,5]Filipe Cardoso

[1] Polytechnic of Viseu, Portugal
[2] CISeD - Research Centre in Digital Services, Polytechnic of Viseu, Portugal
[3] Polytechnic of Coimbra, Portugal
[4] INESC Coimbra, Portugal
[5] Polytechnic of Santarem

[3]maryam@dei.uc.pt,
[1]pv26447@alunos.estgv.ipv.pt, [1]pv26448@alunos.estgv.ipv.pt, [1]pv26721@alunos.estgv.ipv.pt,
[2]pedromom@estgv.ipv.pt, [2,3]filipe.sa@isec.pt, [4,5]filipe.cardoso@esg.ipsantarem.pt

***Abstract*—With the popularity of smartphones and mobile devices, mobile application (a.k.a. "app") markets have been growing exponentially in terms of the number of users and downloads. To increase user satisfaction, app developers invest a lot of work into gathering and utilizing user input. This paper presents an analysis of the mobile app market through the development of a mobile app. The app provides users with an overview of the most important statistics, including the best apps in each category, the categories with the most apps, and the overall statistics. The data was collected through the analysis of publicly available annual reports, and presented in a user-friendly format through the use of graphics. The focus of the study was to provide a useful tool for developers and individuals seeking specific statistics. Although the approach has proven to be effective, the authors suggest potential, such as incorporating live data from the Google Play Store and analysing the App Store. Additionally, comparing the data from multiple years can provide useful insights into the evolution of the market.**

***Index Terms*—Play Store, Mobile app, Data set, Data Analysis, Big Data**

I. Introduction

The mobile app industry is experiencing exponential growth, and this growth has brought with it a new set of challenges for developers. Keeping up with the ever-evolving market and user demands is one of the most significant challenges that developers face. In today's world, most mobile apps are distributed through the Google Play Store, which acts as a vast repository of information about various apps, such as their descriptions, reviews, costs, and downloads. However, gathering and analyzing this information can be an extremely challenging task for developers.

Traditionally, developers have used various standard channels to gather user feedback, such as bug/change repositories, online forums, and emails. However, these methods can be time-consuming and may even increase the project's development costs. This is why developers require a more efficient and user-friendly solution that can simplify the analysis of important app statistics.

To address this challenge, this study presents the development of a mobile app that can make the analysis of app statistics much easier. The app provides users with an intuitive interface that enables them to view and understand key statistics from the Google Play Store, such as the best-performing apps in a given category, the most significant categories, and more. The app is designed to be user-friendly, and it uses graphics and explanations to make it easier for users to get a comprehensive overview of the data, instead of presenting raw data.

The development of this mobile app was focused on creating a tool that is both user-friendly and comprehensive. The app's interface is aimed at developers and those looking for specific statistics, but it still provides a comprehensive research tool. The app enables users to easily retrieve statistics on a specific app, compare two apps, view the most significant categories, and check overall statistics.

The analysis and development of this mobile app has revealed some valuable insights into the app development process. Although the study is limited to the analysis of annual reports instead of current data, it still highlights the significance of having a good overview of key information for app development success. The app provides developers with a comprehensive tool that can help them make informed decisions about their app development projects. It enables them to understand the current market trends and demands, as well as identify the most successful apps and categories in the market.

The remaining of this paper is organized as follows. Section II, covers the related work analysed for this project. Section III, focuses on the architecture of the proposed system. Section IV describes our methods for evaluating data quality. Section

2023 18th Iberian Conference on Information Systems and Technologies (CISTI)
20 – 23 June 2023, Aveiro, Portugal
ISBN: 978-989-33-4792-8

V, contains the final product results. Finally, in Section VI, conclusions are drawn, followed by the introduction of future work guidelines.

II. State of the art / Related work(s)

In recent years, the growth of the mobile app market on platforms such as Google Play and the App Store has presented a unique opportunity and challenge for developers. With an abundance of apps available on these platforms, it can be difficult for developers to stand out and attract a substantial user base. To tackle this challenge, we propose the use of an information system that can provide valuable insights to app developers.

Our system analyses the analytics of top-performing apps and categorizes them into similar groups, offering developers a glimpse into the most appealing features and characteristics for users. By using annual reports from Google Play and the App Store, we aim to identify trends and patterns in the market and provide developers with the necessary tools to succeed in this competitive landscape.

The study [1], presents analysis of data-usage in mobile apps. The authors conducted this study to better understand and explain why and how the data collected is used. They also explored the area of app privacy policy and verified if the app policy aligned with the actual usage of the data. The focus of the paper was on the stated data-usage purposes and the real practices, to enhance trust and increase the lever of transparency in the usage of user data by mobile apps.

Existing studies in this field have explored various aspects of the mobile app market. For example, [2] presents a framework for extracting relevant data from user reviews, [3] focuses on the technical challenges of mobile app testing, and [4] uses a theory-based model to analyse the consumption of mobile apps.

In [5], a literature review of the field of mobile app development. The authors conduct a comprehensive analysis of the area of mobile app development, and they identify key researches and trends. Also, it is presented the challenges and opportunities in the field, as well as some guidelines for future research.

[6] investigates the unique market structure of the mobile app market through an analysis of Google Play transactions.

A systematic review of mobile app development research was performed by [7], emphasizing the requirement for increased user-centered design and development in the field. [8] conducted an empirical study of the adoption and utilization of mobile apps, with results indicating that perceived usefulness and ease of use significantly influence user adoption and usage. The findings of the study offer crucial understanding into how users perceive and embrace mobile apps.

[9] studied users' mobile app behaviour through an empirical study. The results showed that users are influenced by various factors such as app features, functionalities, and the app's reputation. The study provides valuable insights into the factors that influence user behaviour and can be useful in designing mobile apps that meet user needs.

Our project builds upon these prior studies of behaviour by presenting the analysed data in a user-friendly mobile app that facilitates easy access to important statistics, such as the most popular categories, the best-performing apps, and the overall market trends. With our app, developers can gain valuable insights and make informed decisions when developing their own apps.

III. Architecture

The architecture of the proposed solution is a well-integrated system that utilizes multiple technologies and components to deliver an efficient and effective solution for managing and processing data from the Google Play Store. The interaction and communication between components are clearly detailed in Figure 1, providing a comprehensive understanding of the system's design.

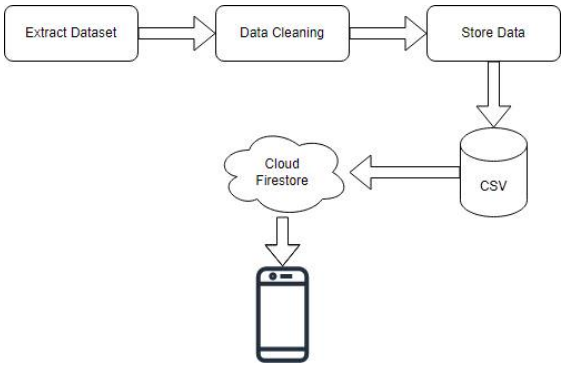

Fig. 1: System architecture

The solution is built upon a filtered CSV data set, which contains essential information for each app available on the Google Play Store, including details such as app name, category, rating, number of installs, price, size, developer, and other relevant data. To manage and process this data, the project employs the use of Cloud Firestore, a flexible and scalable NoSQL database offered by Google.

An authentication system has been implemented using Firebase Authentication, providing secure and reliable email and password verification. Python was utilized to import the data from the CSV file into Firestore, facilitated by the use of the Python Firebase library for data management operations. The mobile app, developed in Kotlin, follows an offline-first approach and an MVVM architectural pattern for maintainability and testability. In addition, Dagger2, a code generation-based dependency injection framework, was utilized for code reuse and efficient dependency management.

To ensure smooth and efficient image loading and caching, the solution employs the use of Glide, an open-source library with a user-friendly API. Firestore was also used for data storage and synchronization, offering real-time updates and offline support through caching a copy of the data on the

2023 18th Iberian Conference on Information Systems and Technologies (CISTI)
20 – 23 June 2023, Aveiro, Portugal
ISBN: 978-989-33-4792-8

device. Efficient retrieval and display of large data sets were ensured through the implementation of pagination, which improves performance and adapts to changing data.

Data analysis and visualization are achieved through the use of Matplotlib and Seaborn, two highly-regarded data visualization libraries in Python. These libraries generate charts and plots to present the data in a clear, concise, and easy- to-understand manner, allowing users to quickly grasp the information and insights provided by the solution.

In conclusion, the proposed solution is a robust and comprehensive approach to managing and processing data from the Google Play Store. By incorporating various technologies and components, the solution delivers a flexible, scalable, and user-friendly solution for developers and marketers to make informed decisions about their app strategies. The system's architecture is well-designed, ensuring reliable and efficient data management, processing, and analysis.

IV. Experimental Setup

In the following sections, we will describe each step of our project, from data collection to final implementation. The first step is to obtain the data set, which we will describe in more detail. We then filtered the data to only include relevant information, which we stored in a new CSV. We implemented indexes in our database to allow for efficient data querying. We also used visual representations to help understand the trends and patterns in the data. Finally, we created a mobile application with a user interface that displays the results of our data analysis.

The data set for this project was collected from the Google Play Store using a Python script called Scrapy. The collection was performed by Gautham Prakash in June 2021 on a cloud virtual machine instance. The data was saved in a CSV file and can be accessed on Kaggle. The data set includes 2.2 million unique values representing 2.2 million distinct apps, each characterized by 24 parameters such as App Name, Category, Rating, Installs, Price, Size, and Developer. This data set serves as the foundation for all analysis in this project [10].

The original data set had 2.2 million rows and 24 columns, but we filtered it to improve accuracy and avoid exceeding the Cloud Firestore system limits. Some data was incorrect or lacked useful information, while the Firebase Spark plan had daily limits of 20,000 writes and 50,000 reads. To reduce the data set, we established criteria that removed apps with 0 rating, no downloads, corrupted data, and kept only those with over 1 million downloads. This resulted in a reduced data set of approximately, 11200 apps, which was sufficient for building an efficient information system with quick response times.

To improve query performance and efficiency, we created various indexes on the collections, as shown in Figure 2, to ensure that our queries would execute quickly and effectively, even as the data set grew in size. This enabled us to filter and sort on specific fields and resulted in faster query performances.

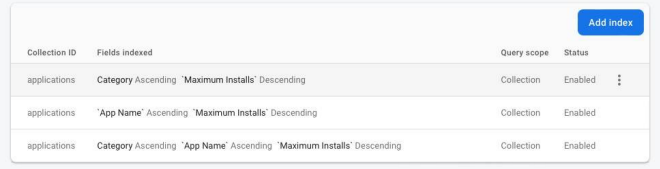

Fig. 2: Indexes created in Firestore

Without indexing, Firestore would have had to scan all documents in the database, leading to slower and more intensive queries. Therefore, indexing was essential for good performance and a smooth user experience in our application. To aid in comprehending the data in our set, we employed Python to generate visualizations. These graphics offer a clearer perspective on data distribution, enabling the app user to identify trends and patterns in the data. This visualization process enables a deeper understanding of the data and informs our analysis and decision-making.

A. User Interface

Our mobile application provides a comprehensive solution for exploring and analysing Google Play Store apps. With interactive visualizations and intuitive navigation, users can easily compare and evaluate various apps to make informed decisions. The interface (Figures 3, 4 and 5) is designed to be user-friendly, providing quick access to detailed information about individual apps, as well as the ability to compare data between two apps.

In addition to these core features, the mobile application also offers several tools to help users better understand the data. These tools include the ability to:

- Compare different apps on the Google Play Store;
- Save favourite apps for future reference;
- Access information on the most popular and highly rated apps;
- Search for specific apps based on name, category, or developer;
- Browse apps by category to discover new and trending apps;
- Keep track of recently viewed apps for quick access;
- Receive daily suggestions for apps to explore, based on user preferences.

By combining these features, our mobile application provides a one-stop-shop for app discovery and analysis. Whether you're looking for a new game to play or comparing features and pricing for a particular app, the app provides all the data you need to make an informed decision. Our app is designed to help users make the most of the Google Play Store by providing an accessible and engaging way to explore and analyse the data.

V. Results and Analysis

This section is split into two sections: first, we will show the results of our analyzes of Google Play Store apps and second shows performance indicators of our mobile app:

2023 18th Iberian Conference on Information Systems and Technologies (CISTI)
20 – 23 June 2023, Aveiro, Portugal
ISBN: 978-989-33-4792-8

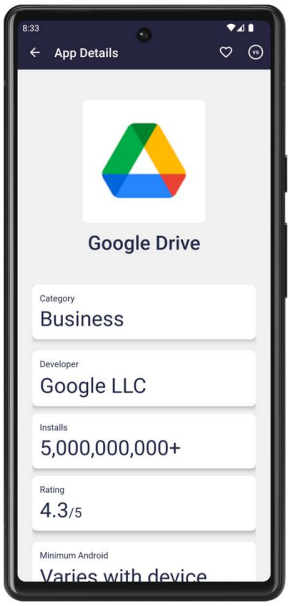

Fig. 3: App Details Screen

Fig. 4: Compare Apps Screen

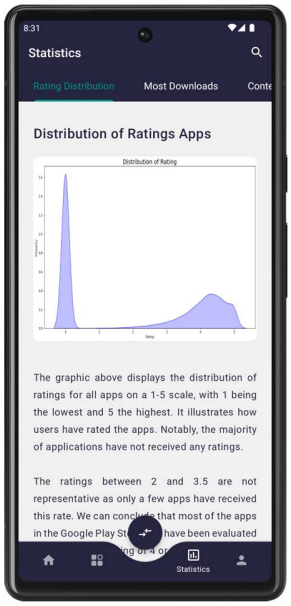

Fig. 5: Statistics Screen

A. Analyzes of Google Play Store Apps

The distribution of ratings for all the apps in the Google Play Store ranges from 1 to 5, with 1 being the lowest and 5 being the highest. It is observed that a majority of the apps have not received any ratings, indicated by the high number of ratings at 0. The majority of the evaluated apps receive a rating between 4 and 5. A limited number of apps receive a rating between 2 and 3.5.

Categories with the most downloads help app developers and marketers understand where they can focus their efforts for growth. In the data set, there are 48 different categories, with education being the most popular, followed by music and tools.

Distribution of content types contained in the data set. The content type is organized by the number of apps, with five types: Everyone, Teen, Mature 17+, Everyone 10+, and Adult only 18+. It is observed that the majority of the apps, 87%, have a content type of ”Everyone”. This is followed by ”Teen”, ”Mature 17+”, ”Everyone 10+”, and ”Adult only 18+”. ”Unrated” refers to apps that do not have this field filled in the data set, while ”Adult only 18+” has a 0% representation in the data set with only 65 apps. This suggests that the majority of these apps are intended for public use. This trend indicates that many developers have designed their apps with the intention of making them accessible to a broad range of users, without any age restrictions. Furthermore, this trend also indicates that many of these apps are ”family-friendly,” which means that the content of these apps is appropriate for all ages. This is because many brands that pay for advertisements are often restrictive about the type of content that they associate with their brand. As a result, app developers who create apps with no restrictions on type or age can increase the number of downloads, and the number of apps, and consequently generate more revenues.

Apps with ”Ad-Supported” represents 50% of the market. That means the app is funded through advertising revenue, so will show advertisements within the app, such as pop-ups, ads, banners etc. The remaining 50% are ”ad not supported”, so are not funded through advertising revenue.

The market division between free (98%) and paid (2%) apps. The size of the section representing free apps indicates that the majority of the app market consists of free apps. The smaller section representing paid apps indicates that a significant, yet smaller, a portion of the app market consists of paid apps (approximately 45,000 apps). This data highlights the diversity of options available to app users, with both free and paid apps offering different features and functionality.

As mentioned before, only 2% of overall apps are paid, however, some of these apps can be really expensive, typically when they are specialized for professionals in software, security medical or architecture. Analysing the 3 most expensive apps, the most expensive ones reached a value of 400$. The 1st is ”MESH Connect” and is related to connection security, category ”Productivity”. The 2nd is ”TEST EGY”, which costs 399.99$, category ”Tools” and the 3rd ”Taxes” one costs also 399.99$ and the category is ”Business”.

Making distribution of paid apps by price tiers is visible in figure 6 the bigger amount of apps costs under 50$ (40 000 apps). The gap between 50-200 $ has no market and there are around 4000 thousand apps that cost more than 200$.

In Graphic 7, the bar chart shows the developers with the most downloads, with the size of each wedge representing the number of downloads. This tool allows for easy comparison of the relative proportions of downloads among different developers, providing insight into the distribution of downloads among the developers.

2023 18th Iberian Conference on Information Systems and Technologies (CISTI)
20 – 23 June 2023, Aveiro, Portugal
ISBN: 978-989-33-4792-8

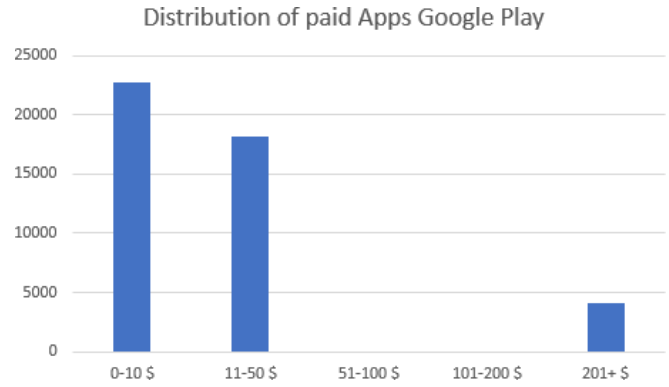

Fig. 6: Distribution of paid apps by price

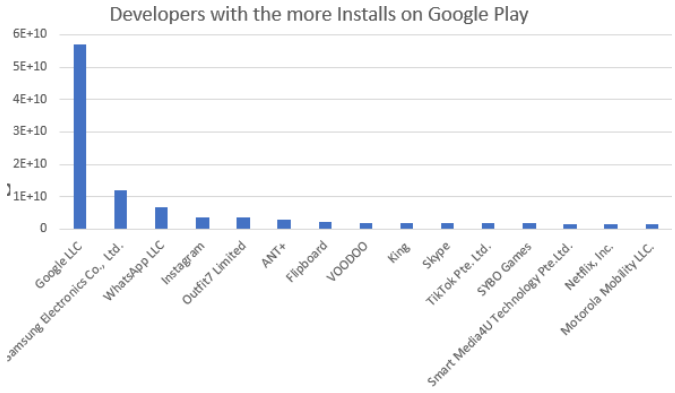

Fig. 7: Top Developers By Number of Downloads

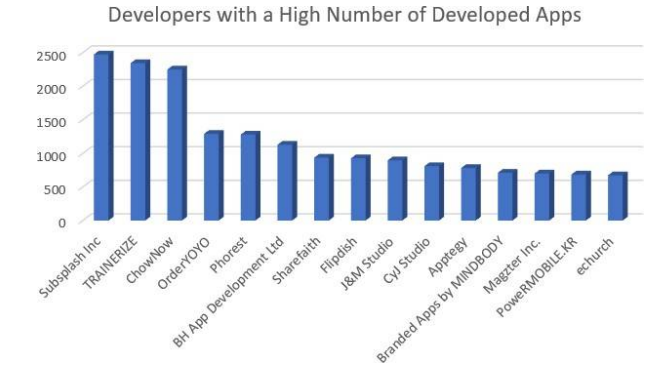

Fig. 8: Top Developers By Number of Apps

Graphic 8 shows a ranking of developers based on the number of apps they have launched in the Google Play Store. The developers are ranked from highest to lowest, with the top developer having launched the most apps. "SubsPlash Inc", "Trainerize" and "ChowNow" are the most relevant developers in terms of the number of apps launched, each having launched over 2,000 apps.

Editor's choice is a designation given to certain apps or games in app stores as Google Play, which means the app has been selected by the app store's team as a high-quality app. Only 853 apps have this distinction and, 2312091 doesn't. if we analyse this number in the total number of existing apps, this value does not reach 1%. So, almost 100% of the apps don't have this certificate.

Some applications have the possibility to buy additional content, and features within an already installed app. This feature allows developers to offer additional value to their users and earn additional revenue from their apps. Examples are unlocking the next levels, game content, premium features etc. It is possible to verify that only 8% of all apps have this ability, and the remaining 92% don't offer the possibility to buy extra features.

The size of apps on Google Play Store can vary greatly, with the majority of applications ranging from a few megabytes to hundreds of megabytes in size. The exact size of an app depends on various factors, including the complexity of the app's functionality, the amount of multimedia content it includes, the platform it is built for and so on. The graphic 9 shows the distribution of apps by size gaps. A large part of apps (70%) has between 1-10 MB and the second bigger group have 29% of apps, with a size of 10-99 MB. This means that almost all apps have less than 100 MB. The remaining 1% is for apps with more than 100 MB. The distribution of application sizes within the Google Play Store serves as an indicator of the concern that developers possess with respect to limiting their applications' size as much as possible. This caution can be attributed because applications with larger sizes will require extended download and installation times, which can result in frustration on the part of users, particularly those with slower internet connections. Additionally, in cases where the app is particularly large, the application may occupy a significant portion of the device's storage, which can limit the number of other applications or files that can be accommodated. It is imperative to have a good balance between providing the necessary features and functionality while optimizing the application's size and download time. Thus, the significant portion of applications available on Google Play is less than 100 MB to improve the number of installs and profitability. This is essential, given that consumers tend to gravitate toward applications that provide essential functionality without consuming excessive amounts of storage on their devices.

B. Our Mobile App

After developing our application, we conducted an extensive research to gain an in-depth understanding of its overall performance and behaviour. Our analysis revealed that the

2023 18th Iberian Conference on Information Systems and Technologies (CISTI)
20 – 23 June 2023, Aveiro, Portugal
ISBN: 978-989-33-4792-8

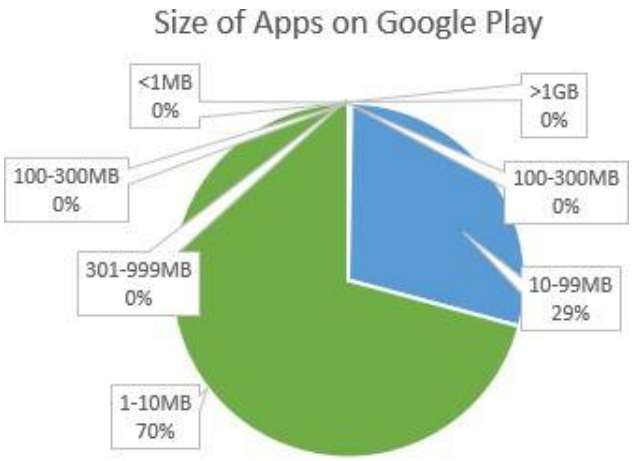

Fig. 9: Size of Apps

application had a total size of 38 MB, which included all the code, assets, and resources required to run the app seamlessly.

In addition to the application size, we also discovered that the app generated user data during usage, which amounted to 3.75 MB. The generated data included user preferences, settings, or any other information that users created while interacting with the application. The user data plays a vital role in enhancing the user experience, and as such, it's crucial to ensure that the app efficiently stores and utilizes this information.

Furthermore, the app utilizes approximately 2 MB of cache memory to improve its performance by storing frequently accessed data for faster retrieval. The use of cache memory helps reduce the app's load time and improves its overall performance, creating a seamless experience for the user.

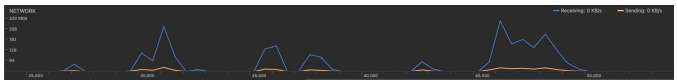

Fig. 10: Network requests performance

It's important to consider the data transfer rate when sending or receiving data over the network. Our research indicated that data sent through network requests can take up to 32 KB/s, while data received can reach up to a maximum of 320 KB/s, as shown in Figure 10. This information is critical, particularly for developers who heavily rely on network communication, to ensure they optimize the application's performance and provide an excellent user experience.

In conclusion, having a clear understanding of the size and performance of our application is crucial for ensuring a positive user experience. By being aware of the app's analytics, developers can optimize its performance and ensure efficient use of device resources.

VI. Conclusions and Future Work

The importance of having access to relevant information and trends in the mobile app market cannot be overstated. With a massive amount of data available in the Google Play Store, developers and app market enthusiasts often struggle to process and analyse it in a timely manner. This is where our app comes into play, providing users with a user-friendly interface to view and understand essential statistics and trends in the app market.

One of the unique features of our app is its ability to compare two apps and view statistics for a specific app. This functionality caters to the needs of app developers, enabling them to make informed decisions based on the performance of their app and the competition. Furthermore, our app offers an overall view of the app market, presenting users with a comprehensive understanding of the industry as a whole. The use of graphics and accompanying explanations enhances the user's ability to comprehend the data and its implications.

Our app and its accompanying research have significantly enhanced the comprehension of the mobile app development process and offered invaluable insights into the app market. Our app serves as a valuable tool for app developers and anyone interested in app market statistics, providing the necessary information and analysis to stay informed and ahead of the competition.

While there are limitations posed by the use of annual reports instead of up-to-date data, our analysis and app development still provide valuable insights into the world of mobile app development. By focusing on the most critical statistics and trends in the app market, our work provides a snapshot of the app industry that can be a valuable resource for app developers and anyone interested in app market data.

Overall, our app and study offer a comprehensive overview of the mobile app market, providing users with a user-friendly interface and valuable insights. Whether you are an app developer or simply interested in app market trends, our app provides the information you need to stay informed and ahead of the curve.

A. Future Work

To further enhance the solution provided by our mobile app, we propose the following improvements:

1) Live data: Implementing a way to access live data from the Google Play Store will provide more up-to-date statistics to users and give a more comprehensive view of the market. Also, it allows the application to have always the most recent data.
2) Market evolution: By using a data set as we did, we can provide valuable information. However, comparing this data to future years can give insights into market trends and how the market is evolving.
3) Analysis of the App Store: The App Store also has a significant market share for the iOS platform. Analysing both the Google Play and App Store data will provide a comprehensive understanding of the app development market and its trends.

2023 18th Iberian Conference on Information Systems and Technologies (CISTI)
20 – 23 June 2023, Aveiro, Portugal
ISBN: 978-989-33-4792-8

4) Add a feature to analyse the same application from different store markets. It will allow drawing some conclusions regarding the type of consumers of each brand.

Furthermore, there is also the possibility of exploring new data sources and analytics methods to gain deeper insights into the app development market. For example, incorporating data from app store reviews and feedback from users can provide additional information on the popularity, user satisfaction, and areas for improvement for specific apps.

In addition to these proposed improvements, the mobile app also has the potential to be further expanded in the future. For instance, incorporating machine learning techniques to analyse the data and provide more accurate and meaningful insights to users can also be explored.

ACKNOWLEDGEMENTS

"This work is funded by National Funds through the FCT – Foundation for Science and Technology, I.P., within the scope of the project Ref. UIDB/05583/2020. Furthermore, we would like to thank the Research centre in Digital Services (CISeD) and the Instituto Politécnico de Viseu for their support." Maryam Abbasi thanks the National funding by FCT - Foundation for Science and Technology, P.I., through the institutional scientific employment program-contract (CEECINST/00077/2021)

REFERENCES

[1] Duc Bui, Yuan Yao, Kang G Shin, Jong-Min Choi, and Junbum Shin. Consistency analysis of data-usage purposes in mobile apps. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 2824–2843, 2021.

[2] Ning Chen, Jialiu Lin, Steven CH Hoi, Xiaokui Xiao, and Boshen Zhang. Ar-miner: mining informative reviews for developers from mobile app marketplace. In *Proceedings of the 36th international conference on software engineering*, pages 767–778, 2014.

[3] Jerry Gao, Xiaoying Bai, Wei-Tek Tsai, and Tadahiro Uehara. Mobile application testing: a tutorial. *Computer*, 47(2):46–55, 2014.

[4] Sang Pil Han, Sungho Park, and Wonseok Oh. Mobile app analytics. *MIS quarterly*, 40(4):983–1008, 2016.

[5] Francisco J Mart ínez-López, Yangchun Li, and Susan M Young. *Social Media Monetization: Platforms, Strategic Models and Critical Success Factors*. Springer Nature, 2022.

[6] Nan Zhong and Florian Michahelles. Google play is not a long tail market: An empirical analysis of app adoption on the google play app market. In *Proceedings of the 28th annual ACM symposium on applied computing*, pages 499–504, 2013.

[7] Q. Bao, M. Lu, and D. Zhang. A systematic review of mobile app development research. volume 130, pages 20–35, 2017.

[8] A. Kankanhalli, B. C. Tan, and K. K. Wei. User adoption and usage of mobile apps: An empirical investigation. volume 66, pages 2419–2434, 2015.

[9] C. L. Hsu and H. P. Lu. Understanding users' mobile app behavior: An empirical study of app usage. volume 48, pages 99–108, 2015.

[10] Gautham Prakash. Google play store app data of 2.3 million+ applications dataset in kaggle. https://www.kaggle.com/datasets/gauthamp10/google-playstore-apps, 2021.

ISBN: 978-989-33-4792-8