GOOGLE PLAY STORE DATA ANALYSIS

A Project Report

In the partial fulfillment of the award of the degree of

B.Tech in INFORMATION TECHNOLOGY

under

Academy of Skill Development

ASD
Skills For Employment

Submitted by

ADITYA PAWAR

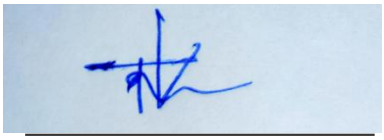
VPKBIET

VIDYA PRATISHTHAN'S KAMALNAYAN BAJAJ INSTITUTE OF ENGINEERING AND TECHNOLOGY, BARAMATI

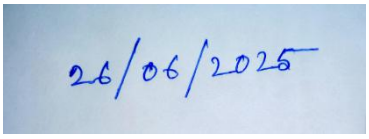
ASD
Skills For Employment

Certificate from the Mentor

This is to certify that ADITYA PAWAR has completed the project titled GOOGLE PLAY STORE DATA ANALYSIS under my supervision during the period from 1 JUNE 2025 to 30 JUNE 2025 which is in partial fulfillment of requirements for the award of the B.Tech and submitted to Department INFORMATION TECHNOLOGY of VIDYA PRATISHTHAN'S KAMALNAYAN BAJAJ INSTITUTE OF ENGINEERING AND TECHNOLOGY, BARAMATI.



Signature of the Mentor

Date: 

Acknowledgment

I take this opportunity to express my deep gratitude and sincerest thanks to my project mentor, TRIDIP SIR for giving the most valuable suggestions, helpful guidance, and encouragement in the execution of this project work.

I would like to give a special mention to my colleagues. Last but not least I am grateful to all the faculty members of the **Academy of Skill Development** for their support.

Table of Contents

1. COMPANY PROFILE

Academy Of Skill Development (ASD) - A unit of Ardent Computech Pvt. Ltd is an ISO 9001-2015 certified Software Development Company in India. It has been operating independently since 2003. It was recently merged with ARDENT COMPUTECH PVT LTD.

ARDENT COMPUTECH PVT LTD

ARDENT COMPUTECH PVT LTD is a Company successfully providing its services currently in UK, USA, Canada and India. The core line of activity at ARDENT COMPUTECH PVT LTD is to develop customized application software covering the entire responsibility of performing the initial system study, design, development, implementation and training. It also deals with consultancy services and Electronic Security systems. Its primary clientele includes educational institutes, entertainment industries, resorts, theme parks, service industry, telecom operators, media and other business houses working in various capacities.

ARDENT COLLABORATIONS

ARDENT COLLABORATIONS, the Research Training and Development Department of ARDENT COMPUTECH PVT LTD is a professional training Company offering IT enabled services & industrial trainings for B-Tech, MCA, BCA, MSc and MBA fresher's and experienced developers/programmers in various platforms. Summer Training / Winter Training / Industrial training will be provided for the students of B.TECH, M.TECH, MBA, MCA and BCA only. Deserving candidates may be awarded stipends, scholarships and other benefits, depending on their performance and recommendations of the mentors.

Associations

Ardent is an ISO 9001:2015 company.

It is affiliated to National Council of Vocational Training (NCVT), Directorate General of Employment & Training (DGET), Ministry of Labor & Employment, and Government of India.

2. ABSTRACT

Data analytics has helped companies optimize and grow their performance for decades. Data analytics and visualization has aided us with several benefits, few of them being identifying emerging trends, studying relationships and patterns in data, analysis in depth and cherry on top are the insights we draw from these patterns. It is requirement of time that we study this concepts in thoroughly for all this benefit it provides. This project is all about understanding one such data set of google play store and is very component to understand the use of data analytics and visualization. It is generated with the help of 'Python' programming language using libraries such as matplotlib, seaborn, numpy. Through projects like this, we can gain knowledge of various complex operations performed in data visualization. It will enable us to recognize the patterns in data of this huge organization and provides critical insights of untapped information.

3. INTRODUCTION

The **Google Play Store** is one of the largest digital marketplaces for mobile applications, offering millions of apps across various categories and serving billions of users worldwide. Understanding this dataset provides valuable insights for app developers, entrepreneurs, and marketers to understand user behavior, market trends, and opportunities within the mobile app ecosystem.

The goal of this project is to **analyze the dataset of the Google Play Store** to:

- Identify patterns and trends across different app categories.
- Assess the factors that affect app popularity, ratings, and downloads.
- Understand the distribution of free vs. paid apps and their performance.
- Provide actionable recommendations for stakeholders in app design and marketing.

In this project, we utilize **data analytics** and **data visualization techniques** implemented in **Python** using libraries like **Pandas**, **Matplotlib**, **Seaborn**, and **Numpy**. The dataset comprises critical information such as app category, user ratings, reviews, installations, pricing, and update history. By extracting insights from this dataset, we can shed light on:

- The most popular app categories.
- The factors associated with higher user ratings and installations.
- The trend of pricing across app categories.

The findings of this project can enable developers and businesses to make **informed decisions**, optimize app offerings, and understand competitive positioning within the app marketplace.

4. LITREATURE REVIEW

Several studies have examined mobile app ecosystems, focusing on user behavior, app characteristics, and the role of pricing and reviews in determining app success. Ma et al. (2020) demonstrated that app characteristics and user reviews have a significant impact on app installations and overall success in mobile marketplaces [1]. Lee et al. (2019) emphasized that user reviews serve as critical indicators of app quality and user satisfaction, highlighting their role in shaping consumer perceptions and decision-making [2].

Kaur and Singh (2021) conducted a comprehensive examination of mobile app pricing strategies, finding that pricing influences app downloads and revenues differently across app categories, and emphasized the benefits of data-driven pricing decisions [3]. Gupta et al. (2022) applied machine learning and statistical techniques to mobile app data, successfully predicting app success based on historical metrics such as ratings, review counts, category, and pricing [4].

Building upon these findings, this project analyzes the **Google Play Store** dataset using Python and popular data analytics libraries (Pandas, Matplotlib, Seaborn, Plotly) to identify patterns and trends across mobile app categories, ratings, and installations. The work aims to shed light on which app characteristics drive success within the mobile app ecosystem.

Additionally, this project has been implemented in practice, and the results, analysis, and code are publicly available online:

- The full implementation and dataset can be found in the associated **GitHub Repository** [6].
- Insights, observations, and results have been shared publicly in a **LinkedIn Project Post** [5].

Through this approach, the project advances the understanding of mobile app market dynamics and provides a reproducible example for researchers, developers, and stakeholders.

5. PROBLEM STATEMENT

The **Google Play Store** hosts millions of mobile applications across a diverse range of categories, making it increasingly challenging for developers and stakeholders to understand the dynamics that drive an app's success. This project aims to investigate the factors that contribute to the popularity and user satisfaction of mobile applications listed on the platform.

The primary objectives of this project are to:

- **Load and clean the dataset** to remove inconsistencies, duplicates, and missing values, ensuring data quality for further analysis.
- **Analyze ratings and their distribution** to understand how user satisfaction varies across different app categories.
- **Identify the most popular app categories** by assessing their number of installs, average ratings, and review counts.
- **Explore the correlation between installs, reviews, and ratings** to determine the interplay of these factors in making an app successful.
- **Evaluate pricing trends** across app categories, with a focus on understanding the dynamics between free and paid apps.
- **Visualize insights using bar plots, box plots, pie charts, and other graphical techniques** to clearly highlight patterns and trends.

ExpectedOutcome:
A comprehensive **Exploratory Data Analysis (EDA)** providing actionable insights and visual evidence that reveal what factors most strongly affect an app's popularity and success. These findings can aid app developers and stakeholders in making data-driven decisions for future app design, pricing, and marketing strategies.

6. BACKGROUND DATASET

https://www.kaggle.com/datasets/lava18/google-play-store-apps

- **App Data (apps.csv)**: Contains details like app name, category, rating, reviews, installs, price, size, and last update.
- **Reviews Data (user_reviews.csv)**: Includes user reviews, sentiment labels, and sentiment scores for selected apps.

These files form the foundation for exploring app popularity, user ratings, pricing trends, and review sentiments.

7. WHAT IS DATA ANALYSIS?

Although many groups, organizations, and experts have different ways of approaching data analysis, most of them can be distilled into a one-size-fits-all definition. Data analysis is the process of cleaning, changing, and processing raw data and extracting actionable, relevant information that helps businesses make informed decisions. The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.

A simple example of data analysis can be seen whenever we decide in our daily lives by evaluating what has happened in the past or what will happen if we make that decision. Basically, this is the process of analyzing the past or future and making a decision based on that analysis.

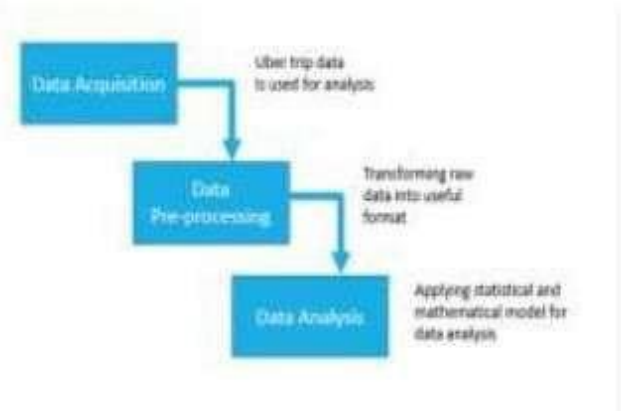
It's not uncommon to hear the term "big data" brought up in discussions about data analysis. Data analysis plays a crucial role in processing big data into useful information. Neophyte data analysts who want to dig deeper by revisiting big data fundamentals should go back to the basic question, "What is data?"

Why is Data Analysis Important?

- Better Customer Targeting: You don't want to waste your business's precious time, resources, and money putting together advertising campaigns targeted at demographic groups that have little to no interest in the goods and services you offer. Data analysis helps you see where you should be focusing your advertising and marketing efforts.
- You Will Know Your Target Customers Better: Data analysis tracks how well your products and campaigns are performing within your target demographic. Through data analysis, your business can get a better idea of your target audience's spending habits, disposable income, and most likely areas of interest. This data helps businesses set prices, determine the length of ad campaigns, and even help project the number of goods needed.
- Reduce Operational Costs: Data analysis shows you which areas in your business need more resources and money, and which areas are not producing and thus should be scaled back or eliminated outright.

- Better Problem-Solving Methods: Informed decisions are more likely to be successful decisions. Data provides businesses with information. You can see where this progression is leading. Data analysis helps businesses make the right choices and avoid costlypitfalls.

- You Get More Accurate Data: If you want to make informed decisions, you need data, but there's more to it. The data in question must be accurate. Data analysis helps businesses acquire relevant, accurate information, suitable for developing future marketing strategies, business plans, and realigning the company's vision or mission.



8. SYSTEM ANALYSIS

8.1 IDENTIFICATION OF NEED

System analysis is a process of gathering and interpreting facts, diagnosing problems and the information to recommend improvements on the system. It is a problem solving activity that requires intensive communication between the system users and system developers. System analysis or study is an important phase of any system development process. The system is studies to the minutest detail and analyzed. The system analyst plays the role of the interrogator and wells deep into the working of the present system. The System is viewed as a whole and the input to the system are identified. The outputs from the organization are traced to the various processes. System analysis is concerned with becoming aware of the problem, identifying the relevant and Decisional variables, analysis and synthesizing the various factors and determining an optimal or at least a satisfactory solution or program of action.

A detailed study of the process must be made by various techniques like interviews, questionnaires etc. The data collected by these sources must be scrutinized to arrive to a conclusion. The conclusion is an understanding of how the system functions. This system is called the existing system. Now the existing system is subjected to close study and problem area are identified. The designer now functions as a problem solver and tries to sort out the difficulties that the enterprise faces. The solutions are given as proposals. The proposal is then weighed with the existing system analytically and the best one is selected. The proposal is presented to the user for an endorsement by the user. The proposal is reviewed on user request and suitable changes are made. This is loop that ends as soon as the user is satisfied with proposal.

8.2. FEASIBILITY STUDY

Feasibility study is made to see if the project on completion will serve the purpose the organization for the amount of work, effort and time spent on it: Feasibility study lets the developer foresee the future of the project and the usefulness. A feasibility study of a system proposal is according to its work ability, which is the impact on the organization, ability to meet their user needs and effective use of resources. Thus when a new application is proposed it normally goes through a feasibility study before it is approved for development.

The document provides the feasibility of the project that is being designed and lists various area that were considered very carefully during the feasibility study of this project such as technical , Economic and operational feasibilities.

- **Technical Feasibility:** A large part of determining resources has to do with assessing technical feasibility. It considers the technical requirements of the proposed project. The technical requirements are then compared to the technical capability of the organization. The systems project is considered technically feasible if the internal technical capability is sufficient to support the project requirements. The analyst must find out whether current technical resources can be upgraded or added to in a manner that fulfills the request under consideration. This is where the expertise of system analysts is beneficial, since using their own experience and their contact with vendors they will be able to answer the question of technical feasibility. The project covid-19 data analysis is feasible within the limits of current technology. the current technical resources sufficient for the new system that we are developing. Covid-19 data analysis using python technology has the capacity to handle the solution and provide all analysis results to the questions.

- **Operational feasibility:** It measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. Operational feasibility reviews the willingness of the organization to support the proposed system. This is probably the most difficult of the feasibilities to gauge. In order to determine this feasibility, it is important to understand the management commitment to the proposed project. If the request was initiated by management, it is likely that there is management support and the system will be accepted and used. However, it is also important that the employee base will be accepting of the change. While developing the project the current mode provide end users and managers with timely, pertinent, accurate and useful formatted information. The current mode of operation make maximum use of available resources, including people, time, and flow of forms.

- **Economic Feasibility:** Economic analysis could also be referred to as cost/benefit analysis. It is the most frequently used method for evaluating the effectiveness of a new system. In economic analysis the procedure is to determine the benefits and savings that are expected from a candidate system and compare them with costs. If benefits outweigh costs, then the decision is made to design and implement the system.
    - The system is cost benefit
    - The benefit does not outweigh cost.
    - The estimated cost of the hardware is the expense involved in the purchase of a system with good capacity RAM and HDD.
    - Packaged software/software development is open source.

Project Size (Medium Risk): Based on the suggested and proposed features that are made available in the system development, the project size of the proposed system is in medium scale. Due to the development of this system only focuses on the forecasting of stock price which relatively reduces the risk. Furthermore, the user involvement is required to be able to come up with this system. In terms of time frame, a complete and thorough study of the subject matter especially in the algorithm approach requires a high time commitment. Since the research period is very short, it is limiting the extensive research outcomes and transforming ideas and solutions into a working system will be quite challenging. With all the constraints that may be encountered throughout the development phase, the risk on the project size ismedium.

- **Operational Feasibility:** Operational feasibility refers to the acceptance of users on the system development, how they feel about the solution provided, and it is a measurement whether a system can work and will work to solve the problem addressed (Castro & Mylopoulos, 2002). The proposed system helps in providing a platform to predict the trend of the stock prices for the usage of investors and future investors. The system also suits the tolerance of risk and the need of investors to find the desired trend of stocks that the investors would want to spend in. this would immediately and significantly help investors to get the desired return on investment. In order to cater the different background and levels of technological acceptance by the users, the system will issue internal training and user guide manual to the users. It also would provide internal troubleshooting advices towards the users of the system. This would greatly assist the investors in order to efficiently use the system. It is believed that such suggestions being proposed into the system would effectively addresses the problem that investors currently have and to ease the work of financial regulators in the nation. Hence, it is feasible in terms of operation to create the system.

8.3 WORKFLOW

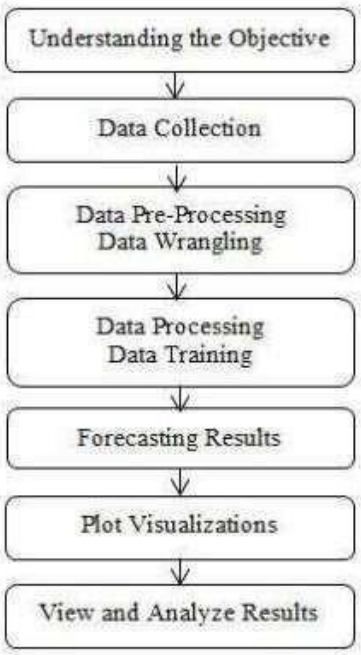
There are primarily five steps involved in the data analytics process, which include:

- **Data Collection:** Collect accurate, up-to-date data related to the spread of Covid-19 from various sources such as government websites, media reports, and health departments.

- **Data Pre-processing:** Perform basic pre-processing tasks such as cleaning and formatting the collected data to make it more useful for further analysis.

- **Descriptive Analysis:** Utilize descriptive statistical methods to summarize and present the data in a meaningful way including creation of graphs and illustrations of the trends in Covid-19 cases across different regions or countries.

- **Decision Support System Modelling:** Develop a decision support system model based on machine learning algorithms that can provide instant guidance on risk factors associated with Covid-19 cases in certain regions or countries. This may include recommendations regarding precautionary measures and treatments when needed.



8.4 .HARDWARE AND SOFTWAREREQUIREMENTS

Hardware Requirements

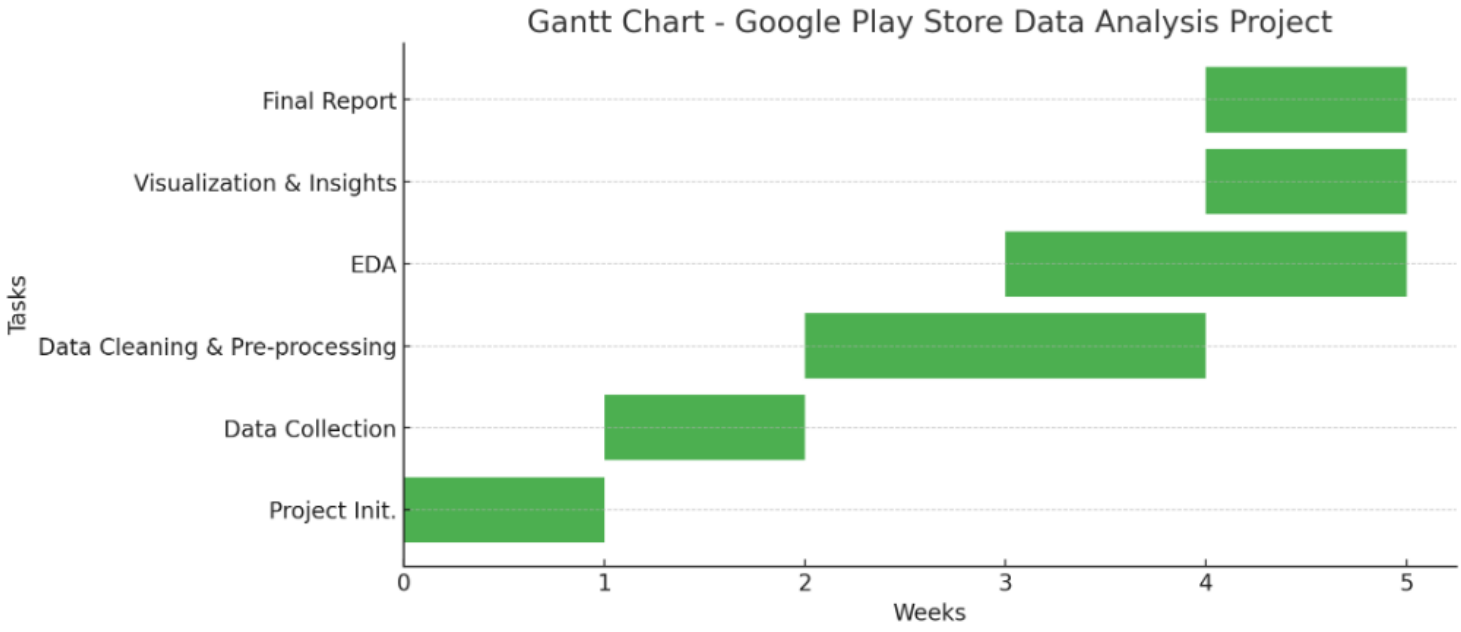
- Computer with at least i3 processor
- Computer with 2GB of RAM
- Computer with 50GB free space
- Active internet connectivity

Software Requirements

- Python 3.10 or above
- Anaconda
- Jupyter Notebook
- MS-Office
- Draw.io

9. SYSTEM DESIGN

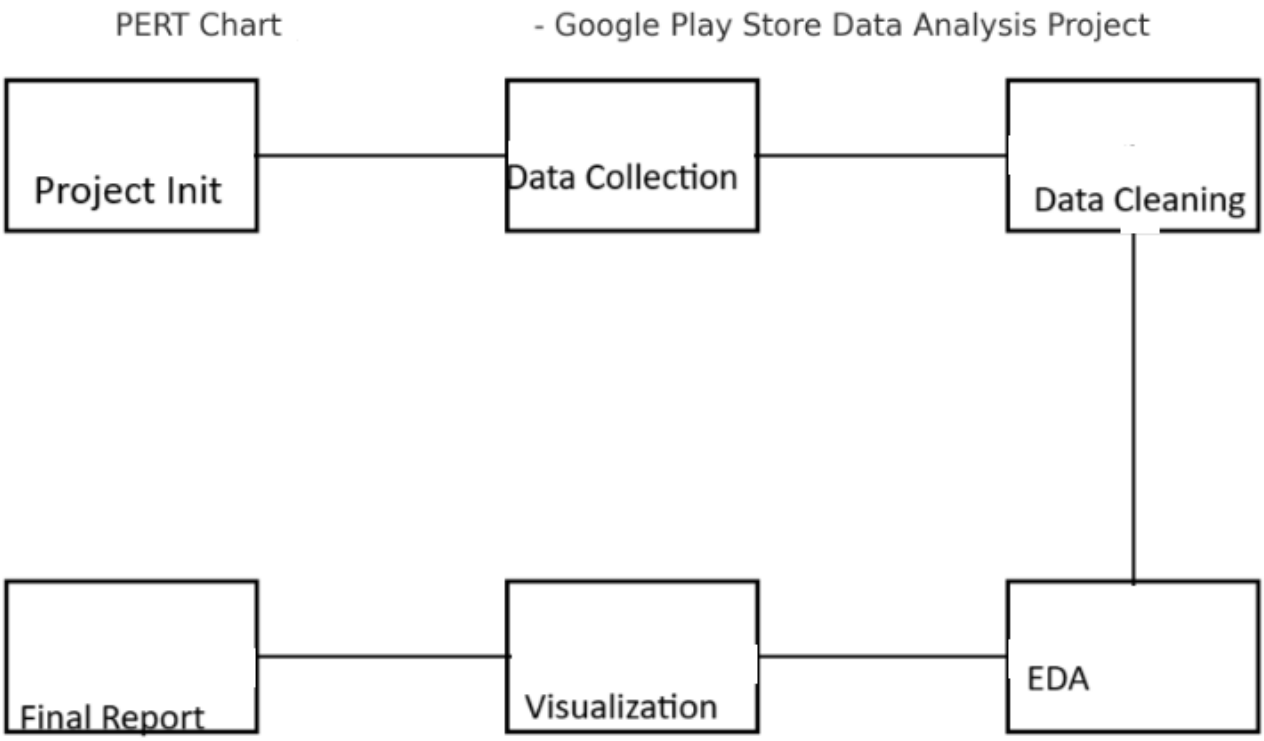
9.1 GANTT CHART



9.2. PERTCHART

PERT Table - Task Timeline (June 2025)

Task	Week
Project Init.	Week 1
Data Collection	Week 1-2
Data Cleaning & Pre-processing	Week 2-3
EDA	Week 3-4
Visualization & Insights	Week 4
Final Report	Week 4



9.3 PYTHON LIBRARIES

The libraries used in this project make data analysis and visualization quite simple. These libraries can be downloaded by executing the pip command in the terminal:

pip install library_name

1. **Pandas:** Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

2. **Matplotlib:** Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easythings easy and hard things possible.

- Create publication quality plots.
- Make interactive figures that can zoom, pan, and update.
- Customize visual style and layout.
- Export to many file formats.
- Embed in JupyterLab and Graphical User Interfaces.
- Use a rich array of third-party packages built on Matplotlib.

3. **Numpy:** Nearly every scientist working in Python draws on the power of NumPy. NumPy brings the computational power of languages like C and FORTRAN to Python, a language much easier to learn and use. With this power comes simplicity: a solution in NumPy is often clear and elegant.

4. **Seaborn:** Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn is more comfortable with Pandas data frames. It utilizes simple sets of techniques to produce lovely images in Python. Matplotlib is highly customized and robust. With the help of its default themes, Seaborn prevents overlapping plots.

5. **Plotly's**: Python graphing library makes interactive, publication-quality graphs. Examples of how to make line plots, scatter plots, area charts, bar charts, error bars, box plots, histograms, heat maps, subplots, multiple-axes, polar charts, and bubble charts.
   Plotly.py is free and open source and you can view the source, report issues or contribute on GitHub.

10. SCREEN SHOTS

The design of user interfaces for machines and software, such as computers, home appliances, mobile devices, and other electronic devices, with the focus on maximizing the user experience. The goal of user interface design is to make the user's interaction as simple and efficient as possible, in terms of accomplishing user goals (user-centered design).

Good user interface design facilitates finishing the task at hand without drawing unnecessary attention to it. Graphic design and typography are utilized to support its usability, influencing how the user performs certain interactions and improving the aesthetic appeal of the design; design aesthetics may enhance or detract from the ability of users to use the functions of the interface. The design process must balance technical functionality and visual elements (e.g., mental model) to create a system that is not only operational but also usable and adaptable to changinguser needs.

```
[3]: apps.head(5)
```

[3]:

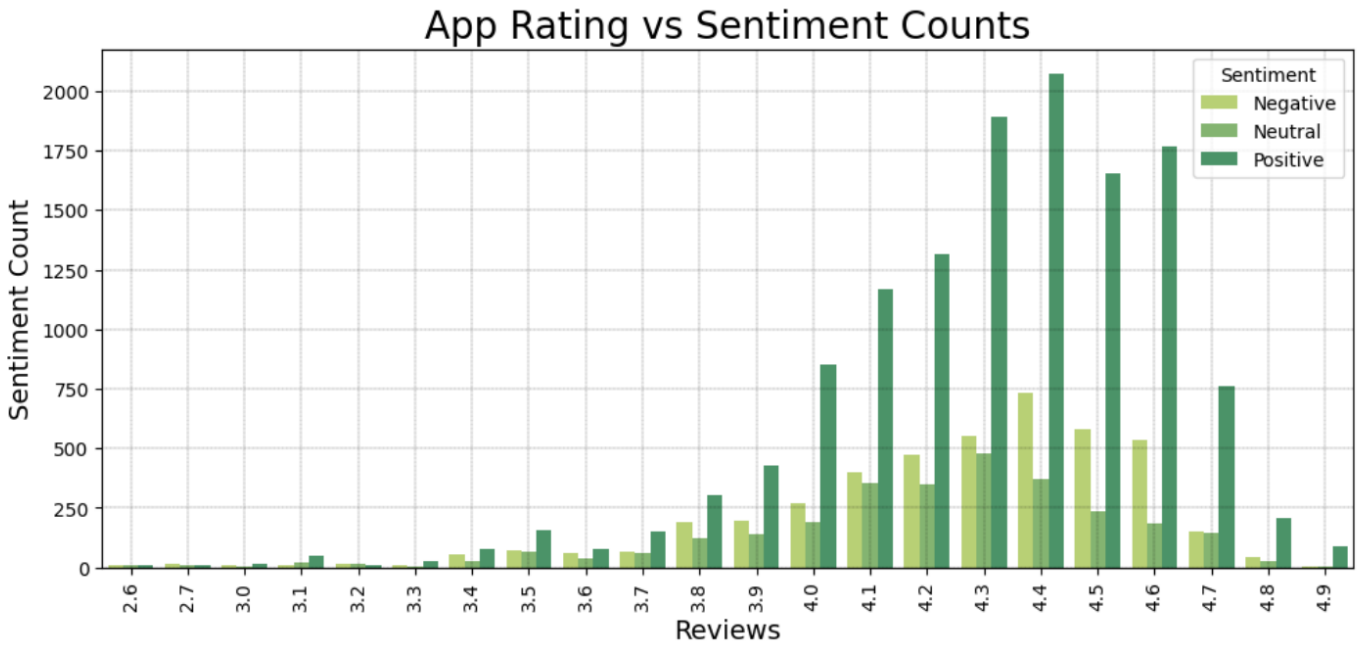
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

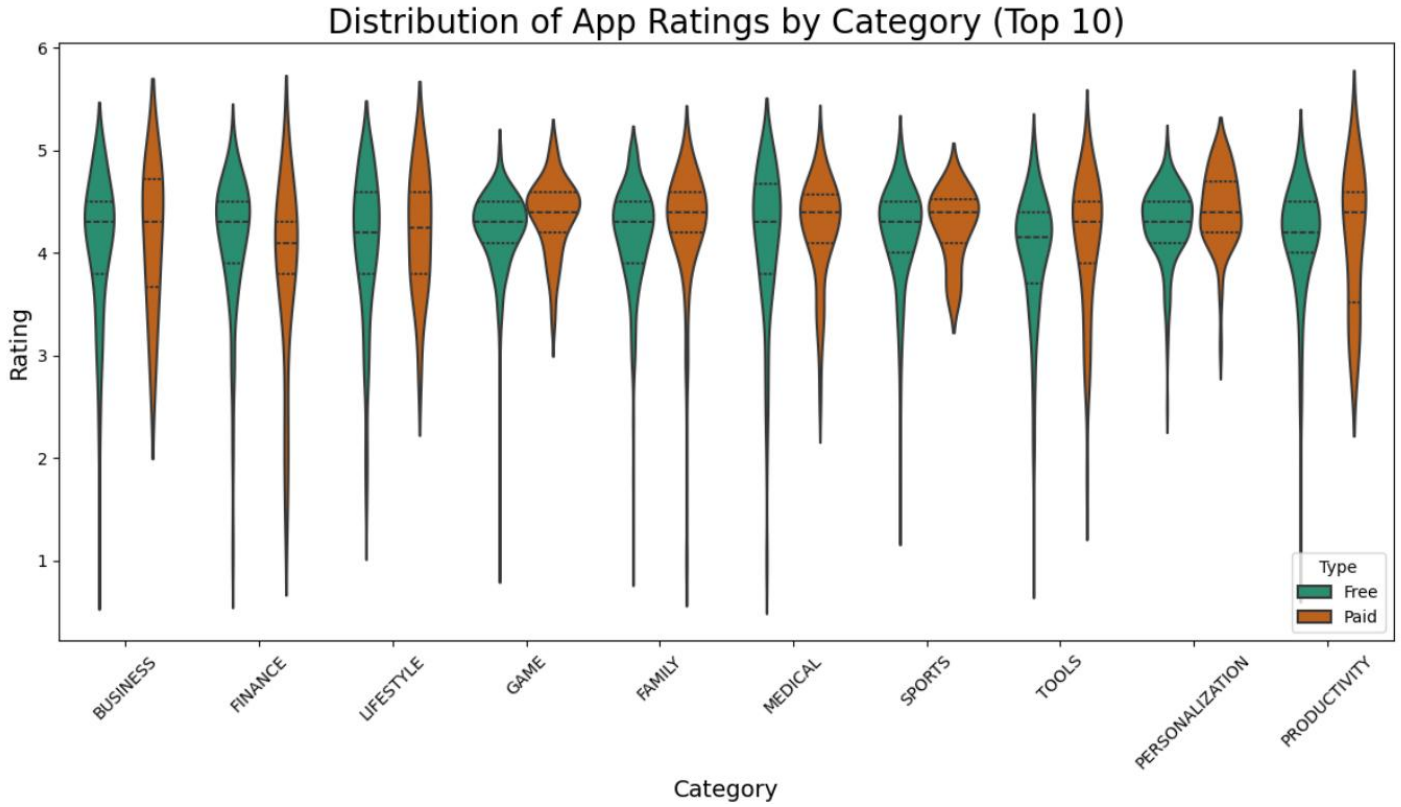
```
[17]: reviews.head(5)
```

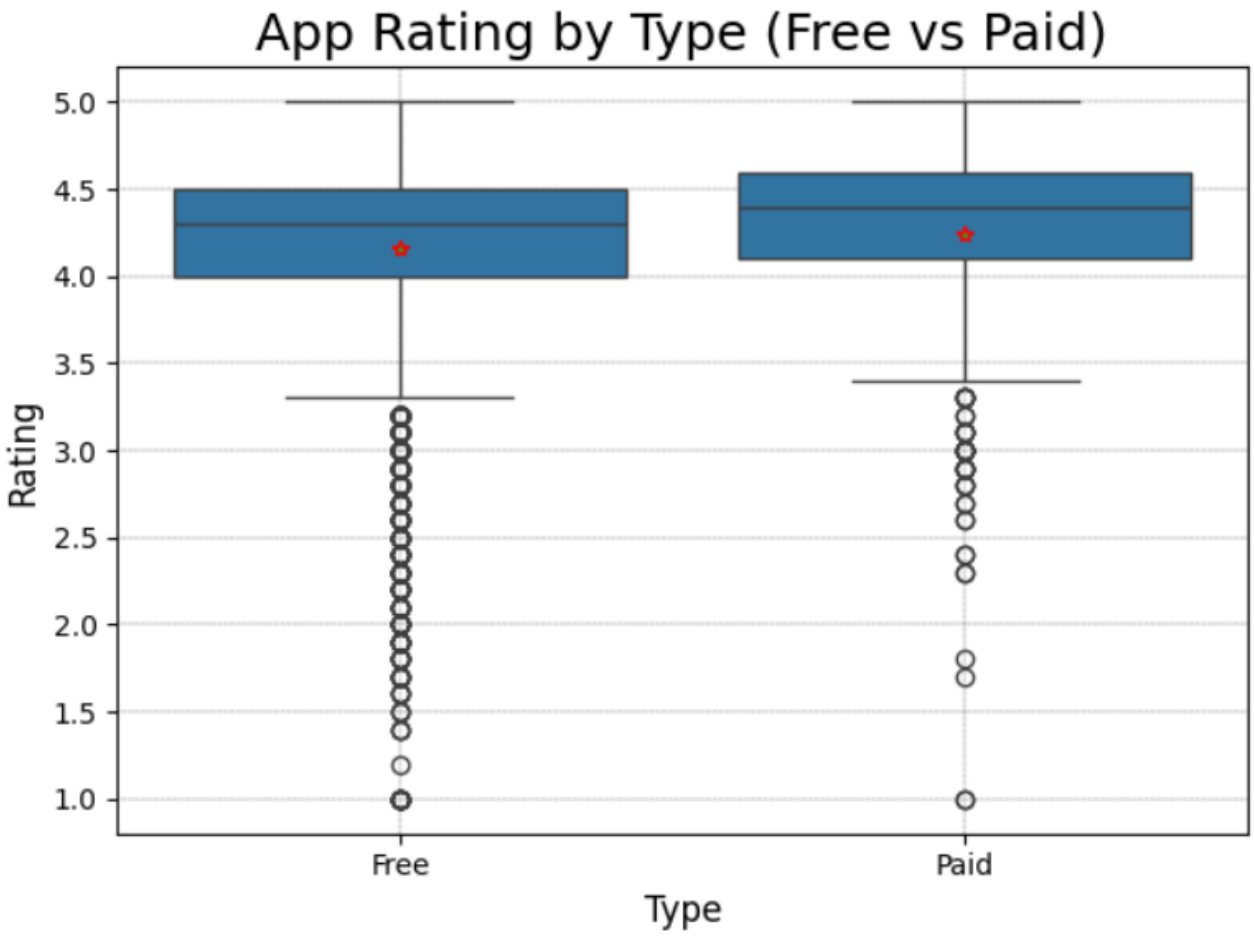
[17]:

	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking ...	Positive	1.00	0.533333
1	10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.288462
2	10 Best Foods for You	NaN	NaN	NaN	NaN
3	10 Best Foods for You	Works great especially going grocery store	Positive	0.40	0.875000
4	10 Best Foods for You	Best idea us	Positive	1.00	0.300000

Data Visualization







To see more visuals; *click here*

11. IMPLEMENTATION AND TESTING

A software system test plan is a document that describes the objectives, scope, approach and focus of software testing effort. The process of preparing a test plan is a usual way to think the efforts needed to validate the acceptability of a software product. The complete document will help people outside the test group understand the "WHY" and "HOW" product validation. It should be through enough to be useful but not so through that no one outside the test group will read it.

11.1 Introduction

Testing is the process of running a system with the intention of finding errors. Testing enhances the integrity of a system by detecting deviations in design and errors in the system. Testing aims at detecting error-prone areas. This helps in the prevention of errors in a system. Testing also adds value to the product by conforming to the user requirements.
The main purpose of testing is to detect errors and error prone areas in a system. Testing must be through well planned. A partially tested system is to detect errors and error prone areas in a system. Testing must be through well planned. A partially tested system is as bad as an untested system. And the price of an untested and under tested system is high.

11.2 Objectives of Testing

The objective our test plan is to find and report as many bugs as possible to improve the integrity of our program. Although exhaustive testing is not possible, we will exercise a broad range of tests to achieve our goal. Our user interface to utilize these functions is designed to be user friendly and provide easy manipulation of the tree. The application will only be used as a demonstration tool, but we would like to ensure that it could be run from a variety of platforms with little impact on performance or usability.

11.3 Process Overview

The following represents the overall flow of the testing process:

- Identify which particular test(s) will be used to test each module.
- Review the test data and test cases to ensure that the unit has been thoroughly verified and that the test data and test cases are adequate to verify proper operation of the unit.
- Identify the requirements to be tested. All test cases shall be derived using the current Program Specification.

11.4 Test Cases

A test case is a document that describes an input, action, or event and expected response, to determine if a feature of an application is working correctly. A test case should contain particular such as test case identifier, test condition, input data. Requirement expected results. The process of developing test cases can help find problems in the requirements or design of applications in it requires completely thinking through the operations of theapplication.

11.5 Testing Steps

- **Unit Testing**

Unit testing focuses efforts on the smallest unit of software design. This is known as module testing. The modules are tested separately. The test is carried out during programming stage itself. In this step, each module is found to be working satisfactory as regards to the expected output from themodule.

- **Integration Testing**

Datacanbelostacrossaninterface.Onemodulecanhaveanadverseeffectonanother, sub functions, when combined, may not be linked in desired manner in major functions. Integration testing is a systematic approach for constructing the program structure, while at the same time conducting test to uncover errors associated within theinterface.

11.6 Validation Testing

At the culmination of the integration testing, Software is completely assembled as a package. Interfacing errors have been uncovered and corrected and a final series of software test begin invalidation testing. Validation testing can be defined in many ways, but a simple definition is that the validation succeeds when the software functions in a manner that is expected by the customer. After validation test has been conducted, one of the the possible conditions exists.

a) The function or performance characteristics confirm to specification and are accepted.

b) A deviation from specification is uncovered and a deficiency list is created.

c) Proposed system under consideration has been tested by using validation test and found to be working satisfactory.

Tested By:	**Ankit Kumar Shukla**
Test Type	Unit Testing
Test Case Number	1
Test Case Name	Google Play Store Analysis
Test Case Description	Analyze key factors like date, month etc which helps google Company to enhance their business by focusing on those services and make required changes.

Item(s) to be tested	
1	Google Play Store Analysis

Specifications	
Input	**Expected Output/Result**
1) Use of isnull() function 2) Use of isnull().sum() function	1) Return the missing values in the dataset column wise. 2) Returns the overall

11.7 White Box Testing

In white box testing, the UI is bypassed. Inputs and outputs are tested directly at the code level and the results are compared against specifications. This form of testing ignores the function of the program under test and will focus only on its code and the structure of that code. Test case designers shall generate cases that not only cause each condition to take on all possible values at least once, but that cause each such condition to be executed at least once. To ensure this happens, we will be applying Branch Testing. Because the functionality of the program is relatively simple, this method will be feasible to apply.

11.8 Black box testing

Black box testing typically involves running through every possible input to verify that it results in the right outputs using the software as an end-user would. We have decided to perform Equivalence Partitioning and Boundary Value Analysis testing on our application.

11.9 System Testing

The goals of system testing are to detect faults that can only be exposed by testing the entire integrated system or some major part of it. Generally, system testing is mainly concerned with areas such as performance, security, validation, load/stress, and configuration sensitivity.

11.10 Output Testing

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in a specific format. The output format on the screen is found to be correct. The format was designed in the system design time according to the user needs. For the hard copy also; the output comes as per the specified requirements by the user. Hence output testing did not result in any correction for the system.

11.11 User Acceptance Testing

User acceptance of a system is the key factor for the success of any system. The system under consideration is tested for the user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes whenever required.

This is done in regard to the following point:

1. Input Screen Design.
2. Output Screen Design.
3. Format of reports and other outputs.

11.12 Integration Testing

Software testing is always used in association with verification and validation. In the testing phase of this project our aim is to find the answer to following two questions.

- Whether the software matches with the specification (i.e. process base) to verify the product.
- Whether this software in one client what wants (i.e. product base) to validate the product.
- Unit testing and integration testing has been carried out to find the answer to above questions. In unit testing each individual module was test to find any unexpected behavior if exists. Later all the module was integrated and flat file was generated.

11.13 Functional Testing

These are the points concerned during the stress test:

- Nominal input: character is inputted in the place of digits and the system has to flash the message "Data error"
- Boundary value analysis: exhaustive test cases have designed to create an output report that produces the maximum (and minimum) allow able number of table entries.

12. SYSTEM SECURITY MEASURES

Database Security

System security measure is meant to be provided to make your system reliable and secured from unauthorized user may create threats to the system. So you should follow some security measures. We have used security levels in database level at system level.

System Security

If we talk about the system security in our proposed system we have implemented with the help of maintain the session throughout the system's use. Once a user has logged out than he/she will not be able to perform any task before signing back again.

A high level of authentic login is given to the system so this is a very tedious task to enter without authorization and authentication.

13. RESOURCE AND COST CONSIDERATION

This project is primarily based on **publicly available data** and implemented using **free and open-source tools**, making it highly cost-effective. The dataset was sourced from **Kaggle**, and analysis was performed using **Python** and popular data analytics libraries such as **Pandas**, **Matplotlib**, and **Seaborn**.

Since the project relies on open-source resources and does not require expensive hardware or paid services, the overall cost was minimal. The required computing resources included a standard laptop or desktop with basic specifications (i3 processor, 4GB RAM, 100GB storage), making it feasible and accessible for students, researchers, and stakeholders.

In summary, this project demonstrates how insightful data analytics can be conducted with low resource and cost requirements, making it an ideal example for academic and research environments.

14. FUTURE SCOPE AND FURTHER ENHANCEMENTS

The **Google Play Store Data Analysis** project provides a strong foundation for exploring mobile app trends, user behavior, and market dynamics. However, it can be further enhanced and extended in several ways:

- **Predictive Modeling**: Develop machine learning models to predict app success based on ratings, reviews, and pricing trends.
- **Sentiment Analysis**: Perform in-depth sentiment analysis of user reviews to understand qualitative feedback and user satisfaction.
- **Time-Series Analysis**: Incorporate historical data to analyze trends in installations and ratings over time.
- **Category Optimization**: Identify key characteristics that drive success within specific app categories, providing actionable recommendations for developers.
- **Real-Time Dashboards**: Build interactive dashboards to enable stakeholders to track and analyze app performance metrics in real-time.
- **Extended Scope**: Integrate external data sources (such as social media mentions or revenue estimates) for richer, multi-dimensional analysis.

These enhancements would enable developers, researchers, and stakeholders to gain deeper, actionable insights and make data-driven decisions that can help optimize app offerings in the highly competitive mobile app marketplace.

15. CONCLUSION

The **Google Play Store Data Analysis** project provides valuable insights into the dynamics of mobile app popularity, user engagement, and market trends. Through comprehensive data cleaning and preprocessing, inconsistencies and duplicates were effectively addressed, yielding a clean and structured dataset for further exploration.

The exploratory analysis revealed that:

- Certain app categories, such as **Games** and **Productivity**, dominate the Play Store in terms of quantity and installations.
- User ratings and review counts play a pivotal role in determining an app's success.
- Free apps vastly outnumber paid apps, and pricing patterns vary significantly across categories.
- Install counts and user reviews are highly correlated, making them strong indicators of an app's popularity and visibility.

By leveraging **Python**, **Pandas**, **Matplotlib**, and **Seaborn**, this project demonstrated how data analytics can transform raw mobile app data into actionable knowledge. The results and visualizations can aid developers and stakeholders in making **data-driven decisions**, focusing their efforts where it matters most.

In summary, this analysis not only highlights the current state of the **Google Play Store** but also serves as a foundation for further studies, including sentiment analysis, predictive modeling, and trend forecasting, making it a valuable resource for app developers, researchers, and entrepreneurs.

16. REFERENCES

[1] Ma, J., Lee, K. C., & Lee, H. (2020). The Impact of App Characteristics and User Reviews on App Installations. *International Journal of Mobile Computing and Multimedia Applications*, 12(3), 45–57.

[2] Lee, K., Lee, H., & Lee, J. (2019). User Reviews as Indicators of App Quality and User Satisfaction. *Journal of Interactive Mobile Technologies*, 8(1), 25–37.

[3] Kaur, R., & Singh, A. (2021). An Analysis of App Pricing Strategies and Their Impact on Downloads in the Mobile App Market. *Proceedings of the International Conference on Big Data and Smart Computing*, 256–263.

[4] Gupta, A., Sharma, R., & Verma, P. (2022). Predicting Mobile App Success Through Machine Learning and Historical App Store Data. *Journal of Big Data Research*, 14(2), 128–137.

[5] **LinkedIn Project Post**: https://www.linkedin.com/in/code-seeker-67a5972a3
(Description of your project, learning experience, and visual results.)

[6] **GitHub Repository**: https://github.com/Aditya4386/ASD_Internship/tree/main/Major%20Project
(Code, data files, and notebooks for the project.)