**Capstone Project Submission**

***Exploratory Data Analysis On Google Play Store Apps Reviews.***

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| * **Name: Ashish Mali.** * **Email Id:** [**abmali81292@gmail.com**](mailto:abmali81292@gmail.com)   + Collecting, Preparing, Processing and Managing the data for EDA.   + Asking an important question to get an inside.   + Preparing and exploring the data for analysis.   + Carry out the data cleaning operation.   + Carry out data analysis operation.   + Get the insides form the data.   + Share the insides through art of visualization and static reports.   + Created the detail technical documentation for EDA.   + Created the project summery reports for the stakeholders.   + Created and managed the GitHub reops for the project.   + Present the finding with stakeholders via PowerPoint presentation. |
| **Please paste the GitHub Repo link.** |
| Github Link:- [**https://github.com/ashish-mali/EDA-On-Playstore-App-Reviews.git**](https://github.com/ashish-mali/EDA-On-Playstore-App-Reviews.git) |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| When it comes to application distribution, Google Play is the market leader. As the world of the internet expands and internet access becomes more widespread, the demand for the application has grown rapidly in recent years. Because Google Play is based on the Android operating system, which controls more than 70% of the smartphone market, it is critical to examine app distribution from the standpoint of the developer, who is a key stakeholder in this project.  This analysis project will be divided into six phases. Whereas in each phase, we deal with a specific aspect of the analysis one at a time. As we progress through these phases, we get closer to the insides that we need for a successful application in the Play Store. Let’s look at these phases bitwise.   * Ask * Prepare * Process * Analyze * Share * Act   During the **ask** phase of the analysis, we determine the questions that will provide us with valuable inside information. Which is then used to contextualize and make data-driven decisions. Spend some time with the dataset and figure out the multiple questions before sorting the most important one for the analysis.  For the **prepare** process we have datasets in hand, which are two files namely play store data.csv and user review.csv.  The most difficult part of the process is **processing** the data because each attribute has different null and duplicated values. The "Rating" attribute has the most null values, 1474, accounting for approximately 13.6% of the total values. In terms of duplicated values, the "App" attribute contains 483 duplicated values, accounting for approximately 4.5% of the total records. Aside from that, we corrected the attribute's data type for further analysis. In terms of the merged data-frame, the 1074 common records account for roughly 10% of the total records available in the play store dataset. The "Translated Reviews" attribute has the most null values, accounting for approximately 42% of the total. Handling all these anomalies and correcting the data for the further analysis.  In the **analysis** phase we mold the data in order to gain the insides. Following are the insides are drowned from the analysis while answering the question asked in the ask phase of the analysis.   1. The app's name should accurately describe its value propositions. because majority of successful application have this quality. 2. Launching the apps in the category which having more and easy user reach such as **"FAMILY"**, **"GAME"**etc. 3. As per our analysis most of the apps are free it's around **92%**, so, if possible, try to **launch the app with "Free" type**. as it's **increasing the user engagement.** 4. We also seen that the number of installs is correlated with the rating of the application as number of installs increases so as the application rating. 5. One important inside we get with respect to the size of the app is that as the size of the application increases the installation of the app decrease. So, if we release a new apps in market **make sure it's under 20MB.** 6. Content rating also affect the user engagement as more restricted your content rating is the more restricted your user engagement. So, try to keep user rating as **"Everyone"**. 7. Make sure the app will get the update at regular interval, as it's an important factor for user engagement and performance of the application. In our analysis we seen that most of the apps will get their app update at **July month**. 8. There are in total 865 apps which are uniquely reviewed. 9. As we seen form subjectivity most of the reviews as the objective, so for successful apps it's more important to keep eye on the user reviews and early resolution of the problems. 10. We also seen our analysis that there is strong relationship between Install and reviews. for the new apps reviews are the important tool for increase the user engagement. 11. For the apps it's more important have the android version compatibility **above version 4**. As we won't explore this attribute much but by using simple sorting, we can confirm this. 12. Sentiment Polarity is not highly correlated with Sentiment Subjectivity.   The next step in the analysis is to **share** the inside with the stakeholders, so we prepare static reports such as technical documentation and a project summary. We also have details on the aspects in the Google colab notebook.  The last step of our analysis is the **act** phase. Where the stockholder user our drown inside and applied it as per their requirement. |