**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Sahib Hingorani.  [Hingorani.saahib@gmail.com](mailto:Hingorani.saahib@gmail.com)  This project on Play store apps review has been done individually by me. |
| **Please paste the GitHub Repo link.** |
| GitHub Link:- https://github.com/SahibHi/Play\_store\_app\_reviews |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Play-store-app-review** This project was done individually. In this we were provided with two excel files as input data, 1) Play Store Data 2) User reviews  So, I started exploring dataset and worked on the same. I started with importing multiple libraries and then mount my google drive and copy the path to the colab. After that I analysed dataset and found the numeric data i.e., Reviews, size, installs etc. were in type object, to make the data presentable data was cleaned by converting them into Float. During cleaning of data, I found multiple odd entries in different columns, I was doing cleaning and replacing with appropriate data in that too but later I found that this was the case with every column for one App named- “Life Made Wi-Fi Touchscreen Photo Frame” on row- 10475. So, I decided to shift all the entries by one and then check. Once it was done, rest of my data cleaning was completed smoothly and successfully.  **CONCLUSION**  The first analysis which I did was to check how many applications are there in a single type of category. which concluded as the maximum number of apps are from family category which was followed by games and then tools.  Further, I tried to analyse the type of audience app developers are trying to target! Through the analysis I can easily conclude that they target the apps which comes under EVERYONE tag. On the same analysis I tried to check the apps distribution between content rating on a sunburst chart, and interestingly apps percentage were different for all.  **EVERYONE** - Family apps were leading followed by Tools and Games.  **TEEN** - Games and then Family.  **MATURE 17+** - Dating apps are leading.  **EVERYONE 10+** - Family and Games are leading to others.  Adding to this I tried to analyse on how many apps in market are free of cost to users and how many are paid. On the same we can check that majority of the apps are free to users which I have plotted in a countplot and also in a scatterplot just to show it in a different way.  Further I tried to analyse the Size of the applications in different kinds of categories, and from that we can check the data usage for apps per category.  Then I tried to check the correlation between the subsets. From the above data it is clear that there is strong positive correlation between reviews and install.  After that I tried to analyse the ratings and can check that maximum ratings were in between 4 and 4.5 out of 5.  And in my last conclusion I tried to analyse second dataset in which I tried to analyse the positivity or negativity of the user while writing the reviews, which turned out to be that users are pretty much write positive reviews for the apps followed by negative reviews and then neutrals one. |