PlayStoreAppReviewAnalysis

Presentedby-Hritik Saini , Arpit Dargan

ABSTRACT

Google play store is simply entertainment at ourfingertips. It's an official app store and a digitalmedia store having enormous things to offer likebooks, movies, programs and music. Applicationsare either free or paid. Our team has worked onplaystoredata.Thisdatasetcontains13featuresand 10840 observations. This information can be usedforpredictingkeyfactorsresponsibleforappenga gement&successstory.

INTRODUCTION

Play store is not just an app store, it's a platformoffering various digital content to its consumers.TheGooglePlayStoreishometoandroida pplications,music,movies,books,gamesandtelevisio n programs. 81% of the apps are free ofcost which has led to immense popularity of thisplatform. As per google survey report 3000+ appsare being added every other day. The Google PlayStore contains applications for the Android systemonly. This document reveals the dynamics of thePlay Store app and gives actionable insights for thedeveloperstoworkonandruletheAndroidmarket.

INTEGRALMETHODOLOGY:

First,weinvestigatedsomebasicinformationofourdat aset. On doing so we found out that our dataneeded some cleaning, some values were missing,andsomedatatypeswereincorrect.Westarted withdatacleaningandcorrectingthedatatypes,follow ed by data visualisation. We removed someunnecessaryfeaturesandmadeitreadyforanalys isusingdifferentplots.

DATASETDESCRIPTION:

Thisdatasethas13featuresand10840observations

App	Nameof the App
Category	Categoryunder whichit falls
Rating	Application'sratingonplaystore
Reviews	Number ofreviewsofthe app
Size	Sizeof theapp
Installs	NumberofInstallationoftheapp
Type	Whetherthe appisfree or paid
Price	Priceoftheappifit'sapaidapp(0if it's afreeapp)
Content Rating	Appropriatetargetaudienceofthea pp
Genres	Genresunderwhichthe appfalls
LastUpd ated	DatewhentheAppwaslastup dated
Current Version	Current version ofthe App
Android Version	Minimum android versionrequiredtos upportthe App.

BREAKDOWNOFDATASETS

Beforeproceedingtodatavisualisation,weneedtoper formthe followingsteps:

1. Importingrequiredpackagesforfutureanalys is.
2. Mountingdriveandreadingdatafilesfrom Go ogledrive.
3. Removingfuturewarningsinseabornplots.
4. Viewingall datainformation.
5. Droppingduplicate.
6. Removingspecialcharacters
7. Checkinguniquevalues,nullcountanddataty pesof each column.

8. Segregationofnumericalandcategoricaldata .

EXAMININGNULL/ MISSINGVALUES

Some values in our dataset are null or missing.These values affect the accuracy and performanceof the models that predict the outcome, so theseneedtobehandled.Whileanalysingourdatasetthefirst thing we will do is to examine the null ormissingvaluesinourdataset.Thismakesourresultaccurate.Missing values are more in Size & Ratingcolumns as can be seen by plotting graphs. Henceseveralmethods areusedtoremovethese values.

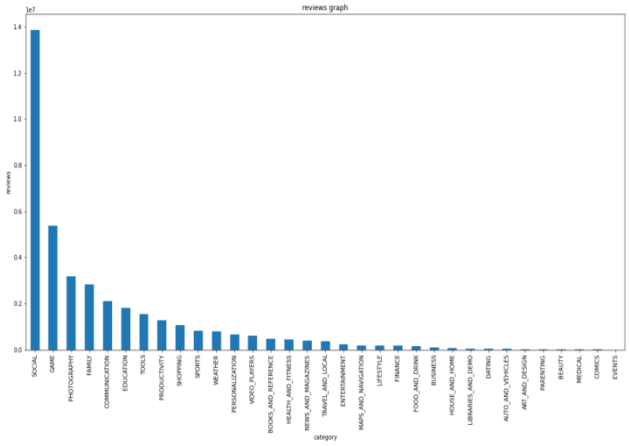
DATACLEANING

Datacleaningistheforemoststepinanydatascienceproject.Cleanerthedata,betteraretheresults.Astheproverbgoesbysaying"MoreDatabeatscleveralgorithm,butbetterdatabeatsmoreData"–PeterNorvig.Tobeginwithourdatacleaning,firstweremovetheduplicatevalues.Thenwe remove unnecessary characters in our dataset.Afterdoingsowefindtheuniquevaluesofeachcolumnandmakethenecessarychangesineachcolumnlikeconvertingdatatypes,removingthenulland'nan' values.Lastly,wehavedoneexploratorydata analysis ofour dataset.

DATAVISUALIZATION

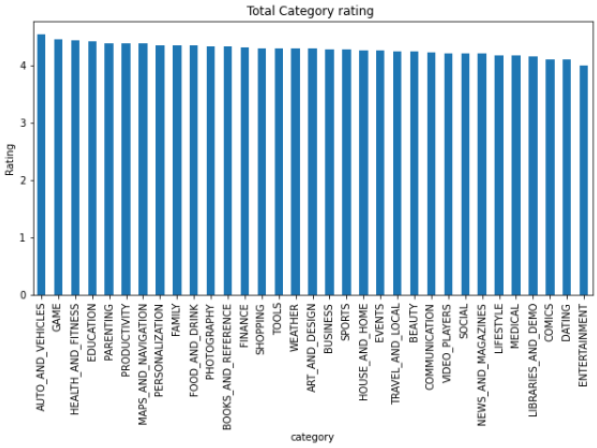
Observation1:

From this barplot graph analysis it tells that category social has the highest and events has the lowest value.



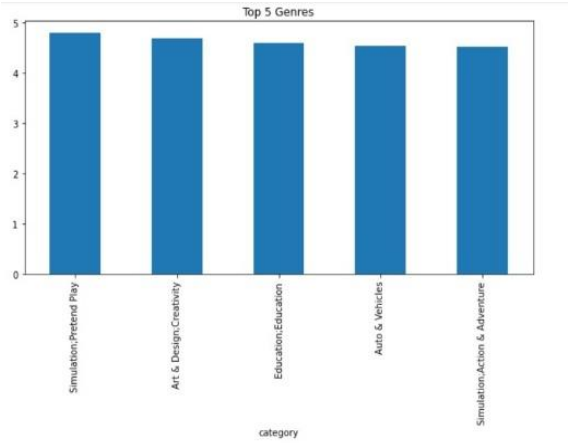
Observation2:

Weplotted agraph oftop categories on playstore



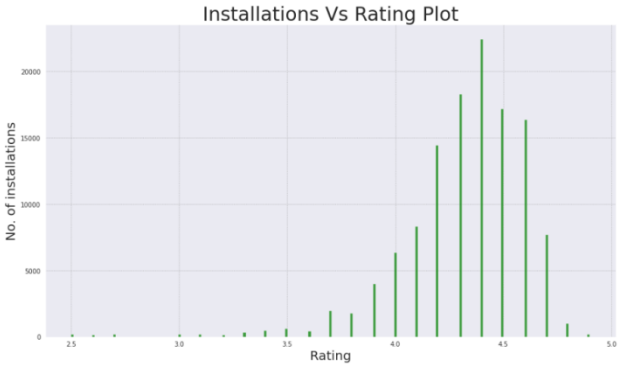
Observation3:

Weplotted agraph oftop five genres.



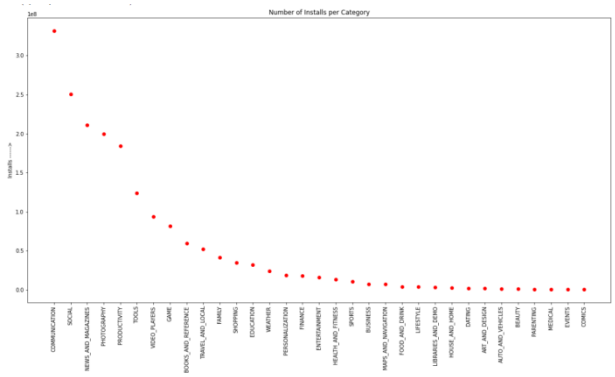
Observation4:

We plotted a graph on installation vs ratings.



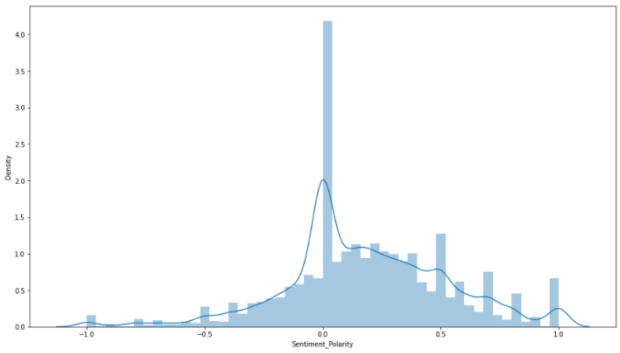
observation5:

Weplottedagraph on installation per category



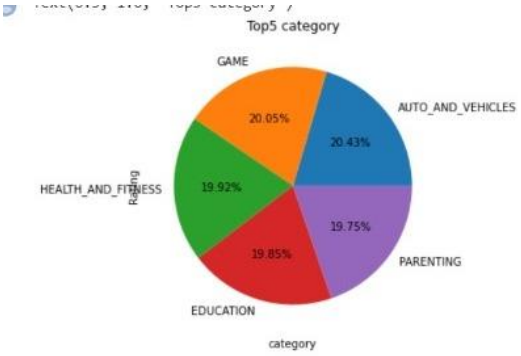
Observation6:

We plotted a graph on sentiment polarity



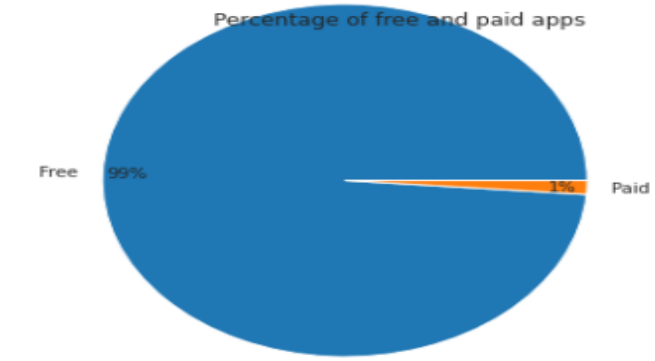
Observation7:

We plotted a pie chart of top five category.



Observation 7:

Percentage of paid and free apps



Advantages of visualisation

Visualiseddataisprocessedfasterandeasier.

- Better insights of the data are drawnwhichmaybe missedintraditionalreports
- Helps us visualise trends whichimproveperformance

CONCLUSIONANDFUTUREWORK

The app developers can predict the outcome ofthe developed apps. Better insights are drawnfrom this visualisation. Apps which need to beimprovedcanbeworkeduponbythedevelopers. Thedatasetcontainsimmensepossibilitiestoimprovebusinessvaluesandhaveapositiveimpact. Wecouldaddasystemthatwouldcreateapplication on its own by using the data set andcreating the best user interface by highly ratedapps.

ACKNOWLEDGEMENT

This project is presented by-Sarthak Arora, JayNandasana,ArshiWani,PranjaliTete,MadhaviMali.
Our sincere efforts have made us to accomplishthetaskofcompletingthisproject.Weareextremelygratefultoourinstructorsandmentors who have helped us to grow in thisfield.Wewouldliketoexpressoursinceregratitudetothecelebratedauthorswhosephenomenalworkhasbeenconsultedandreferred in our project work. We also wish toconveyourappreciationtoourpeerswhoprovided encouragement and timely support inthehourofneed.Thisprojecthelpedusimproveourskillsandenhancedourknowledge.

REFERENCES

https://seaborn.pydata.org/tutorial/color_palettes.html https://www.python-graph-gallery.com/196select-one-color-with-matplotlibhttps://www.analyticsvidhya.com/blog/2021/05/10-colab-tips-and-hacks-for-efficient-use-ofit/