

Analyzing and Predicting Ratings for Apps on Google Play Store

Made by: Nishtha

1

Importance of App Analysis

Why do we need it and who is interested?



Who needs it and why?

- ▶ **Developers:** so that they can build apps most common among users
- ▶ **Marketing Companies:** so that they can advertise in most popular apps
- ▶ **Common People:** to explore apps related to their interests

2

Data Acquisition and Cleaning

From where do we get the data?

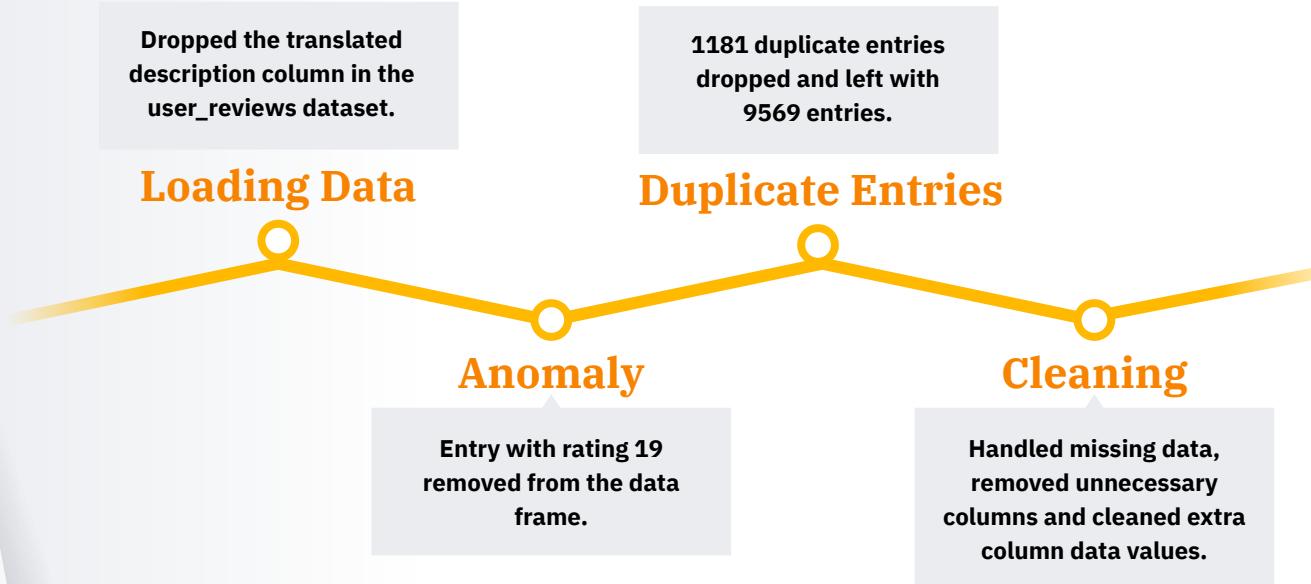
Kaggle

Data Source

(10841, 13)
Google Play Store Dataset

(64295, 5)
Google Play Store User Reviews Dataset

Data Cleaning Process

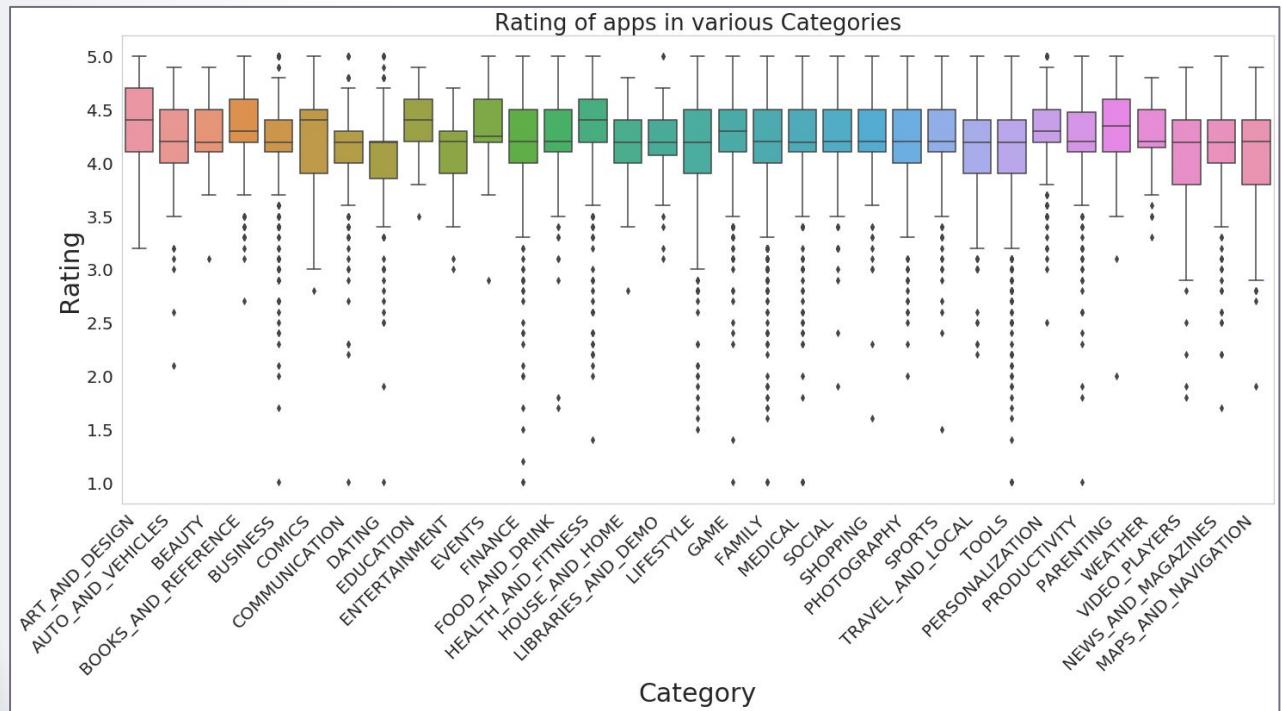


3

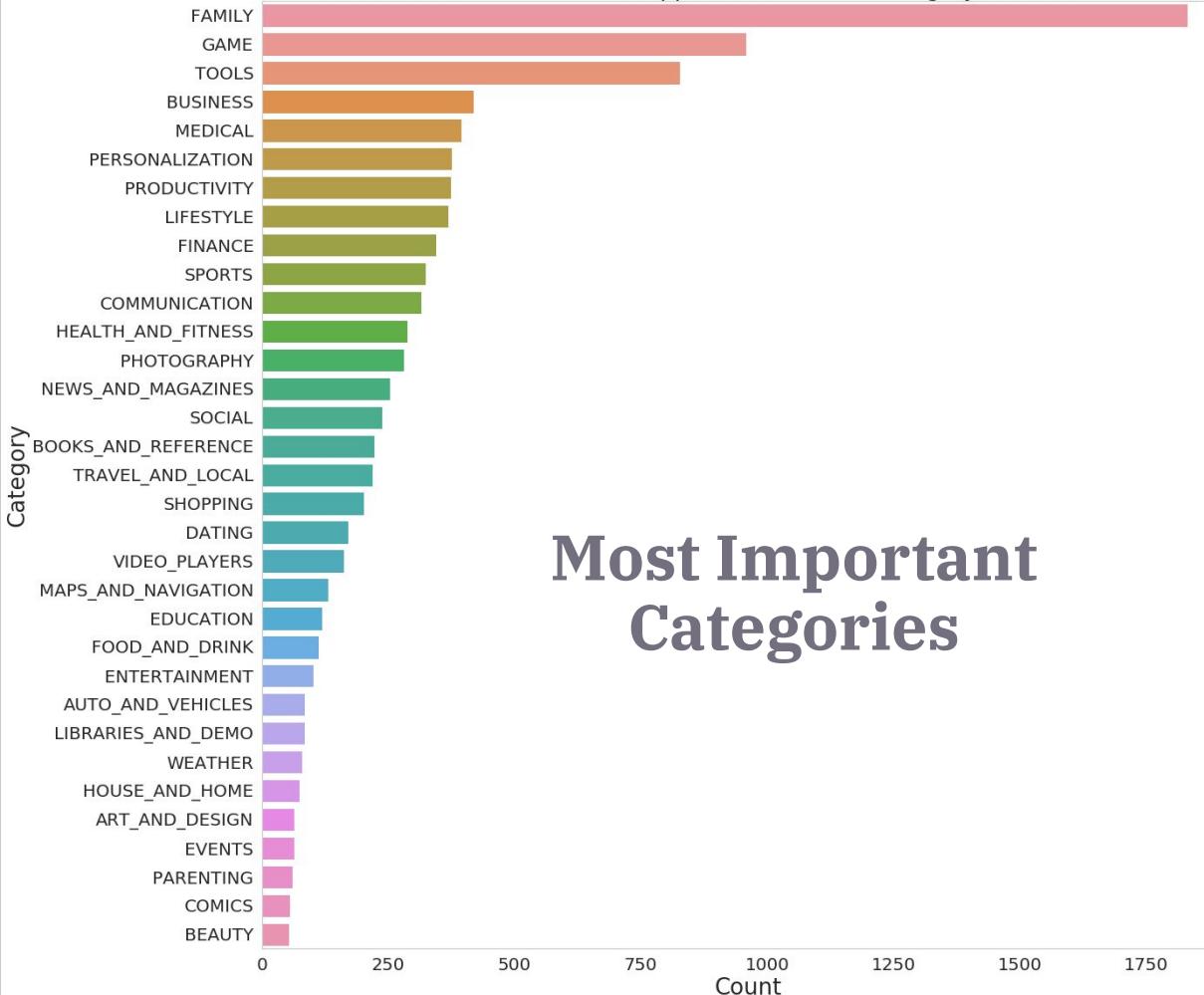
Exploratory Data Analysis

What does the data tell us?

Influence of Categories on Ratings



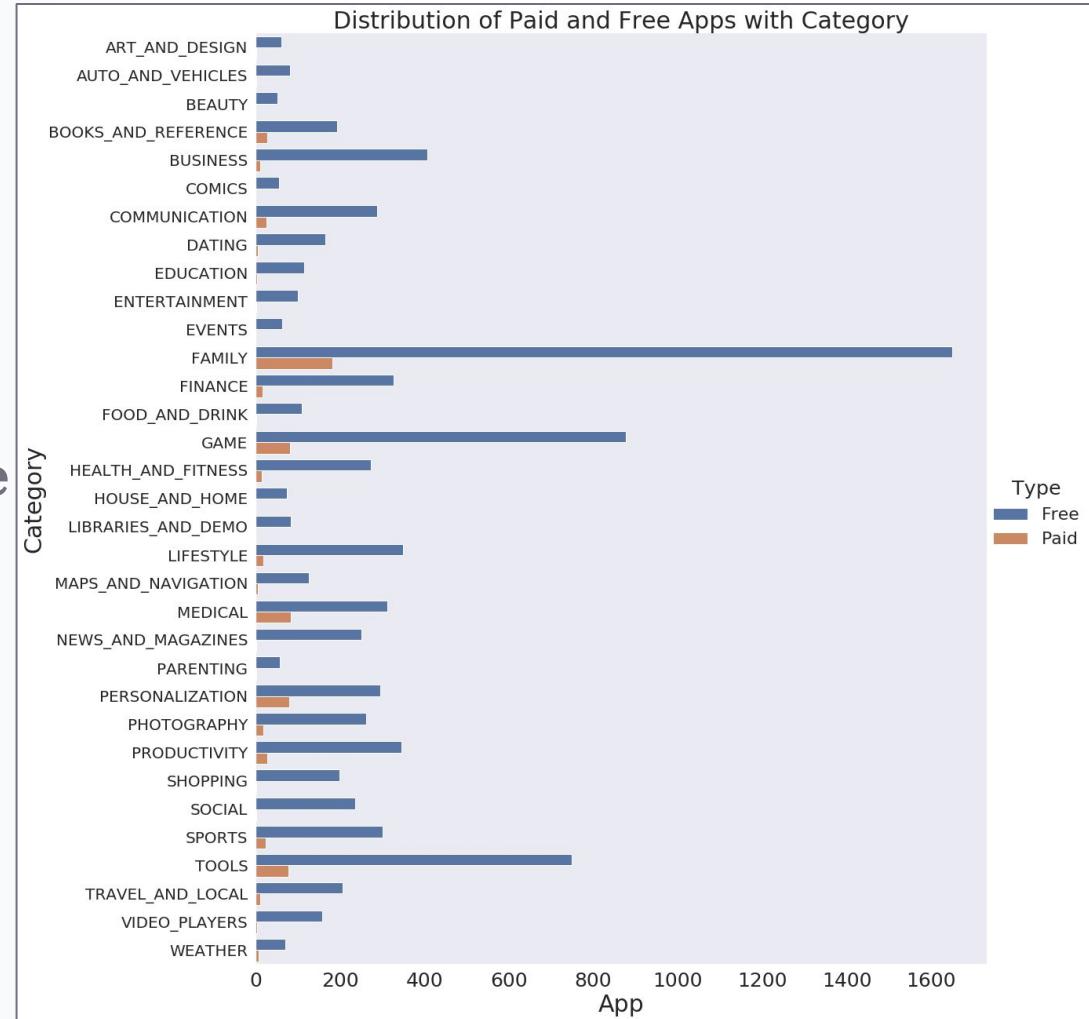
Number of apps of a Particular Category

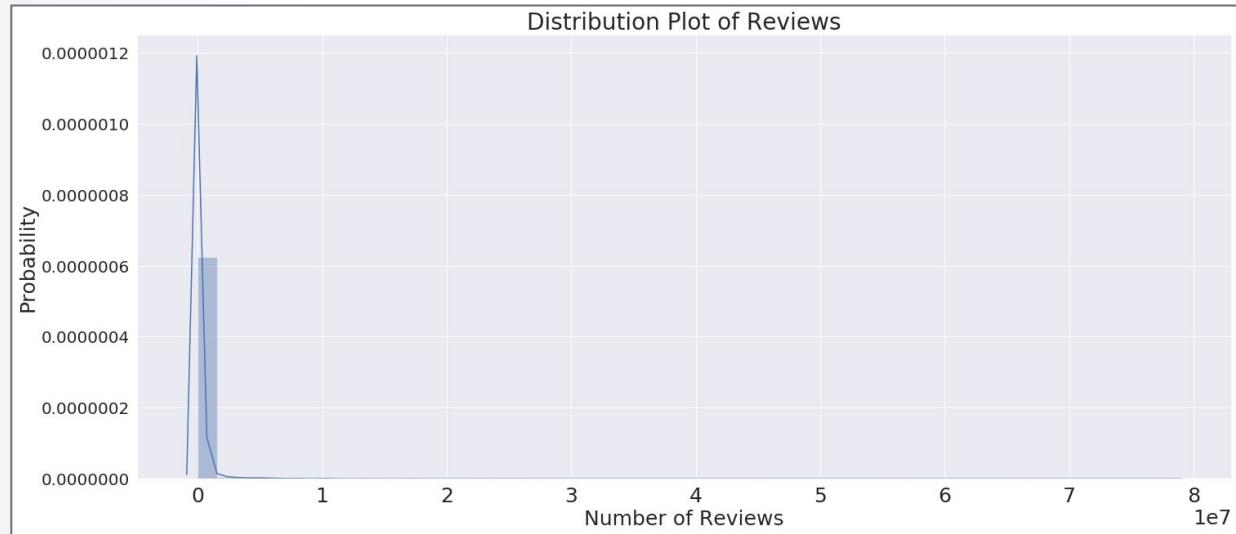


Top 3 Categories

Category	Number of Apps
Family	1,832
Games	959
Tools	827

Effect of Category and Type Combined

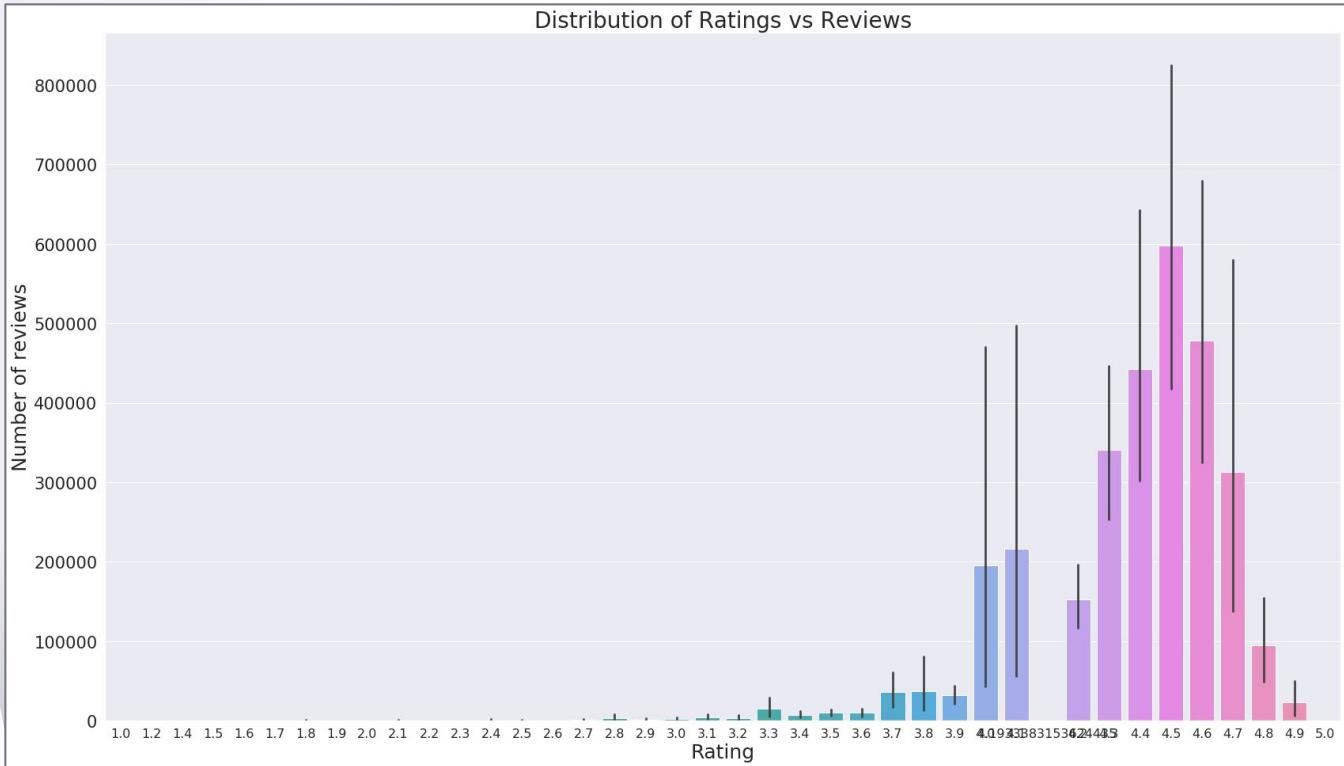


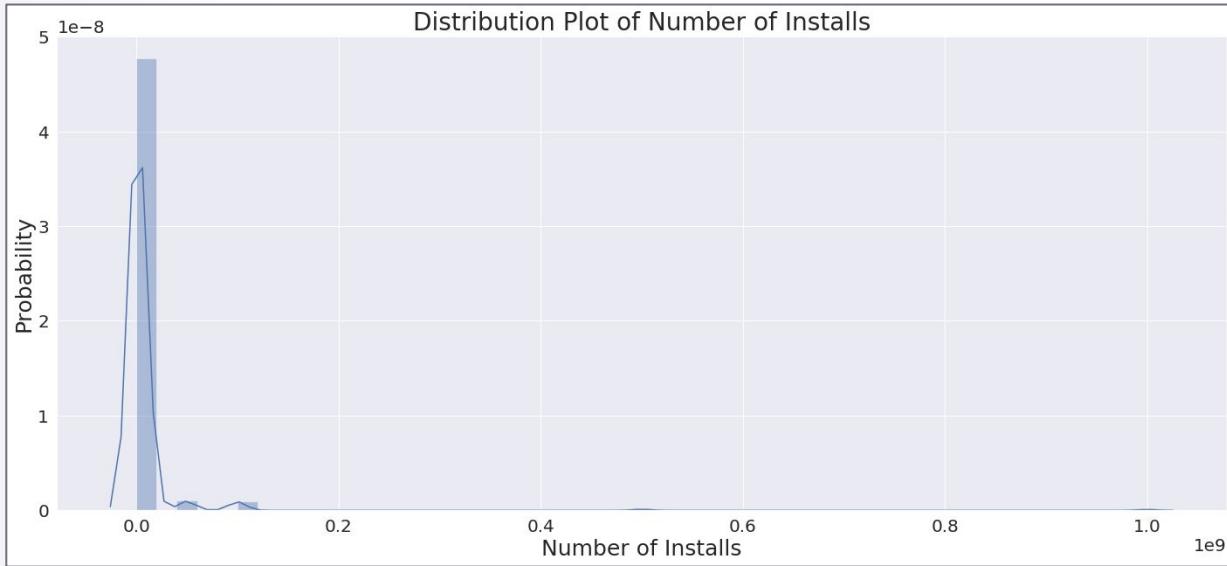


Most Reviewed Apps

- 1. Facebook**
- 2. WhatsApp Messenger**
- 3. Instagram**

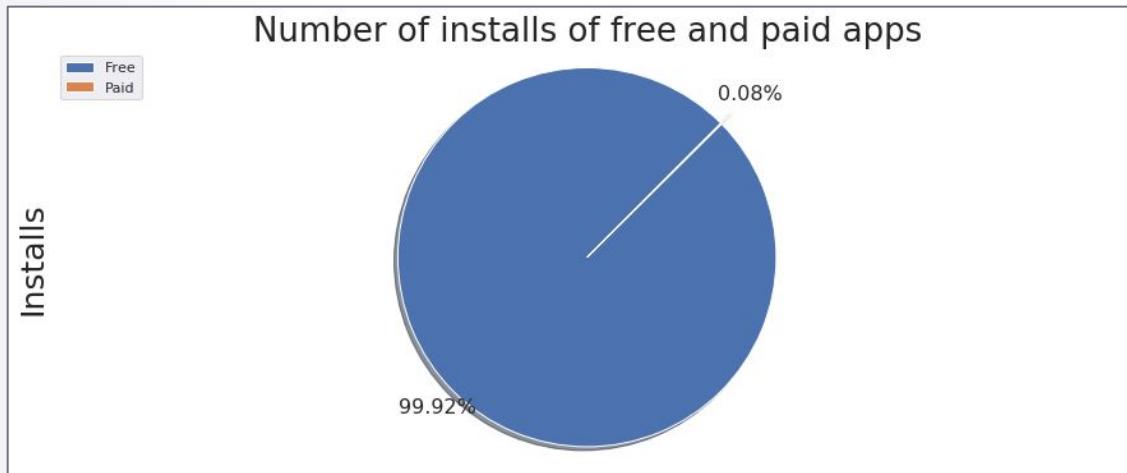
How Ratings and Reviews affect each other?





Most Installed Apps

- **Facebook**
- **WhatsApp Messenger**
- **Instagram**
- **YouTube**
- **Subway Surfers**
- **Google Drive**
- **Google Photos**
- **Google Play Music**
- **Google+**
- **Google**



Installs and Types

TYPE

INSTALLS

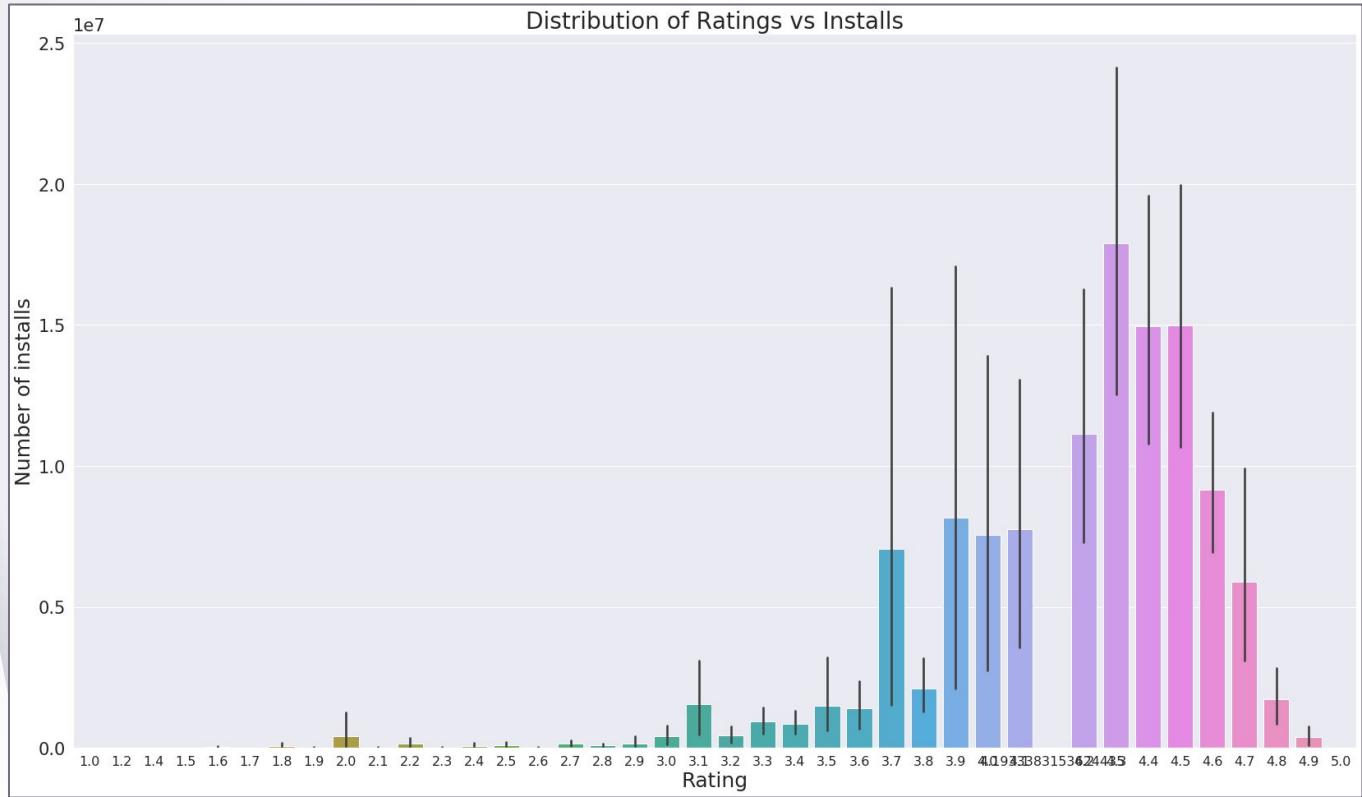
Free

75,065,572,646

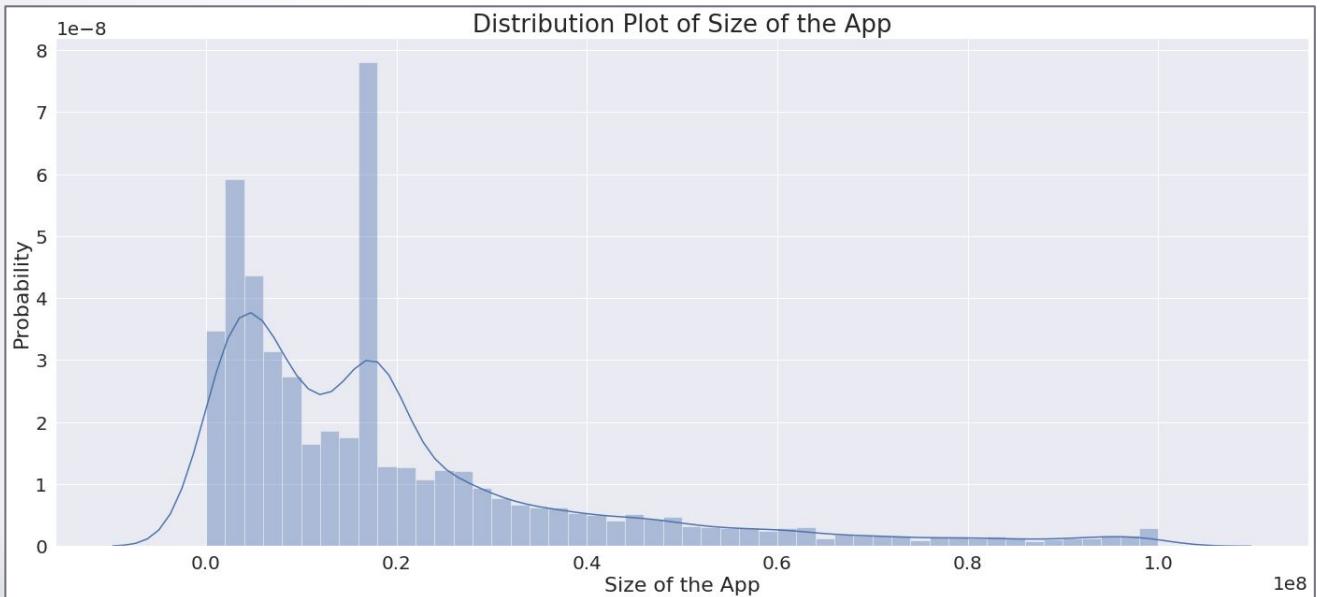
Paid

57,364,881

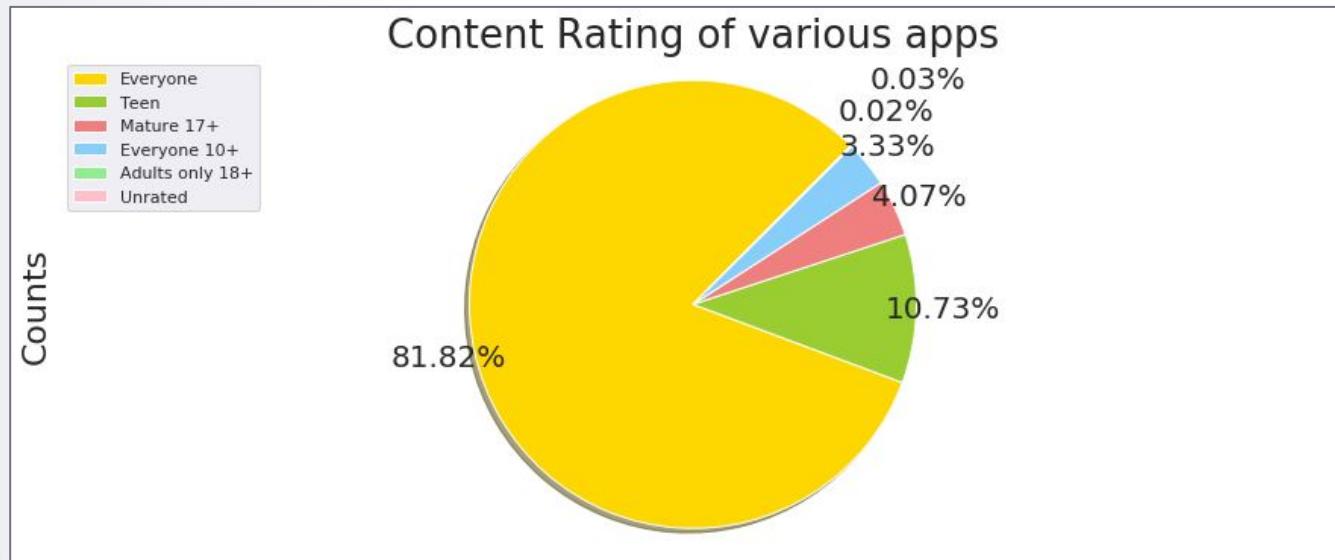
How Ratings and Installs affect each other?



Distribution of Size of App across Data set

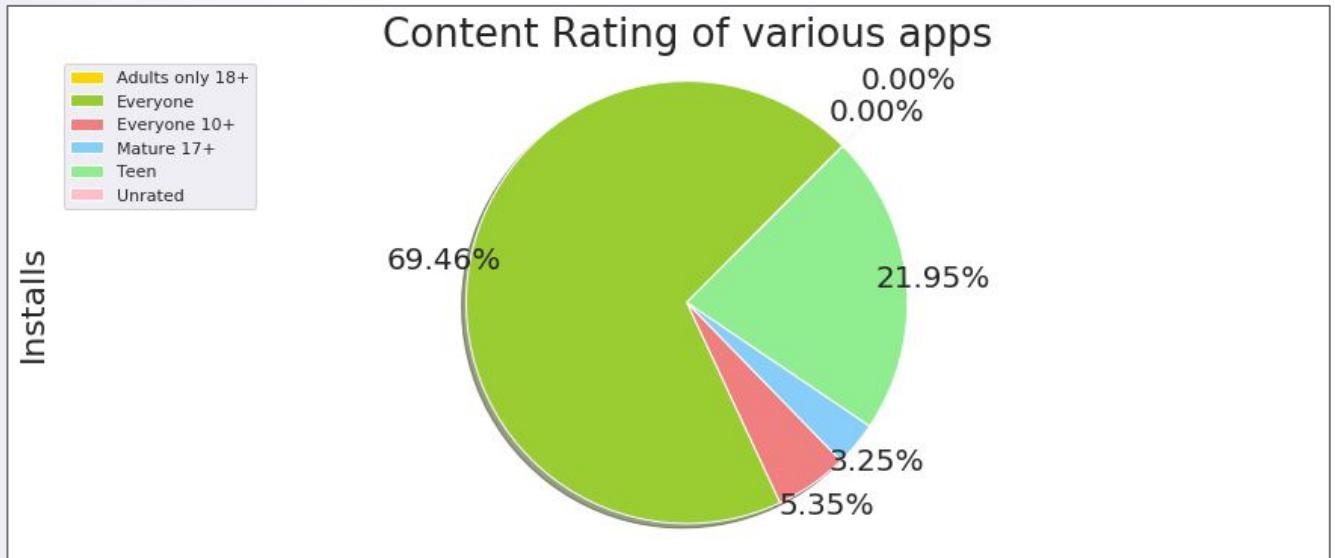


Content Ratings of Apps



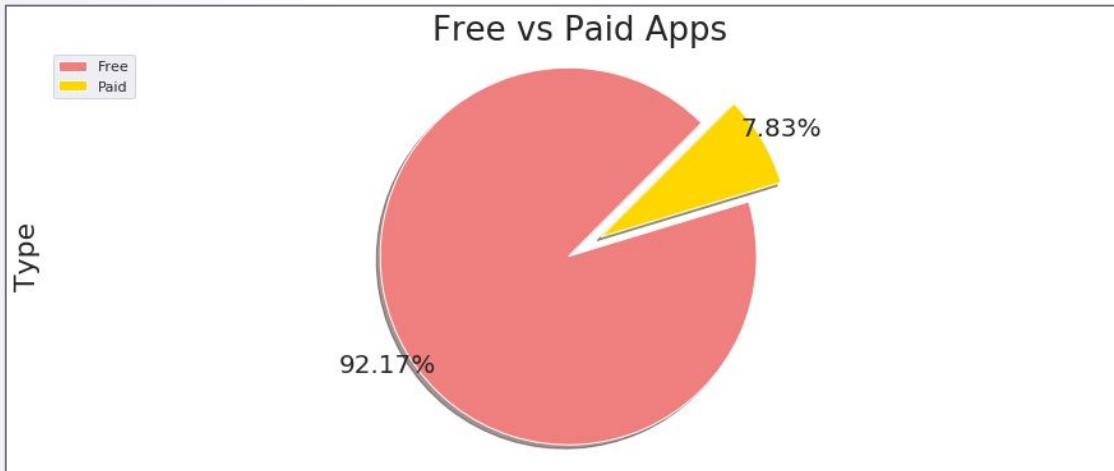
7,903
Content Rating : Everyone

Installs of Apps of different Content Rating



52,179,352,961

Installs for Content Rating : Everyone



Free vs Paid Apps

TYPE

NUMBER OF APPS

Free

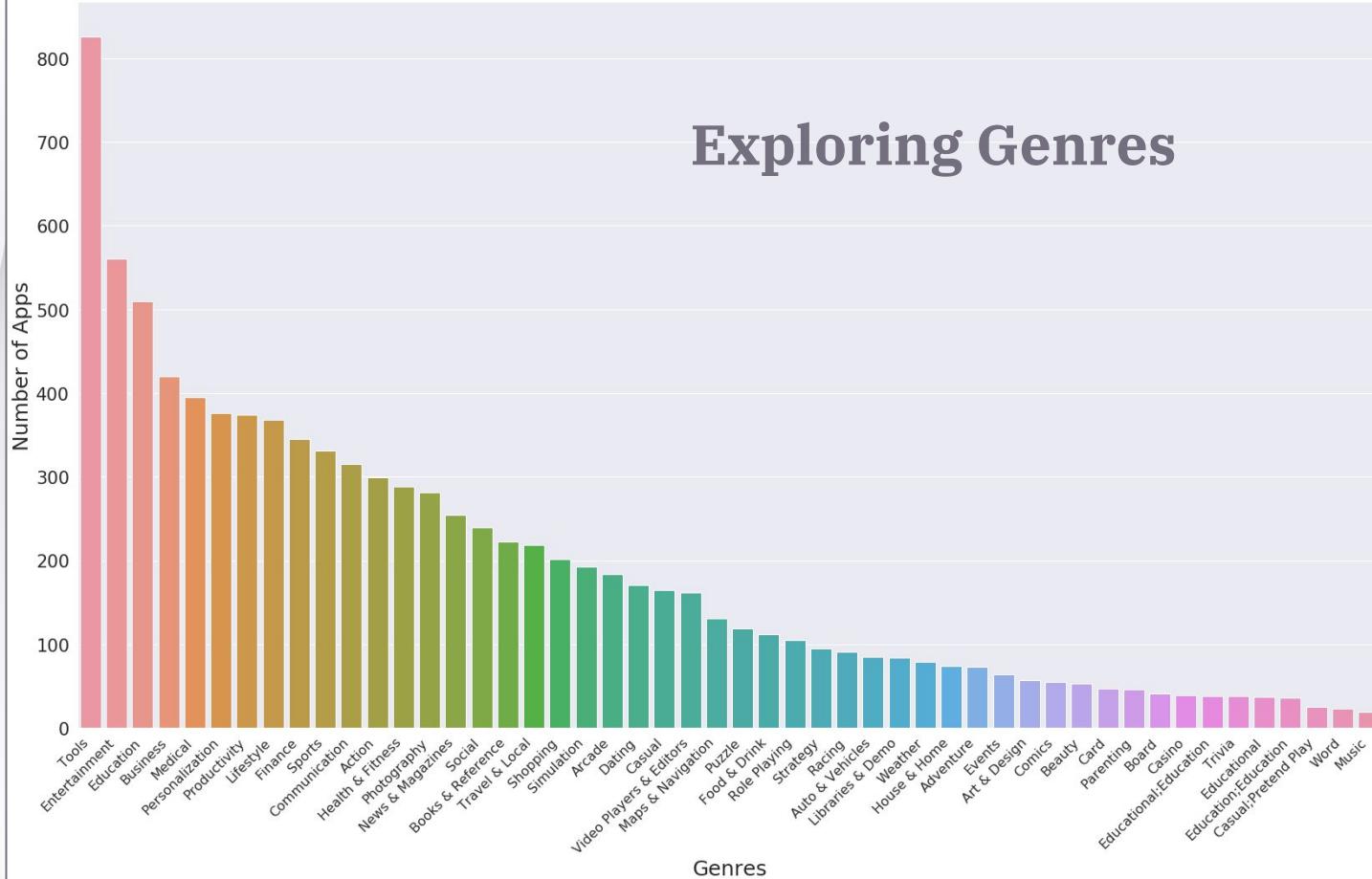
8,903

Paid

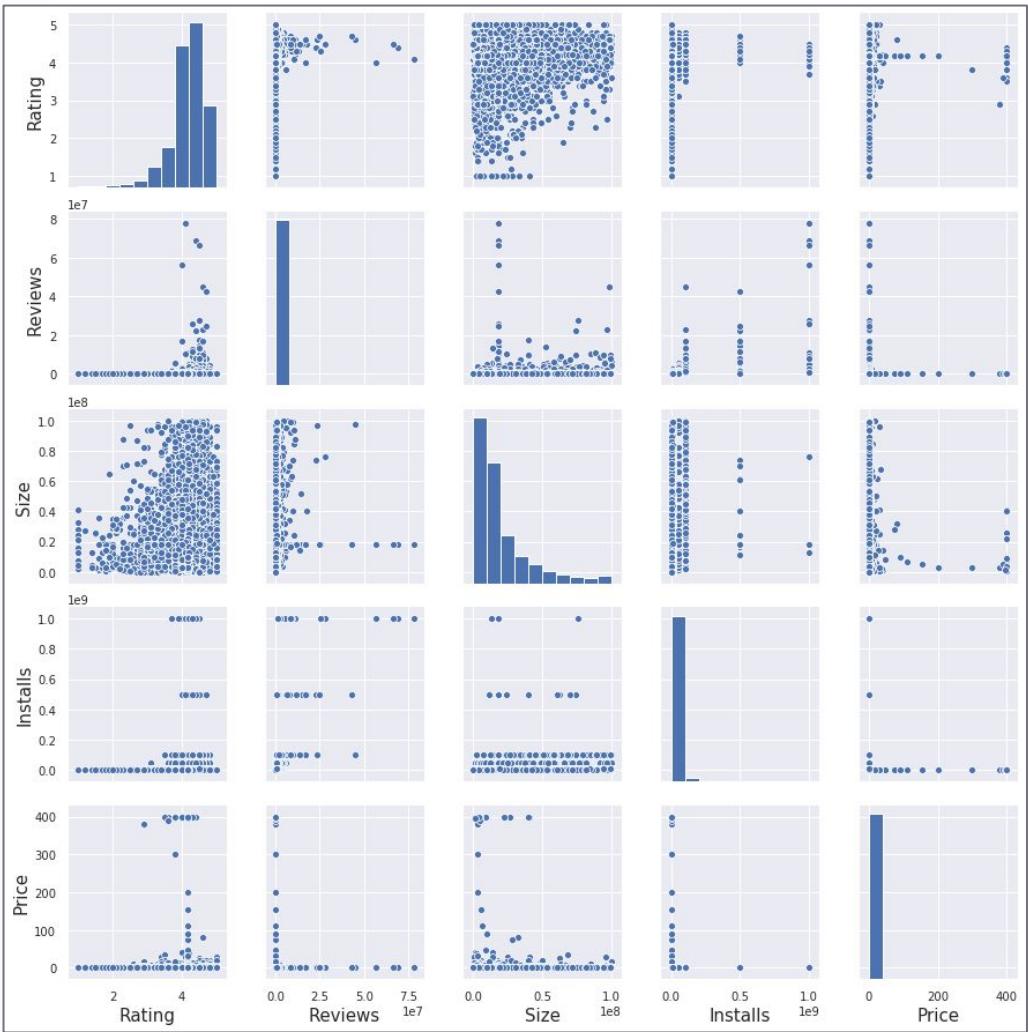
756

Distribution of Genres

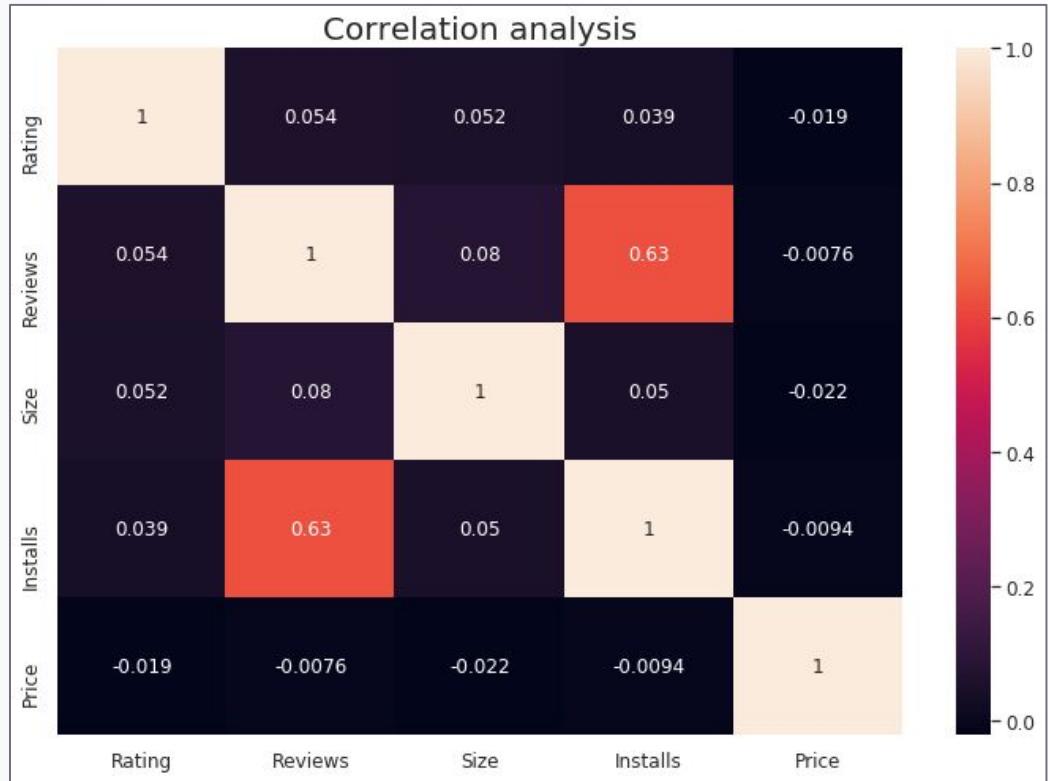
Exploring Genres



Pairwise Plots



Correlation Map

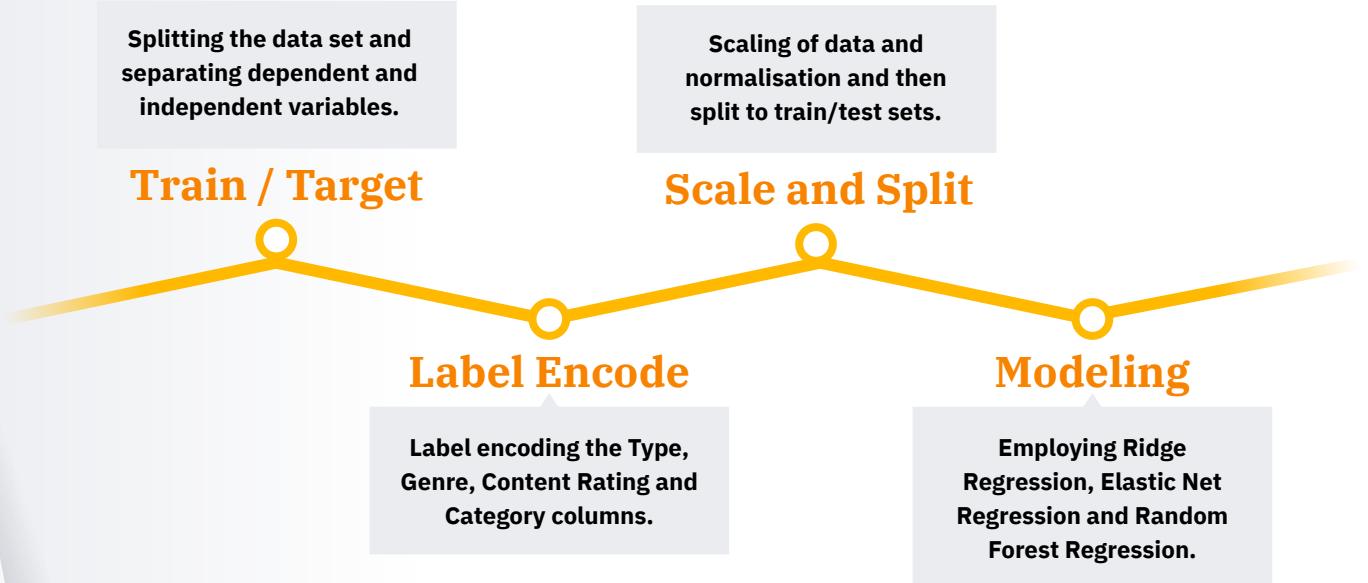


4

Creating a model for Data

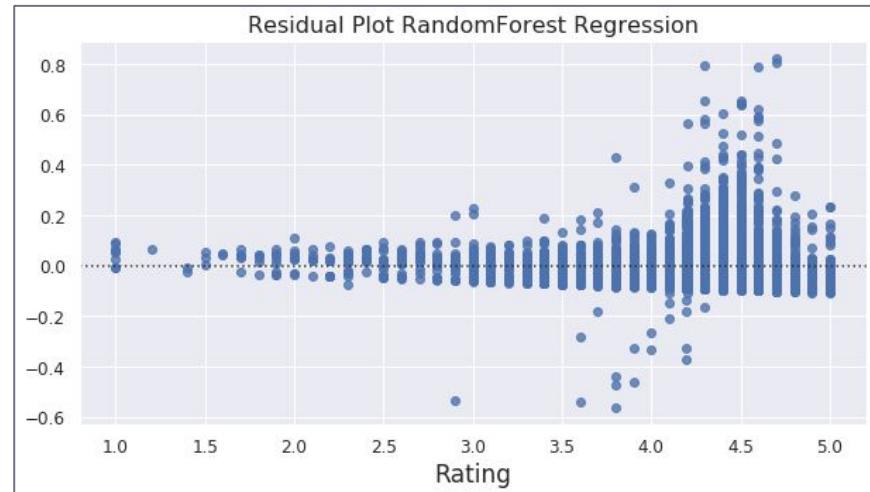
Let's make our model intelligent

Data Modeling Process



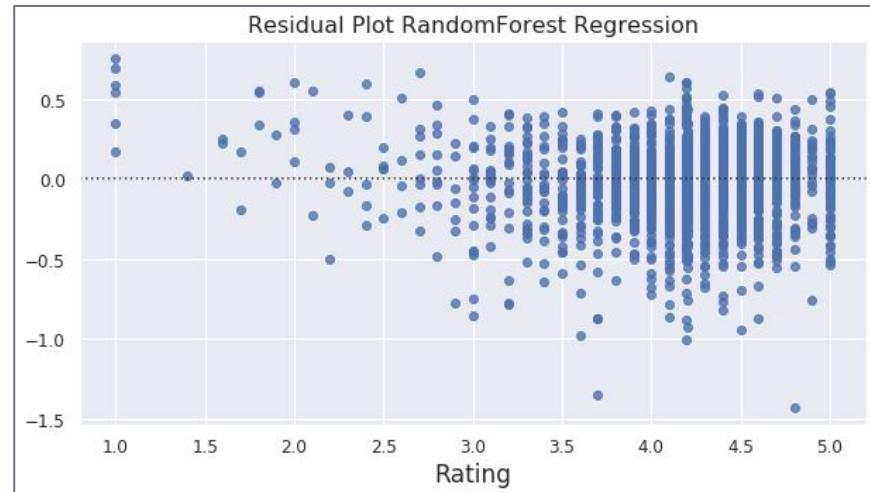
Random Forest Regression (Residual Plots)

Train Set



Train Set

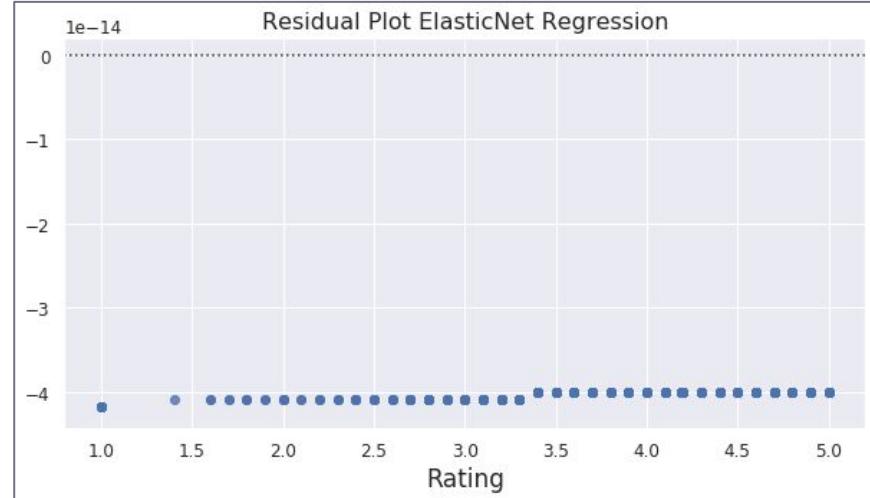
Test Set



Elastic Net Regression (Residual Plots)

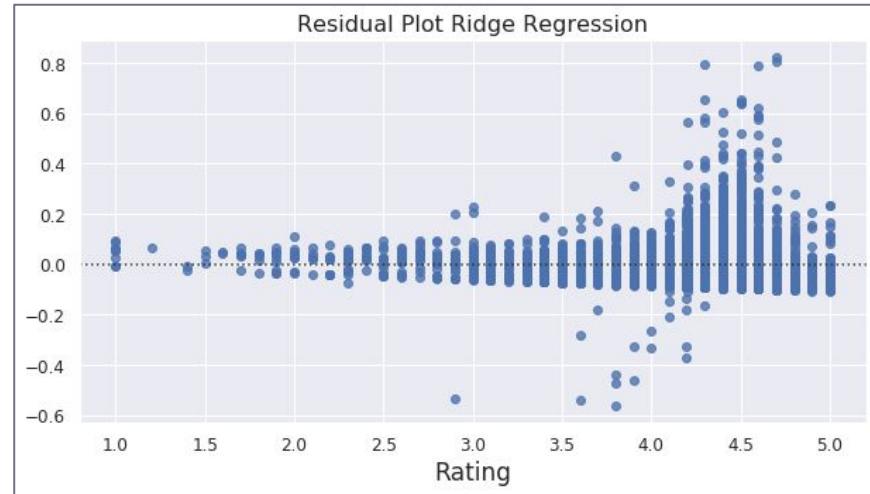
Train Set

Test Set

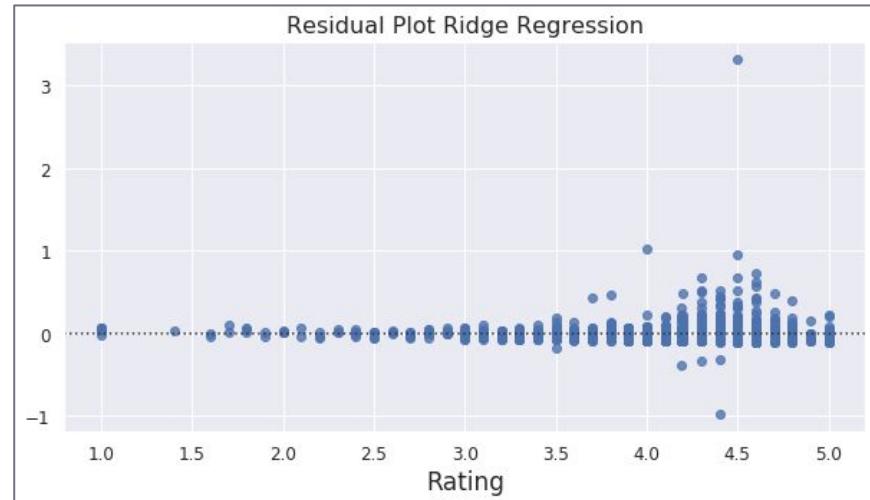


Ridge Regression (Residual Plots)

Train Set



Test Set



Results for various Models

MODEL	MSE* (TRAIN SET)	MSE* (TEST SET)
Random Forest	0.03139	0.22994
Ridge	0.24119	0.25395
Elastic Net	0.23563	0.25423

5

Concluding Remarks

What did we learn ?



Major Findings

- ▶ **Developers:** should build quality apps in most common categories and may also try to target less common categories only if they have good content.
- ▶ **Marketing Companies:** advertise in top 40 most installed apps for best results.
- ▶ **Category and Genre:** affect the app market the most
- ▶ **Paid vs Free:** free apps are most preferred
- ▶ **Size:** small sized apps are preferred by customers
- ▶ **Content Rating:** ‘Everyone’ has the largest viewership

Thank You