GOOGLE PLAY STORE APPS

Group 7
Amey Athaley
Javeria Rangoonwala
Sahana Subramanian
Victor Nguyen

WHY ANALYZE THE GOOGLE PLAY STORE?

Mobile App Market is set to grow 20% by 2023

Android Apps comprise 90% of the Mobile App Market

What makes an App popular? Can we predict how popular it's going to be?

What are some interesting patterns in user behavior related to app usage & feedback?

OVERVIEW OF ANALYSIS

Data Cleaning

Understand the structure of the dataset and clean data before analysis

Data Exploration

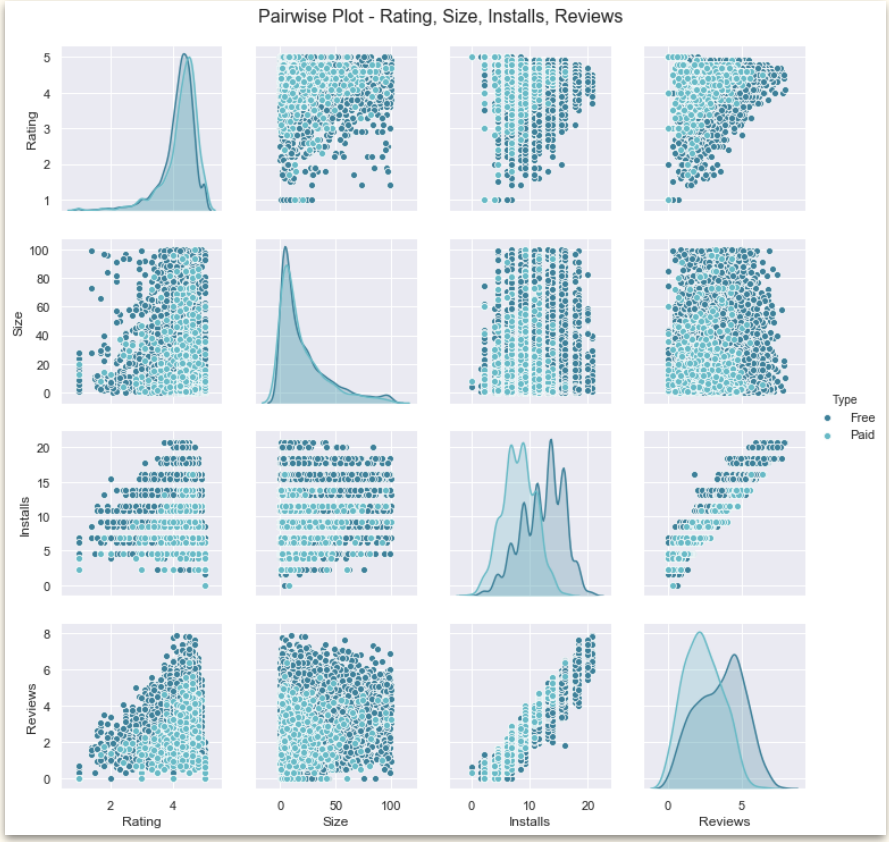
Uncover initial patterns, characteristics, and points of interest using visual exploration

Predictive Modeling

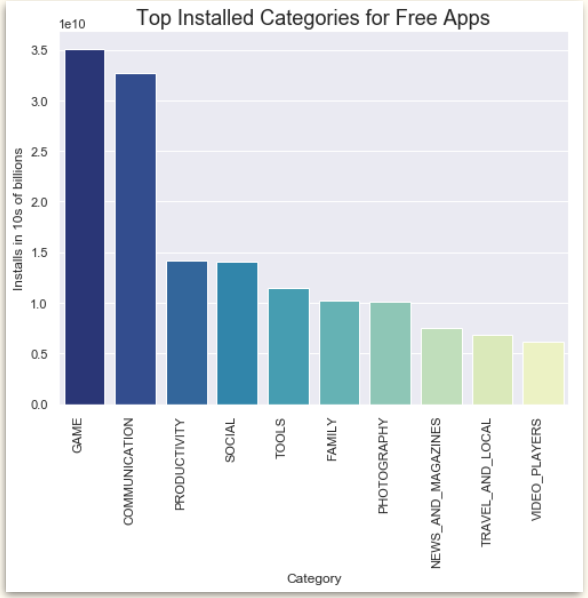
Formulate a statistical model to forecast an outcome using relevant predictors

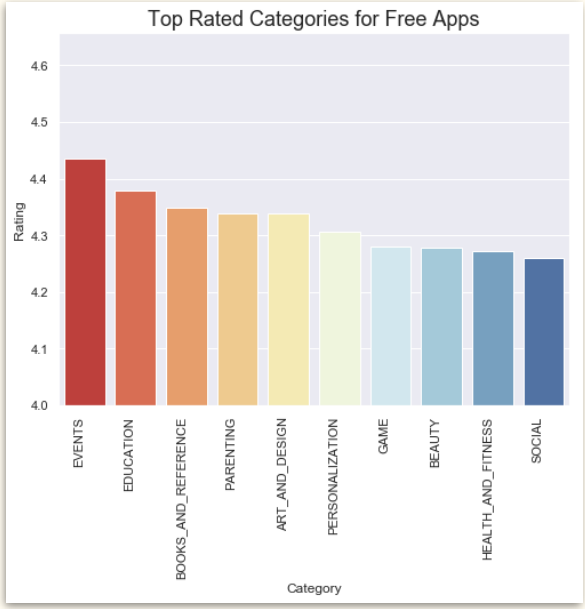
EXPLORATORY ANALYSIS

BIVARIATE ANALYSIS



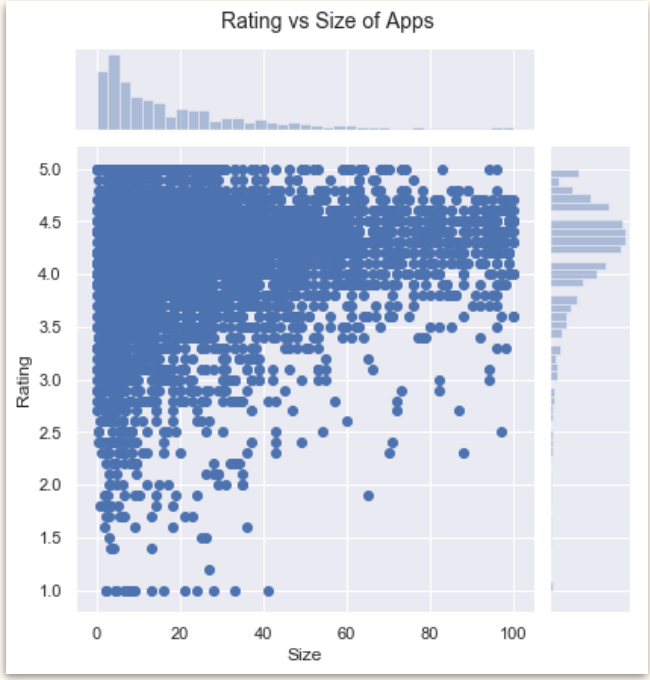
CATEGORIES IN DEMAND

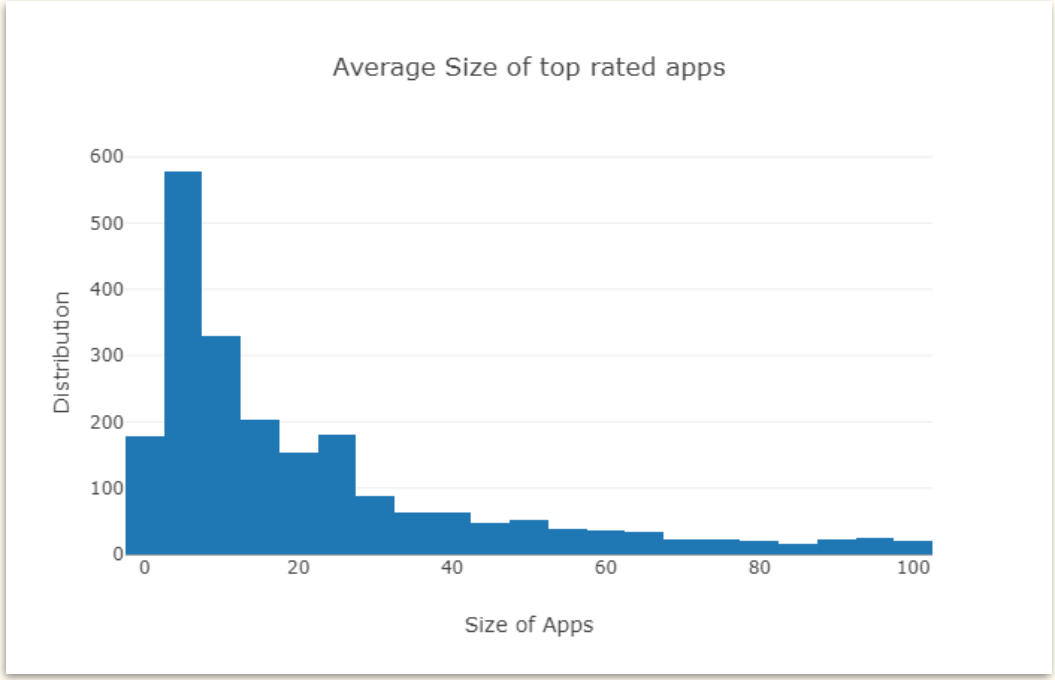




- Communication, Productivity, Tools, Family, Photography, News & Magazines, Travel & Local, Video Players are untapped free app categories
- High Installs because of a high underlying demand
- Low user satisfaction

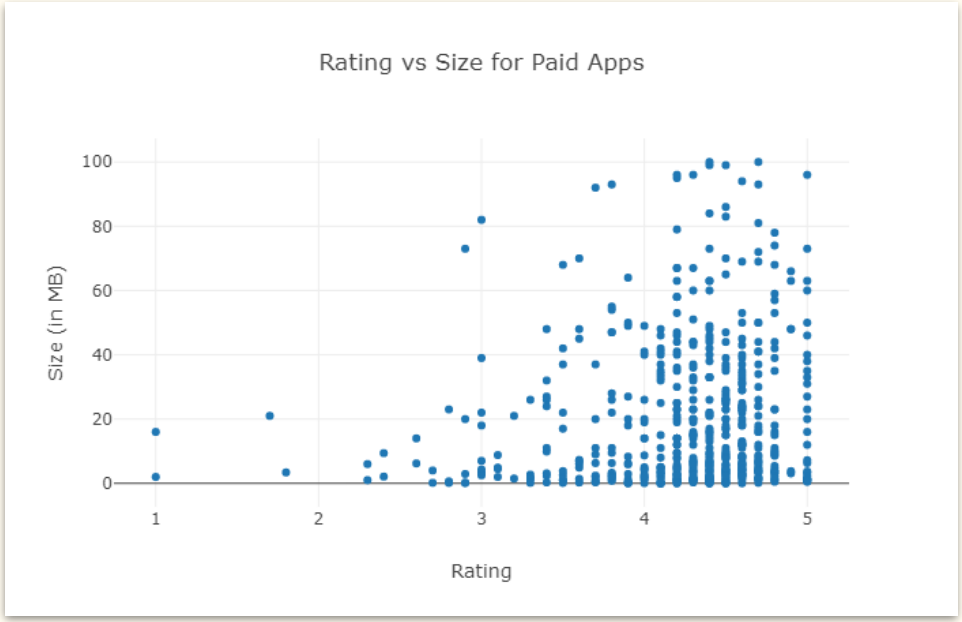
SIZING STRATEGY

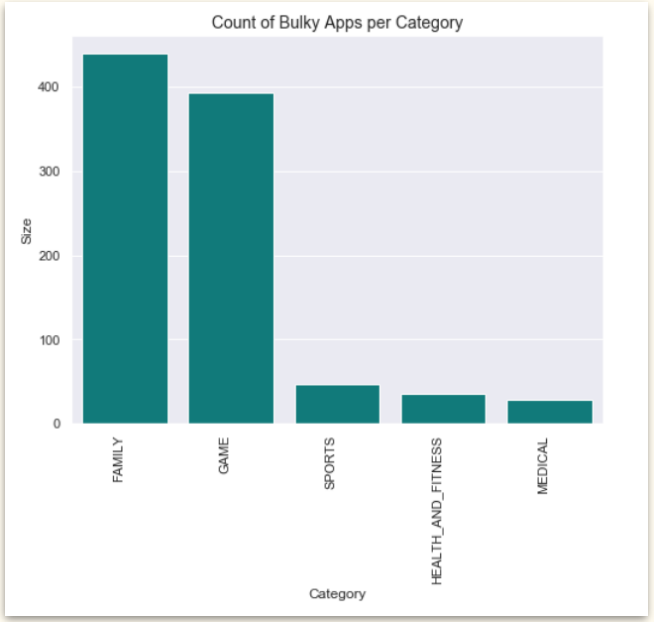




Highly rated apps are optimally sized between ~2 MB and ~40 MB

SIZING STRATEGY

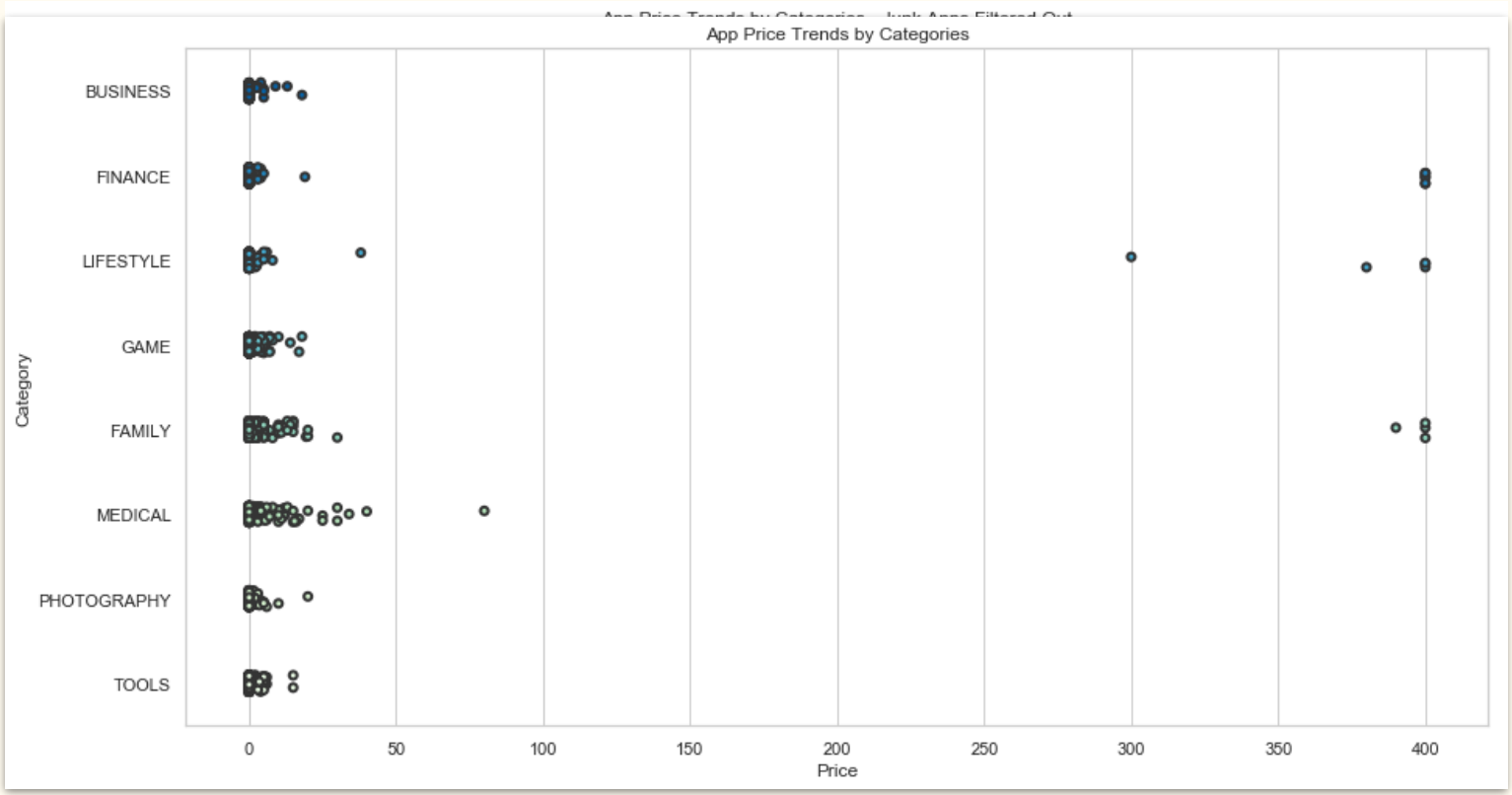




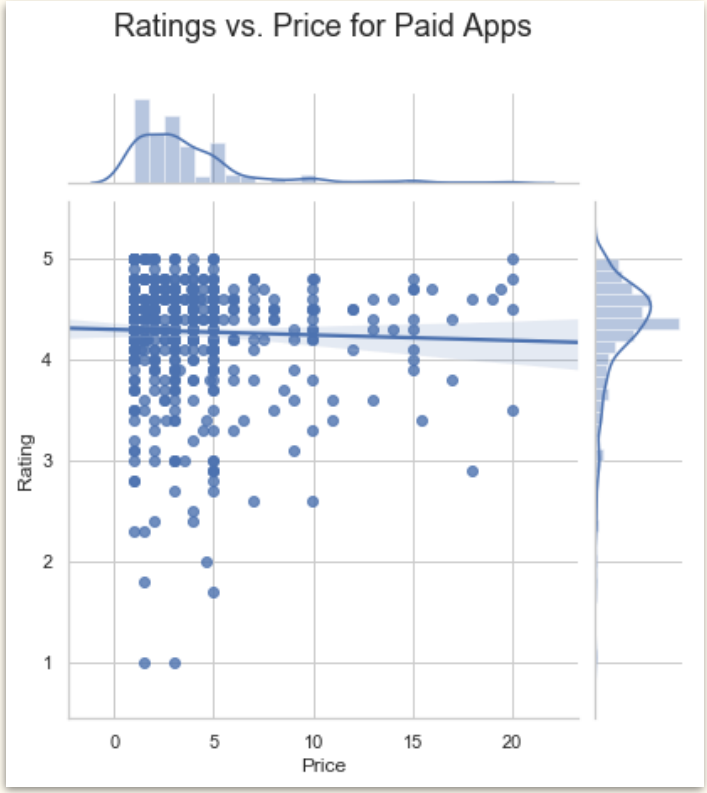
Users prefer to pay for apps that are light-weighted

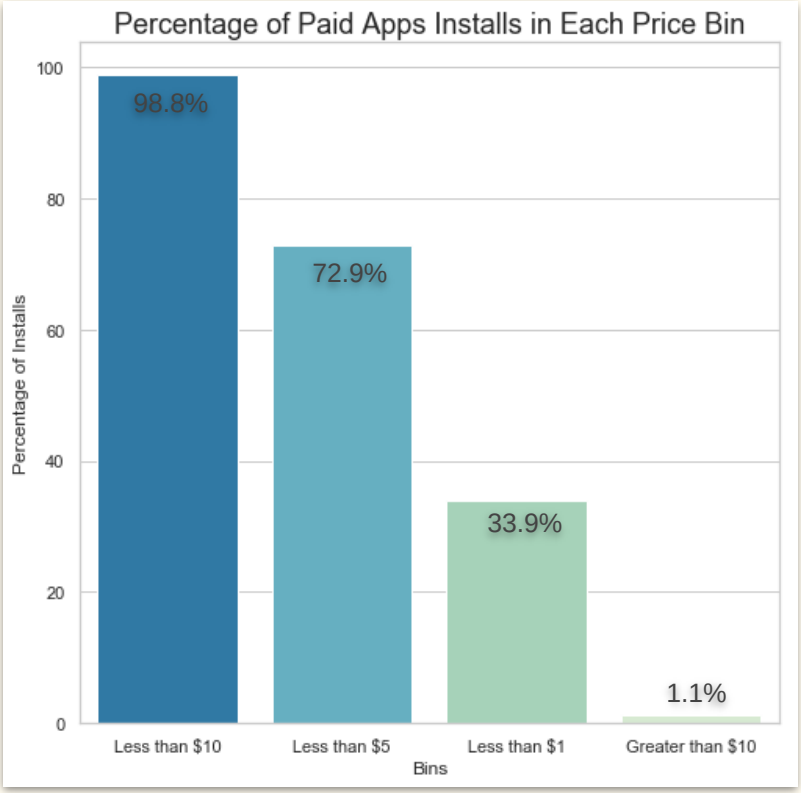
PRICING STRATEGY

App	Price
Vargo Anesthesia Mega App	79.99
[illegible]expensive app	[illegible]
💎 I'[illegible]	[illegible]
[illegible]ump[illegible]	[illegible]0
[illegible]	[illegible].99
[illegible]	[illegible]99.99
[illegible]	299.99
[illegible]	[illegible] 399.99
I[illegible]	[illegible]h 379.99
[illegible]	399.99
[illegible]	[illegible]99.99
[illegible]	[illegible].99
I am[illegible]	[illegible]99
[illegible] I Ar[illegible]	[illegible]
[illegible] I am [illegible]	[illegible]
[illegible]RICH PRO PLU[illegible]	399.99

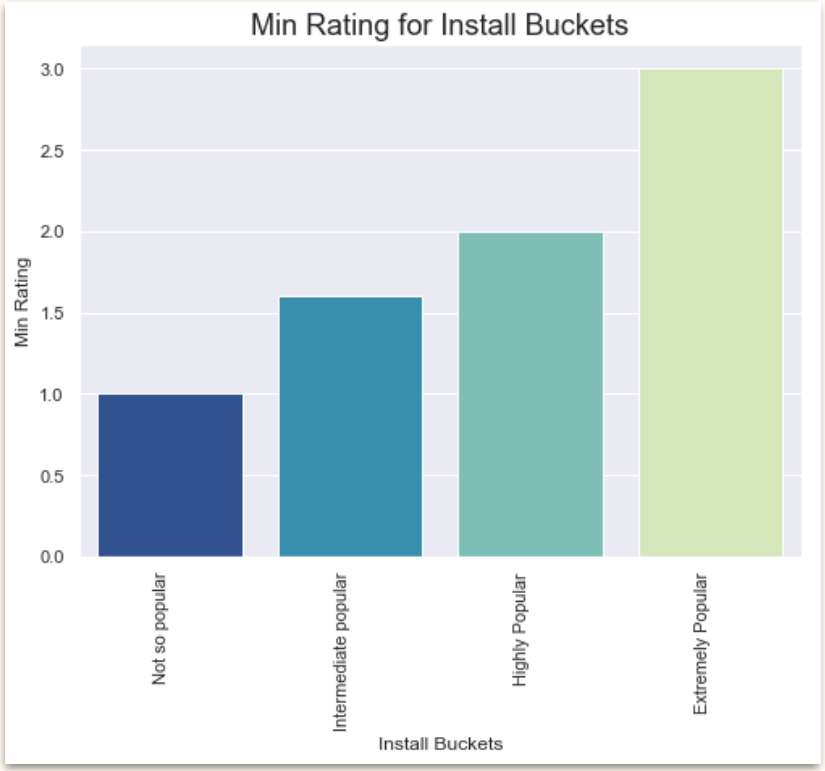


PRICING STRATEGY





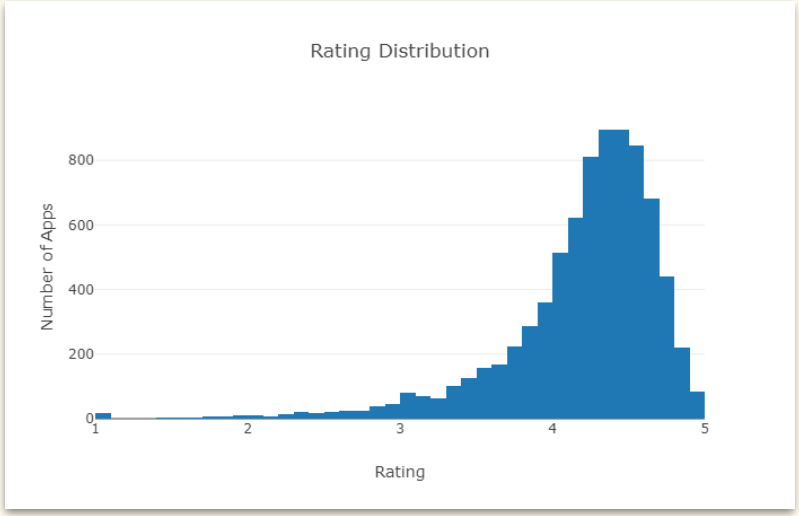
HOW DOES MINIMUM RATING VARY WITH INSTALLS?

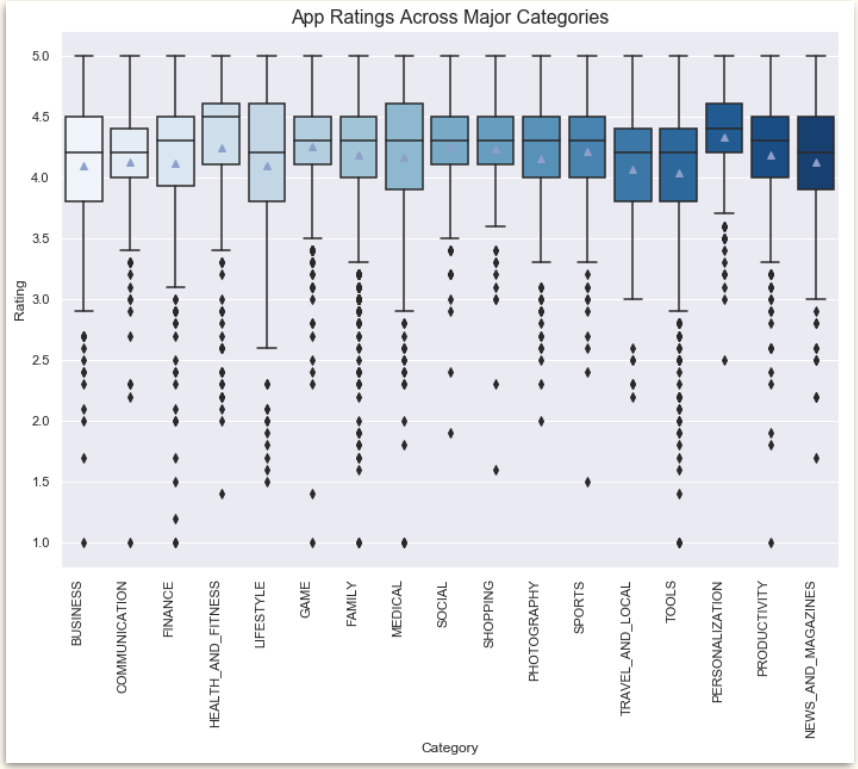


Installs	Install Buckets
0 - 10000	Not so popular
10000 - 100000	Intermediate popular
100000 - 500000	Highly Popular
500000+	Extremely Popular

Finding: An app that finds a popular audience is less likely to be critically rated.

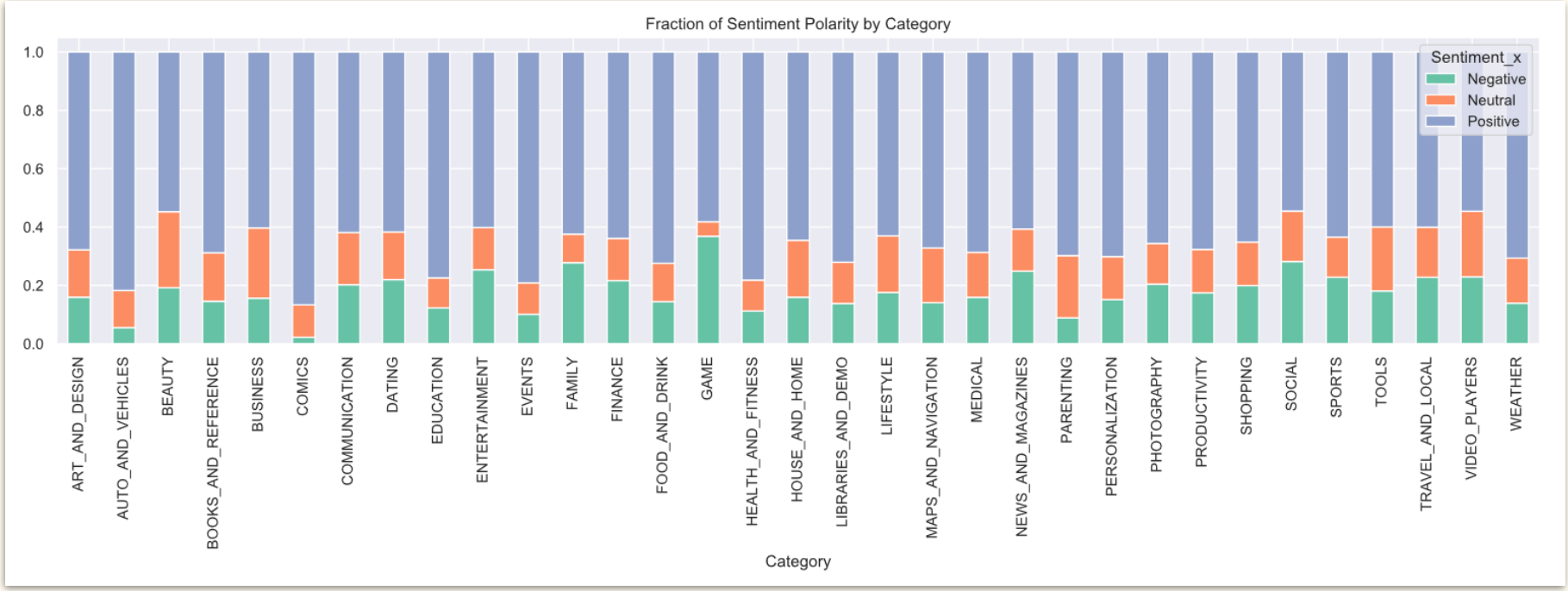
RATING DISTRIBUTION



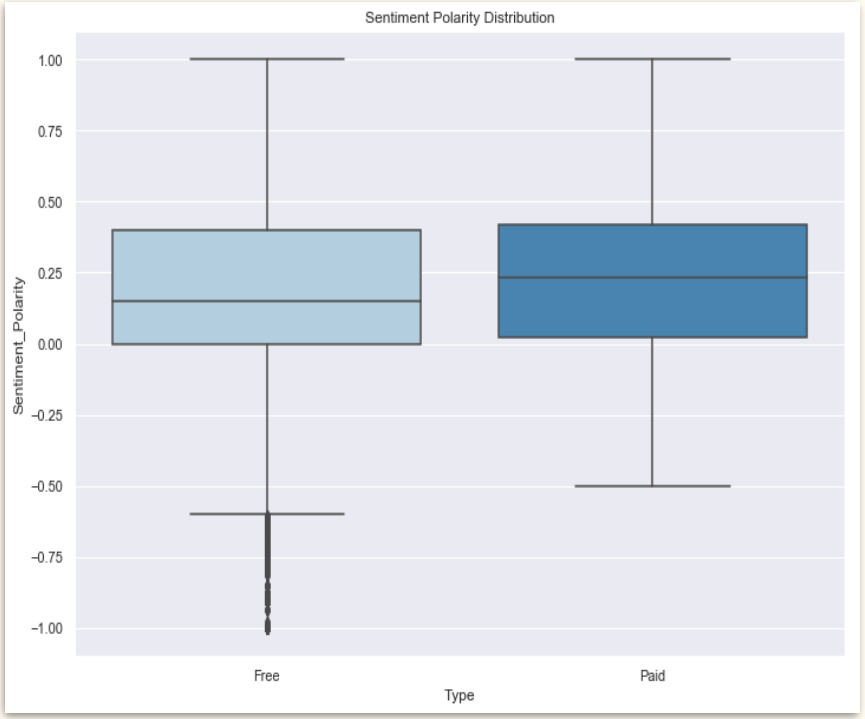


- Average rating of apps is 4.17
- One-way Anova test revealed that the average ratings of categories are statistically different
- Best performing apps - Health and Fitness
- Worst performing apps - Dating

SENTIMENT POLARITY BY CATEGORY



SENTIMENT POLARITY BY TYPE (FREE/PAID)

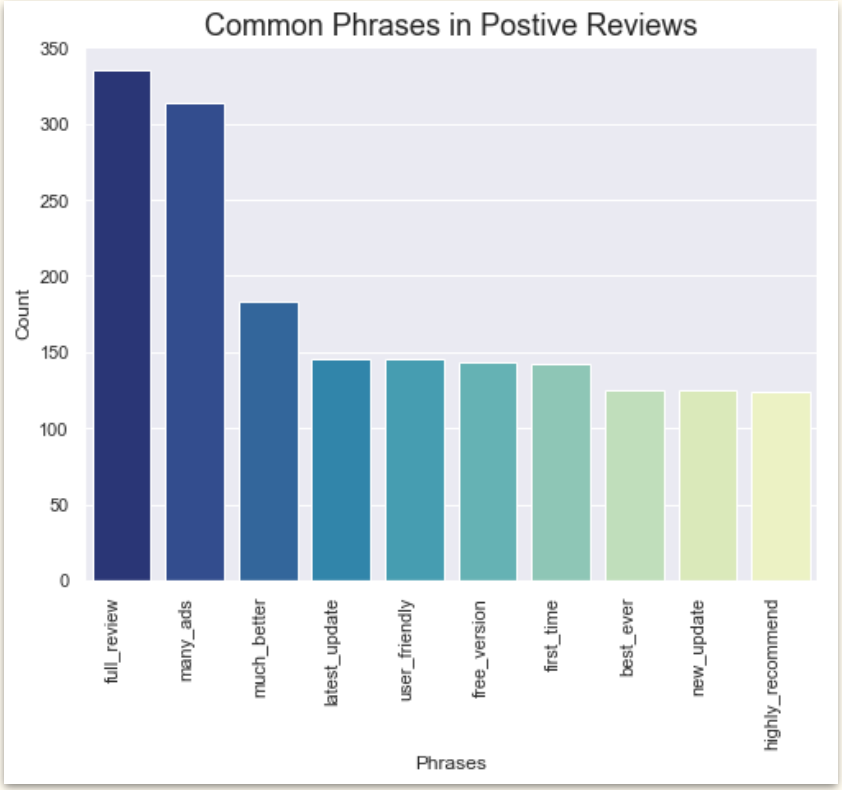


People are harsher towards free apps whereas users are more tolerant when they are paying for it.

WORD CLOUD - POSITIVE REVIEWS

WORD CLOUD - NEGATIVE REVIEWS

COMMON PHRASES - POSITIVE REVIEWS

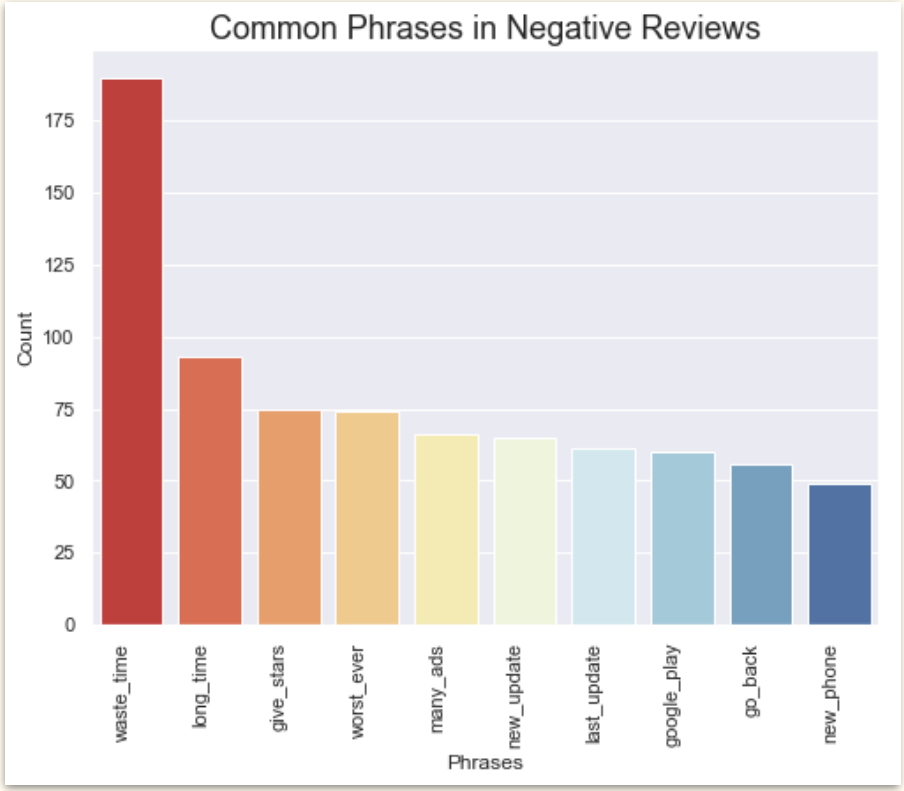


Common phrases in positive reviews:

- User-friendly
- Free version
- Works great
- Highly recommend

Usability is one of the main reasons for positive reviews

COMMON PHRASES - NEGATIVE REVIEWS



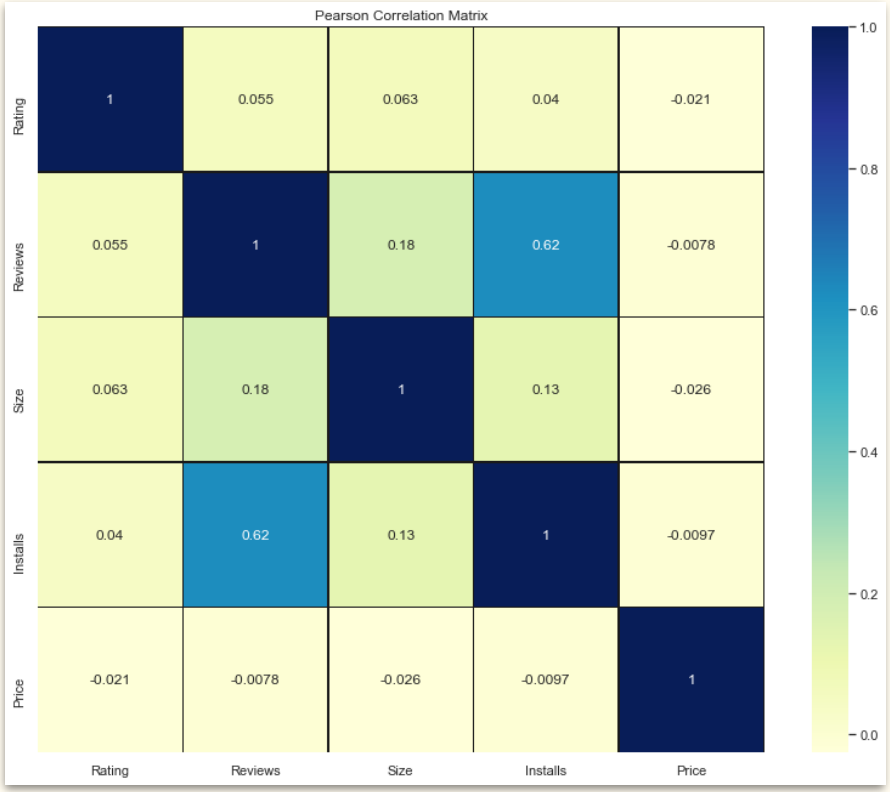
Common phrases in negative reviews:

- Waste time
- Many ads
- Spend money
- Takes forever

Loading time is one of the main reasons for negative reviews

PREDICTIVE MODELING

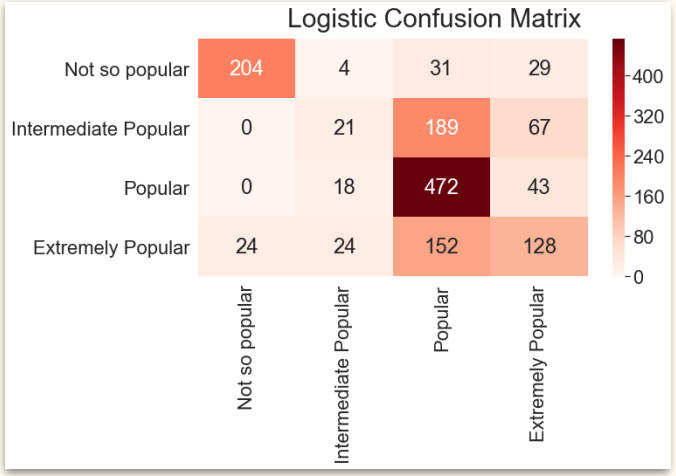
HOW TO PREDICT THE POPULARITY OF AN APP

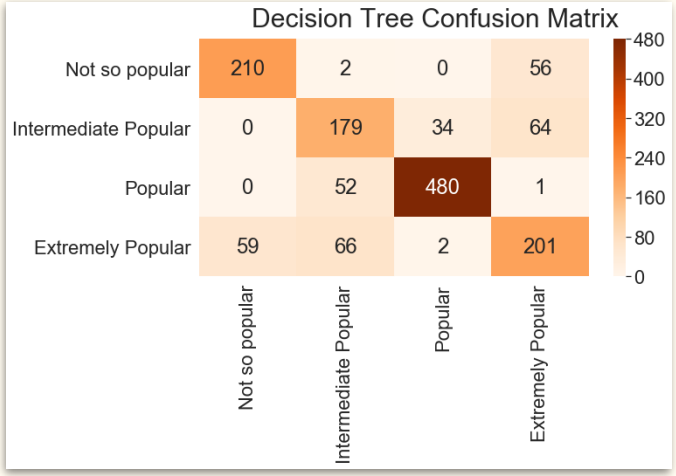


Based on the correlation plot, we identified the following variables that we could use in our classification problem :

1. Category
2. Genre
3. Reviews
4. Content Rating
5. Size
6. Type

LOGISTIC REGRESSION AND DECISION TREES

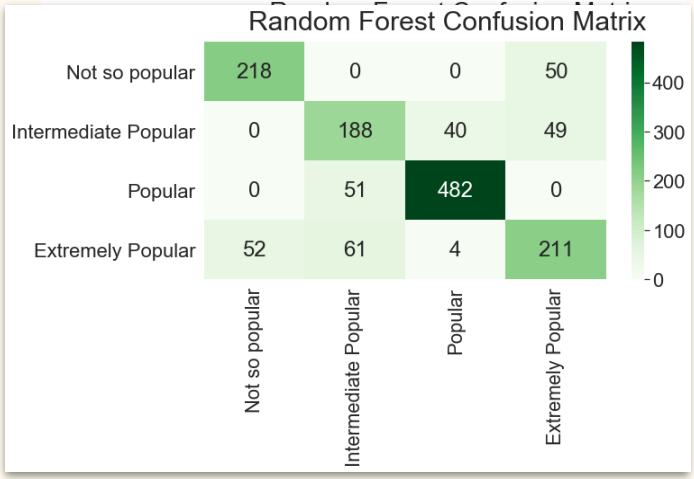


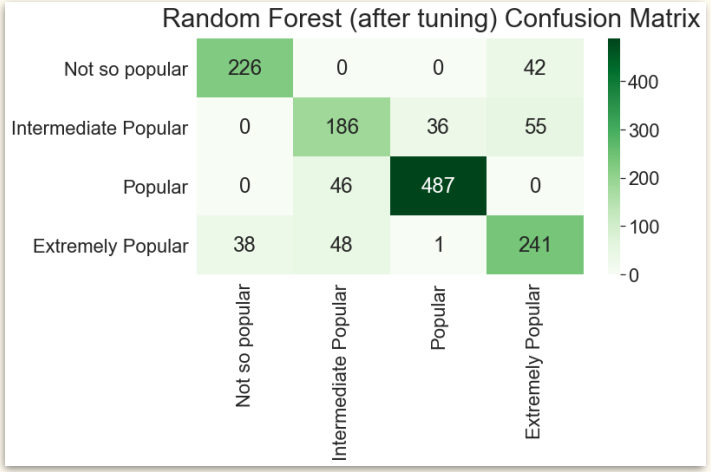


We bucketed installs into 4 bins based on quartile values & used test & hold out sets to perform the following classification algorithms :

1. Logistic Regression ~ 58.67%
2. Decision Trees ~ 76.10%

HYPERTUNING RANDOM FOREST



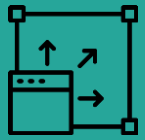


We used hypertuning & 3 fold cross validation to obtain a optimal classification accuracy of 81.08% on Random Forest Classifier

Parameters used:

1. No of trees
2. Level of depth of the trees
3. Minimum number of samples required to split a node
4. No of variables to consider at each split
5. Min/Max no of samples required at each node

SUMMARY



SIZE

Highly rated apps were optimally sized between 2MB to 40 MB

Paid apps with specific functionality were lighter

PRICE

Free apps outperform paid apps

If an app is paid, the likelihood of it being popular is higher if it's priced under $10

RATING

Positive relation between installs and rating.

However, highly installed apps are not always highly rated

REVIEWS

Issues like loading time and positive features like usability were revealed by sentiment analysis.

People tend to review harsher for paid apps

QUESTIONS?