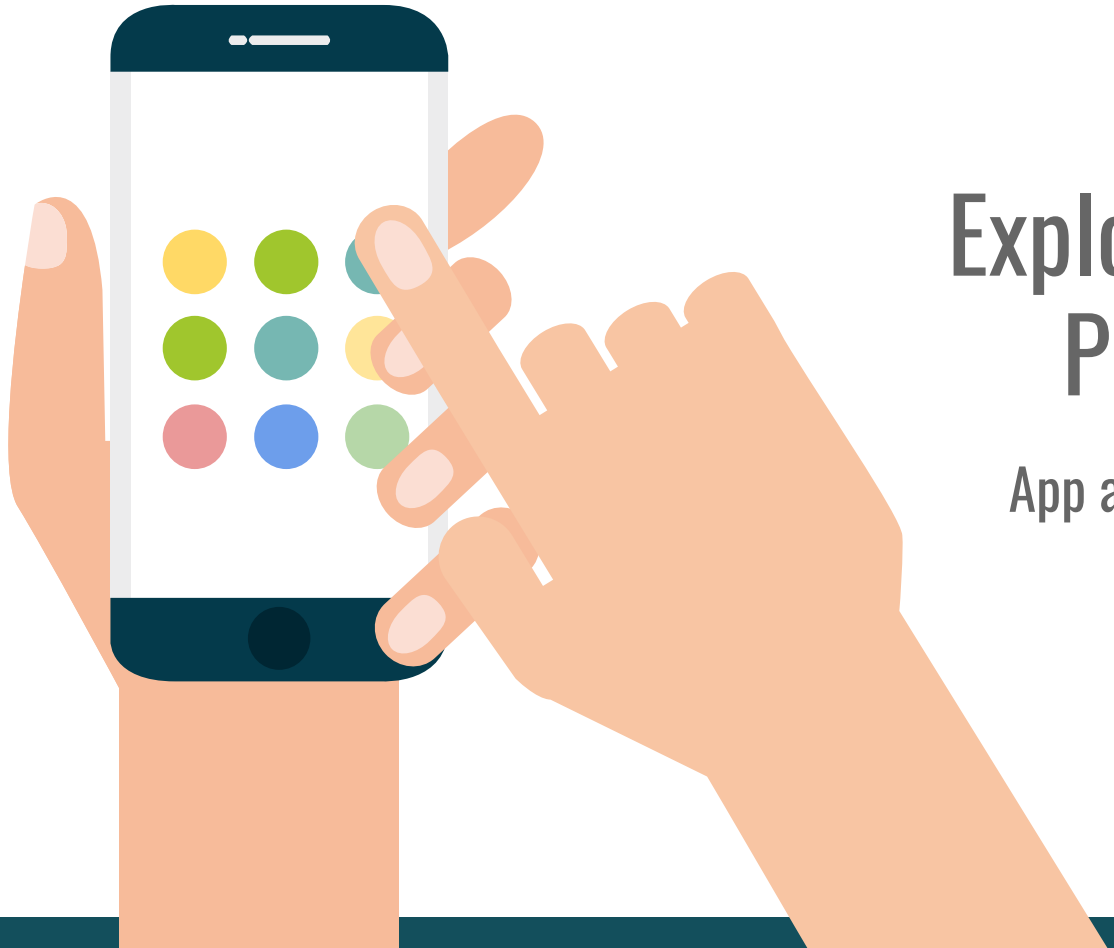
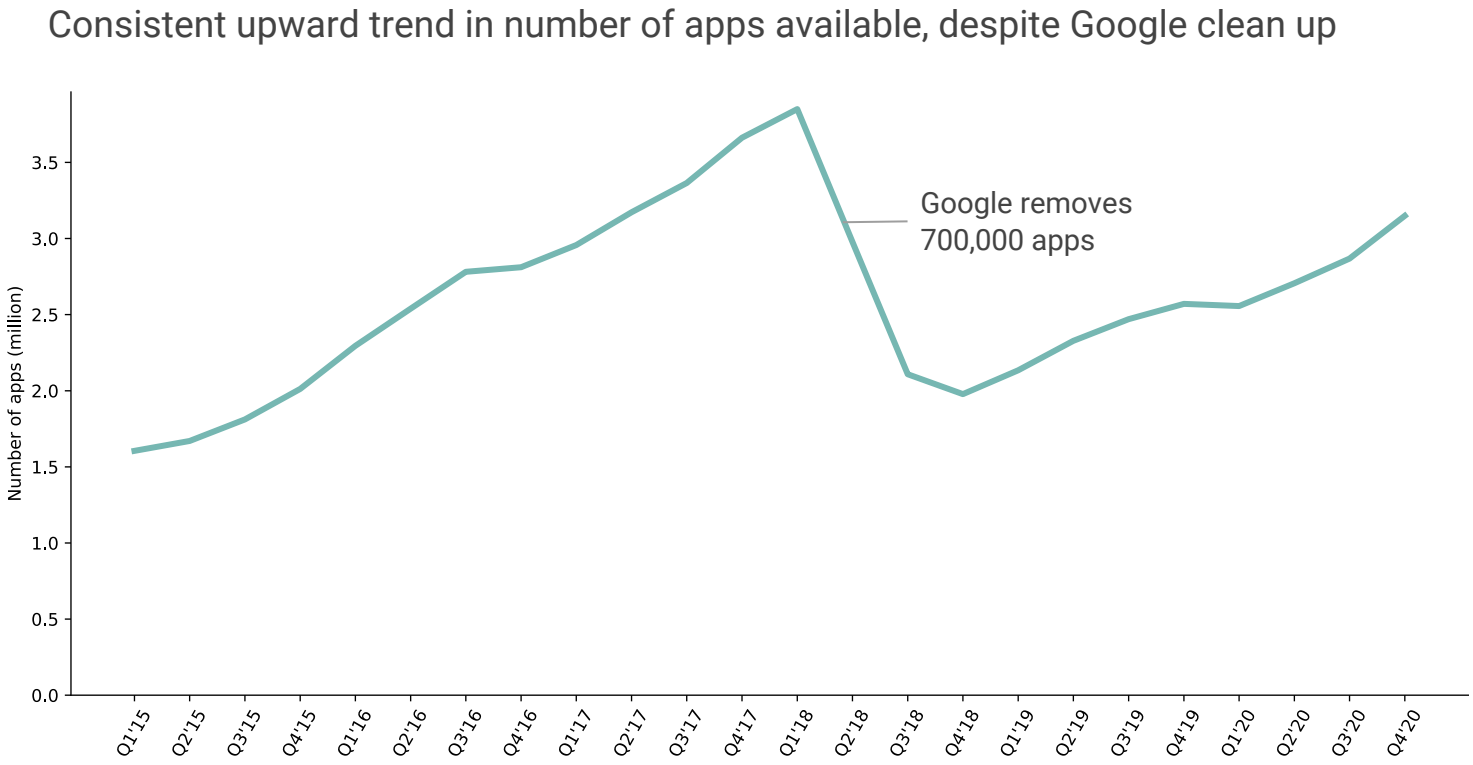


Exploring the Google Play app store

App analysis and recommender

Tawney Lott
22 March 2021

With over 3 million apps, the Google Play store is a rich resource for app developers and users



The analysis drew on quantitative and qualitative methods and utilized a range of tools

Methods and process

1. Desktop review of Google Play Store news and similar studies
2. Exploratory data analysis
3. Topic modeling and sentiment analysis
4. Linear regression to predict app rating
5. Recommender system for app users

Analysis tools

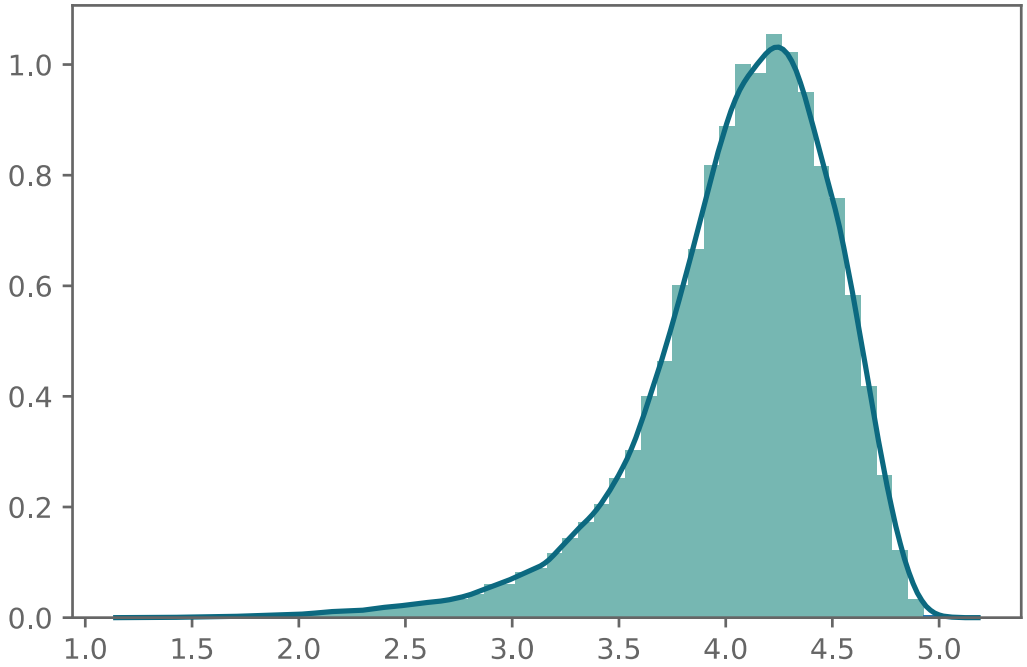
- Python
- Pandas
- Numpy
- Scikit Learn
- Natural Language Toolkit

Visualization tools

- Matplotlib
- Seaborn
- Tableau
- Streamlit

The sample included data for 22,000 apps which were scraped from the Google Play store

There is a left skew in the distribution of app ratings, which is the target for the predictive model



Features

- App name
- Developer
- Description
- Release date
- Category
- Free (yes / no)
- Price
- Number of ratings
- Number of reviews
- User reviews
- Content rating
- Date of last update

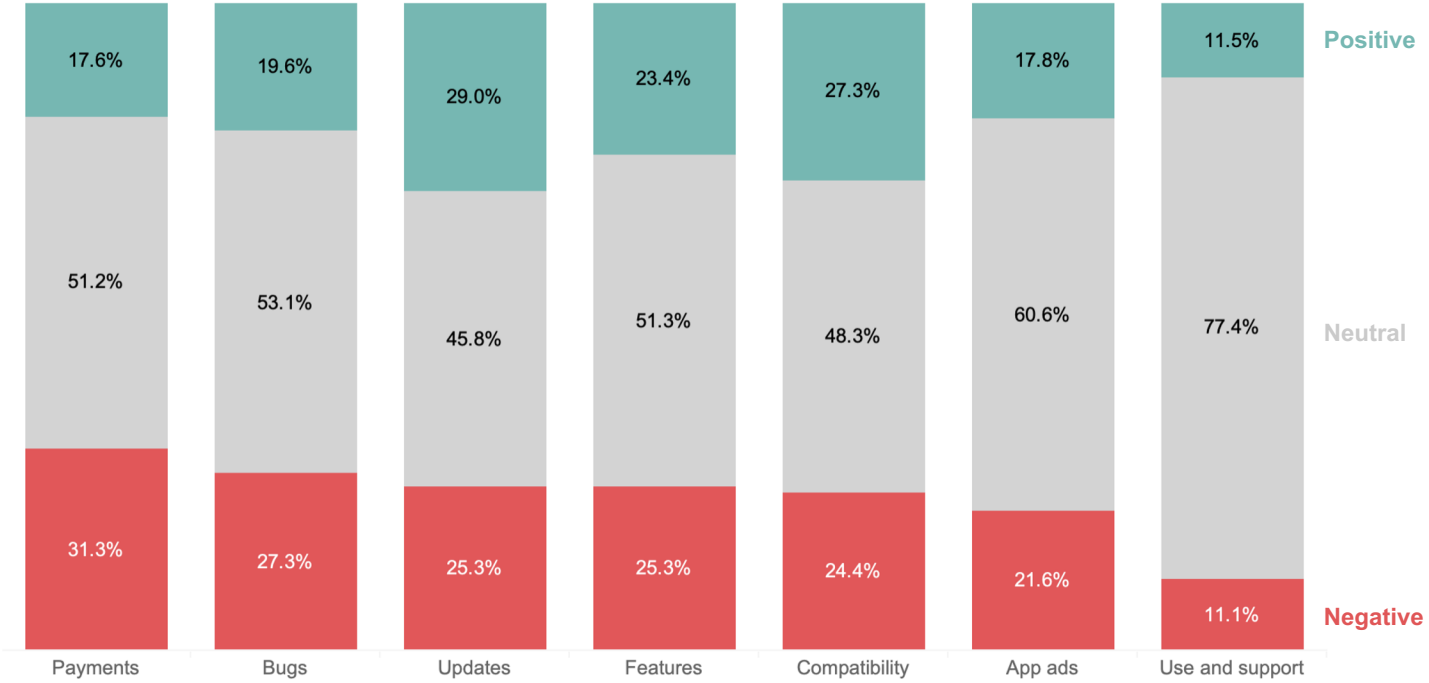
Topic modeling was used to understand the underlying topics of user reviews

The analysis revealed **seven core topics** and the top words within each topic

Features	Option	Change	Start	Find	Feature
App bugs	Fix	Big	Issue	Crash	Load
Payments	Money	Pay	Real	Card	Purchase
In-app ads	Ad	Watch	Remove	Click	Force
Device compatibility	Work	Stop	Device	Reinstall	Button
App updates	Update	Last	Latest	Year	Before
Useability and support	Help	Learn	Find	Account	Support

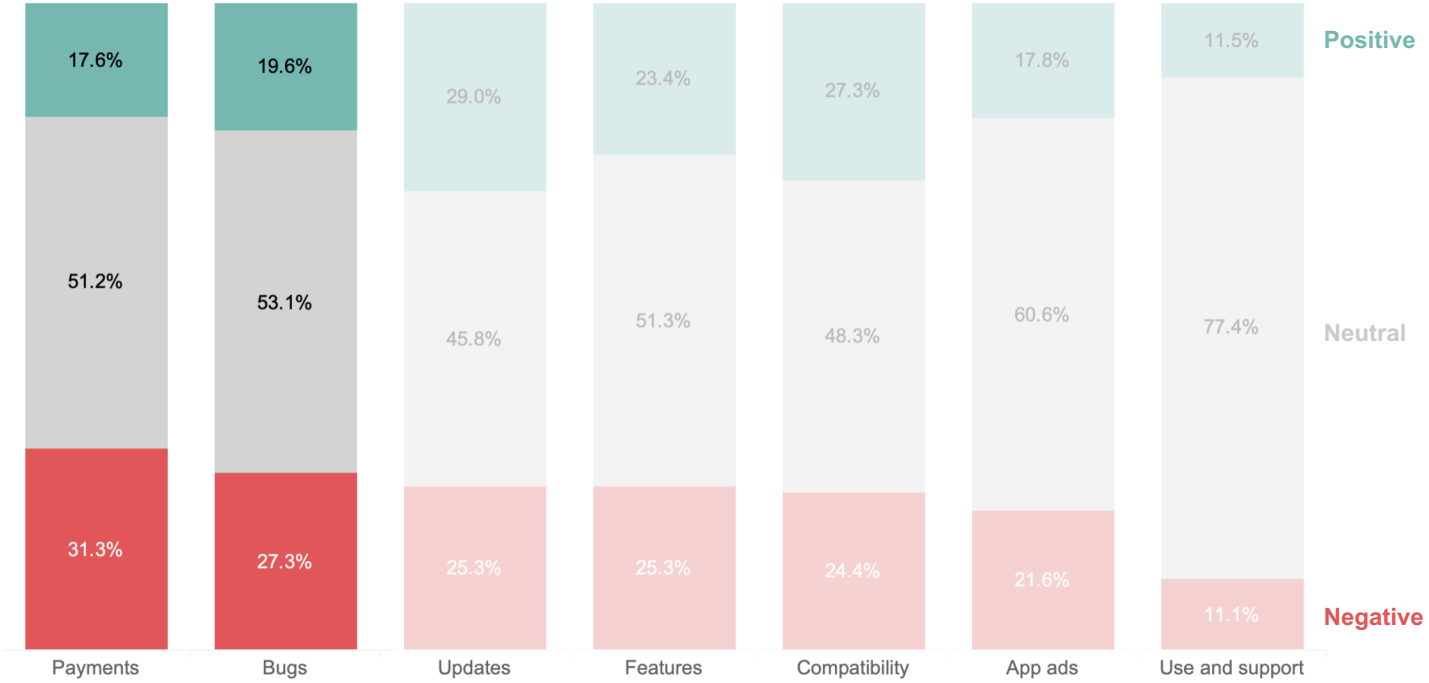
Sentiment analysis of user reviews was used to provide an additional layer of insight across the seven topics

While the majority of reviews appear to be neutral, the share of negative reviews exceed positive reviews across most of the topics



Sentiment analysis of user reviews was used to provide an additional layer of insight across the seven topics

The Payments topic has a larger share of negative reviews than Bugs, suggesting an opportunity to improve Payments experience



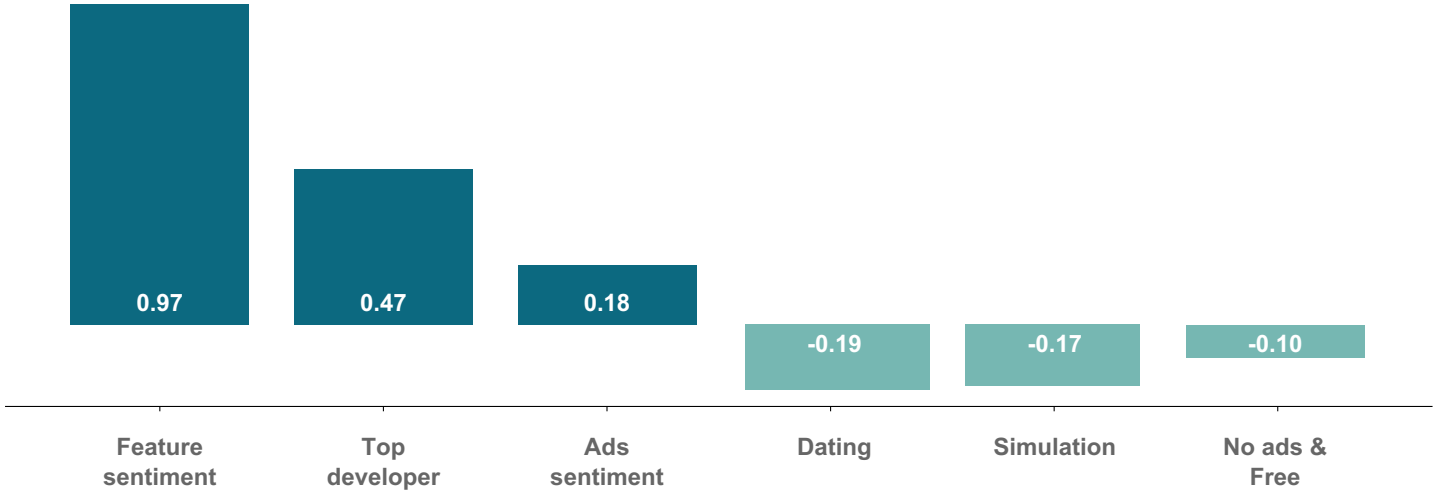
These layers of analysis were combined with the data scraped online to predict average user ratings for each app

Results

- The model explains 31% of the variance in user ratings
- MAE: 0.28
- Best performance on apps with a score in the range of 3.8 to 4.5 average user rating
- Performed less well on lower-rated apps

The model identified the strongest features in predicting app ratings

Positive regression coefficients are associated with higher user ratings, while negative coefficients are associated with lower ratings



Findings from this analysis provide important insights for app developers and a fun tool for users

Opportunities

- User sentiment regarding **Features is strongest predictor** of ratings - give the user what they want

Findings from this analysis provide important insights for app developers and a fun tool for users

Opportunities

- User sentiment regarding **Features is strongest predictor** of ratings - give the user what they want
- Payments results suggests there is an **opportunity to improve payments experience** and process for apps

Findings from this analysis provide important insights for app developers and a fun tool for users

Opportunities

- User sentiment regarding **Features is strongest predictor** of ratings - give the user what they want
- Payments results suggests there is an **opportunity to improve payments experience** and process for apps
- Apps by **Top Developers tend to have better ratings** - important to establish and maintain brand

Findings from this analysis provide important insights for app developers and a fun tool for users

Opportunities

- User sentiment regarding **Features is strongest predictor** of ratings - give the user what they want
- Payments results suggests there is an **opportunity to improve payments experience** and process for apps
- Apps by **Top Developers tend to have better ratings** - important to establish and maintain brand
- Apps that are **free and do not contain ads tend to have lower scores** - serious developers should investigate ads and / or payment models