BUS 256 - MARKETING ANALYTICS

Google Play Store App

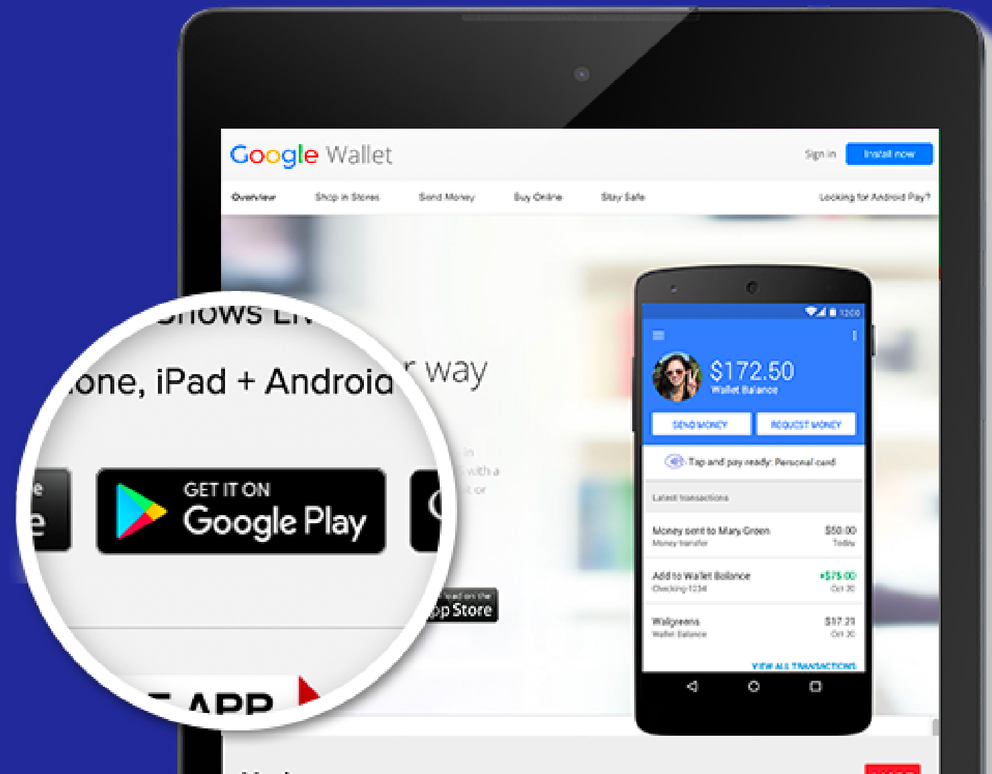
Jasmine, Tina, Noah

In this Presentation

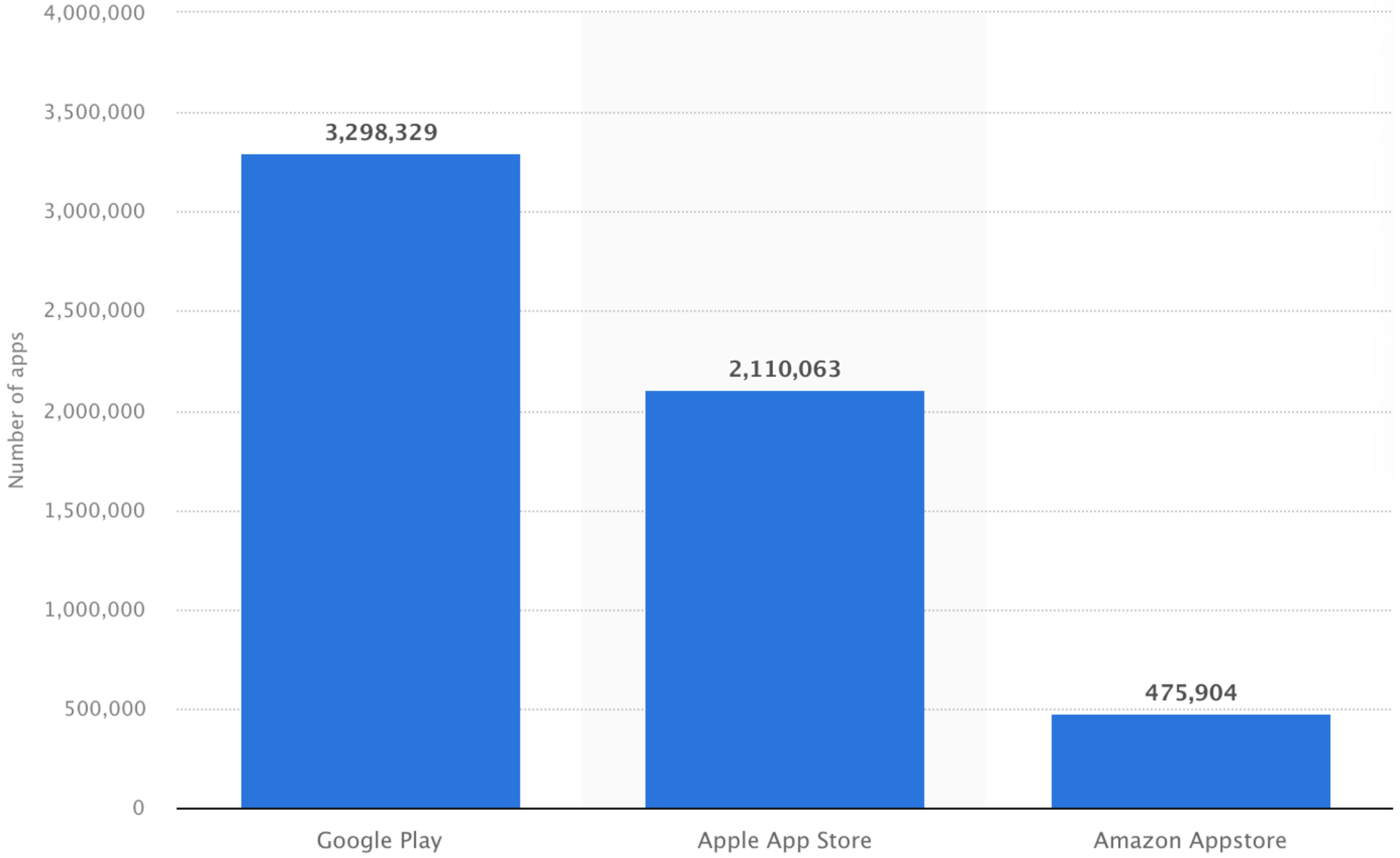
Overview

Google Play Store

- Official app store for Android devices
- Digital media store (music, books, movies, and TV shows, etc.)
- First launched in 2008 as Android Market



Google Play Store has the biggest number of apps (Q1, 2022)



Source: Statista

Project Overview

- Goal: **To identify the features attributable to an app's success.**
- Dataframes:
  Google Play Store Apps
  Google Play Store User Reviews

Hypotheses

- The category, genres, content rating, size, type, and price of an app all have an influence on its success or popularity.
- Users are more likely to leave reviews for apps that they have negative experience with.

Metrics

App (Dimension)

- App ID (Pk)
- App name
- Category
- Genres
- Content rating
- Size
- Type
- Price

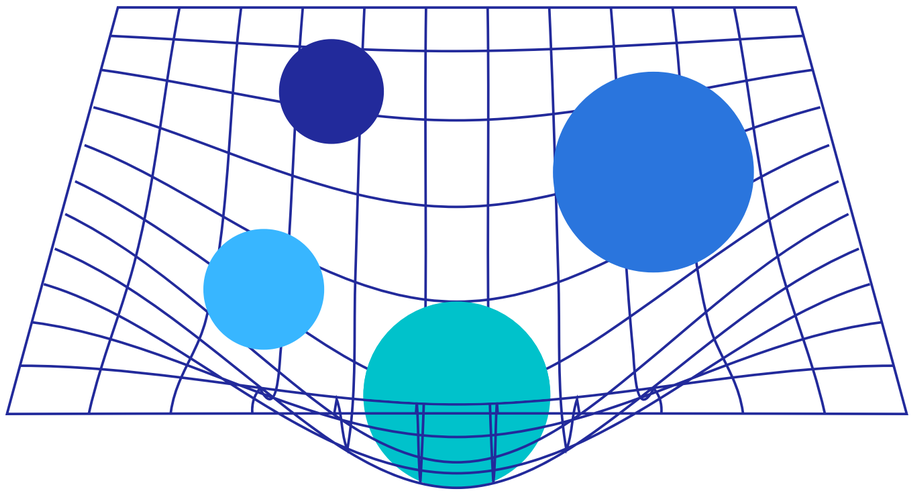
App success (Fact)

- Installs
- Ratings
- Reviews
- App ID
- Review ID

Review (Dimension)

- Review ID (Pk)
- Review sentiment
- Sentiment polarity
- Sentiment subjectivity

Data Cleansing & Standardization



Step 1 Metadata Explanation

Understand each feature

Step 2 Standardize Field Names

Rename column names and check index

Step 3 Keep Relevant Data Fields

Drop NaN values and duplicates

Step 4 Correct Feature Type

Clean the column data and Convert each column to the wanted data type

Step 5 Adjust / Add Necessary Columns

For simpler and quicker analysis purposes

Comparison before and after data cleansing

Before

```
df_1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   App             10841 non-null  object
 1   Category        10841 non-null  object
 2   Rating          9367 non-null   float64
 3   Reviews         10841 non-null  object
 4   Size            10841 non-null  object
 5   Installs        10841 non-null  object
 6   Type            10840 non-null  object
 7   Price           10841 non-null  object
 8   Content_rating  10840 non-null  object
 9   Genres          10841 non-null  object
 10  Last_updated    10841 non-null  object
dtypes: float64(1), object(10)
memory usage: 931.8+ KB
```

After

```
df_1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8892 entries, 0 to 8891
Data columns (total 12 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   App                      8892 non-null   object
 1   Category                 8892 non-null   object
 2   Rating                   8892 non-null   float64
 3   Reviews                  8892 non-null   int32
 4   Size                     8892 non-null   int32
 5   Installs                 8892 non-null   int32
 6   Type                     8892 non-null   int32
 7   Price                    8892 non-null   float64
 8   Content_rating           8892 non-null   object
 9   Genres                   8892 non-null   int32
 10  Last_updated             8892 non-null   datetime64[ns]
 11  Days_after_last_updated  8892 non-null   int64
dtypes: datetime64[ns](1), float64(2), int32(5), int64(1), object(3)
memory usage: 660.1+ KB
```

Data Description

Valid Data - Play Store App Features

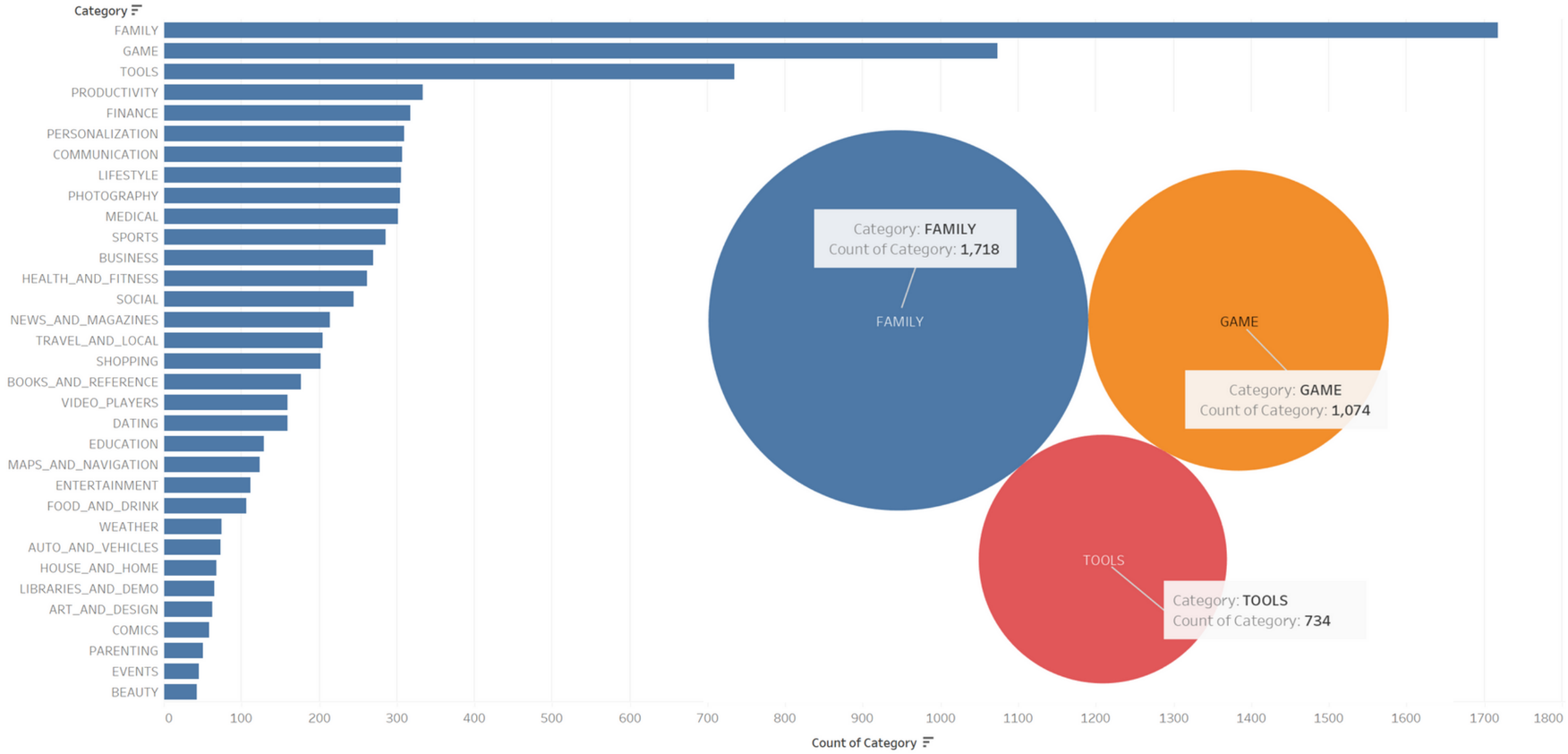
- 8,892 entries
- Total 12 columns

Valid Data - App Actual Reviews

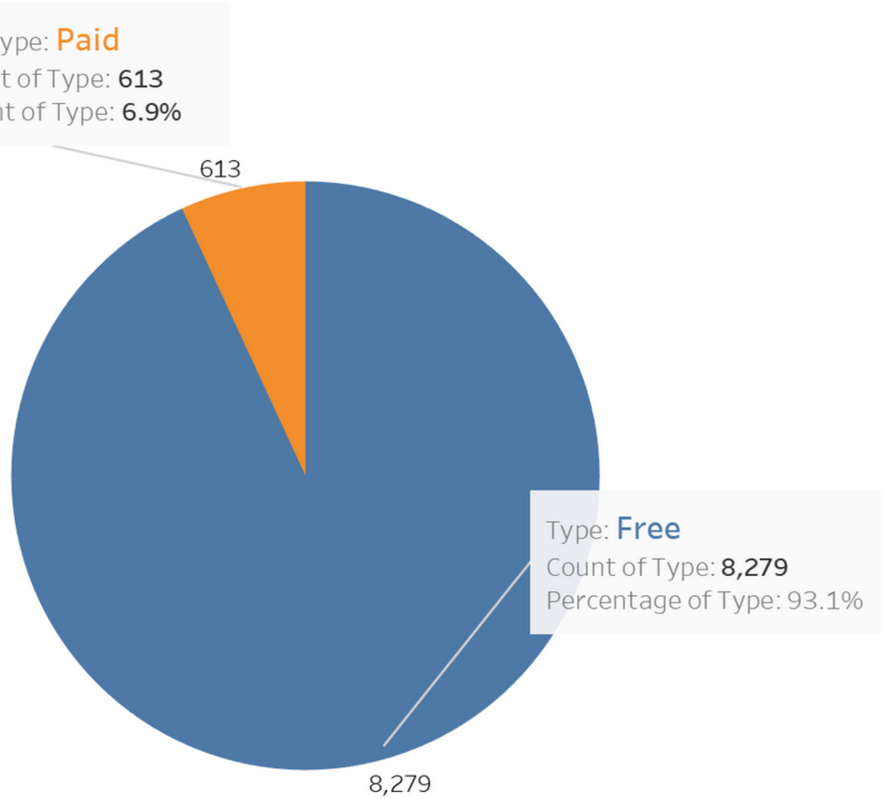
- 40,376 entries of Reviews
- 816 unique apps

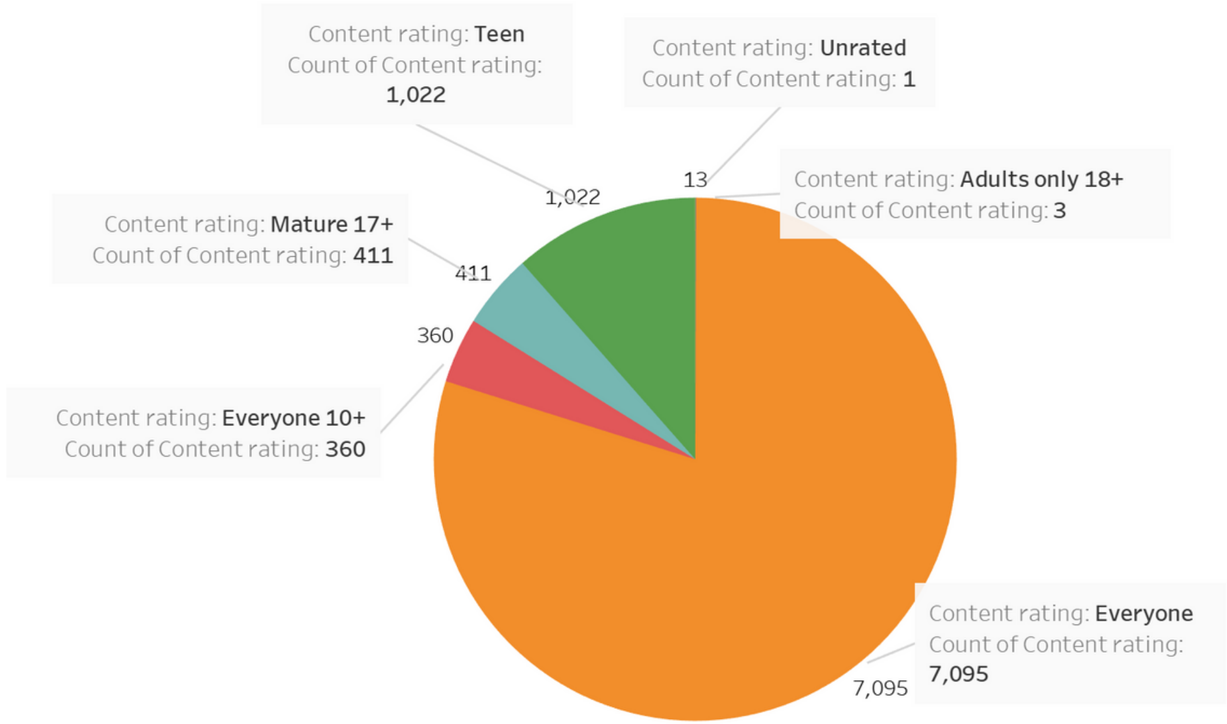
Unique Values

- App: 8,196
- Category: 33
- Type:
  - Free
  - Paid
- Content_rating:
  - 'Everyone'
  - 'Teen'
  - 'Everyone 10+'
  - 'Mature 17+'
  - 'Adults only 18+'
  - 'Unrated'

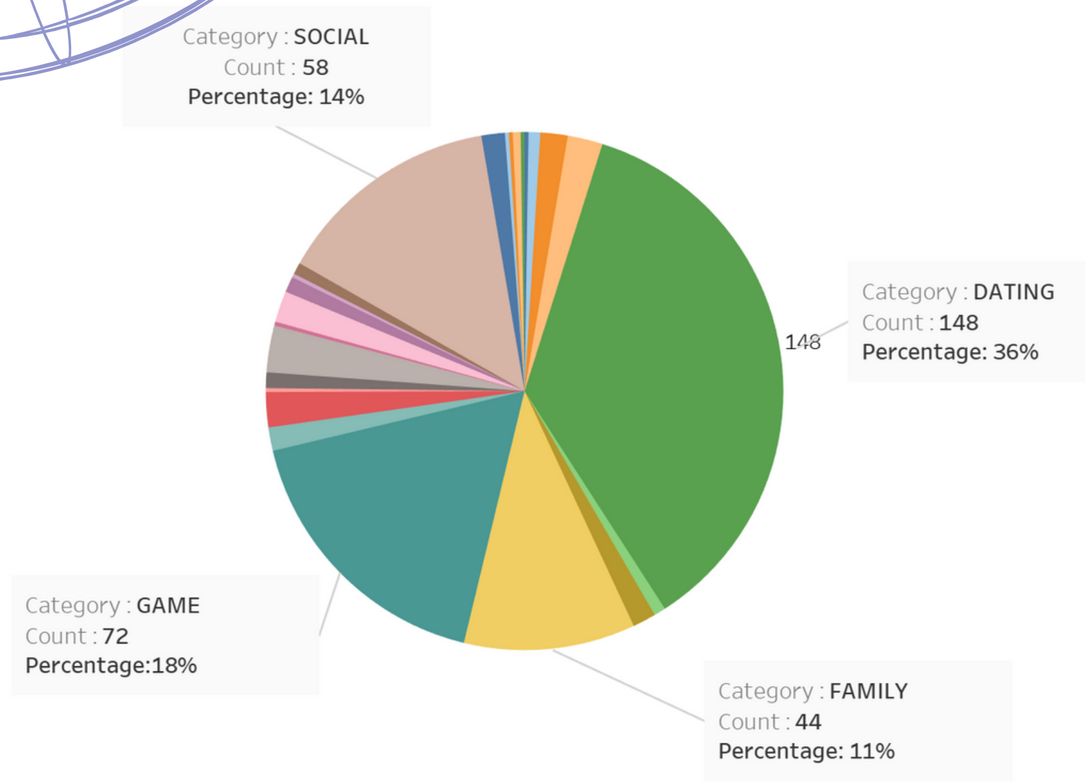


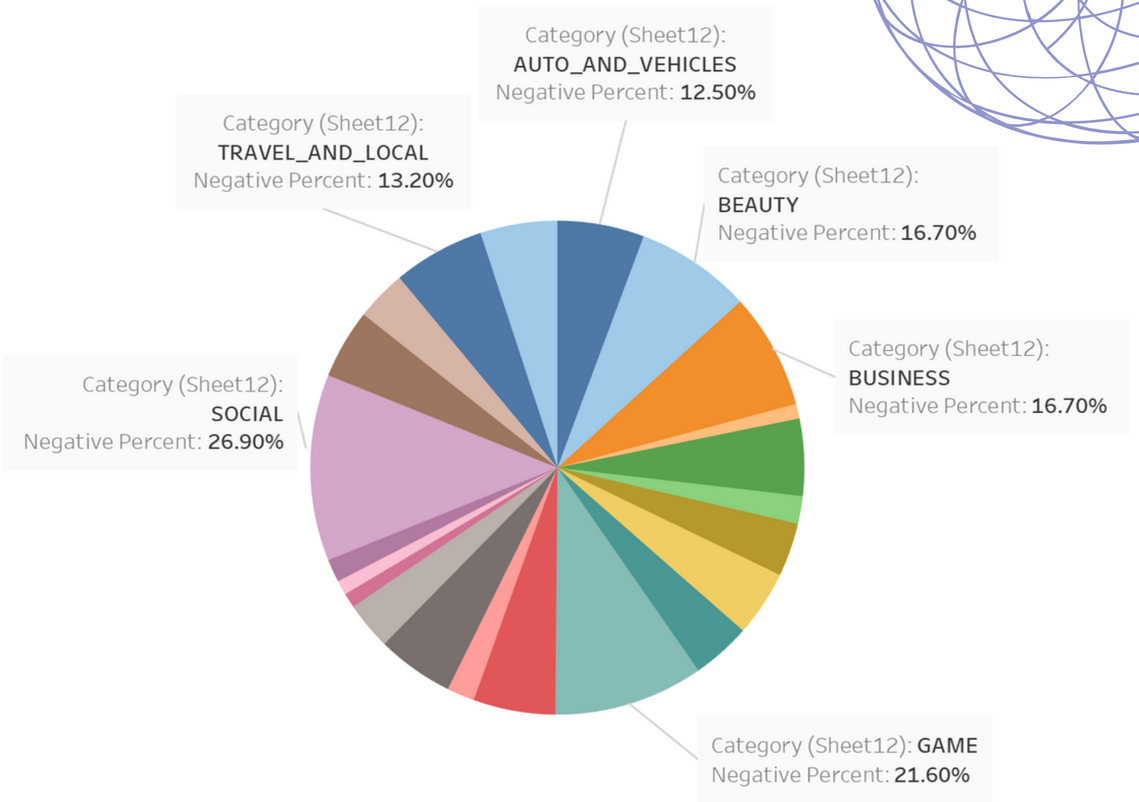
Types and Content Ratings





Category Marks in Different Fields



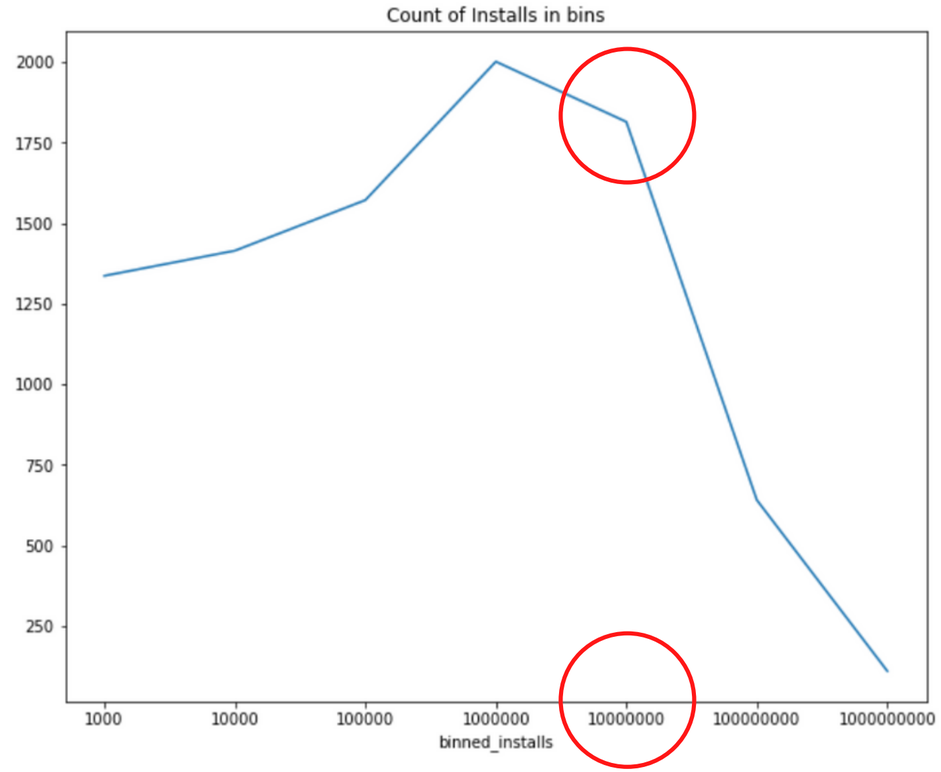


Categories rated Mature 17+

% of apps in different categories that are reviewed negatively

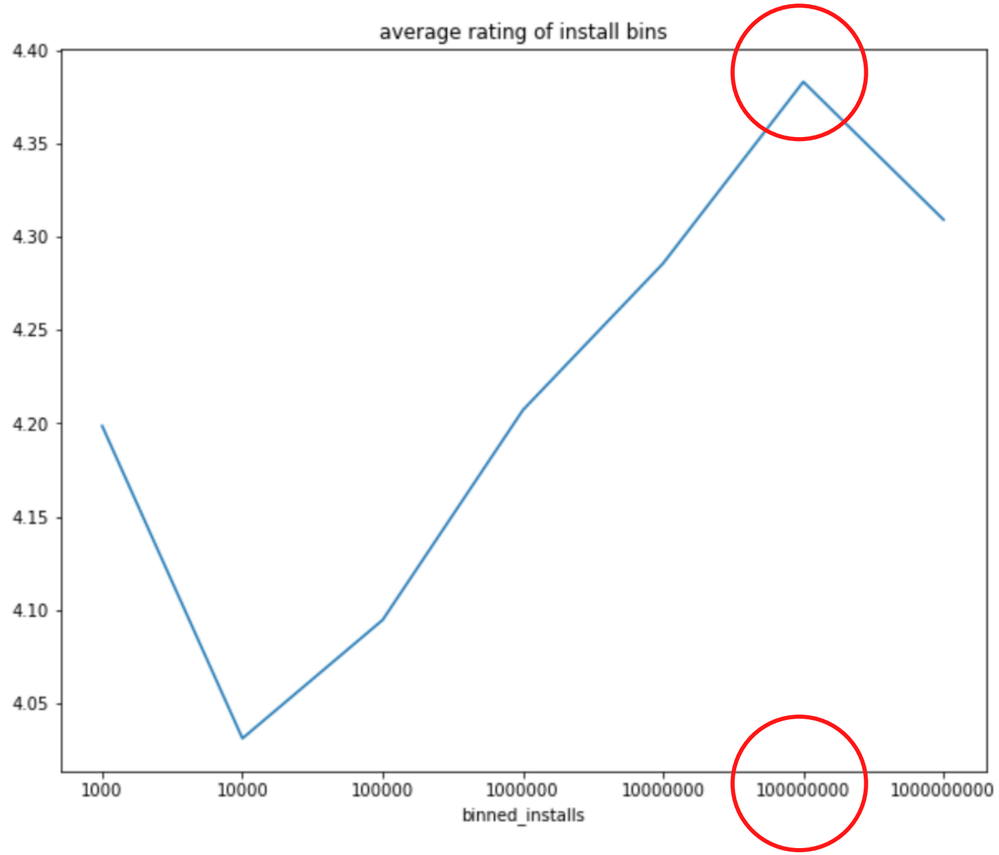
Bins: Installs - Count & Average Rating

<AxesSubplot:title={'center':'Count of Installs in bins'}, xlabel='binned_installs'>

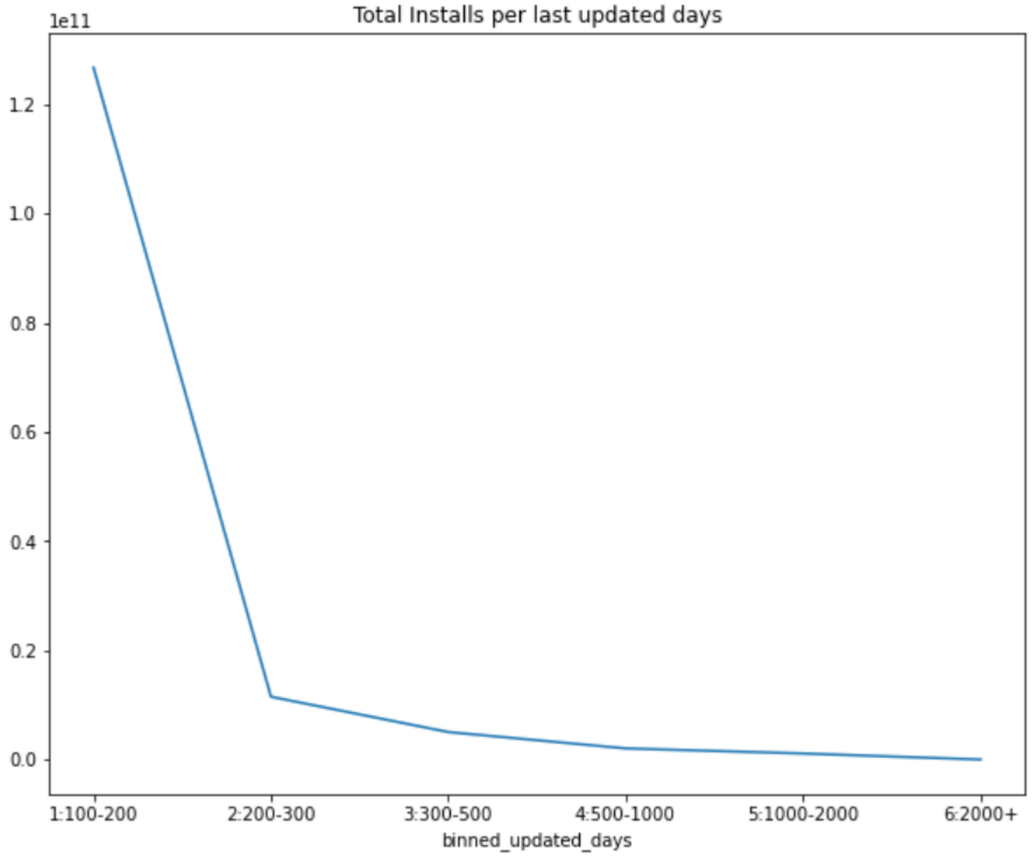


Highest Point: 1,000,000
Turning Point: 10,000,000

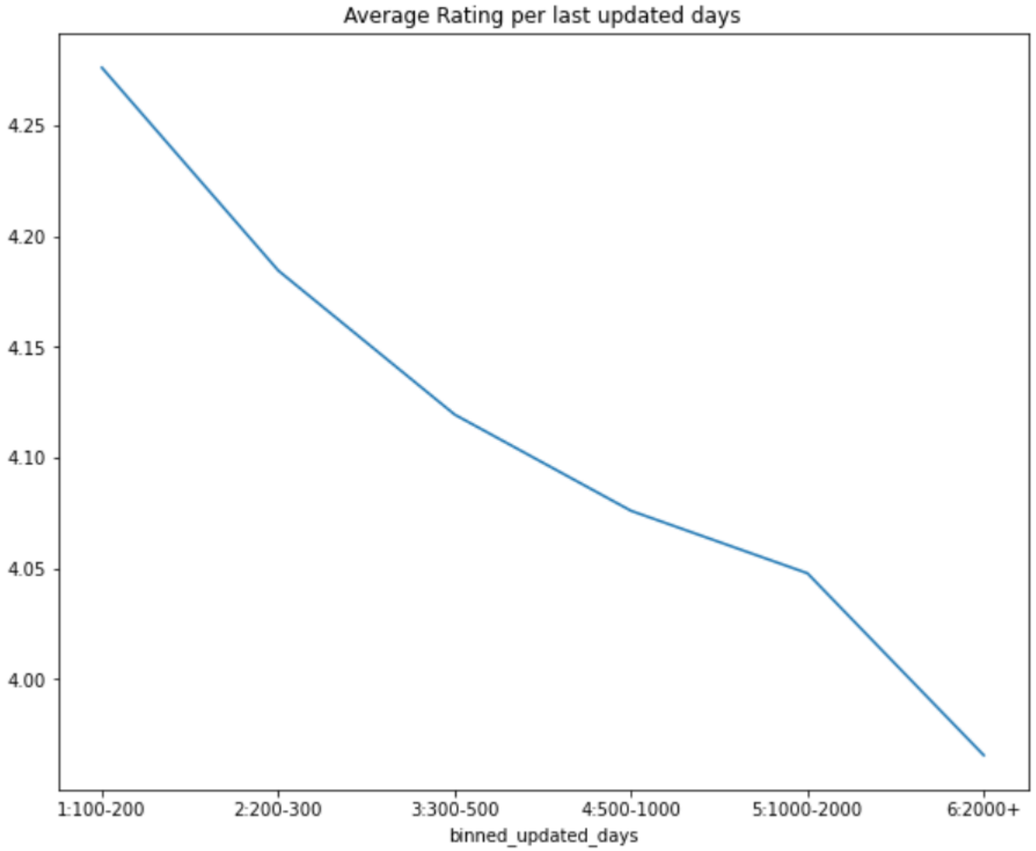
<AxesSubplot:title={'center':'average rating of install bins'}, xlabel='binned_installs'>



Highest Point: 10,000,000
Increasing trend between 10,000-100,000,000

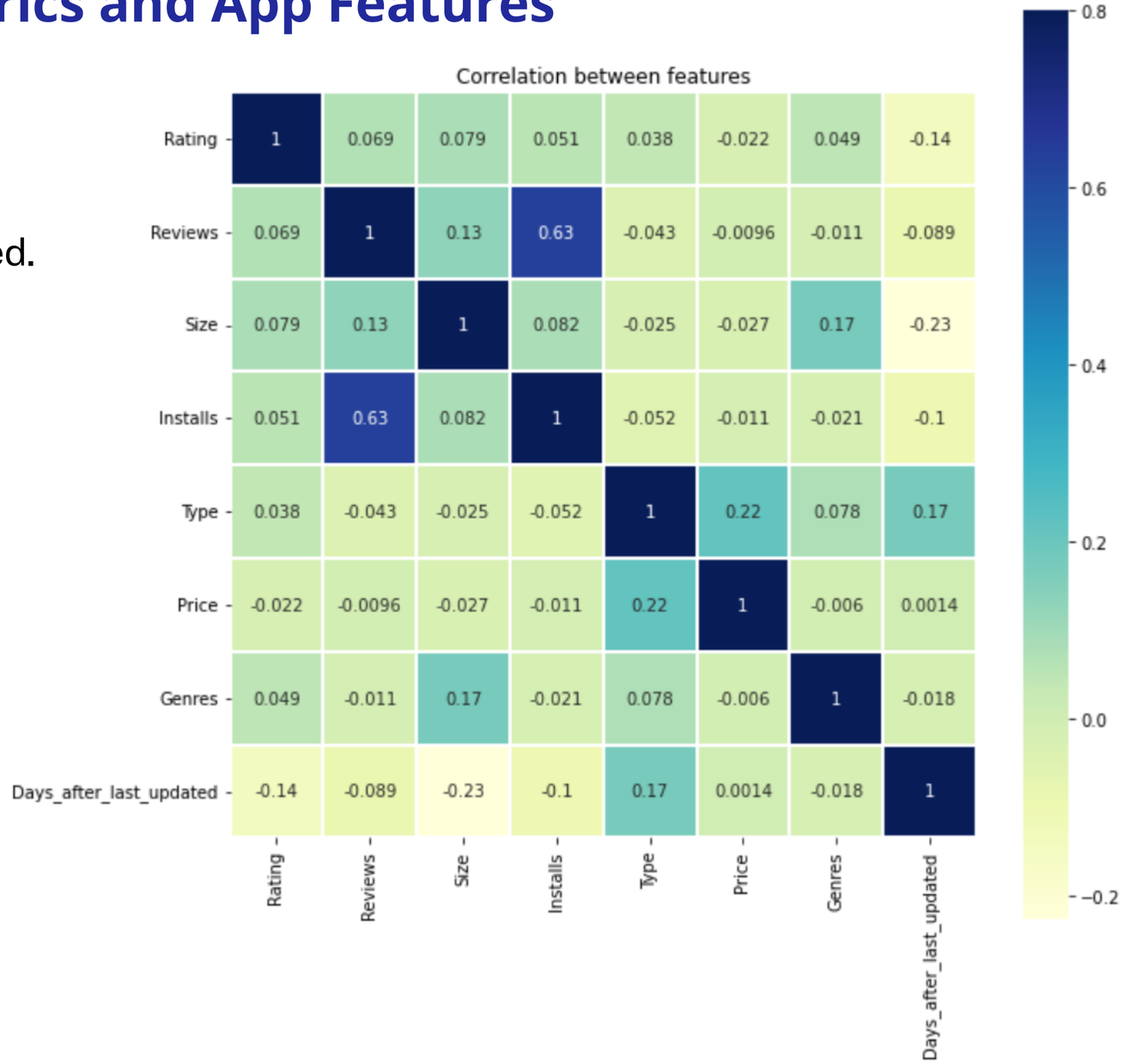


Recently updated apps have more installs and higher ratings.



Correlation among Success Metrics and App Features

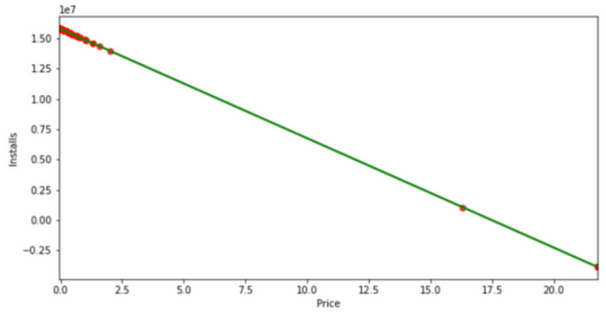
Installs and reviews are moderately correlated.



Simple Linear Regression

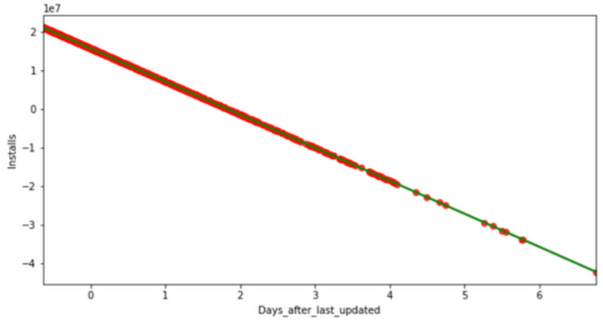
- **Price & Installs**

  As price increases, installs decrease.



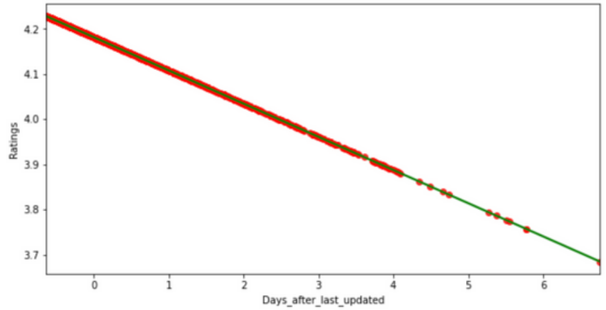
- **Days after last updated & Installs**

  More recently updated apps show a higher installs rate.



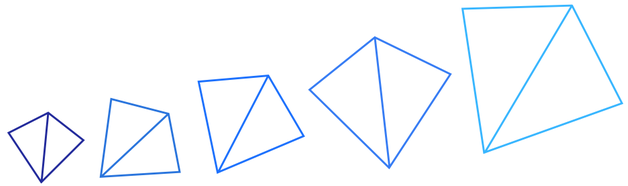
- **Days after last updated & Rating**

  The more recent the update, the higher the rating.



Multiple Linear Regression - Rating

	Coefficients	
Size	0.028	Bigger size, higher rating
Price	-0.011	Larger price, lower rating
Content rating	-0.0095	Larger availablity, lower rating
Days after last updated	-0.068	Recent update, higher rating



Multiple Linear Regressions

- **Installs** with **Rating**, **Size**, **Reviews**, **Price**, **Content Rating**, and **Days after Last Updated**

R2: -0.0296
(+) Rating and Reviews
(-) Size, Price, Content Rating, Days after Last Updated

- **Review Number** with **Sentiment Polarity** and **Subjectivity**

R2: -15.86
(+) Subjectivity
(-) Sentiment Polarity

Conclusion & Recommendations

01

Categories with more content violations and more negative reviews:
- 'Social'
- 'Game'

02

Frequent updates

-> Constantly maintain the app, fix bugs, and add new features

03

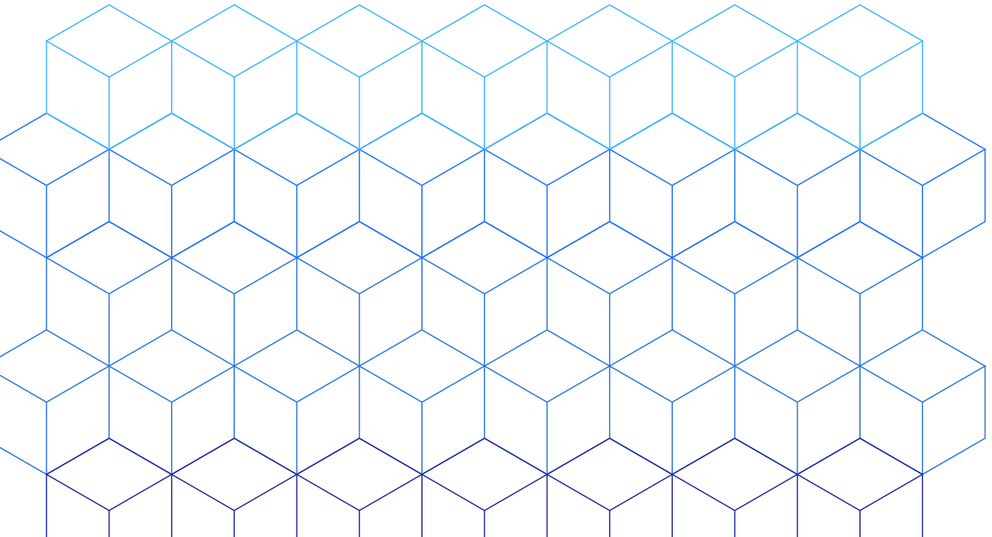
Size, price, and content rating

-> For niche apps with strong value propositions only.

04

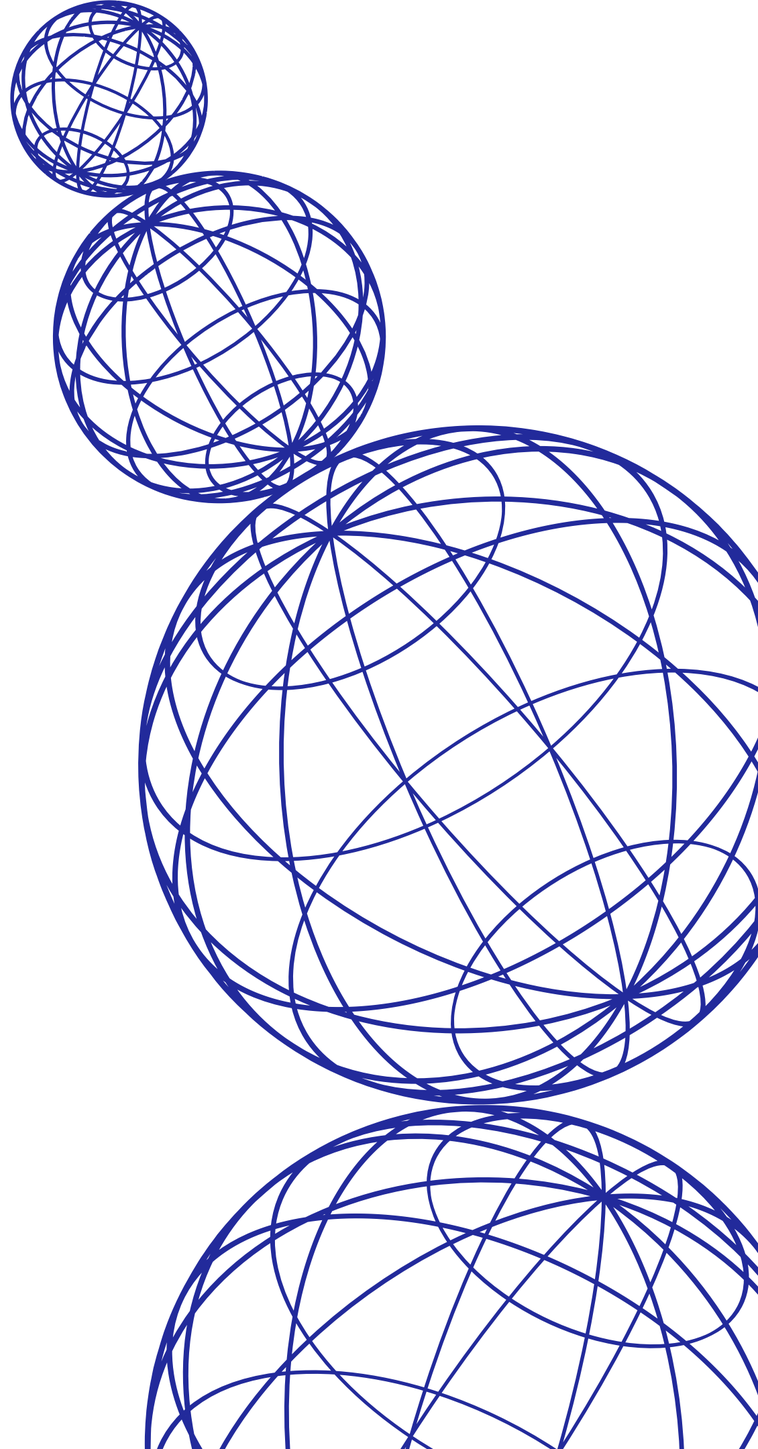
Most reviews are negative

-> Be responsive to negative reviews and try to address users' concerns as soon as possible



Thank you for listening

Questions?



References

“Biggest App Stores in the World 2022.” Statista, April 27, 2022.
“What Is Google Play Store? Everything You Need to Know.” Lucic, Kristijan. Android Headlines, April 3, 2020.
"Google Play Store Apps" and "Google Play Store User Reviews." Lavanya. Kaggle, 2019.