BUS 256: Marketing Analytics
Professor Ashish Mrig
Team: Keyi Jiang, Noah Nguyen, Tina Li

Team Project
Topic: **Google Play Store Apps' Performances**
Project Summary

Problem: We want to find out the features that contribute to the success of an app on Google Play Store. These features could be a guideline for people to design a successful app.

Hypotheses:

- The category, genres, content rating, size, type, and price of an app on the Google Play Store all have an influence on its success or popularity.
- Some apps can have very good ratings and reviews but few installs due to the above factors being not accessible to everyone.
- Users are more likely to leave reviews for apps that they have negative experience with.

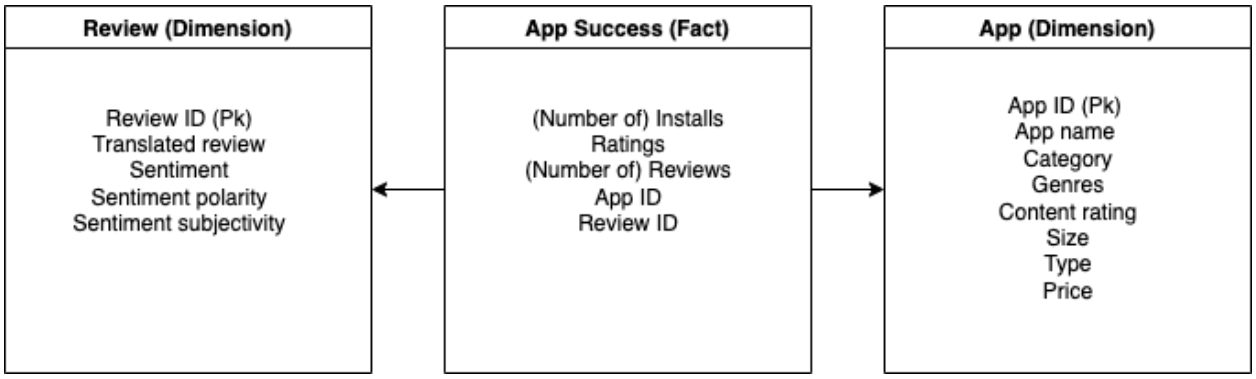
Dataframe:

Google Play Store Apps (https://www.kaggle.com/lava18/google-play-store-apps)
Google Play Store User Reviews (https://www.kaggle.com/lava18/google-play-store-apps?select=googleplaystore_user_reviews.csv)

Fact (Metrics): Number of installs, Ratings, Reviews

We think there are three factors that can measure the success of an app. The number of installs directly shows the popularity of an app. Second, an app of higher ratings means that it's praised by its users, then more potential viewers may install the app. Third, reviews could be controversial, we think an outstanding app should have many but positive reviews.

Dimensional Model:



Approach:

We are going to measure the app success by each factor such as category, genre, content rating, size, type or pricing structure to see if there is a difference between different categories, genres, content ratings, etc. In addition, we will try to find out any differences between the trends of installs, ratings, and reviews. Then, we will create charts that best showcase the trend. For example, an app may have very high ratings but few downloads. This may also be explained by plotting those measurements against the factors since, for instance, an 18+ content rating could restrict the number of people that can access a very highly-rated app.

Moreover, from the second data frame, we can approach the third hypothesis by measuring the number of installs of the app with the positive and negative reviews. An app may have many negative reviews but a high number of installs because people tend to leave negative reviews. We believe sometimes a really popular app may have more negative reviews and a relatively middle-level rating because of the central limit theorem. We hope to approach these problems with our data.

Some possible cleanse & explore:

Add app_id column for efficiency.

Separate Gernes column and combine it with the category column as the main category and affiliated category.

Clean for missing data.

Categorize size by small, medium and large.