

# Google Play Store Data Analysis

Project Proposal – Shankar Mohanathas

## About this Dataset

While many public datasets (on Kaggle and the like) provide Apple App Store data, there are not many counterpart datasets available for Google Play Store apps anywhere on the web. On digging deeper, I found out that iTunes App Store page deploys a nicely indexed appendix-like structure to allow for simple and easy web scraping. On the other hand, Google Play Store uses sophisticated modern-day techniques (like dynamic page load) using JQuery making scraping more challenging.

## Project Description

To analyze and construct understanding from given dataset through data cleaning, with the use of visualizations, maps, graphs and trends.

## Source of Data

The source of raw data was found in <https://www.kaggle.com/datasets/lava18/google-play-store-apps>, which has crucial information in categories such as, category of data, price, reviews, size, version, last updated, and content rating. The original CSV file contained over 10 000 data entries.

## Research Questions

1. What is the distribution of app categories in the Google Play Store?
2. What is the percentage of free vs paid apps in the Google Play Store?
3. What is the distribution of app ratings in the Google Play Store?
4. How do the number of reviews correlate with app ratings?
5. What are the most expensive app categories on the Google Play Store?
6. Which app categories have the highest number of installs?
7. How does app size affect the number of installs?

