# Google Play Store App Popularity Analysis & Prediction

## ECE 143 (FA25) Project Proposal - Team 5

## Problem

Analyzing the top factors that contribute to an app's popularity, and building a model to predict the popularity given these factors

## Dataset

Name: Google Play Store Apps (link)

This dataset contains 10841 entries of different apps on the play store, with 13 separate features like rating, size, genre etc. The 'output' variable for this project is the number of installs.

## Proposed Solution and Real world Application

There are more than 1.5 million apps on the Google Play Store, with more than 1,000 apps launched every day on average. Further, app popularity often tends to be a winner-take-all market, with a handful of apps dominating a majority of the user base. Understanding the features that drive the popularity of an app is critical for developers as well as marketers.

For the purposes of this project, we consider an app "popular" if it has over 1M (one million) installs. Thus, we can boil down our dataset to two classes - 'popular' and 'not popular'. (please note that this is tentative, and we may choose to increase the number of classes based on analysis). Once we have cleaned up the dataset, we plan to use various techniques to identify what are the most important features that make an app

popular. Common techniques include visual exploration, correlation studies, t-tests, and random forests. We shall apply as many techniques as deemed necessary. Once the key features have been identified, we shall train and then test (using an 80-20 split of the dataset) a logistic regression (or some other similar ML technique - to be finalized after analysis) ML model to be able to predict, given those features, what is the chance that an app will be popular.

## Project steps

| **S.No.** | **Step** | **Estimated completion time** | **Person(s) in charge** |
|---|---|---|---|
| 1. | Data extraction & clean-up<br><br>*preprocessing, conversion to numeric categories, dropping incomplete entries etc.* | 1 week | Aditi Pagey, Ke Liu |
| 2. | Data Analysis & Visualization<br><br>*identify the top features that affect popularity with various data analysis techniques* | 2 weeks | Shaurya Chopra, Abdumannon Yovkochov |
| 3. | Prediction Model<br><br>*based on the top features, train a logistic-regression (or other) based model for predicting whether an app will be popular or not* | 1 week | Zijun Zhang, Ke Liu |