GET IT ON

Google Play

Chaquayla Halmon

Overview

What will we be going through?

Business Problem

Data

Process

Final Model

Recommendations

Future Research

	Free	Paid	Grossing
1	911 Emergency Dispatcher SUPERSONIC STUDIOS LTD	Minecraft Mojang	Roblox Roblox Corporation
2	YouTube: Watch, Listen, Str... Google LLC	Bloons TD 6 Ninja Kiwi	YouTube: Watch, Listen, Str... Google LLC
3	TikTok TikTok Pte. Ltd.	Heads Up! Warner Bros.	Tinder - Dating & Meet Pe... Tinder Inc.
4	ESPN Tournament Challen... ESPN	Procreate Pocket Savage Interactive Pty Ltd	Disney+ Disney
5	Finger On The App 2 Beast Interactive, LLC	Geometry Dash RobTop Games AB	Candy Crush Saga King
6	Instagram Instagram, Inc.	HotSchedules HotSchedules	Twitch: Live Game Streami... Twitch Interactive, Inc.
7	Facebook Facebook, Inc.	Monopoly Marmalade Game Studio	Pokémon GO Niantic, Inc.
8	Messenger Facebook, Inc.	Plague Inc. Ndemic Creations	BIGO LIVE-Live Stream, G... BIGO TECHNOLOGY PTE. LTD.
9	Snapchat Snap, Inc.	Grand Theft Auto: San And... Rockstar Games	PUBG MOBILE 3RD ANNIV... Tencent Mobile International Limit...
10	Gmail - Email by Google Google LLC	Papa's Mocharia To Go! Flipline Studios	HBO Max: Stream TV & Mo... WarnerMedia

Business Problem:

As an app developer, you need to find a way to make your apps more popular on Google App Store.

What features need more attention?

DATA

From:

Google dataset from Kaggle - over 10k data points

Web Scrape new data from Google Play Store - obtained over 2.8k data points

Total: 13680 datapoints

What?

Types of columns:

Numerical - Price, app size, ratings,etc

Descriptive - category, content rating, app name, etc

Final DataSet

After cleaning and getting rid of missing information the final dataset has 11,822 data points.

Process

Import/clean data

After combining the two datasets. Clean the missing data and drop any unneeded columns.

Baseline Model

Used four different models:
Linear regression - 32.5%
KNN regressor - 33.9%
Decision Trees - 31.6%
Random Forest - 43.8%

Fine Tuning

To get better results, got rid of outliers, log transform continuous variables, more cleaning, etc.

Re-Model

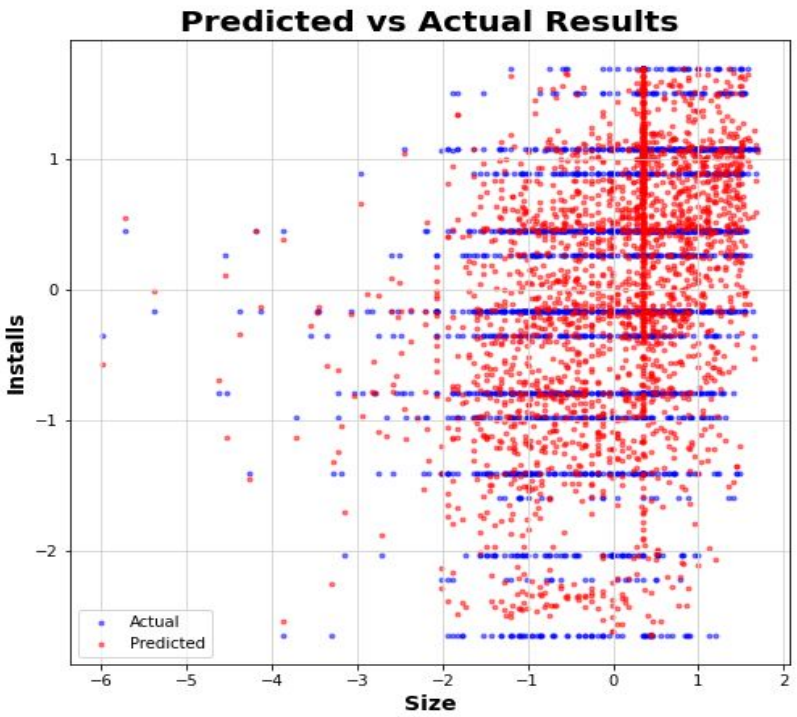
Linear regression - 54.5%
KNN regressor - 58.8%
Decision Trees - 48.6%
Random Forest - 60.8%

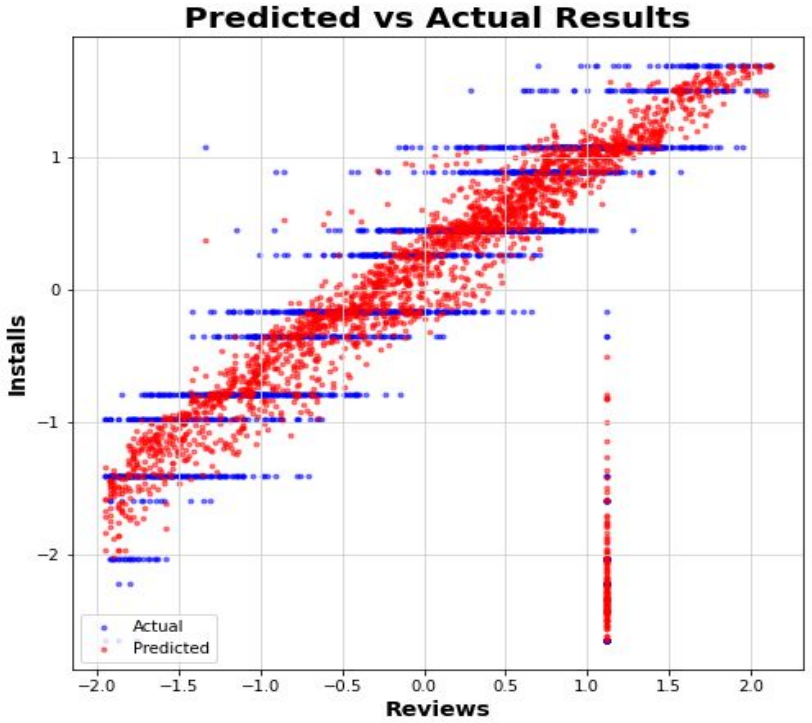
Final Model

Linear regression - 50.4%
KNN regressor - 21.4%
Decision Trees - 88.9%
Random Forest - 90.3%

Final Model

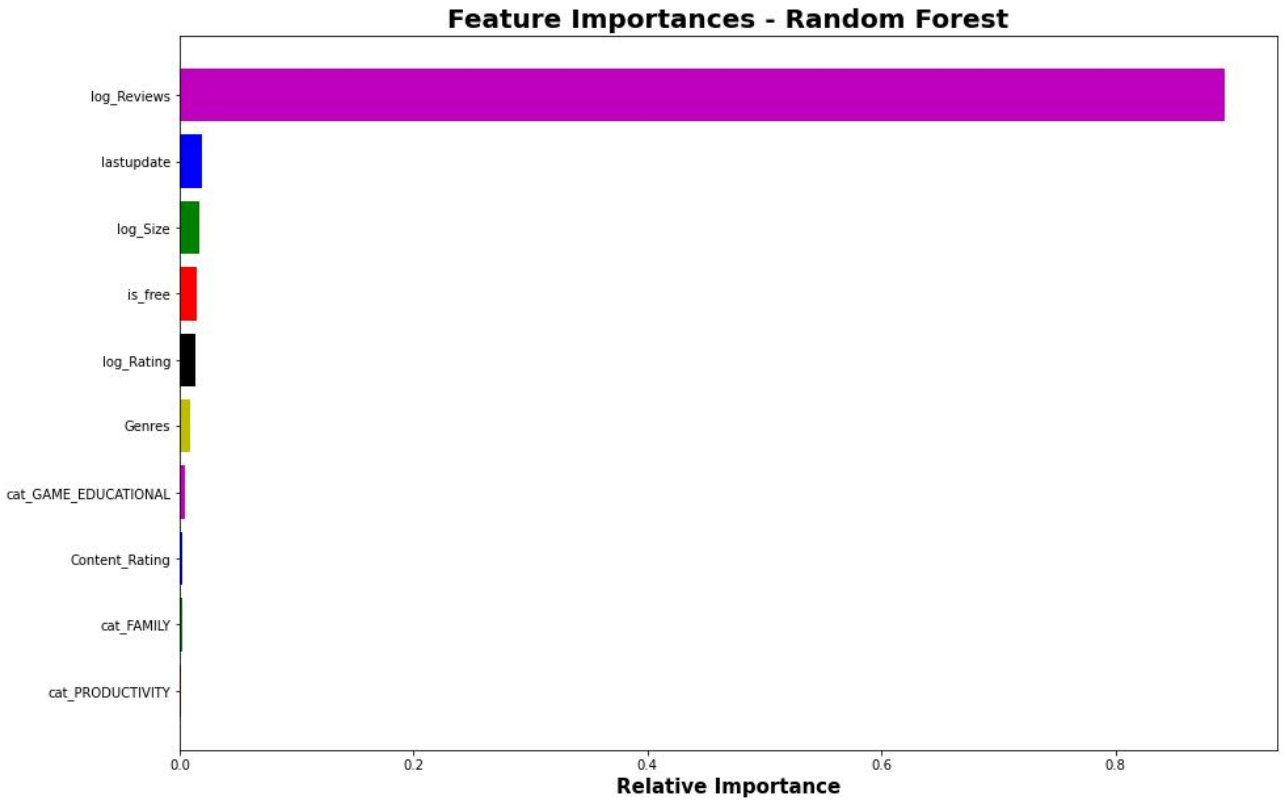
Predictions VS Actual





Final Model

Random Forest 90.3% Accuracy



Reviews

Updates

Size

Recommendations

Start the cycle

Market your App to get more installs that will in return get more reviews that can lead to more installs.

Out with the old

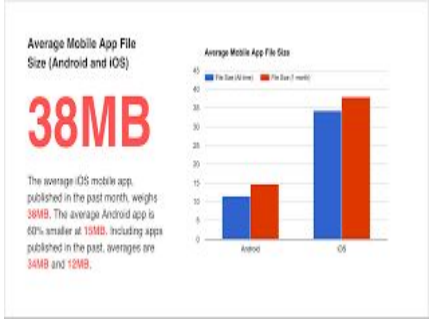
Focused on timely updates for your apps. Eliminate potential bugs and keep the consumers engaged with new content.

Trending

Communication, family, and gaming apps are the way to go, to get high installs, reviews, and rating.

Size

Focus on making apps no bigger than 200 megabytes.



Future Research

- Profits - research the difference in free, freemium, paid, subscriptions, etc.
- Time Series - look into the progress of apps over a period of time and document churn rate.
- Sentiment Analysis - look into what consumers like about certain categories and look for improvements

CONNECT WITH ME

Chaquayla Halmon

Github for other projects:
https://github.com/halmonchaquayla

Linkedin to network:
https://www.linkedin.com/in/chaquayla-halmon-605747201/

Medium to check out my blog:
https://halmonchaquayla.medium.com/