Exploration and Comparison of Applications in App Store and Google Play

In today tech-world, there are many applications are used in our daily life. Some are free, and some are paid. Some are designed for children; some are for teens and adults. Some are entertainment, and some are for educational purposes, etc. In this project, I want to have a general statistical summary and comparison of applications in App store and Google Play such as which categories of application have top downloads, which categories of applications are paid, does price have negative impact on number of installs and rating. That information will help app-developing companies know the publics' demands for applications, so they can know what applications and features they should work on when inviting new applications.

The two CSV files of App Store and Google Play are downloaded from Kaggle. The App store dataset was collected in 2017 with more than 7000 Apple applications, and Google Play dataset was collected in 2018 with more than 10,000 Google Apps. In each dataset, details include App category, age limits, price, number of installs, number of reviews, average rating and etc.

My plan for the project mostly is to clean and make the data in those two datasets consistent so that comparisons are reasonable. Examples include changing continues data such as price, rating, and number of reviews to categorical data and using one standard for age limits.

At the end, I would like to demonstrate

- number of installs and average rating for each category of applications using histogram
- price distribution for each genre using boxplot
- if price has negative impact on installs and ratings using coefficient.
- each category's proportion in App store and Google Play
- proportion of free and paid in App store and Google Play