Best apps store for your operating system

Michael Underwood

December 29th, 2020

The Business Problem

- Problem: Customer needs a major apps store to build into their operating system for the greatest user experience. They have narrowed down the options to Google Play and Apple Store

- Goal: To use user reviews to discover whether Google Play apps have higher reviews on average than Apple Store apps (or vice versa)

Cleaning & Transforming Data

- Loaded in Google and Apple data and filtered for 'Category', 'Rating', 'Reviews', and 'Price' and made sure column labels matched for the different stores.

- Ensured price data was a numeric value with no symbols in the entry, eliminated values that were not numeric (e.g. "Everyone")

- Merged data and dropped entries that had incomplete or missing values for any column

- Filter only those apps that have been reviewed at least once

```
Out[8]:  array(['0', '$4.99', '$3.99', '$6.99', '$1.49', '$2.99', '$7.99', '$5.99',
               '$3.49', '$1.99', '$9.99', '$7.49', '$0.99', '$9.00', '$5.49',
               '$10.00', '$24.99', '$11.99', '$79.99', '$16.99', '$14.99',
               '$1.00', '$29.99', '$12.99', '$2.49', '$10.99', '$1.50', '$19.99',
               '$15.99', '$33.99', '$74.99', '$39.99', '$3.95', '$4.49', '$1.70',
               '$8.99', '$2.00', '$3.88', '$25.99', '$399.99', '$17.99',
               '$400.00', '$3.02', '$1.76', '$4.84', '$4.77', '$1.61', '$2.50',
               '$1.59', '$6.49', '$1.29', '$5.00', '$13.99', '$299.99', '$379.99',
               '$37.99', '$18.99', '$389.99', '$19.90', '$8.49', '$1.75',
               '$14.00', '$4.85', '$46.99', '$109.99', '$154.99', '$3.08',
               '$2.59', '$4.80', '$1.96', '$19.40', '$3.90', '$4.59', '$15.46',
               '$3.04', '$4.29', '$2.60', '$3.28', '$4.60', '$28.99', '$2.95',
               '$2.90', '$1.97', '$200.00', '$89.99', '$2.56', '$30.99', '$3.61',
               '$394.99', '$1.26', 'Everyone', '$1.20', '$1.04'], dtype=object)
```

Visualizing

- Summarize the data visually and analytically (by the column `platform`)
- Added platforms "apple" and "google" for each entry to differentiate between the different sources
- Calculated basic statistical data such as mean and standard deviation
- Created boxplot for the different sources to understand the distribution by rating

Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x210769e2790>

Boxplot grouped by platform
Rating
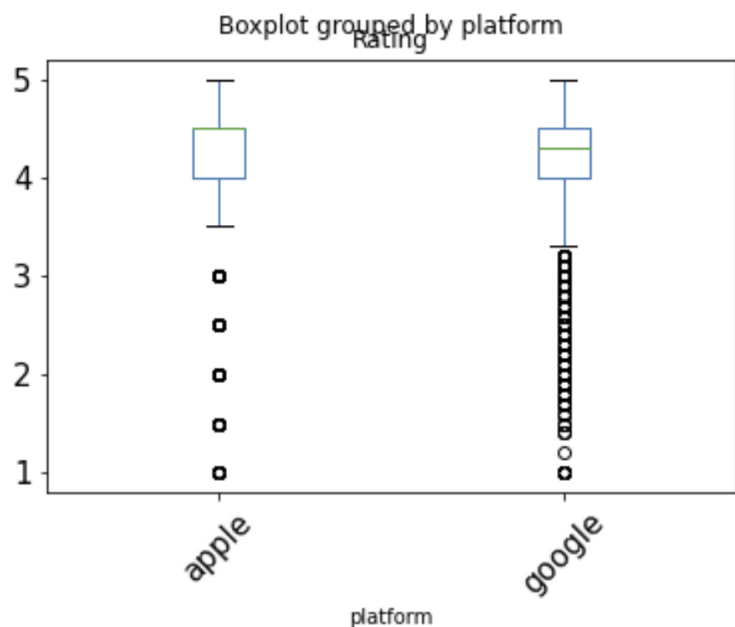
Out[21]:

platform	count	mean	std	min	25%	50%	75%	max
apple	6268.0	4.049697	0.726943	1.0	4.0	4.5	4.5	5.0
google	9366.0	4.191757	0.515219	1.0	4.0	4.3	4.5	5.0

Modeling

- Checked if data was normally distributed by p-value and verified with histogram

- Although mean reviews were similar (4.05 for Apple and 4.19 for Google), proceeded with a non-parametric test

- Randomized the *Ratings* data for 10,000 permutations and plotted the absolute difference in means in a histogram
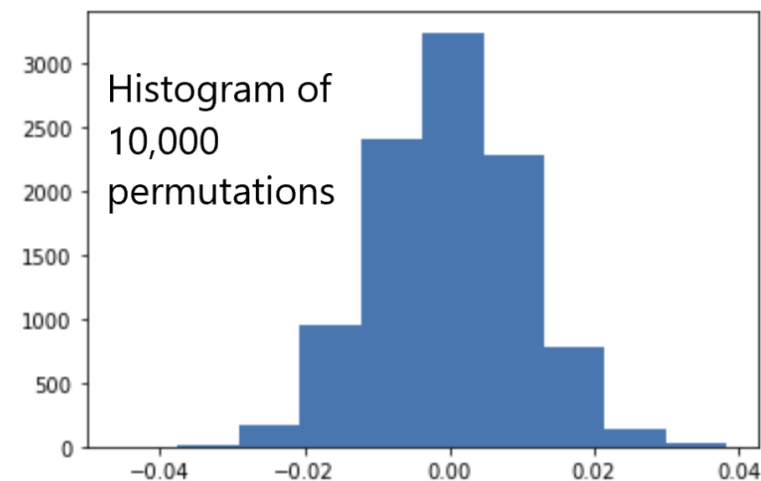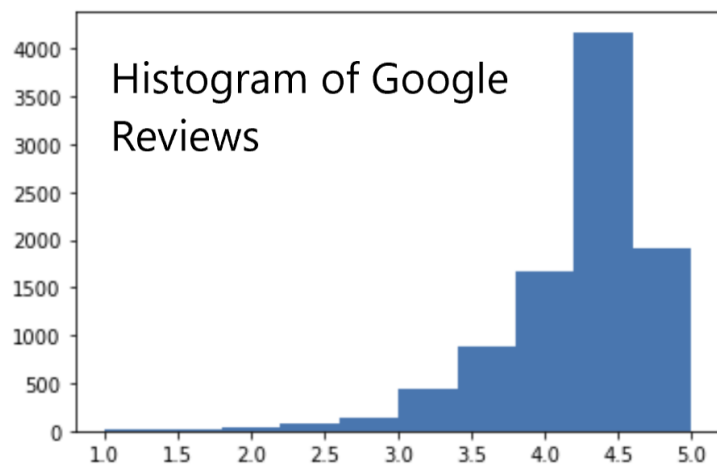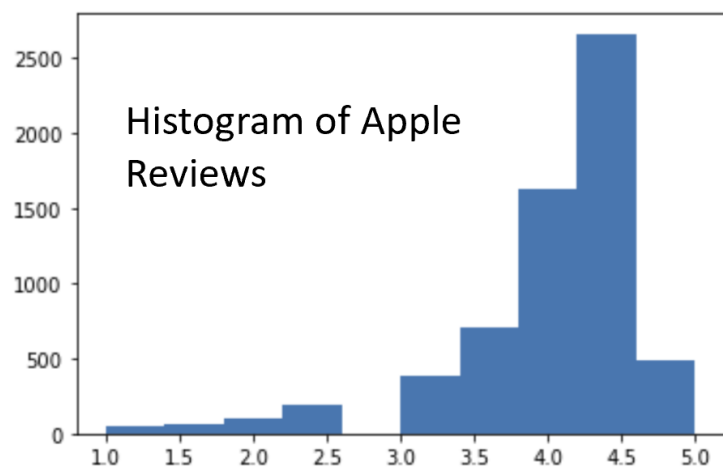
Findings

It was found that shuffling the data 10,000 times by *Ratings* provided many different values for the difference of means. Our significance level was 5% but shuffling our data found that *no values* were as extreme as our original grouping by the Apple and Google platforms (p-value of 0).

Our null hypothesis – the observed difference in the mean rating of Apple Store and Google Play apps is due to chance (and thus not due to the platform) – is rejected, the platform is very significant.

It is my recommendation that the customer only integrate Google Play into their operating system interface because on average the applications received better reviews and provide a better user experience.