Harun Gunes

Reviews in Apple Store vs Google Play

Did Apple Store apps receive better reviews than Google Play apps?

We are going to decide if platform is important to upload apps by checking data from Apple Store and Google Play.

Cleaning / Transforming

1. Checked the data types. The data type of the price column on google. dataframe was object. We fixed it by eliminating 'everyone' and '$' symbol.
2. # Convert the 'Reviews' and 'Price' column to a numeric data type.

```
# Use the unique() pandas method on the Price column to check its unique values.
Google['Price'].unique()
```

```
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)
```

```
# Use the function dtypes.
Google['Price'].unique()
```

```
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.  ,   5.49,  10.  ,
        24.99,  11.99,  79.99,  16.99,  14.99,   1.  ,  29.99,  12.99,
         2.49,  10.99,   1.5 ,  19.99,  15.99,  33.99,  74.99,  39.99,
         3.95,   4.49,   1.7 ,   8.99,   2.  ,   3.88,  25.99, 399.99,
        17.99, 400.  ,   3.02,   1.76,   4.84,   4.77,   1.61,   2.5 ,
         1.59,   6.49,   1.29,   5.  ,  13.99, 299.99, 379.99,  37.99,
        18.99, 389.99,  19.9 ,   8.49,   1.75,  14.  ,   4.85,  46.99,
       109.99, 154.99,   3.08,   2.59,   4.8 ,   1.96,  19.4 ,   3.9 ,
         4.59,  15.46,   3.04,   4.29,   2.6 ,   3.28,   4.6 ,  28.99,
         2.95,   2.9 ,   1.97, 200.  ,  89.99,   2.56,  30.99,   3.61,
       394.99,   1.26,   1.2 ,   1.04])
```

Cleaning / Transforming

3. Changing the column names of Apple to prepare for our join of the two datasets

Form

	prime_genre	user_rating	rating_count_tot	price
0	Games	4.0	21292	3.99

to

Category	Rating	Reviews	Price
ART_AND_DESIGN	4.1	159	0

4. Join the two datasets

	Category	Rating	Reviews	Price	platform
12526	Book	4.0	160	1.99	apple
2253	MEDICAL	4.6	92	79.99	google
15970	Games	5.0	3	0.99	apple
1905	GAME	4.4	5566889	0.00	google
15235	Photo & Video	3.5	2373	0.00	apple
5413	FAMILY	4.5	29495	0.00	google

Cleaning / Transforming

5. Eliminate the NaN values.
We eliminated 929 apps from 16563 because of NaN values.

Visualizing

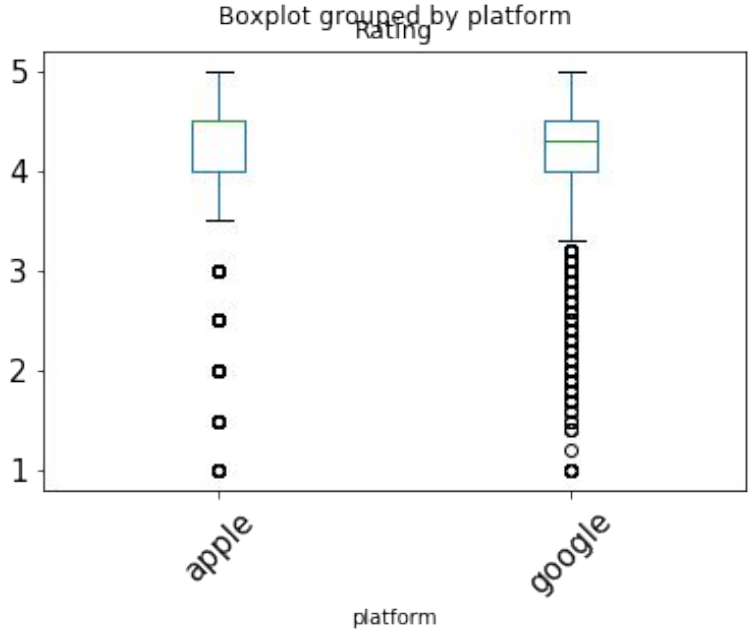
1. We checked means of platforms which are Google and Apple according to ‘Rating Count’.

Mean of Apple’s rating count = **4.049697**
Mean of Google’s rating count =**4.191757**

Means of the both platforms are very close each other.

Visualizing

2. Let's also get a visual summary of the ‘Rating’ column. When we look at the boxplot, their boxplots are similar.



Modelling

1. Our null hypothesis is 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).

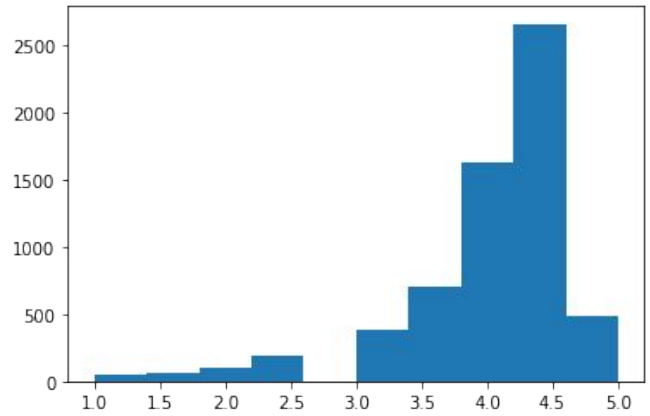
2. Alternative hypothesis is the observed difference in the average ratings of apple and google users is not due to chance (and is actually due to platform)

Modelling

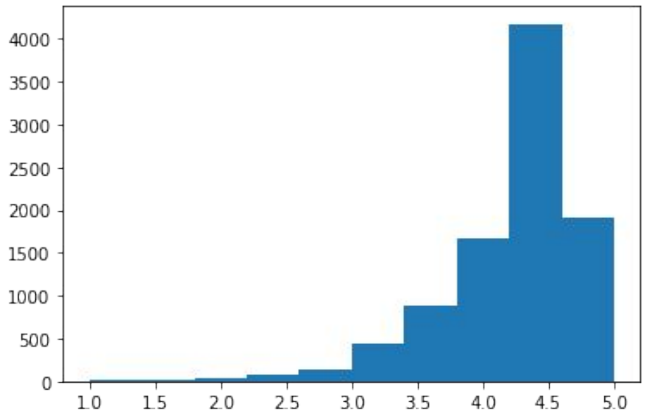
3. Firstly, we decided if the data distribution is normal by checking p value for both platform.

Since the p-values is 0 for both tests, regardless of what we pick for the significance level, our conclusion is that the data are not normally distributed.

Apple



Google



4. When we checked the distribution with a histogram, we see that histogram is not symmetric and unimodal (one hump)

Modelling

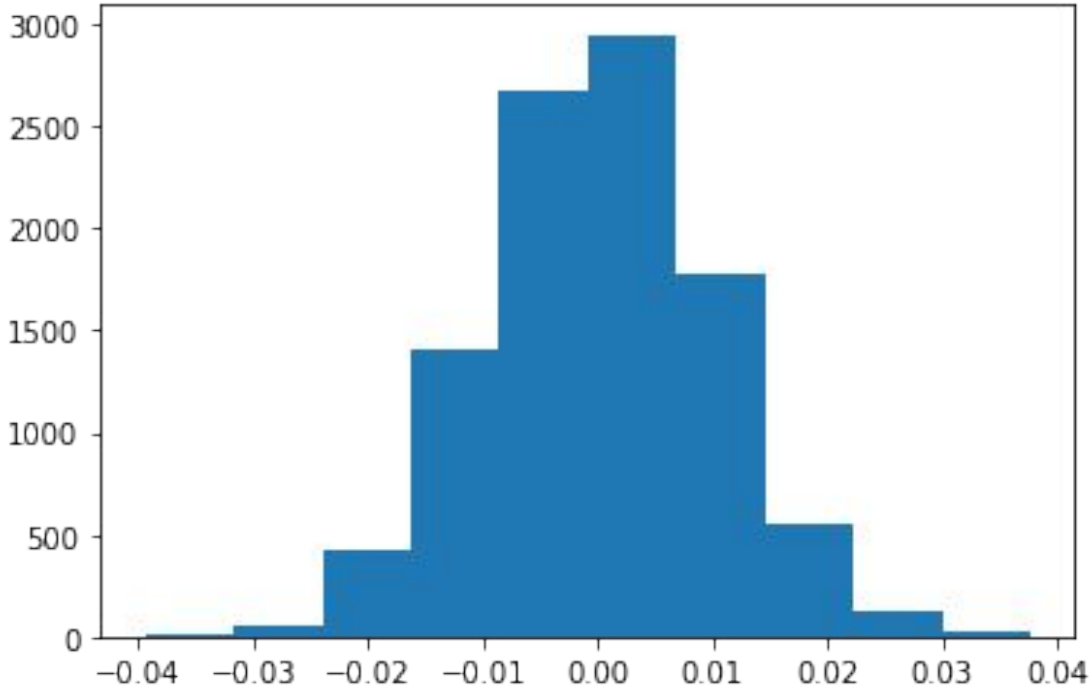
5. **Since the data aren't normally distributed, we're using a permutation test (*non-parametric)* test here.**

Permutation test helps us to distribute ‘Rating’ data randomly but platform columns are not changing with ‘Rating’.We repeated the permutation 10000 times. Difference of means between the two platforms is around 0.00.

Before permutation test the difference was around 0.14.

Modelling

Histogram of the difference



Evaluating and Concluding

1. So actually, zero differences are at least as extreme as our observed difference!

2. So the p-value of our observed data is 0.

3. It doesn't matter which significance level we pick; our observed data is statistically significant, and we reject the Null Hypothesis.

4. We conclude that platform does impact on ratings. Google Play can be preferred.