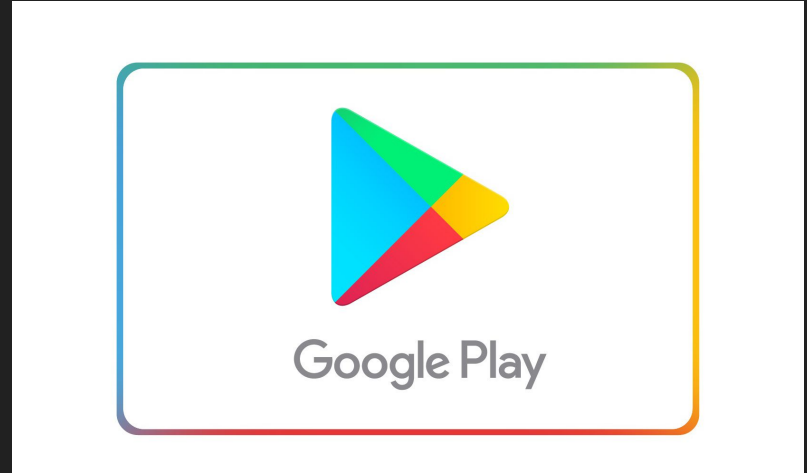


Which App Store is Better: Executive Presentation

# Apple or Google?



In this project, the data science team tried to determine whether Google Play or Apple Store is a safer and better choice to strike a deal with, since we can only make a deal with one of these entities.

Conclusion:

It is better to strike a deal  
with Google Play!



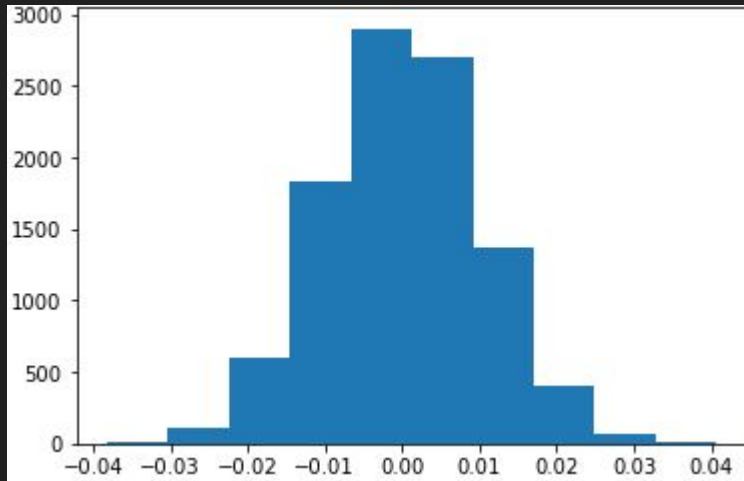
Google Play clearly outperforms the Apple Store in app ratings

How do we know?



Google Play has a greater average app rating by about 0.142 (out of 5). This may seem to be a very small amount, so it could just be random, BUT...

We randomly assigned each app to google or apple to see whether an average rating difference of 0.142 could occur by random chance, and we calculated the average differences after 10000 random assignments.



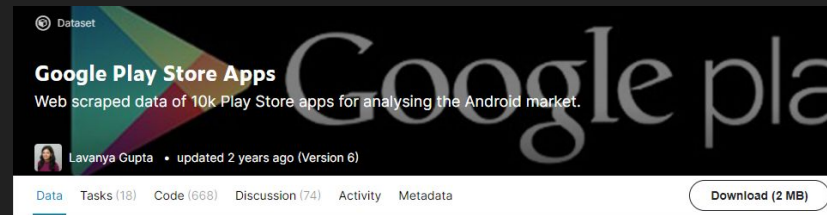
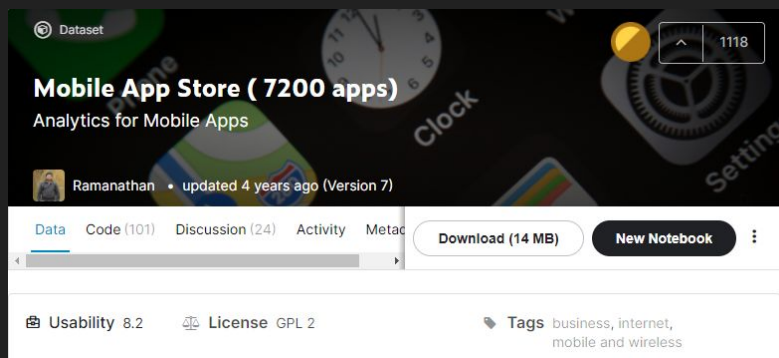
And we found that all of the observed average differences in app rating were less than or equal to about 0.04. 0.142 is literally OFF THE CHARTS!

0.142

Thus, the better ratings of Google Play are certainly NOT due to random chance, so Google Play is certainly the better investment based on app ratings.

That is the essential conclusion and reason:  
Google Play is a better choice than Apple Store because it has a systematically higher average app rating (and more apps) which was verified by a statistical permutation test to be virtually certain truth.

More technical justification to follow, only for interested audiences



Our first step was to load up web-scraped data from about 10000 Google Play apps and over 7000 Apple Store apps, housed in csv's on the reliable website Kaggle. There are many features in this data not relevant to our project:

Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Andri
4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4 anc
3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4 anc

It turned out that we could infer only 2 of these features to be relevant to the question we are asking (which app store is better to strike a deal with):

Rating	Reviews
4.1	159
3.9	967
4.7	87510

1: **Rating**, because it's directly relevant to **EVALUATING** the app store quality

2: **Reviews**, because we **can't** rate apps that have **0 reviews**

```
df.groupby('platform')['Rating'].describe()
```

	count	mean	std	min	25%	50%	75%	max
platform								
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

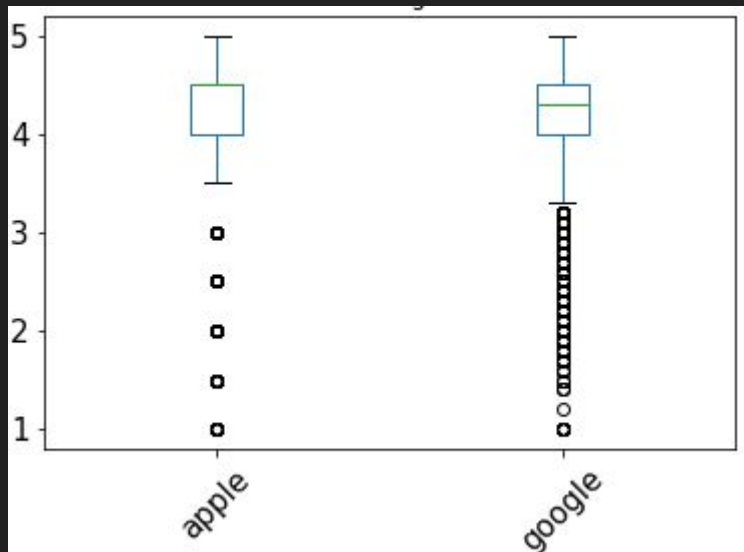
```
df = df[df['Reviews'] != 0]
```

	count	mean
platform		
apple	6268.0	4.049697
google	9366.0	4.191757

By doing a **GroupBy** on each platform and describing the **Ratings** column, we can see the **0.142** difference in average rating that we saw before!

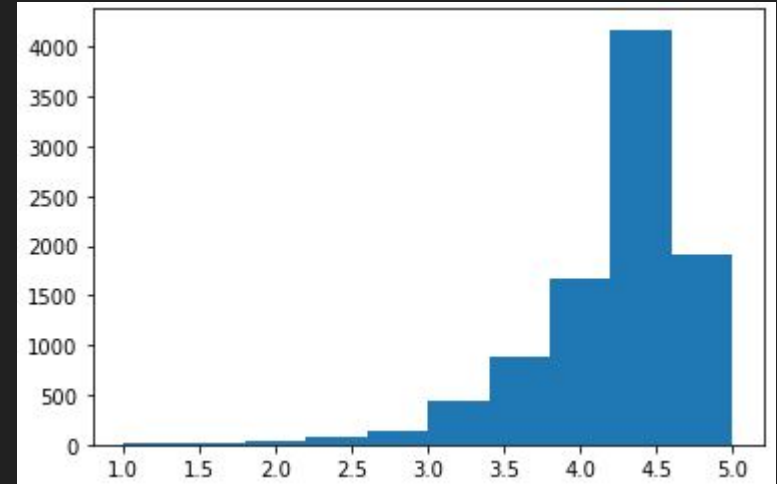
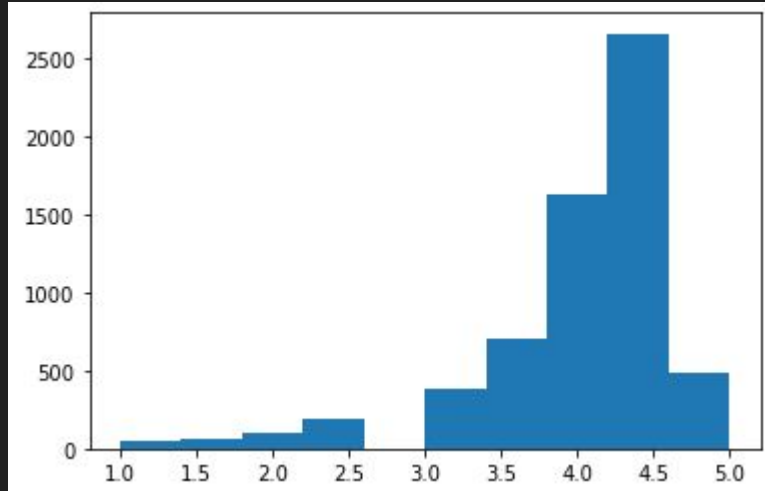


But it's not obvious whether 0.142 is a random chance difference or not, since the app rating distributions look very similar except for relative sparsity.



We need to perform a statistical test to find out whether the 0.142 difference is likely random or not.

To decide **which test** to perform, we need to know whether the **distributions** of app ratings are **normal** or not. This requires that they are **symmetric**!



We can see that they clearly are **NOT NORMAL**!

This means we need to perform a **non-parametric** test, such as a **permutation test**, and check the **statistical significance** of our 0.142 difference against random chance. Here are our **two hypotheses**:

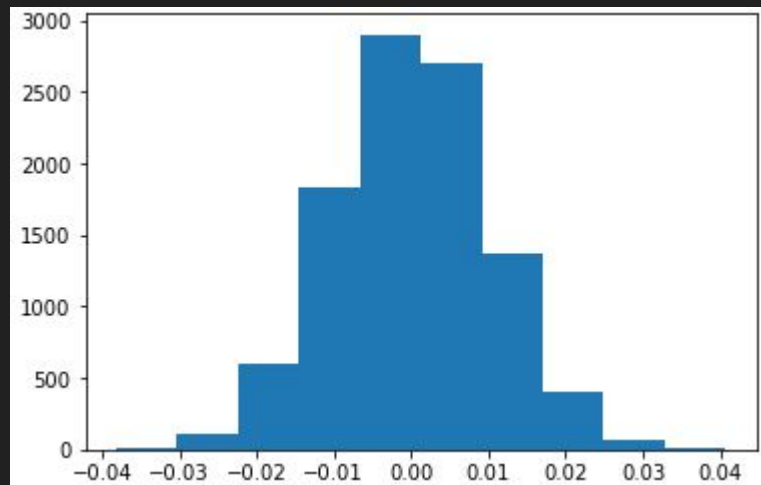
**H<sub>null</sub>**: the observed difference in the mean rating of Apple Store and Google Play apps is due to chance (and thus not due to the platform).

The more interesting hypothesis is called the **Alternate hypothesis**:

**H<sub>alternative</sub>**: the observed difference in the average ratings of apple and google users is not due to chance (and is actually due to platform)

We're also going to pick a **significance level** of 0.05.

We did 10000 permutations of assigning the review ratings randomly to a label of Apple or Google, observing the mean difference each time. **Not even ONE** of the 10000 random permutations had a difference even close to the 0.142 that we observed in our non-randomized data set.



In order to conclude that Google Play or Apple Store reviews are systematically better than the competition, we required a P-value as small as 0.05 (which means that 5% or fewer of the random permutations get results as high as the observed difference of 0.142). Our actual P-value is 0, since not even one of the random trials got a value close to 0.142. Therefore, it is virtually certain that Google Play has better reviews, and thus it is the better choice to sign a deal with.