

Data storytelling

The problem

Problem statement

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

Stages of the process

Sourcing and loading

- **Load the two datasets**
- **Pick the columns that we are going to work with**
- **Subsetting the data on this basis**

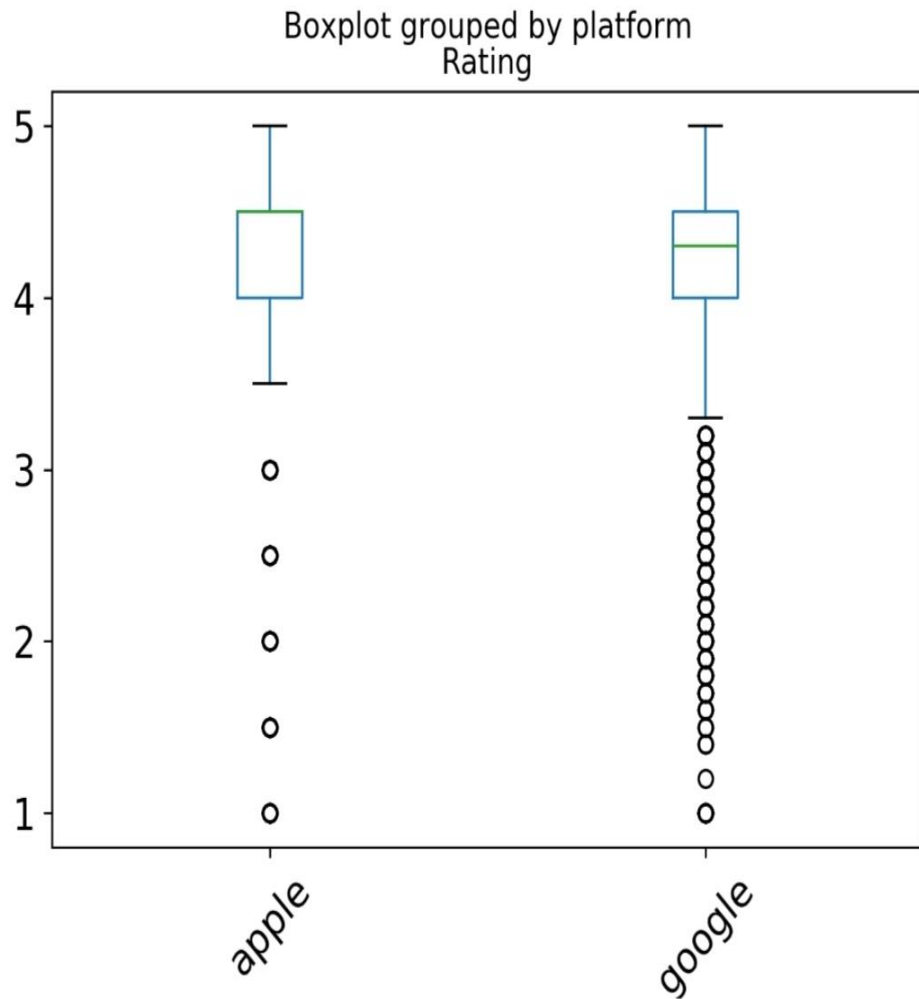
Cleaning, transforming and visualizing

- Check the data types and fix them
- Add a platform column to both the Apple and the Google dataframes
- Changing the column names to prepare for a join
- Join the two data sets
- Eliminate the NaN values
- Filter only those apps that have been reviewed at least once
- Summarize the data visually and analytically (by the column platform)

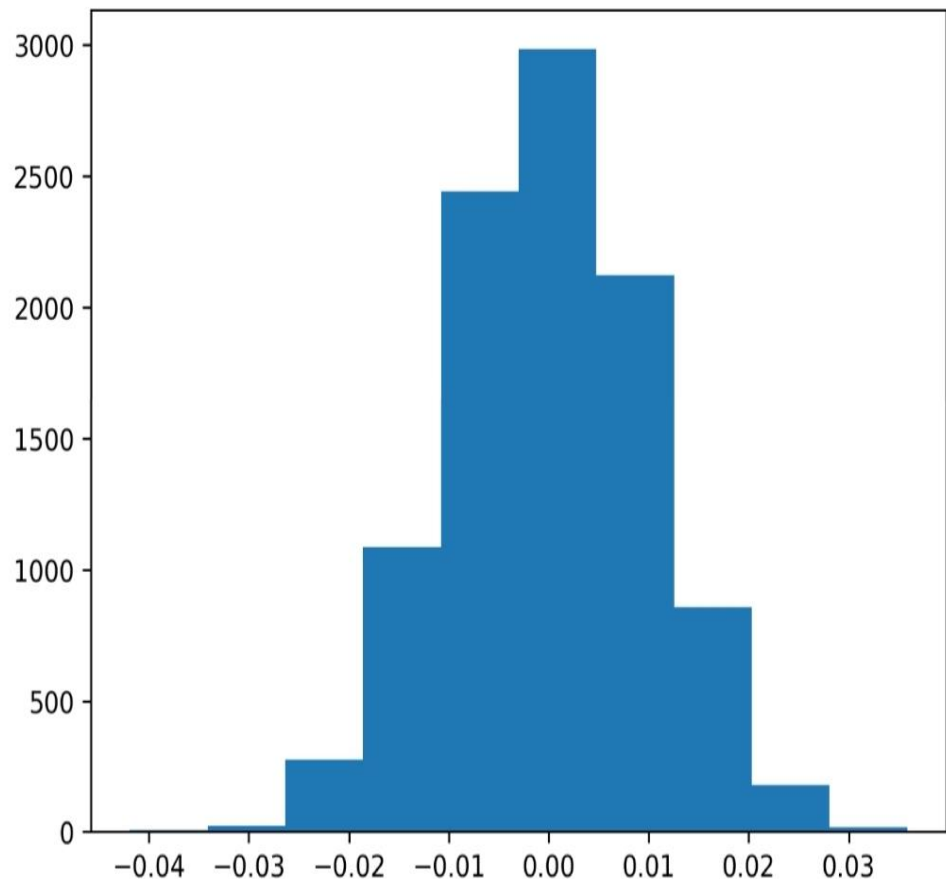
Modelling

- **Hypothesis formulation**
- **Getting the distribution of the data**
- **Permutation test**

Box plot of the data from the apple and google reviews



Difference between
the means of apple
and google reviews
over 10000
permutations



Conclusion

Observed difference is significant,
and that the Null hypothesis is
false; platform does have impact
on ratings_g