

# Assignment 3 – Data Discovery

By Randy Leon

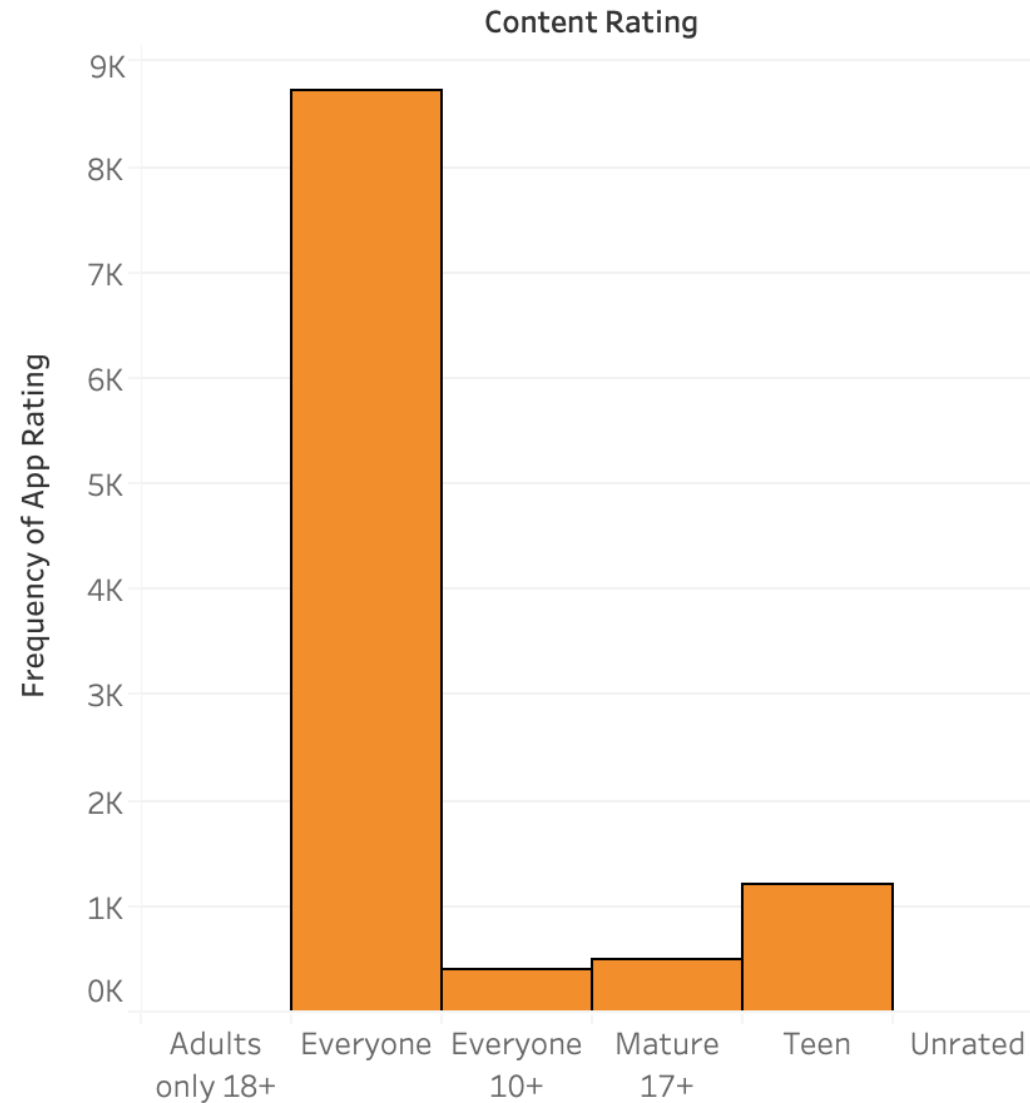
# Percent Distribution of 10k Google Play Apps (2017-2018)



I made a % distribution of ratings of ten-thousand Google Play apps between the years 2017 and 2018.

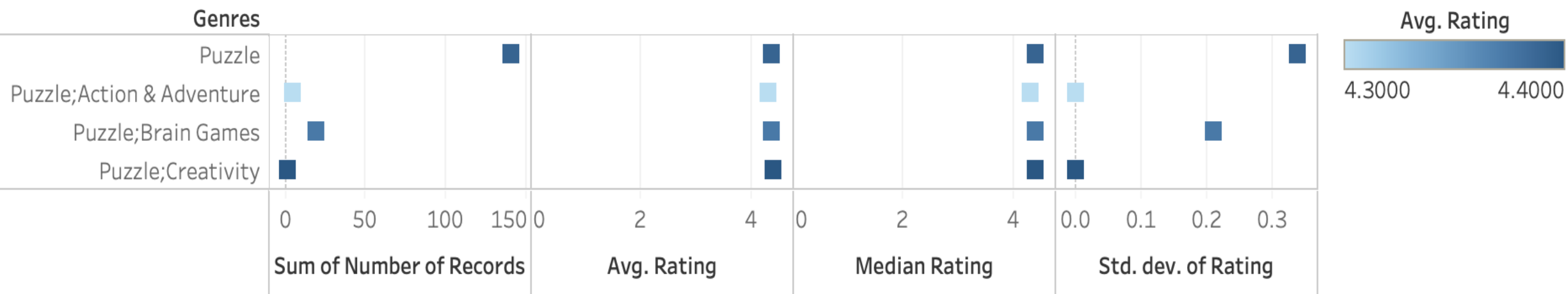
Interesting how my data is skewed left - more people review an app highly than those who reviewed it poorly. This can say that if you are posting a review of an app, chances are it's because you like the app to begin with; that is, unless you absolutely hate it so much, that you go out of your way to rate it one star (outlier explained)

# Frequency of Content Ratings Across 10,000 Google Play Apps (2017-2018)



Most apps are rated for everyone, that being said, those are probably the apps most people are likely to review.

# Statistical Layout of Ratings of Puzzle-Based Apps



I like to point toward the std. deviation portion of this visualization. Clearly, nearly everyone who rates puzzle games feels the same way about them, however there is no split between those that rated puzzle; action and adventure, and puzzle; creativity.

## Story 1

Question 1: Who personally tabulated this data and can I trust that source?

Question 2: Are there similar data sets that exist that I can compare results with? (For example, is this same data found for the year 2013?)

Question 3: What kind of business applications can be used from looking at this data?

Question 4: Are there current visualizations that do a better job of telling a story of the data than my own?

Question 5: How does the data differ than the set from the App Store(Apple) if at all?

Question 6: What can we learn from the landscape of mobile smartphone applications using this data?

DATASET: <https://www.kaggle.com/lava18/google-play-store-apps>