Top Video Game Analysis

Data Science Capstone Project

The Problem

Context:

- A data science project to review how well a game does based on its summary, scores, and user reviews.

For this project a hypothetical client could be a company that publishes and creates video games. Once we take in all of these factors and develop a model, it could be used to produce a game concept that would be a hit and thus profitable for the company.

Who are the stakeholders?

- Me
- Any Video Game company that wants to work with me

Data Information

Source:

Data acquired on Kaggle from user: Deep Contractor

Includes 6 attributes for being a top video game

Contains 18800 Observations.

File format: CSV file

**Each Record:** a unique Top selling video game ranging from 1995 through 2021

Dataset

Top Video Games 1995-2021 Metacritic
Details about the top rated video games 1995-2021

Deep Contractor · updated 2 months ago (Version 3)

Data Wrangling

- 'User_review' column had missing values, but didn't want to fill with things like an average.
- Categorical features such as 'title' and 'date_released' were left as objects.
- Didn't drop any columns because there were only 6.

Distribution of the 'user_reviews'.
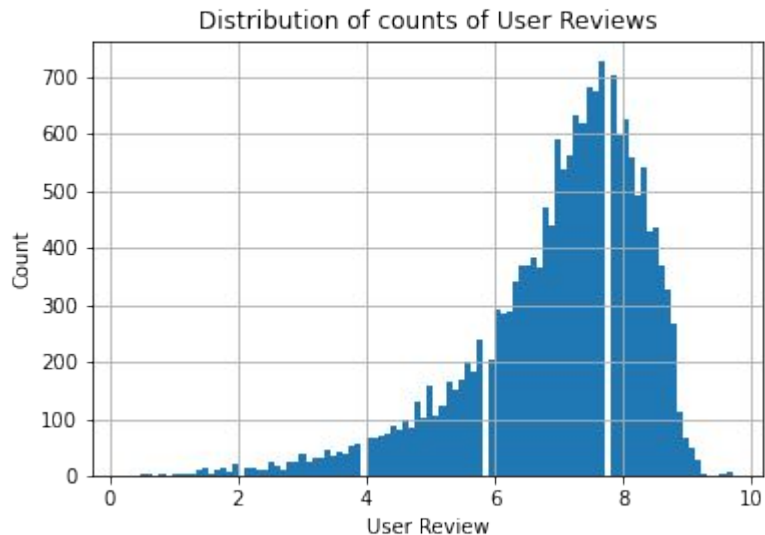
Distribution of the 'meta_score' reviews.
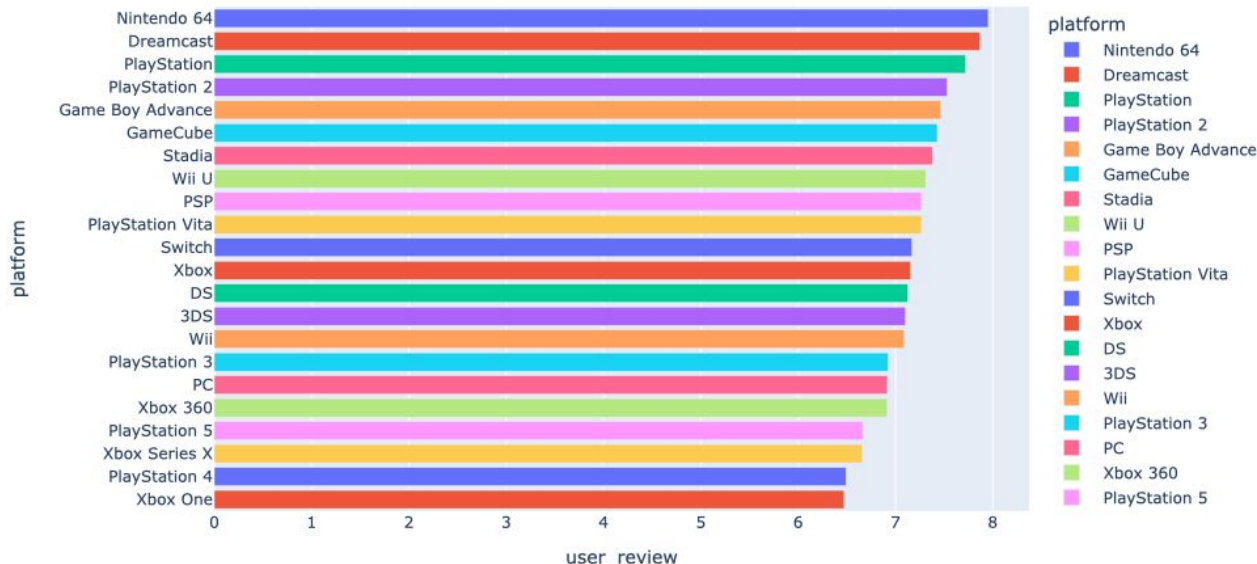

Distribution of counts of User Reviews

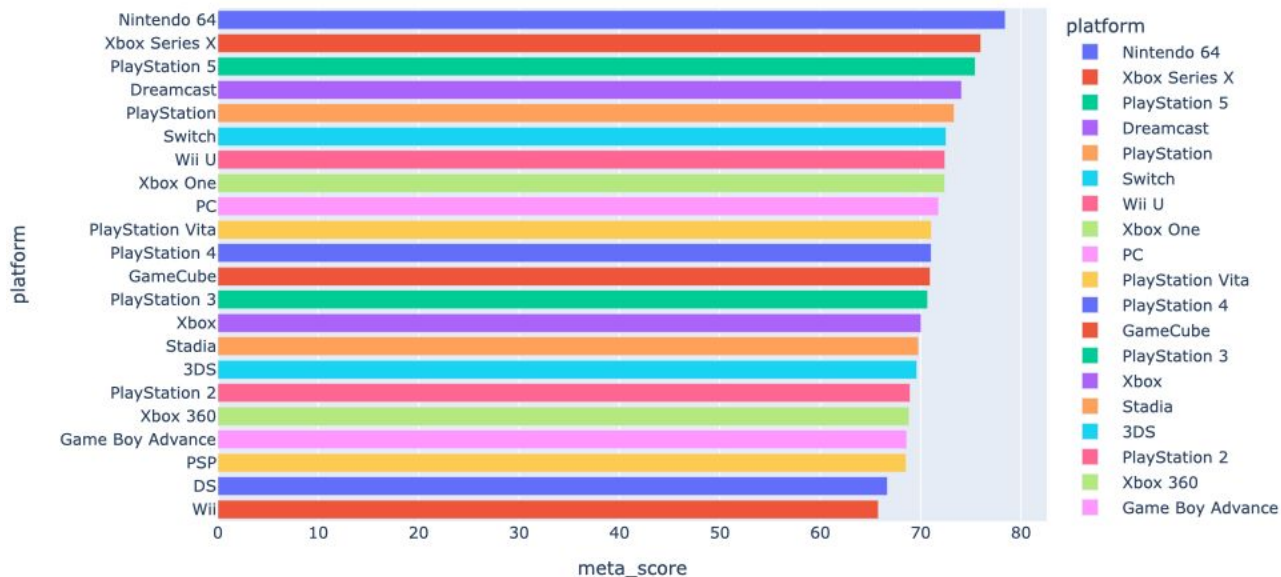Distribution of counts of Meta Website Game Scores
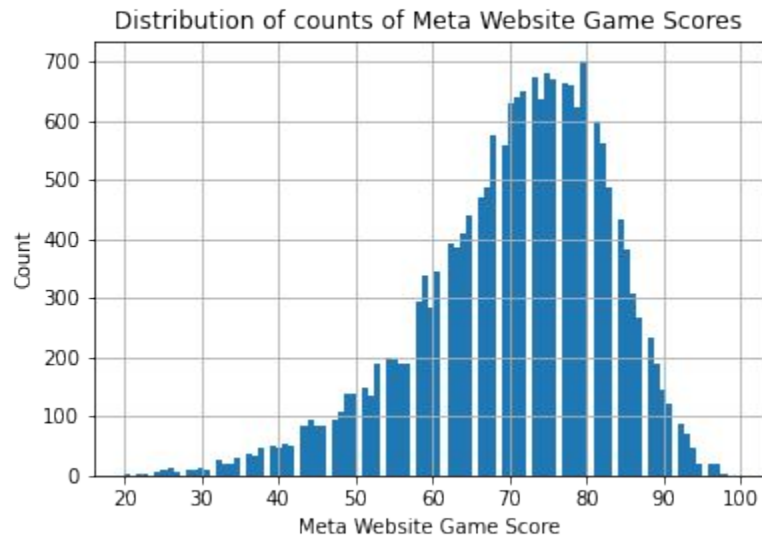
EDA - Plotting Reviews based on Platform - 'user_review'

I wanted to see how a game performed based on the platform it was played on.

EDA - Plotting Reviews based on Platform - 'meta_critic' reviews.

Top reviewed platforms in both columns was the cult classic 'Nintendo 64'!!

EDA - Correlation Heatmap

After distributing these reviews columns, I wanted to look at the correlation heat map and found that the two review columns have a solid 0.53 correlation.

Correlation Heatmap
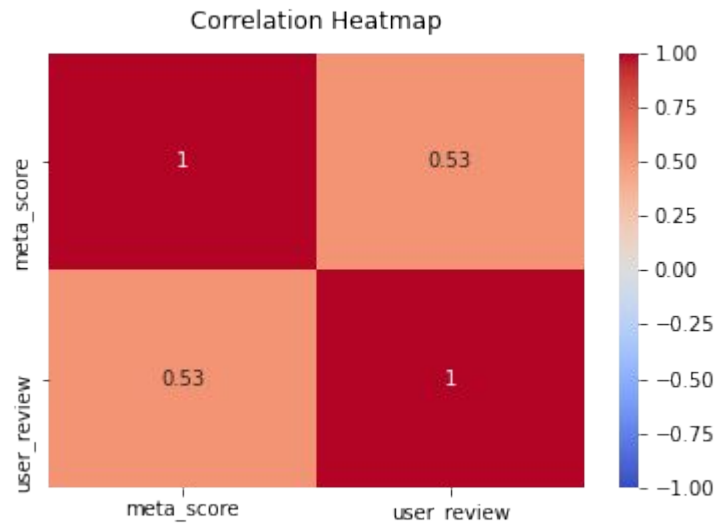
Unsupervised Machine Learning Modeling

Text Analysis / Topic Modeling

Modeling Overview

- Type:
  - Unsupervised Learning
- Description:
  - Text Analysis
  - Topic Modeling
- Tools: Python's Gensim library, and pyLDAvis library
- Modeling Steps:
  - Data Pre-processing (removing stop words, extending stop words)
  - Looping through and fitting the model (unsupervised)
  - Finding the best number for the topic model.

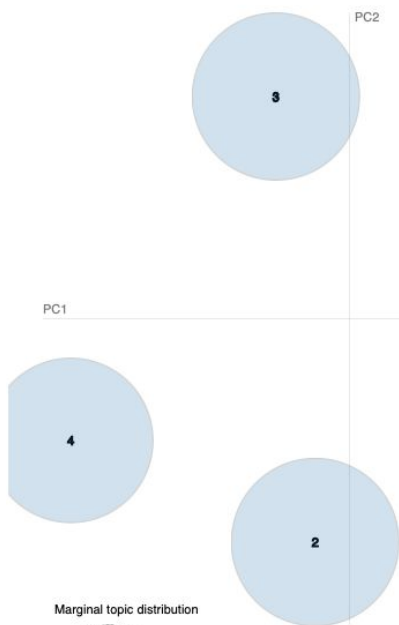# Topic Modeling using Gensim/pyLDAvis -no topic selected
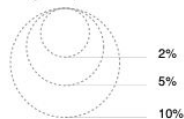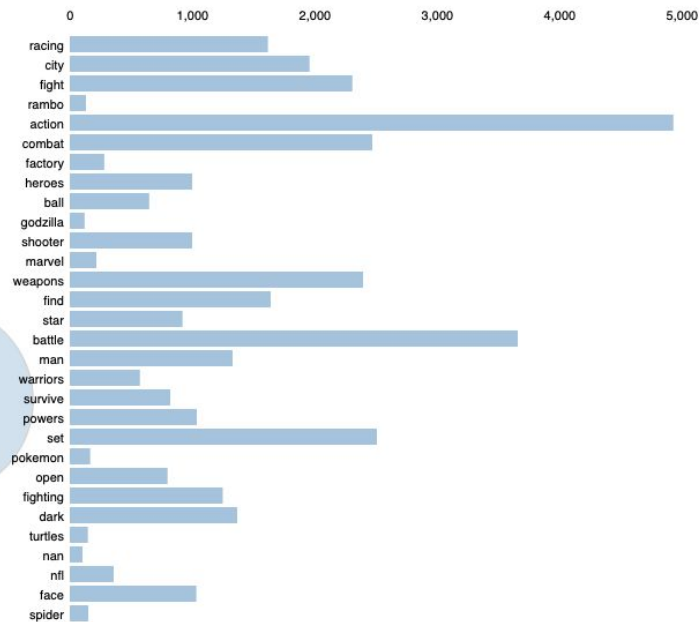
Topic Modeling using Gensim/pyLDAvis - 1st Topic
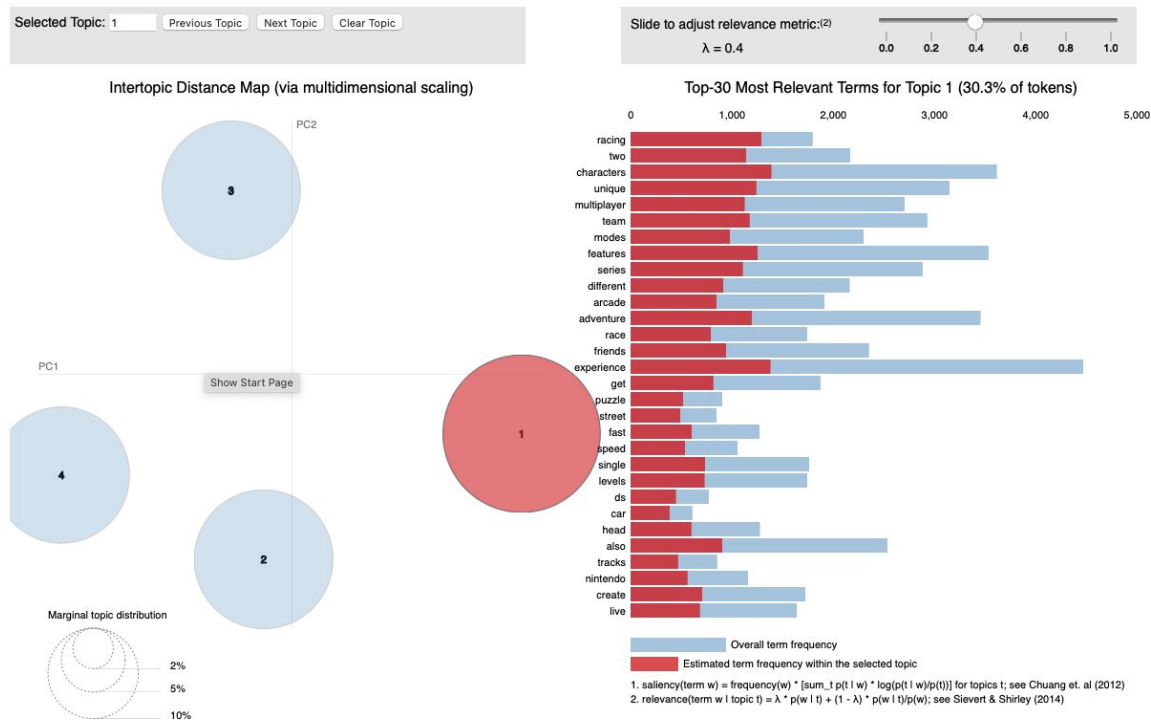
Topic Modeling using Gensim/pyLDAvis - 2nd Topic
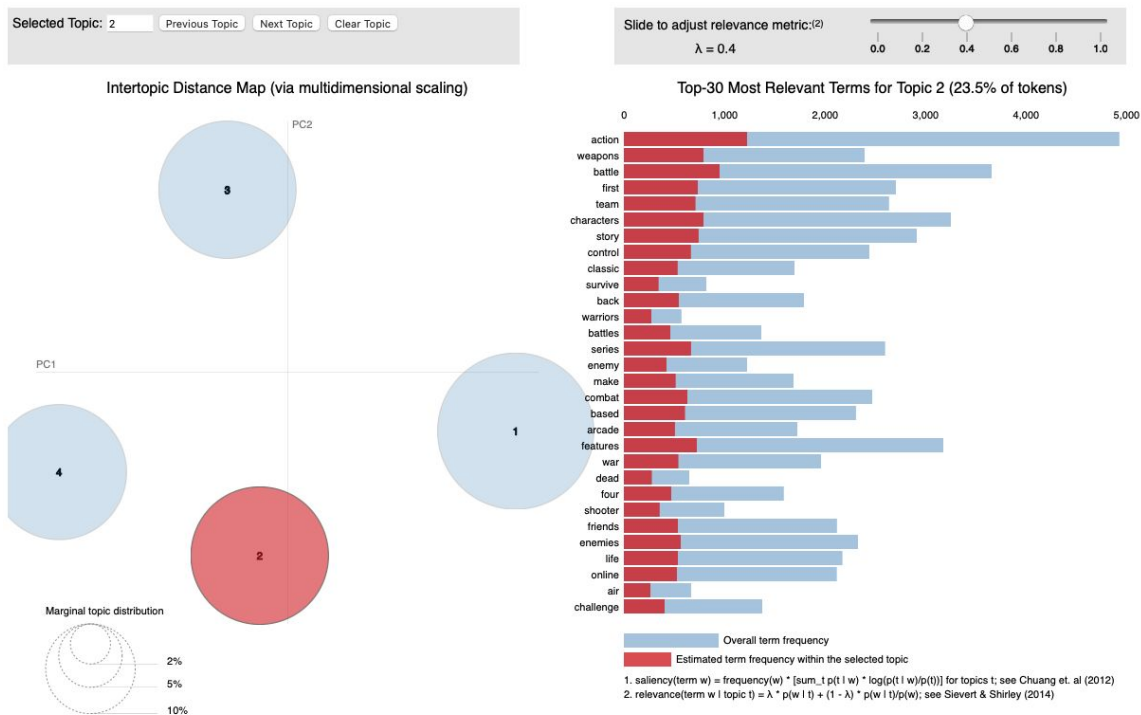
Topic Modeling using Gensim/pyLDAvis - 3rd Topic

Topic Modeling using Gensim/pyLDAvis - 4th Topic
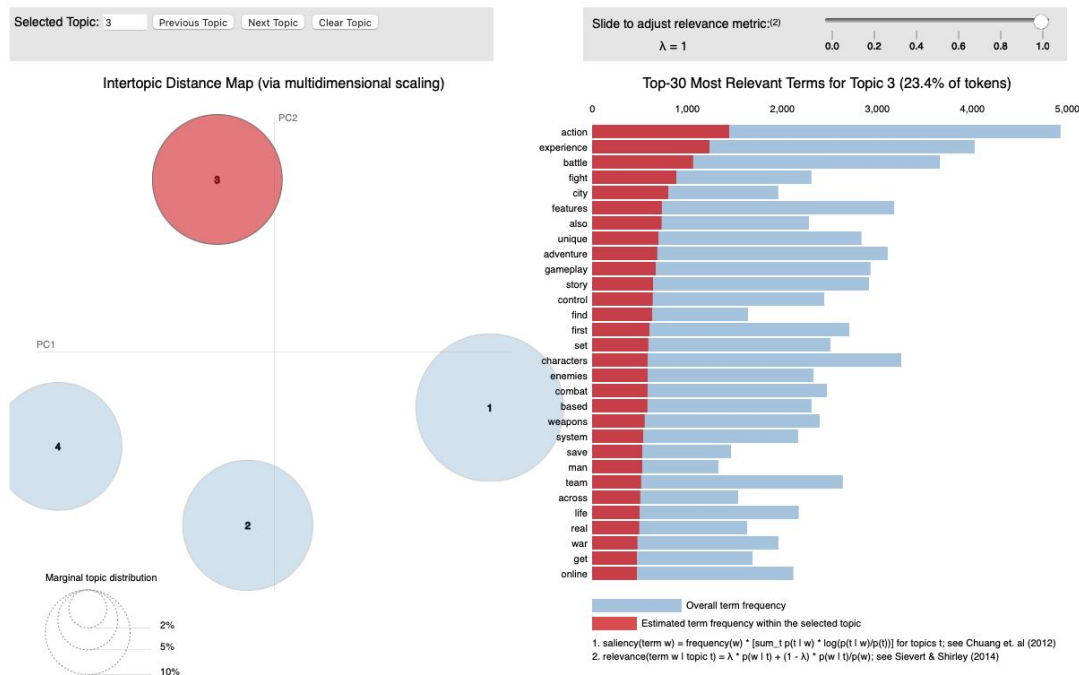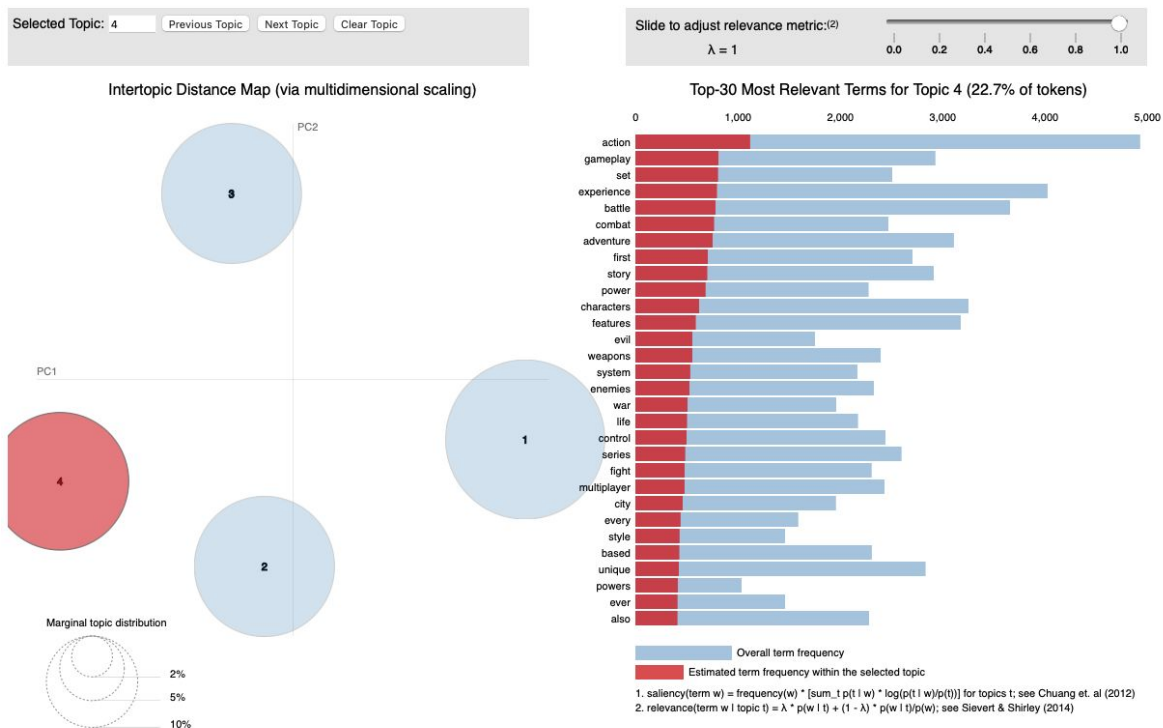
Conclusion

In conclusion it was really fun to see all the different things I did in the EDA and the modeling, specifically seeing the reviews based on the platform. It was amazing to see cult classics like 'Zelda', 'Soul Calibur' and other beloved titles.

So again if a company wants to make a best selling game based off the Top Video Games of 1995-2021 they should have a game that is an experience in action and adventure, but also have a sort of epic battle with lots of characters to interact with.

Resources

- https://www.kaggle.com/deepcontractor/top-video-games-19952021-metacritic
- https://radimrehurek.com/gensim/
- https://github.com/bmabey/pyLDAvis
- https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24
- https://plotly.com/python/
- My Github repo: https://github.com/FayD21/Top-Video-Game-Analysis-Capstone-3
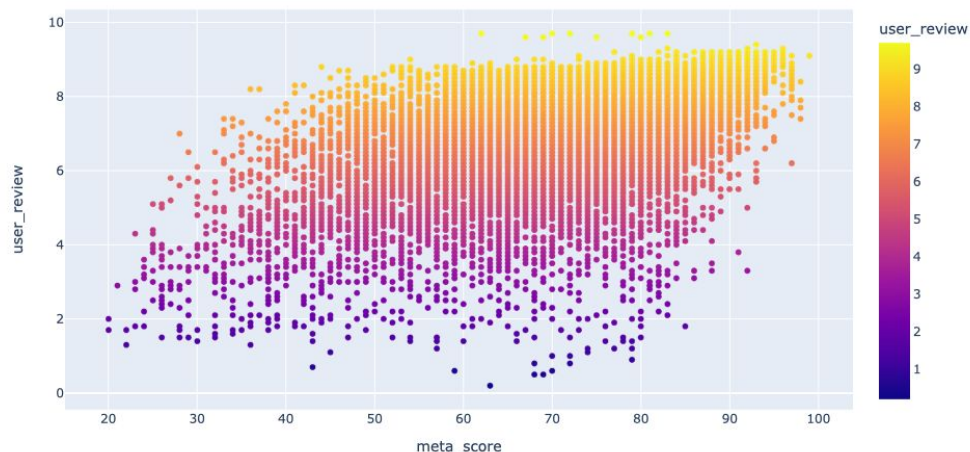
Thank You!

Fay Dennis

Fay Dennis

Email: iamfaydennis@gmail.com

GitHub: https://github.com/FayD21

Special Thanks to my Mentor!

Jeff Hevrin