

Data Mining Disney Reviews

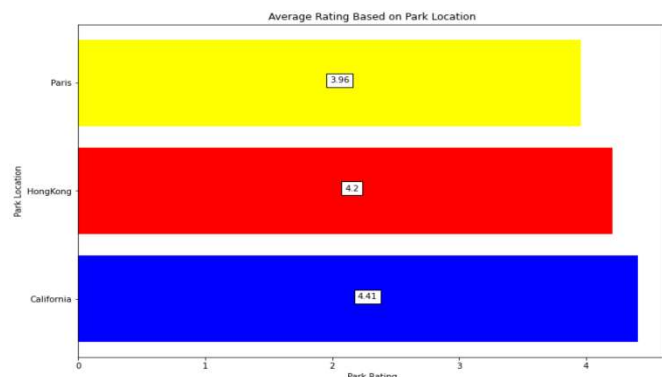
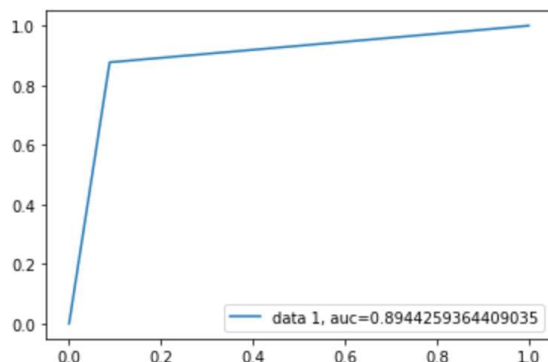
The project within the portfolio repository focuses on addressing the mass amount of reviews Disney receives a day, by using dummy variables to split the different parks to determine where these reviews are coming from, showing us what parks areas, the reviews are referencing, are the reviews negative or positive with sentiment analysis, and determine the steps needed to improve on guest's feedback. Which is completed in a Jupyter Notebook.

Instructions for Jupyter Notebooks and Content

- If you are unfamiliar with Jupyter Notebook, please follow the link for installation and getting started: [Project Jupyter | Home](#)
- Next download notebook [Disney Reviews Jupyter Notebook File](#) and run file in Jupyter Notebook
- Download data from GitHub [DisneylandReviews.csv](#)

Project Overview

A horizontal bar graph showed the average review rating based on park location. California had the highest rating out of the three park locations with a 4.41 and Paris had an average rating of 3.98 out of 5.



When building a model for the Disneyland California a ROC curve visualization can show you the accuracy of one's model as this model came in at 89.44% accurate.

Goals for Project

The goal for this project is to allow users to work through a dataset that is focused on reviews that a business receives. The project takes you through different steps such as visualizations, data preparation and extraction, engineering new features, creation of dummy variables, model building and evaluation. If you found this project useful and have any questions regarding any of the projects, please feel free to contact me at TheodoreKobyHercsky@gmail.com