Project Overview

In this project, I analyzed a dataset and then communicated my findings about it. I used the Python libraries NumPy, pandas, and Matplotlib to make the analysis easier.

This data set contains information about 10,000 movies collected from The Movie Database (TMDb), including user ratings and revenue. Certain columns, like ‘cast’ and ‘genres’, contain multiple values separated by pipe (|) characters. The final two columns ending with “_adj” show the budget and revenue of the associated movie in terms of 2010 dollars, accounting for inflation over time.

Packages Used

I installed python plus the following libraries:

* pandas
* NumPy
* Matplotlib
* csv

Questions Asked?

1. What's the Runtime of the movies?
2. What's the relationship between movie budget and revenue?
3. What is the most profitable movie?
4. What are the number of movies released each year?

What have I learned?

After completing the project i learned:

* All the steps involved in a typical data analysis process
* Knowing how to investigate problems in a dataset and wrangle the data into a format I can use
* Having practice communicating the results of my analysis
* Being able to use vectorized operations in NumPy and pandas to speed up your data analysis code
* how to use Matplotlib to produce plots showing the findings
* Being comfortable posing questions that can be answered with a given dataset and then answering those questions

Findings

I observed that Budget and Revenue have a positively correlated relationship.The revenue has remained higher than the budget throughout the years. Star Wars is the most profitable movie in the dataset and warrior's way has the lowest profitability.there is a significant increase from the number of movies released each year from 1960 to 2015