DrewAlderfer / phase_1_project Public

0 stars 0 forks

Star Unwatch

Code Issues Pull requests Actions Projects Wiki Security Insig

main

Andrew Alderfer 3 ... yesterday 3

View code

README.md

Phase 1 Project Overview

Welcome to my Phase 1 Datascience project repo!

This project was a self directed EDA and data processing project using the Python Pandas library and a collection of datasets from IMDB (the Internet Movie Database) and a few other similar sources. The Intention of the project is to discover, analyze and use patterns and relationships found in the provided datasets to make business recommendations according to the following prompt.

Business Understanding

The client, Microsoft, has decided to explore getting involved in the entertainment industry. Specifically, they are interested in making movies. I have been task with examining the provided data and making three recommendations for how to successfully enter the entertainment industry as a large, established company currently operating in a different industry.

Some questions I attempted to answer with this dataset were: * What types of movies are succesful in today's market * What patterns and trends can be gleaned from the dataset * Given the observable business environment, what practical advice would I provide a client about entering the industry

Data Understanding

The source for the data used was a publicly licensed sql database provided by the website IMDB. The database contained 8 tables:

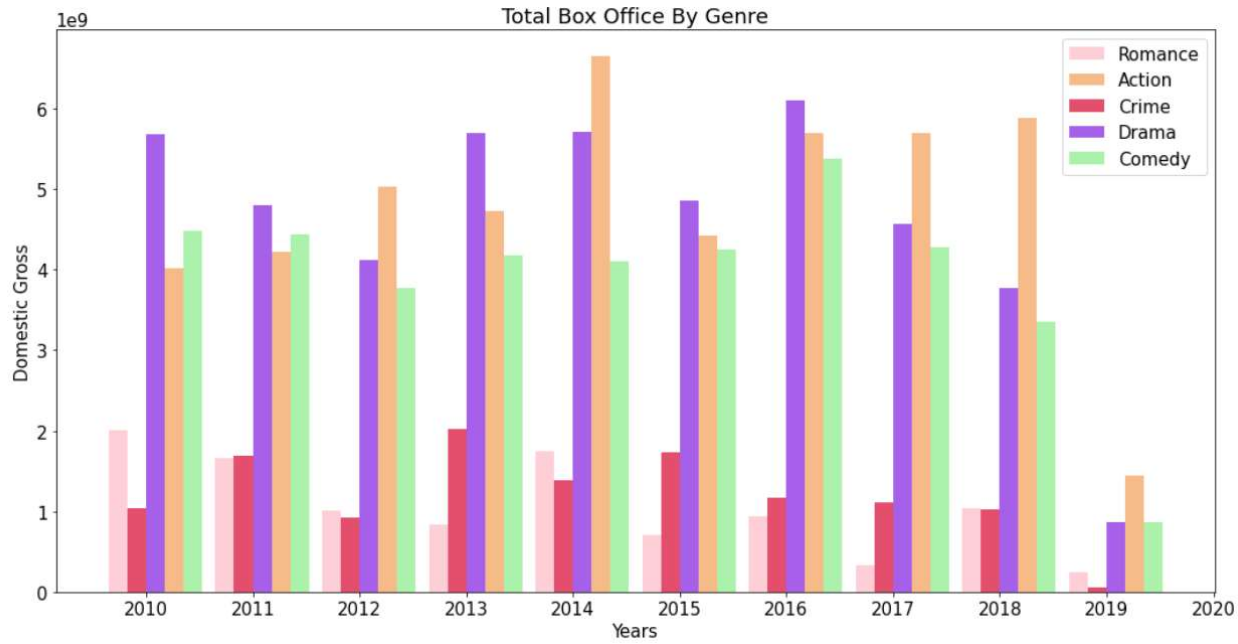
Out[3]:

	Table Names
0	movie_basics
1	directors
2	known_for
3	movie_akas
4	movie_ratings
5	persons
6	principals
7	writers

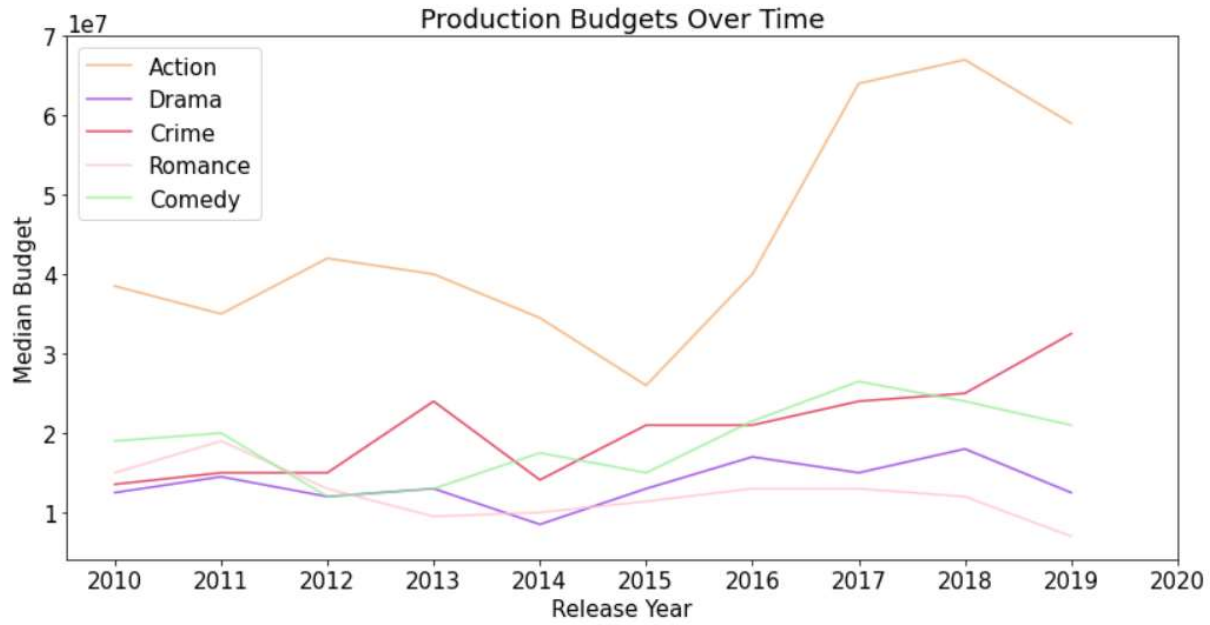
I started off the EDA process by pulling these table out of the sql database and storing them in Pandas DataFrame objects oragnized within a python dict object. I was then able to start observing the structure of and relationships between the tables and begin the process of analyzing there meaning.

Visualizations

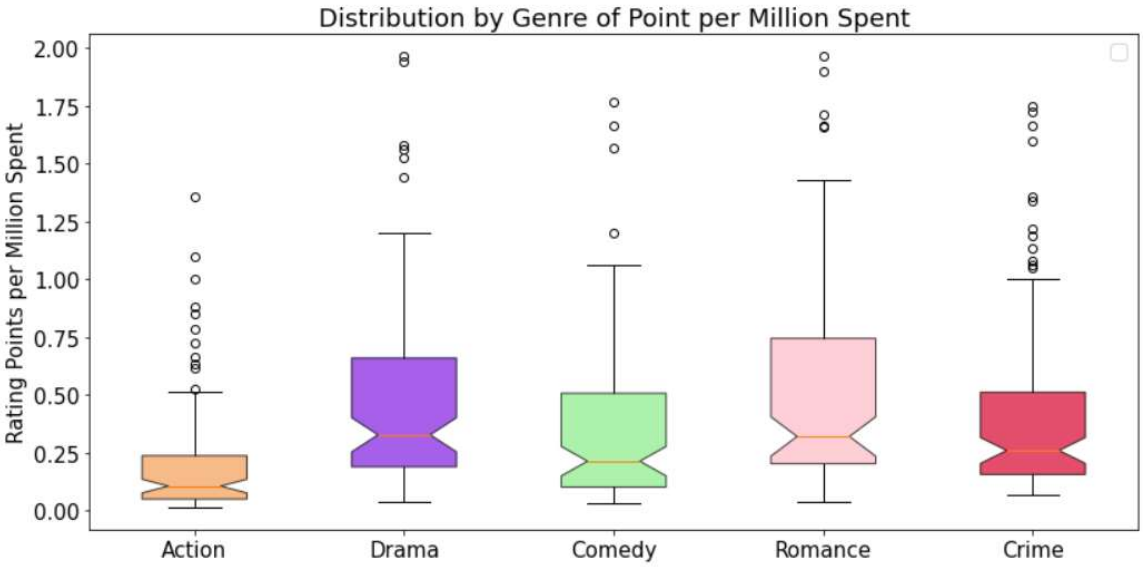
I started by visualizing the total box office grosses for each of the top genres.



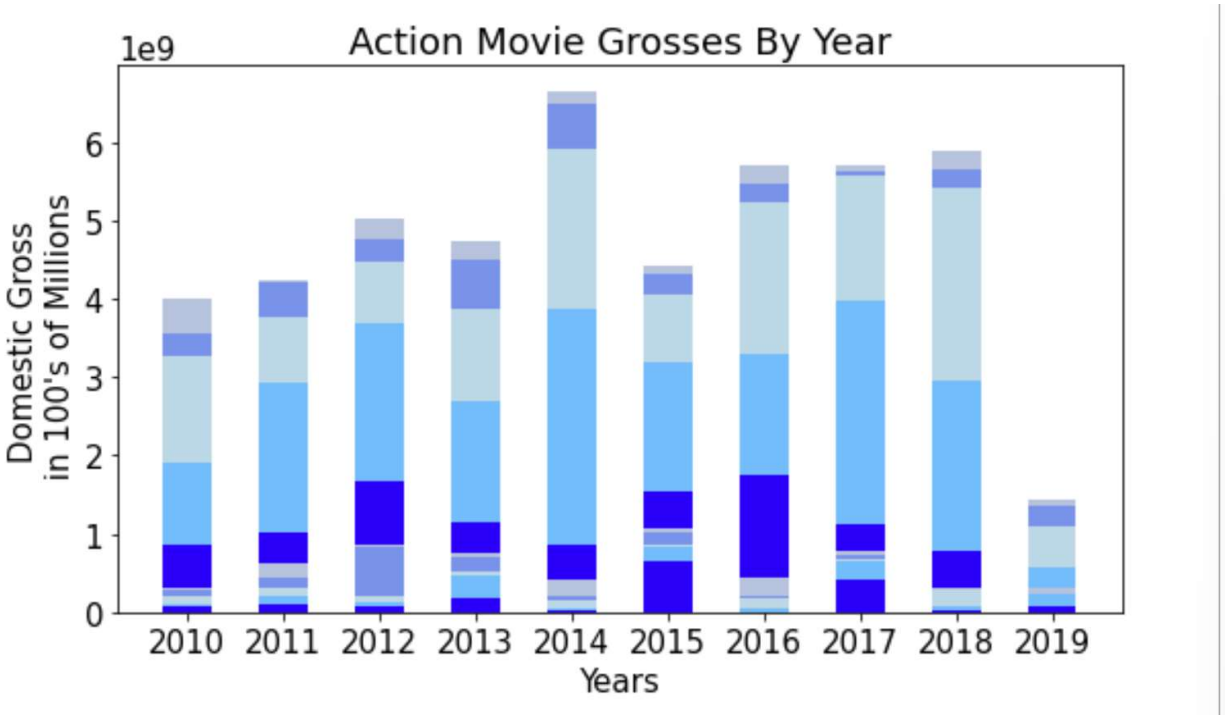
Then I decided to compare the median production budget by each of these genres.

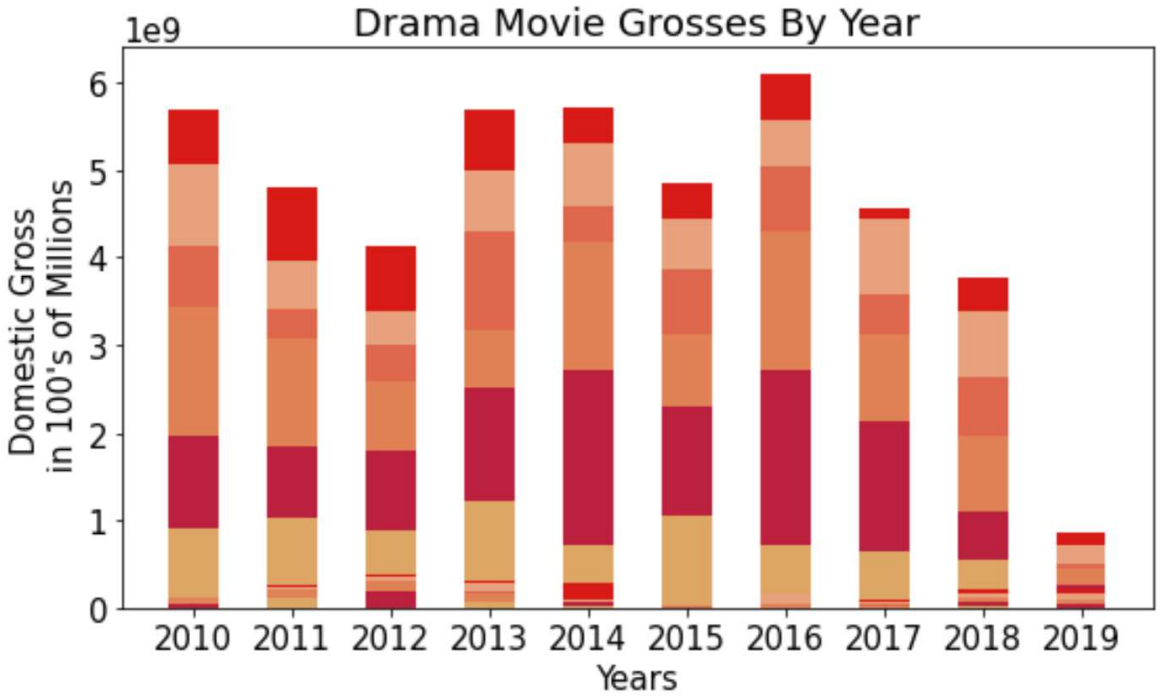


and the distribution within each genre of ratings vs budget:



After identifying that Action and Drama commanded the largest market share among the various genres, I wanted to compare the distribution of the that market share within each repsectively.





Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- Jupyter Notebook 100.0%