Movies Analysis for Microsoft

ANALYSIS BY SHRADHA WADDEPALLI

Summary

From ten database, I sifted through tables to examine the information provided. I used five tables for my project. After inspecting, the quality of data, I realised the need to clean data to plan my future moves towards analysis.

Taking advantage of the data present in each table I cleaned and joined them. Using some statistical methods and graphs to show top performing movies. These movies represent top performing genre's.

My second and third recommendations were extracted in few steps. I am required to Join & merge using some panda conditional data extraction methods to be able to finalise list of writers and directors.

Business Problem

Microsoft is aiming to create new movie studio. They require recommendations on the type of movies doing well at Box office. Recommedations incude actionable insights to help Microsoft with new movie studio.

Data

After sifted through the datasets, I noticed that, out of 10 different tables in CSV and JSON format, I can use 5 tables to extract desired data.

bom.movie_gross.csv.

imdb.title.crew.csv.

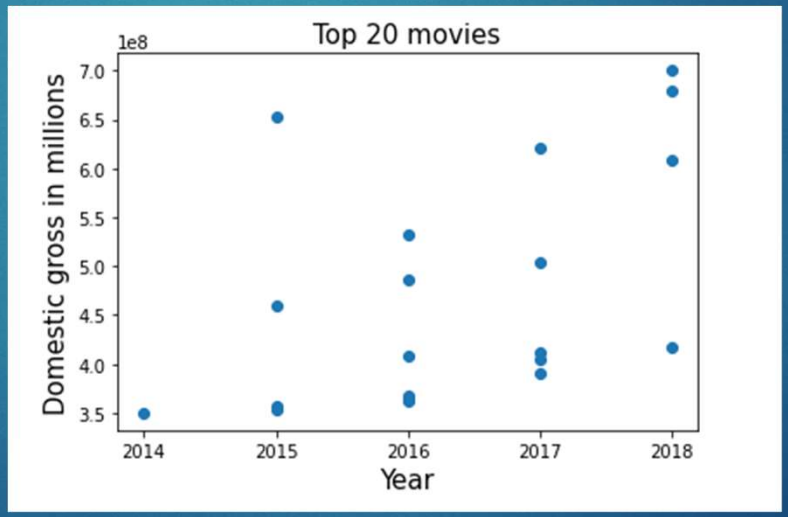
imdb.title.principals.csv.

imdb.title.basics.csv.

imdb.name.basics.csv.gz

Methods

I used "bom.movie_gross" dataset to extract highest domestic gross within 2014 to 2018 by using .sort and .loc method. I narrowed down the list to top 20. Plotted scatter plot by importing Matplotlib showing highest grossed movies in particular year.



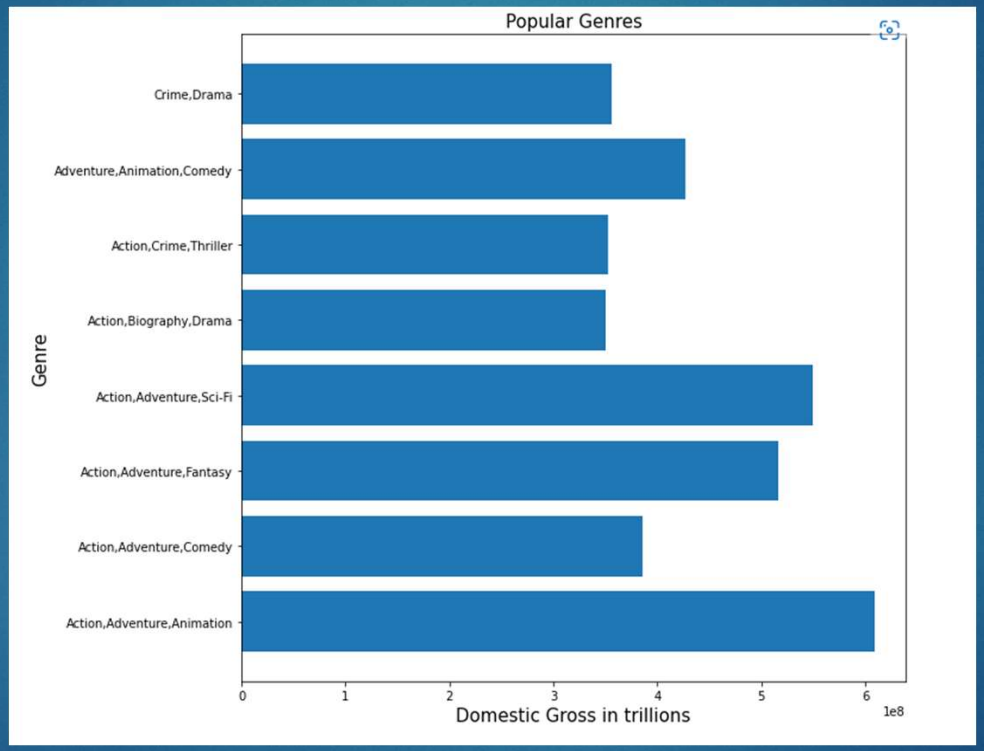
- Joining “imdb.title.basics” with my top 20 movies dataset, I was able to extract Genre’s data. Which formed my **recommendation 1-Top Genre**.
- Employing .isin function to extract data relating to the top 20 movies, which include “tconst” and “Primary Title”
- Merging 'imdb.title.crew.csv’ to newly extracted data table provided with Directors and Writers coded names.
- Again, left merging 'imdb.name.basics.csv’. On “directors” and “nconst” as respective indexes. I obtained the corresponding names of the coded directors, by using .dropna(subset) method. Hence forming my **recommendation 2 – Top Directors list.**
- getting to my third recommendation was a bit of work. I extracted data from my top 20 movie list & “imdb.title.principals.csv” by left merging on “ncont” to get an individual list of writers forming my **recommendation 3- Top writers**

Recommendation -1. Genre

Microsoft should aim to produce a movie within following genres. These movies have been performing very well at the box office.

POPULAR GENRES	
Crime	Drama
Animation	Adventure
Action	Sci-Fi
Fantasy	Thriller

Popular Genres



Recommendation -2. Director

- Microsoft should produce movies with either of following Directors.

Colin Trevorrow	Clint Eastwood	James Wan	J.A. Bayona
Patty Jenkins	Jake Kasdan	Brad Bird	
Tim Miller	Joss Whedon	Gareth Edwards	
Ryan Coogler	Rian Johnson	James Gunn	

Recommendation -3. Writer

- Microsoft should get the stories written by either of following Writers or combination of writers

1	Joe Robert Cole	16	John Knoll
2	Stan Lee	17	Gary Whitta
3	Jack Kirby	18	Victoria Strouse
4	Jack Kirby	19	Bob Peterson
5	Christopher Markus	20	Stan Lee
6	Stephen McFeely	21	Jack Kirby
7	Stan Lee	22	Joe Simon
8	Rick Jaffa	23	Jim Starlin
9	Amanda Silver	24	Derek Connolly
10	Derek Connolly	25	Colin Trevorrow
11	Michael Crichton	26	Michael Crichton
12	George Lucas	27	Harry G. Peter
13	George Lucas	28	Allan Heinberg
14	Chris Weitz	29	Zack Snyder
15	Tony Gilroy	30	Jason Fuchs

Results Recommendation -3. Writer

31	William Moulton Marston	46	Cinco Paul
32	Jack Kirby	47	Ken Daurio
33	Christopher Markus	48	Brian Lynch
34	Stephen McFeely	49	Rhett Reese
35	Joe Simon	50	Paul Wernick
36	Chris Van Allsburg	51	Dylan Schaffer
37	Chris McKenna	52	Chris Morgan
38	Erik Sommers	53	Gary Scott Thompson
39	Scott Rosenberg	54	Jason Hall
40	Jeff Pinkner	55	Chris Kyle
41	Jim Starlin	56	Scott McEwen
42	Dan Abnett	57	Jim DeFelice
43	Andy Lanning		
44	Steve Englehart		
45	Steve Gan		

Email: shradha.waddepalli@gmail.com
GitHub: @shrwad