Exploratory Data Analysis (EDA) Proposal for Top 10000 Popular Movies

1. Introduction

This project is one of the T5 Data Science BootCamp requirements is called the Exploratory Data Analysis (EDA).
Recommendation systems are used everywhere these days. Netflix , Amazon Prime , YouTube , Online shopping sites etc.
This project aimed to provides suggestions by movies Recommendation system. To help people find the most popular movies according to people's ratings. by implement an exploratory data analysis on the dataset for 10000 Popular movies based on the TMDB ratings.

2. Dataset

The dataset contains 10,000 movies divided into 13 columns: ID, language, title, popularity, release date, average rating, number of votes, genre revenue, runtime and tagline.
Datasets like this are a way to start working on the recommendations system and we can extract meaningful data. The dataset was generated from the official API installed by TMDB

3. Source

Data provided by Kaggle has been used in this project
Website:
(https://www.kaggle.com/omkarborikar/top-10000-popular-movies

4. Features

- (10000) rows of movies
- (13) columns :

Column Name	Description
id	Every movie has its unique ID.
original_language	There are total 44 languages present in this column. Total 7771 movies with 'English' as original language. Values in this column are ISO 639-1 codes of languages. I.e 'en' for 'English' , 'hi' for 'Hindi' etc.
original_title	Title of the movie.
popularity	Popularity of movie. Bigger the number , higher the popularity.
release_date	Release date of the movie. If release date is not present for any movie , then that movie is not released yet.
vote_average	Average of rating/vote for the movie.
vote_count	Number of ratings/vote recorded for the movie.
genre	Genre of the movie.
overview	Brief description of movie in string format.
revenue	Revenue of Movie
runtime	Runtime of movie in minutes.
tagline	Tagline of the movie

5. Tools

- Technologies: Jupyter Notebook, Python, SQL and SQLlite.
- Libraries: Pandas, NumPy, Matplotlib and Seaborn.

6. Questions

1. Which type of movies are more popular?
2. What are the most popular languages in movies?
3. What are the Top 10 movies with maximum vote average?
4. Display analysis of the revenue trend over the decades.
5. Display analysis of the trend of movie runtime.
6. Are specific genres associated with higher revenues?

Shaima alzahrani
Shaima.alabedi@gmail.com