DECI project TMDB 5000 movie

Table of Contents

Introduction

Forecasting a movie's success before its release is challenging, given that even high-budget films, sometimes exceeding $100 million, can falter. While studios like Pixar may have consistent hits, there's no fail-safe formula for success. Film enthusiasts may prioritize different aspects, such as commercial viability or critical acclaim. However, by analyzing data on plot, cast, crew, budget, and revenues from thousands of films, researchers can uncover patterns that shed light on what drives success in the industry. This exploration marks the beginning of an intriguing journey into understanding the complexities of filmmaking.

Questions and Exploratory Data Analysis

What kinds of properties are associated with movies that have high revenues?
Answer in code

What is the movie with the highst revenue?
Avatar

Is the revenue always higher than the budget?
No

Does runtime differs in revenue?
Yes, movies under 200 min have more revenue

who is the director responsible for making a movie have high revenue?
Steven Spielberg

What kinds of properties are associated with movies that have high revenues?
year , genre , budget and runtime

Data Wrangling

**NAN values was fill with 'none',
duplicated values was dropped
and I got rid of unneeded columns**

Conclusions

based on data analysis, every year have a popular genre and the movie 'Avatar' has the highest revenue. we figured out that higher budget doesn't always mean higher revenue also movie runtime should be between 0 to 200 so it gets high revenue. the director that most of his movies gets high revenue is 'Steven Spielberg'. to summarize everything properties associated with movies that have high revenue is budget, genre and runtime and that is based on release year and you must use a successful director like 'Steven Spielberg' and in data wrangling the limitation was duplicated values and missing values