

MICROSOFT MOVIE STUDIO PROJECT

1. OVERVIEW

- All big companies are creating original video content whereas microsoft has created none hence lagging behind.
- They want to create a new movie studio but they don't know anything about creating movies

2. BUSINESS UNDERSTANDING

2.1 BUSINESS PROBLEM

- As a data scientist I'm tasked to process data from different film datasets in order to determine what type of films are currently doing the best
- This will be achieved by determining what is required to make a successful film

2.2 PROBLEM STATEMENT

- The problem statement is to determine what type of films microsoft is to create in regards to films that are currently successful
- This would be achieved by determining:
 - ☐ Which genre of movies are the most profitable
 - ☐ Influence of certain actors and directors in success of films and number of films featured
 - ☐ Amount of budget required to produce a successful film.
 - ☐ Best time to release a film

3. DATA UNDERSTANDING

3.1 DATA COLLECTION

- Data is collected from the datasets containing the movie's budget and finances, people's record and bankability. Has various information such as trending movies, actors and the total gross income.

3.2 DATA DESCRIPTION

<u>Column name</u>	<u>Description</u>
movie_year	Movie production year
director	Movie director names
main_actor_1	Movie actor name
budget	Movie budget
worldwide	Movie total gross
genre_1	Movie type

3.3 SAMPLING STRATEGY

- Collecting data from top 10 production companies
- Determining correlation between numeric values and average budget of films

4. DATA PREPARATION

4.1 SELECTING DATA

- We'll use data from the top 10 production companies with it's related columns

4.2 DATA CLEANING

- Done to ensure validity ,reliability, accuracy, uniformity and consistency of the data

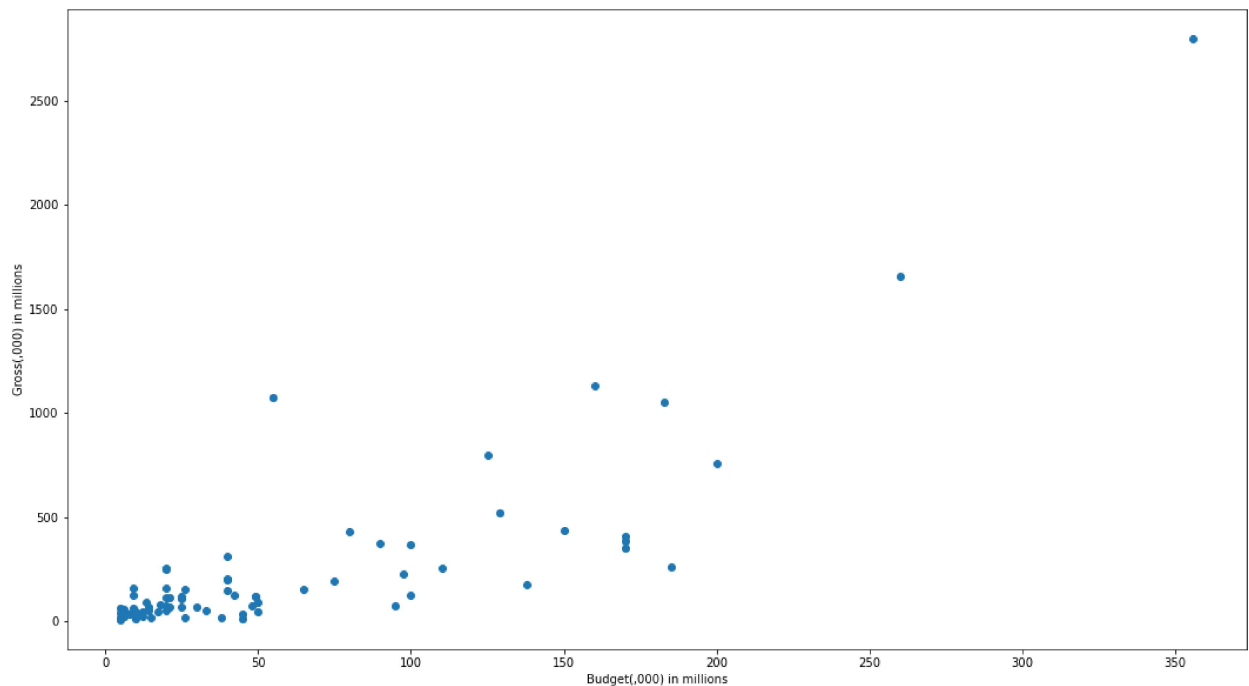
5. DATA ANALYSIS

5.1 EXPLORATORY DATA ANALYSIS

5.1.1 UNIVARIATE DATA ANALYSIS

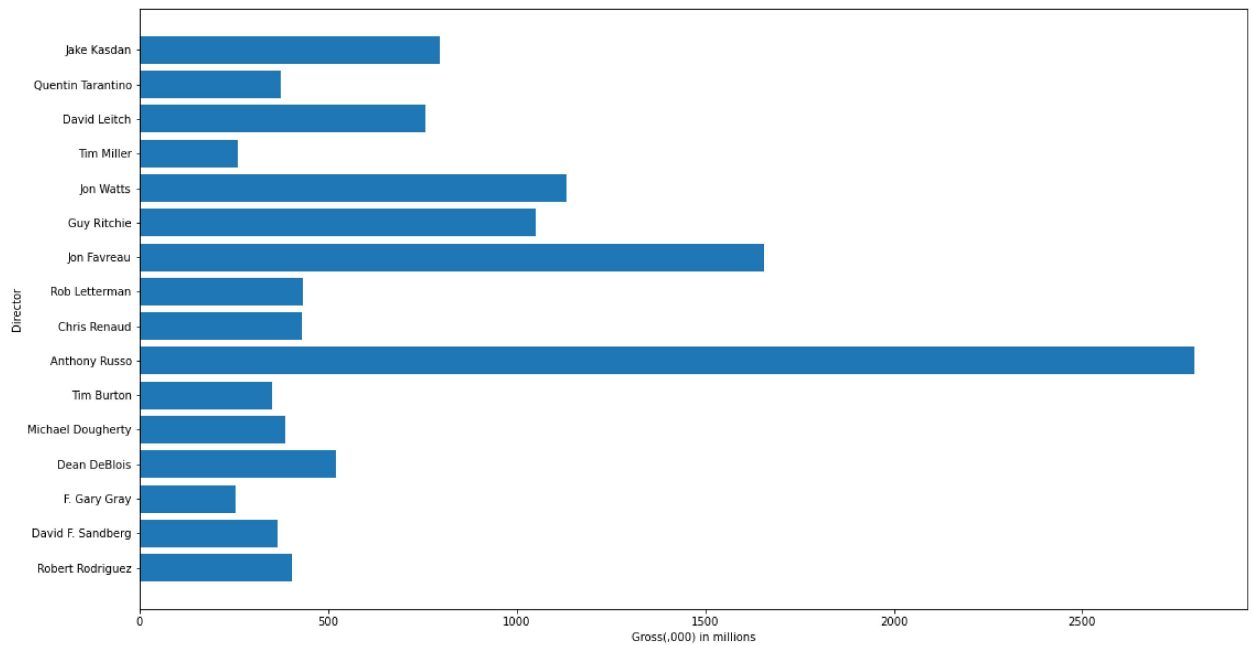
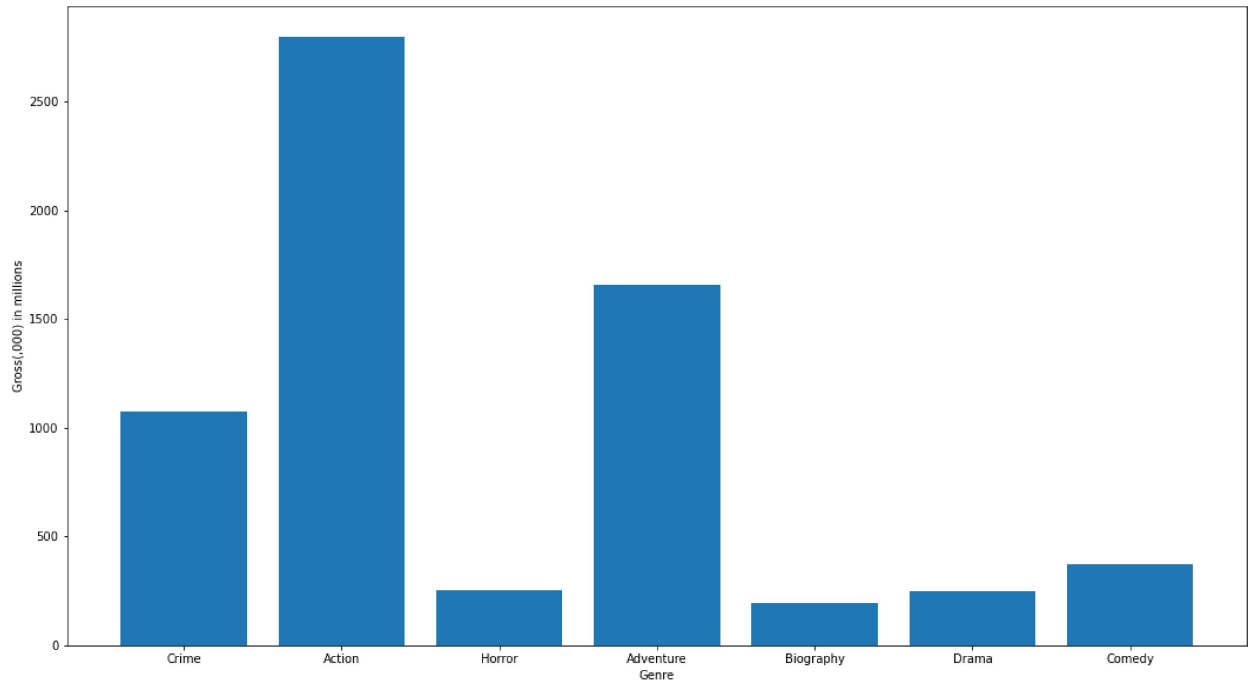
a) Numerical Data

Based on the data the numerical data used was budget and gross to determine the correlation between the two variables and find the budget estimate



b) Categorical Data

Based on the data the categorical data used was genre and director in relation to the gross. This was used to determine the most successful genres and directors



c) Summary statistics

In summary we can conclude that action is the most successful genre and Anthony Russo is the most successful director

d) Univariate analysis recommendation

It's recommended that a budget is stipulated within the mean and either of the most successful genres and directors are selected

6. CONCLUSION

In conclusion:

Budget and gross have a weak correlation hence it's not safe to conclude that a higher budget reflects to a higher gross income
Action is the most successful genre by a mile. Adventure and crime films are also fairly successful

Anthony Russo Is the most successful film maker. Jon Favreau and Jon Watts are also fairly successful

7. RECOMMENDATION

It's recommended:

Having a budget within the mean budget
Production of action adventure or crime films
Hiring either of the three most performing directors