



MORINGA

Discover · Grow · Transform

PHASE 1 PROJECT

PRESENTED BY: CYNTHIA KARUGA

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

DATA ANALYSIS

BY PANDAS

PROBLEM STATEMENT

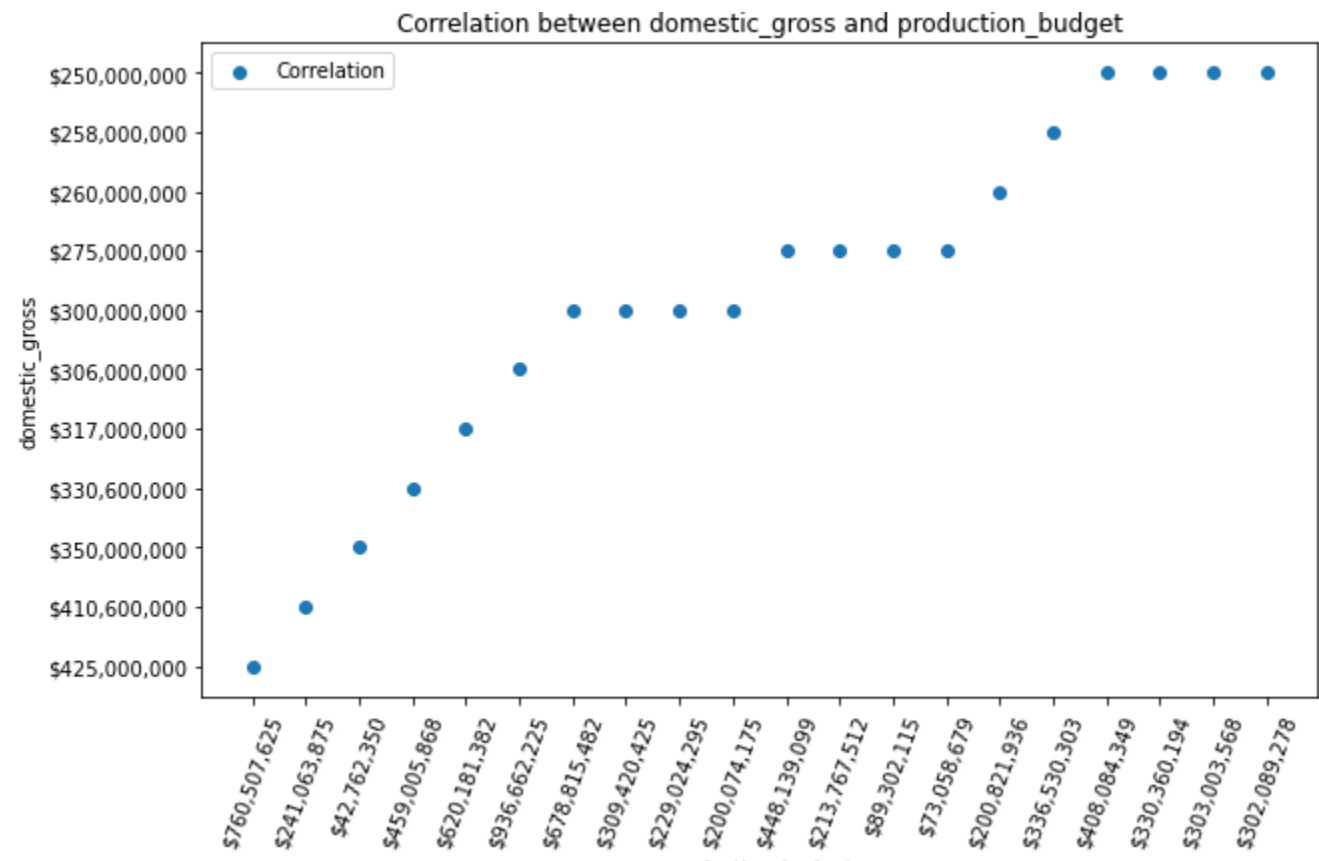
Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

DATA USED

- ▶ [tn.movie_budgets.csv](#)
- ▶ [title.basics.csv](#)
- ▶ [title.ratings.csv](#)

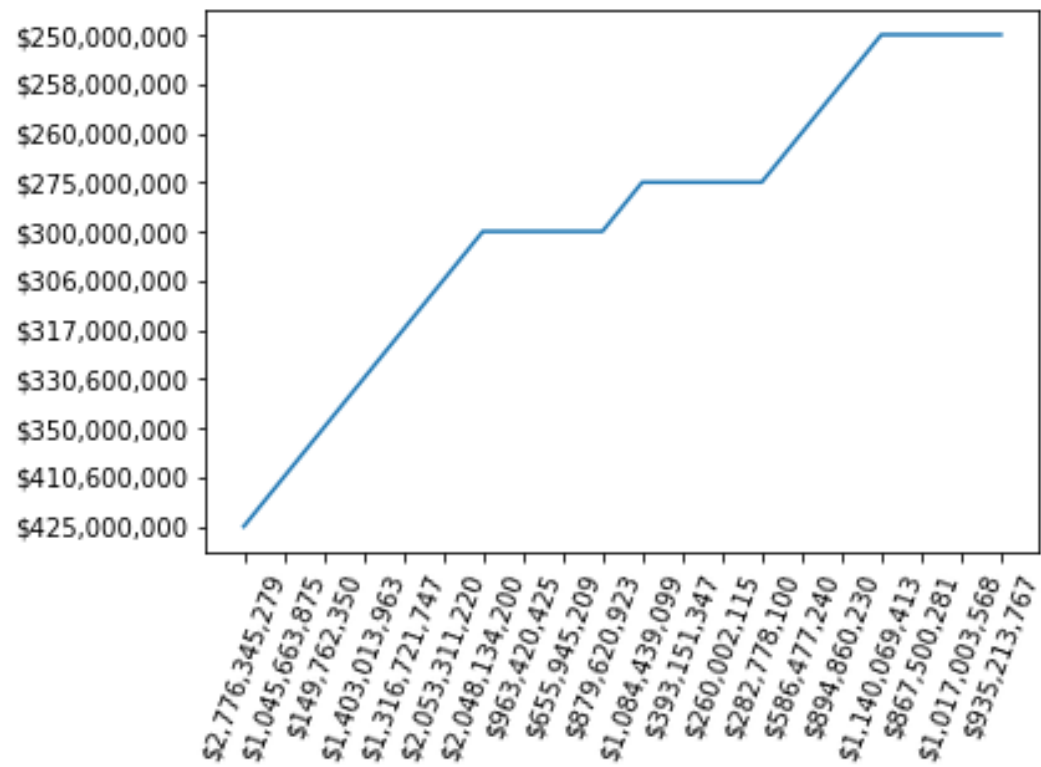
FINDINGS

1.A positive correlation exists between Correlation between domestic gross and production budget



- ▶ A sample size was used rather than the whole population
- ▶ Therefore, as the budget increases so does the gross up to a certain point where it plateaus then picks again

2.A positive correlation exists between Correlation between foreign gross and production budget



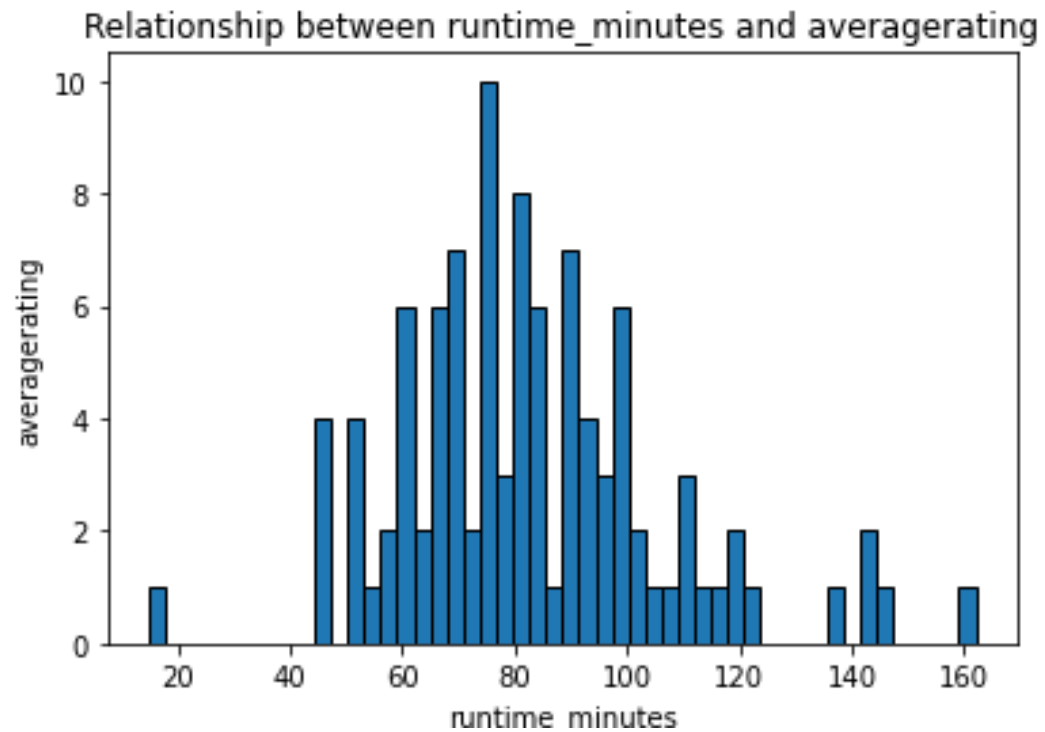
- ▶ Both domestic and gross product behave the same with regards to budget
- ▶ They increase then plateau at a certain level

3.Documentaries have the highest ratings

genres	averagerating	r
Crime,Documentary	10.0	1
Documentary	10.0	2
Documentary	10.0	3
Documentary	10.0	4
Documentary	10.0	5
Documentary	10.0	6
Documentary	10.0	7
Documentary	10.0	8
Comedy,Drama	10.0	9

- ▶ Just a quick look into the sample data shows that Documentaries have the highest rating

4. A relationship does exist between runtime minutes and average rating



- ▶ The data is centralized at the middle meaning an a movie with an average runtime is most likely to receive a high rating

Conclusion

- ▶ Sample data was used rather the population data
- ▶ This was a general view of data but calculation of numerical values e.g median and correlation could be useful for clarity

Recommendations

- ▶ Microsoft should invest more on documentaries
- ▶ The running time of the movies is important and should be kept at an 'average'
- ▶ The budget put forth should be an indication of the gross they are most likely to receive in the short term