DATA ANALYSIS OF FILMS AT THE BOX OFFICE

OBJECTIVES

❖ To find out the top films at the box office

❖ To find out the film genres with the highest ratings and those with the lowest ratings

❖ To find out the relationship between runtime and the rating of the films

❖ To find out the relationship between runtime minutes and number of votes

❖ To find out the relationship between the domestic gross and foreign gross

PROJECT DATA

- In this project, two datasets were used; the bom movie gross data and the im.db sql database.

- In the sql database, the focus was on the movie basics and movie ratings tables.

METHODS USED

- To analyse the two datasets;first, the data was cleaned then exploratory data analysis was perfomed .
- The analysis was done using python and sql

Business understanding

The complex relationship between movie-related data and movie box-office revenues, plus the increasing volume of data in online movie databases, pose challenges for their effective analysis.

This analysis will take into considerations two datasets to try and understand the types of films that are currently doing the best at the box office in order to decide what type of films microsoft should create.

Data understanding

- This project considered two datasets.
- The bom movie gross dataset was a csv file with five columns; title, studio, foreign gross, domestic gross and year.
- The im.db dataset was an sql database . Two tables from the database, movie basics and movie ratings, were used since they were the most relevant tables for the research

❖ Data analysis
The top films at the box office

Sorting the bom movie gross data using the foreign gross, the top 5 movies were;

- Harry Potter and the Deathly Hallows Part 2

- Avengers: Age of Ultron

- Marvel's The Avengers

- Jurassic World: Fallen Kingdom

- Frozen

❖ The top films at the box office

- Sorting the bom movie gross data using the domestic gross, the top 5 movies were;
- Star Wars: The Force Awakens
- Black Panther
- Avengers: Infinity War
- Jurassic World
- Marvel's The Avengers.

❖ The top films at the box office

Using the sql database, the top 5 rated movies were;

- The Dark Knight: The Ballad of the N Word

- The Paternal Bond: Barbary Macaques

- I Was Born Yesterday!

- Ellis Island

- Requiem voor een Boom

❖ Movie genres with the highest ratings

The top rated movies were of genre;

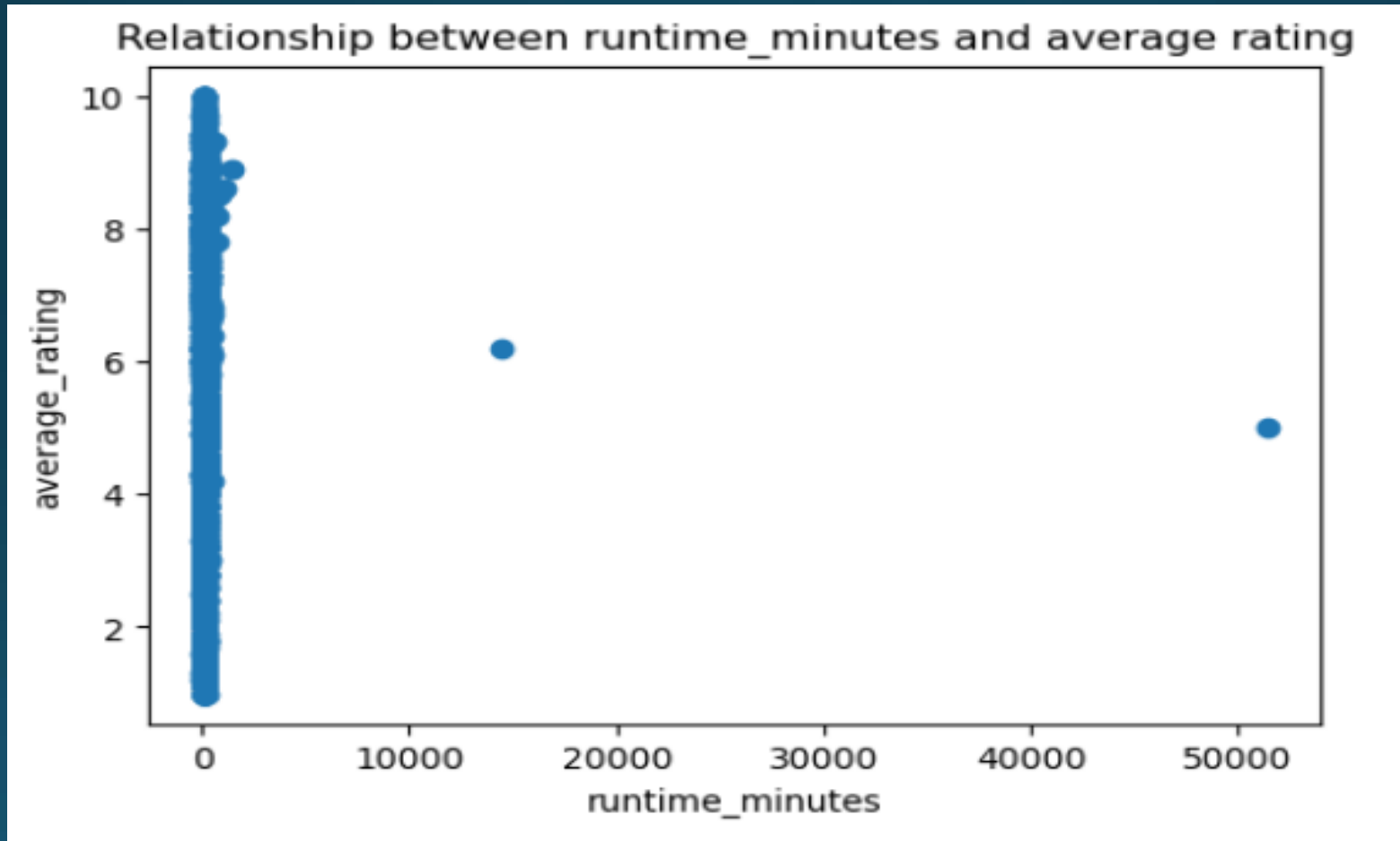- Comedy

- Drama

- Documentary

- History

Movie genres with the lowest ratings

- The poorly rated movies were of genre;
- Documentary
- Drama
- Horror
- Thriller
- Comedy
- Family .

Relationship between runtime minutes and the rating of the films

- The correlation between runtime minutes and the average rating was -0.007076, which shows that they have a negative relationship(but not a strong negative relationship since the correlation is close to 0), that is, increase in runtime minutes lowers the rating.
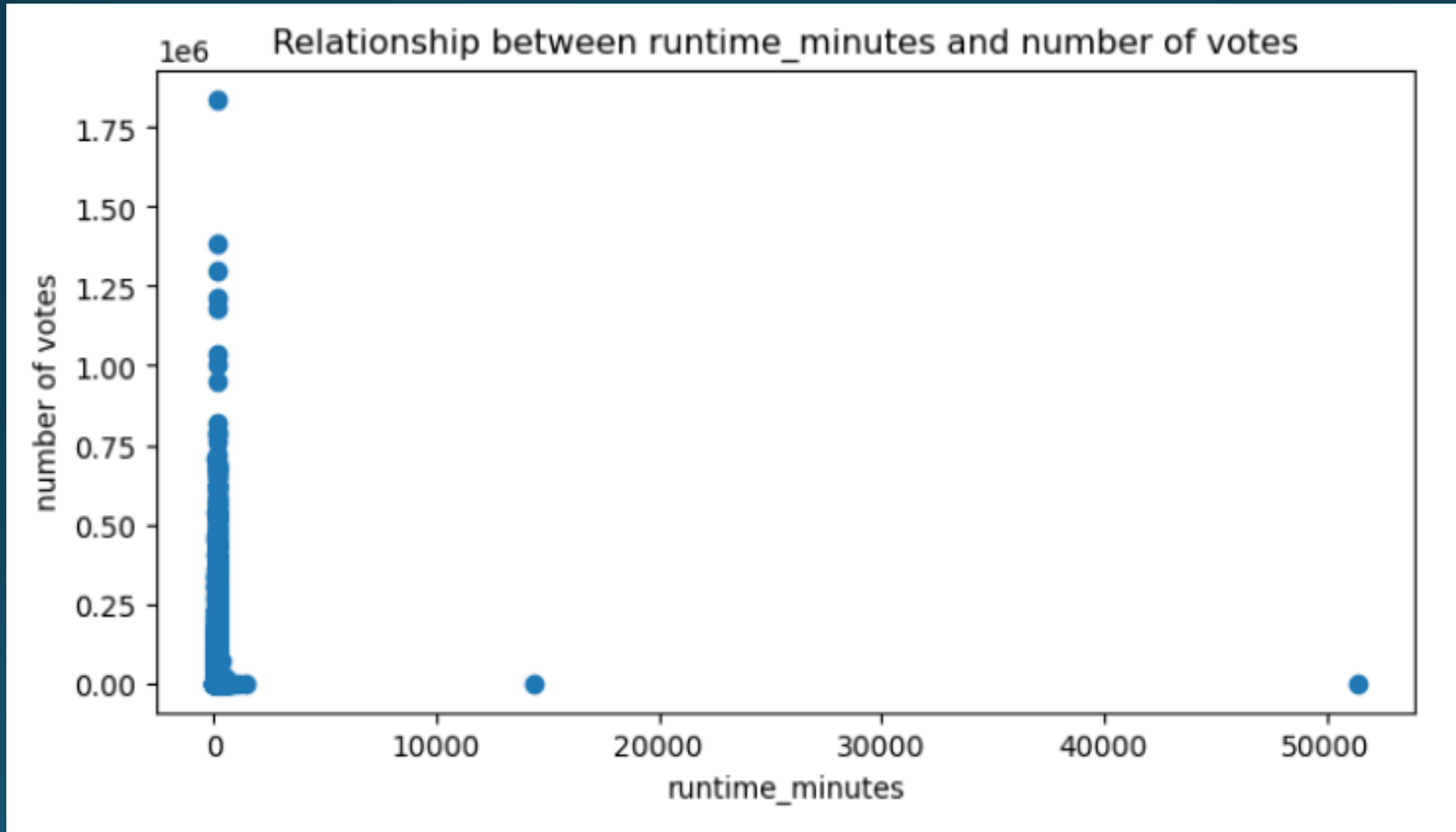
- In the next slide, we can visualize this.

Visualization of the relationship between runtime minutes and rating

Relationship between runtime_minutes and average rating

Relationship between runtime minutes and the number of votes

- The correlation between runtime and number of votes is 0.012428, which shows that they have a positive relationship, though not a strong relationship since the correlation is close to 0, implying that increase in runtime minutes leads to increase in the number of votes.
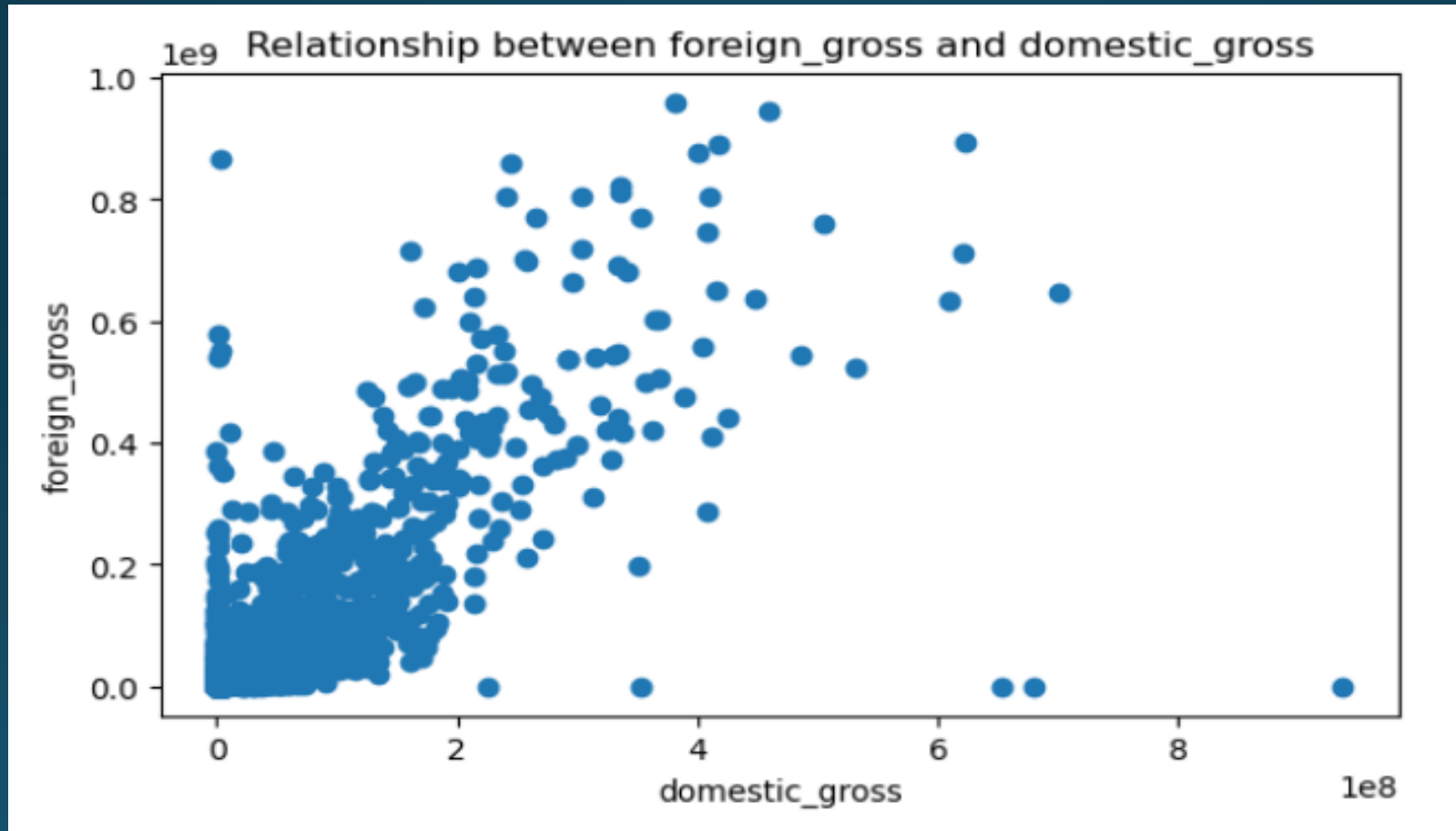
- We can visualize this in the next slide.

Visualization of the relationship between runtime and number of votes

Relationship between the domestic gross and foreign gross

- The correlation between domestic gross and foreign gross was 0.768451 which indicates that the two variables have a strong relationship and that they are positively correlated ,that is, increase in domestic gross leads to increase in foreign gross and vice versa.

- We can visualize this in the next slide.

Visualization of the relationship between domestic and foreign gross

❖ Conclusion Recommendations

- The company should consider creating more movies of genre ; comedy, drama, documentary and history since the top rated movies were of these genres.

- The company should consider creating few movies of genre; horror, thriller and family since the least rated movies were of these genres.

- The company should make movies that are not too long nor too short since increase in runtime minutes increases number of votes but also decreases the average rating.

Next steps

- From this analysis, it is noted that most of the top movies were recorded by BV and Uni studios. More research on both these studios might be necessary so as to know why their movies came up top.

Thankyou for your time and attention.

If there are any questions, feel free to ask.

Contact information;

NAME; Sharon Wanjiku Kimani

LINKEDIN PROFILE; Sharon Kimani

EMAIL; kimanisharon23@gmail.com