DATA SCIENCE PHASE 1

INDIVIDUAL PROJECT

INTRODUCTION.

- Main goal is to create the best selling movies in the industry.

Factors

❖ **Genre Selection**

❖ **Audience Preference**

❖ **Budget Allocation**

❖ **Quality Assurance**:

Data sets available

- Two CSV files

   movie_budgets.CSV

  bom.movie_gross.Csv

- Sqlite database

   'im.db

PARSING THROUGH THE DATA

Columns we need:

Movie_budgets.csv

✓ Production_budget

✓ Domestic_gross

✓ Id

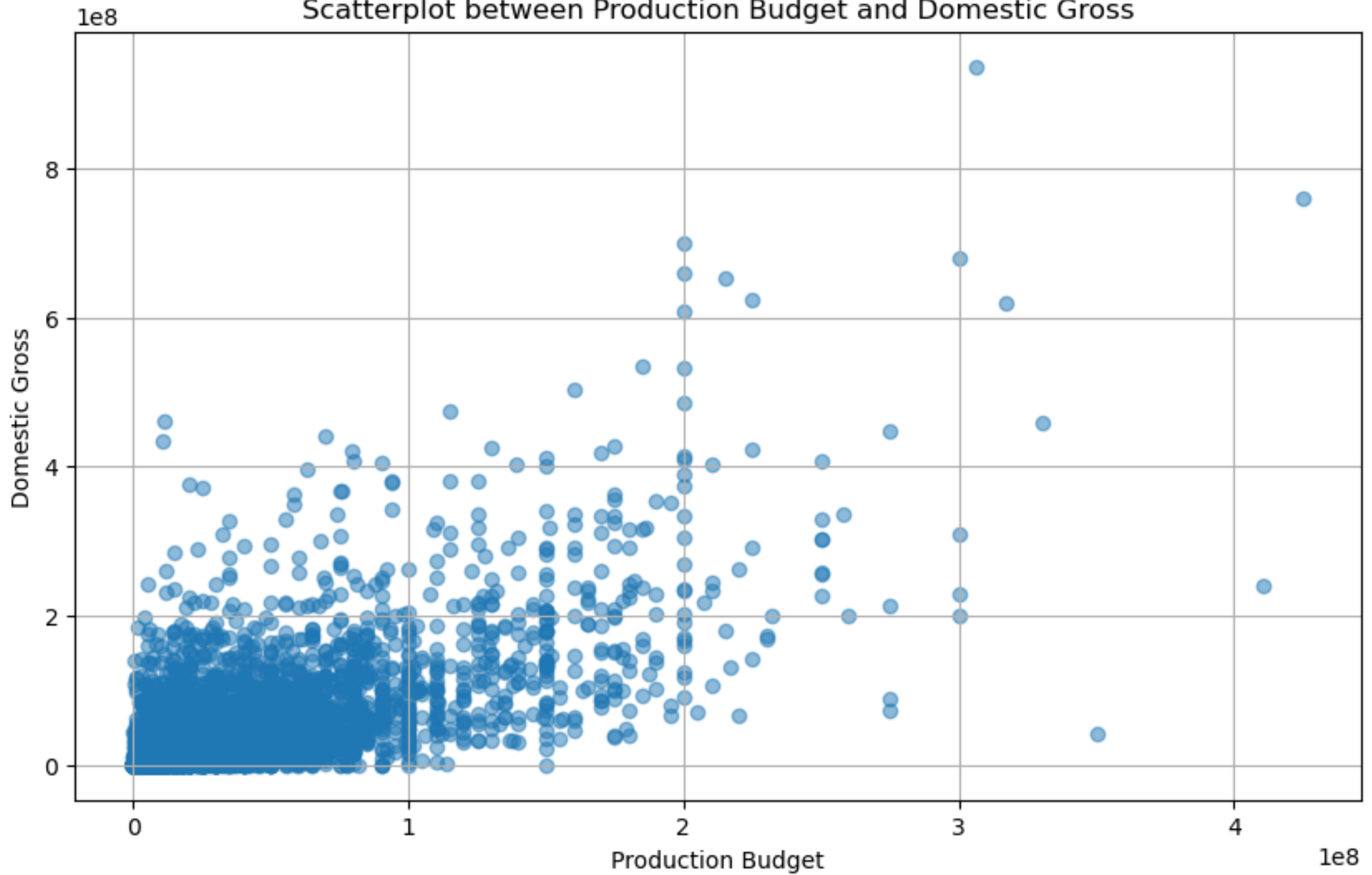Bom.movie_gross

✓ Domestic _gross

Im,db –tables we need

✓ Movie_akas

✓ Writers

✓ Persons

✓ Movie_basics

✓ movie_ratings

✓ principals

Scatterplot between Production Budget and Domestic Gross

Expenses vs profit projected.

- A correlation coefficient of 0.685682 indicates a moderately positive correlation between the two variables.

- A positive covariance suggests that when production_budget increases, domestic_gross also tends to increase and vice-versa.

- strong relationship between the production budget and domestic income.

Comparison between two data sets

Movie_budgets

- The average domestic_gross is 28745850 dollars

Bom_df

- The average production_budget of the movie is:3.158776e+07 translating into 3,1587,760 dollars

- The domestic_gross income is:4.187333e+07 translating to 4,1873,330 dollars.

Quality Assurance.

Average running time of the best movie:

The runtime of the best rated movie is empty hence we use the second best movie has a runtime of is 70.0.

Average running time of the least rated movie is 72.0

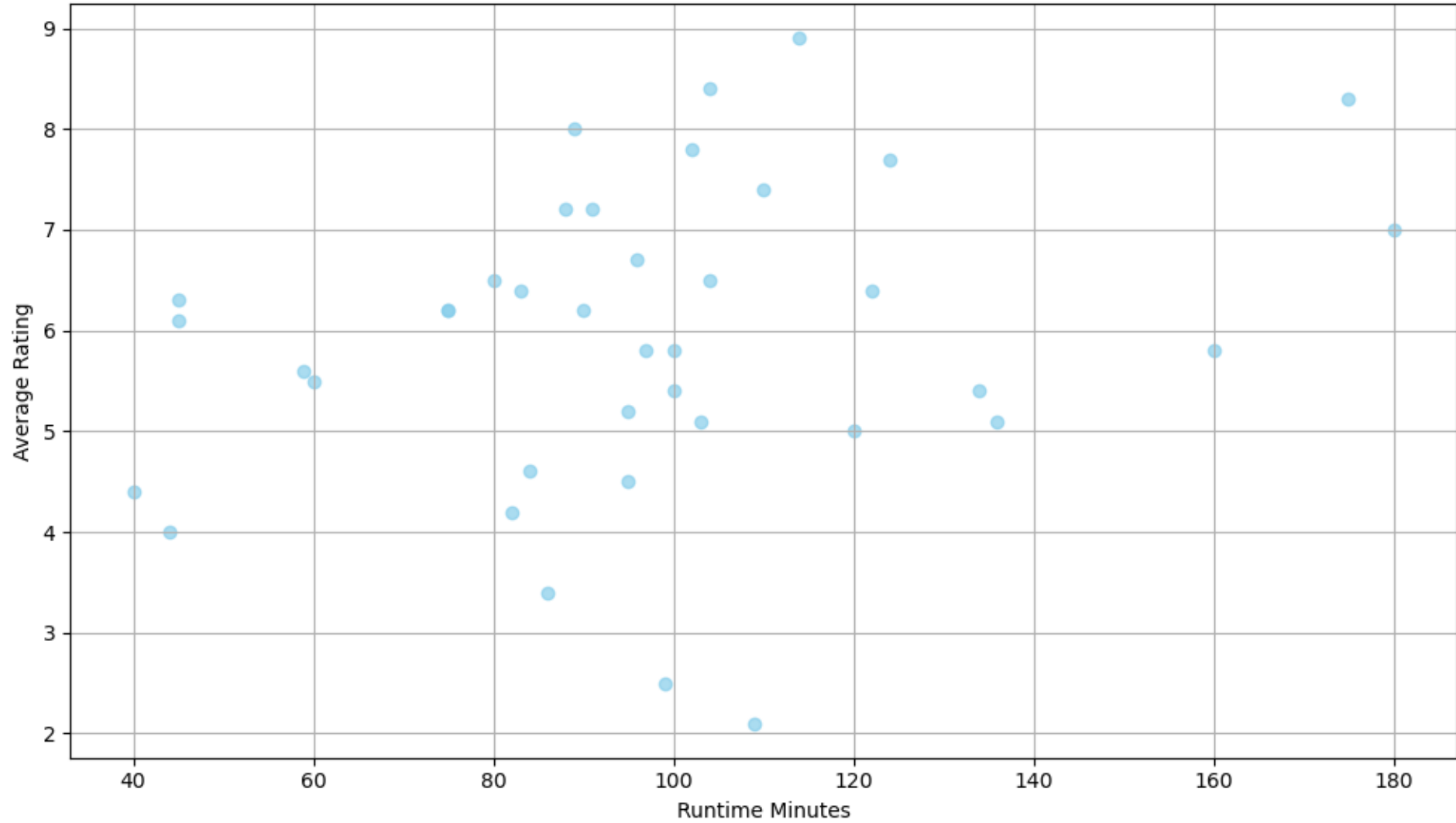| runtime_minutes	
count	114405.000000
mean	86.187247
std	166.360590
min	1.000000
25%	70.000000
50%	87.000000
75%	99.000000
max	51420.000000

Scatter Plot of Average Rating vs Runtime Minutes

Genre Selection.

- **The most watched movies are Documentaries.**

- Movie Title: Exteriores: Mulheres Brasileiras na Diplomacia, Genres: Documentary

- Movie Title: Freeing Bernie Baran, Genres: Crime,Documentary

- Movie Title: Exteriores: Mulheres Brasileiras na Diplomacia, Genres: Documentary

- Movie Title: Foreign - Brazilian Women in Diplomacy, Genres: Documentary

- Movie Title: Hercule contre Hermès, Genres: Documentar

- y Movie Title: Revolution Food, Genres: Documentary

- Movie Title: Dog Days in the Heartland, Genres: Dram

- a Movie Title: Requiem voor een Boom, Genres: Documentary

- Movie Title: Atlas Mountain: Barbary Macaques - Childcaring Is the Man's Job, Genres: Documentary

- The least rated movies are Romance,Horror and comedy.

Movie Crew.

- The best producer is Maria Demeshkina Peek .

- The two best writers are:Brian Baucum and Loreto Di Cesare.

Geographical region.

```
region num_movies
0 None     53293
1 US       51490
2 XWW      18467
3 RU       13817
4 DE       11634
5 FR       10990
6 ES        9007
7 GB        8942
8 CA        8871
9 PL        8691
```

- The largest target audience is from USA

Recommedations

- Given that Documentaries and Drama are the most produced movies, Microsoft should focus on these genre

- Given that Brian Baucum and Loreto Di Cesare are identified as the best writers and Maria Demeshkina Peek as the best producer, Microsoft should continue to collaborate with these talented individuals

- Microsoft may consider analyzing the runtime preferences of the audience and adjusting their movie lengths accordingly to optimize viewer engagement.s.

- With a correlation coefficient of 0.685682 and a positive covariance between production_budget and domestic_gross, it's evident that there's a moderately

Limitations of the project.

- Time constraints

. As a result, the analysis did not cover all relevant aspects of the problem thus not able to provide a high-level overview.

- The findings and conclusions of the analysis may be limited in their generalizability to broader populations or contexts.

For example, Analysis is based on a specific dataset and sample, the results is just a representative of the entire population.

Thankyou!

Winfred Kinya Bundi.

Winfred.kinya@student.moringaschool.com