Microsoft Data Analysis

IN THE SLIDES ABOVE I AM PRESENTING METHODS I USED TO ANALYSIS DATA FOR MICROSOFT

Microsoft

THEY WANTED TO START MOVIE PRODUCTION. SO THEY WANTED ANALYSIS OF THE FOLLOWING FILES TO GET A STEPPING STONE WHERE TO START FROM

The files provided for data analysis :

- 1.RT.REVIEWS.TSV
- 2. RT.MOVIES.INFO.TSV
- 3. MOVIES.CSV
- 4. MOVIES_BUDGET.CSV
- 5. MOVIE_GROSS

I choose the following files to start my work

- RT.REVIEWS.TSV
- MOVIES.CSV
- MOVIES-BUDGET.CSV

WORK PROCEDURE

- 1) I strated my work with converting the files into dataframes.
- eg movie_budget.csv to movie_budgetcsv_df

- A data frame is a structure that organizes data into a two dimensional tasble of rows and columns.

- From there I printed the output to check my organized structure

PROCEDURE TWO:

- I conducted data cleaning to erase
- -duplicates
- - remove the incomplete data types
- -remove the missing files.
- -remove undesired formats

- - this enhances data accuracy while analyzing data.

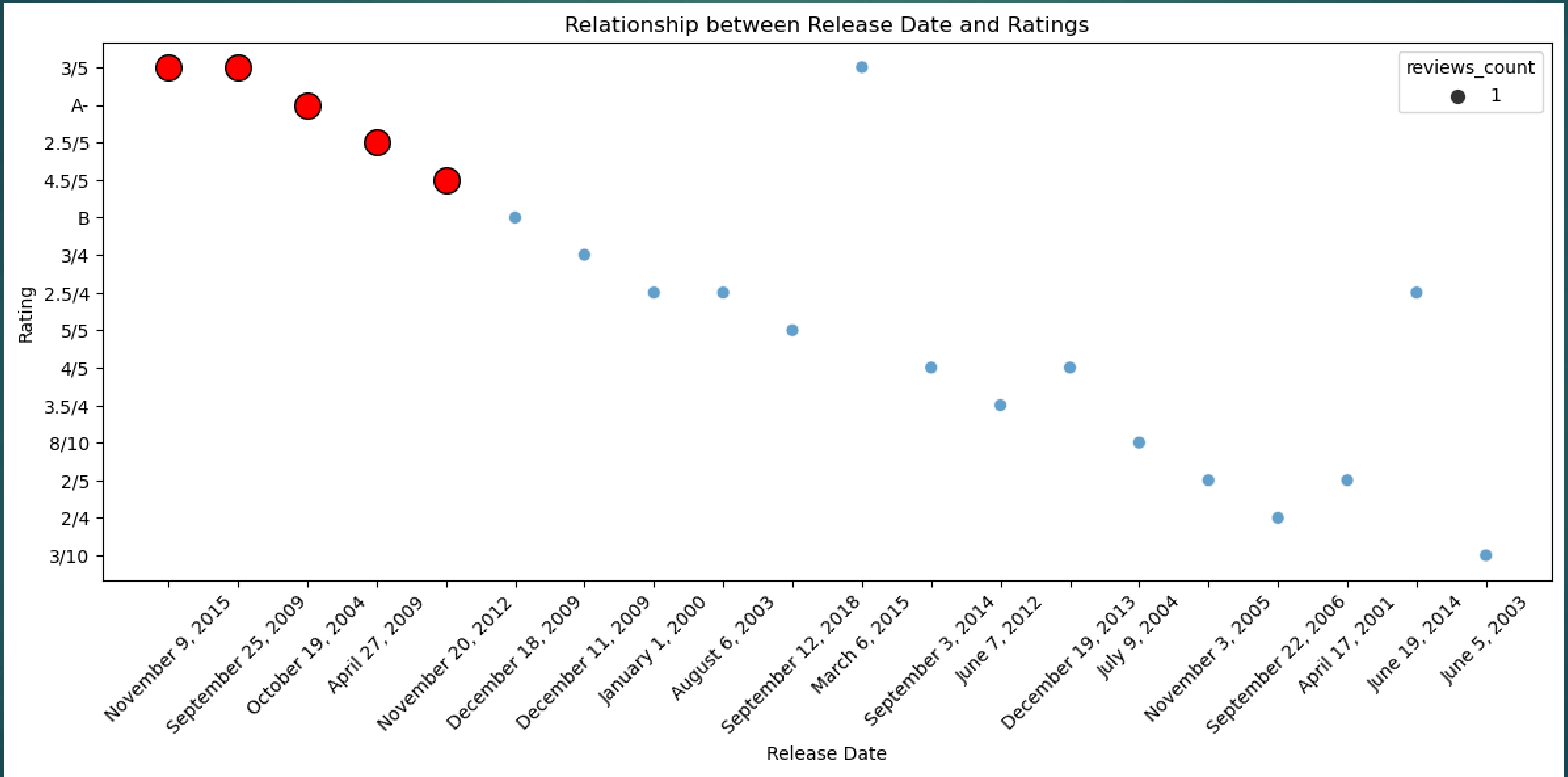
QUESTIONS

- I set my work into questions for easy understanding and flow of work.
- I used the rt_reviews_df
- It contained columns like;
- ‘publisher’, ’ratings,’ ‘release data’ and ‘reviews’
- :The question answered the below arguments:

- I used the question to calculate the reviews of each publisher ,

- from there i used data visualization by using a scatter graph to show the relationship between release date and rating . the last question was counting the number of ratings. –

QUESTION ONE GRAPH

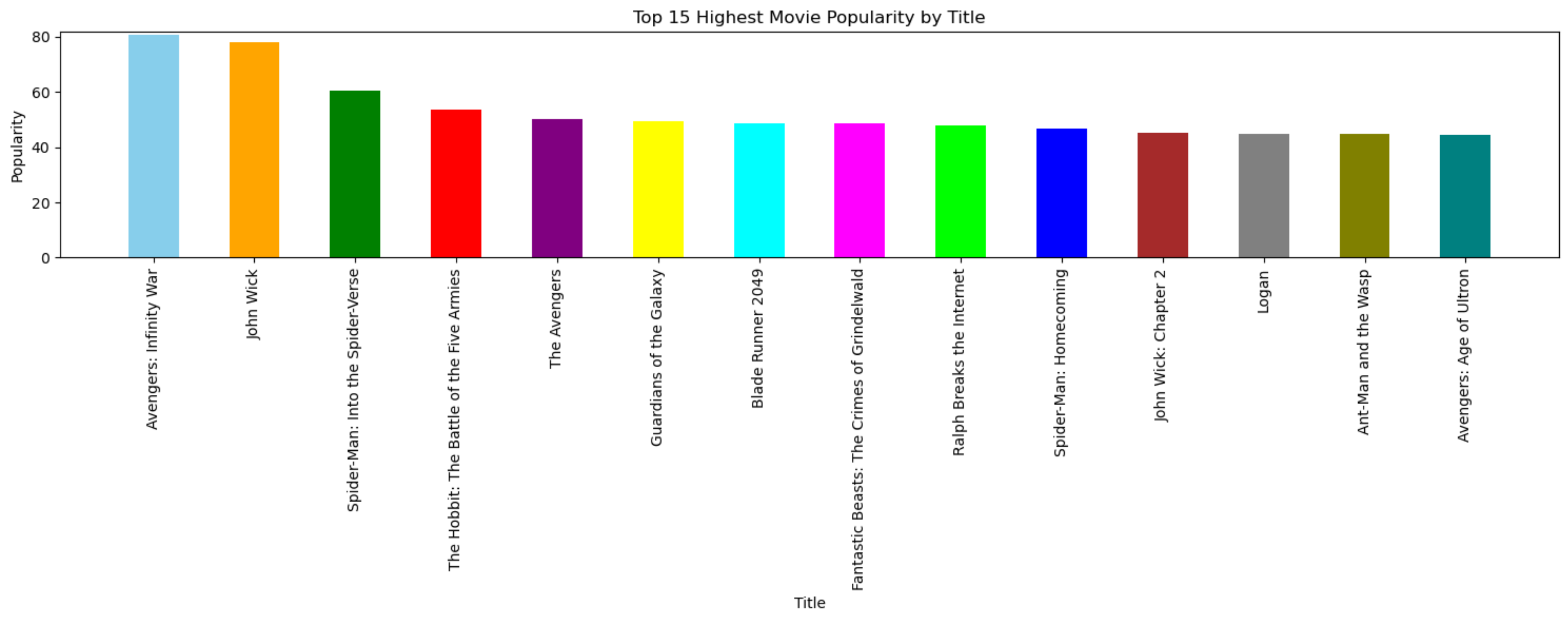


QUESTION TWO

- In question two I followed the same directory as question one
- I used movie_csv_df
- The dataframe consist columns eg: ‘popularity’, ‘genre_ids’, ‘title’

- The question answered :the below arguments:
- -check the movie title with the highest vote_count
- -check the movie title which was the most popular
- -checked the ratings of the movie title

QUEESTION TWO GRAPH



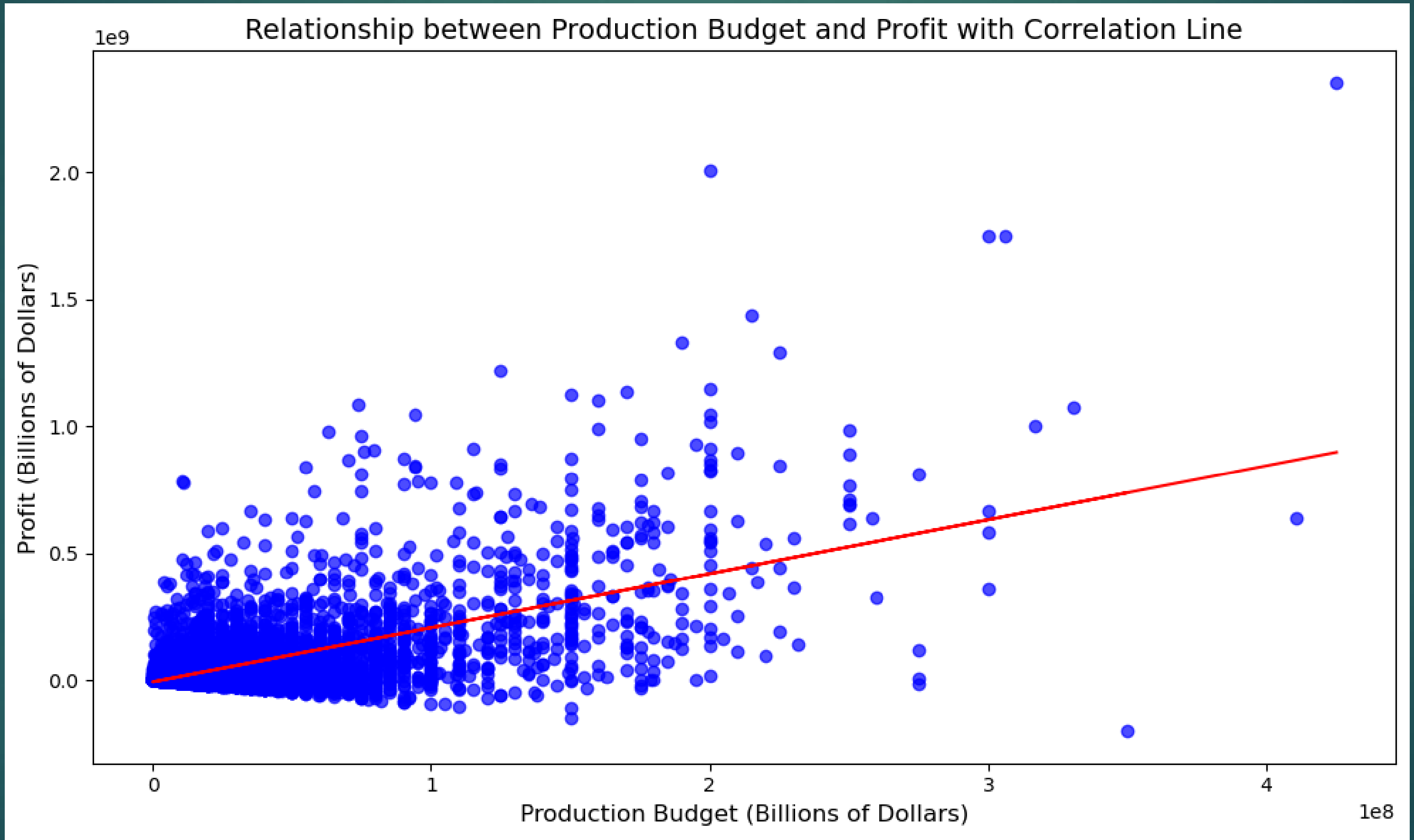
QUESTION THREE

- in the above question I used the movie_budgetcsv_df

- The dataframe consists of the below colums:
- - ‘worldwide_gross’,’ production_budget’ , release date

- The answered the below arguments:
- - calculate the profit margin
- - show the relationship between profit and budget
- - relationship between release date and budget.

QUESTION 3 GRAPH



RECCOMENDATIONS

- to release movies during the last three months of the year.
- to hire publishers with the high rating for the movies,
- to choose a movie title that will be popular and get a high vote_count
- i would advise they allocate a budget that is good enough to yeild high profits

DATA SCIENTIST DETAILS:

ESTHER MUKAMI NJAGI

LINKEDLN PROFILE: Esther Mukami