R and Power BI Project

In my R and Power BI assignment, the primary objective was to delve into the world of Hollywood movies and conduct a thorough analysis of their performance between 2007 and 2012. Armed with a dataset encompassing movie titles, genres, studios, profitability, and ratings sourced from [InformationIsBeautiful.net](https://public.tableau.com/app/sample-data/HollywoodsMostProfitableStories.csv), I embarked on a journey to unearth insights and present them through a dynamic Power BI dashboard. This narrative unfolds through the lens of my code, which involves loading, cleaning, and exploring the data in R, followed by the creation of a visually appealing dashboard in Power BI, catering to the client's specific requests and incorporating the company's distinctive brand colours.

**Step 1: Initial Exploratory Analysis**

**Explanation:**

**A computer screen shot of a computer

Description automatically generated**

* + The code reads the CSV file from the provided URL and stores it in the variable **df**.
  + The **View(df)** command allows you to visually inspect the data in a separate viewer.
  + The code installs and loads the **tidyverse** package, a collection of packages for data manipulation and visualization.
  + **str(df)** provides information about the structure of the data, including data types.

A screenshot of a computer

Description automatically generated

R display of the datasets

**Step 2: Clean Data**

**Explanation:**

**A screenshot of a computer code

Description automatically generated**

* + **colSums(is.na(df))** checks for missing values in each column and prints the sum.
  + **df <- na.omit(df)** or **df <- df %>% drop\_na()** removes rows with missing values.
  + The last line confirms that missing values have been successfully removed.

**Step 3: Exploratory Data Analysis**

**A computer screen with text

Description automatically generated**

* + **summary(df)** provides summary statistics for the dataset.
  + The scatterplot and bar chart are created using the **ggplot2** package, visualizing relationships between variables.

A close-up of a computer screen

Description automatically generated

Summary Statistics for the datasets

A graph with black dots

Description automatically generated

The scatter plot (Rotten tomatoes scores) and Bar chart (Number of films each year) for the data set

A black text with black text

Description automatically generated with medium confidence

**A graph of a number of columns

Description automatically generated**

**Step 4: Export Data**

* + **write.csv (df, "clean\_df.csv")** exports the cleaned data to a CSV file named "clean\_df.csv."

**Power BI Dashboard Instructions**

* Import the "clean\_df.csv" file into Power BI.
* Create visualizations based on the client's requirements, using the brand colours (blue, green, brown).
* Include the following charts in the Power BI Dashboard:
  + Average Rotten Tomatoes ratings of each genre
  + Number of movies produced per year.
  + Audience scores for each film
  + Profitability per studio
  + Worldwide gross per genre

A screenshot of a graph

Description automatically generated

So, this is the dashboard. I had to use the colour green, brown, and blue because that’s the colours of our client’s logo.