Iqra Irshad

Data Cleaning and removing outliers in R Data Visualisation and Building dashboard in PowerBi

R Script and PowerBi

Assignment 3

Table of Contents

1. **R Script2**

Step 12

Step 2,33

Step 4,54

Step 6,75

Step 8,96

Step 10,117

Step 12,138

Step 14,159

Step 16,1710

Step 1811

1. **EER3**

Step 1,212

Step 3,413

Step 5,614

Step 7,815

Step 9,1016

Step 1117

**R Script**

**Task 1: Data Cleaning**

* **Step 1: Import Data Set**

Click Import Data Set from the Environment pane > Select Import CSV file.

A screenshot of a computer

Description automatically generated

A pop-up will appear > Type Name > Click Import.

Graphical user interface, text

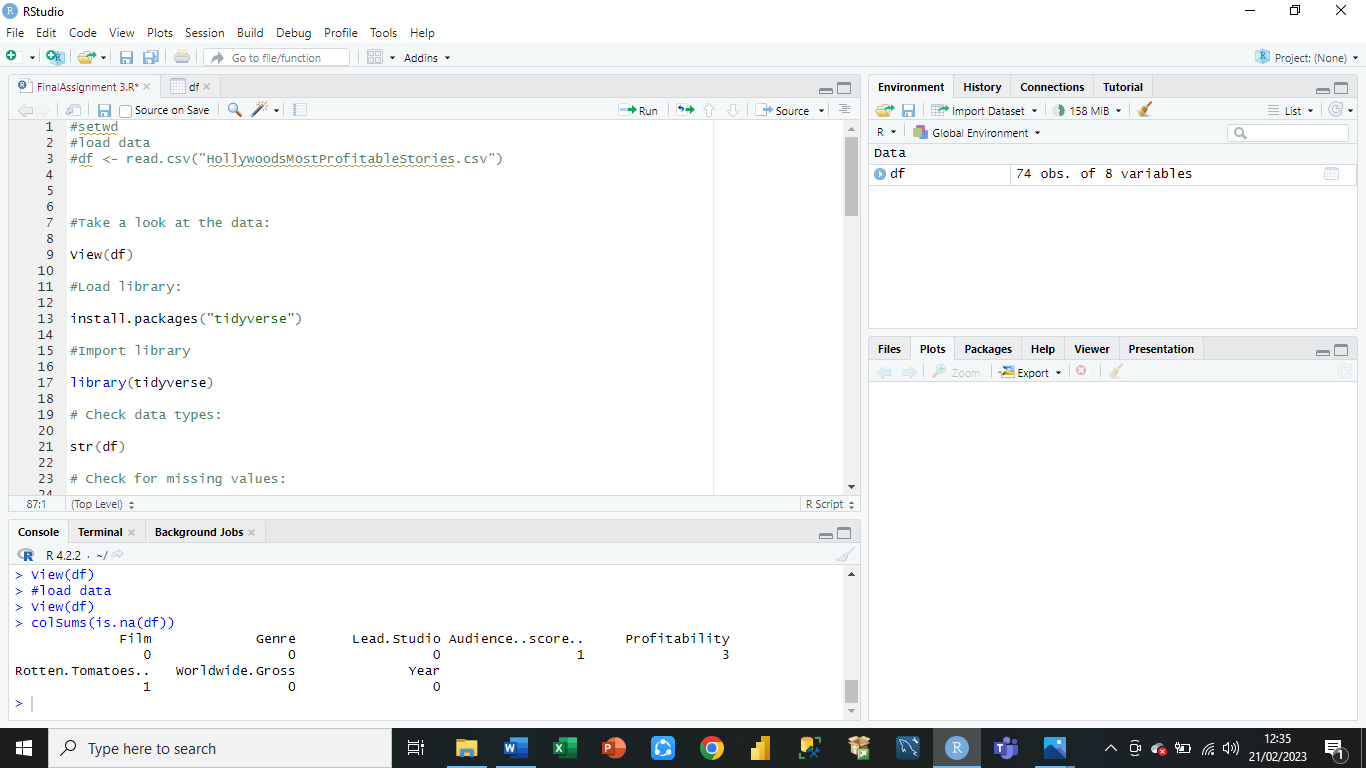
Description automatically generated with medium confidence

* **Step 2: Load and Import the Library**

To Load and Import tidyverse Library > Type

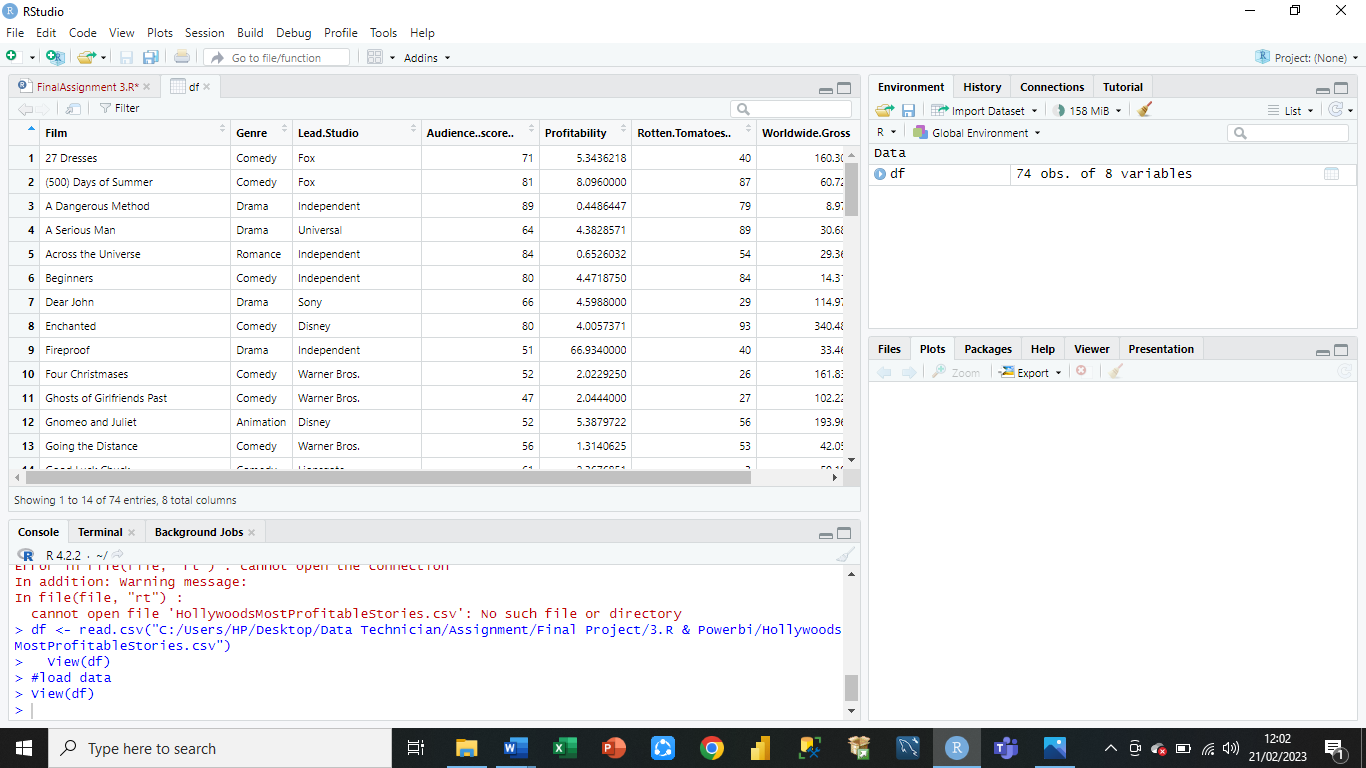
install.packages("tidyverse")

library(tidyverse)



* **Step 3: Look at Data Frame**

To View Data Frame > Type Veiw(df) > Data Set will Be shown in a new Script



* **Step 4: Check Data Types (Structure)**

To check the structure of Data Frame and Data types > Type str()

A screenshot of a computer

Description automatically generated

* **Step 5: Check for Missing Values**

To check for Missing Values > Type colsums(is.na(df))

A screenshot of a computer

Description automatically generated

* **Step 6: Drop Missing Values**

To drop the Missing Values if there is any > Type df <-na.omit(df)

A screenshot of a computer

Description automatically generated

* **Step 7: Check Missing Values removed**

To make sure Missing Values are removed or not > Type colsums(is.na(df))

A screenshot of a computer

Description automatically generated

* **Step 8: Check for Duplicates**

To Check for Duplicates > Type dim(df[duplicated(df$Film),])[1]

A screenshot of a computer

Description automatically generated

* **Step 9: Round off Values to 2 places**

To Round Off Values in two columns Profitability and Worldwide.Gross > Type

df$Profitability <- round(df$Profitability ,digit=2)

df$Worldwide.Gross <- round(df$Worldwide.Gross ,digit=2)

A screenshot of a computer

Description automatically generated

* **Step 10: View Data Frame**

To View Rows in Data Frame > Type dim(df)

A screenshot of a computer

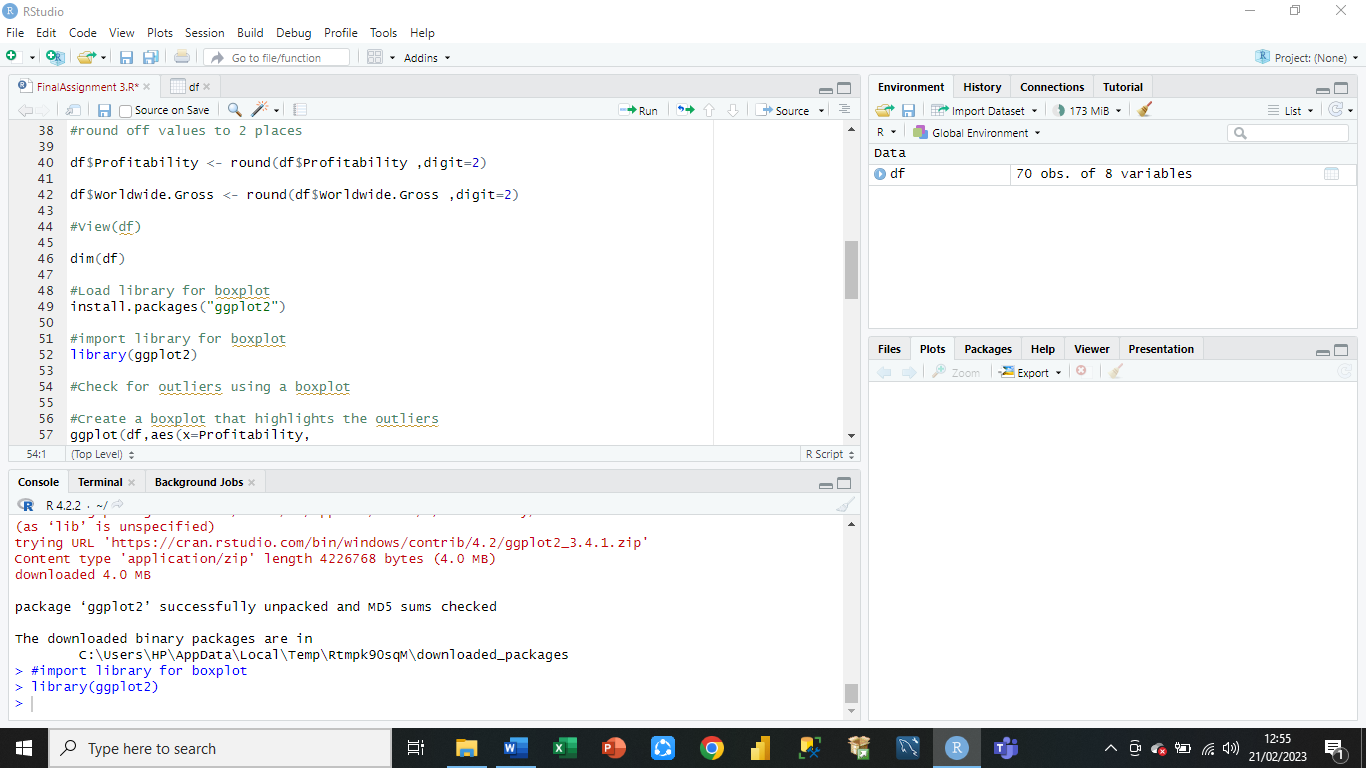
Description automatically generated

* **Step 11: Load and Import Library for Boxplot**

Load and Import “ggplot2” Library > Type

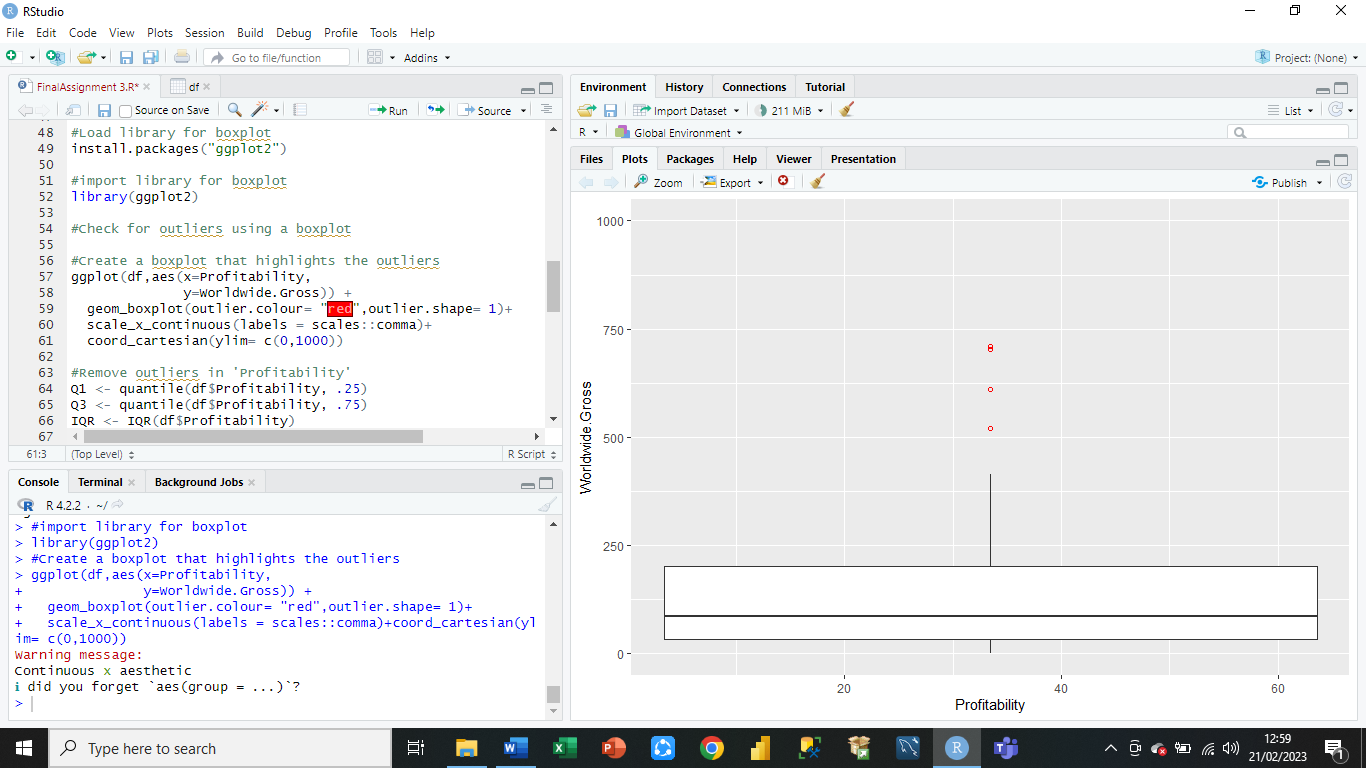
install.packages("ggplot2")

library(ggplot2)



* **Step 12: Create a Boxplot that highlights Outliers**

Boxplot that highlights the outliers for Profitability and Worldwide.Gross on X and Y axis respectively



* **Step 13: Remove outliers in 'Profitability'**

Removing Outliers in Profitability by creating Quantiles of 0.25 And 0.75

Graphical user interface, application

Description automatically generated

* **Step 14: Remove outliers in 'Worldwide. Gross'**

Removing Outliers in Profitability by creating Quantiles of 0.25 And 0.75

Graphical user interface, application

Description automatically generated

* **Step 15: Summary Statistics/Univariate Analysis**

To Check Summary Statistics of Data Frame > Type summary(df1)

Shows Univariate Analysis i-e, Mean, median, Mode, Min, Max, 1st 3rd Quantiles of each Column.

A screenshot of a computer

Description automatically generated

* **Step 16: Bivariate Analysis / Scatterplot**

To Do Bivariate Analysis > Draw Scatterplot > Shows the relationship between Rotten Tomatoes and Lead Studio.

A screenshot of a computer

Description automatically generated with medium confidence

* **Step 17: Bivariate Analysis / Bar Chart**

To Do Bivariate Analysis > Draw Bar Chart > Shows number in each year.

Graphical user interface, application

Description automatically generated

* **Step 18: Export Clean Data**

To save Clean Data Set > Export the clean Data Set > Type

write.csv(df1, "clean\_df.csv")

To Check > Click Files > Clean CSV file will appear.

A screenshot of a computer

Description automatically generated

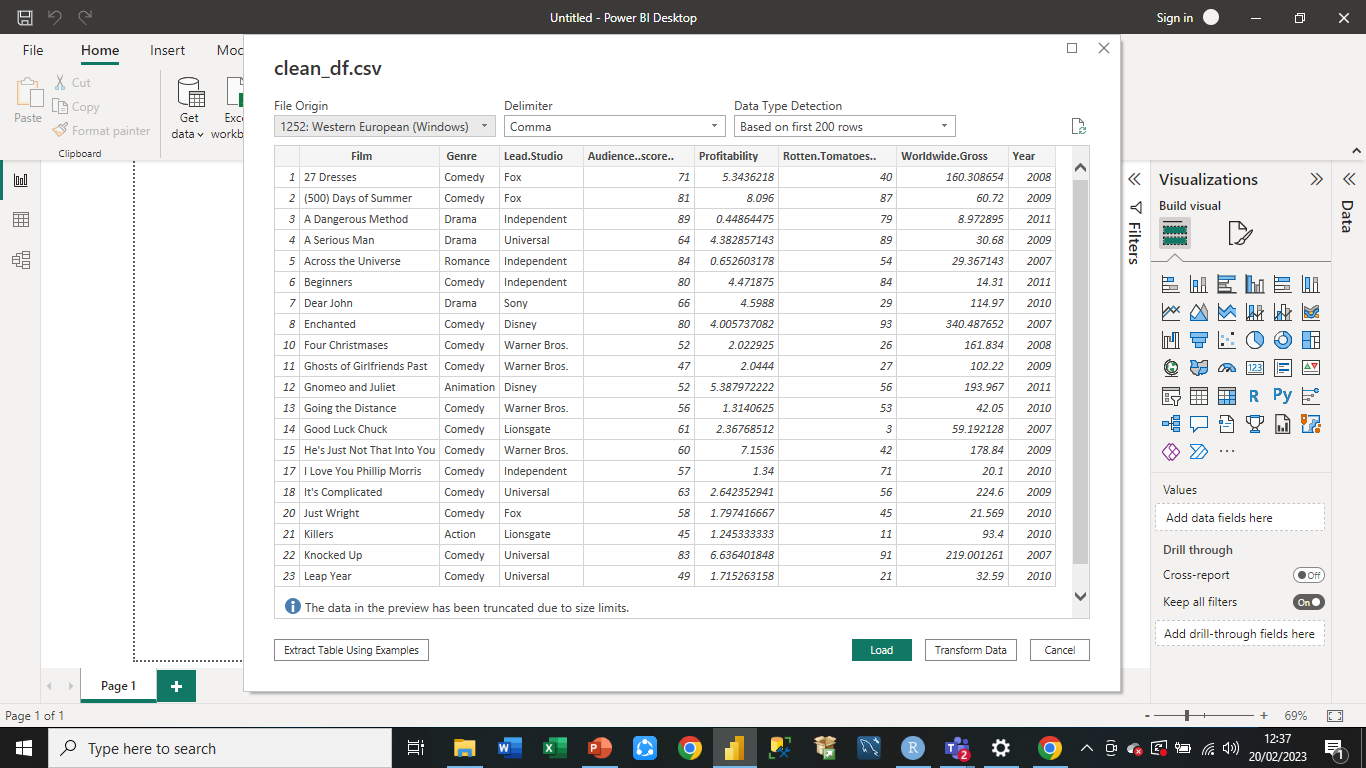
**Power Bi**

**Task 2**

* **Step 1: Import Data Set**

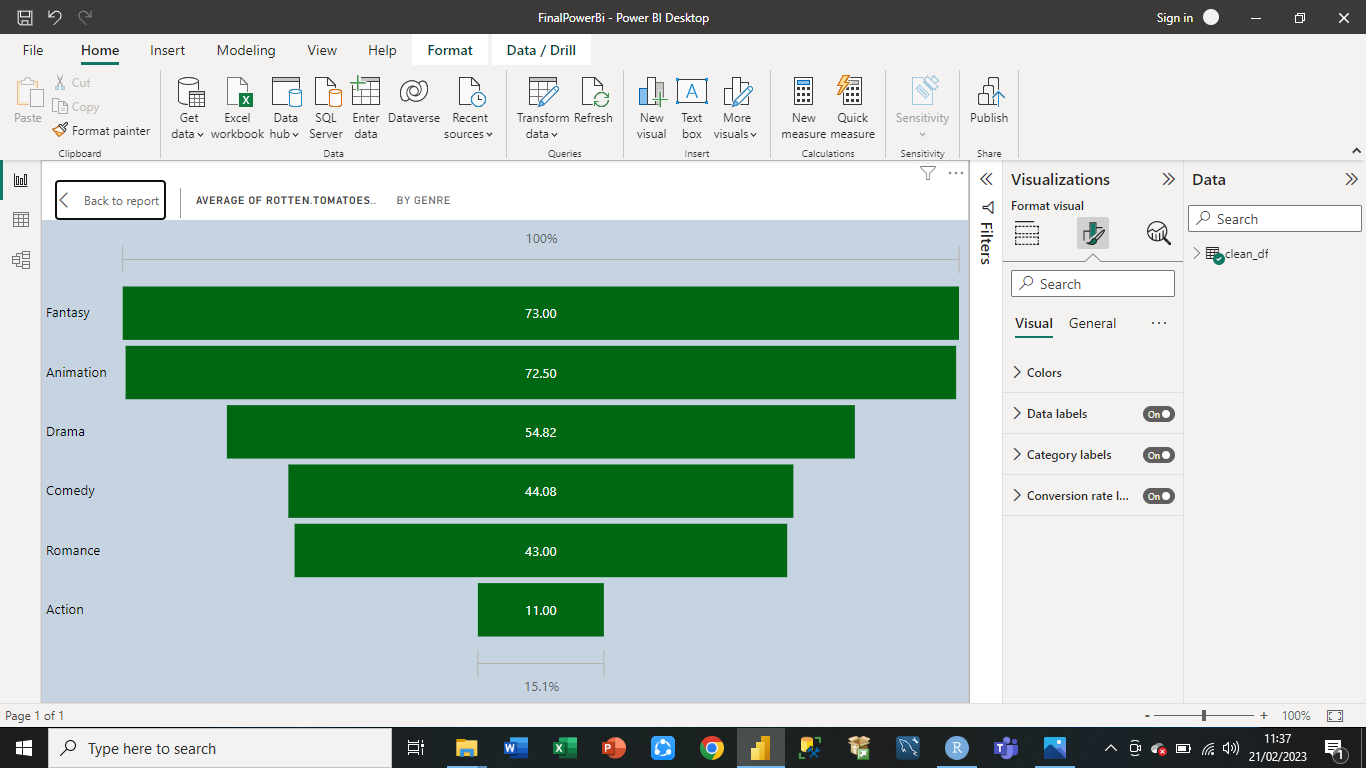
To import Data Set > Click on Get data > Select CSV file > A popup will appear > Click Load

Start creating Graphs in Report Veiw.



* **Step 2: Create Graph 1**

Create 1st Graph > Select Funnel Graph from Visualisation Pane > Shows Average of Rotten Tomatoes by Genre



* **Step 3: Create Graph 2**

Create 2nd Graph > Select Treemap Graph from Visualisation Pane > Shows Sum of Worldwide.Gross by Year

Graphical user interface, chart, treemap chart

Description automatically generated

* **Step 4: Create Graph 3**

Create 3rd Graph > Select Stacked Bar Chart From Visualisation Pane > Shows Sum of Audience.Score by Film

Graphical user interface

Description automatically generated

* **Step 5: Create Graph 4**

Create 4th Graph > Select Area Chart from Visualisation Pane > Shows Sum of Profitability by Lead.Studio

Graphical user interface, application

Description automatically generated

* **Step 6: Create Graph 5**

Create 5th Graph > Select Pie Chart from Visualisation Pane > Shows Count of Film by Year

Graphical user interface, chart, application, pie chart

Description automatically generated

* **Step 7: Create Graph 6**

Create 6th Graph > Select Waterfall Chart from Visualisation Pane > Shows Count of Film by Lead.Studio

A screenshot of a computer

Description automatically generated

* **Step 8: Create Graph 7**

Create 7th Graph > Select Gauge Chart from Visualisation Pane > Total Rotten Tomatoes

Graphical user interface, application

Description automatically generated

* **Step 9: Create Graph 8**

Create 8th Graph > Select Gauge Chart from Visualisation Pane > Total Worldwide.Gross

Graphical user interface, application

Description automatically generated

* **Step 10: Create Graph 9**

Create 9th Graph > Select Gauge Chart from Visualisation Pane > Total Audience Score

Graphical user interface, application

Description automatically generated

* **Step 11: Build a Dashboard**

Graphical user interface, application, treemap chart

Description automatically generatedBuild a Dashboard by plotting all graphs on it and design it by using three colors: Blue, Green, and Brown, as the company requires.