R and power bi project

To analyse the performance of Hollywood movies

Arun Asokan

JustIT

Table of Contents

**Step 1: Initial Exploratory Analysis2**

**Step 2.1: Outlier removal3**

**Step 3: Exploratory Data Analysis 3**

**Step 4: Export data 3**

**Step 1: Initial Exploratory Analysis**

I saved the download to an excel file and imported the dataset as highlighted. I viewed the database (saved as df) and loaded/ installed the library “tidyverse”.

**Graphical user interface, text, application

Description automatically generated**

The library was then imported, and data types checked.

**Text

Description automatically generated**

**Text

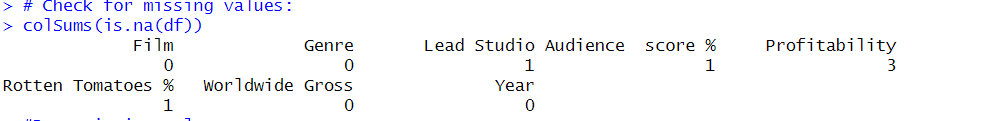
Description automatically generated**

**Step 2: Clean Data**

Missing values were checked for and then dropped/ removed from the database. I then checked/ viewed the database/ table to ensure they had been removed.

Graphical user interface, text, application, chat or text message

Description automatically generated



Duplicates were checked for and then I rounded both the “Probability” and “Worldwide Gross” to 2 decimal places to clean the data. I then viewed the database/ table to ensure this had been completed correctly.

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**Step 2.1: Outlier removal**

I ran a check to see if there were any anomalies/ outliers in my data and then plotted a boxplot to highlight these.

Text

Description automatically generated

Chart

Description automatically generated

I removed the outliers from “Probability”.

A picture containing text

Description automatically generated

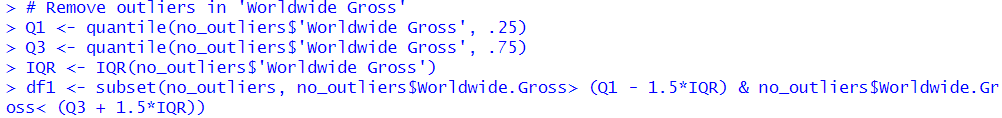
A picture containing graphical user interface

Description automatically generated

Then repeated the process to remove outliers from “Worldwide Gross”

Text

Description automatically generated





**Step 3: Exploratory Data Analysis**

I created a summary of the dataset and viewed this. From this dataset I created a scatterplot graph for “Lead Studio” and “Rotten Tomatoes” and a relevant bar chart for count of films per year.

Text

Description automatically generated

Chart, bar chart

Description automatically generated

**Step 4: Export data**

Finally, the cleaned data was exported as a .csv file.



**Step 5: Create Power BI Dashboard**

Data was imported in to PowerBI

Graphical user interface, text, application

Description automatically generated

Dashboard created using company’s brand colours blue, green and brown. I gave a brief description as to what the dashboard shows and highlighted the main insights.

Graphical user interface, chart

Description automatically generated