|  |  |
| --- | --- |
| Popcorn and drink in an empty red theater  **R AND POWER BI PROJECT Assignment - 3** | Aim: To analyze the performance of Hollywood movies  Data: Title, genre, studio, profitability and ratings for movies released 2007-2012. Source: InformationIsBeautiful.net Download data from this link: https://public.tableau.com/app/sample-data/HollywoodsMostProfitableStories.csv  Gulnar Sohail Data Technician BootCamp |

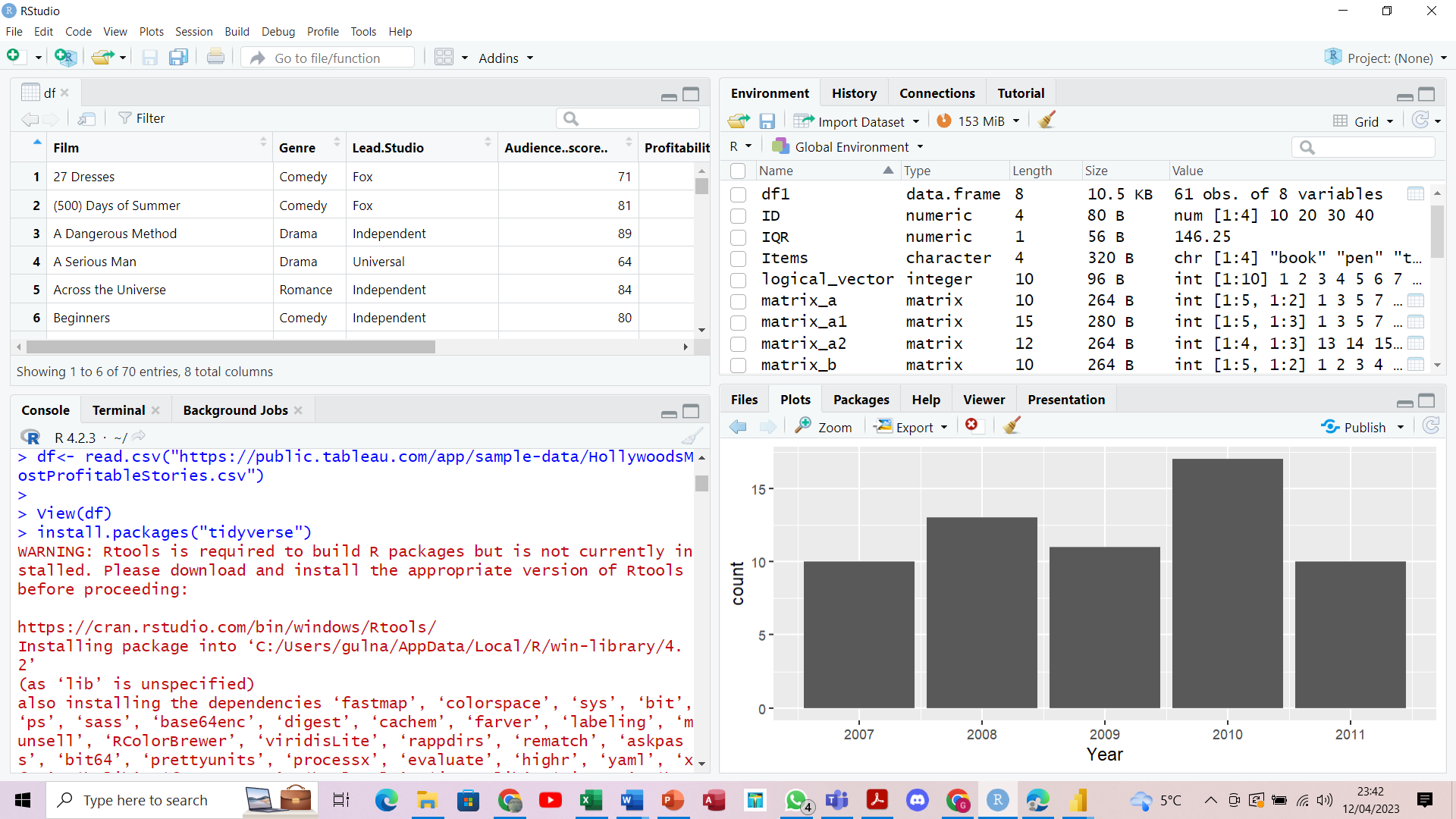
# **Step1: Initial Exploratory Analysis**

## Load data from the link given, view data frame, load and import library and finally check data types.

**Note: loading library will take quite a few time.**

Commands are as below:

> df<- read.csv("https://public.tableau.com/app/sample-data/HollywoodsMostProfitableStories.csv")



> View(df)

> install.packages("tidyverse") Wait for a while

> install.packages("tidyverse")

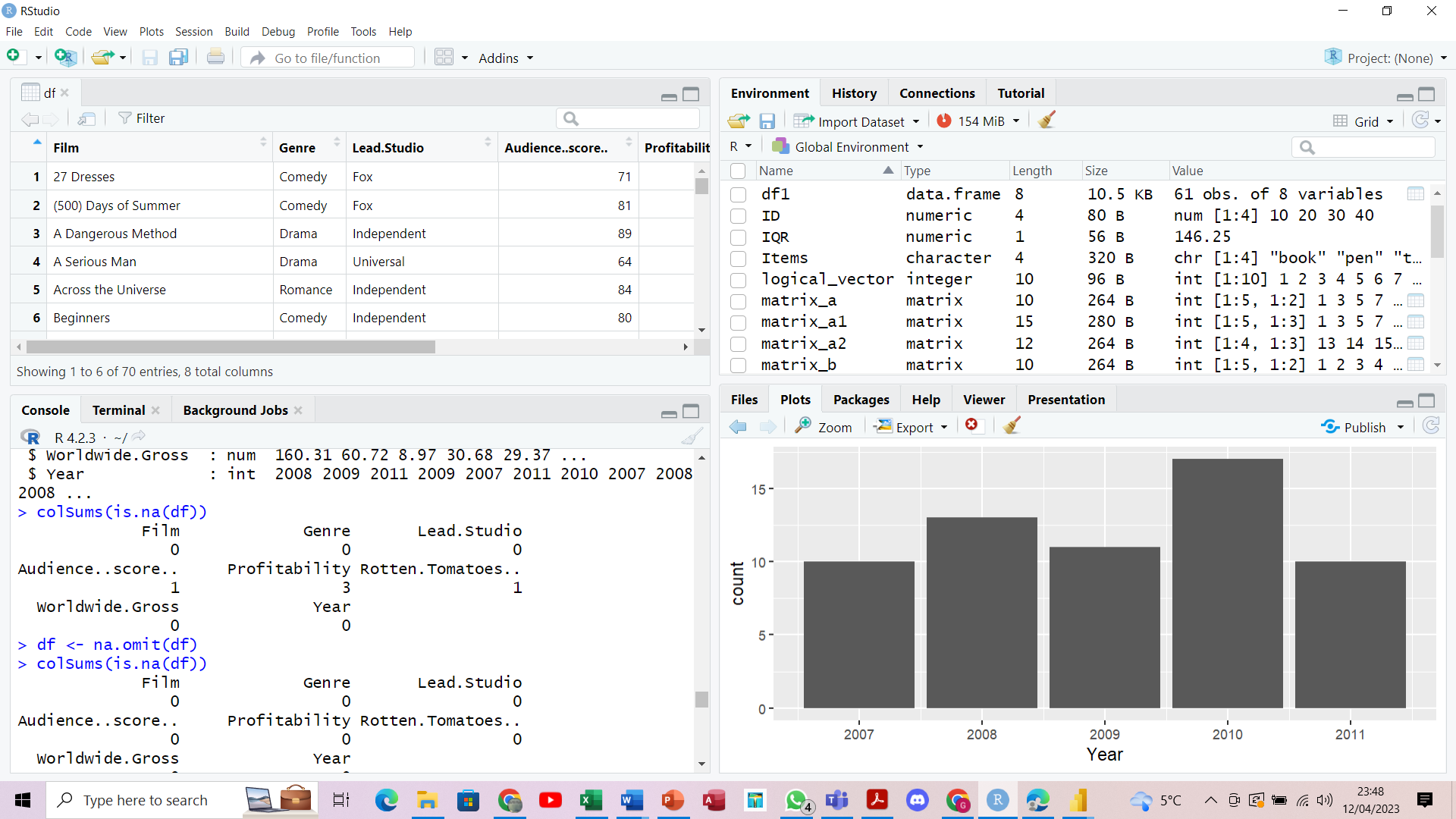
> str(df)

# **Step 2: Clean Data**

## Check for missing values, exclude missing values, empty rows removed, check for duplicates, decimal choice up to 2 digits

For Instance:

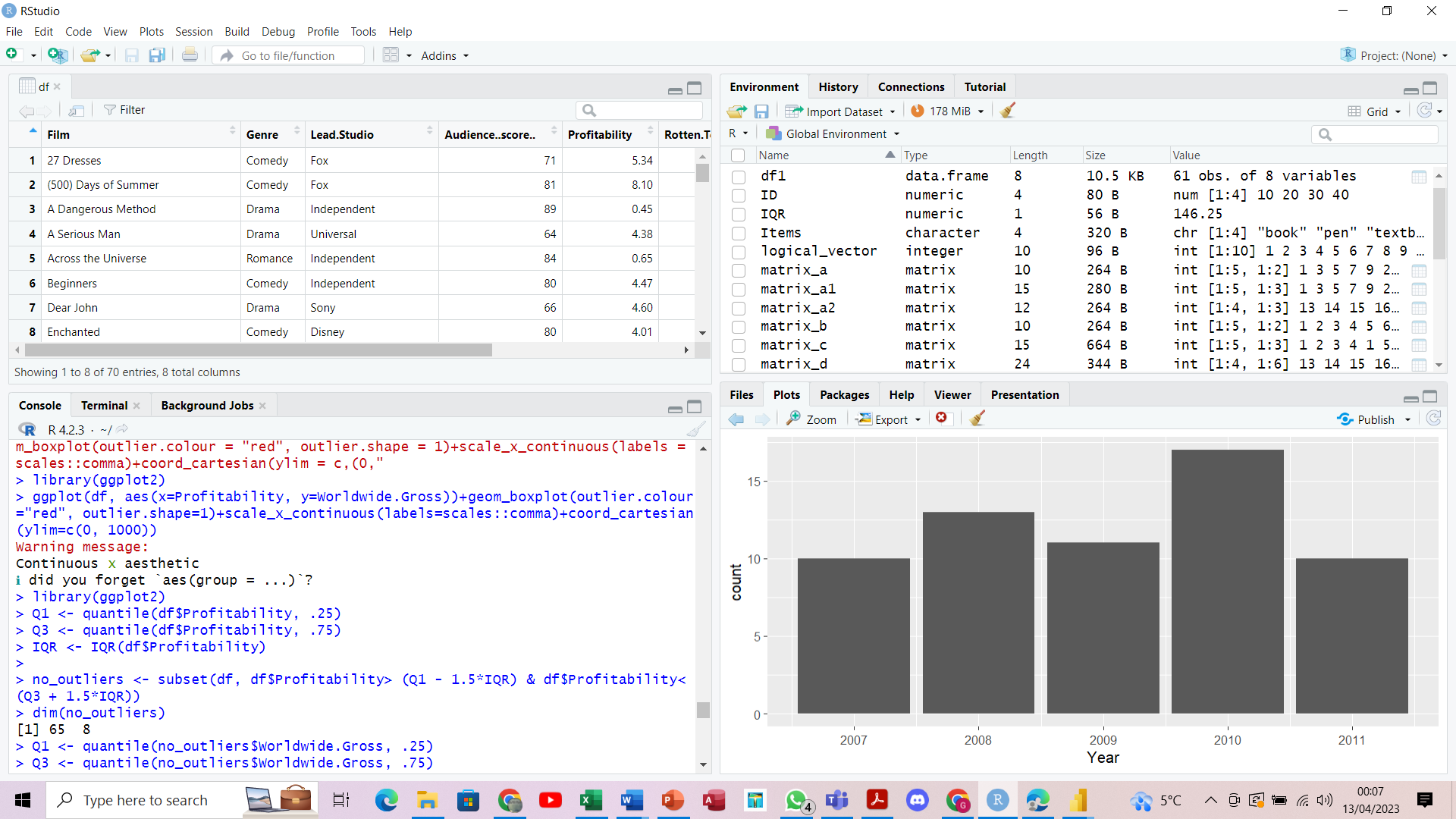
> colSums(is.na(df))



# **Step 2.1: Outlier removal**

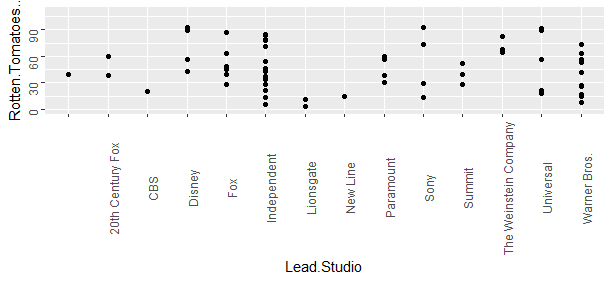
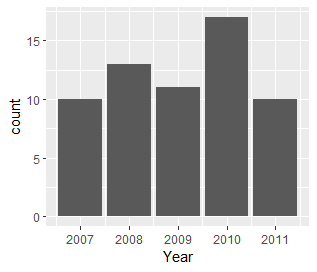
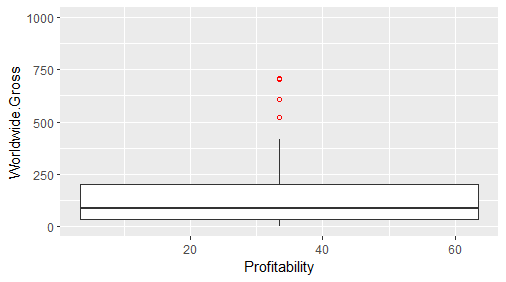
## Check for any abnormality values, outliers and remove them using proper command.

Firstly, we find first (Q1) Q1 <- quantile(no\_outliers$Worldwide.Gross, .25) and third (Q3) Q3 <- quantile(no\_outliers$Worldwide.Gross, .75). Then, we find interquartile range (IQR) IQR <- IQR(no\_outliers$Worldwide.Gross). In addition, we calculate df1 <- subset(no\_outliers, no\_outliers$Worldwide.Gross> (Q1 - 1.5\*IQR) & no\_outliers$Worldwide.Gross< (Q3 + 1.5\*IQR)). Then, we use subset() function to remove outliers.



# Step 3: Exploratory Data Analysis

## Some Analysis now



# Step 4: Export data

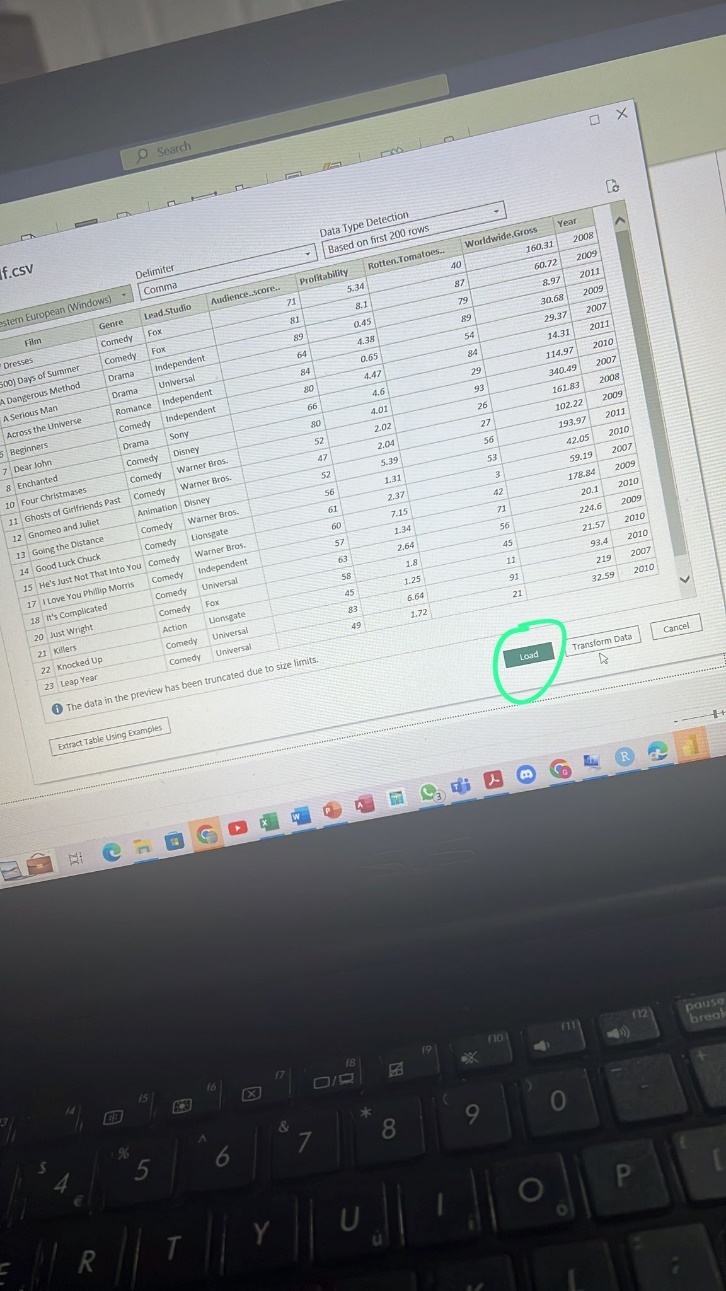
## Once data is fully cleaned and ready, you can export it using below command:

write.csv(df1, “clean\_df.csv")

**Note: Check your document folder to find the csv file.**

# Import clean\_df in Power BI.

## Get Data 🡪 CSV 🡪 Load or Transform Data



# Step 5: Create Power BI Dashboard

## Client’s Requirement

For the dashboard, the company would like you to use their brand colors which are blue, green and brown. You can use light or dark shades of each color. For example, light blue and dark blue are acceptable.

### Click [Here](https://app.powerbi.com/links/iv6C3LIifC?ctid=6efd0f20-57c8-4447-b53f-00d4992ca50b&pbi_source=linkShare) for the Published link

## Further see below points as fulfilled.

## The average Rotten Tomatoes ratings of each genre

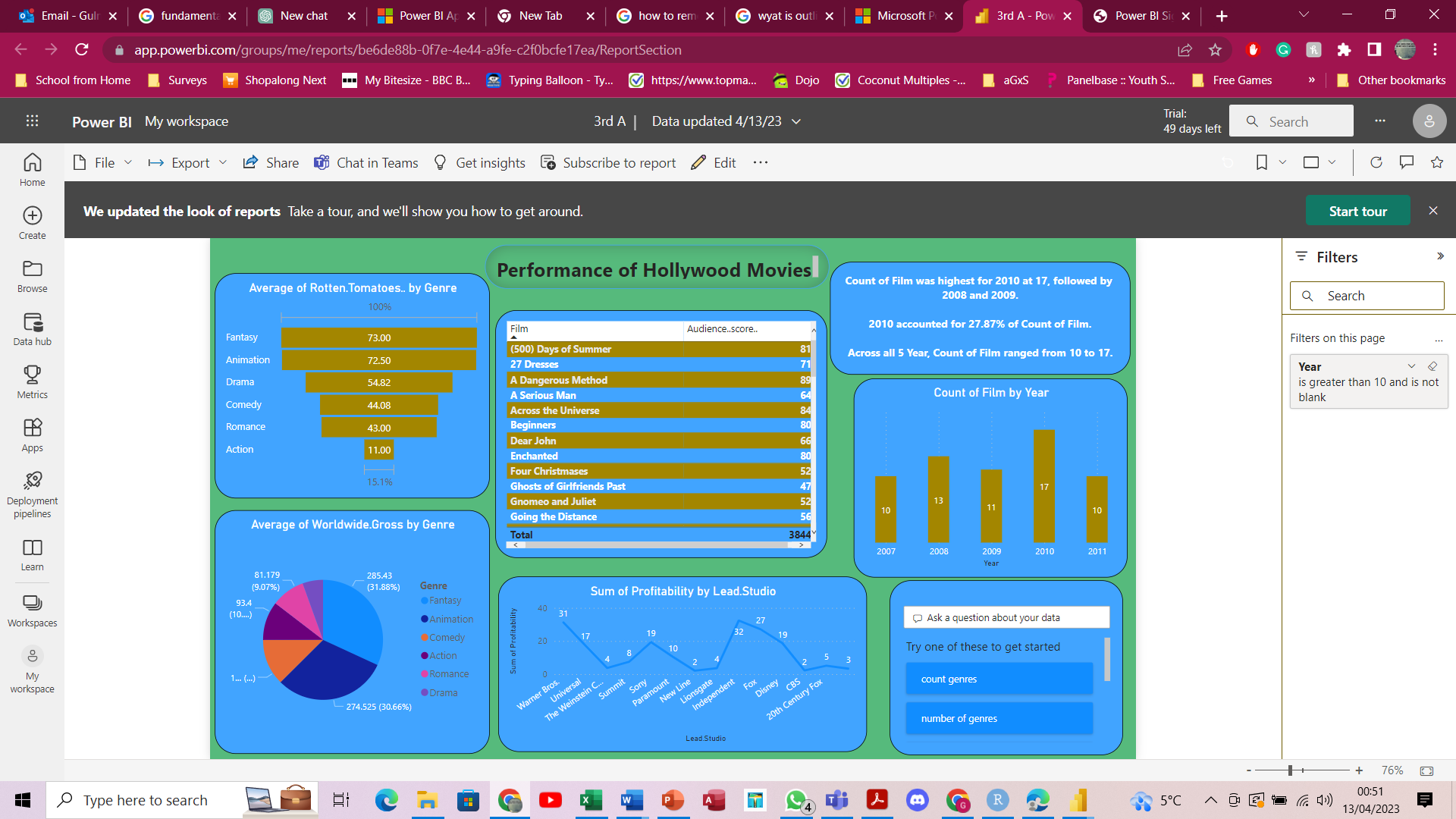
## The number of movies produced per year

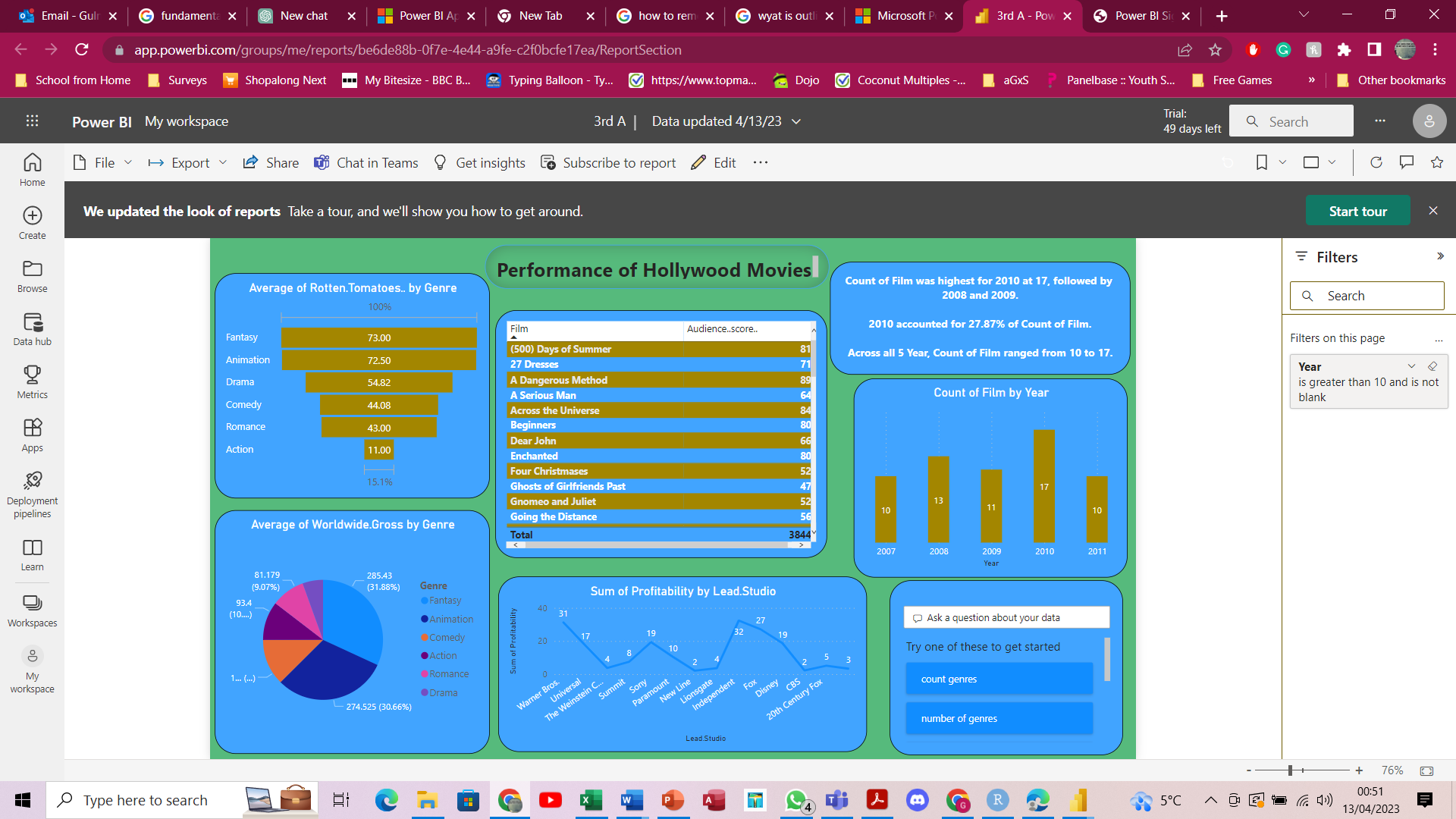
## The audience score for each film

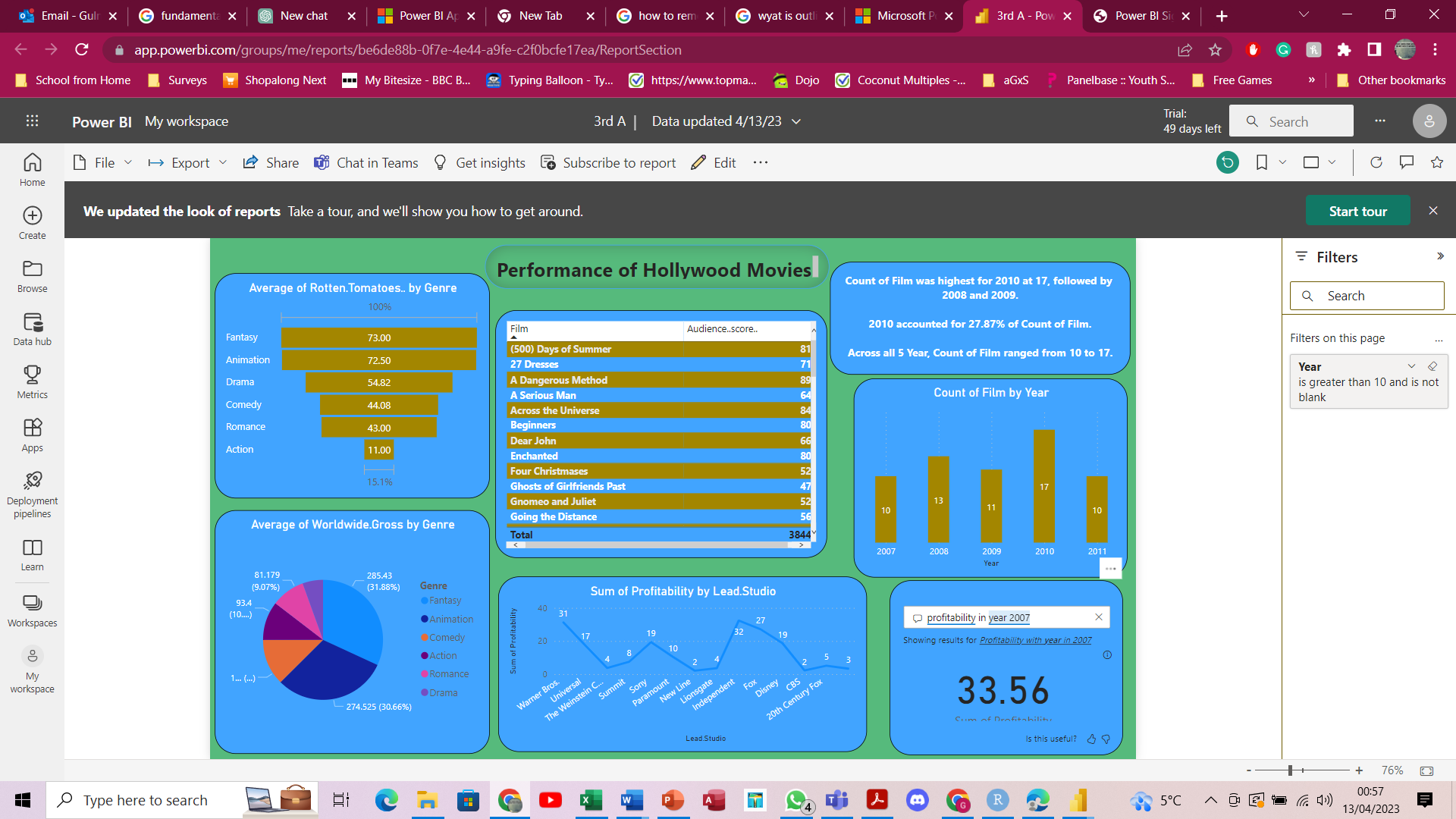
## The profitability per studio

## The worldwide gross per genre

Additionally added a summary section

And ask a question section, where client can quickly find some visibility of their ambiguity.

  
for instance:



Thanks for your time