

Project

R and Power BI Project



Aim: To analyze the performance of Hollywood movies

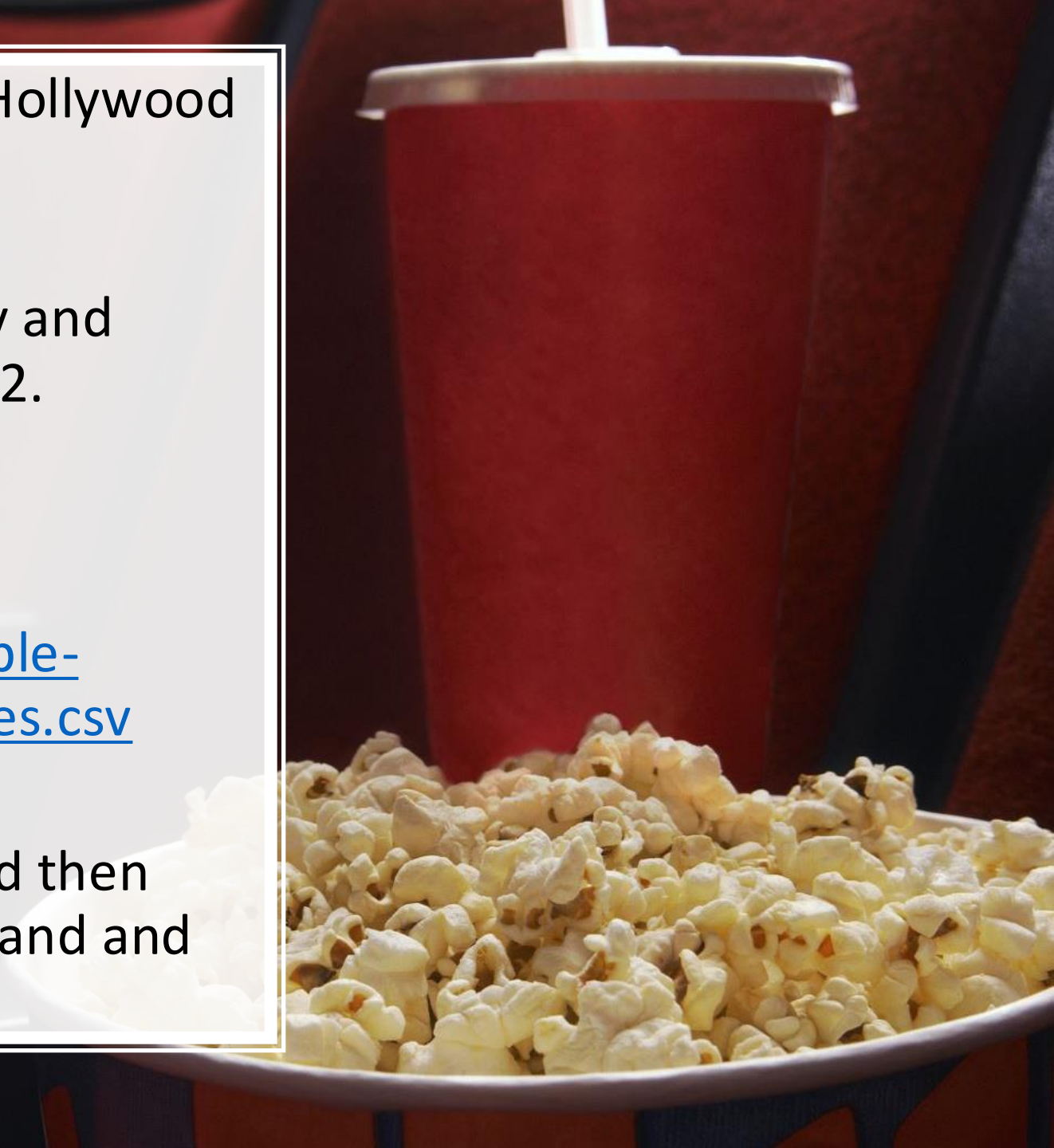
Data: Title, genre, studio, profitability and ratings for movies released 2007-2012.

Source: InformationIsBeautiful.net

Download data from this link:

<https://public.tableau.com/app/sample-data/HollywoodsMostProfitableStories.csv>

Tip: Windows users, press control and then click the link. Mac users, press command and then click the link.



Step1: Initial Exploratory Analysis

#Load data

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

#Take a look at the data:

```
View(df)
```

#Load library:

```
install.packages("tidyverse")
```

```
#Import library
```

```
library(tidyverse)
```

```
# Check data types:
```

```
str(df)
```

Step 2: Clean Data

Check for missing values:

```
colSums(is.na(df))
```

#Drop missing values

```
df <- na.omit(df) or df <- df %>% drop_na()
```

check to make sure that the rows have been removed

```
colSums(is.na(df))
```

Step 3: Exploratory Data Analysis

#Summary Statistics:

```
summary(df)
```

#scatterplot

```
ggplot(df, aes(x=Lead.Studio, y=Rotten.Tomatoes..)) + geom_point()+  
scale_y_continuous(labels = scales::comma)+coord_cartesian(ylim = c(0,  
110))+theme(axis.text.x = element_text(angle = 90))
```

#bar chart

```
ggplot(df, aes(x=Year)) + geom_bar()
```

Step 4: Export data

```
#Export clean data  
write.csv(df, "clean_df.csv")
```

Tips:

- Import `clean_df` in Power BI.
- Feel free to add more charts apart from the ones mentioned in the next slide.
- 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.

You can combine these colors any way that you like. For example, you can use only blue and green if you want to.

Step 5: Create Power BI Dashboard

In your power BI Dashboard, the client would like to see:

- 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