R Programming for Business

In the context of our R Programming project, we selected a movie dataset obtained from Kaggle called “Full TMDB Movies Dataset 2023”. The dataset contains a collection of 930,000 movies from The Movie Database (TMDB), a community-built database for movies and TV shows, collected in October 2023. The dataset offers a comprehensive compilation of movie-related information, encompassing a wide array of variables such as movie titles, ratings, release dates, revenue figures, genres, and more. This dataset has a big analytical potential and allows us to examine financial metrics, including budget and revenue, while also exploring qualitative measures such as audience preferences and movie genres. Additionally, since the dataset includes movies from all over the world, it makes it possible to study the film industry from a global perspective. However, an inherent limitation associated with this dataset arises from its collaborative nature, where any member of the community can contribute or edit the information on the TMDB website. Consequently, the accuracy of the acquired data remains unverifiable.

To enhance the quality of our data, we conducted a data cleaning process using Tableau Prep. This involved the removal of unnecessary columns from the original dataset, such as "backdrop path" (a URL for the movie's backdrop image), "adult" (indicating suitability for adults), "homepage" (a URL for the movie's official website), "imdb_id" (IMDb identifier), "overview" (brief movie summary), "poster path" (a URL for the movie's poster image), "tagline" (movie slogan), and "spoken languages" (languages spoken in the movie). Additionally, we removed all null values. These column exclusions and data cleaning steps, while critical for optimizing our analysis, did not compromise the integrity of our findings, given the dataset's substantial size. Moreover, we chose to focus solely on data spanning the years 2013 to 2023, with the intention of reducing dataset size and conducting our movie analyses within a defined 10-year timeframe. Also, to enrich our dataset, we added an extra column called 'profit' using R which calculates the difference between a movie's revenue and its budget. This new variable enhances our dataset's ability to analyse the financial performance of each movie.

For reference, we have included a data dictionary below, providing a concise description of our dataset and its variables.

Variable	Type	Description	Insights
ID	Numerical	Unique identifier for each movie given from TMDB	
Title	Factor	Title of the movie	
Vote average	Numerical	Average rating of the movie	If 0, no data available
Vote count	Integer	Number of votes for the movie	If 0, no data available
Status	Factor	Status of the movie	3 options: released, post-production, in production
Release Date	Factor	Release date of the movie	Format: yyyy-mm-dd
Revenue	Integer	Total revenue generated by the movie in USD	If 0, no data is available
Runtime	Integer	Duration of the movie in minutes	If 0, no data is available
Budget	Numerical	Budget allocated for the movie in USD	If 0, no data is available
Original language	Integer	Original language of the movie	
Popularity	Numerical	Score of the movie	If 0, no data is available
Genres	Factor	Genres associated with the movie	
Production companies	Factor	Production companies involved in the movie	
Production country	Factor	Countries where the movie was produced	
Profit	Numerical	Revenue-budget	

We explored the dataset to derive insights and recommendations that can guide decision-making in the film and entertainment sector while also shedding light on the landscape of the film industry over the past decade considering factors such as genre trends, financial performance and release strategies.

First, we conducted a thorough check in R, which involved examining the dataset for any missing values and duplicated rows, while also introducing the 'profit' column.

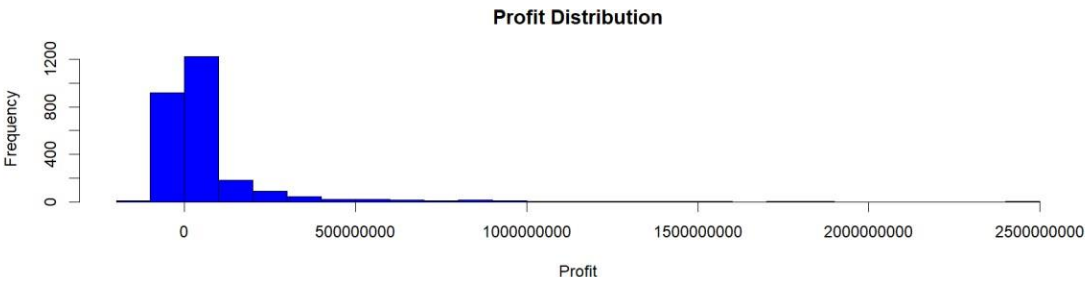
```
#Verifying Data
# Identify missing values
is_missing <- is.na(TMDB_data)
# Count missing values
sum(is_missing) #0
# Identify duplicate rows
duplicated_rows <- duplicated(TMDB_data)
sum (duplicated_rows) #0

#Enriching dataset: adding profit variable
TMDB_data$profit <- TMDB_data$revenue - TMDB_data$budget
head (TMDB_data)
```

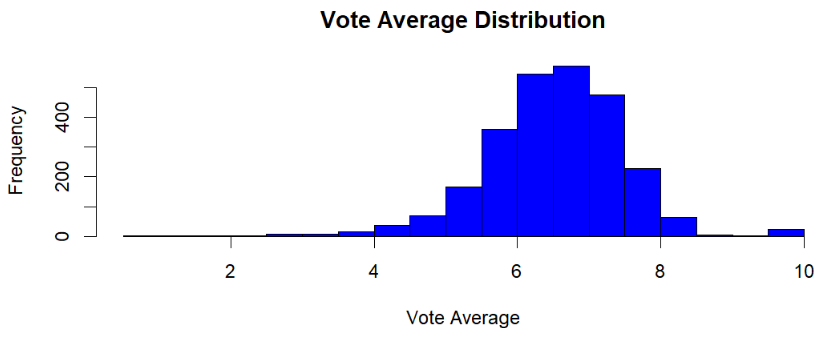
Subsequently, we examined the dataset's structure and generated summary statistics.

```
> str(TMDB_data)
'data.frame':   2583 obs. of  16 variables:
 $ id                  : int  157336 293660 299536 118340 299534 475557 106646 99861 271110 68721 ...
 $ imdb_id             : Factor w/ 2583 levels "tt0293429","tt0315642",..: 36 383 1580 717 1581 2312 56 971 1381 295 ...
 $ title               : Factor w/ 2564 levels "'71","#Alive",..: 948 529 236 801 235 1014 2330 234 419 956 ...
 $ vote_average        : num  8.42 7.61 8.26 7.91 8.26 ...
 $ vote_count          : int  32571 28894 27713 26638 23857 23425 22222 21754 21541 21064 ...
 $ status              : Factor w/ 1 level "Released": 1 1 1 1 1 1 1 1 1 1 ...
 $ release_date        : Factor w/ 1450 levels "1/1/2013","1/1/2014",..: 358 610 802 1177 799 124 447 793 805 777 ...
 $ revenue             : num  7.02e+08 7.83e+08 2.05e+09 7.73e+08 2.80e+09 ...
 $ budget              : int  165000000 58000000 300000000 170000000 356000000 55000000 100000000 365000000 250000000 200000000 ...
 $ runtime             : int  169 108 149 121 181 122 180 141 147 130 ...
 $ original_language   : Factor w/ 51 levels "ar","bn","ca",..: 9 9 9 9 9 9 9 9 9 9 ...
 $ popularity          : num  140.2 72.7 154.3 33.3 91.8 ...
 $ genres              : Factor w/ 859 levels "Action","Action, Adventure",..: 179 8 139 107 199 402 384 29 139 29 ...
 $ production_companies: Factor w/ 2441 levels "100 Bares, Ministerio de Cultura, 369 Productions, Catmandu Branded Entertainment, JEMPSA, Plural - Jemps
a, Pra"| __truncated__,..: 1092 15 1230 1230 1230 2313 661 1230 1230 1230 ...
 $ production_countries: Factor w/ 442 levels "Andorra, Spain, Poland",..: 409 418 418 418 418 130 418 418 418 418 ...
 $ profit              : num  5.37e+08 7.25e+08 1.75e+09 6.03e+08 2.44e+09 ...
> summary(TMDB_data)
       id              imdb_id                         title       vote_average      vote_count          status         release_date 
 Min.   :    189   tt0293429:   1   Gold                    :   3   Min.   : 0.800   Min.   :    1   Released:2583   10/12/2017:   8  
 1st Qu.: 256962   tt0315642:   1   Anna                    :   2   1st Qu.: 6.000   1st Qu.:  128                   12/25/2014:   8  
 Median : 374173   tt0359950:   1   Battle of the Sexes :   2   Median : 6.590   Median :  730                   9/1/2017  :   7  
 Mean   : 389579   tt0365907:   1   Beauty and the Beast:   2   Mean   : 6.537   Mean   : 2205                   10/16/2015:   6  
 3rd Qu.: 496237   tt0369610:   1   Believe                 :   2   3rd Qu.: 7.162   3rd Qu.: 2606                   10/7/2016 :   6  
 Max.   :1123927   tt0385887:   1   Cake                    :   2   Max.   :10.000   Max.   :32571                   12/3/2015 :   6  
                   (Other)  :2577   (Other)                 :2570                                                    (Other)   :2542  
    revenue              budget              runtime      original_language   popularity                 genres    
 Min.   :        70   Min.   :    10000   Min.   :  2.0   en     :1699      Min.   :   0.600   Drama             : 184  
 1st Qu.:   2588495   1st Qu.:  4000000   1st Qu.: 97.0   fr     : 127      1st Qu.:   8.025   Comedy            : 153  
 Median :  17056265   Median : 12000000   Median :108.0   hi     : 121      Median :  14.476   Comedy, Drama     :  80  
 Mean   :  99003650   Mean   : 32439512   Mean   :112.3   ru     :  89      Mean   :  25.673   Drama, Romance    :  57  
 3rd Qu.:  83071024   3rd Qu.: 35000000   3rd Qu.:125.0   es     :  78      3rd Qu.:  26.093   Horror, Thriller:  53  
 Max.   :2800000000   Max.   :460000000   Max.   :218.0   ml     :  53      Max.   :1175.267   Drama, History    :  46  
                                                          (Other): 416                         (Other)           :2010  
                                        production_companies                                           production_countries
 Marvel Studios                                     :  19     United States of America                  : 979  
 Yash Raj Films                                     :   7     India                                     : 256  
 DreamWorks Animation, 20th Century Fox             :   6     United Kingdom, United States of America: 108  
 Pixar, Walt Disney Pictures                        :   6     Russia                                    :  84  
 Walt Disney Pictures, Pixar                        :   6     France                                    :  65  
 Walt Disney Pictures, Walt Disney Animation Studios:   6     Canada, United States of America          :  55  
 (Other)                                            :2533     (Other)                                   :1036  
     profit          
 Min.   :-199545977  
 1st Qu.:  -2190074  
 Median :   4261569  
 Mean   :  66564138  
 3rd Qu.:  49190230  
 Max.   :2444000000  
```

In line with summary statistics, we used histograms to visualize the distribution of profits and the average rating of movies.



This graph shows the distribution of profit. The x-axis shows the profit in billions of dollars, and the y-axis shows the frequency of movies having that profit. Overall, the graph shows that the distribution of profit is unequal. The graph is skewed to the left, which means that there are more movies with lower profits than there are movies with higher profits.



The graph shows the distribution of average movie ratings. The x-axis depicts the average movie ratings, ranging from 1 to 10, while the y-axis illustrates the frequency of movies corresponding to each rating. The graph exhibits a slight rightward skew, indicating a prevalence of movies with higher ratings. In summary, the graphical representation underscores that most movies tend to achieve an average rating of 7/10.

Next, we performed correlation analyses to investigate the relationships between profit, budget, revenue, average movie ratings, and runtime.

```
> cor_matrix <- cor(TMDB_data[, c("budget", "revenue", "profit", "vote_average", "runtime")])
> print(cor_matrix) #all are positively correlated
                budget   revenue    profit vote_average   runtime
budget       1.0000000 0.7643442 0.6378220    0.1346150 0.2421872
revenue      0.7643442 1.0000000 0.9841366    0.1949533 0.2085825
profit       0.6378220 0.9841366 1.0000000    0.1958217 0.1825035
vote_average 0.1346150 0.1949533 0.1958217    1.0000000 0.2217416
runtime      0.2421872 0.2085825 0.1825035    0.2217416 1.0000000
```

- There is a strong positive correlation of approximately 0.76 between budget and revenue. This suggests that, in general, as the budget for a movie increases, its revenue tends to increase as well.
- Budget and profit also exhibit a strong positive correlation of around 0.64. This indicates that as movie budgets increase, profits tend to increase, which is consistent with the relationship seen in the budget-revenue correlation.
- The correlation between vote average and runtime is relatively weak, with a correlation coefficient of about 0.22. This suggests that there is not a strong relationship between a movie's runtime and its vote average. Other factors likely play a more significant role in determining the average movie rating.
- The correlation between profit and vote average is approximately 0.20, which indicates a weak positive correlation. The correlation between profit and runtime is also weak, with a correlation coefficient of about 0.18. The runtime and average rating alone are unlikely to be key drivers of a movie's financial success. Other factors like genre or release date might be more significant.

These findings can guide further analyses and decision-making within the film industry and help understand the factors influencing a movie's financial performance and audience reception.

In that sense, we investigated the potential impact of seasonality on profitability by evaluating the average profit on both a monthly and seasonal basis.

	month	mean_profit
	<chr>	<dbl>
1	01	35934077.
2	02	53743142.
3	03	63314342.
4	04	88651043.
5	05	84207508.
6	06	118399028.
7	07	90200596.
8	08	35664327.
9	09	34894374.
10	10	48425869.
11	11	70368226.
12	12	100773658.

	season	mean_profit
	<chr>	<dbl>
1	Fall	48798150.
2	Spring	78463054.
3	Summer	80122704.
4	Winter	64183908.

We observe that June stands out with the highest profit, amounting to approximately $119 million. Following closely, the second and third most profitable months are December and July. The robust profitability in June and July can be attributed to their positioning at the onset of the summer season, while December's success aligns with the holiday festivities.

However, after a more in-depth examination of the second table, which provides insights into the average profit per season, we find that the summer season emerges as the most lucrative. Interestingly, Winter, despite December's individual success, ranks as the third-best season.

Considering these findings, we propose a strategic recommendation for movie studios. Primarily, we recommend scheduling movie releases during the summer season, as it consistently yields higher profits. Furthermore, even though Winter is not the best overall season, studios should prioritize December for their releases. This approach aligns with maximizing profitability while also accounting for seasonal variations in audience demand.

Following our exploration of seasonality's impact on profitability, our analysis extended to consider other factors that may exert influence on movie profits. We shifted our focus to analyse the influence of movie genres on profit and conducted individual linear regression models for different genres.

Action (p<0.01)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     43297779    4255246   10.18   <2e-16 ***
genres_Action1  81432255    7960836   10.23   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 182800000 on 2581 degrees of freedom
Multiple R-squared:  0.03896,   Adjusted R-squared:  0.03859
F-statistic: 104.6 on 1 and 2581 DF,  p-value: < 2.2e-16
```

Adventure (p<0.01)

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        38083455    3884247   9.805   <2e-16 ***
genres_Adventure1 146545030    8810837  16.632   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 177200000 on 2581 degrees of freedom
Multiple R-squared:  0.09681,   Adjusted R-squared:  0.09646
F-statistic: 276.6 on 1 and 2581 DF,  p-value: < 2.2e-16
```

Science Fiction (p<0.01)

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             52196566    3779214   13.81   <2e-16 ***
genres_ScienceFiction1 136439113   11646065   11.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 181700000 on 2581 degrees of freedom
Multiple R-squared:  0.05049,   Adjusted R-squared:  0.05012
F-statistic: 137.3 on 1 and 2581 DF,  p-value: < 2.2e-16
```

Fantasy (p<0.01)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     58252445    3855148  15.110  < 2e-16 ***
genres_Fantasy1 76131574   11667525   6.525 8.15e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 184900000 on 2581 degrees of freedom
Multiple R-squared:  0.01623,   Adjusted R-squared:  0.01585
F-statistic: 42.58 on 1 and 2581 DF,  p-value: 8.151e-11
```

Drama (p<0.01)

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    94616649    4900883  19.306   <2e-16 ***
genres_Drama1 -61667774    7266371  -8.487   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 183900000 on 2581 degrees of freedom
Multiple R-squared:  0.02715,   Adjusted R-squared:  0.02677
F-statistic: 72.02 on 1 and 2581 DF,  p-value: < 2.2e-16
```

History (p<0.05)

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       68994918    3799925  18.157   <2e-16 ***
genres_History1 -34688984   14354827  -2.417   0.0157 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 186200000 on 2581 degrees of freedom
Multiple R-squared:  0.002257,  Adjusted R-squared:  0.001871
F-statistic:  5.84 on 1 and 2581 DF,  p-value: 0.01574
```

Comedy (P>0.01)

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    69354154    4467374  15.525   <2e-16 ***
genres_Comedy1 -8558922    7824533  -1.094    0.274
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 186400000 on 2581 degrees of freedom
Multiple R-squared:  0.0004634, Adjusted R-squared:  7.611e-05
F-statistic: 1.197 on 1 and 2581 DF,  p-value: 0.2741
```

Documentary (P>0.01)

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         67235901    3686373   18.24   <2e-16 ***
genres_Documentary1 -61970147   35406455   -1.75   0.0802 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 186300000 on 2581 degrees of freedom
Multiple R-squared:  0.001185,  Adjusted R-squared:  0.0007985
F-statistic: 3.063 on 1 and 2581 DF,  p-value: 0.08019
```

TV Movie (P>0.01)

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      66616152    3669750  18.153   <2e-16 ***
genres_TVMovie1 -67176090  131881436  -0.509    0.611
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 186400000 on 2581 degrees of freedom
Multiple R-squared:  0.0001005, Adjusted R-squared:  -0.0002869
F-statistic: 0.2595 on 1 and 2581 DF,  p-value: 0.6105
```

Our results indicate that the presence of “Action”, “Adventure”, “Science Fiction”, “Fantasy”, “Drama” and “History” genres is statistically significant in predicting movie profit. This suggests that there is a strong relationship between the inclusion of those genres in a movie and its profitability and it is highly unlikely that the observed relationship is due to random chance. On the other hand, we observed that the presence of the "Documentary", “Comedy” and "TV Movie" genres does not serve as a predictive factor for movie profits.

Given the strong and statistically significant relationships between the "Action," "Adventure," "Science Fiction," "Fantasy," "Drama," and "History" genres and movie profitability, we recommend movie studios and production companies to consider investing in and producing movies within these genres. Allocate resources, marketing efforts, and creative talent to develop projects that fall within these categories. This can increase the likelihood of generating higher profits. Also, we recommend continuously monitoring market trends and audience preferences. The popularity of genres can evolve over time, and staying informed about shifts in audience tastes can inform production decisions. Use market research and data analysis to adapt your genre choices accordingly. The results indicate that "Documentary," "Comedy," and "TV Movie" genres do not significantly predict movie profits. However, these genres often cater to niche audiences with unique interests. Consider producing niche content for specific target markets, as it may not generate big profits but can build loyal audiences and generate revenue through niche distribution channels.

In summary, our comprehensive analysis of the dataset has yielded valuable insights and recommendations that can inform decision-making within the film and entertainment sector. These insights encompass factors, including genre trends, financial performance, and release date strategies. Based on our research findings, we recommend that movie studios and production companies adopt the following strategies:

- Strategic Release Timing: Opt for movie releases during the summer season. Additionally, prioritize December as a strategic release window. These choices align with maximizing profit while accommodating seasonal fluctuations in audience demand.
- Genre-Focused Production: Our analysis highlights the impact of specific genres on movie profitability. Therefore, consider directing resources, creative talent, and marketing efforts towards producing content within the "Action," "Adventure," "Science Fiction," "Fantasy," "Drama," and "History" genres. This targeted approach increases the likelihood of generating elevated profits.

By implementing these recommendations, movie studios and production companies could position themselves for enhanced profitability and market competitiveness in the dynamic film industry.

R CODE

```
Select dataset
TMDB_data<- read.csv("C:/Users/lidie/OneDrive/Escritorio/R PROJECT CLEAN
DATA.csv", stringsAsFactors=TRUE)
print(TMDB_data)

#Verifying Data
# Identify missing values
is_missing <- is.na(TMDB_data)
# Count missing values
sum(is_missing) #0
# Identify duplicate rows
duplicated_rows <- duplicated(TMDB_data)
sum (duplicated_rows) #0


#Enriching dataset: adding profit variable
TMDB_data$profit <- TMDB_data$revenue - TMDB_data$budget
head (TMDB_data)

# Check the structure of the data
str(TMDB_data)
#print total number of columns and rows
cat("Total Columns: ", ncol(TMDB_data)) #16 columns
cat("Total Rows: ", nrow(TMDB_data)) #2583 rows

# Generate summary statistics
summary(TMDB_data)

#Histograms
options(scipen = 999)
hist(TMDB_data$profit,
    main = "Profit Distribution",
    xlab = "Profit",
    ylab = "Frequency",
    col = "blue",
    border = "black",
    breaks = 30)

options(scipen = 999)
hist(TMDB_data$vote_average,
    main = "Vote Average Distribution",
    xlab = "Vote Average",
    ylab = "Frequency",
    col = "blue",
    border = "black",
    breaks = 20)
```

#Correlation Analyses

```
cor_matrix <- cor(TMDB_data[, c("budget", "revenue", "profit", "vote_average",
"runtime")])
print(cor_matrix) #all are positively correlated
```

#Average profit on both a monthly and seasonal basis

```
library(dplyr)
TMDB_data <- TMDB_data %>%
  mutate(month = format(release_date, "%m"),
         season = case_when(
           month %in% c("12", "01", "02") ~ "Winter",
           month %in% c("03", "04", "05") ~ "Spring",
           month %in% c("06", "07", "08") ~ "Summer",
           month %in% c("09", "10", "11") ~ "Fall"
         ))
mean_profit_by_month <- TMDB_data %>%
group_by(month) %>%
summarize(mean_profit = mean(profit))
mean_profit_by_season <- TMDB_data %>%
group_by(season) %>%
summarize(mean_profit = mean(profit))
print(mean_profit_by_month)
print(mean_profit_by_season)
```

 # Single-factors regression on profit

```
PROJECT<- read_csv("C:/Users/20105/Desktop/R PROJECT CLEAN DATA.csv")

PROJECT$profit<-(PROJECT$revenue-PROJECT$budget)

generics<-as.factor(PROJECT$genres)
generics

PROJECT$genres_Action<-str_detect(PROJECT$genres,"Action")
PROJECT$genres_Adventure<-str_detect(PROJECT$genres,"Adventure")
PROJECT$genres_ScienceFiction<-str_detect(PROJECT$genres,"Science Fiction")
PROJECT$genres_Fantasy<-str_detect(PROJECT$genres,"Fantasy")
PROJECT$genres_Comedy<-str_detect(PROJECT$genres,"Comedy")
PROJECT$genres_History<-str_detect(PROJECT$genres,"History")
PROJECT$genres_Drama<-str_detect(PROJECT$genres,"Drama")
PROJECT$genres_TV<-str_detect(PROJECT$genres,"TV Movie")
PROJECT$genres_Documentary<-str_detect(PROJECT$genres,"Documentary")


library(stringr)
log_data<-
data.frame(profit=PROJECT$profit,genres_Action=PROJECT$genres_Action,genres_Adven
ture=PROJECT$genres_Adventure,genres_Comedy=PROJECT$genres_Comedy,genres_Sci
```

enceFiction=PROJECT$genres_ScienceFiction,genres_Fantasy=PROJECT$genres_Fantasy, genres_Drama=PROJECT$genres_Drama)
log_data$genres_Action<-str_replace_all(log_data$genres_Action,"TRUE","1")
log_data$genres_Action<-str_replace_all(log_data$genres_Action,"FALSE","0")
log_data$genres_Adventure<-str_replace_all(log_data$genres_Adventure,"TRUE","1")
log_data$genres_Adventure<-str_replace_all(log_data$genres_Adventure,"FALSE","0")
log_data$genres_Comedy<-str_replace_all(log_data$genres_Comedy,"TRUE","1")
log_data$genres_Comedy<-str_replace_all(log_data$genres_Comedy,"FALSE","0")
log_data$genres_ScienceFiction<-
str_replace_all(log_data$genres_ScienceFiction,"TRUE","1")
log_data$genres_ScienceFiction<-
str_replace_all(log_data$genres_ScienceFiction,"FALSE","0")
log_data$genres_Fantasy<-str_replace_all(log_data$genres_Fantasy,"TRUE","1")
log_data$genres_Fantasy<-str_replace_all(log_data$genres_Fantasy,"FALSE","0")
log_data$genres_Drama<-str_replace_all(log_data$genres_Drama,"TRUE","1")
log_data$genres_Drama<-str_replace_all(log_data$genres_Drama,"FALSE","0")

log_data
summary(log_data)

#Action-profit single factor regression
fit1<-lm(profit~genres_Action,data=log_data)
fit1Sum<-summary(fit1)
fit1Sum

#Adventure-profit single factor regression
fit2<-lm(profit~genres_Adventure,data=log_data)
fit2Sum<-summary(fit2)
fit2Sum

#Comedy-profit single factor regression
fit3<-lm(profit~genres_Comedy,data=log_data)
fit3Sum<-summary(fit3)
fit3Sum

#ScienceFiction-profit single factor regression
fit4<-lm(profit~genres_ScienceFiction,data=log_data)
fit4Sum<-summary(fit4)
fit4Sum

#Fantasy-profit single factor regression
fit5<-lm(profit~genres_Fantasy,data=log_data)
fit5Sum<-summary(fit5)
fit5Sum

#Drama-profit single factor regression
fit6<-lm(profit~genres_Drama,data=log_data)
fit6Sum<-summary(fit6)

```
fit6Sum

 log_data1<-
data.frame(profit=PROJECT$profit,genres_Documentary=PROJECT$genres_Documentary,
genres_TVMovie=PROJECT$genres_TV,genres_History=PROJECT$genres_History)
log_data1$genres_Documentary<-
str_replace_all(log_data1$genres_Documentary,"TRUE","1")
log_data1$genres_Documentary<-
str_replace_all(log_data1$genres_Documentary,"FALSE","0")
log_data1$genres_TVMovie<-str_replace_all(log_data1$genres_TVMovie,"TRUE","1")
log_data1$genres_TVMovie<-str_replace_all(log_data1$genres_TVMovie,"FALSE","0")
log_data1$genres_History<-str_replace_all(log_data1$genres_History,"TRUE","1")
log_data1$genres_History<-str_replace_all(log_data1$genres_History,"FALSE","0")
log_data1

#Documentary-profit single factor regression
fit7<-lm(profit~genres_Documentary,data=log_data1)
fit7Sum<-summary(fit7)
fit7Sum

#TVMovie-profit single factor regression
fit8<-lm(profit~genres_TVMovie,data=log_data1)
fit8Sum<-summary(fit8)
fit8Sum

#History-profit single factor regression
fit9<-lm(profit~genres_History,data=log_data1)
fit9Sum<-summary(fit9)
fit9Sum
```