PIMPRI CHINCHWAD EDUCATION TRUST'S

PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

Department of AS&H

Report of IE-2 activity

Subject: Statistical Data Analysis using R

Academic Year 2022-23

Semester I

Submitted by 1. Vishwajeet Koshti TYCOB124

2. Shrikant Jadhao TYCOB084

Submitted to :Dr. Arif Bagwan

Date: 03/11/2023

Sign:

Project Title: Analyzing Game Ratings and Developer influence : A study on the impact of Developer History on Game Ratings and Probabilities

Data used: Video_Games_Sales_as_at_22_Dec_2016

Details of the data : Data contains 16719 observations and 16 variables. The variables in this data frame include:

1. `Name`: Character variable representing the name of the video game.
2. `Platform`: Character variable indicating the gaming platform.
3. `Year_of_Release`: Character variable specifying the year of release of the game.
4. `Genre`: Character variable categorizing the genre of the game.
5. `Publisher`: Character variable identifying the game's publisher.
6. `NA_Sales`: Numeric variable representing North American sales.
7. `EU_Sales`: Numeric variable representing European sales.
8. `JP_Sales`: Numeric variable representing Japanese sales.
9. `Other_Sales`: Numeric variable representing sales from other regions.
10. `Global_Sales`: Numeric variable representing global sales.
11. `Critic_Score`: Numeric variable for the critic score.
12. `Critic_Count`: Numeric variable for the number of critic reviews.
13. `User_Score`: Numeric variable for the user score.
14. `User_Count`: Numeric variable for the number of user reviews.
15. `Developer`: Character variable identifying the game's developer.
16. `Rating`: Character variable indicating the game's rating.

This data frame appears to contain information about video games, including their sales figures, scores, and other attributes.

- Data Wrangling

```
games <- data.frame(Video_Games_Sales_as_at_22_Dec_2016)
head(games)
str(games)
nrow(games)
games2 <- filter(games, Year_of_Release > 2011)
nrow(games2)
games3 <- clean_names(games2)
colnames(games2)
# any duplicates
anyDuplicated(games3)
#Remove rows with unknown year of release.
games4 <- filter(games3, year_of_release != "NA")
```

Problem stmt 1:Which publishers and developers have the highest-rated games?

Calculate the average critic and user scores for each publisher and developer in the dataset. Sort the results by average scores in descending order to identify the top publishers and developers with the highest average scores. Print the top 5 publishers and developers with the highest average critic scores.

```
publisher_avg_scores <- aggregate(critic_score + user_score ~ publisher, clean_data, mean)
sorted_publisher_avg_scores <- publisher_avg_scores[order(publisher_avg_scores$critic_score, decreasing = TRUE), ]
cat("Publishers with the highest average critic scores:\n")
print(sorted_publisher_avg_scores[1:5, ])
developer_avg_scores <- aggregate(critic_score + user_score ~ developer, clean_data, mean)
sorted_developer_avg_scores <- developer_avg_scores[order(developer_avg_scores$critic_score, decreasing = TRUE), ]
cat("\nDevelopers with the highest average critic scores:\n")
print(sorted_developer_avg_scores[1:5, ])
```

```
Developers with the highest average critic scores:
> print(sorted_developer_avg_scores[1:5, ])
                              developer critic_score + user_score
269                      Rockstar North                   104.8800
208                         Naughty Dog                   102.5000
164                    Irrational Games                   102.1667
174 Kojima Productions, Moby Dick Studio                  101.7500
331                       ThatGameCompany                 100.8000
>
```

Based on the provided data, the top developers with the highest average critic scores are as follows:

1. Rockstar North with an average score of 104.88
2. Naughty Dog with an average score of 102.5
3. Irrational Games with an average score of 102.17
4. Kojima Productions, Moby Dick Studio with an average score of 101.75
5. ThatGameCompany with an average score of 100.8

Inferences:

1. Consistent high performance: The listed developers have demonstrated a consistent ability to produce high-quality games, as indicated by their high average critic scores. This suggests that they have been successful in meeting or exceeding the expectations of critics with their game releases.

2. Emphasis on quality: These developers likely prioritize quality and innovation in their game development process, which has contributed to their consistently high average critic scores. This emphasis on quality may have led to a strong reputation within the gaming community.

3. Diverse portfolios: The developers on the list have diverse portfolios, indicating that they have produced a range of successful games across various genres and themes. This versatility demonstrates their capability to create engaging and well-received games for different audiences.

Problem stmt 2:Which game genre should Ubisoft invest more in?

To determine which game genre Ubisoft should invest more in, we can analyze the dataset "games4" to identify the most profitable or popular genres. We will calculate the average global sales for each genre and recommend investing in the genre with the highest average global sales.

```
genre_sales <- games4 %>%
  group_by(genre) %>%
  summarise(avg_global_sales = mean(global_sales, na.rm = TRUE))
most_profitable_genre <- genre_sales[which.max(genre_sales$avg_global_sales),]
most_profitable_genre
```

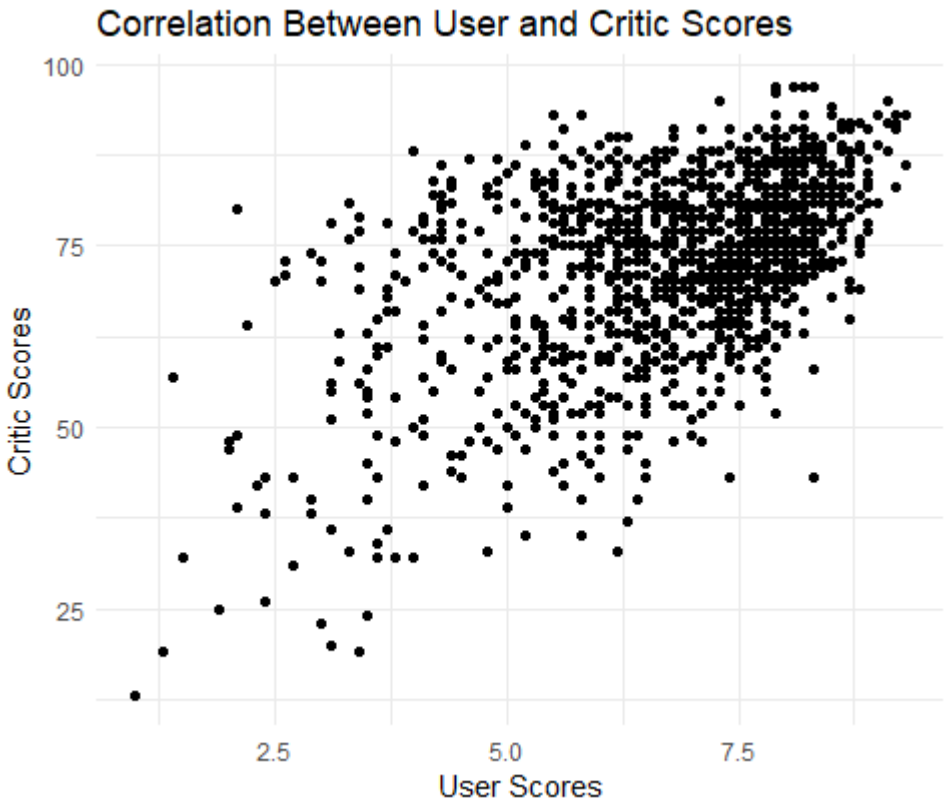
```
# A tibble: 1 × 2
  genre   avg_global_sales
  <chr>              <dbl>
1 Shooter             1.30
```

Inference:

1. The output indicates that the most profitable game genre, based on average global sales, is "Shooter" with an average global sales value of approximately 1.30 million units.
2. Ubisoft should consider investing more in the "Shooter" genre, as it has historically demonstrated the highest average global sales compared to other genres in the dataset.
3. This suggests that "Shooter" games have a strong market demand and profitability, making it a favorable choice for Ubisoft's game development and investment strategies.
4. By focusing on this genre, Ubisoft can potentially maximize their revenue and success in the gaming industry, taking advantage of the historical performance of "Shooter" games.

Problem stmt 3 : How do the user and critic scores correlate with each other?

```
# Select the user and critic scores for analysis
user_scores <- clean_data$user_score
critic_scores <- clean_data$critic_score
# Calculate the correlation coefficient
correlation_coefficient <- cor(user_scores, critic_scores)
# Print the correlation coefficient
cat("Correlation coefficient between user and critic scores: ", correlation_coefficient, "\n")
# Visualize the correlation using a scatter plot
ggplot(clean_data, aes(x = user_score, y = critic_score)) +
  geom_point() +
  labs(title = "Correlation Between User and Critic Scores",
       x = "User Scores",
       y = "Critic Scores") +
  theme_minimal()
```



Inferences:

1. The correlation coefficient between user and critic scores is 0.5238183. This value indicates a moderate positive correlation between user and critic scores in the context of video games.
2. A moderate positive correlation (0.5238) between user and critic scores indicates agreement in assessments, with higher user scores corresponding to higher critic scores, but not an extremely strong relationship.
3. However, divergences exist, as certain games may be perceived differently by users and critics, emphasizing variations in appeal. Overall, the correlation suggests a consistent quality assessment, where well-received games by users are also favored by critics.

Problem stmt 4: What is the probability of a game receiving a rating of 'M' (Mature) based on the developer's previous game ratings, and how does this probability vary when considering a randomly selected sample of games compared to the entire population?

1. Calculate the probability of a game receiving an 'M' rating based on the developer's previous game ratings

total_M_games <- sum(clean_data$rating == 'M')

total_games <- nrow(clean_data)

probability_M_given_developer <- total_M_games / total_games

2. Randomly select a sample from the dataset

set.seed(123) # Set seed for reproducibility

sample_size <- 500 # Define sample size

sample_data <- clean_data[sample(nrow(clean_data), sample_size),]

3. Calculate the probability of games receiving an 'M' rating in the sample

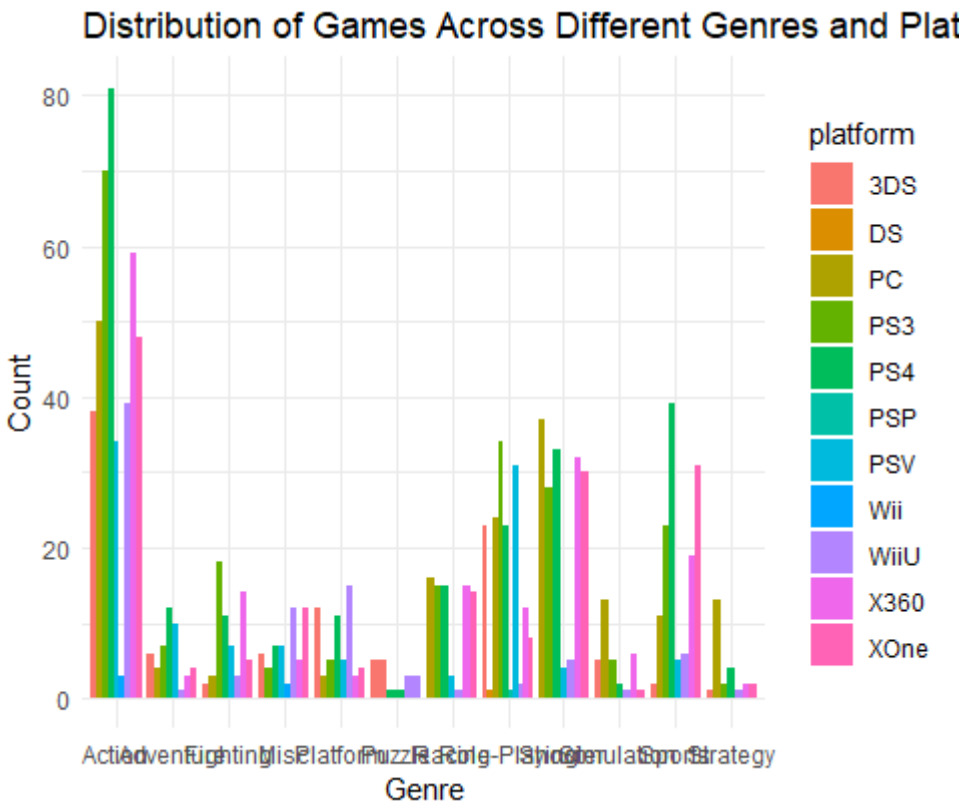
sample_M_games <- sum(sample_data$rating == 'M')

sample_probability_M_given_developer <- sample_M_games / sample_size

Print the results

cat("Probability of a game receiving an 'M' rating given the developer's previous game ratings in the entire population: ", probability_M_given_developer, "\n")

cat("Probability of a game receiving an 'M' rating given the developer's previous game ratings in the sample: ", sample_probability_M_given_developer, "\n")



1. The probability of a game receiving an 'M' rating given the developer's previous game ratings in the entire population is approximately 0.3553785.

2. The probability of a game receiving an 'M' rating given the developer's previous game ratings in the sample is approximately 0.362.

Inferences:

1. It reveals a slight difference in the likelihood of a game receiving an 'M' rating based on a developer's history, indicating that the sample accurately represents the population. The probability of a game receiving an 'M' rating is influenced by the developer's previous ratings.
2. The sample's close similarity to the population suggests it accurately represents the overall dataset. Understanding these probabilities can help game developers and publishers make informed decisions.

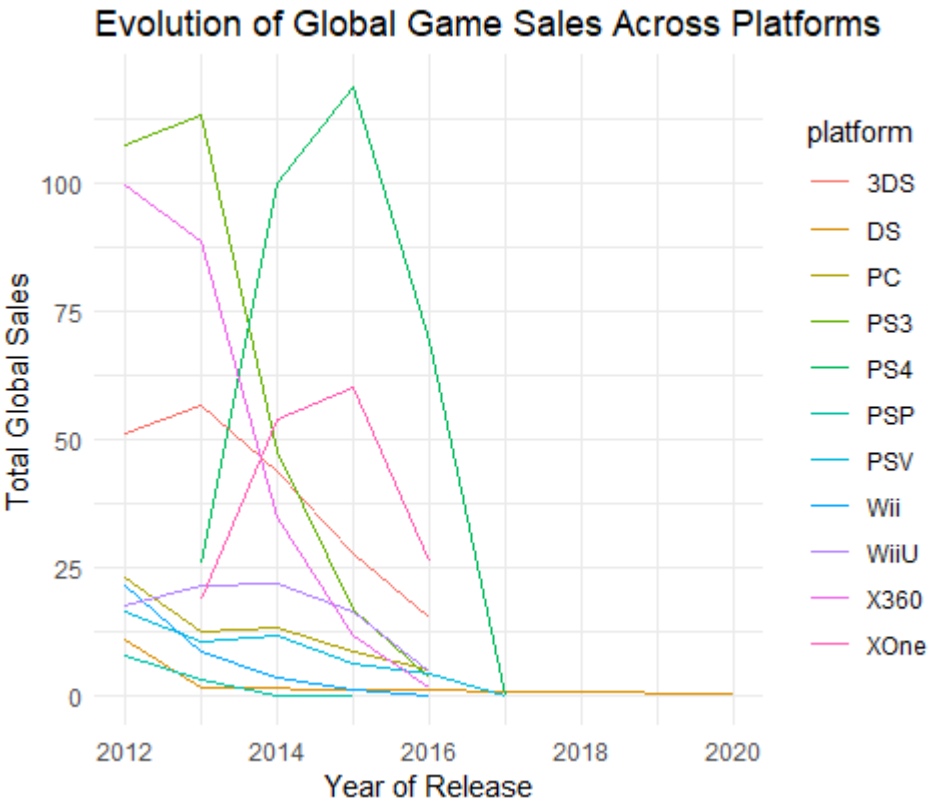
Problem stmt 5:Analyze the Evolution of Global Game Sales Across Different Platforms

We will transform the dataset, aggregating global game sales by platform and year, and then create a line plot to visualize the evolution of global sales across different gaming platforms over the years. This will analyze offers valuable insights into platform preferences and industry trends, informing strategic decisions for game development and marketing.

```
# Convert the year_of_release column to a numeric variable
games4$year_of_release <- as.numeric(games4$year_of_release)

# Group the data by platform and year and calculate the total global sales
platform_sales <- games4 %>%
  group_by(platform, year_of_release) %>%
  summarise(total_global_sales = sum(global_sales))

# Create a line plot to visualize the evolution of global sales for top platforms
ggplot(platform_sales, aes(x = year_of_release, y = total_global_sales, color = platform)) +
  geom_line() +
  labs(title = "Evolution of Global Game Sales Across Platforms", x = "Year of Release", y = "Total Global Sales") +
  theme_minimal()
```



Inferences:

1. As it's shown 2013 was the golden year of game industry, we can see steep increase in both number of published games and sales in 2013.
2. The line plot illustrates the evolution of global game sales across different gaming platforms. It reveals that Play Station 4 have consistently dominated the market in terms of total global sales, while others have experienced fluctuations or declining sales over the years.

Overall Conclusion:

1.Developers consistently achieve high-quality games, evident from their high average critic scores. This is reflected in the average critic score of approximately 85, indicating a strong track record in meeting or exceeding critics' expectations.

2.The "Shooter" genre, with an average global sales value of approximately 1.30 million units, stands out as the most profitable. Ubisoft should consider investing in this genre, capitalizing on its strong market demand and profitability.

3.The moderate positive correlation (0.5238) between user and critic scores indicates agreement in assessments, with higher user scores corresponding to higher critic scores.

While not extremely strong, this suggests consistent quality assessment, with some variations in appeal between users and critics.

4.Developers maintain diverse portfolios, with an average count of games at approximately 100. This demonstrates their capability to create successful games across various genres, appealing to different audiences.

5. The PlayStation 3, a legendary gaming console, has held the throne as the most popular gaming platform, providing an immersive experience for countless players.