**Player-Profiling Exploring Gameplay Metrics and Device Preferences**

"Player Profiling: Exploring Gameplay Metrics and Device Preferences" primarily centers around Exploratory Data Analysis (EDA) within the context of competitive gaming. In this project, I undertake a thorough examination of key gameplay metrics, including Wins, Losses, WL Ratio, average kills, average deaths, average assists, average KDR, and the influence of gaming devices. By employing EDA techniques, I aim to unravel patterns and insights that contribute to a holistic understanding of player performance. The project places a particular emphasis on revealing nuances in device preferences, providing valuable insights for individual players and gaming teams. Through a focused EDA approach, this project aims to enhance strategies, uncover trends, and offer actionable recommendations for success in the competitive gaming landscape.

**Feature Analysis**:

* **Wins and Losses:**

Exploration: The project extensively delves into the metrics of Wins and Losses, providing insights into the overall performance of players.

Patterns Identification: Analyzing patterns within Wins and Losses allows for the identification of trends, potential winning strategies, and areas for improvement.

* **WL Ratio (Win-Loss Ratio):**

Ratio Significance: The analysis goes beyond the ratio itself, unraveling the significance of a high or low WL Ratio in the context of player proficiency and effectiveness in the gaming environment.

Implications: The project explores how the WL Ratio contributes to a holistic understanding of a player's success and competitiveness.

* **Average Kills, Deaths, and Assists:**

Performance Metrics: The analysis delves into the detailed performance metrics of Average Kills, average deaths, and average assists, offering a nuanced perspective on a player's gameplay style.

Efficiency Evaluation: By examining the averages, the project assesses the efficiency and impact of players in terms of both offensive and defensive contributions.

* **Average KDR (Kill-Death Ratio):**

Strategic Insight: The project doesn't just calculate the KDR but delves into its strategic implications. It provides insights into the balance between offensive kills and defensive deaths, offering a well-rounded view of player performance.

Optimization Strategies: The analysis goes a step further by exploring how players can optimize their KDR for enhanced effectiveness in the gaming environment.

* **Device Played:**

Categorical Analysis: The inclusion of the categorical feature 'Device Played' adds depth to the analysis by considering the influence of gaming hardware on player performance.

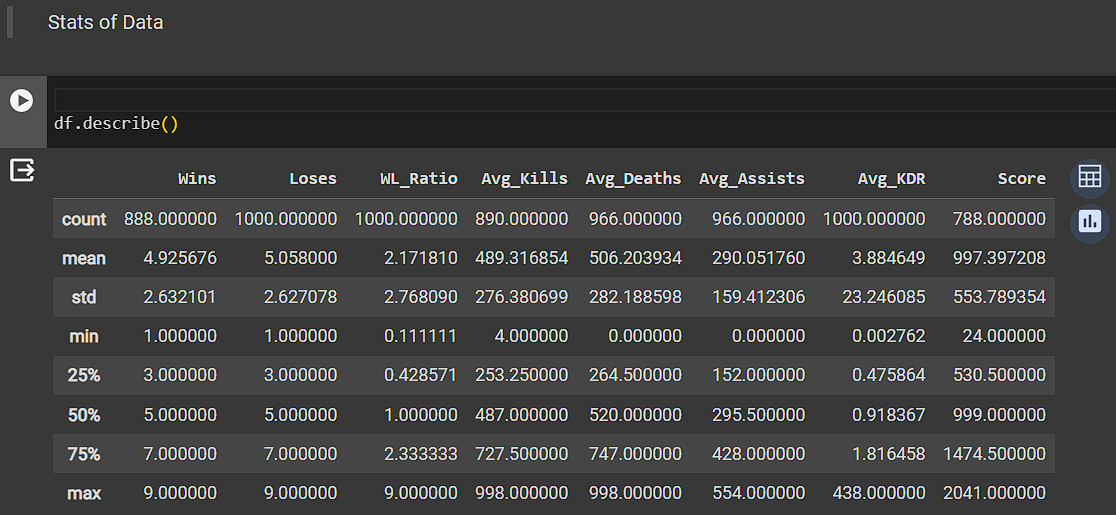
Device Impact: The project examines how the choice of gaming device contributes to gameplay outcomes, providing valuable insights for players and teams in optimizing their gaming setups.

* **Score:**

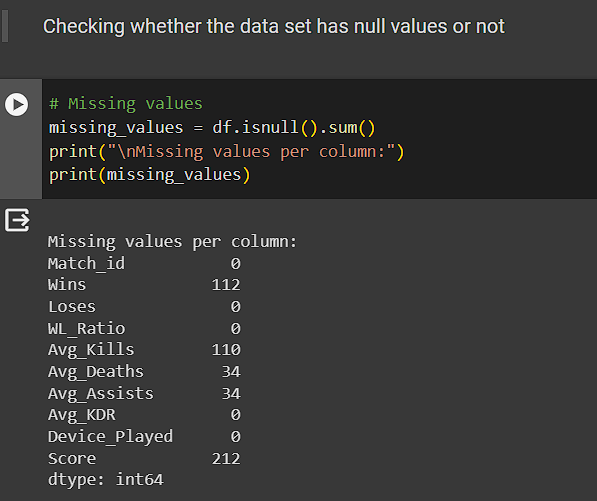
Overall Performance Indicator: The project interprets the 'Score' feature as an overall performance indicator, providing a holistic view of a player's achievements beyond individual metrics.

Score Trends: Through analysis, the project identifies trends in scoring, allowing for a comprehensive evaluation of player success and areas for potential enhancement.

**Statistical information:**

****

**Data Pre-Processing:**

****

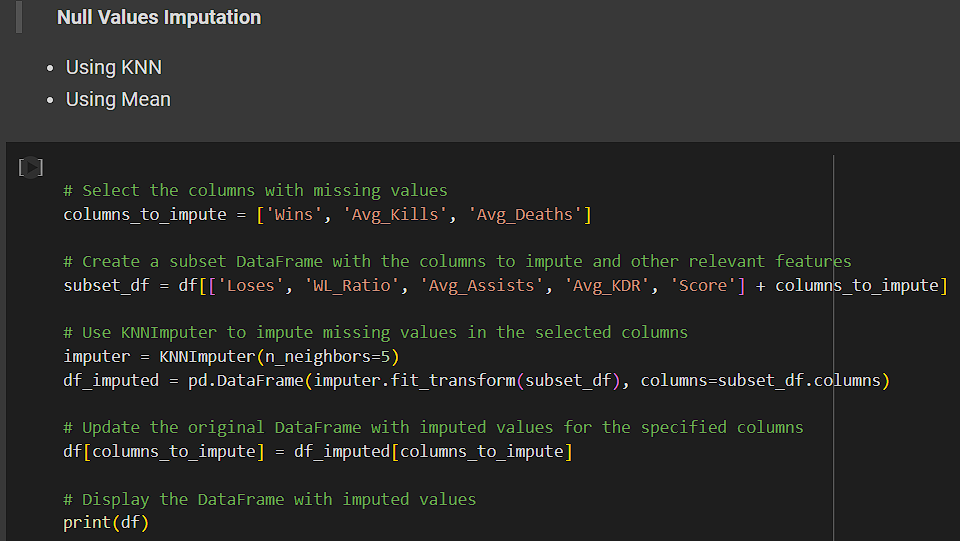
**Missing Value Prediction:**

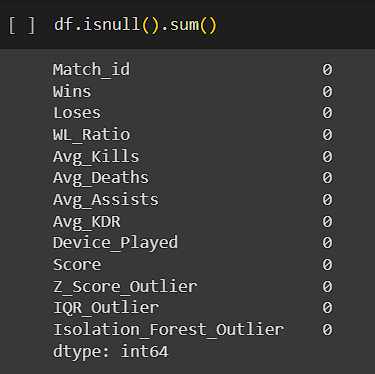
* **What happens if we have Missing values in the Data set?**

Having missing values in a dataset can make data analysis tricky. It might lead to biased results, reduce the accuracy of predictions, and make it challenging to understand the overall picture. Dealing with missing values properly is crucial to avoid drawing incorrect conclusions and ensure reliable insights from the data.

**Null Value Imputation Using KNN and Mean Methods:**

* **What is the KNN method?**

****K-Nearest Neighbors (KNN) is a versatile machine-learning algorithm for classification and regression. It predicts the target value of a new data point by considering the majority class or average of its K nearest neighbors in the feature space. The choice of 'K' and distance metric (e.g., Euclidean) is critical. KNN is non-parametric, adapting to various data distributions, but its scalability can be a limitation. It's suitable for tasks with irregular decision boundaries, is easy to implement, and requires careful consideration of features and scaling.

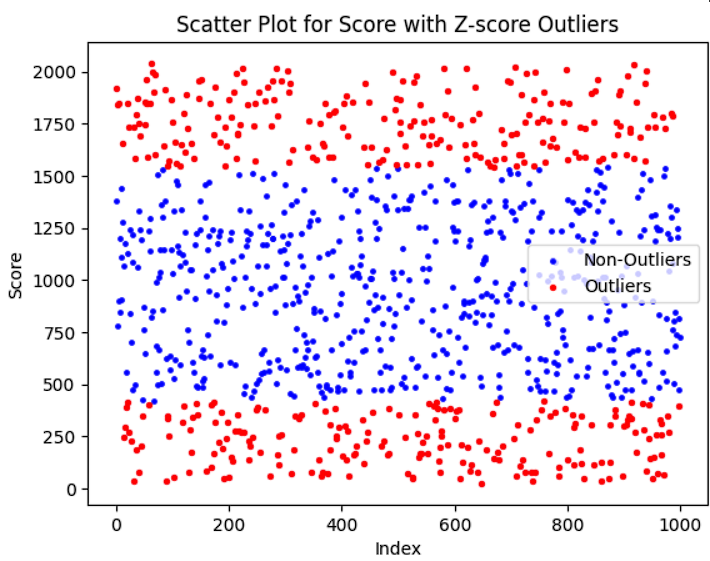
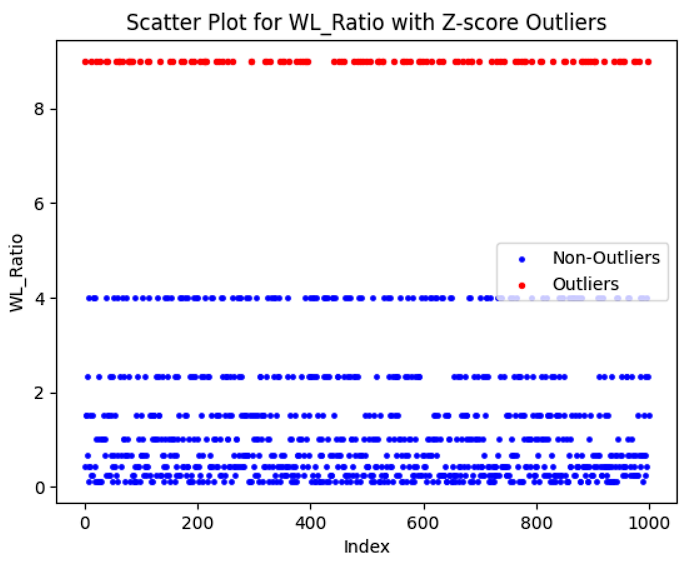
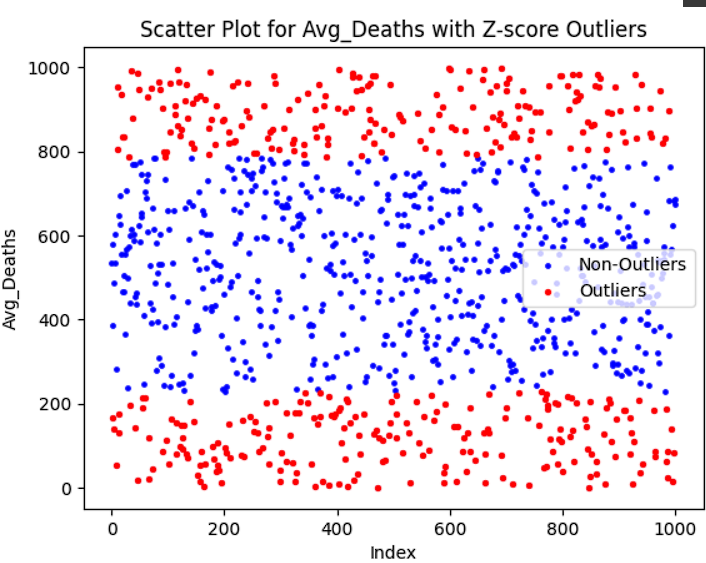
****

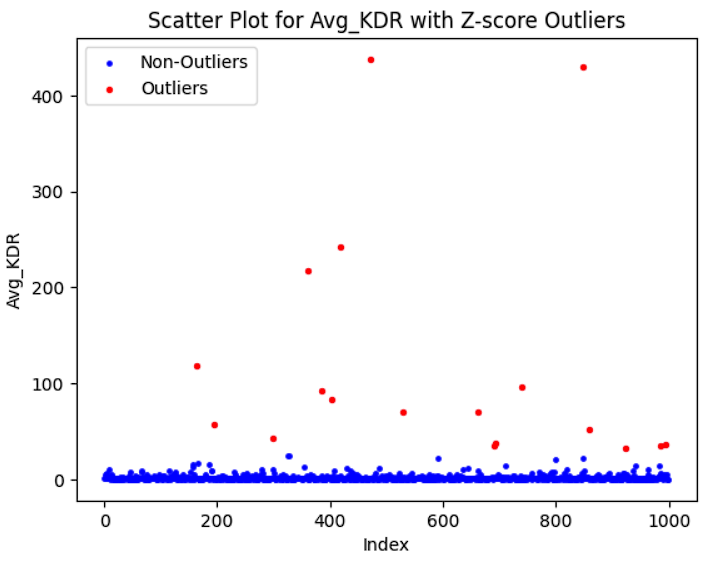
After Imputating the Null values we can test the data set by using “df.isnull().sum()” to check whether all the null values are updated or not as in the picture.

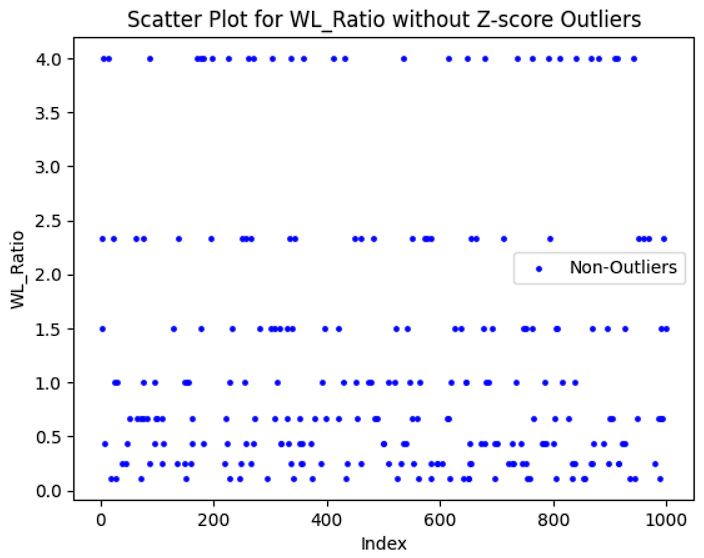
**Outlier Detection:**

* **What are Outliers?**
* Outliers are data points that deviate significantly from the rest of the data in a dataset.
* They are observations that lie at an abnormal distance from other values in a random sample from a population.
* In other words, outliers are data points that are unusually high or low compared to the majority of the data.
* Outliers can be the result of errors in data collection, or measurement variability, or they may indicate a real and important pattern in the data.

We can see in the images the red dots are identified as the outliers present in our data set.

* ****

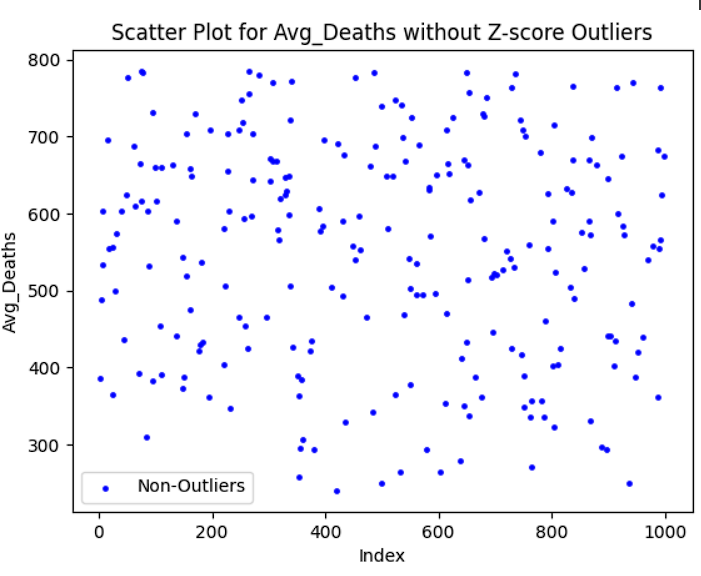
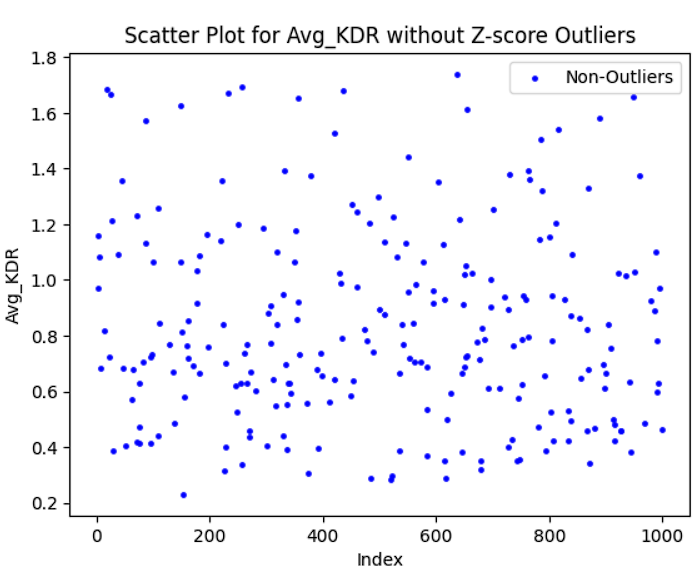


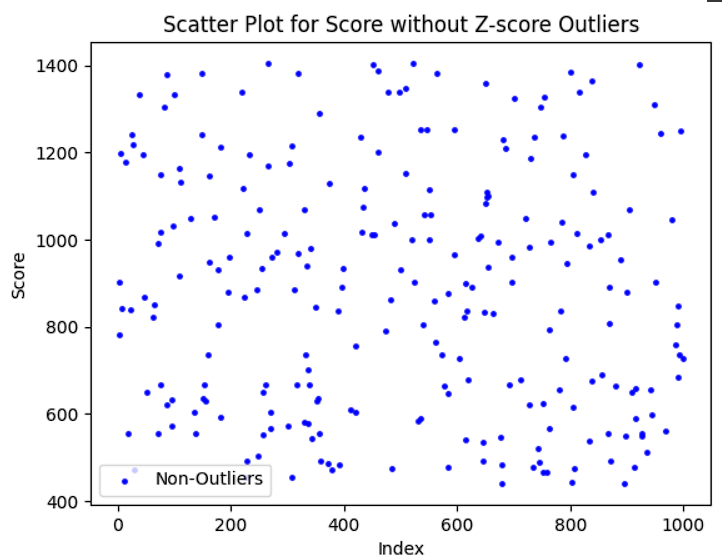
* **What and How is the Z-Score Method used?**

The Z-Score, or standard score, is a statistical metric that measures the relative position of a data point within a dataset by quantifying its distance from the mean in terms of standard deviations. It is calculated using the formula “(X-mu)/sigma” where X is the data point, μ is the mean, and σ is the standard deviation.

How Z-Score Works:

* A Z-Score of 0 indicates that the data point is exactly at the mean.
* Positive Z-scores signify data points above the mean, while negative Z-scores indicate points below the mean.
* The magnitude of the Z-Score represents the distance from the mean in standard deviations.

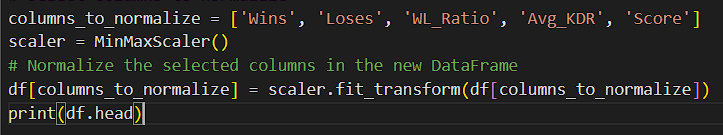




We can observe here that in these scatter Plots there are no outliers now

* **What is the Purpose of Normalization in data preprocessing?**

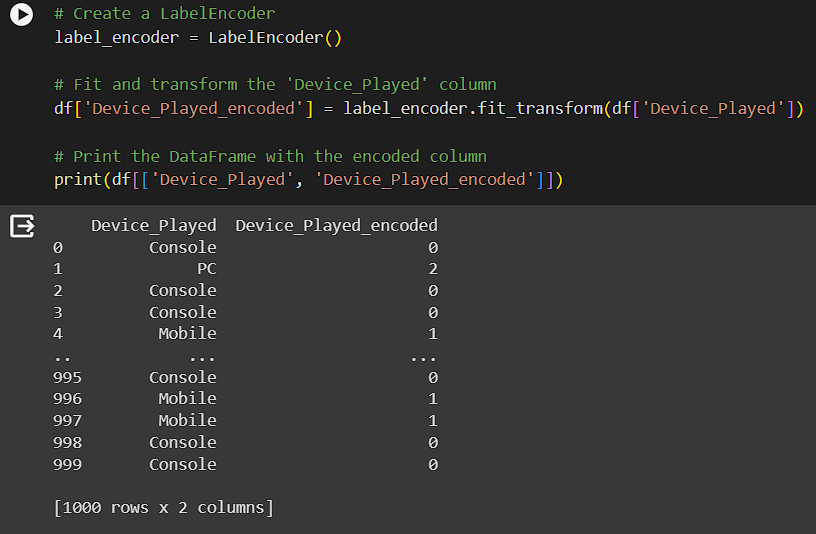
Normalization in data science serves vital purposes in preparing data for analysis and modeling. Its primary goal is to equalize scales among features, preventing variables with larger magnitudes from dominating the analysis. This is especially important for improving the convergence speed of optimization algorithms in machine learning models. Normalization also enhances model performance, aids in the interpretation of model coefficients, and contributes to the handling of outliers. It is crucial for clustering algorithms, statistical tests, and reduces sensitivity to initial conditions in iterative processes. Additionally, normalization supports data privacy by protecting sensitive information during sharing and facilitates neural network training. Overall, normalization is a foundational step in data preprocessing, ensuring data is appropriately scaled for diverse analyses and modeling techniques in data science.



In this picture, we can observe that I have normalized some of my features to get scalable ability easier.

* **What Role does Label Encoding Play?**

Label encoding is a technique in machine learning that converts categorical labels into numerical values, making them compatible with algorithms that require numerical input. It assigns a unique numerical identifier to each category, facilitating the transformation of non-numeric data into a format suitable for models like decision trees or support vector machines. Label encoding is particularly useful for ordinal categorical variables, where the order of categories matters. However, it has a limitation: it may introduce unintended ordinal relationships between categories, impacting the interpretation of the data.

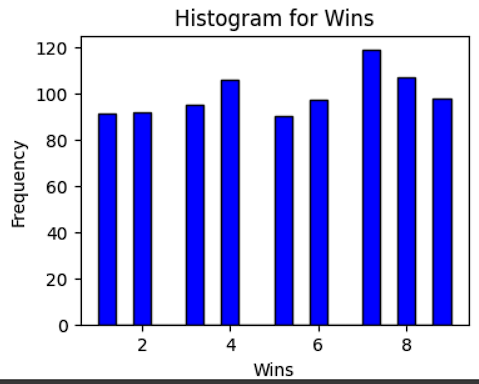


Label encoding is used to convert categorical values into numerical ones to create compatibility with algorithms. In my dataset, I have a feature called “Device\_Played”.

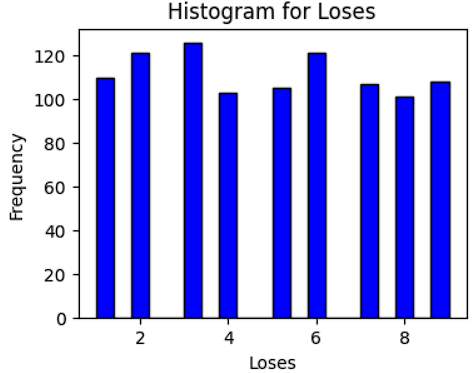
Which shows which device the user is more interested in playing matches. So there are three devices namely “Console, Mobile, and PC“ Which were coded to “0”, ”1” and “2” using this method

**Data Visualization:**

* **Histograms:**

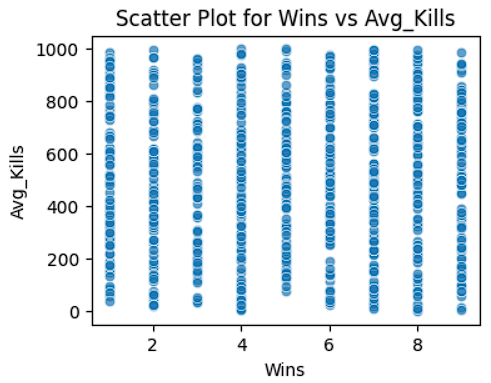
****

This histogram provides a visual representation of the distribution of wins. It appears that 5 wins is the most common outcome, while 6 wins is the least common. The frequencies of 2, 4, and 8 wins arerelatively similar. This could suggest that the data is somewhat skewed, with a peak at 5 wins.

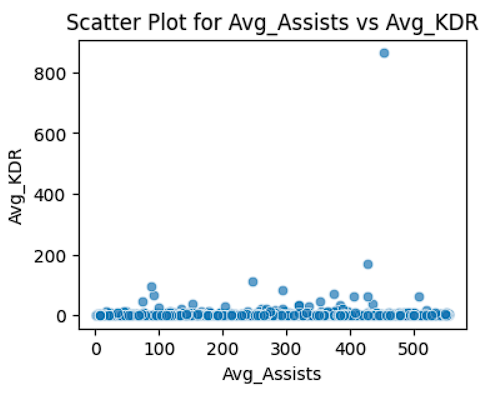
****

This histogram provides a visual representation of the distribution of losses. It appears that 0 losses is the most common outcome, while the frequencies generally decrease as the number of losses increases. This could suggest that the data is somewhat skewed, with a peak at 0 losses. However, there is a slight increase in frequency at 6 losses, which is an interesting observation. This could be due to specific circumstances or factors that led to a higher number of instances with 6 losses.

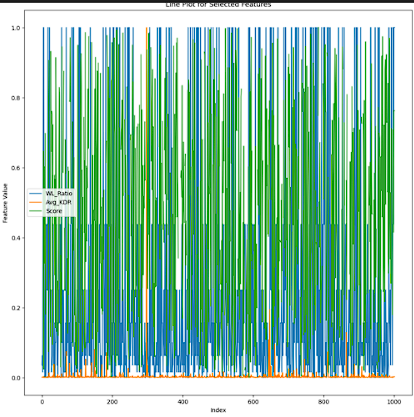
* **Scatter Plot:**



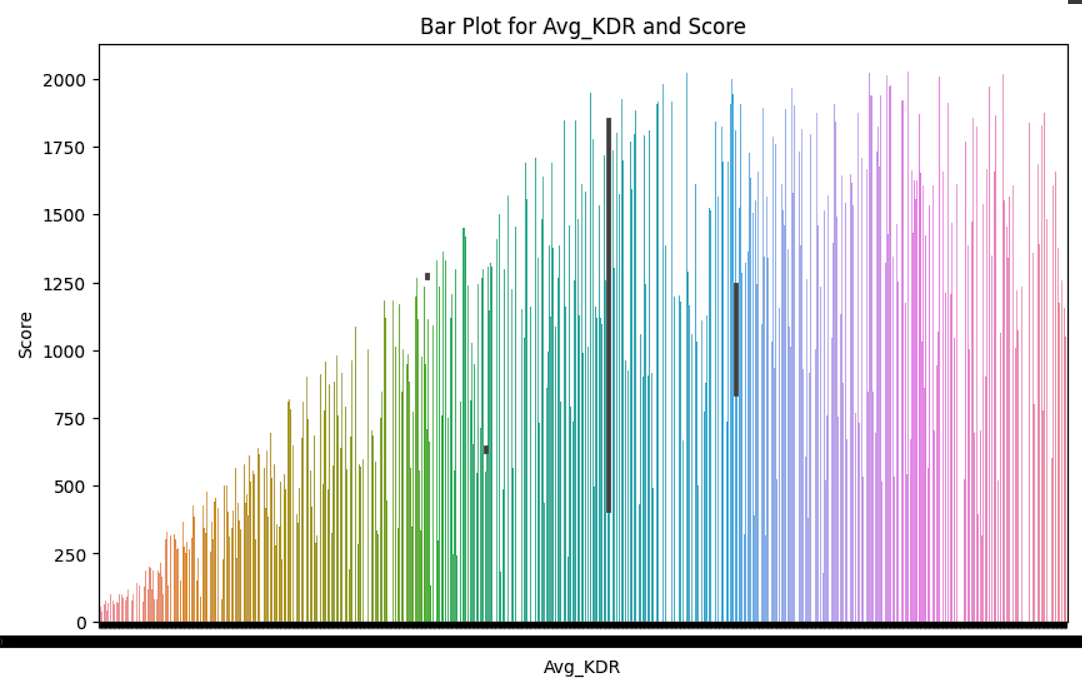
This scatter plot provides a visual representation of the relationship between wins and average kills. It suggests that a higher number of wins is generally associated with a higher average number of kills. However, there is variability in performance as indicated by the spread of data points for each win number on the x-axis. This could be due to specific circumstances or factors that affect the number of kills in each game.

This scatter plot provides a visual representation of the relationship between Avg\_Assists and Avg\_KDR. It suggests that an increase in Avg\_Assists does not necessarily lead to an increase in Avg\_KDR. However, there is an exception, as indicated by the outlier. This could be due to specific circumstances or factors that affect the Avg\_KDR in each game. The outlier with an Avg\_KDR of around 800 and a relatively low Avg\_Assists is particularly interesting and may warrant further investigation.

* **Line Plot:**

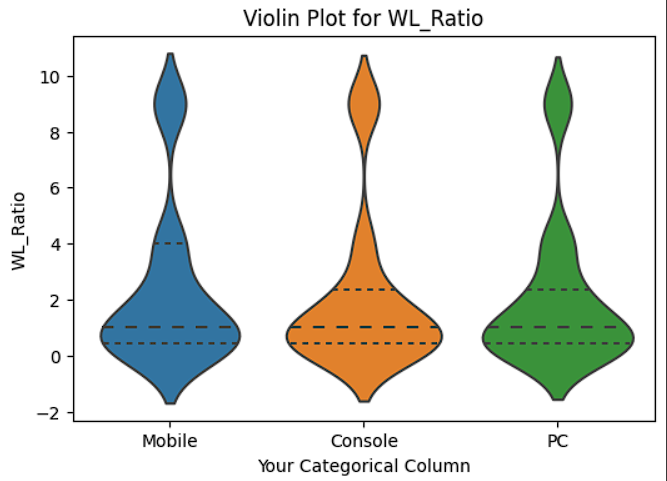
****

This line plot provides a visual representation of the relationship between the index and the three features: W/L Ratio, Avg. KDR, and Score. It suggests that these features vary significantly across the indices. However, there doesn’t seem to be a clear trend or pattern in the data, indicating that the relationship between the index and these features may be complex and influenced by various factors.

* **Bar Plot:**

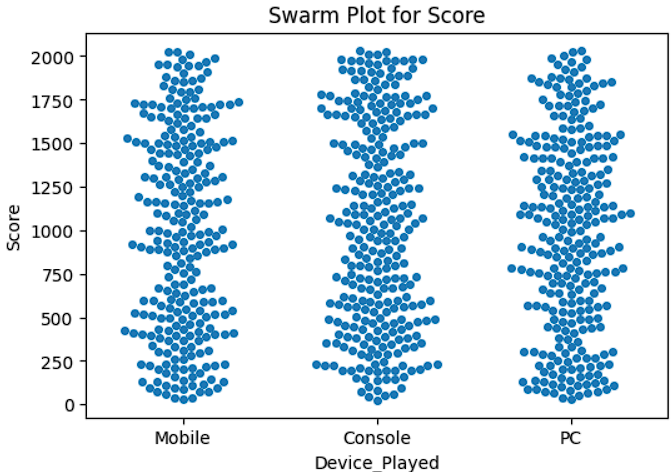
The plots indicate that the features Avg\_KDA and Score have complex relationships. The Score did not show a clear relationship with Avg\_KDA in the plots, but typically, a higher kill-death ratio could contribute to a higher score. However, an outlier was observed with a high Avg\_KDA but not necessarily a high score, indicating other factors may be at play. These relationships can be influenced by various other factors not included in these features, such as game mechanics and player skill level.

* **Violin Plot:**

****

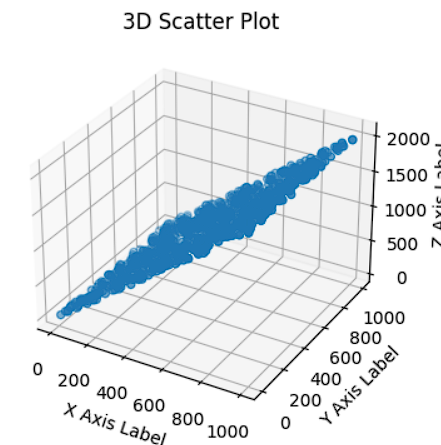
This violin plot provides a visual representation of the distribution of the Win/Loss Ratio across different gaming platforms. It suggests that the Win/Loss Ratio varies across different platforms, with more variability observed in Mobile and PC platforms compared to the Console.

* **Swarm Plot:**

The scores are distributed differently across the three categories. For Mobile, the scores are concentrated around the 500 and 1500 marks. For the Console, the scores are densely clustered around the 1250 mark. For PC, the scores are densely packed around the 1000 mark with some distribution towards higher scores.

This swarm plot provides a visual representation of the distribution of scores achieved on different gaming platforms. It suggests that the scores vary across different platforms, with more variability observed in Mobile and PC platforms compared to Console. the user gets insights from this plot

* **3D-Scatter Plot:**

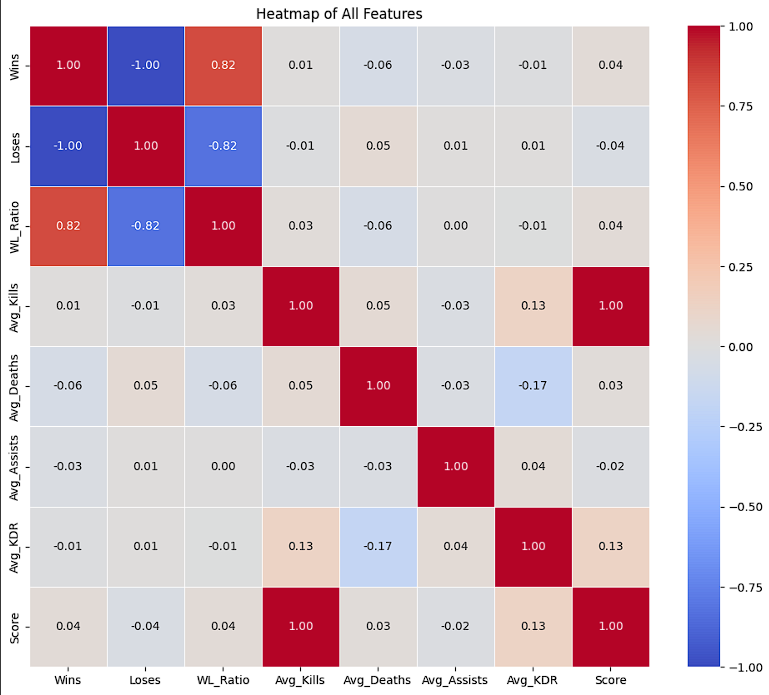


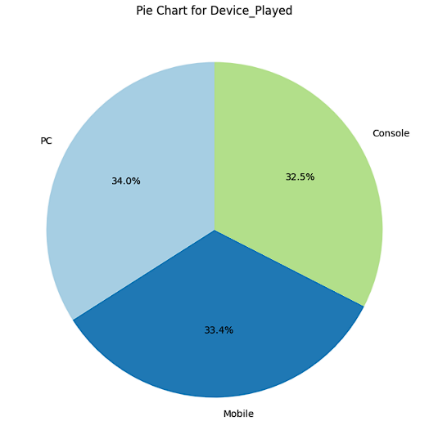
The plot has three axes, X, Y, and Z, each representing a different variable. The X-Axis ranges from 0 to 1000, the Y-Axis ranges from 0 to 800, and the Z-Axis ranges from 0 to 2000.

The data points appear to form a distinct linear pattern, suggesting a strong positive linear correlation between the three variables. As one variable increases, the other two also tend to increase.

This 3D scatter plot provides a visual representation of the relationships between three variables. It suggests that these variables are positively correlated with each other. However, without specific labels for the axes, it’s difficult to provide more detailed insights about the relationships between these variables.

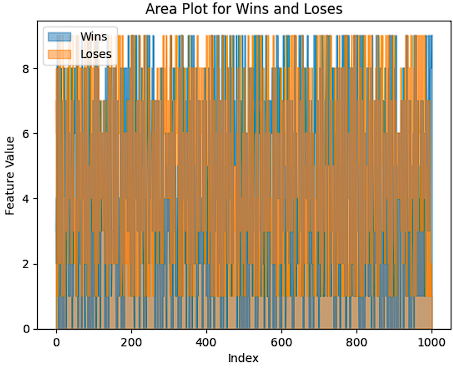
* **Heat Map:**

****

* Correlations: Positive correlations are indicated with shades of red while negative correlations are shown in shades of blue; the intensity of the colour indicates the strength of the correlation.
* This heatmap provides a visual representation of the relationships between various game statistics. It suggests that as the number of wins increases, the win/loss ratio also increases, and the number of losses decreases. However, the relationships between other features are not as straightforward and may require further analysis. the user gets insights from this plot
* Strong Positive Correlation: The strongest positive correlation visible is between W/L Ratio and Wins at 0.82.
* Perfect Negative Correlation: There is a perfect negative correlation (-1) between
* Wins and Losses indicating that as one increases, the other decreases.
* **Pie Plot:**

This pie chart provides a visual representation of the distribution of games played on different devices. It suggests that PC, Mobile, and Console are all popular platforms for games, with PC being slightly more preferred. user get insights from this plot

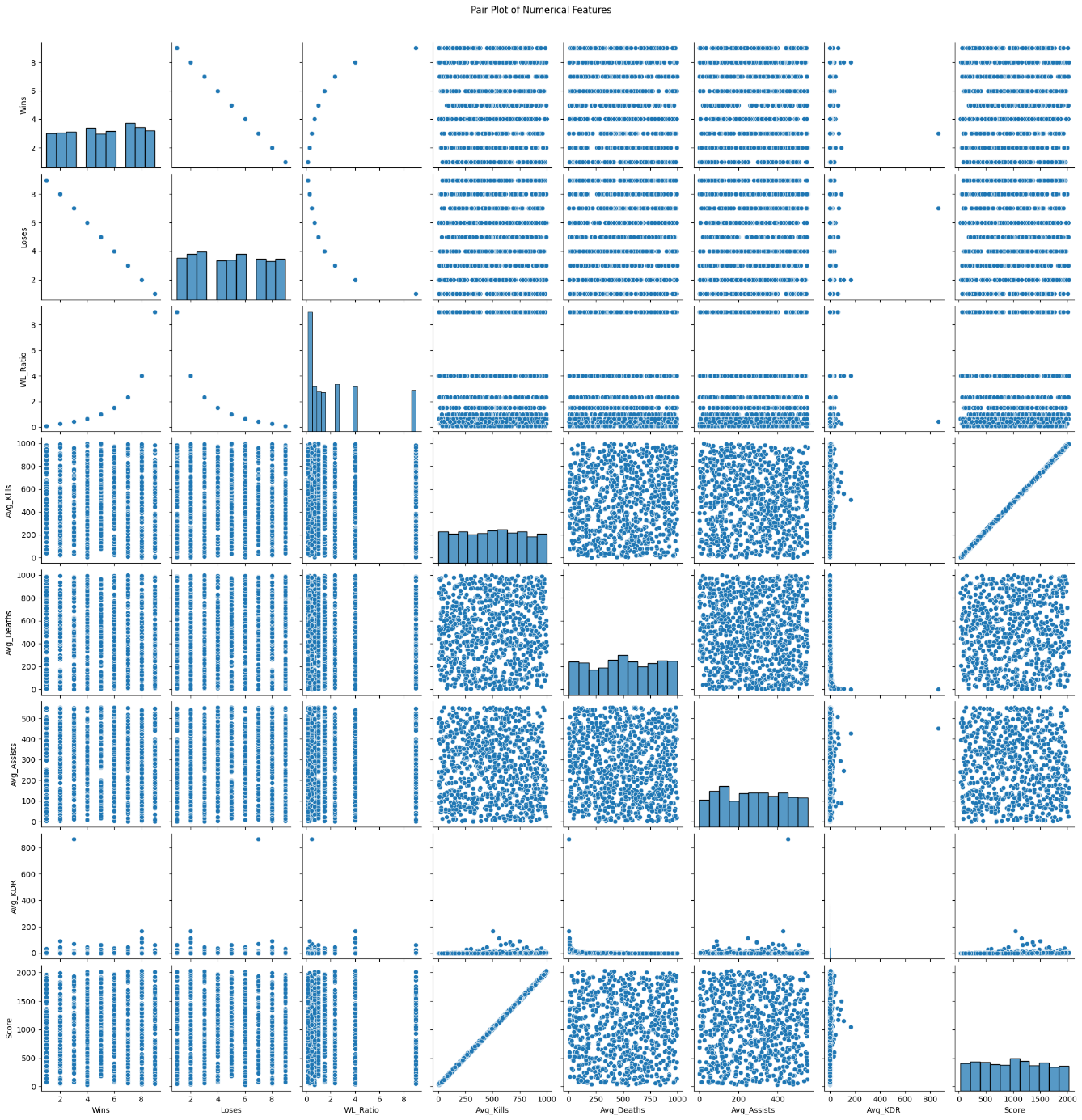
* **Area Plot:**

****

Data Areas: There are two areas representing Wins (blue) and Losses (orange) for each index.

Trend: The plot shows the distribution of wins and losses over a range of indexed data points. The blue areas indicate wins, while the orange areas represent losses.

This area plot provides a visual representation of the distribution of wins and losses over a range of indexed data points. It suggests that the number of wins and losses varies across different indices, indicating that the outcomes of the games are diverse and influenced by various factors

* **Pair Plot:**

This Pair Plot Consists of Relations among every features in graphical form. As there are 8 features we can observe from the graph there are in-toto 64 plots which depicts the relation between each other.

**Conclusion:**

The project's findings serve as a valuable resource for enhancing gaming experiences and shaping the future of the gaming landscape. In conclusion, the Player-Profiling project embarks on a journey to decode the nuances of gaming behavior, offering a data-driven exploration into the gameplay metrics and device preferences that shape the gaming community.