Data Understanding for 900,000 Hands of Blackjack Results

Gathering Data

The goal of this project is to analyze a simulated dataset of 900,000 Blackjack hands to identify patterns and develop strategies to optimize player wins. The dataset must include key attributes such as cards dealt to players and the dealer (up to five for each), the sum of their card values, and the outcomes of each round. Additionally, it must provide information about the monetary metrics for each game, such as the amounts won or lost by players and dealers. These requirements ensure the dataset can support detailed analysis of winning strategies, dealer behaviors, and gameplay patterns.

The dataset provided satisfies these requirements, containing all necessary fields to analyze the outcomes of Blackjack games. It was generated using a simulation designed to mimic real casino conditions, ensuring relevance to real-world scenarios. To streamline analysis and avoid biases, the data selection criteria include only games with complete records for both players and the dealer. Incomplete or ambiguous game outcomes are excluded. If computational efficiency becomes an issue, a random subset of the data will be used, ensuring that the sample remains representative of the overall dataset.

Describing Data

The dataset contains detailed information about each game round, including cards dealt to the players and dealer, the total value of those cards, and the outcome of the game. The outcomes are categorized by whether the player or dealer won, tied, or lost, along with the reasons for those outcomes, such as a bust or achieving a higher total. Monetary fields specify the amount won or lost by players and dealers in each round. Each round is uniquely identified, and data for six players is recorded per round. The structure of the dataset reflects a typical Blackjack game setup,

with clear distinctions between player and dealer actions and outcomes. The inclusion of monetary metrics provides an additional layer of analysis, allowing for evaluations of profitability and strategy effectiveness. Overall, the dataset appears well-suited for analyzing gameplay dynamics and deriving insights into optimal play strategies.

Exploring Data

Initial exploration of the dataset revealed several important patterns and trends. Card distributions align with expectations for a standard deck, with higher-value cards (such as tens and face cards) appearing more frequently due to their prevalence. Player and dealer hands tend to cluster around totals that maximize chances of winning, particularly those near 21. This suggests that the simulation incorporates realistic strategies and decision-making rules for both players and the dealer. But obviously, to be certain a deeper analysis is needed.

Initial analysis of game outcomes showed that players win approximately a bit less than half of the games, a proportion consistent with the house edge typically observed in Blackjack. Losses are most often attributed to players being outscored by the dealer rather than busting. The monetary outcomes, when plotted, show that most games result in small wins or losses, but there are a few games where players experience very large wins or losses. This reflects the high-risk, high-reward nature of certain Blackjack rounds.

Verifying Data Quality

The dataset was examined for completeness, consistency, and accuracy. All required fields, such as cards, totals, outcomes, and monetary values, were present in every record. There were no missing values, and categorical fields were appropriately labeled. Card values were within valid ranges for a standard deck, and the sums matched the individual card values, confirming data integrity.

Additionally, the distribution of outcomes and monetary metrics was consistent with expectations for Blackjack, suggesting that the dataset accurately reflects typical gameplay dynamics. Preliminary checks for outliers revealed a few rounds with extreme monetary values, which appear to be genuine results of high-risk bets rather than data errors. Overall, the dataset meets the necessary quality standards for analysis, providing a reliable foundation.