**Exploratory Data Analysis.**

**Outcome of EDA:**

* We have observed in our analysis that most of the players in our data have salary of less than 1 million dollars while highest salaries are above 6 million dollars for Baseball player.
* Most players have experience of 2 to 10 years in our, while some of the players have up to 20 years of experience.
* Majority of players in our data are while, Hispanic are lowest in number
* The salary of black ethnic players in our data is compatibly higher than other players, while salaries are strongly correlated to number of games played and career runs. It means that black players have better performance record in our data
* Career runs have 64% impact on salaries. While, Hypothesis Test have shown that players with more years of experience have significant impact on salaries.
* The regression model for predicting the salaries of players based on features like career runs, games played, and experience shows that we can accurately predict the salaries of players based on their previous performance and experiences.

**Missed during the analysis:**

* We can have a lot more features relevant to baseball players’ performance, such as batting average, player position, career hits, average score, and career fielding percentage, which can provide important insights into player performances.

**Variable that could help in analysis**

* We had a large number of additional variables in our data, including career at bats, career hits, career doubles, career triples, home runs, runs batted in, career batting average, career walks, career strikeouts, career stolen bases, career fielding percent, and white percent in city, black percent in city, and Hispanic percent in city, that can give important insights and improve the performance of regression model.

**Incorrect Assumptions**

* Our assumptions were correct for all hypothesis tests, probability mass function, and cumulative distributive function. Data analysis reveals significant insights into player performance and how it affects salaries.

**Challenges faced:**

* We faced difficulty in carrying out the hypothesis tests and regression analysis. Our data is highly skewed, so we had to conduct extensive research to find the best statistical test for our analysis. The regression model for forecasting player salaries produces reliable results, and we can improve its performance even further by including more relevant features in our data.