

- **Project Name and Team Member:**
 - **Name:** NBA Players' Stats Analysis
 - **Member:** Chuanzhou Zhang, Kenneth Chan
- **Problem to Solve:**
 - What stats are important to NBA players?
 - Can we predict player salaries based on their stats? If accuracy is low then other factors may contribute to salary. (player's popularity which is not quantifiable)
 -
- **Data Collection:**
 - Scrape players stats from the ESPN player stats API
 - Scrape hoopshype for players current and future salaries.
 - Merge the data from the two websites based on Players Names.
- **Data Cleaning:**
 - Add team tag to each player
 - Merge the two different datasets (Join id is player name)
 - Convert Categorical data. For example, players' positions can be changed into numbers for easier analysis.
 - Handle missing values and outliers that do not make sense. (salary data missing)
 - Tackle name variation problems.
- **Analysis and Visualizations:**
 - **Correlation (player stats - salary):** Determining what player stats affect player's salary.
 - **Predict:** predict player salary base on their stats
 - Implement regression analysis to identify statistically significant factors influencing salaries.
 - Create correlation matrix to identify initially which stat is the most important factor for determining salary.
 - identify which statistic a player needs to improve for a better chance at a higher salary.
 - Time series analysis: player performance over time. Check if the player's stats are consistent overtime.
- **Visualization**
 - Create scatterplots for different stats against the salaries to identify outliers which show which players are underpaid vs which are overpaid.
 - Create a time series chart for some players stats and their salary over time, showing how their salary increases with the improvement of said statistic.
 - Heatmap to show the correlation matrix.
 - Bar Chart to show the importance of each factor.