**GROUP PROJECT PROPOSAL**

**Team members:**

Harini Chaturvedula

Prakash Manne

[Venkat Akshay Reddy Jaggavarapu](https://unt.instructure.com/groups/103671/users/272129)

[Uday Raj Suthapalli](https://unt.instructure.com/groups/103671/users/272427)

**Problem Statement:**

Finding the best player is important for a baseball team, as it is often a relief to avoid overpaying the wrong players. Often, teams take the risk with the wrong players and literally pay for it. We will use a sports case study where data for the year 1986 has been provided to analyze the salary of baseball players based on various factors that impact the salary of the players.

**Description:**

There are multiple factors that decide whether a team or a player has a reasonable salary offer. It is important to analyse past performances and trends to determine which statistic such as hits, home-runs, runs battled in, errors or put outs had the highest impact on earning better salary.

**Discussion:**

The dataset has several independent variables that impact player’s salary. Using statistical analysis, we can understand how changes in an independent variable can have an impact on the dependent variable.

**Independent variables:**

1. Runs battled in 1986
2. Earned Run Average in 1986
3. Errors in 1986
4. Put outs in 1986
5. Home runs in 1986
6. Hits in 1986
7. Career Hits
8. Career Home runs
9. Career RBIs
10. Career runs

**Dependent Variable:**

1. Salary of the players

**Teamwork and Contribution:**

1. Harini Chaturvedula: Will be working on data wrangling and data visualization since the data has missing values and duplicates that can hinder with the analysis of data. Therefore, the data needs to be cleaned up and processed to ensure that all the rows and columns have relevant data. After data processing is done, we will load our data into the SAS tool for data visualization.
2. Prakash Manne: Will be working on statistical and mathematical analysis where correlation between the variables is achieved by running the data in SAS tool using several simulations.
3. [Venkat Akshay Reddy Jaggavarapu](https://unt.instructure.com/groups/103671/users/272129): Will be working on creating multiple patterns using statistical analysis to derive a conclusion.
4. [Uday Raj Suthapalli](https://unt.instructure.com/groups/103671/users/272427): Will be working on documenting and communicating the progress of the project.

**Collaboration plan:**

| Date | Tasks | Mode of Communication |
| --- | --- | --- |
| 11/12/2021 | Project Proposal | Microsoft Teams |
| 11/14/2021 | Data Exploration | Microsoft Teams |
| 11/19/2021 | Data pre-processing, includes data wrangling and cleaning | In person |
| 11/24/2021 | Data Visualization | Microsoft Teams |
| 11/28/2021 | Statistical Analysis | In person |
| 12/01/2021 | Documentation | Microsoft Teams |
| 12/02/2021 | Conclusions | Microsoft Teams |
| 12/03/2021 | Final documentation | Microsoft Teams |
| 12/05/2021 | Review | Microsoft Teams |