Levi Nickerson

109340569

CSCI 4830

Baseball Lineup Projector

Domain Description:

In recent years daily fantasy has become a popular hobby that many people enjoy. In particular, daily fantasy for baseball has become popular. This is because of how many games are played and the fact that there is an MLB game every day for almost two thirds of the year. If you aren’t familiar with daily fantasy it consists of a participant selecting a contest and creating a lineup. There are numerous contests every day with slight variations, but the main idea is that the high score wins. To create a lineup a participant selects a roster of players from the pool of teams playing that day. Each player is given a “salary” and a “total salary” is determined per contest. The contestant must create a roster that does not exceed this “total salary”. Once this is completed, the games are played and a winner is determined based on the statistics of their selected roster. This program is intended to help daily fantasy users select an efficient lineup each and every day. The user of this program is most likely a sports fanatic who is looking for help with daily fantasy.

Dataset Overview:

The core data set that will be utilized is the statistical information regarding MLB players. This data is updated daily and contains statistical data for the entire season. The data consists of various statistical categories for batters and pitchers. This data can be downloaded as an Excel spreadsheet and then analyzed. This is initially a large dataset that consists of every MLB player that has made an appearance at some point in the season. This is over 2000 players, but it can be quickly sorted down to hundreds of players for each day. This data set is also coupled with the data set that is taken from the fantasy league. This data set corresponds to the numerical salary of each player and the total salary allowed for the contest. Similarly, it is downloaded as an Excel spreadsheet and consists of all of the players that have made an appearance during the season that are playing on that specified day. The primary objective of using these two data sets is to be able to compute the most ideal lineup for the day. To compute the ideal lineup the Effectiveness Variable will be computed. This is a derived variable that will be computed using the key statistics for each player. Potential key statistics include: batting average, home runs, singles, doubles, triples, stolen bases, walks, innings pitched, strikeouts, wins, etc. This is the key variable for analysis.

Task Description:

The user of this system will be able to generate graphics that will allow them to pick an ideal lineup. The fun of daily fantasy is having the opportunity to flex your baseball knowledge, but it is impossible to know every lineup and matchup. The program is intended to step in and perform the heavy lifting. The user can input the two data sets and the program will output various things. First, the program will compute an effectiveness value for each player. This is a value that is dependent on a players batting or pitching stats. Second, this value coupled with a player’s salary will be used to create a 3D visualization. This 3D visualization will allow the user to delve into the data and select outliers that may be potential good picks for the day. Finally, the program will produce is theoretical lineups based on altering the computation of the Effectiveness Variable.

Vis Design Ideas:

There are two primary visualizations that will be used. The first (Figure 1 on following drawing) is a 3D representation of Salary versus Effectiveness Variable for each player. Two graphs will be produced, one for batters and one for pitchers. The graph will have sortable axis so that important athletes can be identified. The user can manipulate and maneuver through the visualization to better understand the data. The second visualization (Figures 2 and 3) is a series of tables. These tables will produce ideal lineups conforming to different Effectiveness Variables. The Effectiveness Variable can be altered by changing the weighting of the algorithm. For example, the weighting can favor home runs which would produce a lineup with players who are expected to hit home runs. For pitchers the algorithm can favor pitchers who have a tendency to accumulate a large number of strikeouts. These tools will highlight a great number of athletes that should perform well that day. Using these outputs the user has more than enough information to select a lineup that they would enjoy watching and has a good chance of performing well.

Usage Scenario:

Once the design has been implemented, a typical user can use it to determine the ideal lineup for the day. The user is required to import the current datasets for the day into the program and run it. This requires downloading both datasets and then importing them into Python. Once the datasets are in Python they must be input into the program. Once the program has run it will calculate Effectiveness Values and produce its visualizations. The visualizations will produce the information necessary to generate an ideal lineup. It will produce information regarding the most cost effective pitchers and batters as a 3D visualization. Using this visualization, the user can sort the data and determine effective players. It will also produce theoretical lineups depending on manipulations of the Effectiveness variable. All of this information will quickly inform the user and allow them to select a lineup that will perform and one that the user will enjoy watching.