Weather and Carlos Santana's Effectiveness

Term Project Milestone One

Matthew Heinen

Bellevue University

DSC540: Data Preparation

Professor Williams

7/14/2024

Project Subject Area

This project is inspired by the traditional "slow start" to the year commonly exhibited by Major League Baseball veteran 1st baseman Carlos Santana. Santana has played 17 MLB seasons and often attributes his slow start in the early months of the season to weather conditions. The purpose of this project is to combine Santana's statistics for the balls he puts in play with Statcast data as well as weather conditions to see if we can explain and eventually predict Santana's performance.

Data Sources:

Flat File:

This File is a CSV that contains all available Statcast data for every ball put in play during road games over the course of Carlos Santana's 17-year Major League baseball career. The file contains over 60 different data points over 2,831 events (balls put in play). Data recorded includes the pitcher, teams involved game, date, pitcher handedness, game situation, pitch type, and velocity/ pitch break information. Additionally, this dataset contains various hit information including, hit result, location, exit velocity, launch angle, swing length and hitter handedness (Carlos Santana is a switch hitter).

Carlos Santana Road Balls in Play Statcast Search. baseballsavant.com. (n.d.-a).
https://baseballsavant.mlb.com/statcast_search

API:

The Historical Weather API receives a latitude, longitude and time and will return historical weather information from that specific time or period. The API returns: temperature, feels like temperature, air pressure, humidity, wind direction, wind speed, precipitation as well as a weather condition and description.

Historical weather API - OpenWeatherMap. (n.d.). Openweathermap.org.
https://openweathermap.org/history

Website:

This project will require two website tables. The first from Wikipedia which includes all active MLB team names, the name of their ballpark and if their stadium is open air, retractable or fixed. This page is essential in linking the team's name to the ballpark. It also provides helpful information about the roof type.

The Steamheads page cited below contains a table in which every professional ballpark is listed with corresponding information. The page includes the ballparks name, city, first and last games, seasons, games, latitude, longitude and altitude. Latitude and longitude will be essential for making an API call. Additionally, the city ballpark altitude could prove helpful in our modeling.

Seamheads.com Ballparks Database. (n.d.). Www.seamheads.com. Retrieved June 15, 2024,
    from https://www.seamheads.com/ballparks/index.php

Wikipedia Contributors. (2019, November 8). *List of current Major League Baseball stadiums*.
    Wikipedia; Wikimedia Foundation.
    https://en.wikipedia.org/wiki/List_of_current_Major_League_Baseball_stadiums

Relationships

In order to tie all these relationships between all three data sources several different join methods. The flat file and web pages will be linked by the home team name in the Wikipedia page. The flat file however uses the team's abbreviation code. To link to the full team's name listed on the website a dictionary will need to be created with the team code as the key and the team's full name as the value. Using this dictionary a left outer join can be created to link each event to the corresponding home ballpark. This file will then be linked to the Steamheads page using a left outer join on stadium to pull the latitude and longitude.

To pull the weather conditions at a given time I will use the information from the game date in the flat file and ballpark latitude and longitude to pull weather data from the API for each game event. The working file will then store all the pertinent Statcast and corresponding weather information.

Project Approach

To evaluate Carlos Santana's perceived underperformance in the early season and cold climates we first must establish if he does indeed underperform in those conditions. We will attack this problem using descriptive stats and t-tests to see if there is indeed a statistically significant difference in performance. T-tests could include is there a difference in how hard (exit velocity) Santana hits the ball when it is above or below 70 degrees. Another test may be is Santana more likely to strikeout in wet conditions.

Secondly, we will seek to model Carlos Santana's performance; Can we build a model using historical weather data that will reveal more trends in his performance that may allow a team to better utilize his talents or make more informed roster decisions.

Concerns and Challenges

In executing this project one challenge that I anticipate running into is performing the API call for nearly 3000 game events to get the weather for every ball put in play. The API only allows for 1000 free calls per day, Therefore, I will either need to pay for access or spread my calls across 3 days.

Each game event also only has a date and not a time of day. Therefore, I will have to make assumptions of the game time. Naturally the weather conditions can change throughout a game therefore this will limit the power and credibility of my project. I will likely simply call for the game start time and assume that that will be close enough to the weather at the time of the at bat.

Ethical Implications

Ethically, depending on the success of this project the trends and information revealed could affect Carlos Santana's contract negotiations moving forward. If this project finds that Carlos is an ineffective player in cold climates it could limit his ability to sign in colder markets, thus limiting his earning potential.

On a larger scale, the findings of this projects could easily be replicated and applied to both other major leaguers and prospect players. Potentially effecting the earning potential of already established major league players. Regarding the minor leagues, trends revealed from this project could affect the resources teams are willing to commit to player development either

positive or negative. Widening the scope even further: It is possible that amateur players are passed on in the draft and minor league contracts if this project reveals that some players are ineffective in certain climates. Thus, altering the earning potential of athletes who have dedicated their lives to chasing a career in professional baseball.

Works Cited

Carlos Santana Road Balls in Play Statcast Search. baseballsavant.com. (n.d.-a).
https://baseballsavant.mlb.com/statcast_search

Historical weather API - OpenWeatherMap. (n.d.). Openweathermap.org.
https://openweathermap.org/history

Seamheads.com Ballparks Database. (n.d.). Www.seamheads.com. Retrieved June 15, 2024,
from https://www.seamheads.com/ballparks/index.php

Wikipedia Contributors. (2019, November 8). *List of current Major League Baseball stadiums*.
Wikipedia; Wikimedia Foundation.
https://en.wikipedia.org/wiki/List_of_current_Major_League_Baseball_stadiums