Speaker Bio:

Joe Tucci is a Library and Information Science at the University of Illinois, where is studying to become a public librarian. He recently got his undergraduate degree from Eastern Illinois University in English and had not previously studied any kind of data science or computer science. Currently he works at the iSchool’s Help Desk providing technical assistance to students, faculty, and staff of the iSchool.

Title:

Heat, Dot Dogs, and Happiness: The External Factors that Affect Baseball

Description:

There’s so much data that comes from baseball that transforming that data to determine, for example, the worth of a center fielder can get even more specific with advanced stats like sabermetrics. Teams are often looking at these advanced stats when considering salaries and trades. But what if teams put as much consideration into the quality of their hot dogs at games as they do statistics? So, what are the external factors that contribute to the game of baseball? Those external factors include the food at stadiums, average fan attendance in a given year, weather data from each city the team plays in, and the success of a team’s season and player’s success.

Baseball data comes in droves for any kind of statistic, and this project pairs together baseball team statistics with those external factors. To get data on food and facilities at a stadium, there’s data on health code violations at each stadium. Fan experience was also a factor when think about crowd noise and overall joyful experience of the game and reviews from the internet served as the basis for this data. Weather including average monthly temperature and precipitation was also taken into consideration as well.

The result of combining all this data will involve basic baseball team statistics like their win-loss record, team batting average, and team ERA, combined with the external factors mentioned previously like weather and food. This data will not provide or create any kind of aggregate score or rating of teams or teams’ stadiums but rather provide insight into different kinds of data for each team that will have a different result for each team and stadium.

Who and Why:

This data project was primarily for anyone interested in data aggregation and data mashing. Because the main topic is on the American sport, baseball, a general knowledge of the game. Audience would benefit from knowing the length of a typical season, 162 games, and that only half of those are played at home in a team’s stadium. The attendance is based on season averages and this is reflected in the data as well.

The audience will gain knowledge on data mashing and the different kinds of data that can be mashed with baseball. This project is mostly for fans of baseball and the different kinds of data that baseball brings along with it. Though non-fans of baseball or any sport will appreciate the variety of different datasets that are mashed together with great overlap.

Outline:

Intro (5 minutes)

Here I’ll introduce the project and explain why the baseball data and other data will combine and mash into one dataset.

Baseball and stats (5-7 minutes)

Overview of the baseball dataset and explanation of the slight changes to data like only looking at a team’s home record.

Fan Review Scores (5 minutes)

Explanation of what the numbers mean in this dataset, why I took the average of the four categories, and how it relates to the overall project.

Weather Data (5 minutes)

Brief overview and analysis of the weather data, describing a few of the changes I had to make to the city data, and how it fits into the project.

Analysis of the final dataset (10-15 minutes)

Rundown of the entire dataset, going over most of the columns and a few rows of data going over the row column by column.

Notes:

Most of the presentation will be a rundown of the Jupyter notebook. It should be worth noting that this data is all for the year 2017 and is really only a small sample size to build off of in the future. The data in the final dataset is also not grouped together or organized in any specific manner as well.