Capstone Project Declaration

Introduction

The NFL (National Football League) is one of the most popular and successful sports and entertainment leagues in the world. In 2019 the 32 league teams garnered more than $15 Billion dollars collectively in revenue from various avenues such as ticket sales, media contracts, sponsorships and concessions among others. The success of the league and each of its teams has a huge positive impact on local economies in many ways.

From a host city's perspective there is great benefit from increased sales tax and property tax revenues, increased real estate investment and development, elevated potential to create new jobs and higher prospect of luring corporations to move to the area.

From a fan perspective, the environment around the game is more than just the game itself. NFL fans like to enjoy themselves before and after the game though socialization with other fans. In other words, NFL fans like to eat and drink! Cities that have developed healthy availability of options for their NFL fans within the area of the game are poised for success in keeping a fan-base happy and incurring more revenue for their city coffers.

To have an NFL team franchise is a jewel in the crown of any city lucky enough to have one and a well-sought-after prize for those who do not. Hence, it is critical that cities understand their unique position and the importance of knowing where they stand in comparison other NFL cities on the basis of creating a fan friendly atmosphere that makes the fan happy, and in turn, makes it more enticing for them to want to attend a game in person, thus increasing the revenue for the team.

The goal of this report is to analyze the availability, variety and quality of food related venues within walking distance to the game to help cities understand how they compare to their fellow member cities. As a cursory benefit it is the hopes the analysis will also aid the fan to understand what options are available, especially if they might intend to travel and support their team away. We will also examine if winning has any correlation with how fan friendly a city can be.

IBM Data Science Certification – Capstone Project Declaration
Phillip Todd Danner

Data Analysis Strategy

The analysis will start with collecting information about each team including stadium and team winning percentage for previous seasons. This information will rely on various sources and will be compiled manually, given that there are only 32 teams and no single reliable source available for all information required. Location data for the stadiums (latitude, longitude) will be gathered using geo-location API such as Geopy Nomanatim,The intended stadium data set will resemble something as show below, but may include more fields as the analysis demands:

:

	Team	Division	Stadium	Capacity	Address	latitude	longitude
0	Buffalo Bills	AFC East	Bills Stadium	71,608	One Bills Drive, Orchard Park, NY 14127	42.771341	-78.787363
1	Miami Dolphins	AFC East	Hard Rock Stadium	65,326	2201 NW 196th Terrace, Miami Gardens, FL 33056	25.953790	-80.238164
2	New England Patriots	AFC East	Gillette Stadium	66,829	1 Patriot Pl, Foxborough, MA 02035	42.092269	-71.265460
3	New York Jets	AFC East	MetLife Stadium	82,500	1 MetLife Stadium Dr, East Rutherford, NJ 07073	40.813507	-74.074344
4	Baltimore Ravens	AFC North	M&T Bank Stadium	71,008	1101 Russell Street, Baltimore, MD 21230	39.277966	-76.623809
5	Cincinnati Bengals	AFC North	Paul Brown Stadium	65,515	One Paul Brown Stadium, Cincinnati, OH 45202	39.095464	-84.516049
6	Cleveland Browns	AFC North	FirstEnergy Stadium	67,895	100 Alfred Lerner Way, Cleveland, OH 44114	41.506056	-81.699712
7	Pittsburgh Steelers	AFC North	Heinz Field	68,400	100 Art Rooney Ave, Pittsburgh, PA 15212	40.446688	-80.014015

Location data for each stadium will be used to gathering venue data from Foursquare.com and will include only “Food” related venue categories. The initial radius from the stadium used for gathering venue information will be 1000m (deemed as “walking distance”), but may be expanded depending on results of the dataset.

The data will be cleaned and organized to hopefully produce the following analysis:

Ranking

1) Ranking of Stadiums by Quantity of Venues
2) Ranking of Stadiums by Variety of Venues
3) Ranking of Stadiums by Quality of venues (by average user rating)

Clustering

Further analysis to identify similarities between stadiums will use K-Means clustering using the optimal number of clusters.

Correlation

Correlation and possibly regression analysis will be performed using team winning percentage (and possibly other criteria) against Quantity, Variety and Quality of venues. Correlation analysis will exclude teams that have recently changed venues within the last year (Los Angeles Chargers, Los Angeles Rams, Las Vegas Raiders) since there is no winning percentage data for those teams at their new venues.

- Analysis is based on availability within close proximity (“walking distance”) of each stadium. No analysis is being performed as to how accessible the area around the stadium actually is to pedestrian traffic.
- Venues within the stadiums themselves are not being considered as part of the study