The Battle of the Neighborhoods (Week 1)

1.1. Problem and Background

1.1.1. Problem

- Columbus may be the largest city in Ohio population-wise, but given its dearth of professional sports teams and overall irrelevance to the rest of the state, the main rivalry between in the state is between Cincinnati and Columbus. As a student at Ohio State, I hear this all the time: “Our city is better!” But rarely do I hear concrete facts cited as to why that is. Is one city more diverse than the other? Does one city have more to do than the other? Is there better food in one of the cities? The questions go on… and they continue to go on unanswered. Wouldn’t it be great if someone found a way to *objectively* state whether Cleveland or Cincinnati is better in a given area? I sure do.

1.1.2. Background

- Separated by a 240-mile drive on Interstate-71, Cincinnati and Cleveland each own their respective corners of Ohio: Cincinnati in the Southwest and Cleveland in the Northeast. Including suburbs, the two cities account for around 40% of the state’s total population. Though outsiders typically regard the cities as being “total dumps” and “incredibly boring,” residents of the two cities are unanimously passionate about what their cities stand for. Cincinnatians love their chili on spaghetti, Oktoberfest celebration, cornhole, and beloved Bengals. Clevelanders, on the other hand, can’t get enough of Lake Erie, Cedar Point, Polish-boy sandwiches, and their beloved Tribe. Though it will be difficult to shoot any of those things down (minus the sports teams – go Bengals), comparing the things that both cities pride themselves on may lead to some interesting conclusions.

1.1.3. Interested Parties

- The results of this analysis are for any Cincinnati or Cleveland native who constantly have to defend their hometown against people from the other city. From a business prospective, the results of this study could be used to convince people moving into Ohio to move to either one of the cities depending on what their priorities are.

1.2. Data

1.2.1 Locations

- I will be focusing on the urban/downtown areas of Cincinnati and Cleveland, so a simple google search should suffice for the location data. [1,2] These longitude and latitude values will be the center of a circular region where various venues are explored. The geopy package could also be used to find the longitude and latitude values of Cincinnati and Cleveland.

1.2.2 Venues

- To find numerous venues in downtown Cleveland and Cincinnati, I will be using the Foursquare API and previously mentioned location data.

1.2.3 Sports Team Data

- I thought this would be a fun one to explore. I'll focus on the cities' respective baseball and football teams since only Cleveland has an NBA team and only Cincinnati has an MLS team. Since neither city has won a championship in football or baseball in the past 30 years (Sorry 1990 Reds), I'll focus on wins and losses, which can be found on numerous sports databases. [3,4,5,6]

1.3. Methodology

1.3.1. Exploratory Data Analysis

- After scraping wins and losses data from pro-football-reference.com, I analyzed the teams' winning percentages against each other by taking their total number of wins against each other and dividing that number by their total number of games against each other.
- I also analyzed the teams' performances against the rest of the league by summing their total wins and dividing that number by their total wins and losses.

- After scraping data from Wikipedia, I was able to repeat the above process for the baseball teams.
- With the location data, I simply used the groupby and count functions to group the venues by type and see how many of each type of venue there was in each city.

1.3.2. Inferential Statistics

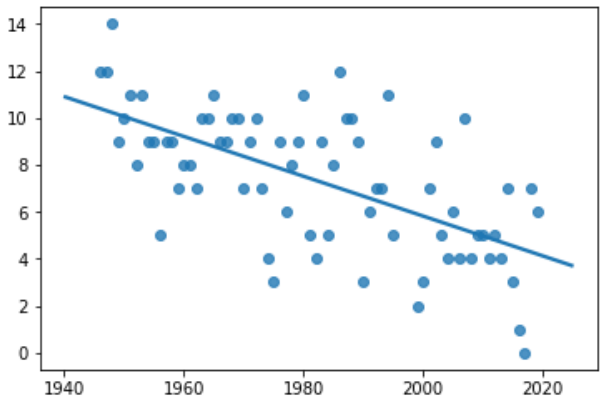
- After analyzing each teams wins and losses, I decided to plot a regression time that mapped time versus wins by using the regplot function from the seaborn package in python.
- I also used the LinearRegression function from the sk_learn package to calculate the relative fit of each of the trendlines and the slope of each of the trendlines to see how the teams have fared over the years.

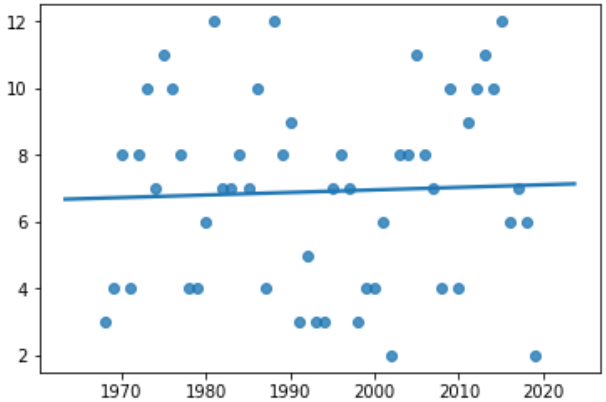
1.3.3. Machine Learning

- Machine Learning was not needed in my project, so no machine learning methods were used.

1.4. Results

- I discovered that the Bengals have won 55% of their games against the Browns and 45% of their games overall, while the Browns have won 45% of their games against the Bengals and 51% of their games overall.
- Throughout their history, the Browns have been declining fairly rapidly while the Bengals have slowly improved. This is evidenced by the Browns win/time slope of -0.0848 and the Bengals win/time slope of 0.0076. The trendlines can be seen below. The Browns' graph is on the left, while the Bengals' graph is on the right.





- After doing some quick research online, I found that the Indians have won 56% of theirs games against the Reds and 51.1% of their games overall, while the Reds have won 44% of their games against the Indians and 50.5% of their games overall.
- By using the folium package and latlong.com, I was correctly able to locate the cities of Cleveland and Columbus by using their geographical coordinates. I then accessed the types of venues in both cities within a 2500 meter radius and discovered that Cleveland had 60 different types of venues while Cincinnati had 57 different types of venues. In order to conduct this analysis, I created dataframes like the one shown below. (This dataframe is for Cleveland)

categories	name	lat	lng
Airport	1	1	1
American Restaurant	4	4	4
BBQ Joint	2	2	2
Bakery	2	2	2
Baseball Stadium	1	1	1
Basketball Stadium	1	1	1
Bavarian Restaurant	1	1	1
Brewery	3	3	3
Burger Joint	2	2	2
Café	3	3	3
Chinese Restaurant	1	1	1
Clothing Store	1	1	1
Coffee Shop	7	7	7
Comedy Club	1	1	1
Credit Union	1	1	1
Deli / Bodega	3	3	3
Dim Sum Restaurant	1	1	1
Diner	1	1	1
Fast Food Restaurant	1	1	1
Flea Market	1	1	1
Football Stadium	1	1	1
Fountain	1	1	1

1.5. Discussion

- This discussion deals solely with the question: “Is Cincinnati or Cleveland the better city?”
- We will first explore the sports teams:
    - Cincinnati leads the all time series against Cleveland in football, meaning that in a head-to-head context, Cincinnati has an edge against Cleveland. However, the Indians also have the head-to-head edge against the Reds.
    - Both the Browns and the Indians have better all time winning percentages than the Cincinnati teams in their respective sports.
    - The Reds have more championships (5) than the Indians (2), but the Browns have won more championships than the Bengals.
    - Since their franchise’s inception in 1966, the Bengals have been much more successful than the Browns.
    - Overall, both cities have categories they “win” in the area of sports, meaning that neither city dominates in the sports category. There is no objective winner here.
- We now explore the location data:
    - Cleveland is a smaller city population-wise than Cincinnati, yet it has a more diverse base of venues to enjoy than Cincinnati. Thus, it can reasonably be inferred that there is more to do in Cleveland than in Cincinnati.

1.6. Conclusion

- Based off my metrics, it appears that the city of Cleveland is slightly superior to the city of Cincinnati. However, the margin is so slim that the debate is likely to continue for a long time.
- In the future, one could take a more in depth look into the cities than I did. Things like cost of living, average age, and quality of venues could be analyzed, and the cities could even be compared on a neighborhood by neighborhood level.

References:

[1] https://www.latlong.net/place/cleveland-oh-usa-2643.html [2] https://www.latlong.net/place/cincinnati-oh-usa-6855.html [3] https://www.footballdb.com/teams/nfl/cincinnati-bengals/opponents [4] https://www.footballdb.com/teams/nfl/cleveland-browns/opponents [5] https://en.wikipedia.org/wiki/List_of_all-time_Major_League_Baseball_win%E2%80%93loss_records [6] http://mcubed.net/mlb/cle/cin.shtml