Battle of the Neighborhoods

# Introduction:

Where there are colleges and universities, there are restaurants.  Dining out, and particularly dining out with friends, is common for college students. Unfortunately for student groups, they often must debate which venue to choose from in order to satisfy multiple tastes. There are many choices, such as Chinese, Indian, Thai, Mexican, and Greek from which to debate and choose. Innovative restaurant entrepreneurs have recognized that groups do not want to visit multiple venues when their tastes span multiple cuisines. As a result, fusion style restaurants are on the rise and expanding to college towns.

While it is common to think of combining coffee with bagels or Thai with Chinese within a venue, more innovative style restaurants are finding niche markets. Two former University of Minnesota students founded Burrigato in 2017, a restaurant unique to the school, that fuses Asian style cuisine with a Mexican style wrap. While it is predictable for a venue to serve both deli sandwiches and ice cream, in today’s eclectic university settings venturing into a Greek and Somalian combination or a meld of Cajun and Baltic would not be extreme.

This exploratory analysis will develop a simple metric to measure the concentration of two cuisine types near a population center. A higher number of restaurants per individual will correlate with a higher concentration score. This metric will rank the chosen locations, higher being better, for potential expansion or development of fusion style restaurants. In order to limit the scope of the analysis, a “Burrigato” scenario will be explored by posing the following question. If the Burrigato owners desired to open a franchise near another Big Ten Conference University, which of the thirteen schools has the highest concentration of Asian and Mexican restaurants and would therefore be considered the optimal expansion candidate?  It should be noted that Burrigato is a real world restaurant, but for the purposes of this presentation should be considered a fictional example.

Multiple groups will be interested in the methods and results of this research. Marketers would be interested in the concentration score in order to target and streamline advertising campaigns. Universities would value the metric to more fully understand their diversity and inclusivity demographics. Most importantly, entrepreneurs would be interested in the score to target future expansion or initial openings of fusion style restaurants.

# Data:

### 2.1 Data Gathering

For this exploratory study, the Big Ten Conference will serve to limit the scope of the analysis.  There are presently 14 schools within the conference and their locations and enrollment are easily obtained from multiple internet sources.  The data for the .csv file used for the analysis was obtained from the Wikipedia page Associated with the Big Ten;

https://en.wikipedia.org/wiki/Big\_Ten\_Conference

The university location data was pulled from Foursqaure and merged with the wikipedia data to form the following table:

Table 1:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Institution** | **Division** | **Location** | **Enrollment** | **Latitude** | **Longitude** |
| Indiana University | East | Bloomington, Indiana | 48,514 | 39.1772 | -86.5154 |
| University of Maryland | East | College Park, Maryland | 38,140 | 38.992 | -76.9461 |
| University of Michigan | East | Ann Arbor, Michigan | 43,625 | 42.2942 | -83.7100 |
| Michigan State University | East | East Lansing, Michigan | 50,085 | 42.7186 | -84.4779 |
| Ohio State University | East | Columbus, Ohio | 58,322 | 40.0057 | -83.0287 |
| Pennsylvania State University | East | State College, Pennsylvania | 45,518 | 40.8034 | -77.8591 |
| Rutgers University | East | New Brunswick, New Jersey | 40,720 | 40.4791 | -74.4317 |
| University of Illinois | West | Champaign-Urbana, Illinois | 49,339 | 40.1007 | -88.2313 |
| University of Iowa | West | Iowa City, Iowa | 33,334 | 41.6659 | -91.5731 |
| University of Nebraska | West | Lincoln, Nebraska | 33,273 | 40.8175 | -96.7045 |
| Northwestern University | West | Evanston, Illinois | 21,208 | 42.0551 | -87.6758 |
| Purdue University | West | West Lafayette, Indiana | 39,464 | 40.4275 | -86.9123 |
| University of Wisconsin | West | Madison, Wisconsin | 49,193 | 43.0798 | -89.4307 |

All location data for restaurants was also pulled from Foursquare. By limiting this ranking to 13 universities, the number of API calls required to pull two restaurant styles per university does not exceed the daily limits of non-fee developer accounts. The number of schools is thirteen because for our fictional scenario the origin location of the University of Minnesota is not included. Likewise, the number of calls was limited by setting the radial distance limit within Foursquare to 10 Km.

### 2.1 Concentration Score Definition:

The concentration score metric is a measure of the number of students per the restaurant styles/cuisine types. Where Up represents the number of enrolled students, and R1 and R2 represent the number of restaurants for the two styles, the value is the product of the two ratios:

Concentration Score = [(Up /10,000) / (1/R1)] \* [(Up /10,000) / (1/R2)]

If there were 15 Asian restaurants and 20 Mexican restaurants within a 10 Km radius of the University of Minnesota, the score would calculate as:

Score = [(51,147 / 10,000) / (1/15)] \* [(51,147 / 10,000) / (1/20)] = 179.0145

Multiple variants of this metric were considered, but set aside due to their undesired complexity. For example, distances between the individual restaurants and the center of the university can be weighted or the enrollment can be adjusted for the city/community general population. The principal of Occam’s razor applies to make this analysis more efficient.

## Methods and Analysis

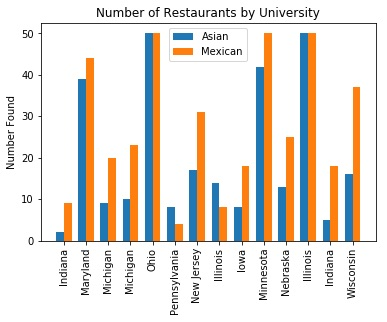
A Jupyter Notebook within IBM’s Watson Studio was developed to retrieve the restaurant location data, create the appropriate data frames, calculate the concentration score, and rank the universities. The notebook has been published to Github and can be located here:

https://github.com/bdwasson/Main\_Repository/blob/master/Battle of the Neighborhoods Project.ipynb

### 3.1 Exploratory Analysis

The initial university data, seen in Table 1, contains the location data for the institutions. Within this first step, the notebook combines a .csv file Wikipedia locations with Foursquare latitude and longitude coordinates. Simple visual confirmation of this data is sufficient to confirm its accuracy. For example, Nebraska has a longitude of -96.70 and Rutgers has a longitude of -74.43. This is clearly correct as Nebraska is located significantly west of New Jersey. With similar logic, the latitude for Wisconsin, 43.07, is further north than Illinois, 40.10. Following the retrieval of restaurant counts, the number of restaurants by university is displayed in chart 1.

Chart 1:



The chart confirms that the counts are reasonable and vary across universities. This not only confirms that a valid concentration score calculation is possible, it also confirms that the limits on the number of calls and the distance radius are appropriate.

As a final visual confirmation of the data, the Mexican restaurants surrounding the University of Wisconsin – Madison were plotted. This map confirms that the locations are varied and being retrieved correctly.

### 

### 3.2 Inferential Testing

The final version of the concentration score is a product of several iterations. As the intent is to keep the calculation as simple as possible, the first formula to evaluate was R1 + R2, the simple sum of the restaurants around the university. This calculation results in tied scores and illogical results. While simplicity is maximized, the calculation must result in a unique score for each school and a logical ranking based on total counts. After multiple iterations, the final formula, as seen above, was determined to be sufficient for the ranking.

## Summary

### 4.1 Results and Recommendation

The final results are summarized in Table 2:

Table 2:



Ohio State University in Columbus, Ohio is the recommended school with a concentration score of 583.22.

### 4.2 Discussion

The recommendation of the University of Ohio is not a surprising result given the similarities with the University of Minnesota. Additional analysis of the overall ranking table (Table 2) reinforces this recommendation with several observations. First, the total number of Asian and Mexican venues summed to 100 for Ohio, which was the maximum for the group. Second, even though Northwestern University also summed to 100 the school scored significantly lower due to a student body that is half the size of Ohio. Third and finally, Indiana University was ranked last with only a fraction of the number of restaurants as Ohio.

### 4.3 Conclusion

The process of ranking universities based on the concentration of two restaurant cuisines appears to be an effective manner to rank potential expansion restaurant locations.

In future uses, the methodology can be expanded along multiple lines of investigation for more in-depth analysis. City, county, or other regional population centers can be used rather than universities as the locus points. Scaling to three, four, or more cuisine types is a logical extension of the analysis. K-means clustering or other machine learning algorithms can also be used to define a weighting factor across population centers to refine the ranking.

In summary, the process is effective in answering the current question and exploring additional inquiries. While this project focuses on a single existing fusion restaurant, the methodology could easily answer whether locations are optimal for a new style restaurant, such as bratwurst empanadas. This initial analysis is basic in nature but rigorous enough to produce accurate, viable, and actionable ranking results.