Week 1

**Introduction**

With on-going COVID-19 shutdowns and people increasingly working from home, there is significantly less demand for downtown restaurants now than before the pandemic. Restauranteurs are increasingly looking to start food trucks or move their restaurants to new locations.

The goal of this capstone project is to identify suburbs around Pittsburgh, PA that are the best choice to move restaurants to (or set up food trucks at). The best suburbs will include those with top venues being other restaurants, showing high demand in that area. The best suburbs will be clustered together to give a list of target areas to restauranteurs, ranked by population.

This will be a case study for a Chinese restaurant where they will want the number of restaurants in a neighborhood to be high (as a signal for demand), but with Chinese restaurants outside the top 10 business in that neighborhood in order to avoid an oversaturated market.

**Data**

Two data sources will be used for this project. The first will be pulling a list of zip codes in Allegheny County (where Pittsburgh resides), zipdatamaps.com/allegheny-pa-county-zipcodes. This data will be scrubbed to remove zip codes with “Pittsburgh” in the name to refine the list of new restaurant locations to the suburbs.

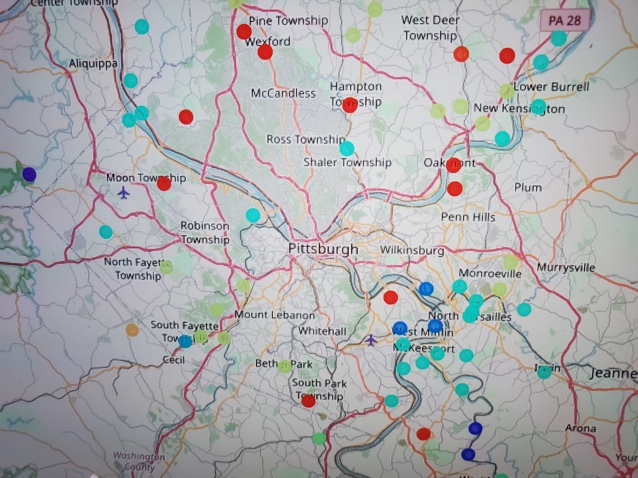
The other data source will be Foursquare API data to determine top businesses and neighborhood similarities. This data will be analyzed to make clusters that have high restaurant volumes (signaling demand) and will be further analyzed to ensure recommended suburbs do not already have a high volume of Chinese restaurants.

Ultimately a third data source was needed to map the Pittsburgh suburban zip codes to a latitude and longitude (https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/).

Week 2

**Methodology**

The zip code list was imported and cleansed for formatting. It was also merged with the latitude and longitude data in order to locate each specific Pittsburgh suburb. To refine the DataFrame to suburbs only, any area with “Pittsburgh” in its name was removed.

Foursquare API data was pulled in for each suburb on the refined list. This data was used to determine the top ten venue types in each area and the DataFrame was broken into 12 clusters using K-Means. K-Means was used in order to identify similar suburbs based on venue data. Overall the clusters would group like areas. The number of 12 clusters was determined after trial and error where lesser values didn’t allow for the best distinction of like areas, but any higher resulted in small, equal clusters.

Two clusters (1 and 9) were relatively similar, but 9 had the higher overall restaurant quantity with the lowest number of existing Chinese restaurants.

Populations were appended back into the final list for ranking.

**Results**

The result is a list of 13 target suburbs ranked by population for a Chinese restaurant owner to explore for relocation or to set up a food truck. This is a reduction of 79% from the refined list of 61 suburbs, allowing for a more targeted destination selection.

|  |  |
| --- | --- |
| **Neighborhood** | **Population** |
| Bethel Park | 29529 |
| Gibsonia | 27049 |
| Bridgeville | 16213 |
| Oakdale | 9956 |
| Cheswick | 9029 |
| Pitcairn | 3294 |
| Presto | 1163 |
| Creighton | 1128 |
| Harwick | 895 |
| Indianola | 461 |
| Morgan | 382 |
| Bakerstown | 323 |
| Warrendale | 300 |

**Discussion**

Being somewhat familiar with the area, the results were expected. K-Means performed well in separating out areas that are the more homogenous, restaurant-friendly suburbs from other suburbs that are more rural, industrial, or restaurant-averse. Rerunning the notebook several times resulted in a few different configurations with two clearly top clusters, but one edging out the other. Each time, the final result accomplishes the goal of providing a Chinese restaurant owner with a population-ranked list of target suburbs based on high restaurant demand, but low Chinese restaurant competition.

**Conclusion**

It is possible to use Foursquare API data to develop a list of target relocation suburbs for a downtown Pittsburgh Chinese restaurant owner. The final list reduces 61 Pittsburgh suburbs down to 13 that are the most restaurant-friendly while having limited Chinese restaurant competition.