

# IBM Applied Data Science Capstone

Week #5 – Research Paper

## *“Opening a New Poke Restaurant in the City of Seattle”*



[https://commons.wikimedia.org/wiki/File:Seattle\\_Kerry\\_Park\\_Skyline.jpg](https://commons.wikimedia.org/wiki/File:Seattle_Kerry_Park_Skyline.jpg)

By: Matt Pierce  
January 2021

# 1. Business Problem

- **Objective:** Find best location in Seattle to open new Poke restaurant
- **Hypothetical Client:** Existing Owner, wants to expand from 1 to 2
- **Target Requirements:**
  - Clustered near other restaurants (e.g.: Game Theory of Gas Stations)
  - No existing Poke place nearby
  - Timely, need answer within 2 weeks
  - Be sure to include assumptions, errors, and omissions at Discussion section

## 2. Data Acquisition

- **Data Needs:**

- List of neighborhoods, zip codes and corresponding geo-coordinates
- Demographic metrics tied to the zip codes and/or neighborhoods
- Venue data such as nearby restaurants, competitors, like ratings, etc.

- **Data Sources:**

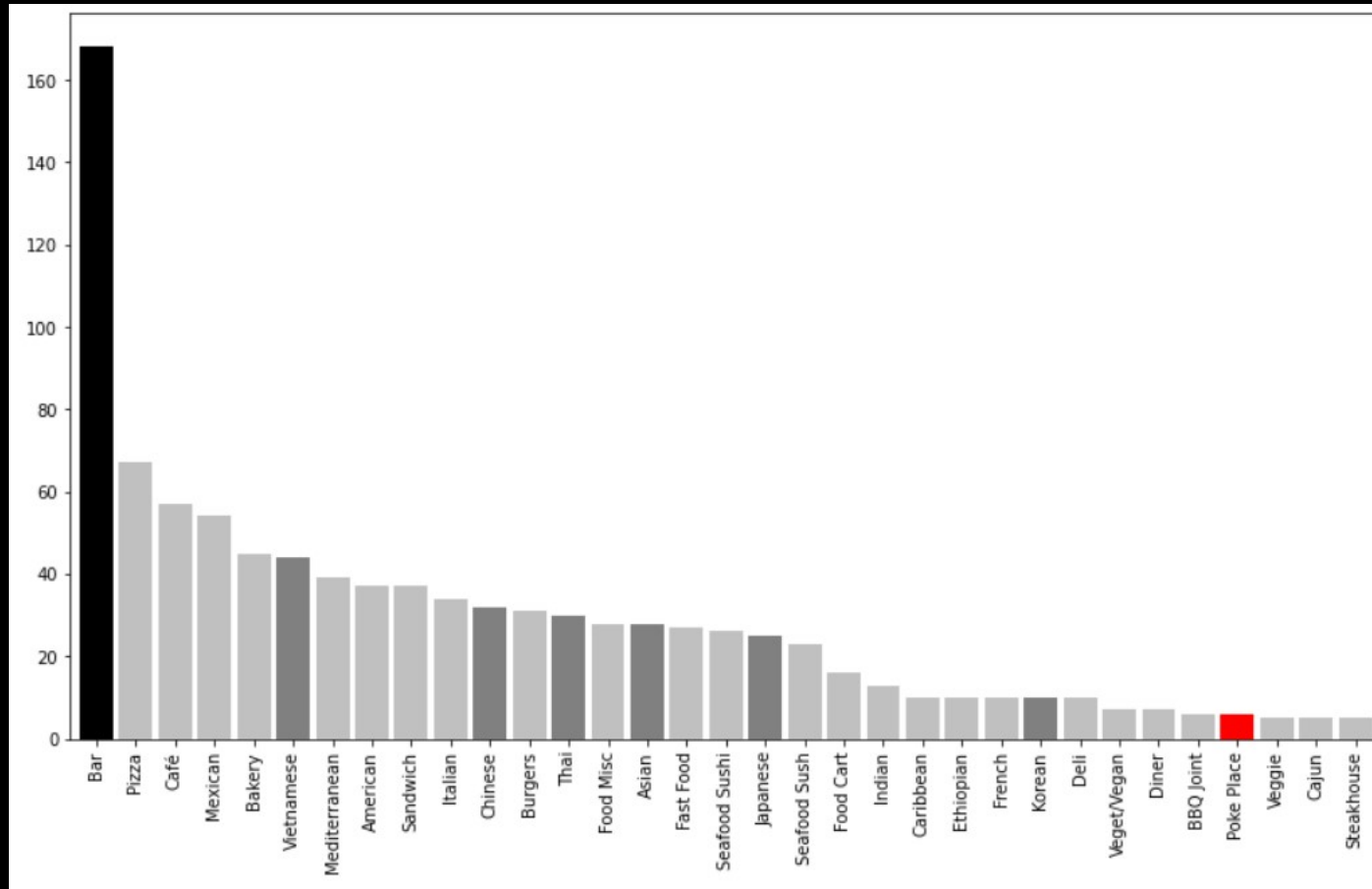
- City of Seattle OpenData Urban Centers list (online, neighborhoods)
- AgingKingCounty.org PDF that cross-walks neighborhoods to zip codes
- OpenDataSoft.com ties zip codes to geo-coordinates (latitude/longitude)
- Python Geocode package to fetch geo-coordinates from addresses in a pinch
- FourSquare API for nearby venue information

### 3. Methodology

- **Data Loading:** Primarily CSV preparation, then used Pandas `read_csv`
- **Data Cleaning:** Significant. Removed null value rows, de-duped and rolled up values to zipcode level, calculated aggregates, calculated population densities and rental percentages, combined multiple sources into single flat table grouped on zip code with neighborhoods aggregated into comma delimited list, validated results at checkpoint
- **Data Exploration:** Significant. Explored base data via Folium map. Configured FourSquare API + loaded venue data.
- **Feature Selection:** Wrote two scrubs to filter and simplify venue categories to restaurant related venues.
- **Modeling:** Performed K-Means clustering to find patterns in locations
- **Visualize:** Clusters in Folium map

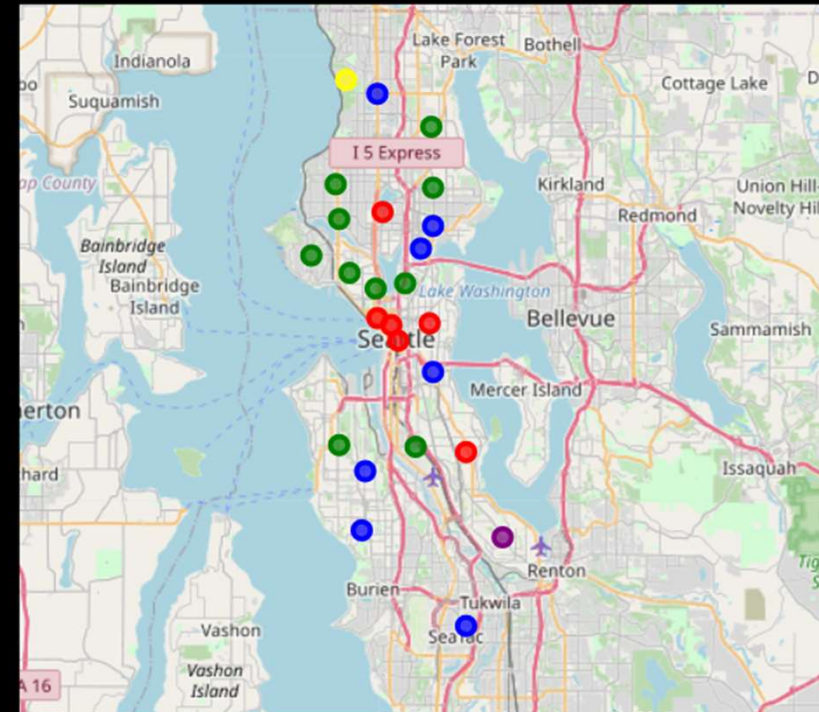
## 4.A Results, Macro

- There is ample room for growth of Poke restaurants in Seattle



## 4.B Results, Modeling

- **Cluster #2 – Green = Target Locations**
  - 10 zip codes, no Poke competition yet
- **Cluster #4 – Blue = Possible Target Zones**
  - 7 zip codes, no Poke competition
- **Cluster #1 – Red = Avoid / Saturated**
  - 6 zip codes – ALL have Poke competition
- **Clusters #3/#5 – Yellow/Purple = Outlier**
  - 2 zip codes (one per cluster) = bad
  - No Poke competition...but few venues either



## 5. Discussion

- **Recommendations:**

- Yes, there is **ample upside** potential still in Seattle for opening a Poke restaurant
- Focus on **Cluster #2** primarily but also #4 because little to no competition yet
- **Avoid Cluster #1** as it is over saturated already with Poke competition
- **Avoid outlier Clusters #3 and #5** – there's probably no market there
- **Suggest** zip codes **98122** (Capital Hill), **98103** (Fremont), and **98121** (Belltown)

- **Areas of Improvement:**

- More rigor w/demographics: factor rental rate, median income into clustering
- Add in bus stations, don't remove them
- FourSquare radius 20 blocks = possible gaps, increase radius and dedup venues

## 6. Conclusions

- **Business Question = Answered**
- Yes, it is a good idea to open a second Poke restaurant
- Suggested Locations = in or around zip codes...
  - **98122** (Capital Hill)
  - **98103** (Fremont)
  - **98121** (Belltown)



# Thank You!



Taken from Sample Presentation that were part of course instructions – I liked this