Urban Diversity Within Seattle's Neighborhoods

Natalia Quintero
March 2019

Contents

Introduction

What makes a neighborhood? Not only residents, usually venues tend to attract diverse groups of people. A neighborhood with a variety of places may turn to more social interactions but it will also attract people who may change it in the long term. When looking for a new place to live, you are not just buying or renting a home; you are becoming a part of that neighborhood and looking to match your needs at that point in time. There are common components to all neighborhoods, some people may be attracted to areas with lots of restaurants and nightlife while others would like a more residential and quiet area. With this project I want to explore and have a better understanding of the neighborhoods in Seattle and their local amenities, and hopefully help those who are looking to move in the near future find a place that is in sync with their current lifestyle.

Methodology

Unsupervised learning involves models that describe data without reference to any labels and partitions it into distinct groups of similar items. This project attempts to categorize and segment data in Seattle's neighborhoods; identify different venues and what a neighborhood offers for newcomers or those who are looking to live close to specific places.

Data Processing Cycle

OSEMN or OSEMiN is an acronym that stands for Obtain, Scrub, Explore, Model and iNterpret. It is a list of tasks that a data scientist should be familiar with, not an expert on each of them. It is also considered as the blueprint for data exploration. These steps don't follow one another, as you go through the analysis is normal to come back to any stage and move back and forth.

Obtain Understand requirements, gather information about the problem to be solved, adopt and understand the tools that will be most suitable to do the job. In this case, the *categories* function of the Foursquare API will be used to retrieve the venues and the Seattle Department of Neighborhoods website to obtain a list of the city's neighborhoods. (https://www.seattle.gov/neighborhoods/neighborhoods-and-districts)

Raw data sets generated

- **sea_neigh_list.csv**, data set generated from extracting the name of the neighborhoods by zone from the website
- **coordinates.csv**, data set with the geographical coordinates of each neighborhood.

	0	Zone
0	Belltown	Central
1	Broadmoor	Central
2	Capitol Hill	Central
3	Central District	Central
4	Denny-Blaine	Central
5	Downtown	Central
6	Eastlake	Central

From sea_neigh_list.csv

Scrub (clean) In general, this is where most of the time is spent; data cleaning has an impact on the accuracy of the results. This is where you check for missing or null values, replace or remove them, extract columns or format data types. At this point the dataset presented should be read in a clean manner, free of irrelevant characters and in a usable format. I have removed the zone column

- **coordinates_shaped.csv**, data file ready to be used without the zone column

Merge both data frames

```
In [40]: data = pd.merge(neighborhoods, coordinates, on='Neighborhood', how='inner')
         data.head()
```

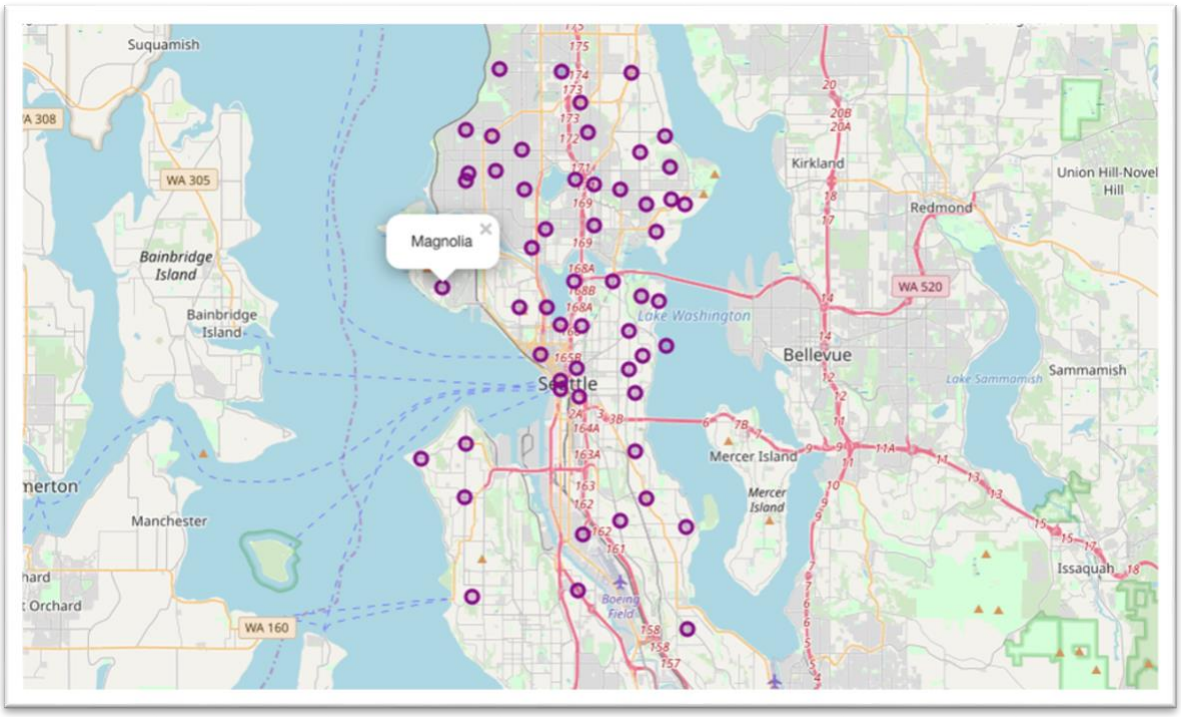
Out[40]:

	Neighborhood	Zone	Latitude	Longitude
0	Belltown	Central	47.6147	-122.3448
1	Broadmoor	Central	47.6362	-122.2898
2	Capitol Hill	Central	47.6253	-122.3222
3	Central District	Central	47.6088	-122.2964
4	Denny-Blaine	Central	47.6178	-122.2758

Data set ready to use

Explore This is where exploratory data analysis happens, where you get to know the data you are working with. No hypotheses are tested and no predictions are made here. The focus is to understand the distribution of the columns, check for multicollinearity,

and make sure the dataset meets what is necessary for the type of model you will apply later on.



Map of Seattle with neighborhoods defined in the data set

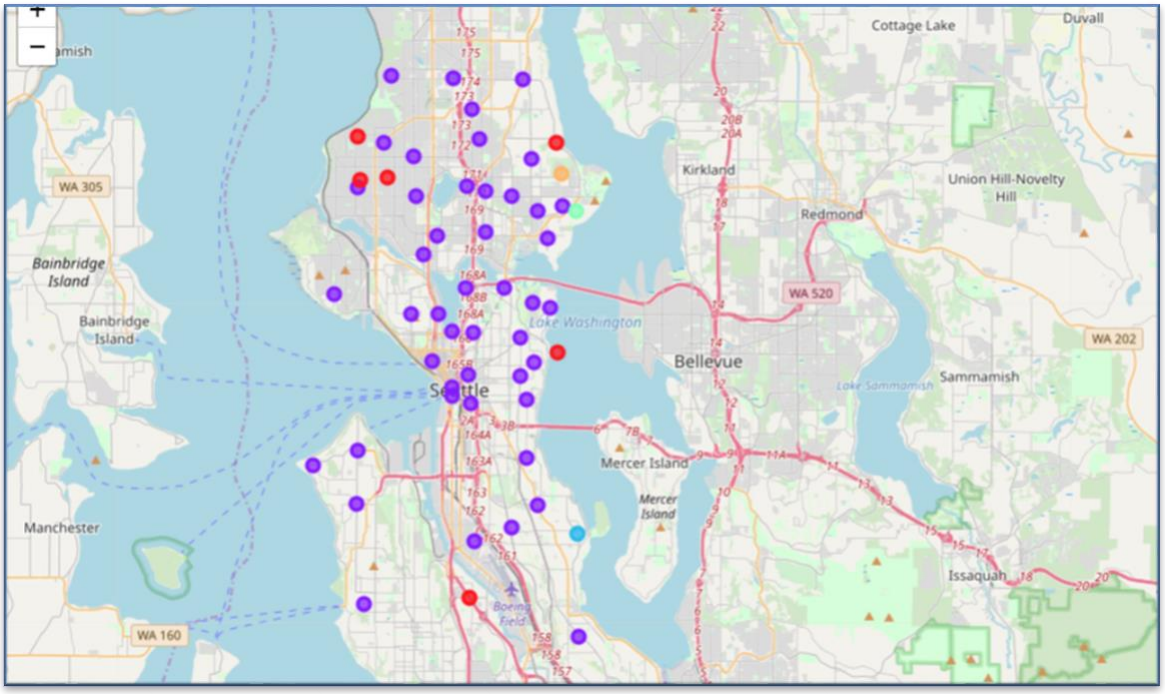
256 unique venues found

	Zone	Neighborhood	Venue	Venue Category
0	Central	Belltown	FOB Poke Bar	Poke Place
1	Central	Belltown	Shiro's	Sushi Restaurant
2	Central	Belltown	Rocco's	Pizza Place
3	Central	Belltown	Umi Sake House	Sushi Restaurant
4	Central	Belltown	Have A Heart Belltown	Marijuana Dispensary
5	Central	Belltown	Jupiter	Bar
6	Central	Belltown	Bangrak Market	Thai Restaurant
7	Central	Belltown	Bathtub Gin & Co.	Speakeasy
8	Central	Belltown	CrossFit Felix	Gym
9	Central	Belltown	The Crocodile	Rock Club

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Admiral	Coffee Shop	Grocery Store	Pub	Chinese Restaurant	Middle Eastern Restaurant	Spa	Shoe Repair	Salon / Barbershop	Post Office	Pizza Place
1	Alki	Coffee Shop	Park	Restaurant	Beach	Italian Restaurant	Art Gallery	Donut Shop	Mexican Restaurant	Sandwich Place	Scenic Lookout
2	Ballard	Hobby Shop	French Restaurant	Coffee Shop	Jewelry Store	Gift Shop	Park	Dog Run	Donut Shop	Drugstore	Dumpling Restaurant
3	Beacon Hill	Bus Station	Light Rail Station	Doctor's Office	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Fair	Eye Doctor	Event Space	Ethiopian Restaurant
4	Belltown	Bar	Coffee Shop	Cocktail Bar	Sushi Restaurant	New American Restaurant	Lounge	Breakfast Spot	Gym	Pizza Place	Bakery

10 most popular venues for the first 5 neighborhoods. From venues_sorted set

Model It refers to the process of using probabilistic methods to try to predict the outcome of an event. Apply a technique(s) or algorithm(s) to predict and interpret, adjust the ones you have to try and increase results.



Model with 5 clusters

Interpretation Draw conclusions, evaluate the meaning of the results, and use visualization tools to communicate in an understandable manner

Cluster 1 (purple) dominant throughout the city, with venues like parks and coffee shops. Refer to the notebook to find the full list