IBM DATA SCIENCE CAPSTONE PROJECT

Introduction: Business Problem

As an individual who is fascinated with very monumental things, it was not absurd that Joy Zel decided to move to Washington State in the United States of America. Known for being the place that houses the Space Needle, National Park and many others, it was a side attraction, she couldn't turn down

As a center for Agriculture, Forestry and Fishery, Washington State contributes to the economy of the United States in a very important manner, and water being its most powerful resource. Washington is categorized by fertile soil and beautiful vegetation, and it is a contributing factor to the state's importance

Washington State, also known as the State of Washington is a state in the Pacific Northwest Region of the United States. It is no surprise that it was named after the famous George Washington and has a population of 7.6 million. It is the 18th Largest state and the 13th most populous state.

The state's coastal location and excellent harbours, give it a leading role in trade with Alaska, Canada, and countries of the Pacific Rim. Washington cities have sister cities in several countries, and their professional and trade associations commonly include Canadian members.

After the final decision on her relocated, there was one thing that had not yet been decided concerning Joy's move which was: What City Would She Live In. There was quite a number of cities to choose from, but Joy wanted to be in the most optimal city and that is the problem we hope to solve.

The only condition she stipulated was that she wanted a city that was surrounded by popular tourist attractions such as; garden, landmarks, art museums and cafes.

Using this information, we proceed to solving the problem.

Data

In order to find the most suitable city for Joy, a list of cities with their corresponding zipcodes was downloaded from http://www.downloadexcelfiles.com/us_en/download-list-zip-codes-washington-state in csv format.

Venue queries will then be made by neighborhood name/cities using FourSquare Api. The resulting data regarding venue category will be used to find similarities in the corresponding cities. With a proper list of top venue categories for each state, a cluster can be made grouping them together for easier understanding.

These are the data needed for the success of this project:

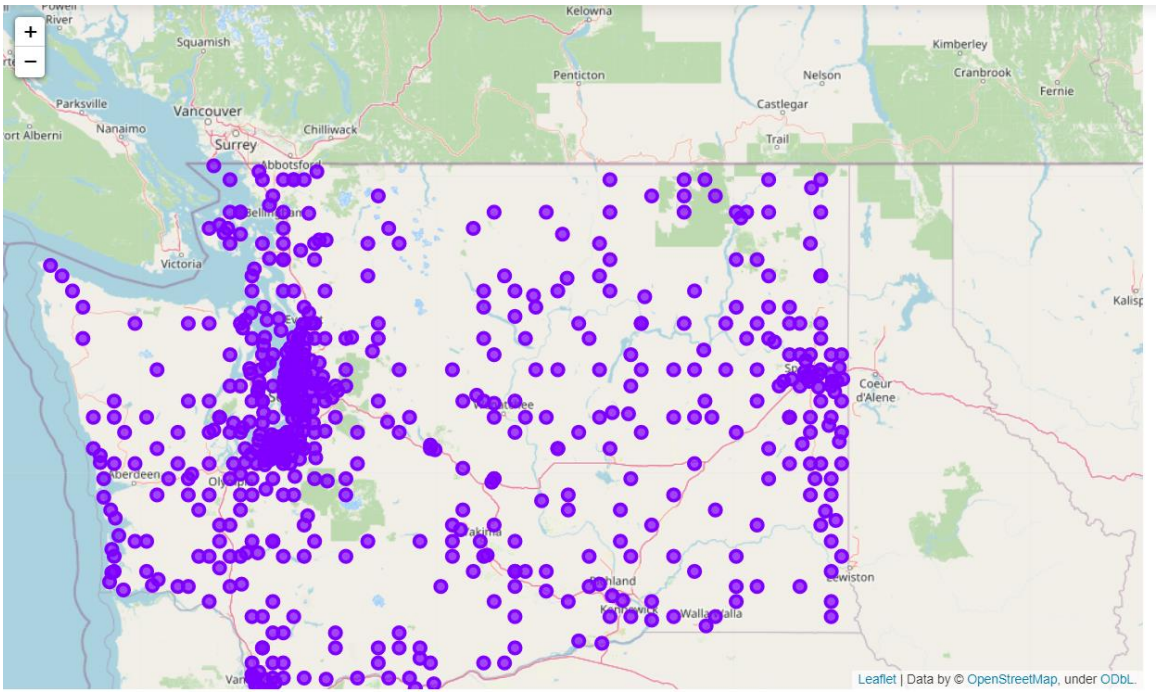
- Cities and Zipcodes in Washington State
- Coordinates of the cities
- Top Venues for each city
- Categories of Venue

Methodology

As a result of the nature of the problem at hand, we will make use of k-means clustering to properly group the cities based on their most in-demand venue.

A one-hot encoding will be the done on the data and then grouped by city. The encoding will return venue categories as a column per city which will then be further grouped. The encoded dataframe will further be filtered into the top venues and then the k-means algorithm will be passed over it. This will return a cluster of cities. Then an adequate recommendation can be given to Joy.

Analysis



Map of clusters.

Cluster 1

Cluster 1

```
In [107]: wash_merged.loc[wash_merged['Cluster Labels'] == 0, wash_merged.columns[[1] + list(range(5, wash_merged.shape[1]))]]
```

Out[107]:

	Neighborhood	Cluster Labels	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	ABERDEEN	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
1	ACME	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
2	ADDY	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
3	AIRWAY HEIGHTS	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
4	ALBION	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
...	...	...	...	...	...	...	...	...	...	...	...	...
581	YAKIMA	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
582	YAKIMA	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
583	YAKIMA	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
584	YELM	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place
585	ZILLAH	0	Garden	American Restaurant	Government Building	Hotel	Monument / Landmark	Art Museum	Bar	Café	Hotel Bar	Sandwich Place

586 rows × 12 columns

Cluster 2

Cluster 2:This Cluster Has No Neighborhood

```
In [108]: wash_merged.loc[wash_merged['Cluster Labels'] == 1, wash_merged.columns[[1] + list(range(5, wash_merged.shape[1]))]]
```

Out[108]:

Neighborhood	Cluster Labels	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

Cluster 3

Cluster 3:This Cluster Has No Neighborhood

```
In [110]: wash_merged.loc[wash_merged['Cluster Labels'] == 2, wash_merged.columns[[1] + list(range(5, wash_merged.shape[1]))]]
```

Out[110]:

Neighborhood	Cluster Labels	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

Result and Discussion

It shows that after sorting venues by most common, the neighborhoods were grouped into just one cluster. Although this was not a possibility we saw coming, it goes to show that this state that Joy has planned to relocate has a number of major requirements in a least a great percentage of the entire state. After reviewing the cluster, the cities analysed for this data presented that it has the major venue category requirement by Joy which was; cafe, art museum, garden. This shows that all the neighborhoods have the same concentration of venue category.

Conclusion

Joy can pick any of the cities cited in this study and she will be quite satisfied.