Capstone Final project: Moving Recommendation

Introduction/Business Problem

Due to the COVID-19 pandemic, being locate near the workplace has become less necessary and working remotely has become essential. As a result working remotely has become our new norm and will most likely continue after the pandemic. This becomes an excellent opportunity to bounce between cities from one year to the next and experience a city in depth instead of shallow glimpse a vacation offers. However, choosing a neighborhood to move into without visiting first becomes quite daunting. Using the Foursquare venue api, we could compare cities and evaluate them before moving.

Data

Using the Foursquare api, compare the top 100 venues between a list of cities and a base city. Rank each of the neighborhoods based on the similarity to the base city. To get a better result, weight each venue category based on preference.

Methodology

The foursquare api provides 135 different venue categories. However, the venue categories are too specific. For example, in Oakland, CA 7 of the top 10 venues are food related (9 including bars). With such a specificity comparing cities will be meaningless.

City	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
Oakland, California	Bar	Sandwich Place	Chinese Restaurant	Clothing Store	Wine Bar	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Indian Restaurant	Dim Sum Restaurant
Portland, Oregon	Hotel	Coffee Shop	Restaurant	Food Truck	Sandwich Place	Park	Food Court	Jewelry Store	Burger Joint	Yoga Studio
San Francisco, California	Coffee Shop	Café	Hotel	Park	Theater	Cocktail Bar	French Restaurant	Wine Bar	Bakery	Poke Place
Seattle, Washington	Coffee Shop	Hotel	Cocktail Bar	Café	Japanese Restaurant	Italian Restaurant	Sandwich Place	Park	Mexican Restaurant	Burger Joint

To make the analysis more useful, map the venue categories to more generic venue categories. In addition, to the frequency preferences for each of the generic venue categories. For this analysis a list of randomly generated preferences were used from [1, 10].

For this analysis no machine learning was necessary since calculating the similarities between the cities of interest and the base city is straightword. The similarity is calculated by cosine similarity: Similarity = < p A | p B > /(|| p A || || p B ||)

City	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
Oakland, California	Restaurant	Bar	Cafe	Hobby	Clothing Store	Arts Venue	Arts venue	Groceries	Cupcake Shop	Sporting Goods Shop
Portland, Oregon	Restaurant	Hotel	Cafe	Sightseeing	Bar	Hobby	Clothing Store	Jewelry Store	Sporting Goods Shop	Arts Venue
San Francisco, California	Restaurant	Cafe	Bar	Hobby	Arts Venue	Sightseeing	Arts venue	Clothing Store	Hotel	Grocery
Seattle, Washington	Restaurant	Cafe	Bar	Sightseeing	Hobby	Hotel	Antique Shop	Arts Venue	Arts venue	Building

Results

Oakland, California is the base city and the list of cities of interest are: Portland, Oregon; Seattle, Washington; and San Francisco, California. The city of interest that is most similar to Oakland, California is Portland, Oregon (0.9327). However, the three cities of interest have comparable similarity: Seattle, Washington (0.9080) and San Francisco, California (0.8395).

City	Similarity
Portland, Oregon	0.9327
Seattle, Washington	0.9080
San Francisco, California	0.8395

Discussion

Further refinement of the data can be done to differentiate between the cities of interest. The mapping between the specific venue categories to the generic venue categories can be further refined by adding back in more categories. Another possibility is to increase the number of venues retrieved from the Foursquare api. Also by supplementing the foursquare data with housing information can also help differentiate. This housing information can come from zillow and can be housing availability and cost. Finally, being more specific to areas within cities can help distinguish between cities of interest.

Although even with these changes, the cities of interest may still be comparable. Since their general demographic are quite similar.

Conclusion

In this study, we compared three cities: Portland, Oregon; Seattle, Washington; and San Francisco, California to Oakland, California to decide what city would be nice to remotely work from for a given year. This was done by comparing the venue categories of the top 100 venues of each city. The specific venue categories were overly specific that would skew this analysis, a mapping between the specific categories to more generic categories were applied. Because of the breadth of the different venue categories, preferences were also given to further refine the similarity. Through the analysis we observed that Portland, Oregon is the most similar to Oakland, California.