DETERMINING THE BEST NEIGHBORHOOD FOR A RESTAURANT

INTRODCUTION

- Silicon Valley is the heart of global innovation, technology and social media. It is home to big tech giants like Apple, Facebook, Google, Microsoft, Linkedin, Uber and many more. With such big names come tremondous job opportunities which pulls people all over the world to migrate and make this silicon valley their home. But this huge influx of people eventually causes its own problems like, high property rates, cost of living and crazy traffic. While this fast paced life has its allure, it does not suite everyone and some prefer to have the advantages of silicon valley balanced with a somewhat quite and relaxing life more closer to the nature.
- For these individuals the "SILICON FOREST" is a godsent! The silicon forest is a group of tech companies eastablished in the Willamett Valley in Oregon. Sice 1980's to date the area around silicon forest has seen a rise in population, infrastructure growth and development, expansion of cities, growth in housing and night life. It has almost all the pros of silicon valley with a more chilled out lifestyle. The Silicon Forest has drawn a crowd similar to silicon valley to its midst.

BUSINESS PROBLEM

- Let us consider an individual in intersted in starting a Restaurant with a unique idea which has not been tested before. They are not ceratin if it will be embraced by the masses or be a big failure. They want to try running a successfull business at a smaller scale, to test the waters before investing huge amounts and opening at a large scale in heavily populous cities like San Francisco and San Jose. They hear about the silicon forest and feel as the client base might be similar it would be a good idea to consider opening a business there. They want to explore Oregon to start a busniess.
- So in this Capstone project I would like to address this problem. We will try to find out which county in Oregon has the most venues so as to select that county as the place of busniess. To explore the neighborhoods in that county, pitting them each other in the battle of neighborhoods.
- This kind of analysis will help not only the people seeking to open a business in popular places but with lower property rates, it will also help people who are seeking otions of places similar to silicon valley to relocate for a quiter life.

DESCRIPTION OF DATA

1.For the list of Counties

- Data Source: https://en.wikipedia.org/wiki/List_of_counties_in_Oregon
- Data Description: Will scrape the wikipidea page with the table contatining the information of all the counties of Oregon.

2. For Latittude and Longitutdes

- Data Source: Geopy
- Data Description: The wikipedia page does not contain the latitudes and longitude information. So the data was obtained by using the Geocoder class of the Geopy client.

3. List of Venues and Venue categories.

- Data Source: Foursquare API's
- Data Description: By using the Foursquare API we can get the information of the venues in the counties. Based on this this data we can further analyse the data by selecting the county with the highest number of venues. We can again use Foursquare API to get data about the venue categories in the neighborhoods of the selected county.

METHODOLOGY

- DATA PREPARATION

As the State of Oregon is the state of interest, I first scraped the wikipidea page for information about counties in Oregon. I used pandas to scrape the table on wikipedia page and convert it to a pandas data frame.

I further cleaned the dataframe by dropping columns which we will not be using for analysis.

	County	County seat[4]	Est.[4]	Population[6]
0	Baker County	Baker City	1862	16510
1	Benton County	Corvallis	1847	91320
2	Clackamas County	Oregon City	1843	404980
3	Clatsop County	Astoria	1844	38225
4	Columbia County	Saint Helens	1854	50795
5	Coos County	Coquille	1853	63190
6	Crook County	Prineville	1882	21580
7	Curry County	Gold Beach	1855	22600
8	Deschutes County	Bend	1916	176635
9	Douglas County	Roseburg	1852	110395

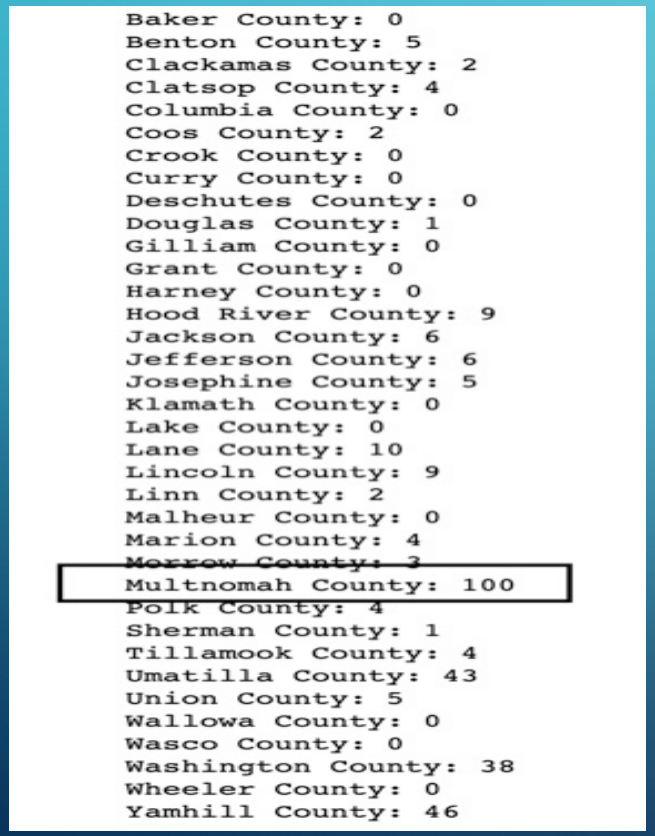
For using the foursquare API to get information of the venues and restaurants in the counties, the latitude and longitude information is required. this information is missing in our dataframe. So lets use the geocoder class of Geopy to get the coordinates of all the counties of Oregon.
This is how our resulting dataframe looks like.

]:

	County	County seat[4]	Est.[4]	Latitude	Longitude
0	Baker County	Baker City	1862	44.725964	-117.620482
1	Benton County	Corvallis	1847	44.494937	-123.406568
2	Clackamas County	Oregon City	1843	45.160882	-122.230504
3	Clatsop County	Astoria	1844	45.980728	-123.668750
4	Columbia County	Saint Helens	1854	45.928020	-123.082293
5	Coos County	Coquille	1853	43.218414	-124.109621
6	Crook County	Prineville	1882	44.146029	-120.383948
7	Curry County	Gold Beach	1855	42.426543	-124.216879
8	Deschutes County	Bend	1916	43.819484	-121.166966
9	Douglas County	Roseburg	1852	43.192211	-123.113596
10	Gilliam County	Condon	1885	45.329874	-120.220220
11	Grant County	Canyon City	1864	44.464824	-119.056528
12	Harney County	Burns	1889	43.055195	-119.024026
13	Hood River County	Hood River	1908	45.531164	-121.647559
14	Jackson County	Medford	1852	42.398015	-122.754095
15	Jefferson County	Madras	1914	44.607250	-121.246314

DATA ANALYSIS

- **Now we have all the information to call the Foursquare API. We will define the url and call the function to get the venues present in these counties of Oregon. The following image gives us the Counties with number of venues present.**



```
Baker County: 0
Benton County: 5
Clackamas County: 2
Clatsop County: 4
Columbia County: 0
Coos County: 2
Crook County: 0
Curry County: 0
Deschutes County: 0
Douglas County: 1
Gilliam County: 0
Grant County: 0
Harney County: 0
Hood River County: 9
Jackson County: 6
Jefferson County: 6
Josephine County: 5
Klamath County: 0
Lake County: 0
Lane County: 10
Lincoln County: 9
Linn County: 2
Malheur County: 0
Marion County: 4
Morrow County: 3
Multnomah County: 100
Polk County: 4
Sherman County: 1
Tillamook County: 4
Umatilla County: 43
Union County: 5
Wallowa County: 0
Wasco County: 0
Washington County: 38
Wheeler County: 0
Yamhill County: 46
```

As seen in the image on the left. Multnomah County promises to be the best candidate for further analysis with maximum number of venues returned as compared to other counties.

- Let us continue our analysis further by using Multnomah County. We will repeat the same procedure of getting data online and converting it to a dataframe. Then using Geopy to get the coordinates of each neighborhood of the County. Lastly using the Foursquare API to get Venue Locations for all the neighborhoods of Multnomah county.
- The API returned 262 unique venue categories. This includes parks, trials etc. But our business problem dictates that we use only restaurant data, so we create a new data frame with only restaurants as the venue category. We now have only 43 restaurant categories and the dataframe look like the image below.

County	County Latitude	County Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Alameda	45.548631	-122.636481	Pine State Biscuits	45.558981	-122.642697	Southern / Soul Food Restaurant
Alameda	45.548631	-122.636481	Kargi Gogo	45.559113	-122.634038	Dumpling Restaurant
Alameda	45.548631	-122.636481	Bollywood Theater	45.559270	-122.644011	Indian Restaurant
Alameda	45.548631	-122.636481	Urdaneta	45.559136	-122.634119	Tapas Restaurant
Alameda	45.548631	-122.636481	Beast	45.562462	-122.634985	French Restaurant

- Let us now analyze each neighborhood to get TOP 5 most common venues. Let do this by first creating a dataframe with one hot encoding.
- Then by using groupby on the neighborhood column we will calculate the frequency of occurrence for each of the unique venues

	Neighborhoods	American Restaurant	Argentinian Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Chinese Restaurant	Cuban Restaurant	...	Restaurant	Russian Restaura
1	Alameda	0	0	0	0	0	0	0	0	0	...	0	0
2	Alameda	0	0	0	0	0	0	0	0	0	...	0	0
4	Alameda	0	0	0	0	0	0	0	0	0	...	0	0
8	Alameda	0	0	0	0	0	0	0	0	0	...	0	0
15	Alameda	0	0	0	0	0	0	0	0	0	...	0	0

Dataframe resulting from One Hot Encoding

	Neighborhoods	American Restaurant	Argentinian Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Chinese Restaurant	Cuban Restaurant	...	Restaurant	Russian Restaurar
0	Alameda	0.034483	0.034483	0.000000	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.000000	0.0
1	Albina	0.041667	0.041667	0.000000	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.000000	0.0
2	Arleta	0.086957	0.000000	0.043478	0.0	0.0	0.043478	0.0	0.086957	0.0	...	0.000000	0.0
3	Barnes Heights	0.058824	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.058824	0.0
4	Bonneville	0.500000	0.000000	0.000000	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.500000	0.0

Dataframe displaying frequency of occurrence of different types of restaurants

- We get the Top 5 Most common Venues for each neighborhood.

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Alameda	Thai Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Southern / Soul Food Restaurant	French Restaurant
1	Albina	Thai Restaurant	Sushi Restaurant	New American Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant
2	Arleta	Vietnamese Restaurant	Mexican Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant
3	Barnes Heights	Italian Restaurant	Mexican Restaurant	Restaurant	French Restaurant	Hawaiian Restaurant
4	Bonneville	American Restaurant	Restaurant	Dumpling Restaurant	Indian Restaurant	Hawaiian Restaurant
...	...	...	...	...	...	...
56	West Portland Park	Mexican Restaurant	American Restaurant	Sushi Restaurant	Seafood Restaurant	German Restaurant
57	Westmoreland	Italian Restaurant	Sushi Restaurant	Seafood Restaurant	Vietnamese Restaurant	American Restaurant
58	Whitwood Court	Mexican Restaurant	Thai Restaurant	Italian Restaurant	Chinese Restaurant	Falafel Restaurant
59	Willamette Heights	Italian Restaurant	Restaurant	French Restaurant	Korean Restaurant	Mediterranean Restaurant
60	Woodstock	Vietnamese Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Sushi Restaurant

- As mentioned in the Business Problem section of the notebook. We are trying to help someone intrerested to open a business in Oregon, determine which neighborhood would be best. They have an idea for a Fusion Restaurant in which they wish to combine the spices and flavours of Thai cuisine with the dishes loved in American cuisine.
- So lets create a new dataframe from the dataframe with top most common venues by selecting those neighborhoods in which Thai cuisine is much loved and features in the 1st Most Common Venue

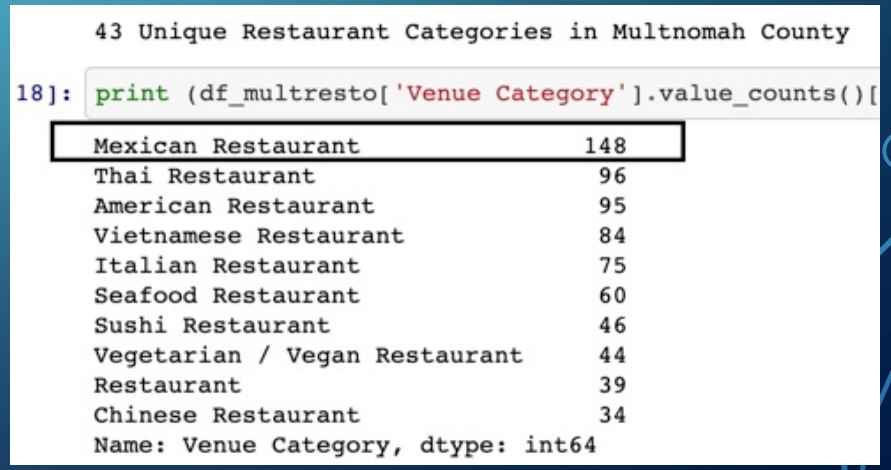
	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Alameda	Thai Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Southern / Soul Food Restaurant	French Restaurant
1	Albina	Thai Restaurant	Sushi Restaurant	New American Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant
25	Irvington	Thai Restaurant	Middle Eastern Restaurant	Southern / Soul Food Restaurant	French Restaurant	New American Restaurant
28	Kenton	Thai Restaurant	Mexican Restaurant	American Restaurant	New American Restaurant	Mediterranean Restaurant
30	Laurelhurst	Thai Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Modern European Restaurant	Cuban Restaurant
47	Saint Johns	Thai Restaurant	Mexican Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant
52	University Park	Thai Restaurant	Mexican Restaurant	Seafood Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant

- Those neighborhoods in which American Cuisine also features in the Most common venues is also of interest to us as the fusion restaurant will also attract customers who relish dishes of American Cuisine
- So let us create a new and final dataframe with the Neighborhoods which satisfy both the criterias and have both Thai as well as American Cuisines in the most common venues. The resulting dataframe will give us the best Neighborhoods for this Fusion Restaurant that our client wants to invest in.

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Albina	Thai Restaurant	Sushi Restaurant	New American Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant
28	Kenton	Thai Restaurant	Mexican Restaurant	American Restaurant	New American Restaurant	Mediterranean Restaurant
28	Kenton	Thai Restaurant	Mexican Restaurant	American Restaurant	New American Restaurant	Mediterranean Restaurant
25	Irvington	Thai Restaurant	Middle Eastern Restaurant	Southern / Soul Food Restaurant	French Restaurant	New American Restaurant

RESULTS

- By this analysis we got an idea about which restaurants are most liked in Portland, the county seat of the Multnomah County.
- We determined the best neighborhoods for the Fusion Thai-American Restaurant.
- We were also able to determine by this analysis that Mexican Restaurants top the list in most number of restaurants in the Multnomah County. This can be of interest to someone who is interested to open a Mexican Restaurant or give a business idea to someone who wants to open a restaurant in Multnomah County but wants a good analysis of which restaurants are most liked.



DISCUSSIONS

- We carried out this analysis only on the basis of data gathered from Foursquare API. We have not considered other factors like demographics, foot traffic, property rates, restaurant prices, if the neighborhood has predominantly businesses or residential complexes, schools, proximity to public transportation system etc. Analysis taking all these factors into consideration will yield a better result of the best neighborhood.
- This analysis is a very basic one but it gives us a good idea of what the businesses are operating in Portland, what kind of food people prefer and which would be good neighborhoods to start a restaurants depending on its Cuisine.

CONCLUSION

- Analysis was done on data to determine the best neighborhoods for a Fusion Thai-American Restaurant in Oregon, USA. This analysis involved data collection by web scraping , Geopy, Foursquare API. Pandas dataframes were used for the analysis. The result of the analysis gave us 3 good candidate neighborhoods for the Restaurant. We were also able to determine the most loved cuisine. The analysis can be improved and made more accurate by considering more parameters rather than just venue data.
- However, in conclusion our business problem was solved. We were able to determine the best neighborhood. This would be a good trial run for our customer who wants to test the waters in a city not as large, costly, populated as San Francisco but comes close with regards to demographics and has potential for tremendous growth.