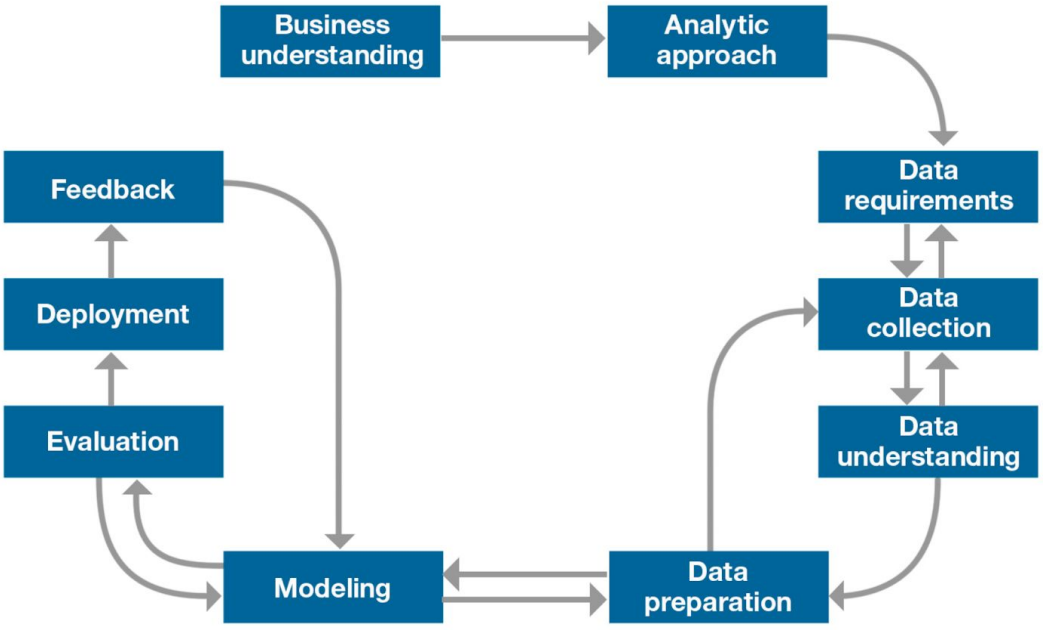
The Battle of Neighborhoods

Pet grooming Service

Introduction

The scope of the project is to find an optimal location for the pet groom shop in Wake county. We will use the approach outlined in the IBM process to arrive at the findings.



Business understanding

Business Understanding and a clear problem statement is critical for analysis. This defines expectations with the stakeholders we are trying to solve. The key driver for this stage is the problem statement.

Business Problem

Times have changed and in this pandemic era people are more connected with one another from their home. The social distancing and public gathering has led people to spend more time with their loved ones at home. The need for companionship has led to a sudden surge of people getting new pet. The companionship of a pet gives them immense joy and a let out. With the increase in adoption of the pet there is a sudden increase in the need for pet related services like grooming, pet care, vet services etc. In this project we will focus on one such service i.e. Pet grooming. We will try to find the best location to open the pet grooming service in the Wake county area, North Carolina, USA. The goal again is to find the ideal location for the new pet grooming store given the Wake county population distributions, median household income and venu popularity.

Data (Requirements, Collections and Understanding)

We need to get two kinds for data to overlay and understand the current market condition, i.e. Townships and the existing popular venue data around these townships. The data sources that we will be using for the analysis are cataloged below

1. Data stats for market research www.iii.org

   Will provide neighbourhood specific details about the income, population and etc to understand the demographic breakdown.

2. Wake county area is one of the largest counties in NC The wake county consists of 20 major townships. Details for this town ships will be gathered from the NC wake county website and wikipedia

   The wake county data provide major townships, their population, latitude and longitude.

3. The Geo coordinates need for the analysis will be sourced from the Wikipedia, google maps and geocode

   The Latitude and Longitude from this will be used to get the popular venue locations around these locations for further analysis.

4. Median household income details will be sourced from statisticalatlas https://statisticalatlas.com/

5. Top venues will be sourced using Four square.

   Foursquare will provide the the popular venues around the townships based on their latitude and longitude. We will use this data to understand the traffic pattern and customer choices available to provide suitable recommendations.

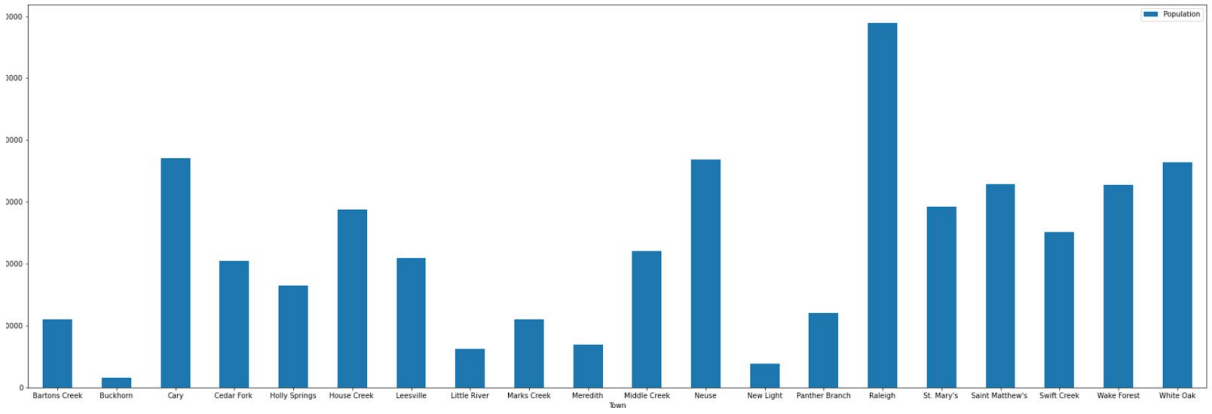
Data Preparation

The neighbourhood data in the wikipedia for the wake county is not available in a tabular format or downloadable CSV. The unstructured data needs to be processed to create a structured data as shown that can be further used for the analysis.

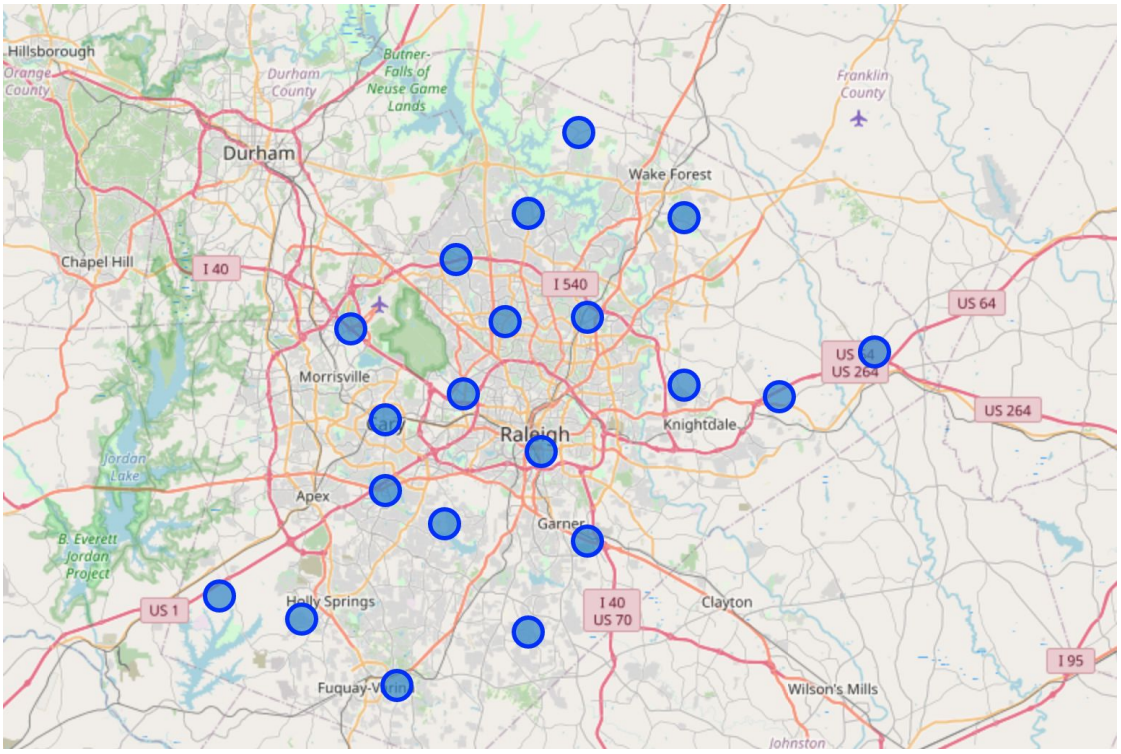
- Township Data

	Town	Latitude	Longitude	Population
0	Bartons Creek	35.950664	-78.645547	22055
1	Buckhorn	35.655147	-78.940138	3251
2	Cary	35.791357	-78.781622	74074
3	Cedar Fork	35.861111	-78.814722	40841
4	Holly Springs	35.636988	-78.860912	33071
5	House Creek	35.866778	-78.668240	57439
6	Leesville	35.914125	-78.713608	41850
7	Little River	35.843485	-78.315941	12528
8	Marks Creek	35.809344	-78.406973	21932
9	Meredith	35.811332	-78.707938	13926
10	Middle Creek	35.585306	-78.770290	44136
11	Neuse	35.869952	-78.588795	73617
12	New Light	36.012532	-78.597331	7591
13	Panther Branch	35.627506	-78.645548	24019
14	Raleigh	35.766667	-78.633333	117838
15	St. Mary's	35.697611	-78.588795	58484
16	Saint Matthew's	35.818162	-78.497925	65731
17	Swift Creek	35.710555	-78.724947	50225
18	Wake Forest	35.947230	-78.497925	65491
19	White Oak	35.736724	-78.781006	72894

- Population Distribution



- Township Map view



Using the FourSqure API we were able to fetch the popular venues close to the center of the townships. The foursquare venue data returned by the API is in JSON format. This JSON data is coverage to table data for processing using Pandas with the help for JSON library.

- JSON Data

```
{"meta":{"code":200,"requestId":"5fb55ee0cce9aa1674ccab87"},"response":{"suggestedFilters":
{"name":"$-$$$$","key":"price"}]},"headerLocation":"Raleigh","headerFullLocation":"Raleigh'
Places","name":"recommended","items":[{"reasons":{"count":0,"items":[{"summary":"This spot
Rd","lat":35.927308562873975,"lng":-78.62172058023836,"labeledLatLngs":[{"label":"display",
["10809 Honeycutt Rd","Raleigh, NC 27614","United States"]},"categories":[{"id":"4bf58dd8d4
{"count":0,"groups":[]}},"referralId":"e-0-4bb761e06edc76b0d316321c-0"},{"reasons":{"count'
Church Rd","lat":35.969983391878316,"lng":-78.64017963409424,"labeledLatLngs":[{"label":"di
27614","United States"]},"categories":[{"id":"4bf58dd8d48988d163941735","name":"Park","plur
4ba629a0f964a520393839e3-1"},{"reasons":{"count":0,"items":[{"summary":"This spot is popula
Rd","lat":35.907475795244515,"lng":-78.65719431273305,"labeledLatLngs":[{"label":"display",
["9216 Baileywick Rd (at Lead Mine Rd)","Raleigh, NC 27615","United States"]},"categories":
```


- Pandas Dataframe

Preprocessed jason data with venue data combined with Twonship data as panda data frame

	Town	Population	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Bartons Creek	22055	35.950664	-78.645547	Homewood Nursery & Garden Center	35.927309	-78.621721	Hardware Store
1	Bartons Creek	22055	35.950664	-78.645547	Blue Jay Point County Park	35.969983	-78.640180	Park
2	Bartons Creek	22055	35.950664	-78.645547	A.E. Finley YMCA	35.907476	-78.657194	Gym
3	Bartons Creek	22055	35.950664	-78.645547	Sassool	35.903908	-78.656748	Mediterranean Restaurant
4	Bartons Creek	22055	35.950664	-78.645547	Goodberry's Frozen Custard	35.902244	-78.656899	Ice Cream Shop
5	Bartons Creek	22055	35.950664	-78.645547	Great Harvest Bread Company	35.905211	-78.656980	Bakery
6	Bartons Creek	22055	35.950664	-78.645547	Whole Foods Market	35.899209	-78.650282	Grocery Store
7	Bartons Creek	22055	35.950664	-78.645547	Taylor's Wine Shop	35.915844	-78.650133	Wine Shop
8	Bartons Creek	22055	35.950664	-78.645547	Poppyseed Market Café & Wine Bar	35.905084	-78.657466	Café
9	Bartons Creek	22055	35.950664	-78.645547	Char-Grill	35.901271	-78.655816	Burger Joint

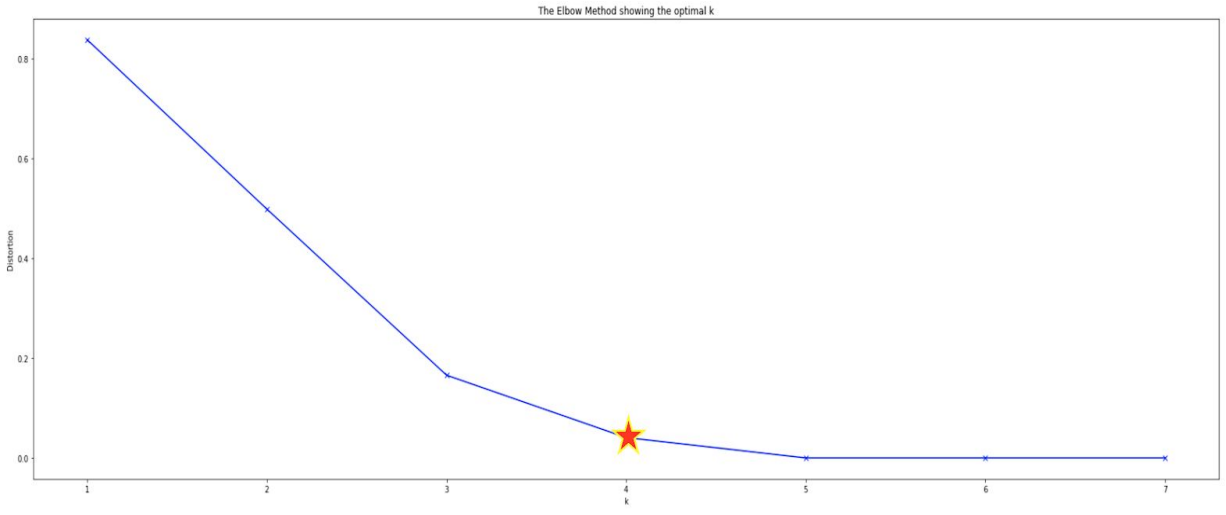
- Filtered data for relevant venues and categorized them for clustering

	key_0	Town	Population	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory	Department Store	Pet Service	Pet Store	Shopping Mall	Veterinarian
0	787	Marks Creek	21932	35.809344	-78.406973	T.J. Maxx	35.797542	-78.512520	Department Store	1	0	0	0	0
1	1162	New Light	7591	36.012532	-78.597331	Kohl's	35.987050	-78.530997	Department Store	1	0	0	0	0
2	1539	Saint Matthew's	65731	35.818162	-78.497925	Kohl's	35.798671	-78.511095	Department Store	1	0	0	0	0
3	692	Little River	12528	35.843485	-78.315941	Belk	35.832183	-78.316864	Department Store	1	0	0	0	0
4	65	Bartons Creek	22055	35.950664	-78.645547	Kohl's	35.905490	-78.602343	Department Store	1	0	0	0	0
5	722	Little River	12528	35.843485	-78.315941	Kannon's Of Wendell	35.781359	-78.369921	Department Store	1	0	0	0	0
6	1561	Saint Matthew's	65731	35.818162	-78.497925	T.J. Maxx	35.797542	-78.512520	Department Store	1	0	0	0	0
7	1105	Neuse	73617	35.869952	-78.588795	Kohl's	35.905490	-78.602343	Department Store	1	0	0	0	0
8	1459	St. Mary's	58484	35.697611	-78.588795	Kohl's	35.693394	-78.580289	Department Store	1	0	0	0	0
9	1035	Neuse	73617	35.869952	-78.588795	Dillard's	35.865909	-78.577609	Department Store	1	0	0	0	0

Modelling

Based on the business requirement, We need to analyze location based data and Venu popularity to arrive at the recommendations. Grouping the venues around around key location and analysing them as clusters will be a natural fit for this kind for the business problem in hand. We will be using K means algorithm to classify and analyze to arrive at the recommendation.

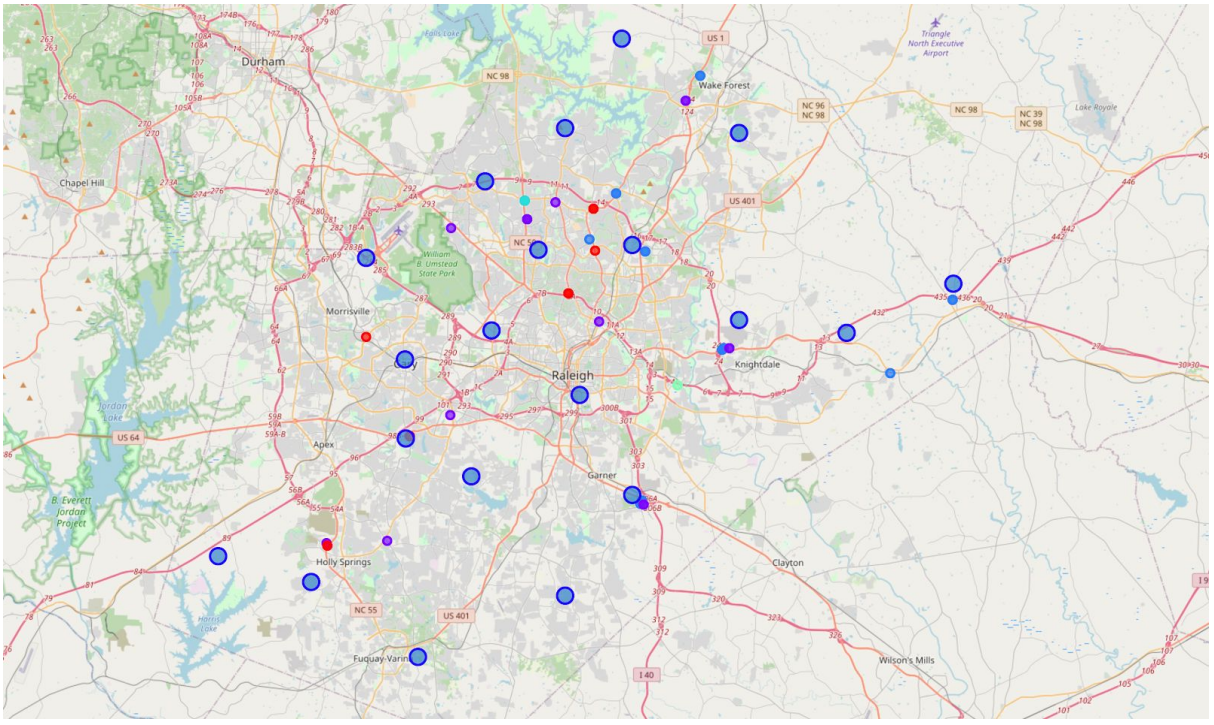
K-means algorithm is one of the most common cluster methods of unsupervised learning. The K-means will be run for 4 clusters. When the optimal elbow method showed 4 as the optimal k value.



Using 4 cluster configuration for k-means, clustering analysis was run using the ssklearn library. The results from the cluster analysis is plotted over the township data using the folium api.

When the initial data set was run only for PETs, Vets and Services the foursquare API propopular venue result set was too small the model was reanalysed and the additional category where pet service could be and option like Vets, Malls were included in the analysis. This kind of modeling, evaluation and fine tuning the mode based on the analysis feedback is an important aspect in the framework.

Cluster Analysis



Results

Based on the clustering analysis we can infer the following

1. Small towns Panther Branch, buckhon, swift creek, New light , Barton creek doesnt have any services
2. Cary one of the hightest population does not have any service
3. Raleigh one of the hightest population does not have any service
4. Cary shopping mall is an option
5. Shopping malls and department stores are are high traffic areas

Discussion

1. Raleigh and Cary looks like a good option for the Standalone Pet grooming shops. These cities are large and lack this kind of service.
2. Mall and department stores are the most popular venues based on the foursq api. These places attract lots of foot traffic. Having a pet grooming location will be an ideal option inside a mall in Cary and Raleigh. The clients will be able to drop off their pets for grooming service while they can shop in the mall is an additional benefit that will attract lots of customers.

3. The second best option for having a standalone location is in between Leesville and Neuse . This location has lots of popular venues that will attract foot traffic

Conclusion

As people are always looking for companionship, pets are becoming very popular options. The growth in the pets presents a tremendous market as the current market is having a void in catering the demands for the pet needs. The analysis data show there pet service market in wake county has great potential.