Your Road

Abstract

This project is describing the implementation of machine learning in the way of how the tourists could choose their trip. Most of web sites might give you a recommendation includes the price of hotels, shortest road and other useful things. However, it is still missing giving you a most convenient recommendation which fit your style, requirements, mode, needs or even gives you something difference regardless the time or the cost.

The project will use unsupervised machine learning method, Clustering, using K means as a way to cluster the venues of the predetermined latitude or longitude of the proposed trip. then evaluate if it fit tourist profile or not. a map with the clusters will help the tourist to choose other better roads passing in areas fit his requirements.

Data is available by Foursquare location data provider. By exploring the venues around the located latitude and longitude. the category for each venue will be provided then form a data frame including these venues and categories. Data exploratory and analysis are a long process to build a data structure fit to the model.

In conclusion, the tourist or the new commers will get a simple tool which will help the tourist to choose different roads for his target through a recommendations of machine learning clustering application **" *Your Road*".**

Introduction

Planning a trip is not become a mysterious as before with the available location data providers includes google maps, foursquare, and others. However, choosing the convenient trip still a little bit fuzzy not because of the rare of information but the toooo much information you can get. Machine learning is a magnificent tool to solve such problem and gives you what fits your profile and character if you need.

"Your Road" will gives you a simple map with colored clustered like **Figure-1**. those colored areas will help you to choose the road which fit your type or profile. The tourist should have a profile with a lot of features which will help the system to compare between yours features and others. and in this project, the supposed trip will be started from **Vancouver, CA** and end at **Los Angeles, US**.

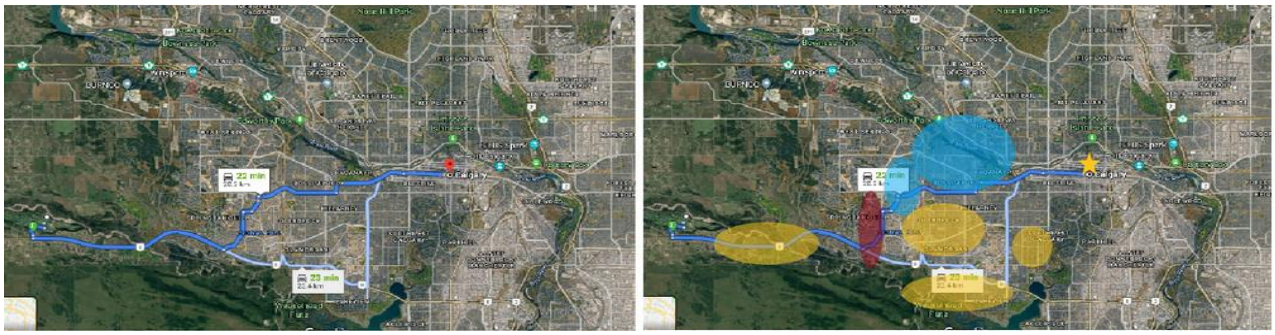

Figure-1: Example of the project expectations

Methodology

Unsupervised machine learning approach, Clustering, has been chosen for this project. K means is one of the methods which commonly used in this kind of application. The Location data provider ***FourSquare*** will be the data source for the venues located in the predetermined coordinates. In addition, Tourist profile will full of input features which will be processed to be able to be clustered with the available Venues data.

Data pre-processing

Data requirements

First, Tourist or user profile will have a much of features which will help in unsupervised learning and comparison. All these data will be uploaded by the user itself using a drop list choice to standardize and categorize the data. in this project this data will be estimated from our side.

Second, Coordinates of the trip of the start point and destination point are also will be provided by the user. However, the available roads will be suggested by the location data provider like Google map. in this project these coordinates will be estimated from our side like the user profile.

Third, Foursquare location data provider will be used for exploring the venues located in the predetermined coordinates in the all estimated roads stations. **Figure-2** is an example of data used in project from Foursquare.

	station	station Latitude	station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Vancouver,CA	49.260872	-123.113953	Whole Foods Market	49.264018	-123.115303	Grocery Store
1	Vancouver,CA	49.260872	-123.113953	Vij's	49.257525	-123.115083	Indian Restaurant
2	Vancouver,CA	49.260872	-123.113953	La Taqueria Pinche Taco Shop	49.263534	-123.112795	Taco Place
3	Vancouver,CA	49.260872	-123.113953	33 Acres Brewing Company	49.263857	-123.105330	Brewery
4	Vancouver,CA	49.260872	-123.113953	Aperture Coffee Bar	49.263252	-123.110438	Coffee Shop

Figure-2: Data retrieved from Foursquare

Data understanding and preparation

Data understanding, figuring out the data resources and mislead data are vital in the understanding process of data. In addition, Cleaning mis values and mislead values then standardization and categorization are a cornerstone to help the model to be healthy avoiding irrationality. **Figure-3** is an example of the process applied in our project.

```
# one hot encoding
Trip_1_onehot = pd.get_dummies(Trip_1_venues[['Venue Category']], prefix="", prefix_sep="")


# # add neighborhood column back to dataframe
Trip_1_onehot['station'] = Trip_1_venues['station']

fixed_columns = [Trip_1_onehot.columns[-1]] + list(Trip_1_onehot.columns[:-1])
Trip_1_onehot = Trip_1_onehot[fixed_columns]

Trip_1_onehot.head()
```

Figure-3 Normalization of the venues data

Modeling

After prepare the data to be fit as model require, start building an unsupervised machine learning model has done using one of the available methods like K means. Modelling the Data using a K means method is an iterative way to get the best K assumption and other input parameters. However, in this project, the process has been iterating once as in **Figure-5**. Other more advanced tuning parameters are out of scope of this project.

```
kclusters = 4

total_Trip_grouped_clustering = total_Trip_grouped.drop('station', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(total_Trip_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Figure-5: K MEANS, Clustering process

Results and discussion

As a result of Modelling, the user will have another parameter to take in consideration while preparing the trip. **Figure-6** is showing a result of the model, the different colored clusters gives a new information for the travelers to decide which road will take beside other features available like distance, cost,..etc.

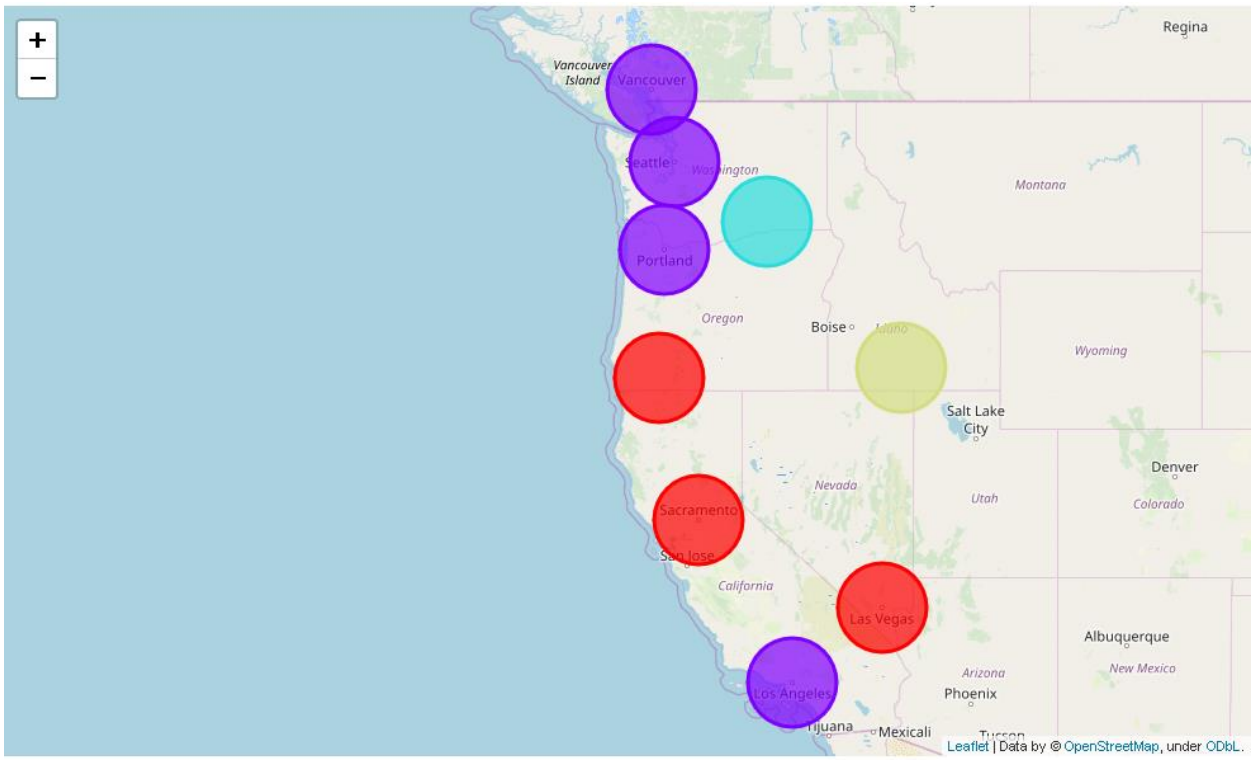

Figure-6: Clustering process result on the Map

Finally, we did our job and ***now it's your turn to entertain***!