Sushi Bar in Saint Petersburg, Russia

June 21, 2019

0.0.1 Introduction/Business Problem

Sushi is traditional Japanese food is gaining great popularity for many people especially in **Saint Petersburg, Russia**. Saint Petersburg is Russia's second-largest city after Moscow, with about 5,4 million inhabitants in 2019 and with lots of business opportunities and business friendly environment. Since the number of sushi bars in Saint Petersburg is rather small, customers may be interested in opening additional sushi bars in the most favorable neighborhoods of the city. However, any new business venture or expansion in the country needs to be reviewed carefully and strategically targeted so that the return on investment will be sustainably reasonable and more importantly the investment can be considerably less risky. Particularly, the location of the sushi bar is one of the most important decisions that will determine whether the bar will be a success or a failure.

The objective of this capstone project is to analyze and select the best locations in the city of Saint Petersburg, Russia to open a new sushi bar. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the following business question: **Which neighborhoods would be a good choice for opening a new sushi bar in Saint Petersburg, Russia?**

0.0.2 Data description

To solve this problem, we will need the following data:

- List of neighbourhoods in Saint Petersburg.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to sushi bars. We will use this data to perform clustering on the neighbourhoods.

Unfortunately, the Saint Petersburg neighborhood data is not widely available on the Internet in the structured format, hence we need to scrap it through an existing Wikipedia page https://en.wikipedia.org/wiki/Category:Districts_of_Saint_Petersburg that has all the information we need to explore and cluster the neighborhoods in Saint Petersburg.

Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

And after that, we will use Foursquare API to get the venue data for those neighbourhoods. Since Foursquare has one of the largest database used by many developers around the world, we will use it to get information about *Sushi Restaurant* category of the venue data in order to help us to solve the business problem put forward.

The data before feature engineering step looks like as follows:

In [4]:
```
import pandas as pd
pd.read_csv('Saint Petersburg_venues.csv').head()
```

Out[4]:

	Neighborhood	Latitude	Longitude	VenueName	\
0	Admiralteysky District	59.92659	30.3056	.	
1	Admiralteysky District	59.92659	30.3056	Chao, mama!	
2	Admiralteysky District	59.92659	30.3056	CUP IN CUP	
3	Admiralteysky District	59.92659	30.3056		
4	Admiralteysky District	59.92659	30.3056	Pacman	

	VenueLatitude	VenueLongitude	VenueCategory
0	59.926702	30.307921	Bakery
1	59.926993	30.308474	Hotel
2	59.928074	30.302705	Coffee Shop
3	59.926154	30.310403	Clothing Store
4	59.923537	30.307985	Hookah Bar