Exploring the Cuisines of Los Angeles City

1. Introduction and a discussion of the background

Los Angeles City (also simply known as L.A., and nicknamed the "City of Angels") is California's most populous city. The metropolitan area is the second most populated area in the U.S. and home to more than 17 million residents from all over the globe.

Los Angeles is known for its locations and its food. There are many food trucks, hotels, and restaurants that are renowned for their different delicacies. The cuisines of many cultures have taken root and it is filled with restaurants and delicacies that are made only in LA and also served with various dishes. With more restaurants opening every week, L.A. is topping the to-do lists of every food obsessive in America right now.

The aim of this project is to segment L.A's neighbourhoods categorically. In major clusters and test their cuisines. A good purpose is to look at the food preferences and taste of the neighborhood cluster. Further examination could reveal whether food has any relation with the neighborhood's diversity.

This project will help to understand the diversity of a neighborhood by leveraging venue data from Foursquare's 'Places API' and 'k-means clustering' machine learning algorithm. Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighborhood.
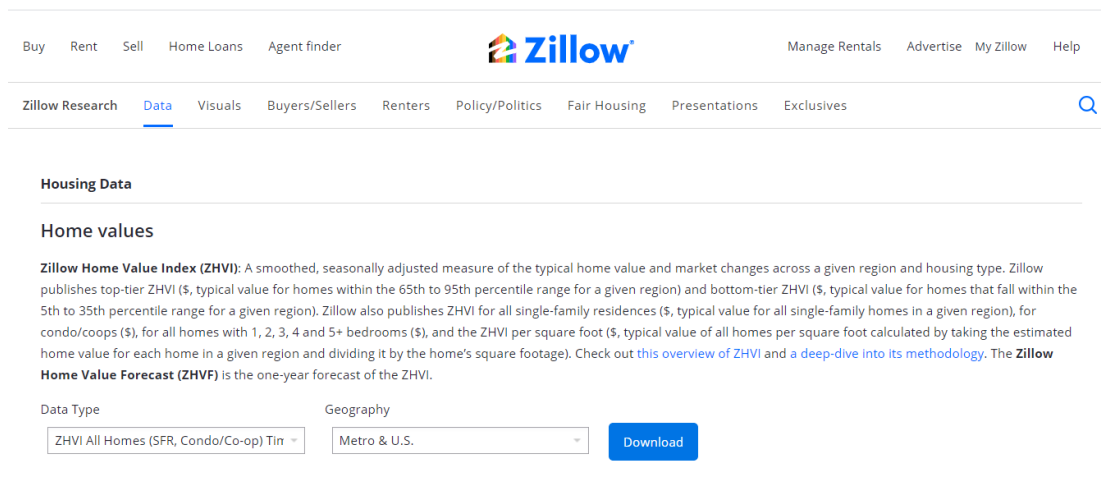

Target Audience

- Stakeholders will be the one involved in using this quantifiable research to consider how various cultures and cuisines are spread over LA. Also, a new food vendor willing to open his or her restaurant can make use of this project. Or by a government authority to better examine and study the diversity of their city's culture.
- Business Analyst or Data Scientists who want to analyze LA's neighborhoods using Exploratory Data Analysis and other statistical & machine learning techniques to obtain all the data they need, perform some operations on it and finally be able to tell a story from it.

2. Source of the Data

- **Zillow**

Link - https://www.zillow.com/research/data/

A complete list of neighborhoods of Los Angeles was downloaded in CSV format from Zillow Housing Data.

- **Foursquare API –**

Link - https://developer.foursquare.com/docs

To get location and other information about various Food venues in LA I'm using Foursquare's explore API. Using the Foursquare's venue API with mentioned category (which gives venues recommendations), I'm fetching details about the food venues up present in LA and collected their names, categories and locations (latitude and longitude).

Import required packages

All the required packages for analyzation and clustering are imported

```
In [2]:  1  import pandas as pd # library for data analsysis
         2  import numpy as np # library to handle data in a vectorized manner
         3
         4  # import wget
         5  import os
         6
         7
         8  import json # library to handle JSON files
         9  from pprint import pprint # data pretty printer
         10
         11  import requests # library to handle requests
         12  from bs4 import BeautifulSoup  # library to handle web scraping
         13
         14  from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
         15
         16  import folium # map rendering library
         17
         18  import matplotlib.cm as cm # Matplotlib and associated plotting modules
         19  import matplotlib.colors as colors # Matplotlib and associated plotting modules
         20  import matplotlib.pyplot as plt
         21
         22  from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe
         23
         24  from collections import Counter # count occurrences
         25
         26  from sklearn.cluster import KMeans # import k-means from clustering stage
```
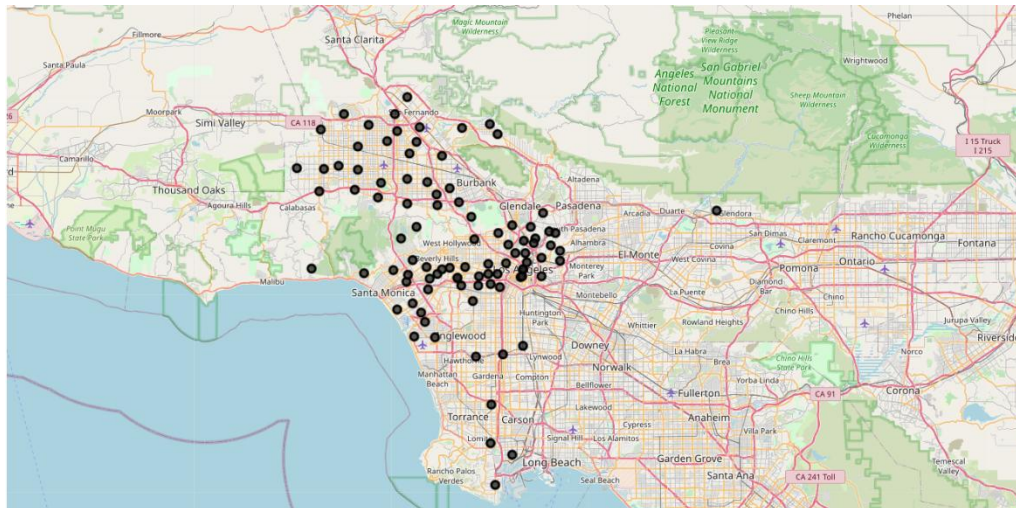
After creating a dataframe for the neighborhoods of LA, use geopy's nominatum library to fetch location details

Use **geopy's nominatum** library to find latitude & longitude for the given location

```
In [4]:  1  #Finding coordinates of Los Angeles city using geolocator
         2  address = 'Los Angeles City, California'
         3  location = None
         4
         5  # define an instance of the geocoder -> foursquare_agent
         6  while location == None:
         7      try:
         8          geolocator = Nominatim(user_agent="foursquare_agent")
         9          location = geolocator.geocode(address)
         10         latitude = location.latitude
         11         longitude = location.longitude
         12     except:
         13         pass
         14  print('The geograpical coordinate of Los Angelse City are {}, {}.'.format(latitude, longitude))
```

The geograpical coordinate of Los Angelse City are 34.0536909, -118.2427666.

Once the location details of nieghborhoods of LA fetched, create a folium MAP to visualize it

Foursquare API Credentials and Version

```
In [11]:   1  CLIENT_ID = '                                    ' # your Foursquare ID
           2  CLIENT_SECRET = '                                ' # your Foursquare Secret
           3  VERSION = '20180605' # Foursquare API version
           4
           5  print('Your credentails:')
           6  print('CLIENT_ID: ' + CLIENT_ID)
           7  print('CLIENT_SECRET:' + CLIENT_SECRET)

           Your credentails:
           CLIENT_ID:
           CLIENT_SECRET:
```

Explore available categories in Foursquare API

```
_____
Arts & Entertainment
categories 36

_____
College & University
categories 23

_____
Event
categories 12

_____
Food
categories 92

_____
Nightlife Spot
categories 7

_____
Outdoors & Recreation
categories 62

_____
Professional & Other Places
categories 43

_____
Residence
categories 5

_____
Shop & Service
categories 145

_____
Travel & Transport
categories 34
```

Explore the sample response format of food category

```
Out[18]:  {'id': '4b8736bef964a520c4b631e3',
          'name': 'Jack in the Box',
          'location': {'address': '11400 S Figueroa St',
           'lat': 33.93072024913725,
           'lng': -118.28219890594482,
           'labeledLatLngs': [{'label': 'display',
             'lat': 33.93072024913725,
             'lng': -118.28219890594482},
            {'label': 'entrance', 'lat': 33.93078, 'lng': -118.282333}],
           'distance': 413,
           'postalCode': '90061',
           'cc': 'US',
           'city': 'Los Angeles',
           'state': 'CA',
           'country': 'United States',
           'formattedAddress': ['11400 S Figueroa St',
            'Los Angeles, CA 90061',
            'United States']},
          'categories': [{'id': '4bf58dd8d48988d16e941735',
            'name': 'Fast Food Restaurant',
            'pluralName': 'Fast Food Restaurants',
            'shortName': 'Fast Food',
            'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/fastfood_',
             'suffix': '.png'},
            'primary': True}],
          'referralId': 'v-1592069339',
          'hasPerk': False}
```

Explore all Neighborhoods of LA to get FOOD venues within 5 miles range and with a limit of 1000 venues

Now lets repeat this process for each and every Neighborhood of LA to get FOOD venues within 5 miles range and with a limit of 1000 venues

```
In [20]:   1  def getNearbyFood(names, latitudes, longitudes, radius=8046, LIMIT=1000):
           2      not_found = 0
           3      CLIENT_ID = 'OEVQYGFMA22NPP0GVV4TEGS4CDNU2UWJXZK1FV2YV4ZLKKLH' # your Foursquare ID
           4      CLIENT_SECRET = 'YEZ4GJ2OBG1ODWZ1PQ53EAMUTZFXHKM45ORK5OGEMOVV5PQ2'
           5      VERSION = '20180605' # Foursquare API version
           6      venues_list=[]
           7      for name, lat, lng in zip(names, latitudes, longitudes):
           8          # create the API request URL
           9          url = 'https://api.foursquare.com/v2/venues/search?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&categoryId={}&limit={}'.format(
          10              CLIENT_ID,
          11              CLIENT_SECRET,
          12              VERSION,
          13              lat,
          14              lng,
          15              radius,
          16              "4d4b7105d754a06374d81259", # "Food" category id
          17              LIMIT)
          18
          19          try:
          20              # make the GET request
          21              results = requests.get(url).json()['response']['venues']
          22
          23              # return only relevant information for each nearby venue
          24              venues_list.append([(
          25                  name,
          26                  lat,
          27                  lng,
          28                  v['name'],
          29                  v['location']['lat'],
          30                  v['location']['lng'],
```

Create dataframe with neighborhood and venue details

	index	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	0	South Los Angeles	33.928291	-118.278813	Carl's Jr.	33.902674	-118.351993	Fast Food Restaurant
1	1	South Los Angeles	33.928291	-118.278813	McDonald's	33.902275	-118.335027	Fast Food Restaurant
2	2	South Los Angeles	33.928291	-118.278813	Taco Bell	33.901042	-118.344162	Fast Food Restaurant
3	6	South Los Angeles	33.928291	-118.278813	Popeyes Louisiana Kitchen	33.976057	-118.357958	Fried Chicken Joint
4	10	South Los Angeles	33.928291	-118.278813	Chipotle Mexican Grill	34.016956	-118.282584	Mexican Restaurant

```
n [30]:    1  la_venues_df.shape
```

Out[30]: (1818, 8)

Perform one hot encoding and aggregate with MEAN for neighborhood

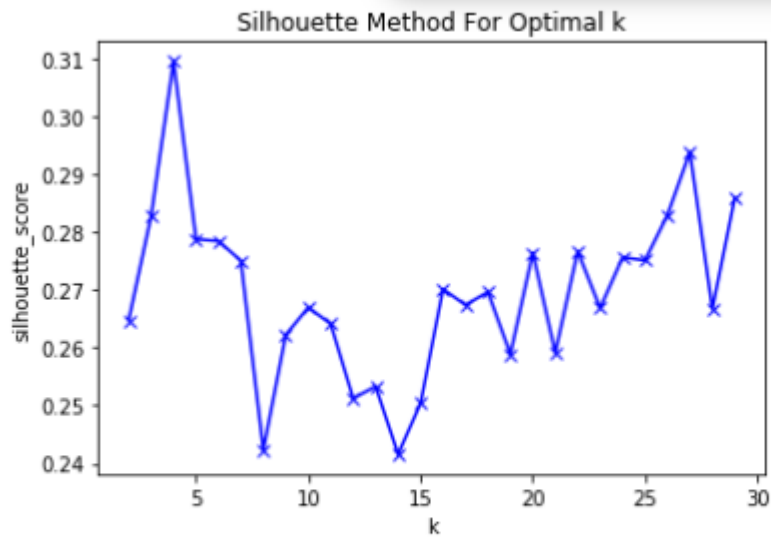|	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Asian Restaurant	Australian Restaurant	BBQ Joint	Brewery	Chinese Restaurant	Cuban Restaurant	...	Steakhouse	Sushi Restaurant	Taco Place	Taiwanese Restaurant	Temple	T Res
0	Adams-Normandie	0.0	0.0	0.000000	0.0	0.0	0.000000	0.052632	0.000000	0.0	...	0.0	0.000000	0.052632	0.0	0.000000	
1	Arleta	0.0	0.0	0.000000	0.0	0.0	0.043478	0.000000	0.000000	0.0	...	0.0	0.000000	0.000000	0.0	0.043478	
2	Arlington Heights	0.0	0.0	0.000000	0.0	0.0	0.000000	0.062500	0.000000	0.0	...	0.0	0.000000	0.000000	0.0	0.000000	
3	Atwater Village	0.0	0.0	0.000000	0.0	0.0	0.000000	0.058824	0.000000	0.0	...	0.0	0.058824	0.058824	0.0	0.000000	
4	Bel Air	0.0	0.0	0.055556	0.0	0.0	0.000000	0.000000	0.055556	0.0	...	0.0	0.000000	0.000000	0.0	0.000000	

5 rows × 51 columns

Identified top 9 Food categories

	Neighborhood	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
0	Adams-Normandie	Fast Food Restaurant	Fried Chicken Joint	Vegetarian / Vegan Restaurant	Seafood Restaurant	Taco Place	Brewery	Southern / Soul Food Restaurant	Shopping Mall	Dumpling Restaurant
1	Arleta	Fast Food Restaurant	Mexican Restaurant	Vietnamese Restaurant	Falafel Restaurant	Fried Chicken Joint	Middle Eastern Restaurant	Noodle House	Pizza Place	Italian Restaurant
2	Arlington Heights	Fast Food Restaurant	Fried Chicken Joint	Vegetarian / Vegan Restaurant	Brewery	Shopping Mall	Japanese Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant
3	Atwater Village	Fast Food Restaurant	Fried Chicken Joint	Vegetarian / Vegan Restaurant	Greek Restaurant	Taco Place	Sushi Restaurant	Brewery	Shopping Mall	Vietnamese Restaurant
4	Bel Air	Fast Food Restaurant	Udon Restaurant	Japanese Restaurant	American Restaurant	Greek Restaurant	Fried Chicken Joint	Mediterranean Restaurant	Chinese Restaurant	Mexican Restaurant

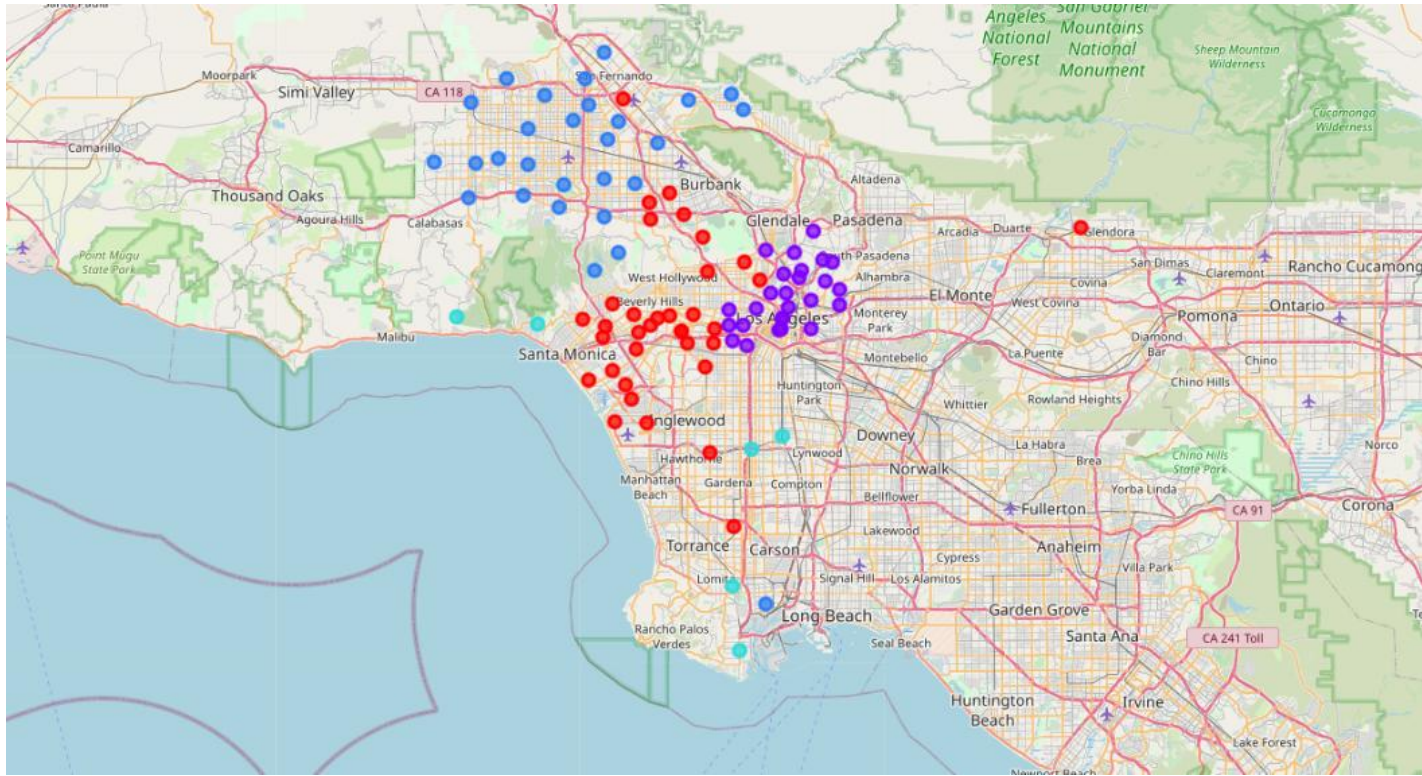## Optimized k-value using Silhouette score

Silhouette Method For Optimal k

Since there is a peak at K=4, lets set the optimized K values to 4

Clustered neighborhoods

	Cluster Labels	Neighborhood	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	Latitude	Longitude
0	1	Adams-Normandie	Fast Food Restaurant	Fried Chicken Joint	Vegetarian / Vegan Restaurant	Seafood Restaurant	Taco Place	Brewery	Southern / Soul Food Restaurant	Shopping Mall	Dumpling Restaurant	34.031788	-118.300247
1	2	Arleta	Fast Food Restaurant	Mexican Restaurant	Vietnamese Restaurant	Falafel Restaurant	Fried Chicken Joint	Middle Eastern Restaurant	Noodle House	Pizza Place	Italian Restaurant	34.241327	-118.432205
2	0	Arlington Heights	Fast Food Restaurant	Fried Chicken Joint	Vegetarian / Vegan Restaurant	Brewery	Shopping Mall	Japanese Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant	34.043494	-118.321374
3	1	Atwater Village	Fast Food Restaurant	Fried Chicken Joint	Vegetarian / Vegan Restaurant	Greek Restaurant	Taco Place	Sushi Restaurant	Brewery	Shopping Mall	Vietnamese Restaurant	34.118698	-118.262392
4	2	Bel Air	Fast Food Restaurant	Udon Restaurant	Japanese Restaurant	American Restaurant	Greek Restaurant	Fried Chicken Joint	Mediterranean Restaurant	Chinese Restaurant	Mexican Restaurant	34.098883	-118.459881

Visualizing clusters of similar food categories of LA Neighborhoods

Click here for Github link