# OPENING A NEW SHOPPING COMPLEX

Location – San Diego, California

# Problem Statement

- One of the major projects that have acquired attention in the past decade is shopping complex. This project is capital intensive and requires a lot planning.
- San Diego is one of those cities where the growth of commercial places has peaked in past few years and there is a lot scope of these projects.
- There are a lot of investors and builders looking to collaborate and build these complexes but the problem is that they want to find the best location where they can setup their project.
- There are different kind of investors some wants to take risk and setup their projects in areas which are already quite competitive on the other hand some want to setup in areas which are growing and might become a commercial hub in future.

# Neighbourhoods Data

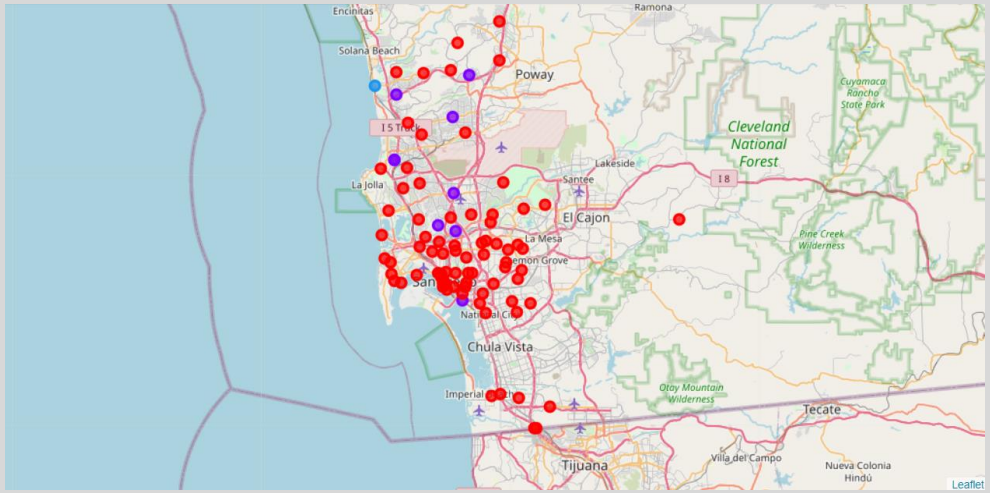
- The whole data collection is done is three broad steps as defined below in detail.
- Web Scraping – Collecting data about the neighbourhoods in San Diego is done with help of the python packages and converted into a pandas dataframe.
- Geological Co-ordinates – The latitude and longitude details of every neighbourhood is acquired with the help of the geocoder package In python.
- Foursquare API – Each neighbourhood in our dataset must have some venues which the people visit a lot for example a mall, cinema, shops or even a park. With the help of the foursquare API we can get the top venues associated with each neighbourhood which will form a major part of our analysis.

# Methodology

- The main element is the data collection. The neighborhood's data is freely available on the Wikipedia page (https://en.wikipedia.org/wiki/List_of_communities_and_neighborhoods_of_San _Diego). This is the list of locations and now we use the geocoder package in-order to find out the latitude and longitude of these places.
- Now we will use the foursquare API to obtain the top venues near a place in our dataset. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude.
- The final stage is the unsupervised learning in which we first identify the cluster where there is high commercial activities and then we find the cluster where the shopping malls are situated, and which places can be new hotspots for shopping complex. We will use the K-means clustering algorithm and test it out on various values of K.

# Results

- The result from the k-means clustering first recognizes the places where there are any commercial activities. The second cluster recognizes the density of the shopping complex.
- • Cluster 0: Neighborhoods with moderate number of shopping malls
- • Cluster 1: Neighborhoods with low number to no existence of shopping malls
- • Cluster 2: Neighborhoods with high concentration of shopping malls
- The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, and cluster 2 in blue color.



# Predictions

- Shopping Label 0 - indicates that the area has very less proportion of shopping complex and has other features.
- Shopping Label 1 - indicates that the area has moderate amount of shopping complex and that they are flourishing
- Shopping Label 2 - indicates that the area has good shopping complex and it is known for them. This tells us the areas which are commercial and where you might find a place for your business. The labels have a deep meaning which depends upon the consumer. So, the consumer might be afraid of competition or he may find some cheap land in cluster 0 of shopping that is where there are low number of complex. This decision depends upon the mindset you have and the assets. An apt choice would be built in an area where there is some shopping complex which are there to tell us that the area recognizes them and that they are flourishing. The third kind is for people who are more confident or risk takers because they already know that the
- cluster 2 is famous for shopping and if we are able to build something extraordinary then we might end up stealing away a lot of consumers which have been built by the other complex. This area is sensitive to competition and ups and down.