Opening a Shopping Mall in Los Angeles

Author: Shiyun Lu

Business Problem and Target Audience

There are plenty of malls in Los Angeles. However, there will never be enough of them. If someone with a lot of money wants to open a new shopping mall in L.A., data science should come in to help.

Let's assume we are helping some major stakeholders who want to pick an address in L.A. to oopen up a new shopping mall. We are trying to find out locations which are the most profitable. We will use methodologies like data science and machine learning to solve this problem based on some models we can build (e.g. clustering models).

So what is an ideal location for Los Angeles to open a new shopping mall? This will be the problem we stick to in the following process of exploring.

This report will be helpful for stakeholders who are planning to invest in opening a new shopping mall in the greater Los Angeles area in a way to assist to determine the location of the mall.

Data

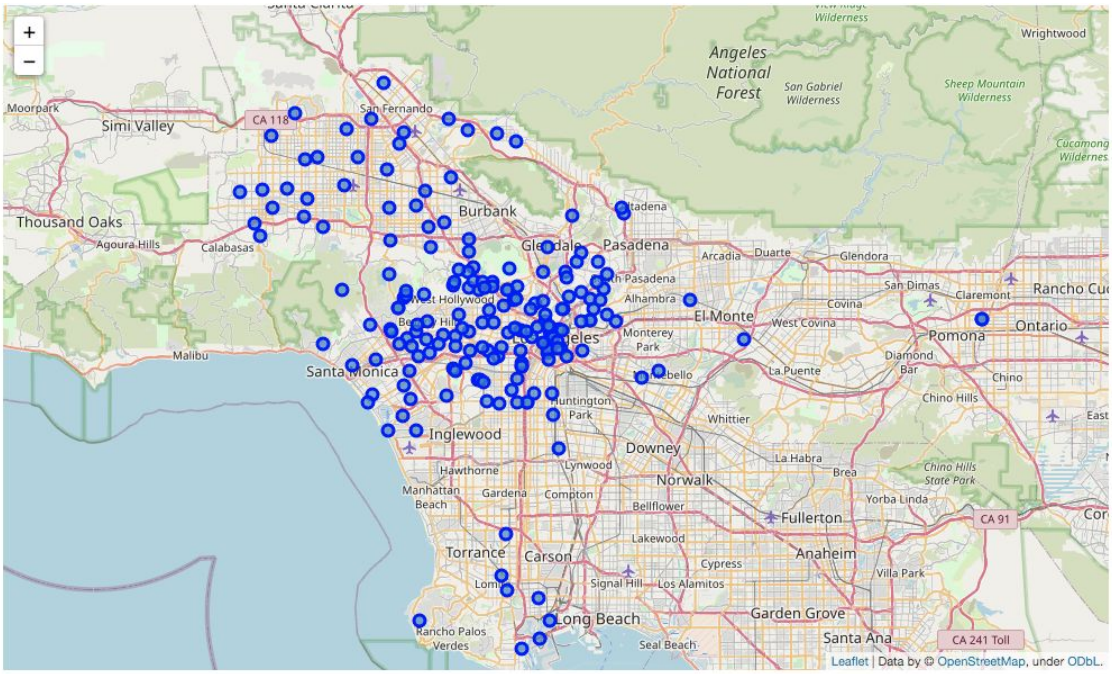
We need below data for this project to be done:

1. Neighbourhood information of the greater Los Angeles area.
2. 2.Geographical information (latitude and longitude) of the greater Los Angeles area.
3. 3.Venue data of the greater Los Angeles area.

The Los Angeles government offered a list of zip codes of L.A.: (http://file.lacounty.gov/SDSInter/lac/1031552_MasterZipCodes.pdf) from where we are able to extract L.A. neighbourhood information. For Geographical information (latitude and longitude) we can get it from Python package and for venue data we are able to get it from Foursquare API according to the methods

Methodology

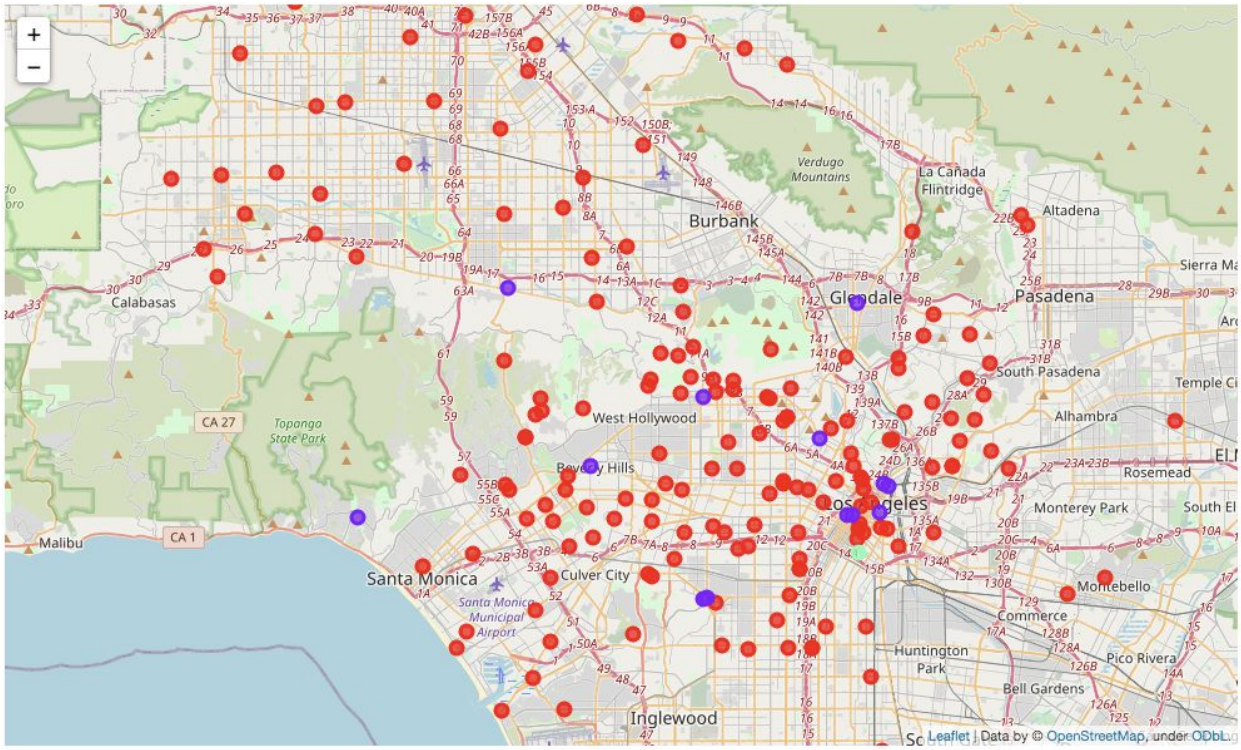
First we use Folium to get the map of Los Angeles, then use a Json file to get coordinates for neighbourhoods of Los Angeles. And below map figure can be shown:



This figure has displayed all neighbourhoods we extracted from wikipedia geographically on the L.A. map. Then using the Foursquare API we can get venue information of L.A. area like below:

	Neighborhoods	ATM	Acai House	Accessories Store	Adult Boutique	Alternative Healer	American Restaurant	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	A Restau
0	Angelino Heights	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Angelino Heights	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Angelino Heights	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Angelino Heights	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Angelino Heights	0	0	0	0	0	0	0	0	0	0	0	0	0	

After having all the venue information, we are ready to do clustering of the L.A. areas. We set cluster number as 3 and ran K-means clustering so we are able to get below results (shown on map figure):



Result and Conclusion

I have clustered Los Angeles into clusters using and the result shows cluster 2 has higher density of shopping malls while cluster 1 has lower one. Thus I might advise stakeholders to open a new shopping mall in cluster 1 area to get potentially high customer inflow.

This project has made use of Folium, sklearn and other python libraries to visually determine the best location of a new shopping mall in the L.A. area. Also we used Foursquare API and other public information extracted from the internet to complete the data collection process.We are able to find areas with lower density of malls and give advice to stakeholders regarding the business problem mentioned above.