Final Project (CoWorking Space in Los Angeles, CA)

The Battle of the Neighborhoods - Week 2

By Oleksiy Lomakin

Table of contents

1. Business Problem :

Looking for a place for a new COWORKING space in Los Angeles County, CA

The coworking industry is becoming a lucrative worldwide phenomenon, so now is the perfect time to break into the business. There needs to be a niche for a new collaborative workspace business, and thankfully there are plenty. Perhaps the area has a high demand for coworking spaces beyond coffee shops and libraries. Maybe the existing spaces in the area are too focused on one type of worker. On the other hand, maybe the other spaces are too broad.

We need to envision our target market's coworking environment. Should we consider a space in a premium facility for companies or an affordable solution for freelancers? Is our primary audience big business or small teams? Will our future facility integrate with the city's business environment?

2. Description of the data:

1. Los Angeles Times Mapping Los Angeles County Neighborhoods
2. Foursqure City Guide
3. Census Tract Locations (LA)
4. Latlong.net

2.1 Los Angeles Times Mapping Los Angeles County Neighborhoods

This regional view is your portal to indvidual maps and statistics for 158 cities and unincorporated places and 114 neighborhoods within the city of Los Angeles. In addition, there are maps of 42 unincorporated areas that we have collapsed inside of adjacent cities. With this source we can analyze LA County Neighborhoods depending on their locations and population features such as average income, education, diversity, age e.t.c.

2.2 Foursqure City Guide

We will need to work with Foursquare City Guide in order to get information regarding operating coworking spaces and nearby venues.

2.3 Census Tract Locations (LA)

The Neighborhood Data for Social Change (NDSC) platform is a project of the USC Price Center for Social Innovation. NDSC is a free, publicly available online resource for civic actors to learn about their neighborhoods.The platform helps tell the stories of neighborhoods through maps, charts, data analysis, and storytelling; helping community stakeholders track measurable change, improve local policies and programs, and ultimately advocate for a better quality of life within their communities.

2.4 Lat Long

We will use this saurce in case we need to get latitude and longitude of some geographical objects. Latlong.net is an online geographic tool that can be used to lookup latitude and longitude of a place, and get its coordinates on map. You can search for a place using a city's or town's name, as well as the name of special places, and the correct lat long coordinates will be shown at the bottom of the latitude longitude finder form. At that, the place you found will be displayed with the point marker centered on map. Also the gps coordinates will be displayed below the map.

3. Methodology

In this project we will direct our efforts on detecting neighborhoods of Los Angeles County that have good conditions to open a new Coworking Space, particularly those with high number of currently operating ones.

In first step we have collected the required neighborhood data with features such as locations, population features, average income, diversity, age e.t.c. We will perform exploratory data analysis.

Second step in our analysis will be identifying currently operating Coworking Spaces (according to Foursquare categorization). We will use Folium maps to identify and analyze the Neighborhoods with Coworking Spaces information. We will try to find Correlations between neighborhood features.

In third step we will need data preparation Clustering for futere neighborhood segmentation. After, we will perform clustering using K-Means method with a certain number of clusters. Then, we will analyze the data based on cluster segmentation and compare it with the data of neighborhoods that have the highest number of currently operating Coworking Spaces.

In fourth step, by selecting important and meaningful information of features, we will narrow down the neighborhood list to only promising locations. We will be able evaluate our analysis by comparing promising location list the neighborhood list with currently operating Coworking Spaces.

4. Analysis

Import csv file with basic features constracted in 'Data Collection.ipynb'

	NEIGHBORHOOD	POPULATION PER SQMI	AVERAGE HOUSEHOLD SIZE	DIVERSITY INDEX	FOREIGN BORN POPULATION	MEDIAN INCOME	MEDIAN AGE
0	Koreatown	42611	2.7	0.602	68.0	30558.0	30.0
1	Westlake	38214	3.0	0.430	67.6	26757.0	27.0
2	East Hollywood	31095	3.0	0.578	66.5	29927.0	31.0
3	Pico-Union	25352	3.3	0.264	64.6	26424.0	27.0
4	Maywood	23638	4.1	0.069	55.2	41203.0	23.0

4.3 Exploring the data

	POPULATION PER SQMI	AVERAGE HOUSEHOLD SIZE	DIVERSITY INDEX	FOREIGN BORN POPULATION	MEDIAN INCOME	MEDIAN AGE
count	265.000000	265.000000	265.000000	265.000000	265.000000	264.000000
mean	7571.471698	2.972075	0.492234	31.817358	68647.022642	33.178030
std	6925.987567	0.643715	0.160948	14.401015	31740.466505	5.942255
min	2.000000	1.500000	0.065000	4.400000	15003.000000	21.000000
25%	1601.000000	2.500000	0.380000	20.300000	48518.000000	28.000000
50%	6459.000000	2.900000	0.526000	31.800000	63039.000000	33.500000
75%	11266.000000	3.400000	0.628000	42.400000	81279.000000	37.000000
max	42611.000000	4.600000	0.755000	72.400000	207938.000000	48.000000

With info function we can see the types of values and NaNs

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 265 entries, 0 to 264
Data columns (total 7 columns):
NEIGHBORHOOD              265 non-null object
POPULATION PER SQMI      265 non-null int64
AVERAGE HOUSEHOLD SIZE   265 non-null float64
DIVERSITY INDEX          265 non-null float64
FOREIGN BORN POPULATION  265 non-null float64
MEDIAN INCOME            265 non-null float64
MEDIAN AGE               264 non-null float64
dtypes: float64(5), int64(1), object(1)
memory usage: 14.6+ KB
```

4.4 Searching Coworking Spaces in LA County with Foursquare

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighbo
0	FlipWork CoWorking	Fair	448 S Hill St	US	Los Angeles	United States	5th St	781	[448 S Hill St (5th St), Los Angeles, CA 90013...	[{'label': 'display', 'lat': 34.049073, 'lng':...	34.049073	-118.251249	
1	Kleverdog Coworking	Coworking Space	418 Bamboo Ln Ste A	US	Los Angeles	United States	between Hill St and Broadway Ave	1640	[418 Bamboo Ln Ste A (between Hill St and Broa...	[{'label': 'display', 'lat': 34.06600345686554...	34.066003	-118.237353	
2	Beach House Coworking	Coworking Space	2219 Main St	US	Santa Monica	United States	NaN	23001	[2219 Main St, Santa Monica, CA 90405, United ...	[{'label': 'display', 'lat': 34.005062, 'lng':...	34.005062	-118.486420	Ocea
3	Coworking Plus	Coworking Space	2293 W 190th St	US	Torrance	United States	Western	22661	[2293 W 190th St (Western), Torrance, CA 90504...	[{'label': 'display', 'lat': 33.85869110821988...	33.858691	-118.319778	
4	Ofis Coworking	Coworking Space	NaN	US	Los Angeles	United States	NaN	18130	[Los Angeles, CA, United States]	[{'label': 'display', 'lat': 34.002665, 'lng':...	34.002665	-118.430875	
5	LAUNCH Coworking	Coworking Space	3525 Eastham Dr	US	Culver City	United States	NaN	12719	[3525 Eastham Dr, Culver City, CA 90232, Unite...	[{'label': 'display', 'lat': 34.025877, 'lng':...	34.025877	-118.377853	

Plotting Coworking Space locations on a map to visualize for better understanding.
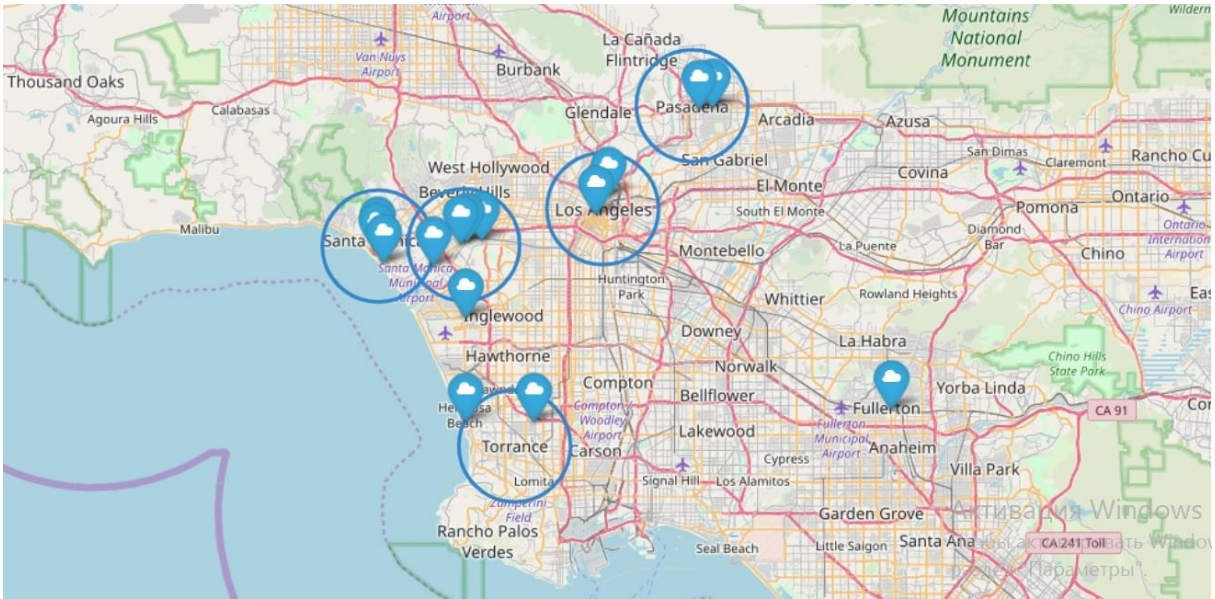
Combining this information to a dataframe including columns: NEIGHBORHOOD and Number Of Coworking places

## 4.5 Analyzing Neighborhoods with Coworking Spaces information

Let's go back to our exploratory data analysis with the information. We need to add a new column to our dataset with the number of Coworking current operating locations in each neighborhood. We will set other values as zero.

4.6 Correlations
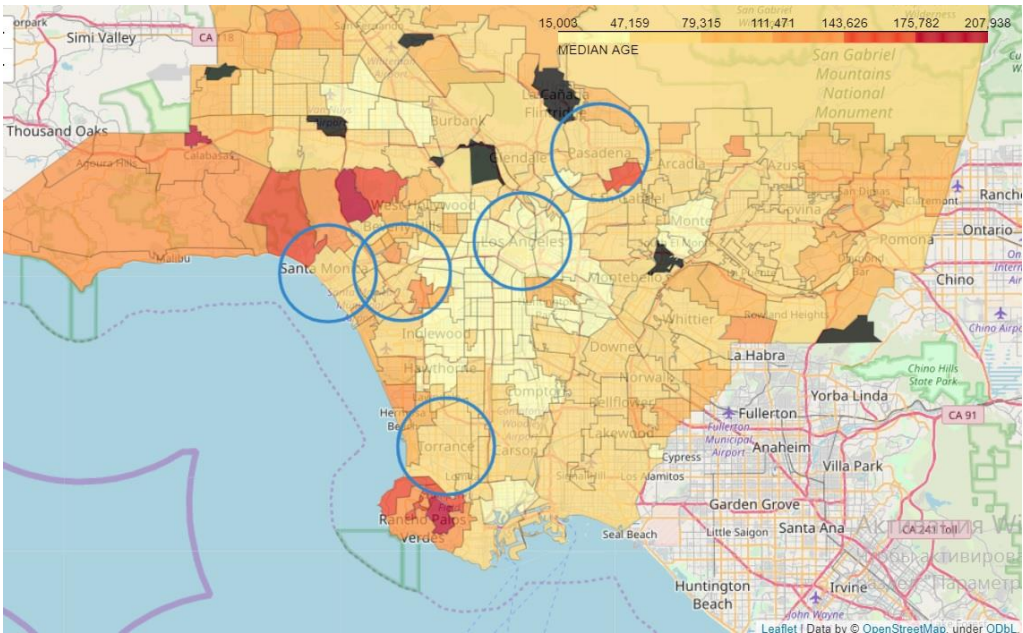
Performing a dataframe with Features Correlations

	Unnamed: 0	POPULATION PER SQMI	AVERAGE HOUSEHOLD SIZE	DIVERSITY INDEX	FOREIGN BORN POPULATION	MEDIAN INCOME	MEDIAN AGE	CW_N
Unnamed: 0	1.000000	-0.941971	0.228393	0.532824	-0.188778	-0.043282	-0.074942	-0.039431
POPULATION PER SQMI	-0.941971	1.000000	-0.011151	-0.590618	0.060270	0.266449	-0.083081	0.130123
AVERAGE HOUSEHOLD SIZE	0.228393	-0.011151	1.000000	-0.048903	0.243529	0.097772	-0.395829	-0.233369
DIVERSITY INDEX	0.532824	-0.590618	-0.048903	1.000000	0.181470	-0.454514	0.430113	0.124096
FOREIGN BORN POPULATION	-0.188778	0.060270	0.243529	0.181470	1.000000	-0.874109	0.435546	-0.127935
MEDIAN INCOME	-0.043282	0.266449	0.097772	-0.454514	-0.874109	1.000000	-0.585762	0.067016
MEDIAN AGE	-0.074942	-0.083081	-0.395829	0.430113	0.435546	-0.585762	1.000000	0.412021
CW_N	-0.039431	0.130123	-0.233369	0.124096	-0.127935	0.067016	0.412021	1.000000

Features distribution visualization
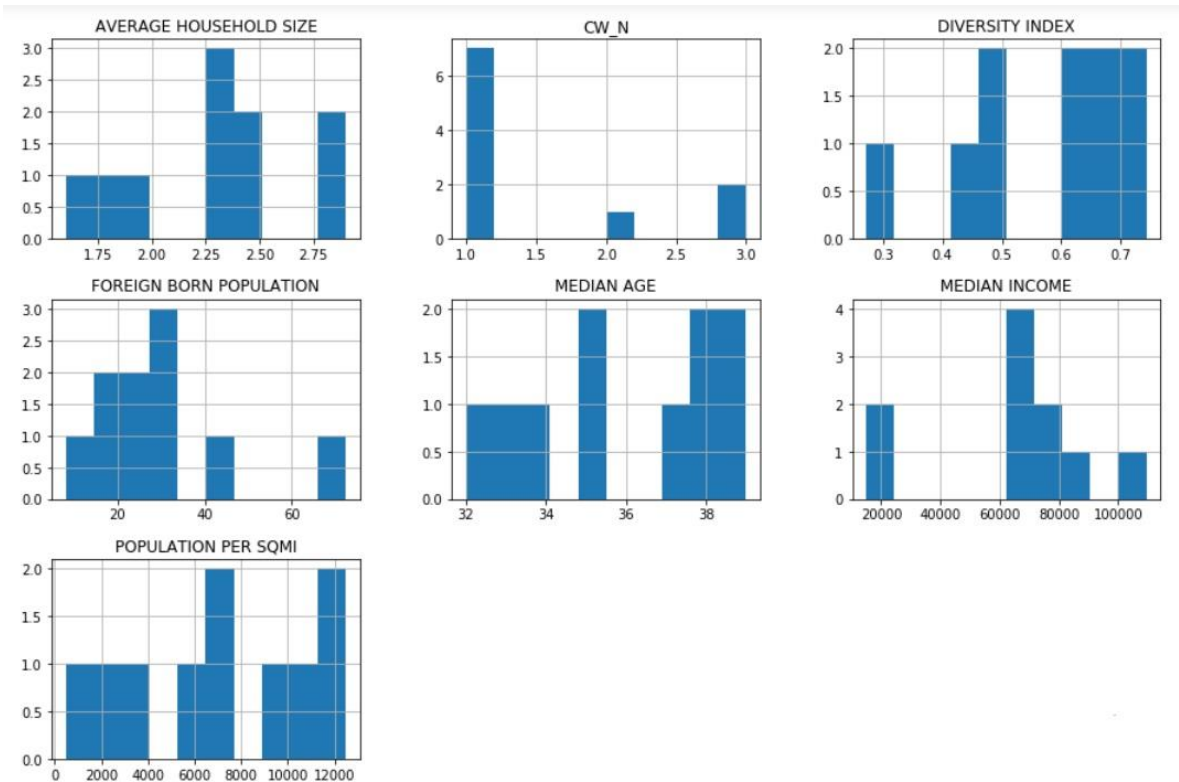
4.7 Visualizing neighborhoods based on median income

Using the url in order to to get GEO information about the neighborhoods in LA County
Generating a choropleth map using the NEIGHBORHOOD and MEDIAN INCOME information.

4.8 Clustering data preparation

What are the similar neighborhoods?
Also, we add a column with numerical value with the number of Coworking places

	NN	MEDIAN INCOME	MEDIAN AGE	POPULATION PER SQMI	AVERAGE HOUSEHOLD SIZE	DIVERSITY INDEX	FOREIGN BORN POPULATION
0	0	83983.0	37.0	166.0	3.0	0.316	7.1
1	1	29606.0	26.0	21848.0	3.2	0.545	46.7
2	2	117608.0	37.0	2495.0	3.0	0.304	13.5
3	3	106078.0	40.0	99.0	2.9	0.282	4.4
4	4	53224.0	35.0	11275.0	2.8	0.631	50.8

Preprocessing scaling. Checking how the dataframe looks

	NN	MEDIAN INCOME	MEDIAN AGE	POPULATION PER SQMI	AVERAGE HOUSEHOLD SIZE	DIVERSITY INDEX	FOREIGN BORN POPULATION
0	0	0.479693	0.644407	-1.067935	0.049670	-1.095225	-1.715863
1	1	-1.239629	-1.210258	2.074514	0.361884	0.327607	1.034063
2	2	1.542867	0.644407	-0.730385	0.049670	-1.169784	-1.271431

4.8 Clustering data preparation

Let's perform clustering using K-Means method with number of clusters 3. Also, we will need to add a column with clusters labels to our dataset
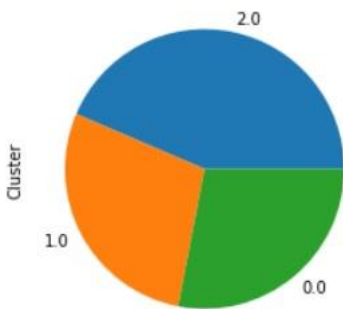
Visualize clustering decomposition
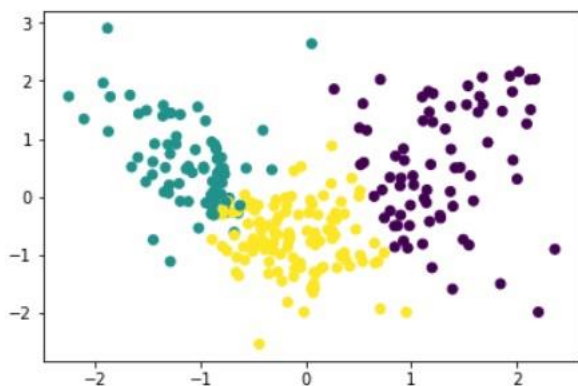
4.9 Analyzing clusters

Clusters median information

| Cluster	NN	MEDIAN INCOME	MEDIAN AGE	POPULATION PER SQMI	AVERAGE HOUSEHOLD SIZE	DIVERSITY INDEX	FOREIGN BORN POPULATION
0	131.229730	43446.621622	26.283784	14130.378378	3.674324	0.393297	45.368919
1	131.173333	102893.200000	39.000000	2921.200000	2.500000	0.398000	17.333333
2	132.069565	62906.608696	33.817391	6298.747826	2.819130	0.617443	32.524348

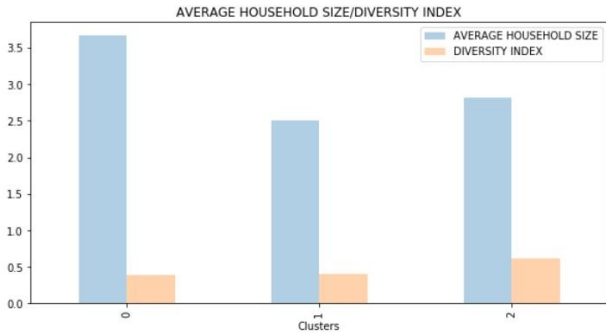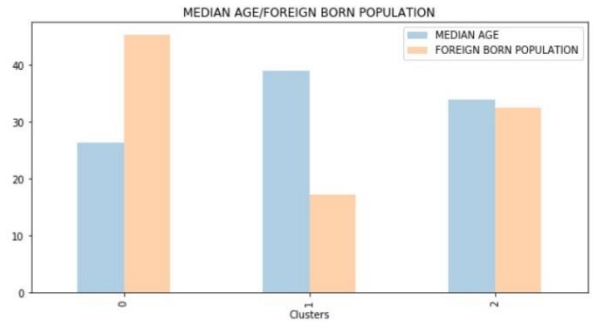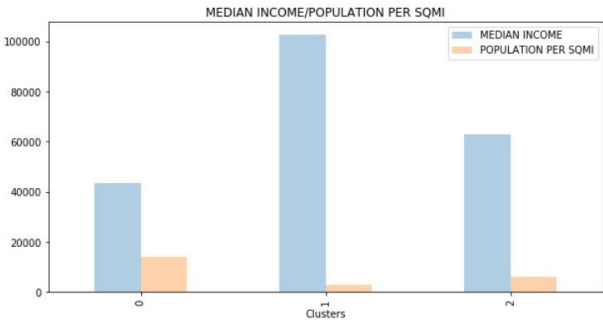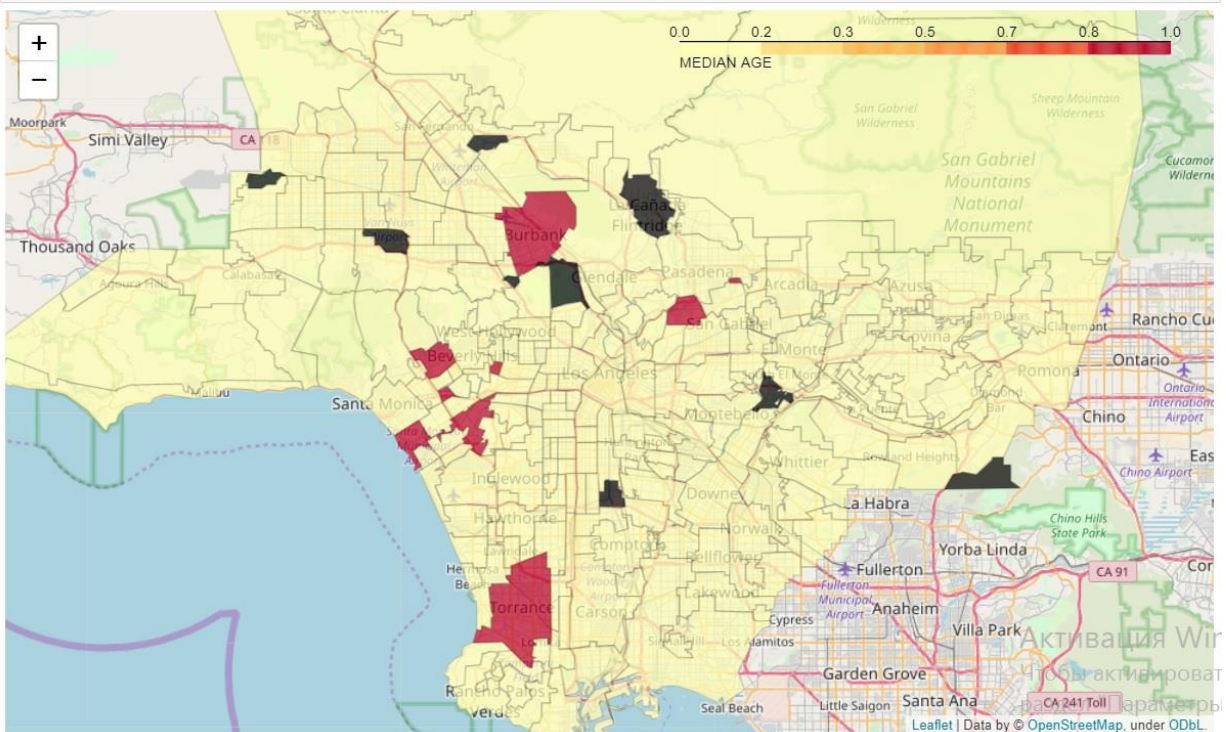Visualizing clusters with pie plot.   Coworking Locations in Neighborhood Clusters
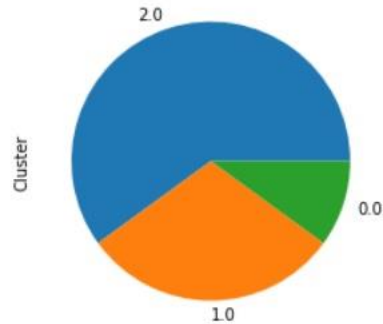
Plotting all clusters on the map
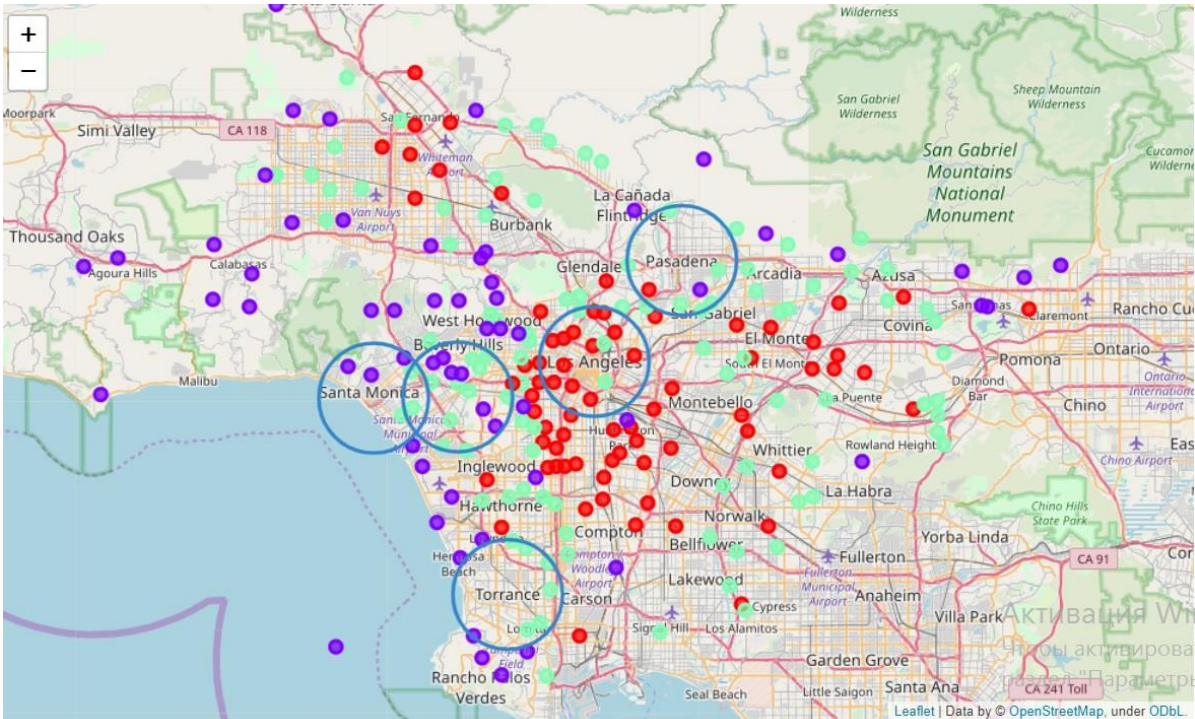
Features distribution by clusters

Plotting theNeighborhoods selected by Clusters and key features

5. Results

Here is the list of promising locations in Los Angeles County for a new Coworking Space. Some of them already have currently operating collaborative work space businesses, some of them not. The similarity of these neighborhoods features is high. The list has been narrowed down to top 10 promising locations.

	NEIGHBORHOOD	POPULATION PER SQMI	AVERAGE HOUSEHOLD SIZE	DIVERSITY INDEX	FOREIGN BORN POPULATION
0	Westwood	13036.0	2.0	0.543	31.3
1	Venice	11891.0	1.9	0.534	22.3
2	Carthay	9642.0	2.1	0.615	25.1
3	San Pasqual	8036.0	2.3	0.601	32.6
4	Culver City	7475.0	2.3	0.681	26.6
5	Rancho Park	7169.0	2.2	0.598	28.5
6	South Pasadena	7114.0	2.3	0.651	24.4
7	Torrance	6701.0	2.5	0.628	27.6
8	Burbank	5785.0	2.4	0.581	31.1

The main neighborhoods features are: High Median Income, Average household size, Foreign Born Population, Diversity, and Density. The Age Median value and Population Density will require extra attention.

6. Discussion

Our analysis shows that some features of Los Angeles County Neighborhoods are important and meaningful and some can cause a false positive correlations. We discovered that High Median Income neighborhood feature should be one of the first to be considered. According to over observations the value should be higher than average or in our case more than 62000 as in cluster "1". Average household size correlates very well Median Income and the suggested value should be less than 2.6 as in Pasadena. Population Diversity Index value was almost same in some cluster segments, but higher in all Neighborhoods with the big numbers of Coworking places. The value should be set to at least higher than average. There was correlation in Foreign Born Population feature and the value may be set as low. The Population per SqMi feature may cause false positive conclusions. The best suggestion would to narrow down the neighborhood list by Median Income first and eliminate the neighborhoods with low median income that may have a high population density. Then from the narrowed data we can drop wealthy suburbs that may have a low population density. It would be a good idea to compare this information with the Average household size. The median Age feature can be set higher in some cases. However, it requires a specific analysis for university campuses with the high number of foreign born people, high density and low median age.

Also, we can consider adding some information and data to our analysis such as residential and commercial real estate rates, office space availability, libraries and coffee shops nearby.

7. Conclusion

The main purpose of this project was to identify the best neighborhoods to open a new Coworking Space. We have created a list of the most promising locations for this business by collecting the data from multiple sources, providing exploratory data analysis, data cleaning and preparation, performed clustering segmentation using K-Means method, analyzed important neighborhood features and provided suggested values for them. There is a room for a deeper analysis using other methods and adding new information data.

Final decision on optimal Coworking Space location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone. Taking into consideration additional factors.