Searching where live in Lima

Capstone project by Diana Quintanilla

Introduction

For the last few years I have been living in the city of Lima. I want find a apartment according to my preferences about the neighborhood. For example, one might want to look at all areas which has proximity to pubs, cafes, public transport etc.

¿Where in Lima city will I buy or alquile a apartment?

To figure this out I want to compare the various neighbourhoods in Lima to see how similar they are, and finally figure out if there is a neighbourhood with my preferences.

Data

I used two datasets for the above problem:

- A list of neighbourhoods in Lima at [http://www.movistar.com.pe/documents/10182/24285/3G.pdf/4a97babb-9b45-4c39-bcc4-d4acf6155ca8]. This includes
    i. Name of neighbourhood
    ii. Name of district
    iii. their geographic coordinates
- The second dataset I used is Foursquare venues data with categories, this will be used in collaboration with the above dataset.

Methodology - Data clean

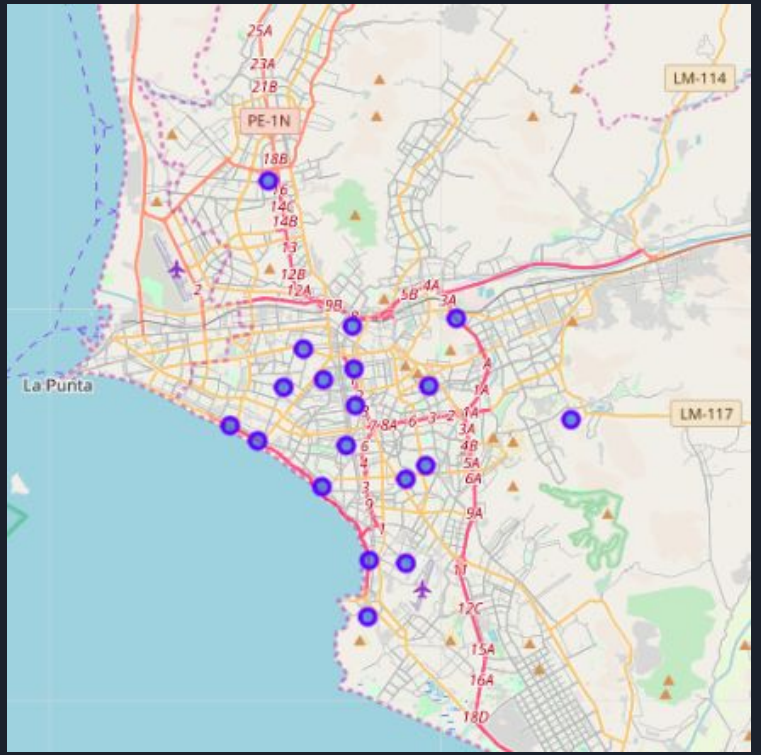
Get this dataset:

	Código UBIGEO	Centro poblado	Distrito	Provincia	Región	LATITUD	LONGITUD	CLASIFICACIÓN INEI
717	1501010001	LIMA	LIMA	LIMA	LIMA	-12.046679	-77.03230	URBANO
719	1501040001	BARRANCO	BARRANCO	LIMA	LIMA	-12.149599	-77.02474	URBANO
720	1501050001	BREÑA	BREÑA	LIMA	LIMA	-12.056910	-77.05366	URBANO
737	1501080001	CHORRILLOS	CHORRILLOS	LIMA	LIMA	-12.174429	-77.02482	URBANO
741	1501130001	JESUS MARIA	JESUS MARIA	LIMA	LIMA	-12.069999	-77.04524	URBANO

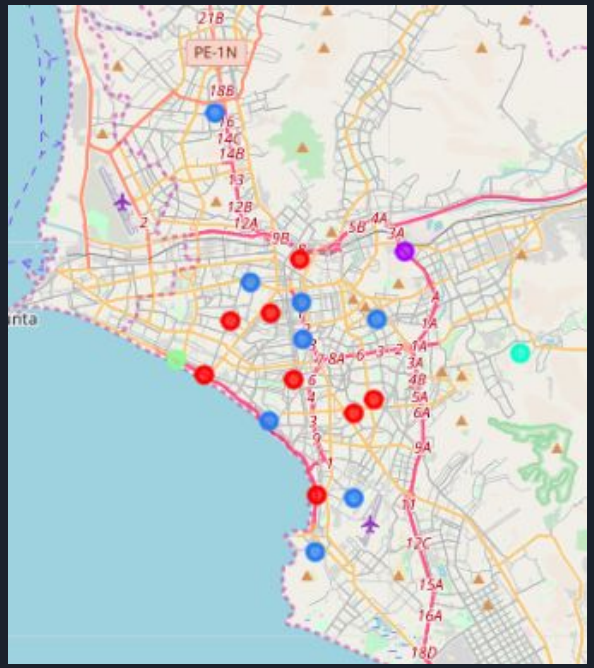
Methodology - Data Exploration

Neighborhoods in Lima:

rhood	Zone	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
LIMA	URBANO	-12.046679	-77.0323	Plaza Mayor de Lima	-12.045983	-77.030565	Plaza
LIMA	URBANO	-12.046679	-77.0323	Casa Bernardo O'Higgins	-12.047588	-77.032498	Art Gallery
LIMA	URBANO	-12.046679	-77.0323	Palacio Municipal de Lima	-12.045283	-77.030917	City Hall
LIMA	URBANO	-12.046679	-77.0323	Teatro Municipal de Lima	-12.045770	-77.034839	Theater
LIMA	URBANO	-12.046679	-77.0323	Galería Municipal Pancho Fierro	-12.045495	-77.031341	Art Gallery



Methodology - Clustering



I use K means Clustering for the problem because after the exploratory analysis it was obvious that we can categorise neighbourhoods into different homogenous clusters based on the venue categories.

Results

I obtained 5 clusters from K means clustering executed in the previous step.

Cluster 1. Contains many restaurants and diners, as well as some other services, like art gallery

Cluster 2. Seems to be dominated by coffee shop and mallS

Cluster 3. Is mostly linked to open areas for bar,parks and music

Cluster 4. Is the most homogeneous one (it only contains one neighborhood)

Cluster 5. Is characterized by shops, gym and restaurants

Conclusion

We can effectively help in the apartment search process for our customers by clustering the area based on the venue categories and filtering the cluster to narrow down their search area. Customers can then prefer to focus their apartment search by analysing various trade offs. Those who are confused can efficiently compare various areas and choose the one which is most suitable for them.