Coursera

IBM Data Science
Certificado profesional

Data Science Capstone Final:

"Analysis and location of Technical Service venues for computers and laptops in Lima, Peru"

By Rubén García

Lima, Perú, Jun 2020

Index

1. Introduction

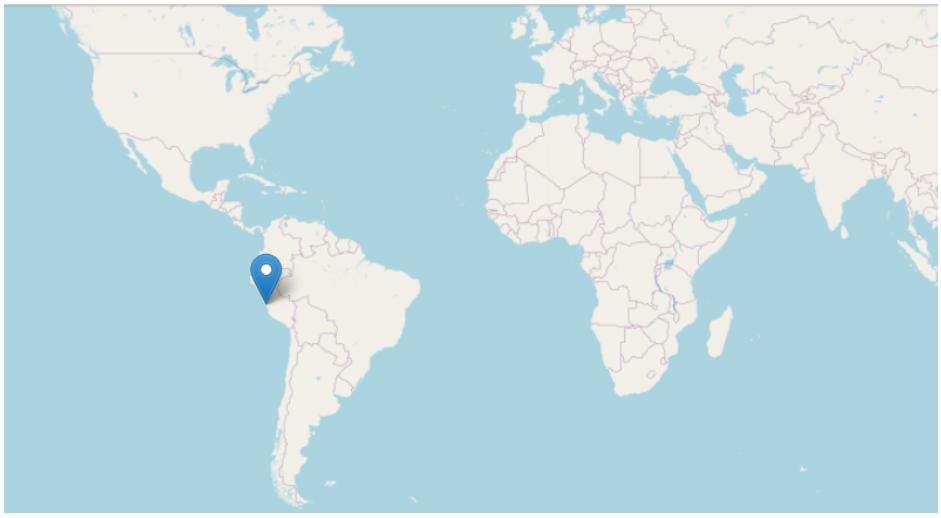
Lima city, where I live, is the capital and the most populous city in the Republic of Peru. It is located on the central coast of the country, on the shores of the Pacific Ocean, with approximately 11 million inhabitants representing 32% of the Peruvian population. It was founded on January 18, 1535 with the name of the City of Kings by the Spanish conquistador Francisco Pizarro. Both the city of Lima and those of the interior of the country have a gastronomic, cultural, mining and cultivation wealth.

Metropolitan Lima is divided into 50 districts (Boroughs) in total, where the province of Lima has 43 districts and the constitutional province of Callao with 7 districts. It is also organized in cones or axes of the city: North Lima, Downtown Lima, Modern Lima, East Lima, South Lima, Callao and health resort..

Currently in the world, the Covid-19 pandemic has brought damage in health, job dismissal and the economy of many countries, confining us into our houses in social isolation; in this way, families have changed: their consumption habits, remote work, online classes for children, deliverys services, use of streaming platforms (for example Netflix, Youtube, Spotify, etc.), learning platforms and others. All of this make us purchase, repair computers, laptops, tablets, printers and mobile phones this way.

This is the reason why we could take advantage of the location data of “Foursquare API” which would help families to solve this problem of having a technical service nearby and, at the same time, the opportunity to invest in a similar business where this service does not exist, especially in the districts with higher economic incomes.

1.1 Ubication of Lima, Perú



2. Data Description

To consider the problem I can list the datas as below:

* INEI: National Institute of Statistics and Informatics, which is an autonomous constitutional body of Peru. He is also in charge of population, housing, business, agrarian, university, etc. censuses.

(https://www.inei.gob.pe/)

* CPI: Peruvian company dedicated to research, market research and public opinion.

(http://cpi.pe/)

* APEIM: Peruvian Association of Market Intelligence Companies is a non-profit association that groups market research and public opinion companies with the purpose of promoting, developing, protecting and defending market research activity in Peru.

(http://apeim.com.pe/)

* Forsquare API: To get the closest places to locate technical service for computers, laptops, printers and accessories.

(https://foursquare.com/)

* Geodata: It allows obtaining the latitude and longitude of a point in the world.

(https://www.geodatos.net/)

3. Methodology

For solving the problem it will be usage the data sources indicated in the previos point. Data wrangling will include pre-processing, clean, transformation and data formatting.

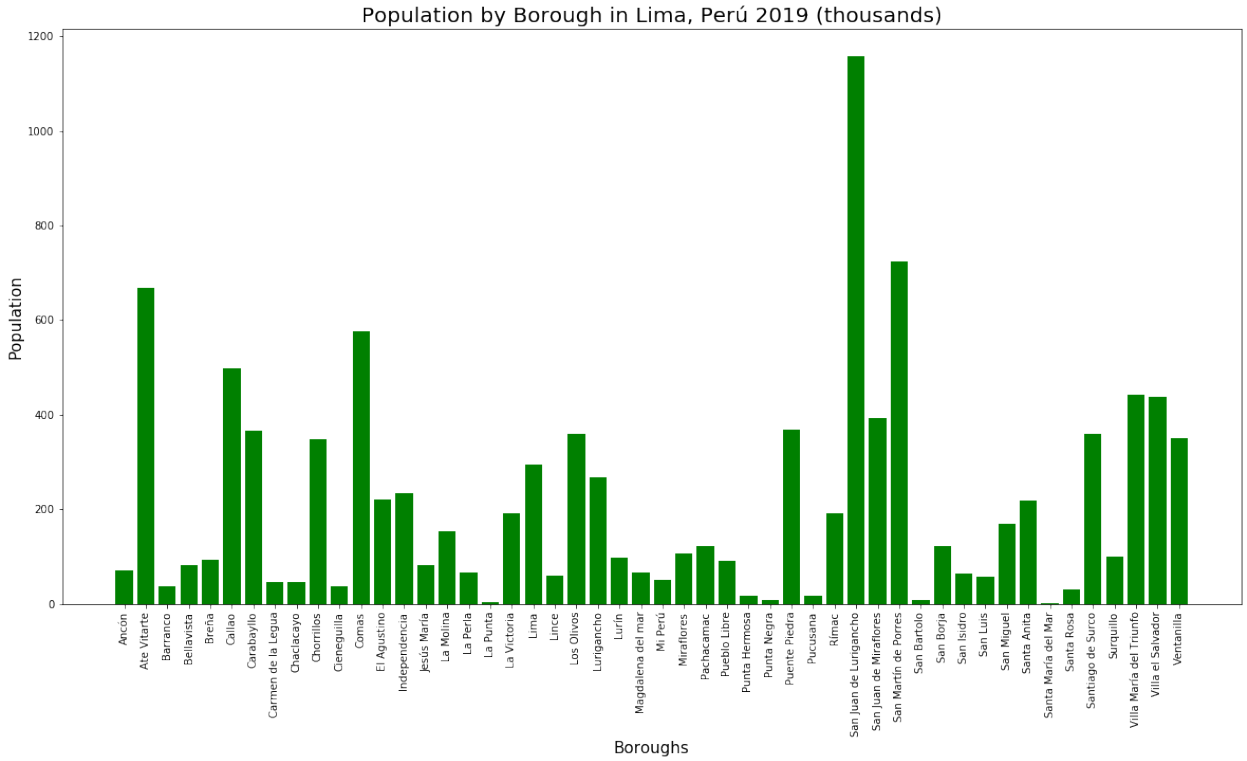
The master data of Lima city is divided into 50 districts (Boroughs). It is also organized in cones or axes of the city: North Lima, Downtown Lima, Modern Lima, East Lima, South Lima, Callao and health resort. The master data has the folowing columns:

Borough	Axes	Latitude	Longitude	Population
Ancón	Balnearios	-11.7736	-77.1761	70.1
Ate Vitarte	Lima Este	-12.0261	-76.9192	667.2
Barranco	Lima Moderna	-12.1492	-77.0217	37.5
Bellavista	Callao	-12.0601	-77.1116	81.7
Breña	Lima Centro	-12.0569	-77.0536	93.4

3.1. Show the population by Borough

In the next graph, I show the population by Borough in Lima City. The column Population is expressed in thousand:

Borough	Population
Ancón	70.1
Ate Vitarte	667.2
Barranco	37.5
Bellavista	81.7
Breña	93.4

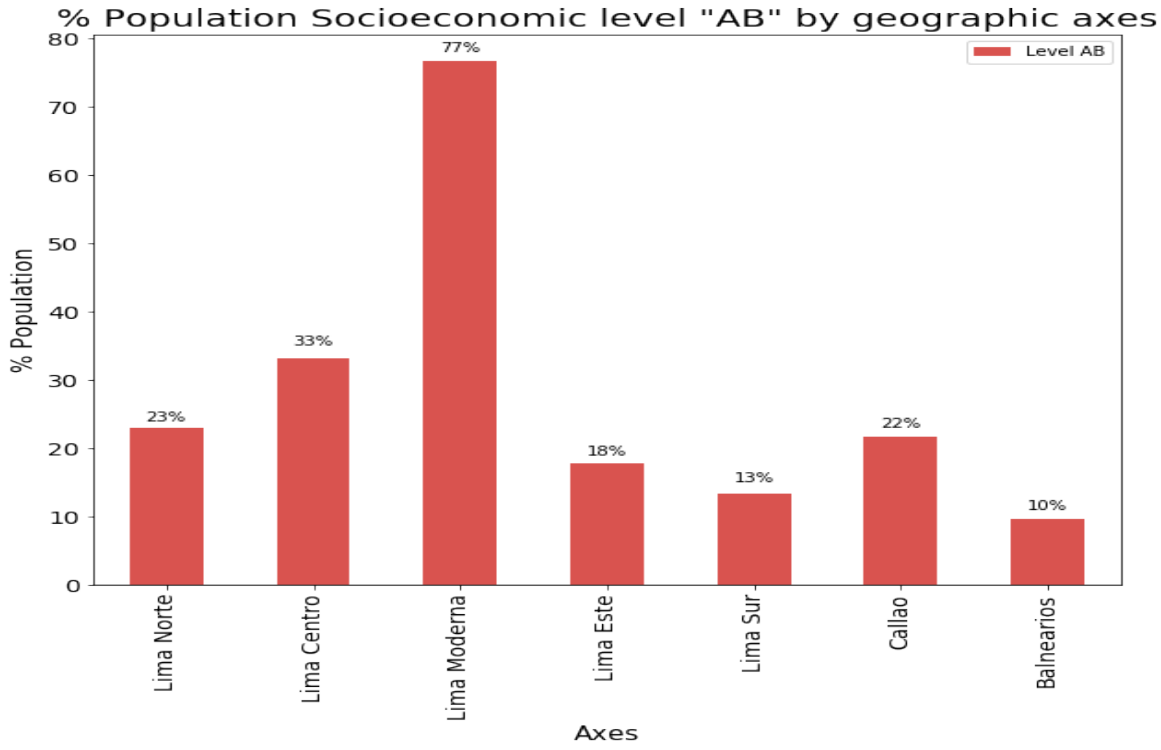


3.2. Borughs grouped by Socioeconomic level

In this case the next master data show the Socioeconomic level of the population by geographic axes, the most high level is AB, C, D, E respectively:

Axes	Population	Level AB	Level C	Level D	Level E
Lima Norte	2627.6	22.9	44.1	27.6	5.4
Lima Centro	828.4	33.1	43.3	20.2	3.5
Lima Moderna	1416.0	76.8	17.4	4.5	1.3
Lima Este	2616.4	17.7	45.7	29.6	7.0
Lima Sur	1839.8	13.3	53.4	27.4	5.9
Callao	1100.4	21.7	45.9	23.6	8.8
Balnearios	152.4	9.7	39.9	37.7	12.7

[8]:



For our analysis we will focus on the highest socio-economic level "AB", its "Lima Moderna" axe:

Axes	Population	Level AB	Level C	Level D	Level E
Lima Moderna	1416.0	76.8	17.4	4.5	1.3

This axe involved to next Boroughs:

Borough	Axes	Latitude	Longitude
Barranco	Lima Moderna	-12.1492	-77.0217
Jesús María	Lima Moderna	-12.0700	-77.0453
La Molina	Lima Moderna	-12.0789	-76.9169
Lince	Lima Moderna	-12.0833	-77.0317
Magdalena del mar	Lima Moderna	-12.1002	-77.0647
Miraflores	Lima Moderna	-12.1207	-77.0299
Pueblo Libre	Lima Moderna	-12.0753	-77.0659

Borough	Axes	Latitude	Longitude
San Borja	Lima Moderna	-12.1072	-76.9992
San Isidro	Lima Moderna	-12.0991	-77.0375
San Miguel	Lima Moderna	-12.0763	-77.0909
Santiago de Surco	Lima Moderna	-12.1464	-77.0067
Surquillo	Lima Moderna	-12.1186	-77.0217

4. Results

I utilized the “Foursquare API” to explore the boroughs and “Technical Services” (reparation and sale computers, laptops and accessories) venues. I aplied, as limit, 100 venues and the radius 1500 meter for each borough from their given latitude and longitude informations, indicated in the last table.

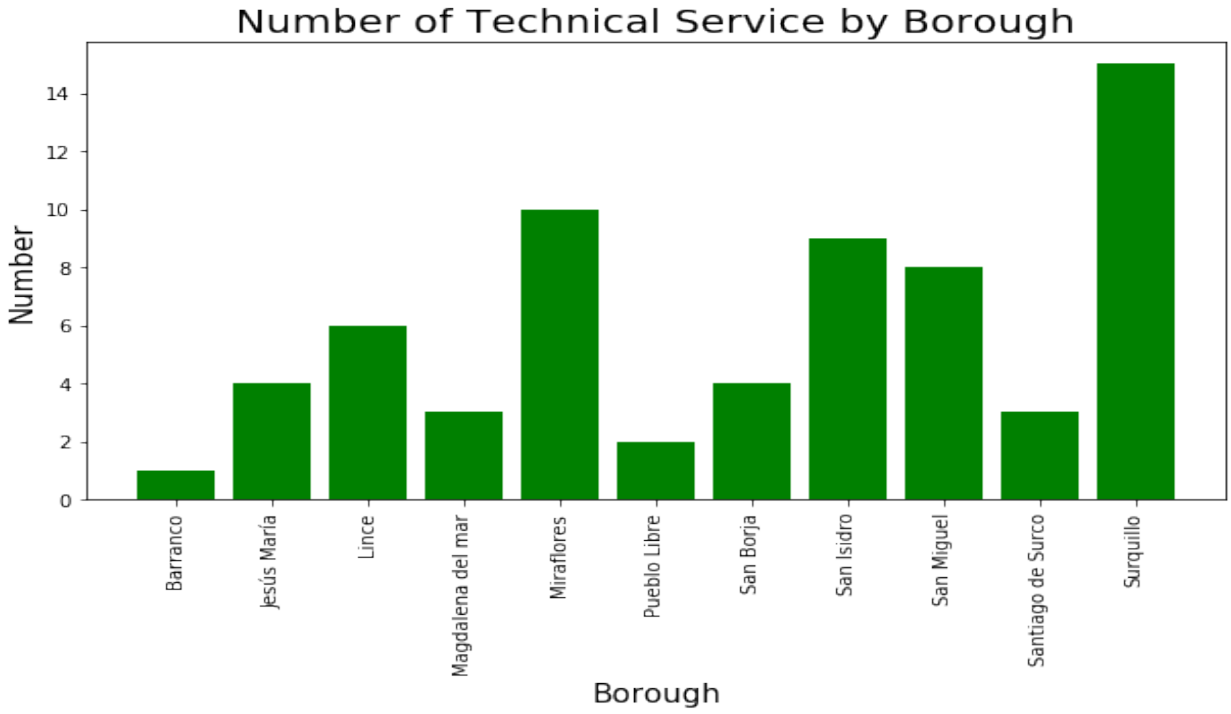
4.1 Explore Technical Services and define information of interest :

name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	Borough	crossStreet	id
BYD Servicio Técnico	Automotive Shop	Av. Venegas 898 B Santiago de Surco	-12.140631	-77.012553	[{'label': 'display', 'lat': -12.1406309181500...	1378.0	LIMA 33	PE	Lima	Lima	Perú	[Av. Venegas 898 B Santiago de Surco, Lima, LI...	Barranco	NaN	507979f6e4b07018b64fda6c
Servicio tecnico consolas	Electronics Store	Grl Santa Cruz 777	-12.073605	-77.041020	[{'label': 'display', 'lat': -12.073605, 'lng'...	614.0	NaN	PE	NaN	NaN	Perú	[Grl Santa Cruz 777, Perú]	Jesús María	NaN	50cce8bbe4b0f94b338da9a2
Soporte técnico PC, Laptops	Electronics Store	Pasaje Velarde 165	-12.055529	-77.038811	[{'label': 'display', 'lat': -12.055529, 'lng'...	1758.0	NaN	PE	Cercado Lima	Perú	Perú	[Pasaje Velarde 165 (Tienda 108), Cercado Lima...	Jesús María	Tienda 108	4fbaa20fe4b0a2484fbefd91
Servicio Tecnico B&C	Coworking Space	NaN	-12.055025	-77.037513	[{'label': 'display', 'lat': -12.05502472, 'ln...	1870.0	NaN	PE	NaN	NaN	Perú	[Perú]	Jesús María	NaN	4eaec070cc21b06a58251592
Sociedad Nacional de Técnicos	Factory	NaN	-12.072120	-77.041090	[{'label': 'display', 'lat': -12.0721200919896...	515.0	NaN	PE	NaN	NaN	Perú	[Perú]	Jesús María	NaN	4f14d52ce4b03f4c03162058
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	Borough	crossStreet	id
Soporte Tecnico	College Lab	Marsano 242	-12.104859	-77.017614	[{'label': 'display', 'lat': -12.1048588496179...	1592.0	051	PE	Lima	Lima	Perú	[Marsano 242 (Surquillo), Lima, 051, Perú]	Surquillo	Surquillo	51b74ca3498e88b3b83243fd
Soporte Tecnico - Usmp (Fcctp)	Assisted Living	Marsano 242	-12.104851	-77.017569	[{'label': 'display', 'lat': -12.1048512571751...	1595.0	051	PE	Lima	Lima	Perú	[Marsano 242 (Surquillo), Lima, 051, Perú]	Surquillo	Surquillo	51b74d75498e8ae18ca1426a
loreal servicio técnico	Building	NaN	-12.102933	-77.019302	[{'label': 'display', 'lat': -12.1029329299926...	1763.0	NaN	PE	NaN	NaN	Perú	[Perú]	Surquillo	NaN	512b7978e4b092e6017e6de7
Servicio Técnico de Philips	Electronics Store	Av Galvez Barnechea 1085 San Borja	-12.106954	-77.010478	[{'label': 'display', 'lat': -12.1069536820870...	1781.0	NaN	PE	Lima	Lima	Perú	[Av Galvez Barnechea 1085 San Borja, Lima, Perú]	Surquillo	NaN	505dfa2c498e46c28b3301c2
Asociacion mutualistas de tecnicos y oficiales...	Auditorium	NaN	-12.116933	-77.026043	[{'label': 'display', 'lat': -12.1169332271954...	507.0	NaN	PE	NaN	NaN	Perú	[Perú]	Surquillo	NaN	50cb6416e4b044fa48be814b

4.2 Show graph for Technical Services by Borough

4.3 Show how many venues (Technical Service) were returned for each Borough



4.4 Analyze each Borough along with the top 5 most common venues

----Barranco----

	venue	freq
0	Automotive Shop	1.0
1	Assisted Living	0.0
2	Furniture / Home Store	0.0
3	Professional & Other Places	0.0
4	Other Repair Shop	0.0

----Jesús María----

	venue	freq
0	Electronics Store	0.50
1	Factory	0.25
2	Coworking Space	0.25
3	Furniture / Home Store	0.00
4	Professional & Other Places	0.00

----Lince----

	venue	freq
0	Electronics Store	0.50
1	Office	0.33
2	Factory	0.17

3 Furniture / Home Store 0.00
4 Professional & Other Places 0.00

----Magdalena del mar----

	venue	freq
0	Electronics Store	1.0
1	Assisted Living	0.0
2	Furniture / Home Store	0.0
3	Professional & Other Places	0.0
4	Other Repair Shop	0.0

----Miraflores----

	venue	freq
0	Electronics Store	0.2
1	Shop & Service	0.1
2	Auto Garage	0.1
3	Other Repair Shop	0.1
4	Business Service	0.1

----Pueblo Libre----

	venue	freq
0	Auto Garage	0.5
1	Other Repair Shop	0.5
2	Assisted Living	0.0
3	Furniture / Home Store	0.0
4	Professional & Other Places	0.0

----San Borja----

	venue	freq
0	Electronics Store	0.50
1	Factory	0.25
2	Convenience Store	0.25
3	Furniture / Home Store	0.00
4	Professional & Other Places	0.00

----San Isidro----

	venue	freq
0	Office	0.22
1	Shop & Service	0.11
2	Professional & Other Places	0.11
3	Automotive Shop	0.11
4	Camera Store	0.11

----San Miguel----

	venue	freq
0	Building	0.25
1	Furniture / Home Store	0.12
2	Meeting Room	0.12
3	IT Services	0.12

4 Gym 0.12

----Santiago de Surco----

	venue	freq
0	Automotive Shop	0.33
1	Home Service	0.33
2	Electronics Store	0.33
3	Assisted Living	0.00
4	Furniture / Home Store	0.00

----Surquillo----

	venue	freq
0	Electronics Store	0.20
1	Assisted Living	0.07
2	Department Store	0.07
3	Other Repair Shop	0.07
4	Medical Lab	0.07

4.5. Create the new dataframe and display the top 10 venues for each Borough.

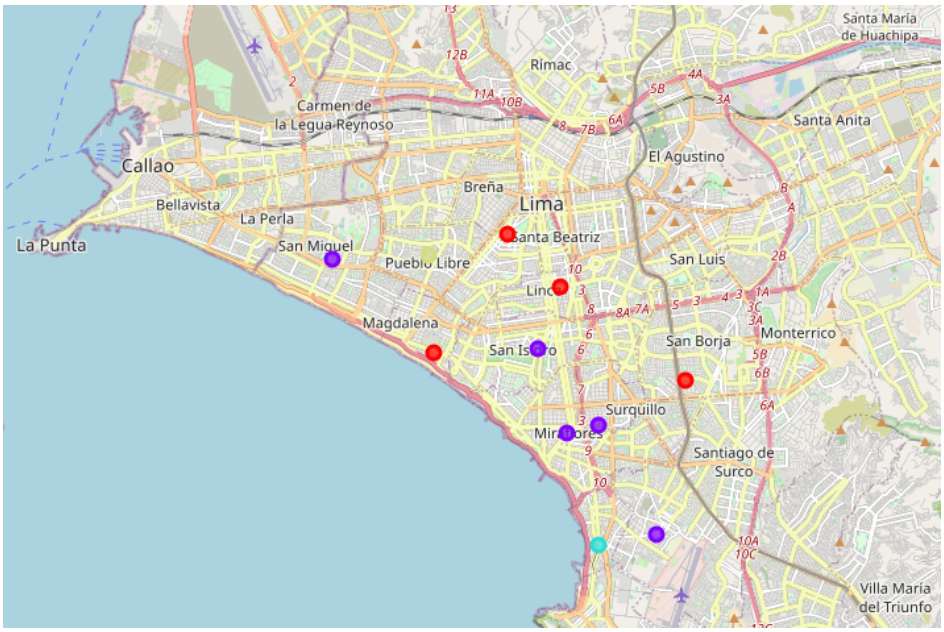
Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Barranco	Automotive Shop	Shop & Service	Professional & Other Places	Auditorium	Auto Garage	Building	Business Service	Camera Store	College Lab	Convenience Store
Jesús María	Electronics Store	Coworking Space	Factory	Department Store	Auditorium	Auto Garage	Automotive Shop	Building	Business Service	Camera Store
Lince	Electronics Store	Office	Factory	Department Store	Auditorium	Auto Garage	Automotive Shop	Building	Business Service	Camera Store
Magdalena del mar	Electronics Store	Professional & Other Places	Auditorium	Auto Garage	Automotive Shop	Building	Business Service	Camera Store	College Lab	Convenience Store
Miraflores	Electronics Store	IT Services	Auditorium	Auto Garage	Business Service	Department Store	Shop & Service	Other Repair Shop	Medical Lab	Home Service

4.6. Cluster Boroughs

I used Kmeans to cluster the Borouhs with k=4, because the venues are the most important in the "Lima Moderna" axe:

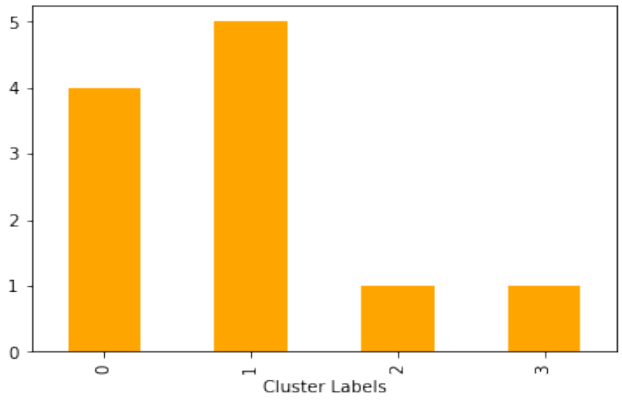
Borough	Axes	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Barranco	Lima Moderna	-12.1492	-77.0217	2	Automotive Shop	Shop & Service	Professional & Other Places	Auditorium	Auto Garage	Building	Business Service	Camera Store	College Lab	Convenience Store
Jesús María	Lima Moderna	-12.0700	-77.0453	0	Electronics Store	Coworking Space	Factory	Department Store	Auditorium	Auto Garage	Automotive Shop	Building	Business Service	Camera Store
Lince	Lima Moderna	-12.0833	-77.0317	0	Electronics Store	Office	Factory	Department Store	Auditorium	Auto Garage	Automotive Shop	Building	Business Service	Camera Store
Magdalena del mar	Lima Moderna	-12.1002	-77.0647	0	Electronics Store	Professional & Other Places	Auditorium	Auto Garage	Automotive Shop	Building	Business Service	Camera Store	College Lab	Convenience Store
Miraflores	Lima Moderna	-12.1207	-77.0299	1	Electronics Store	IT Services	Auditorium	Auto Garage	Business Service	Department Store	Shop & Service	Other Repair Shop	Medical Lab	Home Service
Pueblo Libre	Lima Moderna	-12.0753	-77.0659	3	Other Repair Shop	Auto Garage	Shop & Service	Department Store	Auditorium	Automotive Shop	Building	Business Service	Camera Store	College Lab
San Borja	Lima Moderna	-12.1072	-76.9992	0	Electronics Store	Convenience Store	Factory	Department Store	Auditorium	Auto Garage	Automotive Shop	Building	Business Service	Camera Store
San Isidro	Lima Moderna	-12.0991	-77.0375	1	Office	Shop & Service	Professional & Other Places	Automotive Shop	Camera Store	Coworking Space	Electronics Store	Meeting Room	Home Service	College Lab
San Miguel	Lima Moderna	-12.0763	-77.0909	1	Building	Meeting Room	IT Services	Gym	General Entertainment	Furniture / Home Store	Shop & Service	Convenience Store	Auditorium	Auto Garage
Santiago de Surco	Lima Moderna	-12.1464	-77.0067	1	Electronics Store	Automotive Shop	Home Service	Department Store	Auditorium	Auto Garage	Building	Business Service	Camera Store	College Lab
Surquillo	Lima Moderna	-12.1186	-77.0217	1	Electronics Stor	Department Stor	Auditorium	Auto Garag	Automotive Sho	Building	Business Serv	College Lab	Shop & Serv	IT Servi

4.7 Visualize the resulting clusters



5. Discussion

I can estimate the number of 1st Most Common Venue in each cluster. Thus, I created a bar chart which may help us to find proper labels for each cluster:



When we examine above graph we can label each cluster as follows:

Cluster 0 : Diferent services (repair of computer, laptops and accesories)

Cluster 1 : Electronics Store (sale and repair of computer, laptops and accesories)

Cluster 2 : Other reparacion services

Cluster 3 : Other services

The 4 identified clusters help us to view the types of business of technical service, sale and repair of computers currently offered in these districts of "Modern Lima" with high economic income per family.

I set the optimum k value to 4. However, only 13 district coordinates were used. For future studies, these data can also be accessed dynamically from specific platforms or packages, using coordinates in real time.

6. Conclusion

The city of Lima, Peru, is divided into 50 districts (Boroughs), and organized into 7 axes or zones, and 5 socioeconomic segments (A, B, C, D, E). In the present study, I focused only on segment AB, which is the one with the highest economic income and belongs to the axe of "Modern Lima" with 13 districts.

The objective of this study is to inform citizens that, due to the Pandemic and the fact that they are confined to their homes, the option of locating the closest place to their homes to find technical repair service and sale of computers, laptops and accessories, is necessary for working, studying and entertaining at home.

At the same time, it helps to identify potential investors of the places where you can invest in a business of these types, since the number of devices in families is increasing over and over again, therefore it is a great opportunity.

This study has allowed me to put into practice and deepen my knowledge about Phyton, the different libraries for analysis used and learning algorithms.

I hope that this study will guide future students in their learning in this world of Data Science, since every day job opportunities increase.