Tourism in Chile and Peru post COVID-19: clustering touristic neighborhoods for boosting countries' economies

Author: Diego Ignacio López Orellana

Date: September 18th 2020

Table of Contents

1. Introduction / Business Problem

1.1 Discussion of the Background

Latin America is one of the most famous and visited regions worldwide, by millions of tourists from the continent and around the globe. According to the prestigious worldwide Market Research Agency ***'Euromonitor International'*** on its report *'Top 100 City Destinations 2018'*, there are seven Latin American cities among the 100 most visited cities around the world in 2017 **[1]**. The cities, in descending order based on their number of visitors, are:

- **Cancun, Mexico (35th place).**
- **Punta Cana, Dominican Republic (58th place).**
- **Mexico City, Mexico (88th place).**
- **Buenos Aires, Argentina (89th place).**
- **Lima, Peru (91th place).**
- **Santiago de Chile, Chile (93th place).**
- **Rio de Janeiro, Brazil (94th place).**

In consequence, as shown in the 2020 '*Economic Impact Reports'* from the ***'World Travel & Tourism Council'***, local and foreign tourists report millions, and even billions of US dollars (in Argentina, Brazil and Mexico) **[2]**, to each of the mentioned countries in the previous list. Therefore, the Travel and Tourism industry is key for the economic development of these Latin American countries; its impact shown **as a percentage of the total GDP [2] (Gross Domestic Product) [3] by country** is the following:

- **Argentina: 9.2% of total GDP** (+0.2%, *2019 Travel & Tourism GDP Growth*).
- **Brazil: 7.7% of total GDP** (+3.0%, *2019 Travel & Tourism GDP Growth*).

- **Chile: 10.0% of total GDP** (+0.3%, *2019 Travel & Tourism GDP Growth*).
- **Dominican Republic: 16.3% of total GDP** (+2.0%, *2019 Travel & Tourism GDP Growth*).
- **Mexico: 15.5% of total GDP** (+2.0%, *2019 Travel & Tourism GDP Growth*).
- **Peru: 9.3% of total GDP** (+2.9%, *2019 Travel & Tourism GDP Growth*).

1.2 Description of the Business Problem

As it is well known by everybody, COVID-19 has become one of the worst pandemics of the last 100 years, with **more than 23 million confirmed cases and over 800.000 confirmed deaths worldwide** [4]. The novel COVID-19 virus not only has taken innocent lives, it has also caused the worst economic crisis since the *Great Depression in 1929* [5]. The ***World Tourism Organization (UNWTO)*** point that international tourism, between January and May 2020, has suffered a fall of *"… **300 million tourists and US$320 billion lost in international tourism receipts** – more than three times the loss during the Global Economic Crisis of 2009"* [6].

Latin America has become one of the most affected regions worldwide during the pandemic. As of August 24th, there are **more than 6.7 million confirmed cases, 259.065 confirmed deaths** and several countries are battling major outbreaks [7]. Thus, in each country all the productive and service sectors have been affected in their incomes. According to the UN Economic Commission for Latin America and the Caribbean, COVID-19 pandemic "… **may cause total GDP in the Caribbean and Latin America to fall by 8 percentage points and 1 percentage point**, respectively" [8]. This percentage, seemingly insignificant, would cause every country loss thousands of millions of US dollars, thus deepening the economic crisis that the Latin American countries will face in the upcoming years.

Once the COVID-19 is over, it will be key for Latin American countries to boost their tourism countrywide to help recover their economies as soon as possible. From this background, and for the sake of simplicity, it is formulated as the first problem to solve, the **clustering (*using k-Means Machine Learning algorithm*) of the different neighborhoods only from Lima (Peru) and Santiago de Chile (Chile), to find which are the most touristic ones. Hence, it will help the local and international touristic agencies to promote tourism in those areas** to increase the flow of visitors, and with it, the money that this economic item reports both to Chile and Peru.

And as a second problem to tackle, we **will use the same k-Means clustering algorithm to group these touristic neighborhoods between Lima and Santiago de Chile**, with the objective of **finding similar (or dissimilar) touristic neighborhoods between the mentioned cities.** This valuable information **could help Peruvian and Chilean touristic agencies, for example, to create alliances or a collaborative network between each other and create touristic packages across these cities to support tourism between different countries**, attract more Latin American tourists and other ones from other continents such as Africa, Asia or Europe; in addition to increase the tourism profits and the percentage of GDP that tourism contributes to each country.

Later, in a future study, this analysis would be deepened by adding more Latin American cities to analyze. This more detailed study **could even help the local governments adopting cooperation policies to promote cooperative tourism plans and boost this item in Latin America, recover faster their economy and the GDP percentage that tourism contributes to each Latin American country.**

In these problems, we will only **analyze Lima and Santiago de Chile metropolitan areas, according to its definition and a population density of at least 1.000 inhabitants per square mile [9], which are among the 100 most visited in the world**. This decision was taken because smaller cities such as Cancun and Punta Cana are much smaller and have fewer neighborhoods, which are mainly defined in advance as touristic sectors, since tourism is their main source of incomes. And also, because this project seeks to carry out a pilot study of analysis and measurements for cooperative tourism in Latin America (starting with Chile and Peru), which is suggested to deepen in a second study.

1.3 Stakeholders interested in the Project

The potential stakeholders that this Data Science project contemplates are the following:

I. Chilean and Peruvian touristic agencies which are interested in improving their touristic packages post COVID-19 pandemic to capture more tourists around the world, with the objective of increasing the number of visitors in their countries and the revenues perceived by the tourism. The project **also contemplates touristic agencies of both countries interested in cooperating between each other and promote greater packages that can boost in a broader and more cooperative way the tourism in this region,** always with the focus in increasing the visits and GDP percentage that tourism contributes to each country.

II. Governments and their Tourism Ministries from Chile and Peru which are **interested in adopting collaborative policies to boost faster and more effectively the tourism in their countries, under a post COVID-19 pandemic context, to contribute for a faster economical recovering for their countries and subregion**: more local/foreign tourists and an important increase in Tourism & Travel revenues that goes to each State coffers.

2. Description of the Data

2.1 Datasets used in the Project

In this project, we will gather reliable data and preprocess it using different techniques to finally get the boroughs, neighborhoods, coordinates and the most common venues, in Lima and Santiago de Chile, to be analyzed (mentioned in the Business Problem). The sources from where we will obtain our datasets and the information that each one contains, according to each city, are the following:

2.1.1 Lima, Peru

For Peru's capital, we work with the boroughs and neighborhoods from Lima and Callao, as both provinces are merged in the urban Lima Metropolitan Area [10]. In the case of its neighborhoods (known in Peru as *'barrios'*), we work with the *'centros poblados'* as they are the smallest political-administrative circumscriptions of the country; due to the lack of data about Lima's neighborhoods.

The information about Lima's *'centros poblados'* is gathered from the *'Plataforma Nacional de Datos Abiertos'* webpage [11], which belongs to the Peruvian government. From that page, we get a **.xlsx file** called *'ListadoCentroPobladosMTC.xlsx'*. It contains the following relevant columns from the Peruvian *'centros poblados'*:

Column (Feature)	Description
Provincia	Each of the Peruvian provinces. ***Data type: String.***
Distrito	Each of the Peruvian districts. ***Data type: String.***
CCPP	Each Peruvian *'centro poblado'*. ***Data type: String.***

Latitud (coord X)	Latitude coordinate of each Peruvian *'centro poblado'*. ***Data type: Float.***
Longitud (coord Y)	Latitude coordinate of each Peruvian *'centro poblado'*. ***Data type: Float.***
CLASIFICACIÓN INEI	Represents if each Peruvian *'centro poblado'* is rural or urban. ***Data type: String.***

Table 1: *ListadoCentroPobladosMTC.xlsx* file columns and their descriptions.

2.1.2 Santiago de Chile, Chile

The capital of Chile is divided into 32 boroughs or communes (called *'Comunas'*) [12], which are subdivided into hundreds of neighborhoods, known as *'Barrios'*. For this project, we will also add four boroughs located in Santiago conurbation: Padre Hurtado, Peñaflor, Puente Alto and San Bernardo.

The dataset containing its boroughs will be retrieved from Wikipedia [13] by applying Web Scraping to get table from an HTML webpage. Unfortunately, there is no information available about their neighborhoods nor latitude or longitude values. Therefore, we will retrieve both coordinates by using the ArcGIS World Geocoding Service [14], to convert each neighborhood/borough name into their coordinates.

On this operation, we make calls to its database by sending **Request/GET** sentences to retrieve this information. It will be stored in a **.json file**, which will be parsed and converted to a Pandas data frame to start working with its information.

After creating both datasets, we will join them to create a ***Pandas data frame*** containing all the information required. The data frame will contain the following columns:

Column (Feature)	Description
Neighborhood	Name of each neighborhood or *'barrio'* in Santiago. Here, we assume boroughs as neighborhoods. ***Data type: String.***
Location	Position in a cartesian plane with respect to the direction of each neighborhood. ***Data type: String.***
Latitude	Latitude coordinate from each neighborhood in Santiago. ***Data type: Float.***
Longitude	Longitude coordinate from each neighborhood in Santiago. ***Data type: Float.***

Table 2: *Pandas* data frame columns and their descriptions, for Santiago de Chile.

The latitude and longitude coordinates for the neighborhoods in Santiago de Chile, will be retrieved making calls to a database, which will be explained next.

2.2 APIs used to gather venues and their coordinates

As mentioned before, to gather the coordinates of each neighborhood for Santiago de Chile, where this information is missing, we will make calls to the **ArcGIS database by its World Geocoding REST API [14],** to get the latitude and longitude values required per neighborhood.

On the other hand, to cluster the neighborhoods based on their venues, we will use the **Foursquare API [15]** to make calls to its database and retrieve a **.json file** containing the **venues of the different**

neighborhoods of each city within a radius of 2500 meters for Lima, and 7000 meters for Santiago de Chile, to determine which ones are the most common based on their categories.

2.3 Python libraries used for the Capstone

To import the data, preprocess, make an exploratory analysis and then model it and evaluate the results, we need to import several Python libraries to perform these several tasks that we will carry out during the final project. This is because these libraries contain the functions and methods needed to perform the segmentation of the neighborhoods. The following table summarizes each Python library that will be used in the Final Capstone and its description:

Python Library	Description
NumPy	Math library to work with N-dimensional arrays.
Pandas	Library for importing, manipulating and analyzing data in data frames.
JSON	Contain methods to handle JSON files.
GeoPy	Python client which contains geocoder class for several geocoding services to retrieve coordinates of different places.
BeautifulSoup 4	Popular Python library, widely used for Web Scraping: it allows to parse and pull data out of HTML and XML files.
Requests	Library to handle HTTP requests.
Matplotlib	Popular Python plotting package with contains several modules for 2D and 3D plotting.
Scikit Learn	Free Machine Learning library to work with several ML algorithms, performs most of the tasks in a ML pipeline.
Folium	Data visualization library, it is used to visualize geospatial data by the creation of maps at any location in the world.

Table 3: Python libraries used in the Capstone and their descriptions.

3. Methodology for Capstone Project

This project is based on a structured Data Science working methodology involving a series of steps from Business and Data understanding to sketch the suggestions from the outcomes of the project to its potential stakeholders: **Chilean and Peruvian touristic agencies, the Governments and their Tourism Ministries from Chile and Peru**.

The following image [16] briefly resume these steps:

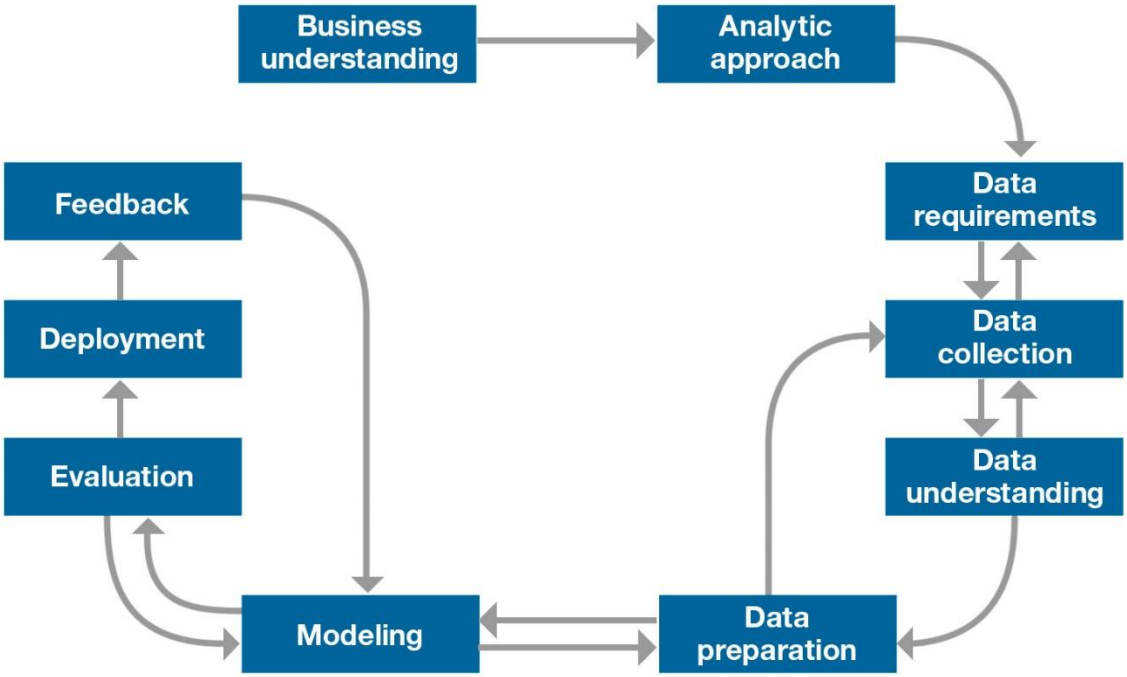

Figure 1: Data Science Methodology used for the project. Source: IBM Big Data and Analytics Hub.

We can summarize the tasks to perform in this project as follows:

1. Business and Data understanding to come up with a problem and define the data requirements to solve it.
2. Data Preparation: wrangling, formatting and preprocessing it to prepare data for further analysis.

3. Exploratory Data Analysis: this step is done to summarize the main venues from each neighborhood to later group them based on the categories of their most common venues.
4. Modelling: with the datasets from Lima and Santiago de Chile, we apply the k-Means Clustering Algorithm per each city to cluster their neighborhoods. This is done to find which ones are more similar based on their venues' categories.
5. Evaluation: each model is evaluated to segment and label which neighborhoods are the most touristic ones.
6. Deployment: write and show to the potential stakeholders the touristic neighborhoods and governments of Chile and Peru, to improve their touristic packages post COVID-19 pandemic.

After this step, the data frames from the touristic neighborhoods of both cities are merged into one. Then, this data frame is used in the following steps:

7. Re-Modelling: the data frame which groups the touristic neighborhood is used to run again the k-Means Clustering Algorithm to cluster and find which neighborhoods are more similar based on their most popular venues' categories.
8. Re-Evaluation: from this model, we segment and label the touristic neighborhoods across the cities studied: Lima & Santiago de Chile.
9. Re-Deployment: again, we deliver a stack of proposals to the touristic neighborhoods and governments from Lima and Santiago de Chile. But now, the focus is to propose them cooperative alliances and policies that can help their countries to recover the touristic industry, in a faster and more effectively way, their economic recovering.

This project is divided into three main parts:

1. Exploring, clustering and segmentation for neighborhoods, to detect touristic zones in Lima, Peru.

2. Exploring, clustering and segmentation for neighborhoods, to detect touristic zones in Santiago de Chile, Chile.
3. Re-cluster and re-segment touristic neighborhoods from Lima and Santiago de Chile, in a joint analysis of both cities.

Therefore, the methodology, results and discussion sections will be presented according to each part of this Capstone Project.

4. Exploring, clustering and segmentation for neighborhoods in Lima, Peru

4.1 Data cleaning and understanding

First, the .xlsx file '*ListadoCentroPobladosMTC.xlsx*' is converted to a *Pandas* data frame, considering that we filtered the neighborhoods to retrieve urban neighborhoods from Callao and Lima.

	Province	Borough	Neighborhood	Latitude	Longitude	Rural/Urban Neighborhood
0	CALLAO	CALLAO	CALLAO	-12.051558	-77.134117	URBANO
1	CALLAO	BELLAVISTA	BELLAVISTA	-12.062462	-77.128632	URBANO
2	CALLAO	CARMEN DE LA LEGUA REYNOSO	CARMEN DE LA LEGUA REYNOSO	-12.053540	-77.098017	URBANO
3	CALLAO	LA PERLA	LA PERLA	-12.070180	-77.118287	URBANO
4	CALLAO	LA PUNTA	LA PUNTA	-12.072643	-77.164256	URBANO

Figure 2: Data frame of Lima's neighborhoods and its coordinates. Source: Own Elaboration.

From the preliminary information retrieved from the data set, we conclude that in Lima and Callao provinces, there are **50 boroughs and 51 neighborhoods *('centros poblados')* in their urban zones**.

4.2 Exploratory Data Analysis

Next, we use Foursquare API to explore and retrieve venues for each neighborhood in Lima, within a **radius of 2500 meters and a maximum of 100 venues per zone**.

To do this, we create a function called ***getNearestVenues***, which makes repeated **calls to the Foursquare API, to retrieve information about each venue** from each neighborhood from Lima, within a radius of **2500 meters**.

Later, this function **reorders the data from the neighborhoods and its respective venues**, appending it **into a new data frame** called ***nearest_venues*** as shown below:

```
def getNearestVenues(names, latitudes, longitudes, radius=2500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        #print(name) Optional -> Show Lima's neighborhood names

        # Create the API request URL (called 'url')
        url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # Make the GET request to call the Foursquare database
        results = requests.get(url).json()['response']['groups'][0]['items']

        # Return only relevant information (latitude, longitude, category, name) for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearest_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearest_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearest_venues)
```

Figure 3: Function *getNearestVenues* to retrieve venues from Lima's neighborhoods. Source: Own Elaboration.

<u>Note:</u> we will **use this function again** to retrieve the **venues from the neighborhoods of Santiago**, but using a radius of **7000 meters**.

Here we get that, considering all the neighborhoods from Lima, we retrieve **a total of 3009 venues, with 227 distinct venues' categories (types of places)**. Then, one hot encoding is performed and a new data frame is created to calculate the relative frequencies of each venue category in Lima, Peru.

Now, to study the trend of which are the most common venues in Lima, we built a bar plot which shows the top 20 venues across all the neighborhoods from Peruvian capital. This plot is shown next:

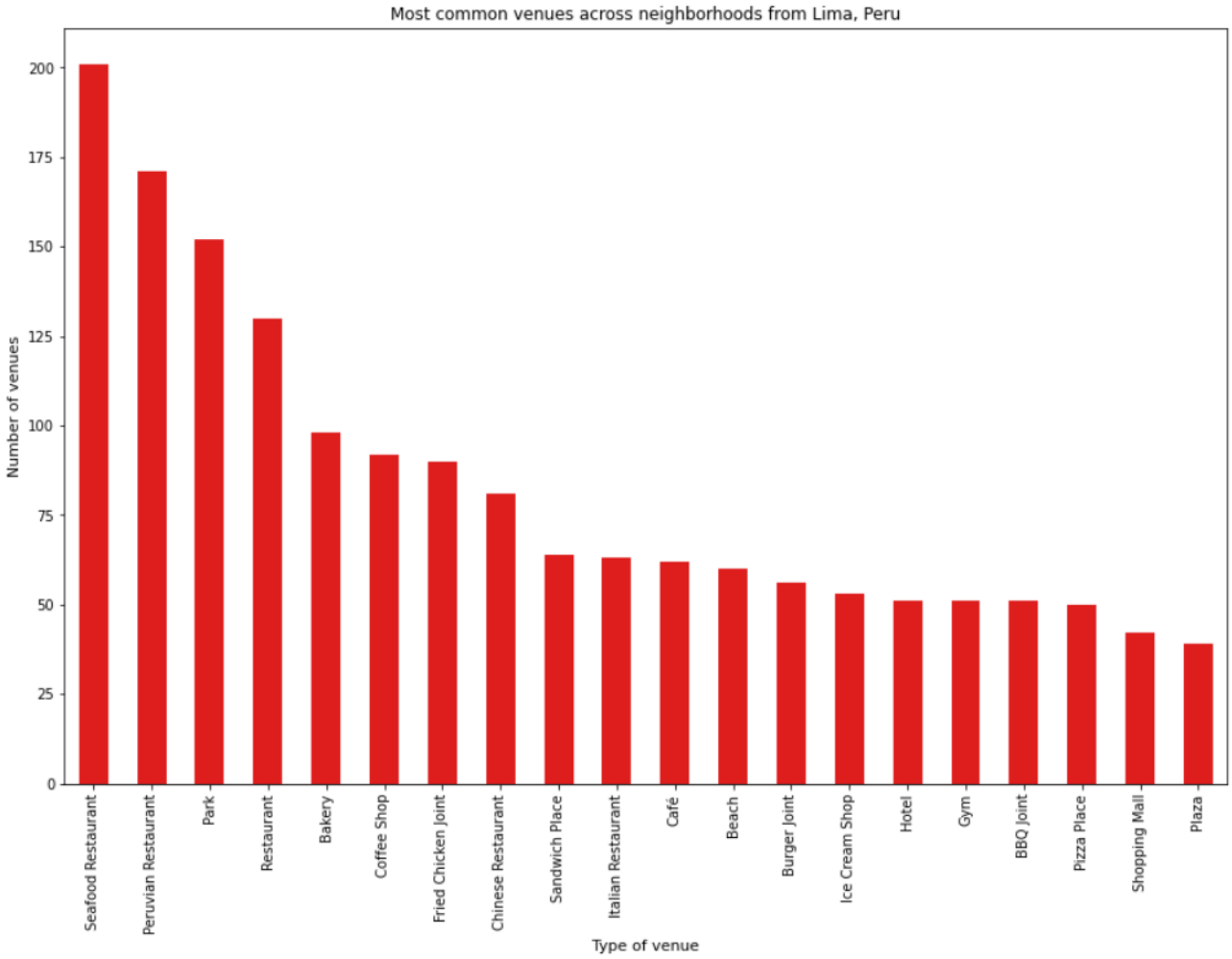

Figure 4: Bar plot with top 20 most common venues in Lima, Peru. Source: Own Elaboration.

Here we can see that in Lima, the most popular venues are: **Restaurants (Seafood and Peruvian), Parks and Bakery**. *Limeños* (Lima's people) and tourists also love to eat **Chinese food**.

Lima is a **huge coastal city and very famous for its local food fusioned with Chinese food**, a heritage from Chinese people who emigrated to Peru between the 19th and 20th centuries [17]. Therefore, the **Peruvian cuisine is considered among the best ones worldwide** [18], so people love to eat their delicious dishes. Also, locals and foreigners **love to eat delicious cakes and breads in their bakeries**.

On other hand, as **Lima was planified to take advantage of its extensive coastline**, which combined with a **nice and mild weather** in the zone, make **Parks one of the most preferred venues to take a walk**. Next, we retrieve from the relative frequency of venues, the names of the 10 most common ones per neighborhood. This data will be used for modeling stage.

4.3 Modeling with k-Means Clustering Algorithm

In this part, we cluster the neighborhoods in Lima into different groups, according to the results of the k-Means Clustering Model. We determine to use this model, **because it can cluster big amounts of data in small time, in a very accurate way**.

To determine which the optimal K to use in the model, we use the **Elbow method**. It consists in running the k-Means Model for a range of K different clusters.

Then we calculate, for each number of clusters, its respective **Sum of Squared Errors (SSE)**, between each point in a cluster and its

respective centroid within it. Finally, we sum up the errors of all the clusters to get the SSE value and plot them (x coordinates for K values, y coordinates for Mean distances to data points from cluster centroids) [19].

When we apply the Elbow method, as we know in advance that increasing K will always decrease the error or SSE value, then the **elbow point** (and therefore the K value to use in k-Means Model) is determined **where the rate of decrease sharply shifts**.

After running k-Means algorithm **20 times** and plotting each K value with its respective SSE value, we plot them in the next figure:

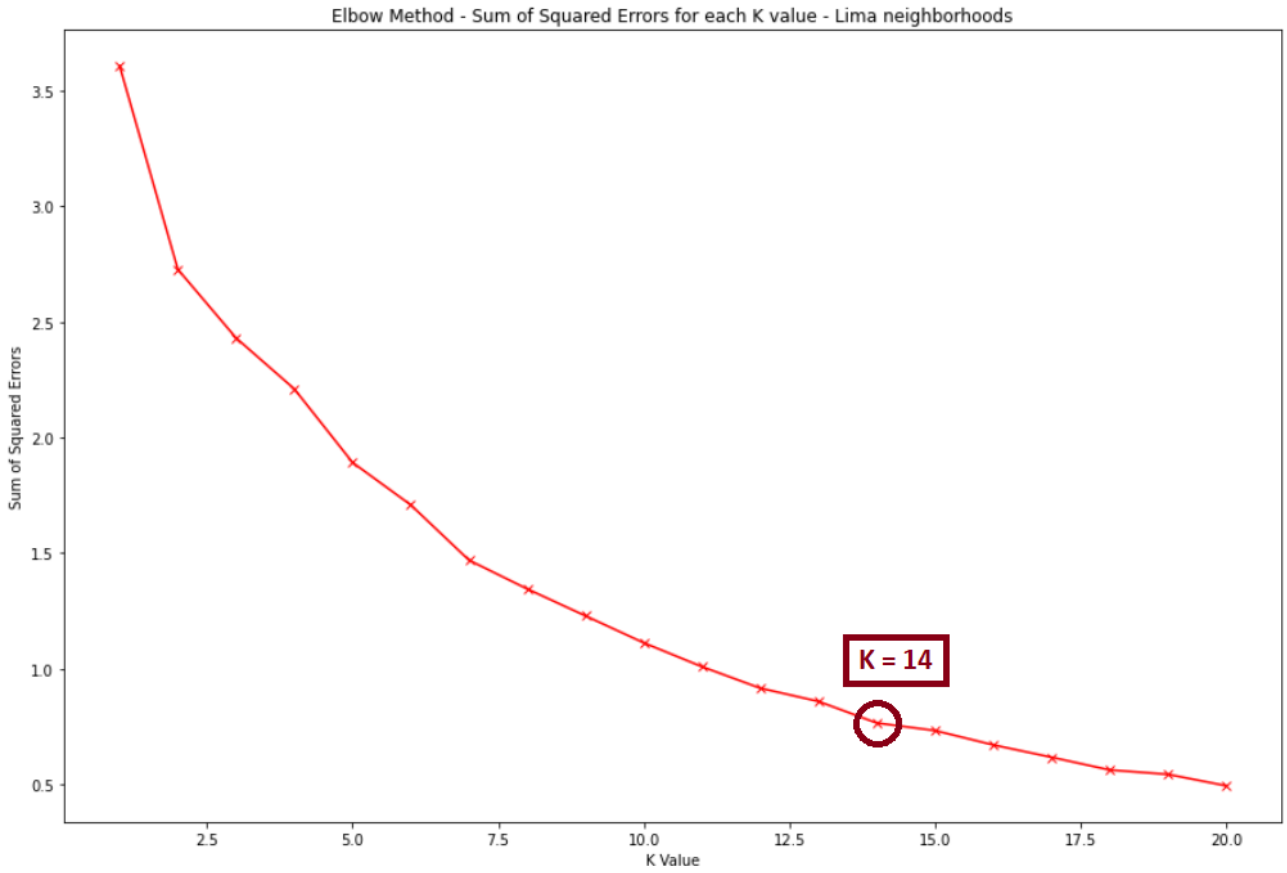

Figure 5: Elbow Method – SSE for each K value – Lima neighborhoods. Source: Own Elaboration.

As the rate of SSE decreases after **K=14, this is our optimal K value**.

Then, as **K represents the number of neighborhood clusters** in Lima, we plot the 14 clusters in the following **Folium map**:

Figure 6: Folium map of clustered Lima neighborhoods. Source: Own Elaboration.

4.4 Results: Segmenting and Labeling Lima's neighborhoods

Next, we **segment Lima's neighborhoods according to their cluster labels and name each one** of them. Here we retrieve only the touristic neighborhoods from the city, which will be used again when we do the joint analysis between Lima and Santiago de Chile.

The touristic neighborhoods were chosen considering their **connectivity**, **variety of venues to visit** and specially their **security**; because as Latin American people know, our countries and cities have experienced uneven growth and increasing segregation of their neighborhoods between rich and poor ones.

Although we divided Lima's neighborhood in 14 clusters, **only 6 of them were considered touristic (4 if we consider that clusters 1, 6, 13 are merged into one due to their similarity).** This result is based on the previous criteria. In the following table, we resume these clusters with respect to their label, neighborhoods and their main features:

N° of cluster	Cluster's Label	Neighborhoods
0	Touristic, cultural, green areas, gastronomic and financial	Barranco, Breña, Chaclacayo, Jesus María, Lince, Magdalena del Mar, Pueblo Libre, Miraflores, San Francisco de Borja, San Isidro, San Miguel, Santiago de Surco, Surquillo
1, 6, 13	Touristic beaches in Northern and Southern Lima	Santa Rosa, Pucusana, Punta Negra
5	Touristic beaches and small coastal towns in Northern and Southern Lima	Ancón, Punta Hermosa, San Bartolo, Santa María del Mar
11	Historical, cultural and coastal, located in Lima and Callao	Bellavista, Chorrillos, Callao, La Punta, La Victoria, Lima

Table 4: Segmenting and labeling of touristic neighborhoods in Lima.

4.5 Discussion: Recommendations to improve Lima's tourism, after COVID-19 pandemic

The background and recommendations for every touristic cluster in Lima, to recover faster after COVID-19 pandemic, are the following:

As we can appreciate in the clusters (1, 5, 6 and 13), Lima has an extensive coastline with several seaside touristic neighborhoods, which have a wide offer of restaurants, nightclubs, surf spots, parks and scenic lookouts.

Therefore, it is recommended for touristic agencies and Peruvian government to: **1) Promote more Lima beaches on their tourist brochures and mass media campaigns, 2) Improve road infrastructure to get to these neighborhoods faster and safer, and 3) Expand the hotel offer to receive more tourists.**

On the other hand, we have the most affluent and developed neighborhoods of Lima on Cluster 0, which combine green areas, cultural, gastronomic and touristic venues surrounded by extensive beaches and parks. Although this cluster counts with a wide amount of attractions to visit, **it is necessary to decongest its streets to generate a more sustainable development and expedite traffic** [20]. As a measure to improve this issue, is to **promote the use of public transport and improve its infrastructure, to decrease the use of particular vehicles**. It is also recommended to the government to **create extensive bicycle lanes and encourage the use of bicycles to get around these neighborhoods**; this would also help to improve the air and life quality for all the people from those Lima's neighborhoods.

Finally, we have the historical and cultural Cluster 11, which **include some of the oldest neighborhoods in Callao and Lima**. These zones they inherited some of the oldest buildings of the Inca and Spanish culture, including the **Cathedral of Lima, Government Palace** [21] **and Real Felipe Fortress** [22]. Not only that, these neighborhoods have a wide cultural offer, with several venues for tourists who love the artistic and musical exhibitions shown there. Although these are beautiful neighborhoods, they are very unsafe and, especially at night, there are many robberies on tourists [23]. Therefore, one recommendation for the Peruvian government is to **strengthen police control and through security cameras to reduce robberies and the level of insecurity in the historic Cluster of Lima and Callao**.

5. Exploring, clustering and segmentation for neighborhoods in Santiago de Chile, Chile

5.1 Data cleaning and understanding

Here, neighborhood names and their locations are retrieved from applying Web Scraping to the following webpage: https://es.wikipedia.org/wiki/Anexo:Comunas_de_Santiago_de_Chile. This **.json file** is converted to a Pandas data frame.

The previous data frame does not contain the Latitude and Longitude coordinates, for each neighborhood in Santiago. Hence, we **call the ArcGIS API by using the geocode method to get the coordinates values mentioned** [24], by making calls to ArcGIS database. We pass this **.json file** to a Pandas data frame. Both data sets are merged.

	Neighborhood	Location	Latitude	Longitude
0	Cerrillos	Southwest	-33.48862	-70.70079
1	Cerro Navia	Northwest	-33.43472	-70.72991
2	Conchali	North	-33.39659	-70.67104
3	El Bosque	South	-33.55566	-70.66589
4	Estacion Central	Southwest	-33.45100	-70.67941

Figure 7: Data frame of Santiago's neighborhoods and its coordinates. Source: Own Elaboration.

From the information retrieved out of this data set, **for Santiago de Chile we work with 36 neighborhoods (*'comunas'*)**.

5.2 Exploratory Data Analysis

Again, we use Foursquare API to retrieve the venues from each neighborhood in Santiago de Chile. Here, we define a **radius of 7000 meters and a maximum of 100 venues retrieved per zone**.

Considering all the neighborhoods from Lima, Foursquare retrieved **a total of 3504 venues, with 217 distinct venues' categories (types of places)**. As it was done for Lima's neighborhoods, One Hot Encoding and a new data frame is created to calculate the relative frequencies of each venue category in Santiago de Chile, Chile.

We are interested in studying the trending of the most common venues, across all the neighborhoods in Santiago de Chile. To do that, we graph a bar plot which shows the 20 most common venues from this city, as shown in the following figure:

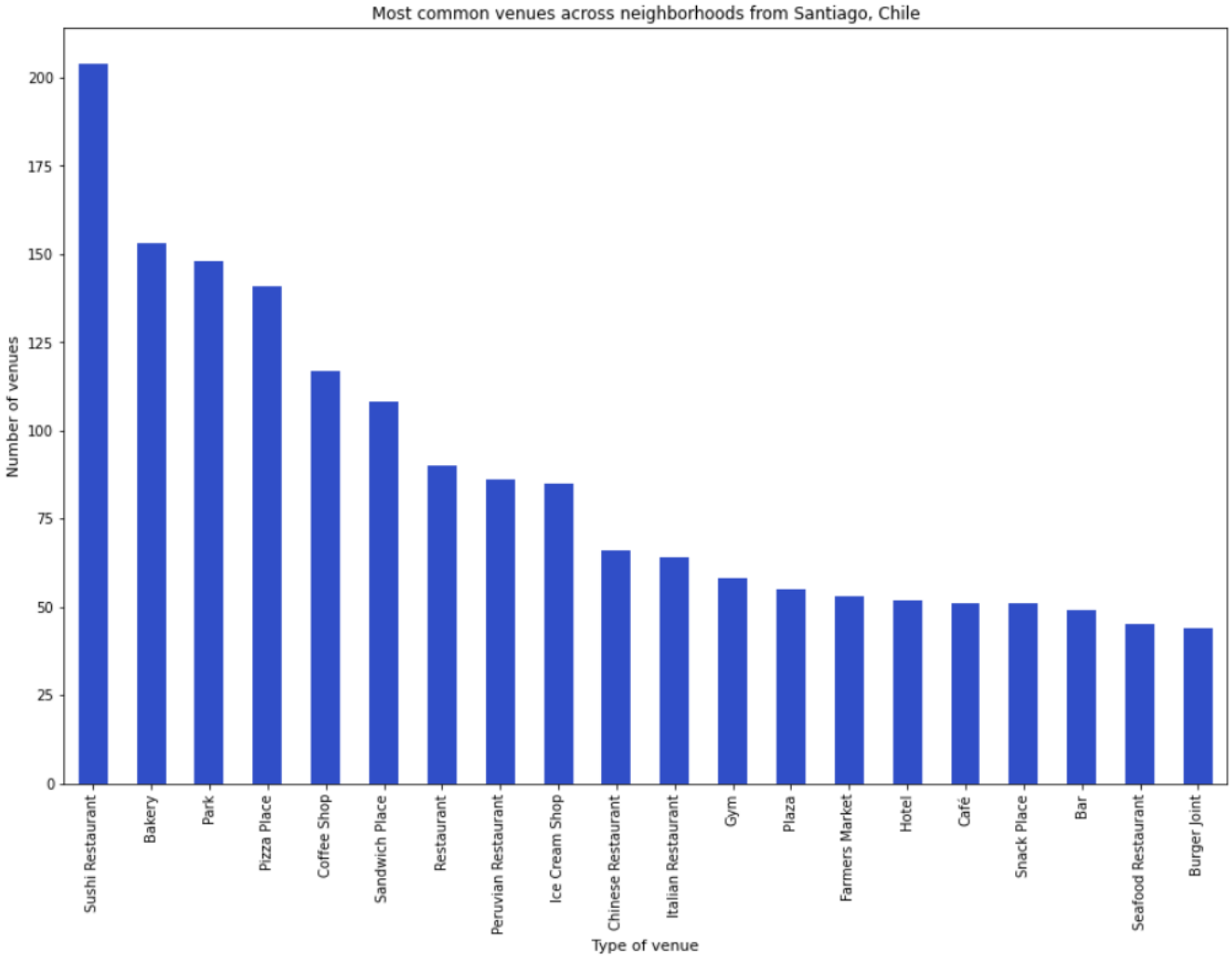

Figure 8: Bar plot with top 20 most common venues in Santiago, Chile. Source: Own Elaboration.

According to the venues returned by Foursquare API, the **five most popular venues** in Santiago de Chile are:

1. **Sushi Restaurant**
2. **Bakery**
3. **Park**
4. **Pizza Place**
5. **Coffee Shop**

Chile is a **country highly influenced by foreign cultures, and each year receives a large number of immigrants from various countries** [25]. That explains why **Sushi restaurants are far more popular than Chilean ones** [26]; this boom started in the last decade with the expansion of venues to eat Sushi in Chile. Today is very typical to see a lot of Asian restaurants in every neighborhood from Santiago.

But the *santiaguinos* (Santiago de Chile's people) and tourists **also love Chilean bread and their varieties**: Chile is the **second country behind Germany which consumes the most bread (in kilograms) per person in the world** [27]. Therefore, bakeries are highly popular in Santiago.

People in Santiago **love to visit their numerous parks** while they take a walk with the dog, practice sports, meet with their friends or want to escape from the noise of the city.

Next, **as *santiaguinos* and every tourist love pizza**, you can visit and enjoy Pizza places in practically all the neighborhoods in Santiago.

Finally, Chileans and their tourists **love to socialize while they enjoy a Cappuccino with a cake or sandwich**. This explains the popularity of Coffee Shops where people enjoy a nice time with their friends or family, and enjoy a delicious meal.

5.3 Modeling with k-Means Clustering Algorithm

In this part, we will cluster the neighborhoods in Santiago de Chile into different groups, according to the results of the k-Means Clustering Model. The **decision to use this algorithm and the explanation of elbow point to choose K were explained before**.

Again, we run k-Means algorithm **20 times** and plot each K value with its respective SSE value. The result is shown here:

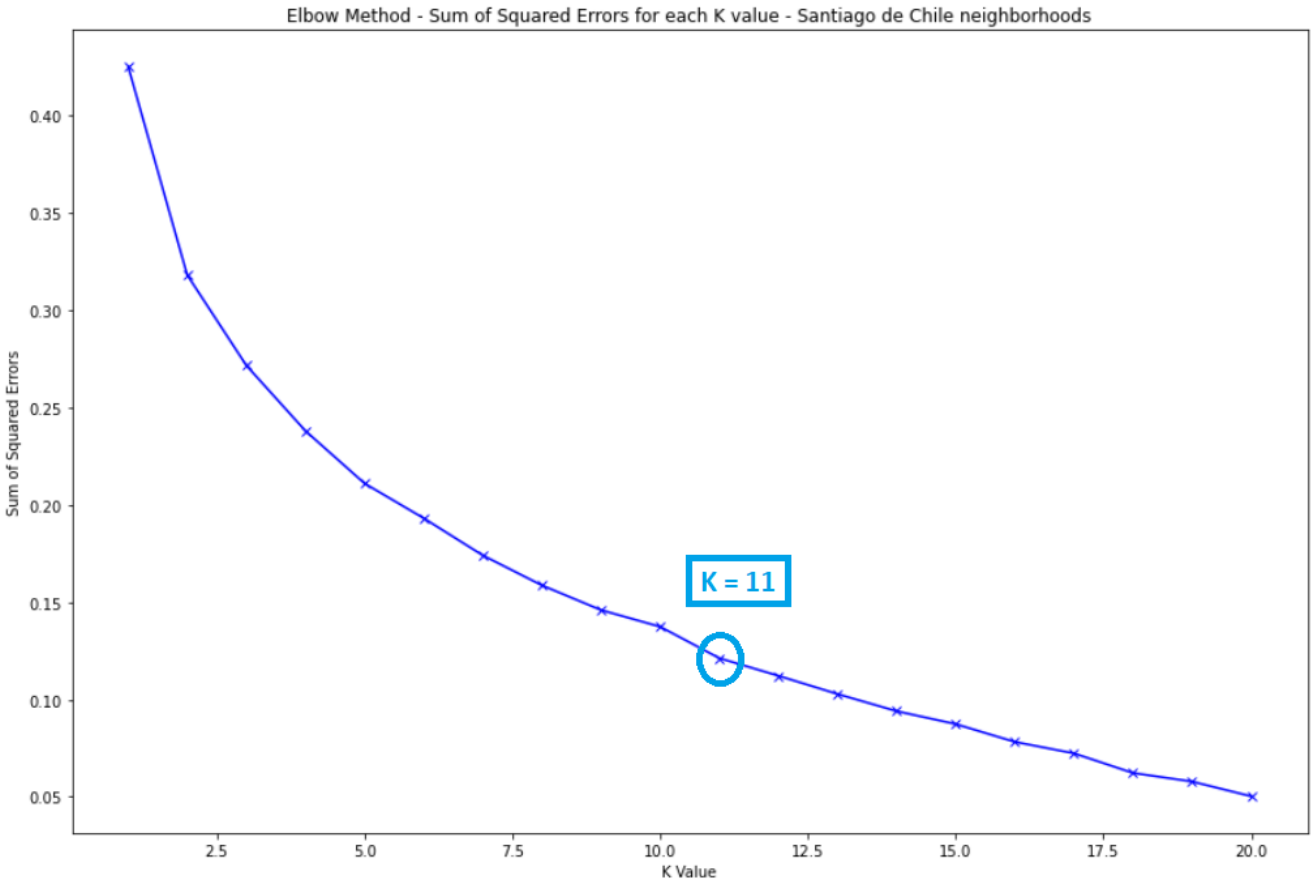

Figure 9: Elbow Method – SSE for each K value – Santiago neighborhoods. Source: Own Elaboration.

As the rate of SSE decreases after **K=11, this is our optimal K value**.

Next, as **K represents the number of neighborhood clusters** in Santiago de Chile, we plot the 11 clusters in the following **Folium map:**

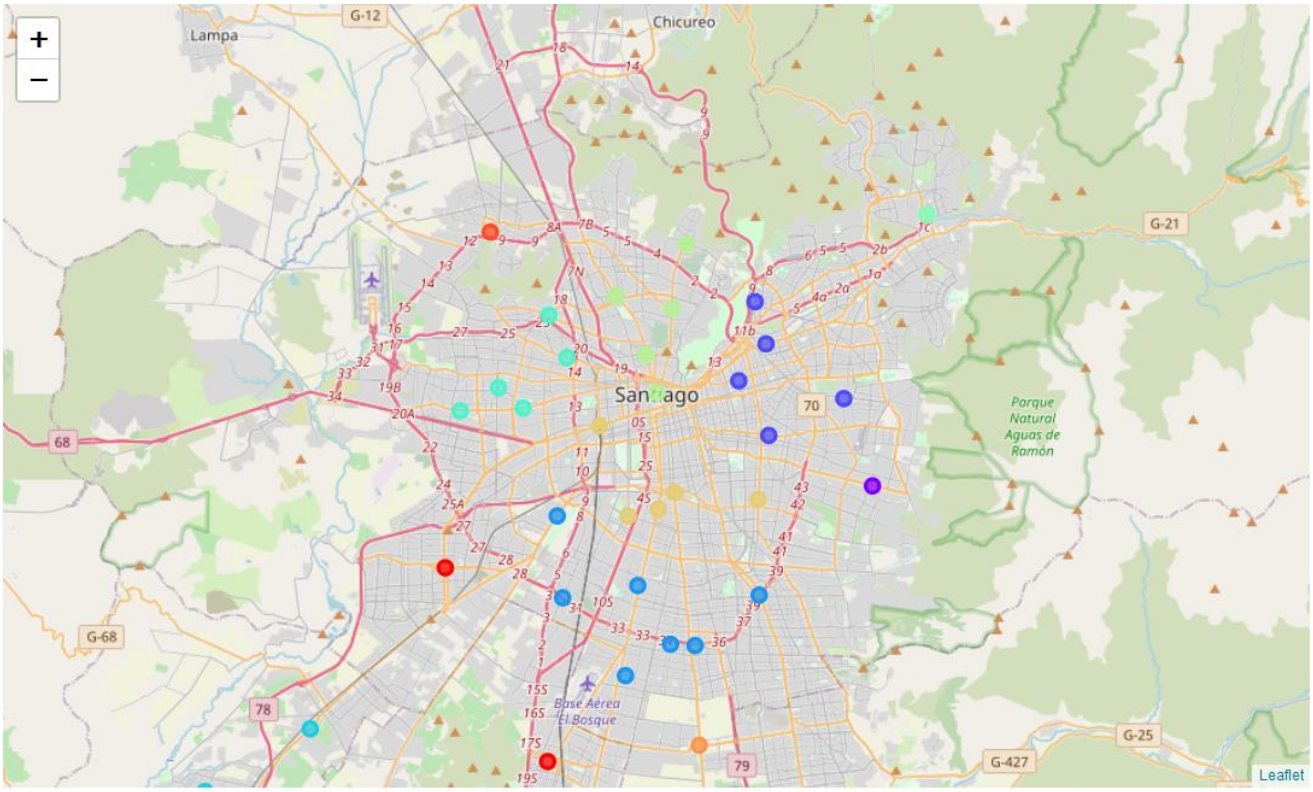

Figure 10: Folium map of clustered Santiago neighborhoods. Source: Own Elaboration.

5.4 Results: Segmenting and Labeling Santiago's neighborhoods

Next, we **segment and label Santiago de Chile's touristic neighborhoods according to their clusters and name each one of them**. This will be done with respect to their most common venues.

As we are interested in which are the touristic neighborhoods from Santiago de Chile to boost the tourism on that city, we only focus on the clusters which contain those neighborhoods. This decision was

taken considering the same factors as Lima: connectivity of the neighborhoods, variety of venues to visit and specially their security.

Considering these constrains, from the 11 clusters which divide Santiago's neighborhoods, we have only **4 clusters considered touristic (3 if we consider that clusters 1, 6 are merged into one due to their similarity)**. The next table, summarize these clusters with respect to their labels, neighborhoods and their main characteristics:

N° of cluster	Cluster's Label	Neighborhoods
1, 6	Residential and outdoor in Andes' foothills	Lo Barnechea, Peñalolén
2	Gastronomic, financial and green areas in Northeast Santiago	La Reina, Las Condes, Ñuñoa, Providencia, Vitacura
7	Historical, commercial, bohemian and gastronomic cluster in Downtown and North Santiago	Conchalí, Huechuraba, Independencia, Recoleta, Santiago

Table 5: Segmenting and labeling of touristic neighborhoods in Santiago.

5.5 Discussion: Recommendations to improve Santiago's tourism, after COVID-19 pandemic

The background and recommendations for touristic clusters in Santiago de Chile, to recover faster post COVID-19, are the following:

In the foothills of Peñalolén and Lo Barnechea neighborhoods, are located **extensive natural parks where tourists can go on long walks, practice outdoor sports and even visit shrines such as the Bahai Temple** [28]. But accessing these parks is very difficult by cars or even buses, due to the steep slope of the roads that exist to get there. Hence, one solution for Chilean government to attract more tourists to the Andes foothills is to **build cableways that go from these neighborhoods to the natural parks mentioned**. By that way, more Chileans and tourists could visit that parks and a more sustainable tourism would be developed by helping to reduce air pollution in Santiago de Chile. If to the above **we add more publicity by tourist agencies**, tourism would develop much more in these neighborhoods.

At the Northeast sector of Santiago de Chile, **their neighborhoods enjoy some of the best quality of life indices in Chile** [29] **and are visited by thousands of tourists every year**. Huge parks, tall and modern buildings and a lot of gastronomic venues are some of the most common venues there. But this development has had consequences on the road infrastructure of the communes. **Roads and Subway (Metro) congestion is critical in rush hours** [30] and travel times in this cluster are delayed a lot. Therefore, as in Lima, one practical solution is that the government can **create extensive bicycle lanes and encourage the use of bicycles to get around these neighborhoods**; this would also help to improve the poor air quality of Santiago de Chile and improve their citizens' quality of life.

And in the third Cluster (7), are located the **historical, bohemian, gastronomic and commercial neighborhood of Santiago de Chile**. These zones are some of the most visited ones in Santiago and Chile: tourists are delighted to know the history of Santiago and its historical buildings, such as **La Moneda (Government Palace) or National Museum of Fine Arts** [31]. Not only that, Chileans and foreigners **also love their wide gastronomic offer (from Peruvian to Italian food), go to shopping at their numerous commercial venues and go to their scenic lookouts to appreciate Andes mountain range and the city**. But as in Lima, Santiago's downtown has a serious crime and insecurity problem, especially late at night [32]. Therefore, once again, the recommendation to Chilean government is to **strengthen police control and through security cameras to reduce robberies and the level of insecurity in the historic Cluster of Santiago de Chile**.

6. Re-clustering and re-segmentation for touristic neighborhoods in Lima and Santiago de Chile

6.1 Data cleaning and understanding

Before we re-cluster and analyze the similarity between touristic neighborhoods from Lima and Santiago de Chile, we start by converting to Pandas data frames, **the .csv data sets** containing the **neighborhoods from Lima and Santiago, and its respective venues retrieved by Foursquare**. These sets are retrieved from the ones we used for the previous analysis of the neighborhoods from both cities.

To do a joint cluster analysis, first we filter the data to keep only the **names and coordinates of the touristic neighborhoods**. Then, we merge both data frames into one, where we get that there are **38 touristic zones in Lima and Santiago**.

On the other hand, we do another filter to the venues data sets from the two Latin American capital cities, keeping only the **venues belonging to touristic neighborhoods**. After we do this and merge the data frames, we get that **in these 38 touristic zones, there are 2996 venues in total, with 229 unique venues categories**.

6.2 Exploratory Data Analysis

As in the previous analyses done for each city separately, we perform **One Hot Encoding to separate each venue category into a column** in a new data frame. But in this case, we **only do this with the touristic neighborhoods** from Lima and Santiago de Chile.

Next, we create another data frame by **grouping the different venues by Neighborhood Name column** and then **take the relative frequency** of occurrence of **each category**.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CALLAO	-12.051558	-77.134117	Obelisco	-12.052351	-77.134222	Plaza
1	CALLAO	-12.051558	-77.134117	Francesco	-12.058225	-77.130595	Seafood Restaurant
2	CALLAO	-12.051558	-77.134117	Panadería Olcese	-12.061041	-77.143544	Bakery
3	CALLAO	-12.051558	-77.134117	Chifa El Osito	-12.059813	-77.135372	Chinese Restaurant
4	CALLAO	-12.051558	-77.134117	CasaCor Callao	-12.060004	-77.147282	Public Art
2991	Vitacura	-33.398990	-70.601010	Club de Planeadores de Santiago	-33.379745	-70.578790	Airport
2992	Vitacura	-33.398990	-70.601010	Parque Casa Piedra	-33.384268	-70.590440	Park
2993	Vitacura	-33.398990	-70.601010	Sublime - Sandwich & Bar	-33.385422	-70.555144	Sandwich Place
2994	Vitacura	-33.398990	-70.601010	Santiago Papperchase	-33.386364	-70.610017	Racetrack
2995	Vitacura	-33.398990	-70.601010	Plaza Inés de Suárez	-33.438078	-70.611917	Playground

Figure 11: Data frame containing venues and neighborhood from touristic zones of Lima and Santiago.
Source: Own Elaboration.

Then, we build a bar plot which shows the 20 most common venues, across all the touristic neighborhoods in Lima and Santiago de Chile:

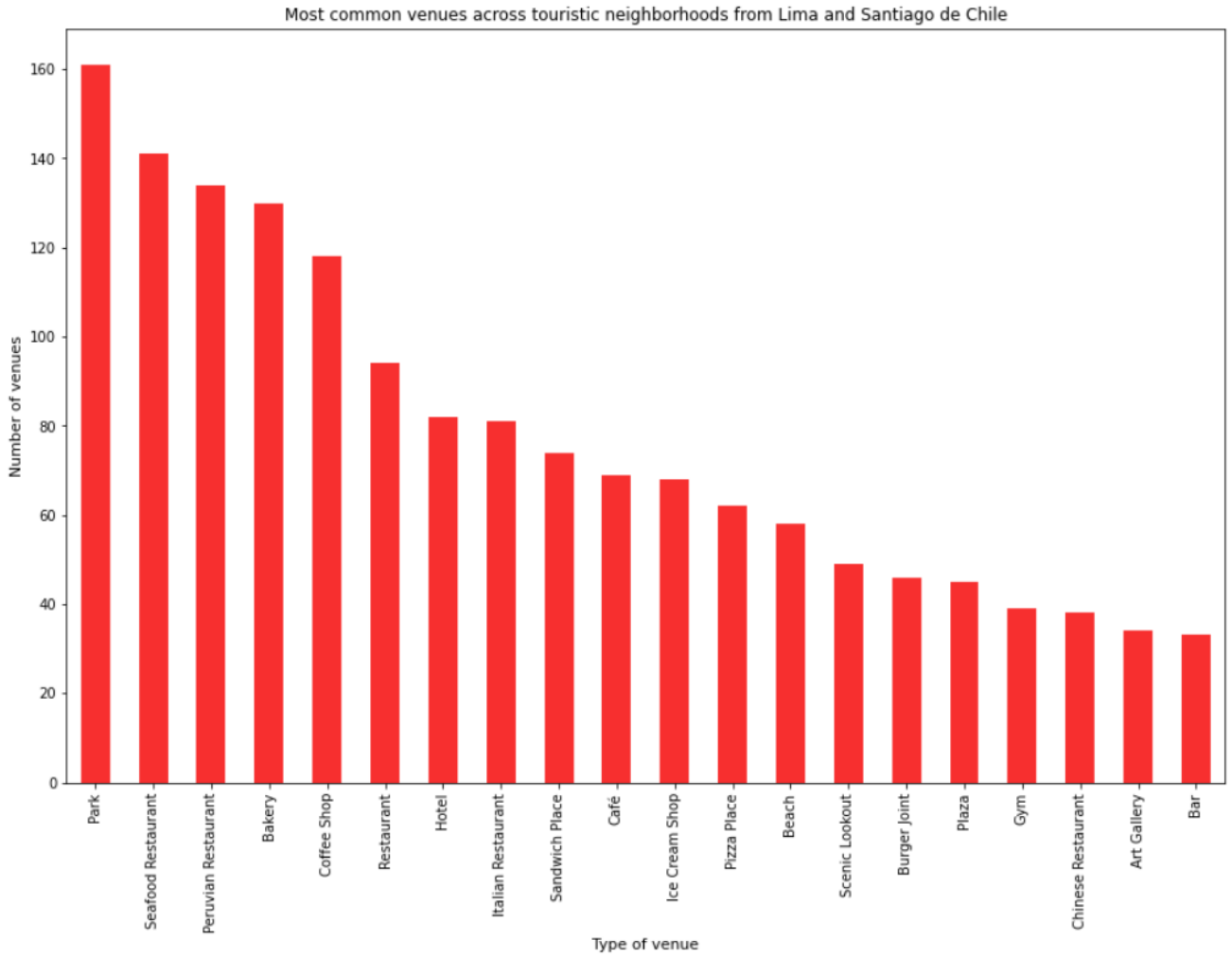

Figure 12: Bar plot with top 20 most common venues from Lima and Santiago. Source: Own Elaboration.

From this bar chart, we infer that in these touristic neighborhoods there is a prevalence of Parks, Restaurants, Bakeries and Coffee Shops.

First, in Lima and Santiago de Chile, people love to **visit outdoor places specially parks, which are the biggest open places in the big cities**. The reason of this choice, is mainly to escape from the stress of the city, practice sports, share and enjoy a picnic with family, friends, take the pet for a walk or just relax lying on the grass.

On the other hand, inhabitants and tourists of both capital cities **enjoy eating national and international food from the best restaurants in Lima and Santiago de Chile**. Here we notice the influence of **Seafoods and Peruvian gastronomy** on Chileans and Peruvians.

The first one is explained because **both are coastal countries and base an important part of their gastronomic culture on the consumption of seafood**. The influence of Peruvian gastronomy is due to:

1) Its cuisine is considered **one of the best in the world**. In 2019, Peru has awarded as **World's Leading Culinary Destination** [33].

2) Is **one of the main Peruvian export products**; in Lima, people are deeply rooted in their culture and that has **led many Peruvian chefs to open many restaurants of their local food**. Chile has received **hundreds of thousands of Peruvian immigrants** [34], who have installed **restaurants of their local cuisine** in that country.

Finally, because **in Chile specially people love to eat bread** [35], and in both countries people often go to drink coffee while enjoying the company of friends and family, that explains the **popularity of Bakeries and Coffee Shops** which are **among the most popular venues in Lima and Santiago de Chile**.

Now, we will compare these results with the **most common venues from Lima and Santiago de Chile separately, only considering their touristic neighborhoods**:

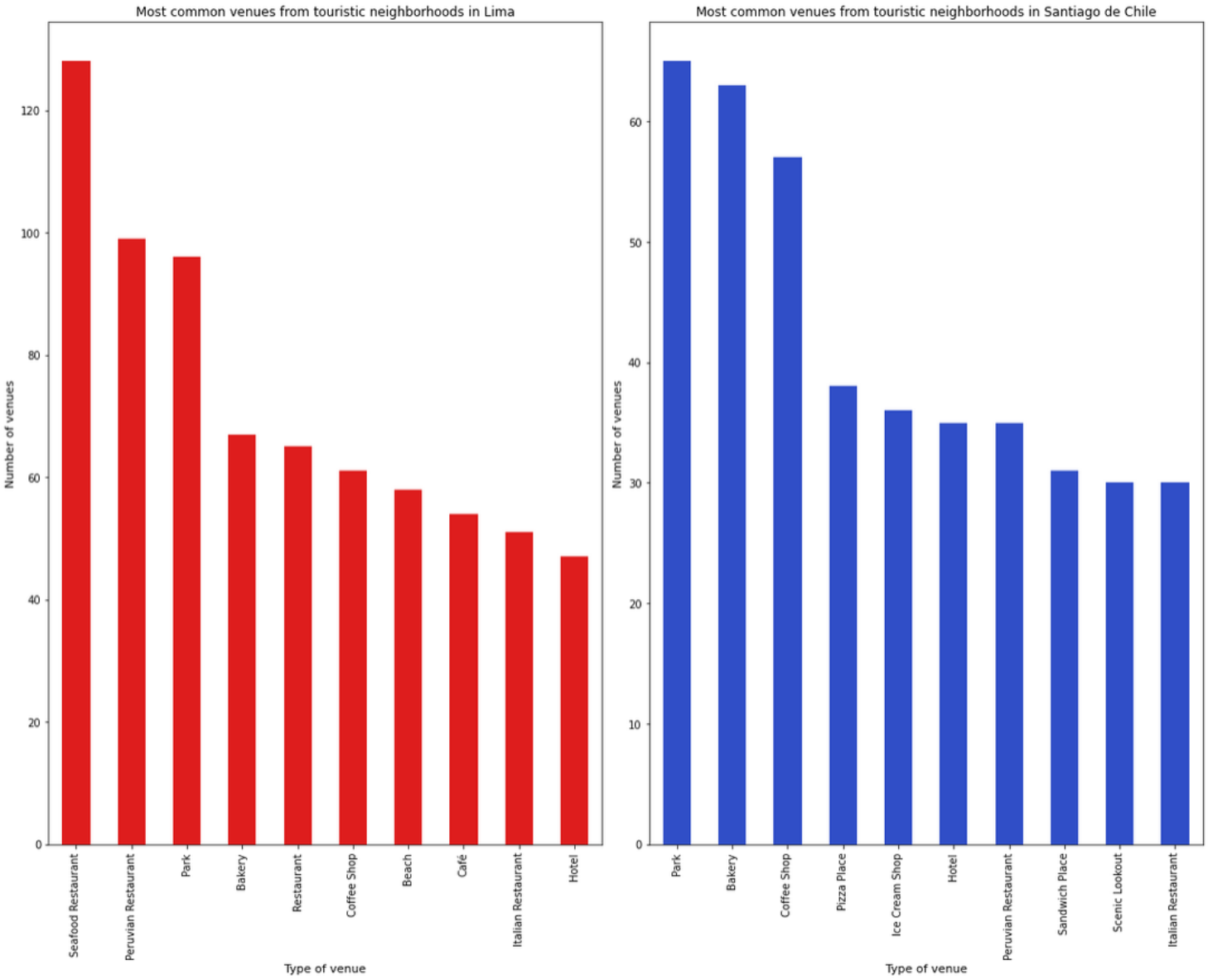

Figure 13: Bar plots with top 10 common venues from Lima and Santiago. Source: Own Elaboration.

We **notice similarities in some of the 10 most common places visited**: **Parks, Bakeries, Coffee Shops, Hotels, Peruvian and Italian Restaurants**. This would indicate us that **several neighborhoods between these cities will be grouped in the same clusters**. When

neighborhoods are grouped and segmented *by running the k-Means clustering algorithm, this hypothesis will be tested*. But it's needed to **highlight that the number and order of each venue category are different**, so this *hypothesis may not be correct*.

Notice that there are also a couple of different venues, **which characterize the neighborhoods of each city separately**. In the case of **Lima**, we appreciate the influence of its proximity to the Pacific Ocean, because *some of the most common and distinct venues* with respect to Santiago de Chile are: **Seafood Restaurant and Beach**. Like every coastal city, **Lima** has taken advantage of the **harvest of Seafood products, the proximity to its beaches and development of attractions around them**, to show the world the wealth of its natural resources in favor of local and foreign tourism.

Meanwhile in **Santiago de Chile**, as this city is located in the Central Valley surrounded by hills and mountains, different governments **have promoted the construction of Scenic Lookouts**: such as the ones in *Santa Lucía or San Cristóbal Hills*, to take advantage of **panoramic views from the heights of its mountains to the city**. These places have become some of the most favorite ones on Santiago de Chile.

On the other hand, as Chile is the second country with highest bread consumption **[35]**, local people and tourists love to visit **Sandwich Places** and eat one of the hundreds of sandwiches varieties, whose offer grows thanks to new creations by chefs and new ingredients brought from immigrants. Finally, as the temperatures in Santiago de Chile are very high, especially during Spring and Summer **[36]**, **locals and tourists love to buy an Ice Cream at one of their many shops** on these neighborhoods, to escape from the heat of this warm city.

6.3 Modeling with k-Means Clustering Algorithm

In this part, we will cluster and segment touristic neighborhoods in Lima and Santiago de Chile into different groups, according to the results of the **k-Means Clustering Model**.

As it has been done and explained previously in this Capstone Project, we will run **k-Means** 20 times (K from 1 to 20), to determine the **optimal K according to the Elbow method**.

When we plot these the SSE (sum of squared errors) values, with respect to **each K** to find its optimal value, we get the following figure:

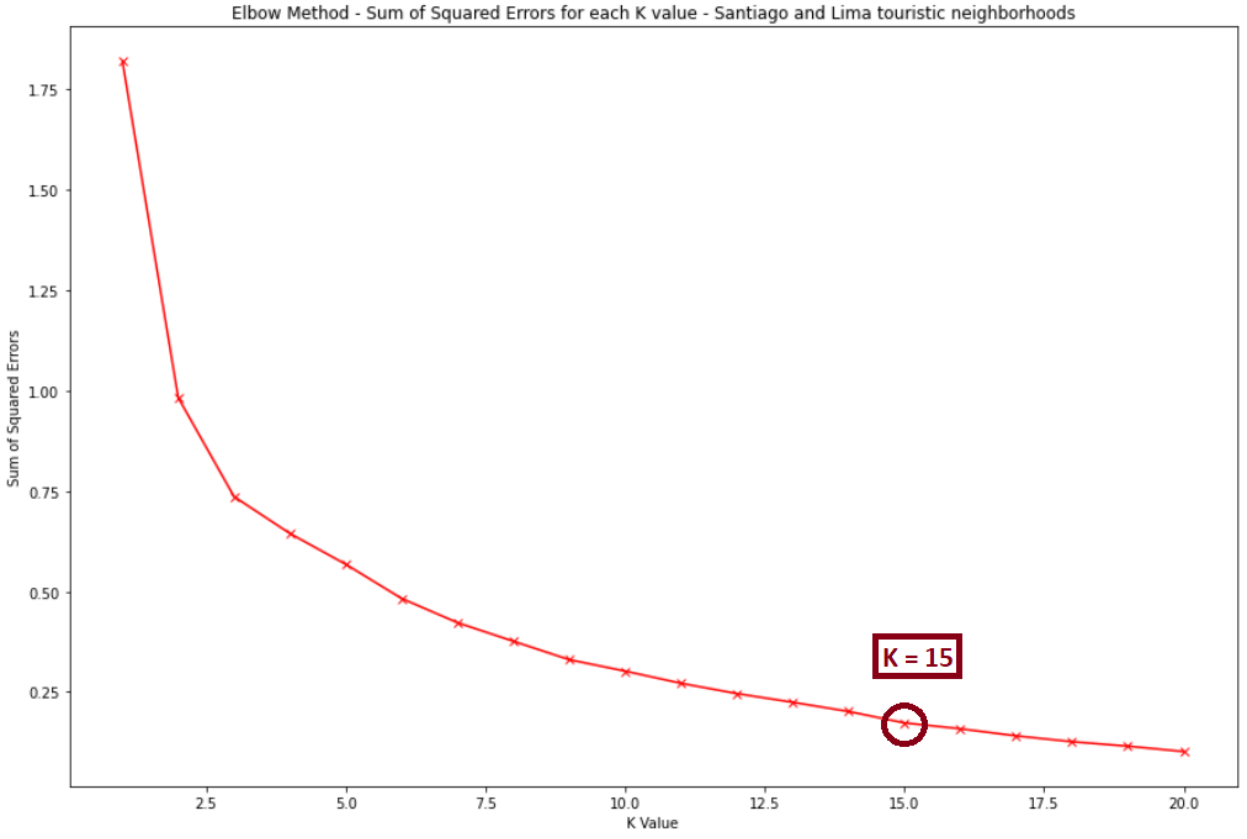

Figure 14: Elbow Method – SSE for each K value – Santiago and Lima neighborhoods. Source: Own Elaboration.

As the rate of SSE decreases after **K=15, this is our optimal K value**. Next, as **K represents the number of neighborhood clusters** in Lima and Santiago, we plot the 15 clusters in the next **Folium maps:**

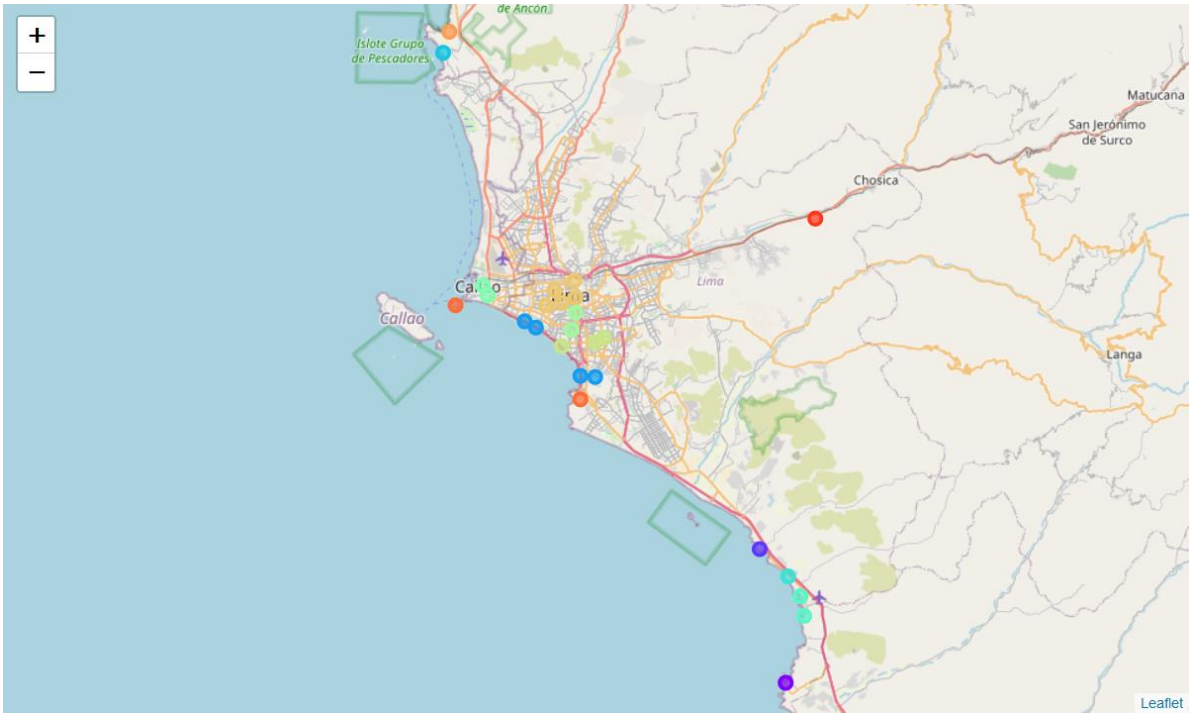

Figure 15: Folium map of clustered Lima neighborhoods for joint analysis. Source: Own Elaboration.

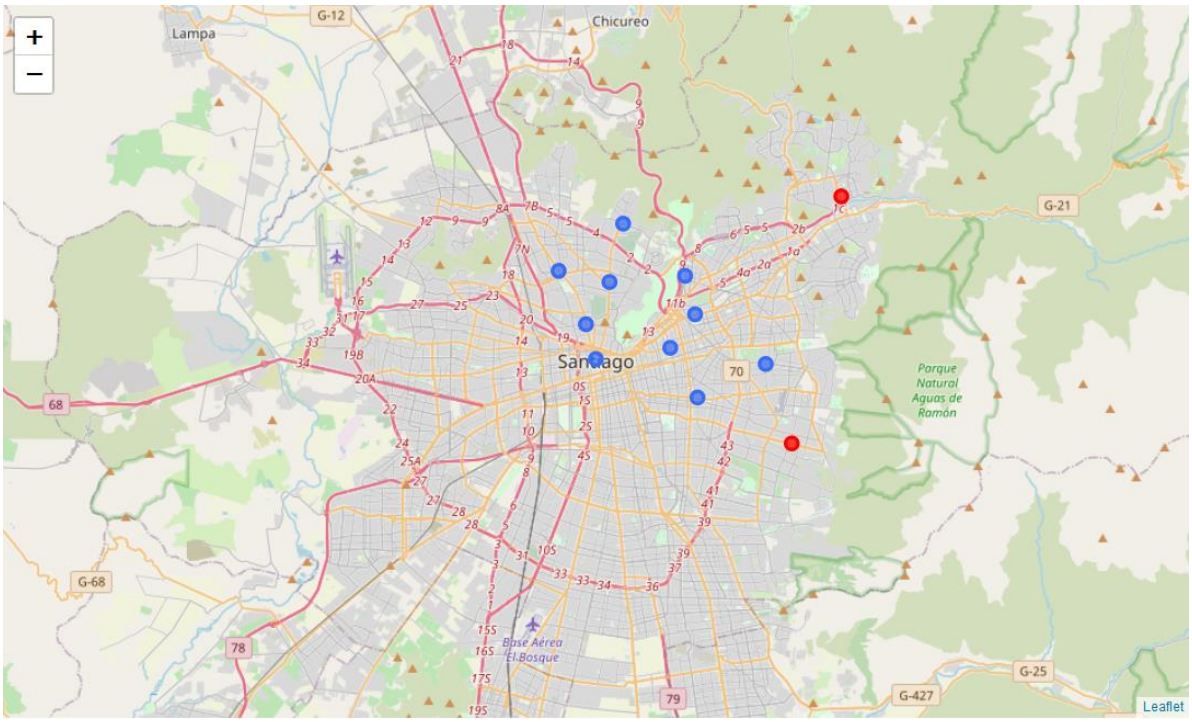

Figure 16: Folium map of clustered Santiago neighborhoods for joint analysis. Source: Own Elaboration.

Opposite to our hypothesis, the **touristic neighborhoods from Lima and Santiago de Chile are dissimilar between them**, because **there are not similar neighborhoods from different cities clustered**. We will **show the common venues per cluster & analyze the differences**.

6.4 Results: Re-segmenting and re-labeling Lima and Santiago's touristic neighborhoods

To show the causes, we **compare the venues categories which exist in more neighborhoods**, considering Lima and Santiago separately.

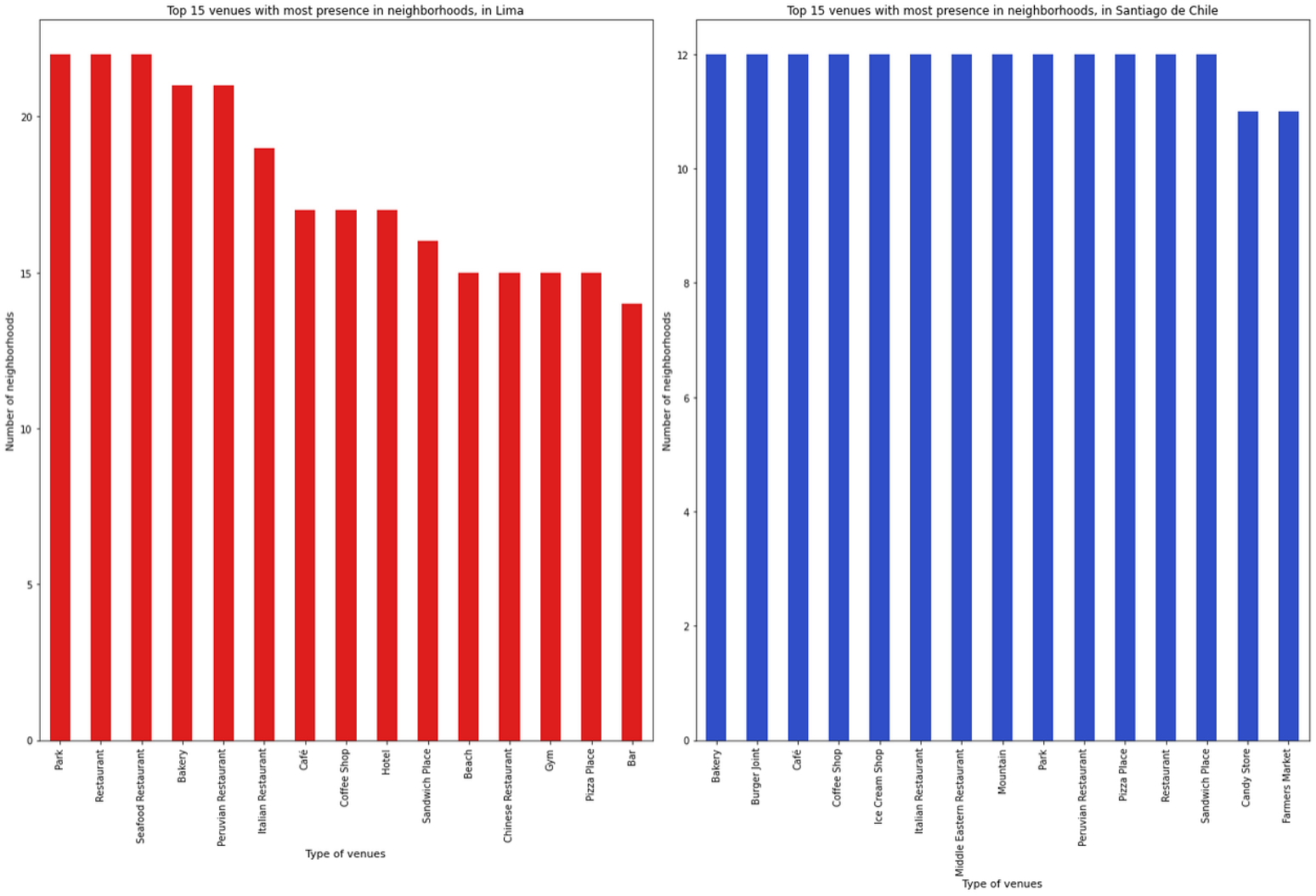

Figure 17: Bar plots showing top 15 venues with most presences in touristic neighborhoods from Lima and Santiago. Source: Own Elaboration.

From these two bar plots, we **notice important differences, in the frequency and distribution, of the 15 venues with most presences in touristic neighborhoods** of Lima and Santiago de Chile. These are explained by the following factors:

Top 15 venues with most presence in neighborhoods, in Lima

- In the case of touristic neighborhoods in Lima, we notice an important presence of local restaurants (Peruvian, Seafood and Chinese), which indicates us that inhabitants and tourists love local gastronomy and that explains its important presence.
- On the other hand, when people visit Lima they enjoy hanging with friends or family at their beaches, taking advantage of the extensive Lima coastline. There is also an important presence of Bars, where tourists love to take local and foreign drinks (including the famous Peruvian **Pisco Sour**), as one of the venues with most presences in touristic neighborhoods.
- Thirdly, in Lima's touristic neighborhoods there is a more important presence of Gyms, with presence in 15 zones. This would indicate that there is a higher demand and offer of places to train and practice indoor sports.
- Finally, as Lima is a much bigger city with more districts than Santiago de Chile, it has a bigger hotel offer available to receive much more tourists from Peru and all over the world. Of course, taking advantage of their tourist attraction, the zones we are analyzing concentrate the biggest tourist demand in Lima and Callao.

Top 15 venues with most presence in neighborhoods, in Santiago de Chile

- Opposite to Lima, as we explained previously, Santiago de Chile and its inhabitants are highly influenced by foreign cultures, and this includes the cuisine too. We can appreciate this by noticing that the restaurants with most presences in touristic zones are: Burger Joints, Middle Eastern, Peruvian, Italian, Pizza and Sandwich places. This foreign influence can also be seen in the important presence of Candy Stores, which are present in 11 touristic neighborhoods in the city.

- Secondly, as Santiago de Chile is located in Chile's central valley, people can't enjoy the sea but the mountains. Locals and tourists love to practice outdoor sports near these places (e.g. hiking), and appreciate from the mountain ranges, the city from an aerial view.
- It is also important to remark that, due to the high temperatures in Santiago de Chile (specially in Spring and Summer [4]), people in the city love to enjoy Ice Creams and go to the shops to eat one. This explains the popularity of the Ice Cream shops and, by that way, people from Santiago de Chile and tourists can cool off during the hot days that are very frequent in this city.
- As a last point, Farmers Market are also one of the venues with most presence in Santiago's touristic neighborhoods. This is explained from the old tradition of Chileans, dating from the time when Chile was a Spanish Colony, to buy fruits and vegetables to these markets **[37]**.

Next, we **segment Santiago de Chile's and Lima's touristic neighborhoods according to the re-cluster performed: label and name each one of them**. This is done with respect to their most common venues. Now, we will **analyze the differences between each separated analysis** from each city and the **joint analysis of both cities**.

According to k-Means Clustering Model, we have **15 distinct clusters (13 if we considered that clusters 1, 5 and 6 were merged, due to their similarity)**. The clusters, their labels and neighborhoods included in each one, are shown in the next table:

N° of cluster	Cluster's Label	City	Neighborhoods
0	Residential and outdoor in Andes' foothills	Santiago de Chile	Lo Barnechea, Peñalolén
3	Gastronomic, financial and green areas, merged with historical commercial and bohemian cluster (North, Downtown and Northeast zones)	Santiago de Chile	Conchalí, Huechuraba, Independencia, La Reina, Las Condes, Ñuñoa, Providencia, Recoleta, Santiago, Vitacura
1, 5, 6	Touristic beaches in Northern and Southern zones	Lima	Pucusana, Punta Negra, Santa Rosa
2	Punta Hermosa coastal zone	Lima	Punta Hermosa
4	Gastronomic and green areas	Lima	Barranco, Magdalena del Mar, San Miguel, Santiago de Surco
7	Touristic beaches, outdoor sports and services cluster, in Southern zone	Lima	San Bartolo, Santa María del Mar
8	Callao centric zone	Lima	Callao, Bellavista
9	Financial and commercial cluster in Metropolitan zone	Lima	Lince, San Isidro
10	Touristic, green areas and outdoor sport in Metropolitan zone	Lima	Miraflores, San Francisco de Borja, Surquillo
11	Historic, cultural and gastronomic cluster in Downtown zone	Lima	Breña, Lima, Jesus María, La Victoria, Pueblo Libre
12	Touristic beach and small coastal town in Northern zone	Lima	Ancón
13	Coastal, gastronomic and green areas from Callao and Lima	Lima	Chorrillos, La Punta
14	Gastronomic and green areas, at the countryside area of Chaclacayo	Lima	Chaclacayo

Table 6: Segmenting and labeling of touristic neighborhoods in Lima and Santiago.

Here we can see that there was an **important rearrange of clusters**, comparing the analysis of **each separated city, versus the joint analysis** of touristic neighborhoods. It jumps out the fact that in the **joint analysis, Lima's touristic clusters were much more segregated and well defined now**.

This is shown in the **increased number of clusters (from 6 to 13), and how close neighborhoods were better and more precisely separated**; for example, in Downtown Lima and coastal neighborhoods. According to this information, we would think that k-Means clustering algorithm performed much better on this data.

But if we analyze the **touristic clusters from Santiago de Chile, its number was reduced from 3 to 2**. Here we highlight the fact that the **financial and gastronomic cluster** was **merged with the historical, commercial and bohemian** one. Considering that these neighborhoods are very different, we can think that **k-Means performed worse in Santiago de Chile**. But why?

Because it is possible that **k-Means algorithm arrived to a suboptimal result**, due to:

1) **Lack of venues data to perform a correctly clustering** of the touristic neighborhoods, in Lima and Santiago de Chile.

2) **Elbow method and optimal K** value chosen, led to just a **local optimal clustering** [38].

3) Because **k-Means algorithm can't separate the points according to their density**, and due to the fact that it is **sensitive to outliers**, k-Means itself **would not be** the recommended **algorithm to tackle problems involving geospatial data**.

A good advice to get better results is to **perform a deeper study on this project**, by performing the same analysis **using DBSCAN clustering** due to the fact that it can **find arbitrarily shaped clusters**. Also, because can **prevent joining outlier points in previously optimal clusters**. And finally, due to the fact that you **do not need to precompute K** to run the algorithm [39]. The following illustrative figure, show the difference by clustering with k-Means vs DBSCAN:

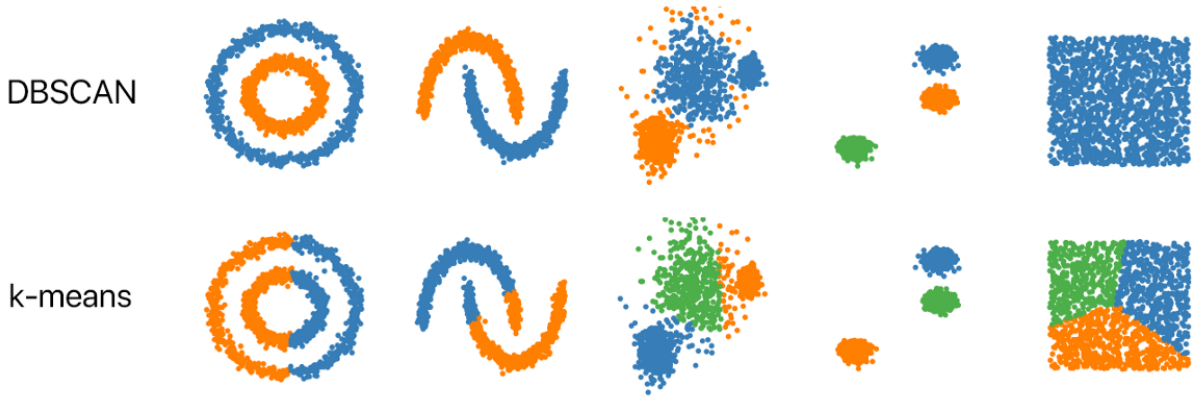

Figure 18: Comparison of k-Means vs DBSCAN clustering models. Source: Towards Data Science.

6.5. Discussion: Recommendations to improve Santiago and Lima cooperative tourism, after COVID-19 pandemic

From the analysis, we infer that **Lima and Santiago de Chile touristic neighborhoods are very dissimilar** between each other.

This **heterogeneity** between both cities, opens a **big opportunity to promote a consolidate alliances for a cooperative tourism**, that can help both cities to recover their Tourism and Economies.

Some recommendations for touristic agencies and governments from Chile and Peru, are shown below:

1) Chilean and Peruvian governments can sign a **cooperative tourism treaty**. This must address the **creation of cooperative guilds of touristic agencies, advertising** of Lima and Santiago de Chile in **mass communication media**, to promote touristic venues and neighborhoods. And also, **promote around the world, create a web page and do advertising in social media**.

2) Create a **Chilean-Peruvian tourism organization**, to create a **common tourism fund** and establish an **organization to direct cooperative tourism projects**. Its objective is to **create new projects for: improve infrastructure, accesses and creating new venues** to attract tourists. Is also necessary to **give money to the tourist agencies** that work with tourists (in Lima and Santiago), **finance ads, promotion tourism tours** and **promote online touristic marketing**.

3) Security is a topic that worries local and foreign tourists a lot. To **ensure their safety**, is necessary to **invest material and monetary resources to reinforce police control**, with the objective of ensuring tourists safety. And also, **improve the remote surveillance systems**

that can **alert the police, timely and accurately, about criminal acts** in tourist neighborhoods and prevent crimes.

4) Touristic agencies must work in **creating better tourist packages**, where the different goodnesses of Lima and Santiago de Chile are highlighted. Not only this, is also important that **these tourist packages can adjust to different tourists**, too. The idea is to offer packages to Lima, Santiago de Chile or both cities according to each tourist interest. It is also important when selling these packages, to **encourage tourists who only want to visit one city, to visit both**. So, the agencies will **deliver special benefits and offers, for those who visit Lima and Santiago de Chile**: more tourists will visit both places.

5) The work of touristic agencies is key to promote tourism in streets, and attract more tourists to Lima and Santiago de Chile. Thus, is **necessary that these agencies travel the touristic neighborhoods, to discover new venues and make more accurate descriptions** of the important ones. Is also important that they **raise information about new potential tourist neighborhoods, and propose work plans to the governments of Chile and Peru**. Thus, they can **invest on new projects to convert more neighborhoods into touristic ones, helping to generate more visits and tourism revenues**. And also, **to improve life quality of thousands of neighbors from those zones** which today, are far away from a desired development and security level.

7. Conclusions from the Capstone Project

From this project we can conclude the following main points:

Touristic neighborhoods from Lima are very dissimilar from the ones of Santiago de Chile. Although there share several common venues as parks, bakeries, coffee shops and restaurants, the order and frequency of the most common venues differ a lot. In the **first city, there is a predominance of Local Cuisine Restaurants, Gyms, Bars and Beaches**. Lima takes advantage of their coast, Peruvian cuisine and the promotion of Peruvian and international cocktails to promote these venues in touristic neighborhoods. In the case of **Santiago de Chile, is distinguished by their Mountains, Foreign Restaurants, Ice Cream Shops and Farmers Markets**. This can be explained because is a hot city located in Central Valley, with a culture highly influenced by foreign countries, but which still preserves old traditions from the time of the Spanish Colony, such as buying at the Farmer Market.

When we perform the cluster and segmentation of each city separately versus a joint analysis, we notice that the **number of touristic clusters change drastically in Lima and Santiago de Chile**. In the **first city, the number of clusters increase from 6 to 13, and in the second one, this number decreases from 3 to 2**. Here we noticed that, while **in Lima the neighborhoods were more precisely segmented, in Santiago de Chile** the historical and financial clusters were **merged into a single one, giving a poor clustering**. This indicates that **k-Means maybe found a suboptimal solution**. Therefore, it is recommended in a further analysis to **run the DBSCAN clustering algorithm**. Their advantages are: **does not requires K** value to run the algorithm, **prevent joining outlier points** for more optimal clusters, and can **find arbitrarily shaped clusters** (useful for geospatial data).

To **reactivate Chile and Peru economies** after COVID-19 pandemic, several **measures were proposed for tourist agencies and governments**. In the case of Lima, were proposed these solutions: **promote more Lima beaches** on brochures and mass media, **improve road infrastructure**, promote the **use of bike and public transport** in Metropolitan Lima, **strengthen police control** and **upgrade security cameras** to reduce crimes in Downtown Lima. For **Santiago de Chile**, is proposed to **build cableways** and **advertise more natural parks** near Andes foothills, encourage the **use of bikes** at Northeast Sector and also **strengthen police control** to protect tourist from being robbed, to reduce the level of insecurity in Downtown Santiago.

And as last point, **at a joint level**, it is recommended for government and tourists' agencies, to **create cooperative guilds of agencies and sign a cooperative tourism treaty**. This is done **to promote cooperative tourism** by working jointly between Chile and Peru, **deliver bigger and better projects** in infrastructure, publicity, tourist packages and **discover/promote potential touristic neighborhoods**, to **recover the number of tourists** visiting both countries **and their tourism/economic GDP**.

8. Bibliography

[1] Geerts, W. (2018). *Top 100 City Destinations 2018.* London, United Kingdom: Euromonitor International.

[2] World Travel & Tourism Council. (2020, March 1). *Economic Impact Reports*. Retrieved from World Travel & Tourism Council: https://wttc.org/Research/Economic-Impact

[3] Bondarenko, P. (2006, June 9). *Gross Domestic Product – Economics.* Retrieved from Britannica: https://www.britannica.com/topic/gross-domestic-product#info-article-history

[4] New York Times. (2020, August 24). *The Coronavirus Outbreak – Global Death Toll From Virus Surpasses 800,000*. Retrieved from New York Times: https://www.nytimes.com/2020/08/22/world/covid-19-coronavirus.html

[5] Gopinath, G. (2020, April 14). The Great Lockdown: Worst Economic Downturn Since the Great Depression. Retrieved from IMFBlog - International Monetary Fund Blog: https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/

[6] United Nations World Tourist Organization. (2020, July 28). Impact of COVID-19 on Global Tourism Made Clear as UNWTO Counts the Cost of Standstill. Retrieved from UNWTO - World Tourism Organization: https://www.unwto.org/news/impact-of-covid-19-on-global-tourism-made-clear-as-unwto-counts-the-cost-of-standstill

[7] Horton, J. (2020, August 10). Coronavirus: What are the numbers out of Latin America? Retrieved from BBC News: https://www.bbc.com/news/world-latin-america-52711458

[8] Economic Commission for Latin America and the Caribbean (ECLAC) - United Nations. (2020). COVID-19 Reports. Santiago de Chile: Economic Commission for Latin America and the Caribbean (ECLAC).

[9] The Editors of Encyclopaedia Britannica. (1998, July 20). Metropolitan Area - Demography. Retrieved from Britannica: https://www.britannica.com/topic/metropolitan-area

[10] Wikipedia. (2020, July 18). *Lima metropolitan area*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Lima_metropolitan_area

[11] OTI - Ministerio de Transportes y Telecomunicaciones Perú. (2018, March 23). *MTC - Centros Poblados*. Retrieved from Plataforma Nacional de Datos Abiertos - Gobierno de Perú: https://www.datosabiertos.gob.pe/dataset/mtc-centros-poblados

[12] Wikipedia. (2019, December 29). *Santiago Province, Chile*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Santiago_Province,_Chile

[13] Wikipedia. (2020, August 22). *Anexo:Comunas de Santiago de Chile*. Retrieved from Wikipedia: https://es.wikipedia.org/wiki/Anexo:Comunas_de_Santiago_de_Chile

[14] ArcGIS for Developers. (2020, January 01). *Geocoding with ArcGIS*. Retrieved from ArcGIS for Developers: https://developers.arcgis.com/features/geocoding

[15] Foursquare. (2020, January 1). *Foursquare Developers - Endpoints*. Retrieved from Foursquare: https://developer.foursquare.com/docs/places-api/endpoints/

[16] Rollins, J. (2015, August 24). *Blogs - Why we need a methodology for Data Science*. Retrieved from IBM Big Data and Analytics Hub: https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science

[17] Wikipedia (2020, September 14). *Chinese Peruvians*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Chinese_Peruvians

[18] Wikipedia (2020, September 7). *Peruvian Cuisine*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Peruvian_cuisine

[19] Robert Block's Glove (2017, December 26). *Using the elbow method to determine the optimal number of clusters for k-means clustering*. Retrieved from Wikipedia: https://bl.ocks.org/rpgove/0060ff3b656618e9136b

[20] Peru Telegraph (2019, June 5). *Lima – the third worst congested city in the world*. Retrieved from Peru Telegraph: https://www.perutelegraph.com/news/peruvian-economy/lima-the-third-worst-congested-city-in-the-world

[21] Wikipedia (2020, May 19). *Historic Centre of Lima*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Historic_Centre_of_Lima

[22] Wikipedia (2019, December 25). *Real Felipe Fortress*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Real_Felipe_Fortress

[23] TeleSUR (2018, September 2). *Peru: Insecurities Rise in Lima Amid High Crime Rates*. Retrieved from TeleSUR: https://www.telesurenglish.net/news/Peru-Insecurities-Rise-in-Lima-Amid-High-Crime-Rates-20180902-0007.html

[24] ArcGIS for Developers (2020, January 1). *ArcGIS API for Python*. Retrieved from ArcGIS for Developers: https://developers.arcgis.com/python/guide/using-the-geocode-function/

[25] Wikipedia (2020, September 16). *Immigration to Chile*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Immigration_to_Chile

[26] EFE (2014, August 5). *La comida asiática ocupa el mayor número de restaurantes en Santiago*. Retrieved from El Mostrador: https://www.elmostrador.cl/cultura/2014/08/05/la-comida-asiatica-ocupa-el-mayor-numero-de-restaurantes-en-santiago-de-chile/

[27] Chile Travel (2019, July 30). *Chile, The Sandwich Paradise*. Retrieved from Chile Travel: https://chile.travel/en/chile-the-sandwich-paradise

[28] Bahá'í Temple (2018, January 1). *Bahá'í House of Worship of South America*. Retrieved from Bahá'í Temple: https://templo.bahai.cl/en/

[29] Cámara Chilena de la Construcción (2018, May 7). *Índice de Calidad de Vida Urbana (ICVU) 2018*. Retrieved from Cámara Chilena de la Construcción: https://www.cchc.cl/comunicaciones/noticias/indice-de-calidad-de-vida-urbana-icvu-2018

[30] Alanna Petroff (2017, February 20). *The 15 worst cities for rush hour traffic*. Retrieved from CNN Business: https://money.cnn.com/2017/02/20/autos/traffic-rush-hour-cities/index.html

[31] Beth Delthony (2014, April 7). *30 Things To Do When Visiting Santiago, Chile*. Retrieved from Santiago Tourist: http://santiagotourist.com/30-things-to-do-when-visiting-santiago-chile/

[32] Nina Hobson (2018, April 4). *Is Santiago de Chile Safe?* Retrieved from The Expater: https://theexpater.com/2018/04/04/is-santiago-de-chile-safe/

[33] World Travel Awards (2019, January 1). *World's Leading Culinary Destination 2019*. Retrieved from World Travel Awards: https://www.worldtravelawards.com/award-worlds-leading-culinary-destination-2019

[34] Wikipedia (2020, September 16). *Immigration to Chile*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Immigration_to_Chile

[35] Chile Travel (2019, July 30). *Chile, The Sandwich Paradise*. Retrieved from Chile Travel: https://chile.travel/en/chile-the-sandwich-paradise

[36] Holiday Weather (2020, January 1). *Santiago, Chile: Annual Weather Averages*. Retrieved from Holiday Weather: https://www.holiday-weather.com/santiago/averages/

[37] Eduardo Cavieres F. (2011, January 31). *Mercados y comercio informal en el Chile de la transición de colonia a república*. Retrieved from OpenEdition Journals – Nuevo Mundo Mundos Nuevos: https://journals.openedition.org/nuevomundo/60545

[38] Khyati Mahendru (2019, June 17). *How to Determine the Optimal K for K-Means?* Retrieved from Medium – Analytics Vidhya: https://medium.com/analytics-vidhya/how-to-determine-the-optimal-k-for-k-means-708505d204eb

[39] Fisseha Berhane (2015, January 1). *Data distributions where Kmeans clustering fails – Can DBSCAN be a solution?* Retrieved from Fisseha Berhane's webpage (Data Science Enthusiast): https://datascience-enthusiast.com/Python/DBSCAN_Kmeans.html