Investment in the best business opportunity

Italo Vila Arévalo

September 08, 2020

1. Introduction

1.1. Background

Within the investment world, the type of business to be bet on always goes hand in hand with key aspects such as the strength of demand, the concept of the brand, innovation and a good cost and budget plan. In the South American region, the gastronomic and tourist capital par excellence is Peru, this is demonstrated by being awarded by The World Travel Awards for the eighth consecutive year as "Best culinary destination in the world" in addition to having also been awarded as "Best cultural destination in the world".

Tourism in Peru is the third largest industry in the nation, behind fishing and mining. It is mainly directed towards archaeological monuments, as it has more than one hundred thousand sites of interest, ecotourism in the Peruvian Amazon, cultural tourism in colonial cities, gastronomic tourism, adventure tourism and beach tourism. (Wikipedia, s.f.)

Tourism is of vital importance because it gives people historical facts through cultural, social, gastronomic experiences, among others, to entertain the stay in the country where tourism is being carried out, whether native or foreign.

1.2. Business Problem

In this capstone project we will try to provide a viable and profitable solution for a favorable touristic business in a very popular and tourist section through statistics of preferences, clustering, etc. In this context, the investor needs to know accurate data in which the decision he makes is based on the assertiveness of the treatment of this data. It is for these reasons that through this capstone we will try to give a compass or a north on investment in tourism based on the distribution of the business categories and their demands.

2. Data acquisition and cleaning

2.1. Data Sources

For this project, it will use 4 main sources to obtain information that will allow its manipulation and subsequent analysis, which are:

- **Foursquare API**: The pertinent consults will be made to determine the relevant venues to identify the nearest venues around of specific location, places, ratings from customer and clients and venues categories. (Foursquare, s.f.)

- **Google Maps API**: It Allow to make requests about the coordinates, places, and specific routes (Google, s.f.).

- **INEI:** The INEI is the National Institute of Statistics and Informatics for its acronym in Spanish that it can extract information about economics, social and cultures topics that help to understand the actual context.

- **Websites:** Different web pages will give a vision from another perspective in favor of the contribution of the capstone.

2.2. Data cleaning

The first step is to justify the development of the economy in the relevant sector such as tourism. For this reason, the main tourist indicators such as the departure or entry of tourists (national and foreign) and income versus income per capita in the INEI were consulted. The relevant file was downloaded and read for processing.

The downloaded file was an excel file, some labels were modified because there were incomplete cells and the column labels were overlapping. After these modifications, the file was read and processed on python in a Jupiter notebook like the conversion of the monetary and populational amounts to float with 2 decimals, creating a list for the years in question for graphic processing, the table of this shown below:

Año	Turistas de Entrada	Turistas de Salida	Ingreso (Mill. US$)	Egreso (Mill. US$)	Ingreso per cápita (US$)	Egreso per cápita (US$)	Saldo Turistas	Saldo Divisas (Mill. US$)	Saldo Per cápita (US$)
1995	475120.00	513716.00	428.00	297.00	900.83	578.14	-38596.00	131.00	322.68
1996	590545.00	532328.00	670.02	349.98	1134.58	657.46	58217.00	320.04	477.12
1997	659567.00	579776.00	816.36	433.76	1237.72	748.16	79791.00	382.59	489.56
1998	735191.00	615857.00	844.80	452.82	1149.09	735.27	119334.00	391.98	413.82
1999	799946.00	646830.00	889.52	443.47	1111.98	685.61	153116.00	446.05	426.37

Table 1. Data frame of the main tourism indicators (5 firsts data)

A graph was made comparing the entry and exit of national and foreign tourists with the intention of seeing the favorable dynamic behavior of this sector of the economy, with the notable result that as of the new millennium (year 2000) inbound tourism became more notable due to the implementation of receptive tourism around the continent in order to encourage visits to Peru through fairs, promotions and advertising of different entertainment options highlighting its culture, gastronomy and architecture, among others.

	Turistas de Entrada	Turistas de Salida
1995	475120.00	513716.00
1996	590545.00	532328.00
1997	659567.00	579776.00
1998	735191.00	615857.00
1999	799946.00	646830.00

Table 2. Data frame of incoming and outgoing tourists
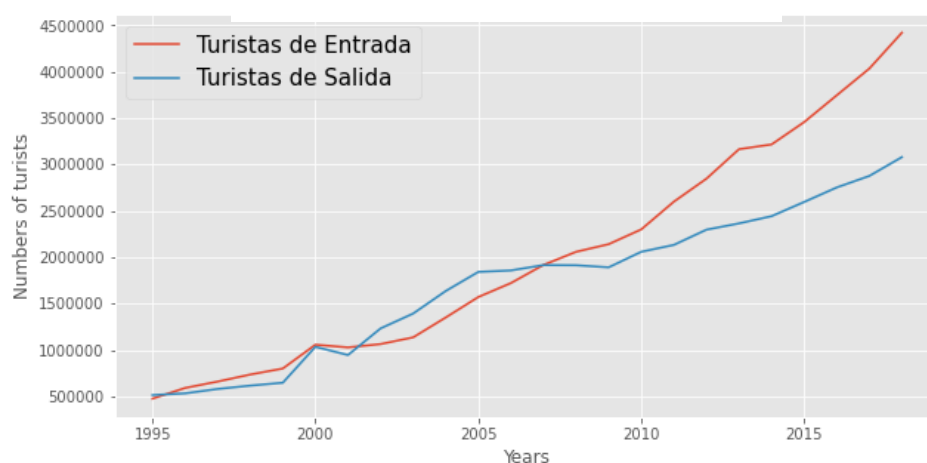*(5 first data)*

Figure 1. Incoming vs Outcoming tourists

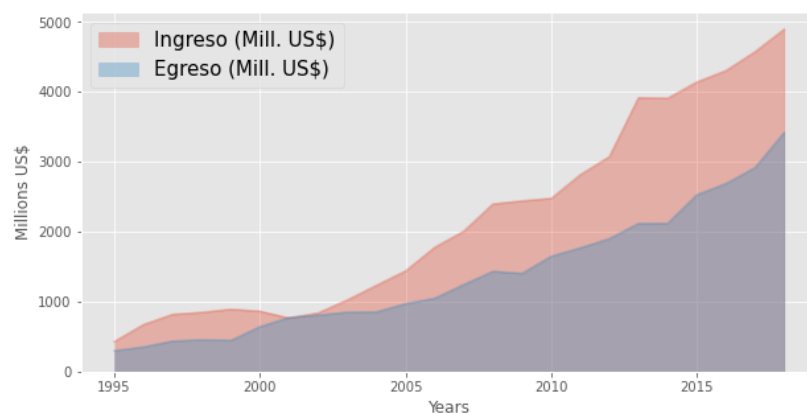The consequence of this tourist boom is reflected in tourism revenues as shown below:

Figure 2. Income vs Expenses

In this way it can be ensured that the investment will not have a negative benefit when choosing the type of business that is required to invest, because the tourism sector and all entrepreneurship that is around it, is profitable.

Following the objective of the project, It is going to consider that the Peruvian political-administrative-territorial division of Lima has 3 levels, such as *provinces, cones and districts* similar to those that are state, municipality and neighborhood respectively, so we work with them for future figures so it going to consider Metropolitan Lima as a region for the initial purpose, without the constitutional province of Callao, this area is made up of five subregions, the extremes of which are locally named Cones. These subregions, are as follows:

- **Downtown Lima**: Also known as the City of Lima or Downtown Lima; It includes the districts of Barranco, Breña, Jesús María, La Victoria, Lima, Lince, Magdalena del Mar, Miraflores, Pueblo Libre, Rímac, San Borja, San Isidro, San Miguel, Santiago de Surco and Surquillo.
- **East Lima:** Also known as the East Cone, it includes the districts of: Ate, Chaclacayo, Cieneguilla, El Agustino, La Molina, Lurigancho, San Juan de Lurigancho, San Luis and Santa Anita.
- **North Lima:** Also known as North Cone, it includes the districts of Ancón, Carabayllo, Comas, Independencia, Los Olivos, Puente Piedra, San Martín de Porres and Santa Rosa.
- **South Lima:** Also known as the Southern Cone, it includes the districts of: Chorrillos, Lurín, Pachacámac, Pucusana, Punta Hermosa, Punta Negra, San Bartolo, San Juan de Miraflores, Santa María del Mar, Villa el Salvador and Villa María del Triunfo.

By scraping web pages, it was possible to obtain the necessary data to identify each district and then mix them with their respective cones to later find relevant data thanks to different libraries made by python collaborators such as Josua Schmid's, which implemented a type of scraping where he translates tables from HTML to pandas data frames in python, this way you can match data already made. (Schmid, 2014)

	Cone	Districts	UBIGEO	Area (km²)	Population	Population density (/km²)	Postal code
0	North Lima	Ancón	150102	299.22	29,419	98.3	02
1	East Lima	Ate	150103	77.72	419,663	5,399.7	03
2	Downtown Lima	Barranco	150104	3.33	45,922	13,790.4	04
3	Downtown Lima	Breña	150105	3.22	94,808	29,443.5	05
4	North Lima	Carabayllo	150106	346.88	188,764	544.2	06

Table 3. Data frame of district's cones with their basic information (5 first data)

2.3. Exploratory Data Analysis

With the use of the Geopy python library, the coordinates of the different districts of Metropolitan Lima are found, that is, the districts are geocoded through their formatted address to give as a result their respective latitude and longitude coordinates and then append it with the table created earlier.

	Cone	Districts	UBIGEO	Area (km²)	Population	Population density (/km²)	Postal code	Latitude	Longitude
0	North Lima	Ancón	150102	299.22	29,419	98.3	02	-11.70	-77.11
1	East Lima	Ate	150103	77.72	419,663	5,399.7	03	-12.04	-76.90
2	Downtown Lima	Barranco	150104	3.33	45,922	13,790.4	04	-12.14	-77.02
3	Downtown Lima	Breña	150105	3.22	94,808	29,443.5	05	-12.06	-77.05
4	North Lima	Carabayllo	150106	346.88	188,764	544.2	06	-11.79	-76.99

Table 4. Data frame of district coordinates by cones

Other libraries for geographic visualization is folium which provides us with tools for this purpose, where the different districts of Metropolitan Lima will be shown as shown below:
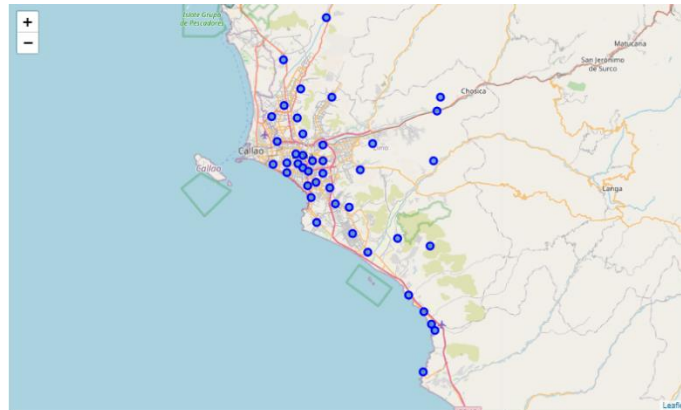

Figure 3. Map of districts distribution

In this case Lima Downtown was chosen because this Cone is very attractive for tourists also to be a central place to enjoy all kind of venues, restaurant, bar, etc.

In addition to having many places to eat, drink and hangout, it is an ideal place to invest in a business because the traffic of people is adequate to promote it.

Lima Downtown has practically everything, history, food, beaches, and very good music. Also, Lima Downtown have the most per capita incomes with 22,3% in the high, 52.8% in middle high and 24.9% in middle level of economic strata average (INEI, 2020).

Once the cone with which to work is chosen, we will use the Foursquare API to make requests about places or venues in a specific location, which in this case would be the Lima Downtown districts within a radius of 1500 meters and a maximum of 200 venues per district resulting a JSON file with different data such as name, venues category, latitude, longitude, etc. giving a total of 1256 venues and 189 venues categories.

	Districts	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barranco	-12.14	-77.02	FUGA	-12.14	-77.02	Nightclub
1	Barranco	-12.14	-77.02	Pizerrìa La Linterna	-12.14	-77.02	Pizza Place
2	Barranco	-12.14	-77.02	Sandwiches Monstruos	-12.14	-77.02	Sandwich Place
3	Barranco	-12.14	-77.02	Hotel B	-12.14	-77.02	Hotel
4	Barranco	-12.14	-77.02	Belgicius	-12.14	-77.02	Belgian Restaurant

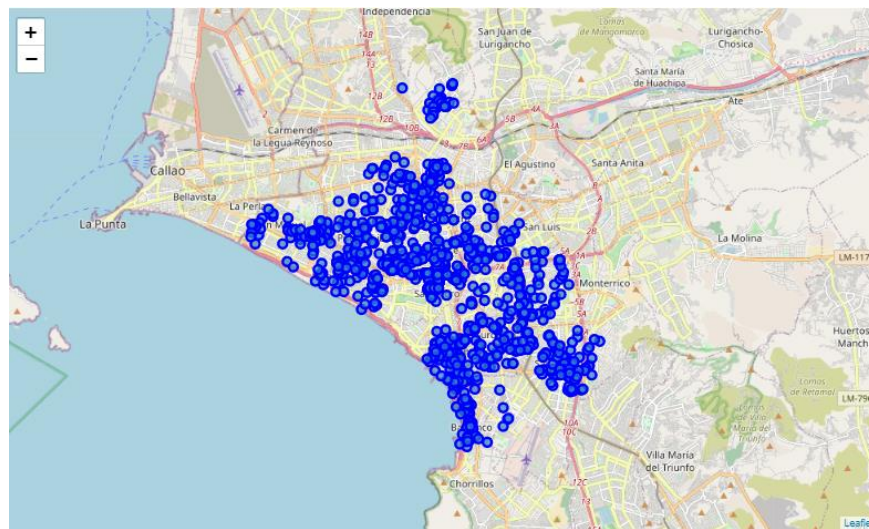*Table 5. Data frame of venues coordinates*

Figure 4. Map of districts venues

The idea of grouping is to determine in which places a guided tourist activity can be carried out throughout the center of Metropolitan Lima as part of the experience of being in a unique place that can offer its clients and consumers a diversity of options to try. That is why clustering will be done by categories of venues found before. The 10 most common venues by category were found, finding that about 45% are restaurants, and the rest are bars or parks,

this is taken into account when clustering in 5 partitions where those that stand out the most for their quantity and venues ratings are Breña, Barranco, Miraflores, San Isidro, Santiago de Surco, Surquillo, Jesús María, Lince and Lima.

Now with the use of the Google maps API (Places API, Directions API) you can search for the ideal route which will be carried out with a tourist guide around the main places in Metropolitan Lima.

Within the syntax of Directions API of Google Maps, the coordinates of the waypoints where you want the route to pass are required, which is why the districts, except those of origin and destination, will be found with the Geocoding API which is the similar to Geopy library.

The Breña district was established as the origin since it is the northernmost district compared to the others (Rímac was not considered since its number of venues is lower than the average), the district of destination to Barranco and the waypoints, the other districts considered.

The result is stored in a JSON file which will be normalized and formatted to analyze it in a panda's data frame for better manipulation.

	Instructions	Distance (m.)	Duration (seg.)	start_location.lat	start_location.lng	end_location.lat	end_location.lng
0	Head southeast on Jirón Gral. Vidal<...	10	2	-12.06	-77.05	-12.06	-77.05
1	Turn left at the 1st cross street onto ...	287	81	-12.06	-77.05	-12.06	-77.05
2	Turn right at the 2nd cross street onto...	1103	194	-12.06	-77.05	-12.06	-77.04
3	At the roundabout, take the 2nd exit on...	596	111	-12.06	-77.04	-12.06	-77.04
4	At the roundabout, take the 4th exit on...	227	59	-12.06	-77.04	-12.06	-77.04
5	Turn right onto Av. República de Chi...	366	52	-12.06	-77.04	-12.07	-77.04
6	Turn left onto Av. Arenales	71	8	-12.07	-77.04	-12.07	-77.04
7	Turn left at the 1st cross street onto ...	250	70	-12.07	-77.04	-12.07	-77.04
8	Turn left at the 2nd cross street onto ...	436	87	-12.07	-77.04	-12.06	-77.04
9	Continue straight to stay on Av. Petit Thou...	26	9	-12.06	-77.04	-12.06	-77.04
10	Turn left onto Av. 28 de Julio	305	61	-12.06	-77.04	-12.06	-77.04
11	At the roundabout, take the 4th exit on...	1438	176	-12.06	-77.04	-12.08	-77.04

Table 6. Data frame of touristic route

Finally, the tourist route is done, the only thing missing is the elaboration of the visualization on a map by folium where the instructions of the streets, avenues and highways were placed where mobility must pass to travel through clusters throughout Metropolitan Lima and make the most of the different entertainment options that Lima can offer.
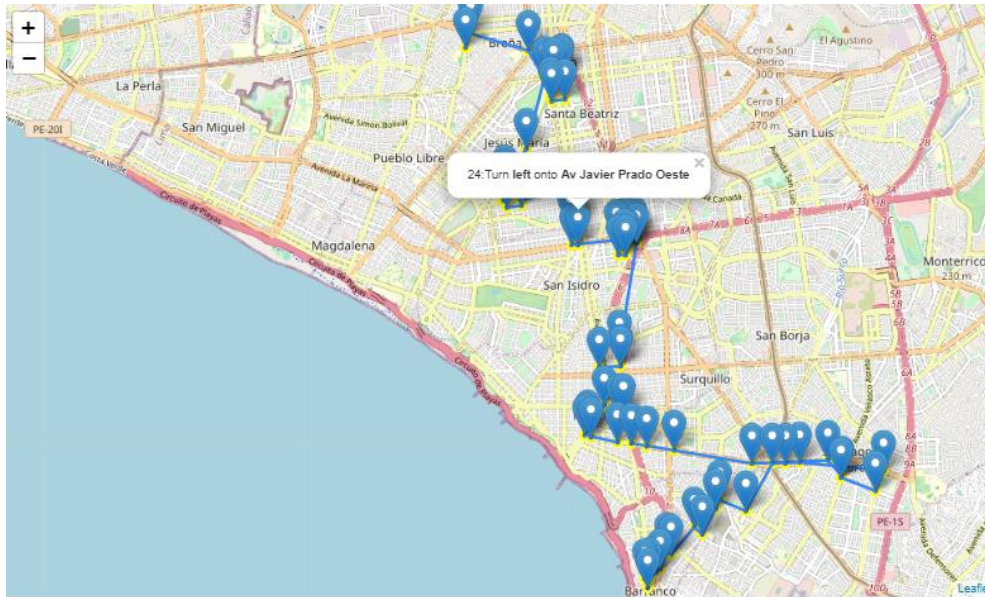
Figure 5. Recommended route by clusters

3. Results and Discussion

Thanks to the official data that can be found in government country's portals, in this case Peru, it was possible to verify that the tourism indices, not counting the current situation of the pandemic in order to be as practical as possible in this capstone , are very favorable measures when observing the positive trend in foreign currency inflows to the country as well as the crowded transit to the country in matters of tourism, gastronomy and business. In this way, any business that focuses on these areas mentioned above will have a favorable profitability.

After verifying the summary of the internal analysis, we made the web scraping to search of data from the web where they would give us information about the different districts within Metropolitan Lima without counting Callao, and then look for the venues around each district analyzed which we It will serve to identify the main representative businesses in each district, reaffirming the gastronomic potential of the country, resulting in that almost 90% of the venues categories are Creole restaurants, marine or bars.

After the clustering was done indicating the center of all venues in each district, this in turn helped us to make a tourist route that goes from the district of Breña, where 100 venues were found, to Barranco, where it ends the limit of the districts within Lima Downtown thanks to the Google maps API, which allowed us to find the shortest route

through all the clusters found previously, in this way this route will help us together with our stakeholders to carry out a profitable business full of food, tradition, culture, gastronomy and above all, diversity.

The result was 9 crowded districts where there is a diversity of venues (189 categories of venues). Its main categories were Creole food restaurants, seafood restaurants, bars and parks. A suitable route was obtained for the purpose of the business to be carried out.

The purpose of this analysis is to give the investor a north of the places where he can project a profitable business with a high degree of competition and then what better way than to give him various options that both he and the client can have, giving a recommended route for this purpose

4. Conclusion

In this capstone we concluded the compass or north of a profitable business based on official data such as the INEI, Google API for the proposed route and web pages for the primary collection of data concerning the districts of Lima. The implementation of all operations such as hiring tour guides, mobility, business plans, prices and others is left in the hands of the stakeholders.

5. Bibliography

Foursquare. (n.d.). *Foursquare developers*. Retrieved from
https://es.foursquare.com/developers/apps

Google. (n.d.). *Google Cloud Platform*. Retrieved from
https://console.cloud.google.com/home/dashboard?project=big-iridium-288418

INEI. (2020). *Planos Estratificados de Lima Metropolitana a Nivel de Manzanas.* Lima.

Schmid, J. (2014, 03 05). *Github*. Retrieved from Github: https://github.com/schmijos/html-table-parser-python3

Wikipedia. (n.d.). Retrieved from https://es.wikipedia.org/wiki/Turismo_en_el_Per%C3%BA