

Clustering and Analyzing Panamá Smart City Locations

Author: Kenry A. Sanchez Vento.

June 16th, 2019.

1. Introduction

Let's say *The Marketing Department of Liberty Technologies Corp.* wants to know which industries, stores or Malls are being benefited of the *Paradas Smart City* which are some places of the city with Free Wi-Fi and radio of 7 meters. The company is responsible of the installation of any Free Point around the city and are the owners of the access points.

They want to know which places are around the Free Wi-Fi point for that way offer to those companies their portal access as a point for gain more clients. For Example:

- Offer discounts to be used on local stores.
- Offer some events like expositions.
- Visit tourist places.

In this way, they want to know the answer to these questions:

1. How are distributed the points around the city?
2. How are segmented the points?
3. Is there a way to clustering or segmenting the points using attributes like kind of place, store, etc.?
4. Which are the most commons kind of place around the points?

1.1 Who can be interested?

In a Business Focus, not just for the Liberty Technologies Corp, understand the segmentation of any store or merchandise around the place can help to other companies (usually, most be the businesses around the Wi-Fi points) to make marketing strategies in based on those places. In fact, they could support a business strategy in alliance with *The Department of Liberty Technologies Corp.* which is the responsible of the wifi points.

2. Data Requirements

The data we're going to use to solve the problem we have discuss will be from 2 sources:

- The site of *Portal de datos Abiertos de la alcaldía de Panamá* (Open Data Portal from Panama City town Hall). From this source we are going to recover all the localizations of any Smart City Point around the City.
- Exploration data. This data will be recover using *Foursquare API*, which is a platform using for put and pull exploration data from all the cities around the world.

2.1 Pre-Processing Data

The data recover from the portal site must be verified, which means we need to clean it first. We need to remove any Nan value inside of it. The data recover from the site come with 4 attributes

- Nombre (Names).
- GROUPID.
- X (LONGITUD).
- Y (LATITUDE).

We need to remove the row with **Nan** values in the location attributes (latitude and longitude) because we are not able to discover missing values for both attributes in every row. Also we need to remove any duplicate row. In fact, any row represent just one smart point in the table.

Fig 1. Smart table cleaned

	X	Y	OBJECTID	Nombre
0	-79.535111	8.962863	1	Cinta Costera
1	-79.528691	8.972058	2	Cinta Costera
2	-79.521034	8.974748	3	Cinta Costera
32	-79.533981	9.009629	47	Tumba Muerto, Centro Comercial al Dorado
34	-79.534493	9.009785	49	Tumba Muerto, Centro Comercial al Dorado

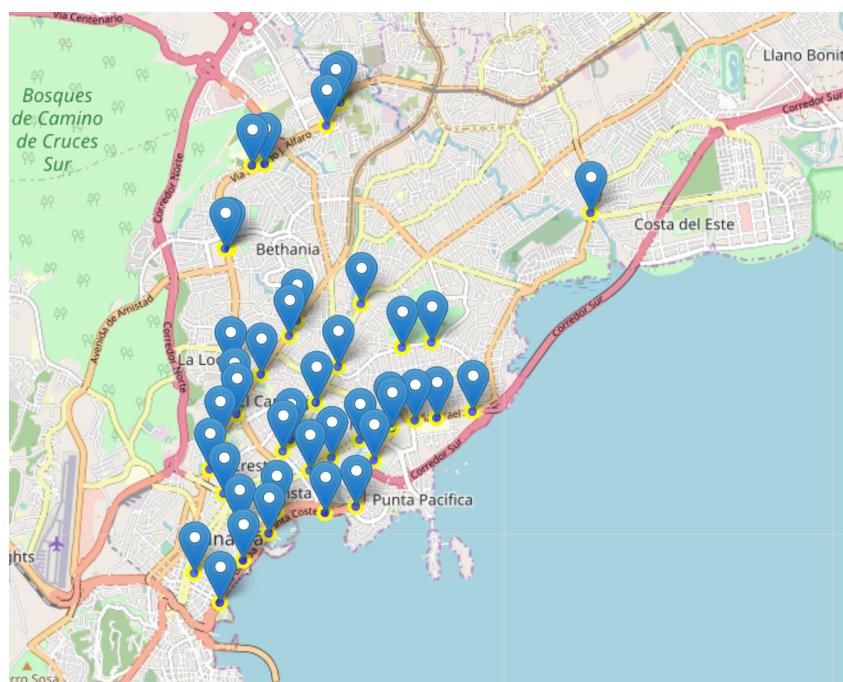
3. Exploration Data

The Smart Point table will be used to create a map of Panama City with all the Smart Point Locations. First, we need to define a radius for our search avenue. We want to select any place close to our points. For this initial approach, taking into consideration we don't have any studio about which places people decide to visit close to the Wi-Fi Points and how often they decide it, we're going to choose a default radius for our studio. We're going to say (*based on The Hurwicz Criterion*) people are able to walk to from the wifi point to any place near by 250 meters when there is a special occasion, as an example, a special discount or an event. As every Wi-Fi Point has a radius of 7 meters, that means people don't need to be in the center of the point to access to the portal and then see the advertisement. So, we are going to start with a radius of 257 meters.

Then, with our map we defined some approaches:

- The WiFi Points are distributed from the west to east of the city.
- Some WiFi Points share avenues with another. As an example Cinta Costera and Avenida Balboa.
- There are 2 Points far away of the center of the city which are at Av. José A. Arango.
- The majority of points are at the west of the city.

Fig 2. Smart Point Locations.



3.1 Foursquare API Data

Foursquare API help us to recover the next points:

- A. Search avenues close to the smart point locations.
- B. Recover the categories of the places close to the avenues and in consequence, close to smart points.

With this data, we create an initial group, just for have an idea about how are the places grouped around the points. We create a score card, when we show the percentage of category for each row (smart point).

Indeed , we can define a list of most commons category places around the smart points.

Fig 3. Sample Most Common Categories Table.

	Smart Point Name	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category	8th Most Common Category	9th Most Common Category	10th Most Common Category
0	Av. José A. Arango, frente al Macdonals	Department Store	Restaurant	Fast Food Restaurant	Shopping Mall	Latin American Restaurant	Sporting Goods Shop	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Court
1	Av. José A. Arango, junto al Super Xtra de Los...	American Restaurant	Sporting Goods Shop	Fast Food Restaurant	Restaurant	Department Store	Shopping Mall	Italian Restaurant	Bakery	Gift Shop	Factory
2	Avenida Balboa, frente a Multicentro	Bar	Hotel	American Restaurant	Cocktail Bar	Lounge	Pool	Pizza Place	Pharmacy	Park	Nightclub
3	Avenida Balboa, junto a el Hospital de Niño	Diner	Indie Movie Theater	Gay Bar	Drugstore	Electronics Store	Factory	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Food Court
4	Avenida Balboa, junto a él Parque Urracá	Mexican Restaurant	Bar	Park	Café	Buffet	Restaurant	Tapas Restaurant	Gym	Pub	Hotel
5	Avenida Justo Arosemena, frente a el Hospital ...	Sandwich Place	Hotel	Gastropub	Pharmacy	Food Court	Gas Station	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant

In the exploratory data collection from Foursquare API, we discover that point **Transístmica, junto a Escuela de Artes y Oficios** was not present in Foursquare DB (Database). It was not possible to obtain data from that point, so we must remove it from the initial data table.

To our analysis, the goals are:

- Discover commons categories and avenues.
- If we notice a point that is useless for our final model, remove it.
- Prepare an idea about how could be segmented the smart points around the city.

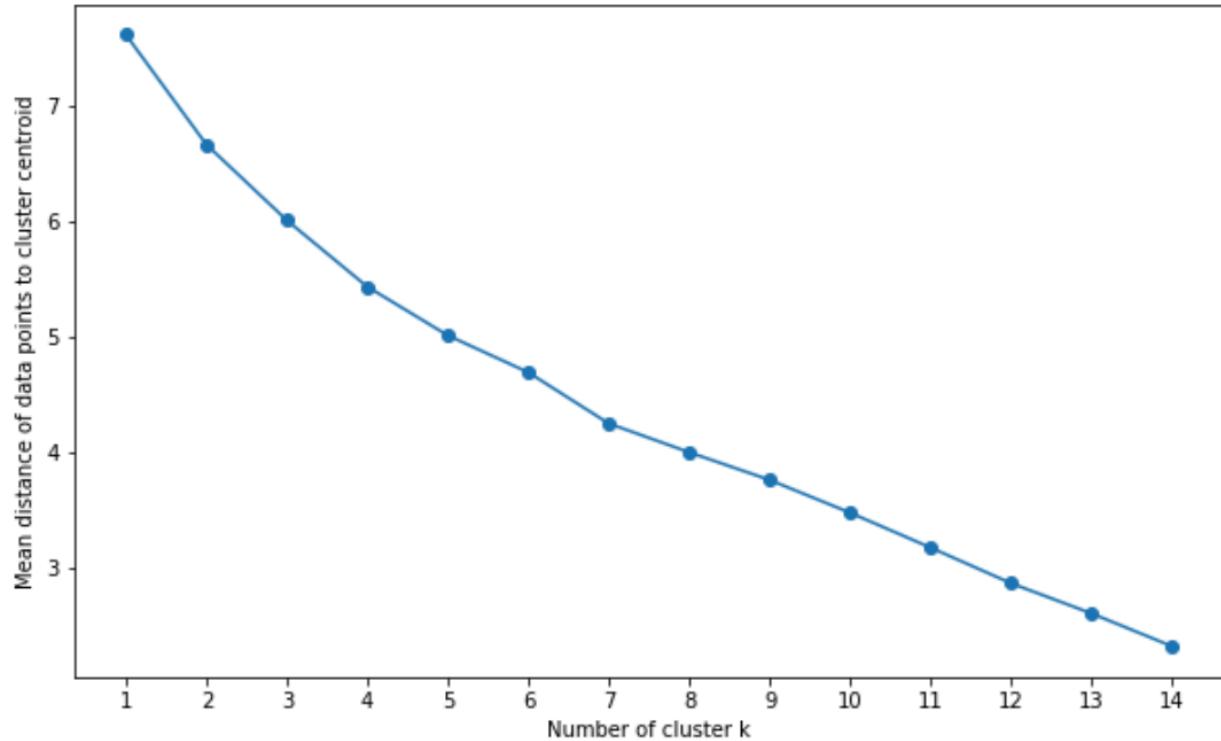
4. Data Analysis and Modeling

The Machine Learning Model to use to cluster the data will be **K-Means**, which can cluster the smart point locations in segmentations using similar attributes and patterns by every row in the table.

We have used the **Elbow Method** to define the best “k centroids” which is the best number of centroids for the clustering. Every time that k-means use one more centroids, the mean distance of data points to cluster will be smaller, so is not just a good idea to choose the k centroids which can reduce the distance between the points and any centroid. That’s way the elbow method can help us to choose the most closet “k centroid”.

After create our ML Model and use the elbow method, we choose “**K = 6**” as the best number of centroids for our cluster.

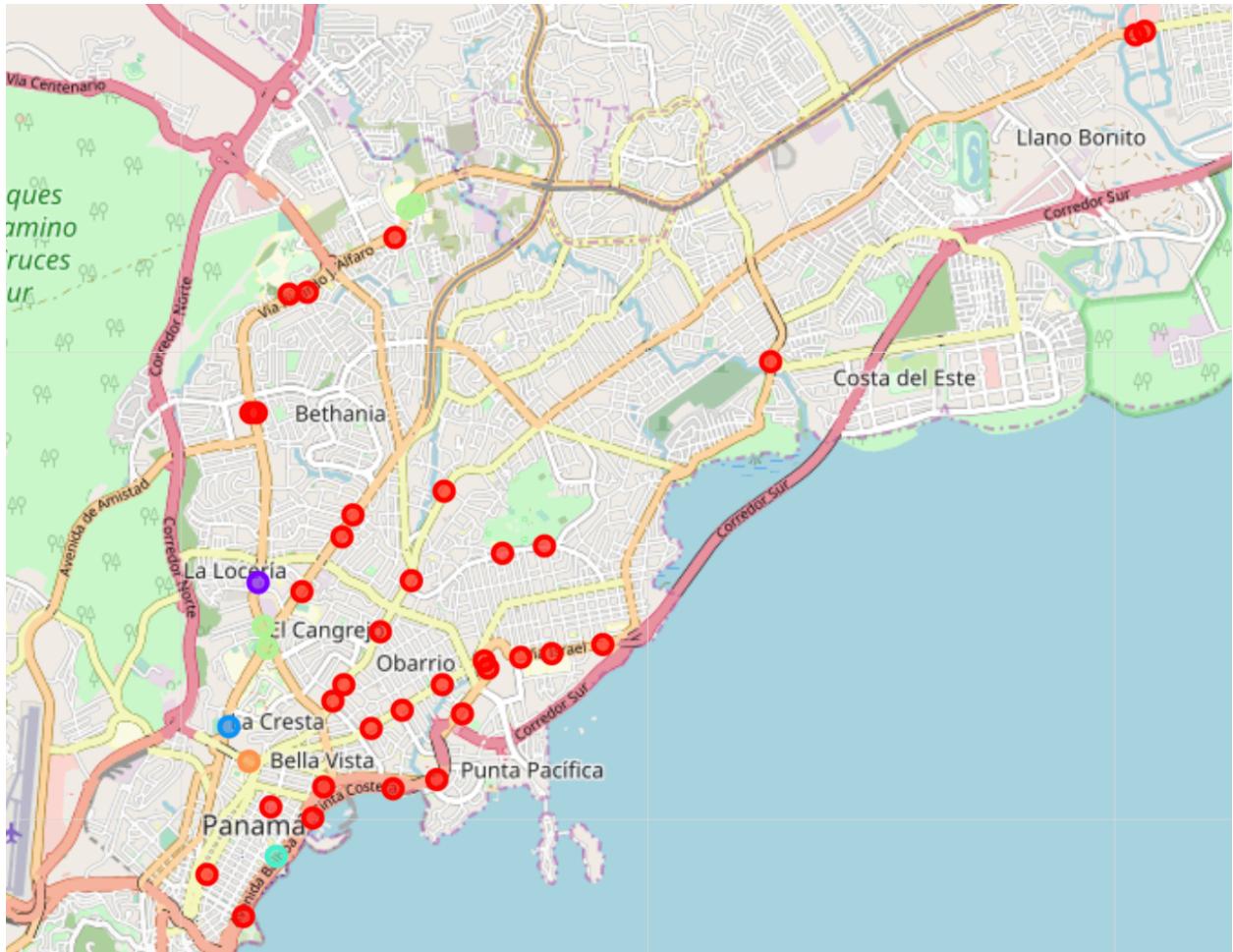
Fig 4. Elbow method result. The elbow is most present in $k = 6$.



4.1 Mapping Cluster into Map

Using our ML Model, we can cluster our smart point table as the following map.

Fig 5. Smart Points Clustered.



Where the difference colors means the following:

- Red color represent cluster number one.
- Blue color represent cluster number three.
- Purple color represent cluster number two.
- Turquoise color represent cluster number four.
- Green color represent cluster number five.
- Orange color represent cluster number six.

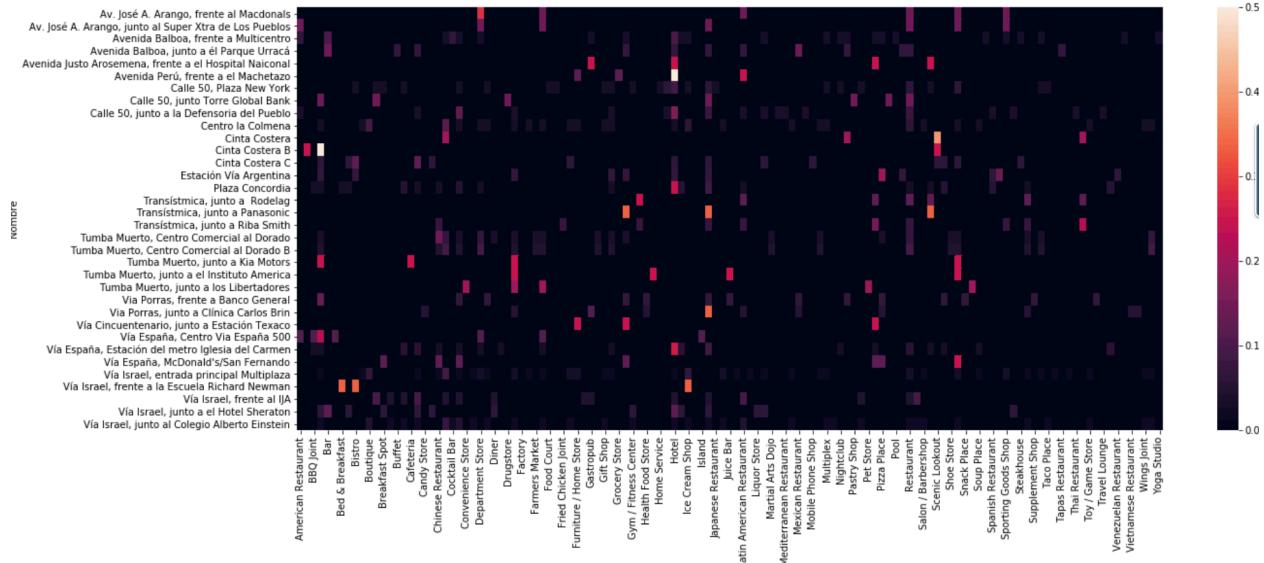
If we make a quickly analysis, it's not complicate to discover the majority of them represent cluster 1. Also, the whole cluster 1 represent the center of the city (which is the zones where people use to been in the city).

4.2 Analyzing Cluster and Results section

We use a Heat-Map to discover insights from our clusters.

A. From Cluster One:

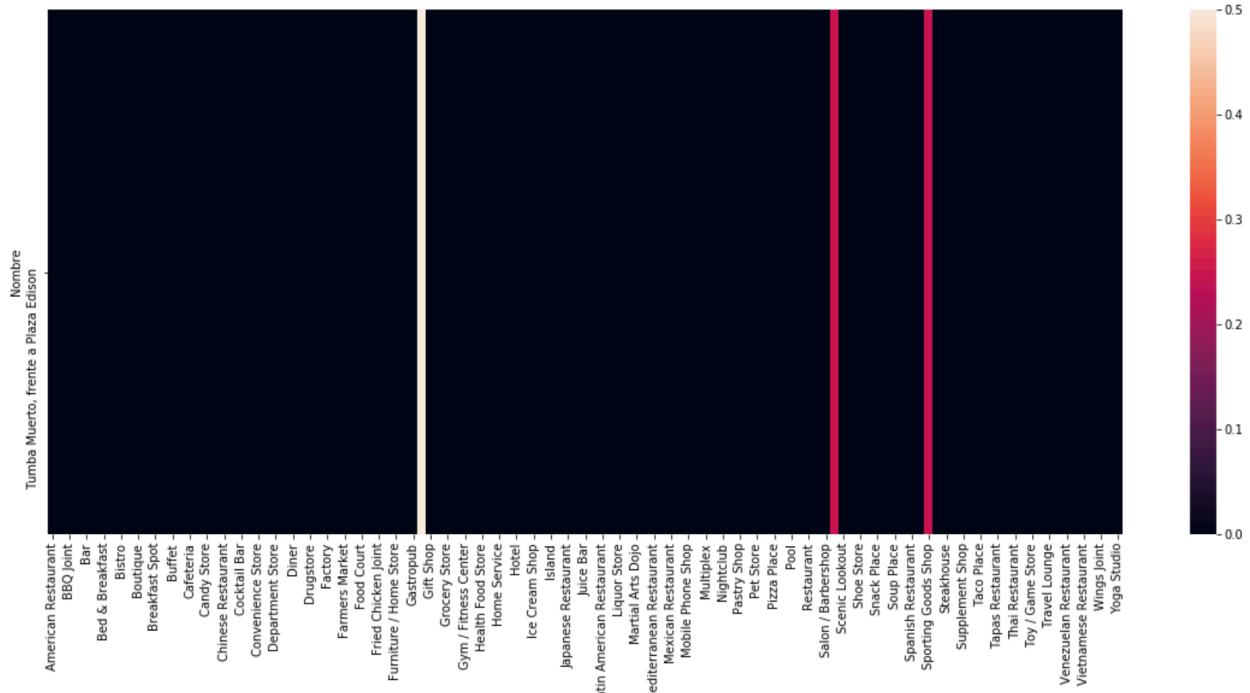
Fig 6. Heat-Map Cluster 1.



This the first and more difficult group to understand. If we see the final Map, we see that the red colors (cluster 1) are in the whole center of the city. They represent the majority of points and this is a great insight which means the majority of points are in the same category of places. What kind of places? The heatmap shows a particular pattern that others clusters don't have. If we see the majority of the avenues have places for Food and drinks. No so many are about hotel, barbers, shops, etc. So maybe people are more prone to be in these places, spending their times with friends and drinking or having a lunch than in another side of the city. This is a great insight for business because this cluster represent the first option for create business strategies using The Smart Points.

B. From Cluster Two:

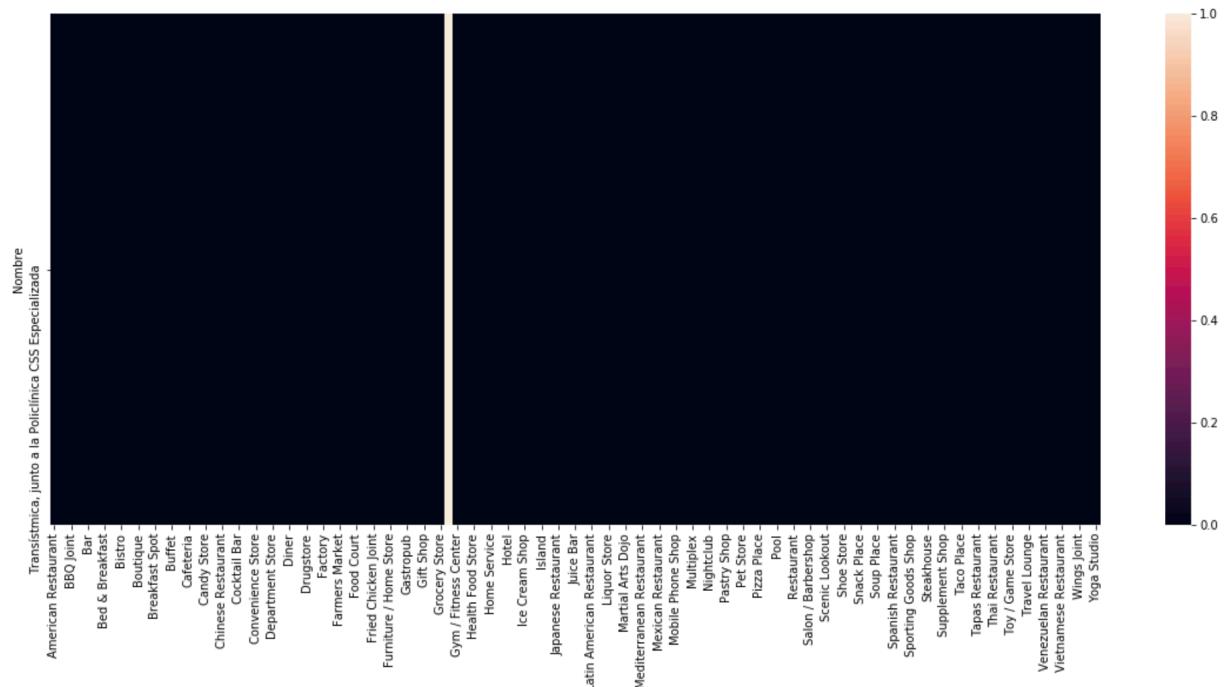
Fig 7. Heat-Map Cluster 2.



In this cluster, there is just one avenue which is *Tumba Muerto frente a plaza edison*. Analyzing the heat-map we can understand that this is a place just for make shopping. So, in the smart point people that are close to this point could be here just for that.

C. From Cluster Three:

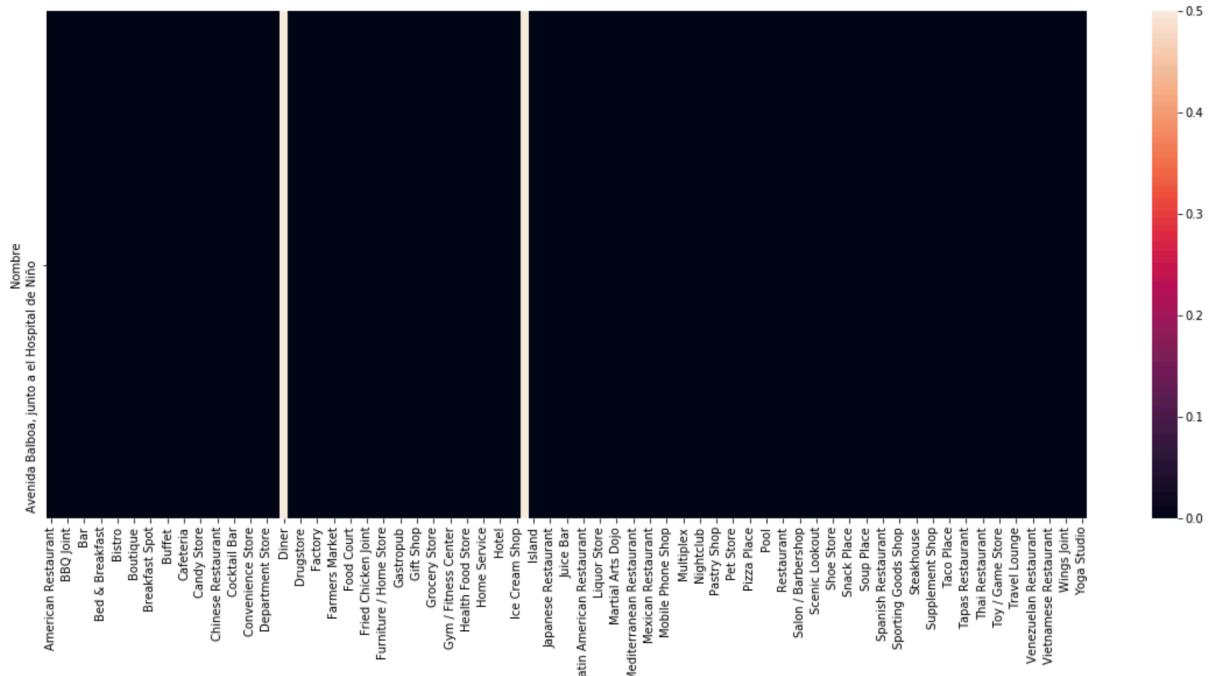
Fig 8. Heat-Map Cluster 3.



In this cluster, the heat-map tell us that this is a zone just for grocery stores. The measure is in 50%, so i think it's a good zone for strategies with groceries.

D. From Cluster Four:

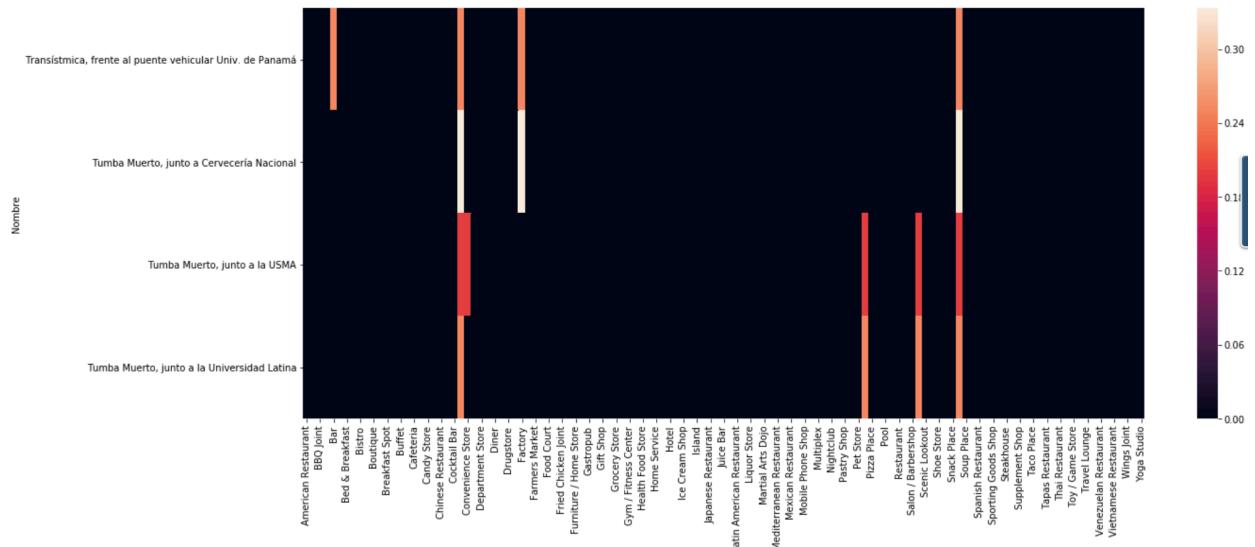
Fig 9. Heat-Map Cluster 4.



This cluster have just one avenue too which is *Transistmica junto a policlinica..*. It also doesn't seems like a special point. I don't think this could be a good place for marketing strategies either.

E. From Cluster Five:

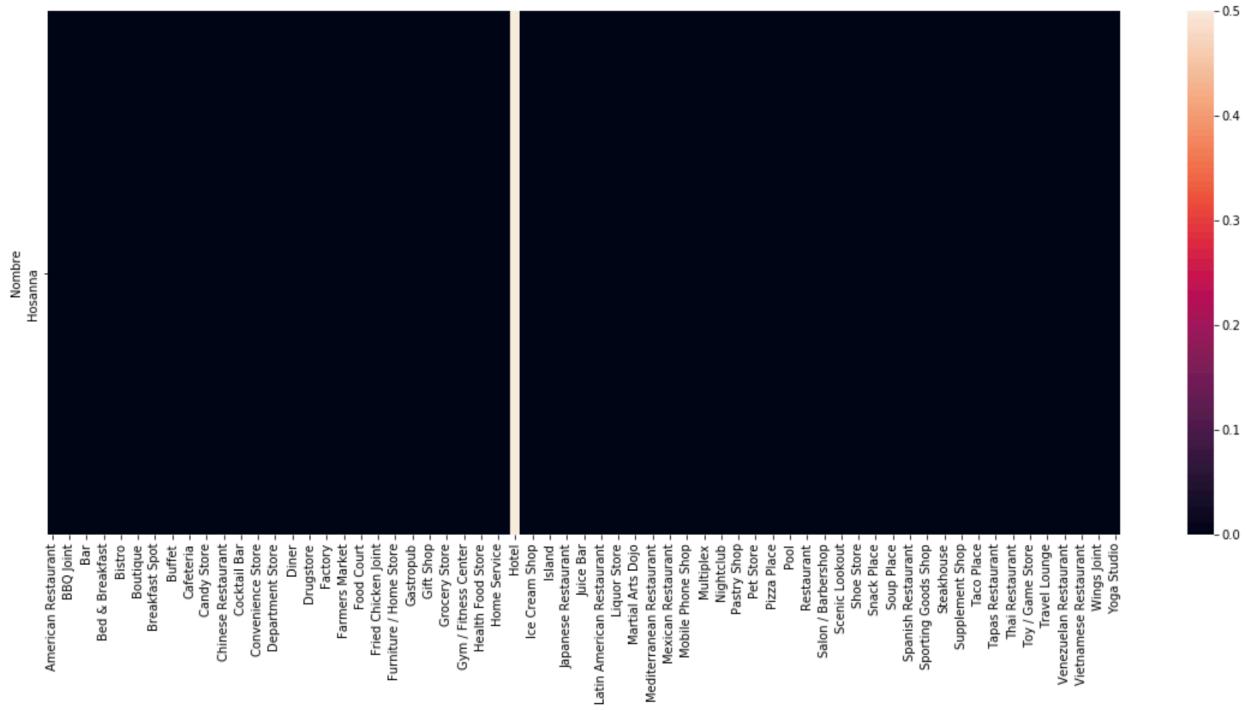
Fig 10. Heat-Map Cluster 5.



According to the cluster, this area is for places that offer drinks and food, very similar to cluster one. It is also important to comply with the largest number of avenues, after cluster one.

F. From Cluster Six:

Fig 11. Heat-Map Cluster 6.



In this cluster is just one avenue, 'Hosanna'. But, according to the heat-map, this is a place close to hotels or just one hotel.

5. Conclusions

Our analysis shows the distribution of smart points around the city using 6 groups of segmentations. For each group we have collected the patterns about how many kinds of places are around them.

According to this studio, we can define the following:

- The majority of Smart Points belongs or are close to places where people are prone to be. Places like bars, restaurants, groceries, hotels. This is important to understand because we can be sure that maybe at the beginning, the company did a great job trying to distributed the points with a strategy.
- We have now 6 kinds of clusters and we can define it in the following way: places where people are prone to be --> Places where people are not use to being.
- The most commons places are about: Food, Bars and groceries.

6. Discussion Section

Our studio was created to understand how the smart points are distributed around the city. The measure of the studio was define by two points:

1. The main data, which is the smart points locations.
2. Foursquare API.

Using these tools, we were able to made a complete analysis about the smart points. We discover the most commons places and also a way to create a model group for future distributions, which means how can be locate it a new point based on this studio.

However, it's important to clarify just one concern.

While we have been using Foursquare API to create our exploration data, we need to use a better API with more accuracy data for next studios. Foursquare API doesn't has the whole locations of the city or even all the places (stores, hotels, parks, etc.). So, for future studios, we need to use another API with more accuracy data about The city of Panamá. Foursquare API was not able to recover data from the point Transístmica, junto a Escuela de Artes y Oficios, which could be important for the final analysis.