CUNDINAMARCA POTENTIAL MARKET

Gabriel Gerardo Amaya Becerra

CONTENIDO

INTRODUCTION

BUSINESS CONTEXT

- A technology company has a new business line know as Analytics – AI.
- The reason that company formed this new line was because companies does not only need an engineer structure but also an implementation of AI technologies.

Resource: Flaticon

DATA SET

DATA BASES

- Cundinamarca data set: Cundinamarca is formed by 116 towns and 1 special district that is Bogotá D.C., the data set was download by DANE and has the following structure:
  - Codigo_Depto: Code to identify each department.
  - Nombre_Depto: Name of each department.
  - Provincia: Group by each province where town is located inside each department.
  - Codigo_Municipio: Codes of each town.
  - Nombre_Municipio: Name of each town.

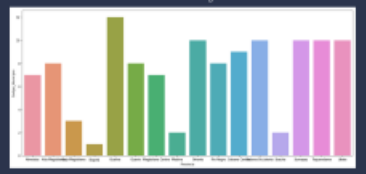
Resource: Google

EXPLORATORY DATA ANALYSIS

UNDERSTANDING DATA

- The first analysis was to know how many towns are in each province, it is important to say that province is a way to aggregate towns in groups.
  - Gualiva is the province that has the greatest number of towns following by Oriente, Sabana Occidente, Sumapaz, Tequendama and Ubate.
  - Bajo Magdalena, Medina and Soacha are the towns with lesser number of towns.
  - Bogotá is a special district in that way there do not have more towns or cities inside it.

Distribution of towns by each Province

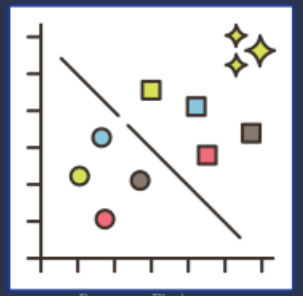


Resource: Own develop

PREDICTIVE MODELING

ASSUMPTIONS

- We create a k-means model in the way to understand what kind of similitudes has the towns, and also to get some industries where the company could begin.
- We think that 15 clusters is a good way to understand towns, we expect to get 7 towns inside each cluster.



Resource: Flaticon

CONCLUSIONS

GENERAL

- The company will focus on analyse in deep the idea to make a solution for food companies, specially for restaurants where they can create a technological solution for getting orders but also to learn more about their clients.
- The company will evaluate in deep the cluster 3 specially venues bodegas, because in that places they will find medium and big companies that would be clustering on similar purposes.
- The company would see what kind of venues are more interesting, that would help to create better cluster and summarize, because as we see before the cluster 1 has a lot of different places were the company will create a solution like multiplex and history, but this cluster do not give the possibility to know the real potential market that company will have.
- Finally, the company will think how smes technology they can help outdoor activities, however it is important to say that normally when people want to do these activities, they prefer to be disconnected (not use technology) in that way this would be a rick idea.

Resource: Flaticon

INTRODUCTION

BUSINESS CONTEXT

- ✓ A technology company has a new business line know as Analytics – AI.
- ✓ The reason that company formed this new line was because companies does not only need an engineer structure but also an implementation of AI technologies.

Resource: Flaticon

INTRODUCTION

Resource: Flaticon

BUSINESS PROBLEM

- ✓ The company wants to know what will be the first companies to develop a solution that's because they know that once the product has been developed it is easier to adjust to similar company and sold it.
- ✓ Also, the first approach will help the company to know potential markets where the company will sell its solution.

DATA SET

DATA BASES

- Cundinamarca data set: Cundinamarca is formed by 116 towns and 1 special district that is Bogotá D.C., the data set was download by DANE and has the following structure:
  - ✓Codigo_Depto: Code to identify each department.
  - ✓ Nombre_Depto: Name of each department.
  - ✓Provincia: Group by each province where town is located inside each department.
  - ✓Codigo_Municipio: Codes of each town.
  - ✓Nombre_Municipio: Name of each town.

Resource: Google

DATA SET

DATA BASES

Resource: Google

- Foursquare Cundinamarca data set:
  - ✓ Codigo_Municipio: Codes of each town.
  - ✓ Codigo_Municipio Latitude: Latitude where town is located.
  - ✓ Codigo_Municipio Longitude: Longitude where town is located.
  - ✓ Venue: Name of the company.
  - ✓ Venue Latitude: Latitude where company is located.
  - ✓ Venue Longitude: Longitude where company is located.
  - ✓ Venue Category: Group or sector where company take part.

DATA SET

CLEANING PROCESS

- The Cundinamarca data set from DANE has no need more modifications based on it we get the latitude and longitude using geocoder.
- Foursquare data set was cleaning removing some common things like Parks, central plaza and farms, based on the structure that towns were building.
- After cleaning process both data base were merged to create a new full data set.

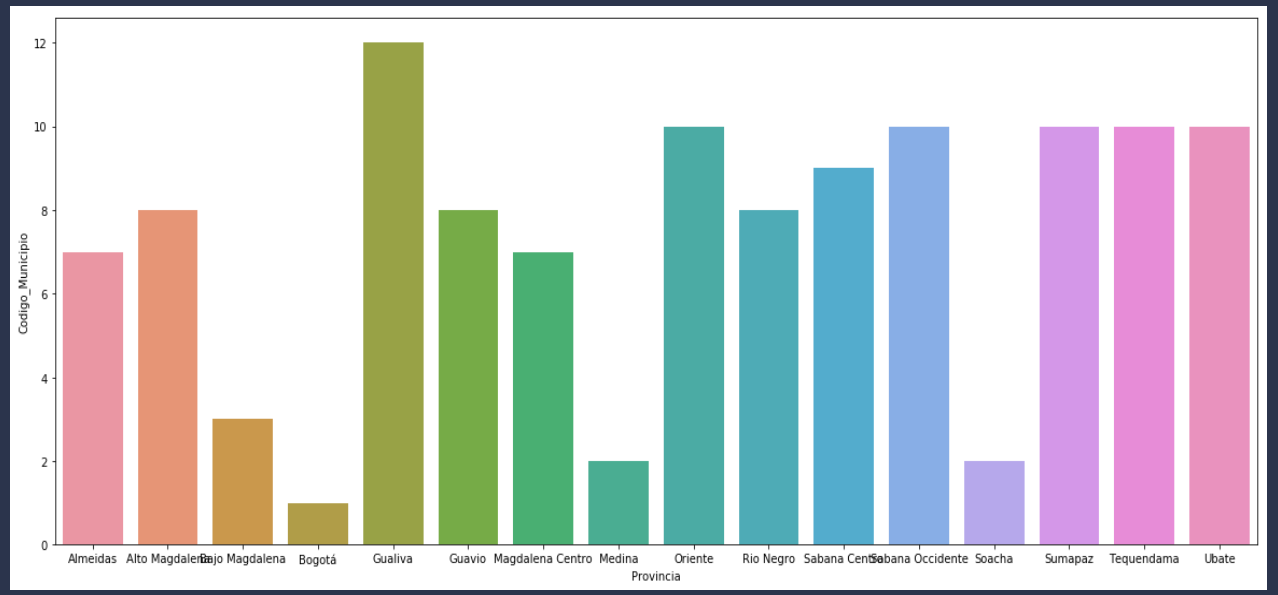
Resource: Flaticon

EXPLORATORY DATA ANALYSIS

UNDERSTANDING DATA

- ✓ The first analysis was to know how many towns are in each province, it is important to say that province is a way to aggregate towns in groups.
  - ✓ Gualiva is the province that has the greatest number of towns following by Oriente, Sabana Occidente, Sumapaz, Tequendama and Ubate.
  - ✓ Bajo Magdalena, Medina and Soacha are the towns with lesser number of towns.
  - ✓ Bogotá is a special district in that way there do not have more towns or cities inside it.

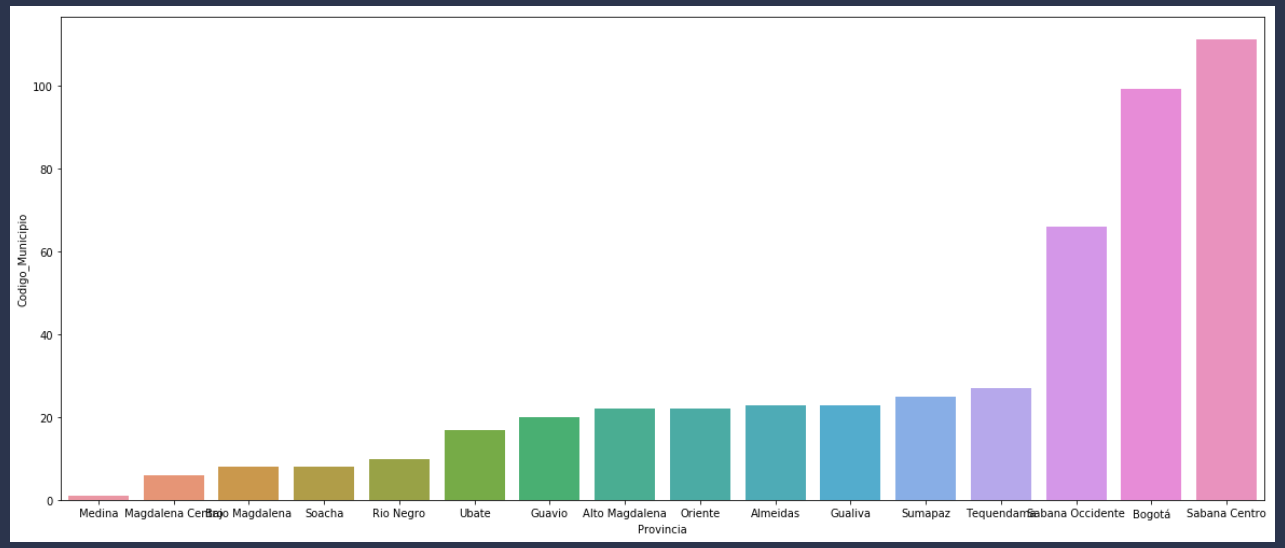
Distribution of towns by each Province



Resource: Own develop

EXPLORATORY DATA ANALYSIS

Quantities of venues collected by each Province.



Resource: Own develop

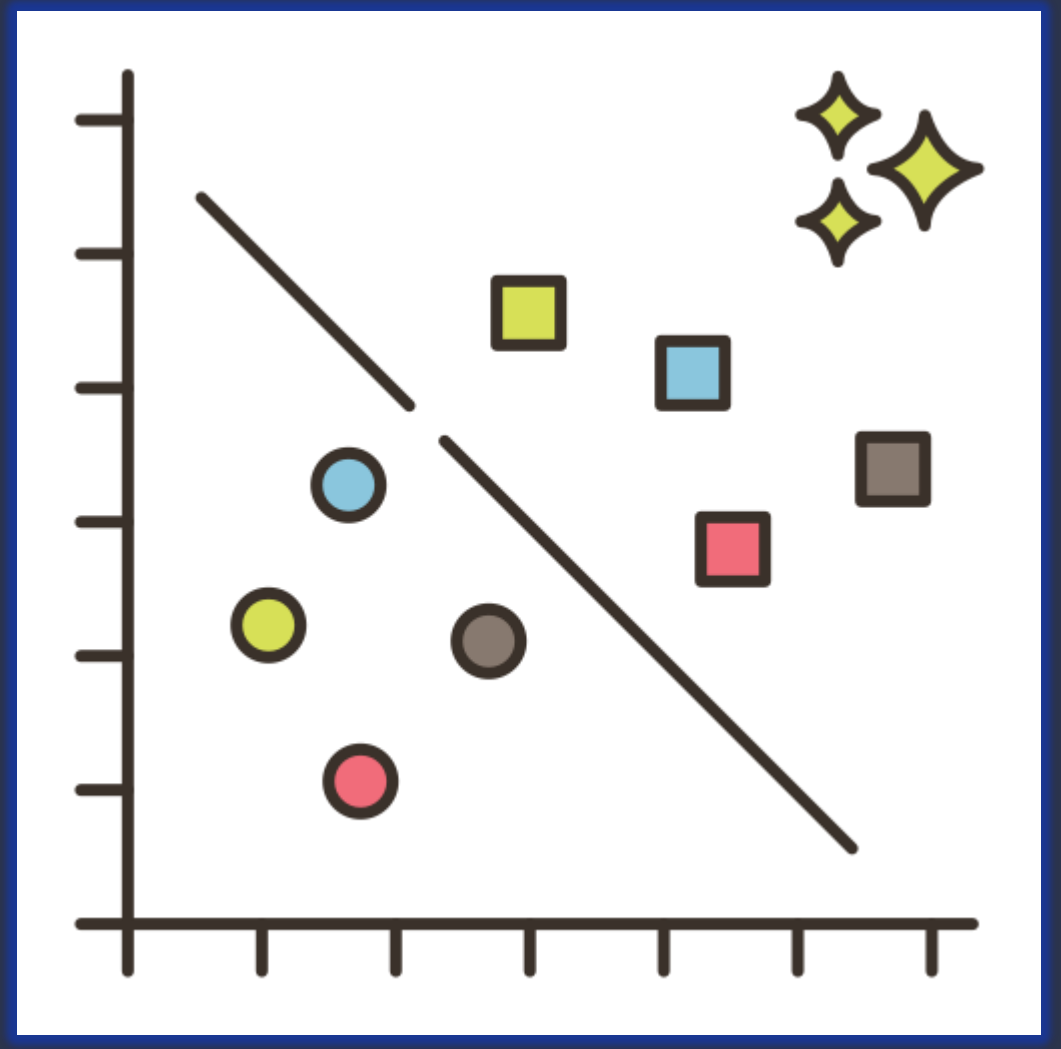
UNDERSTANDING DATA

- ✓ Once we get the information from foursquare API, we repeat the previous process.
  - ✓ It is important to remember that we get venues information as a 1 km from latitude and longitude.
  - ✓ Also, it is important to take in mind that Bogota is a big city, however we only get information from a unique point location.

PREDICTIVE MODELING

ASSUMPTIONS

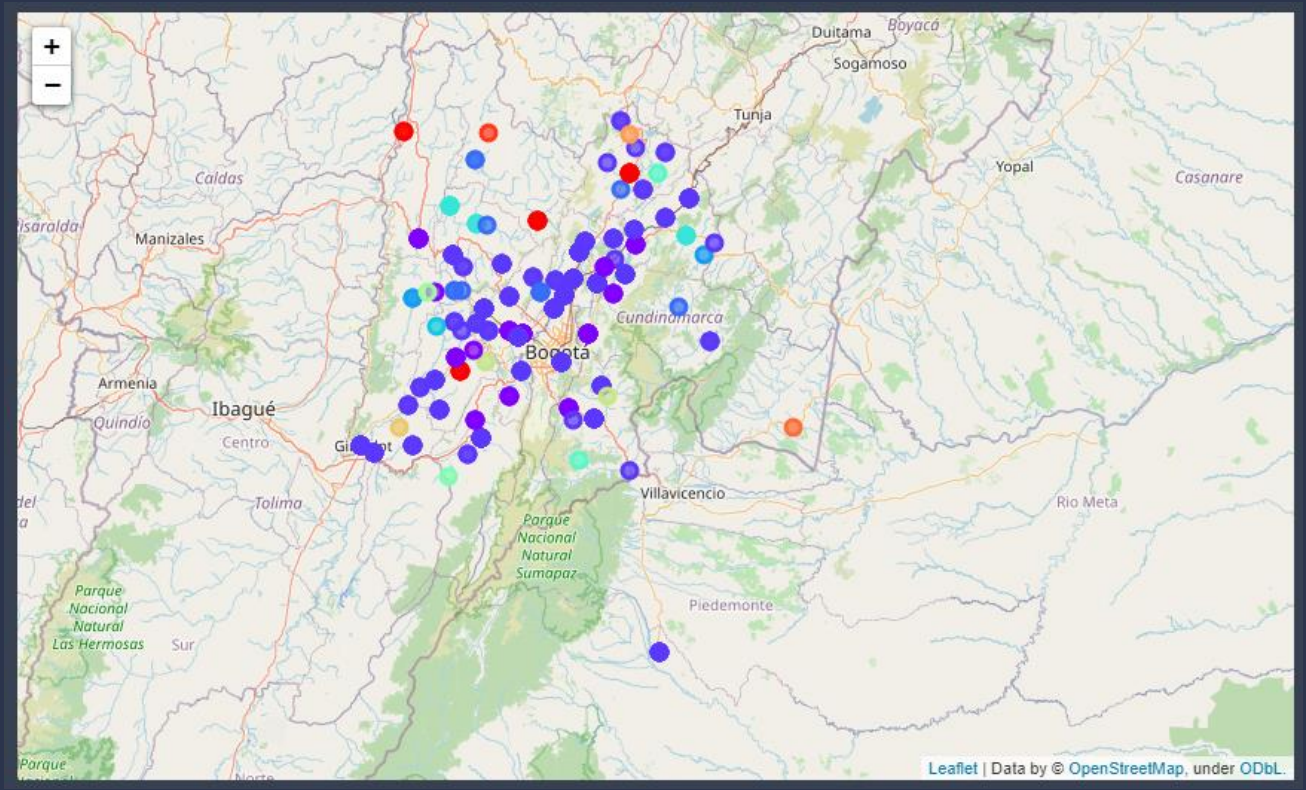
- We create a k-means model in the way to understand what kind of similitudes has the towns, and also to get some industries where the company could begin.
- We think that 15 clusters is a good way to understand towns, we expect to get 7 towns inside each cluster.



Resource: Flaticon

PREDICTIVE MODELING

Cluster results inside Cundinamarca´s map



Resource: Own develop

RESULTS

- ✓ Cluster 0 is based on food venues, the most common are Latin American Restaurant, Soup place, Food court and Diner places.
- ✓ Cluster 1 has multiple kinds of places like multiplex, shops, foods, mountain, and history, in that way it will be a good idea to filter these places and run again a cluster to get better ideas about kinds of venues and segmentation for each group of towns.
- ✓ Cluster 2 is more related with restaurants and coffee places, also another important thing to highlight is related with Bodega which will give us to think that here we will find some medium and big companies.
- ✓ Cluster 3 is more related with home shops and wings joint.
- ✓ Cluster 6 is more related with outdoor activities like rafting, mountain, lake, and pool.

CONCLUSIONS

GENERAL

- ✓ The company will focus on analyze in deep the idea to make a solution for food companies, specially for restaurants where they can create a technological solution for getting orders but also to learn more about their clients.
- ✓ The company will evaluate in deep the cluster 2 specially venues bodegas, because in that places they will find medium and big companies that would be clustering on similar purposes.
- ✓ The company would see what kind of venues are more interesting, that would help to create better cluster and summarize, because as we see before the cluster 1 has a lot of different places were the company will create a solution like multiplex and history, but this cluster do not give the possibility to know the real potential market that company will have.
- ✓ Finally, the company will think how since technology they can help outdoor activities, however it is important to say that normally when people want to do these activities, they prefer to be disconnected (not use technology) in that way this would be a risk idea.

Resource: Flaticon

CONCLUSIONS

Resource: Flaticon

RECOMMENDATIONS

- Based on this first approach to potential market the company will help to create filters based on their interest, these filters will help to run again the model and would create better clusters and potential market.
- Independent the sector that they select to create a solution, it is very important to make an evaluation about needs, movements and the client disposition to acquire and implement a new TI solution, because as we discuss before venues related with outdoor activities not always are related with technology connection.

THANK YOU SO MUCH!!!