




# *The Battle of Neighborhoods*


*An IBM-coursera capstone project*

*Submitted by: Samuel P. Polanco, Jr.*





*DEAR READER THANK YOU FOR  
YOUR TIME IN READING MY  
REPORT, I HOPE YOU LIKE IT*



# Table of Contents

INTRODUCTION	*****	1-3
DATA	*****	4-6
METHOD	*****	7
BUSINESS UNDERSTANDING	**	8
ANALYTIC APPROACH	*****	9
DATA REQUIREMENTS AND COLLECTION	*****	10
DATA UNDERSTANDING	*****	11
DATA PREPARATION	*****	12
MODELING	*****	13-17
INFERENTIAL TESTING	*****	18-19
DISCUSSION	*****	20-22
CONCLUSION	*****	23-24
THANK YOU	*****	25

# INTRODUCTION

## **PROBLEM:**

What is the best kind of business to put up in the Province of Negros Occidental, the Philippines, and where should I place it in order to maximize profits?

## **BACKGROUND:**

Due to the current administration's efforts to enhance the economy of the Philippines, the Duterte administration decided to allocate most of its resources to infrastructure projects. Roads and bridges have been built, government buildings have been erected, and government transactions have been streamlined in order to improve efficiency and ease of use.

Because of this numerous businesses have been established and construction projects are on full-swing.



## WHO ARE THE TARGET AUDIENCE?:

### 1. ENGINEERING FIRMS:

-in order to prepare as to what kind of engineering techniques will they use in order to put up the structure(s) safely and effectively

### 2. INVESTORS:

-in order to strategize their marketing and finances on how to allocate resources for the success of the project

### 3. GOVERNMENT:

-in order to pass laws that would moderate the trade and competition so that illicit practices can be prevented

### 4. SMALL BUSINESS OWNERS:

-in order to adjust their businesses as to how they can take advantage of the upcoming competition and demand





## 5. SERVICE BUSINESSES:

-in order to market themselves to different sectors as to how they can be of service to them and so that they can price themselves accordingly according to demand

## 6. ENERGY SECTOR:

-in order for the electric cooperatives to prepare their lines and power supply to different kinds of electrical loads in order to ensure smooth grid operations

## 7. CITIZENS:

-citizens can be aware of their options in case they need something to have

## 8. MEDIA:

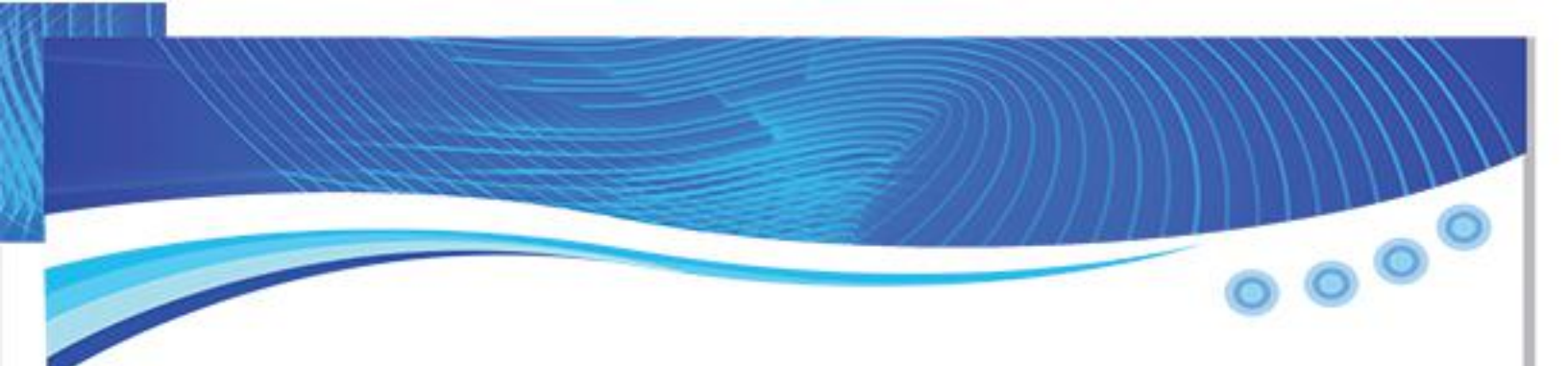
-media can take advantage of advertising so that they can strategically place their ads to the target customers

## 9. TOURISTS:

-tourist can take advantage of surrounding amenities in order for them to enjoy and make the most of their vacation

## **Negros\_Occidental\_Cities.csv:**

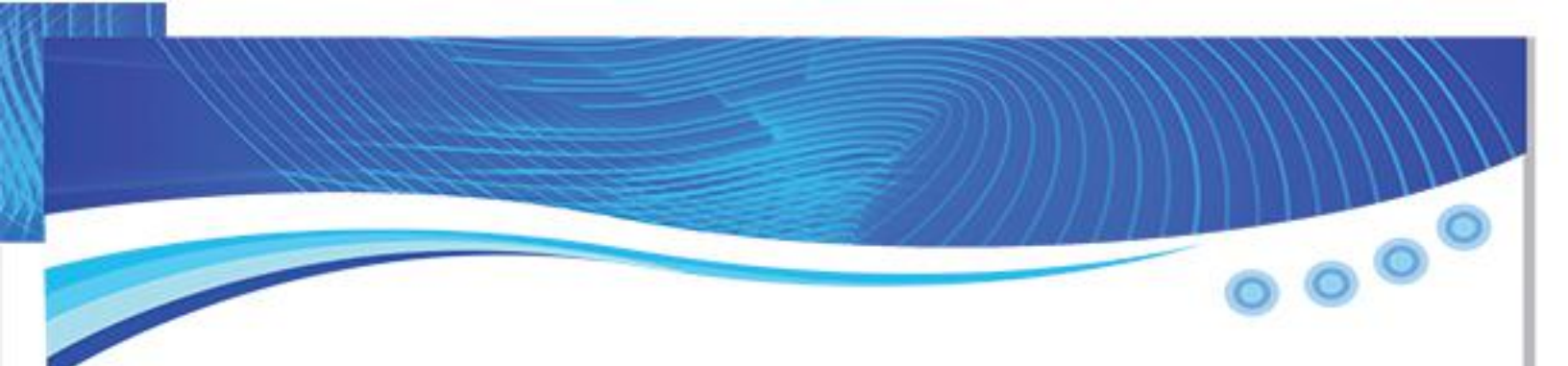
I got this dataset by extracting the list of cities in Negros Occidental from the Philippines Statistics Authority website, and at the same time using wikipedia in order to get their decimal coordinates. And after getting the name of the cities and their corresponding coordinates I turn them into a csv by googlesheets. After that I import the csv to my jupyter notebook and process them further in order to get the detailed informations (latitude, longitude venues and each of their respective latitudes and longitudes), after getting their detailed information, I create a bar chart to see which city have the most venues, and from their onwards make my comprehensive analysis about that city.



## **Bacolod City Article.txt:**

This is a text file which I will later use for my word cloud, my word cloud will help me to gain additional insights in order reinforce my decision at the conclusion.

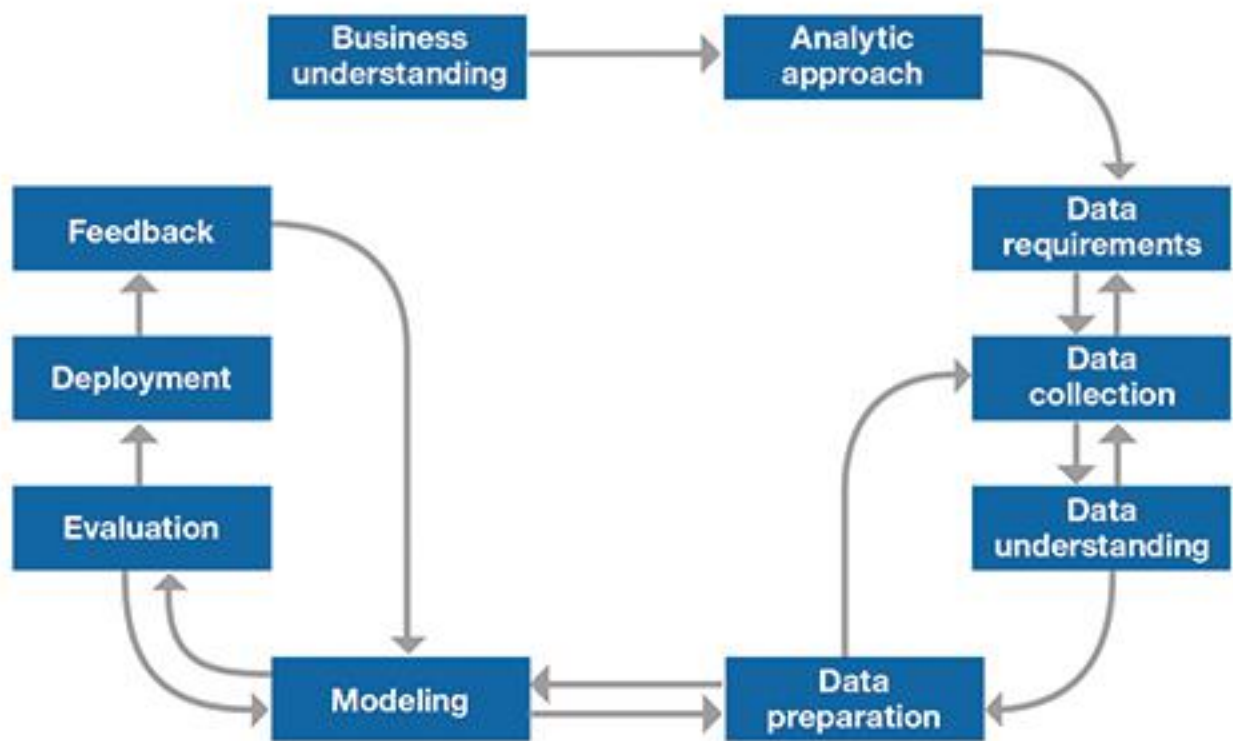




## **negros\_cities\_with\_venues (extracted via foursquare API) .CSV:**

This dataset will be my main data for analysis, I will use a k-means algorithm in order to cluster them and after clustering them I will see which cities have similarities and have the most number of venues, after that, I will make a pie chart in order to see what percentage each venue contributes to the total count, after that I will map each venue by using folium and decide on what type of business that I would want and where should I place it in order to maximize profits.

# METHODOLOGY



MY GRADE AND YOUR COMMENTS  
WILL SERVE AS MY FEEDBACK



## **Business Understanding:**

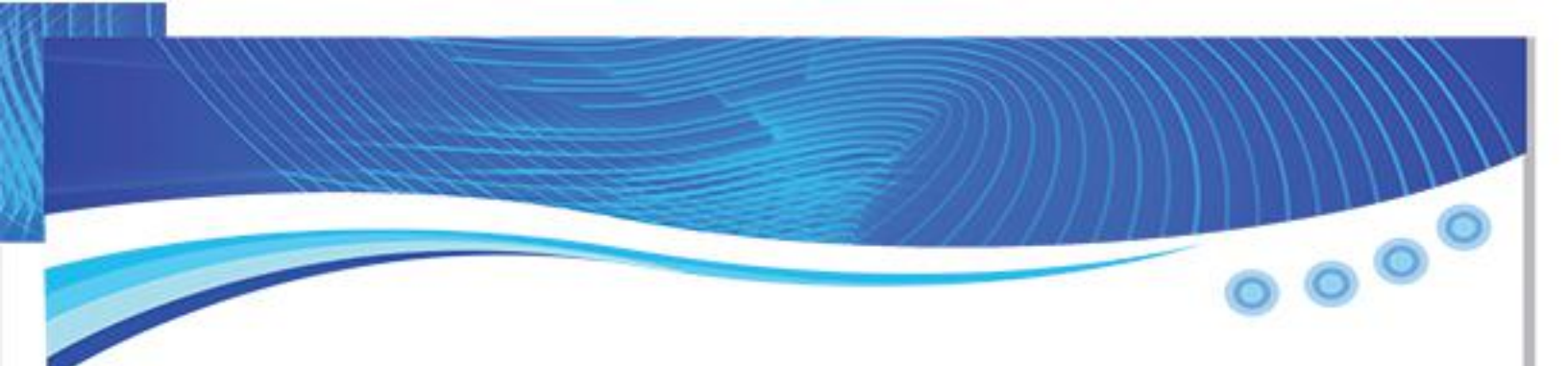
This is where I brainstormed about the problem, and I wanted the business problem to be applicable to my country and at the same time related to the latest trend. I decided that my main target audience would be engineering firms and investors, since they are currently the sectors that my business problem needs, for they have the resources and skills necessary for the completion of the project. In this phase of the project I decided to set my goals, and as to where should I get my data.



## **Analytic Approach:**

In this phase of the project I brainstormed about possible approaches as to how I would use the data in order to solve the problem. I also have thought about what visualisation techniques I would use in order to better relate my findings to the stakeholders involved.

I decided to use a relationship modeling technique, namely k-means since my problem required clustering in order to distinguish the ideal candidate(s).



## **Data requirements:**

In this part of the project I brainstormed on what type of data that I need that would be highly relevant to my problem.

I decided to use geographical (decimal coordinates of each city and venue) and textual data in order to derive insights for my final analysis.

## **Data collection:**

In this phase of the project I decided to get my geographical data from the Philippine Statistics Authority website, my coordinates from wikipedia, article from business insider website and descriptive data from the foursquare API.





## **Data understanding:**

I decided to format the initial data into a csv file by using google sheets since they are highly fragmented and come from multiple sources, also the initial data were not that many therefore manually encoding them would be the most efficient method.

I copied the article into my notepad and saved it as a txt file format for further processing.



## **Data preparation:**

After gathering the initial data I first refined the “Negros Occidental\_Cities.csv” data set in order to extract the latitude and longitude of each city. Next by using the foursquare API I extracted additional data as json, and refined it further so that I can convert it to a pandas dataframe “negros\_cities\_with\_venues (extracted via foursquare API).csv” with the necessary data for modeling later on.

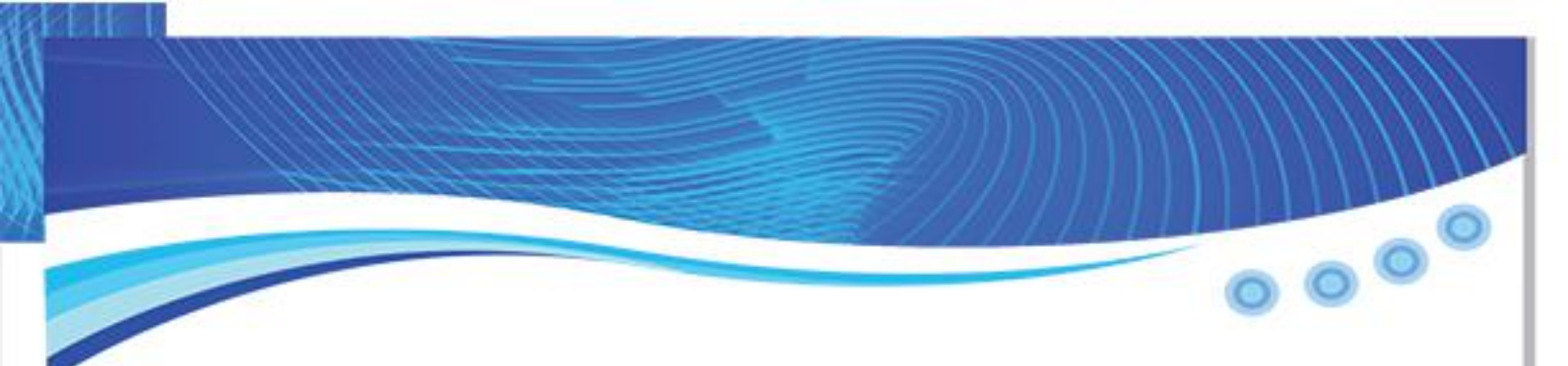
Next is I process the “Bacolod City Article.txt” so that I can model it later on for my word cloud.



## Modeling:

Using the “negros\_cities\_with\_venues (extracted via foursquare API).csv” I manipulate the data in order to use a k-means algorithm in order to cluster the cities, and after clustering them I found out that there is a particular cluster that had similar amenities, and to further reinforce my findings, I decided to plot each city with their corresponding number of venues. Also to further inspect if geography plays a role, I decided to map each city using folium.

After seeing that Bacolod City has the most number of venues, I decided to generate a pie graph in order to visualise the contribution of each venues to the total venue count, in percent. This will serve as the basis as to what type of business shall I establish.



## **Modeling continued:**

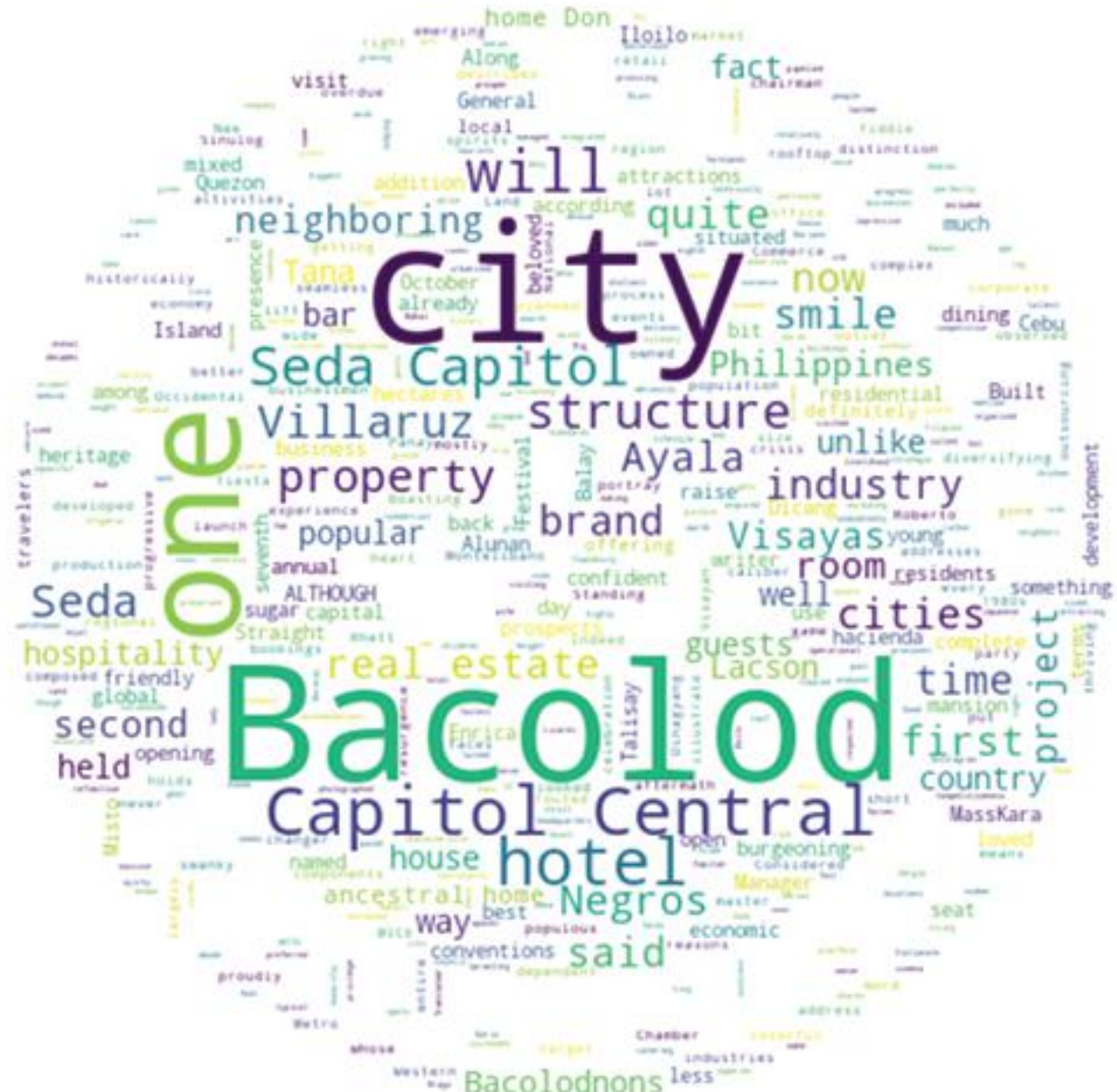
Next, I use a word cloud in order to gain additional insights as to what are the mindset of the citizens that live in the area, the most anticipated events, a peek into their daily routines, the people's mentality, and the city's economic climate in general. This will also serve to reinforce my final decision as to what type of business should I establish and, where should I place it within the city.

Lastly, I map again all the venues via folium so that I can see which location is the most favourable spot for my business.





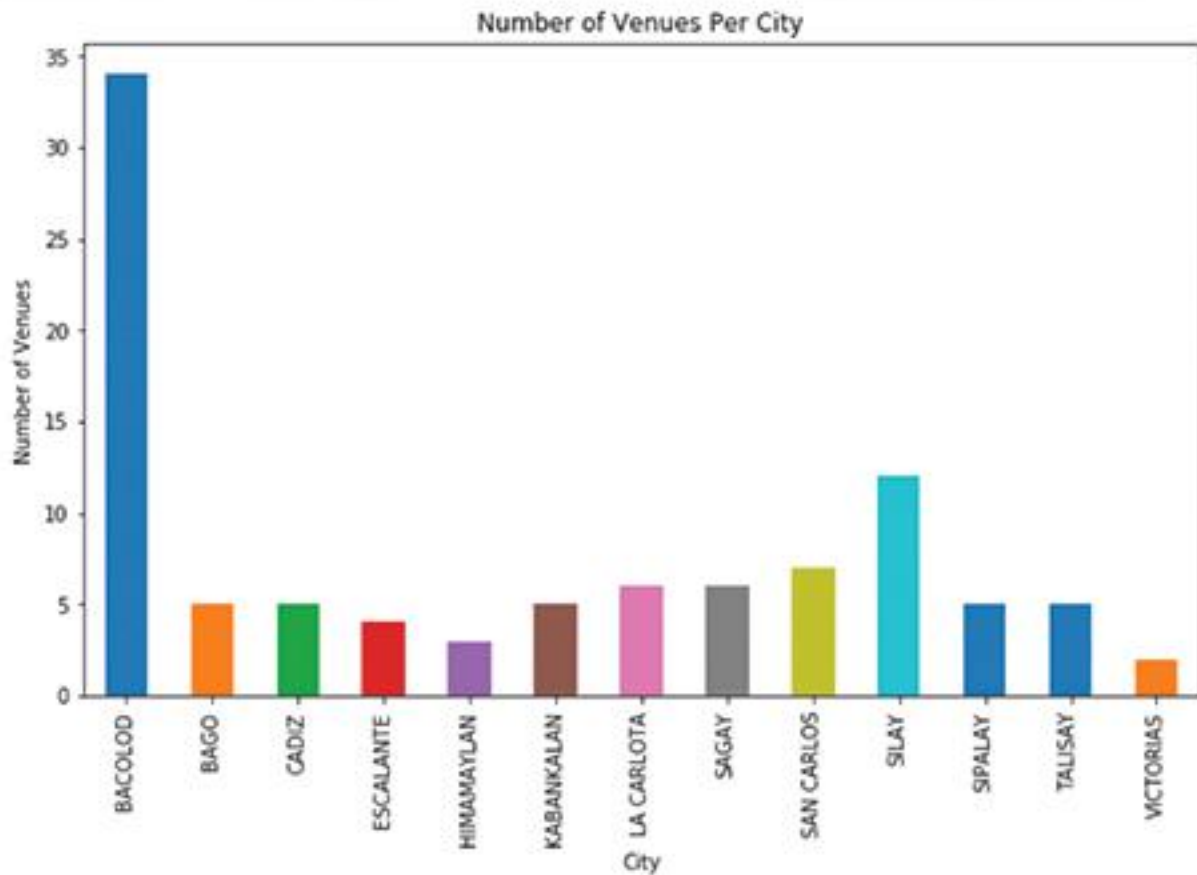
# MODELS



## WORD CLOUD



# MODELS



## BAR GRAPH

# MODELS

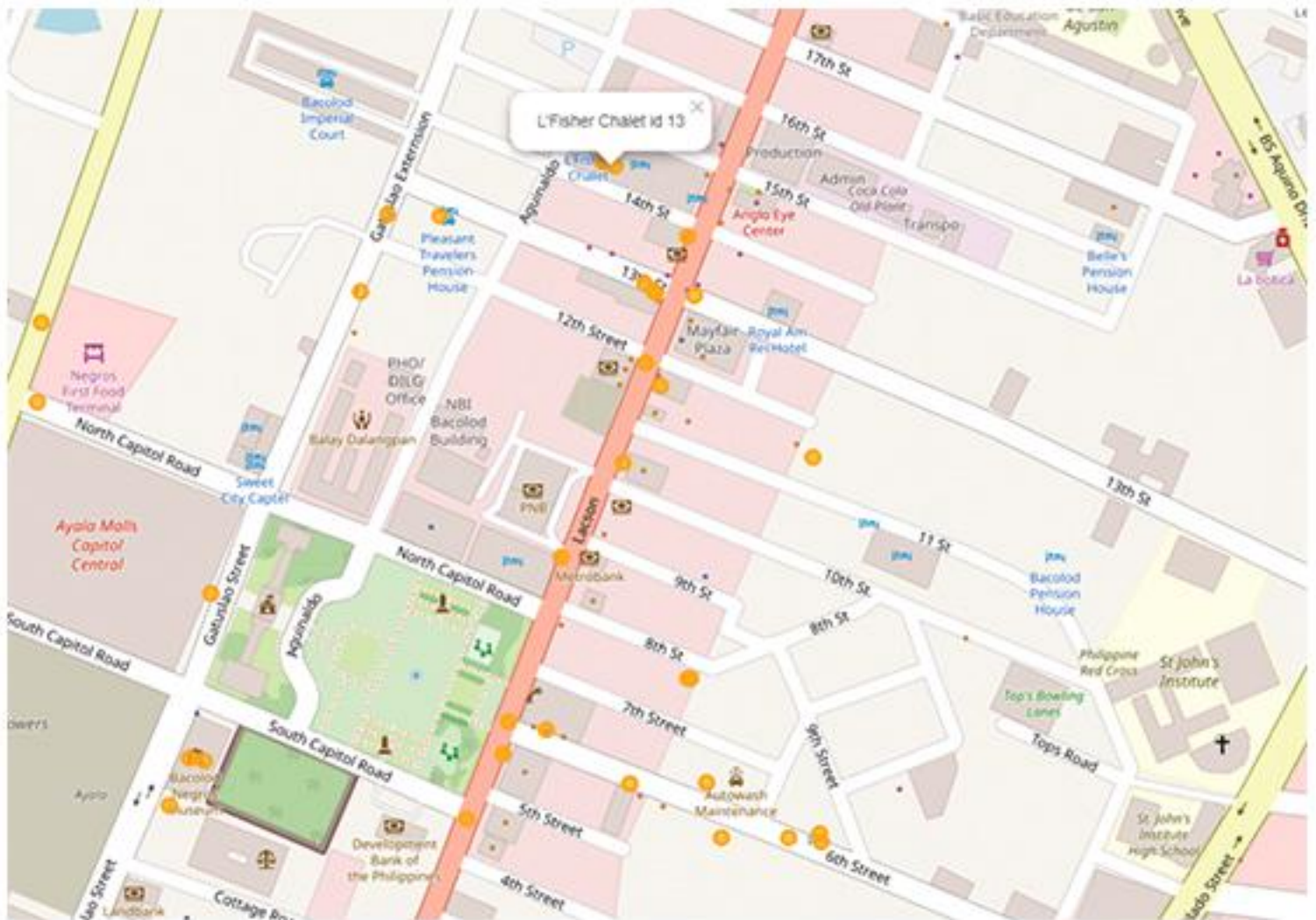


## NEGROS OCCIDENTAL MAP



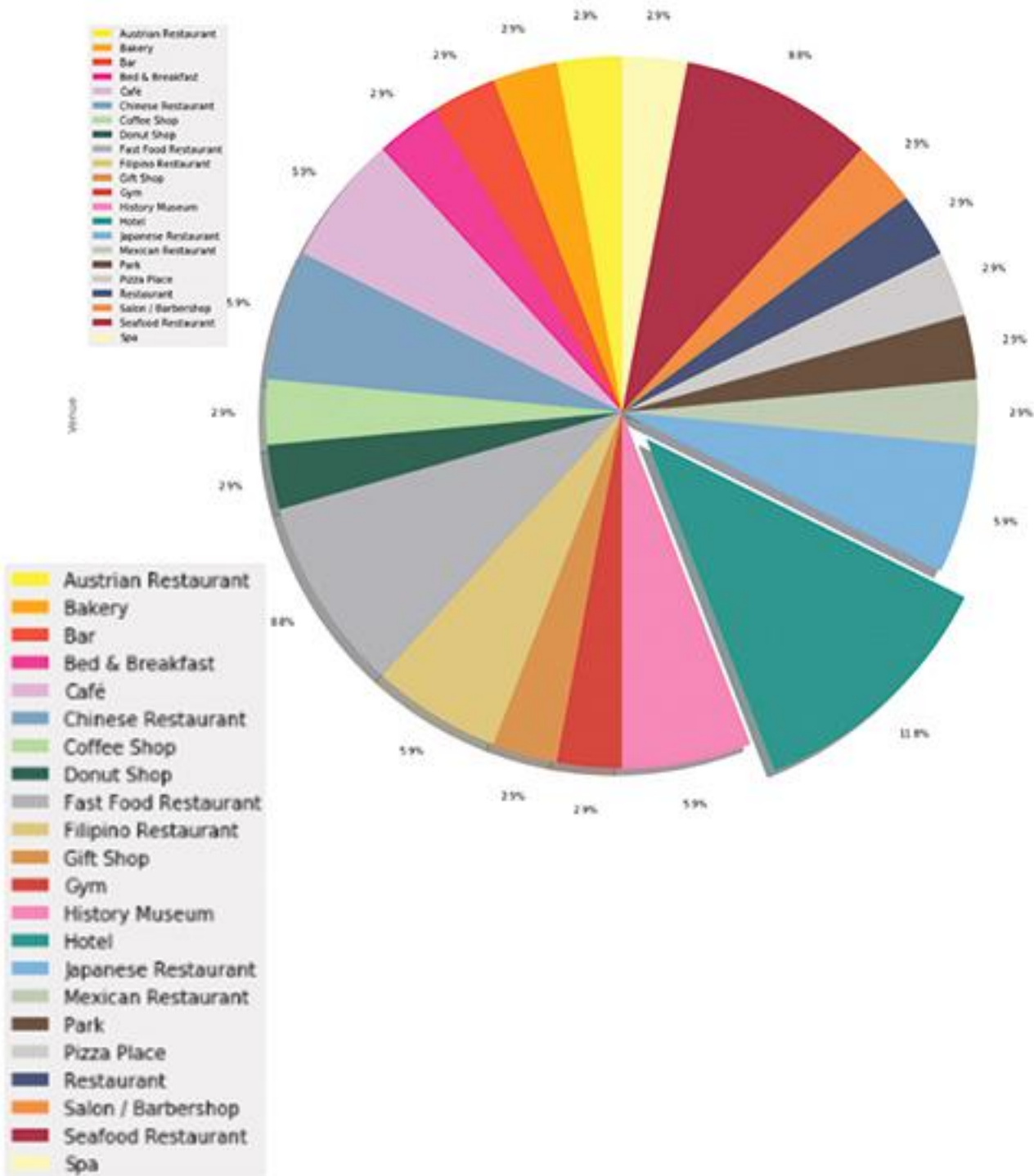
## Inferential Testing, Result or Evaluation section:

Based on the Model that was generated, we could see that the proximity of venues are not that far from each other, and based on the frequency of venues we could infer that they usually would cater to tourists, young adults, office workers and business persons.



# Inferential Testing, Result or Evaluation section:

What are the most common types of venues in Bacolod City





# Discussion and Deployment

Based on the model that was generated and from our inference, my suggestion would be that a laundry wash service business would be ideal and its location near capitol lagoon, since most of the businesses venues here are for socializing, tourists, and employees therefore it is most likely that their stay within the city is only temporary, and restaurants would need a quick cleaning service for their tablecloths, and other fabrics that they would use in their business.

Also I observed that there is no laundry business that is available nearby, and hotel laundry services are likely to be expensive. Therefore you could offer a more competitive price and also you could ask for partnerships with different businesses nearby e.g. hotels and restaurants, giving them discount perks and priority based on their volume.



# Discussion



**Lacson Street**

# Discussion



	Austrian Restaurant
	Bakery
	Bar
	Bed & Breakfast
	Café
	Chinese Restaurant
	Coffee Shop
	Donut Shop
	Fast Food Restaurant
	Filipino Restaurant
	Gift Shop
	Gym
	History Museum
	Hotel
	Japanese Restaurant
	Mexican Restaurant
	Park
	Pizza Place
	Restaurant
	Salon / Barbershop
	Seafood Restaurant
	Spa

## Venues in Bacolod City

# Conclusion

Using data science methodologies, I was able to derive to my laundry business decision near capitol lagoon. After seeing the data and examining them carefully with the aid of vusal models, I was able to infer that a laundry business would be ideal.

However if we would use traditional survey and study methods, there is a big possibility that I would not be able to come up with this idea, or the result would be too late since traditional techniques are time consuming and costly. Also traditional methods are prone to error since people who would conduct the study are prone to personal biases, and they have lapses in concentration. Lastly the aid of visualization tools and machine learning methodologies, have greatly accelerated the the decision making process since pictures speak a thousand words and computers compute way faster than humans, and are less prone to error.



# Conclusion

The main takeaway in this project is the value of data science in the decision making process. Data Science if done right does not only solve big and important problems, they may also give us a new perspective on things, since we may fail to see the big picture due to our personal biases, data science is also fast in calculations, precise in producing results and accurate in their predictions. Data Science truly is the future not only of businesses but of the world.

# SPECIAL THANKS:



THANK YOU COURSERA AND IBM FOR  
MAKING THIS COURSE POSSIBLE

## Instructors



**Joseph Santarcangelo**

Ph.D., Data Scientist at IBM  
IBM Developer Skills Network



**Alex Aklson**

Ph.D., Data Scientist  
IBM Developer Skills Network



**Rav Ahuja**

Data Science Program Manager  
IBM



**SAEED AGHABOZORGI**

Ph.D., Sr. Data Scientist  
IBM Developer Skills Network



**Polong Lin**

Data Scientist  
IBM Developer Skills Network

THANK YOU INSTRUCTORS FOR IMPARTING  
YOUR PRECIOUS KNOWLEDGE AND FOR GUIDING  
ME ALONG THE WAY!!!