



RESTAURANTS ANALYSIS

A Report on the Analysis of Restaurants in Chandigarh

Abstract

Restaurant business has seen a great surge after the booming of online services such as recommendation system, Online Order, Home Delivery etc. Due to increased business it is paramount to know the existing market for Restaurants in a city.

Recommendation to business owners about places perfect for opening a new Restaurant with a particular Cuisine where the owner would get best traffic for the Restaurant.

I build an Analysis system using Zomato Data for the business owners to analyze the market based on Cuisines, Nightlife, Popularity and Restaurants types. The Insights derived from the analysis is helpful in understanding the existing business and how or which area lacks the service.

Amit Maindola
Maindola.amit@gmail.com

Introduction

Data Science is a process of using the data to understand various parameters associated with the Hypothesis/Models of a problem, which we try to validate through the data. It's like to use storytelling to generate insight, insights which can be used to make strategic choices by an Organization or an Institution. In Data Science we perform the data extraction from various source so we can study the data.

Since the Dawn of cloud computing Data Science changed the way people find the products, information, solutions and even other people. Earlier when it was very tedious job to find the information/solution to a problem, now the good news is we have lots and lots of data available mostly in the web. As for the availability of data we can perform various analysis and visualization to the problem statement for which we are working.

As we have lots of data available it is the problem to download, clean and format data from various available sources either from Database, web, csv or Excel files. We need data to be in a meaningful format before making the necessary analysis and visualization. The goal is to generate meaningful insights or a recommendation system or a product solution to the customers for a richer experience based on the Business requirement/Problem. With burgeoning consumerism buoyed by emergence of web, buyers are presented with increasing range of choices while sellers are being faced with the challenge of personalizing their advertising efforts.

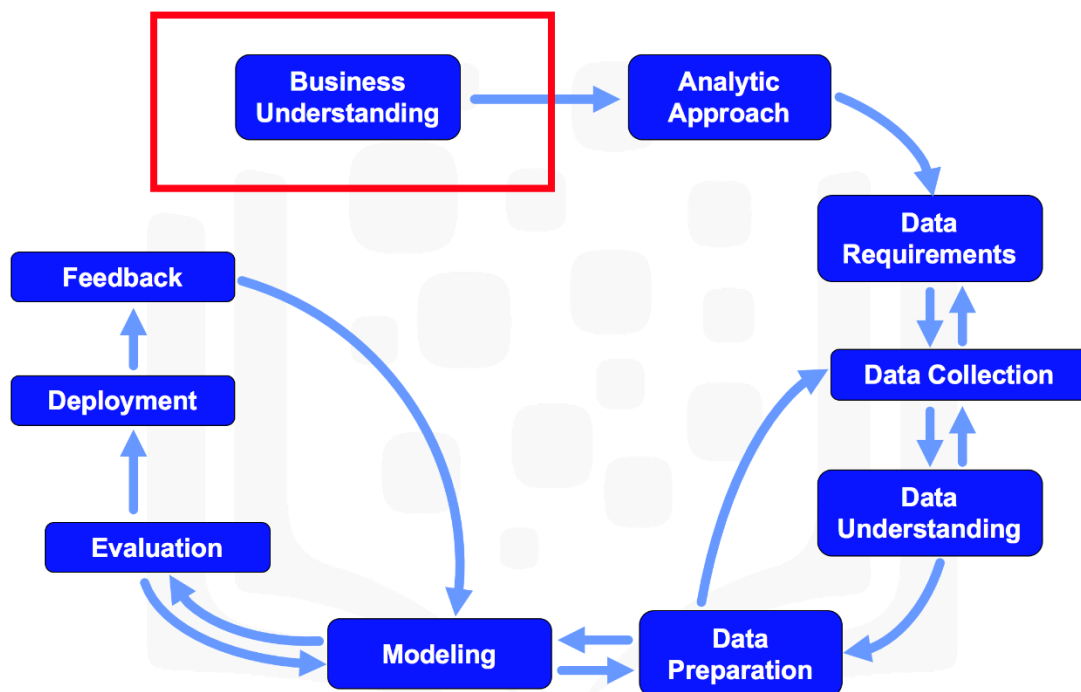
Data science is relevant today because we have tons of data available. We used to worry about lack of data. Now we have a data deluge. In the past, we didn't have algorithms, now we have algorithms. In the past, the software was expensive, now it's open source and free. In the past, we couldn't store large amounts of data, now for a fraction of a cost, we can have gazillions of datasets for a very low cost. So, the tools to work with data, the variability of data, and the ability to store and analyze data, it's all cheap, it's all available, it's all ubiquitous, it's here. There's never been a better time to be a data scientist.

Business Problem

Due to Increased connectivity and availability of Data it is now possible to extract and derive the useful Insights from the available data. We can meticulously study the existing Market of our Hypothesis/Model/Product. We have to establish our facts with the data and visualization.

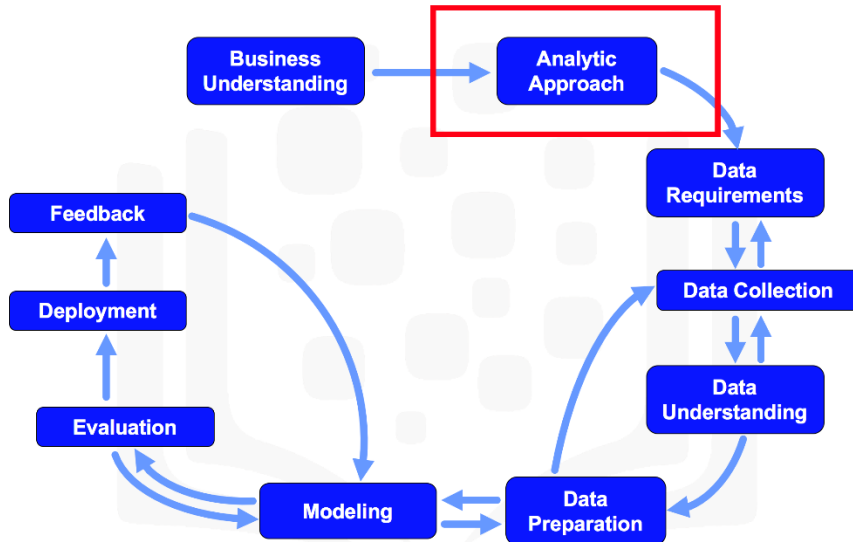
Our requirement is to build a model which can analyze the existing Restaurants in a city and can suggest where a new Restaurant can be opened depending on the type of Cuisine or the Customer category or the Initial Budget we have for the Business.

A detailed models has to be build considering the popularity of the locality, cuisines, price range and the customer category we can target. To fulfill the Business Requirement we have to study the existing data about the market, finding out the various parameters associated with the Business which are to be considered. The derived insights can help us in providing better customer service which will lead to the better customer satisfaction. Based on the Model we can determine which Cuisines we can involve in the Menu which can affect the Traffic in the new Restaurant.

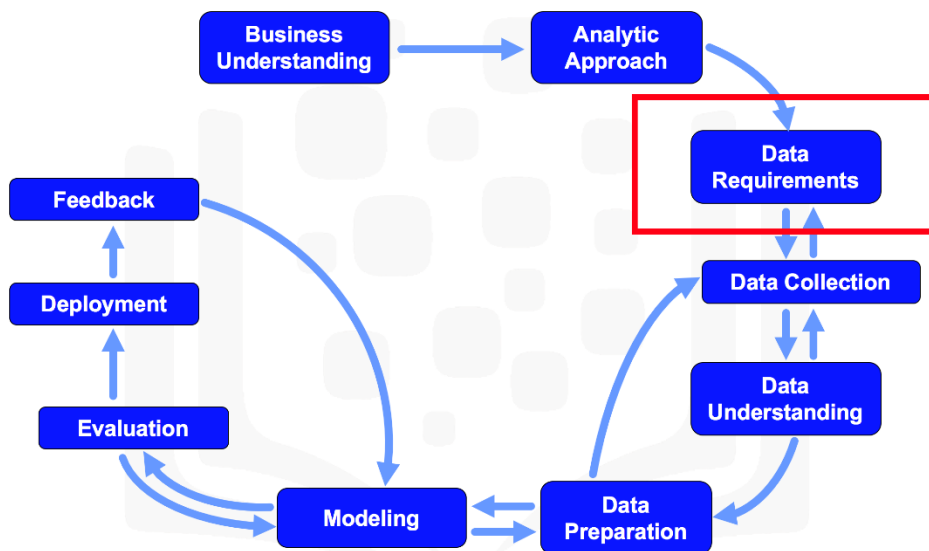


Methodology

At this point, we realize that automating the process of determining the cuisine of a given dish is not a straightforward problem as we need to come up with a way that is very robust to the many cuisines and their variations.



We have to perform various analytics approach to get the right data for our Model. Once we have got the right data, we can go ahead and apply the appropriate Algorithm.



Data

Fetching the data:

1. City areas has been fetched from Website : mapsofindia
<https://www.mapsofindia.com/chandigarh/localities/>
2. Location of the areas are derived using the Google API.
3. Restaurant data has been collected from the Zomato API in the form of .json files.

Data Processing:

Data Processing has been done on the following categories:

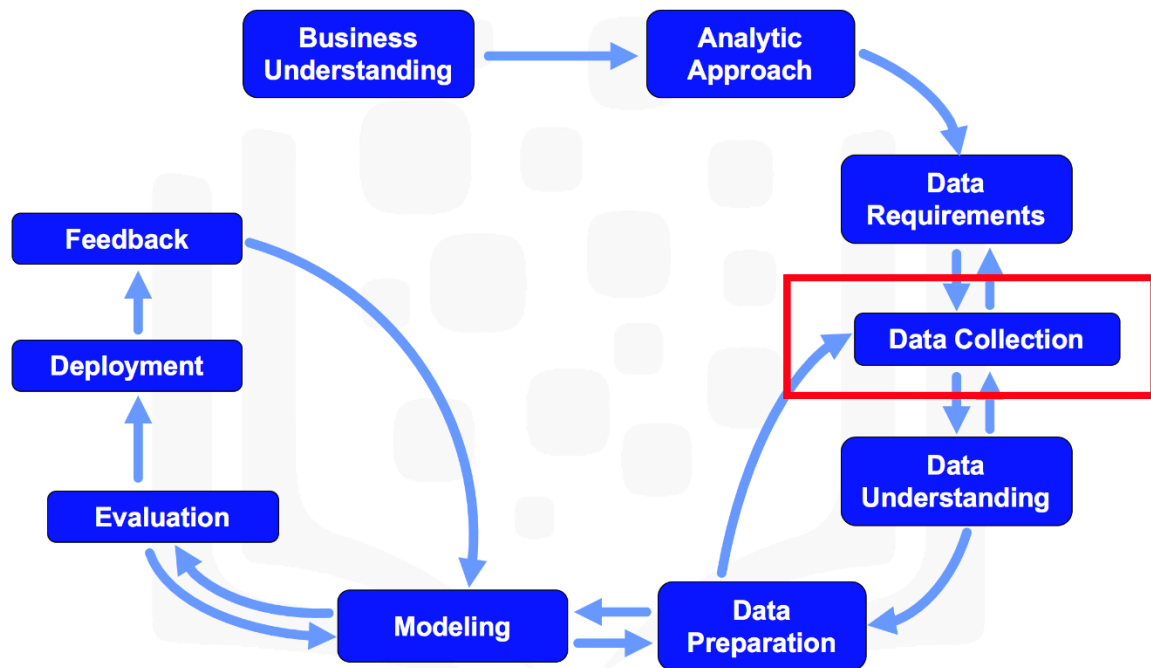
1. Cuisine
2. City
3. Location
4. Popularity of the location
5. Price range of Restaurant

The collected data has been stored in the Comma Separated Value File (CSV). Each restaurant in the dataset is uniquely identified by its Restaurant Id. Every Restaurant contains the following variables:

- **City:** City in which restaurant is located
- **Area:** Area of the City in which restaurant is located
- **Area Latitude:** Latitude coordinate of the restaurant's location
- **Area Longitude:** Longitude coordinate of the restaurant's location
- **Popularity Index:** Popularity of the area of city
- **Nightlife Index:** Popularity at night of the area of city
- **Top Cuisines:** Famous Cuisines of the area of city
- **Restaurant Id:** Unique id of every restaurant across various cities of the world
- **Restaurant Name:** Name of the restaurant
- **Cuisines:** Cuisines offered by the restaurant
- **Average Cost for two:** Cost for two people in different currencies
- **Price range:** range of price of food
- **Has Online delivery:** yes/ no
- **Has Table booking:** yes/no
- **Aggregate Rating:** Average rating out of 5
- **Rating text:** text on the basis of rating of rating
- **Votes:** Number of ratings casted by people
- **Locality:** Number of ratings casted by people
- **Address:** Location in the city
- **Latitude:** Latitude coordinate of the restaurant's location
- **Longitude:** Longitude coordinate of the restaurant's location

Now that the data collection stage is complete, data scientists typically use descriptive statistics and visualization techniques to better understand the data and get acquainted with it. Data scientists, essentially, explore the data to:

1. understand its content,
2. assess its quality,
3. discover any interesting preliminary insights, and,
4. Determine whether additional data is necessary to fill any gaps in the data.



Below is the Sample of our collected Data in the JSON format.

```

▼ root: {} 4 keys
  ► location: {} 9 keys
  ► popularity: {} 9 keys
  link: "https://www.zomato.com/chandigarh/sector-34-chandigarh-restaurants"
  ▼ nearby_restaurants: [] 9 items
    ▼ 0: {} 1 key
      ▼ restaurant: {} 32 keys
        ► R: {} 1 key
          apikey: "f4c850268294012ef371d71202772caa"
          id: "123806"
          name: "Chandigarh Flavours"
          url: "https://www.zomato.com/chandigarh/chandigarh-flavours-sector-46?utm_source=api_basic_user&utm_medium=api&utm_campaign=v2.1"
        ► location: {} 9 keys
          switch_to_order_menu: 0
          cuisines: "North Indian, Chinese"
          average_cost_for_two: 550
          price_range: 2
          currency: "Rs."
          offers: [] 0 items

```

```
# View the categories DataFrame
categories_df = pd.read_csv(os.path.join(csv_path, "zomato_categories.csv"), index_col='id')
categories_df.head()
```

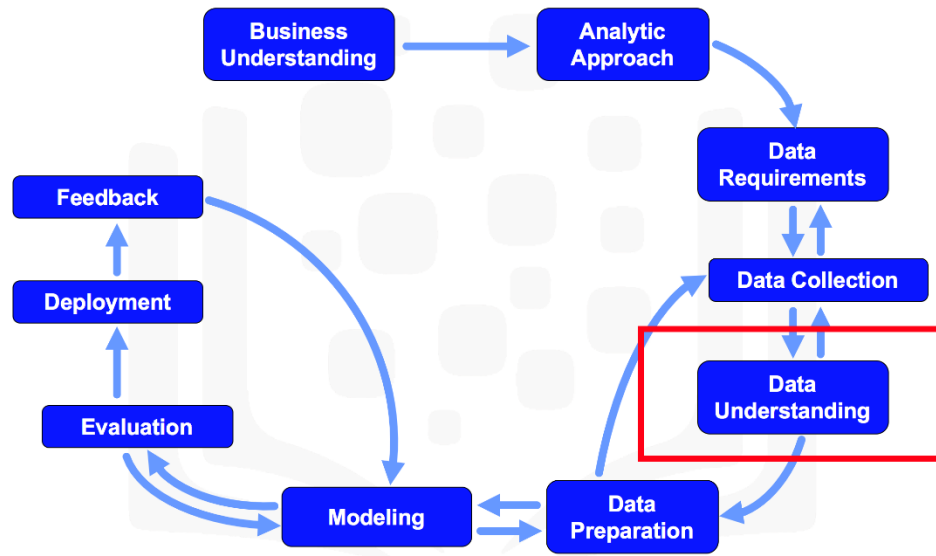
	name
id	
1	Delivery
2	Dine-out
3	Nightlife
4	Catching-up
5	Takeaway

Filter...

```
▼ root: {} 4 keys
  results_found: 257
  results_start: 0
  results_shown: 20
▼ restaurants: [] 20 items
  ▼ 0: {} 1 key
    ▼ restaurant: {} 34 keys
      ▶ R: {} 1 key
        apikey: "f4c850268294012ef371d71202772caa"
        id: "121820"
        name: "24/7 All Day Dining - The Lalit Chandigarh"
        url: "https://www.zomato.com/24-7RestaurantChandigarh?utm_source=api_basic_user&utm_medium=api&utm_campaign=v2.1"
      ▶ location: {} 9 keys
        switch_to_order_menu: 0
        cuisines: "Italian, North Indian"
        average_cost_for_two: 3000
        price_range: 4
        currency: "Rs."
        offers: [] 0 items
        opentable_support: 0
        is_zomato_book_res: 1
        mezzo_provider: "ZOMATO_BOOK"
        is_book_form_web_view: 0
        book_form_web_view_url: ""
        thumb: "https://b.zmtcdn.com/data/res_imagery/121820_RESTAURANT_1411202922_3f50d1a685ff60293c1e0633fd4a4bda_c.jpg?fit=around%7C200%3A200&crop=200%3A200%3B%2A%2C%2A"
```

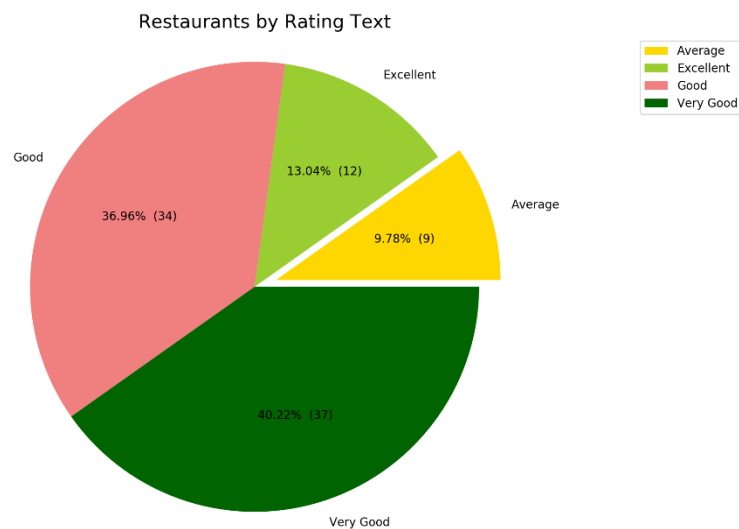
Data Understanding

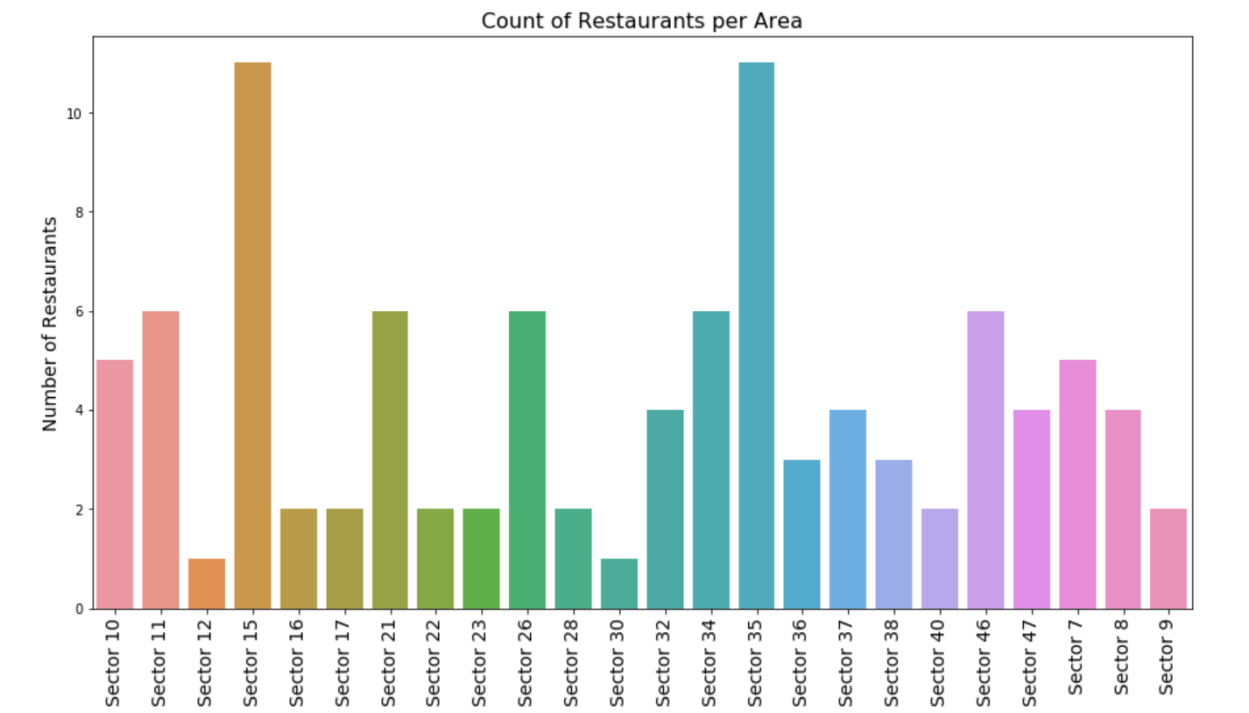
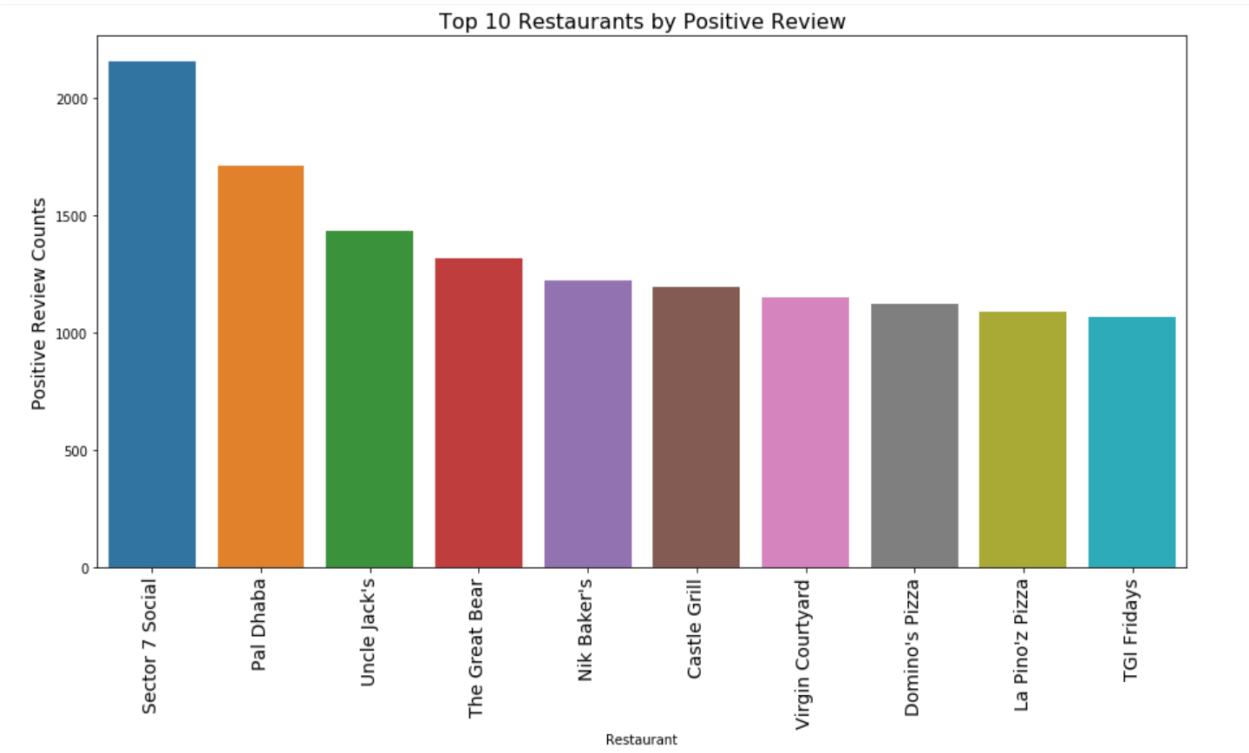
Now as we have collected the Data from Various sources we have to understand the Data through various methods. We will use Python for the Data Wrangling, cleaning and manipulation.



We will be doing various techniques to convert the JSON data into Tabular format so we can perform the visualization on it.

```
# Extract data from the JSON file
categories = [cat['categories'] for cat in json.load(open(os.path.join(raw_path, 'zomato_categories.json'), "r"))['result']['categories']]
# Save data to a CSV file
pd.DataFrame(data=categories).to_csv(os.path.join(csv_path, "zomato_categories.csv"), index=False)
print("File saved.")
```





Data Preparation

Now as we have the understanding of the data, we can extract the data into the format which will be easy for the Clustering Algorithm. We would be performing One Hot Encoding and other manipulations to bring the data into the format along with the required parameters.

Observation

Conclusion