Defining the best area for an Italian restaurant

Paulo Vitor Peclat de Oliveira Tavares

April 16, 2021

1. Introduction

1.1. Backgroud

The objective of this work is to apply the tools learned during the IBM Data science course. This report will present the steps developed during this project and the results obtained.

1.2. Problem

For this project, I chose a hypothetical business problem. The question that we are trying to answer is the following. A successful owner of multiple mid to high-end restaurants decided to open a new restaurant in São Paulo, Brasil. Having visited the city many times in recent years, he couldn't disregard the big boom in gastronomy. He is keen on opening a new unit, which will focus on the Italian kitchen. Taking into account the price level at which the restaurant will operate, the intent is to find an optimal location in an area, where gastronomy is booming and which is easily accessible for tourists and for wealthier local citizens as well.

2. Data acquisition and cleaning

2.1. Data sources

To perform this analysis, we will need the following data:

1. List of the districts of São Paulo;

2. Geo-coordinates of the districts in São Paulo;

3. Top venues of districts.

Listo f districts will be obtained from wikipedia:

(https://pt.wikipedia.org/wiki/Lista_dos_distritos_de_São_Paulo_por_população)

Geo-coordinates of districts will be obtained with the help of the geocoder tool in the notebook. Top venues data will be obtained from Foursquare through an API.

3. Methodology

For this Project , the K-means machine learning technique was applied for creating clusters of districts and determine the best cluster for the italian restaurant. The silhouette score was used for choosing the optimal number of clusters.

After data collection, a list of 96 districts was obtained, using the geocode the latitude and longitude of these districts were collected. After an evaluation, it was observed that the districts of Pedreira, Marsilac, Paralheiros and Grajaú were outside the area of the city of São Paulo. Therefore, it was decided to exclude these districts. The districts can be seen in figure 1.

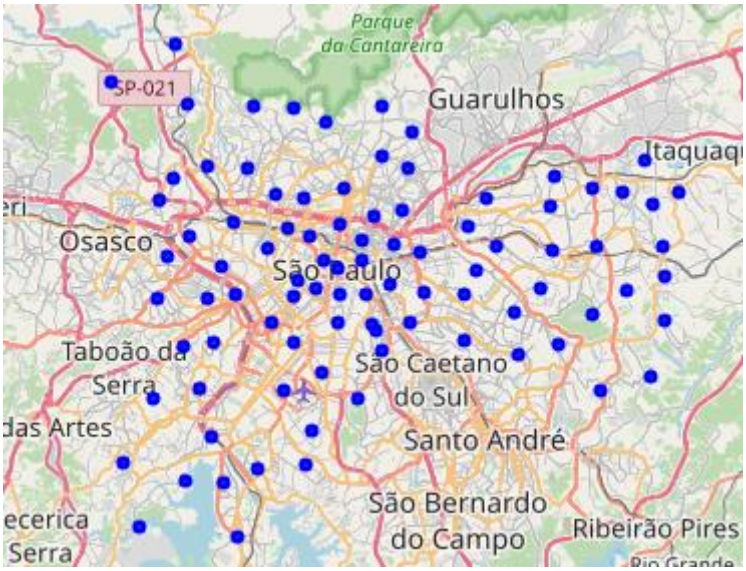

Figure 1: Districts of São Paulo

In the next step of the analysis, the districts were explored in greater detail. It means venues were collected for each district via Foursquare API. The data from Foursquare is received in json format. After arranging the data, we have up to 100 venues for each district. Venues are collected within a radius of 2000 meters from the point of district coordinates.

To improve the model and reduce the data obtained, venues that appeared only once were excluded. Applying that reduction, there were 286 uniques venues categories, compared to 359 categories before reduction,

Considering that São Paulo is one of the biggest city in the world, it may be wiser to reduce the area to be analyze, firstly the districts were clustered based on their locations. As it can be seen in figure 2. After analyzing the data, it was decided to use the districts contained in cluster in red, purple and acquamarine. The final districts can be seen in figure 3.

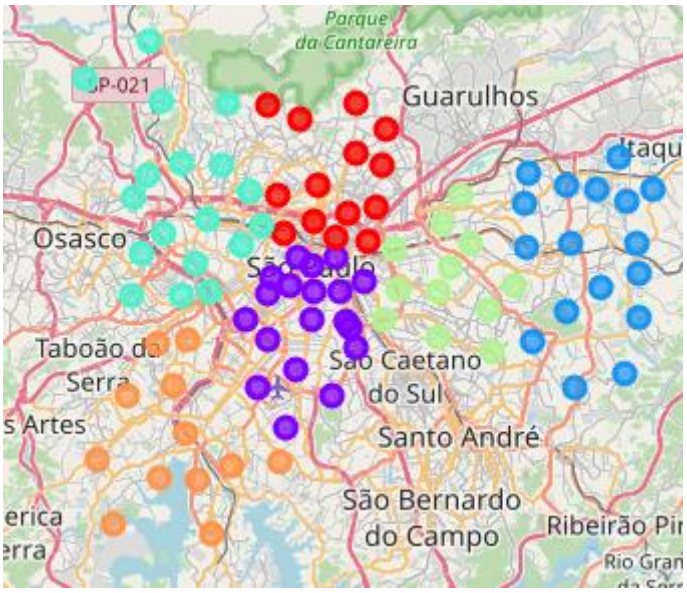

Figure 2: Clusters based on locations

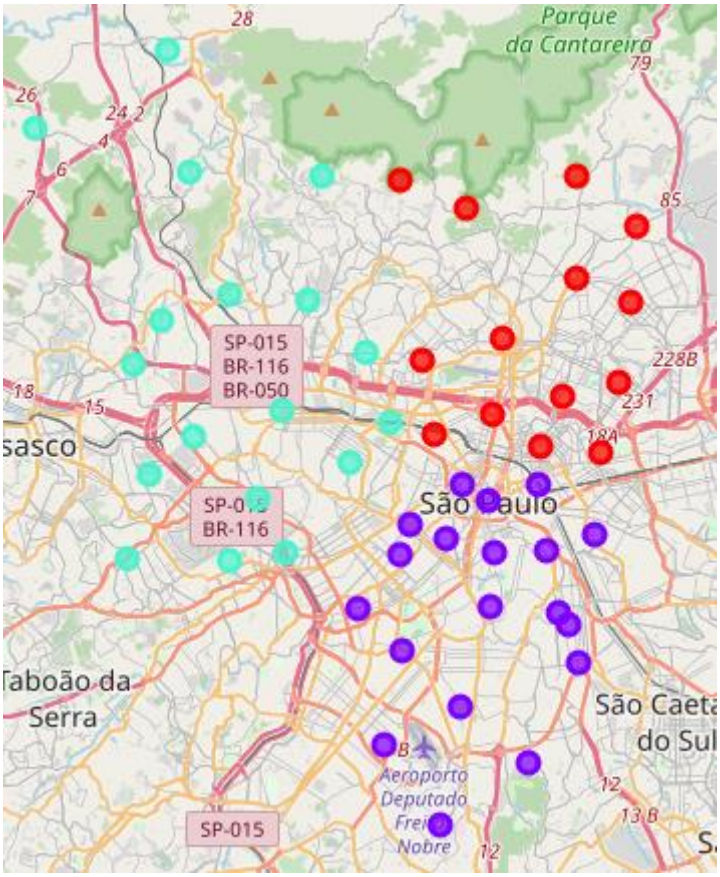

Figure 3: Final districts

For analyzing the districts, we focus on venue categories. For that purpose, we use the one-hot encoding. This creates dummy variables for categories so the data set could be used for

machine learning. After performing manipulations with the dataset, we get the following table, which shows the top ten most common venues for each district (first four shown in the table).

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alto de Pinheiros	Restaurant	Plaza	Athletics & Sports	Gym / Fitness Center	Dog Run	Clothing Store	Cosmetics Shop	Brazilian Restaurant	Track	Health Food Store
1	Anhanguera	Bakery	Grocery Store	Convenience Store	Park	Food & Drink Shop	Gym / Fitness Center	Snack Place	Campground	Supermarket	Ice Cream Shop
4	Barra Funda	Pizza Place	Italian Restaurant	Motel	Pet Store	Dessert Shop	Restaurant	Hotel	Theater	Farmers Market	Café
5	Bela Vista	Ice Cream Shop	Theater	Italian Restaurant	Coffee Shop	Café	Hotel	Movie Theater	Cultural Center	Bakery	Bookstore
6	Belém	Pizza Place	Bar	Restaurant	Snack Place	Café	Chocolate Shop	Churrascaria	Ice Cream Shop	Pet Store	Bakery

Figure 4: Ten most common venues for each district.

3.1. Clustering

Now that we have the dataset ready, we perform clustering. For this, unsupervised machine learning technique will be used based on K-means. For K-means clustering, we need to decide on the number of clusters that we want to use. To avoid the trial and error approach, the silhouette score was used. The following graph shows the silhouette scores for a range of clusters variations. For this Project, It was select 3 clusters.

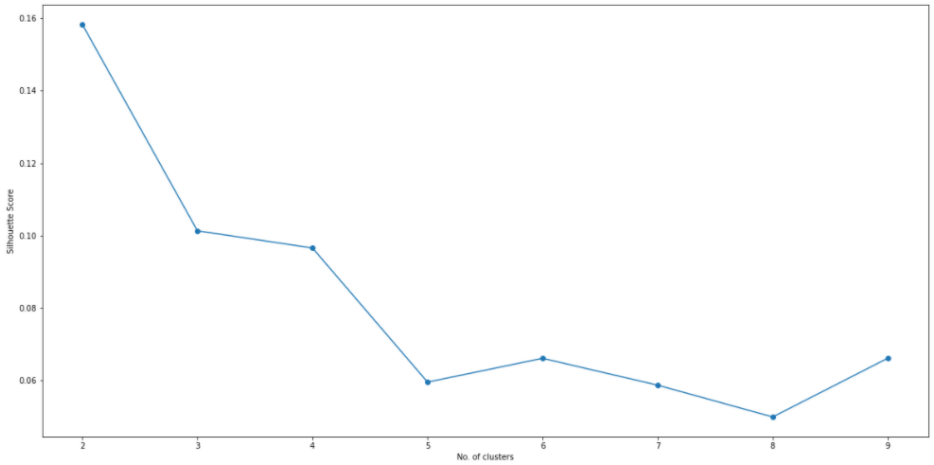

Figure 5: Number of clusters.

The clusters based on the venues can be visualized on the figure 6.

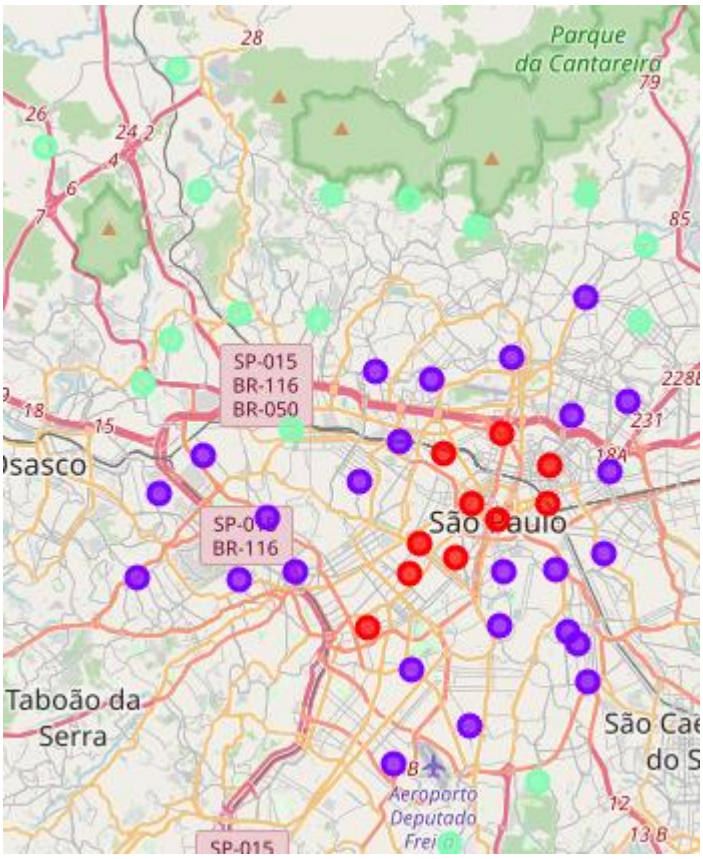

Figure 6: Clusters based on venues

4. Results

Looking at the cluster data, we can see that cluster 1 is the one that we are the most interested in.

Cluster 2 is a district where top gastronomy is not really represented, it is more common fast food and regular restaurant.

	Distrito	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Sacomã	Pizza Place	Snack Place	Gym	Bakery	Bar	Burger Joint	Farmers Market	Dessert Shop	Middle Eastern Restaurant	Candy Store
29	Saúde	Bakery	Dessert Shop	Pizza Place	Japanese Restaurant	Burger Joint	Pharmacy	Fruit & Vegetable Store	Coffee Shop	Restaurant	Brazilian Restaurant
30	Vila Mariana	Burger Joint	Hostel	Gym / Fitness Center	Bakery	Pizza Place	Food Truck	Pet Store	Dance Studio	Italian Restaurant	Ice Cream Shop
37	Santana	Burger Joint	Japanese Restaurant	Pizza Place	Dessert Shop	Bar	Gym / Fitness Center	Bakery	Middle Eastern Restaurant	Fruit & Vegetable Store	Italian Restaurant
38	Rio Pequeno	Pizza Place	Pharmacy	Bakery	Gym	Japanese Restaurant	Dessert Shop	Brazilian Restaurant	Bar	Italian Restaurant	Candy Store
40	Vila Maria	Gym / Fitness Center	Brazilian Restaurant	Pharmacy	Bar	Bakery	Pizza Place	Restaurant	Snack Place	Chocolate Shop	Churrascaria
41	Perdizes	Dessert Shop	Pet Store	Pharmacy	Pizza Place	Bar	Gym	Italian Restaurant	Plaza	Ice Cream Shop	Bakery
45	Ipiranga	Pet Store	Pizza Place	Bar	Coffee Shop	Candy Store	Snack Place	Brazilian Restaurant	Dessert Shop	Pharmacy	Bakery
51	Raposo Tavares	Bakery	Burger Joint	Pet Store	Bar	Pizza Place	Coffee Shop	Candy Store	Snack Place	Brazilian Restaurant	Supermarket
52	Tucuruvi	Pizza Place	Bakery	Bar	Dessert Shop	Burger Joint	Gym / Fitness Center	Pet Store	Italian Restaurant	Dance Studio	Market
61	Casa Verde	Gym / Fitness Center	Japanese Restaurant	Bar	Bakery	Restaurant	Gym	Dessert Shop	Pet Store	Burger Joint	Music Venue
65	Moema	Supermarket	Ice Cream Shop	Snack Place	Dessert Shop	Health & Beauty Service	Bar	Plaza	Salon / Barbershop	Chocolate Shop	Argentinian Restaurant
67	Limão	Bakery	Pizza Place	Bar	Gym / Fitness Center	Gym	Motel	Dessert Shop	Restaurant	Soccer Field	Pet Store
71	Liberdade	Bakery	Burger Joint	Dance Studio	Grocery Store	Gym / Fitness Center	Cosmetics Shop	Japanese Restaurant	Ice Cream Shop	Farmers Market	Pet Store
75	Pinheiros	Coffee Shop	Pizza Place	Bakery	Vegetarian / Vegan Restaurant	Art Gallery	Italian Restaurant	Cosmetics Shop	Chocolate Shop	Restaurant	Japanese Restaurant
76	Mooca	Bar	Pizza Place	Bakery	Burger Joint	Gym	Italian Restaurant	Brazilian Restaurant	BBQ Joint	Gym / Fitness Center	Pet Store
77	Campo Belo	Brazilian Restaurant	Dessert Shop	Bar	Pet Store	Bakery	Spa	Middle Eastern Restaurant	Chocolate Shop	Italian Restaurant	Pharmacy
79	Vila Guilherme	Churrascaria	Middle Eastern Restaurant	Brazilian Restaurant	Italian Restaurant	Bar	Pet Store	Snack Place	Café	Ice Cream Shop	Dessert Shop
80	Butantã	Pizza Place	Gym / Fitness Center	Martial Arts School	Ice Cream Shop	Athletics & Sports	Brazilian Restaurant	Restaurant	Science Museum	Burger Joint	Track
82	Jaguaré	Restaurant	Dessert Shop	Pizza Place	Gym	Japanese Restaurant	Park	Food Truck	Brazilian Restaurant	Steakhouse	Bar
84	Belém	Pizza Place	Bar	Restaurant	Snack Place	Café	Chocolate Shop	Churrascaria	Ice Cream Shop	Pet Store	Bakery
85	Alto de Pinheiros	Restaurant	Plaza	Athletics & Sports	Gym / Fitness Center	Dog Run	Clothing Store	Cosmetics Shop	Brazilian Restaurant	Track	Health Food Store
86	Vila Leopoldina	Dessert Shop	Pizza Place	Flower Shop	Music Venue	Restaurant	Brazilian Restaurant	Burger Joint	BBQ Joint	Supermarket	Tennis Court
88	Cambuci	Bakery	Italian Restaurant	Pizza Place	Gym / Fitness Center	Restaurant	Burger Joint	Brazilian Restaurant	Dance Studio	Dessert Shop	Park
94	Barra Funda	Pizza Place	Italian Restaurant	Motel	Pet Store	Dessert Shop	Restaurant	Hotel	Theater	Farmers Market	Café

Figure 7: Cluster 1

Cluster 3 s an outer district and it seems to be residencial areas, where are common to have bakeries.

	Distrito	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Brasilândia	Bakery	Gym / Fitness Center	Grocery Store	Snack Place	Big Box Store	Gym	Market	Pharmacy	Pizza Place	Plaza
9	Jabaquara	Gym / Fitness Center	Bakery	Brazilian Restaurant	Pizza Place	Dessert Shop	Park	Farmers Market	Pharmacy	Japanese Restaurant	Soccer Field
13	Tremembé	Bakery	Gym	Supermarket	Ice Cream Shop	Gym / Fitness Center	Pharmacy	Fast Food Restaurant	Pizza Place	Plaza	Grocery Store
15	Jaraguá	Bakery	Gym / Fitness Center	Brazilian Restaurant	Pizza Place	Fast Food Restaurant	Miscellaneous Shop	Grocery Store	Gym	Clothing Store	Market
17	Pirituba	Bakery	Pizza Place	Pharmacy	Gym / Fitness Center	Gym	Japanese Restaurant	Park	Sandwich Place	Bar	Tea Room
22	Cachoeirinha	Bakery	Pizza Place	Pharmacy	Farmers Market	Supermarket	Grocery Store	Gym	Italian Restaurant	Gym / Fitness Center	Print Shop
25	Freguesia do Ó	Pizza Place	Bakery	Gym / Fitness Center	Dessert Shop	Ice Cream Shop	Gym	Brazilian Restaurant	Burger Joint	Farmers Market	Italian Restaurant
31	Vila Medeiros	Bakery	Gym / Fitness Center	Pizza Place	Brazilian Restaurant	Pet Store	Gym	Pharmacy	Farmers Market	Chocolate Shop	Clothing Store
42	Cursino	Gym / Fitness Center	Bakery	Pizza Place	Brazilian Restaurant	Farmers Market	Cosmetics Shop	Coffee Shop	Bar	Fruit & Vegetable Store	Pharmacy
44	Mandaqui	Bakery	Pizza Place	Restaurant	Gym	Fast Food Restaurant	Bar	Pharmacy	Brazilian Restaurant	Pool	Park
54	Jaçanã	Bakery	Gym / Fitness Center	Brazilian Restaurant	Restaurant	Soccer Field	Pizza Place	Grocery Store	Supermarket	Chocolate Shop	Fast Food Restaurant
63	São Domingos	Bakery	Gym	Pizza Place	Gym / Fitness Center	Pharmacy	Restaurant	Brazilian Restaurant	Dessert Shop	Sushi Restaurant	Sandwich Place
68	Perus	Bakery	Snack Place	Brazilian Restaurant	Plaza	Pizza Place	Convenience Store	Toll Plaza	Gym / Fitness Center	Gymnastics Gym	Chocolate Shop
73	Anhanguera	Bakery	Grocery Store	Convenience Store	Park	Food & Drink Shop	Gym / Fitness Center	Snack Place	Campground	Supermarket	Ice Cream Shop
74	Lapa	Bakery	Pizza Place	Gym / Fitness Center	Bar	Restaurant	Pastelaria	Pharmacy	Gym	Brazilian Restaurant	BBQ Joint
91	Jaguara	Bakery	Gym	Brazilian Restaurant	Pizza Place	Restaurant	Burger Joint	Market	Bar	Sandwich Place	Grocery Store

Figure 8: Cluster 2

Cluster 1 is where we see lot's of gastronomy related venues (Italian Restaurant,Middle Eastern Restaurant and Pizza Place, for example). Also, cluster 2 presents the largest number of Italian restaurants as the most commun options in the districts. Based on that, we can advise the restaurant owner to consider the districts from this cluster as the potential location for the new restaurant.

	Distrito	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
56	Itaim Bibi	Italian Restaurant	Restaurant	Hotel	Ice Cream Shop	Argentinian Restaurant	Cycle Studio	Café	French Restaurant	Japanese Restaurant	Art Museum
60	Jardim Paulista	Ice Cream Shop	Italian Restaurant	Hotel	Brazilian Restaurant	Art Museum	Restaurant	Cosmetics Shop	Bookstore	Pizza Place	Bakery
64	Santa Cecília	Pizza Place	Italian Restaurant	Theater	Restaurant	Middle Eastern Restaurant	Café	Bakery	Women's Store	Brazilian Restaurant	Fruit & Vegetable Store
70	Bela Vista	Ice Cream Shop	Theater	Italian Restaurant	Coffee Shop	Café	Hotel	Movie Theater	Cultural Center	Bakery	Bookstore
78	Consolação	Bakery	Ice Cream Shop	Italian Restaurant	Hotel	Café	Theater	Pizza Place	Coffee Shop	Brazilian Restaurant	Bookstore
81	República	Theater	Italian Restaurant	Bakery	Sandwich Place	Cultural Center	Grocery Store	Middle Eastern Restaurant	Café	Brazilian Restaurant	Coffee Shop
89	Bom Retiro	Middle Eastern Restaurant	Korean Restaurant	Café	Women's Store	Dessert Shop	Restaurant	Brazilian Restaurant	Bar	Art Museum	Italian Restaurant
90	Brás	Middle Eastern Restaurant	Food & Drink Shop	Pizza Place	Restaurant	Italian Restaurant	Brazilian Restaurant	Bar	Market	Sandwich Place	Arts & Crafts Store
92	Sé	Grocery Store	Bakery	Market	Theater	Food & Drink Shop	Italian Restaurant	Cultural Center	Sandwich Place	Brazilian Restaurant	Middle Eastern Restaurant
93	Pari	Middle Eastern Restaurant	Clothing Store	Churrascaria	Food & Drink Shop	Pet Store	Restaurant	Italian Restaurant	Sandwich Place	Shopping Mall	Market

Figure 9: Cluster 1

5. Discussion and conclusion

Based on what we learned about the clusters, we can advise the restaurant owner to consider the districts from cluster 1as a potential location for the new restaurant. These are the districts where gastronomy is well represented and also hotels are frequent. These satisfy the two original criteria that the location should be in a gastronomical centre and in a location that is easily accessible for tourists. The districts “Itaim Bibi” and “Jardim Paulista” would be a good choice since they are close to each other and have Italian restaurants as the most common venue.

This paper discussed the process of coming up with an answer for a hypothetical though real-life like business problem. The analysis was performed based on the toolset of data science and relied heavily on the use of Python and Python libraries such as Pandas, Scikit, Folium to name a few. Data was collected from a different type of sources and in different formats. For analysis, machine learning technique was used. The output of the analysis provided a thorough base for the recommendation for the business problem in question.