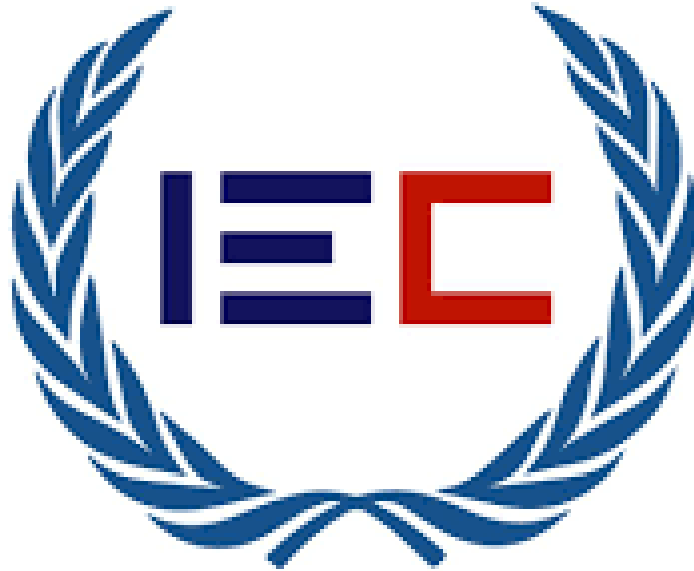# Capstone Project 1

Capstone Project 1

Group (H) Members

Muhammad Wajeeh

Areeba Siddique

Zubair

Mohsin Moltani

Hina Motiani

M.Taimoor Khan

Submitted to

Sir Faizan

Sir Misbbah

Capstone Project 1

Project Overview

In our **first Capstone Project** we selected a dataset and then export it to the **MySql Workbunch**. After exporting we select some business problems related to our dataset and then solve these queries and get some results. For the datatypes and data description we made a **Data Dictionery**. For better understanding of stackholders we designed a **ER-Diagram** and performed **Visualization** of our dataset.

Now we are going to explain each step one by one thoroughly.

1-MySql Workbunch

For the installation of MySql Workbunch open the website

www.mysql.com and go to downlaods option and scroll down and select **MySql Community(GPL) downloads**
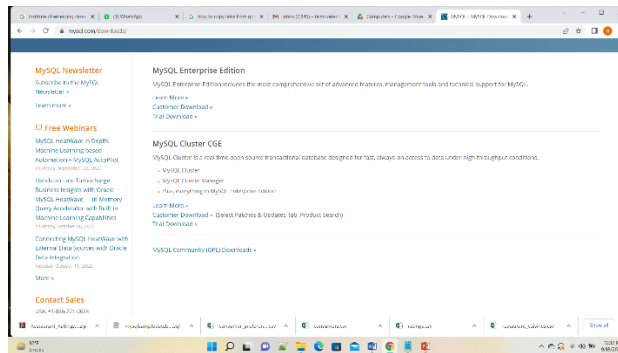

Figure 1: MySql Workbunch

After clicking on the Mysql Community (GPL) downloads , a screen pop up and then select **Mysql Installer for Windows**. Download **5.5Mb file.**
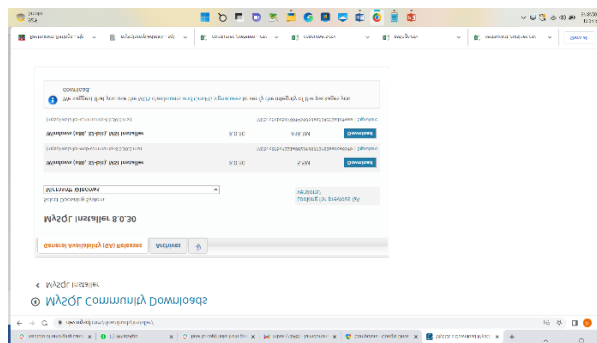

Figure 2: MySql Installer

Capstone Project 1

After downloading the file open it and select " **Custom**" showing at the last.

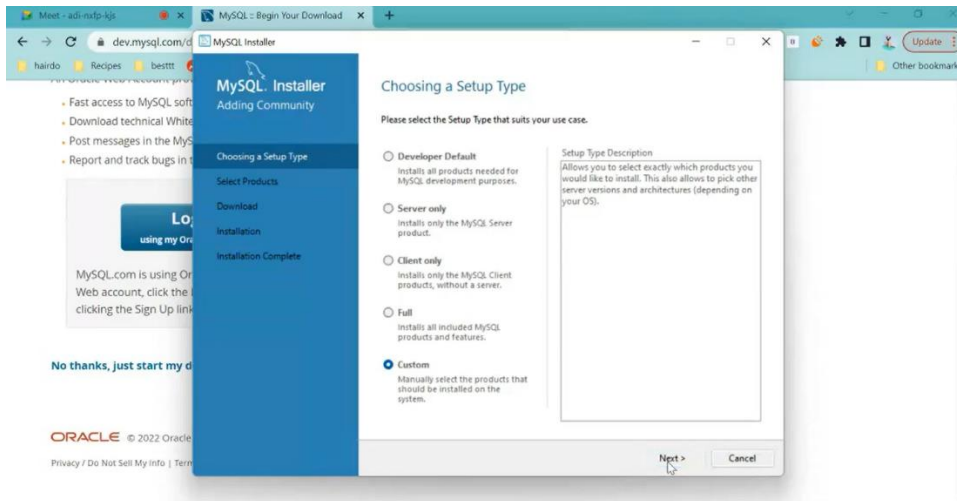*Figure 3: Select Custom*

After selecting custom press next and install there pacakges.

1- MySql Server 8.0.30
2- MySql Workbunch 8.0.30
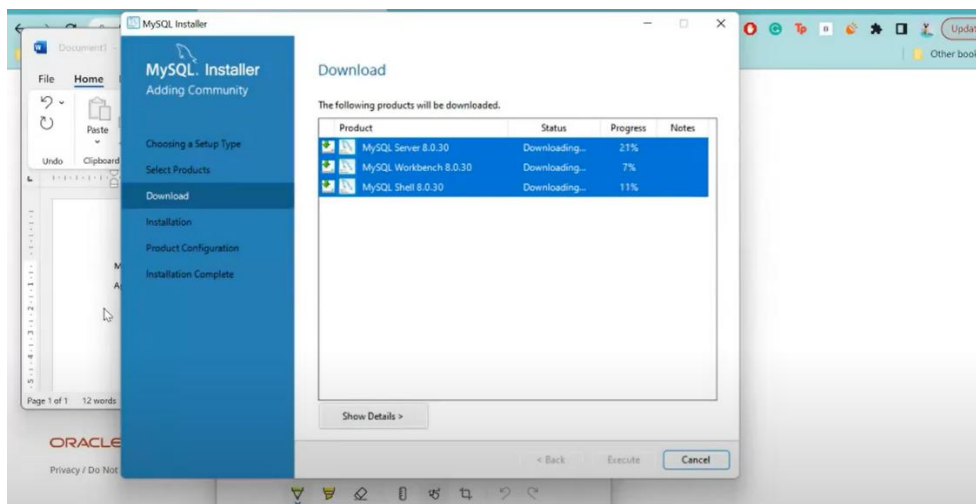3- MySql Shell 8.0.30

Figure 4: Installation of Pacakages

After finishing the downloading , execute it for the installation. Once it done select a **password** and then click on finish. Your MySql Workbunch is ready to work.

Warning : You should select a **Password** that is easy to remember.

Capstone Project 1

<u>About Dataset</u>

We selected a **"Restaurant Rating"** dataset. This data collected in **Mexico** by real users in 2012. There are many restaurants and data contain additional information about each restaurant. Data includes cuisines , customers and their preferences. Consumers are highly focused because without them business can't run.

<u>Why We Preferred that Dataset</u>

We selected **"Restaurant Rating"** dataset to analyze the gap between consumers demand and restaurants products they offered. In todays world mostly businesses run on customer demand and these businesses preferred their customers first priority. So, we select that dataset to get insights to business problems and also make some suggestions for restaurant owners

<u>Data in Dataset</u>

Our dataset has five files.

1- Consumer Rating 2-Consumer Preferences 3-Consumer 4-Restaurants

5-Restaurants Cuisines

Capstone Project 1

Data Dictionary

A **Data Dictionary** is a collection of names, definitions and attributes about data elements that are being used or captured in a database , information system or a part of a research project.

Why We Use a Data Dictionary

Data Dictionary is useful for a number of reasons. Some important reasons are mentioned below.

1- Make Data easier to analyze
2- Assist in avoiding data inconsistencies across a project
3- Help define conventions that are to be used in a project
4- Enforce the use of data standards
5- Provide consistencies in the collection and use of data across multiple members of a research team

Our Dataset Dictionary

Our dataset consist of 5 tables. Each table has several fields of data. We briefly describe each field and make it easier for our stakeholders for understanding the data. Below is our Data Dictionary

Table	Field	Description	Data type	Constraint	Required	Contains NULL	Acceptable Values	Business Rule
Consumers	Consumer_ID	Unique identifier for each consumer	NVARCHAR(5)	PRIMARY KEY	YES	NO	(U1—)-(U1—)	Consumer_ID must be alphanumeric
	City	City where the consumer lives	VARCHAR(15)	NOT UNIQUE	NO	NO	located in Mexico	this field must not be numeric
	State	State where the consumer lives	VARCHAR(15)	NOT UNIQUE	YES	NO	located in Mexico	this field must not be numeric
	Country	Country where the consumer lives	VARCHAR(6)	NOT UNIQUE	YES	NO	Only Mexico	this field must not be numeric
	Latitude	Latitude where the consumer lives	DECIMAL(8,6)	NOT UNIQUE	NO	NO	N/A	this field contain decimal with negative or positive sign
	Longitude	Longitude where the consumer lives	DECIMAL(9,6)	NOT UNIQUE	NO	NO	N/A	this field contain decimal with negative or positive sign
	Smoker	whether consumer smokes or not	VARCHAR(3)	NOT UNIQUE	YES	YES	YES or NO	this field must be alphabetic
	Drink_Level	Consumer's Drinking Status	VARCHAR(14)	NOT UNIQUE	YES	NO	Abstemious Or Casual Drinker Or Social Drinker	this field must be alphabetic
	Transportation_Method	Way through consumer transports	VARCHAR(7)	NOT UNIQUE	YES	YES	by Foot Or Public Transport Or Car	this field must be alphabetic
	Marital_Status	Consumer's Marital Status	VARCHAR(7)	NOT UNIQUE	YES	YES	Single Or Married	this field must be alphabetic
	Children	Children's dependency status	VARCHAR(11)	NOT UNIQUE	NO	YES	Dependent Or independent Or Kids	this field must be alphabetic
	Age	Consumer's Age	INT(2)	NOT UNIQUE	YES	NO	1 Digit Or 2 Digit	this field must be numeric
	Occupation	Consumer's Occupation	VARCHAR(10)	NOT UNIQUE	YES	YES	Student Or Employed Or Unemployed	this field must be alphabetic
	Budget	Consumer's Budget	VARCHAR(6)	NOT UNIQUE	YES	YES	Low Or Medium Or High	this field must be alphabetic
Consumer_Preferences	Consumer_ID	Unique identifier for each consumer	NVARCHAR(5)	FOREIGN KEY	YES	NO	(U1—)-(U1—)	Consumer_ID must be alphanumeric
	Preferred_Cuisine	Types of food consumer prefers	VARCHAR(16)	NOT UNIQUE	YES	NO	N/A	this field must be alphabetic
Ratings	Consumer_ID	Unique identifier for each consumer	NVARCHAR(5)	FOREIGN KEY	YES	NO	(U1—)-(U1—)	Consumer_ID must be alphanumeric
	Restaurant_ID	Unique identifier for each restaurant	INT(6)	FOREIGN KEY	YES	NO	6-digit Number	this field muust be numeric
	Overall_Rating	Overall rating by the consumer for the restaurant	INT(1)	NOT UNIQUE	YES	NO	0=Unsatisfactory, 1= Satisfactory, 2=Highly Satisfactory	this field muust be numeric
	Food_Rating	Food's rating by the consumer for the restaurant	INT(1)	NOT UNIQUE	YES	NO	0=Unsatisfactory, 1= Satisfactory, 2=Highly Satisfactory	this field muust be numeric
	Service_Rating	Service rating by the consumer for the restaurant	INT(1)	NOT UNIQUE	YES	NO	0=Unsatisfactory, 1= Satisfactory, 2=Highly Satisfactory	this field muust be numeric
Restaurants	Restaurant_ID	Unique identifier for each restaurant	INT(6)	PRIMARY KEY	YES	NO	6-digit Number	this field must be numeric
	Name	Restaurant's Name	VARCHAR(52)	NOT UNIQUE	YES	NO	N/A	this field must be alphabetic
	City	Restaurant's City	VARCHAR(15)	NOT UNIQUE	NO	NO	located in Mexico	this field must be alphabetic
	State	Restaurant's State	VARCHAR(15)	NOT UNIQUE	YES	NO	located in Mexico	this field must be alphabetic
	Country	Restaurant's Country	VARCHAR(6)	NOT UNIQUE	YES	NO	Mexico	this field must be alphabetic
	Zip_Code	Restaurant's Zip Code	INT(5)	NOT UNIQUE	NO	YES	5-digit Number	this field must be alphabetic
	Latitude	Restaurant's Latitude	DECIMAL(9,7)	UNIQUE	NO	NO	N/A	this field contain decimal with negative or positive sign
	Longitude	Restaurant's longitude	DECIMAL(10,7)	UNIQUE	NO	NO	N/A	this field contain decimal with negative or positive sign
	Alcohol_Service	whether restaurant serves alcohol or not(if yes, then define way)	VARCHAR(15)	NOT UNIQUE	YES	NO	No Alcohol Or Wine and Beer Or Full Bar	this field must be alphabetic
	Smoking_Allowed	whether smoking is allowed or not(if yes, then define way)	VARCHAR(15)	NOT UNIQUE	YES	NO	Yes Or No Or Smoking Section	this field must be alphabetic
	Price	Restaurant's Price	VARCHAR(6)	NOT UNIQUE	YES	NO	Low Or Medium Or High	this field must be alphabetic
	Franchise	Whether restaurant is franchise or not	VARCHAR(3)	NOT UNIQUE	YES	NO	Yes Or No	this field must be alphabetic
	Area	Restaurant's Occupation Status	VARCHAR(6)	NOT UNIQUE	NO	NO	Open Or Closed	this field must be alphabetic
	Parking	Whether restaurant offers any sort of parking	VARCHAR(6)	NOT UNIQUE	NO	NO	None Or Yes Or Public Or Valet	this field must be alphabetic
Restaurant_Cuisines	Restaurant_ID	Unique identifier for each restaurant	INT(6)	FOREIGN KEY	YES	NO	6-digit Number	this field muust be numeric
	Cuisine	Types of food restaurant serves	VARCHAR(13)	NOT UNIQUE	YES	NO	N/A	this field must be alphabetic

Figure 5: Data Dictionary

Capstone Project 1

ER-Diagram :

ER Diagram stands for Entity Relationship Diagram, also known as ERD is a diagram that displays the relationship of entity sets stored in a database. In other words, ER diagrams help to explain the logical structure of databases. ER diagrams are created based on three basic concepts: entities, attributes and relationships.

Restaurant Rating ER-D :

There are five entities in our ER-Diagram. PK and boldface denote the attribute(s) that constitute the entity type's **unique identifier**. All entities are connected in the form of One-to-Many binary relationship.
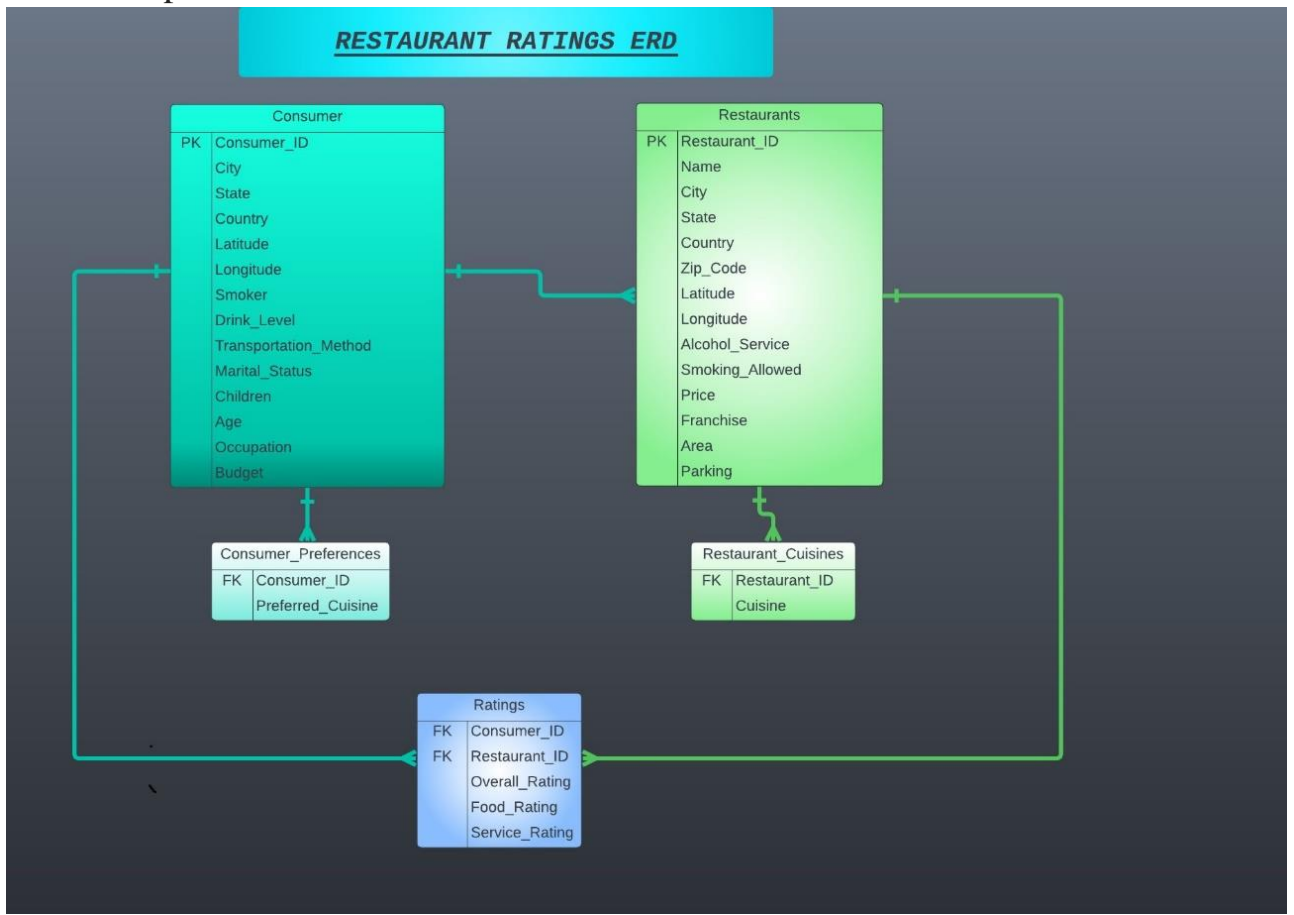
Figure 6: ER-Diagram

Capstone Project 1

Queries And Visualization

Queries are the business problems arises from the dataset. From our dataset we find 12 queries which we will discuss one by one.

Visualization is the technique of presenting our results in a graphical manner for better understanding.

Query 1: How many times restaurants get lowest and highest rating ??

Solution: First we find out the total ratings which are 1161 then we choose overall ratings from rating table. We found out that there are 486 high and 254 low overall ratings for restaurants.
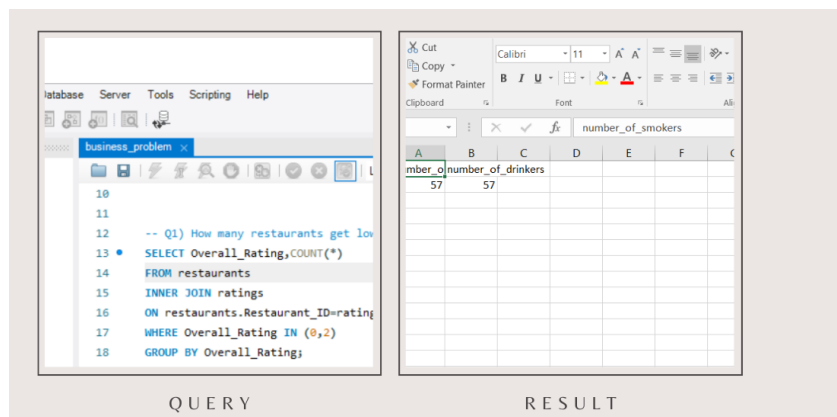
Figure 7: Query 1 And Result

Visualization:
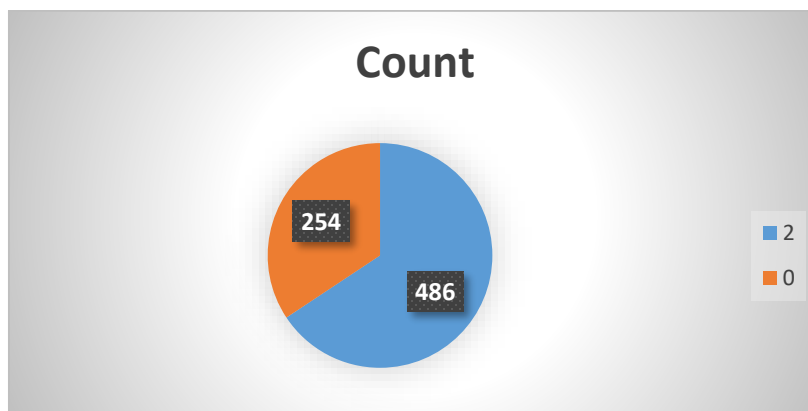
Figure 8: Query 1 Visualization

Capstone Project 1

Query 2: How many times restaurants get lower and higher ratings and by which consumers?

Solution: This query is same as earlier, only difference is that here we also mentioned the customers which give rating. There are 136 customers who give overall ratings and ratings are given by same user multiple times.
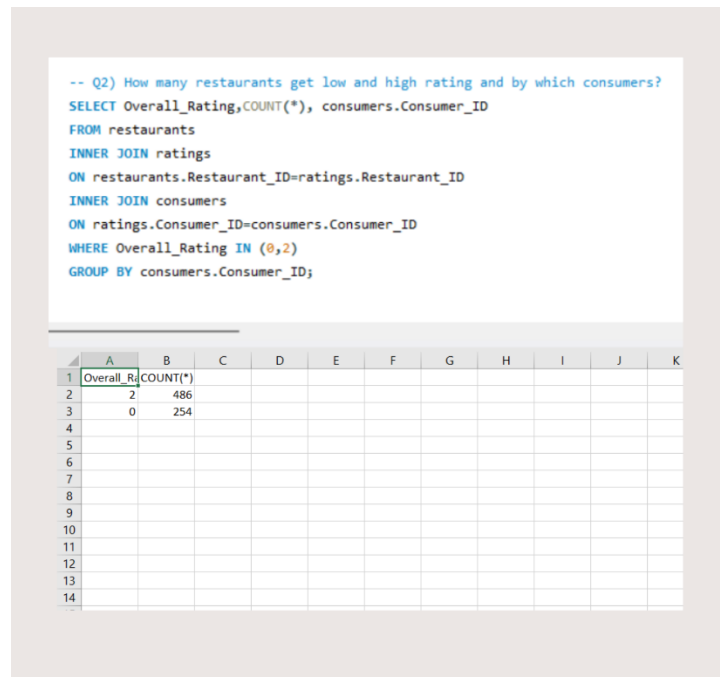
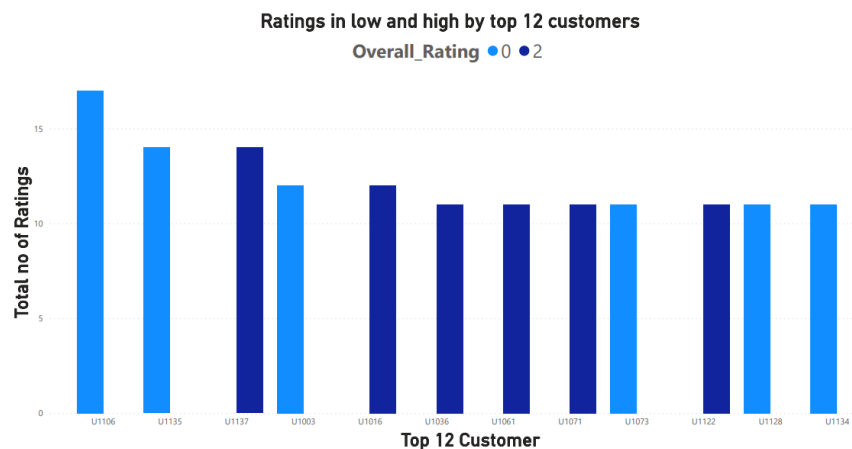*Figure 9: Query 2 and Result*

Visualization

Figure 10: Query 2 Visualization

Capstone Project 1

Query 3: Select all restaurants where public parking is available and their budget is low?

Solution: There are 7 restaurants out of 130 give public parking although their budget is low.

```
-- Q3) Select all restaurants where public parking is
SELECT Name,Parking,Price
FROM restaurants
WHERE Parking="Public" AND Price ="Low";
```

	A	B	C	D	E	F	G	H
1	Overall_Ra	COUNT(*)	Consumer_ID					
2	2	3	U1001					
3	2	4	U1002					
4	2	12	U1003					
5	2	7	U1004					
6	2	5	U1005					
7	2	4	U1006					
8	0	1	U1007					
9	2	1	U1008					
10	0	6	U1009					
11	2	3	U1010					
12	2	1	U1011					
13	2	4	U1012					

Figure 11: Query 3 and Result

Query 4: What can you learn from the highest rated restaurants? Do consumer preferences have an effect on ratings?

Solution: Top 3 high rated restaurant from all are Tortas Locas Hipocampo, Puesto De Tacos, Cantina Restaurants. Through these findings we analyzed that where smoking is not allowed, have no franchise and area is closed people give high ratings.

```
36      -- Q4) What can you learn from the highest rated restaurants? Do consumer preferences hav
37 •    SELECT Name,COUNT(Overall_Rating)
38      FROM restaurants;
39 •    SELECT * FROM ratings;
40 •    SELECT * FROM consumer_preferences;
41 •    SELECT MAX(Overall_Rating) FROM ratings;
42      -- ANs
43 •    SELECT * FROM restaurants;
44 •    SELECT Name,COUNT(Overall_Rating),COUNT(Food_Rating),COUNT(Service_Rating)
45      FROM restaurants
46      INNER JOIN ratings
```

Figure 12: Query 4

Capstone Project 1

Query 5: Are there any demand & supply gaps that you can exploit in the market?

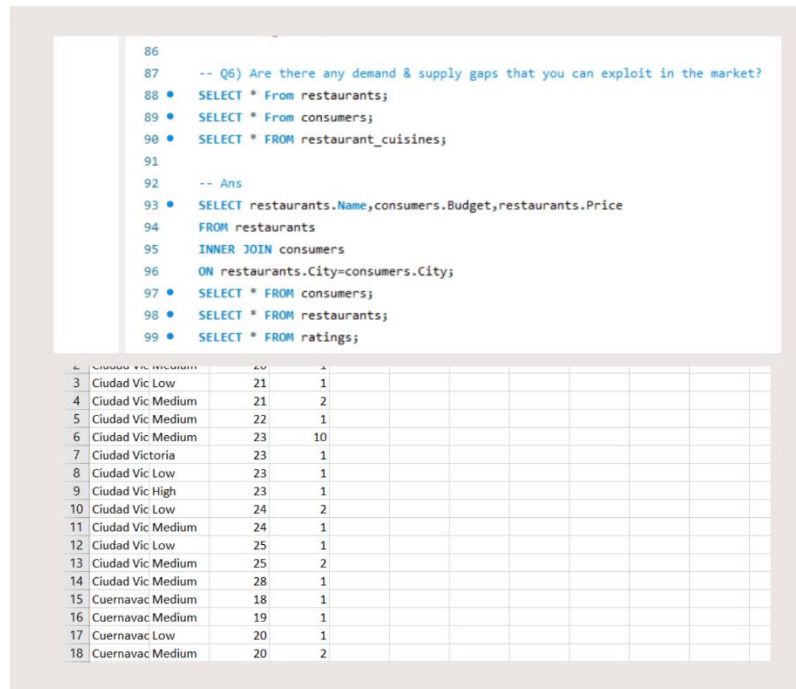**Solution:** Yes there are some gaps between demand and supply that we can exploit in the market.

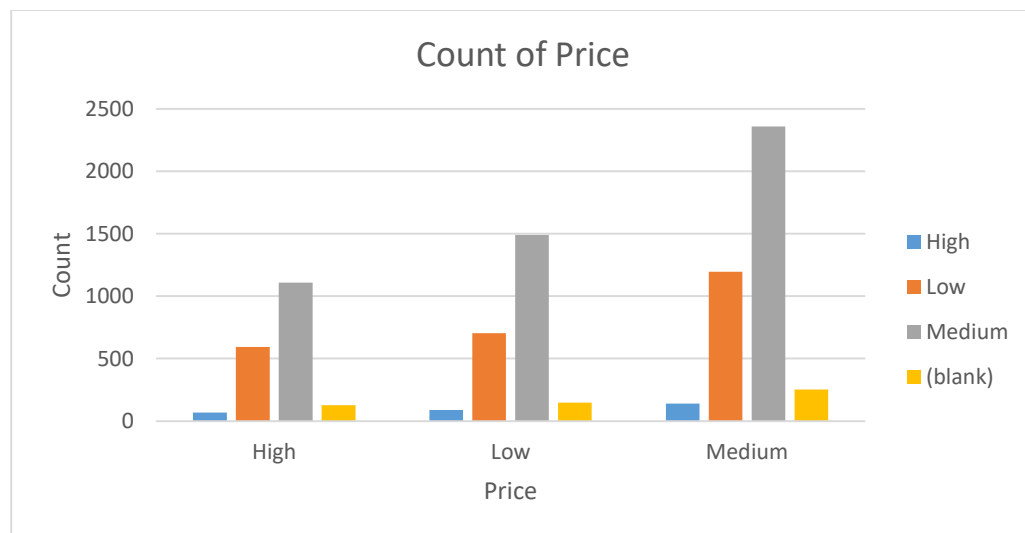*Figure 13: Query 5 and Result*

Visualization:

Figure 14: Visualization Query 5

Capstone Project 1

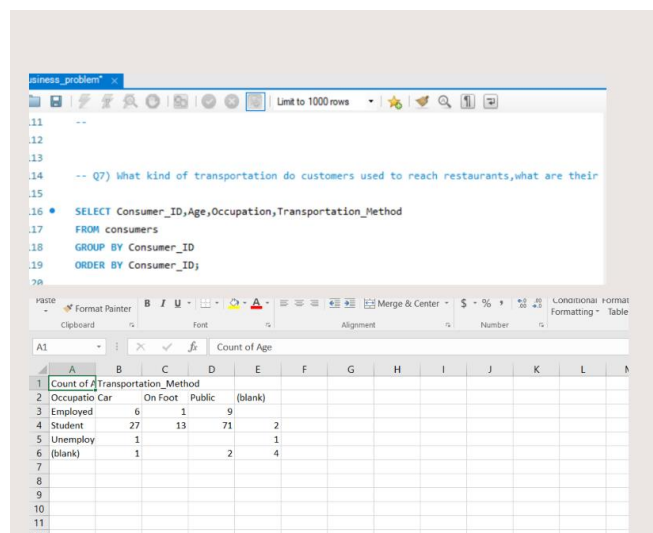**Query 6**: If you were invest into restaurants which characteristics would you be looking for?

Solution: By our findings we first check through ratings

Ratings	Alcohol	Smoking	Price	Franchise	Area	Parking
High	Full Bar, Wine & Beer	Allowed	High, Medium	Mostly No	Mostly Closed	Mostly Yes
Medium	Full Bar, Wine & Beer	Smoking Section	High, Medium	No	Closed	Mostly Yes
Low	Full Bar, Wine & Beer	Smoking Section	High, Medium	Either Yes or No	Either Open or Closed	Mostly Yes

Here we can clearly analyzed that even restaurants having low or high ratings but these services are common.

Query 7: What kind of transportation do customers used to reach restaurants, what are their ages and occupation?

Solution: We found out most of customers using public transport around 82 customer out of 138. Customers are in wide range of ages from 18 to 82, out of which most of customers are 23 years old. Most customers are students, 113 out of 138.

Figure 15: Query 7 and Result

Capstone Project 1
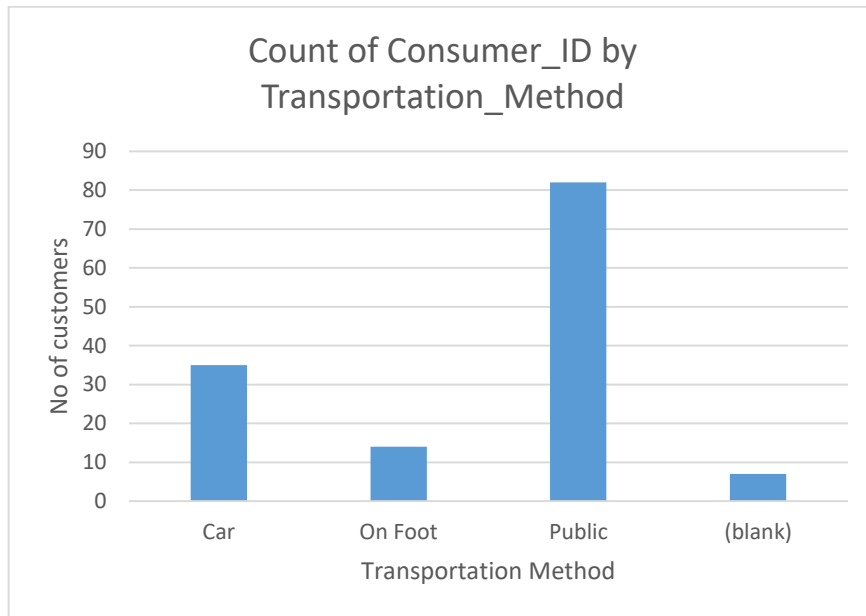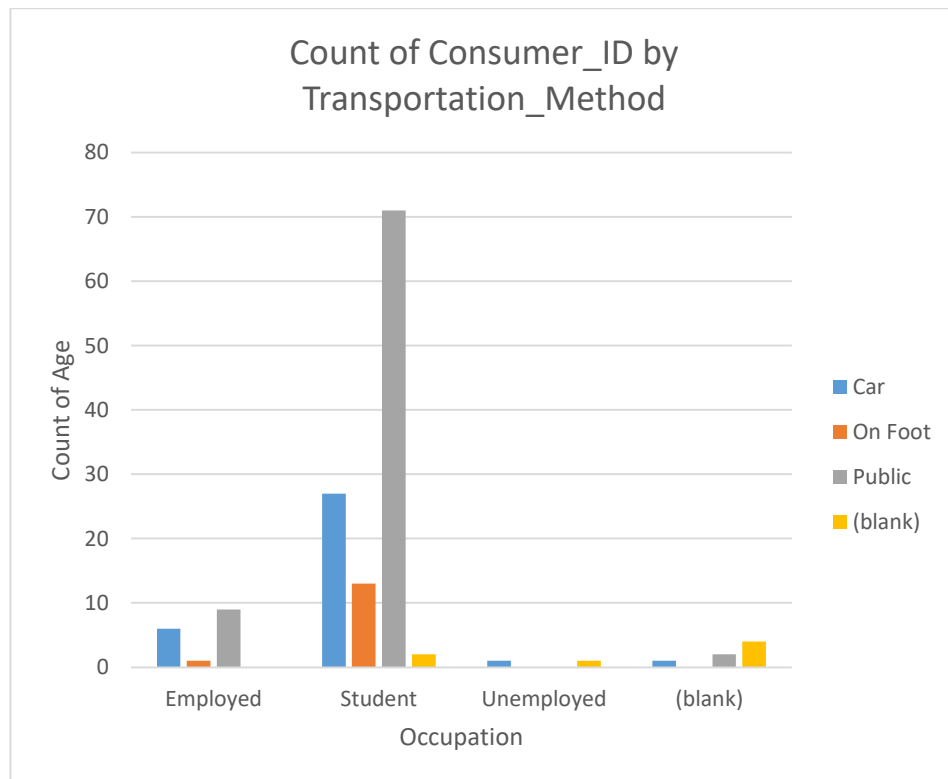
Visualization

Figure 16: Q7 Part a

Figure 17: Query 7 part B

Capstone Project 1

Query 8: How many consumers prefer both drinking and smoking in restaurants? Among them which kind of restaurants cuisine do they prefer?

Solution: There are 26 out 138 customers who prefer both smoking and drinking, also among 26 most prefer Mexican food.
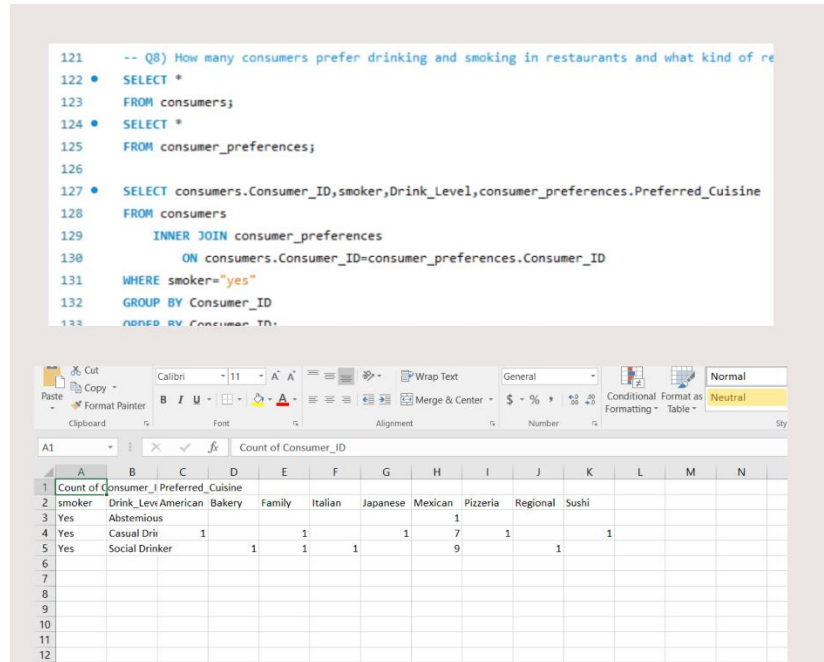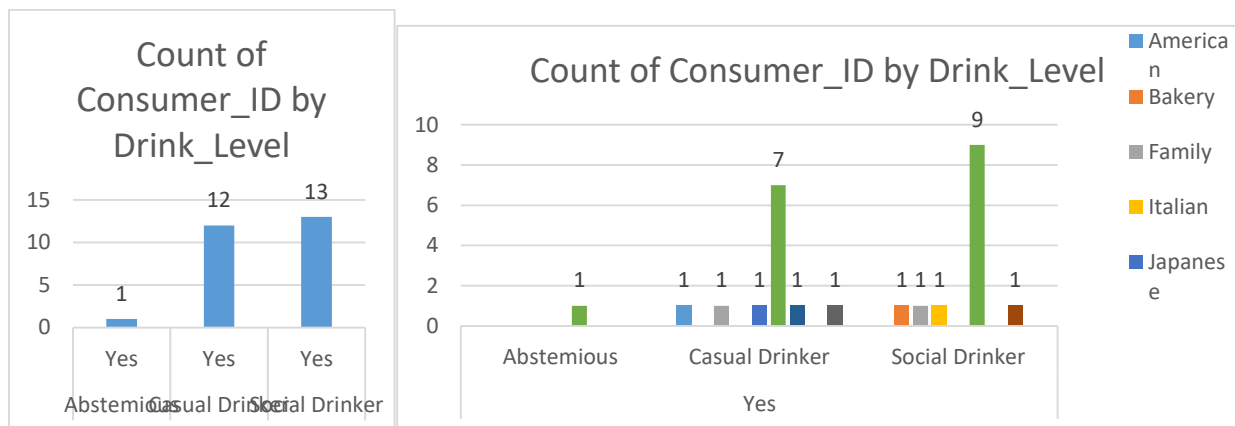

Figure 18: Query 8 and Result

Visualization

Figure 19: Q8 Visualization

Capstone Project 1

Query 9: What kind of restaurants allows smoking and drinking service in their restaurants?

Solution: There are 15 restaurants out of 130 who allows both smoking and drinking and most of restaurants are Bars
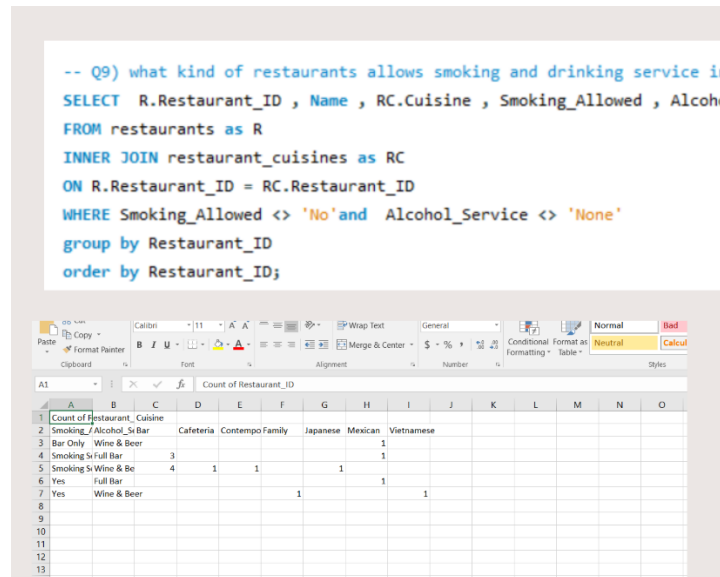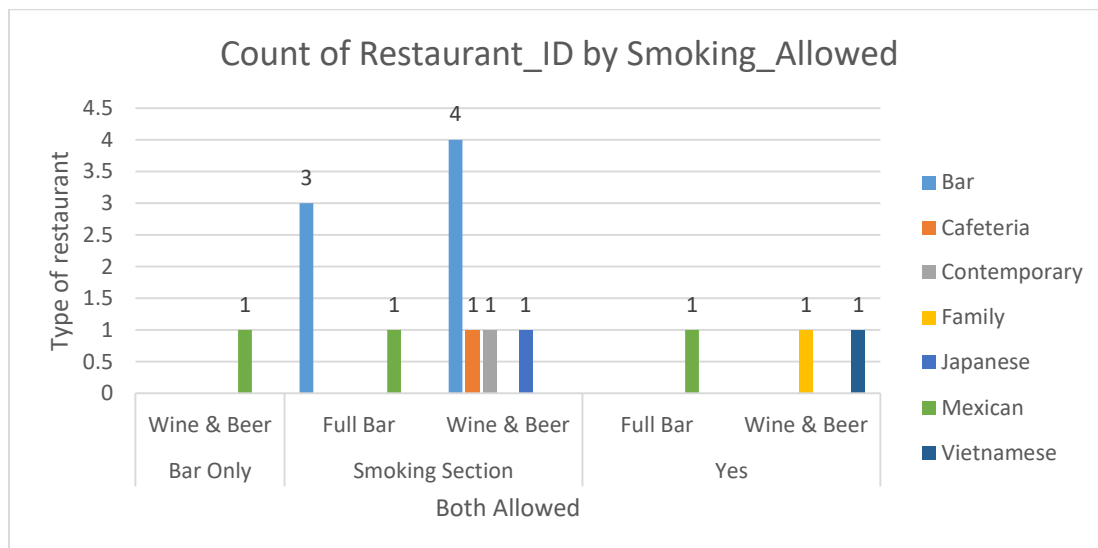

Figure 20: Q9 Query and Result

Visualization

Figure 21: Query 9 Visualization

Capstone Project 1

Query 10: Which city has maximum no of restaurants and which of them are highly rated?

Solution: Most of our restaurants are located in San Luis Potosi state 84 out of 130, and out of 84 restaurants 46 restaurants that is 2 > 1 > 0.
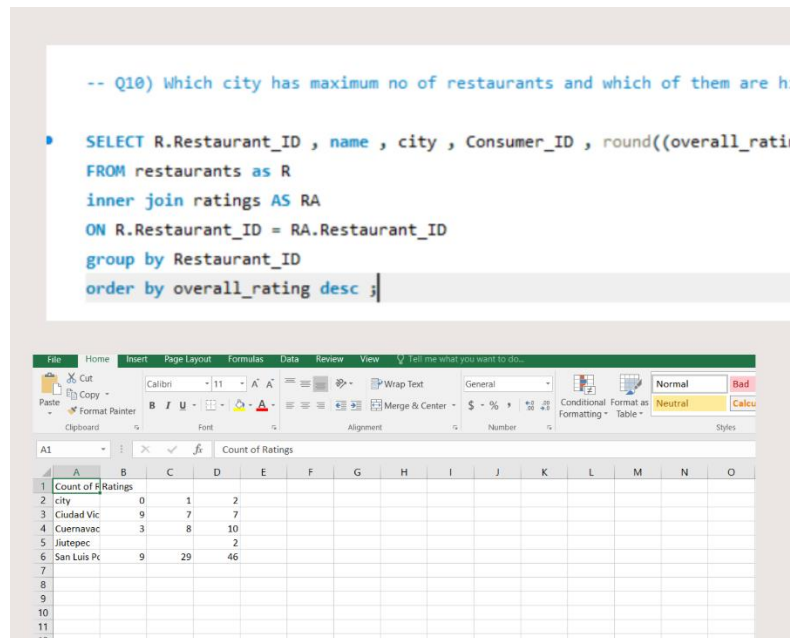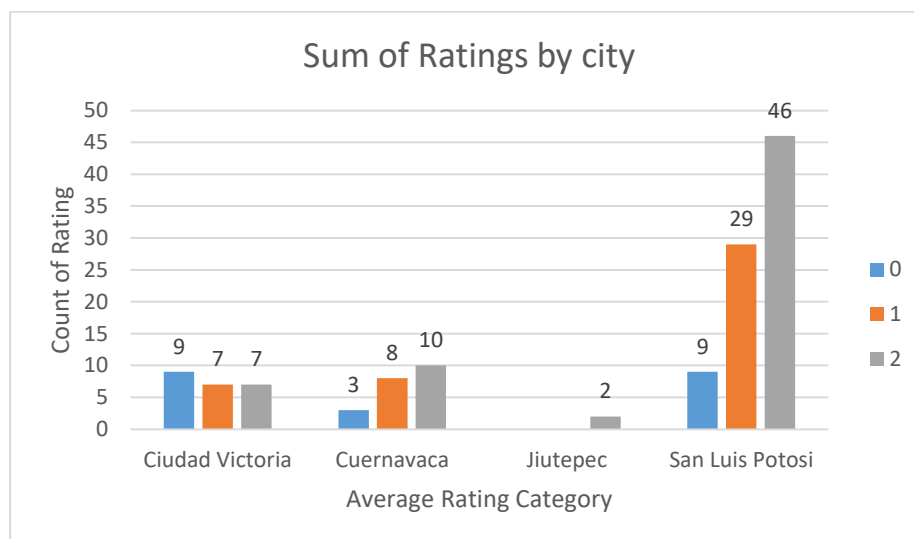

Figure 22: Query 10 and Result

Visualization

Figure 23: Query 10 Visualization

Capstone Project 1

Query 11: What are the consumer demographics? Does this indicate a bias in the data sample?

Solution: Consumer Demographics include Age, City and Budget of Consumer. Through graph, the bias we find out is that San luis Potosi is the only city where maximum 21 age group people has medium budget and 23 age group has relatively dominant low budget whereas in Ciudad Victoria, lesser 23 age group people earns medium budget than Potosi. People having high budget is no where dominant in any city. This indicates biasness in Consumer Demographics.

```
67      -- Q11) What are the consumer demographics? Does this indicate a bias in the data sample?
68 •    SELECT * FROM consumers;
69 •    SELECT * FROM consumer_preferences;
70
71 •    SELECT City,Age,Marital_Status,Budget,Occupation
72      FROM consumers;
73 •    SELECT DISTINCT City FROM consumers;
74 •    SELECT DISTINCT Age FROM consumers;
75 •    SELECT DISTINCT Marital_Status FROM consumers;
76
77 •    SELECT  City,Budget,Age,COUNT(Age)
78      FROM consumers
79      GROUP BY City,Budget,Age
80      ORDER BY City ,Age;
81 •    SELECT  City,Age From consumers WHERE City="San Luis Potosi"  AND Age = 20;
82 •    SELECT DISTINCT Age
83      FROM consumers
84      WHERE City="San Luis Potosi"
85      ORDER BY Age DESC;
86
```
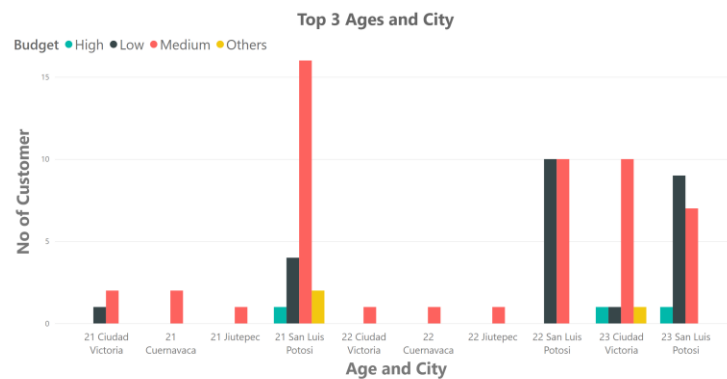
Figure 24: Query 11 and Results

Visualization:

Figure 25: Query 11 Visualization

Capstone Project 1

Query 12: How much do customers prefer spending in restaurants and what kind of cuisine?

Solution: In our dataset, there are three budgets as well as three cuisines. Mostly consumers preferred Mexican rather than American and Pizzeria. Mostly consumers have a medium budget for their expenses.
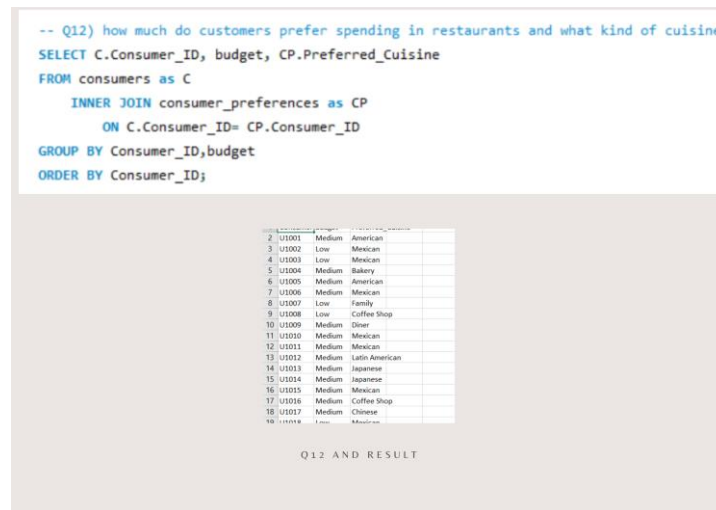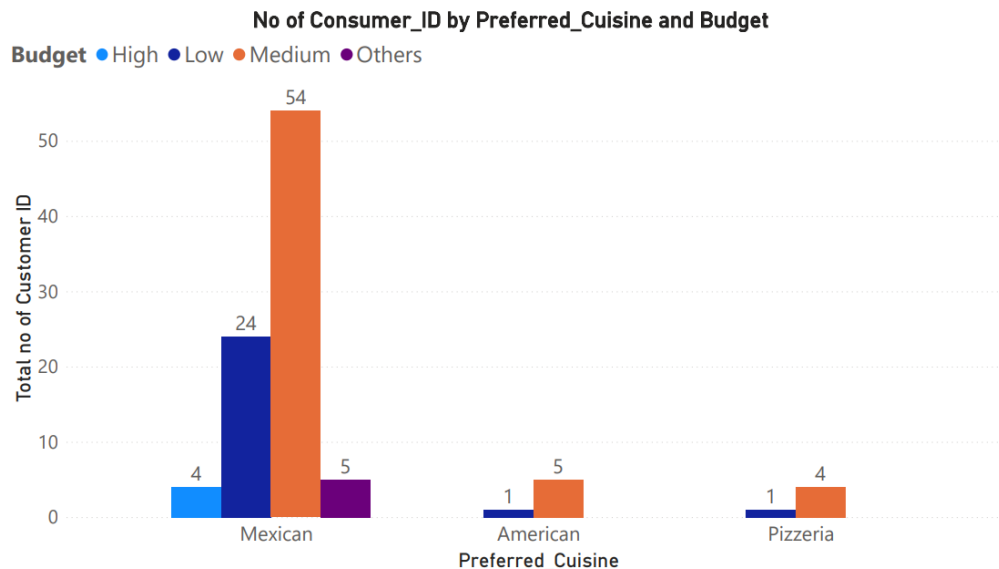

Figure 26: Q12 Query and Solution

Visualization

Figure 27: Query 12 Visualization

Capstone Project 1

Query 13: Which Restaurants allowed smoking and allowed parking; their Consumer ratings vs Which Restaurants doesn't allow smoking and doesn't allow parking; their Consumer ratings?

Solution: There are some restaurants where all 3 parking and smoking service is available and the restaurants that are allowing 'Smoking Section' and their Parking is medium, rated (1) and by narrow margin dominating

There are some restaurants who don't allow smoking and parking, in which medium rated restaurants are on top
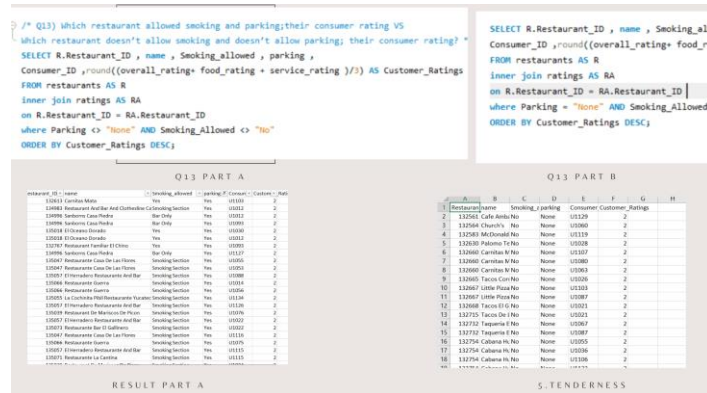

Figure 28: Q13 query and solution

Visualization

Figure 29: Q 13 Visualization

Capstone Project 1

Recommendations:

　　　　　　　　We have added recommendation in terms of investing in Q7 (We can have look again).

- Those restaurants which are being rated low, they need to work on the cuisines which are now a day's preferences of consumers. If they are allowing smoking and alcohol services, then they must work on their service rating than Food Rating.

- According to trends, those who are renders above services they allot valet parking which standardize restaurant.

- If the restaurants are not allowing any of these services then they must focus on food quality rather than service quality, they must have open area for consumers to interact with environment and enjoy their meal.

- But preferably, it is to mention that those restaurants which allow smoking and alcohol services are not on favorable side because they are missing the trick of attracting variety of consumers to their restaurant because due to allowing those services, you are limiting your consumers.

- ✓ Hope this helps restaurant owners of Mexico to maximize their profit.

Conclusion

　　　　　　　From the selected dataset we find many business problems that try to solve all these queries logically in MySql Workbunch. Through this project we did practice of MySql and we came to know about more ways to solve out a business problem. This Project helped us in learning of MySql Bunch , team work , time management , Data Visualization , ER-Diagram and Data Dictinories. And these all are the key methods in data analytics