

# PROJECT REPORT

## CONTENTS:

- PROBLEM STATEMENT
- RESTAURANTS DATASET DATA
  - DATA DESCRIPTION
  - OBJECTIVES
  - ANALYSIS
    - QUERY
    - OUTPUT
    - DESCRIPTION
    - INFERENCE
  - CONCLUSION

# PROBLEM STATEMENT

To analyze and seek insights for different scenarios from the dataset of the Restaurant Industry. Analyses is done on different aspects including state-wise analysis.

## Data Description:

Dataset of different restaurants was used for drawing different insights about restaurants in various aspects such as ratings, parking availabilities, cuisines, etc. It includes following tables:

- Chefmozaccepts-Location Wise availability of Payment Modes
- Chefmozcuisine -Location Wise availability of Cuisine
- Chefmozhours4 - Working Hours of Restaurant
- Chefmozparking - Parking availability at restaurants at different places
- Geoplaces2 - Location Wise Summary of dress code, country, state, etc.
- Rating\_final - User wise rating to the restaurants in diff locations
- Usercuisine - User had which Cuisine.
- User payment - User used which payment mode.
- Userprofile - Users personal details like a smoker, drink level, interest, religion, etc.

### List of Tables:

#### 1) geoplaces2:

| Column name:   | Data Type | Describe  |
|----------------|-----------|---|
| Place id       | Integer   | ID for place of the restaurants                                       |
| Latitude       | real      | Latitude measurement values of the place                              |
| Longitude      | real      | Longitude measurement values of the place                             |
| the_geom_meter | text      | The geo meter letters   |
| Name           | Varchar   | Restaurants name  |
| Address        | Text      | Address of place (restaurants)  |
| City           | Varchar   | Name of the city  |
| State          | Varchar   | Name of the State   |
| Country        | Varchar   | Name of the Country   |
| Fax            | varchar   | No data inside it.  |
| Zip            | integer   | Zip code number   |
| Alcohol        | Varchar   | Whether alcohol is served or not and what type of alcohol served      |
| Smoking_area   | varchar   | Whether smoking allowed or not and where they can smoke.              |
| Dress_code     | Varchar   | Type of dress   |
| Accessibility  | Varchar   | Tells the accessibility whether have complete access or partially etc |
| Price          | Varchar   | Tells price medium, Low and High.                                     |

|                |         |   |
|----------------|---------|---|
| URL            | Varchar | USR of the place restaurant               |
| Rambience      | Varchar | What type of ambiance                     |
| franchise      | varchar | Type of franchise                         |
| Area           | Varchar | Type of area whether it is closed or open |
| Other_Services | varchar | Other services provided by the restaurant |

| Field          | Type   | Null | Key | Default | Extra |
|----------------|--------|------|-----|---------|-------|
| placeID        | int    | YES  |     | NULL    |       |
| latitude       | double | YES  |     | NULL    |       |
| longitude      | double | YES  |     | NULL    |       |
| the_geom_meter | text   | YES  |     | NULL    |       |
| name           | text   | YES  |     | NULL    |       |
| address        | text   | YES  |     | NULL    |       |
| city           | text   | YES  |     | NULL    |       |
| state          | text   | YES  |     | NULL    |       |
| country        | text   | YES  |     | NULL    |       |
| fax            | text   | YES  |     | NULL    |       |
| zip            | text   | YES  |     | NULL    |       |
| alcohol        | text   | YES  |     | NULL    |       |
| smoking_area   | text   | YES  |     | NULL    |       |
| dress_code     | text   | YES  |     | NULL    |       |
| accessibility  | text   | YES  |     | NULL    |       |
| price          | text   | YES  |     | NULL    |       |
| url            | text   | YES  |     | NULL    |       |
| Rambience      | text   | YES  |     | NULL    |       |
| franchise      | text   | YES  |     | NULL    |       |
| area           | text   | YES  |     | NULL    |       |
| other_services | text   | YES  |     | NULL    |       |

This table consists of 21 columns. Out of these, the columns latitude and longitude consist of float values. Placeid contains the values with integer datatype. The rest of the columns are of text datatype. All columns can contain null values.

#### Chefmozaccepts:

| Column Name | Data type | Description            |
|-------------|-----------|------------------------|
| Place ID    | Integer   | Place id for the place |
| payment     | Varchar   | Type of the payments   |

| Field    | Type | Null | Key | Default | Extra |
|----------|------|------|-----|---------|-------|
| placeID  | int  | YES  |     | NULL    |       |
| Rpayment | text | YES  |     | NULL    |       |

This table consists of 2 columns. First column has integer datatype and the other column has text datatype.

#### Chefmozcuisine:

| Column Name | Data type | Description            |
|-------------|-----------|------------------------|
| Place ID    | Integer   | Place id for the place |
| Rcuisine    | Varchar   | Type of the cuisine    |

| Field    | Type | Null | Key | Default | Extra |
|----------|------|------|-----|---------|-------|
| placeID  | int  | YES  |     | NULL    |       |
| Rcuisine | text | YES  |     | NULL    |       |

#### Userprofile:

| Column name:     | Data Type | description   |
|------------------|-----------|---|
| User ID          | Integer   | ID for User   |
| Latitude         | real      | Latitude measurement values of the place of the user  |
| Longitude        | real      | Longitude measurement values of the place of the user |
| Smoker           | Varchar   | Boolean type True or False                            |
| Drive Level      | Varchar   | Type of the drinking condition                        |
| dress_preference | Varchar   | Type of dress preference                              |
| ambience         | Varchar   | Type of ambience                                      |
| transport        | Varchar   | Type of transport the user is using                   |
| marital_status   | Varchar   | Type of Martial Status                                |
| hijos            | varchar   | Type of Hijos like independent or kids                |
| birth_year       | integer   | Year of the birth                                     |
| interest         | Varchar   | User is interested                                    |
| personality      | varchar   | Type of personality of the user                       |
| religion         | Varchar   | Type of user religion                                 |
| activity         | Varchar   | User activity what he does                            |
| color            | Varchar   | User interest color                                   |
| Weight           | Integer   | Weight of the user                                    |
| Budget           | Varchar   | Type Budget medium, Low and High.                     |
| Height           | Integer   | Height of the user in meters                          |

| Field            | Type   | Null | Key | Default | Extra |
|------------------|--------|------|-----|---------|-------|
| userID           | text   | YES  |     | NULL    |       |
| latitude         | double | YES  |     | NULL    |       |
| longitude        | double | YES  |     | NULL    |       |
| smoker           | text   | YES  |     | NULL    |       |
| drink_level      | text   | YES  |     | NULL    |       |
| dress_preference | text   | YES  |     | NULL    |       |
| ambience         | text   | YES  |     | NULL    |       |
| transport        | text   | YES  |     | NULL    |       |
| marital_status   | text   | YES  |     | NULL    |       |
| hijos            | text   | YES  |     | NULL    |       |
| birth_year       | int    | YES  |     | NULL    |       |
| interest         | text   | YES  |     | NULL    |       |
| personality      | text   | YES  |     | NULL    |       |
| religion         | text   | YES  |     | NULL    |       |
| activity         | text   | YES  |     | NULL    |       |
| color            | text   | YES  |     | NULL    |       |
| weight           | int    | YES  |     | NULL    |       |
| budget           | text   | YES  |     | NULL    |       |
| height           | double | YES  |     | NULL    |       |

#### chefmozhours4:

| Column Name | Data type | Description            |
|-------------|-----------|------------------------|
| Place ID    | Integer   | Place id for the place |

|       |      |                              |
|-------|------|------------------------------|
| Hours | Time | Time of the working in a day |
| Days  | Text | Working days                 |

| Field   | Type | Null | Key | Default | Extra |
|---------|------|------|-----|---------|-------|
| placeID | int  | YES  |     | HULL    |       |
| hours   | text | YES  |     | HULL    |       |
| days    | text | YES  |     | HULL    |       |

#### Chefmozparking:

| Column Name | Data type | Description                       |
|-------------|-----------|-----------------------------------|
| Place ID    | Integer   | Place id for the place            |
| Parking lot | Varchar   | Describes the type of the parking |

| Field       | Type | Null | Key | Default | Extra |
|-------------|------|------|-----|---------|-------|
| placeID     | int  | YES  |     | HULL    |       |
| parking_lot | text | YES  |     | HULL    |       |

#### rating\_final:

| Column Name    | Data type | Description                      |
|----------------|-----------|----------------------------------|
| User ID        | Varchar   | Id for the user                  |
| Place ID       | Integer   | Place id for the place           |
| Rating         | integer   | Rating given by the user         |
| Food_Rating    | Integer   | Food rating given by the user    |
| Service_Rating | integer   | Service rating given by the user |

| Field          | Type | Null | Key | Default | Extra |
|----------------|------|------|-----|---------|-------|
| userID         | text | YES  |     | HULL    |       |
| placeID        | int  | YES  |     | HULL    |       |
| rating         | int  | YES  |     | HULL    |       |
| food_rating    | int  | YES  |     | HULL    |       |
| service_rating | int  | YES  |     | HULL    |       |

#### Usercuisine:

| Column Name | Data type | Description          |
|-------------|-----------|----------------------|
| User ID     | Varchar   | User id for the user |
| Rcuisine    | Varchar   | Type of the cuisines |

| Field    | Type | Null | Key | Default | Extra |
|----------|------|------|-----|---------|-------|
| userID   | text | YES  |     | HULL    |       |
| Rcuisine | text | YES  |     | HULL    |       |

#### Userpayment:

| Column Name | Data type | Description          |
|-------------|-----------|----------------------|
| User ID     | Varchar   | User id for the user |
| Upayment    | Varchar   | Type of the payment  |

| Field    | Type | Null | Key | Default | Extra |
|----------|------|------|-----|---------|-------|
| userID   | text | YES  |     | NULL    |       |
| Upayment | text | YES  |     | NULL    |       |

## ANALYSIS

### 1) Objective:

To find out the total visits to all restaurants under all alcohol categories available.

### Query:

```
SELECT a.placeID,a.name, a.alcohol, COUNT(b.userid) AS total_visits
FROM geoplaces2 a JOIN rating_final b
ON a.placeID = b.placeID
where a.alcohol not like "%NO_Alcohol%"
GROUP BY alcohol, a.placeID, a.name
ORDER BY total_visits DESC;
```

### Output:

| placeID | name                               | alcohol   | total_visits |
|---------|------------------------------------|-----------|--------------|
| 135032  | Cafeteria y Restaurant El Pacifico | Wine-Beer | 28           |
| 135052  | La Cantina Restaurante             | Full_Bar  | 25           |
| 132862  | La Posada del Virrey               | Wine-Beer | 18           |
| 135041  | Luna Cafe                          | Wine-Beer | 17           |
| 132921  | crudalia                           | Wine-Beer | 17           |
| 135028  | La Virreina                        | Wine-Beer | 15           |
| 135057  | El Herradero Restaurante and Bar   | Wine-Beer | 15           |
| 132856  | Unicols Pizza                      | Wine-Beer | 14           |
| 135076  | Restaurante Pueblo Bonito          | Wine-Beer | 13           |
| 135045  | Restaurante la Gran Via            | Wine-Beer | 13           |
| 132723  | Gordas de morales                  | Full_Bar  | 12           |
| 135066  | Restaurante Guerra                 | Wine-Beer | 12           |
| 135069  | Abondance Restaurante Bar          | Wine-Beer | 12           |
| 135026  | la Cantina                         | Full_Bar  | 11           |

Number of rows: 43

### Explanation:

We used inner join between the geoplaces2 table and the rating\_final table on the common field of the place ID. WHERE clause filters the result set to only include places that serve alcohol. GROUP BY groups the result set by the alcohol availability status, place ID, and name to ensure that the aggregate function (COUNT) is applied to each group separately.

### Inference:

This analysis helps in analyzing the popularity of places that serve alcohol and identifying the number of users who have rated those places.

## 2) Objective:

Find out the average rating according to alcohol and price so that we can understand the rating in respective price categories as well.

| placeID | name  | alcohol   | price  | rating according to alcohol | rating according to price |
|---------|---|-----------|--------|-----------------------------|---------------------------|
| 135026  | la Cantina  | Full_Bar  | high   | 1.2625                      | 1.3091                    |
| 134975  | Rincon del Bife                                   | Full_Bar  | high   | 1.2625                      | 1.3091                    |
| 134983  | Restaurant and Bar and Clothesline Carlos N Ch... | Full_Bar  | high   | 1.2625                      | 1.3091                    |
| 135052  | La Cantina Restaurante                            | Full_Bar  | high   | 1.2625                      | 1.3091                    |
| 135018  | El Oceano Dorado                                  | Full_Bar  | medium | 1.2625                      | 1.2416                    |
| 135071  | Restaurante la Cantina                            | Full_Bar  | medium | 1.2625                      | 1.2416                    |
| 132723  | Gordas de morales                                 | Full_Bar  | medium | 1.2625                      | 1.2416                    |
| 135104  | vips  | Full_Bar  | medium | 1.2625                      | 1.2416                    |
| 132937  | rockabilly  | Full_Bar  | low    | 1.2625                      | 1.0893                    |
| 134986  | Restaurant Las Mananitas                          | Wine-Beer | high   | 1.2445                      | 1.3091                    |
| 135073  | Restaurante Bar El Gallinero                      | Wine-Beer | high   | 1.2445                      | 1.3091                    |
| 134992  | Restaurant Teely                                  | Wine-Beer | high   | 1.2445                      | 1.3091                    |
| 135045  | Restaurante la Gran Via                           | Wine-Beer | high   | 1.2445                      | 1.3091                    |
| 135066  | Restaurante Guerra                                | Wine-Beer | high   | 1.2445                      | 1.3091                    |

## Query:

```
SELECT distinct b.placeID,b.name, b.alcohol, b.price,
  AVG(a.rating) OVER(PARTITION BY b.alcohol) as `rating according to alcohol`,
  AVG(a.rating) OVER(PARTITION BY b.price) as `rating according to price`
FROM rating_final a JOIN geoplaces2 b
ON a.placeID = b.placeID
WHERE b.alcohol NOT LIKE "%NO_Alcohol%"
ORDER BY AVG(a.rating) OVER(PARTITION BY b.alcohol) DESC,
  AVG(a.rating) OVER(PARTITION BY b.price) DESC;
```

Output: number of rows=43

## Explanation:

The result set is sorted by the average rating for each alcohol availability and by the average rating for each price group in descending order allowing the user to identify the places with the best ratings for each group.

The AVG function is used with the OVER clause to calculate the average rating for each alcohol availability group and each price group. This creates a separate group of rows for each group, and the AVG function is applied to each group to calculate the average rating.

## Inference:

There are 4 restaurants which top the rating list with respect to both price and alcohol, which include la Cantina, Rincon del Bife, Restaurant and Bar and Clothesline Carlos N Charlies and La Cantina Restaurante even though the price is high.



Lowest ratings are given to particularly 5 restaurants which belong under the wine-beer category. These restaurants have low ratings despite having low prices.

Overall, this query is helpful in analyzing the relationship between the rating score, alcohol availability, and the price of the places. The use of the AVG function allows for the efficient calculation of the average rating for each group, making the query faster and more efficient.

### 3) Objective:

To quantify that what are the parking availability as well in different alcohol categories along with the total number of restaurants.

Query:

```
SELECT a.alcohol AS alcohol_type, b.parking_lot,
       COUNT(DISTINCT a.placeID) AS total_restaurants,
       SUM(b.parking_lot IN
           ('public', 'yes', 'valet parking', 'fee', 'street', 'validated parking')) AS ParkingAvailable_count,
       SUM(b.parking_lot = 'none') AS NoParking_count
FROM geoplaces2 a LEFT JOIN chefmozparking b
ON a.placeID = b.placeID
WHERE a.alcohol NOT LIKE '%NO_Alcohol%'
GROUP BY a.alcohol, b.parking_lot;
```

Output:

| alcohol_type | parking_lot   | total_restaurants | ParkingAvailable_count | NoParking_count |
|--------------|---------------|-------------------|------------------------|-----------------|
| Full_Bar     | none          | 3                 | 0                      | 3               |
| Full_Bar     | public        | 1                 | 1                      | 0               |
| Full_Bar     | valet parking | 1                 | 1                      | 0               |
| Full_Bar     | yes           | 4                 | 4                      | 0               |
| Wine-Beer    | none          | 12                | 0                      | 12              |
| Wine-Beer    | public        | 2                 | 2                      | 0               |
| Wine-Beer    | valet parking | 2                 | 2                      | 0               |
| Wine-Beer    | yes           | 18                | 18                     | 0               |

Explanation:

This SQL query retrieves the total number of restaurants for each alcohol availability and parking lot combination, and the count of restaurants that have parking available or no parking at all.

Inference:

It can be inferred that there are a total of three Full-bar restaurants with no parking availability. In case of wine-beer restaurants there are 12 restaurants which do not have parking facilities.



#### 4) Objective:

To Also take out the percentage of different cuisine in each alcohol type.

#### Query:

```
SELECT
  a.alcohol AS alcohol_type,
  b.rcuisine AS cuisine_type,
  COUNT(DISTINCT a.placeid) AS total_restaurants,
  SUM(p.parking_lot
    IN ('public', 'yes', 'valet parking', 'fee', 'street', 'validated parking')) AS parking_available_count,
  SUM(p.parking_lot = 'none') AS no_parking_count,
  ROUND(COUNT(DISTINCT a.placeid) / SUM(COUNT(DISTINCT a.placeid)) OVER (PARTITION BY a.alcohol) * 100, 2) AS cuisine_percentage
FROM geoplaces2 a
JOIN chefmozCuisine b ON a.placeid = b.placeid
JOIN chefmozparking p ON a.placeid = p.placeid
WHERE a.alcohol NOT LIKE '%NO_Alcohol%' AND a.country <> '?'
GROUP BY a.alcohol, b.rcuisine
ORDER BY a.alcohol, cuisine_percentage DESC;
```

#### Output:

| alcohol_type | cuisine_type    | total_restaurants | parking_available_count | no_parking_count | cuisine_percentage |
|--------------|-----------------|-------------------|-------------------------|------------------|--------------------|
| Full_Bar     | Bar             | 5                 | 2                       | 3                | 45.45              |
| Full_Bar     | Bar_Pub_Brewery | 4                 | 2                       | 2                | 36.36              |
| Full_Bar     | Mexican         | 2                 | 2                       | 0                | 18.18              |
| Wine-Beer    | Bar             | 6                 | 5                       | 1                | 31.58              |
| Wine-Beer    | International   | 3                 | 3                       | 0                | 15.79              |
| Wine-Beer    | Mexican         | 3                 | 0                       | 3                | 15.79              |
| Wine-Beer    | Cafeteria       | 2                 | 1                       | 1                | 10.53              |
| Wine-Beer    | Contemporary    | 2                 | 2                       | 0                | 10.53              |
| Wine-Beer    | Bar_Pub_Brewery | 1                 | 1                       | 0                | 5.26               |
| Wine-Beer    | Italian         | 1                 | 0                       | 1                | 5.26               |
| Wine-Beer    | Japanese        | 1                 | 1                       | 0                | 5.26               |

#### Inference:

Maximum varieties of cuisines are present in wine-beer type of restaurants. Majority of full bar and wine-beer restaurants have bar cuisine. The cuisines with the least percentage include bar pub brewery, Italian, Japanese, and Mexican.

It can also be seen that all the restaurants with Mexican cuisine have parking availability. All the international and contemporary restaurants also have parking availability.

#### 5) Objective:

To take out the average rating of each state.

#### Query:

```

update geoplaces2 set state= replace(state,"san luis potos","San Luis Potosi");
update geoplaces2 set state= replace(state,"San Luis Potosii","San Luis Potosi");
SELECT
    a.state, ROUND(AVG(b.rating), 2) AS average_rating
FROM
    geoplaces2 a
    INNER JOIN
    rating_final b ON a.placeid = b.placeid where a.state <> "?"
GROUP BY state order by average_rating desc;

```

Output:

| state           | average_rating |
|-----------------|----------------|
| s.l.p.          | 1.38           |
| Morelos         | 1.36           |
| mexico          | 1.24           |
| SLP             | 1.23           |
| San Luis Potosi | 1.13           |
| Tamaulipas      | 0.91           |

Explanation:

The first operation updates the "state" column in the "geoplaces2" table by replacing any instances of "san luis potos" and "San Luis Potosii" with "San Luis Potosi". The strings "san luis potos" and "San Luis Potosii" were believed to be the same as the string "San Luis Potosi". Hence, they were corrected.

The second operation calculates the average rating of restaurants in each state, rounding the result to two decimal places. The result set is sorted in descending order by the average rating, which is rounded to two decimal places using the ROUND function.

Inference:

It is seen that the highest rating is given to the state s.l.p which is presumed to be San Luis Potosi but due to uncertainty, it is left unchanged since this has to be discussed and cross-checked. The state with the least rating is Tamaulipas.

6) Objective:

'Tamaulipas' Is the lowest rated state. Quantify the reason why it is the lowest rated by providing the summary on the basis of State, alcohol, and Cuisine.

Query:

```

SELECT
    COUNT(gp.placeid) AS number_of_restaurants,
    Rcuisine,
    alcohol
FROM
    geoplaces2 gp
    INNER JOIN
    chefmozcuisine cc ON gp.placeid = cc.placeid
WHERE
    gp.state LIKE '%Tamaulipas%'
GROUP BY Rcuisine , alcohol;

```

Output:

| number_of_restaurants | Rcuisine  | alcohol           |
|-----------------------|-----------|-------------------|
| 9                     | Mexican   | No_Alcohol_Served |
| 1                     | Pizzeria  | No_Alcohol_Served |
| 1                     | Armenian  | No_Alcohol_Served |
| 1                     | Italian   | No_Alcohol_Served |
| 1                     | Fast_Food | No_Alcohol_Served |
| 1                     | Regional  | No_Alcohol_Served |

### Inference:

None of the restaurants in Tamaulipas serve alcohol. There are no international restaurants in Tamaulipas. Further, there are only 6 types of cuisines in Tamaulipas state, compared to other states like san luis potos which have 11 type of cuisines.

### 7) Objective:

To find the average weight, food rating, and service rating of the customers who have visited KFC and tried Mexican or Italian types of cuisine, and also their budget level is low.

### Query:

```
select b.userid,avg(b.weight) as avg_weight , a.food_rating,a.service_rating from userprofile b,
(SELECT
  placeid, userid, food_rating, service_rating
FROM
  rating_final
WHERE
  placeid IN (SELECT
    placeid
  FROM
    geoplaces2
  WHERE
    placeid IN (SELECT
      placeid
    FROM
      chefmozcuisine
    WHERE
      Rcuisine LIKE '%mexican%'
      OR Rcuisine LIKE '%italian%') and name like "%kfc%"
  AND price LIKE '%low%'))a where a.userid=b.userid group by b.userid ,food_rating,service_rating
order by avg_weight desc;
```

### Output:

| userid | avg_weight | food_rating | service_rating |
|--------|------------|-------------|----------------|
|--------|------------|-------------|----------------|

### Explanation:

The query retrieves the following information:

- b.userid: the user ID from the "userprofile" table.
- AVG (b. weight): the average weight from the "userprofile" table.
- a. food\_rating: the food rating from the "rating\_final" table.
- a. service\_rating: the service rating from the "rating\_final" table.

The query used a subquery to filter the data the only include rows where “placeid “is associated with a restaurant that serves Mexican or Italian cuisine, is named “KFC” and has a low price.

But it is seen that there are no customers who have visited KFC and tried Mexican or Italian types of cuisine.

#### CONCLUSION:

- Majority of the restaurants serve the wine-beer type of alcohol.
- The highest rating with respect to price, as well as the alcohol is given to the restaurants under the category of full-bar type.
- Even though majority of the restaurants are wine-beer type, in case of ratings, full-bar restaurants top the rating list. Specific restaurants such as la Cantina have high ratings because of which it is the highest visited restaurant. Whereas 'Rincon del Bife' is least visited and still has the highest rating.
- There are only 3 restaurants in full bar category which do not provide parking facilities. Whereas, in case of wine-beer restaurants, there are 12 restaurants without parking facility. This might be one of the reasons for high ratings of full bar restaurants compared to the wine beer restaurants.
- Maximum varieties of cuisines are present in wine-beer type of restaurants. Majority of full bar and wine beer restaurants have bar cuisine. The cuisines with least percentage include bar pub brewery, Italian, Japanese and Mexican.
- It was seen that the state with lowest rating is “Tamaulipas”. There is uncertainty with respect to the highest rated state as it is inferred to be s.l.p and unclear about what it represents.
- The low ratings of Tamaulipas might be because None of the restaurants in Tamaulipas serve alcohol. Further, there are only 6 types of cuisines in Tamaulipas state, compared to other states like san luis potos which have 11 types of cuisines. Hence the state must work on these aspects to increase its ratings.