**Zomato Restaurants Analysis**

**Tasks**

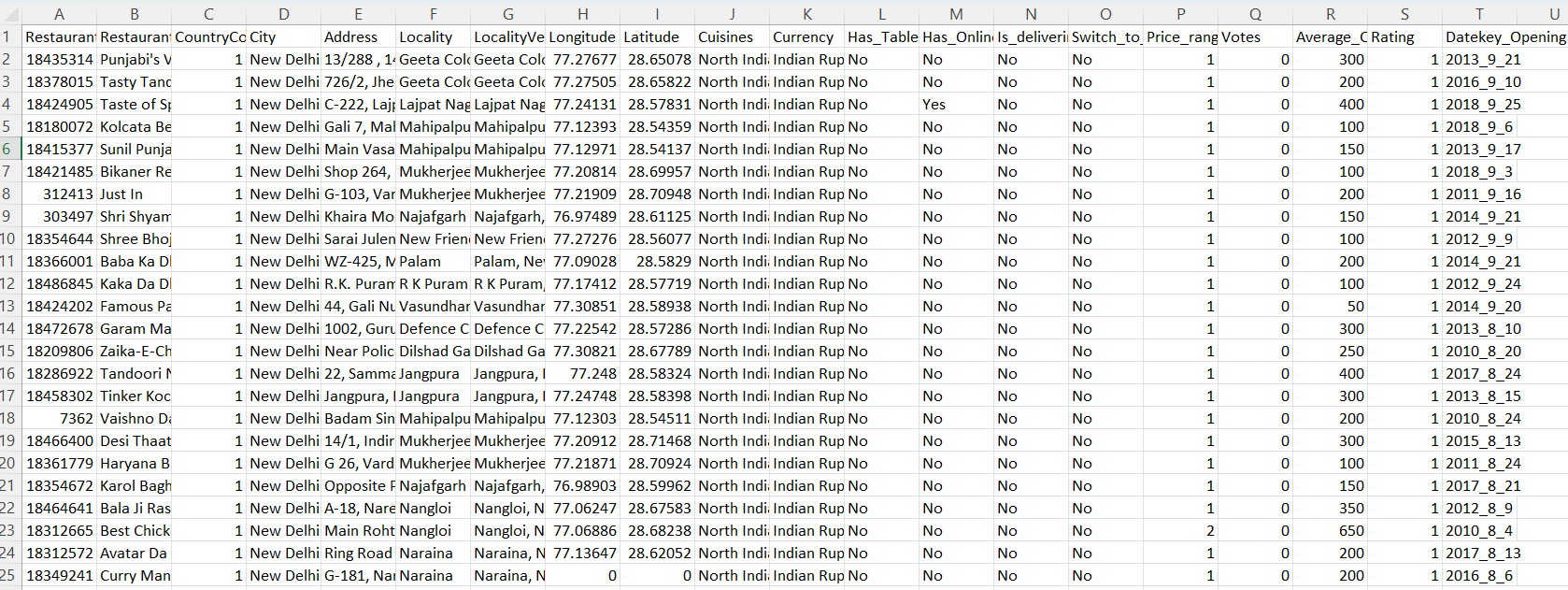
**Objective Questions:**

**1. What is the total no. of tables present in the data?**

**A.** There are 2 tables present in the data. One in the “Raw Data” Sheet and other in the “country description” Sheet.

**2. What is the total no. of attributes present in the data?**

**A.**

****

20 attributes are present in the data.

**3. How many categorical columns are there in the data? [Search about categorical and continuous** **data, and try to answer this question]**

**A.** Ther are 13 categorical columns. They are RestaurantName, CountryCode, City, Address, Locality, LocalityVerbose, Cuisines, Currency, Has-Table-Booking, Has-Online-Delivery, Is-Delivering-Now, Price-range and Rating.

We didn’t include the column “Switch-to-order-menu” in the categorical columns because it has only one value, that is, “No”. So, this will be considered as “Constant Column”.

**4. The data consists of some inconsistent and missing values so ensure that the data used for further analysis is cleaned.**

**A.** The correct data types are provided to all the columns.

There are 9 missing values in Cuisines Column. Used formula “=COUNTBLANK(K1:K9552)” to know the missing values in each columns. There are 9551 records in the data. So, removing 9 records won’t affect the data greatly. To remove 9 rows from the data,

1. Select Cuisines Column -> Press Ctrl+G -> Click Special -> Click Blanks -> Click ok

2. In the Home Ribbion, Click Delete -> Click Delete Sheet Rows

There are no duplicate values in the data. To check for duplicate values,

Select entire Data -> Click Data Tab -> Click Remove Duplicates -> Click OK.

**5. Using the LookUp functions, fill up the countries in the original data using the country code.**

**A.** Created column “Country” in Raw Data Sheet and filled it using formula “=VLOOKUP($C2,'country description'!$A$1:$B$16,2,0)”.

**6. Create a table to represent the number of restaurants opened in each country.**

**A.**

|  |  |
| --- | --- |
| **Country** | **Count of RestaurantID** |
| Canada | 4 |
| Qatar | 20 |
| Singapore | 20 |
| Sri Lanka | 20 |
| Indonesia | 21 |
| Philippines | 22 |
| Australia | 24 |
| Turkey | 34 |
| New Zealand | 40 |
| Brazil | 60 |
| United Arab Emirates | 60 |
| South Africa | 60 |
| United Kingdom | 80 |
| United States of America | 425 |
| India | 8652 |
| **Grand Total** | **9542** |

**Insights:**

Canada has lowest number of restaurants whereas India has highest number of restaurants.

So, Canada is the number one choice for opening new restaurants.

**Steps:**

1. Go to insert tab -> Click PivotChart -> Click PivotChart & PivotTable.

2. Drag “Country” in rows and “ResturantID” in values.

3. Click “RestaurantID” in Values -> Click on Value Field Settings -> Click on Count -> Click OK.

**7. Also, the management wants to look at the number of restaurants opened each year, so provide them with something here.**

**A.**

|  |  |
| --- | --- |
| **Opening Year** | **Count of RestaurantID** |
| 2010 | 1079 |
| 2011 | 1096 |
| 2012 | 1022 |
| 2013 | 1059 |
| 2014 | 1049 |
| 2015 | 1023 |
| 2016 | 1026 |
| 2017 | 1086 |
| 2018 | 1102 |
| **Grand Total** | **9542** |

**Insights:**

The lies in the range 1021 – 1102, meaning nearly 1060 opens every year.

**Steps:**

1. Go to insert tab -> Click PivotChart -> Click PivotChart & PivotTable.

2. Drag “Datekey-Opening” in rows and “ResturantID” in values.

3. Click “RestaurantID” in Values -> Click on Value Field Settings -> Click on Count -> Click OK.

**8. What is the total number of restaurants in India in the price range of 4?**

**A.** Total number of restaurants in India in the price range of 4 = 388.

This suggests that out of all the 8652 only 388 restaurants are charging high price.

Used formula “=COUNTIFS('Raw Data'!$D$2:$D$9543,"India",'Raw Data'!$Q$2:$Q$9543,4)”

**9. What is the average number of voters for the restaurants in each country according to the data?**

**A.**

|  |  |
| --- | --- |
| **Country** | **Average of Votes** |
| Australia | 111 |
| Brazil | 20 |
| Canada | 103 |
| India | 137 |
| Indonesia | 772 |
| New Zealand | 243 |
| Philippines | 407 |
| Qatar | 164 |
| Singapore | 32 |
| South Africa | 315 |
| Sri Lanka | 146 |
| Turkey | 431 |
| United Arab Emirates | 494 |
| United Kingdom | 205 |
| United States of America | 428 |
| **Grand Total** | **157** |

The above table shows the number of votes received by restaurants in different countries.

**Insights:**

Indonesia has highest (772) Average Votes and Brazil has lowest (20) Average Votes.

From this, we can consider Brazil as one of the Counties where new Restaurant can be opened.

**Steps:**

1. Go to insert tab -> Click PivotChart -> Click PivotChart & PivotTable.

2. Drag “Country” in rows and “Votes” in values.

3. Click “Votes” in Values -> Click on Value Field Settings -> Click on Average -> Click OK.

**10. Calculate the average rating for all the restaurants that have price\_range < 4 and provide online delivery. Use only the “IF” function, Logical Operators, and Aggregation functions to solve this problem. [Note: Don’t use Conditional aggregation in this question.]**

**A.**

**Insights:**

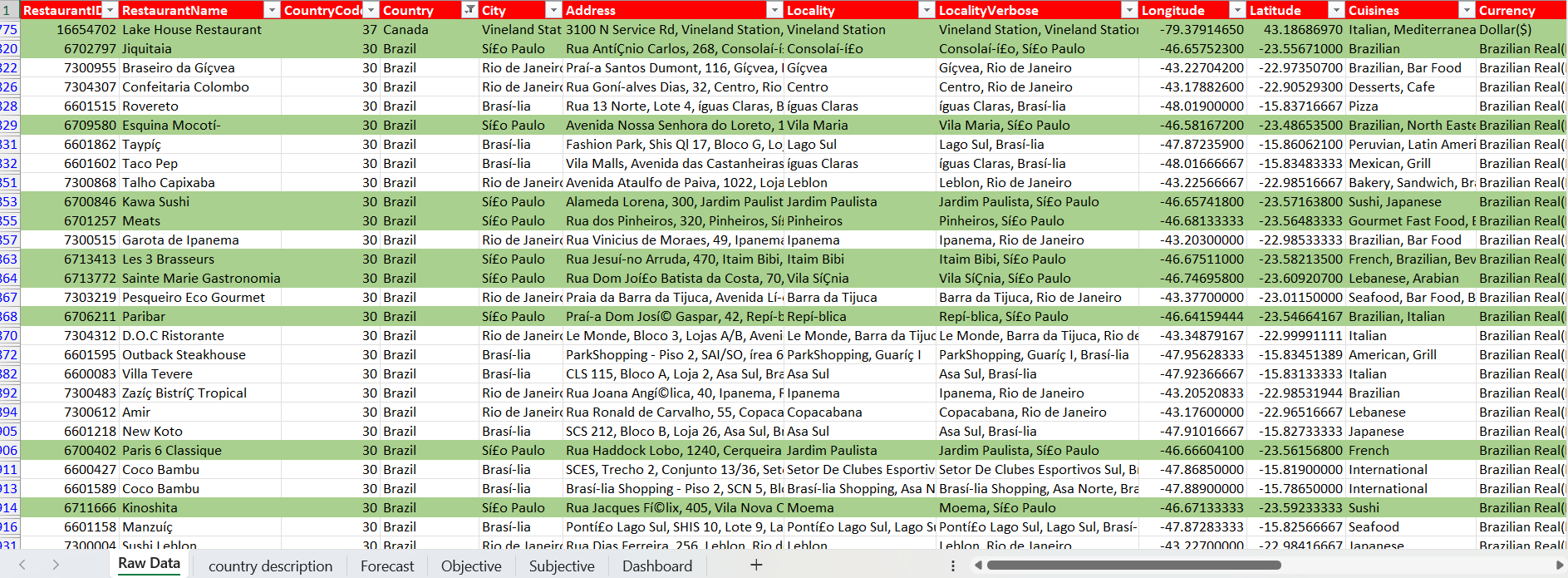
The average rating for all the restaurants that have price\_range < 4 and provide online delivery = 3.27.

**Steps:**

Used the formula “=AVERAGE(IF(('Raw Data'!$Q$2:$Q$9543<4)\*('Raw Data'!N2:N9543="Yes"),'Raw Data'!$T$2:$T$9543))” and press Ctrl+Shift+Enter. This will automatically calculate as array formula.

**11. Using Conditional formatting highlight the rows of restaurants that are located in the countries or cities that you’ve suggested to the management for opening new restaurants.**

**A.**



**Insights:**

Conditional formatting is applied on the Sí£o Paulo, Chatham-Kent, Consort, Vineland Station and Yorkton cities. This highlights the rows based on the formula given in the steps. Above image shows a small portion how the conditional formatting looks.

**Steps:**

1. Click Conditional Formatting -> Click New Rule -> Click Use a formula to determine which cells to format.

2. Enter the formula “=OR($E1 = ($G$35:$G$39)) -> Click Ok

3. Click Conditional Formatting -> Click Manage Rules -> Click on the applies to -> Select range $A$1:$W$9543 -> Click Apply -> Click Ok.

**12. Create a new customized price column that consists of the abbreviation/symbol of the currency along with the Average\_cost\_for\_two value. [Use string operations to do this task]**

**A.** Created a new column named ‘Price’ in the Raw Data Sheet andused formula

“=MID(L2,FIND("(",$L2)+1,FIND(")",$L2)-FIND("(",$L2)-1) &" " & $S2”

for this task.

**13. How can you create an array formula in Excel or Google Sheets to count the number of restaurants listed that do not offer online delivery, are in the lowest price range, and have an average cost for two people less than or equal to 250 Indian Rupees?**

**A.**

**Insights:**

The number of restaurants listed that do not offer online delivery, are in the lowest price range, and have an average cost for two people less than or equal to 250 Indian Rupees is 1691.

**Steps:**

Used array formula,

“=SUM(IF((('Raw Data'!$N:$N="No") \* ('Raw Data'!$Q:$Q=MIN('Raw Data'!$Q$1:$Q$9543)) \* ('Raw Data'!$W:$W<=250)), 1, 0))” and press Ctrl+Shift+Enter.

Or, the below formula can also be used

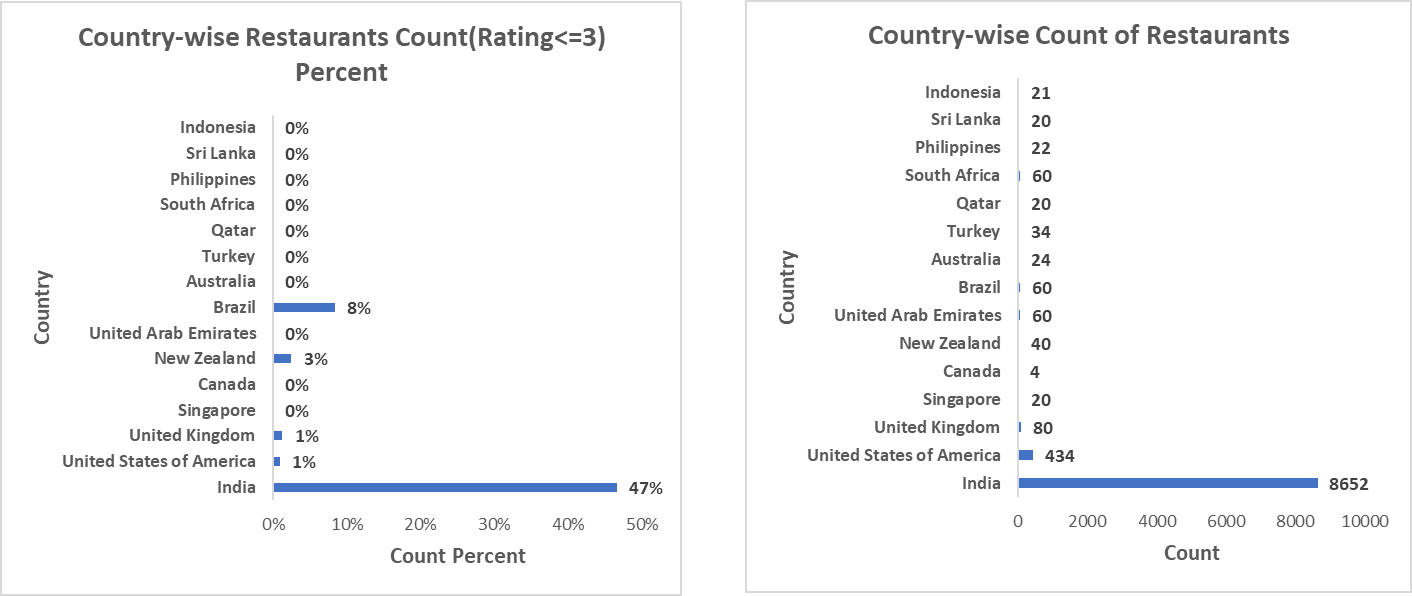
Used formula, “=COUNTIFS('Raw Data'!$N:$N,"No",'Raw Data'!$Q:$Q,MIN('Raw Data'!$Q$1:$Q$9543),'Raw Data'!$W:$W,"<=250")” and press Ctrl+Shift+Enter.

**Subjective Question:**

**1. Suggest a few countries where the team can open newer restaurants with lesser competition. Which visualization/technique will you use here to justify the suggestions?**

**A.** Few countries where the team can open newer restaurants with lesser competition are-

**Canada and Brazil**.



**Insights:**

In the Column “Restaurants Count(Rating<=3) Percent” checked for the values and found that **Brazil and India** were 8% and 47% respectively, which we highest among the others. But India was excluded from the list as its restaurant count is very high.

**Canda** was also chosen as Cananda has 1 restaurant per City provided in data.

**Steps:**

Calculated Percent of Count of Restaurants where Rating is less than equal to 3.

Created a Column Named “Country” using formula “==UNIQUE('Raw Data'!D2:D9543)”.

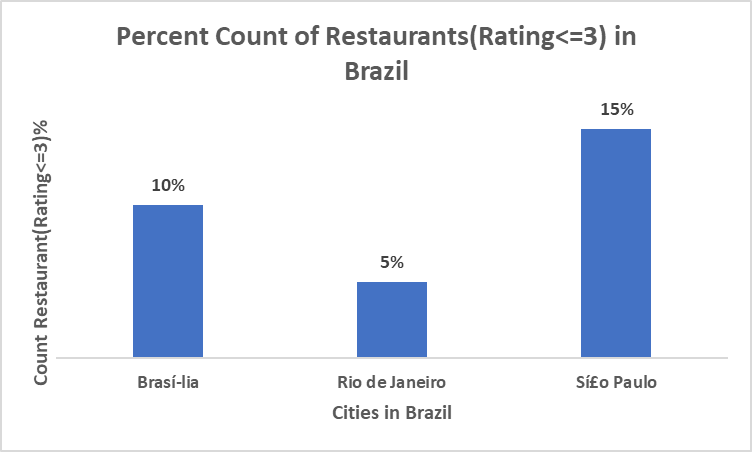
Created a Column Named “Restaurant Count(Rating<2.4)” using formula “=COUNTIFS('Raw Data'!$D$2:$D$9543,Subjective!$A2,'Raw Data'!$T$2:$T$9543,"<=3")”.

Created Column Named “Count of Restaurants” using formula “=COUNTIF('Raw Data'!$D$2:$D$9543,Subjective!$A2)”.

Created Column Named “Restaurants Count(Rating<=3) Percent” using formula “=ROUND($E2/$F2,4)” and then converting to Percent.

**2. Come up with the names of States and cities in the suggested countries suitable for opening restaurants.**

**A.**

****

|  |  |
| --- | --- |
| **Country (City)** | **Count of Restaurant** |
| **Canada** | |
| Chatham-Kent | 1 |
| Consort | 1 |
| Yorkton | 1 |
| Vineland Station | 1 |

**Insights:**

Cities in Brazil for opening suitable for opening new restaurant include **Sí£o Paulo** as it stands at top in the count restaurants where the rating is less than equal to 3.

Cities in for opening suitable for opening new restaurant Canada include **Chatham-Kent, Consort, Vineland Station** and **Yorkton**. These cities are chosen because all the cities have only 1 restaurant.

**Steps:**

For creating the columns chart

1. Select all the data in the Raw Sheet -> Click Insert Table -> Click PivotCharts & PivotTable.

2. Drag “Country” in rows and then again drag “City” in rows.

3. Copy and paste the cities of Brazil in another column.

4. Calculating Restaurant Count(Rating<=3) using the formula

“=COUNTIFS('Raw Data'!$E$2:$E$9543,Subjective!$A27,'Raw Data'!$T$2:$T$9543,"<=3")”

5. Calculating Count of Restaurant using formula

“=COUNTIFS('Raw Data'!$E$2:$E$9543,Subjective!$A27)”

6. Calculating Restaurants(Rating<=3) Percent using formula

“=IFERROR(ROUND(C27/D27,4), "")”.

7. Select the new table created -> Click Insert tab -> Click Insert Column or Bar Chart -> Select Clustered Column.

For creating the table

1. 1. Select all the data in the Raw Sheet -> Click Insert Table -> Click PivotCharts & PivotTable.

2. Drag “Country” in rows and then again drag “City” in rows.

3. Copy and paste the cities of Canada in another column.

4. Calculating Count of Restaurants using formula

“=COUNTIF('Raw Data'!$E$1:$E$9543,Subjective!$A34)”.

**3. According to the countries you suggested, what is the current quality regarding ratings for restaurants that are open there?**

**A.**

**Insights:**

The maximum, minimum and average ratings for Brazil are 4.9, 1 and 3.8 respectively.

The maximum, minimum and average ratings for Canada are 4.3, 3 and 3.6 respectively.

**Steps:**

1. Created a column ‘Country’ and used formula

“=UNIQUE('Raw Data'!D2:D9543)”.

to get the names of all the countries.

2. Created column ‘Max of Rating’ and used formula

“=MAXIFS('Raw Data'!$T$2:$T$9543,'Raw Data'!$D$2:$D$9543,Subjective!$A3)”.

3. Created column ‘Min of Rating’ and used formula

“=MINIFS('Raw Data'!$T$2:$T$9543,'Raw Data'!$D$2:$D$9543,Subjective!$A3)”.

4. Created column ‘Average of Rating’ and used formula

“=AVERAGEIF('Raw Data'!$D$2:$D$9543,Subjective!$A3,'Raw Data'!$T$2:$T$9543)”.

5. Select the above 4 columns. Go to Insert Tab -> Click Insert Column or Bar Chart -> Click Stacked Column.

**4. Also, what is the current expenditure on food in the suggested countries, so we can keep our financial expenditure in control?**

**A.**

**Insights:**

In Canada and Brazil, the avg. expenditure is Rs. 2207.99 and Rs. 2185.64 respectively.This shows that there are people who spends on food in restaurants.

**Steps:**

1. Created a column ‘Country’ and used formula

“=UNIQUE('Raw Data'!D2:D9543)”.

2. Created column ‘Avg. Expenditure(in Rs.)’ and used formula

“=ROUND(AVERAGEIF('Raw Data'!$D$2:$D$9543,Subjective!$A3,'Raw Data'!$W$2:$W$9543),2)”.

3. Select the above 2 columns. Go to Insert Tab -> Click Insert Column or Bar Chart -> Click Clustered Bar.

**5. Come up with the names of restaurants from the recommended states that are our biggest competitors and also those that are rated in the lower brackets, i.e. 1-2 or 2-3.**

**A.** Names of restaurants from the recommended states that are our biggest competitors:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The restaurants that could be our biggest competitors** | | | | |
| **Sí£o Paulo** | **Chatham-Kent** | **Consort** | **Yorkton** | **Vineland Station** |
| Skye - Hotel Unique | Tokyo Sushi | Consort Restaurant | Arigato Sushi | Lake House Restaurant |

Name of the restaurants from the recommended states that are in the lower bracket (Rating<3):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The lower bucket restaurants** | | | | |
| **Sí£o Paulo** | **Chatham-Kent** | **Consort** | **Yorkton** | **Vineland Station** |
| Cantinho da Gula |  |  |  |  |
| Super Grill |  |  |  |  |
| Divino Fogí£o |  |  |  |  |

The first table contains the names of the restaurants with highest rating in the city.

The second table contains the names of the restaurants with ratings less than 3.

**Insights:**

In Sí£o Paulo, the top competitor is Skye - Hotel Unique.

In Chatham-Kent, Consort, Yorkton and Vineland Station there are only 1 restaurant per city. So, our top competitors are Tokyo Sushi, Consort Restaurant, Arigato Sushi and Lake House Restaurant respectively.

**Steps:**

1. To get the first table, we arranged the city names in a row

2. Used formula

“=FILTER('Raw Data'!$B:$B,('Raw Data'!$E:$E=Subjective!A$37)\*('Raw Data'!$T:$T=MAXIFS('Raw Data'!$T:$T,'Raw Data'!$E:$E,Subjective!A$37)),"")”

for getting the results.

3. To get the second table first we arranged the city names in a row.

4. Used formula

“=FILTER('Raw Data'!$B1:$B9543,('Raw Data'!$E1:$E9543=Subjective!A$43)\*('Raw Data'!$T$1:$T$9543<3),"")”

for getting the results.

**6. Which cuisines should we focus on in the newer restaurants to get better feedback? Does the choice of cuisines affect the restaurant ratings?**

**A.** The cuisines we should focus on in the newer restaurants to get better feedback are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The highest rated restaurants have the following cuisines** | | | | |
| **Sí£o Paulo** | **Chatham-Kent** | **Consort** | **Yorkton** | **Vineland Station** |
| Brazilian | Japanese, Sushi | Chinese, Canadian | Asian | Italian, Mediterranean, Pizza |
| Brazilian, North Eastern |  |  |  |  |
| Gourmet Fast Food, Burger |  |  |  |  |
| French, Brazilian, Beverages |  |  |  |  |
| Lebanese, Arabian |  |  |  |  |
| Brazilian, Italian |  |  |  |  |
| Italian, Pizza |  |  |  |  |
| Steak, BBQ |  |  |  |  |
| Beverages, International |  |  |  |  |
| BBQ, Grill, Brazilian |  |  |  |  |
| Italian |  |  |  |  |
| Brazilian, Bar Food, Beverages |  |  |  |  |

The above table gives the names of the Cuisines provided by the restaurants having highest ratings in their respective cities.

Yes, the choice of cuisine affects the restaurant rating. But we should also consider the taste of the cuisine because at the end taste is what matters the most.

**Insights:**

Sí£o Paulo is a city in Brazil, so Brazilian food is must in the new restaurant. Other than that, we can use Beverages, Italian and fast food like Pizza and Burger in the cuisine.

In the Canadian, cities we can exactly use the cuisines used by our competitor. But we can also include cuisines provided by the restaurant in city that is far away from that location.

**Steps:**

1. To get this table first we arranged the city names in a row.

2. For Sí£o Paulo, formula used is

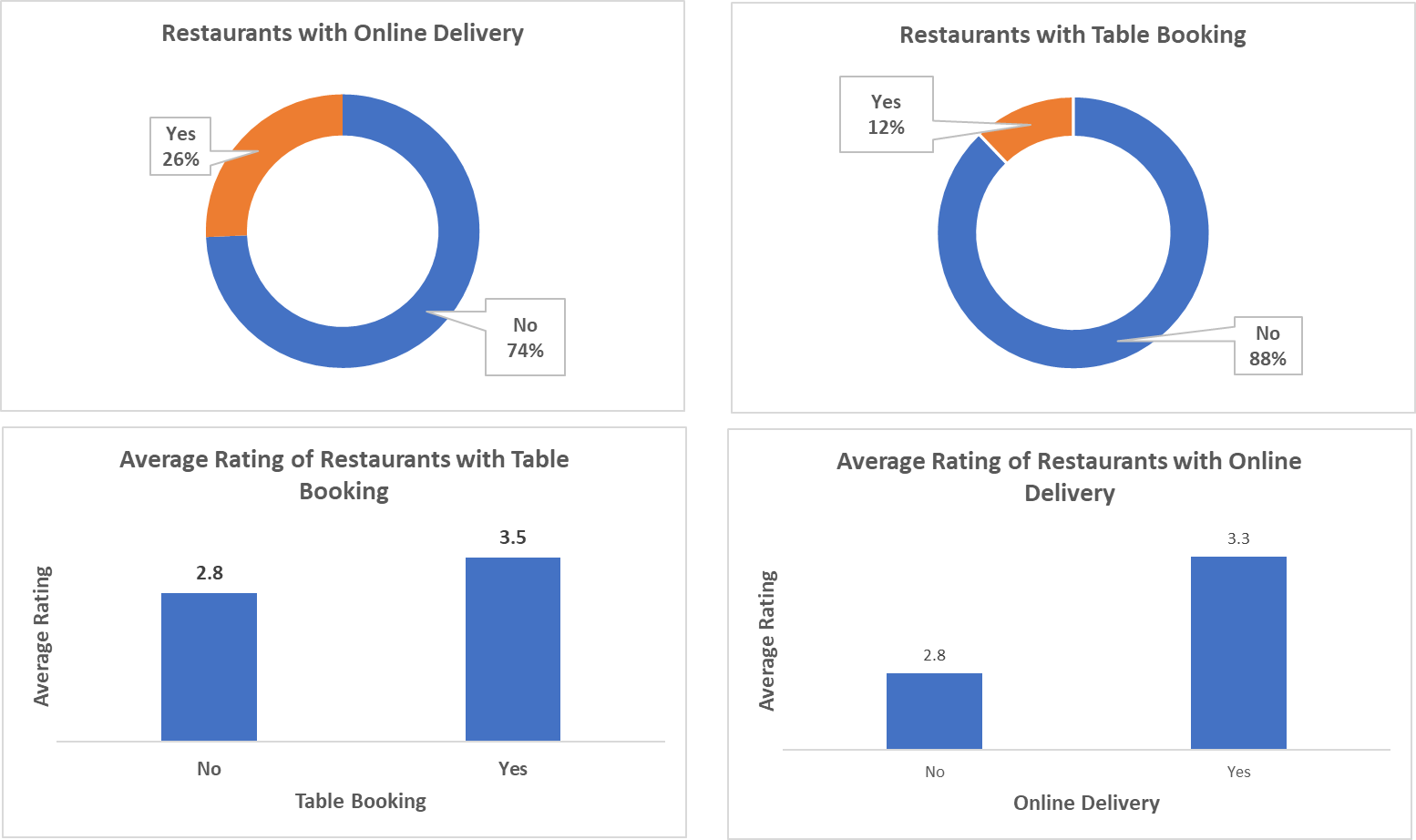
“=FILTER('Raw Data'!$K:$K,('Raw Data'!$E:$E=Subjective!B$56)\*('Raw Data'!$T:$T=MAXIFS('Raw Data'!$T:$T,'Raw Data'!$E:$E,Subjective!B$56)),"")”

3. For the cities of Canada formula used is

“=FILTER('Raw Data'!$K:$K,('Raw Data'!$E:$E=Subjective!B$56)\*('Raw Data'!$T:$T=MAXIFS('Raw Data'!$T:$T,'Raw Data'!$E:$E,Subjective!B$56)),"")”.

**7. According to our current data, should we go for online delivery and table booking? Does that affect the customer’s ratings?**

**A.**



**Insights:**

The above donut chart shows that there are 26% of restaurants having Online Delivery and only 12% have Table Booking.

The average rating for restaurants with Table Booking and Online Delivery is higher than those who do not have these services.

So, it can be concluded that restaurants having Online delivery and Table Booking are rated high.

**Steps:**

1. Go to insert tab -> Click PivotChart -> Click PivotChart & PivotTable.

2. Drag Has-Online\_delivery to Rows, and RestaurantID to Values.

3. Click RestaurantID in values -> Click Value Field Settings -> Click on Count -> Click Ok.

4. Click on Design tab -> Click Change Chart Type -> Select Pie -> Click on Doughnut and you get the Restaurants with Online Delivery Chart.

5. For Restaurants with Table Booking Chart, follow steps 1 – 4, just replace Has-Online-Delivery with Has-Table-Booking in step 2.

6. For Average Rating of Restaurants with Online Delivery Chart, follow steps 1 – 3 and in step 3 instead of Count, Click Average in the Value Field Settings.

7. For Average Rating of Restaurants with Table Booking Chart, follow steps 1 – 3, replace Has-Online-Delivery with Has-Table-Booking in step 2 and in step 3 instead of Count, Click Average in the Value Field Settings.

**8. Should the team keep the rate of cuisines higher? Will that affect the feedback? According to our data are the rates of cuisines and ratings, correlated?**

**A.**

|  |  |
| --- | --- |
| **Average rates of restaurants** | |
| **Country** | **Average of average  cost for two** |
| Brazil | 134.67 |
| Canada | 36.25 |

**Insights:**

The team should keep the rates a little higher than R$ 134.67 in Brazil and a little higher than $ 36.25 in Canada.

According to our data, the rates of cuisines and ratings are not correlated. We got the correlation coefficient of 0.059.

According to our correlation data, the rates of the cuisines does not affect the ratings.

**Steps:**

1. For getting the Average of average cost for two, the formula used is

“=ROUND(AVERAGEIFS('Raw Data'!$S:$S,'Raw Data'!$D:$D,Subjective!$A95,'Raw Data'!$T:$T,">=3"),2)”.

2. For finding the correlation between rates of cuisines and rating, the formula used is

“=ROUND(CORREL('Raw Data'!S2:S9552,'Raw Data'!T2:T9552),3)”.

**9. What is the distribution of the number of restaurants of different price ranges in all the countries?**

**A.** The distribution of the number of restaurants of different price ranges in all the countries is as:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of RestaurantID** | **Price-range** | | | |
| **Country** | **1** | **2** | **3** | **4** |
| Australia | 4 | 14 | 5 | 1 |
| Brazil | 2 | 7 | 16 | 35 |
| Canada |  | 3 |  | 1 |
| India | 4295 | 2858 | 1111 | 388 |
| Indonesia |  | 1 | 20 |  |
| New Zealand | 3 | 4 | 17 | 16 |
| Philippines |  | 1 | 12 | 9 |
| Qatar |  | 1 | 5 | 14 |
| Singapore |  | 1 | 5 | 14 |
| South Africa |  | 4 | 17 | 39 |
| Sri Lanka |  | 6 | 11 | 3 |
| Turkey |  | 11 | 18 | 5 |
| United Arab Emirates |  | 9 | 29 | 22 |
| United Kingdom | 4 | 28 | 32 | 16 |
| United States of America | 136 | 165 | 110 | 23 |

**Insights:**

In Brazil majority of the restaurants are in price-range 3 and 4. So, in beginning the price-range should be 3.

In Canada majority of the restaurant are in the price-range 2. So, in new restaurant the price-range should be 2 or 3.

**Steps:**

1. Go to insert tab -> Click PivotChart -> Click PivotChart & PivotTable.

2. Drag Country to Rows, Price-range to Columns and RestaurantID to Values.

3. Click RestaurantID in values -> Click Value Field Settings -> Click on Count -> Click Ok.

**10. Explain your approach in brief for suggesting countries/cities in order to open new restaurants, if the objective and subjective questions would have not been given to assist you. [you have to give bullet pointers in order to answer this question]**

**A.** If objective and subjective questions were not provided, then my approach would be:

1. The first thing that comes to mind is the rating. So, I would surely consider countries based on rating.

2. While I was doing the project, I thought that Expenditure can also be treated as a factor for considering the countries. More is the expenditure, more people ordering from restaurants. But other factors should be considered too.

3. The columns Country, City, Latitude and Longitude could be used for making a Map Chart which could give us a clear view of the restaurant density within a city and could have used that data for suggesting the location for the opening of the new restaurant. This would have been the most interesting part of the analysis.