

**BY : Sidharth Thakur**

Problem Statement

You are hired as a consultant data analyst by zomato where the team is looking for expansion andopening more restaurants. Your task is to come up with strategies/suggestions about opening newer restaurants.



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

**Method Used for cleaning the data :**

* Formatting Of Date Column **- Datekey\_opening - This column of the date is not set in appropriate date format. For formatting the date, formula used here is =DATEVALUE(SUBSTITUTE(U2,”\_”,”-”,)) - By using this formula date changed from (**2019\_9\_21 to 21-9-2019) .

And after this I have extracted 3 more column from the datekey opening column. The 3 new columns are DAY, MONTH , YEAR.

For DAY formula used is : =DAY(V2)

For MONTH formula used is : =TEXT(V2,”mmmm”)

For YEAR formula used is : =YEAR(V2)

* Border - The data provided did not had any border, so I added borders, so that the data looks good.
* WRAP TEXT - The text in the data were overlapping and due to this it was not understandable. And for overcoming this, wrap text was applied to the data.

### **Missing Values in ‘Average Cost for Two**

### **Observation:** Missing values appeared mostly for restaurants with a rating of 1.

* **Solution:** These were filled using **median imputation**, as this method is robust to outliers.
* **Reasoning:** Since the missing values correlated with low-rated restaurants, median was a better representative than mean.

### **Currency Conversion to INR**

* **Issue:** 'Average Cost for Two' was in multiple currencies.
* **Solution:**
  + A live **currency conversion table** was created using **Power Query** and an **API to fetch real-time USD rates**, which were then converted to INR.
  + **Formula Used:** =S736\*'country description'!$E$2 (For Excel-based transformation)

### **Standardizing Currency Names**

* **Currency names were updated to follow ISO universal naming:**
  + **Indian Rupee → INR**
  + **Dollar → USD**

### **Handling Geographic Coordinates**

* **Columns:** Longitude, Latitude
* **Observation:** Some values were 0 or negative.
* **Decision:** These were **not deleted**, as they may represent valid locations or simply require further geographic filtering.

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

**Function used for extracting country from the country code:**

* Vlookup Function - As the common value in raw data sheet and country description sheet is country code. Therefore I have used Vlookup ( as it lookups the value in vertical column ).

Formula used is :

=VLOOKUP(C2,'country description'!$A$2:$B$16,2,FALSE)

C2 - It is the country code selected for searching.

country description'!$A$2:$B$16 - This is the range where

the code will be searched.

2 - This is column number where country will be searched.

False - This is used to denote the match type. And false means

exact match.

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

**Function used to here to count the restaurants opened in each country:**

* COUNTIF - AS the question ask to create a table and show no. of restaurants in each country. For this we have to count the number of restaurants therefore I have used the countif formula to extract the number of restaurants opened in each country. The formula used is : =COUNTIF('Raw Data'!$D$1:$D$9552,'Analysis '!$A2) .

Raw Data'!$D$1:$D$9552 - This denotes the range selected

Analysis '!$A2 - This is criteria selected, that is country.

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

**Function used to here to count the restaurants opened in each Year :**

* COUNTIF - As the question ask to create a table and show no. of restaurants in each year. For this we have to count the number of restaurants opened in each year therefore I have used the countif formula to extract the number of restaurants opened in each year. The formula used is : =COUNTIF('Raw Data'!$Y$1:$Y$9552,'Analysis '!D2)

'Raw Data'!$Y$1:$Y$9552 - This denotes the range selected

Analysis '!D2 - This is criteria selected, that is year.

* Location of the Table : Sheet named as Analysis in Excel File.

1. What is the total number of restaurants in India which are in the price range 4 ?

**Function used here to count total restaurants in india in price**

**Range of 4 :**

* COUNTIFS - As the question ask to show no. of restaurants in India in price ranga of 4.As there are two conditions,so I have used the countifs formula to extract the number of restaurants in price range of 4 in India. The formula used is : =COUNTIFS('Raw Data'!$D$1:$D$9552,'Analysis '!D22,'Raw Data'!$Q$1:$Q$9552,4)

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

**Function used here to count total restaurants in india in price**

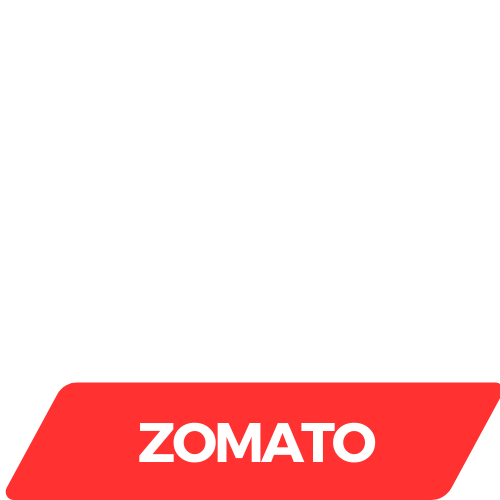
**Range of 4 :**

* AVERAGEIF - As the question ask to show average number of voters for restaurant in each country, I have used Averageif function.The formula used is: = AVERAGEIF('Raw Data'!$D:$D,$A22,'Raw Data'!$R:$R)

Raw Data'!$D:$D - This is the criteria range selected

$A22 - This the criteria for taking average of

Raw Data'!$R:$R - This is the range of voters selected for average





1. Suggest few countries where the team can open newer restaurants with lesser competition. Which visualization/technique will you use here in order to

justify the suggestions?

Method used : I have used Pivot table, in which rows are country and value field is count of restaurant id and average rating.After creating the pivot table, I applied the filter on value fields and selected bottom 8 values. The reason is I want to select countries with less competition and ratings above 3 but less than 4.

* Analysis: After observation, I would suggest countries which comes under the condition of less competition with average rating. The reason for choosing average rating is that, if the people in that country are not satisfied with the restaurants and hence the ratings are less. And this thing can be an advantage, we can do a market survey and analyze the reasons why people are less satisfied and we can focus on those things while opening the restaurant.
* Countries Suggested for opening new restaurant : AUSTRALIA , CANADA , SINGAPORE , SRI LANKA.
* Visualization method used : Line chart
* Location -

Excel file - sheet name - 4) New opening - state and cities

Table - 1) Analysis of Countries with lesser competition

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

Name the chart/spreadsheet function you will use for solving the problem?

Method used : I have used Pivot table, in which rows are country and City and value field column is count of restaurant id and average of rating.

* Cities selected :

AUSTRALIA - Armidale , Balingup, Flaxton, Macedon, Penola

CANADA - Consort , Yorkton

SINGAPORE - Singapore

SRI LANKA - Colombo

* Location -

Excel file - sheet name - 4) New opening - state and cities

Table - 2) Cities in the suggested countries.

3. According to the countries you suggested, what is current quality in terms of ratings for restaurants that are opened there?

Will you use any aggregation function or a visualization here to solve the problem?

Method used : I have used aggregated function that is Averageif function to find the average rating of the suggested countries.

Formula used : =AVERAGEIF('Raw Data - Filtered'!$D$735:$D$9551,"Australia",'Raw Data - Filtered'!$T$735:$T$9551)

* Visualization method used : Waterfall chart

Location :

Excel file - sheet name - 4) New opening - state and cities

Table - 3) Suggested country with rating

4. Also what is the current expenditure on the food in the suggested countries, so that we can keep our financial expenditure in control?  
Mention the functionality which you will use for giving the suggestions, will it be any aggregated function or a visualization?

Method used : I have used aggregated function that is SUMIF function to find the total expenditure in the suggested country.

* Formula used : =SUMIF('Raw Data - Filtered'!$D$735:$D$9551,"Australia",'Raw Data - Filtered'!$Z$735:$Z$9551)

Current Expenditure On Food :

Australia - 32,281

Canada - 8,098

Singapore - 1,94,033

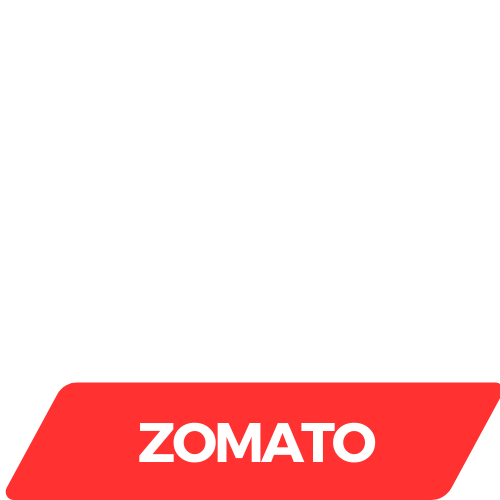
Sri Lanka - 11,875

* Visualization method used : Pie Chart

Location :

Excel file - sheet name - 4) New opening - state and cities

Table - 4) Total expenditure on food



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

How do you decide if anyone is a competitor?

Method used : I have used four pivot tables for all the four countries. In the row section Country and restaurant name is there, in the filter section city, and in the value field setting average of rating and average of cost of two in indian currency.

On the basis of rating and cost the competitors are evaluated.

* Restaurants who are biggest competitors : Are marked in green in the excel file. These biggest competitors as they have highest ratings.
* Restaurants which are rated low : This category has been sub divided into two. One which are marked in yellow are the restaurants which are having average ratings. Second category are marked in red are the restaurants which are in the lowest bracket with lowest ratings.
* Location -

Excel file - sheet name - 5) Competitor Analysis

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

What is the basis for the suggestions? And mention how did you decide if the cuisines affect the ratings?

CUISINES TO FOCUS - Pizza, Italian, Seafood, Mediterranean, Modern australian, Bakery , Desert , American, Chinese.

* The choice of cuisines - affect the ratings as every country as their own local food choices and preferences. For example if you are selling food in other country it would not sell as much as it would sell in australia.
* Basis for suggestion - The basis I have taken the ratings. The restaurants with cuisines which are local to that country are high in rating and other than this other food preferences which are famous in all countries are like seafood , italian.
* Decision - As by observing the pivot table, I analyze that some cuisines like seafood are having very high rating due to its popularity in that country. The decision is that cuisine affect the rating is based on the observation.These analyses provide a starting point for understanding trends and patterns in the data, allowing to make informed decisions and improvements in the restaurant business based on customer feedback and preferences.
* Location :

Excel file - sheet name - 6) Cuisines Analysis

7. According to our current data, should we go for online delivery and table booking? Does that affect the customer’s ratings? Mention your approach and spreadsheet function for the answer.

Method used : I have used COUNIFS function to calculate the table booking and online delivery for all the country.

Formula Used : =COUNTIFS('1) Raw Data - Filtered'!$D$735:$D$9551,"Australia",'1) Raw Data - Filtered'!$N$735:$N$9551,"Yes") - FOR TABLE BOOKING

=COUNTIFS('1) Raw Data - Filtered'!$D$735:$D$9551,"Australia",'1) Raw Data - Filtered'!$M$735:$M$9551,"Yes") - For ONLINE DELIVERY

Decision - We should go for table booking and online delivery. As none of the restaurants are providing this facility. This can be an edge over our competitors. But for doing this we can go for a survey to know that people in that country are willing to go for online delivery and table booking or not.

Cost vs. Services: Explore if there is a relationship between the cost for two and the availability of table booking or delivery services. Analyze whether customers are willing to pay more for the convenience of these services.

Geographical Trends: Consider exploring if the prevalence of table booking and delivery services varies by country or region.

Location -

Excel file - sheet name - 7) Delivery Analysis.

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

Distribution means the numbers of different price ranges, how will you show this using a chart?

* Method used : I have used Pivot table, in which rows are price range and value field is count of restaurant ids.
* Distribution of restaurant in different price range :

1-2 : 4444

2-3 : 3113

3-4 : 1408

4-5 : 586

* Visualization method : Histogram
* Location : Excel File - sheet named : 9) Price range

