**BY : Lokesh Sharma**

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

**Objective Questions**:

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

• 2 Tables are in the data.

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

• In table-1 20 attributes & In table-2 2 attributes

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

• 13 categorical columns in data.

• 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, used Text to column to extract the Year with this date format**

**• Handling Null Values in Cuisine Column:**

**Identified null values in the Cuisine column.**

**Press CTRL + G selected blanks and selected the upper value of the with the help of arrow button.**

**Location – RAW Data Sheet – Column name – Update Cuisines**

• Using the LookUp functions, fill up the countries in the original data using 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$16,2,0)

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

**Restaurant Count In Each country**

**Row Labels**

**Count of RestaurantID**

Australia

24

Brazil

60

Canada

4

India

8652

Indonesia

21

New Zealand

40

Philippines

22

Qatar

20

Singapore

20

South Africa

60

Sri Lanka

20

Turkey

34

United Arab Emirates

60

United Kingdom

80

United States of America

434

**Grand Total**

**9551**

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

**Restaurants Opened In Each Year**

**Row Labels**

**Count of RestaurantID**

2010

1080

2011

1098

2012

1022

2013

1061

2014

1051

2015

1024

2016

1027

2017

1086

2018

1102

**Grand Total**

**9551**

• What is the total number of restaurants in India in the 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('1. Raw Data'!$D:$D,"India",'1. Raw Data'!Q:Q,4

Location – Analysis Sheet – question no. 8 mention

**India has 388 number of restaurants in price range 4.**

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

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('1. Raw Data'!$D:$D,$B23,'1. Raw Data'!$S:$S)

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

**Average Voters in Each Country**

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

• 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**

**3.27381151**

**Formula – Using array formula {=AVERAGE(IF(('Raw Data'!$P:$P<4)\*('Raw Data'!$M:$M="Yes"),'Raw Data'!$T:$T))}**

**Location – Analysis sheet – objective no. 10**

• 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.

**That are highlighted in excel sheet name 4. New Opening states & cities**

• 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]

**I first compiled a list of current currency symbols manually. Following that, I utilized the VLOOKUP function in conjunction with the CONCATENATE function. This approach allowed me to dynamically associate each currency description with its respective symbol, resulting in a customized column with the desired concatenated information.**

**Formula - =** **=CONCATENATE(VLOOKUP($C2,'2. Country Description'!$A$1:$E$16,4,0),'1. Raw Data'!$T2)**

**Location – Raw data Sheet – Column Name Customized Price Column**

• 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?

**=COUNTIFS('1. Raw Data'!O:O,"No",'1. Raw Data'!R:R,1,'1. Raw Data'!T:T,"<=250",'1. Raw Data'!M:M,"Indian Rupees(Rs.)")**

**Loction – Analysis sheet – Objective Q no. 13**

**Subjective Question:**

• 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?

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 - Suggested Country with ratings

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

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 - Cities in the suggested countries.

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

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

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

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

• 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.

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

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

Method used : I have used CORREL Function to find the correlation between rate of cuisines and ratings

• Sign (-): The negative sign indicates a slight negative correlation. This means that as restaurant ratings increase, there is a very small tendency for the cost for two to decrease slightly, and vice versa. However, the correlation is so close to zero that it's essentially negligible.

• Magnitude (Absolute Value): The absolute value (ignoring the negative sign) is very close to zero (0.0057), indicating an extremely weak correlation. In practical terms, the correlation is so minimal that it's unlikely to be practically significant or meaningful

• Decision: We can keep the rate of cuisines higher, as the correlation is very negligible between rating and rate of cuisines.

Location **:** Excel File - sheet named - 8) Correlation

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

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-: 4444

2-: 3113

3-: 1408

4-: 586

• Visualization method : Histogram

• Location : Excel File - sheet named : 9) Price range

• Explain your approach in brief for suggesting countries/cities in order to open new restaurants, if the objective and subjective questions would have been given to assist you. **[you have to give bullet pointers in order to answer this question]**

**Countries Suggested (Australia, Canada, Singapore, Sri Lanka):**

**Analytical Criteria:**

**• Utilized Pivot tables with restaurant count and average ratings.**

**• Applied filters for low competition and ratings below 4.**

**Strategic Insights:**

**• Identified regions with both low competition and potential for improvement in average ratings.**

**• Focused on countries where market entry could yield substantial benefits.**

**City Selection:**

**Analytical Criteria:**

**• Employed similar Pivot table approach on city-level data.**

**• Focused on cities within suggested countries with low competition and ratings less than 4.**

**Strategic Insights:**

**• Chose cities aligning with the overall country criteria.**

**• Aimed for a balanced selection of cities within the recommended countries based on data analysis.**

**The dashboard must consist of Year-wise and country slicers.**