

BY:ANIL R

**Problem Statement**

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

OBJECTIVE QUESTIONS

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

**Steps used for cleaning data:**

* **BORDER: given data did not have any borders, I added the border so the data looks decent.**
* **HIDING COLUMNS: columns F and H will be hidden because there are other columns as well which were providing information about the city and state. Longitude and latitude columns are not useful for preparing reports so columns I and j are hidden.**
* **WRAP TEXT: all the text in the given data is wrapped so that text will not overlap again.**
* **Formatting date column: Datekey\_opening column was not having a proper date format,all the “ \_ ” was replaced by the “/” so that the Dtaekey\_opening column data will look in proper format.**
* **COST IN INR: This column is added to the given data which gives the average of two cost column currency in INR.I have created currency exchange rate table which is used while converting other currencies into INR.**

**formula used is:** **=S2\*'2.country description'!$E$4.**

* **Filled missing data’s: A few cuisines that belong to USA restaurants were missing, these cells are filled by using a mode of cuisines in the USA(pivot table used) and a few average cost-for-two column cells were missing which belong to the countries India and USA those cells are filled with the average of the cost for two cells (pivot table used).**

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

**Function used for fetching country using country code**

**Vlookup function**: **Country code is the common value found in the source data and the country description sheet. vlookup is the lookup function I have used to extract country name.**

**Formula:** =VLOOKUP(C2,'2.country description'!$A$2:$B$16,2,0)

**C2=lookup value that is country code.**

**'2.country description'!$A$2:$B$16 : range that is used to search country code.**

**2:2 is the column number where country will be searched.**

**0:0 for the exact match.**

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

**The pivot table is used to count the number of restaurants opened in each country where the country is taken in the row field, the restaurant-id is taken in the value field, and value field is summarized using the count function**

|  |  |
| --- | --- |
| NUMBER OF RESTAURANTS OPENED IN EACH COUNTRY | |
| **Country** | **count of Restaurant** |
| 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** |

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

**Pivot table is the function used to find the number of restaurants opened in each year. Where the year field is taken in the row and the restaurant-id is taken in the value and the value field is summarized by the count function**

|  |  |
| --- | --- |
| RESTAURENT OPENEND IN EVER YEAR | |
| **YEAR** | **Count of Restaurant** |
| 2010 | 1080 |
| 2011 | 1098 |
| 2012 | 1022 |
| 2013 | 1061 |
| 2014 | 1051 |
| 2015 | 1024 |
| 2016 | 1027 |
| 2017 | 1086 |
| 2018 | 1102 |
| **Grand Total** | **9551** |

**5 What is the total number of restaurants in India which are in the price range 4?**

**The function used to count total restaurants in India in the price range of 4 :**

**COUNTIFS: As per the question there are two conditions, so I have used the countifs formula to extract the number of restaurants in the price range of 4 in India.**

**Formula used:** **=COUNTIFS(source!$D$2:$D$9552,source!$D$2,source!$Q$2:$Q$9552,4)**

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

**Function used to get the average number of voters for the restaurants in each country**

**AVERAGEIF: As per the question average number of voters for restaurants in each country, I have used Averageif function.**

* **Formula used :**

**=AVERAGEIF(source!$D$2:D9552,'4.analysis'!A29,source!$R$2:$R$9552)**

**source!$D$2:D9552: It is the criteria range.**

**'4.analysis'!A29: It is the criteria.**

**'4.analysis'!A29,source!$R$2:$R$9552: Range of voters for to get average.**

Subjective Questions

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 utilized a pivot table where the rows represent countries, and the value field includes the count of restaurant IDs and the average rating. After creating the pivot table, I applied the sort function (A-Z) on the count of restaurant IDs in the value field. Through this, I identified countries with fewer restaurants. Subsequently, I selected countries with lower ratings, specifically in the range of 3-4. Following this process, I identified countries with less competition.**
* **Suggested Countries Are: AUSTRALIA, CANADA, SINGAPORE, SRI LANKA**
* **Visualization method used: column chart**
* **Location-**

**Excel file-sheet name-5.new openings**

**Table-1 Restaurant with minimal competition.**

1. **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 utilized a pivot table in which the rows represent the country and city, and the value field column includes the count of restaurant IDs and the average rating.**
* **Cities of the suggested countries:**

**AUSTRALIA: Aramidale,Balingup,Flaxton,Forrest,Lorn,Macedon,Penola,Lakes entrance.**

**CANADA: Consort,Yorkton.**

**SINGAPORE :Singapore.**

**SRI LANKA: Colombo.**

* **Location-**

**Excel file-sheet name-5.new openings**

**Table-2 Cities in the suggested Countries**

1. **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 utilized an aggregated function, specifically the AVERAGEIF function, to determine the average rating for restaurants of the suggested**

**countries.**

* **Visualization method used:3-D pie-chart.**
* **Formula used: =AVERAGEIF(source!$D$2:$D$9552,I3,source!$T$2:$T$9552)**

**source!$D$2:$D$9552: it is the criteria range.**

**I3: criteria.**

**source!$T$2:$T$9552: range of rating to get average.**

* **Location:**

**Excel file-sheet name-5.new openings**

**Table-3 Restaurant Average rating of suggested Countries.**

1. **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 employed a pivot table, organizing countries in the column and specifying the cost for two in the value field. The summation function (SUM) was applied to aggregate values in the specified field. Additionally, a filter was implemented to narrow down the dataset to include only the suggested countries.**

|  |  |
| --- | --- |
| **COUNTRY** | **Sum of COST IN INR** |
| **Australia** | **27,254.80** |
| **Canada** | **8,997.25** |
| **Singapore** | **1,65,592.35** |
| **Sri Lanka** | **11,440.00** |

* **Visualization method used: Clustered column.**
* **Location:**

**Excel file-sheet name-5.new openings**

**Table-4 Expenditure on food for the suggested countries.**

1. **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 employed four different pivot tables for the suggested country. In the row field, I included both the country and the restaurant name. In the value field, I calculated the average rating and average cost in INR. Based on these average ratings and costs, competitors are Identified.**
* **Restaurants That Are Biggest Competitors: Restaurants that provide significant competition are highlighted in cream color. These restaurants are considered the biggest competitors based on average ratings.**
* **Restaurants with Low Ratings: It is divided into two categories. The restaurants with average ratings are marked with white color, while the other restaurants with low ratings are marked with blue color.**
* **Location:**

**Excel file-sheet name-6.competitor analysis**

1. **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 on: Pizza, Bar Food,** **Mediterranean, Seafood,** **Modern Australian, Australian,** **Café,** **Italian,** **Bakery,** **American, Steak,** **Juices, Desserts,** **Sri Lankan.**
* **Choice of Cuisines: The choice of cuisines certainly affects restaurant ratings because it is important to select cuisines that are local to specific countries. For instance, highly-rated cuisines in Canada may not be suitable for Sri Lankan restaurants, as the preferences of Sri Lankan people differ, and they may prefer different cuisines.**
* **Basis for suggestion:** **The basis I have used is the rating. Restaurants featuring cuisines that are local to the country receive higher ratings, while other cuisines chosen are those that are widely renowned across countries, such as café, Italian, and seafood.**
* **Decision:** **After analyzing the pivot table, I understood that cuisines local to specific countries receive higher ratings, whereas cuisines not familiar to the countries are rated lower. Observing this, I have concluded that cuisines directly impact the ratings**.
* **Location:**

**Excel file-sheet name-7.Analysis of cuisines**

1. **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: I had used Countifs function to calculate table booking and online delivery for the suggested countries**
* **Formula: =COUNTIFS(source!$D$2:$D$9552,"Australia",source!$N$2:$N$9552,"Yes") -for online delivery**

**=COUNTIFS(source!$D$2:$D$9552,"Australia",source!$M$2:$M$9552,"yes")-for table booking**

* **Decision: According to our records, there are currently no suggested countries offering both online delivery and table booking services. As we enter the market, we can promote ourselves by offering these services with a minimal convenience fee.**
* **This is expected to positively impact the ratings, as customers will appreciate having options such as online delivery and table booking available.**
* **Location:**

**Excel file-sheet name-8.Delivery and booking**

1. **Should the team keep the rate of cuisines higher? Will that affect the feedback? According to our data are the rate of cuisines and ratings, correlated? Back the suggestion with proper insights and visualization.**

* **Method used: I had used CORREL function to find the correlation between the ratings and cost of cuisines**

* **Formula used:** **=CORREL($G$2:$G$69,$H$2:$H$69)**
* **After conducting a correlation analysis, I obtained a negative correlation value. This indicates that when ratings increase, the cost of cuisines tends to decrease, and conversely, when ratings decrease, the cost of cuisines tends to increase. The relationship works in both directions.**
* **It is a weak correlation which is closer to 0. The correlation is so minimal that is unlikely to be practically meaningfull, so we can neglect the correlation between them.**
* **Decision:** **We can increase the price of the cuisine, as there is a negligible correlation between the ratings and the cost of cuisines.**
* **Visualization method used: Scatter chart**
* **Location:**

**Excel file-sheet name-9.correlation**

1. **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 had used pivot table, price range is used in row field and count of restaurant-id is used in value field**
* **Distribution of resaturent based on price range:**

|  |  |
| --- | --- |
| **Price Range** | **Count of RestaurantID** |
| **1** | **4444** |
| **2** | **3113** |
| **3** | **1408** |
| **4** | **586** |

* **Visualization method used:3-D Clustered Column Chart.**
* **Location:**

**Excel file-sheet name-10.Price range**