

BY:MALLLESH H

**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 QUETIONS**

**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, This was done to enhance overall appearance and ensure more polished look for data.**
* **HIDING COLUMNS: columns F and H were 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.**
* **WRAPTEXT: all the text in the given data is wrapped so that prevents the potential overlap of data and more organized manner of information.**
* **Formatting date column: Date\_key\_opening column was not having a proper date format, all the “\_ “were 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 curriences 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 United States of America (pivot table used) and a few average cost-for-two column cells were missing which belong to the countries India and United States of America 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**

|  |  |
| --- | --- |
| NO OF RESTAURANTS OPENED IN ECAH COUNTRY | |
| **Country** | **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** |

**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 value field is summarized by the count function**

***RESTAURANTS OPENED IN EACH YEAR***

|  |  |
| --- | --- |
| **Years** | **Count of RestaurantID** |
| 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(‘Original data’!$D$2:$D$9552, ‘Original data’!$D$2, ‘Original data’!$Q$2:$Q$9552,4)

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

**Method: I created pivot table, including countries in the row field and voters in value field.**

|  |  |
| --- | --- |
| **COUNTRY** | **Average of Votes** |
| Australia | 111.4 |
|  |  |
| Brazil | 19.6 |
| Canada | 103.0 |
| India | 137.2 |
| Indonesia | 772.1 |
| New Zealand | 243.0 |
| Philippines | 407.4 |
| Qatar | 163.8 |
| Singapore | 31.9 |
| South Africa | 315.2 |
| Sri Lanka | 146.5 |
| Turkey | 431.5 |
| United Arab Emirates | 493.5 |
| United Kingdom | 205.5 |
| United States of America | 428.2 |
| **Grand Total** | **156.9097477** |

**SUBJECTIVE QUESTIONS**

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?**

* **Suggested countries are**🡪 **Australia, Canada, Singapore, Srilanka**
* **Method used🡪 I utilized a pivot table where the rows represent the countries, and value field represntd the count of Restaurant IDs and Average rating. After creating pivot table, I applied a sort function (bottom 4) on count of restaurant IDs in the value field. Through this, I identified the countries with fewer restaurants.**
* **Visualization method🡪 Column chart**
* **Location: Excel fie🡪new openings🡪Table1(Restaurants with minimum compition).**

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?**

* **Cities of the suggested countries**

**AUSTRALIA🡪Aramidale,Balingup,Flaxton,Macedon,Mayfield,Montville,Paynsville,Penola.**

**CANADA🡪 Consort,Yorkton.**

**SINGAPORE 🡪Singapore.**

**SRI LANKA🡪Colombo.**

* **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.**
* **Location🡪Excel file🡪new openings🡪Table2 (cities in 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 created pivot table, including countries and ratins in row and value field respectively**
* **Visualization method used:3-D pie-chart.**

|  |  |
| --- | --- |
| **Country** | **Average of Rating** |
| Australia | 3.658333333 |
| Canada | 3.575 |
| Singapore | 3.575 |
| Sri Lanka | 3.87 |

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

|  |  |
| --- | --- |
| EXPENDITURE IN EACH COUNTRY | |
| **Country** | **Sum of Average\_Cost\_for\_two\_in\_INR** |
| Australia | 32281.3 |
| Canada | 8992.9 |
| Singapore | 194220.25 |
| Sri Lanka | 12350 |
| **Grand Total** | **247844.45** |

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

**Excel file🡪sheet name🡪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🡪 I created individual pivot tables for the suggested countries. In the row field, I included country and restaurant name. average rating and average cost of products are included in the value field. Based on the averahe rating and the average cost of products the compitators are identified.**
* **Restuarants that are biggest compitators 🡪 The restaurants having more than 4 rating are marked with brown color(excel sheet analysing compitators). The restauramts having more than 3.5 rating with less than 500 of average cost.**
* **Restaurants with less Ratings🡪 The restaurants with rating range 2-3 are colored with yellow color**
* **Location🡪 Excel file🡪compitataor analysis🡪Table no 5 to 8**

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?**

* **Suggested cuisines:**
* **Australia🡪Pizza, Bar food, Mediterranean, Sea food, Modern Australian, Australian.**
* **Canada🡪 Italian, Mediterranean, Pizza.**
* **Singapore🡪American, Bakery, Italian, Steak.**
* **Srilanka🡪 Deserts, Juices, Sea food.**
* **Choice of Cuisines: Choice of cuisines definitely affect the rating of a restaurant, so it is important to select the cuisines that are local to specific countries. For example, highly rated cuisines in Singaopore may not be suitable for the Australian resturants, Australian people preffers deferent cuisines.**
* **Basis for the suggetions: I created pivot table, included country and cuisines name in row, average rating in value field. After observing pivot table I identified that the average rating of some cuisine is varies in different countries. For example rating of sea food cuisine is 4.9 and 3 in Srilanka and Singapore respectively.**
* **Decision: After analyzing pivot table, the choice of cuisine is directly affect on the ratiings.**
* **Location: Excel file🡪cuisine analysis🡪Table 9**

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.**

* **Decision:Given our findings that none of the suggested countries currently provide both online delivry and table booking services,we plan to differentiate ourselves in the market by introducing these offerings, accompained by a nominal convince fee to enhance customer satisfaction and convinience**
* **Method: I used countifs function to calculate to check the number restaurants provide online delivery and table bokings**

**Formula**

**COUNTIFS(‘Original data’!$M$2:$M$9552,”country name”, Original data’!$M$2:$M$9552,”yes”)🡪 online delivery**

**COUNTIFS(‘Original data’!$N$2:$N$9552,”country name”, Original data’!$N$2:$N$9552,”yes”)🡪 table booking**

* **Location: excel file🡪booking and delivery**

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.**

* **Decision:** **Following a correlation analysis, it was revealed that there exists a negative correlation. This implies that as ratings rise, there is a tendency for cuisine costs to decline, and conversely, as ratings decrease, the costs of cuisines tend to rise. This bidirectional relationship underscores the dynamic interplay between customer satisfaction and pricing in the culinary offerings.**
* **Method: I Pivot table to find correlation between rate of cuisines and ratings.**
* **Visualization method: Scatter chart**
* **As there is a negligble correlation between rate of cuisines and rating,we can increase the rate of cuisines**

**9. 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?**

|  |  |
| --- | --- |
| **Row Labels** | **Count of RestaurantID** |
| 1 | 4444 |
| 2 | 3113 |
| 3 | 1408 |
| 4 | 586 |

**Method: I created pivot table, included Price range in rows and count of restaurant id in value field.**

**Visualization method: Column chart**