# Zomato Restaurants Analysis

# --------Ekta

# Objective Questions

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

### The data set given had some inconsistencies which were corrected as following.

* **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).**

## 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**

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

The below table represents the number of restaurants opened in each country.

|  |  |
| --- | --- |
| Countries | 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** |

The above data has been visualised in the Column chart below.

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

From the analysis it was understood that highest number of restaurants were opened in 2018, where 1102 restaurants where opened. Similarly, the lowest number of restaurants were opened in 2012, where 1022 restaurants where opened.

There is no visible trend in the number of restaurants opened each year.

The table given below gives information about the number of restaurants opened in each year.

|  |  |
| --- | --- |
| Opening Year | Count of RestaurantID |
| 2010 | 1080 |
| 2011 | 1098 |
| 2012 | 1022 |
| 2013 | 1061 |
| 2014 | 1051 |
| 2015 | 1024 |
| 2016 | 1027 |
| 2017 | 1086 |
| 2018 | 1102 |
| Grand Total | **9551** |

The above data has been visualised using a Column chart as shown below.

## What is the total number of restaurants in India in the price range of 4?

There is a total number of 388 restaurants opened in India in the price range of 4

The Chart below shows details when country slicer is applied for INDIA

## According to the data, what is the average number of voters for the restaurants in each country?

The table given below shows the average number of voters for each country.

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

The above data has been visualised as given below.

# Subjective Questions

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

**Suggested countries are**🡪 **Australia, Canada, Singapore, Srilanka**

**Method used**🡪

* **I utilized a pivot table where the rows represent the countries, and value field**

**represented the count of Restaurant IDs and Average rating.**

* **After creating pivot table, I applied a sort function that sorts CountryName by Count**

**of RestaurantID in descending order. Through this, I identified the countries with fewer restaurants and their ratings.**

* **On the basis of this visualization we can say that the countries having less number of restaurants will be having lesser competition as new restaurants can be opened in those**

**.**

**Visualization method🡪 Combination Chart**

**Location: Excel fie**🡪**Sheet 1**🡪Countries with lesser competition

**The chart below shows the number of restaurants in each country which shows AUSTRALIA, CANADA, SINGAPORE, SRI LANKA have the less number of restaurants**

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

* **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 : Armidale ,Balingup ,Flaxton ,Forrest ,Lorn ,Macedon ,Penola ,Lakes entrance**

**CANADA : Consort ,Yorkton**

**SINGAPORE : Singapore**

**SRI LANKA : Colombo**

**The table below shows the number of Cities in the suggested Countries :**

|  |  |  |
| --- | --- | --- |
| Row Labels | Count of RestaurantID | Average of Rating |
| Australia | **24** | **3.658333333** |
| Armidale | 1 | 3.5 |
| Balingup | 1 | 3.2 |
| Beechworth | 1 | 4.6 |
| Dicky Beach | 1 | 3.6 |
| East Ballina | 1 | 4.1 |
| Flaxton | 1 | 3.5 |
| Forrest | 1 | 3.7 |
| Hepburn Springs | 2 | 3.8 |
| Huskisson | 1 | 4.1 |
| Inverloch | 1 | 3.7 |
| Lakes Entrance | 1 | 3.8 |
| Lorn | 1 | 3.6 |
| Macedon | 1 | 3.5 |
| Mayfield | 1 | 2.9 |
| Middleton Beach | 1 | 3.8 |
| Montville | 1 | 2.4 |
| Palm Cove | 1 | 4.4 |
| Paynesville | 1 | 2.6 |
| Penola | 1 | 3.4 |
| Phillip Island | 1 | 3.7 |
| Tanunda | 1 | 4.4 |
| Trentham East | 1 | 4.1 |
| Victor Harbor | 1 | 3.6 |
| Canada | **4** | **3.575** |
| Chatham-Kent | 1 | 3.7 |
| Consort | 1 | 3 |
| Vineland Station | 1 | 4.3 |
| Yorkton | 1 | 3.3 |
| Singapore | **20** | **3.575** |
| Singapore | 20 | 3.575 |
| Sri Lanka | **20** | **3.87** |
| Colombo | 20 | 3.87 |
|  |  |  |

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

* **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(RawData!$D$2:$D$9552,I3, RawData!$T$2:$T$9552)**

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

**I3: criteria.**

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

|  |  |
| --- | --- |
| Row Labels | Average of Rating |
| Australia | 3.66 |
| Canada | 3.58 |
| Singapore | 3.58 |
| Sri Lanka | 3.87 |

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

* **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.**

|  |  |
| --- | --- |
| Countries | Sum of Average\_Cost\_for\_two\_INR |
| Australia | 578 |
| Canada | 145 |
| Singapore | 3115 |
| Sri Lanka | 47500 |
| Grand Total | **51338** |

**Visualization method used: Clustered column**

## 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 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.**

## 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 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**.

## According to our current data, should we go for online delivery

and table booking? Does that affect the customer’s ratings?

* **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.**

## 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 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**

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

The distribution of number of restaurants of different price range can be observed in the following chart.

It is understood that the number of restaurants is highest in the price range of ‘1’ which is 4444 restaurants and the lowest is in the price range of ‘4’ which is 586 restaurants.

|  |  |
| --- | --- |
| Ratings | 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: Clustered Bar**