**Zomato Case Study – Zomato Restaurant Analysis**

-By Farseen Ahammed

**Objective Questions**:

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

There are 2 tables present in the data provided.

* *Raw data* sheet has 1 table with all the information about different restaurants all across the globe.
* *Country description* sheet has a helper table to refer the country code for the respective country from the *Raw Data* sheet

A screen shot of a computer

Description automatically generated

*Fig 1.1.1 Raw data sheet*

A screenshot of a computer

Description automatically generated

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

Total no. of attributes present in Raw Data + Total no. of attributes present in country description table

=COUNTA(Raw\_Data[#Headers]) + COUNTA('country description'!A1:B1) , i.e; 20 + 2 = 22 attributes.

A screenshot of a computer

Description automatically generated

1. **How many categorical columns are there in the data?**

* Columns which can be used to divide data into categories are called categorical columns. The following are the categorical columns in the data given:
  + Restaurant ID
  + Restaurant Name
  + CountryCode
  + City
  + Address
  + Locality
  + Locality Verbose
  + Cuisines
  + Currency
  + Has\_Table\_booking
  + Has\_Online\_delivery
  + Is\_delivering\_now
  + Switch\_to\_order\_menu
  + Price\_range
  + Country Code
  + Country Name
  + Datekey\_opening
* Columns which give us measurements for the data are called continuous columns. These are almost all the time numerical. The continuous columns in the given data are:
  + Longitude
  + Latitude
  + Votes
  + Average\_Cost\_for\_two
  + Rating

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

***Handling Missing Values***

* + Create a new sheet called Cleaned Data, for cleaning Raw Data.
  + Select the Data that I needed to clean and go to the ‘Data’ tab and click ‘Filter’.
  + Check if there are any blanks, using ‘Ctrl + F’ or ‘Find & Select’ or using Filter dropdown.
  + There were blank cells on cuisine column.
  + So I have decided to delete entire row who is having blank cells.
  + Press ‘Ctrl + G’ to open the ‘Go To’ dialog box
  + Click on the ‘Special’ Button at the bottom.
  + Deleting Rows with Blank Cells in Excel using ‘Go To Special’ dialog box,.
  + Select ‘Blanks’ and click ‘Ok’ and Delete the rows by right clicking on the blank cells

and click delete ‘Table Rows’.

A screenshot of a computer

Description automatically generated

* We have also removed restaurants that have 0 *votes* since we actually want to launch restaurants by studying restaurants that have received feedback from customers.
* Also, *average\_cost\_for\_two* cannot ever be 0, this seems to be data entry error and in-order to clean it, we have removed those rows.

***Remove Duplicates***

* After cleaning cuisine columns, I selected all columns and go to the “Data” tab , click “Remove Duplicates”, and ensure all columns are checked.

A screenshot of a computer error

Description automatically generated

***Standardize Data Formats***

* Select “Datekey\_Opening” column and format it as a date.
* I have inserted a new 2 columns called “Date\_opening” to store the date format. And “Year” column for storing the year for future analysis purpose.
* In “Date\_opening” column I used Date() and string() to convert to date format.
* Since I don’t want Date column twice, I have decided to copy the “Date\_opening” column and use “Paste special” , select “Values” to “Datekey\_opening”. And Delete “Date\_opening” column.
* And to get “Year” out of “Datekey\_opening” I used “=Year()” function.
* Finally I have sorted the cleaned data on the basis of “RestaurantID”.

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

Inorder to fillup the countries I have used VLOOKUP().

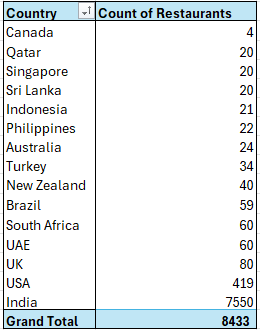
* First I have created a new column called “Country” .
* And fill the names of the countries using VLOOKUP().
* From sheet “country description”.

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

A screenshot of a computer

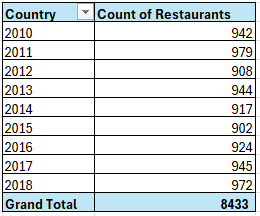
Description automatically generated

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

****

* + We can see that majority of our data comes from India which is around 90% of restaurants in our data.
  + After India the biggest country is USA followed by UK in third position.

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



We can see from above table:

* 2015 was the year with the least number of restaurants opened.
* 2017 was the year with the most number of restaurants opened.

**8. What is the total number of restaurants in India in the price range of 4?**

To find the total number of restaurants in India in price range of 4 I have used COUNTIFS() function.

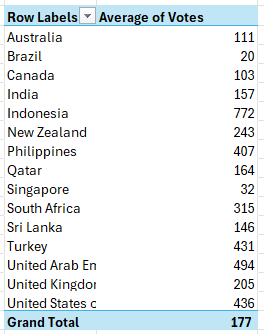


The total no. of restaurants in the price range of 4 is **383**.

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

Firstly, Create a pivot table by selecting the Cleaned Data.

* Drag the ‘Country’ column to Rows and ‘Votes’ to Values make the votes into average field using Value Field Settings.
* We can see that Indonesia has the highest number of voters followed by United Arab Emirates and United States of America.



* Brazil on the other hand has the least amount of voters after Singapore and Canada.

1. **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.**

* We can use the following array formula to get the average of rating for restaurants with price range < 4 and online delivery = “Yes”.

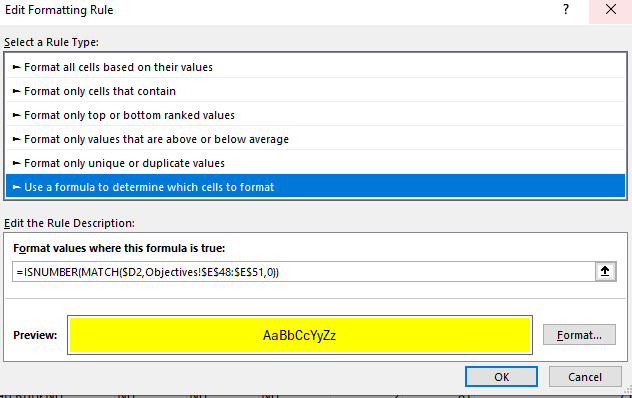
**A screen shot of a computer

Description automatically generated**

* The average rating for all restaurants with price range < 4 and online delivery is **3.30** after rounding the answer to 2 decimal digits.

1. **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.**

* My approach was, first analyse the data of country vs count of restaurants, from where the competition is less by checking less number of restaurants in countries that are Canada, Qatar, Singapore, Sri Lanka.
* Created a new formula to reference suggested countries from another table and use that reference to highlight rows with those countries in the Clean Data table

****

1. **Create a new customized price column that consists of the abbreviation/symbol of the currency along with the Average\_cost\_for\_two value.**

To do this ;

* + Create a column named ‘Customized\_price’.
  + In ‘Customized\_price’ column first try to find the position ‘(‘ from “Currency “ column and moves 1 character to the right :

=FIND(“(“ , $L2)+1

* + Find the position of closing parenthesis:

=FIND(“)”, $L2)

* + Now by using MID() extract the text between the parenthesis:

=MID($L2,FIND("(",$L2)+1,FIND(")",$L2)-FIND("(",$L2)-1)

* + Finally use ‘&’ to concatenate the symbol to Average\_cost\_for\_two:

=MID($L2,FIND("(",$L2)+1,FIND(")",$L2)-FIND("(",$L2)-1)& " " &$S2

1. **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?**

|  |  |
| --- | --- |
| **Country Name** | **INR Exchange Rate** |
| India | 1 |
| Australia | 56.18 |
| Brazil | 15.43 |
| Canada | 61.03 |
| Indonesia | 0.0052 |
| New Zealand | 50.48 |
| Philippines | 1.43 |
| Qatar | 22.91 |
| Singapore | 62.11 |
| South Africa | 4.62 |
| Sri Lanka | 0.28 |
| Turkey | 2.53 |
| United Arab Emirates | 22.75 |
| United Kingdom | 108.25 |
| United States of America | 83.56 |

* First, we will convert all the average values from their native currencies to INR.
* For this we will use combination of VLOOKUP() and mathematical operator “/”.
* We have to get a new table which will have the exchange rate for all currencies with INR.



* Now we have a new column “*Average\_cost\_for\_two\_INR*” with cost for two converted to INR.
* Now create a new column named Array\_Formula in Clean Data.
* In Z2 cell use the array formula to count the number of restaurants listed that do not offer online delivery, are in the lowest price, and have an average cost for two people less than or equal to 250 Indian Rupees.

****

* Then use Sum() to sum Array\_formula column to get the number restaurants listed.



From the above formula we can see that there are **1210** restaurants that has no online food delivery service, fall under the lowest price range of 1 and average cost for two people is at most 250.

**Subjective Question:**

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?**

**Strategy:**

* If you have the number of restaurants per country or city, you can focus on countries with lower restaurant density, even without external demographic data. Countries with fewer restaurants would be potential targets for expansion.
* Use clustering techniques to identify geographic areas that have fewer restaurants in proximity. This spatial analysis can reveal underserved regions within a country. Areas with fewer clusters of restaurants would show up as potential locations.
* Tourists’ locations are one of the great options to open a new restaurant.
* Since there is no population column is not given, I have chosen the lesser number of restaurants in a country.

**Factors:**

* Countries with fewer restaurants suggest less competition.
* A lack of online delivery in certain countries could indicate an untapped opportunity for introducing modern restaurant models.
* Restaurants with higher votes and fewer restaurants typically have high engagement from the public, indicating weaker competition or newer restaurants. Countries with high average votes and fewer restaurants suggest less competition.

**Technique:**

* + By using Pivot Table, we can summarize the data by country and count restaurants in each country, to know which country is suitable for opening new restaurants according to our strategy.

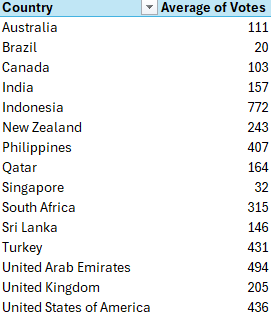
A screenshot of a computer

Description automatically generated

A blue bar chart with white text

Description automatically generated

* + We can observe that **India** has the highest number of restaurants indicating a highly saturated market.
  + We can see that **Canada**, **Qatar**, **Singapore** and **Sri** **Lanka** fits our strategy the best.
  + By using one more Pivot Table to confirm our suitable locations to open a new restaurant by checking the average votes by country.



A graph with numbers and lines

Description automatically generated

* We can observe that ;

**Countries with fewer restaurants and higher votes**:

* **Canada**: Only 4 restaurants with an average of 103 votes.
* **Qatar**: 20 restaurants and 164 votes.
* **Sri Lanka**: 20 restaurants with 146 votes.
* **Singapore:** 20 restaurants with 32 votes. (Exceptional case, Singapore is a beautiful country and have a beautiful city where tourists love to visit, One of the Worlds beautiful airport is situated in Singapore. Since there are fewer restaurants compared to other countries this is a best choice)

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

**Strategy:**

* City which has less restaurants and high average cost for 2 people are the best choices for new restaurants.

**Technique:**

* We can use Pivot Table to find solution for this :

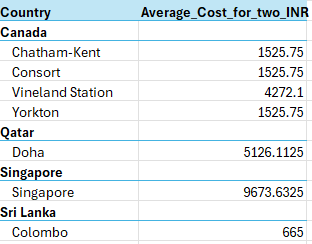
A screenshot of a table

Description automatically generated

**Visualization:**

A graph of restaurants in suggestive

Description automatically generated



**Visualization:**

A blue graph with white text

Description automatically generated

**Suggestions:**

* + We can see according to our data, **Vineland Station** is the best fit city in **Canada**, meanwhile **Doha** and **Colombo** also are good fits for **Qatar** and **Sri** **Lanka** respectively.
  + Singapore itself is a city-state which we can see already satisfy our requirements.
  + We have converted all the currencies into INR for better understanding and simplicity.

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

* To find the current quality regarding ratings for restaurants that are opened in suggested countries is by finding average rating of each country – city.

A screenshot of a computer

Description automatically generated

A graph of a number of countries/regions

Description automatically generated

* + We can see from the above chart that the average rating varies from **3.58** to **4.06** on a scale of **5.0**.
  + Qatar has the highest average rating of 4.06 followed by Sri Lanka with **3.87**.
  + Singapore and Canada have the lowest rating out of all the suggested countries with an average rating of **3.58**.

**Suggestions:**

* + By checking these insights I would say these are all above average ratings , so we should learn few and implement similar way so that we can also get similar results for our new restaurant.

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

A screenshot of a computer

Description automatically generated

A graph with numbers and a bar

Description automatically generated

* + We can see that the average cost for two is low in Sri Lanka , it is just around **₹665**.
  + Singapore leads with **₹** **9**,**674** followed by Qatar with **₹** **5126** and Canada with **₹2212.**

**Suggestions:**

* + By checking these insights I would say Singapore, Qatar and Canada are developed Countries compared to Sri Lanka, so the average cost for two differs from place to place, so we should implement prices of the food accordingly to satisfy the customers.

1. **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.**

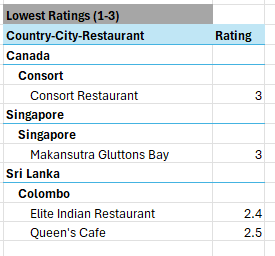
**Approach:**

* We should aim for rating of 3.1-3.5 for our new restaurants.
* This will also help us in attracting wide range of customers and will help in lowering our expenditure.
* This strategy is relatively low risk.
* Our biggest competitors are as follows.

A screenshot of a computer

Description automatically generated

* + Our biggest competitors will be rating of 3.1-3.5 for our new restaurants.
  + Above 3.5 rating most of them having high expenditures.
  + This will also help us in attracting wide range of customers and will help in lowering our expenditure.
  + This strategy is relatively low risk.
  + Our biggest competitors are as follows:
* Restaurants with lowest ratings (1- 3) are follows:

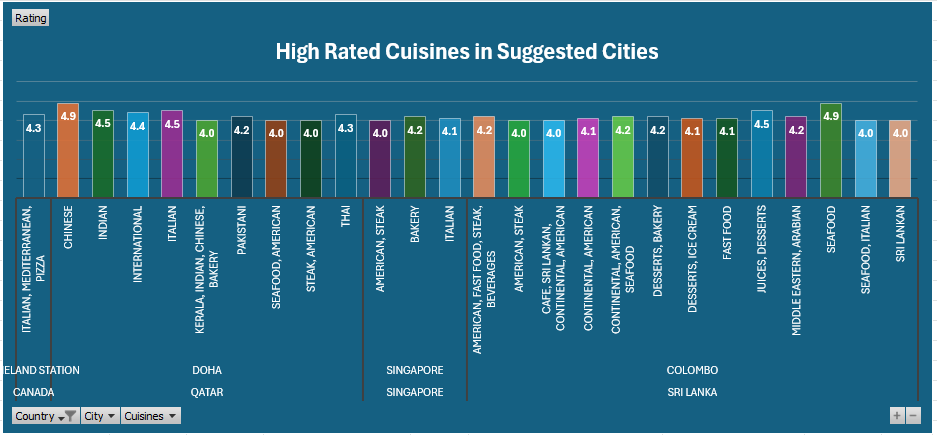


* + We can see that Vineland Station does not have a restaurant with mid-rating (3.1-3.5)
  + Also, we can observe that even though Singapore has high number of mid-rating restaurant meanwhile Qatar and Colombo only has 2.
  + Low rated restaurants (1.0-3.0) are only present in Colombo and Singapore out of all the cities suggested.
  + We can observe that in Vineland station, Colombo and Doha there is a very low saturated market to capture for mid rating restaurant and in Singapore we can aim for a restaurant with marginally high ratings.

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

* In order to find which cuisines should we focus in the newer restaurants to get better feedback , we should find the cuisines of suggested cities having higher ratings using Pivot Table. A screenshot of a computer

  Description automatically generated



**Observations & suggestions:**

* + In the suggested cities we can see that Colombo has restaurants which are highly rated commonly serve Seafood, Juices, Deserts, Bakery and American.
  + For Doha, Chinese, Italian, Thai and Indian are highly rated cuisines.
  + Singapore has Bakery, Italian and American which are served in high-rated restaurants.
  + There is only data for 1 restaurant in Vineland Station which serves Italian/Mediterranean.
  + In conclusion, we should focus more on cuisines which are popular regionally.

1. **According to our current data, should we go for online delivery and table booking? Does that affect the customer’s ratings?**

|  |  |  |
| --- | --- | --- |
| Has table booking | Ratings | Count of Restaurants |
| No |  |  |
| Colombo | 3.9 | 20 |
| Doha | 4.0 | 19 |
| Singapore | 3.6 | 20 |
| Vineland Station | 4.3 | 1 |
| Yes |  |  |
| Doha | 4.7 | 1 |

|  |  |  |
| --- | --- | --- |
| Has online delivery | Ratings | Count of Restaurants |
| No |  |  |
| Colombo | 3.9 | 20 |
| Doha | 4.1 | 20 |
| Singapore | 3.6 | 20 |
| Vineland Station | 4.3 | 1 |

**Observations And Suggestions:**

* None of the restaurants from the suggested countries provide online food delivery
* Only 1 restaurant provides table booking.
* This indicates that online delivery and table booking does not affect ratings.
* Only 1 restaurant from the targeted cities provide table booking service, which is in Doha with exceptional rating of 4.7.
* The rest of cities do not seem to be affected by table booking service.
* By looking at the data, we should provide online delivery and table booking services for our new restaurants so that this will be our unique selling point to attract more customers to our new restaurants.

A screenshot of a computer screen

Description automatically generated

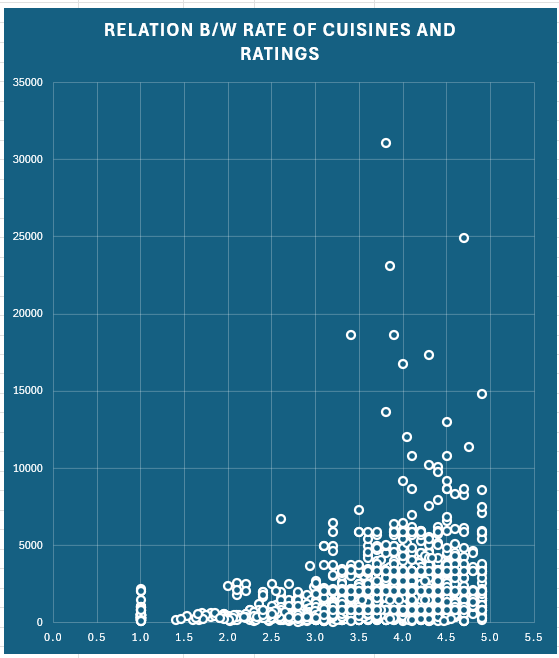
A pie chart with text and numbers

Description automatically generated

A screenshot of a computer

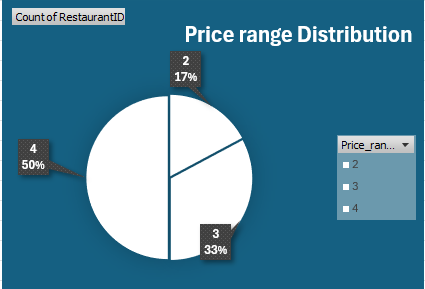
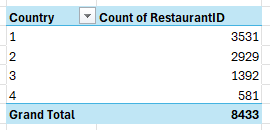
Description automatically generated

1. **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?**



* + We can observe from the above scatter plot created from our whole data, that there is a weak positive correlation between ratings and average cost.
  + The highly-rated restaurants are the ones that offer more expensive food.
  + The variability in price even in the high-rated restaurants bracket indicates that correlation is weak.
  + We can say that if our restaurants are rated highly, we can keep our cuisine rates high but it also indicates that we should have options with lower prices too.

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

****

* + We can see that our data set contains 3531 restaurants which amounts to 42% falling into lowest price range of 1.
  + Price range 2 has about 35% of restaurants which amounts to 2929 restaurants.
  + Followed by price range 3 and 4 having 16% and 7% of restaurants which amounts to 1392 and 581 restaurants respectively.

1. **Explain your approach in brief for suggesting countries/cities in order to open new restaurants, if the objective and subjective questions would have not been given to assist you.** 
   * I would first look at countries with less count of restaurants compared to population of the country, i.e. India
   * After narrowing down the countries that fit the above criteria, I would further narrow down my search for cities by prioritizing cities which are high population centres of the shortlisted countries. Such as Mumbai, Delhi, Bangalore etc.
   * The third criteria to try and make the restaurants more successful, I would focus on cuisines which are more popular among the locals.
   * Additionally, I would provide extra services such as free delivery, reservations, catering services and attractive launch offers.
   * I would give same importance as cuisines to the ambience of the restaurant, which makes more attraction to the customers.
   * For couples, bachelors, and family have separate sections in the restaurant.
   * Music makes the restaurant standout from the crowd, so in weekends there will artists playing songs for good vibes.
   * I will use all the above strategies in order to identify new locations for my restaurants and maximise their profitability.