

**ZOMATO**

**RESTAURANT   
ANALYSIS**

**Problem Statement:**

The team is looking for expansion and opening more restaurants. Your task is to develop strategies/suggestions for opening newer restaurants.

**Objective Questions:**

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

There are two tables present in the data provided one is the **Raw Data** table and other is the **Country Description** table.

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

There are **20** attributes present in the data and all of them are mentioned below.

1. RestaurantID
2. RestaurantName
3. CountryCode
4. City
5. Address
6. Locality
7. LocalityVerbose
8. Longitude
9. Latitude
10. Cuisines
11. Currency
12. Has\_Table\_booking
13. Has\_Online\_delivery
14. Is\_delivering\_now
15. Switch\_to\_order\_menu
16. Price\_range
17. Votes
18. Average\_Cost\_for\_two
19. Rating
20. Datekey\_Opening
21. How many categorical columns are there in the data? [Search about categorical and continuous data, and try to answer this question]

**Categorical Data**

The categorical data consists of categorical variables which represent the characteristics such as a person’s gender, hometown etc. Categorical measurements are expressed in terms of natural language descriptions, but not in terms of numbers. Sometimes categorical data can take numerical values, but those numbers do not have mathematical meaning. Some of the examples of the categorical data are as follows:

* Birthdate
* Favourite sport
* School Postcode

There are **14** categorical columns in the data, all are mentioned below.

1. RestaurantName
2. CountryCode
3. City
4. Address
5. Locality
6. LocalityVerbose
7. Longitude
8. Latitude
9. Cuisines
10. Has\_Table\_booking
11. Has\_Online\_delivery
12. Is\_delivering\_now
13. Switch\_to\_order\_menu
14. Average\_Cost\_for\_two
15. Datekey\_Openin
16. The data consists of some inconsistent and missing values so ensure that the data used for further analysis is cleaned.

For cleaning the whole data, We first selected the whole **Raw Data** table and then by **Data Cleanup** function, we will remove whitespaces present in some cells and then, we will check for duplicate data, here all rows are unique. After checking we found that some city names were not correct, those were **ÛÁstanbul** and **Brasí\_lia** I corrected that as **İstanbul** and **Brasília**. We found in the Cuisines Column there were some BLANKS values, so I changed blanks to **FOOD**.

1. Using the Lookup functions, fill up the countries in the original data using the country code.

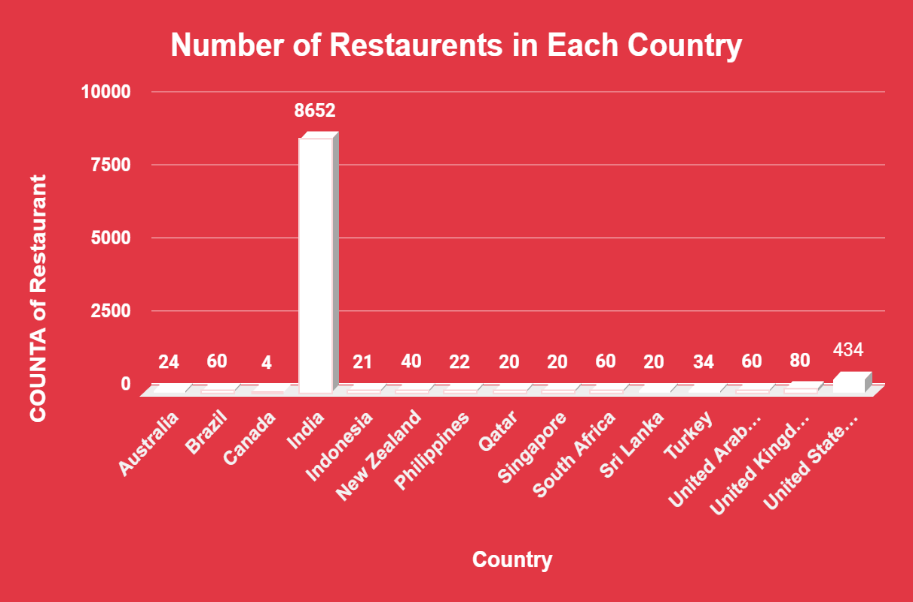
Created a column named **Country** in the Raw Data table using lookup function

=**XLOOKUP($C2,'country description'!$A$2:$A,'country description'!$B$2:$B)**

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

* According to our analysis we find that India has highest number of restaurants **8652.**
* And the Canada have least number of restaurants that is **4.**
* And the total number of restaurants in the given data set that is **9551.**

Select all Raw data and create a pivot table, and in row we add the country column and in value we add the Restaurant Name column and summaries by count function.



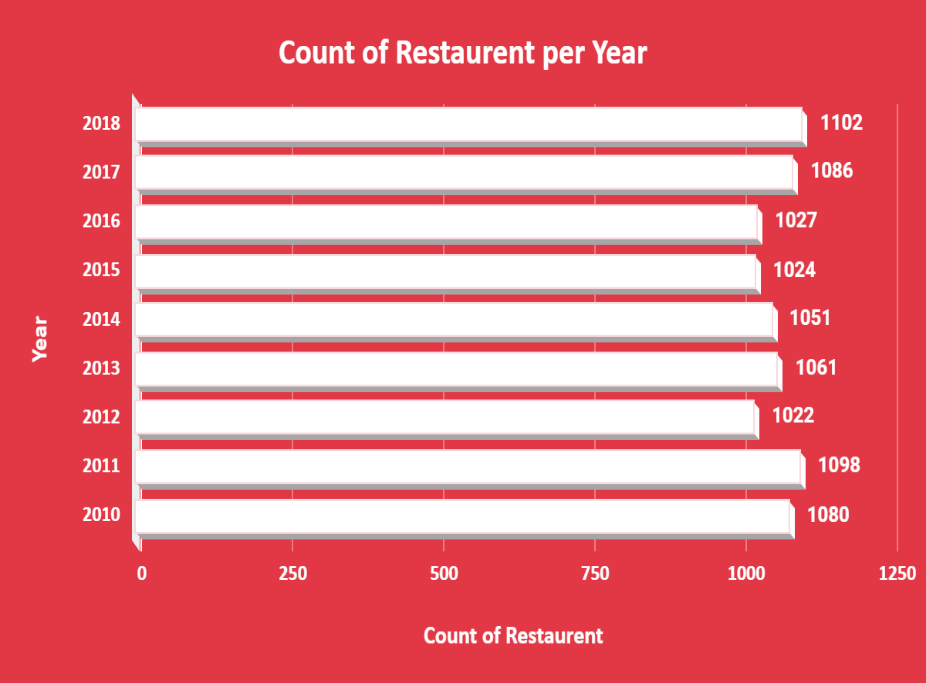
| *Country* | COUNT of Restaurant |
| --- | --- |
| Australia | 24 |
| Brazil | 60 |
| Canada | 4 |
| India | 8652 |
| Indonesia | 21 |
| New Zealand | 40 |
| Philippines | 22 |
| Qatar | 20 |
| Singapore | 20 |
| South Africa | 60 |
| Sri Lanka | 20 |
| Turkey | 34 |
| United Arab Emirates | 60 |
| United Kingdom | 80 |
| United States of America | 434 |

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

* According to our data, every year more than 1000 restaurants are open all over the country from 2010 to 2018.
* As per our analysis the maximum number of restaurants open in 2018 and the minimum number of restaurants open in 2012.

Create the pivot table add Year column in row and Restaurant Name column in value and summaries by count function.

| *Opening\_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** |



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

**=COUNTIFS ('Raw Data’! D2:D,"INDIA",'Raw Data’! R2:R,"4")**

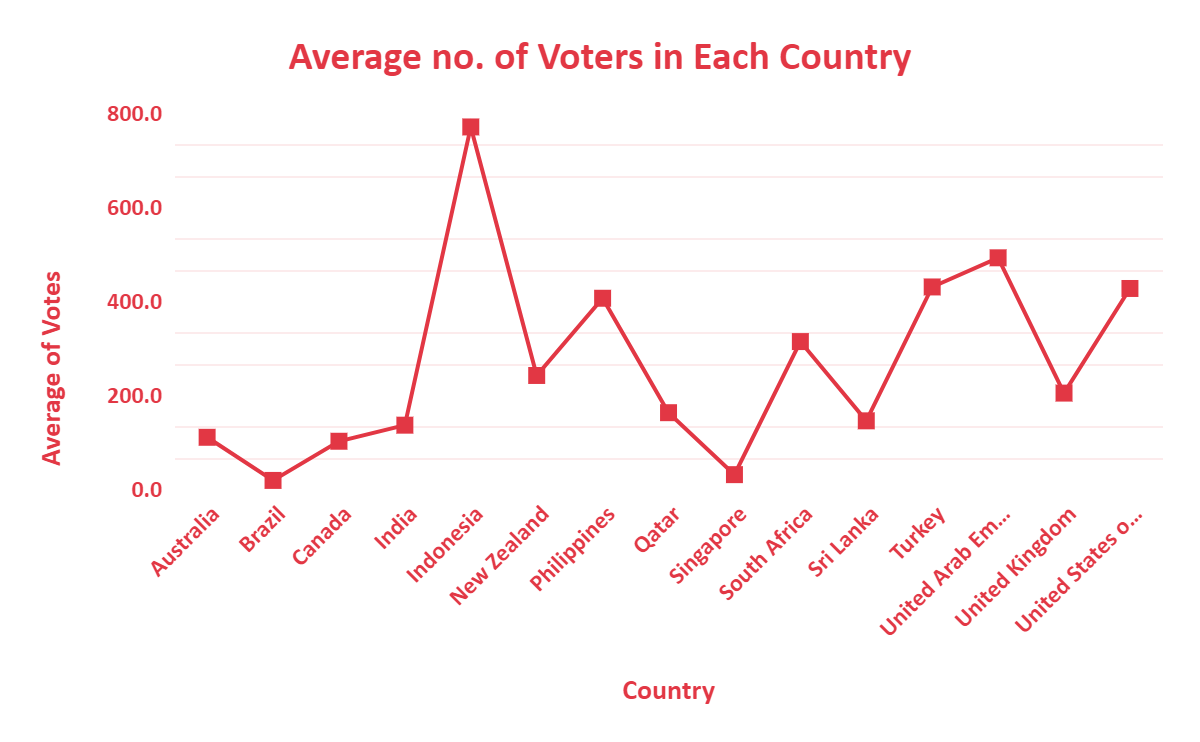
Using the COUNTIFS function on the Raw data, we get the number of restaurants in India which are in the (Price range of 4) **388.**

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

Create the pivot table and add country column in row and votes column in value and summaries by Average function.

With the help of pivot table, we can easily define the Average votes of country. According to our analysis **Brazil** got minimum average of votes and **United Arab Emirates** maximum average of votes.

| *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 |



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. [Note: Don’t use Conditional aggregation in this question.]

We will create a **Helper Column** with using this formula **=IF(AND(R2<4,O2="Yes"),V2,"")** and after that, we will take the average of that column by using **=AVERAGE('Raw Data'!W:W)** and the average of rating was **3.27381151.**

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.

First we will make a Suggested cities column with the help of vlookup formula =VLOOKUP(E2,'SQ2- Name of cities'!B:B,1,0) and from that we got #N/A in every column accept suggested cities and after that we will make a Suggested Country column and filter it by Yes for the suggested cities and no for others.

And after that we will put this formula for in conditional formatting custom formula **=$AA$2:$AA="Yes".**

1. 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]

First we will make **Symbol** column at the right of the **Currency** column at we will extract symbols by using this formula **=REGEXEXTRACT(L3, "\((.\*?)\)").**

Secondly we will make a **Average Cost with Currency** column and will fill that column by applying concatenation of **Average\_cost\_for\_two** column and **Symbol** column by using this formula **=CONCATENATE(M2,T2) .**

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?

We wil apply **COUNTIFS** condition where **Country** is equal to India and **Has\_Online\_delivery** is equal to **NO** and are in the lowest **Price\_range** that is **1,** The count we got is **1685**.

**=ArrayFormula(COUNTIFS('Raw Data'!D2:D,"India",'Raw Data'!O2:O,"No",'Raw Data'!R2:R,"1",'Raw Data'!T2:T,"<=250"))**

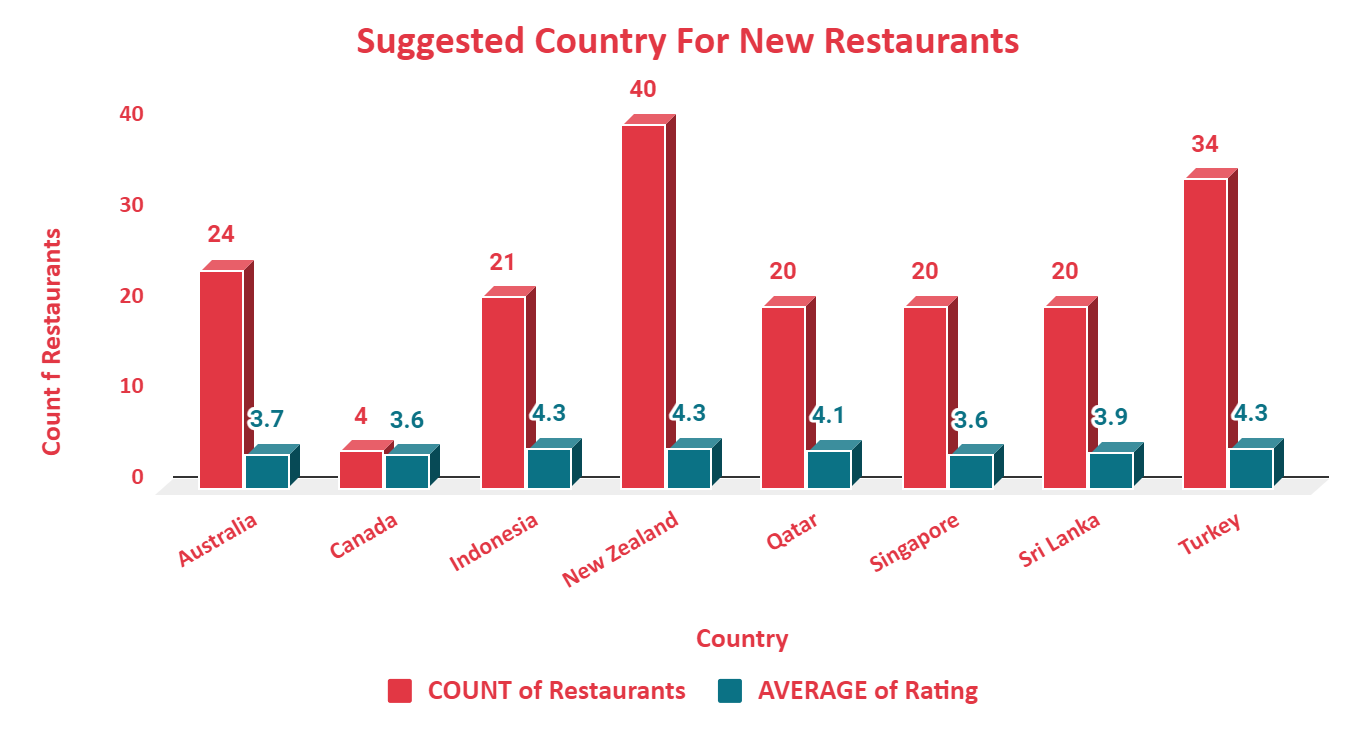
**Subjective Questions:**

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?

We are suggesting few countries for opening new restaurants in Data Analysis sheet in excel project. That are **Australia, Canada, Indonesia, New Zealand, Qatar, Singapore, Sri Lanka, Turkey.** Number of restaurants in this country is very less and **rating is below than 4.5.**

On my visualization and technique create pivot table and add country column in row and restaurants name in value summaries by count function. We fetch the list of number of restaurants in each county after that we add filter on count of restaurants column and select only up to 40 restaurants and we get result of few countries where we open new restaurants in lesser competition.

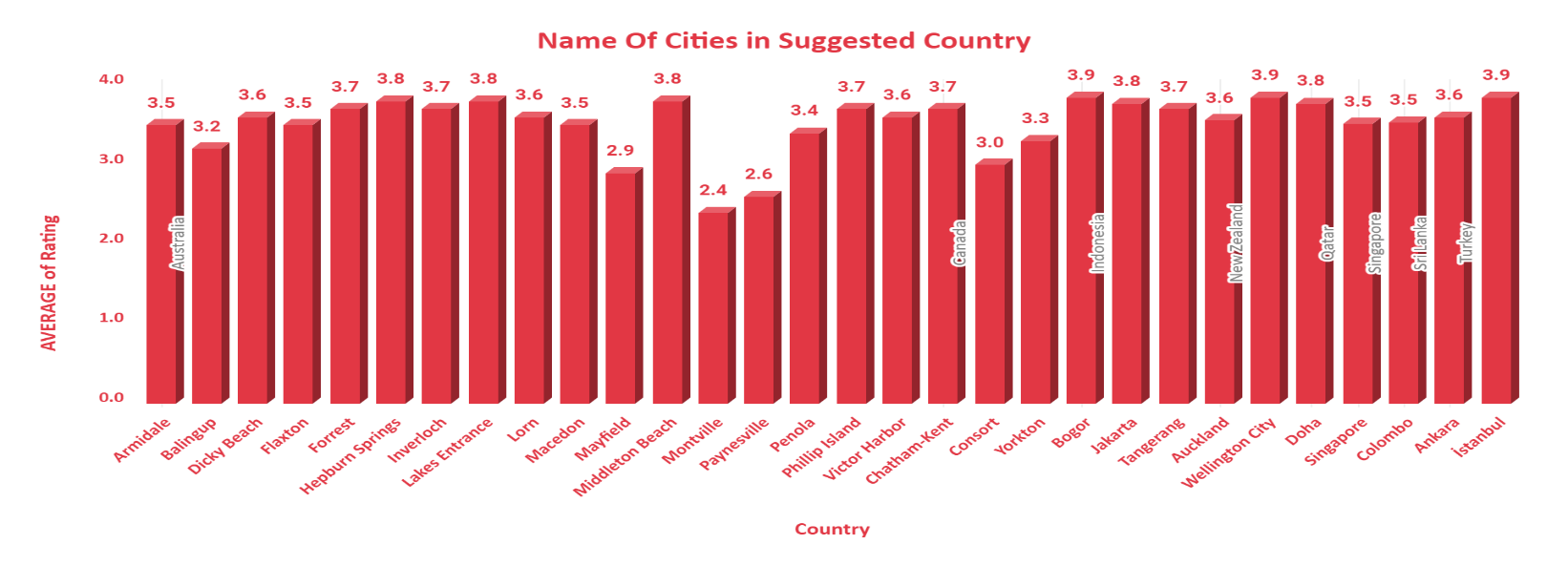
| ***Country*** | **COUNT of Restaurants** | **AVERAGE of Rating** |
| --- | --- | --- |
| Australia | 24 | 3.7 |
| Canada | 4 | 3.6 |
| Indonesia | 21 | 4.3 |
| New Zealand | 40 | 4.3 |
| Qatar | 20 | 4.1 |
| Singapore | 20 | 3.6 |
| Sri Lanka | 20 | 3.9 |
| Turkey | 34 | 4.3 |



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

There are **30** cities in the suggested countries which have average ratings less than **4** and suitable for opening restaurants.

| *Country* | *City* | COUNT of Restaurant | AVERAGE of Rating |
| --- | --- | --- | --- |
| Australia | Armidale | 1 | 3.5 |
|  | Balingup | 1 | 3.2 |
|  | Dicky Beach | 1 | 3.6 |
|  | Flaxton | 1 | 3.5 |
|  | Forrest | 1 | 3.7 |
|  | Hepburn Springs | 2 | 3.8 |
|  | 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 |
|  | Paynesville | 1 | 2.6 |
|  | Penola | 1 | 3.4 |
|  | Phillip Island | 1 | 3.7 |
|  | Victor Harbor | 1 | 3.6 |
| Canada | Chatham-Kent | 1 | 3.7 |
|  | Consort | 1 | 3.0 |
|  | Yorkton | 1 | 3.3 |
| Indonesia | Bogor | 2 | 3.9 |
|  | Jakarta | 3 | 3.8 |
|  | Tangerang | 1 | 3.7 |
| New Zealand | Auckland | 5 | 3.6 |
|  | Wellington City | 2 | 3.9 |
| Qatar | Doha | 12 | 3.8 |
| Singapore | Singapore | 18 | 3.5 |
| Sri Lanka | Colombo | 11 | 3.5 |
| Turkey | Ankara | 2 | 3.6 |
|  | İstanbul | 4 | 3.9 |

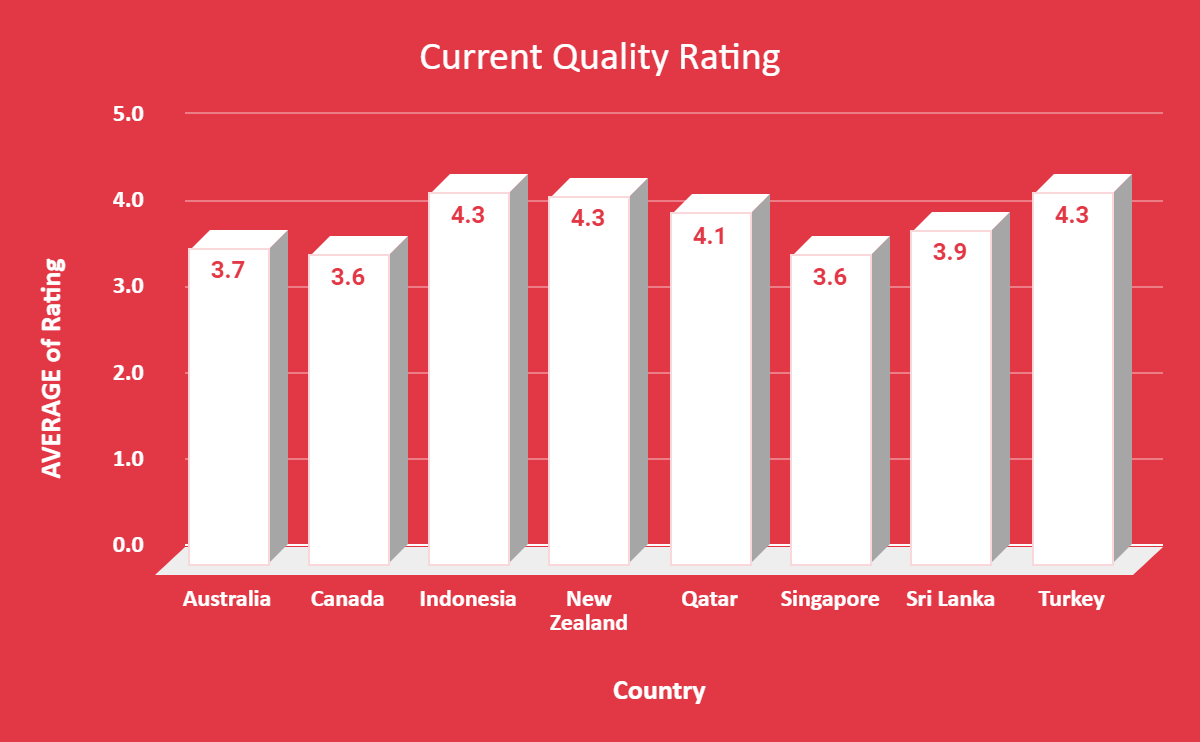


1. According to the countries you suggested, what is the current quality regarding ratings for restaurants that are open there?

According to my analysis the average rating of suggested country is minimum **3.6** and country are **Canada, Singapore**. The maximum average rating is 4.3 and country are **Indonesia, New Zealand, and Turkey.**

To fetching the quality in term off Rating we add the rate column in value and summaries by Average function on same pivot table of Data analysis sheet.

| ***Country*** | **COUNT of Restaurant** | **AVERAGE of Rating** |
| --- | --- | --- |
| Australia | 24 | 3.7 |
| Canada | 4 | 3.6 |
| Indonesia | 21 | 4.3 |
| New Zealand | 40 | 4.3 |
| Qatar | 20 | 4.1 |
| Singapore | 20 | 3.6 |
| Sri Lanka | 20 | 3.9 |
| Turkey | 34 | 4.3 |



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

The average money we spend on food is called as Expenditure on food. For opening the new restaurant, we know the average money we spend on food to control the financial expenditure.

As per my analysis we find the average cost of two person as per given data and change the Country currency in US Doller $ for convenience.

And we find the in Tukey we spend average amount like 2.80$ and in Singapore we need to spend 116.34 $ to control the financial expenditure.

For fetching the current expenditure on food in the suggested countries we add the average cost of two column in value and summaries by Average function to get the average cost in that country. And for new restaurants we can keep our financial expenditure in control.

| *Country* | AVERAGE of Average\_Cost\_for\_two | All units converted to US dollar |
| --- | --- | --- |
| Australia | 24.1 | $15.87 |
| Canada | 36.3 | $27.03 |
| Indonesia | 281190.5 | $17.85 |
| New Zealand | 69.8 | $42.79 |
| Qatar | 223.8 | $61.47 |
| Singapore | 155.8 | $116.34 |
| Sri Lanka | 2375.0 | $7.52 |
| Turkey | 84.9 | $2.80 |
| *Country* | *Currency* |
| Australia | Dollar($) |
| Canada | Dollar($) |
| Indonesia | Indonesian Rupiah(IDR) |
| New Zealand | NewZealand($) |
| Qatar | Qatari Rial(QR) |
| Singapore | Dollar($) |
| Sri Lanka | Sri Lankan Rupee(LKR) |
| Turkey | Turkish Lira(TL) |



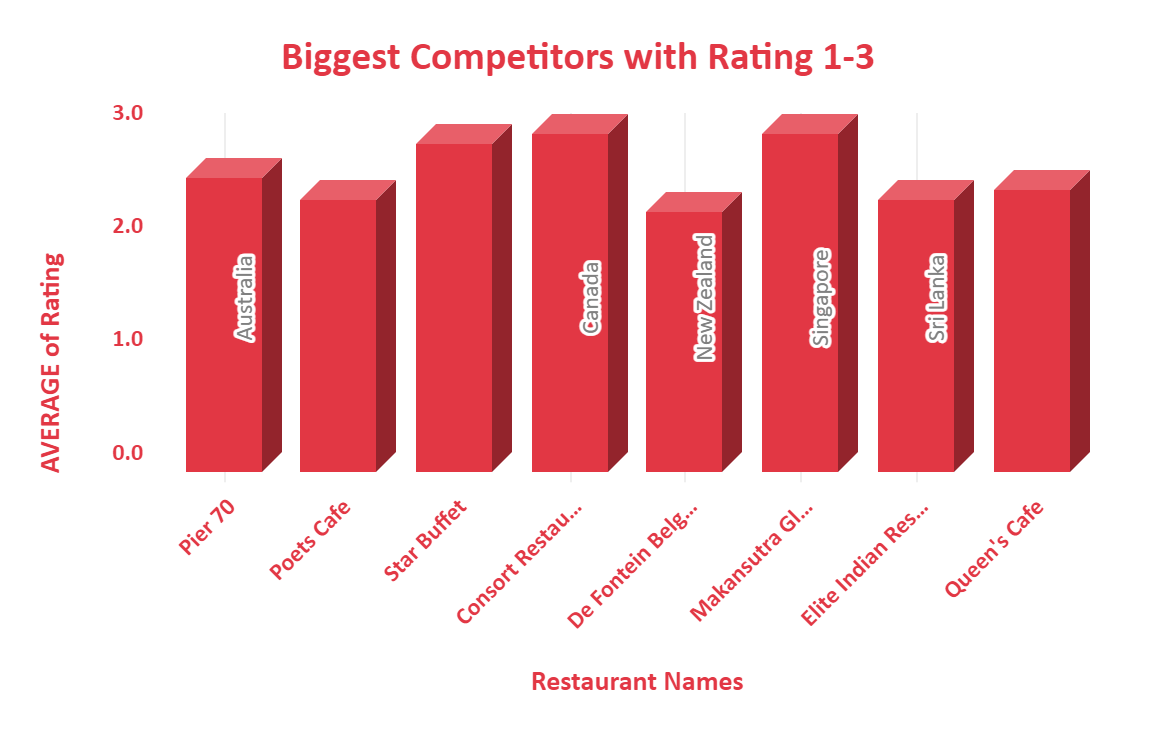
1. Come up with the names of restaurants from the recommended states that are our biggest competitors and also those that are rated in the lower brackets, i.e. 1-2 or 2-3.

There are total **5** countries(**Australia,Canada,New Zealand,Singapore,Srilanka**) and **8** restaurants who are coming under the lower brackets i.e. 1-2 or 2-3 , And the **Biggest Competitor** is **New Zealand’s *De Fontein Belgian Beer Cafe***

with **2.3** rating.

Create a Pivot table and add Country,Rating,Restaurant Name in Rows and Rating in values and summarise by average function and at last add filters on Country and select suggested **8** countries only i.e.(**Australia, Canada, Indonesia, New Zealand, Qatar, Singapore, Sri Lanka, Turkey**) and on Rating select it as **less than 3** and group rating as 1-2, 2-3.

| *Country* | *Grouped Rating* | *RestaurantName* | AVERAGE of Rating |
| --- | --- | --- | --- |
| Australia | 2 - 3 | Pier 70 | 2.6 |
|  |  | Poets Cafe | 2.4 |
|  |  | Star Buffet | 2.9 |
| Canada | 2 - 3 | Consort Restaurant | 3.0 |
| New Zealand | 2 - 3 | De Fontein Belgian Beer Cafe | 2.3 |
| Singapore | 2 - 3 | Makansutra Gluttons Bay | 3.0 |
| Sri Lanka | 2 - 3 | Elite Indian Restaurant | 2.4 |
|  |  | Queen's Cafe | 2.5 |



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

We focus on High rating cuisines for newer restaurants because different country has different cuisines and only some cuisines have high rating. And other cuisines have low rating and it also affect the rating of the cuisines.

As per my analysis we select **118** different cuisines to serve in our newer restaurants. Name of all 118 cuisines are written in a table in google sheet named **SQ6-Cuisines Name For Newer Restaurants.**

Create pivot table and add country and cuisine in row and add Cuisines in value and summarise by count and also add Rating in value and summaries by average.

| *Country* | COUNT of Cuisines |
| --- | --- |
| Australia | 6 |
| Canada | 1 |
| Indonesia | 17 |
| New Zealand | 37 |
| Qatar | 11 |
| Singapore | 3 |
| Sri Lanka | 13 |
| Turkey | 30 |
| **Grand Total** | **118** |

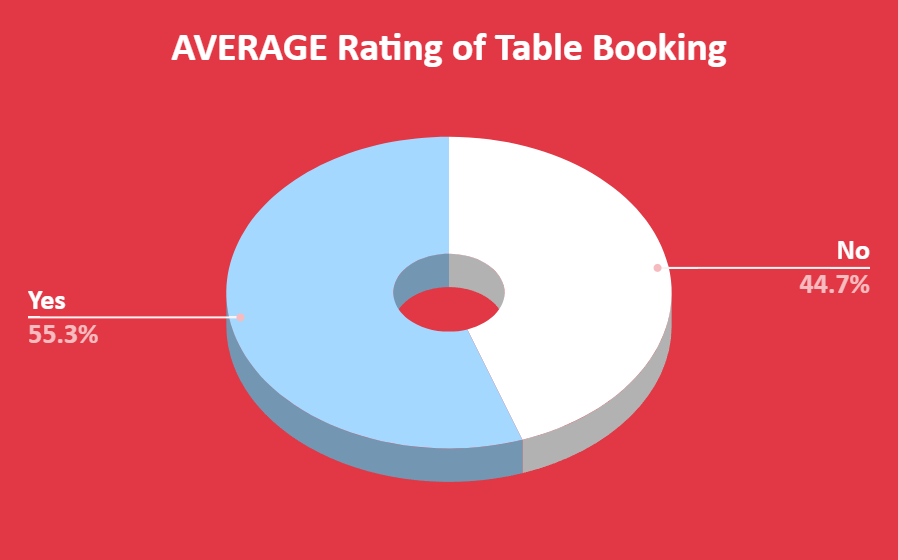


1. According to our current data, should we go for online delivery and table booking? Does that affect the customer’s ratings?

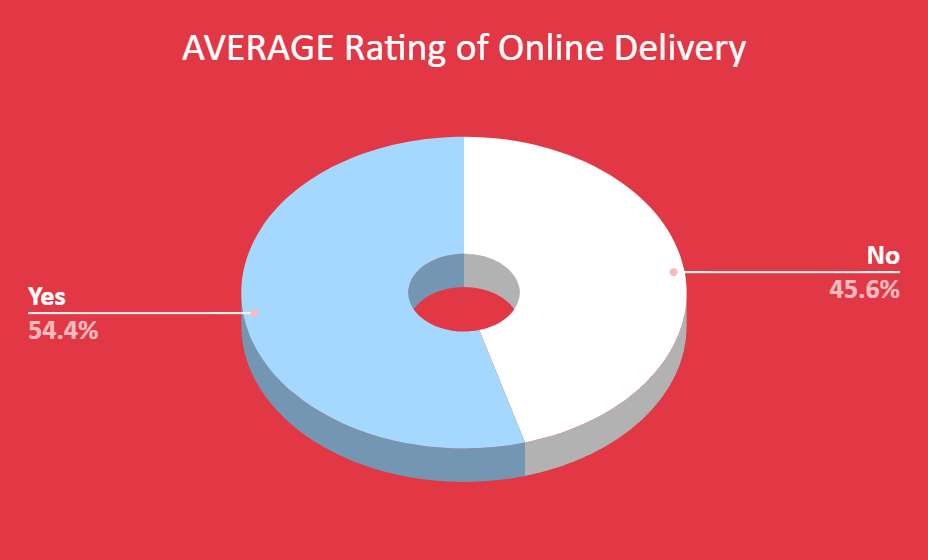
Yes, we can go for online delivery and table booking because the average rating is higher in those restaurants having both online delivery and table booking. And it also affects the rating.

We created two Pivot tables, one for Table\_Booking and other for Online\_Delivery for finding average ratings in both **Yes** and **No** scenarios.

| **Table Booking** |  |
| --- | --- |
| *Has\_Table\_booking* | AVERAGE of Rating |
| No | 2.809686644 |
| Yes | 3.482556131 |



| **Online Delivery** |  |
| --- | --- |
| *Has\_Online\_delivery* | AVERAGE of Rating |
| No | 2.754309859 |
| Yes | 3.288004896 |



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?

As per my analysis you can increase the number of cuisines but for the Rate of cuisines put it as average not more than that. By this you can capture the market and control your financial expenditure.

Yes, the price range and rate are correlated that is **0.46** this not Good strength but it comes in moderate section.

**Formula is – CORREL(A:A,B:B)**

We create the pivot table and add countries and cuisines in row, after that we add rating in value and summarize by average and add average of two in value and summarize by average.

| **Correlation** | **Scale** |
| --- | --- |
| 0.00-0.19 | Very weak |
| 0.20-0.39 | Weak |
| 0.40-0.59 | Moderate |
| 0.60-0.79 | Strong |
| 0.80-1.00 | Very Strong |
| **Correlation of Price Range and Rating** | 0.46 |
| **Formula - CORREL(A:A,B:B)** |  |

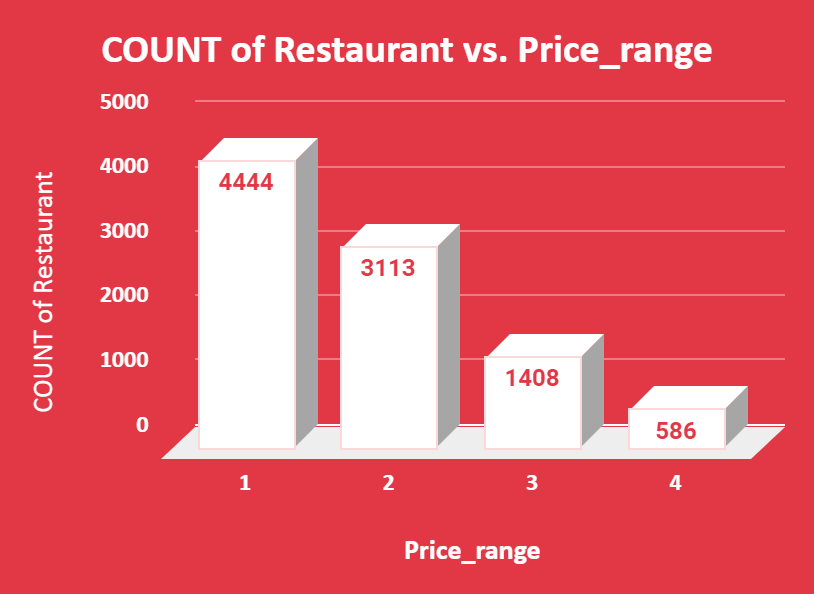


1. What is the distribution of the number of restaurants of different price ranges in all the countries?

According to given data **9551** restaurants available in all countries and with different price ranges like in Price Range 1 there are **4444** restaurants, and Price Range 2 there are **3113** restaurants, Price Range 3 there are **1408** restaurants, and Price Range 4 there are **586** restaurants present.

Once again, we create a pivot table and add Price range in row and Restaurants Name in value and summaries by Count function. So, we easily fetch the different price range of restaurants of all the countries.

| *Price\_range* | COUNT of Restaurant |
| --- | --- |
| 1 | 4444 |
| 2 | 3113 |
| 3 | 1408 |
| 4 | 586 |
| **Grand Total** | **9551** |



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. **[you have to give bullet pointers in order to answer this question]**

My approach for the analysis will be like that :-

* I’ll clean the data first and fill in the blanks if any .
* I’ll check the country in the data , if there is no country I’ll extract country to a new column from the table ,because we have to expand the restaurants all across the world so country wise data would be needed
* After that I’ll see the no. of restaurants in each country.
* Next I’ll see where there are lesser no. of restaurants .
* Then I’ll see the ratings of those restaurants.
* I’ll try to sort the city according to the rating and restaurant count in the country.
* I’ll try to find out where no restaurant is less and ratings are low.
* After all this I’ll see the price\_range of the restaurant and compare them according to their ratings.
* I’ll find the correlation between Rates and Ratings ,where it would fall. We will try to set the Rates according to that.
* I’ll suggest opening in those countries/Cities where the number of restaurants is less and ratings are not good .

**DASHBOARD**

