# **Tasks**

**Learners have to develop a dashboard to support the answers to the following questions and suggestions for places for newer restaurants.**

**Objective Questions**:

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

A table is generally represented in the data as a collection of data-sets or data-points, along with their headers, that we use to study or analyse. Tables can be counted by checking for distinct ranges of cells without any break in data-points within an entire row or an entire column.

**Observation:**

In this case, the total number of tables can be counted manually to see that each of the worksheets named ‘Raw Data’ & ‘country description’ contain exactly one table each

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

An attribute is the characteristic or a feature representing various aspects of a data.

**Observation:**

In this given dataset, attributes of the raw data are indicated by the column header representing individual pieces of information about the Zomato data.

To count the total number of attributes in this case, we can utilise the =COUNTA(starting column header:ending column header) function to count the number of headers in a given range of the dataset (here, our raw data).

Alternatively, we can also select all the headers, and see the count at the bottom right of the google sheet



The count of attributes in the given raw data equals 20.

1. **How many categorical columns are there in the data? [Search about categorical and continuous data, and try to answer this question]**

Categorical or qualitative variables tell us about the group or quality of a particular variable in a data-point.

**Observation:**

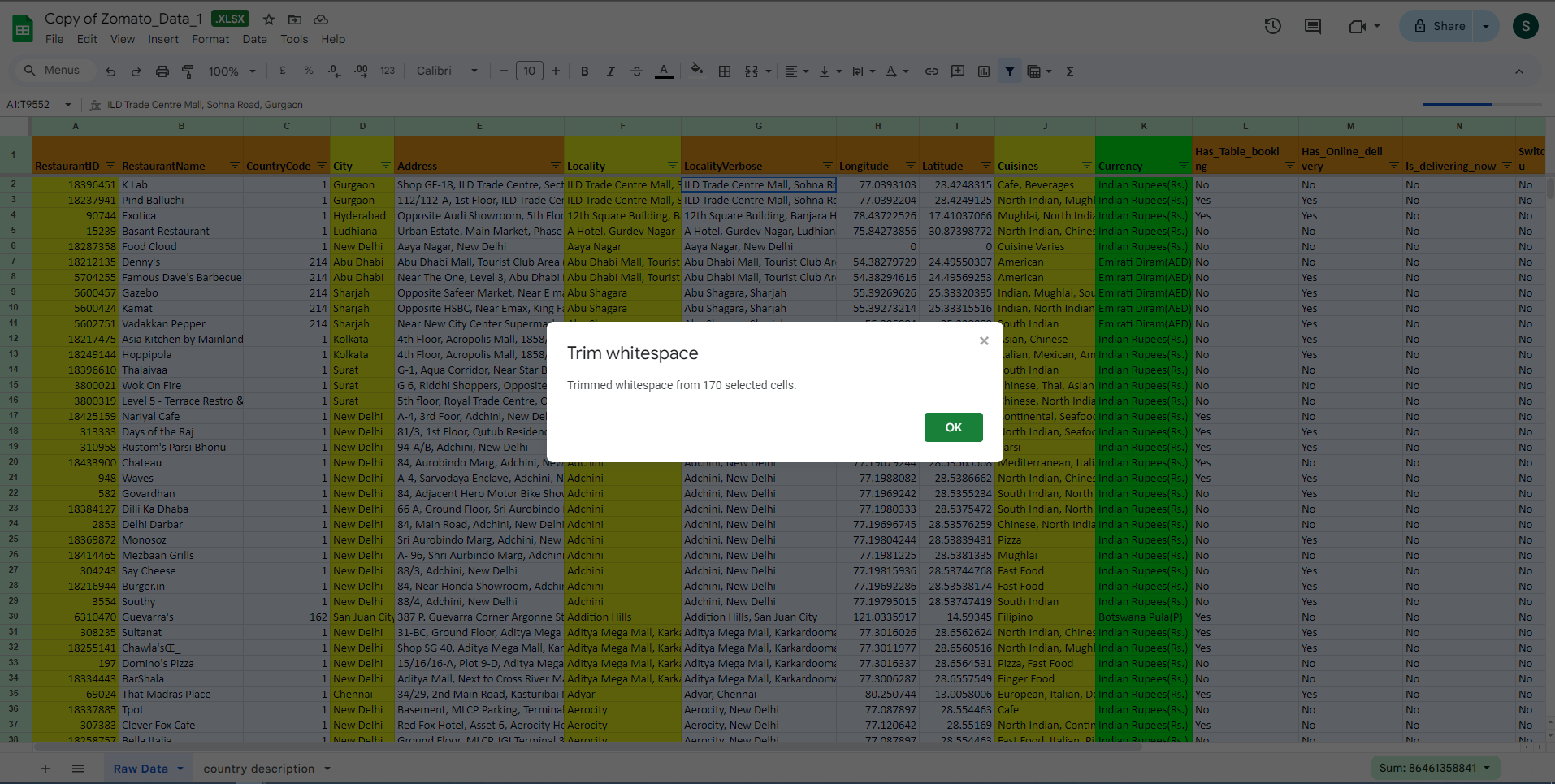
There are sixteen (16) categorical columns within the raw data.These include Restaurant Id & name, Country code, city, Address, Locality & Locality verbose, Longitude, Latitude, Cuisines, Currency, Has Table Booking, Has Online Delivery, Is delivering now, Switch to Order menu and Date Key opening.

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

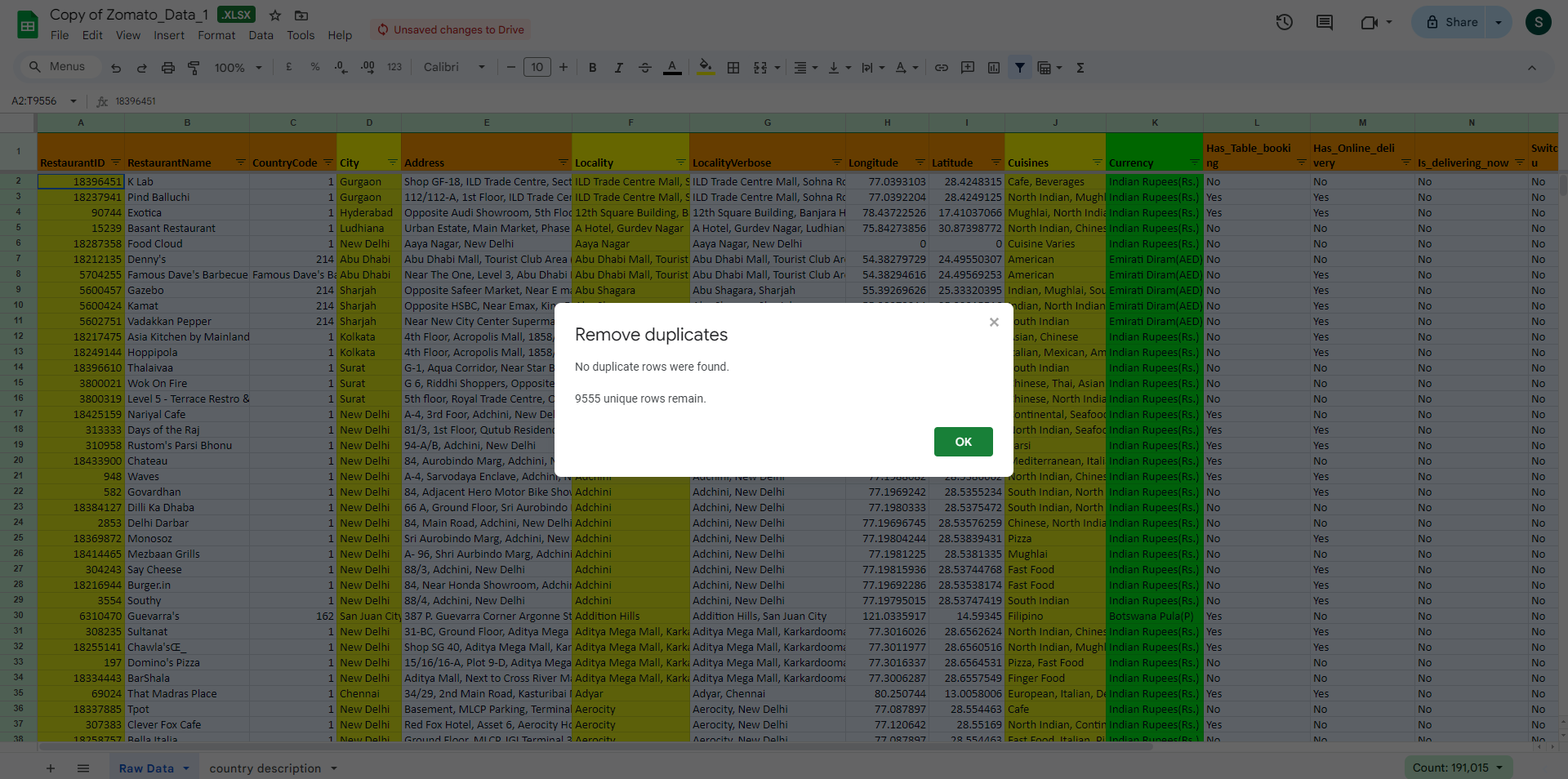
Guideline: To clean the data for inconsistent and missing values for better analysis and interpretation.

**Observation:**

* The data consists of several cells with extra spaces. To remove of the possibility of white spaces affecting the data anaylses, they were trimmed by Selecting Data> Data-cleanup > Trim white spaces



* No duplicates row entries were found within the raw data



* Certain values in the cuisines column were found to be missing. =COUNTBLANK() function was utilised for each column using mixed cell referencing to count blank cells in each column.
  + The raw data consists of nine missing values under the Column header ‘Cuisines’.
  + An additional column (Cuisines\_Imputed) was created in order to handle these missing values. Data was populated using the IF(ISBLANK()) function, where the missing values were being addressed as ‘Data not available’.
* Average cost for dining is given in different currencies. Hence, a currency conversion chart was created in the ‘country description’ worksheet and the average cost for dining was normalised into Indian rupees.
* Values in the opening date key were not found in a proper spreadsheet recognized date order. This was normalised using the following function
  + =DATE(VALUE(REGEXEXTRACT(V2, "\d{4}")), VALUE(REGEXEXTRACT(V2, "\d{1,2}")), VALUE(REGEXEXTRACT(V2, "\d{1,2}")))
  + The =Regexextract(text,regular\_expression) function matches and extracts a regular pattern within a string.

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

Guideline: Create a separate column in the raw data to populate the name of the countries using the country code, by using a LOOKUP function

**Observation:**

A new ‘Country’ column was created and country names were filled in the table using the =VLOOKUP AND MATCH function, by tallying with the country code table present in the ‘country description’ worksheet.

=VLOOKUP(C2,'country description'!$A$1:$B$16,MATCH('country description'!$B$1,'country description'!$A$1:$B$1,0),0)

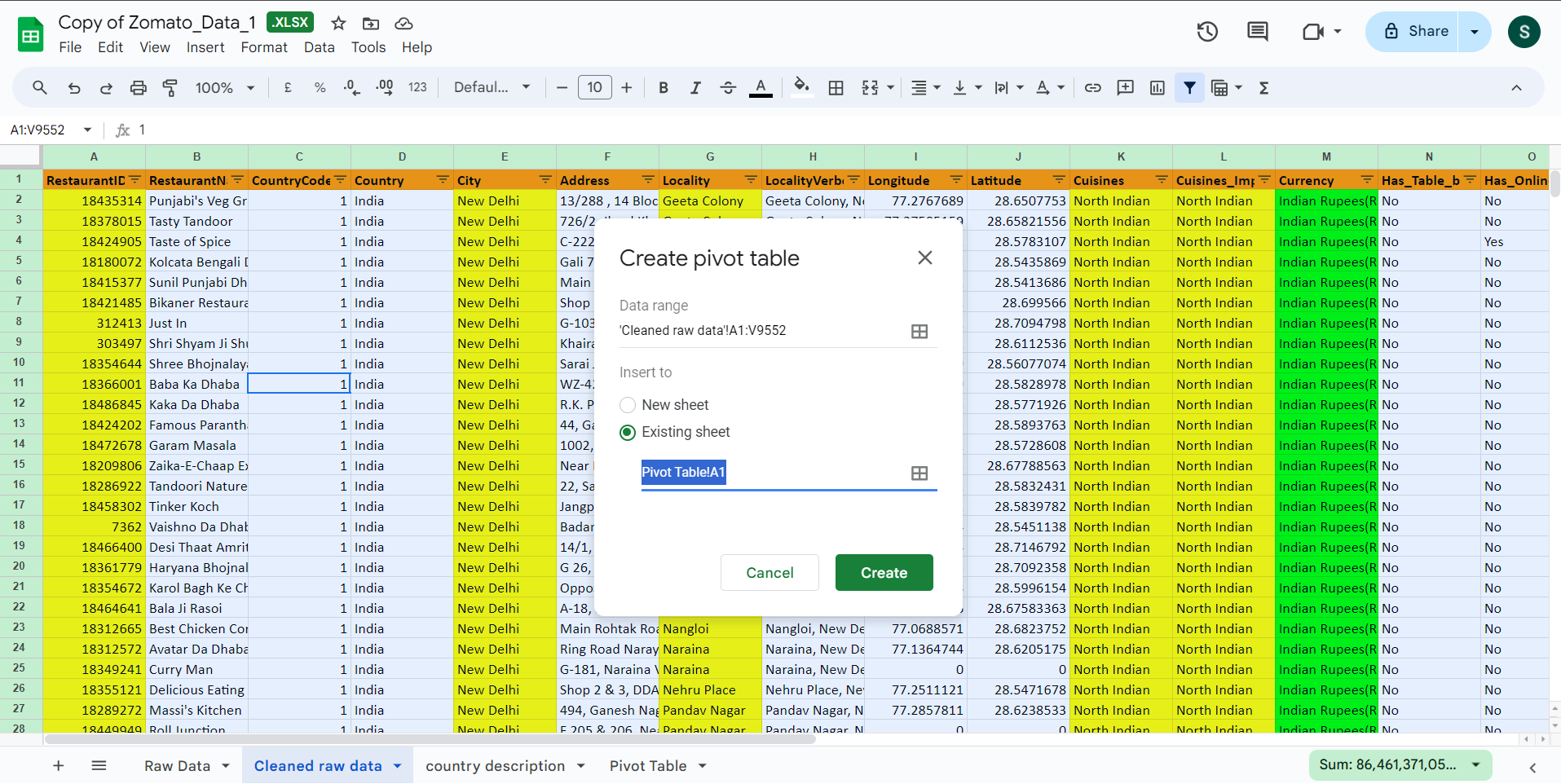
1. **Create a table to represent the number of restaurants opened in each country.**

Guideline: To create a table to represent the total number of restaurants opened in each country

**Observation:**

Total numbers of restaurants sorted country-wise can be displayed using a Pivot Table. A pivot table is a way of summarising, visualising, and analysing data.

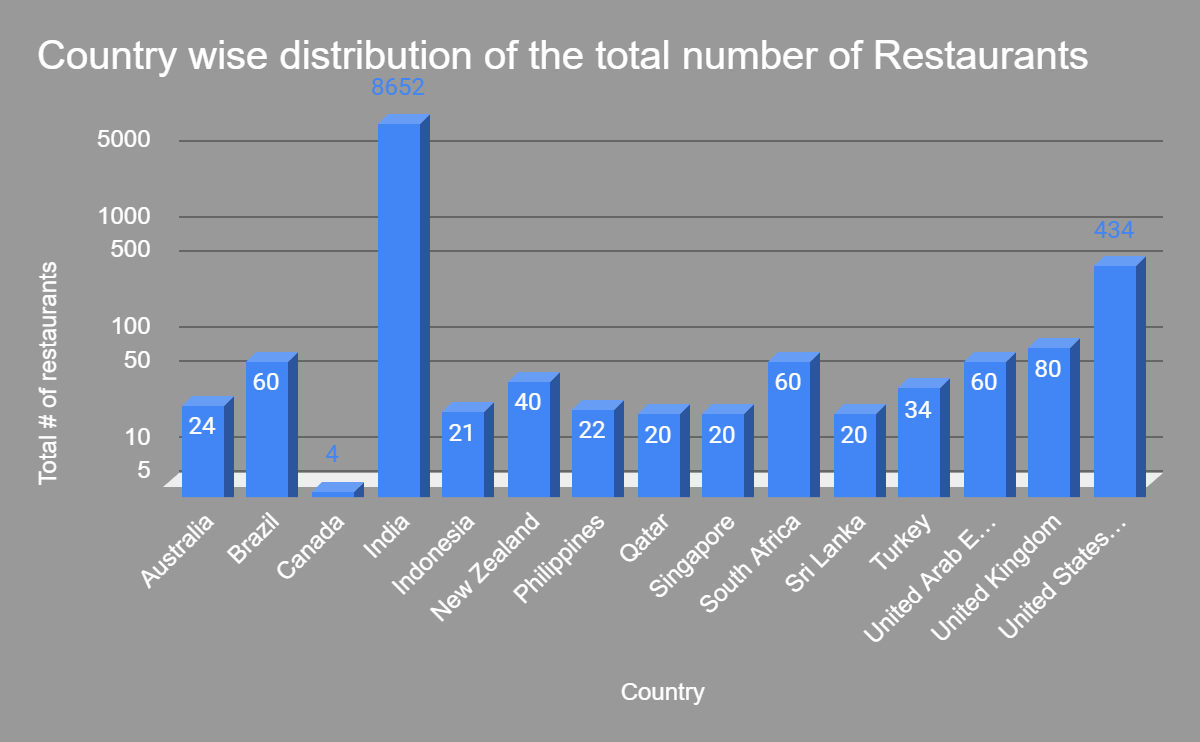
A Pivot table can be created by selecting the desired data > Go to Insert > Pivot Table > Entering the desired location on the popup.



Country names were added into the rows column, and restaurant ids in the values column, where they were summarised by the CountA function to count the unique restaurants associated with that id.

Note: This data was cleaned for duplicates. However, in case if there are any duplicates, data could be summarised using COUNTUNIQUE.

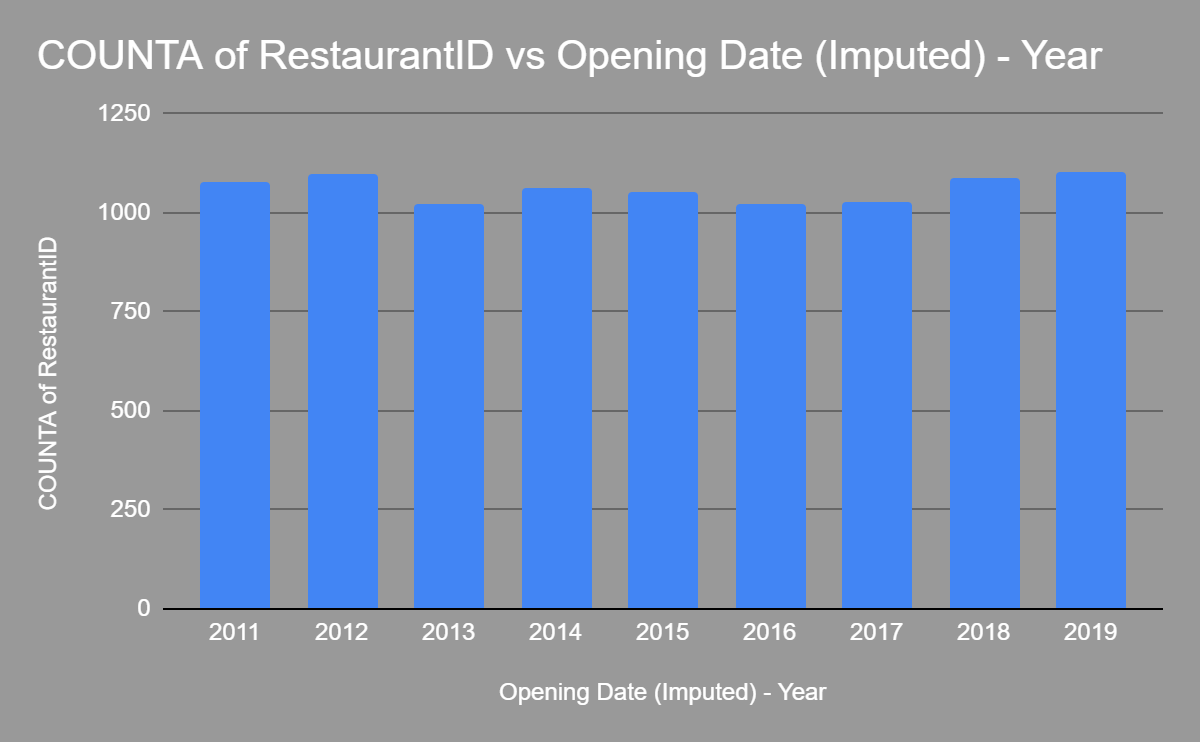
| *Country* | COUNTA of RestaurantID |
| --- | --- |
| Australia | 24 |
| Brazil | 60 |
| Canada | 4 |
| India | 8652 |
| Indonesia | 21 |
| New Zealand | 40 |
| Philippines | 22 |
| Qatar | 20 |
| Singapore | 20 |
| South Africa | 60 |
| Sri Lanka | 20 |
| Turkey | 34 |
| United Arab Emirates | 60 |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



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

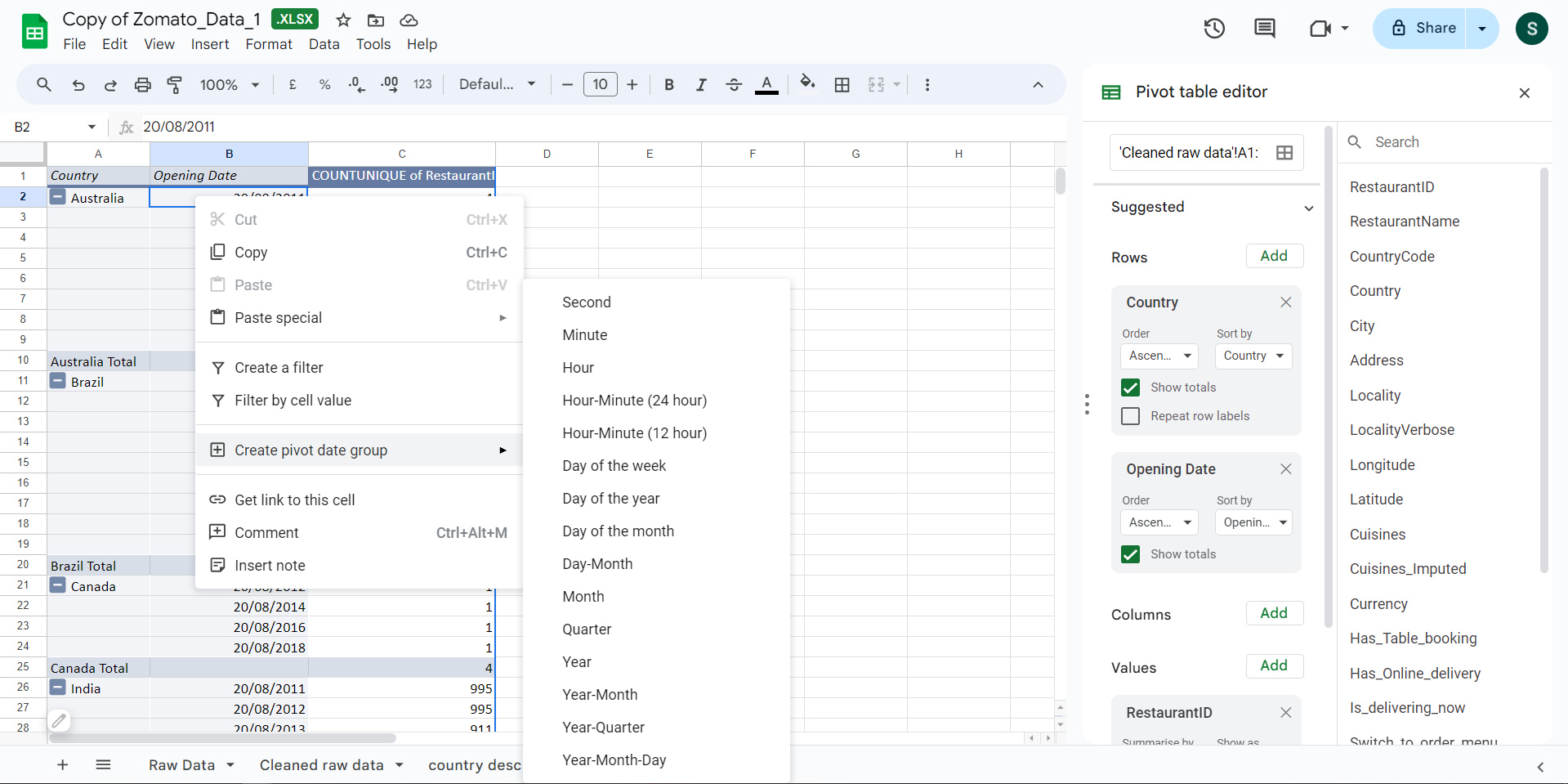
Guideline: To create a pivot table and a column chart to represent the total number of restaurants opened in each year

**Visualisation**:



**Observation:**

To group the number of restaurants according to the opening date, we can push the opening date into the rows column, and group them into years by creating a pivot table rule. Values is filled by COUNTA of restaurant id

Note: To group the date year-wise, we would right-click on any date > Go to Create pivot date group > Year

And finally we would drag the ‘Opening date’ in rows field above the country field to get year wise opening data grouped into different countries

* Almost over a thousand restaurants are opened each year across the world

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

There are two ways counting the total number of restaurants in India in the range of 4. Either we add a price range column in rows in the pivot table and put a filter of Country as India & Price as 4.

Alternatively, we can use the =COUNTIFS('Cleaned raw data'!$D$2:$D,"India",'Cleaned raw data'!$R$2:$R,"4") function by counting the criteria 1 as country as India and criteria 2 as Price Range =4.

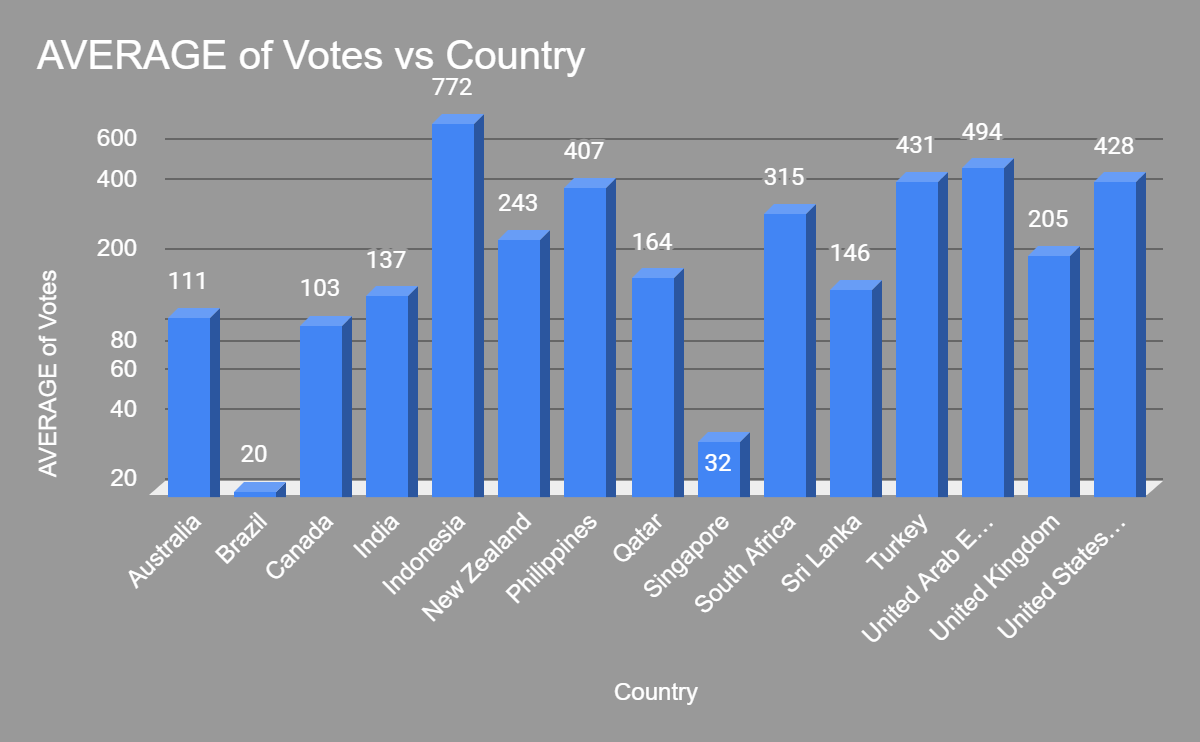
**Observation:**

Total number of restaurants in India 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?**

Guideline: To count the average number of voters in each country

**Visualisation**

****

**Observation:**

* Average number of voters in country can be tabulated in a pivot table by puttings country in the rows column and Votes in the values column summarised as average
  + The average number of voters in each country are given in the chart above. For example, Australia has an 111 number of voters, while Brazil has only 20, for example.

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.]**

Guideline: Calculate the rating of restaurants within a price of 4, with an online delivery using only the IF function with logical operators and the aggregation functions

**Observation:**

The average rating for all the restaurants that have a price range of < 4 & provide online delivery is ~3.3. This can be counted as:

**=ROUND(AVERAGEIFS('Cleaned raw data'!$U$2:$U,'Cleaned raw data'!$R$2:$R,"<4",'Cleaned raw data'!$O$2:$O,"Yes"),2)**

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.**

To highlight any rows using conditional formatting, we go to Format > Conditional Formatting > Select a Format Rule > Done

In this case, we are suggesting cities certain in Canada, Australia, Indonesia and the Philippines as places where Zomato can explore opening of new restaurants, so we would select the Format rule > Format cells if > Custom formula is > =OR($E1="Armidale", $E1="Balingup", $E1="Dicky Beach", $E1="Flaxton", $E1="Forrest", $E1="Hepburn Springs", $E1="Inverloch", $E1="Lakes Entrance", $E1="Lorn", $E1="Macedon", $E1="Mayfield", $E1="Middleton Beach", $E1="Montville", $E1="Paynesville", $E1="Penola", $E1="Phillip Island", $E1="VIctor Harbor", $E1="Chatham-Kent", $E1="Consort", $E1="Yorkton", $E1="Bogor", $E1="Santa Rosa")

1. **Create a new customised 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 have to extract the currency abbreviation from the column ‘currency’ using the MID & FIND Function

=MID(M2, FIND("(", M2) + 1, FIND(")", M2) - FIND("(", M2) - 1)

* +1 extracts from the character that is immediately after the opening parenthesis and -1 would is used to adjust the length of the substring.

Now, we can combine the first string to concatenate with the average cost for two using

=CONCATENATE(MID(M2, FIND("(", M2) + 1, FIND(")", M2) - FIND("(", M2) - 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?**

Guidelines: Count the number of restaurants 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 using an array formula

**Observations**:

An =ARRAYFORMULA(f(x)) is used to apply the formula to an entire column or range at once, instead of having to copy the formula to each cell individually.

We can create it by creating a separate column segregating just the numerical values within the currency column using the formula

=ArrayFormula(COUNTIFS('Cleaned raw data'!$O$2:$O$9552,"No",'Cleaned raw data'!$R$2:$R$9552,"1",'Cleaned raw data'!$Z$2:$Z$9552,"<=250"))

Restaurants 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 are a **total of 1694**.

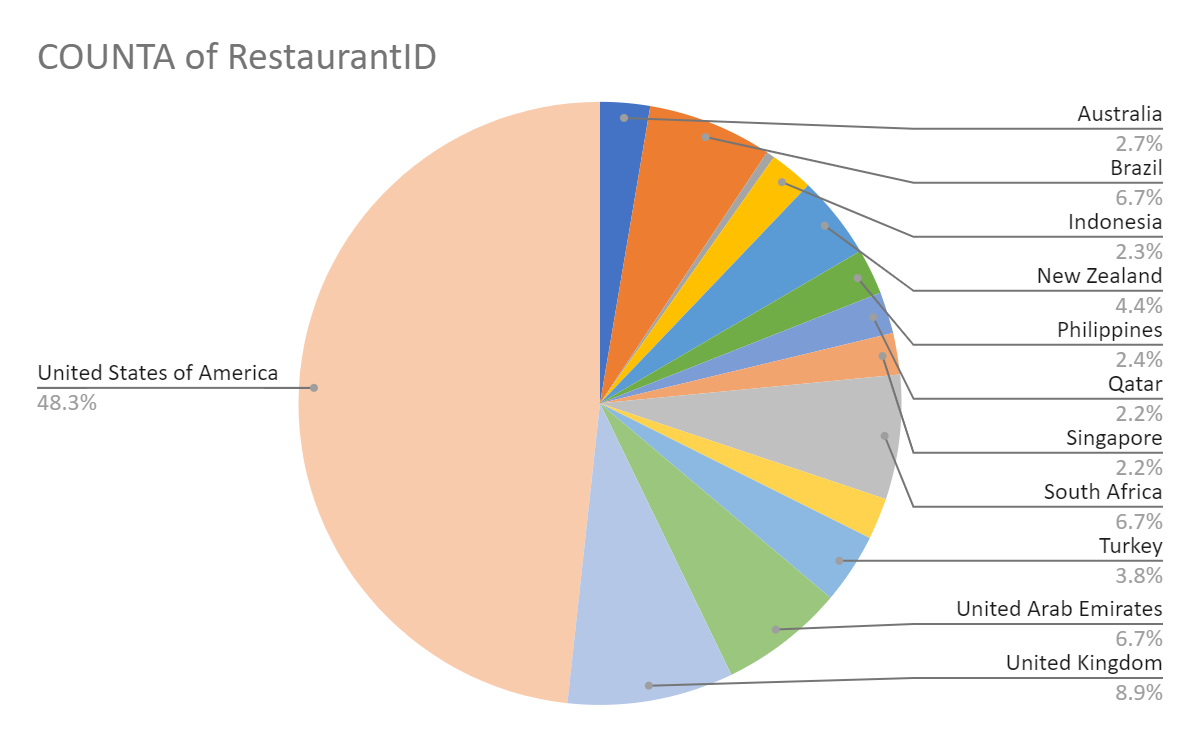
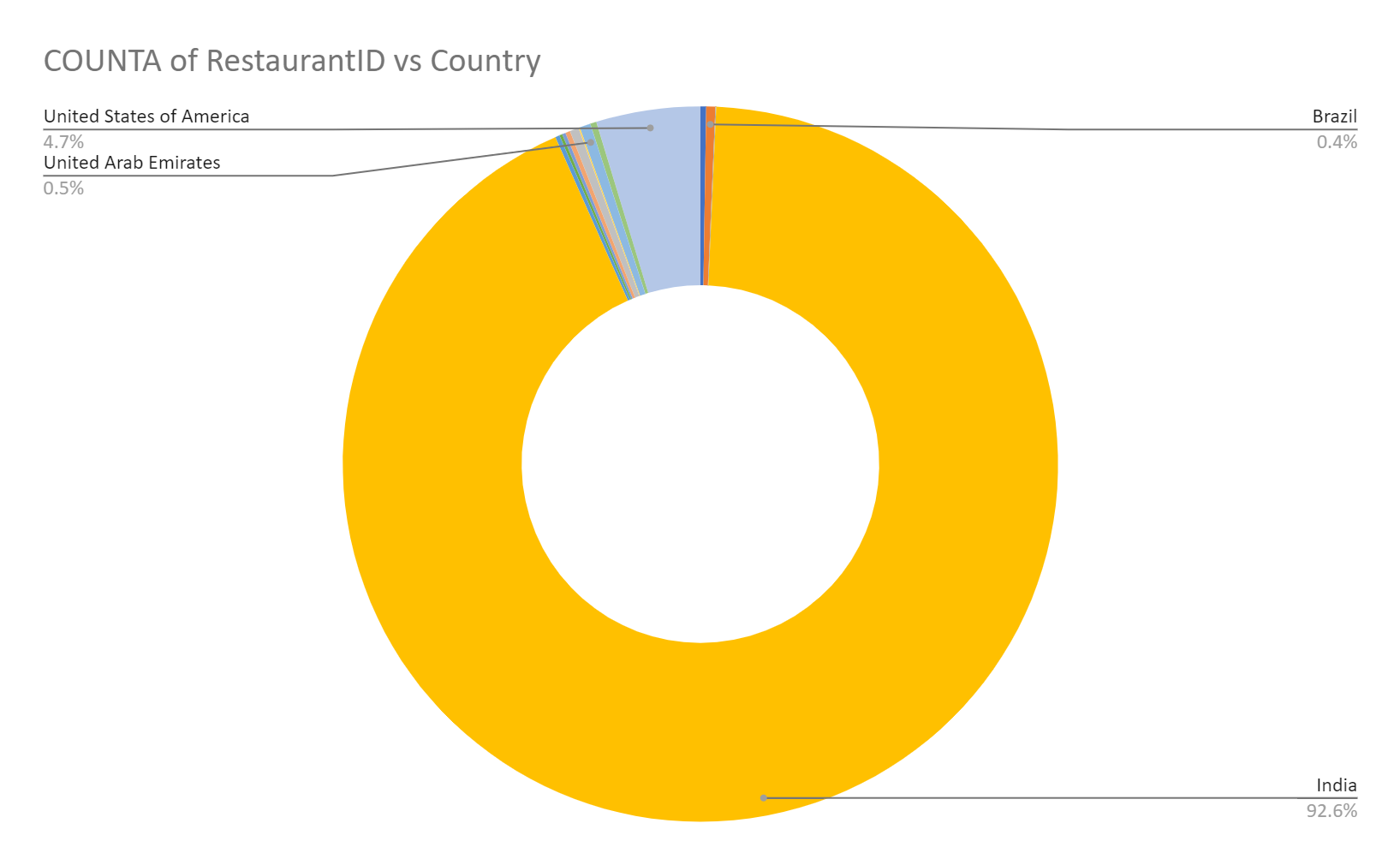
**Subjective Question:**

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?**

Guidelines:Countries are suggested on the basis of the spread/clustering of the total number of restaurants in their respective cities. A smaller cluster with a count of less than 10 restaurants per city could imply a market with lesser competition.

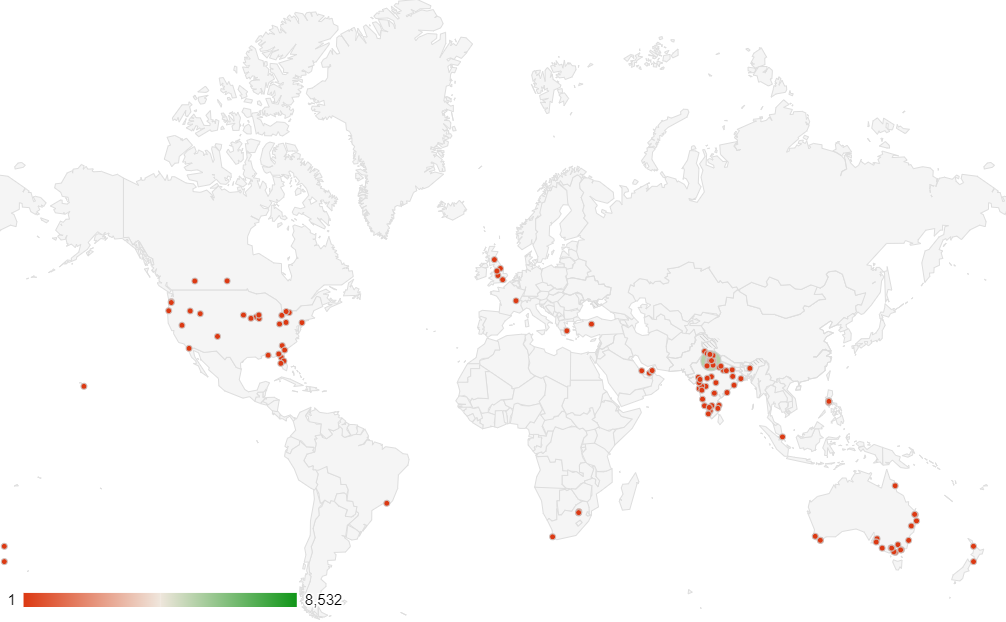
* Further, restaurants are filtered out by setting a criteria for average rating < 4. Assuming that ratings are directly indicative of customer satisfaction with the current choices of restaurants

We can utilise a pivot table to display the total number of restaurants in each country. This information can be displayed using a pie chart, allowing for a quick access of market share across different countries. To analyze deeper, we can create a Geo chart with markers to get a city specific data across a map

**Visualization:**

**Observation:**

* Dominant market: India represents Zomato's largest market worldwide with a restaurant share of approx. 92% with a total of 8652 restaurants, spanning across 43 cities. To expand in a market with less competition, zomato would need to either explore newer cities within India, or explore other unsaturated markets.
* Unsaturated markets: Outside of India, the USA captures about 48% of the zomato’s market with 434 openings spanning across 35 cities. Hence, a more detailed demographic analysis would be required to explore both India and the USA as a possible market.
* All other countries could represent a possible opening for Zomato to expand its market share where there might not be a stiff competition.
* Alternate approach: Although India is the most dominant market, new opportunities could be presented here. However, a more detailed market penetration analysis would be required here.
* Digging deeper: City and the count of restaurant ids in each city (of countries except India and the USA) were plotted on a Geo chart with markers to see the extent/congregation of the places where most restaurants are situated in each country.



Since India is the most dominant market, and USA a dominant market outside India, we will put a filter on our pivot table to remove both of them from our Geo chart.

* Canada, Australia, and the Philippines are the only countries that do not have any city with a large cluster of restaurants together in them (Barring a few cities in other countries).
* Further, filtering with an average rating of < 4 cities yields the following cities (17 in Australia, 3 in Canada and 1 each in Indonesia and the Philippines) where the count of restaurants is less than 10 AND average rating of restaurants per city < 4.

| ***Country*** | ***City*** | ***Has\_Table\_booking*** | ***Has\_Online\_delivery*** | **COUNTA of RestaurantID** | **AVERAGE of Rating** | **AVERAGE of Avg. Cost for two\_Currency (in Rs.)** |
| --- | --- | --- | --- | --- | --- | --- |
| Australia | Armidale | No | No | 1 | 3.5 | 1654.2 |
| Australia | Balingup | No | No | 1 | 3.2 | 1654.2 |
| Australia | Dicky Beach | No | No | 1 | 3.6 | 578.97 |
| Australia | Flaxton | No | No | 1 | 3.5 | 2481.3 |
| Australia | Forrest | No | No | 1 | 3.7 | 1654.2 |
| Australia | Hepburn Springs | No | No | 2 | 3.8 | 1116.585 |
| Australia | Inverloch | No | No | 1 | 3.7 | 578.97 |
| Australia | Lakes Entrance | No | No | 1 | 3.8 | 578.97 |
| Australia | Lorn | No | No | 1 | 3.6 | 1654.2 |
| Australia | Macedon | No | No | 1 | 3.5 | 1654.2 |
| Australia | Mayfield | No | No | 1 | 2.9 | 1654.2 |
| Australia | Middleton Beach | No | No | 1 | 3.8 | 2481.3 |
| Australia | Montville | No | No | 1 | 2.4 | 2481.3 |
| Australia | Paynesville | No | No | 1 | 2.6 | 9925.2 |
| Australia | Penola | No | No | 1 | 3.4 | 1654.2 |
| Australia | Phillip Island | No | No | 1 | 3.7 | 1654.2 |
| Australia | Victor Harbor | No | No | 1 | 3.6 | 1654.2 |
| Canada | Chatham-Kent | No | No | 1 | 3.7 | 2067.75 |
| Canada | Consort | No | No | 1 | 3.0 | 2067.75 |
| Canada | Yorkton | No | No | 1 | 3.3 | 2067.75 |
| Indonesia | Bogor | No | No | 2 | 3.9 | 848 |
| Philippines | Santa Rosa | No | No | 2 | 3.8 | 5032.5 |

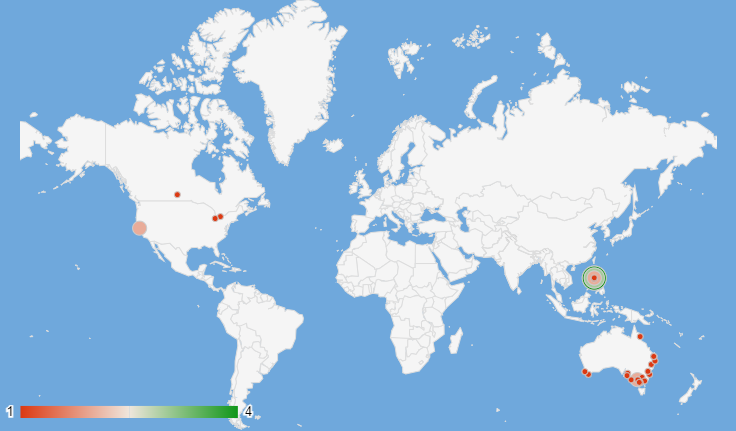
* **Hence, the above-mentioned cities in Australia, Canada, Indonesia and the Philippines are where management could explore possibilities of opening newer restaurants with fewer competition.**
  + **All the above suggested cities do not offer either the table booking option or the online delivery option. This can also be further explored as data suggests that overall, restaurants that offer both online delivery and table booking tend to have a higher rating on an average (Refer Q7 - subjective)**

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

Guidelines: Country and cities are suggested on the basis of the current list of restaurants where market competition is less (restaurants per city > 10 with an average rating per city < 4) with a higher growth potential of new openings.

Visualisation:

All presently listed cities in Canada and Australia where a restaurant already operates could be explored more for opening up of new restaurants, as all of them only operate one restaurant at present.



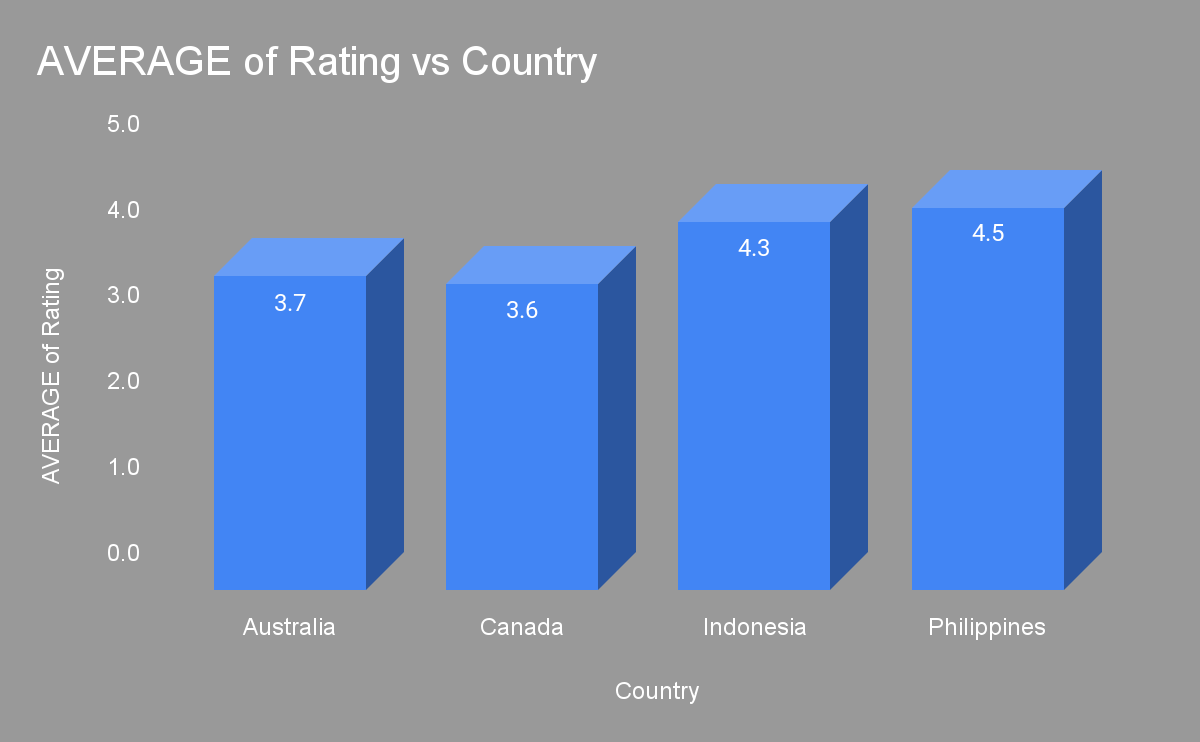
**Observations**:

Almost all the cities in both Canada and Australia have not more than 1 restaurant currently operating within its boundaries. The Philippines too have a smaller cluster (not more than 4) of restaurants in each city. Filtering all the restaurants by count per city < 10 and average rating < 4 gives us the following names of cities that can be taken up for further market for opening up of new restaurants.

| ***Country*** | ***City*** | ***Has\_Table\_booking*** | ***Has\_Online\_delivery*** | **COUNTA of RestaurantID** | **AVERAGE of Rating** | **AVERAGE of Avg. Cost for two\_Currency (in Rs.)** |
| --- | --- | --- | --- | --- | --- | --- |
| **Australia** | **Armidale** | No | No | 1 | 3.5 | 1654.2 |
| Australia | **Balingup** | No | No | 1 | 3.2 | 1654.2 |
| Australia | **Dicky Beach** | No | No | 1 | 3.6 | 578.97 |
| Australia | **Flaxton** | No | No | 1 | 3.5 | 2481.3 |
| Australia | **Forrest** | No | No | 1 | 3.7 | 1654.2 |
| Australia | **Hepburn Springs** | No | No | 2 | 3.8 | 1116.585 |
| Australia | **Inverloch** | No | No | 1 | 3.7 | 578.97 |
| Australia | **Lakes Entrance** | No | No | 1 | 3.8 | 578.97 |
| Australia | **Lorn** | No | No | 1 | 3.6 | 1654.2 |
| Australia | **Macedon** | No | No | 1 | 3.5 | 1654.2 |
| Australia | **Mayfield** | No | No | 1 | 2.9 | 1654.2 |
| Australia | **Middleton Beach** | No | No | 1 | 3.8 | 2481.3 |
| Australia | **Montville** | No | No | 1 | 2.4 | 2481.3 |
| Australia | **Paynesville** | No | No | 1 | 2.6 | 9925.2 |
| Australia | **Penola** | No | No | 1 | 3.4 | 1654.2 |
| Australia | **Phillip Island** | No | No | 1 | 3.7 | 1654.2 |
| Australia | **Victor Harbor** | No | No | 1 | 3.6 | 1654.2 |
| **Canada** | **Chatham-Kent** | No | No | 1 | 3.7 | 2067.75 |
| Canada | **Consort** | No | No | 1 | 3.0 | 2067.75 |
| Canada | **Yorkton** | No | No | 1 | 3.3 | 2067.75 |
| **Indonesia** | **Bogor** | No | No | 2 | 3.9 | 848 |
| **Philippines** | **Santa Rosa** | No | No | 2 | 3.8 | 5032.5 |

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

**Visualization:**



**Observations**:

Avg ratings (Australia) = 3.7

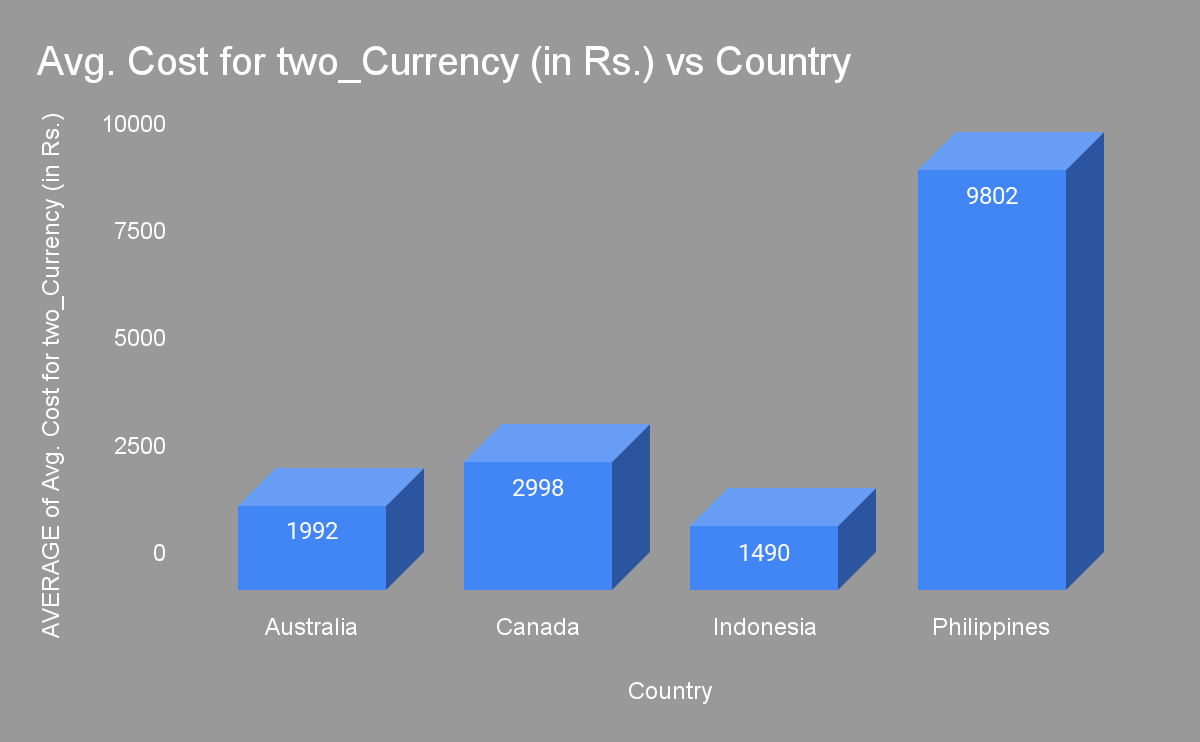
Avg. rating (Canada) = 3.6

Avg. rating (Philippines) = 4.5

Avg. rating (Indonesia) = 4.3

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

**Visualization:**

****

**Observations**:

Avg. price in Indian currency (Australia) = 1992

Avg. price in Indian currency (Canada) = 2998

Avg. price in Indian currency (Philippines) = 9802

Avg. price in Indian currency (Indonesia) = 1490

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**

Guidelines: Any restaurant with a rating >= 4 & >=3 to <4 will be considered rated as the one that is considered to provide an excellent & good experience respectively to its customers.

* Restaurants with a lower price rating and a higher rating would also be considered as a strong competitor

**Visualization**:

| ***Country*** | ***City*** | ***RestaurantName*** | **AVERAGE of Rating** |
| --- | --- | --- | --- |
| Australia | Armidale | Whitebull Hotel | 3.5 |
| Australia | Balingup | Taste of Balingup | 3.2 |
| Australia | Dicky Beach | The Giggling Goat | 3.6 |
| Australia | Flaxton | Flaxton Gardens | 3.5 |
| Australia | Forrest | Bespoke Harvest | 3.7 |
| Australia | Hepburn Springs | Blue Bean Love Cafe | 3.8 |
| Australia | Hepburn Springs | La Trattoria of Lavandula | 3.8 |
| Australia | Inverloch | Beach Box Cafe | 3.7 |
| Australia | Lakes Entrance | Funkey Monkey | 3.8 |
| Australia | Lorn | Stillwater on Belmore | 3.6 |
| Australia | Macedon | Mr. | 3.5 |
| Australia | Middleton Beach | Three Anchors | 3.8 |
| Australia | Penola | DiVine | 3.4 |
| Australia | Phillip Island | Mad Cowes Cafe | 3.7 |
| Australia | Victor Harbor | Anchorage Cafe Restaurant Wine Bar | 3.6 |
| Canada | Chatham-Kent | Tokyo Sushi | 3.7 |
| Canada | Consort | Consort Restaurant | 3 |
| Canada | Yorkton | Arigato Sushi | 3.3 |
| Indonesia | Bogor | Lemongrass | 4 |
| Indonesia | Bogor | Momo Milk | 3.7 |

**Observations**:

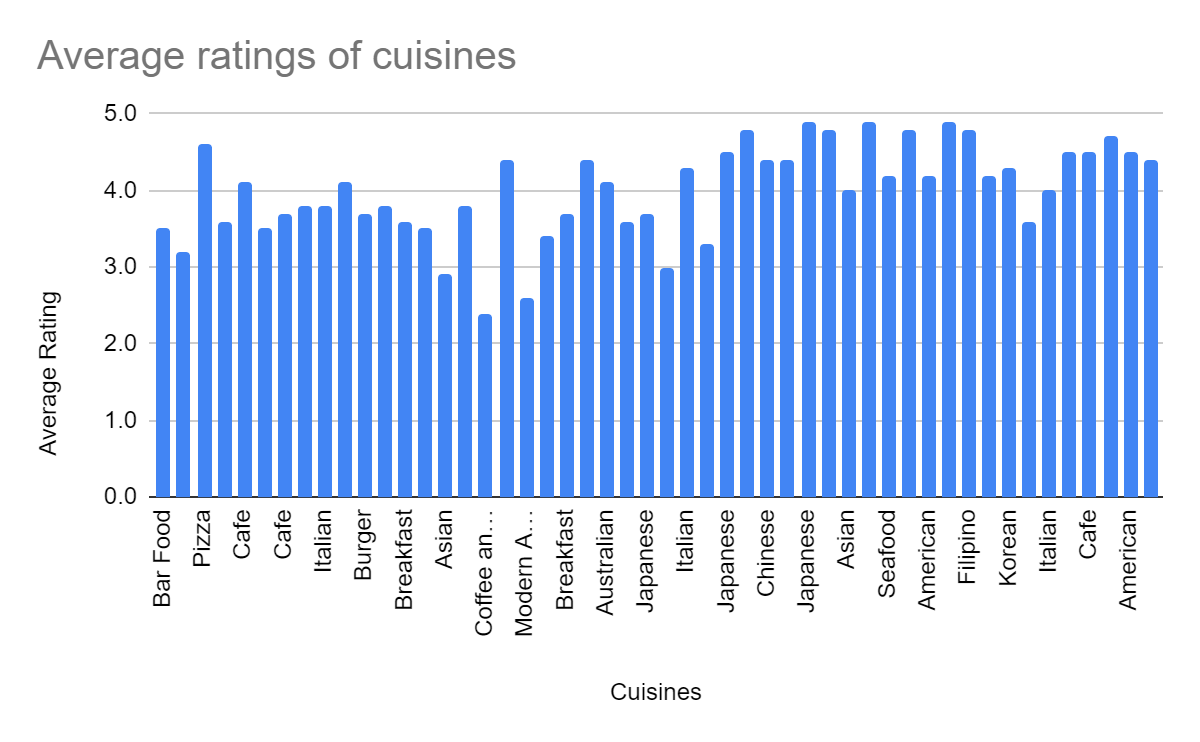
* There is only one restaurant in the suggested cities with a rating >=4 i.e. Lemongrass (rating = 4) from Bogor, Indonesia
  + Bogor is the only city amongst the suggested cities within Indonesia, hence a review of Lemongrass’ strengths needs to be evaluated in detail as compared to the other restaurant (Momo milk; rating = 3.7)
* Other listed restaurants in the table are to considered as competitors as they lie in the higher range of reviews amongst their customers
  + An analysis of their price range, cuisines offered and other matrices is desirable to chart our their strengths
* Restaurants in the lower bracket of ratings, i.e. 1-2 & 2-3 (upper limit exclusive) include:

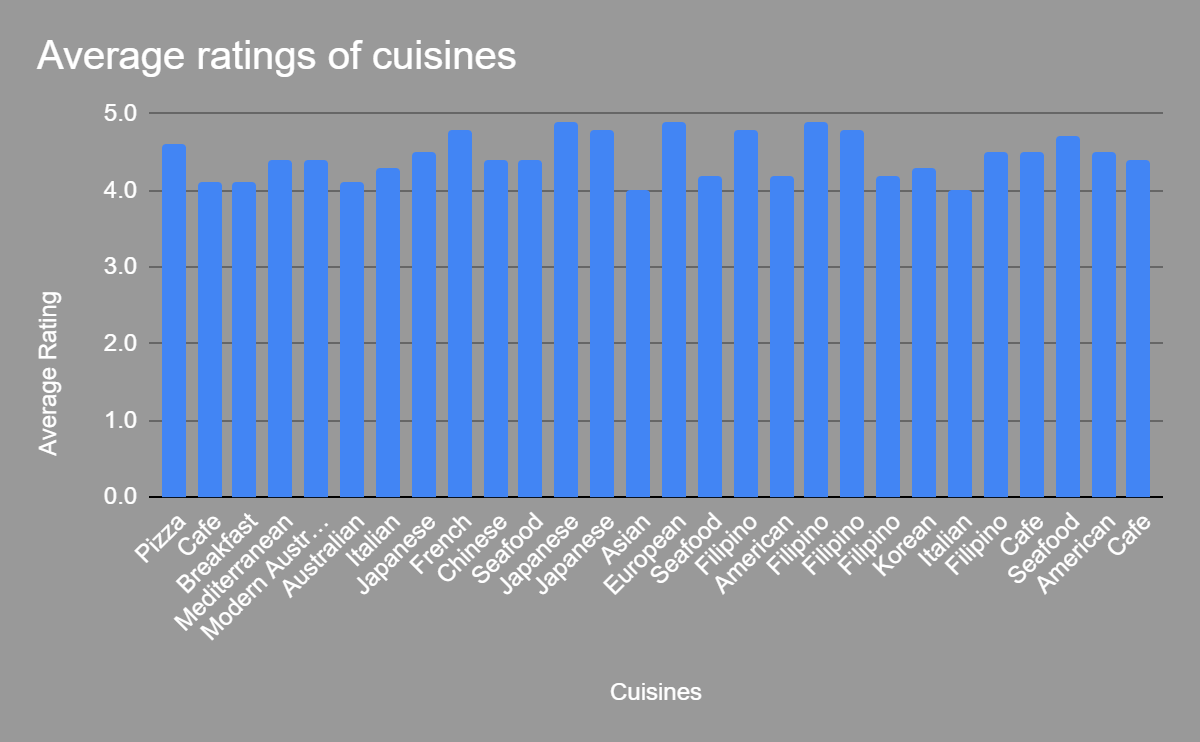
| ***Country*** | ***City*** | ***RestaurantName*** | **AVERAGE of Rating** |
| --- | --- | --- | --- |
| Australia | Mayfield | Star Buffet | 2.9 |
| Australia | Montville | Poets Cafe | 2.4 |
| Australia | Paynesville | Pier 70 | 2.6 |

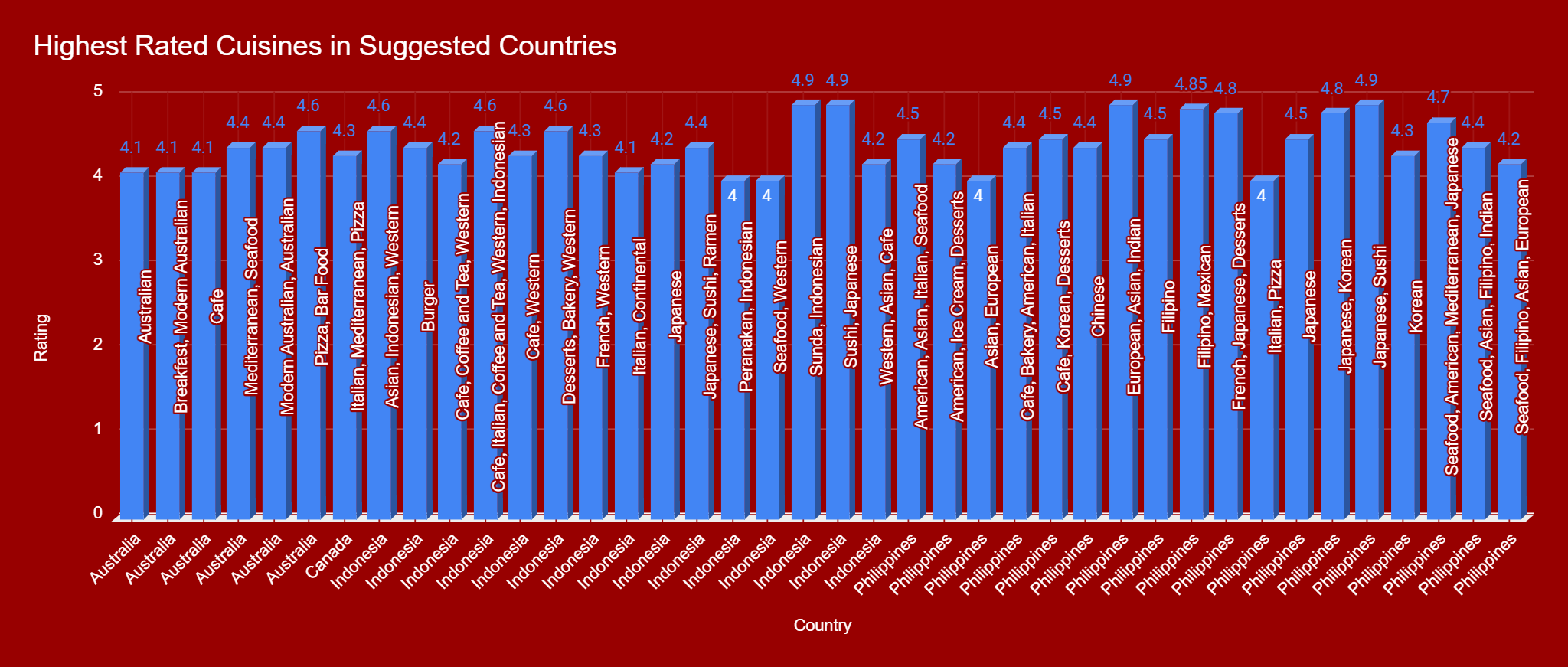
1. **Which cuisines should we focus on in the newer restaurants to get better feedback? Does the choice of cuisines affect the restaurant ratings?**

Guidelines: Cuisines list in the countries of interest were can be plotted against rating to see if choice of cuisine has any effect on the ratings.

**Visualization**:







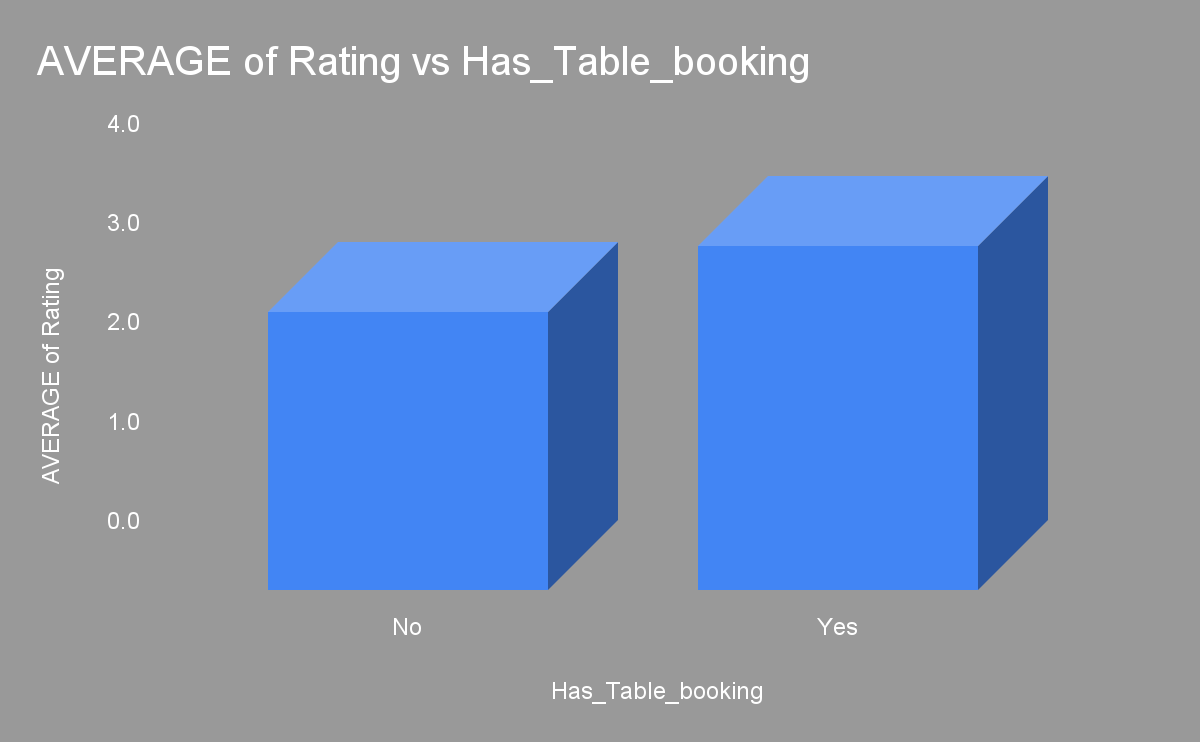
**Observations**:

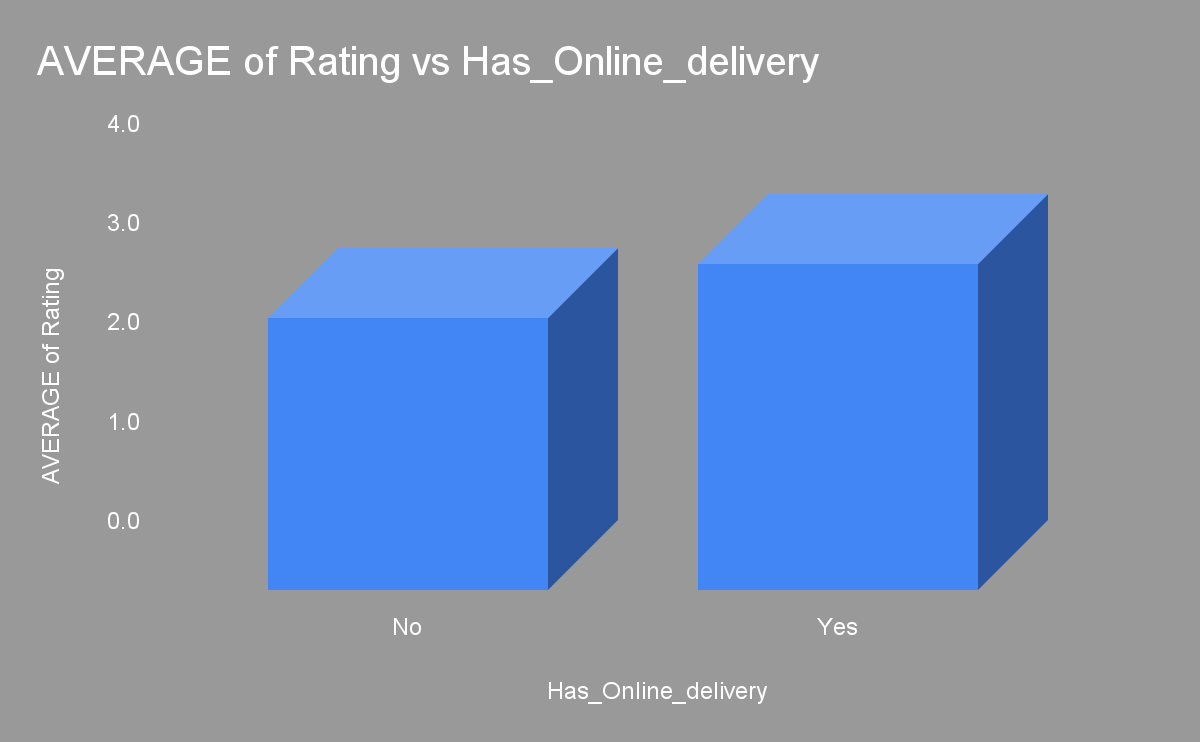
Cuisines list was split into different columns to be considered as one separate entity. This can be plotted against the average rating column. As can be seen, certain cuisines have lower ratings that other

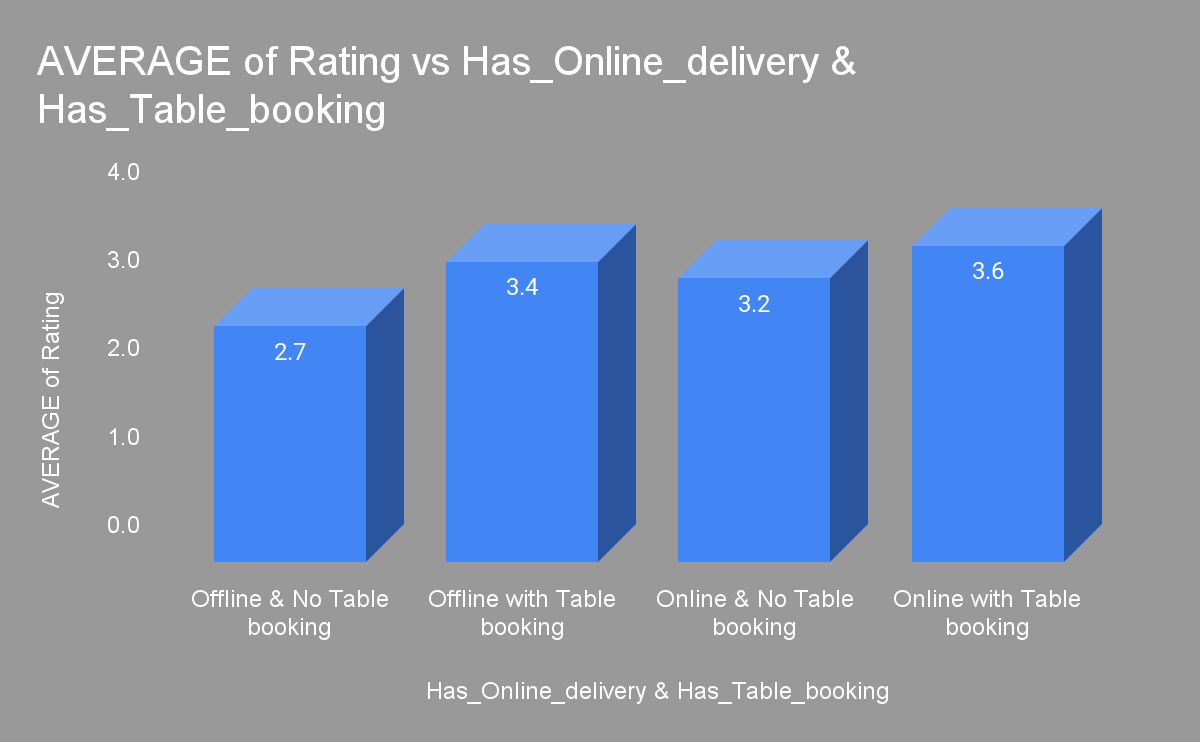
* Restaurants that included a choice of cuisine with a rating of >=4 could be considered as the ones that are most loved by the customers

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

Guidelines:To see the effect of table booking and online delivery on ratings, we can stack the average ratings graph with the choice of having table booking and online delivery.





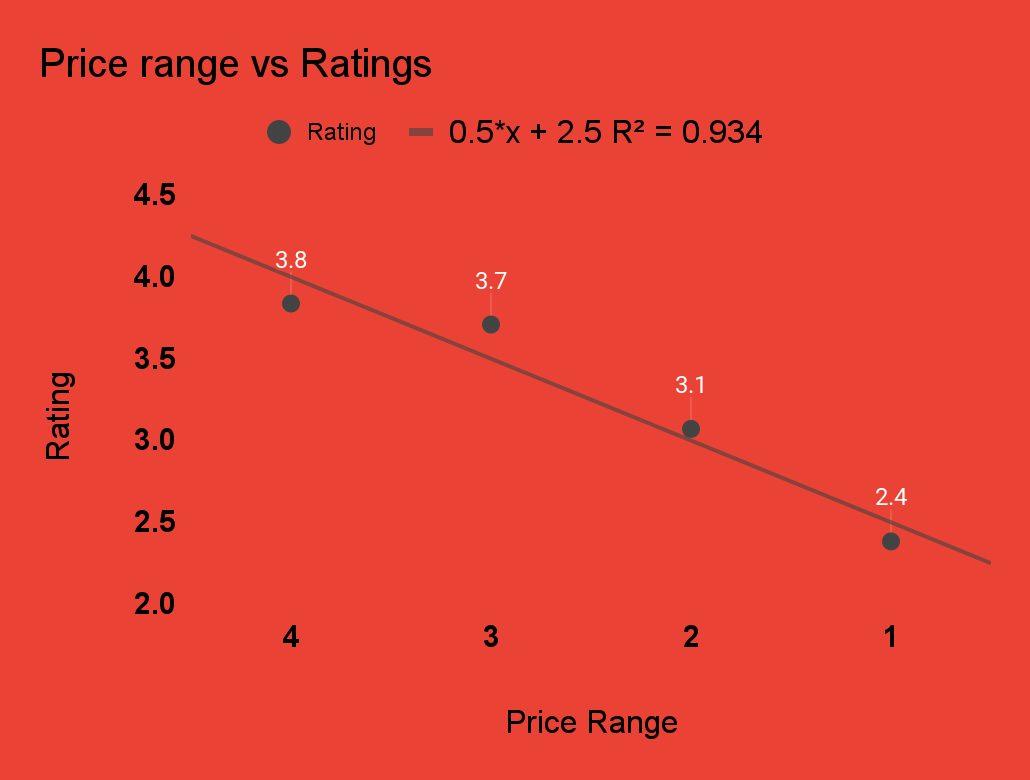


**Observations**:The restaurants that allowed both an option of table booking and online delivery were generally considered to be amongst the highest rated on an average

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?**

**Guidelines:** Price range can be plotted against ratings chart to see their effect on the ratings

**Visualisation**



**Observations**:

There is a strong correlation between average rating vs the price range of restaurants in general.

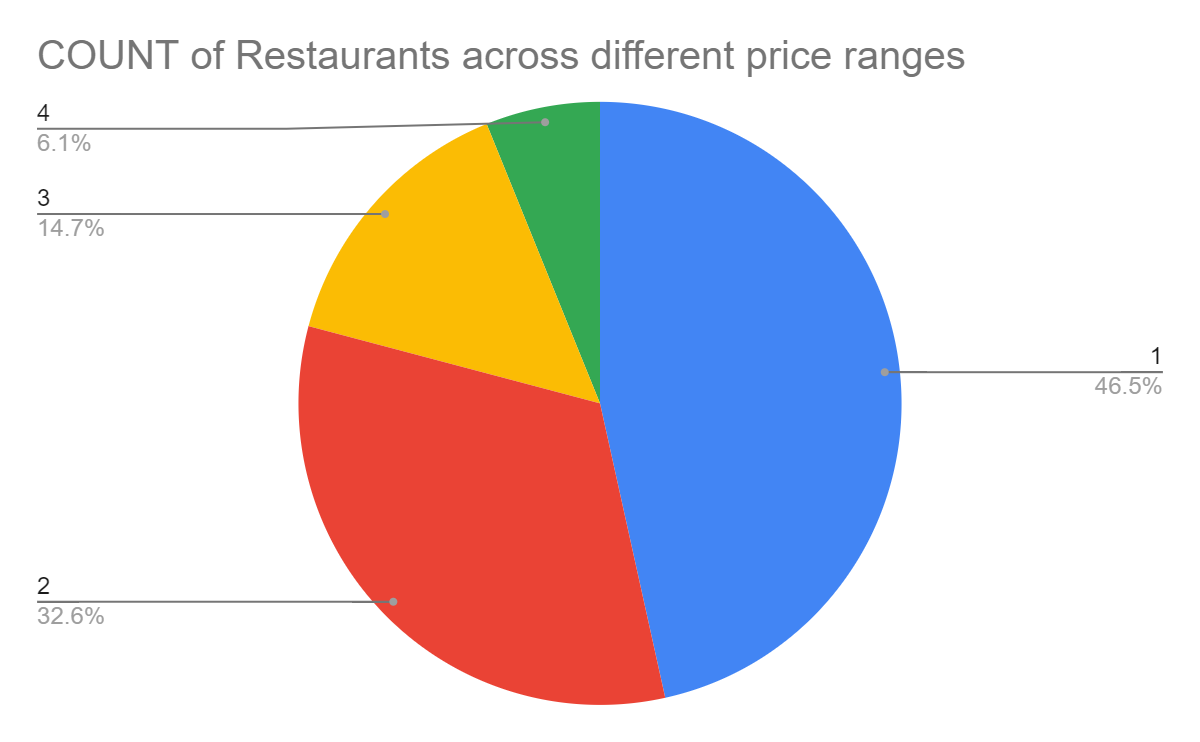
* On average, the price range of restaurants are directly proportional to their rating

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

**Guidelines:** To find the distribution of restaurants of different price ranges across all countries, we can create a pivot table by putting price range in rows and countA of restaurant ids in values

**Visualisation**:

| *Price\_range* | COUNTA of RestaurantID |
| --- | --- |
| 1 | 4444 |
| 2 | 3113 |
| 3 | 1408 |
| 4 | 586 |



**Observation**:

Price range of 1 represents the highest number of restaurants (4444 restaurants, approx. 47% of total) across countries, followed by price range 2 (3113 restaurants, ~33%), price range 3 (1408 restaurants, ~15%) & price range 1 (586 restaurants, ~6%) 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. **[you have to give bullet pointers in order to answer this question]**

To suggest opening of new restaurants we may follow the following procedure:

* Segregating restaurants country-wise to check for the number of restaurants present in each country
* Checking for the average price and rating and to see if they have any correlation
* To see if the average cost for two has any effect on the number of restaurants and their ratings to identify countries with a lower number of restaurants with lower average costs and perhaps higher rating(s).
* Geomarking the number of restaurants city-wise to check for clustering of restaurants in different cities and countries.
* To correlate whether different attributes like price range, rating, average cost for two have any effect on the number of restaurants or on each other to identify potential delimiting factors in opening up of new restaurants
* To identify restaurants in the countries suggested having very high rating, and if they have lower costs or not
* In countries with lower clustering, to obtain population data from external sources to check if major urban centres with higher per capita spending are already covered in the raw data or not.