**** *analysis*

**by Yash Kumar**

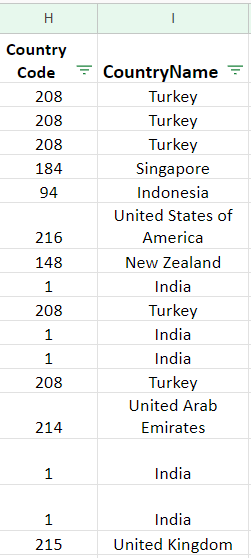
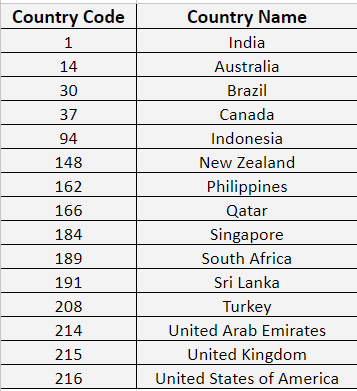
**Excel file link:** [Zomato\_Analysis](https://docs.google.com/spreadsheets/d/1dIbOZeXjQyoLh3JodiFN7BLuKm1bXj-ctDZxc4cTaxQ/edit?usp=sharing) **Objective Ones**

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

* The DateKey\_Opening had dates in inconsistent format. So I fixed it using ‘Find and Replace’ in that column and applied date format generally used in India. Also extracted Year and Month from that.
* Fixed name of city ÛÁstanbul (in Turkey) to Istanbul in all of the sheets.
* Missing values in the Cuisine column have been assigned with ‘No data’.
* Formatted RawData dataset in a proper way (by grouping columns) for general information to be seen.
* Rest of attributes have been grouped and collapsed and can be accessed by clicking **+** in RawData sheet.
* Conditional formatting to highlight other changes made in data or any new data created in the Raw Data Sheet.

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

* It has been done using VLOOKUP in the Raw Data sheet, in column I and H.
* **Country Name** =VLOOKUP($H2,CountryDescription!$A$1:$B$16,2)

****

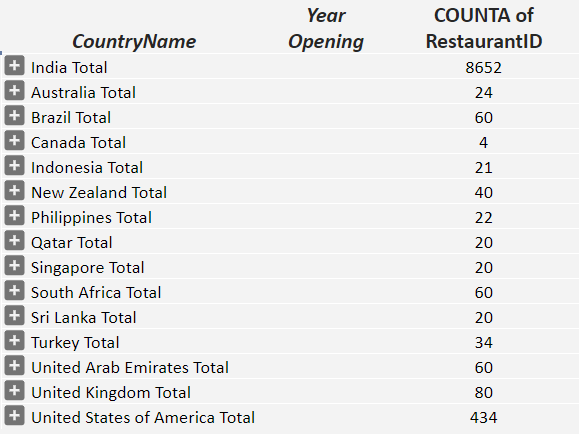
1. Create a table to represent the number of restaurants opened in each country.

* **Here’s the Table representing the number of restaurants in each country:**

| **COUNTRY Name** | **COUNT of Restaurants** |
| --- | --- |
| 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 |
| UAE | 60 |
| UK | 80 |
| USA | 434 |

* It can be seen in pivot tables in Dashboard or Pivot Table (Global) sheet of excel file.

***It has been done using the COUNTA function for RestaurantID as values in the pivot table.***

****

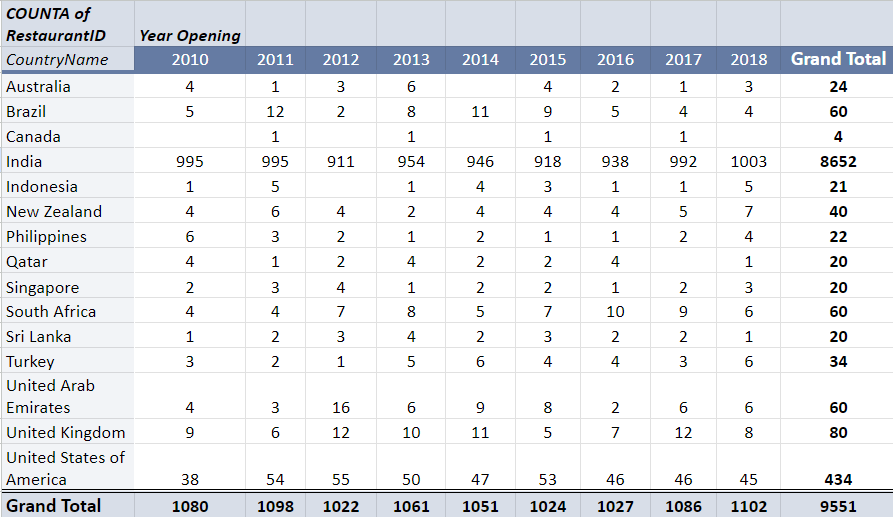
1. Also, the management wants to look at the number of restaurants opened each year, so provide them with something here.

* **Here’s the table to represent no. of restaurants each year:**

| **COUNTRY** | **YEAR** | | | | | | | | | **Total** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** |
| **Australia** | 4 | 1 | 3 | 6 | 0 | 4 | 2 | 1 | 3 | **24** |
| **Brazil** | 5 | 12 | 2 | 8 | 11 | 9 | 5 | 4 | 4 | **60** |
| **Canada** | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | **4** |
| **India** | 995 | 995 | 911 | 954 | 946 | 918 | 938 | 992 | 1003 | **8652** |
| **Indonesia** | 1 | 5 | 0 | 1 | 4 | 3 | 1 | 1 | 5 | **21** |
| **New Zealand** | 4 | 6 | 4 | 2 | 4 | 4 | 4 | 5 | 7 | **40** |
| **Philippines** | 6 | 3 | 2 | 1 | 2 | 1 | 1 | 2 | 4 | **22** |
| **Qatar** | 4 | 1 | 2 | 4 | 2 | 2 | 4 | 0 | 1 | **20** |
| **Singapore** | 2 | 3 | 4 | 1 | 2 | 2 | 1 | 2 | 3 | **20** |
| **South Africa** | 4 | 4 | 7 | 8 | 5 | 7 | 10 | 9 | 6 | **60** |
| **Sri Lanka** | 1 | 2 | 3 | 4 | 2 | 3 | 2 | 2 | 1 | **20** |
| **Turkey** | 3 | 2 | 1 | 5 | 6 | 4 | 4 | 3 | 6 | **34** |
| **UAE** | 4 | 3 | 16 | 6 | 9 | 8 | 2 | 6 | 6 | **60** |
| **UK** | 9 | 6 | 12 | 10 | 11 | 5 | 7 | 12 | 8 | **80** |
| **USA** | 38 | 54 | 55 | 50 | 47 | 53 | 46 | 46 | 45 | **434** |

**To see this data in excel file:**

* Go to the Obj/Sbj sheet in the excel file.
* Then go to the first pivot table
* Once clicked, we can get each year's distribution as seen below

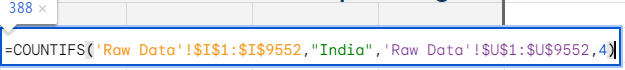
****

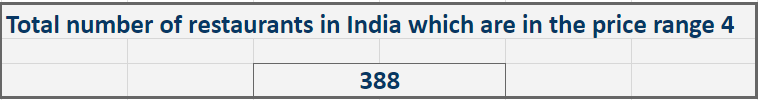
1. What is the total number of restaurants in India which are in the price range 4?

* **Total such restaurants are = 388**

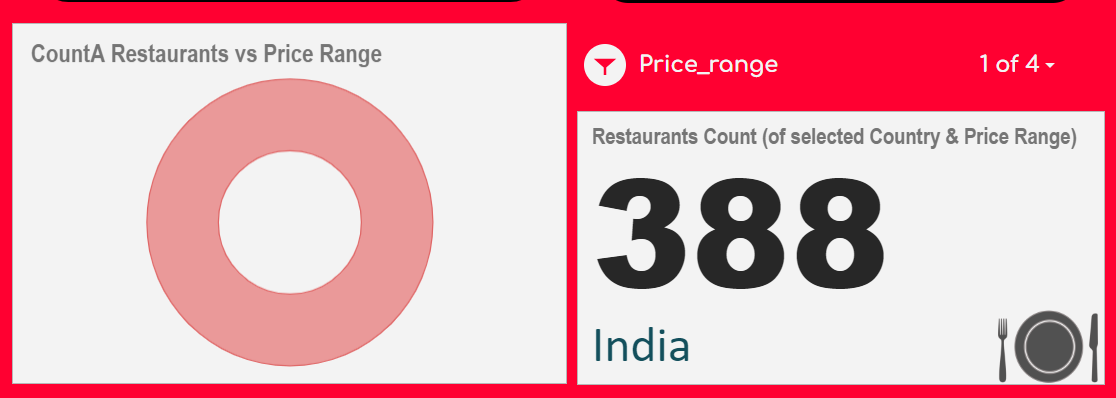
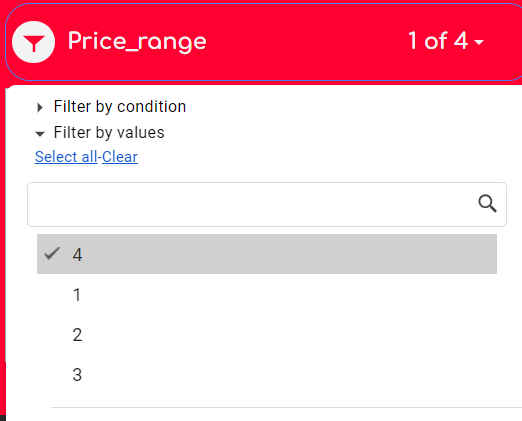
**Formulae Approach:**We can use a conditional aggregation function to get the result.  
Thus, here we used **COUNTIFS** for multiple criteria.

* **No. of Restaurants** =COUNTIFS('Raw Data'!$I$1:$I$9552,"India",'Raw Data'!$U$1:$U$9552,4)





**Chart Approach:**

* Open Excel file
* Go to Dashboard sheet
* Find pivot table related to Price Range
* Click on **+** icon for country to be searched for
* From Slicer above the Scorecard, apply filter ****

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

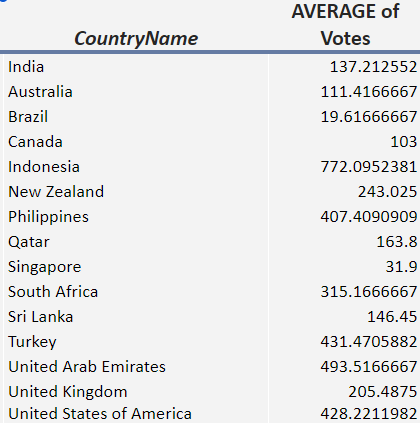
* **Here’s the table to represent no. of average voters in each country:**

| **COUNTRY Name** | **Average of VOTES** |
| --- | --- |
| Australia | 111 |
| Brazil | 20 |
| Canada | 103 |
| India | 137 |
| Indonesia | 772 |
| New Zealand | 243 |
| Philippines | 407 |
| Qatar | 164 |
| Singapore | 32 |
| South Africa | 315 |
| Sri Lanka | 147 |
| Turkey | 432 |
| UAE | 494 |
| UK | 206 |
| USA | 428 |

**To achieve this:**

* Go to Dashboard sheet
* Find pivot table related to average votes
* And we can see the table below.

***It has been done using the Average function for Votes as values in the pivot table.***

******

**Subjective Ones**

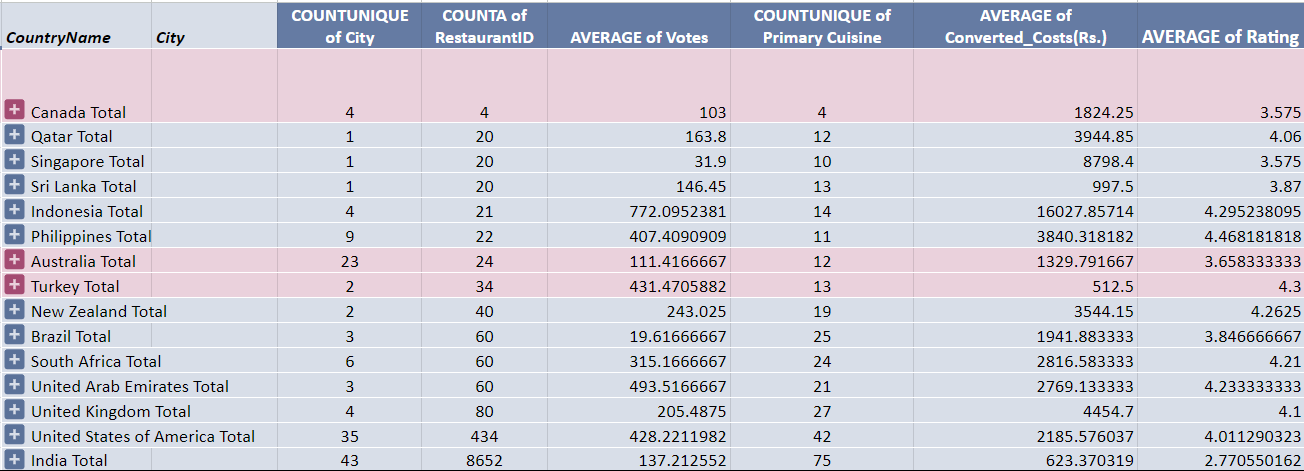
**Reference sheet:** [Zomato\_Analysis](https://docs.google.com/spreadsheets/d/1dIbOZeXjQyoLh3JodiFN7BLuKm1bXj-ctDZxc4cTaxQ/edit#gid=333726490&range=A1:I16)

1. Suggest a few countries where the team can open newer restaurants with lesser competition. Which visualization/technique will you use here in order to justify the suggestions?

* Australia, Canada and Turkey as major 3, as they have
* comparatively less number of restaurants
* comparatively less variety of cuisines (especially when compared to the number of Indian Cuisines).
* Significant number of Voters (people) to try with.
* Reasonable amount of Cost cut in that area.

| **Country** | **No. of Restaurants** | **Unique Cuisines** | **Average Votes** |
| --- | --- | --- | --- |
| **Australia** | 24 | 12 | 111 |
| **Canada** | 4 | 4 | 103 |
| **Turkey** | 34 | 13 | 432 |

* Pivot Tables and Simple Column Chart as we are showing this in a categorical way of countries.

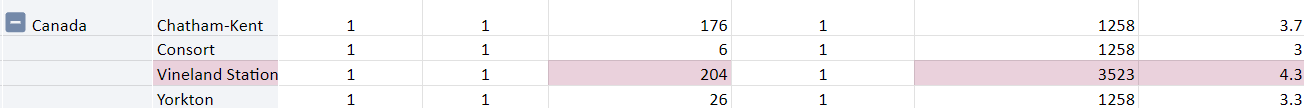
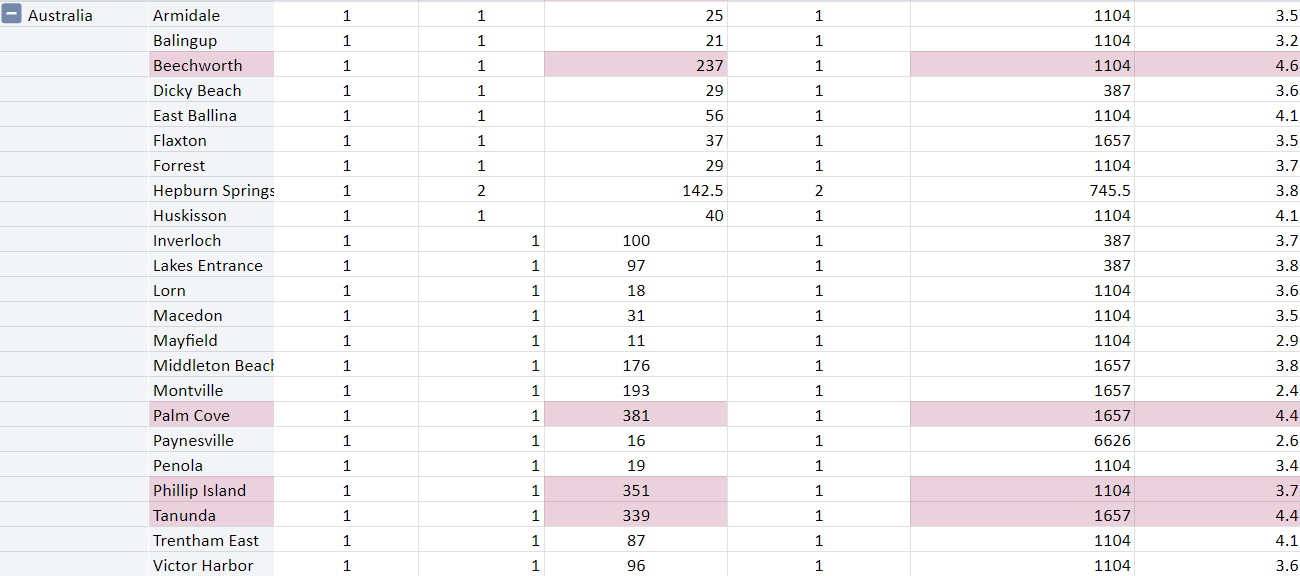
******

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

* For this task, we used conditional formatting in the Pivot Tables (Global) sheet.
* Sorted on the basis of**:**
* Average Voters: **>=200**
* Cost Range: **500 - 2000** (Indian Rupees)

*which means people are ones who like trying out new restaurants and new experiences given a reasonable cost range.*

| **Country** | **City** |
| --- | --- |
| **Australia** | Beechworth |
| Palm Cove |
| Phillip Island |
| Tanunda |
| **Canada** | Vineland Station |
| **Turkey** | Istanbul |

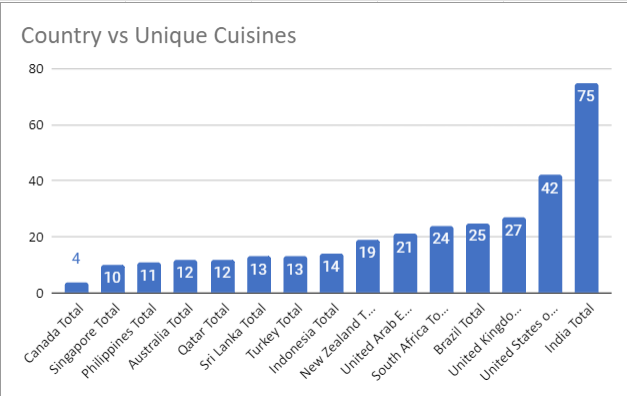
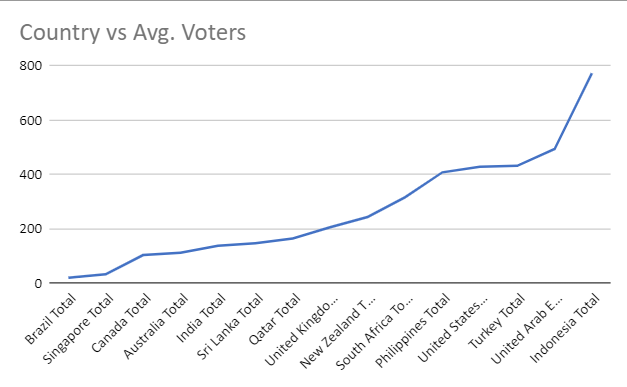
******

1. Name the chart/spreadsheet function you will use for solving the problem?

* We used **Pivot Tables**, **Column & Line charts**, with suitable sorting and formatting.

We can select suitable attributes in pivot tables like:

* *Rating*
* *Cost*
* *Voters*
* *Count of Restaurants*  
    
  to categorize the data in pivot table and then format accordingly and visualize using bar charts to get insight.

******

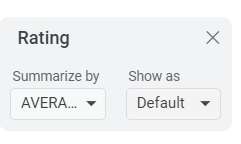
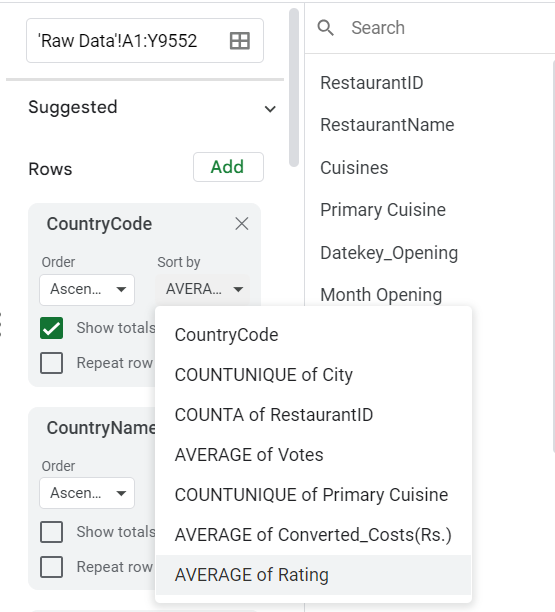
1. According to the countries you suggested, what is the current quality in terms of ratings for restaurants that are opened there?

* Suggested countries rating are in a fairly good region which makes it a preferable environment to succeed by opening a Good restaurant with Good Cuisines.
* **Here’s the table representing them:**

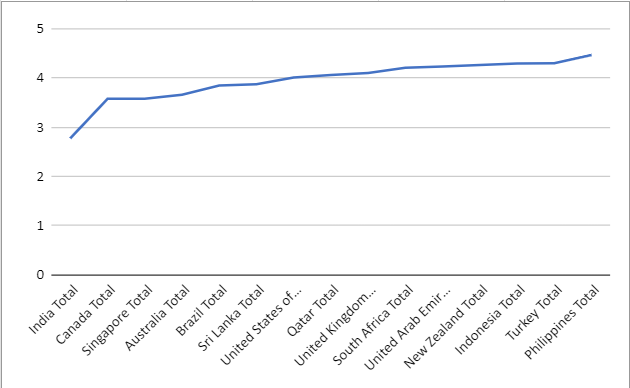
| **Country** | **Rating** |
| --- | --- |
| **Australia** | **3.66** |
| **Canada** | **3.56** |
| **Turkey** | **4.3** |

1. Will you use any aggregation function or a visualization here to solve the problem?

* Yes, we obviously need to use aggregation function, because we need to summarize large dataset and extract meaningful data
* **Functions Used:**
* **AVERAGE** to find the Average of ratings
* **COUNTA** to find Count of ratings
* and relation between them.

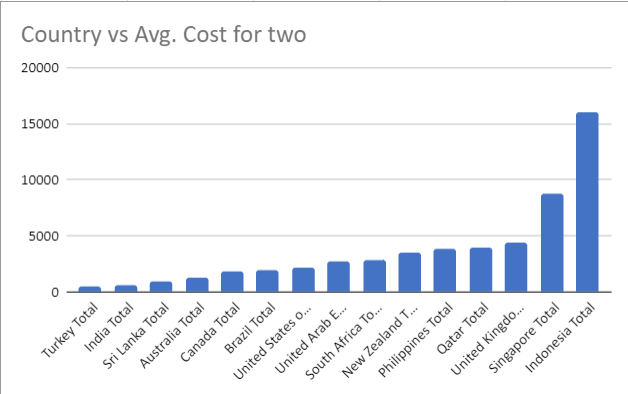


* As for visualization, Line charts are used to get it done.



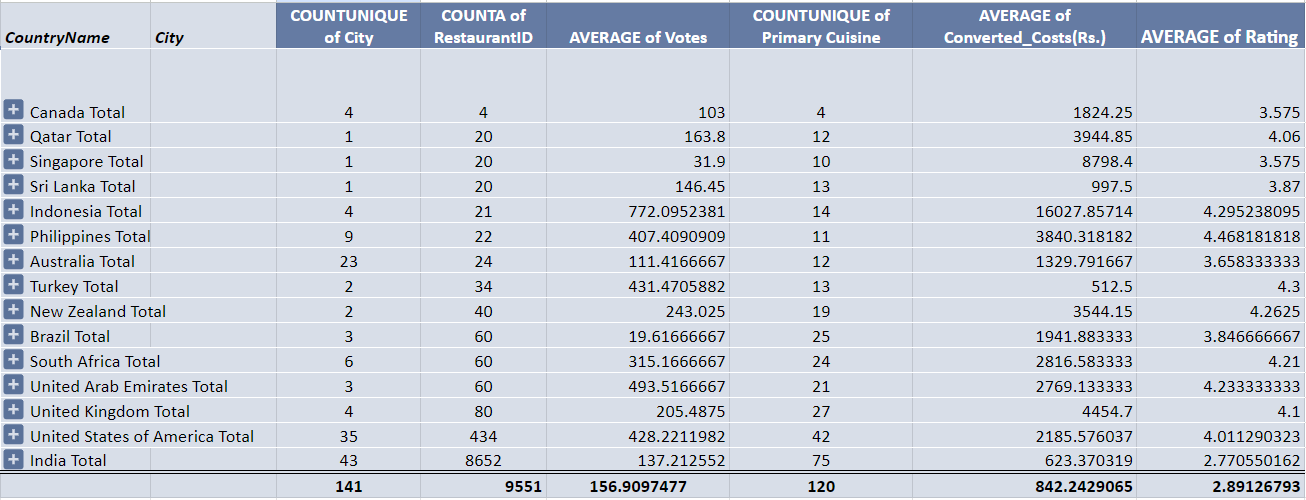
1. Also what is the current expenditure on the food in the suggested countries, so that we can keep our financial expenditure in control?

| **Country** | Roundup **Cost (Rs.)** |
| --- | --- |
| **Australia** | **1330** |
| **Canada** | **1830** |
| **Turkey** | **520** |



1. Mention the functionality which you will use for giving the suggestions, will it be any aggregate function or a visualization?

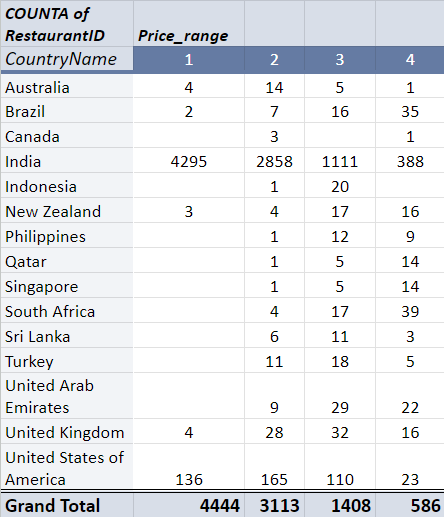
* Will use **pivot table aggregation functions** to extract the **Average\_cost\_for\_two** in each country.
* Can be seen in the **Pivot Tables(Global)** sheet (provided above as reference sheet).



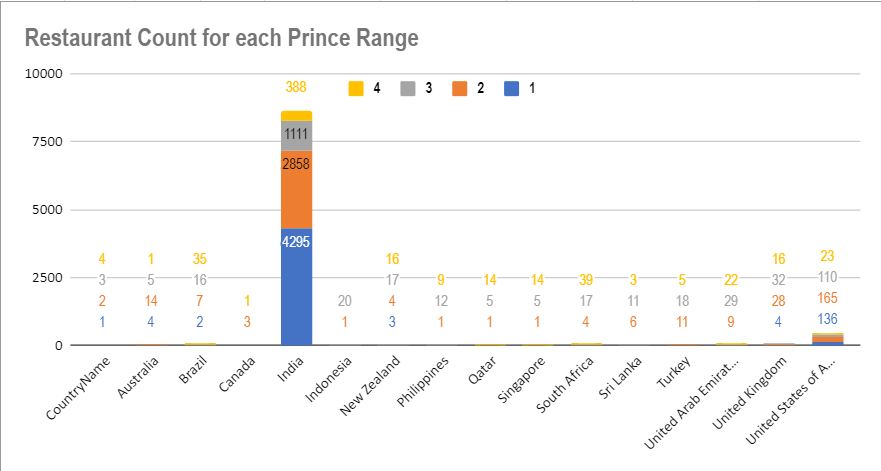
1. What is the distribution of the number of restaurants of different price ranges in all the countries?

* To create the table:
* Create Pivot Table
* Put **Row** with **CountryName**
* **Column** with **Price Range**
* **Value** with **RestaurantID**
* Here’s table representing the number of restaurants of different price ranges in all the countries**:**

| **COUNTRY** | **Price Range** | | | |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** |
| **Australia** | 4 | 14 | 5 | 1 |
| **Brazil** | 2 | 7 | 16 | 35 |
| **Canada** | 0 | 3 | 0 | 1 |
| **India** | 4295 | 2858 | 1111 | 388 |
| **Indonesia** | 0 | 1 | 20 | 0 |
| **New Zealand** | 3 | 4 | 17 | 16 |
| **Philippines** | 0 | 1 | 12 | 9 |
| **Qatar** | 0 | 1 | 5 | 14 |
| **Singapore** | 0 | 1 | 5 | 14 |
| **South Africa** | 0 | 4 | 17 | 39 |
| **Sri Lanka** | 0 | 6 | 11 | 3 |
| **Turkey** | 0 | 11 | 18 | 5 |
| **UAE** | 0 | 9 | 29 | 22 |
| **UK** | 4 | 28 | 32 | 16 |
| **USA** | 136 | 165 | 110 | 23 |
| **Total** | **4444** | **3113** | **1408** | **586** |

******

Distribution means the numbers of different price ranges, how will you show this using a chart?

* Used a stacked column chart (with Data Labels) for this task
* *Insert Chart*
* *Select Stacked Column Chart*
* *X-axis****:*** *CountA of RestaurantID*
* *Y-axis(Series)****:*** *All Price Ranges Columns*
* *Legends****:*** *Price Range*
* **Reason:** It shows perfectly the count of restaurants for each country in each price range distinctly. ******