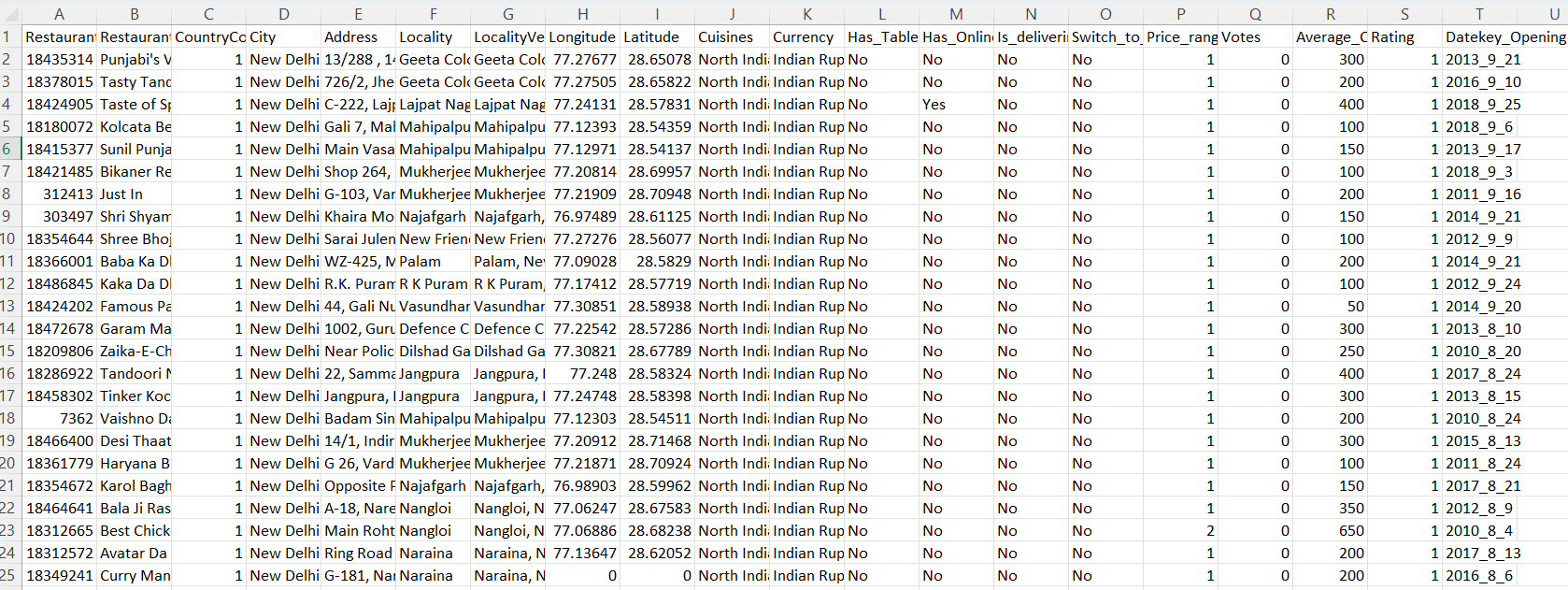
**Zomato Assignment**

**Newton School – By Arti Awasthi**

**Problem Statement**

You are hired as a consultant data analyst by Zomato where the team is looking for expansion andopening more restaurants. Your task is to come up with strategies/suggestions about opening newer restaurants.



**Tasks**

**Objective Questions**

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

**Ans** – There was only one column with a missing value, the **Cuisine** column. To solve this, I’ve created another column i.e., Cuisine Updated where I’ve replaced the missing value with the “Missing” for us to identify which column had having missing component.

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

**Ans** – Created the column “Country” next to CountryCode from the country description sub-sheet with the help of the VLOOKUP function (=VLOOKUP(C2,'country description'!$A$2:$B$16,2,0)).

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

Ans – To solve this question, I’ve created a Pivot table, where the Rows contain Country and Values as Count of Restaurant ID. This data is attached in the **Analysis** sub sheet in the Main Excel File.



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

**Ans** – To solve this, I duplicated the column “**Datekey\_Opening**” in the Raw data file and then applied split text to the column property on the column with delimiters of “\_”. With this, I have extracted the Year from the Datekey\_Opening and labelled it as **Year.** After this, I applied the Pivot table where Rows contain Year and Values contain Count of Restaurant ID. This data is attached in the **Analysis** sub sheet in the Main Excel File. Along with this, I’ve also created a Bar chart for the better representation of data on how many restaurants have opened in each year.A table with numbers and text

Description automatically generated

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

**Ans** – For this, I’ve created another Pivot Table where I placed the Country in Rows and Count of Restaurant ID in Values. After this I’ve also placed the Filter on Country where I only selected the India from the list. Apart from this, I also applied the Price Range as Filter where I selected the 4 only. After applying this, we can achieve the total number of restaurants in India, who has the price range of 4. This data is attached in the **Analysis** sub sheet in the Main Excel File.

A white and blue rectangle with black text

Description automatically generated

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

**Ans** – For this, I’ve created a Pivot table where I placed Country in Rows and Avg of Votes in Values. Then I rounded the avg of votes to 2 decimals so that the data can be readable. This data is attached in the **Analysis** sub sheet in the Main Excel File. Along with this, I’ve also created a Bar chart for a better representation of data on average voters are there in each country.



**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?**

**Ans** – To Approach this problem, I’ve created a Pivot Table where I can visualize which country has how many restaurants and arranged the data in ascending order of restaurant counts. Now, with this data, I’ve only filtered out the countries where the number of restaurants was less than 50. This will help me understand where the competition is less. But there is also the chance that despite the smaller number of restaurants in the country the people there actually like the foods from there, which will make it hard for us to penetrate the market. To resolve that, I also inserted the average rating in the Pivot Table to understand the market sentiment. After this, I only filtered those countries where the average rating was less than 4. The reason for doing this was that the rating of restaurants was low, so people from those countries were not very satisfied with the quality of the food which they were serving. This will give us a good chance to penetrate the market easily if we understand the market needs and the quality of food which the users are looking for. In the end, there were only 4 countries' names which came out to us, which were Canada, Singapore, Sri Lanka and Australia.

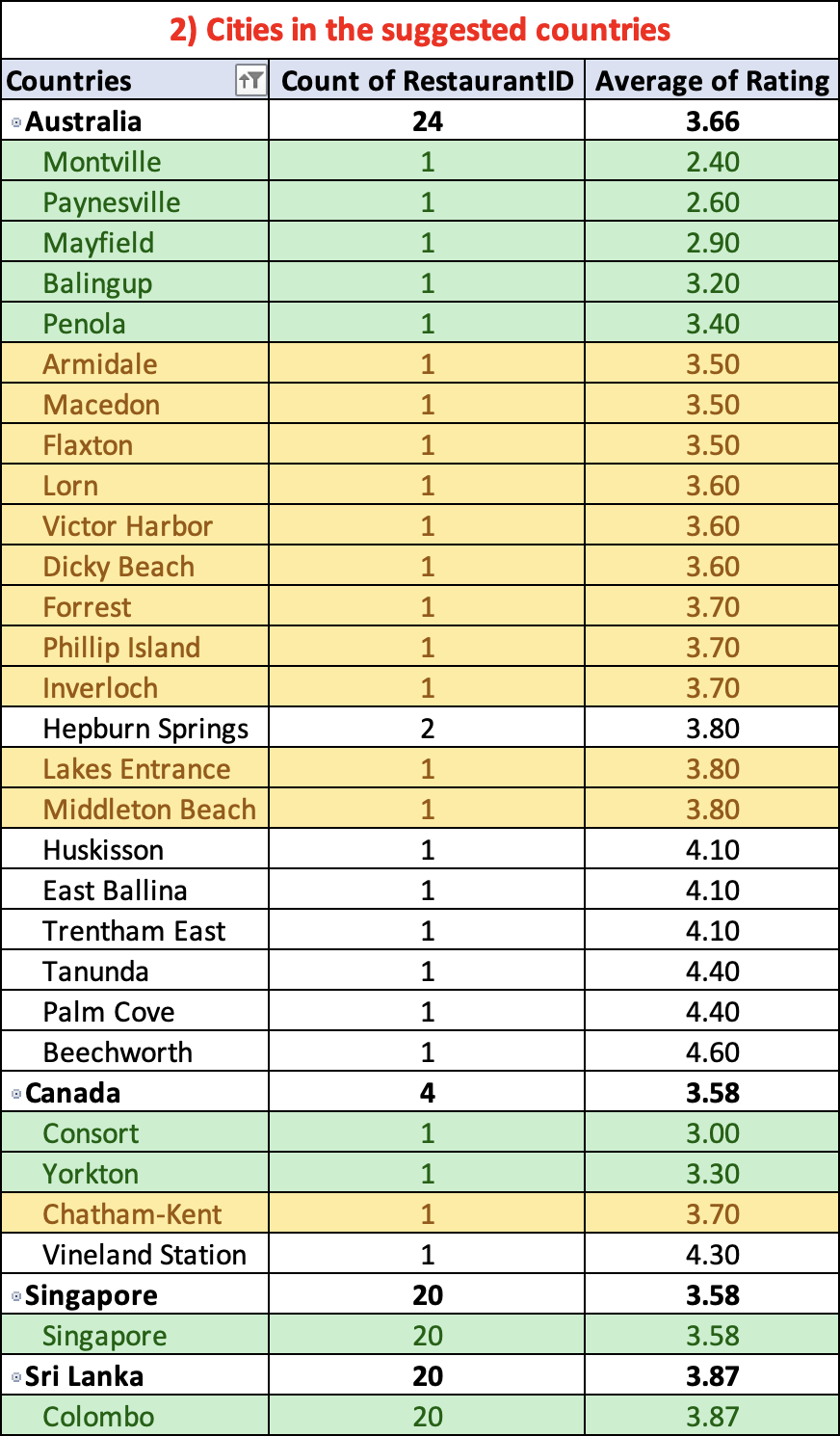
This data can be found in the “**New Openeing-Country and Cities**” sub sheet where I have made the Pivot Table basis on which I took the decision. Apart from this, I’ve also create a Line Graph where I plotted the Count of Restaurants and Avg Ratings for the countries which I selected.

A table with numbers and a number of tables

Description automatically generated with medium confidence

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

**Ans** – For this, I’ve created another Pivot Table where I listed down all the cities in the selected countries along with the number of Restaurants along with the average rating for the city and arranged all of the data in ascending order of Rating. In this Pivot Table, I’ve also highlighted the list of cities where we need to target First and where we need to target later. To do this, I’ve highlighted all the cities in Green where the Rating is lower than 3.5 and the cities in Yellow where the rating is between is in between 3.5-4. For the cities where multiple restaurants are opened, I’ve excluded those cities given the number of city options in this is more. Otherwise, in Countries like Sri Lanka and Singapore, where there is more than 1 restaurant are there but the option is for only 1 city, I’ve considered those cities as well and marked them as Green. This can also be found in the “**New Openeing-Country and Cities**” sub sheet where I have made the Pivot Table basis on which I took the decision.



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

**Ans** – For this, I’ve used the AVERAGEIF function (=AVERAGEIF('Raw Data'!D:D,'New Openeing-Country and Cities'!L5,'Raw Data'!W:W)) for the Four countries which I selected. With this, I can easily see what is the Average Rating of Restaurants in the Country. This can also be found in the “**New Openeing-Country and Cities**” sub-sheet. Apart from this, I’ve also created the visualization with the Bar Chart where I’ve listed the average rating corresponding to the Country.

A table with numbers and text

Description automatically generated

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

Ans – For this, I’ve created a separate sheet “**Currency Conversion**”, where I listed down the current conversion rate for all the different Currency types in the Raw Data. After that, I’ve inserted a new column next to the “Average\_Cost\_for\_two” column with the name of the column as “Average\_Cost\_for\_two\_in\_INR”. In this column, I’ve converted the values which are present in Average\_Cost\_for\_two column with the currency conversion to convert the whole thing in INR by multiplying the value with the VLOOKUP value from the Currency conversion. After this, I’ve created a Pivot Table where I’ve place Country in Rows and Average of Average\_Cost\_for\_two\_in\_INR in Values and then filtered only the selected countries. This can also be found in the “**New Openeing-Country and Cities**” sub-sheet. Apart from this, I’ve also created the visualization with the Pie Chart where we can see what is the average cost for two in INR in the restaurants in the selected countries.

A white and black text with red text

Description automatically generated with medium confidence

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.**

Ans – For this, I’ve created separate Pivot Tables for each selected country where we listed down the Restaurants which are available in the country along with the average rating and average cost for two in INR. To find out the competitors I’ve broken each Pivot table into 3 parts, Highly competition where the rating of the restaurant is 4 or above, Medium competition where the rating of the restaurant is in between 3.2 and 4, and finally less competition where the rating of the restaurant is less than 3.2. I’ve also color coded the Pivots into Green, Yellow and Red for High, medium, and low competition respectively. This data can also be found in the **Competitor Analysis** sub sheet in the main Excel file.

A close-up of words

Description automatically generated

A graph with numbers and text

Description automatically generated with medium confidence

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

Ans – To solve this issue, I’ve created a Pivot Table where I’ve listed all the Cuisines which are getting served in the selected country along with the average of Rating and the total votes which through which the rating has been decided and arranged the data into ascending order of average of rating. With this dataset, I’m trying to find out the cuisines which are rated by a large number of people. This was done to understand what kind of cuisine is currently in demand. For this, I’ve only taken the cuisines where the number of votes is more than 100. But in a country like Singapore where there is no cuisine with more than 100 votes, but the number of votes is evenly distributed, we have to consider all cuisines into consideration. Now, we have taken the cuisines where the rating are below 3.8 rating into high consideration and if the rating is between 3.8 and 4.6 into low consideration. If the rating of any cuisine is above 4.6 along with more than 100 votes, then we should avoid the cuisines as there is a high chance that the users are liking the cuisine and are accustomed to the restaurant. This analysis can be found in the “**Cuisines Analysis**” sub-sheet in the main Excel file where I’ve bifurcated the cuisines based on the above criteria. In my understanding, the choices of cuisines do affect the rating of the restaurant because if you provide the customer with the type of food which is generally bad in the country with the good quality of food then the users are highly likely will like the food and rate the restaurant higher than the peer. But if we serve food which is already highly rated, then the smallest of the mistakes will lead to a bad rating.

A screenshot of a spreadsheet

Description automatically generated

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

Ans – Currently in the given data set, for the selected countries there are no restaurants which are available in this the seat booking or online delivery is being happening. Due to this, we have to look around for the other countries where these facilities are already present, so that we can use them to benchmarking and to arrive at conclusion for the above question. For this, I’ve created a Pivot table where I’ve placed the Has\_Table\_booking and Has\_Online\_delivery in Rows and Average of Rating and Average of Votes in the Values. Here we can clearly see that the Online Delivery and Table booking has good impact on the Rating of the restaurant and we should implement this system in the newer restaurants. This data can also be found in the “**Delivery Analysis**” sub-sheet in the Main Excel file.

A table with numbers and a list of items

Description automatically generated with medium confidence

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?**

Ans – To solve this, I’ve used the correlation function (CORREL function) to find out the correlation between rating and Pricing of the restaurant for the restaurants which are present in the selected countries;. The correlation which came out was -0.011786293, which says that there is no effect on rating if the cuisine price is higher. This analysis can also be found in the “**Correlation**” sub-sheet which is present in the Main Excel file.

A screenshot of a calculator

Description automatically generated

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

Ans – For this, I’ve inserted an additional column in the dataset named Price Brackets where I divided the whole dataset into smaller pieces with the help of If conditions on average cost for two in INR column. After that, I’ve create a Pivot Table where I placed Price Bucket in Rows and Count of restaurant in Values and then Converted the absolute values for each bucket into the %age of value in overall column. After that I’ve created a Visualisation with a Pie Chart to show the distribution of Restaurant in Price Buckets. This data can be found in the “**Pricing Distribution**” subsheet in the Main Excel file.

A table with numbers and numbers

Description automatically generated

**Dashboard**

For this Dashboard, I’ve added all the major elements in the dashboard which were mainly used to decide the which country to select for opening of new restaurants. Along with that, I’ve added few of the basis visualizations which can help us understand how restaurants are performing all over the world. In this I’ve also added few of the slicers like Country and Year which can help us in pinpointing the requirement of the stakeholders and see various country at once without changing the visualization every time. These Slicers will only work on the Dashboard, which is present in the “Dashboard” sub-sheet in the Main Excel file.

A screenshot of a restaurant dashboard

Description automatically generated