# **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?

There are total 3 tables present in the data, namely - Raw Data about the restaurants (Raw Data) table & Country Description (country description) table and cuisines table (Unpivoted Cuisine).

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

The total number of attributes in the cleaned dataset Raw Data, after creating new columns, is 27.

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

**Categorical Data**:

Refers to variables that can be divided into different categories or groups. These categories are often non-numeric and represent qualitative characteristics. In this data, there are 18 categorical columns in the data,

1. **Restaurant ID**: It may seem like an identifier, but since it's a unique identifier, it’s often treated as categorical.
2. **Restaurant Name**: Represents different restaurant names, which are categorical.
3. **Country Code**: Represents different country codes, which are categorical.
4. **Country**: The country name, which is a category.
5. **City**: The name of a city, which is categorical.
6. **Address**: This is typically treated as a categorical field since it's a specific location description.
7. **Locality**: Refers to a neighbourhood or locality, which is categorical.
8. **Locality Verbose**: This would be a more descriptive categorical value.
9. **Cuisines**: Represents categories of food types like "Indian," "Italian," etc.
10. **Currency**: The currency used, such as "INR," "USD," which are categorical.
11. **Has\_Table\_booking**: A binary categorical variable (Yes/No).
12. **Has\_Online\_delivery**: A binary categorical variable (Yes/No).
13. **Is\_delivering\_now**: A binary categorical variable (Yes/No).
14. **Switch\_to\_order\_menu**: A binary categorical variable (Yes/No).
15. **Price range**: Although it’s a numeric scale, it often represents a category (1-5).
16. **Opening Year**: Depending on how it’s used, it could be treated as a categorical variable if specific years are considered discrete categories.
17. **Opening Month**: A categorical variable representing the month the restaurant opened.
18. **Opening Quarter**: Represents categorical values (Q1, Q2, Q3, Q4).

**Continuous Data:**

Refers to variables that can take an infinite number of values within a given range. These values are typically numeric and can be measured with precision.

1. **Longitude**: A continuous variable representing geographic coordinates.
2. **Latitude**: A continuous variable representing geographic coordinates.
3. **Votes**: A numeric value, so it's continuous.
4. **Average\_Cost\_for\_two**: A numeric value, so it's continuous.
5. **INR\_Average\_Cost\_for\_two**: A numeric value, so it's continuous.
6. **Rating**: A numeric value, usually a decimal or integer (e.g., 4.5), so it’s continuous.
7. **Datekey\_Opening**: A date value, treated as continuous in some cases (depending on how it’s used, like tracking time).
8. The data consists of some inconsistent and missing values so ensure that the data used for further analysis is cleaned.

* Removed duplicates by selecting whole range and searching by all the selected columns and no Duplicate values found.
* Checked for any N/A values with filter and using “=COUNTIF ($A$1: A9552,"N/A")” and missing values using “=COUNTBLANK ($A$1: A9552)”. And found the missing rows and fill the data using average, mean and median methods.
* Longitude and Latitude of cities from (India, Sri Lanka, Indonesia) were missing and were filled using the average as per respective cities.
* Cuisine was missing for cities Miller, Orlando and filled it using the maximum rating cuisine from the city.
* Average cost for two was missing for city Miller and filled it using the median of Average cost for two of the country (USA) since there was no other restaurants from the city.

**DateKey Opening Correction:**

* The "DateKey\_Opening" column had inconsistent formats. Updated it using a formula-based approach: =TEXT(DATEVALUE(SUBSTITUTE(X2,"\_","-")),"dd-mm-yyyy"). All dates are now in the correct "dd-mm-yyyy" format.
* This method was chosen to avoid creating additional columns and instead directly convert the dates into the correct format.

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

=XLOOKUP ($C69,'country description'! $A$1:$A$16,'country description'!$B$1:$B$16,0,0)

* Use the XLOOKUP function to create a new column, "Country," in the "Raw Data" worksheet by matching the country code with the corresponding country name.

Explanation:

* The "Raw Data" and "Country Description" worksheets both have a "Country Code" column that can be used to match data. By using the XLOOKUP function, I was able to create a new "Country" column in the "Raw Data" worksheet, which fills in the country names based on the codes.
* The formula used is: =XLOOKUP ($C2,'country description'! $A$1: $A$16,'country description'! $B$1:$B$16,0,0).
* $C2: Refers to the cell with the country code in the "Raw Data" worksheet.
* 'country description'! $A$1: $A$16: Refers to the range in which the look up value is present.
* 'country description'! $B$1:$B$16: Refers to the range in which the return value is present.
* 0: Specifies the value to be returned incase of an error in lookup.
* 0: Specifies an exact match for the lookup.

Output:

* A new column, "Country," has been successfully added to the "Raw Data" worksheet. This column correctly displays the country name corresponding to each country code, as retrieved using the XLOOKUP function.

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

Approach:

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

Explanation:

* **Step 1**: Select all the cells in the "Raw Data" worksheet.
* **Step 2**: Click on Insert -> Pivot Table and specify the location where you want to insert the pivot table.
* **Step 3**: Once the pivot table is created, drag the "Country" field into the Rows section.
* **Step 4**: Drag the "RestaurantID" (or another unique identifier for each restaurant) into the Values section. By default, it might display as a sum, so change it to Count to get the number of restaurants.

**Pivot:**

|  |  |
| --- | --- |
| Country wise data Restaurant# | |
| **Row Labels** | **Count of RestaurantID** |
| India | 8652 |
| United States of America | 434 |
| United Kingdom | 80 |
| Brazil | 60 |
| South Africa | 60 |
| United Arab Emirates | 60 |
| New Zealand | 40 |
| Turkey | 34 |
| Australia | 24 |
| Philippines | 22 |
| Indonesia | 21 |
| Sri Lanka | 20 |
| Singapore | 20 |
| Qatar | 20 |
| Canada | 4 |
| **Grand Total** | **9551** |

**Chart:**

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

**Approach:**

* Create a table to represent the number of restaurants opened in each Year using a pivot table.

**Explanation:**

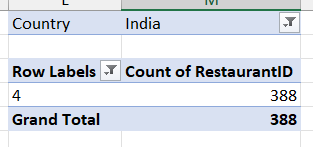
* **Step 1**: Select all the cells in the "Raw Data" worksheet.
* **Step 2**: Click on Insert -> Pivot Table and specify the location where you want to insert the pivot table.
* **Step 3**: Once the pivot table is created, drag the "Year" field into the Rows section.
* **Step 4**: Drag the "RestaurantID" (or another unique identifier for each restaurant) into the Values section. By default, it might display as a sum, so change it to Count to get the number of restaurants.

**Pivot:**

|  |  |
| --- | --- |
| Year wise data Restaurant# | |
| **Row Labels** | **Count of RestaurantID** |
| 2010 | 1080 |
| 2011 | 1098 |
| 2012 | 1022 |
| 2013 | 1061 |
| 2014 | 1051 |
| 2015 | 1024 |
| 2016 | 1027 |
| 2017 | 1086 |
| 2018 | 1102 |
| **Grand Total** | **9551** |

**Chart:**

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



=COUNTIFS('Raw Data'!D:D,"India",'Raw Data'!Q:Q,"4")

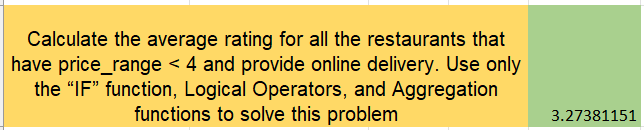
This formula counts the number of rows in the "Raw Data" worksheet where the country is India (column D) and the price range is 4 (column Q). The result is 388, indicating that there are 388 restaurants in India with a price range of 4.

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

**Pivot:**

|  |  |
| --- | --- |
| Country wise Average# Voters | |
| **Row Labels** | **Average of Votes** |
| Indonesia | 772 |
| United Arab Emirates | 494 |
| Turkey | 431 |
| United States of America | 428 |
| Philippines | 407 |
| South Africa | 315 |
| New Zealand | 243 |
| United Kingdom | 205 |
| Qatar | 164 |
| Sri Lanka | 146 |
| India | 137 |
| Australia | 111 |
| Canada | 103 |
| Singapore | 32 |
| Brazil | 20 |
| **Grand Total** | **156.90** |

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.]**

****

**Without Conditional Aggregation:**

To find the average rating of restaurants with a price range of less than 4 and that provide online delivery, you can use

**Formula:**

=AVERAGE(IF( ('Raw Data'!Q:Q<4)\*('Raw Data'!$N:N="Yes"), 'Raw Data'!V:V))

**Explanation:**

* + 'Raw Data'!V:V: The column containing ratings.
  + 'Raw Data'!Q:Q, <4: The condition to filter restaurants with a price range less than 4.
  + 'Raw Data'!N:N, "Yes": The condition to filter restaurants that provide online delivery.

Used IF() condition to consider the rows in column “V” , which meets both the condition using “\*” which acts as a unary AND operator , and those considered rows were averaged using Average().

**Result:**

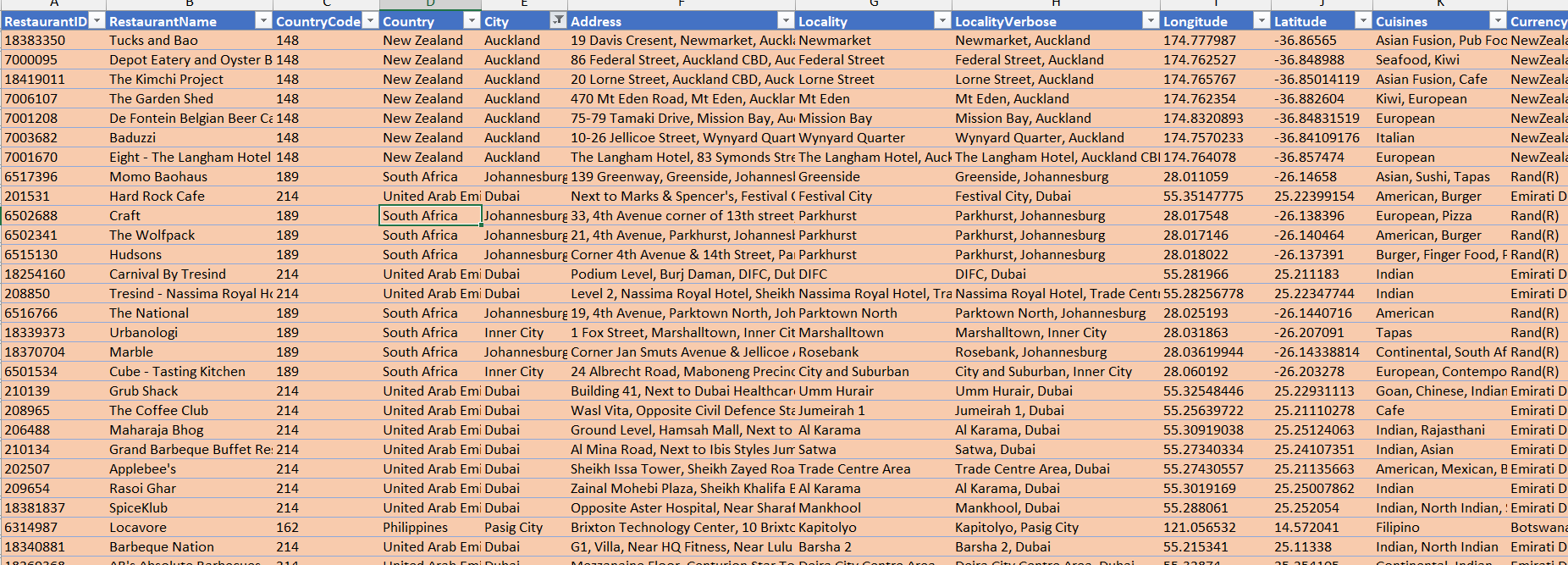
The formula calculates the average rating of the restaurants that meet both conditions, resulting in an average rating of 3.27.

**With Conditional Aggregation**

=SUMIFS('Raw Data'!$U:$U,'Raw Data'!$Q:$Q, "<4", 'Raw Data'!$N:$N, "Yes") / COUNTIFS('Raw Data'!$Q:$Q, "<4", 'Raw Data'!$N:$N, "Yes")

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 the rows of restaurants located in the suggested cities of new restaurant openings, use conditional formatting.



**Explanation:**

* Select the entire “Raw Data” Table.
* Go to Home -> Conditional Formatting -> New Rule.
* Choose Use a formula to determine which cells to format.
* Enter the formula =$E1 ="Dubai” then click Apply and OK. Similarly add same formula for all the suggested cities and will get all the highlighted rows where city name matches the condition

This will highlight the column when the city is Inner City, Johannesburg, Pasig City, Randburg,San Juan City, Tangerang, Auckland, Dubai, Jakarta, Sandton.

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]

To create a new column that combines the currency abbreviation/symbol with the "Average\_cost\_for\_two" value, use the following formula

* =CONCATENATE (MID ($L2, SEARCH("(",$L2)+1,

SEARCH(")",$L2)-1-SEARCH("(",$L2)) ," ",$S2 )

* This formula extracts the currency symbol from the "Currency" column (L) and combines it with the "Average\_cost\_for\_two" value from column (S), creating a customized price format.
* “SEARCH(")",$L2)-1-SEARCH("(",$L2))” helps to extract number of character.
* “SEARCH("(",$L2)+1” finds the first open parenthesis in symbol and takes next char as start index.
* “MID ( $L2 , start index , no’s of char)” ,helps to get data in middle of data and “CONCATENATE” helps to combine symbol with average cost of 2’s.

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?

* To 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,
* use the following array formula:

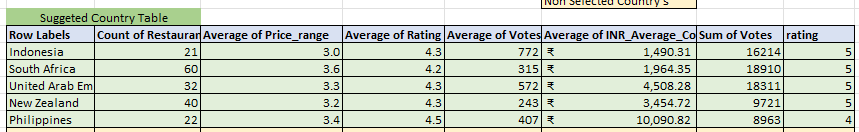
{=SUM((Table1[Has\_Online\_delivery]="No") \* (Table1[Price\_range]=1) \* (Table1[INR\_Average\_Cost\_for\_two]<=250) ) }

* This formula multiplies the conditions to filter the data accordingly and then sums the results to get the total count.
* The Final count was 1694.

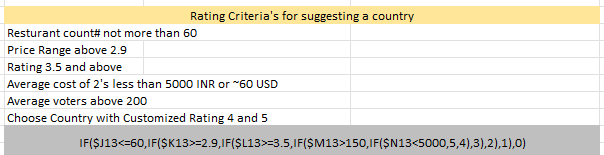
**Subjective Question:**

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?

Indonesia, South Africa, UAE, New Zealand, Philippines are some of the suggested countries where the customer engagement is high and cost of two’s less than Rs.5000 and less No.s of restaurants and other criteria,



Citeria used for Countries selection:

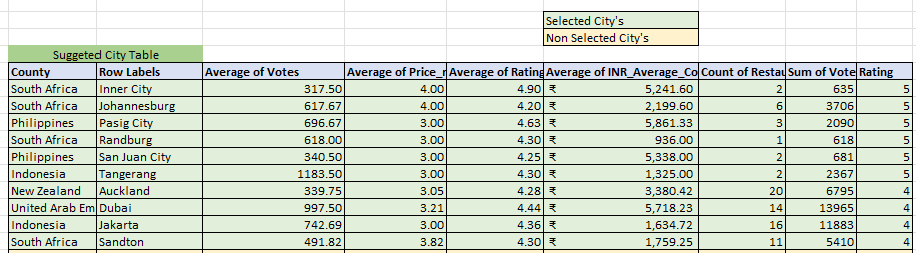


Formula used for Rating column:

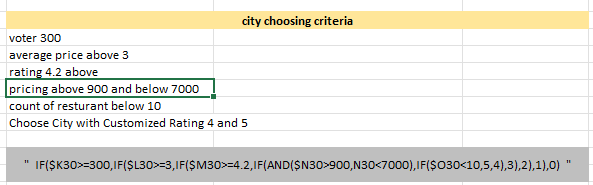
IF($J13<=60,IF($K13>=2.9,IF($L13>=3.5,IF($M13>150,IF($N13<5000,5,4),3),2),1),0)

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

Inner City, Johannesburg, Pasig City, Randburg,San Juan City, Tangerang, Auckland, Dubai, Jakarta, Sandton are some of the suggested Cities where the customer engagement is high and cost of two’s less than Rs.5000 and choose based on other Criteria.



Citeria used for Countries selection:



Formula used for Rating column:

IF($K30>=300,IF($L30>=3,IF($M30>=4.2,IF(AND($N30>900,N30<7000),IF($O30<10,5,4),3),2),1),0)

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

Approach:

* Create a pivot table connected with a chart and a slicer to visualize the data effectively.

Explanation:

* Select the relevant data in the "Raw Data" worksheet.
* Go to Insert -> Pivot Table and choose the desired location.
* In the pivot table:
* Drag the "Country" field into the Rows section.
* Drag the "Rating" field into the Values section and change the aggregation from Sum to Average.

Add a slicer to filter the data by country

**Chart**

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

**Approach:**

* To evaluate the current expenditure on food in the suggested countries and ensure financial expenses are controlled, a pivot table is used.

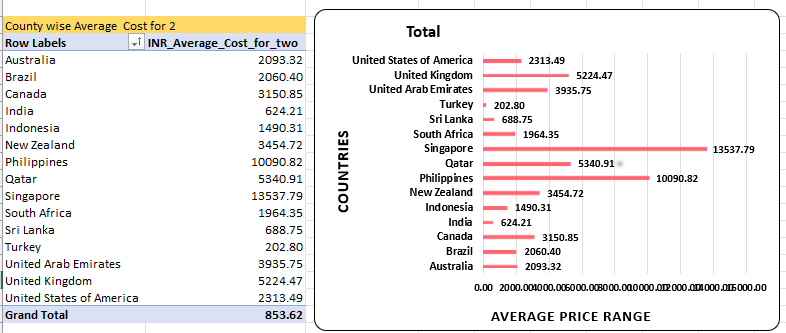
**Explanation:**

* Select the relevant data in the "Raw Data" worksheet.
* Go to Insert -> Pivot Table and choose the desired location for the pivot table.
* In the pivot table:
* Drag the “Country” field into the Rows section.
* Drag the “INR\_Average\_Cost\_for\_two” section.
* Use filters to refine the data as needed.

**Insights:**

* Singapore has the highest average cost for dining for two people, followed by the Philippines, and so on.
* However, countries like South Africa, Sri Lanka, Turkey, New Zealand, Qatar, and the UAE remain affordable, which is very helpful in controlling financial expenditure.

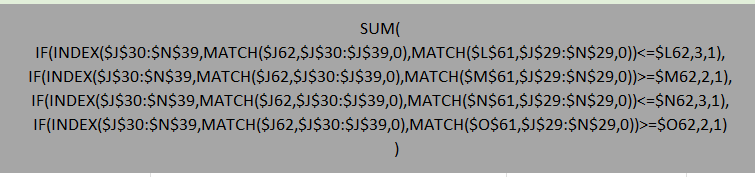
**Visualization:**



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.

Found the Competitors Who are performing better than the City's Average of Voters,Price Range, Rating, and Cost of two's , Found the below mentioned formula (average votes and rating shows customer engagements, this should be higher than average of city and similarly price range and cost of two's lesser that average means cost is low)( with Both this comparision we found restaurants that provider Better food for lesser Price ) , where it compares with the average metrics of "suggested city table" with respective cities and rates the restaurants, and restaurants with higher rating will be our competitors

**Formula used to point restaturants:**

****

Based on points found the restaurants name with highest points who are performing better than cities average metrics,



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

Approach:

* By analysing various insights from the data, I can determine which cuisines to focus on in new restaurants to potentially improve customer feedback and ratings.

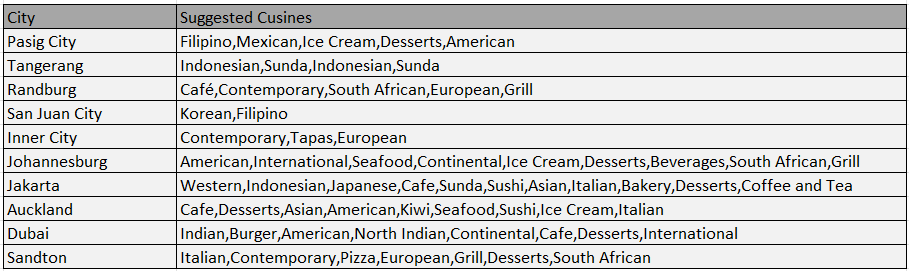
Explanation:

To identify suitable cuisines for new restaurants:

* Create a pivot table filtering by each country.
  + Set the pivot table to display Cuisines in rows and Average Rating in values.
  + Sort the average ratings in descending order to highlight the top 5 cuisines with the highest ratings for each country.
  + Compare the top-rated cuisines across countries to find common ones that consistently receive high ratings.
* Also separated cuisines name using delimiter “,” and using unpivoted column option using power query option found city wise cuisine and their popularity of cuisines across cities. And created pivot,
* With both the pivots found cuisines that are popular having high rating,

* By focusing on these top-rated cuisines:
  + Cuisines that appear frequently among the top-rated options are likely to be well-received across different regions.
  + Incorporate these cuisines into the menus of new restaurants to align with customer preferences and potentially enhance ratings.

**Suggested cuisines for suggested cities:**

****

**Result:**

These insights will provide a data-driven basis for selecting cuisines that are likely to receive better feedback and improve overall restaurant ratings.

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

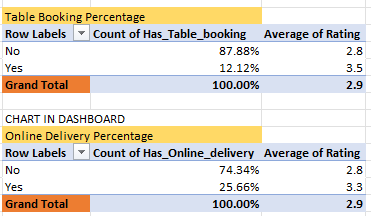
**Approach:**

* Created a pivot table with "Online Delivery" and "Table Booking" in the Rows section and their counts in the Values section.
* Added a slicer for country filtering and converted the data into a chart for visualization.

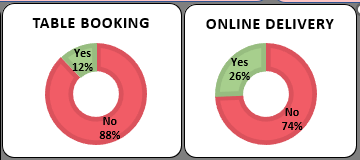
**Explanation:**

* The pivot table provided insights into how the presence of online delivery and table booking affects restaurant ratings. By analysing the data, it was observed that restaurants offering both services generally received higher ratings compared to those that did not. This trend was consistent across different countries, as visualized through the chart.
* The slicer allowed for filtering by country, helping to identify specific regions where these features had a notable impact on ratings. This visualization made it clear that implementing online delivery and table booking enhances customer convenience and satisfaction, which in turn improves ratings.

**Pivots:**



**Charts:**

****

**Result:**

Incorporating online delivery and table booking features is likely to positively affect customer ratings and overall restaurant performance. It is recommended to include these features to enhance customer experience and increase satisfaction

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?

Approach:

* Calculated the correlation coefficient between cuisine rates and ratings using the formula =CORREL(Table1[INR\_Average\_Cost\_for\_two],Table1[Rating])

Explanation:

* The correlation coefficient between cuisine rates and ratings is approximately 0.31, indicating a weak to moderate positive correlation. This suggests that while there is a slight relationship between higher cuisine rates and improved ratings, the impact is minimal.
* A higher rate for cuisines may lead to a slight increase in ratings, but the effect is not significant enough to rely on pricing adjustments alone for improving customer feedback.



Suggestion:

* Given the weak correlation between cuisine rates and ratings, adjusting rates alone is unlikely to significantly impact customer feedback. Focus on improving quality and service to enhance overall customer satisfaction.

Result:

* Adjusting cuisine rates is unlikely to significantly affect customer feedback. It is recommended to consider other factors, such as quality and service, for enhancing overall customer satisfaction.

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

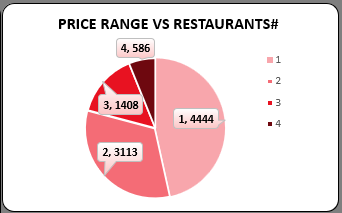
Approach:

* Created a pivot table with Price Range in rows and Count of RestaurantID in values.
* Added a slicer to filter data by country.

Explanation:

* This setup allows for analysing the distribution of restaurants across different price ranges.
* The slicer enables viewing the distribution by country.

Visualization:



* Provides a clear view of how restaurants are distributed across various price ranges in each country. Helps in understanding pricing strategies and market segmentation for new restaurant openings.

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]**

* **Identify key questions:** Determine what questions need to be answered to make informed decisions about new restaurant openings. Document these questions to guide the analysis process.
* **Clean the data:** Ensure that the raw data is accurate and consistent by addressing any issues such as missing values, formatting errors, or inconsistencies.
* **Analyse Consumer Spending Trends:** Leverage the "Price Range" column to create a Pivot Table. Identify trends in food expenses across various locations. Focus on areas with higher spending for potential profitability.
* **Evaluate Competition and Restaurant Ratings**: Look for areas with high demand but low competition. Identify regions with fewer restaurants but higher votes as an opportunity. Assess restaurant ratings in existing establishments to align with customer expectations.
* **Create pivot tables:** Use pivot tables to dynamically analyse and summarize the data. This allows for easy exploration of different dimensions and metrics.
* **Build charts:** Visualize the data with charts derived from pivot tables to identify trends, patterns, and insights more effectively.
* **Compile into a dashboard:** Assemble the charts and insights into a comprehensive dashboard. This centralizes information and facilitates a more holistic analysis to support decision-making.

**The dashboard must consist of Year-wise and country slicers**

****