# **Tasks**

**Objective Questions**:

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

Solution:

* The Zomato\_Analysis\_Saran workbook contains four worksheets, but the cleaned data are organized in a single table for dynamic use. This table is located in the "Raw Data" worksheet, so the total number of tables is one.

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

Solution:

* The total number of attributes in the cleaned dataset, 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]

Solution

* **Categorical data** refers to variables that can be divided into different categories or groups. These categories are often non-numeric and represent qualitative characteristics.
* **Continuous data**, on the other hand, 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.
* Categorical columns are those with discrete and finite values. In this data,there are 19 categorical columns in the data: RestaurantID, RestaurantName, CountryCode, City, Address, Locality, LocalityVerbose, Updated\_Cuisines, Currency, Has\_Table\_booking, Has\_Online\_delivery, Is\_delivering\_now, Updated\_Datekey\_Opening, Switch\_to\_order\_menu, Price\_range, Months,Years,Quarter and Country.

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

Approach:

1. Correct the "DateKey\_Opening" Format:

* Use a formula to correct the date format without creating extra columns.

1. Fill in Missing Values in the "Cuisines" Column:

* Identify and fill in blank spaces with the mode of the "Cuisines" based on the country.

Explanation:

DateKey\_Opening Correction:

* The "DateKey\_Opening" column had inconsistent formats. While there were multiple ways to address this, such as using the "Text to Columns" feature, I opted for a formula-based approach: =TEXT(DATEVALUE(SUBSTITUTE(X2,"\_","-")),"dd-mm-yyyy").
* This method was chosen to avoid creating additional columns and instead directly convert the dates into the correct format.

Cuisines Column Cleaning:

* Applied a filter to the "Cuisines" column and found blank spaces, all within the same country.
* Created a pivot table with "Cuisines" in the row and their repetition in value to identify the most frequent cuisine type for that country.
* Identified "Mexican" as the most common cuisine in the United States and used the formula =IF(ISBLANK($K2),"Mexican",$K2) to fill the blanks with "Mexican."

Visualization:



Output:

* DateKey\_Opening Column: All dates are now in the correct "dd-mm-yyyy" format.
* Cuisines Column: Blank spaces in the "Cuisines" column have been filled with the most frequent cuisine type for the respective country, such as "Mexican" for the United States.

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

Approach:

* Use the VLOOKUP 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 VLOOKUP 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: =VLOOKUP($C2, 'country description'!$A$1:$B$16, 2, 0)
* $C2: Refers to the cell with the country code in the "Raw Data" worksheet.
* 'country description'!$A$1:$B$16: Refers to the range that includes both country codes and country names in the "Country Description" worksheet.
* 2: Indicates that the function should return the value from the second column in the range.
* 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 VLOOKUP 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.

Visualization:



Result:

* This process will create a table that represents the number of restaurants opened in each country, showing the count of restaurants for each country.

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

Approach:

* + 1. Extract the Year from Dates:
  + Create a new column that extracts the year from the "DateKey\_Opening" using the YEAR function.
    1. Use a Pivot Table:
  + Represent the number of restaurants opened each year by creating a pivot table.

Explanation:

1. Step 1: Create a New Column for the Year:
   * Use the formula =YEAR($Y2) to extract the year from the "DateKey\_Opening" column. Drag the formula down to fill the entire column.
2. Step 2: Create a Pivot Table:
   * Select all the cells in the "Raw Data" worksheet.
   * Click Insert -> Pivot Table, specify the location where you want to place the pivot table.
   * In the pivot table, drag the "Year" field (from the new column) into the Rows section.
   * Drag the "RestaurantID" (or another unique identifier for each restaurant) into the Values section. By default, it may show as a sum, so change it to Count to display the number of restaurants.

Visualization:



Output:

* This process will generate a table that shows the number of restaurants opened each year, providing the management with a clear view of yearly restaurant openings.

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

Approach:

* To determine the total number of restaurants in India with a price range of 4, we use the COUNTIFS function.

Formula Explanation:

* The formula =COUNTIFS('Raw Data'!T:T, 4, 'Raw Data'!AC:AC, "India") counts the rows where column T (price range) is 4 and column AC (country) is "India".

Output:

* The result is 388, indicating 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?

Approach:

* To calculate the average number of voters for restaurants in each country, create a pivot table.

Explanation:

* Select all cells in the "Raw Data" sheet.
* Go to Insert -> Pivot Table, and specify the location for the pivot table.
* Drag "Country" to the Rows section and "Votes" to the Values section. By default, it will show the sum; change this to average.

Visualization:



Result:

* This process will display the average number of votes for restaurants in each country.

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.]**

Approach

* To calculate the average rating for restaurants with a price range less than 4 and that offer online delivery, use the IF function along with logical operators and aggregation functions.

Explanation

* The idea is to apply the conditions within an IF function using the AND operator. If both conditions are met, the function will calculate the average rating.
* The formula is =AVERAGE(IF(('Raw Data'!T:T < 4) \* ('Raw Data'!N:N = "Yes"), 'Raw Data'!W:W))

Result

* The result is an average rating of 3.27381151 for all restaurants that meet the specified criteria.

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.

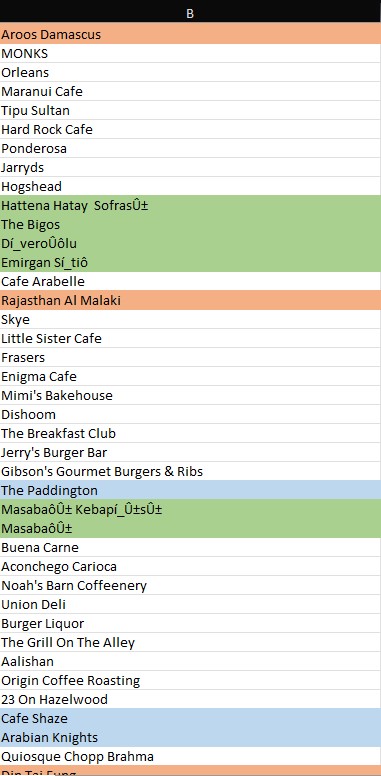
Approach

* To highlight the rows of restaurants located in the countries or cities suggested for new restaurant openings, use conditional formatting.

Explanation

* Select the "RestaurantName" column.
* Go to Home -> Conditional Formatting -> New Rule.
* Choose Use a formula to determine which cells to format.
* Enter the formula: =$AC1="United Arab Emirates", then click Apply and OK.
* And again reapt the same for “Turkey”,”South Africa”, “News Zealand”, “UAE” and “Sri Lanka”.

Visualizaion:



Result

* This will highlight the "RestaurantName" column when the country is South Africa, Sri Lanka, Turkey,News Zealand,Qatar and UAE..

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]

Solution:

* To create a new column that combines the currency abbreviation/symbol with the "Average\_cost\_for\_two" value, use the following formula:=MID($L2, FIND("(", $L2) + 1, FIND(")", $L2) - FIND("(", $L2) - 1) & " " & $V2
* This formula extracts the currency symbol from the "Currency" column (L) and combines it with the "Average\_cost\_for\_two" value from column V, creating a customized price format.

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?

Solution:

* 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: =SUMPRODUCT(('Raw Data'!N:N="No") \* ('Raw Data'!T:T=1) \* ('Raw Data'!S:S<=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?

Approach:

* By analyzing various insights from the data, I can suggest countries where the team can open new restaurants with lesser competition.

Explanation:

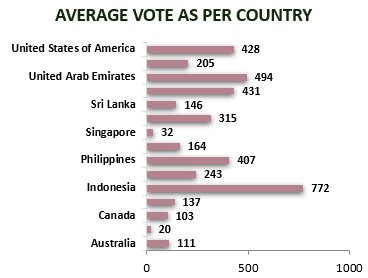
The most suitable countries for opening new restaurants are those where:

* The total number of restaurants is less than 100.
* The average ratings are 3.9 or higher.
* The average expenditure for two people is less than 60 USD.
* The number of voters is higher than 140.

Insights:

Based on these criteria, the suggested countries are South Africa, Sri Lanka, Turkey,News Zealand,Qatar and UAE.

Visualization:

A bar graph with numbers and names

Description automatically generated

A graph showing the country's average rating

Description automatically generatedA graph of the country's expenditure

Description automatically generated

Result:

These visualizations will provide a clear, data-driven basis for recommending the most promising countries for new restaurant openings.

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

Approach:

* I identified suitable states and cities in the suggested countries by analyzing factors such as average rating, the count of restaurants, and average price range. This helped in selecting locations with favorable conditions and lower competition.

Result:

The most suitable locations are

* South Africa: Cape Town, Johannesburg, Pretoria, Randburg, Sandton
* Sri Lanka: Colombo
* Turkey: Ankara, Istanbul
* News Zealand: Auckland, Wellington City
* Qatar: Doha
* UAE: Abu Dhabi, Sharjah

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

Visualization:

A graph showing the country's average rating

Description automatically generated

Insights:

* We can able to the analysis that the selected country are having better rating above 3.9
* New Zealand – 4.3
* Qatar – 4.1
* South Africa – 4.2
* Sri Lanka – 3.9
* Turkey - 4.3
* United Arab Emirate – 4.2

Suggestion:

* In Sri Lanka, where the average restaurant rating is 3.9, enhancing services by implementing online delivery, table booking, and expanding the cuisine options could significantly improve customer satisfaction, making it a strong candidate for opening new restaurants.

Result:

* This method allows for easy analysis of restaurant ratings across different countries

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 “Price in USD” field into the Value 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:

A graph of the country's expenditure

Description automatically generated

Suggestion:

* Given that Singapore and the Philippines have very high expenditure levels, it would be challenging to maintain costs in these markets. However, focusing on countries like South Africa, Sri Lanka, Turkey, and New Zealand, which offer more reasonable pricing and strong restaurant ratings, presents a more viable option for expansion.

Result:

* By using this method, you can effectively monitor and control the food expenditure in the suggested countries, ensuring that financial expenses are managed efficiently.

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.

Approach:

* To identify the biggest competitors and those rated in the lower brackets (1-2 or 2-3) from the recommended states, we'll follow this approach:
* Filter Competitors: First, identify restaurants marked as competitors.
* Check Ratings: From this list, further filter those rated in the lower brackets (1-2 or 2-3).
* Compile Results: Compile the final list of restaurants that meet both criteria.

Explanation:

* Competitors Identification: We already have a list of competitors based on whether they meet the specific criteria related to price range and average amount spent.
* Rating Check: For each restaurant identified as a competitor, we need to check their average rating to ensure they fall into the lower rating brackets.
* Compilation: Gather the names of these restaurants and present them as the final result.

Visualization:

* Here are the restaurants from the recommended states that are our biggest competitors and also rated in the lower brackets (1-2 or 2-3):



Suggestion:

* Since competition is intense in South Africa and Turkey, it would be strategic to focus on countries with less competition, such as New Zealand, Qatar, and Sri Lanka, where there is greater potential to establish a strong market presence.

Results:

* These restaurants are considered significant competitors in their respective regions and have been rated in the lower brackets, indicating potential areas of focus or concern.

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:

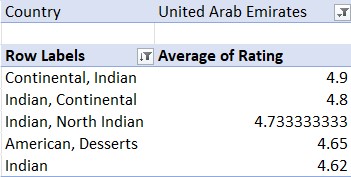
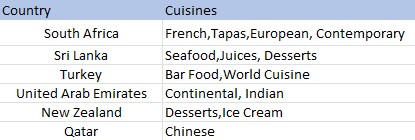
* By analyzing various insights from the data, I can determine which cuisines to focus on in new restaurants to potentially improve customer feedback and ratings.

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.
* 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.

Visualization:

Suggestion:

* Based on the analysis of cuisine ratings, desserts, ice cream, Indian seafood, and juices have received high ratings. Incorporating these types of cuisines into your restaurant offerings could enhance customer satisfaction and make your restaurant more appealing.

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 analyzing 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.

Visualization:

A pie chart with numbers and a number

Description automatically generatedA pie chart with numbers and a few words

Description automatically generated

Suggestion:

* Since South Africa, Qatar, and the UAE benefit from table booking and online delivery, it’s wise to implement both features to improve customer satisfaction and align with successful practices.

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('Raw Data'!W2:W9552,'Raw Data'!S2:S9552).

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 analyzing the distribution of restaurants across different price ranges.
* The slicer enables viewing the distribution by country.

Visualization:

A diagram of a distribution of price range

Description automatically generatedA screenshot of a cellphone

Description automatically generated

Result:

* 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.
* **Create pivot tables:** Use pivot tables to dynamically analyze 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.