**SUBMISSION BY: ROSHAN CHAWLA**

**PROJECT: ZOMATO**

**SKILL USED: MICROSOFT EXCEL**

**BATCH: DATA SCIENCE COURSE MAY 2024**

**Objective Questions**:

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

Ans: Total No of Tables – 2

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

Ans: Total No of attributes present in the data: 20

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

Ans:

**Categorical Data: 4**

(Country,

city,

cuisines,

currency)

**Continuous Data: 7**

(Longitude,

Latitude,

Price\_range,

Votes,

Average\_Cost\_for\_two, Rating,

Year\_Opening)

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

**Ans:** Following below mentioned approached followed to deal with inconsistent and missing values to ensure that data is cleaned for further analysis:

1. **Missing Values:** Column K (Cuisines) in Sheet Raw data has some missing values for country United States of America. Since the missing values are only 9, therefore I have filled those missing values randomly, as it will not impact the results of analysis significantly. Those empty cells are coloured green for future references.
2. **Formatting of Date Column:** I have added a new ColumnX (Datekey\_Opening\_New) to organise the dates given in column W. I have applied the following formula for X2 and copied for the rest of the cells through drag option.

Formula: =DATEVALUE(SUBSTITUTE(W2,"\_","-")).

By using this formula date is changed from (2013\_9\_21 to 21-09-2013).

1. **Added 3 colomns for Day (Column y), Month (Column Z), Year (Column AA):** Following formulas used:

For Day formula used is =Day(Y2) and copied to all the cells.

For Month formula used is =Month(Z2) and copied to all the cells.

For Year formula used is =Year(AA2) and copied to all the cells.

1. **Changing Average Cost in INR in Column AB:**

To convert it to INR, First I have created a data where I have mentioned currency exchange rates in 2nd Sheet Country Description. The rates are taken as on 6th June 2024 from internet.

There are 14 foreign currencies and 1 local currency (India). I have used filter option to filter each country and applied the formula where I have multiplied (Amount in foreign currency \* its rate in another sheet using absolute referencing.

I followed the same procedure for all 13 foreign currencies to convert them into Indian Rupees.

1. **Border:**  Zomato data shared did not have any border. So, I added the border to make the data look aligned, cleans and easy to read.
2. **Helper Column AD:** Created a helper column to determine the average rating all the restaurant that price range <4 and provide online delivery.

First, I have used IF & AND to determine only those values which satisfies the condition of price range and online delivery and then used that column to find out Average. Which is the desired output.

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

**Ans:**  Excel function used to extract Country names:

**XLOOKUP FUNCTION:**

Formula used = XLOOKUP('1. Raw Data'!C2,'2. country description'!$A$2:$A$16,'2. country description'!$B$2:$B$16)

Where,

C2: Lookup Value

'2. country description'!$A$2:$A$16 : Lookup Array

'2. country description'!$B$2:$B$16 : Return Array

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

**Ans:**  I have created Pivot table to represent the number of restaurants opened in each country:

I have taken country in row and Count of Restaurant Id in Values. The output is present at the above-mentioned path.

**Location**: Sheet 3. Analysis => Cell B 25

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

**Ans:** I have created Pivot table to represent number of restaurants opened in each year.

I have taken Years in row and count of Restaurant Id in values. The output is present at the above-mentioned path. I have also created a Bar graph to visualize it.

**Location**: Sheet 3. Analysis => Cell B 43.

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

**Ans:** Total Restaurant in the price range of 4 (India) = 388

Function used is Countifs :

= COUNTIFS('1. Raw Data'!$D$1:$D$9552,"India",'1. Raw Data'!$R$1:$R$9552,">=4")

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

**Ans: Pivot Table:** I have created a Pivot Table to understand Average number of voters in each country. I have put Country in Row and Votes in Values to get the desired output.

**Location**: Sheet 3. Analysis => Cell G3.

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.]

**Ans:** I have created a Helper Column in Column AD.

I have Used IF along with AND operator to find out ratings as per the conditions given in the question.

After that I have calculated the average rating for all the restaurants by using the **Average Formula** in **Sheet 3. Analysis**.

=Average('1. Raw Data'!AD2:AD9552)

So, the average rating for the restaurant that have price\_range < 4 and provide online delivery is 3.273812.

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.
2. Create a new customized price column that consists of the abbreviation/symbol of the currency along with the Average\_cost\_for\_two values. [Use string operations to do this task]

**Ans:** I have created a new column – Currency\_symbol in Column M in Sheet 1. Raw Data.

Steps undertaken to do the same:

1. Applied Text to Column on Column N (with delimiter ‘(‘ )
2. Column M had ‘Rs.)’ as value after text to column.
3. After that I have used Replace Function where

Find : )

Replace : Blank

And click on Replace All.

This way I have got the desired output in Column M.

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?

**Ans:** 1685 such Restaurants which fulfils the conditions.

**Formula used:**

{=SUMPRODUCT(('1. Raw Data'!$D$1:$D$9552="India")\*('1. Raw Data'!AC1:AC9552<=250)\*('1. Raw Data'!$O$1:$O$9552="No")\*('1. Raw Data'!R1:R9552=1))}

**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: Pivot Table:**  I have created a Pivot Table where I have put Country in Row and Count of Restaurant ID & Average of Rating in Values. After analysing the Pivot, I have selected only those countries where there is less competition and ratings are between 3 and 4.

Countries Suggested for New Restaurant: Australia, Canada, Singapore and Sri Lanka.

I have also created a line chart to visualise the same.

**Location**: Sheet 4. Recommendation – New Opening => A3

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

**Ans: Pivot Table:**  I have created Pivot Table where I have put Country and City in Row and Count of Restaurant Id and average of rating in Values.

Cities Shortlisted:

1. **Australia**: Armidale, Balingup, Dicky Beach, Flaxton, Lorn, Macedon, Penola, Victor Harbor.
2. **Canada**: Consort, Yorkton
3. **Singapore**: Singapore
4. **Sri** **Lanka**: Colombo

**Basis of Selection:** I have taken only those cities which have lesser competition and are below Country’s average rating. This is because due to less competition and below average rating, there is a potential to understand the market and open new restaurants based on preferences and past data on highly rated cuisines to attract more customers.

**Location**: Sheet 4. Recommendation – New Opening => E3

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

**Ans: Pivot Table:**  I have created a Pivot Table to understand the current quality regarding ratings of suggested countries where restaurants should be opened.

I have Taken country in rows and Average Ratings in Values and applied filter to get only the values of Australia, Canada, Singapore, and Sri Lanka.

**Location**: Sheet 4. Recommendation – New Opening => I3

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

**Ans: Table:** I have created aTable to understand current expenditure on food in the suggested countries.

I have used **SumIf Function:**

=SUMIF('1. Raw Data'!$D$1:$D$9552,"Australia",'1. Raw Data'!$AB$1:$AB$9552)

Following are the readings:

1. Australia: Rs. 32,177
2. Canada: Rs. 8,841
3. Singapore: Rs. 1,93,053
4. Sri Lanka: Rs. 13,134

I have used a Pie Chart to visualize the same in **Sheet - 4. Recommendation - New Opening.**

**Location**: Sheet 4. Recommendation – New Opening => L3

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:** I have used Pivot Table where I have put Country and restaurant name in rows and in values, I have taken average of rating and average cost of two in Indian Rupees. I have taken City in Filter section.

**Restaurant with biggest competitors:** I have marked the cities in Green which have the biggest competition in the Market.

**Restaurant with weak competition:** I have marked the cities in red which have week competition in the market.

**Location**: Sheet 5. Competitor Analysis => A5

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:**

**Market Dynamics:** There are lot of cuisines in different countries/ cities. If one cuisine worked in 1 country maybe not work in another. So, studying the market and past data is very important to make decision regarding the menu of new restaurants in the suggested cities of suggested countries.

**Suggestion:** I have observed that local cuisines in the country has higher ratings than the other one which are imported. There are few cuisines which are popular in all the suggested countries: Seafood, Italian.

In other words, in each country I have selected the local cuisine and some foreign cuisines which had better ratings. In order to establish a new restaurant, it is important that we cater to the local people by offering local cuisines to make it work. Once the restaurant takes off we can add more foreign cuisines to improve the offerings.

**Verdict:** After observing the data and the Tables, it is clear that some cuisines which are local have higher ratings. So, inclusion of local cuisine will be no brainer if we want to establish new restaurants having less competition but good ratings. We can try including some other cuisines and analyse the trend by the ratings which we are getting from people.

**Visualization:** Used a Line Chart to visualise the Average ratings in the suggested counties.

**Location:** Sheet 6. Cuisine Analysis

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

**Ans:** I have used Countifs function to find out if the suggested countries have online delivery and table bookings.

**Formula used :**

**For Online Delivery** =COUNTIFS('1. Raw Data'!$D$1:$D$9552,"Australia",'1. Raw Data'!$O$1:$O$9552,"Yes")

**For Table Bookings** =COUNTIFS('1. Raw Data'!$D$1:$D$9552,"Australia",'1. Raw Data'!$N$1:$N$9552,"Yes")

I have also created a table where I have ascertained the percentage of online delivery out of total orders to analyse the trend and see if the opportunity can be ceased to increase the market share. Similarly, for Table Books as well.

Also, would like to explore the opportunity to study if average cost of 2 can be increased based on Table booking or delivery services. Sometimes tapping the pain areas works wonder. Similarly, if we tap the areas where there are not able bookings and no online delivery, will the customer be willing to spend more for convenience of service.

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:** I have used CORREL function to analyse the correlation between the average cost of two and ratings.

**Following are my Findings:**

1. R^2 = -0.003942111, meaning it has negative correlation. Negative correlation usually means inverse relationship. Average cost of two has inverse relation with Ratings. So, if the Average cost of 2 increases the Ratings will reduce.

In this case, since the value is closer to 0, the effect of inverse correlation is negligible, meaning the reduction in Rating by increasing the Average cost of two is negligible.

1. We can keep the rates of cuisine higher to optimise profitability at the areas of new Restaurant Opening.
2. What is the distribution of the number of restaurants of different price ranges in all the countries?

**Ans: Pivot Table:** I have created a pivot table to understand the distribution of the number of restaurants of different price ranges in all the countries. I have put Country in rows, Price\_range in Colomn and number of restaurants in values to get the desired Pivot Table.

I have also created a range manually to understand better, and to use the histogram visualization technique to visualise it effectively.

**Location:** Sheet 9. Price Analysis => A1

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 must give bullet pointers in order to answer this question]

**Ans:** 

* Analysed existing data to identify countries with less competition and potential for growth.
* Evaluated average ratings and number of competitors in each country to gauge market saturation.
* Focused on countries with average ratings between 3 and 4 to ensure a good starting point for new restaurants.
* Looked for cities within selected countries that have lower competition and ratings below the country's average.
* Prioritized cities with potential for market understanding and customer attraction based on past data.
* Considered local cuisines and preferences while planning the menu for new restaurants to appeal to the target market.
* Used visualization techniques like pivot tables, charts, and histograms to present data effectively and support decision-making.
* Analysed factors like online delivery and table booking availability to understand customer preferences and market trends.
* Investigated the correlation between cuisine rates and customer ratings to optimize profitability without compromising feedback.
* Utilize techniques like conditional formatting and array formulas to efficiently analyse and interpret data for decision-making.