**ZOMATO ANALYSIS: A exploration of finding and selecting locations.**

**Objective Questions**:

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

There are 2 table provided in the excel:

1. Raw Data: In this sheet we have information about restaurants.
2. Country Description: In this sheet we have country code with the name of the countries.
3. **What is the total no. of attributes present in the data?**

There are 27 attributes present in total of 2 sheets:

**Raw Data:**

* + **Restaurant ID:** Unique identifier for each restaurant.
  + **Restaurant Name:** The name of the restaurant.
  + **CountryCode:** Country code of the location where the restaurant is situated.
  + **City:** The city where the restaurant is located.
  + **Address:** The specific address of the restaurant.
  + **Locality:** The locality or neighborhood where the restaurant is situated.
  + **Locality Verbose:** Detailed information about the locality.
  + **Longitude:** The geographical longitude coordinate of the restaurant.
  + **Latitude:** The geographical latitude coordinate of the restaurant.
  + **Cuisines:** The type of cuisine offered by the restaurant.
  + **Currency:** The currency used for transactions in the restaurant.
  + **Has\_Table\_booking:** Indicates whether the restaurant has a table booking option (Yes/No).
  + **Has\_Online\_delivery:** Indicates whether the restaurant offers online delivery (Yes/No).
  + **Is\_delivering\_now:** Indicates whether the restaurant is currently delivering (Yes/No).
  + **Switch\_to\_order\_menu:** Indicates whether users can switch to the order menu (Yes/No).
  + **Price\_range:** A numeric value indicating the price range category of the restaurant.
  + **Votes:** The number of votes or ratings/(feedback) received by the restaurant.
  + **Average\_Cost\_for\_two:** The average cost for two people dining at the restaurant.
  + **Rating:** The overall rating of the restaurant is based on user reviews.
  + **Datekey\_opening:** The date when the restaurant was opened.
  + **Day:** Day taken from Datekey\_opening.
  + **Month:** Month taken from Datekey\_opening.
  + **Year:** Year taken from Datekey\_opening.
  + **Cost\_for\_two\_in\_Rs.:** Average cost for two converted in Indian rupees.

**Country Description:**

* + **Country Code:** The code assigned to a specific country that helps identify the country in raw data.
  + **Country Name:** The respective country for the given country code.

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 a form of information that can be stored and identified based on their names or labels. It is a type of qualitative data that can be grouped into categories instead of being measured numerically.

* + Restaurant Name
  + City
  + Address
  + Locality
  + Locality Verbose
  + Cuisines
  + Currency
  + Has\_Table\_booking
  + Has\_Online\_delivery
  + Is\_delivering\_now
  + Switch\_to\_order\_menu
  + Country Name

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

**Method Used for cleaning the data :**

* **Formatting Of Date Column -** Column(U): Datekey\_Opening has dates in 21\_9\_2013 format. I used find and replace and replaced “\_” with “-”.

I also extracted 3 more columns from Datekey\_Opening named as

Day:🡪 =DAY(U2)

Month:🡪 =TEXT(U2,”mmm”)

Year:🡪 =YEAR(U2)

* **Deleting Blank Rows:**  Removing the rows Which contained blank cells

First I selected all the blank cell by GoTo and then right click on the blank cells and selecting delete whole rows.

* **Changing the Average\_cost\_of\_two into Indian currency:** I converted the currency in Indian rupees. So that it is easy to use later in project.

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

**VLOOKUP function:** In the table named country description, there are given country name with their country code. I used VLOOKUP function to extract the country out of country code.

=VLOOKUP(C2,'2)country description'!$A$1:$B$16,2,0)

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

**COUNTIF:** In sheet Analysis table named Number of Restaurants Opened in each Country I have counted the number of restaurants by using countif function.

=COUNTIF('1)Raw Data'!$D$2:$D$9528,'3)Analysis'!A24)

|  |  |
| --- | --- |
| **6) Number of restaurants Opened in each country** | |
| **Column1** | **Column2** |
| India | 8643 |
| Australia | 24 |
| Brazil | 60 |
| Canada | 4 |
| Indonesia | 21 |
| New Zealand | 40 |
| Philippines | 22 |
| Qatar | 20 |
| Singapore | 20 |
| South Africa | 60 |
| Sri Lanka | 20 |
| Turkey | 34 |
| United Arab Emirates | 60 |
| United Kingdom | 80 |
| United States of America | 419 |

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

**Pivot Table:** In the sheet name Analysis I have created a pivot table putting year in the rows and count of restaurant\_id in values.

|  |  |
| --- | --- |
| **7) Restaurants Opened in each Year** | |
| **Row Labels** | **Count of Restaurant** |
| 2010 | 7 |
| 2011 | 7 |
| 2012 | 10 |
| 2013 | 12 |
| 2014 | 4 |
| 2015 | 10 |
| 2016 | 5 |
| 2017 | 6 |
| 2018 | 7 |
| **Grand Total** | **68** |

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

**COUNTIFS:** In the sheet name Analysis I have counted the total number of restaurants in India with price range of 4.

=COUNTIFS('1)Raw Data'!$D$2:$D$9528,'3)Analysis'!F6,'1)Raw Data'!$Q$2:$Q$9528,4)

|  |  |
| --- | --- |
| **8)Restaurants in India in Price Range of 4** | |
| India | 388 |

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

**AVERAGEIF:** In the sheet name Analysis I have calculated the average number of voters in each country by using AVERAGEIF.

=AVERAGEIF('1)Raw Data'!$D$2:$D$9528,'3)Analysis'!F24,'1)Raw Data'!$R$2:$R$9528)

|  |  |
| --- | --- |
| **9)Average Voters in each Country** | |
| **Column1** | **Column2** |
| India | 137 |
| Australia | 111 |
| Brazil | 20 |
| Canada | 103 |
| Indonesia | 772 |
| New Zealand | 243 |
| Philippines | 407 |
| Qatar | 164 |
| Singapore | 32 |
| South Africa | 315 |
| Sri Lanka | 146 |
| Turkey | 431 |
| United Arab Emirates | 494 |
| United Kingdom | 205 |
| United States of America | 436 |

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.]**

* I have used the following formula to calculate the average rating for restaurants:

**Sheet Name:** Raw Data

Column🡪 AA

**=AVERAGE(IF(($Q$2:$Q$9528 < 4) \* ($N$2:$N$9528 = "Yes"), $T$2:$T$9528))**

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.**

* Using Conditional Formatting I have highlight the rows of four countries that I have suggested named: Australia, Canada, Singapore, Sri Lanka.
* In Conditional Formatting go to new new rule, click on “ Use a formula to determine which cells to format”.
* In formula field entered the formula

=$D2= “Sri Lanka”

=$D2= “Singapore”

=$D2= “Canada”

=$D2= “Australia”

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]**

**Method:**

* To create a new column with symbol of currency we will use MID function to extract the symbol.
* **=MID(L2, FIND("(", L2) + 1, FIND(")", L2) - FIND("(", L2) - 1)& S2**
* MID function extracts a given number of characters from the middle of a supplied text string based on the provided starting location.
* We gave the starting location through FIND, which will get the location of “(“ and start from next number

**Sheet Name:** Raw Data🡪 column:Z

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?**

**ARRAYFORMULA:** In sheet Raw Data I have counted the number of restaurants that do not offer online delivery, are in the lowest price range, and average cost for two people is less than 250 indian rupees. Using COUNTIFS function in cell AF3

{=COUNTIFS(N2:$N$9528,"No",$Q$2:$Q$9528,1,$Y$2:$Y$9528,"<=250")}

**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?**

**STRATEGY:**

* Country where the restaurants can be opened will be where the competiton is less and performance of restaurants are average i.e a lot of market can be captured.

**Analysis:**

* I used a pivot table, put countries in rows and count of restaurantsID in values and average of restaurants in values.
* I sorted count of restaurantsID in ascending order to see the countries where competiton is less.
* I also payed attention to countries where ratings of restaurants are average (above 3.5 and less than 4) because there will be huge scope of improvement.

**RESULT:**

* Australia
* Canada
* Singapore
* Sri Lnaka

**Visualization Method used:**

* Line Chart

**Sheet Name:** 4) Suggested Cities

|  |  |  |
| --- | --- | --- |
| **1) Analysis of Countries with lesser competition** | | |
| **Row Labels** | **Count of RestaurantID** | **Average of Rating** |
| Canada | 4 | 3.58 |
| Singapore | 18 | 3.59 |
| Sri Lanka | 19 | 3.87 |
| Australia | 20 | 3.74 |
| **Grand Total** | **61** | **3.73** |

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

**Strategy:**

* To find the cities which are suitable for opening restaurants in suggested countries, we should understand the performance of restaurants in the different cities

**Analysis:**

* Use Pivot Table.
* Countries and city in rows, count of restaurantID in values and average of ratings in values
* Cities whose ratings are between 3 to 3.5 are taken in consideration as it will be easy to capture the market there because the performance of others restaurants is average

**Sheet Name:** 4) Suggested Cities

|  |  |  |
| --- | --- | --- |
| **2) Cities in the suggested country** | | |
| **Row Labels** | **Count of RestaurantID** | **Average of Rating** |
| **Australia** | **24** | **3.66** |
| Armidale | 1 | 3.5 |
| Balingup | 1 | 3.2 |
| Beechworth | 1 | 4.6 |
| Dicky Beach | 1 | 3.6 |
| East Ballina | 1 | 4.1 |
| Flaxton | 1 | 3.5 |
| Forrest | 1 | 3.7 |
| Hepburn Springs | 2 | 3.8 |
| Huskisson | 1 | 4.1 |
| Inverloch | 1 | 3.7 |
| Lakes Entrance | 1 | 3.8 |
| Lorn | 1 | 3.6 |
| Macedon | 1 | 3.5 |
| Mayfield | 1 | 2.9 |
| Middleton Beach | 1 | 3.8 |
| Montville | 1 | 2.4 |
| Palm Cove | 1 | 4.4 |
| Paynesville | 1 | 2.6 |
| Penola | 1 | 3.4 |
| Phillip Island | 1 | 3.7 |
| Tanunda | 1 | 4.4 |
| Trentham East | 1 | 4.1 |
| Victor Harbor | 1 | 3.6 |
| **Canada** | **4** | **3.58** |
| Chatham-Kent | 1 | 3.7 |
| Consort | 1 | 3 |
| Vineland Station | 1 | 4.3 |
| Yorkton | 1 | 3.3 |
| **Singapore** | **20** | **3.58** |
| Singapore | 20 | 3.575 |
| **Sri Lanka** | **20** | **3.87** |
| Colombo | 20 | 3.87 |
| **Grand Total** | **68** | **3.69** |

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

**Analysis:**

* Calculated the average of ratings according to the suggested cities using aggregated function AVERAGEIF

**Formula used:**

* =AVERAGEIF('1)Raw Data'!$D$2:$D$9528,"Canada",'1)Raw Data'!$T$2:$T$9528)

**Sheet Name:** 4) Suggested Cities

**Visualization:** Doughnut Chart

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

**Analysis:**

* To calculate the total expenditure on food, I have used a pivot table
* Countries in columns and filter to suggested countries. Average of cost of two in rs in values.
* We can see that Singapore and Australia has higher expenditure on food.

**Sheet Name:** 4) Suggested Cities

**Visualisation:** Column Chart

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.**

**Analysis:**

* I have used pivot table. Put Countries and Restaurant Name in the rows, Average of ratings in values and average of cost for two in rs in values.
* Sort the table in descending order of ratings.
* Restaurants who are our biggest competitor(Rating >4) are highlighted by green.
* Restaurants that are in the lower bracket are highlighted by yellow and red colors.

**Sheet Name:** 5) Competitor Analysis

|  |  |  |
| --- | --- | --- |
| **Australia Competitor Analysis** | | |
| **Row Labels** | **Average of Rating** | **Average of Cost\_for\_two\_in\_Rs.** |
| **Australia** | **3.658333333** | **1312.541667** |
| Bridge Road Brewers | 4.6 | 1090 |
| 1918 Bistro & Grill | 4.4 | 1635 |
| Vivo Bar and Grill | 4.4 | 1635 |
| 5 Little Pigs | 4.1 | 1090 |
| The Belle General | 4.1 | 1090 |
| Pig and Whistle | 4.1 | 1090 |
| Funkey Monkey | 3.8 | 381.5 |
| Blue Bean Love Cafe | 3.8 | 1090 |
| Three Anchors | 3.8 | 1635 |
| La Trattoria of Lavandula | 3.8 | 381.5 |
| Bespoke Harvest | 3.7 | 1090 |
| Beach Box Cafe | 3.7 | 381.5 |
| Mad Cowes Cafe | 3.7 | 1090 |
| Stillwater on Belmore | 3.6 | 1090 |
| The Giggling Goat | 3.6 | 381.5 |
| Anchorage Cafe Restaurant Wine Bar | 3.6 | 1090 |
| Mr. | 3.5 | 1090 |
| Whitebull Hotel | 3.5 | 1090 |
| Flaxton Gardens | 3.5 | 1635 |
| DiVine | 3.4 | 1090 |
| Taste of Balingup | 3.2 | 1090 |
| Star Buffet | 2.9 | 1090 |
| Pier 70 | 2.6 | 6540 |
| Poets Cafe | 2.4 | 1635 |
| **Grand Total** | **3.658333333** | **1312.541667** |

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

**Analysis:**

* I have used a pivot table. Put country and cuisines in rows and average of ratings in values
* Cuisines that have rating more than 4 are most in demand so we should also focus in those cuisines.

**Cuisines:**

* Pizza, Seafood, Italian, Bakery.

**Sheet Name:** 5) Competitor Analysis

|  |  |
| --- | --- |
| **Cuisine Analysis** | |
| **Row Labels** | **Average of Rating** |
| **Australia** | **3.7** |
| Asian | 2.9 |
| Australian | 4.1 |
| Bar Food, Modern Australian | 3.8 |
| Bar Food, Steak | 3.5 |
| Breakfast, Coffee and Tea | 3.7 |
| Breakfast, Coffee and Tea, Modern Australian | 3.7 |
| Breakfast, Modern Australian | 4.1 |
| Burger, Coffee and Tea, Modern Australian | 3.7 |
| Cafe | 3.8 |
| Cafe, Australian | 3.7 |
| Cafe, Coffee and Tea, Modern Australian | 3.8 |
| Cafe, Coffee and Tea, Sandwich | 3.4 |
| Coffee and Tea, Modern Australian | 2.4 |
| Coffee and Tea, Tapas, Australian | 3.6 |
| Coffee and Tea, Tea, Modern Australian | 3.6 |
| Italian, Fusion, Cafe | 3.8 |
| Mediterranean, Seafood | 4.4 |
| Modern Australian | 2.9 |
| Modern Australian, Australian | 4.4 |
| Pizza, Bar Food | 4.6 |
| Tea, Modern Australian | 3.5 |
| **Canada** | **3.6** |
| Asian | 3.3 |
| Chinese, Canadian | 3.0 |
| Italian, Mediterranean, Pizza | 4.3 |
| Japanese, Sushi | 3.7 |
| **Singapore** | **3.6** |
| American | 3.1 |
| American, Bakery, European, Burger, Fusion | 3.8 |
| American, Japanese, Singaporean | 3.2 |
| American, Mexican | 3.2 |
| American, Steak | 4.0 |
| Asian, Continental, Seafood | 3.8 |
| Bakery | 4.2 |
| Cafe | 3.7 |
| Cafe, Spanish, Turkish, Greek | 3.2 |
| Chinese, Continental, Singaporean | 3.4 |
| Chinese, Seafood, Cantonese, Dim Sum | 3.9 |
| Finger Food | 3.9 |
| French | 3.9 |
| French, Mediterranean, European | 3.8 |
| Italian | 4.1 |
| Italian, French, Bakery, Cafe | 3.2 |
| Singaporean, Australian, German | 3.1 |
| Singaporean, Chinese, Seafood, Malay, Indian | 3.0 |
| Western, Fusion, Fast Food | 3.2 |
| **Sri Lanka** | **3.9** |
| American, Chinese, North Indian | 2.5 |
| American, Fast Food, Steak, Beverages | 4.2 |
| American, Steak | 4.0 |
| Cafe, Fast Food, Beverages | 3.8 |
| Cafe, Italian | 3.6 |
| Cafe, Sri Lankan, Continental, American | 4.0 |
| Chinese | 3.4 |
| Continental, American | 4.1 |
| Continental, American, Seafood | 4.2 |
| Desserts, Bakery | 4.2 |
| Desserts, Ice Cream | 4.1 |
| Fast Food | 4.1 |
| Italian, Cafe, Desserts | 3.7 |
| Juices, Desserts | 4.5 |
| Malaysian, North Indian, Sri Lankan | 3.5 |
| Middle Eastern, Arabian | 4.2 |
| North Indian, Chinese, Sri Lankan | 2.4 |
| Seafood | 4.9 |
| Seafood, Italian | 4.0 |
| Sri Lankan | 4.0 |
| Grand Total | 3.7 |

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

**Method Used:** Used COUNTIFS to count the number of restaurants that offer online delivery and table booking.

**Formula:**

**Online Delivery🡪**

**=COUNTIFS('1)Raw Data'!$D$2:$D$9528,"Australia",'1)Raw Data'!$N$2:$N$9528,"Yes")**

**Table Booking🡪**

**=COUNTIFS('1)Raw Data'!$D$2:$D$9528,"Australia",'1)Raw Data'!$M$2:$M$9528,"Yes")**

**Observation:** From the suggested countries no one offers online delivery and table booking facilities. It would be an advantage for us to have these facilities as this will give us an upper hand over the competition.

**Sheet Name:** 5) Competitor Analysis

|  |  |  |
| --- | --- | --- |
| **Delivery Analysis** | | |
|  | Online Delivery | Table Booking |
| **Australia** | 0 | 0 |
| **Canada** | 0 | 0 |
| **Singapore** | 0 | 0 |
| **Sri Lanka** | 0 | 0 |

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?**

**Analysis:**

* I have used a pivot table. Put ratings in rows and average of cost for two in rupees in values
* Used CORREL to calculate the correlation between ratings and cost
* Value is positive that means restaurants with higher ratings has higher cost.
* By taking a scatter plot rating and cost we can see that ratings is directly proportional to cost. That means we can keep the rates higher of cuisines but we have to provide high class facilities also.

**Sheet Name:** 5) Competitor Analysis

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

**Analysis:** I have used a pivot table. Put price range in rows and count of restaurants in values. I have used a Duoghnut chart for visualization.

**Sheet Name:** 5) Competitor Analysis

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]**

**TASK:**

* Zomato wants to open new restaurants in different countries. They are asking us to suggest a few countries by looking through the data and give reasons for decisions.

**APPROACH:**

**Data Cleaning:**

* Check for missing data in the table, handle the missing values by using mode, average etc.
* Remove any duplicate values.
* Check for any spelling error using filter.

**Preparing the Data:**

* Convert the datekey column into date format using find and replace.
* Add columns like day month year using the datekey column.
* Add a country column based on the country codes given in data.
* Convert the average cost of two in a single currency.

**Analysing:**

* Find countries which have less comptetion by counting the number of restaurants.
* I will also consider the countries which have lower ratings.
* Geographical area of country will also be taken in consideration as, if a country has less restaurants but the area is small and another country has a little more number of restaurants but more geographical area. So per kilometre wise the restaurants are less.
* I will check the most loved cuisines for my restaurants.
* Pick the cities in the respective countries which have above average ratings.
* Check how restaurants have opened up in previous year in the selected countries to know the competiton better.
* At what price should we open the restaurants.

**Visualisation:**

* Interpret the analysis above.
* Use different charts to analyse the above for a better understanding
* Use filterin country, price range, year and average cost.

**Conclusion:**

* Write conclusion based on the above interpretation.

**THE END**