

Objective Questions:

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

◆ Inconsistent/blank data found using filter option

LocalityVerbose	Longitude	Latitude	Cuisines	✕	Currency	Has_Table_bookin	Has_Online
Miller, Miller	-98.9891	44.5158			Dollar(\$)	No	No
Albany, Albany	-84.1759	31.5882			Dollar(\$)	No	No
Albany, Albany	-84.154	31.5772			Dollar(\$)	No	No
Dahlonega, Gainesville	-83.9858	34.5318			Dollar(\$)	No	No
Macon, Macon	-83.627979	32.83641			Dollar(\$)	No	No
Winter Park, Orlando	-81.36526	28.596682			Dollar(\$)	No	No
Albany, Albany	-84.1534	31.5751			Dollar(\$)	No	No
Kaimuki, Rest of Hawaii	-157.813432	21.284586			Dollar(\$)	No	No
Tybee Island, Savannah	-80.848297	31.99581			Dollar(\$)	No	No

To Clean the data for further analysis, we filled it with taking the most occurring cuisine in the corresponding column.

[illegible]

Here, match function was used to find the position of each element in the range. The “IF” function excludes any empty cells from consideration. Mode function was used to calculate the mode positions(occurring). Index function was used to retrieve the corresponding text based on the mode position. Next, we filled the empty blanks with the most occurring, i.e “North Indian” to make the data complete.

- 2) Using the LookUp functions, fill up the countries in the original data using the country code.

- ◆ Using the xlookup formula from the given sheets,

=XLOOKUP(D2,Sheet2!\$A\$2:\$A\$16, Sheet2!\$B\$2:\$B\$16)

Here we used the Xlookup formula while matching the country code from the referenced second sheet and filled the country column(added a new column) in the main datasheet.

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

◆ Herewith lies the pivot table,

question 3, 4											
Count of RestaurantName	Column Labels										
Row Labels	2010	2011	2012	2013	2014	2015	2016	2017	2018	Grand Total	
Australia	4	1	3	6		4	2	1	3	24	
Brazil	5	12	2	8	11	9	5	4	4	60	
Canada		1		1		1		1		4	
India	995	995	911	954	946	918	938	992	1003	8652	
Indonesia	1	5		1	4	3	1	1	5	21	
New Zealand	4	6	4	2	4	4	4	5	7	40	
Philippines	6	3	2	1	2	1	1	2	4	22	
Qatar	4	1	2	4	2	2	4		1	20	
Singapore	2	3	4	1	2	2	1	2	3	20	
South Africa	4	4	7	8	5	7	10	9	6	60	
Sri Lanka	1	2	3	4	2	3	2	2	1	20	
Turkey	3	2	1	5	6	4	4	3	6	34	
United Arab Emirates	4	3	16	6	9	8	2	6	6	60	
United Kingdom	9	6	12	10	11	5	7	12	8	80	
United States of America	38	54	55	50	47	53	46	46	45	434	
Grand Total	1080	1098	1022	1061	1051	1024	1027	1086	1102	9551	

Here we've taken the pivot of the entire datasheet. Using rows as Country and values as count of restaurant Name we've found the total Count in the right most section of the pivot table.

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

◆ Using the same datasheet as above:

question 3, 4											
Count of RestaurantName	Column Labels										
Row Labels	2010	2011	2012	2013	2014	2015	2016	2017	2018	Grand Total	
Australia	4	1	3	6		4	2	1	3	24	
Brazil	5	12	2	8	11	9	5	4	4	60	
Canada		1		1		1		1		4	
India	995	995	911	954	946	918	938	992	1003	8652	
Indonesia	1	5		1	4	3	1	1	5	21	
New Zealand	4	6	4	2	4	4	4	5	7	40	
Philippines	6	3	2	1	2	1	1	2	4	22	
Qatar	4	1	2	4	2	2	4		1	20	
Singapore	2	3	4	1	2	2	1	2	3	20	
South Africa	4	4	7	8	5	7	10	9	6	60	
Sri Lanka	1	2	3	4	2	3	2	2	1	20	
Turkey	3	2	1	5	6	4	4	3	6	34	
United Arab Emirates	4	3	16	6	9	8	2	6	6	60	
United Kingdom	9	6	12	10	11	5	7	12	8	80	
United States of America	38	54	55	50	47	53	46	46	45	434	
Grand Total	1080	1098	1022	1061	1051	1024	1027	1086	1102	9551	

Here, we've taken the same datasheet as above with having columns in year. Hence we could find the count of restaurants opened in each year. Additionally, the year format in the main datasheet was unusable so it was rectified using the left formula.

7) Suggest few countries where the team can open newer restaurants with lesser competition.

- ◆ Using the concept of normalization and building a composite score.

Row Labels	Sum of Price_range	Sum of Votes	Count of RestaurantID	Sum of Rating	Normalization				
Australia	51	2674	24	87.8	price range	votes	rating	composite score	
Brazil	204	1177	60	230.8	0.002754266	0.001906044	0.003068	0.016618758	
Canada	10	412	4	14.3	0.013032379	0.000644617	0.009037	0.053821016	
India	14896	1187163	8652	23970.8	0	0	0	0	
Indonesia	62	16214	21	90.2	1	1	1	6	
New Zealand	126	9721	40	170.5	0.003493215	0.013315346	0.003168	0.029806503	
Philippines	74	8963	22	98.3	0.007792557	0.007844105	0.00652	0.042989672	
Qatar	73	3276	20	81.2	0.004299342	0.007205387	0.003506	0.026323136	
Singapore	73	638	20	71.5	0.004232164	0.002413312	0.002793	0.019255325	
South Africa	215	18910	60	252.6	0.004232164	0.000190436	0.002388	0.015817748	
Sri Lanka	57	2929	20	77.4	0.013771329	0.015587095	0.009947	0.072971734	
Turkey	96	14670	34	146.2	0.003157329	0.002120917	0.002634	0.016337397	
United Arab Emirates	193	29611	60	254	0.00577724	0.012014315	0.005506	0.040086233	
United Kingdom	220	16439	80	328	0.01229343	0.02460415	0.010006	0.079207916	
United States of America	888	185848	434	1740.9	0.014107215	0.013504939	0.013095	0.08100307	
Grand Total	17238	1498645	9551	27614.5	0.058981593	0.156255187	0.072072	0.490435267	

Since competition can't only be based on a single criteria, three criterias were used in tandem with each other. What is normalization? Normalization is the process of scaling and transforming data into a standard range(composite score). Normalization formula:

	G	H
price range	0.002754266	0.001

Figuratively, (current cell value - min value of entire range)/ (max value - min value).

Here, Normalization was used for sum(Overall Impact) and not average(collective opinion). After normalizing all the criterias (price range, votes, rating), we weight them according to subjective importance. Given rating as the maximum weightage, we multiply it by 3 and 2 for price and 1 for votes and form a standard single column. The formula used here:

	J	K
	composite score	
	0.016618758	

Figuratively, $\text{rating} \times 3 + \text{price} \times 2 + \text{votes} \times 1$.

Based on the composite score, we've built a rank in the order of lesser to higher competition using the rank formula.

countries	composite scores		rank		
Australia	0.016618758		4		canada
Brazil	0.053821016		10		australia
Canada	0		1		indonesia
India	6		15		
Indonesia	0.029806503		7		
New Zealand	0.042989672		9		
Philippines	0.026323136		6		
Qatar	0.019255325		5		
Singapore	0.015817748		2		
South Africa	0.07297134		11		
Sri Lanka	0.016337397		3		
Turkey	0.040086233		8		
United Arab Emirates	0.079207916		12		
United Kingdom	0.08100307		13		
United States of America	0.490435267		14		

The lowest number indicates the least competition. According to this record, the least competitive country to set up a restaurant would be in Canada and the most would be in India. Similarly If I'd have taken the rating as average and not sum, India would've been the least competitive because of similar numbers all around.

Considering this list, I've chosen Canada, Australia and Indonesia as the least competitive while rushing past the obvious numbers in the rank. Canada is the prime choice because it has only 4 restaurants up to date. Second as Australia because the sum of votes and price range are in the lower bracket giving the new restaurants the opportunity to transcend limits. Indonesia being the last choice. Although expensive, the cuisines are limited and so are the number of restaurants.

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

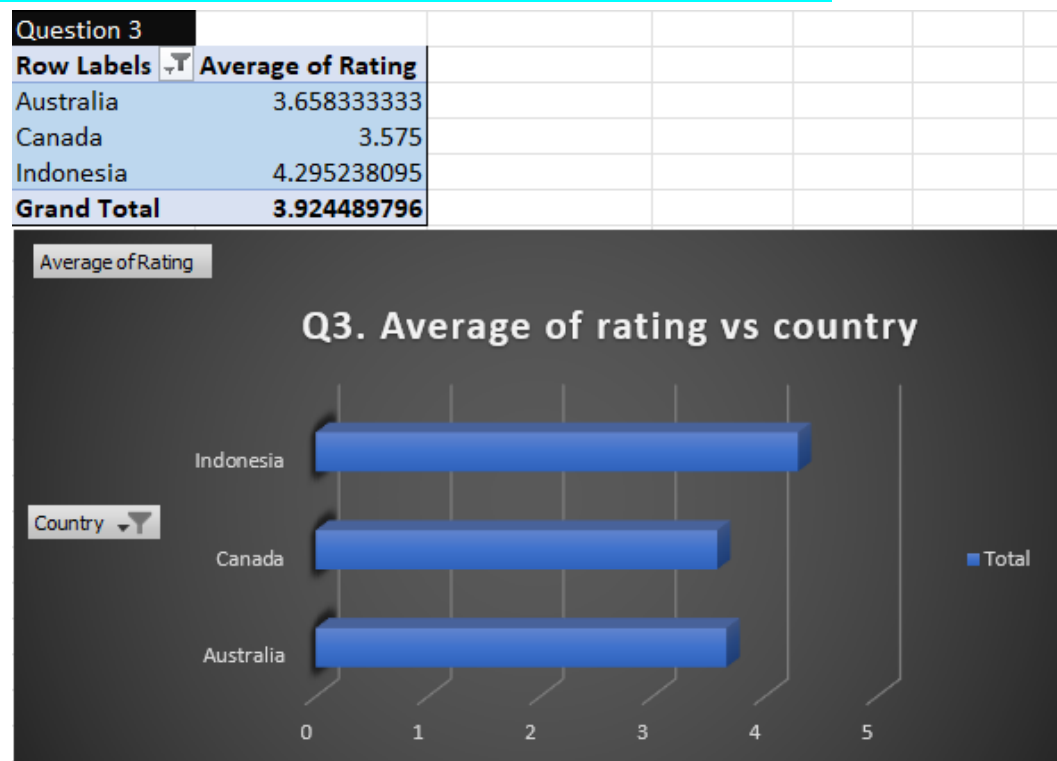
◆ Pivot Function was used and the normalization explained above. In this scenario though, the least competitive cities weren't taken but the most suitable instead.

Question 2			
Row Labels	Sum of Rating	Sum of Price_range	Sum of Votes
Hepburn Springs	7.6	3	285
Palm Cove	4.4	3	381
Tangerang	8.6	6	2367
Tanunda	4.4	3	339
Yorkton	3.3	2	26
Grand Total	28.3	17	3398

Two cities were taken from Australia and one each from Indonesia and Canada.

9) According to the countries you suggested, what is the current quality in terms of ratings for restaurants that are opened there?

- ◆ Average of rating was taken here for the countries selected.



- 10) Also what is the current expenditure on the food in the suggested countries, so that we can keep our financial expenditure in control?

- ◆ Here, average cost for two was taken as the factor for deciding current expenditure.

Question 4

Row Labels	Average of Average_Cost_for_two	Hepburn Springs	Palm Cove	Tangerang	Tanunda	Yorkton	Grand Total
Australia		13.5	30		30		21.75
Canada						25	25
Indonesia				250000			250000
Grand Total		13.5	30	250000	30	25	71444.57143

Here, Country was taken in rows, Average of average cost for two in values and City in columns.

- 11) Come up with the names of restaurants from the recommended states who are our biggest competitors and also those which are rated in the lower brackets, i.e. 1-2 or 2-3.

- ◆ Competition will be decided based on ratings that lie between 4-5.

Question 5	
City	(Multiple Items)
Row Labels	Max of Rating
1918 Bistro & Grill	4.4
Arigato Sushi	3.3
Blue Bean Love Cafe	3.8
La Trattoria of Lavandula	3.8
Onokabe	3.7
Talaga Sampireun	4.9
Grand Total	4.9

Here, we've the darker shade as our biggest competitor, i.e, 1918 bistro & Grill and Talaga Sampireun and we have no restaurants in the lower bracket, 1-3.

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

- ◆ By filtering the ratings to 4-5 we can focus on the cuisines that have effectively higher ratings on cuisines.

Row Labels	Count of Cuisines
1-2	2151
2-3	1430
3-4	4583
4-5	1378
Grand Total	9542

Here, there are 1378 cuisines that lie within the spectrum of our focus. Now another question arises. Are there cuisines that lie within the same 4-5 rating as well as 1-2 rating? Yes there are. But we still have more cuisines in the 4-5 range that are not common with cuisines in the 1-2 range. Hence, choice of cuisines may affect the restaurant ratings.

Here is the number of cuisines in the 4-5 rating range: 785

table booking? Does that affect the customer's ratings?

- ◆ Depending on the data we can figure out the average rating for both the conclusions.

f_x	=AVERAGE(Sheet1!T2:T9552)
H	
Question 7 - Yes	
	2.89126793

Here, we've taken the average rating of the entire main datasheet as the threshold.: 2.812
Then we've utilized the average if function to determine average rating for online delivery and table booking. If they're greater than the average number of the entire datasheet and also greater than the other counterpart then yes, otherwise no.

	2.89126793
	Table booking average rating
Yes	3.482556131
No	2.809686644
	Online delivery average rating
Yes	3.288004896
No	2.754309859

Here, we've the rating for "Yes" for both online delivery and table booking above the threshold signifying its effect on customer ratings.

14) Should the team keep the rate of cuisines higher? Will that affect the feedback? According to our data are the rate of cuisines and ratings, correlated?

- ◆ If correlated, we can find the correlation between price and ratings.

	A	B	C
13			
14	Question 8		
15	0.058957033		

Since value is close to 0, the data has little to no correlation; negligible linear correlation. Hence, the team shouldn't make the rate of cuisines higher.

15) What is the distribution of the number of restaurants of different price ranges in all the countries?

We can find the distribution using pivot table and a 100% stacked column chart to find the differences.

Question 9						
Count of RestaurantID		Column Labels				
Row Labels		1	2	3	4	Grand Total
Australia		4	14	5	1	24
Brazil		2	7	16	35	60
Canada			3		1	4
India		4295	2858	1111	388	8652
Indonesia			1	20		21
New Zealand		3	4	17	16	40
Philippines			1	12	9	22
Qatar			1	5	14	20
Singapore			1	5	14	20
South Africa			4	17	39	60
Sri Lanka			6	11	3	20
Turkey			11	18	5	34
United Arab Emirates			9	29	22	60
United Kingdom		4	28	32	16	80
United States of America		136	165	110	23	434
Grand Total		4444	3113	1408	586	9551

Here, we've countries in rows, count of restaurants in values and price range in columns. Effectively, we can use a stacked column chart.

