7/10/2023

Market Saturation Report on Restaurants in

Bengaluru (aka Bangalore), India

Tabitha Hagen
UNCW BAN 530 APPLICATIONS IN BUS ANALYTICS

Market Saturation: Restaurants in Bengaluru, India

Objectives –

- Analyze Current Trends in Cuisine by city using Tableau or JMP.
  - What is the cuisine that appears most within a given city?
  - What are the reviews about the restaurants in each city?
- Explore Customer Preferences of attributes by city using Tableau or JMP.
  - Do most restaurants have delivery in each city?
  - Do most restaurants have online reservations in each city?
  - What are the most common things Customer reviews talk about in each city?
- Examine Types of restaurants in the area by city using Tableau or JMP.
  - Which type of restaurant is most present in each city?
    - What happens if more of this type of restaurant is added?
    - What happens if some of this type of restaurant closes?
  - Which type of restaurant is least present in each city?

Descriptive Analytics –

Data Cleaning and Understanding

- Preprocessing Data Analysis

Initial status of data – The data set was initially given in the comma-separated values (CSV) format which is a text file that uses a delimiter to separate values. The format did not translate well in Excel with its almost half a gigabyte size, primarily due to the Restaurant Reviews. So, the CSV file was opened in RStudio using the R Language to be cleaned with the name, approximate cost for 2 people, online order, reservation, type of restaurant, liked dishes, cuisines, type of listing and city listed variables.

Translation Corrections – When the data was mined from the original websites and databases, there was a reoccurring error in the titles of certain names with foreign accent marks. The translation errors did present themselves in a reoccurring manner so that a “quick fix” could be implemented using an expected formula. In R Studio, a manual inspection of the titles in ABC order revealed many types of additional characters in the titles which were systematically removed using manual code (see *Figure 1*) so that all that remained was the name of the Restaurant.

```
# Replace String with Another String to correct data import due to accents in names
```{r message=FALSE, error=FALSE}
zomato_r_working$name[zomato_r_working$name == "100Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â°C"] <-
"100°C"

zomato_r_working$name[zomato_r_working$name == "Andy's
CafÂ·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â©"] <- "Andy's Cafe"

zomato_r_working$name[zomato_r_working$name == "Brother?s Bar & Kitchen"] <- "Brother's Bar &
Kitchen"

zomato_r_working$name[zomato_r_working$name ==
"BrotherÂ·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·s Bar & Kitchen"] <-
"Brother's Bar & Kitchen"

zomato_r_working$name[zomato_r_working$name ==
"BrotherÂ·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·s Bar & Kitchen"] <-
"Brother's Bar & Kitchen"

zomato_r_working$name[zomato_r_working$name == 'Bohra Bohra
CafÂ·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â·Â©'] <- 'Bohra Bohra'
```

Figure 2 Correcting Translation Errors

<u>Missing Values</u> – Using the R Studio, missing values were analyzed for replacement by median values, averages, or simply deleting the values. The Summary function quickly noted that there were 346 out of the 51,717 Approx. Cost (for 2 people) values were NA's or missing values. Yet, when the function was applied to create a new median value for each of these missing values, it replaced 28,282 values which was more than half the observations. It was determined that the average would do the same and decided that it would be a more valid result to simply look at this data set without the missing information being imputed. After the simple function to exclude the NA's, there were still valid. 23,435 observations to be analyzed.

```
# Remove missing data, create binary balues, and create factors
```{r message=FALSE, error=FALSE}
# Excludes every row containing NA
zomato_clean = zomato_r_working %>% na.exclude(zomato_clean)
# 23,435 observations or rows of data

# Change no to 0 and yes to 1 for online_ordering and book_table
zomato_clean = zomato_clean %>%
  mutate (online_order= case_match(online_order,
    c("No", "no", "NO") ~ 0,
    .default = 1)) # otherwise have yes equal 1

zomato_clean = zomato_clean %>%
  mutate (book_table= case_match(book_table,
    c("No", "no", "NO") ~ 0,
    .default = 1)) # otherwise have yes equal 1

# Convert character variable to the type factor
zomato_clean = zomato_clean %>% mutate_if(is.character, as_factor) # mutate if it is a character

str(zomato_clean) #compactly displaying the internal structure of a R object
```
```

*Figure 1 Missing values, Binary variables & Factors.*
```

<u>Creating Factors</u> – Since most of the original variables were of the character type, a simple reclass was made to turn the Yes's and No's into binary values for the purpose of exploring the data (see *Figure 2*).

<u>Save new File for other Software Programs</u> – This cleaned file will be used in other Software applications, so it was necessary to save it in a usable format, so it was saved as ***zomato_clean.csv***.

<u>Explore Outlier Data</u> - in JMP Analysis, which can only look at Outliers in numerical values, the Distribution and the Explore Outliers features were used to look at values far away from the Approx. Cost (for 2 people) median of 600. Considering that Fine Dining restaurants would have a cost far different than Quick Bites, it was determined that 3 data points would be treated as outliers and not calculated in the analysis (Approx cost at $5,000 -$6,000).

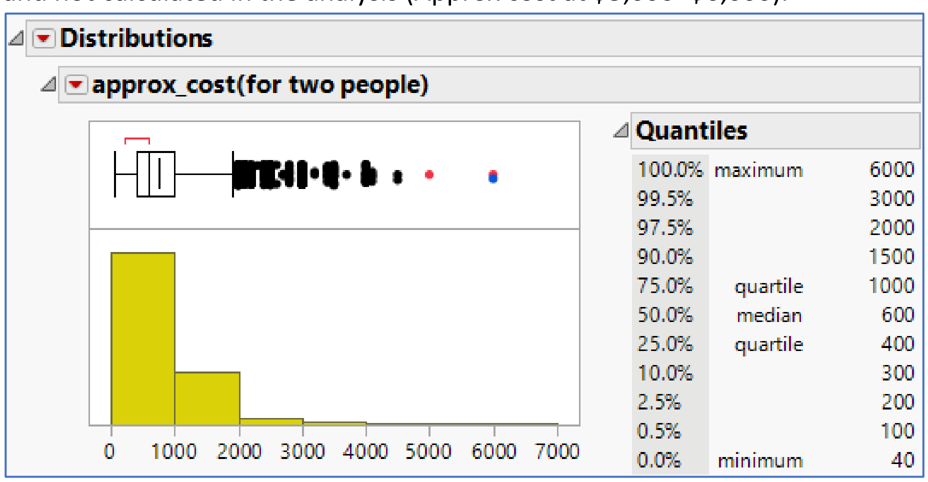

Figure 3 The Distribution shows the Outliers in red/blue on the left which are specified in the chart below.

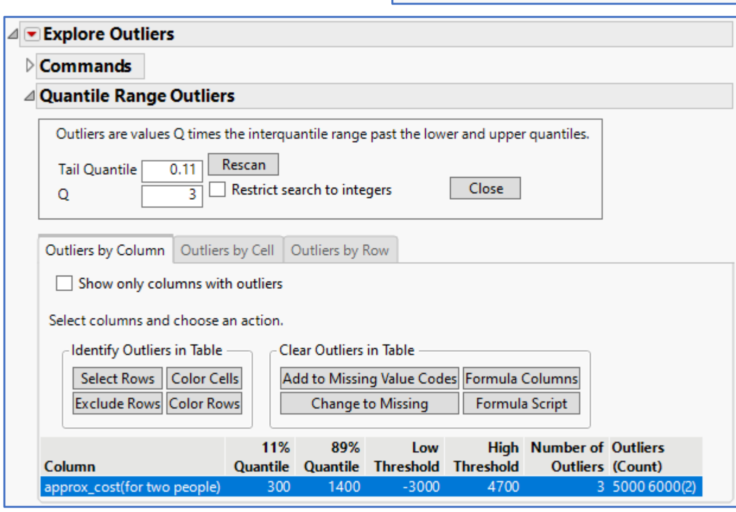

Figure 4 IDENTIFYING Specific Outliers

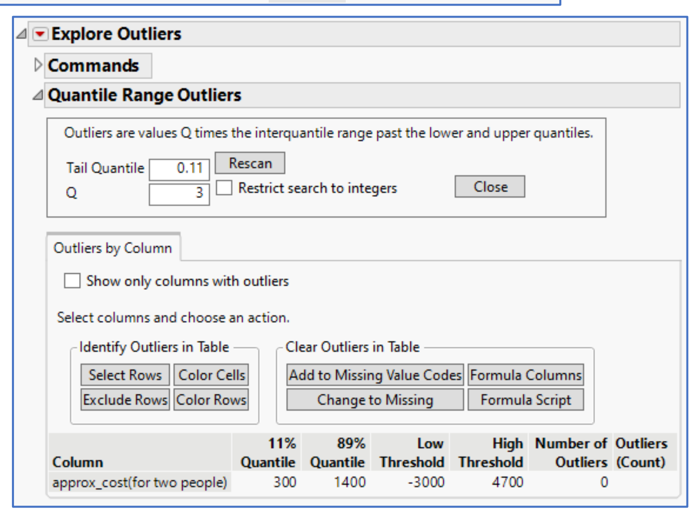

Figure 5 AFTER Outliers Added to Missing Value Codes

Market Saturation: Restaurants in Bengaluru, India

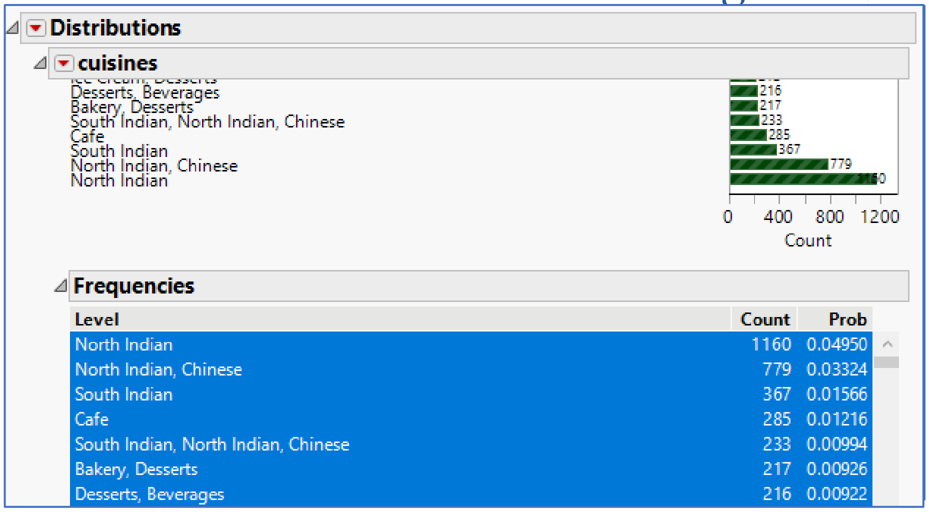

Figure 6 Overall Analysis shows North Indian and North Indian Chinese Restaurants far outnumber the others.

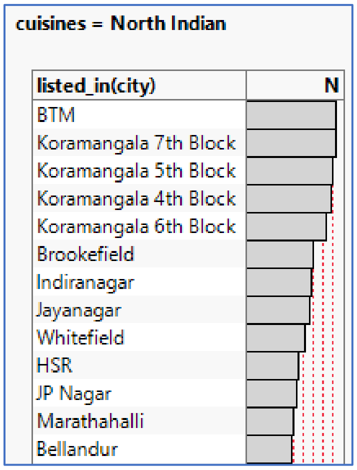

Figure 7 Cities where North Indian Food is most popular.

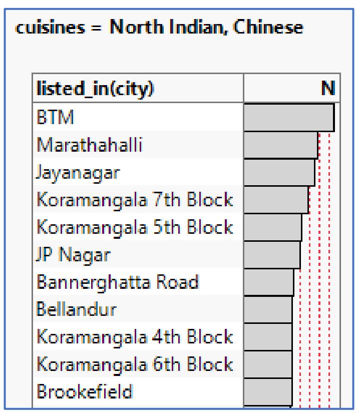

Figure 8 Cities where North Indian, Chinese Food is most popular.

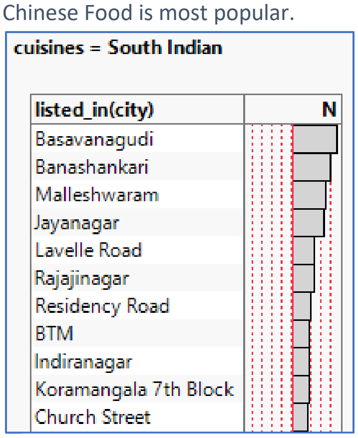

Figure 9 Cities where South Indian Food is most popular.

- Cuisine Segmentation Analysis

Viewing the general distribution of the types of Cuisines (see Figure 6 above), both the vertical bar chart and the Frequency table show that North Indian cuisine is the most popular by almost double that of North Indian combined with Chinese Cuisine (4.95% of the total restaurants reported). Chinese food is widely popular because commodity trading is prevalent due to being in the southern peninsula where a large portion of the population lives (3.32% of the total restaurants reported). This is followed in third place by South Indian cuisine. Considering where Bengaluru is located in the densely populated southern area of the Indian Continent, this makes sense (1.57% of the total restaurants reported). (note: the separation of North Indian and Nort Indian, Chinese cuisine is probably due to a customer survey where multiple cuisine options where allowed to be chosen creating dozens of instances of overlap in the cuisine categorical variable).

In JMP's Tabulate feature, with the appropriate Data Filter, it is easy to see in Figure 7 that North Indian Cuisine is most popular in the cities of BTM, Koramangala (4th, 5th, 6th, & 7th blocks), Brookefield, Indiranagar and Jayanagar (see Figure 7) with a total of 1160 restaurants in Bengaluru. North Indian, Chinese cuisine was second most popular with 779 restaurants (see Figure 8). Additionally, South Indian cuisine was third most popular in Bengaluru with 367 restaurants (see Figure 9).

- Customer Preferences Analysis

When the online ordering data was cleaned in the preprocessing step, values were transformed into binary values in order to easily visualize the results (see figure 10). It can be easily seen that 70% of the restaurants reported have online ordering. As the COVID-19 pandemic hit, citizens were ordered to stay at home and thus, more online ordering was demanded by customers. As shown on the left in Figure 11, the city HSR took this to heart and leads the other cities reported by a large margin with an 81% online ordering. However, cities such as MG Road, Church Street, Brigade Road, Lavelle Road, and Residency Road are falling behind with only 54% to 58% online ordering.

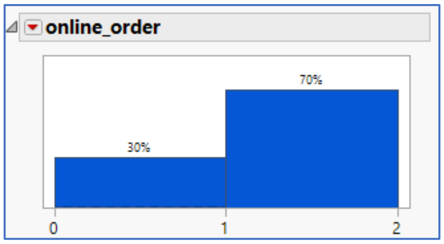

Figure 10 Online ordering where "no" = 0-1 and "yes" = 1-2

listed_in(city)		N	Mean	% of Total
BTM	online_order	1439	0.76	6.66%
Koramangala 7th Block	online_order	1391	0.71	6.04%
Koramangala 4th Block	online_order	1359	0.74	6.12%
Koramangala 5th Block	online_order	1324	0.73	5.85%
Koramangala 6th Block	online_order	1265	0.73	5.64%
Jayanagar	online_order	1090	0.77	5.12%
Indiranagar	online_order	1025	0.64	4.01%
MG Road	online_order	1017	0.57	3.51%
Church Street	online_order	983	0.56	3.36%
Brigade Road	online_order	970	0.58	3.41%
Lavelle Road	online_order	898	0.55	2.99%
Residency Road	online_order	846	0.54	2.76%
JP Nagar	online_order	829	0.76	3.83%
Old Airport Road	online_order	760	0.67	3.09%
HSR	online_order	750	0.81	3.69%
Whitefield	online_order	635	0.70	2.70%
Marathahalli	online_order	622	0.74	2.81%
Basavanagudi	online_order	606	0.75	2.76%
Brookefield	online_order	600	0.77	2.82%
Bannerghatta Road	online_order	550	0.76	2.54%
Frazer Town	online_order	546	0.77	2.56%
Kammanahalli	online_order	534	0.77	2.49%
Kalyan Nagar	online_order	521	0.78	2.47%
Bellandur	online_order	508	0.78	2.40%
Sarjapur Road	online_order	508	0.75	2.32%
Malleshwaram	online_order	498	0.69	2.08%
Rajajinagar	online_order	378	0.66	1.51%
Banashankari	online_order	373	0.78	1.77%
Electronic City	online_order	326	0.70	1.38%
New BEL Road	online_order	284	0.76	1.32%
All	online_order	23435	0.70	100.00%

Figure 11 Online ordering where the larger the mean, the more online ordering is used by city.

Similar to above, the book_table or reservations data was also turned into binary values for data visualizations. On the left in Figure 12, you can see that only 26% of the restaurants reported have online reservations. Customers were reluctant to eat in a public establishment during the pandemic as they were afraid their wait staff, customers or the cleanliness of the restaurant would cause them to get sick. Looking at Figure 13, we can see that only 5 cities reported having online reservations of 30% or higher. More importantly, the 5 cities of Basavanagudi, Bannerghatta Road, Kammanahalli, Benashankari, and New BEL Road have this as an almost forgotten amenity and should seriously look at changing this attribute.

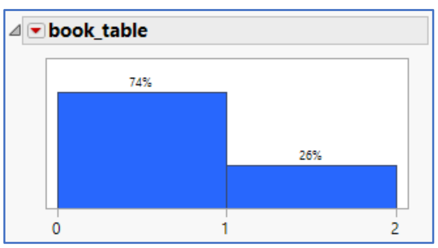

Figure 12 Online ordering where "no" = 0-1 and "yes" = 1-2

listed_in(city)		N	Mean	% of Total
BTM	book_table	1439	0.26	5.99%
Koramangala 7th Block	book_table	1391	0.27	6.03%
Koramangala 4th Block	book_table	1359	0.26	5.77%
Koramangala 5th Block	book_table	1324	0.27	5.75%
Koramangala 6th Block	book_table	1265	0.25	5.23%
Jayanagar	book_table	1090	0.23	4.00%
Indiranagar	book_table	1025	0.30	5.05%
MG Road	book_table	1017	0.34	5.64%
Church Street	book_table	983	0.35	5.60%
Brigade Road	book_table	970	0.34	5.43%
Lavelle Road	book_table	898	0.34	5.02%
Residency Road	book_table	846	0.35	4.76%
JP Nagar	book_table	829	0.21	2.86%
Old Airport Road	book_table	760	0.29	3.62%
HSR	book_table	750	0.22	2.70%
Whitefield	book_table	635	0.29	2.99%
Marathahalli	book_table	622	0.23	2.34%
Basavanagudi	book_table	606	0.19	1.88%
Brookefield	book_table	600	0.20	1.98%
Bannerghatta Road	book_table	550	0.19	1.69%
Frazer Town	book_table	546	0.22	1.95%
Kammanahalli	book_table	534	0.19	1.64%
Kalyan Nagar	book_table	521	0.21	1.79%
Bellandur	book_table	508	0.26	2.16%
Sarjapur Road	book_table	508	0.24	2.00%
Malleshwaram	book_table	498	0.25	2.00%
Rajajinagar	book_table	378	0.23	1.43%
Banashankari	book_table	373	0.11	0.65%
Electronic City	book_table	326	0.25	1.32%
New BEL Road	book_table	284	0.16	0.75%
All	book_table	23435	0.26	100.00%

Figure 13 Reservations where the larger the mean, the more reservations are used by city.