7/10/2023

Tabitha Hagen

UNCW BAN 530 APPLICATIONS IN BUS ANALYTICS

Market Saturation Report on Restaurants in

Bengaluru (aka Bangalore), 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.

**A screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generatedMissing Values** – 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.

Figure 2 Missing values, Binary variables & Factors.

Figure 1 Correcting Translation Errors

**Creating Factors** – 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*). Additionally, on the Rates that customers scored 1 to 5 on, there needed to be a combination of “4.8/5” and “4.8 /5” with an extra space plus the rates were rounded to whole numbers.

**Save new File for other Software Programs** – 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***.

A screenshot of a computer

Description automatically generated**Explore Outlier Data** - 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 The Distribution shows the Outliers in red/blue on the left which are specified in the chart below.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Figure AFTER Outliers Added to Missing Value Codes

Figure IDENTIFYING Specific Outliers

* + A screenshot of a computer

    Description automatically generatedA screenshot of a graph

    Description automatically generatedA screenshot of a computer

    Description automatically generatedCuisine Segmentation Analysis

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

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generatedViewing 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 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 were allowed to be chosen creating dozens of instances of overlap in the cuisine categorical variable).

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

Figure 7 Cities where North Indian Food is most popular.

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).

Figure 9 Cities where South Indian Food is most popular.

* A screenshot of a table

  Description automatically generatedCustomer Preferences Analysis

A screenshot of a computer screen

Description automatically generatedA screenshot of a table

Description automatically generatedWhen 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

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

A screenshot of a graph

Description automatically generatedSimilar 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 13 Reservations where the larger the mean, the more reservations are used by city.

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