Zomato Restaurant Data Visualization and RapidMiner

Srinivas Rao Prithvi Nag Kolla

University of North Texas

Author Note

Srinivas Rao PrithviNag Kolla, Department of Data Science, University of North Texas.

Contact: SrinivasRaoKolla@my.unt.edu

Goals with Dates: (Task A)

1. How many restaurants have less than 1 year of age?
2. Which restaurant is the oldest, if there are more than two, then how many restaurants provide reasonable price too with more rating?

Results: (Task B & C)

1) Answer: There are 1881 New Restaurants with less than 1-year age.

2) Answer: Restaurant “Le Petit Souffle” is oldest of all restaurants with 4 years and average cost of $ 21.62 and rating of 4.8.

* I created a list of random dates using Date() function and found the restaurants age in years using DatedIf() function.
* The Pivot chart shows number of restaurants that are grouped into their respective ages.

For example: there are 1881 restaurants in the cleaned data that are less than 1 year of age group.

* The pie chart shows the similar count of the number of restaurants that are grouped into their respective ages. For example: there are 3 Restaurants in the Cleaned data that are above 4 years of age.

Descriptive Statistics Results: (Task D)

* With the help of Data Analysis ToolPak tool I was able to create a Histogram values like Bins and Frequency of the ratings occurring for the Ratings column given for the restaurants
* The Table of values were helpful in creating the Histogram.

RapidMiner

Step wise details of Section(ii):

1. First, I imported the “Chapter03DataSet-1.csv” dataset into the RapidMiner with the import tool and we have to click next, next in all the windows till finish button.
2. After importing the dataset, I took the retrieve operator into the design window using the Search bar in the operator to find it.
3. Changed the value types of the columns from Polynomial to Value using Modify tool.
4. In order to handle missing data, I used the Replace missing values operator and selected attribute filter type: Single , attribute: Online\_Gaming, default: value, replenishment value: ‘N’ and run to see that all the missing values in attribute Online\_Gaming is replaced with ‘N’.
5. In order to handle inconsistent data present in the data set , we need to apply the Replace operator into the design window then select attribute filter type: Single, attribute: Twitter, Replace What: 99, Replace by: ‘N’ as it is the most common one used in the attribute.
6. In order to perform attribute reduction, I took the operator attribute selector into the design window and then selected attribute filter type: Subset, attribute:
7. Birth\_Year
8. Gender
9. Marital\_Status
10. Race
11. Years\_on\_Internet.
12. In order to filter, we use the Filter Examples operator then select filter as Gender equals to Male(M). We can select one or more filters at a time.
13. Connect all the operators with splines to the end port.
14. Then Run by clicking the Play button. It was exiting working on RapidMiner.

A picture containing screenshot

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a social media post

Description automatically generatedA screenshot of a social media post

Description automatically generatedA screenshot of a cell phone

Description automatically generated

References:

# [MJ Tube].MS Excel: How to Convert Days & Dates into Years Months & Days. Referenced from: <https://www.youtube.com/watch?v=LuhiyQPPKHQ>