Zomato Questionaire

Dataset link: <https://www.kaggle.com/datasets/shrutimehta/zomato-restaurants-data?select=zomato.csv>

1. Are there any missing values in the dataset, and how should they be handled?

2. Should any columns with a large number of missing values be dropped?

3. Do categorical variables need encoding (e.g., one-hot encoding) for machine learning models?

4. Are there outliers in numerical columns that need to be addressed?

5. Should any columns be converted to a more appropriate data type?

6. Do columns with date or time information need to be formatted correctly?

7. Are there duplicate rows in the dataset, and should they be removed?

8. Should any columns with constant values be dropped?

9. Do text-based columns require cleaning or tokenization for natural language processing?

10. Should any categorical variables be combined or simplified to reduce cardinality?

11. Do any columns contain inconsistent or erroneous data that needs correction?

12. Should columns with currency information be converted to a consistent currency format?

13. How should missing values in categorical columns be imputed, if necessary?

14. Are there any columns that need to be renamed for clarity or consistency?

15. Should columns with percentages or ratios be standardized to a common scale?

16. Do numerical columns need scaling for certain machine learning algorithms?

17. Are there columns with unnecessary information that can be dropped from the dataset?

18. Should any columns be created by combining or transforming existing columns?

19. How should text-based columns with multiple entries (e.g., cuisines) be processed?

20. Should any columns be reordered for better readability or analysis?

21. Do columns with numerical values need to be binned or discretized for analysis?

22. Should any columns be converted to datetime objects for time-based analysis?

23. Are there any issues with inconsistent capitalization or formatting in text columns?

24. Should any columns be dropped due to redundancy or multicollinearity?

25. How should columns with hierarchical or nested information be flattened for analysis?