**Data Exploration and Preprocessing Report**

3.1 Introduction:

This report summarizes the exploratory data analysis (EDA) and preprocessing steps performed on the collected restaurant data. EDA provides valuable insights into the data distribution, identifies potential relationships between variables, and helps inform subsequent data preprocessing and model building.

3.2 Exploratory Data Analysis (EDA):

Summary Statistics:

Calculated summary statistics (e.g., mean, median, standard deviation, min, max) for numerical variables (e.g., ratings, price range).

Analyzed the distribution of categorical variables (e.g., cuisine type, location, user demographics).

Data Visualization:

Created visualizations (e.g., histograms, box plots, scatter plots, bar charts) to explore the distribution of variables and identify potential relationships.

Visualized the geographical distribution of restaurants and user activity.

Explored the relationship between user ratings and other factors (e.g., cuisine type, price range, location).

Correlation Analysis:

Investigated correlations between different variables (e.g., user age and dining frequency, restaurant ratings and review sentiment).

3.3 Data Preprocessing:

Data Cleaning:

Handled missing values using appropriate imputation techniques (e.g., mean imputation, median imputation, k-Nearest Neighbors).

Removed duplicate records and corrected inconsistencies in data.

Addressed outliers and anomalies as identified in the Data Quality Report.

Data Transformation:

Transformed numerical variables (e.g., standardization, normalization) to improve model performance.

Converted categorical variables into numerical representations (e.g., one-hot encoding, label encoding).

Performed text preprocessing on textual data (e.g., reviews, descriptions) including cleaning, stemming, and tokenization.

Feature Engineering:

Created new features based on existing data (e.g., distance from user location, day of the week, time of day).

Extracted relevant features from textual data (e.g., sentiment analysis, topic modeling).

3.4 Key Findings:

[Summarize key findings from the EDA, such as the most popular cuisines, the distribution of user ratings, the relationship between price and ratings, etc.]

3.5 Conclusion:

The EDA and data preprocessing steps have provided valuable insights into the characteristics of the data and have prepared the dataset for subsequent model building and evaluation. The cleaned and transformed data will serve as the foundation for developing an effective and personalized restaurant recommendation system.