**Data Collection and Preprocessing Phase**

**Data Exploration and Preprocessing Report**

This report details the data exploration and preprocessing steps to prepare inputs for the Nutrition AI application. Unlike traditional projects, this application primarily utilizes simulated data and pre-trained model inputs to generate personalized dietary recommendations.

**Section Description**

**Data Overview**  
Since no traditional dataset is used, the model relies on user inputs such as text prompts, images of meals, and optional fitness data from APIs. The data processing involves extracting meaningful features from these inputs to provide accurate nutritional analysis.

* **Dimension:** Simulated inputs generated for testing varied dietary scenarios.
* **Descriptive Statistics:** Analysis is conducted directly on processed inputs from the Gemini Pro model, focusing on the accuracy and relevance of nutritional breakdowns.

**Univariate Analysis**

* Conducted on user inputs (e.g., calorie count, macronutrient distribution) to ensure consistency in analysis outcomes.

**Bivariate Analysis**

* Analyzed relationships between different inputs, such as the correlation between fitness data (activity levels) and recommended dietary adjustments.

**Multivariate Analysis**

* Combined multiple inputs (image data, text prompts, and fitness tracker data) to evaluate the model's ability to provide holistic dietary advice.

**Outliers and Anomalies**

* Checked for inconsistencies in simulated data outputs, such as extreme calorie counts or implausible nutritional recommendations.

**Data Preprocessing Code Screenshots**

**Loading Data**

* User inputs (text, images) are processed directly within the Streamlit interface, and any fitness tracker data is fetched through API calls.

**Handling Missing Data**

* Ensured proper handling of missing or incomplete inputs by providing default recommendations for insufficient user data.

**Data Transformation**

* Applied text and image preprocessing to standardize inputs before feeding them into the Gemini Pro model.

**Feature Engineering**

* Extracted key nutritional features from the input data, such as macronutrient percentages, to enhance the accuracy of the recommendations.

**Save Processed Data**

* Processed data and analysis results are saved to CSV logs, maintaining a record of user inputs and model outputs for future reference and analysis.