**Data Collection and Preprocessing Phase**

**Data Collection Plan & Raw Data Sources Identification Report:**

This report outlines the data collection plan and identifies raw data sources used in the Nutrition AI project, ensuring data quality and integrity for providing accurate and personalized dietary recommendations.

**Data Collection Plan:**

**Section Description**

**Project Overview**  
The Nutrition AI project aims to provide personalized dietary recommendations using the Gemini Pro model. The application utilizes user-generated inputs such as images of meals, textual descriptions, and optional fitness data from APIs. The primary objective is to analyze these inputs and offer precise nutritional advice to users.

**Data Collection Plan**

* Gather images of meals uploaded by users for analysis.
* Collect textual descriptions provided by users to guide the model's understanding.
* Integrate optional fitness data from external fitness tracker APIs to enhance the context of recommendations.
* Simulate various dietary scenarios for model testing and validation.

**Raw Data Sources Identified**  
The raw data sources for this project primarily include user-provided inputs and simulated data scenarios. No traditional datasets are used due to the unique nature of the application.

**Raw Data Sources Report:**

| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| --- | --- | --- | --- | --- | --- |
| **User Uploaded Images** | Photos of meals uploaded by users for analysis, providing the basis for calorie and nutritional assessment. | Direct input via Streamlit app | JPG, PNG | Varies | User-provided |
| **Text Prompts** | User-generated text inputs that describe the meal, specify dietary needs, or ask specific nutritional questions. | Direct input via Streamlit app | Text | Varies | User-provided |
| **Simulated Data** | Artificially created meal scenarios to test and validate the model's performance in various dietary conditions. | Internally generated by the development team | CSV | 20 kB | Internal use |
| **Fitness Tracker API** | External API providing fitness data (e.g., activity levels) to enhance personalized dietary recommendations. | <https://api.fitness-tracker.com> | JSON | Varies | Requires user consent |

This Data Collection Plan and Raw Data Sources Report detail the inputs used for the Nutrition AI project, highlighting the innovative approach of utilizing user-generated and simulated data to drive personalized dietary recommendations.