**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| Team ID | xxxxxx |
| Project Title | Nutrition App Using Gemini Pro : Your Comprehensive Guide To Healthy Eating And Well-Being |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Overview of nutrition dataset: dimensions, data types, basic statistics. |
| Univariate Analysis | Analysis of individual variables: calorie content, protein, carbohydrates, fats. |
| Bivariate Analysis | Relationships between variables: correlation between nutrients, scatter plots of calories vs. protein. |
| Multivariate Analysis | Patterns involving multiple variables: nutrient composition analysis. |
| Outliers and Anomalies | Identification and treatment of outliers in nutrient values. |
| **Data Preprocessing Code Screenshots** | |
| Loading Data | df = pd.read\_csv('nutrition\_data.csv') |
| Handling Missing Data | # Code for identifying and handling missing values missing\_values = df.isnull().sum() df\_cleaned = df.dropna() |
| Data Transformation | # Code for transforming variables (scaling, normalization) from sklearn.preprocessing import StandardScaler scaler = StandardScaler() df\_scaled = scaler.fit\_transform(df[['calories', 'protein', 'carbohydrates', 'fats']]) |
| Feature Engineering | # Code for creating new features or modifying existing ones df['calorie\_per\_protein'] = df['calories'] / df['protein'] |
| Save Processed Data | # Code to save the cleaned and processed data df.to\_csv('cleaned\_nutrition\_data.csv', index=False) |