Final Project Proposal
Shouvik Ahmed Antu
Thinking Machines

Project Title: Analyzing Nutritional Value and Caloric Content of Common Foods

Research Question:
How do different food categories vary in terms of calorie density and nutritional content? Specifically, can we identify food groups that offer the highest nutritional value for the lowest calorie intake?

Significance of the Study:
With growing concerns over obesity and diet-related diseases, understanding which foods provide high nutritional value without excessive calories is essential. This project aims to offer insights into the nutritional quality of common foods, helping people make more informed dietary choices.

Data Source:
The dataset chosen for this project is the **"Top 100 Healthiest Foods in the World"** dataset available on Kaggle.

- Dataset Link: Top 100 Healthiest Foods in the World

Data Overview:

- **Columns**: Food name, Category (e.g., fruits, vegetables, grains), Calories per 100g, Protein content, Fat content, Vitamins, Minerals, and more.
- **Number of Records**: 100 rows (each representing a food item).

Research Methods and Tools:

1. **Data Cleaning & Preprocessing**:
    - Use **Pandas** to clean and structure the dataset for analysis, focusing on handling missing values, standardizing nutritional metrics, etc.
2. **Data Analysis**:
    - Conduct exploratory data analysis (EDA) using **Pandas** and **NumPy** to calculate summary statistics (mean, median, standard deviation) for each nutrient across food categories.
3. **Modeling**:
    - Use clustering algorithms like **K-means** to group foods by their nutritional profiles.
    - Perform regression analysis to see if there's a relationship between caloric density and protein/vitamin content.
4. **Visualization**:
    - Use **Matplotlib** and **Plotly** to create visualizations, including:
        - A scatter plot of calories vs. protein content across food categories.
        - A heatmap of nutrient density (nutrition-to-calorie ratio) across food items.