**Group 2. Project Proposal**

**Project Title**: Exploring Obesity Risk Factors

**Team Members**:

1. Kavita Lakhani
2. Monique Reid
3. Andrea Nimako
4. Lynda Sempele
5. Kevin Ngala

**Project Description/Outline:**

This project aims to analyze available datasets to identify primary risk factors linked to obesity. By examining demographic, dietary, and lifestyle factors, we seek to inform preventive interventions and health policy.

**Research Questions**:

1. What is the relationship between dietary habits and obesity rates?
2. How does lifestyle, particularly leisure time physical activity, affect the risk of obesity?
3. How are various dietary patterns, such as frequent intake of high-caloric foods and low consumption of vegetables, associated with levels of obesity?

**Dataset Sources**:

1. Obesity or CVD risk (Classify/Regressor/Cluster) [Obesity or CVD risk (Classify/Regressor/Cluster)](https://www.kaggle.com/datasets/aravindpcoder/obesity-or-cvd-risk-classifyregressorcluster)
2. Geoapify Location Platform: [API Documentation, Playgrounds, code samples](https://apidocs.geoapify.com/#docs)

**Tasks and Requirements / Division of Labor:**

1. ***Data Collection and Preprocessing***: Exploratory Data Analysis, Statistical Analysis, Data Visualization, Interpretation of Findings and Reporting
2. ***Pandas Data Cleaning*:** Use Pandas to clean and format the NHANES dataset.
3. ***Jupyter Notebooks*:** Document data exploration and cleanup in Jupyter notebooks.
4. ***Matplotlib Visualizations*:** Create 6 to 8 visualizations using Matplotlib, with at least 2 per research question.
5. ***PNG Images*:** Save visualizations as PNGs for distribution and presentation.
6. ***Write-Up*:** Summarize findings with headings for each research question, including descriptions and relevant plots.

Through this analysis, we aim to provide insights that support effective strategies for combating obesity and promoting child health.