**DATA 6550 - Module 2 Group 4 Discussion Summary, Week A**

**Key Points Discussed:**

* The project focuses on **bias, representation, and fairness in datasets**, with a deliverable due **February 12**.
* The team explored **biased datasets** to analyze, sharing potential options.
* Weekly discussions were divided into **Week A and Week B**, with **GitHub used for collaboration**.

**Decisions Made:**

* The **Heart Attack Risk Prediction dataset** was strongly considered due to its **AI-generated nature and gender imbalance (70% male, 30% female)**.
* The **Life Expectancy dataset** was also discussed, highlighting **selection bias** due to missing data from certain countries.
* Team members were encouraged to suggest alternative datasets before making a final decision.

**Challenges Identified & Solutions:**

* **Dataset Selection**: Multiple options were proposed, but consensus was needed.
  + *Solution:* Continued discussion and evaluation of datasets based on relevance to bias analysis.
* **Collaboration & Organization**: Need for structured teamwork and version control.
  + *Solution:* A **GitHub repository** was created to streamline work.

**Contributions by Team Members:**

* **Savi Alwis**: Proposed the **Heart Attack Risk Prediction** dataset and initiated dataset discussions.
* **Muqtasid Qureshi**: Suggested an **obesity dataset** and set up the **GitHub repository**. Also shared a bias analysis on a brain stroke dataset as a reference.
* **Pallavi Suram**: Supported dataset discussions and asked for additional recommendations.
* **Moses Kipkemboi**: Reviewed discussions and confirmed alignment with project objectives.

**Outlook & Next Steps:**

* Finalize the **dataset selection** based on group consensus.
* Use **GitHub for version control and documentation**.
* Continue weekly discussions to refine the analysis.