**Title**: Patient Readmission Prediction

**By**: Alice Gwyn, Lynn Soors, Malika Shrestha, Will Plaisance & Silvio Zabala

**Project Overview**: Develop a machine learning model to predict the likelihood of hospital readmission within 30 days of discharge. This model will help healthcare providers across the nation identify high-risk patients early, prioritize interventions, and potentially reduce readmission rates and associated cost.

**Objective**:

This project will focus on the following goals:

* Identify possible cost reduction and propose readmission rates solutions.
* Data Exploration and Analysis
* Apply Machine Learning Techniques
* Visualize results for insights
* Use of Additional Technologies
* Documentation and Presentation

**Possible Datasets**:

* <https://data.cms.gov/provider-data/dataset/9n3s-kdb3#data-table>
* <https://huggingface.co/datasets/aai540-group3/diabetes-readmission>

**Research Questions**:

1. What variables has the strongest correlation with readmission risk?
2. What demographics has the strongest correlation with readmission risk?
3. Does the length of stay influence the probability of readmission?
4. How does readmission rates compare across regions?
5. How does hospital volume, as measured by the number of admissions and discharges, affect patient readmission rates?
6. How do readmission rates vary across regions that could benefit from a predictive model to reduce readmissions in highest readmission rates areas?
7. Which facilities have the highest readmission rates, where a readmission prediction model might have the greatest reduction in readmission rates and healthcare costs?

**Tools and Technologies**:

* Scikit-learn
* Databricks
* Tableau

**Timeline**:

1. Week 1:
   1. Complete project proposal
   2. Source data
   3. Clean and process data for exploratory analysis.
   4. Draft the visualizations
2. Week 2:
   1. Polished visualizations
   2. Complete README.md file.
   3. Create presentation