**Collaboration Plan for Fraud Detection Project**

**Participants:** Ryan Tang and Santiago von Straussburg

**Overview:**  
We are collaboratively analyzing two fraud datasets (Kaggle credit card fraud and synthetic fraud detection from Data World) to explore fraud patterns, feature importance, and machine learning model evaluation.

**Technologies Used:**

1. **GitHub Repository:** For version control, code collaboration, and final project hosting.
2. **Google Colab/Jupyter Notebooks:** For ETL, EDA, and model development.
3. **Discord:** Primary communication platform for real-time discussions.
   1. Weekly meetings at 08:00 PM CST on Thursdays for progress reviews and planning.

**Current Focus:**  
Both team members are currently working together on data exploration, including extraction, transformation, load (ETL), and exploratory data analysis (EDA).

**Roadmap & Milestones:**

1. **Milestone 1 – Initial Dataset Selection & ETL:**
   1. Identify datasets. - Completed
   2. Perform initial ETL on datasets. - Completed
   3. Establish a GitHub repository and GitHub Pages site. - Completed
   4. Develop a basic statistic and initial graph for dataset understanding. - Completed
2. **Milestone 2 – Additional ETL & Exploratory Data Analysis:**
   1. Continue data cleaning and transformation.
   2. Conduct comprehensive EDA with 3-5 key graphs.
   3. Present the first project pitch with initial findings.
3. **Deliverable 1 – In-Class Presentation:**
   1. Finalize and present a 5-7 slide deck covering problem statement, ETL, EDA, and project progress.
4. **Milestone 3 – Model Development & Evaluation:**
   1. Select machine learning models (Random Forest, XGBoost, Logistic Regression).
   2. Begin model training and evaluation.
   3. Analyze model performance and feature importance.
5. **Deliverable 2 – Final Website & Presentation:**
   1. Finalize project with the deployment of results to the GitHub Pages site.
   2. Prepare the final presentation summarizing the project lifecycle.