**DATA 602 - Final Project Proposal Joyce Aldrich**

**1.Research Question**

What patterns or characteristics can help me identify fraudulent transactions in a dataset? By analyzing transaction attributes such as categorical features and available data in hand, I aim to uncover the factors contributing to fraud detection accuracy.

**2.Justification - why is this relevant to you or industry?**

I have been working in the banking industry for the past 7 years, primarily focusing on Compliance-related roles. Compliance and fraud detection often overlap, as both aim to ensure the integrity of financial systems and safeguard against risks.

This project will help me deepen my understanding of fraud detection from a developmental perspective, allowing me to gain hands-on experience in analyzing patterns, designing solutions, and interpreting results. It aligns with my professional interests and equips me with the skills necessary to address evolving challenges in fraud prevention within the banking industry.

**3.Data Sources - did you find this data online or collect yourself? Provide links.**

The dataset for this project is sourced from Kaggle: [Fraud Detection Dataset](https://www.kaggle.com/datasets/gopalmahadevan/fraud-detection-example/data)

https://www.kaggle.com/datasets/gopalmahadevan/fraud-detection-example/data

**4.Libraries potentially being used.**

* Pandas: To load and manipulate tabular data.
* Numpy: For performing numerical calculations and managing multidimensional arrays.
* Matplotlib/Seaborn: For data visualization, including histograms, scatter plots, heatmaps, and bar charts, to understand relationships between features and fraud.
* Scikit-learn: For machine learning tasks such as data splitting, model building, and performance evaluation.

**5.EDA and summary statistics.**

Try to understand class distribution, feature correlation, transaction time and amount, categorical features, and outliers.