# Capstone Project Data Science: Fraud Detection Analysis

**DATASET**

## Project Objectives and Scope

1. **What secondary goals does the fraud detection model aim to achieve?**
2. **How does the model align with the business objectives of the organization?**

## Data Analysis

1. **What are the most significant features contributing to fraud detection?**
2. **How does the correlation matrix help in understanding feature relationships?**

## Data Preprocessing

1. **Why is it necessary to handle missing values before model training?**
2. **What impact do outliers have on the model's performance, and how are they addressed?**

## Model Training

1. **What assumptions does the Gaussian Naive Bayes algorithm make about the data?**
2. **How do you handle imbalanced datasets during model training?**

## Model Evaluation

1. **What is the significance of the ROC curve in evaluating the model?**
2. **How do you interpret the F1 score in the context of fraud detection?**

## Results and Interpretation

1. **How do you interpret the confusion matrix for your model's predictions?**
2. **What does the lift curve tell you about your model's performance?**

## Model Improvement

1. **How does feature engineering enhance the performance of your fraud detection model?**
2. **What role does hyperparameter tuning play in improving the model?**

## Practical Implementation

1. **What infrastructure is needed to deploy the model in a live environment?**
2. **How do you monitor the performance of the deployed model?**

## Technical Implementation

1. **What libraries and tools are essential for implementing Naive Bayes in Python?**
2. **How does stratified cross-validation differ from regular cross-validation?**