# Project Proposal

**Name of the team:** Group 21

**Team members:**

Ding Jiayi, 3036032906

Huang Binqian, 3036032724

Huang Chengdong, 3036033340

Wu Tianchi, 3036033974

Zhang Miao, 3036032839

**Type of financial fraud selected for this project:** Credit card fraud

**Background of the selected case:**

This is a simulated credit card transaction dataset containing legitimate and it was generated using Sparkov Data Generation | Github tool created by Brandon Harris. This simulation was run for fraud transactions from the duration of 1st Jan 2019 - 31st Dec 2020. It covers the credit cards of 1000 customers doing transactions with a pool of 800 merchants.

**Define scope of fraud data analytics:**

Create algorithms and predictive models for credit card fraud detection based on historical transaction data from customer and merchant pools, provide cost-benefit analysis models to stakeholders, and provide them with appropriate recommendations to mitigate fraud risks

**Identify fraud scenarios:**

- Who is the committing person?

A fake merchant or a fake cardholder.

- What are the possible entities involved?

The Merchant, Cardholder

- What are the possible fraudulent actions?

Fraudsters create fraudulent merchant accounts after stealing an e-commerce business' s identification.The newly created "business" then posts charges to customers' credit cards, collects the money and closes the accounts.

Fraudsters successfully impersonates the legitimate owner of a credit card by

providing enough accurate personal data about a cardholder to convince merchants and payment processors the cardholder placed the order.

Fraudsters uses the stolen personal information of account holder to fraudulently gain access to the account. They then use the account to make purchases the actual account holder did not authorize.

## Plan to obtain of the dataset for building the fraud detection model:

The dataset is related to credit card transactions. To find a proper dataset, several conditions are supposed to be considered. The data size cannot be too small. Moreover, the incomplete data should take up less than 10%. It's better if all transactions happened in the past ten years. Then we find a suitable dataset from kaggle.

https://www.kaggle.com/datasets/kartik2112/fraud-detection