**Data Science Practical Project**

**Project Overview**

Abc company is a global leader and pioneer in delivering affordable solar-powered solutions designed for the two billion people in the developing world without access to reliable energy. Abc company provides distributed solar energy solutions for households and small businesses that are transforming the way people all over the world use and pay for energy. A key development in the energy markets Abc company services is the rise of the pay-as-you-go (PAYG) business model for energy assets. This suite of technology solutions enables remote control of products through keypad-entered tokens, allowing customers to pay for expensive assets over time without heavy cost of servicing for the company.

The business model works as a lease-to-own or hire-purchase model, where customers pay a deposit and then small increments over time before paying enough to purchase the product fully, when payments are no longer required.

Abc company employs PAYG services in four markets in America. The largest market for PAYG services is currently in XYZ(Country), where we have registered over 1 million customers to date for our service. We offer a wide range of product and payment plans in XYZ(Country), and we have started experimenting with new models to qualify these customers pre-sale based on a proprietary score.

A new tool has been deployed to assess fit for our customers, and we would like you to assess the performance of this tool. You have been provided an anonymized dataset from our PAYG business in XYZ(Country). This dataset shows performance of a select group of contracts in the first days post-sale.

**Here are the questions we would like you to answer:**

1. **Is there a correlation between the Score and repayment performance?**
2. **Is there a correlation between the Score and delinquency status?**
3. **Are there other interesting insights you can glean from the dataset provided?**

We judge repayment performance based on the total tokens customers have purchased compared to the number of days they have been on the book. Each Token in this case corresponds to one day of energy service. Therefore a customer paying for 7 tokens at 7 days old would have 100% repayment, while a customer with 5 tokens at 10 days old would have 50% repayment, etc.

Each ContractId in the dataset has one row for each day it has been on the book. The age of these contracts are not uniform across the data that has been provided to you. You can assume the last record for each ContractId is the latest day available for that customer. Below are further definitions for the data provided to you.

**Data Definitions**

* ContractId Anon – Unique key for the individual customer account.
* Reliability – Was a Score collected?
* Score – Integer of customer fit between 1-100.
* BASEUNIT\_PRODUCTNAME (copy) – Product number and description of the product that was purchased by the individual customer.
* ACCESORY\_NAME (copy) – Accessory identifier.
* UPSELL – Boolean expression for whether this ContractId is a follow-on sale to an existing customer. If TRUE, individual has purchased another product from Abc company in the past. If FALSE, this is the first purchase recorded from this individual. Customers who are TRUE for Upsell have benefited from more attractive pricing, such as a lower deposit.
* DAYS\_ON\_BOOKS – Age of the contract in days for that specific day of record.
* BASEUNIT\_CONTRACTSTATUS – Contract status including ACTIVE (is in active repayment) or PAID OFF (has completed the full repayment) for that day of record.
* DELINQUENCY – Measure of current days in the status of “delinquency” (a state of no credit on the product) for that day of record. If positive number, the product has been X days without active credit since energy last expired. If negative, the contract has [-X] number of days until the current credit paid will expire on the product. (A contract with -2 delinquency has 2 days of credit remaining to be used on the product). If 0, then it is the first day without electricity.
* TOTAL\_TOKEN\_PAID – Sum total of Tokens paid for (in days) since the contract was opened for that specific day of record. 1 token = 1 day of electricity.