**Midterm Hackathon proposal solution**

Customer Churn Prediction

Customer churn prediction identifies which customer is likely to patronize their product or as high risk at high risk of cancelling their subscription.

This project seeks to incorporate socio-economic characteristics of customers. This means we would be adding additional features that represent socio economic characteristics. This will help us reveal a deeper insight into the reason behind customer churn.

Even though some socio-economic characteristics can be personal it helps in developing more accurate and comprehensive predictive models and data can be kept private The environment of the customer also plays a role in influencing how the customer will relate with the product or engage the product..

Most predictive models depend on transactional and behavioral data. Transactional data refers to information that records the interactions between customers and the business. Example can be purchasing history. Behavioral data captures features like the customers engagement metrics and many more. It focusses on the customers behavior during interactions.

Such socio-economic data can be difficult to get hence we would use or generate a randomized synthetic data for training purposes. Our dataset is going to be trained based on supervised learning. The proposal solution will focus mainly on three socio-economic characteristics, namely:

Income level  
Education  
Employment status

Dataset: [Dataset](https://www.kaggle.com/datasets/blastchar/telco-customer-churn?resource=download)