## Customer churn Prediction for Ecommerce and data integration with data engineering

## Problem

Customer churn is the phenomena in where customers stop interacting with an online business or store. This provides a challenge for most business as it will interfere with organization’s revenue. In the dynamic landscape of e-commerce retaining customers is essential for the businesses as it is directly related to growth and profitability. On the other hand data integration is the development, maintenance, and implementation of systems and processes that take in raw data and produce high quality consistent information that support downstream use cases such as machine learning and or analysis.

## Solution

Implement a system that predicts customer churn by evaluating and analyzing user behavior and purchase history and also implementing data pipelines to collect and preprocess data then use machine learning for predictive analysis

## Abstract

Customer Churn is a phenomena where customers gradually stop interacting with an online business be it buying something from a retail store or a service that is being offered. It is a major issue as it will directly impact growth and profits for all businesses. Data integration is another major issue as businesses are now relying on data to make decisions and it is now important to have that data in a well-designed ecosystem that will allow the ease of accesses, proper structure and security for the data. This will support all downstream use cases such as data analysts or data scientist. The data can also be used for reporting to various stakeholders that are non-technical

Customer churn may lead to business stagnating and revenues dropping as online businesses need user interaction in their site and with their product in order to make profits. In relation with data integration data needs to be made available when needed in a safe a reliable way in order to streamline analysis and reporting processes. This problem can be seen in online businesses that are at the high data maturity stage where data is integral to their decision making.

This brings us to the people directly affected by this problem, we have online businesses such as retail stores and also all people who are the target of downstream data such as data analysts and data scientists among others

Proposed solution will be the implementation of data pipelines for processing raw data and using machine learning to provide predictive analysis. Approach used will be prototyping as the system will be made through several iteration to optimize the model and data pipelines

