



# Northeastern University

## College of Engineering

### **IE 6750 Data Warehousing and Integration**

Project Title

**Policy Lapsation in Life Insurance**

Milestone 1

Problem Definition

Group 4

Aayush Amrute  
Saurabh Chavan  
Paritosh Vyawahare

[amrute.a@northeastern.edu](mailto:amrute.a@northeastern.edu)

[chavan.sau@northeastern.edu](mailto:chavan.sau@northeastern.edu)

[vyawahare.p@northeastern.edu](mailto:vyawahare.p@northeastern.edu)

Submission Date: 01-23-2025

## Problem Setting

Policy lapsation is one of the major challenges in the life insurance industry. It happens when customers fail to pay their premiums within the given due dates, resulting in the termination of their policy benefits. This issue impacts both customers and the company. Customers lose their coverage and financial protection, while the company faces reduced revenue, increased churn rates, and additional costs for reactivating policies or acquiring new customers. Companies often struggle to identify lapsation risks in time due to fragmented data from multiple channels, such as web sales, agents, and bank. These challenges hinder timely interventions and effective decision-making.

## Problem Definition

Life insurance companies face a big challenge when it comes to managing policy lapsation. This happens because the data they rely on is often incomplete or inconsistent, coming from multiple sources like web sales, agents, and bank relationship managers. As a result, companies struggle to intervene on time, leading to delays and higher policy churn. For example, customers who miss their premium payments whether they are on autopay or non-autopay often end up lapsing their policies without receiving proper follow-ups.

To solve this problem, it's important to focus on two key types of data: static reference data and transactional data.

- Static reference data includes foundational, less frequently changing details such as branch information (name, type, and address), agent details (like agent ID, name, and commission rate), and policy types (minimum and maximum terms, sum assured limits). This type of data provides the structure that helps companies organize their operations.
- On the other hand, transactional data is dynamic and constantly updated. It includes details like premium payments, due dates, policy statuses (whether active, lapsed, or matured), and payment methods. This data is crucial for tracking customer activity and identifying trends related to lapsation.

By effectively managing these two types of data, companies can build a solid foundation for accurate reporting and analysis. This, in turn, enables timely interventions, helping to reduce lapsation and improve customer retention.

## Objective

The goal of this project is to create a centralized data engineering solution to integrate, transform, and analyse data from multiple sources such as web sales, agents, and bank/RM channels. By addressing lapsation risks proactively, the system will ensure accurate, consistent, and actionable data for analysis and decision-making.

To achieve this:

1. We will build a robust ETL pipeline to handle static reference data (e.g., customer demographics, policy details) and dynamic transactional data (e.g., premium payments, due dates).

2. Synthetic data will simulate real-world scenarios, such as:
  - Customers entering their details and policy preferences via the website.
  - Agents and banks contributing additional customer and policy information.
  - Dynamic updates from premium payments, including payment methods and statuses.
3. The data will be transformed and stored in a data warehouse, creating a reliable foundation for consistent reporting and interactive dashboards.

## End Goal

The end goal of this project is to develop a complete data engineering solution that integrates and transforms data from multiple sources into a centralized data warehouse. This will enable consistent reporting, interactive dashboards to analyse lapsation trends, and provide a foundation for predictive models to identify high-risk customers, demonstrating how data-driven strategies can reduce lapsation and improve business outcomes.

As we progress through the milestones, the project will focus on analysing key areas that directly impact business performance:

1. **Customer Segmentation and Retention:** Using customer demographics and payment behaviours, we will create segments based on their propensity to lapse. This segmentation will allow the company to proactively target high-risk customers with personalized retention strategies, such as payment reminders, grace period extensions, and special offers.
2. **Agent and Branch Productivity Analysis:** By evaluating agent performance (e.g., policies sold, lapsation rates, and commissions) and branch-level metrics (e.g., policies issued, customer satisfaction), we can identify top-performing agents and branches. This insight can guide training programs for underperforming agents and optimize resource allocation to maximize policy sales.
3. **Optimization of Payment Channels:** Analysing payment methods (e.g., autopay vs. non-autopay) and failure rates will help the company optimize its payment systems. For example, customers on non-autopay could be encouraged to switch to autopay through incentives, reducing the likelihood of missed payments.
4. **Long-Term Strategic Insights:** The dashboards and reports generated from this solution will provide insights into lapsation trends, helping the company refine its long-term strategies. For instance, identifying regions with higher lapsation rates can guide marketing and operational focus in those areas.

By the end of this project, the company will have a robust foundation to make smarter decisions that enhance its competitiveness in the life insurance market.