# 1. Model Logic

We propose using a Logistic Regression model to predict customer delinquency risk.   
This model is ideal for binary classification tasks like predicting whether a customer will become delinquent or not.   
It works by estimating the probability that a given input belongs to the positive class (delinquent = 1).  
  
The model will use the following five features as inputs:  
1. Income – Lower income may correlate with higher risk.  
2. Credit Score – A key indicator of financial health.  
3. Credit Utilization – Higher utilization often signals financial stress.  
4. Missed Payments – A direct indicator of past delinquency.  
5. Debt-to-Income Ratio – Helps assess ability to manage payments.  
  
Workflow Overview:  
- Load the dataset from Excel.  
- Handle missing values using SimpleImputer.  
- Scale numeric features using StandardScaler.  
- Split into training and test sets.  
- Fit a logistic regression model within a pipeline.  
- Predict and evaluate using confusion matrix, classification report, AUC, and ROC curve.

# 2. Model Justification

Logistic Regression is chosen due to its balance between performance and interpretability.   
It provides probabilistic outputs, is easy to deploy, and aligns with regulatory expectations for transparency in financial services.   
Unlike complex models like neural networks, logistic regression clearly shows how each feature contributes to the prediction,   
making it easier for Geldium’s risk and compliance teams to understand and monitor decisions.   
It is computationally efficient, works well on structured data, and helps prioritize high-risk customers with transparency and fairness.

# 3. Evaluation Strategy

To ensure the model performs effectively and ethically, we will evaluate it using the following metrics:  
  
1. Accuracy – Measures overall correctness of predictions.  
2. Precision – Evaluates how many predicted delinquents were actually delinquent.  
3. Recall – Ensures the model catches most actual delinquents.  
4. F1 Score – Balances precision and recall.  
5. AUC-ROC – Visualizes and quantifies model’s ability to separate classes.  
6. Confusion Matrix – Helps identify false positives/negatives.  
  
Additionally, we perform missing value imputation to reduce data bias and use interpretability tools to explain model outputs.   
Fairness checks, such as checking demographic parity or disparate impact (manually or using tools like SHAP),   
will be conducted to ensure no group is unfairly impacted.