# TASK 3

Build a model using Python that will estimate the probability of default for a borrower

# OBJECTIVE

* How to choose appropriate independent variables from a data set that will accurately predict the outcome of a chosen dependent variable
* The importance of using available data to predict customer trends and actions

# CODE

# A simple program to predict loan defaults

# Import necessary libraries

import pandas as pd

from sklearn.linear\_model import LogisticRegression

from sklearn import metrics

# Read the loan data from a CSV file

data = pd.read\_csv('loan\_data\_created.csv')

# Print the first few rows of the data to check it

print("First few rows of the data:")

print(data.head())

# Calculate some financial ratios

data['payment\_to\_income'] = data['loan\_amt\_outstanding'] / data['income']

data['debt\_to\_income'] = data['total\_debt\_outstanding'] / data['income']

# Define the features we'll use to predict defaults

features = ['credit\_lines\_outstanding', 'debt\_to\_income', 'payment\_to\_income', 'years\_employed', 'fico\_score']

# Create and train the model

model = LogisticRegression()

model.fit(data[features], data['default'])

# Print the model coefficients

print("\nModel coefficients:")

for feature, coef in zip(features, model.coef\_[0]):

print(f"{feature}: {coef}")

# Make predictions

predictions = model.predict(data[features])

# Calculate the accuracy of the model

accuracy = (data['default'] == predictions).mean()

print(f"\nModel accuracy: {accuracy:.2f}")

# Calculate and print the AUC score

fpr, tpr, \_ = metrics.roc\_curve(data['default'], predictions)

auc\_score = metrics.auc(fpr, tpr)

print(f"AUC score: {auc\_score:.2f}")

# Example: Predict for a new borrower

new\_borrower = [[5, 0.4, 0.2, 3, 700]] # Example values for the features

default\_probability = model.predict\_proba(new\_borrower)[0][1]

print(f"\nProbability of default for new borrower: {default\_probability:.2f}")