TAIWAN COMMERCIAL BANK CREDIT DEFAULT ANALYSIS

March 9, 2025

OUTLINE

- Project goal
- Business Understanding
- Data and Methods
- Data analysis
- Modelling and Evaluation
- Conclusions

PROJECT GOAL

The bank aims to leverage machine learning and AI-driven classifications model to accurately assess the likelihood of a customer to default on credit card payments. This will improve the banks lending decision making to approve loans for reliable customers and adjusting credit limits based on a client's repayment behavior and enhancing customer experience.

BUSINESS UNDERSTANDING

Key Objectives:

- Develop a model to classify customers as likely to default (1) or not (0)
- Identify key factors contributing to credit default risk.
- Improve decision making on credit card approvals and limit adjustment based on risk

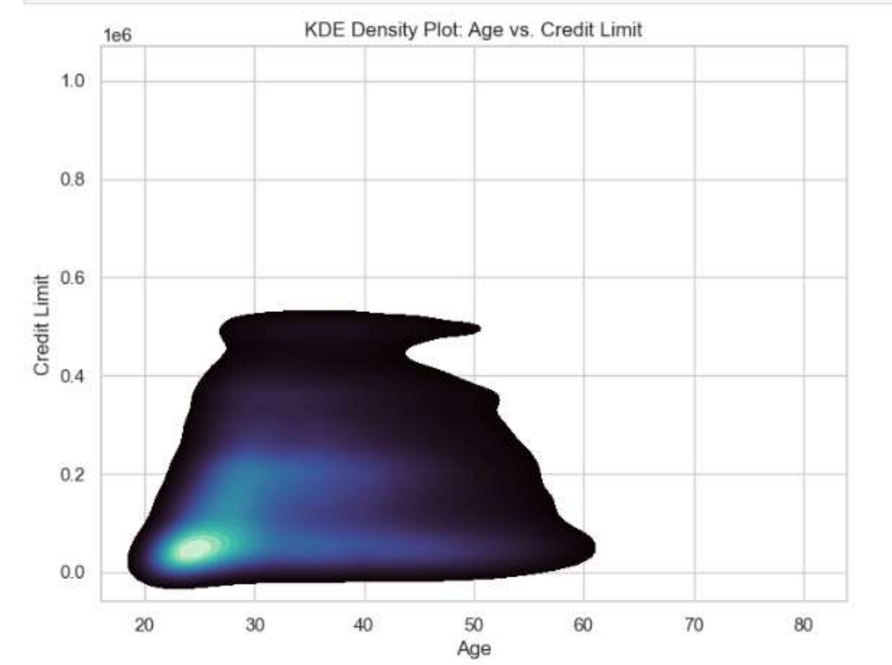
BUSINESS UNDERSTANDING

- **Key Questions:**
- Do demographic features such as sex, education, marriage affect the probability of credit card default
- What is the relationship between credit card limit, bills, payments and payment delays.
- Which customers pose the highest risk of default according to age group, education e.t.c
- Can we build a model to predict default from given data

DATA AND METHODS

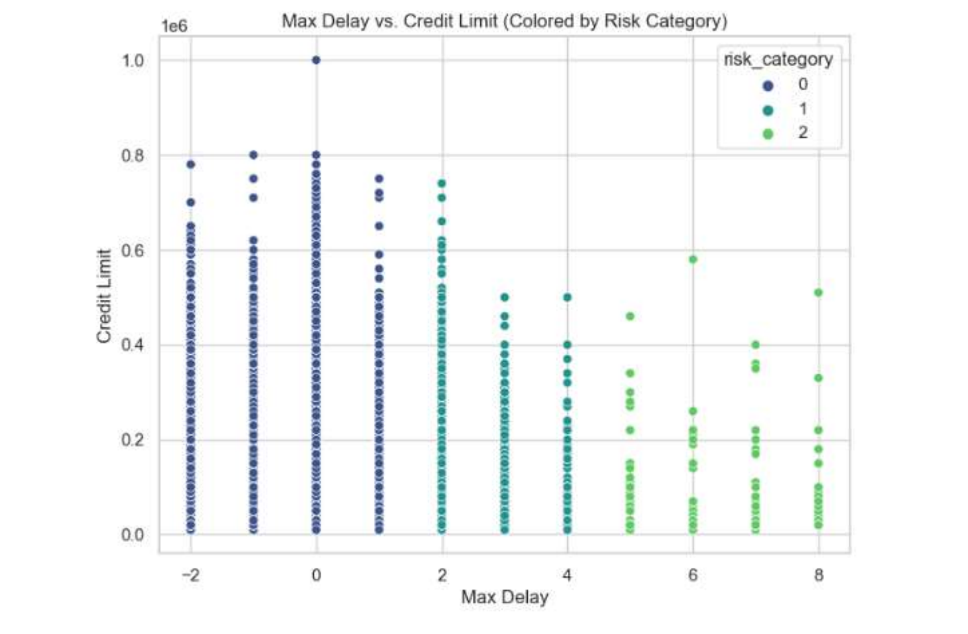
- Data was obtained from the UCI machine learning repository.
- Contains over 20,000 records.
- Data includes: age, gender, bills for past 6 months, payments for past 6 months, education level and delays in payments.

DATA ANALYSIS



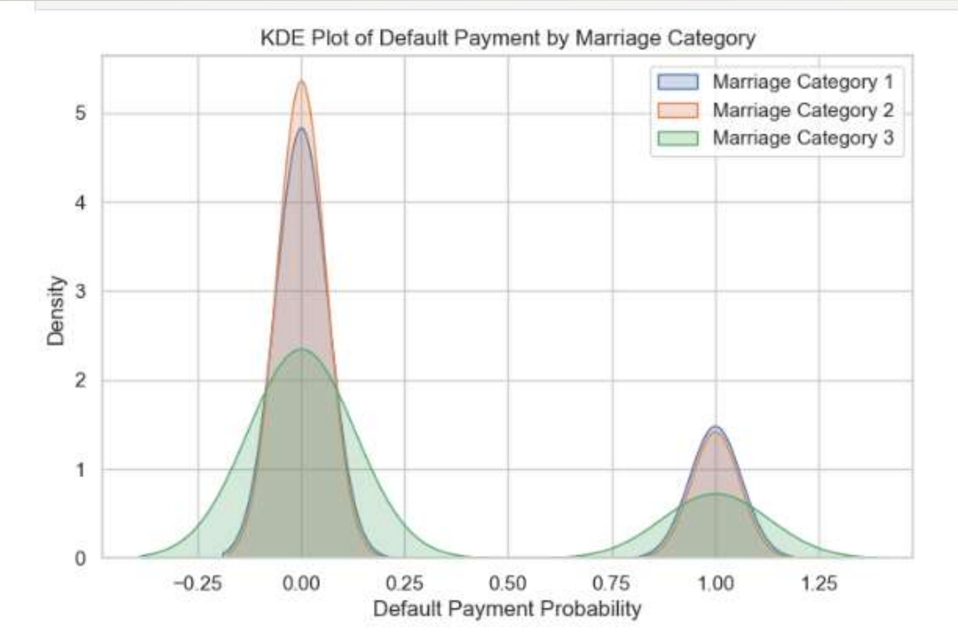
Most individuals ages 25 to 40 have credit limits below 200,000. Older individuals have higher credit limits, but this levels off

DATA ANALYSIS



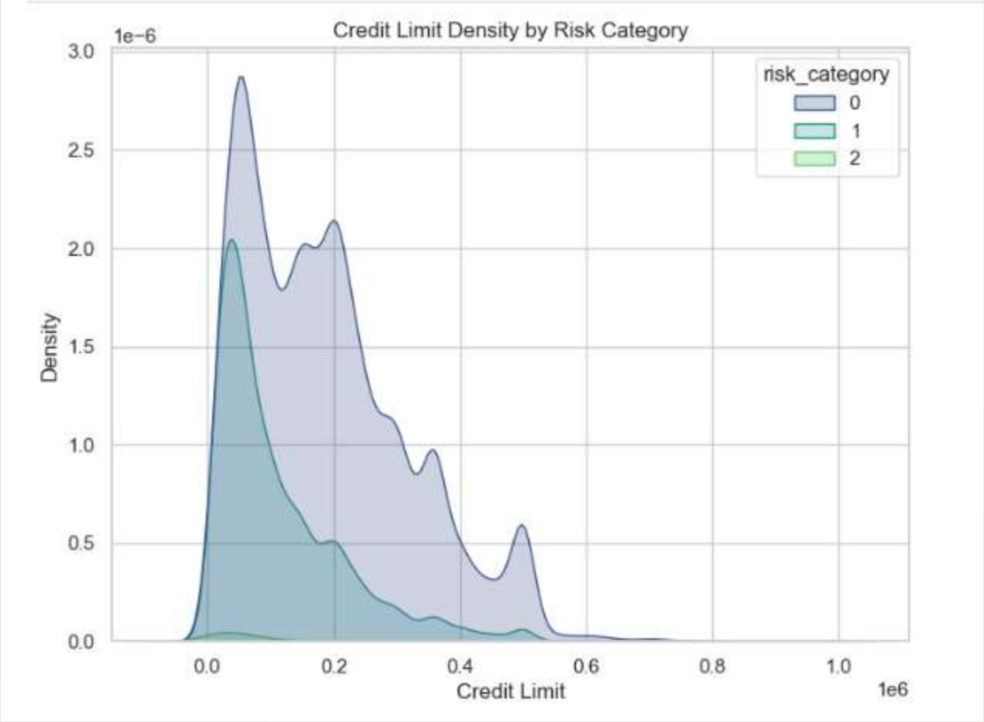
Higher delays are more frequent among **low credit limit customers**.
Credit limit may act as a risk buffer, as customers with very high limits rarely have extreme delays.

DATA ANALYSIS



Marriage status seems to play a role with singles (Category 2) having a slightly lower tendency to default. However, there’s no drastic difference between the three groups hen.

DATA ANALYSIS



The majority of individuals have low credit limit and are in risk category 0.
High credit limit individuals are in all categories but are less common.

MODELLING

- We built and fit the following models:
  - ✓ Logistic Regression model
  - ✓ K Nearest Neighbor Classification model
  - ✓ XGBoost classification model
  - ✓ LightGBM classification model
  - ✓ Random Forest model
  - ✓ Multi-Layer Perceptron Classifier

MODELLING

The best performing model was LightGBM which was better suited to handle categorical data without encoding. The model had a 73% accuracy and 77% ROC AUC score. It could recall defaulters 67% of the time.

CONCLUSIONS

1. Higher credit limits are associated with lower default risk.
2. Lower credit limit individuals are more prone to defaulting.
3. Higher delays are more frequent among low credit limit customers.
4. The majority of defaulters have credit limits below 400k.

RECOMMENDATIONS & NEXT STEPS

- Test the model on more recent and enriched data.
- Explore more classification models to improve our predictions.
- Better predict defaulters by creating interface terms.

Q+A

THANK YOU !!

Evans Makau

GitHub Repo: https://github.com/makau99

E-mail: makauevans01@gmail.com