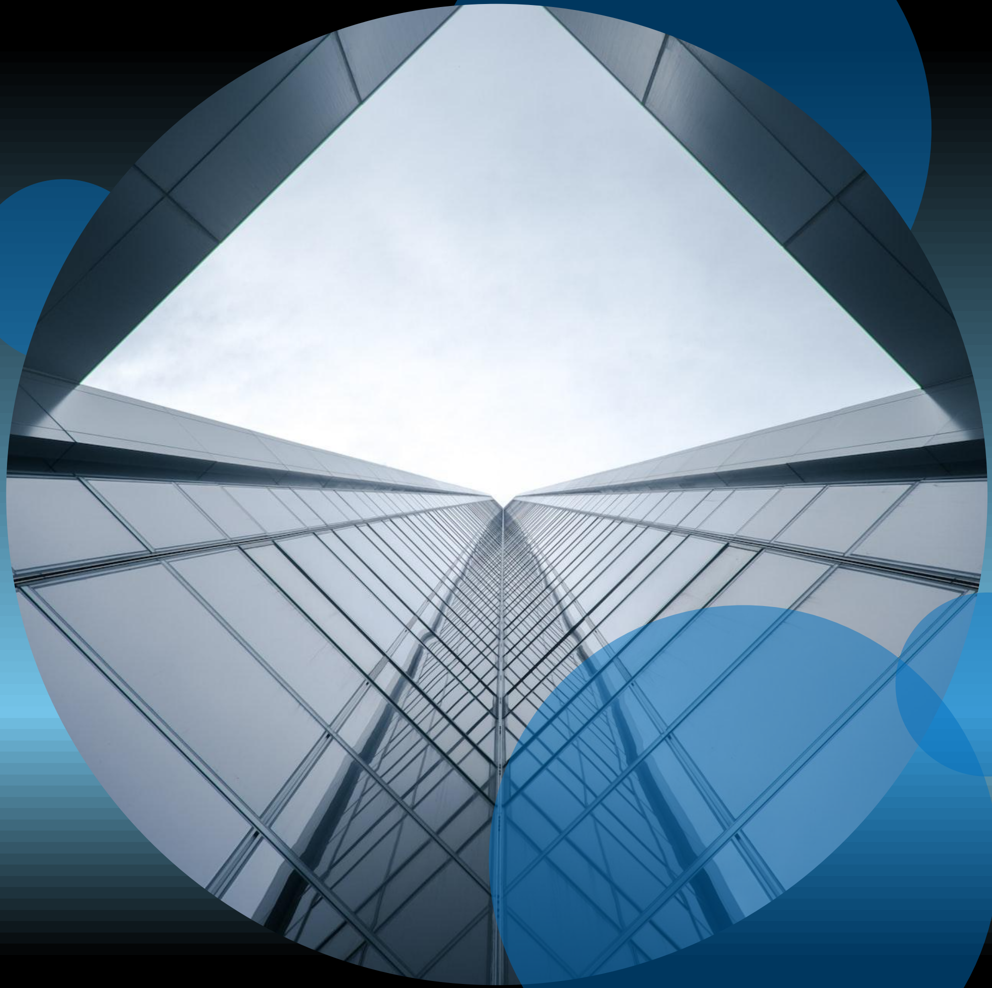
Loan – Default Prediction

A Profiling Approach

Georgios Panos – Royal Institute of Technology (KTH)



Loan Default and its Hallmarks

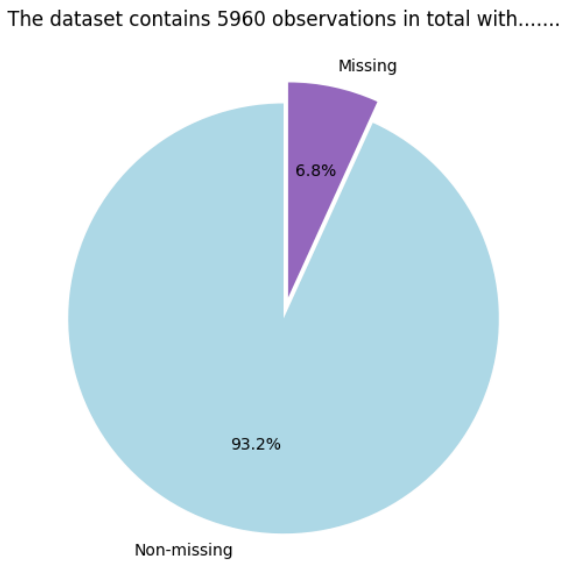
Problem Space

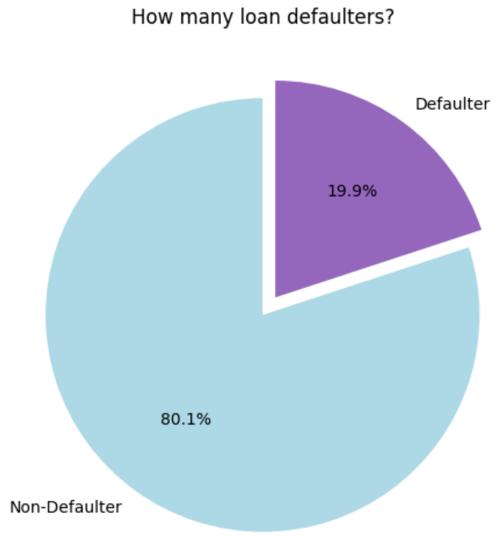
Problem Statement

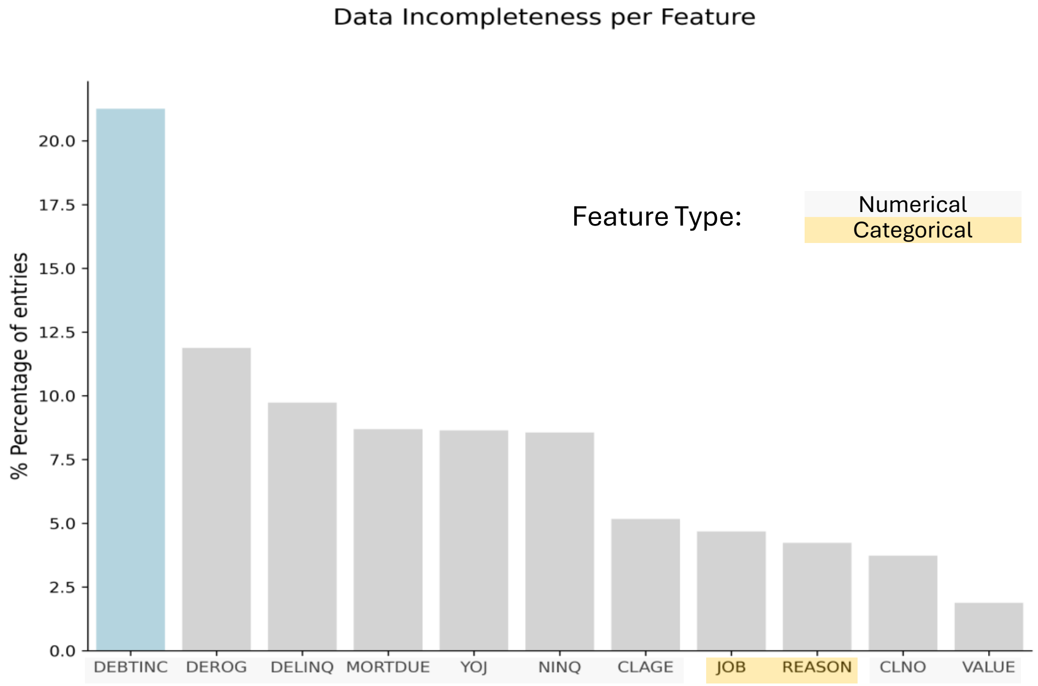
Loan interests is a source of income for the banks and their defaults compound damages to their revenues.

- Which **loan applicant** is most certainly going to **default** in the future, given a loan approval?
- Can we identify the **driving forces** and use **prediction models** to guide us through the loan approval decision?
- Can we **profile** the **loan applicants** in relation to the data management system of the bank institution?

The “Quirks” of the Dataset





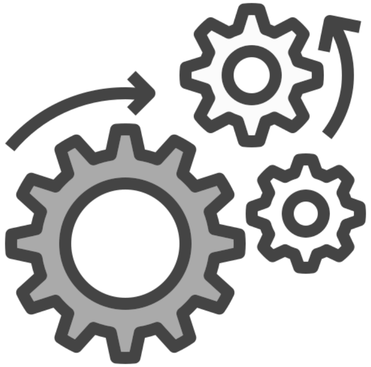


Key – Points from EDA

- **BAD** variable is the class of defaulters/non-defaulters.
- The features **DEBTINC, DEROG, DELING** seem to be correlated with loan defaults.
- **VALUE** and **MORTDUE** are correlated and might be **redundant.**
- Dataset contains **high variation** among its numerical features.
- Many **outliers** to the features, especially those above in the **defaulter class.**
- The JOB variable is **not stratified enough.**

Learning From Data

Moderate Imputation

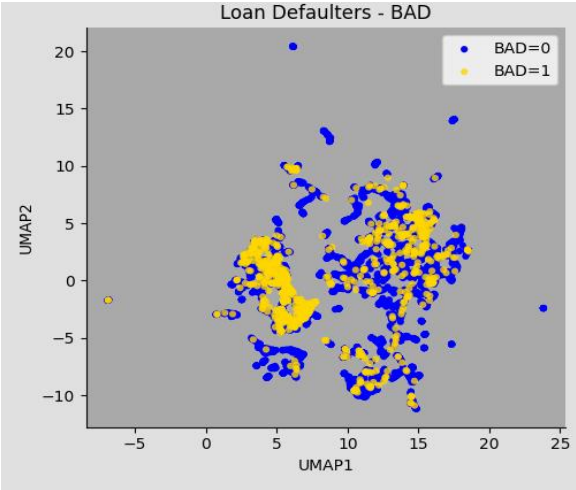


- Kept entries with maximum one **NA**
- **Median** imputation for numeric
- **Mode** imputation for the categories
- Big threshold in outlier filtering.

The boy who cried wolf

- Test and optimize three different ML algorithms based on recall, and ROC auc. Weight – imbalance aware.
- Predict as **many True Positives** as possible.

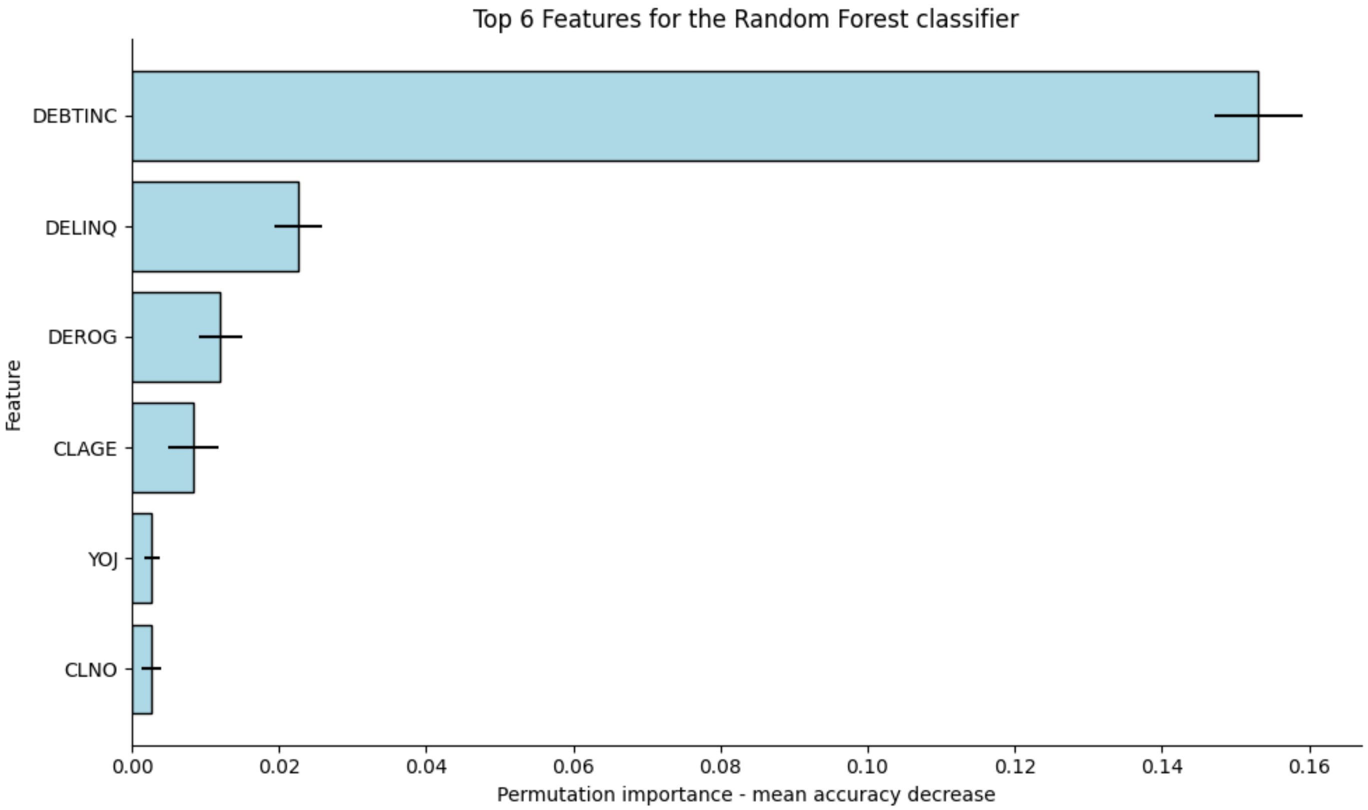
Global Profiling Analysis



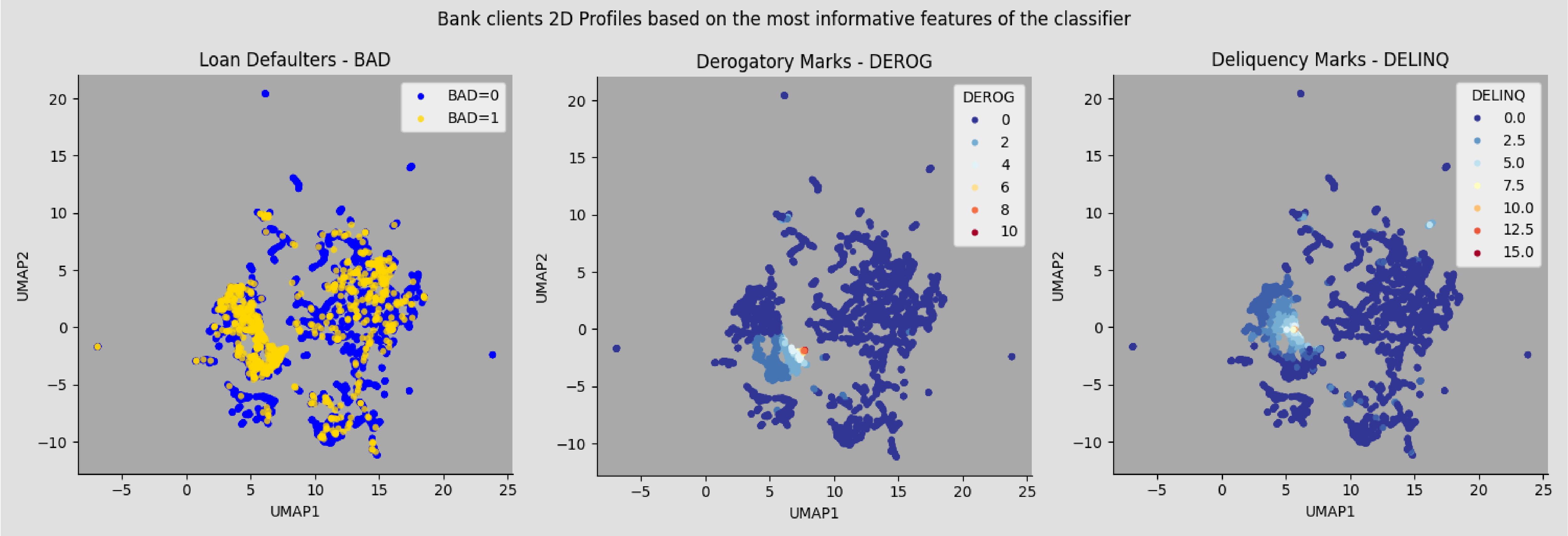
- Identification of **highly informative** features.
- **Profiling** of the loan defaulters and the bank handling system.

Classifiers and Feature Importance Results

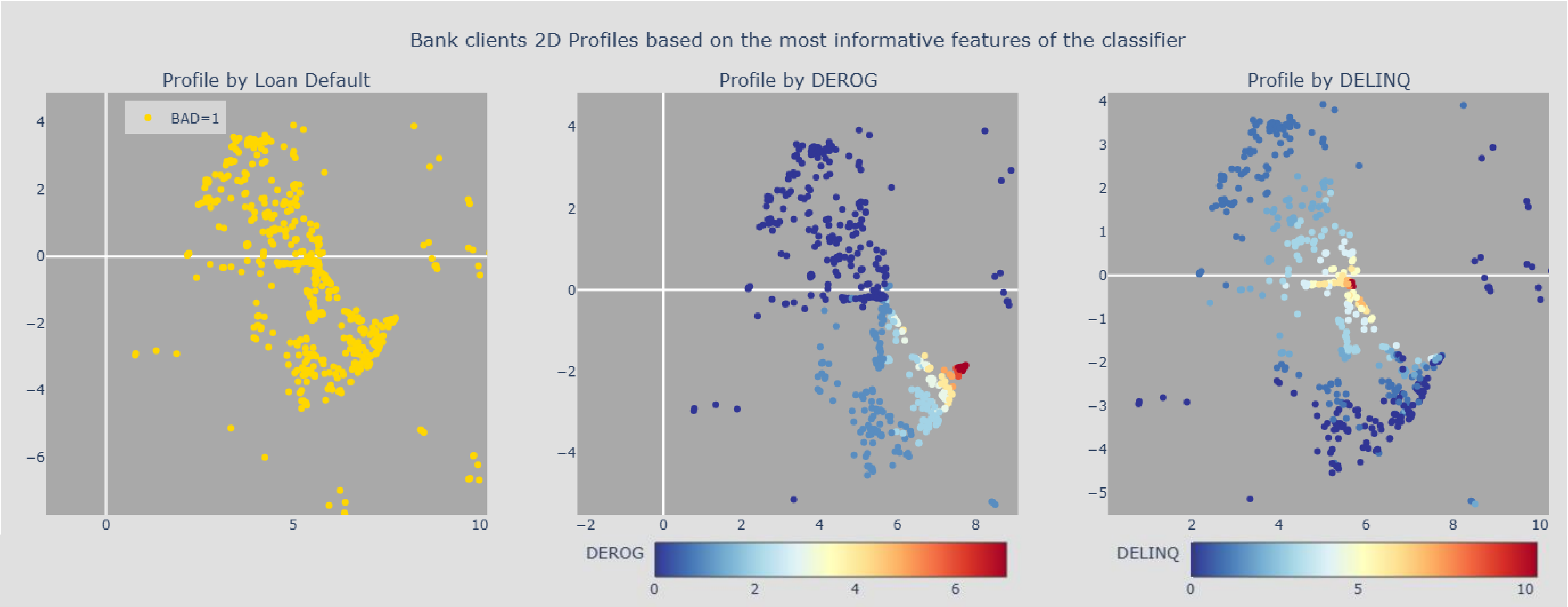
Model	True Positive Rate/ Recall (avg)	True Positive Rate/ Recall (positive class)	Precision	ROC-auc
L2-Logistic	0.72	0.56	0.7	0.81
Decision Tree	0.8	0.78	0.73	0.87
Random Forest	**0.84**	**0.8**	**0.77**	**0.92**



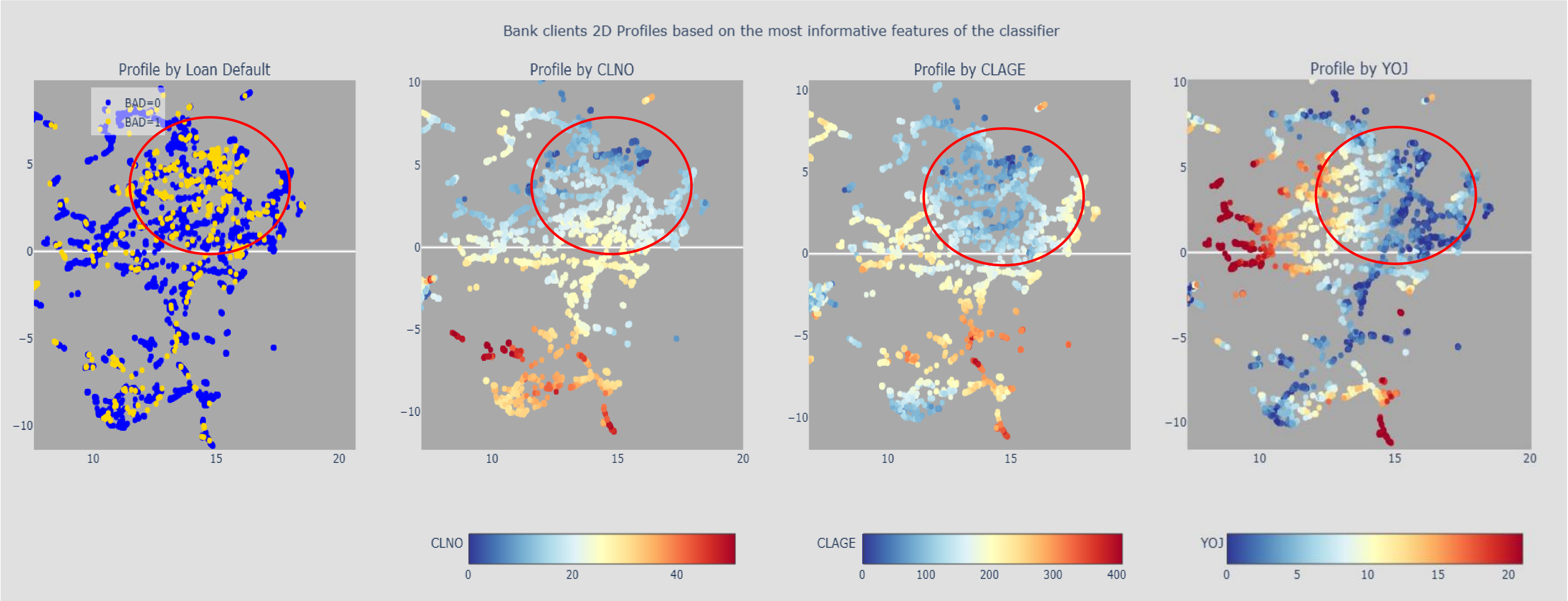
Profile analysis using top 6 features



In-depth exploration of the small island



What about the Big island?



Recommendations

Then bank should investigate the **missing records** regarding the client/borrower handling. Investment in better **data management practices** to cover the **incompleteness** is an important step.

Since some defaulters have delinquency marks, the bank should consider **internal handling practices**, if they are applicable. The model is not aware of them.

Need to expand the information retrieval regarding **age**, **income**, **profession** and **political & marital status**. These parameters might have an influence loan default in **absence of credit marks**.

Given **appropriate data curation,** the model will aid the loan approval decision, and the profiles could unearth hidden relationships.