



PROJECT IDEAS

QUANTITATIVE METHODS

DR. PHILIPPE J.S. DE BROUWER

ALL RIGHTS RESERVED BY THE AUTHOR
prepared for: AGH University of Krakow

NOVEMBER 27, 2024

Contents

1 Credit Risk	2
2 Insurance	3
3 Other Ideas	3

1 Credit Risk

Retail Credit Risk

For retail credit risk modeling, several datasets are available that include borrower demographics, loan details, and whether loans defaulted. These datasets often contain information like loan amount, repayment history, and more.

- A. **UCI Machine Learning Repository - Credit Card Default Dataset:** This dataset contains information about credit card clients and their default status. The task is to predict whether a client will default based on their demographic and financial information. Available at: [UCI Credit Card Default Dataset](#).
- B. **Kaggle - Loan Default Prediction Dataset:** This dataset offers a comprehensive overview of loan applicants, including demographic data (e.g., age, income) and loan details such as loan amount, loan type, and whether the loan defaulted. It can be accessed from Kaggle at [Kaggle Loan Default Dataset](#).
- C. **Lending Club Data:** Lending Club provides detailed data on loan defaults and borrower information, including income, loan amounts, loan purpose, and payment status. This dataset is widely used for predicting loan defaults. Access it at [Github](#).

These datasets contain key variables such as borrower demographics, loan amounts, purposes, and default statuses, which are ideal for retail credit risk modeling and predictive analysis.

Corporate Credit Risk

For corporate lending data, several datasets are available that provide detailed information about borrowers, loan terms, and defaults. These datasets include financial metrics, company size, loan amount, repayment status, and more.

- A. **UCI Machine Learning Repository - German Credit Data:** This dataset contains information about credit applicants, including financial data and loan details. It can be used for corporate credit scoring and default prediction. Available at: [UCI German Credit Data](#).

- B. **Data.gov - Business Lending Dataset:** This government dataset includes information on business loans issued by U.S. banks, including loan amounts, company size, and default status. This is valuable for corporate lending analysis. You can access it at: [Data.gov Business Lending](#).
- C. **LendingClub Corporate Lending Data:** LendingClub offers detailed loan and financial data for businesses, including data on default rates, loan amounts, and borrower information. Access the data here: [LendingClub Corporate Lending Data](#).

These datasets provide key variables such as company type, loan size, purpose, and financial health, which can be leveraged for credit risk modeling in corporate lending.

2 Insurance

Insurance Ideas

- A. **Kaggle - Car Insurance Claim Prediction:** These dataset include features like customer demographic details, driving history, and accident data to predict the likelihood of a claim. Available at: [Kaggle Car Insurance Claim Dataset](#).
- B. **Kaggle - Insurance:** Search for insurance data at: [Kaggle Insurance Datasets](#).

3 Other Ideas

Further Ideas

- A. Use the [search engine](#) of **UCI** to find a dataset of interest.
- B. **Kaggle:** Search for data of your interest at: [Kaggle](#).
- C. **Kaggle - Heart Disease Dataset:** The goal is to predict whether a patient has heart disease based on various health attributes. Available at: [Kaggle Heart Disease Dataset](#).
- D. **Data.gov:** Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more. 301,772 datasets available. It's available from [Data.gov Small Business Lending Dataset](#).