

Homework task for data scientist

Task

Our client is a car loan company and they reached out to Civitta to help them with the data analysis. The main goal of the data analysis is to better understand their customer base and the riskiness of different client groups. The client has not analyzed this topic before and is very interested in your insights.

The client wants the analysis to focus on the following:

- Explore the dataset and identify any interesting trends or patterns related to loan defaults (or other topics).
- Investigate factors that might contribute to the risk of defaulting on a loan. Which customer groups have the highest risk of defaulting? What is a typical risky/not-risky customer profile?
- Build a machine learning model that would predict the probability of default (model results should also be presented in the slide). What is the accuracy of different models and which one do you recommend to use?
- Provide strategic recommendations for the client on how to reduce loan defaults.

Dataset

The dataset contains 3 files:

- client_information file contains all the information about the customers.
- loan_information file contains data about loans (type, term, etc.).
- loan_outcome_information file contains the outcome of each loan (default, paid early, paid on time).

Deliverable

Visualize and present the analysis results in PowerPoint slides in English (up to 10 slides with main insights). Imagine that this will be the slide deck that will be sent to the client.