Unknown Groups of Bank Clients: Project Description

To successfully complete the project, process the steps below for your analysis. Apply concepts known from the lecture and tutorial. If applicable, the use of method extensions are allowed.

1. Explorative data analysis and data cleaning.
2. Data preprocessing.
3. Fit a model.
4. Model diagnostics.
5. Plotting.

Summarize your results within 8-10 pages. Your report should describe your analysis as well as the used methods. Cite papers if reasonable. A literature review or deeper evaluation of the topic is not necessary. Hand in your report by June 30th. Additionally, prepare a presentation of your results. The presentation should not exceed 20 minutes.

The Project "Unknown Groups of Bank Clients" uses clustering techniques with the aim to improve banks' risk management. For this purpose, a customer base is given and the algorithm tries to identify unknown patterns that form groups of customers. To do this the historical payment behavior and some demographic features are available for each customer.

- **ID**: Customer ID.
- **Credit**: Amount of given credit.
- **Sex**: Gender (1=male, 2=female).
- **Education**: Education level of the customer (1=graduate school, 2=university, 3=high school, 4=others).
- **Status**: Marital status (1=married, 2=single, 3=others).
- **Age**: Age in years.
- **Repayment 1-6**: Repayment status in January to June (Repayment 6 for June, Repayment 5 for May, Repayment 3 for April,...) with the following values: -2=no credit consumption, -1=repayed duly, 0=use of revolving credit, 1=payment delay of 1 month, 2=payment delay of 2 months, ... , 8=payment delay of 8 months, 9= payment delay of ≥ 9 months.
- **Amount 1-6**: Amount of the account balance from January to June (Amount 6 for June, Amount 5 for May, ...).
- **Previous 1-6**: Amount of the previous repayment in the corresponding month (Previous 6 for June, Previous 5 for May, ...).

Hint: Think about creating new columns based on the historical payment behavior (e.g. central tendency measures).