# Project description

Your project is to prepare a report for a bank's loan division. You'll need to find out if a customer's marital status and number of children have an impact on whether they will default on a loan. The bank already has some data on customers' credit worthiness.

Your report will be considered when building a **credit score** for a potential customer. A **credit score** is used to evaluate the ability of a potential borrower to repay their loan.

# Instructions for completing the project

**Step 1**. Open the data file `/datasets/credit_scoring_eng.csv` and have a look at the general information.

**Step 2**. Preprocess the data:

  Identify and fill in missing values

  Replace the real number data type with the integer type

  Delete duplicate data

  Categorize the data

Be sure to explain:

  Which missing values you identified

  Possible reasons these missing values were present

  Which method you used to fill in missing values

  Which method you used to find and delete duplicate data and why

  Possible reasons why duplicate data was present

  Which method you used to change the data type and why

  Which dictionaries you've selected for this dataset and why

The data may contain artifacts, or values that don't correspond to reality (for instance, a negative number of days employed). This kind of thing happens when you're working with real data. You need to describe the possible reasons such data may have turned up and process it.

**Step 3.** Answer these questions:

  Is there a connection between having kids and repaying a loan on time?

  Is there a connection between marital status and repaying a loan on time?

Is there a connection between income level and repaying a loan on time?

How do different loan purposes affect on-time loan repayment?

Interpret your answers. Explain what the results you obtained mean.

**Step 4.** Write an overall conclusion.

**Format:**

Complete the project in the Jupyter Notebook (it will open when you click Next). Write code in **code** cells and notes with explanations and interpretations in **markdown** cells. Use formatting and headings.

# Description of the data

`children` : the number of children in the family

`days_employed` : how long the customer has been working

`dob_years` : the customer's age

`education` : the customer's education level

`education_id` : identifier for the customer's education

`family_status` : the customer's marital status

`family_status_id` : identifier for the customer's marital status

`gender` : the customer's gender

`income_type` : the customer's income type

`debt` : whether the customer has ever defaulted on a loan

`total_income` : monthly income

`purpose` : reason for taking out a loan

# How will my project be assessed?

Your reviewer will use assessment criteria to evaluate your project. Before getting started, read the criteria carefully. This will help ensure that you do a good job.

Here's what mentors look at when assessing your project:

How do you describe the problems identified in the data?

What methods do you use to replace data types and to process missing values and duplicate data?

Did you categorize the data? Why did you do it the way you did?

Did you export the final data into pivot tables?

Did you use try-except clauses in your code to handle any unexpected errors?

Did you follow the correct project structure and write clean code?

Did you draw accurate and useful conclusions?

Did you leave comments on each step you took?

When you click "Let's go!" you will be taken to Practicum's JupyterHub. You can complete your project there.

Another option: you can install Python on your computer and do your project locally. There are many guides to installing Python and its libraries on your computer.

An easy way to install everything you need at once is to download the Anaconda distribution kit (https://www.anaconda.com/distribution/). During installation, check the following options in advanced settings:

`Add Anaconda to the system PATH environment variable`

`Register Anaconda as the system Python 3.n` (3.n will be replaced by the version number for the Python software included with your Anaconda download)

There are two ways to open the Jupyter Notebook:

Find the Jupyter Notebook in Anaconda Navigator and launch it by clicking the icon

Write *jupyter notebook* in Command Prompt (Windows) or Terminal (macOS)

The knowledge and skills you acquired in previous lessons will help you successfully complete your project.

Remember, you can always use the Knowledge Base as reference.

Good luck!