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 has 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