# **Home\_Credit\_EDA\_+\_ML**

## Importing Libraries

Firstly started Importing all necessary libraries.

## Data Reading and Exploration

* Then read the train and test csv files in python notebook with the help of pandas libraries.
* Showed descriptive statistics for training data and checked for presence of null values.
* Since our Target variable is binary so there is possibility of biased dataset. So I created a balanced dataset for analysis by concatenating two subsets of the traindf dataframe, where one subset contains all the rows with TARGET value equal to 1, and the other subset contains a random sample of rows with TARGET value equal to 0, with the same number of rows as the subset with TARGET value equal to 1.

## ****EDA & Visualization****

* Plotted histograms of categorical variables in the **traindf dataframe** using the sns.histplot() to visualize the distribution of target variables in the dataset and identify any imbalances or outliers.
* Plotted second histogram similar to the previous one, but it adds a new feature to show the class distribution for each category in the Target variables. It uses **balanceddf dataframe** instead of **traindf**, which is the balanced version of traindf with equal number of samples for each class. Purpose is to visualize the distribution of categorical variables in a dataset and identify any differences in class distributions across categories.
* Plotted a kernel density plot (kde) of Credit Amount.
* Plotted kernel density estimate (KDE) of the distribution of ages in a dataset, with the data represented by the "DAYS\_BIRTH" column.