**Include both the number of records, number of variables and a list of specific variables of interest with the data type of each.**

For the project, we will combine two datasets. The first one is application record dataset, there are 18 variables and 438557 observations. Another one is credit record dataset having 3 variables and 1,048,575 observations. These columns included Client number, Gender, Is there a car, Is there a property, Number of children, Annual income, Income category, Education level, Marital status, Way of living, Birthday, Start date of employment, Is there a mobile phone, Is there a work phone, Is there a phone, Is there an email, Occupation, and Family size. Client number, Record month, and Status are included in the credit record data. We think all the variables are interesting, except having the phone and email. Our main analysis items are annual income and whether there is a property, the number of children, and status. The type of annual income and number of children columns are numerical variables, and the property column is a factor variable.

Here is the list of variables that we are interested in to do Exploratory Data Analysis and perform Data Mining Techniques accordingly. This includes;

|  |  |
| --- | --- |
| **Features** | **Type** |
| ID | INTEGER |
| CODE\_GENDER | FACTOR |
| FLAG\_OWN\_CAR | FACTOR |
| FLAG\_OWN\_REALTY | FACTOR |
| CNT\_CHILDREN | INTEGER |
| AMT\_INCOME\_TOTAL | NUMERIC |
| NAME\_INCOME\_TYPE | FACTOR |
| NAME\_EDUCATION\_TYPE | FACTOR |
| NAME\_FAMILY\_STATUS | FACTOR |
| NAME\_HOUSING\_TYPE | FACTOR |
| DAYS\_BIRTH | INTEGER |
| DAYS\_EMPLOYED | INTEGER |
| OCCUPATION\_TYPE | FACTOR |
| CNT\_FAM\_MEMBERS | NUMERIC |

**Reference:**

[1] *Credit card Approval Prediction*. Kaggle. (n.d.). Retrieved September 26, 2021, from https://www.kaggle.com/rikdifos/credit-card-approval-prediction?