Description of the train dataset that used

1. Application train

* Rows : 307511 rows
* Columns : 122 cols (used 27 cols)

1. Application test

* Rows : 307511 rows
* Columns : 122 cols (used 26 cols)

1. Description of the datasets attributes:
2. SK\_ID\_CURR : ID of loan in our sample
3. TARGET : Target variable

- 1 client with payment difficulties;

* 0 all other cases

1. NAME\_CONTRACT\_TYPE : Identification if loan is cash or revolving

* Cash loan
* Revolving loans

1. CODE\_GENDER : Gender of the client

* F (Female)
* M (Male)

1. FLAG\_OWN\_CAR : Flag if the client owns a car

* Y (Yes)
* N (No)

1. FLAG\_OWN\_REALTY : Flag if client owns a house or flat

* Y (Yes)
* N (No)

1. CNT\_CHILDREN : Number of children the client has
2. AMT\_INCOME\_TOTAL : Income of the client
3. AMT\_CREDIT : Credit amount of the loan
4. AMT\_ANNUITY : Loan annuity
5. NAME\_INCOME\_TYPE : Clients income type

* Working
* State Servant
* Pensioner
* Unemployed
* Student
* Businessman
* Maternity Leave
* Commercial Associate

1. NAME\_EDUCATION\_TYPE : Level of highest education the client achieved

* Academic Degree
* Higher Education
* Incomplete Higher
* Lower Secondary
* Secondary/ Secondary Special

1. NAME\_ FAMILY\_STATUS: Family status of the client

* Single/Not Married
* Separated
* Married

1. NAME\_HOUSING\_TYPE : What is the housing situation of the client

* Co-op Apartment
* House/Apartment
* Municipal Apartment
* Office Apartment
* Rented Apartment
* With Parent

1. DAYS\_ID\_PUBLISH : How many days before the application did client change the identity document with which he applied for the loan
2. AGE : Client's age at the time of application
3. REGION\_RATING\_CLIENT : Our rating of the region where client lives (1,2,3)
4. REGION\_RATING\_CLIENT\_W\_CITY : Our rating of the region where client lives with taking city into account (1,2,3)
5. REG\_CITY\_NOT\_WORK\_CITY : Flag if client's permanent address does not match work address

* 1 (different)
* 0 (same at city level)

1. Ext\_source\_1 : Normalized score from external data source
2. Ext\_source\_2 : Normalized score from external data source
3. Ext\_source\_3 : Normalized score from external data source
4. DAYS\_LAST\_PHONE\_CHANGE : How many days before application did client change phone
5. AGE\_GROUP : Client's age group at the time of application

* Above Twenty
* Above Thirty
* Above Fourty
* Above Fifty
* Above Sixty

1. HAVING\_ACTIVE\_CREDIT : Client's has an active credit at the time of application

* Y (Yes)
* N (No)

1. NAME\_CONTRACT\_STATUS\_PREV : Contract status of previous application

* Approved
* Canceled
* Refund
* Unused Offer