DATA DICTIONARY

|  |  |  |
| --- | --- | --- |
| Slno | Col name | Description |
| 1 | SK\_ID\_CURR | ID of loan in our sample |
| 2 | TARGET | 1. Unable to pay the loan   0 - Able to pay the loan |
| 3 | NAME\_CONTRACT\_TYPE | Identification if loan is cash or revolving |
| 4 | CODE\_GENDER | Gender of the client |
| 5 | FLAG\_OWN\_CAR | Does the client own a car |
| 6 | FLAG\_OWN\_REALTY | Does the client has a property or not |
| 7 | CNT\_CHILDREN | Number of children the client has |
| 8 | AMT\_INCOME\_TOTAL | Income of the client |
| 9 | AMT\_CREDIT | total amount of money that is borrowed by the client |
| 10 | AMT\_ANNUITY | Amount paid / year |
| 11 | AMT\_GOODS\_PRICE | For consumer loans it is the price of the goods for which the loan is given |
| 12 | NAME\_TYPE\_SUITE | Who accompanied client when applying for the previous application |
| 13 | NAME\_INCOME\_TYPE | Clients income type (businessman, working, maternity leave,…) |
| 14 | NAME\_EDUCATION\_TYPE | highest education level the client achieved |
| 15 | NAME\_FAMILY\_STATUS | Marital status of the client |
| 16 | NAME\_HOUSING\_TYPE | What is the housing situation of the client (renting, living with parents, ...) |
| 17 | REGION\_POPULATION\_RELATIVE | Popoulation normalized value, higher value indicates more population in the area |
| 18 | DAYS\_BIRTH | Client's age in days at the time of application |
| 19 | DAYS\_EMPLOYED | How many days before the application the person started current employment |
| 20 | DAYS\_REGISTRATION | How many days before the application did client change his registration |
| 21 | DAYS\_ID\_PUBLISH | How many days before the application did client change the identity document with which he applied for the loan |
| 22 | OWN\_CAR\_AGE | Age of client's car |
| 23 | FLAG\_MOBIL | Did client provide mobile phone (1=YES, 0=NO) |
| 24 | FLAG\_EMP\_PHONE | Did client provide work phone (1=YES, 0=NO) |
| 25 | FLAG\_WORK\_PHONE | Did client provide home phone (1=YES, 0=NO) |
| 26 | FLAG\_CONT\_MOBILE | Was mobile phone reachable (1=YES, 0=NO) |
| 27 | FLAG\_PHONE | Did client provide home phone (1=YES, 0=NO) |
| 28 | FLAG\_EMAIL | Did client provide email (1=YES, 0=NO) |
| 29 | OCCUPATION\_TYPE | What kind of occupation does the client have |
| 30. | CNT\_FAM\_MEMBERS | How many family members does client have |
| 31 | REGION\_RATING\_CLIENT | Our rating of the region where client lives (1,2,3) |
| 32 | REGION\_RATING\_CLIENT\_W\_CITY | Our rating of the region where client lives with taking city into account (1,2,3) |
| 33 | WEEKDAY\_APPR\_PROCESS\_START | On which day of the week did the client apply for the loan |
| 34 | HOUR\_APPR\_PROCESS\_START | Approximately at what hour did the client apply for the loan |
| 35 | REG\_REGION\_NOT\_LIVE\_REGION | Flag if client's permanent address does not match contact address (1=different, 0=same, at region level) |
| 36 | REG\_REGION\_NOT\_WORK\_REGION | Flag if client's permanent address does not match work address (1=different, 0=same, at region level) |
| 37 | LIVE\_REGION\_NOT\_WORK\_REGION | Flag if client's contact address does not match work address (1=different, 0=same, at region level) |
| 38 | REG\_CITY\_NOT\_LIVE\_CITY | Flag if client's permanent address does not match contact address (1=different, 0=same, at city level) |
| 39 | REG\_CITY\_NOT\_WORK\_CITY | Flag if client's permanent address does not match work address (1=different, 0=same, at city level) |
| 40 | LIVE\_CITY\_NOT\_WORK\_CITY | Flag if client's contact address does not match work address (1=different, 0=same, at city level) |
| 41 | ORGANIZATION\_TYPE | Type of organization where client works |
| 42 | EXT\_SOURCE\_1 | Normalized score from external data source |
| 43 | EXT\_SOURCE\_2 | Normalized score from external data source |
| 44 | EXT\_SOURCE\_1 | Normalized score from external data source |
| 45 | APARTMENTS\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 46 | BASEMENTAREA\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 47 | YEARS\_BEGINEXPLUATATION\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 48 | YEARS\_BUILD\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 49 | COMMONAREA\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 50 | ELEVATORS\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 51 | ENTRANCES\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 52 | FLOORSMAX\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 53 | FLOORSMIN\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 54 | LANDAREA\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 55 | LIVINGAPARTMENTS\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 56 | LIVINGAREA\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 57 | NONLIVINGAPARTMENTS\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 58 | NONLIVINGAREA\_AVG | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 59 | APARTMENTS\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 60 | BASEMENTAREA\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 61 | YEARS\_BEGINEXPLUATATION\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 62 | YEARS\_BUILD\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 63 | COMMONAREA\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 64 | ELEVATORS\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 65 | ENTRANCES\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 66 | FLOORSMAX\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 67 | FLOORSMIN\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 68 | LANDAREA\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 69 | LIVINGAPARTMENTS\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 70 | LIVINGAREA\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 71 | NONLIVINGAPARTMENTS\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 72 | NONLIVINGAREA\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 73 | APARTMENTS\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 74 | BASEMENTAREA\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 75 | YEARS\_BEGINEXPLUATATION\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 76 | YEARS\_BUILD\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 77 | COMMONAREA\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 78 | ELEVATORS\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 79 | ENTRANCES\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 80 | FLOORSMAX\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 81 | FLOORSMIN\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 82 | LANDAREA\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 83 | LIVINGAPARTMENTS\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 84 | LIVINGAREA\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 85 | NONLIVINGAPARTMENTS\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 86 | NONLIVINGAREA\_MEDI | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 87 | FONDKAPREMONT\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 88 | HOUSETYPE\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 89 | TOTALAREA\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 90 | WALLSMATERIAL\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 91 | EMERGENCYSTATE\_MODE | Normalized information about building where the client lives, What is average (\_AVG suffix), modus (\_MODE suffix), median (\_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor |
| 92 | OBS\_30\_CNT\_SOCIAL\_CIRCLE | How many observation of client's social surroundings with observable 30 DPD (days past due) default |
| 93 | DEF\_30\_CNT\_SOCIAL\_CIRCLE | How many observation of client's social surroundings defaulted on 30 DPD (days past due) |
| 94 | OBS\_60\_CNT\_SOCIAL\_CIRCLE | How many observation of client's social surroundings with observable 60 DPD (days past due) default |
| 95 | DEF\_60\_CNT\_SOCIAL\_CIRCLE | How many observation of client's social surroundings defaulted on 60 (days past due) DPD |
| 96 | DAYS\_LAST\_PHONE\_CHANGE | How many days before application did client change phone |
| 97 | FLAG\_DOCUMENT\_2 | Did client provide document 2 |
| 98 | FLAG\_DOCUMENT\_3 | Did client provide document 3 |
| 99 | FLAG\_DOCUMENT\_4 | Did client provide document 4 |
| 100 | FLAG\_DOCUMENT\_5 | Did client provide document 5 |
| 101 | FLAG\_DOCUMENT\_6 | Did client provide document 6 |
| 102 | FLAG\_DOCUMENT\_7 | Did client provide document 7 |
| 103 | FLAG\_DOCUMENT\_8 | Did client provide document 8 |
| 104 | FLAG\_DOCUMENT\_9 | Did client provide document 9 |
| 105 | FLAG\_DOCUMENT\_10 | Did client provide document 10 |
| 106 | FLAG\_DOCUMENT\_11 | Did client provide document 11 |
| 107 | FLAG\_DOCUMENT\_12 | Did client provide document 12 |
| 108 | FLAG\_DOCUMENT\_13 | Did client provide document 13 |
| 109 | FLAG\_DOCUMENT\_14 | Did client provide document 14 |
| 110 | FLAG\_DOCUMENT\_15 | Did client provide document 15 |
| 111 | FLAG\_DOCUMENT\_16 | Did client provide document 16 |
| 112 | FLAG\_DOCUMENT\_17 | Did client provide document 17 |
| 113 | FLAG\_DOCUMENT\_18 | Did client provide document 18 |
| 114 | FLAG\_DOCUMENT\_19 | Did client provide document 19 |
| 115 | FLAG\_DOCUMENT\_20 | Did client provide document 20 |
| 116 | FLAG\_DOCUMENT\_21 | Did client provide document 21 |
| 117 | AMT\_REQ\_CREDIT\_BUREAU\_HOUR | Credit Bureau is an agency that collects and maintains credit information on individuals. |
| 118 | AMT\_REQ\_CREDIT\_BUREAU\_DAY | Number of enquiries to Credit Bureau about the client one day before application (excluding one hour before application) |
| 119 | AMT\_REQ\_CREDIT\_BUREAU\_WEEK | Number of enquiries to Credit Bureau about the client one week before application (excluding one day before application) |
| 120 | AMT\_REQ\_CREDIT\_BUREAU\_MON | Number of enquiries to Credit Bureau about the client one month before application (excluding one week before application) |
| 121 | AMT\_REQ\_CREDIT\_BUREAU\_QRT | Number of enquiries to Credit Bureau about the client 3 month before application (excluding one month before application) |
| 122 | AMT\_REQ\_CREDIT\_BUREAU\_YEAR | Number of enquiries to Credit Bureau about the client one day year (excluding last 3 months before application) |

1. Get details about the persons having the oldest cars and latest cars

df[df["OWN\_CAR\_AGE"]==df["OWN\_CAR\_AGE"].max()]

check multicollinarity excists or not if yes go for PCA- AMT\_REQ\_CREDIT\_BUREAU\_DAY, AMT\_REQ\_CREDIT\_BUREAU\_WEEK

Removing non essential features will result in a

1. Better performing model
2. In an easier to understand model
3. In a model that runs faster