**Project-6**

**Decision Making**

Columns that has missing values:

OWN\_CAR\_AGE:

OCCUPATION\_TYPE:

EXT\_SOURCE\_1:

EXT\_SOURCE\_2:

EXT\_SOURCE\_3:

APARTMENTS\_AVG:

BASEMENTAREA\_AVG:

YEARS\_BEGINEXPLUATION\_AVG:

YEARS\_BUILD\_AVG:

COMMONAREA\_AVG:

ELEVATORS\_AVG:

ENTRANCES\_AVG:

FLOORSMAX\_AVG:

FLOORSMIN\_AVG:

LANDAREA\_AVG:

LIVINGAPARTMENTS\_AVG:

LIVINGAREA\_AVG:

NONLIVINGAPARTMENTS\_AVG:

NONLIVINGAREA\_AVG:

APARTMENTS\_MODE:

BASEMENTAREA\_MODE:  
YEARS\_EXPLUATION\_MODE:

YEARS\_BUILD\_MODE:

COMMONAREA\_MODE:

ELEVATORS\_MODE:

ENTRANCES\_MODE:  
FLOORSMAX\_MODE:

FLOORSMIN\_MODE:

LANDAREA\_MODE:

LIVINGAPARTMENTS\_MODE:

LIVINGAREA\_MODE:

NONLIVINGAPARTMENTS\_MODE:

NONLIVINGAREA\_MODE:

APARTMENTS\_MEDI:

BASEMENTAREA\_MEDI:

YEARSBEGIN\_EXPLUATION\_MEDI:

YEARS\_BUILD\_MEDI:

COMMON\_AREA\_MEDI:

ELEVATORS\_MEDI:

ENTRANCES\_MEDI:

FLOORSMAX\_MEDI:

FLOORSMIN\_MEDI:

LANDAREA\_MEDI:

LIVINGAPARTMENTS\_MEDI:

LIVINGAREA\_MEDI:

NONLIVINGAPARTMENTS\_MEDI:

NONLIVINGAREA\_MEDI:

FONDKAPREMONT\_MODE:

HOUSETYPE\_MODE:

TOTALAREA\_MODE:

WALLSMATERIAL\_MODE:

EMERGENCYSTATE\_MODE:

OBS\_30\_CNT\_SOCIAL\_CIRCLE:

DEF\_30\_CNT\_SOCIAL\_CIRCLE:

OBS\_60\_CNT\_SOCIAL\_CIRCLE:

DEF\_60\_CNT\_SOCIAL\_CIRCLE:

DAYS\_LAST\_PHONE\_CHANGE: only 1 column has missing values

AMT\_REQ\_CREDIT\_BUREAU\_HOUR:

AMT\_REQ\_CREDIT\_BUREAU\_DAY:

AMT\_REQ\_CREDIT\_BUREAU\_WEEK:

AMT\_REQ\_CREDIT\_BUREAU\_MON:

AMT\_REQ\_CREDIT\_BUREAU\_QRT:

AMT\_REQ\_CREDIT\_BUREAU\_YEAR:

Based on the significance of each column to the project and the nature of the missing values, here are some decisions I made for handling the missing data:

1. **OCCUPATION\_TYPE:**
   * Occupation type could potentially be an important factor in determining a client's ability to repay a loan. Since there are missing values, you could consider:
     + Imputing missing values with the most common occupation type.
2. **EXT\_SOURCE\_1, EXT\_SOURCE\_2, EXT\_SOURCE\_3:**
   * These columns represent normalized scores from external data sources, which could be crucial for predicting loan default.
     + I imputed the missing values with the median score for each column.
3. **AMT\_GOODS\_PRICE:**

* I imputed this column with its average price