# Application:

* Insignificant Variables:
  + FLAG\_DOCUMENT\_2 - 21
* Variables with many missing values (**Dropped**):
  + OWN\_CAR\_AGE
  + EXT\_SOURCE\_1
  + EXT\_SOURCE\_3
  + 'APARTMENTS\_AVG'
  + 'BASEMENTAREA\_AVG’
  + YEARS\_BEGINEXPLUATATION\_AVG
  + 'YEARS\_BUILD\_AVG
  + COMMONAREA\_AVG
  + 'ELEVATORS\_AVG'
  + 'ENTRANCES\_AVG'
  + 'FLOORSMIN\_AVG'
  + 'FLOORSMAX\_AVG'
  + 'LANDAREA\_AVG'
  + 'LIVINGAPARTMENTS\_AVG'
  + 'LIVINGAREA\_AVG'
  + 'NONLIVINGAPARTMENTS\_AVG'
  + 'NONLIVINGAREA\_AVG'
  + APARTMENTS\_MODE
  + BASEMENTAREA\_MODE
  + YEARS\_BEGINEXPLUATATION\_MODE
  + 'YEARS\_BUILD\_MODE'
  + COMMONAREA\_MODE
  + 'ELEVATORS\_MODE'
  + 'ENTRANCES\_MODE'
  + FLOORSMAX\_MODE
  + FLOORSMIN\_MODE'
  + LIVINGAPARTMENTS\_MODE
  + LIVINGAREA\_MODE
  + NONLIVINGAPARTMENTS\_MODE
  + 'NONLIVINGAREA\_MODE'
  + 'NONLIVINGAREA\_MODE'
  + 'APARTMENTS\_MEDI'
  + 'BASEMENTAREA\_MEDI'
  + YEARS\_BEGINEXPLUATATION\_MEDI
  + 'YEARS\_BUILD\_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'
* Constant Variables:
  + FLAG\_MOBIL
* Columns with some missing variables - Numeric ( **Replaced NAN values with mean**):

'AMT\_ANNUITY', 'AMT\_GOODS\_PRICE',

, 'EXT\_SOURCE\_2',

'OBS\_30\_CNT\_SOCIAL\_CIRCLE', 'DEF\_30\_CNT\_SOCIAL\_CIRCLE',

'OBS\_60\_CNT\_SOCIAL\_CIRCLE', 'DEF\_60\_CNT\_SOCIAL\_CIRCLE',

'DAYS\_LAST\_PHONE\_CHANGE', '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'

* Columns with some missing variables - Categorical (**REPLACED NAN VALUES WITH MOST FREQUENT OBS**)
  + 'NAME\_TYPE\_SUITE', 'OCCUPATION\_TYPE',

# Bureau Balance:

* Insignificant Variables:
  + MONTHS\_BALANCE

# Bureau:

* Insignificant Variables:
  + CREDIT\_CURRENCY, DAYS\_CREDIT,'CREDIT\_DAY\_OVERDUE',

'DAYS\_CREDIT\_ENDDATE', 'DAYS\_ENDDATE\_FACT', 'DAYS\_CREDIT\_UPDATE', 'AMT\_ANNUITY', ‘STATUS\_0’

# Installments\_payments:

* Variables with many missing values:
  + 'DAYS\_ENTRY\_PAYMENT', 'AMT\_PAYMENT'

# Credit\_card\_balance

* Imbalanced data (kept):
  + 'AMT\_DRAWINGS\_ATM\_CURRENT', 'AMT\_DRAWINGS\_CURRENT',

'AMT\_DRAWINGS\_OTHER\_CURRENT', 'AMT\_DRAWINGS\_POS\_CURRENT',

'AMT\_RECIVABLE', 'AMT\_TOTAL\_RECEIVABLE'

# POS\_CASH\_balance:

* Imbalanced data (kept):
  + 'SK\_DPD', 'CNT\_INSTALMENT\_FUTURE'