# Application data

1. Total distinct applications: **307511**
2. Duplicate rows: **0**
3. Applicants who have earlier taken the loan elsewhere: **263491 (86.69%)**
4. 2 types of contract (NAME\_CONTRACT\_TYPE)
5. Revolving loans
6. Cash loans
7. 3 genders (CODE\_GENDER)
8. M
9. XNA
10. F
11. Target variable highly skewed:

|  |  |  |
| --- | --- | --- |
| Target | TotalRows | Percentage |
| 0 | 282686 | 91.92711805 |
| 1 | 24825 | 8.07288194 |

# Bureau

1. No Duplicates
2. Total rows: 1716428
3. Join with Bureau\_Balance, only 774354 IDs
4. Logic to consider active applications:

|  |
| --- |
| SELECT COUNT(\*)  FROM  (  SELECT \*  FROM Final.Bureau  WHERE CREDIT\_ACTIVE = 'ACTIVE' AND (DAYS\_CREDIT\_ENDDATE IS NULL OR DAYS\_CREDIT\_ENDDATE > 0)  ) x  JOIN  (  SELECT SK\_ID\_BUREAU  FROM [Final].[BureauBalance]  GROUP BY SK\_ID\_BUREAU  HAVING MIN(CASE STATUS WHEN 'C' THEN 0 ELSE 1 END) != 0  ORDER BY MONTHS\_BALANCE  )y  ON x.SK\_ID\_BUREAU = y.SK\_ID\_BUREAU |

# Bureau\_Balance

1. Total Rows: 27299925
2. 43041 IDs not present in bureau

# Previous\_Application

1. Total Rows: 1670214
2. There are 4 categories of loan in this table whereas in main table there are 2.
   1. Consumer loans
   2. Revolving loans
   3. Cash loans
   4. XNA

# POS\_CASH\_BALANCE

1. Total Rows: 10001358
2. Distinct Ids: 936325
3. Joining with previous application results in: 898903 distinct Ids
4. Ids not presen tin previous application: 37422
5. **0 rows when joined with Credit\_Card\_Balance**
6. **924949 distinct Ids when joined with installments\_Payments**
7. Distinct contract types:
   1. Consumer loans
   2. Cash loans
8. For all Active POS applications where there is an installment pending, we have the annuity amount
9. Active application logic:

|  |
| --- |
| SELECT  SK\_ID\_CURR  ,SUM(CASE NAME\_CONTRACT\_TYPE WHEN 'Consumer loans' THEN 1 ELSE 0 END) TOTAL\_PREVIOUS\_ACTIVE\_CONSUMER\_LOANS  ,SUM(CASE NAME\_CONTRACT\_TYPE WHEN 'Cash loans' THEN 1 ELSE 0 END) TOTAL\_PREVIOUS\_ACTIVE\_CASH\_LOANS  ,SUM(CC.AMT\_ANNUITY) AVG\_PAYMENT\_PER\_MONTH\_POS  ,SUM(PENDING\_INSTALLMENTS\_POS) PENDING\_INSTALLMENTS\_POS  FROM Final.PreviousApplication CC  JOIN  (  SELECT SK\_ID\_PREV  ,MIN(CNT\_INSTALMENT\_FUTURE) AS PENDING\_INSTALLMENTS\_POS  FROM Final.POSCashBalance  --WHERE NAME\_CONTRACT\_STATUS = 'Active'  GROUP BY SK\_ID\_PREV  HAVING MIN(CNT\_INSTALMENT\_FUTURE)!=0  )Tmp  ON Cc.SK\_ID\_PREV = Tmp.SK\_ID\_PREV  GROUP BY SK\_ID\_CURR |

# Installments\_Payments

1. Total Rows: 13605401
2. DISTINCT Ids: 997752
3. Joining with previous application results in: 958905 distinct Ids
4. Ids not present in previous application: 38847
5. 337 Distinct Ids that are not present in credit\_Card\_Balance and POS\_Cash\_Balance
6. Distinct Contract types
   1. Cash loans
   2. Consumer loans
   3. Revolving loans

# Credit\_Card\_Balance

1. Total Rows: 3840312
2. DISTINCT Ids: 104307
3. Joining with previous application results in: 92935 distinct Ids
4. Ids not present in previous application: 11372
5. **0 rows when joined with pos\_Cash\_Balance**
6. **72466 distinct Ids when joined with installments\_Payments**
7. Distinct contract types:
   1. Revolving loans