# Bank Loan Case Study

## A. Identify Missing Data and Deal with it Appropriately:

Here we are removing unwanted/irrelevant columns which are unfit for analysis and the unwanted columns are specified by calculating the number of null values in the respective columns.

The unwanted columns are as follows

OWN_CAR_AGE,EXT_SOURCE_1,APARTMENTS_AVG,BASEMENTAREA_AVG,YEARS_BUILD_AVG, COMMON_AREA_AVG ,ELEVATORS_AVG, _AVG ,FLOORSMAX_AVG ,FLOORSMIN_AVG, LANDAREA_AVG,LIVINGAPARTMENTS_AVG,LIVINGAREA_AVG,NONLIVINGAPARTMENTS_AVG, NONLIVINGAREA_AVG,APARTMENTS_MODEBASEMENTAREA_MODE,YEARS_BUILD_MODE, COMMON_AREA_MODE,ELEVATORS_MODE,ENTRANCES_MODE,_MODE,FLOORSMIN_MODE, LANDAREA_MODE,LIVINGAPARTMENTS_MODE,LIVINGAREA_MODE, NONLIVINGAPARTMENTS_MODE,NONLIVINGAREA_MODE,_MEDIAN,BASEMENTAREA_MEDIAN, COMMON_AREA_MEDIAN 2,ELEVATORS_MEDIAN ,ENTRANCES_, FLOORSMAX_MEDIAN,FLOORSMIN_MEDIAN,LANDAREA_MEDIAN,_MEDIAN, LIVINGAREA_MEDIAN,NONLIVINGAPARTMENTS_MEDIAN,NONLIVINGAREA_MEDIAN, FONDKAPREMONT_MODE,HOUSETYPE_MODE , WALLSMATERIAL_MODE, FLAG_MOBIL, LAG_EMPLOY_PHONE,FLAG_WORK_PHONE,FLAG_CONT_MOBILE,FLAG_PHONE, FLAG_EMAIL,CNT_FAMILY_MEMBERS,REGION_RATING_CLENT,REGION_RATING_CLENT_W_CITY, EXT_SOURCE_3,YEAR_BEGINEXPLUATATION_AVG,YEAR_BEGINEXPLUATATION_MODE, YEAR_BEGINEXPLUATATION_MEDIAN,TOTAL_AREA_MODE,EMERGENCYSTATE_MODE,DAYS_LAST _PHONE_CHANGE, FLAG DOC 2 ,FLAG DOC 3, FLAG DOC 4 ,FLAG DOC 5 ,FLAG DOC 6 ,FLAG DOC7,FLAG DOC 8 ,FLAG DOC 9, FLAG DOC 10 ,FLAG DOC 11 ,FLAG DOC 12, FLAG DOC 13, FLAG DOC 14 , FLAG DOC 15 ,FLAG DOC 16 ,FLAG DOC 17 ,FLAG DOC 18 ,FLAG DOC 19, FLAG DOC 20, FLAG DOC 21.
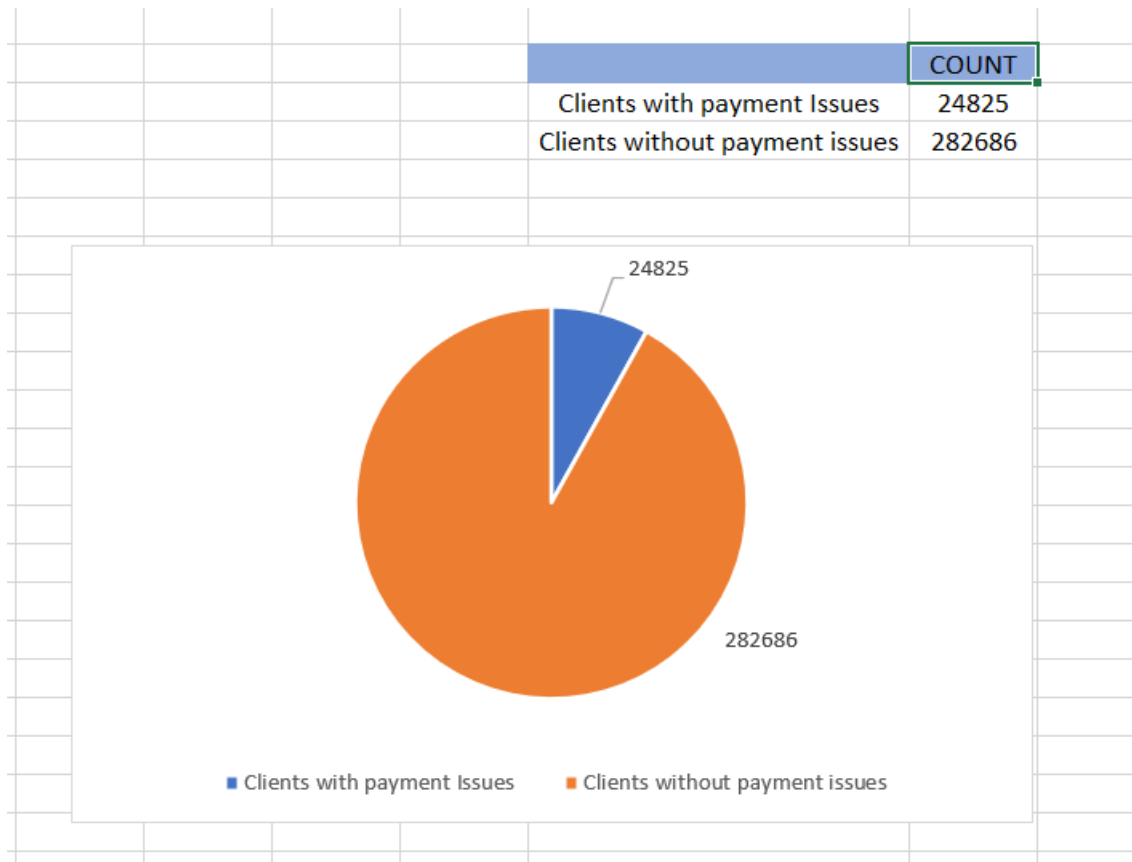
Link For cleaned Data

## B. Identify Outliers in the Dataset:

Here the outliers are defined by using Box plots.

## C. Analyze Data Imbalance:

Analysis based on payment issues

| | COUNT |
|---|---|
| Clients with payment Issues | 24825 |
| Clients without payment issues | 282686 |



Analysis based on gender

| Gender | Count |
|---|---|
| Male | 105059 |
| Female | 202448 |

| TYPE | Count |
|---|---|
| Co-op apartment | 1122 |
| House / apartment | 272868 |
| Municipal apartment | 11183 |
| Office apartment | 2617 |
| Rented apartment | 4881 |
| With parents | 14840 |

## Type of Housing

| Co-op apartment | House / apartment | Municipal apartment | Office apartment | Rented apartment | With parents |
|---|---|---|---|---|---|
| 1122 | 272868 | 11183 | 2617 | 4881 | 14840 |
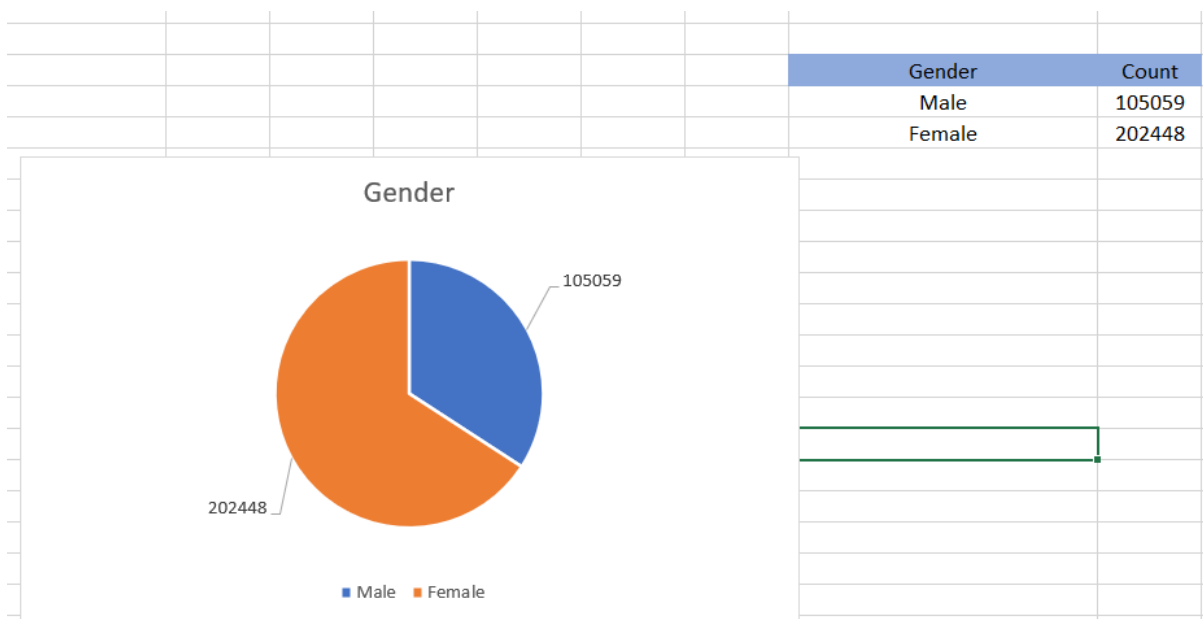
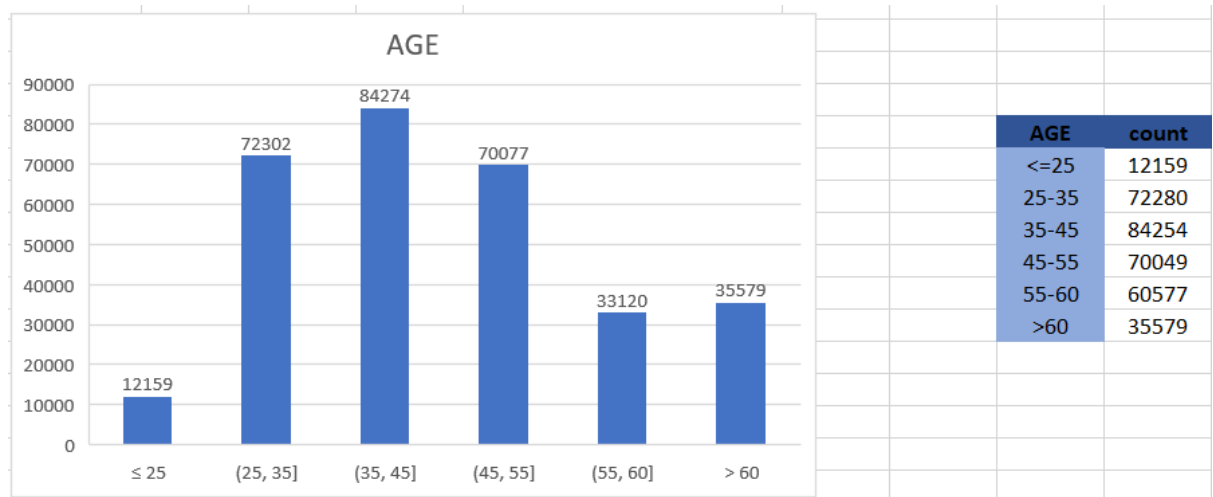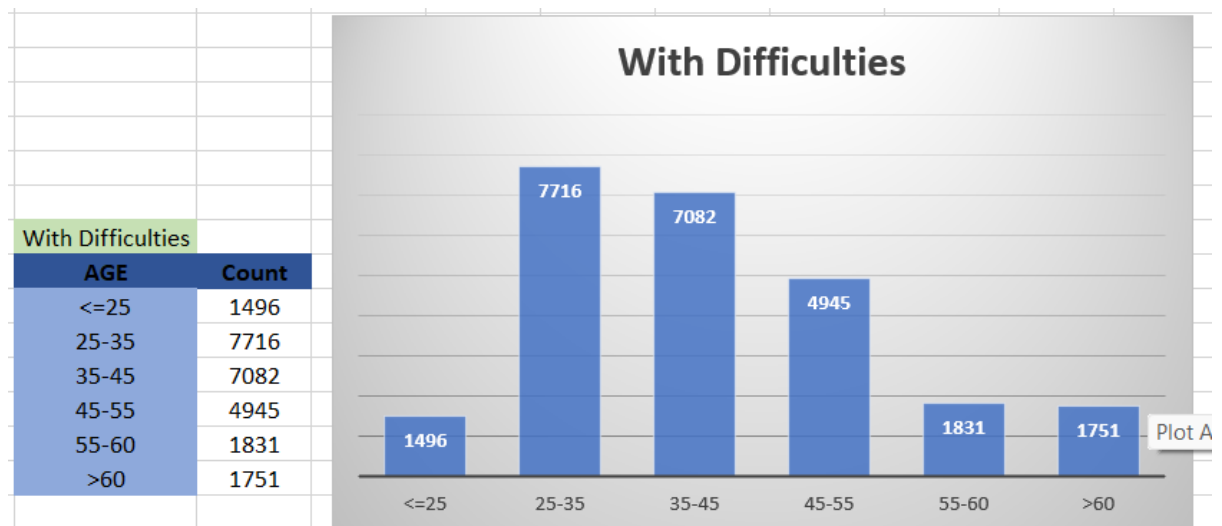## D. Perform Univariate, Segmented Univariate, and Bivariate Analysis:

## Univariate analysis: AGE

Here we are counting the persons of particular age group using countifs function in Excel.



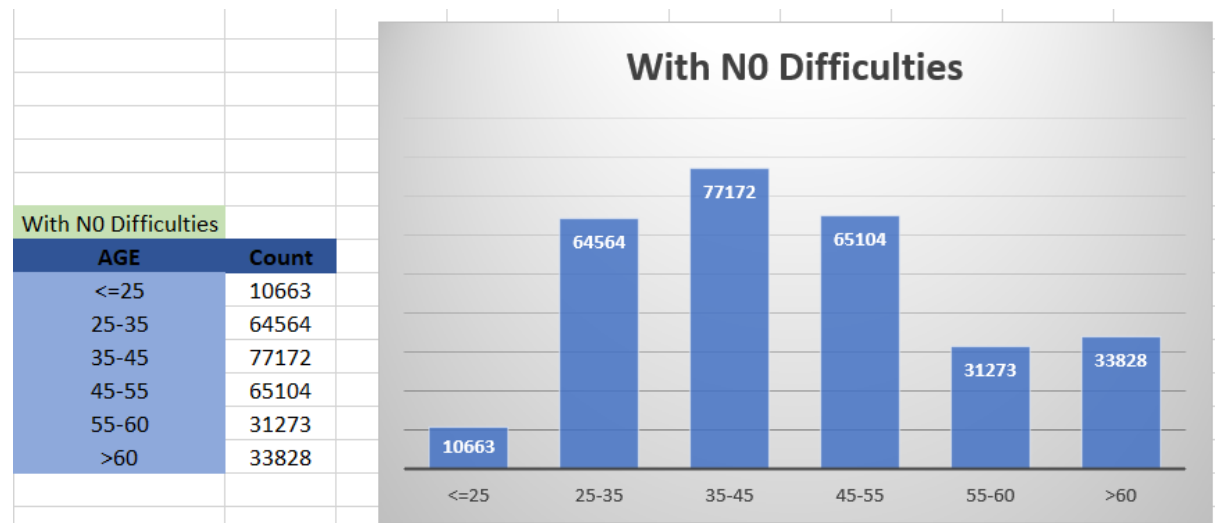| AGE | count |
|-----|-------|
| <=25 | 12159 |
| 25-35 | 72280 |
| 35-45 | 84254 |
| 45-55 | 70049 |
| 55-60 | 60577 |
| >60 | 35579 |

By the above graph we can understand that age from 35-45 has highest number of applications.

## With Payment Difficulties.

With Difficulties

| AGE | Count |
|-----|-------|
| <=25 | 1496 |
| 25-35 | 7716 |
| 35-45 | 7082 |
| 45-55 | 4945 |
| 55-60 | 1831 |
| >60 | 1751 |



By the above graph we can understand that age group between 25-35 has more payment difficulties.

## With No payment Difficulties/ Other Cases.

| With N0 Difficulties | |
|---|---|
| **AGE** | **Count** |
| <=25 | 10663 |
| 25-35 | 64564 |
| 35-45 | 77172 |
| 45-55 | 65104 |
| 55-60 | 31273 |
| >60 | 33828 |



By the above graph We can say that age group in between 35-45 are having no issues in payments.

## Based on Occupation

| Count of TARGET | Column Labels | |
|---|---|---|
| Row Labels | 0 | Grand Total |
| Businessman | 10 | 10 |
| Commercial associate | 66257 | 66257 |
| Maternity leave | 3 | 3 |
| Pensioner | 52380 | 52380 |
| State servant | 20454 | 20454 |
| Student | 18 | 18 |
| Unemployed | 14 | 14 |
| Working | 143550 | 143550 |
| Grand Total | 282686 | 282686 |
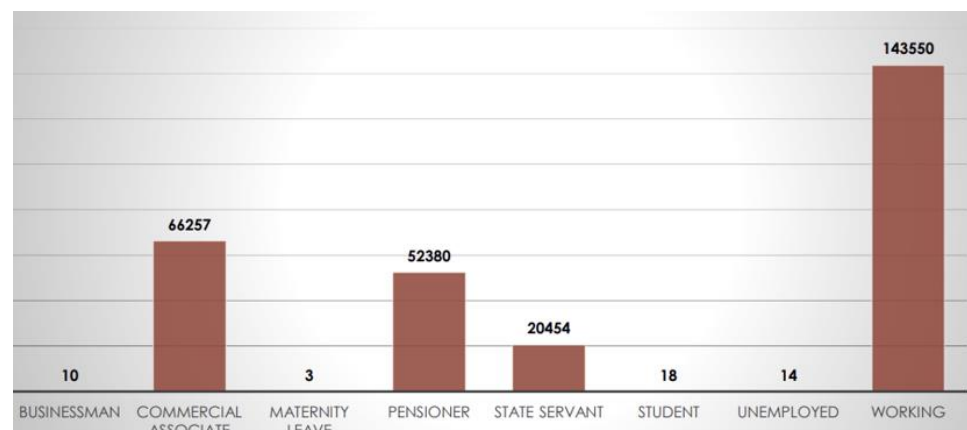
Based on AMT_CREDIT:
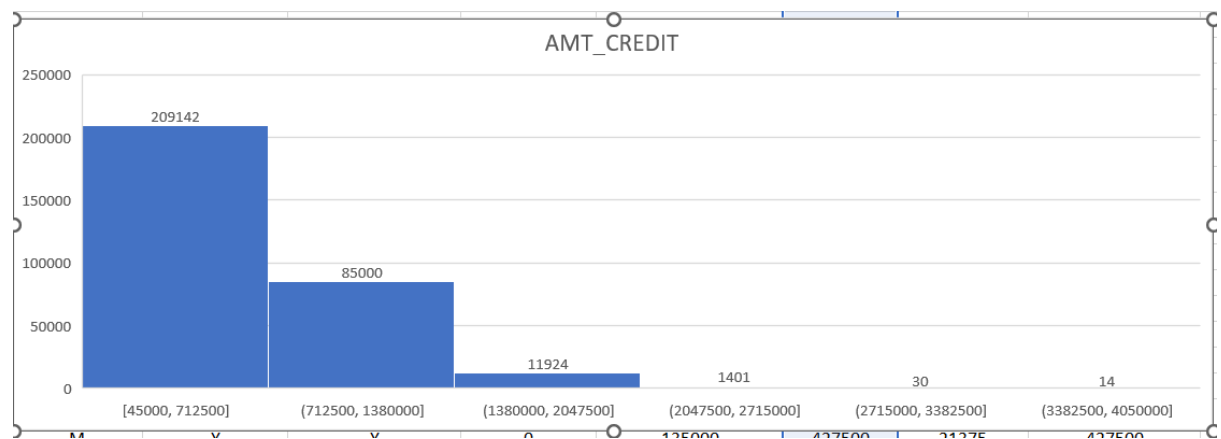


The common amount credited in between 45000-712500.

BIVARIATE ANALYSIS:

Types of Loans VS Gender:

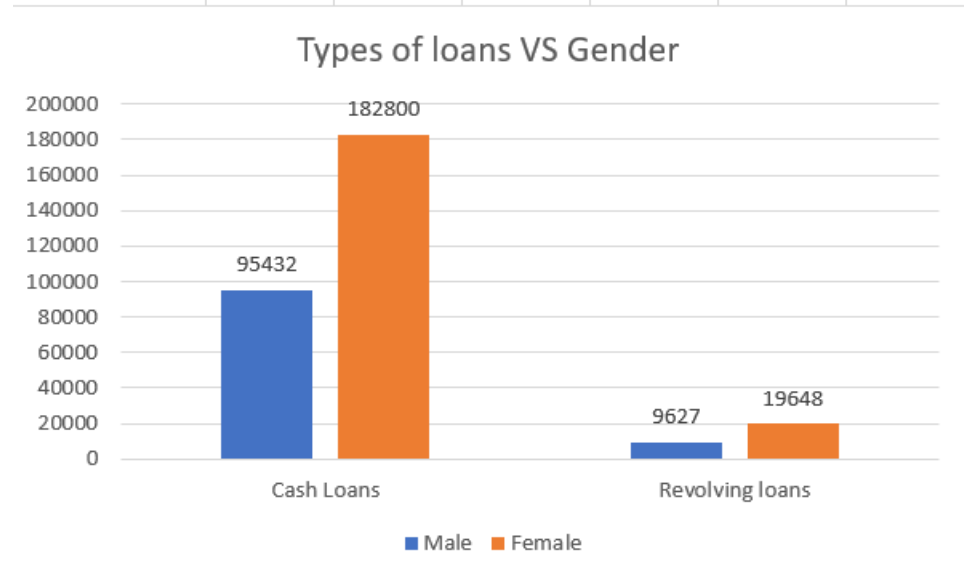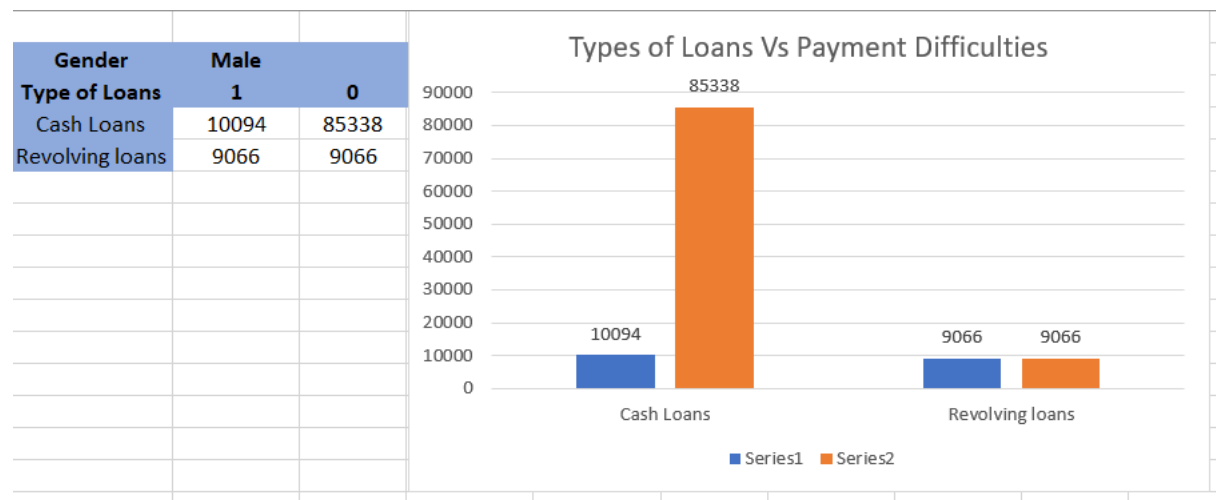| Type of Loans | Gender | | |
| --- | --- | --- | --- |
| | Male | Female | count |
| Cash Loans | 95432 | 182800 | 278232 |
| Revolving loans | 9627 | 19648 | 29279 |

## Types of loans VS payment Difficulties of Men:

Here Series 1 represents 1 i.,e Difficulties in making the payments and Series 2 represents Other Cases/ No difficulties in making the payments.

| Gender | Male | |
|---|---|---|
| **Type of Loans** | **1** | **0** |
| Cash Loans | 10094 | 85338 |
| Revolving loans | 9066 | 9066 |

**Types of Loans Vs Payment Difficulties**

| | Cash Loans | Revolving loans |
|---|---|---|
| Series1 | 10094 | 9066 |
| Series2 | 85338 | 9066 |

## Types of loans VS payment Difficulties of Women:

Here Series 1 represents 1 i.,e Difficulties in making the payments and Series 2 represents Other Cases/ No difficulties in making the payments.

| Gender | Female | |
|---|---|---|
| **Type of Loans** | **1** | **0** |
| Cash Loans | 13127 | 169673 |
| Revolving loans | 1043 | 18605 |

**Types of loans VS payment difficulties**

| | 1 | 2 |
|---|---|---|
| Cash Loans | 13127 | 169673 |
| Revolving loans | 1043 | 18605 |
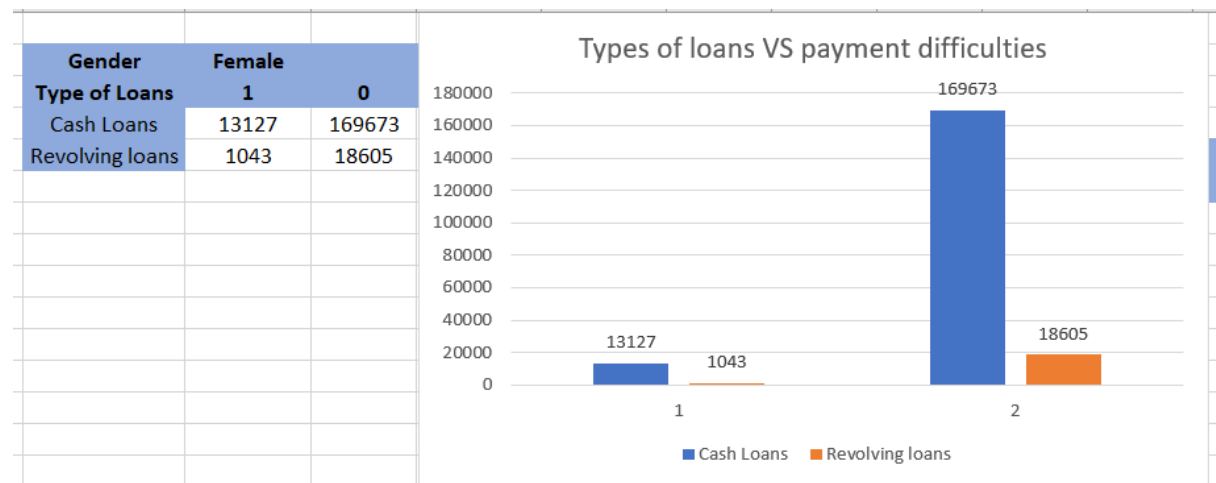
## E. Identify Top Correlations for Different Scenarios:

Here we are removing the irrelevant columns which are unfit for Analysis.

We are removing the following columns:

O HOUR_APPR_PROCESS_START , WEEKDAY_APPR_PROCESS_START_PREV, FLAG_LAST_APPL_PER_CONTRACT , NFLAG_LAST_APPL_IN_DAY , SK_ID_CURR, WEEKDAY_APPR_PROCESS_START.

Link for cleaned data:

https://docs.google.com/spreadsheets/d/10FH6FIhqMI1NoFO5Q1D8bZp_E5y-4wS_/edit?usp=sharing&ouid=108880336182281145657&rtpof=true&sd=true

| Count of NAME_CONTRACT_STATUS | Column Labels | | | | |
| --- | --- | --- | --- | --- | --- |
| Row Labels | Approved | Canceled | Refused | Unused offer | Grand Total |
| Building a house or an annex | 434 | 60 | 1188 | | 1682 |
| Business development | 78 | 12 | 164 | | 254 |
| Buying a garage | 28 | 5 | 51 | | 84 |
| Buying a holiday home / land | 91 | 13 | 230 | | 334 |
| Buying a home | 130 | 23 | 393 | | 546 |
| Buying a new car | 139 | 29 | 465 | 4 | 637 |
| Buying a used car | 552 | 57 | 1166 | 9 | 1784 |
| Car repairs | 223 | 14 | 256 | | 493 |
| Education | 481 | 14 | 476 | 4 | 975 |
| Everyday expenses | 732 | 8 | 740 | 7 | 1487 |
| Furniture | 210 | 15 | 250 | | 475 |
| Gasification / water supply | 75 | 3 | 125 | | 203 |
| Hobby | 11 | | 20 | | 31 |
| Journey | 329 | 10 | 404 | 2 | 745 |
| Medicine | 676 | 25 | 696 | 5 | 1402 |
| Money for a third person | 10 | | 6 | | 16 |
| Other | 4106 | 186 | 5310 | 62 | 9664 |
| Payments on other loans | 189 | 45 | 973 | 3 | 1210 |
| Purchase of electronic equipment | 357 | 4 | 280 | 3 | 644 |
| Refusal to name the goal | 1 | | 7 | | 8 |
| Repairs | 5385 | 381 | 8973 | 28 | 14767 |
| Urgent needs | 2228 | 83 | 2998 | | 5309 |
| Wedding / gift / holiday | 248 | 10 | 336 | | 594 |
| Grand Total | 16713 | 997 | 25507 | 127 | 43344 |

| Category | Approved | Canceled | Refused | Unused offer |
|---|---|---|---|---|
| Wedding / gift /... | 248 | 36 | | |
| Urgent needs | 2228 | 83 | 2998 | |
| Repairs | 5385 | 381 | 8973 | 28 |
| Refusal to name the... | 7 | | | |
| Purchase of... | 357 | 283 | | |
| Payments on other... | 186 | 973 | 8 | |
| Other | 4106 | 186 | 5310 | 62 |
| Money for a third... | 10 | | | |
| Medicine | 676 | 25 | 696 | 5 |
| Journey | 329 | 1042 | | |
| Hobby | 20 | | | |
| Gasification / water... | 725 | | | |
| Furniture | 241 | 250 | | |
| Everyday expenses | 732 | 8 | 740 | 7 |
| Education | 481 | 476 | | |
| Car repairs | 225 | 256 | | |
| Buying a used car | 552 | 57 | 1166 | 9 |
| Buying a new car | 1394 | 654 | | |
| Buying a home | 1036 | 993 | | |
| Buying a holiday... | 92 | 80 | | |
| Buying a garage | 29 | | | |
| Business development | 784 | | | |
| Building a house or... | 4360 | 1188 | | |

Legend: Approved, Canceled, Refused, Unused offer