[**Bank Loan Case Study**](https://trainity.link/data/project06)

MUSKAN BANSAL

**INTRODUCTION:**

This project involves applying EDA (Exploratory Data Analysis) in a real business scenario.The goal of this project is to develop basic understanding about risk management in banks by ensuring maximum profit.

Banks receive a lot of loan applications and it is a big task to analyse everything and understand if loan can be provided or not. Since if the applicant is likely to repay the loan then in that case, not giving loan can result in business loss to the company and if the applicant is unable to repay the loan then in that case, approaving a loan can lead to financial loss to the company.

Aim of this project is to understand all the patterns present in the dataset provided, apply EDA and make a necessary report explaining who are likely to be the defaulters and who can be provided with the loans, ensuring to not reject the applicants who are capable of repaying the loans and not providing loans who are likely to default.

## **BANK LOAN CASE STUDY**

Study aims to identify patterns which indicates if the client has problems paying installments that may be used for taking required actions like denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. Ensuring that consumers capable of repaying the loan aren’t rejected. Basically, company wants to understand driving factors behind loan defaults, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

To understand driving factors behind loan defaulters.

* I, as a Data Analyst have to provide with the problem statement and analyse the approach briefly.
* Identify the missing patterns and appropriate methods to deal with them by either removing or deleting the values.
* I also identified all the outliers present in the dataset and for any data imbalance.
* Then explained results acquired using univariate, segmented univariate and bivariate analysis.
* And finally the top 10 correlations for the client with payment difficulties.

DATASET EXPLAINED:  
The dataset given has information about the loan application at the time of applying for the loan and it has two types of scenarios:

The client with payment difficulties and, client with payment done on time.

It is necessary to note that when a client applies for the loan, there are usually 4 types of decisions that could be taken by the client or by the company:

APPROVED: The company approved the loan.

CANCELLED: Client cancelled the application during approval.

REFUSED: The company rejected the loan.

UNUSED OFFER: Loan was cancelled by the client.

We have 3 csv files provided:

1. application\_data.csv: Contains all information about the client at the time of application. It contains whether the client has payment difficulties or not.

2. previous\_application.csv: Contains information about the client’s previous loan data. It contains whether the previous application was approved, cancelled, refused or unused.

3. columns\_description.csv: It describes all the columns present in the above two csv files.

**TECH\_STACK USED:** Ms Office Home and Student 2019

## **Data Cleaning**

Application Data and previous application data

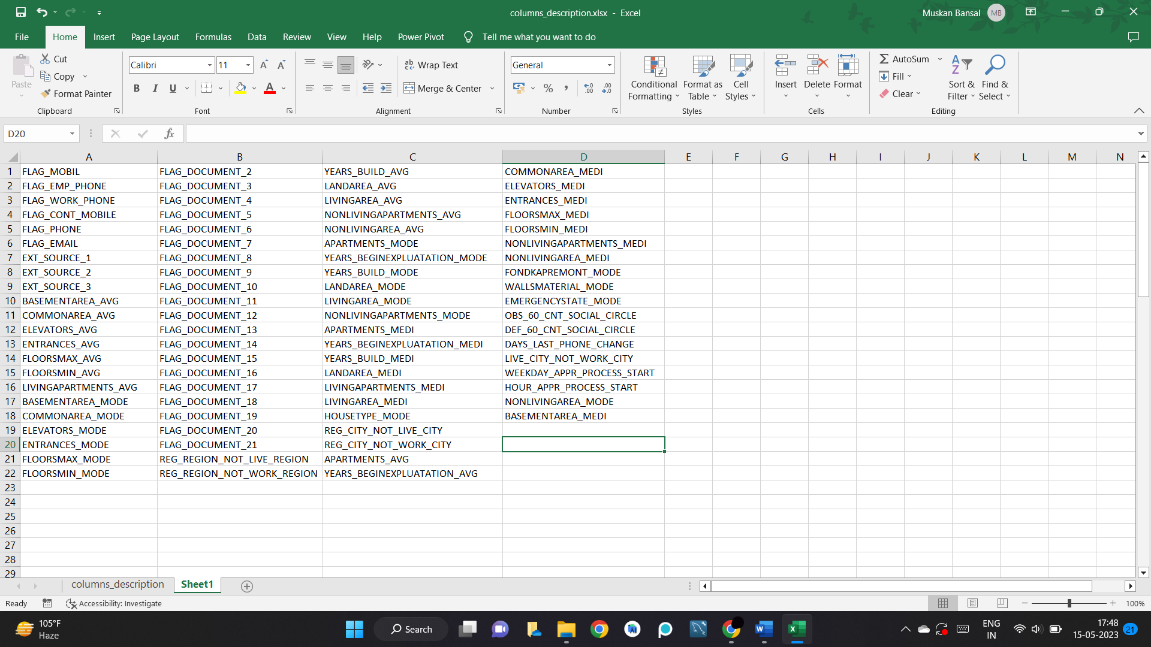
* Checking columns for null values and removng those with null values greater than 40% for application data and removing columns having null values greater than 30% for previous application data and the columns which are not necessary for analysis from both the datasets.
* Imputing null values with suitable values (mean or median).
* Binning certain columns into categorical columns which will be used for analysis.

(Binning are of three types, mean, median and boundaries which is a data cleaning method.)

* Checking datatype of variables and converting them into suitable datatypes.
* Checking dataset for duplicates and outliers.

**REMOVING UNWANTED COLUMNS:**

I deleted all these names presented below in the excel screenshot present in the Application\_Data as they were useless for the analysis purposes.

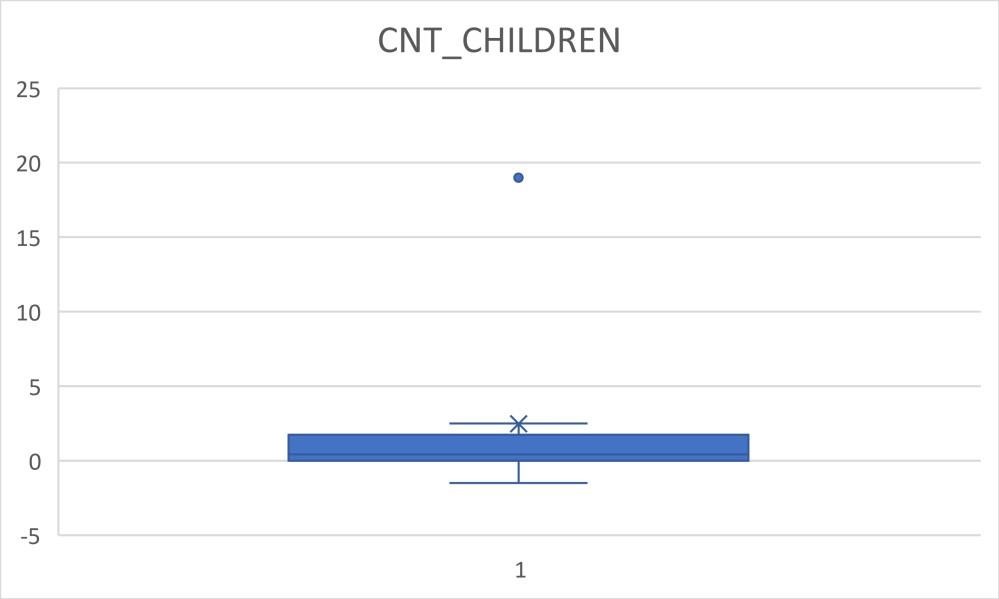
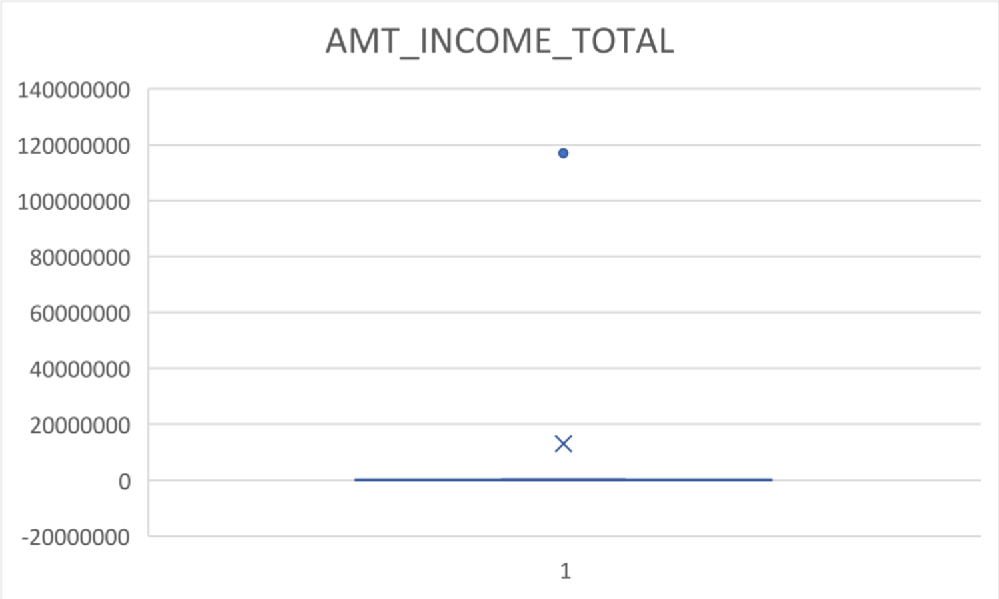


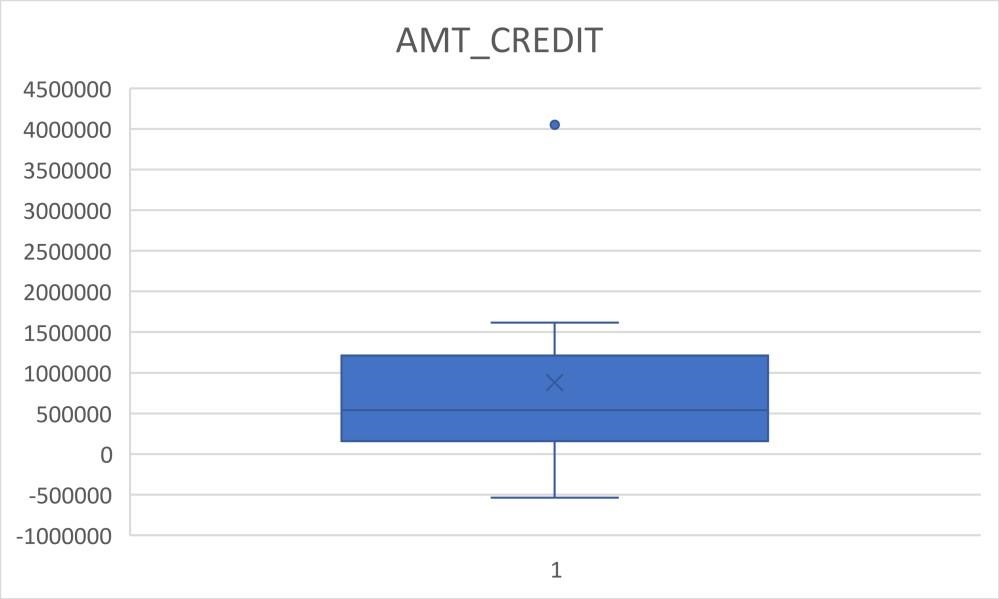
## **Identification Of Outliers:**

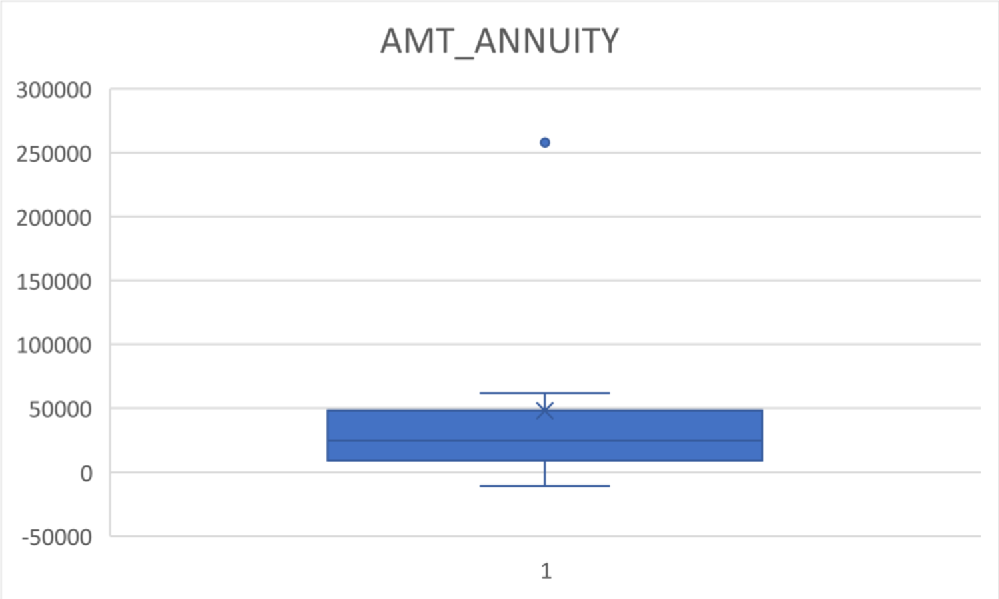
**SEARCHING FOR OUTLIERS USING IQR:**

* By using both IQR method and by visualizing boxplots we can see that the columns CNT\_CHILDREN, AMT\_CREDIT, AMT\_INCOME\_TOTAL, AMT\_ANNULTY, EMPLOYED\_YEARS, AMT\_GOODS\_PRICE have outliers.
* We can see that the maximum outlier value for the column CNT\_CHILDREN is 19 which is highly unlikely.

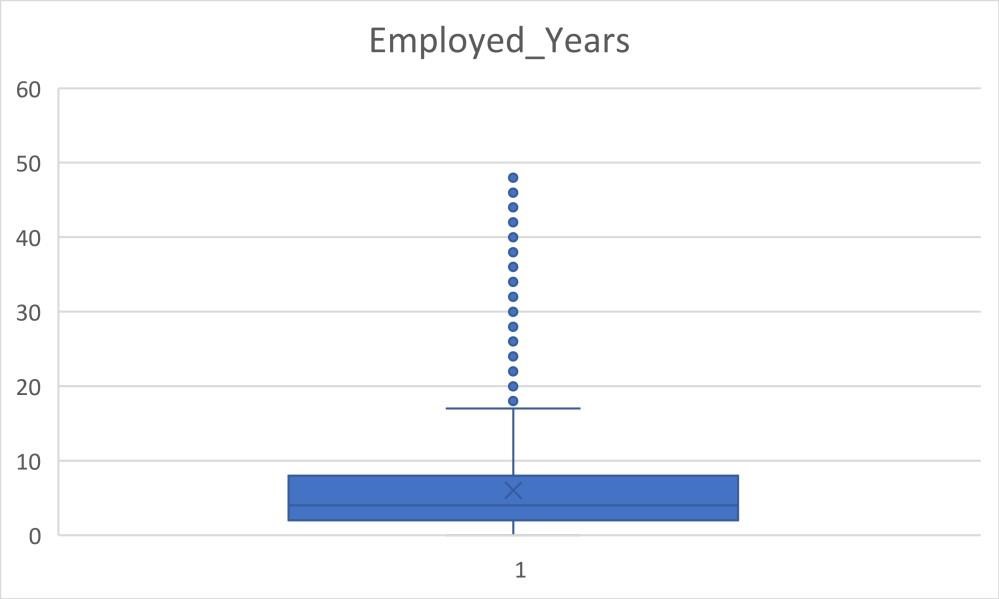
For Application\_data:



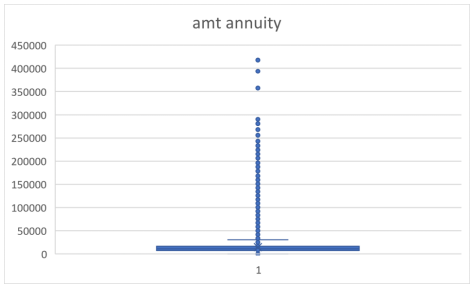


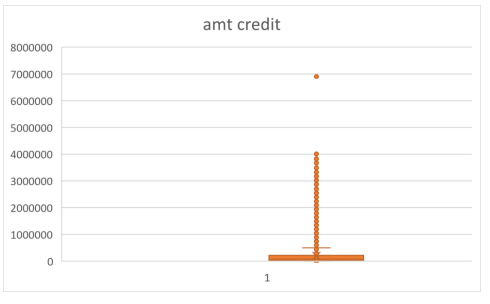






For Previous\_Data:





## **Data Imbalance:**

## So, after analysis I observed that there is a huge data imbalance as explained below:

1. The data has huge imbalance with regards to the TARGET variable. We have extensive data for applicants who are likely to repay the loan but relatively very less data for applicants likely to have difficulty in repaying. 1 stands for the people who had difficulty paying loan.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **TARGET** |
| 0 | 282686 |
| 1 | 24825 |
| **Grand Total** | **307511** |

282686

24825

0

50000

100000

150000

200000

250000

300000

0

1

Target

Total

1. There is also data imbalance in the NAME\_CONTRACT\_TYPE variable. There is more data for Cash loans than Revolving loans.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **NAME\_CONTRACT\_TYPE** |

Cash loans 278232

Revolving

### loans 29279

|  |  |
| --- | --- |
| **Grand Total 307511** |  |
| Contract\_Type  0  50000  100000  150000  200000  250000  300000  Cash loans  Revolving loans  Total | |

3. In Code\_Gender, you can see there is huge amount of female data but not for male. And from the table it is clearly visible that twice number of female took loan than male population.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **CODE\_GENDER** |
| F | 202448 |
| M | 105059 |
| XNA | 4 |
| **Grand Total** | **307511** |

0

50000

100000

150000

200000

250000

F

M

XNA

**Count of CODE\_GENDER**

Total

4. In NAME\_HOUSING\_TYPE most of the applicants live in House/Apartments.

|  |  |  |  |
| --- | --- | --- | --- |
| **Row Labels** | | **Count of**  **NAME\_HOUSING\_TYPE** | |
| Co-op apartment | 1122 |
| House / apartment Municipal | 272868 |
| apartment | 11183 |
| Office apartment | 2617 |
| Rented apartment | 4881 |
| With parents | 14840 |
| **Grand Total 307511** | | | |  |

|  |
| --- |
| Count of NAME\_HOUSING\_TYPE  0  50000  100000  150000  200000  250000  300000  Total  Co-op House / Municipal Office Rented With  apartment apartment apartment apartment apartment parents |

5. More than double the amount of people own realty.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **FLAG\_OWN\_REALTY** |

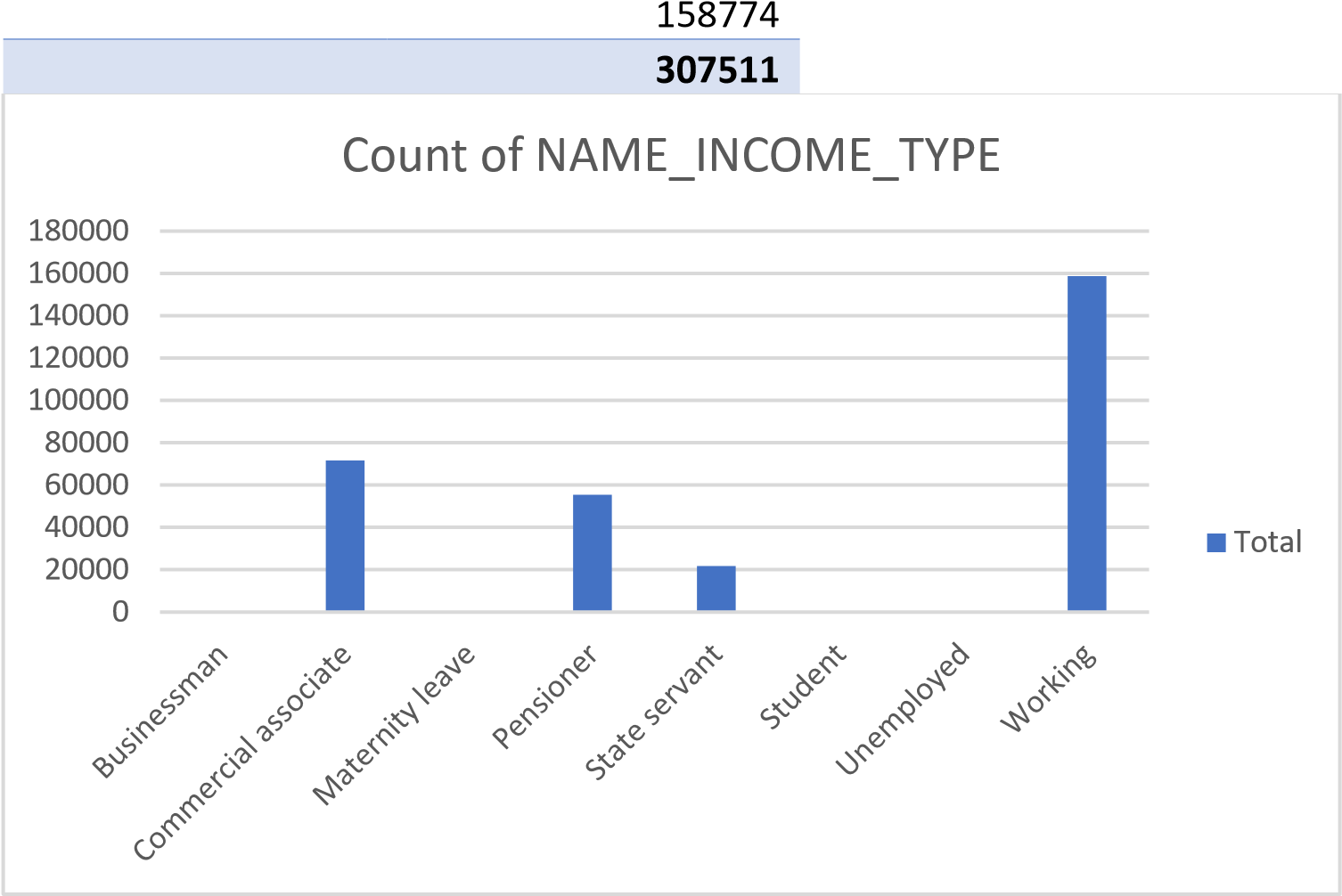
N 94199

## Y 213312

|  |  |  |
| --- | --- | --- |
| **Grand Total 307511** |  |  |
| Count of FLAG\_OWN\_REALTY  0  50000  100000  150000  200000  250000  N Y | | Total |

6. Most of the people work somewhere.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | | **Count of**  **NAME\_INCOME\_TYPE** | | |
| Businessman Commercial | | 10 |
| associate | | 71617 |
| Maternity leave | | 5 |
| Pensioner | | 55362 |
| State servant | | 21703 |
| Student | | 18 |
| Unemployed | | 22 |
| Working | | 158774 |
| **Grand Total** | | 307511 |



7. Majority of the people are students that could be because they arent earning or if they are then its not much for sustaining and certainly they took loans for their education only.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **NAME\_EDUCATION\_TYPE** |
| Academic degree | 164 |
| Higher education | 74863 |
| Incomplete higher | 10277 |
| Lower secondary | 3816 |
| Secondary/secondary special | 218391 |
| **Grand Total** | **307511** |

0

50000

100000

150000

200000

250000

Academic

degree

Higher

education

Incomplete

higher

Lower

secondary

Secondary /

secondary

special

Count of NAME\_EDUCATION\_TYPE

Total

8. Majority of people who took loans dont have kids that could also be related to the factor since majority of people are still studying.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **CNT\_CHILDREN** |
| 0 | 215371 |
| 1 | 61119 |
| 2 | 26749 |
| 3 | 3717 |
| 4 | 429 |
| 5 | 84 |
| 6 | 21 |
| 7 | 7 |
| 8 | 2 |
| 9 | 2 |
| 10 | 2 |
| 11 | 1 |
| 12 | 2 |
| 14 | 3 |
| 19 | 2 |
| **Grand Total** | **307511** |

0

50000

100000

150000

200000

250000

0

1

2

3

4

5

6

7

8

9

10

11

12

14

19

Count of CNT\_CHILDREN

Total

9. Majority of the people are married that is visible from the bar graph below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | |  | **Count of**  **NAME\_FAMILY\_STATUS** | |
| Civil marriage | | 29775 |
| Married | | 196432 |
| Separated Single / not | | 19770 |
| married | | 45444 |
| Unknown | | 2 |
| Widow | | 16088 |
| **Grand Total 307511** | | | | |  |
| Count of NAME\_FAMILY\_STATUS  250000  0  50000  100000  150000  200000  Civil  Married  Separated  Single / not  Unknown  Widow  Total  marriage married | | | | | |

10. Most of the people have only 2 members in their family, followed by 1, 3 and 4.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **CNT\_FAM\_MEMBERS** |

1. 67847
2. 158357
3. 52601
4. 24697
5. 3478
6. 408
7. 81
8. 20
9. 6
10. 3
11. 1
12. 2
13. 1
14. 2
15. 1
16. 2

20 2

### (blank) 2

|  |  |
| --- | --- |
| **Grand Total 307511** |  |
| Count of CNT\_FAM\_MEMBERS  0  20000  40000  60000  80000  100000  120000  140000  160000  180000  1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  20  )  (  blank  Total | |

11. Most of the people live in the city.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **REGION\_RATING\_CLIENT\_W\_CITY** |

1. 34167
2. 229484

#### 3 43860

|  |  |  |
| --- | --- | --- |
| **Grand Total 307511** |  |  |
| Count of REGION\_RATING\_CLIENT\_W\_CITY  0  50000  100000  150000  200000  250000  1  2  3 | | Total |

12. Most of the people own a car.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **FLAG\_OWN\_CAR** |

N 202924

## Y 104587

|  |  |
| --- | --- |
| **Grand Total 307511** |  |
| Count of FLAG\_OWN\_CAR  250000  0  50000  100000  150000  200000  Total  N Y | |

**FOR PREVIOUS\_DATA.CSV FILE:**

13. Most of the data is for Cash loans and consumer loans. There are few records for revolving loans.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | | **Count of**  **NAME\_CONTRACT\_TYPE** | | |
| Cash loans Consumer | | 465861 |
| loans Revolving | | 461049 |

loans

121441

XNA

224

**Grand Total**

**1048575**

0

50000

100000

150000

200000

250000

300000

350000

400000

450000

500000

Cash loans

Consumer loans

Revolving loans

XNA

Contract\_Type

Total

14. Mostly, laons are approved while few are cancelled and unused. And very few are unused.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | | **Count of**  **NAME\_CONTRACT\_STATUS** | | |
| Approved | | 652486 |
| Canceled | | 197231 |
| Refused  Unused | | 182083 |
| offer | | 16775 |
| **Grand Total 1048575** | | | | |  |
| Contract\_Status  0  100000  200000  300000  400000  500000  600000  700000  Approved  Canceled  Refused | | | | | Total  Unused offer |

15. There are large number of data for payment type cash through banks than other payment modes.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **NAME\_PAYMENT\_TYPE** |
| Cash through the bank | 650604 |
| Cashless from the account of the employer | 677 |
| Non-cash from your account | 5233 |
| XNA | 392061 |
| **Grand Total** | **1048575** |

0

100000

200000

300000

400000

500000

600000

700000

Cash through the

bank

Cashless from the

account of the

employer

Non-cash from

your account

XNA

Payment\_Type

Total

16. Loans that are rejected, mostly due to XAP.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | | **Count of**  **CODE\_REJECT\_REASON** | | |
| CLIENT | | 16775 |
| HC | | 109441 |
| LIMIT | | 35248 |
| SCO | | 23472 |
| SCOFR | | 7968 |
| SYSTEM | | 465 |
| VERIF | | 2192 |
| XAP | | 849712 |
| XNA | | 3302 |
| **Grand Total 1048575** | | | | |  |
| Reject\_Reason  0  100000  200000  300000  400000  500000  600000  700000  800000  900000  CLIENT  HC  LIMIT  SCO  SCOFR  SYSTEM  VERIF  XAP  XNA  Total | | | | | |

17. There are a large number of repeater applicants than refreshed and new applicants.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **NAME\_CLIENT\_TYPE** |

New 189990

Refreshed 85411

Repeater 771957

## XNA 1217

|  |  |
| --- | --- |
| **Grand Total 1048575** |  |
| Client\_Type  0  100000  200000  300000  400000  500000  600000  700000  800000  900000  New  Refreshed  Repeater | Total  ater XNA |

18. There are large number of data for POS and almost no data for cars.

|  |  |
| --- | --- |
| **Row Labels** | **Count of**  **NAME\_PORTFOLIO** |
| Cards | 90935 |
| Cars | 279 |
| Cash | 287509 |
| POS | 436846 |
| XNA | 233006 |
| **Grand Total** | **1048575** |

0

50000

100000

150000

200000

250000

300000

350000

400000

450000

500000

Cards

Cars

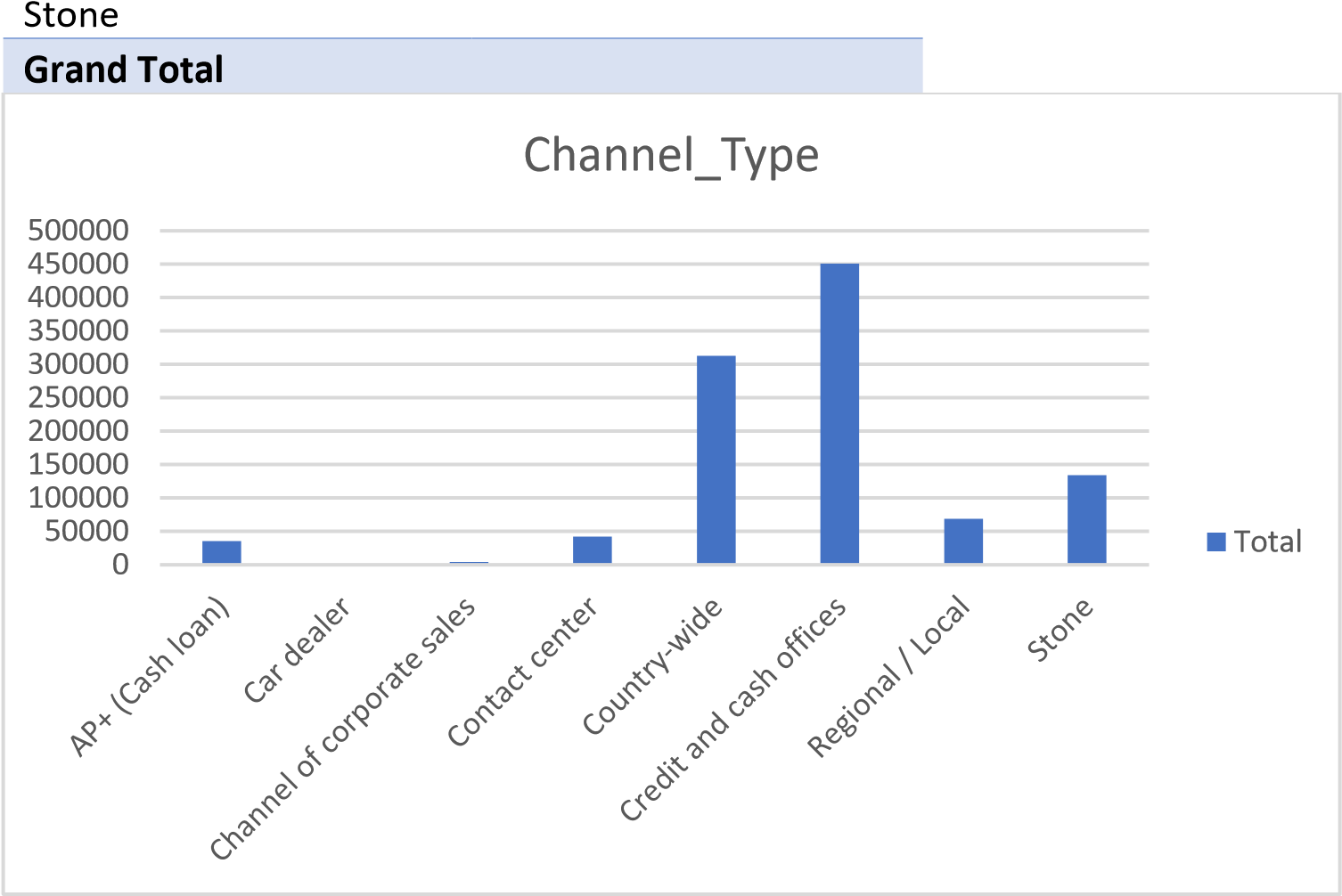
Cash

POS

XNA

Portfolio\_Name

Total

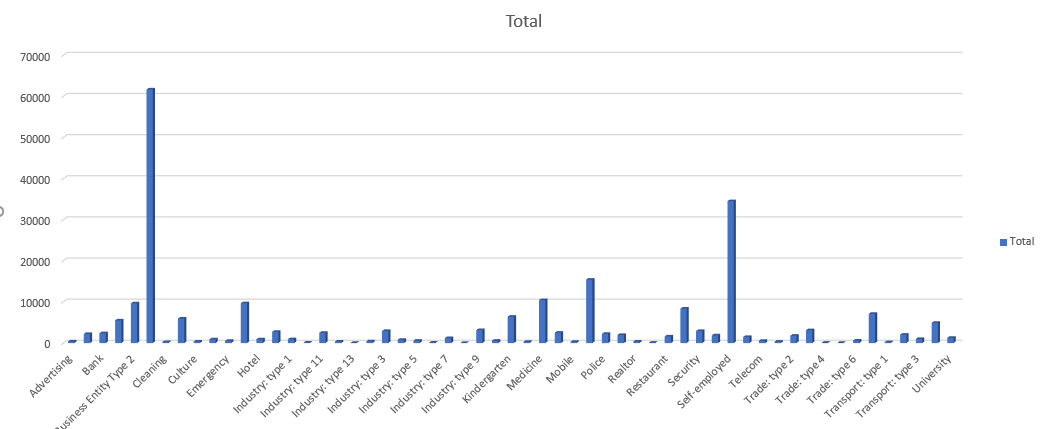
 19. There are large number of data for credit and cash offices, followed by country-wide and stone than other channel type.

|  |  |  |  |
| --- | --- | --- | --- |
| **Row Labels** | | **Count of**  **CHANNEL\_TYPE** | |
| AP+ (Cash loan) | 35806 |
| Car dealer  Channel of corporate | 295 |
| sales | 3958 |
| Contact center | 41962 |
| Country-wide | 312690 |
| Credit and cash offices | 451047 |
|  |  |
| Regional / Local | 68621 |

20. Name\_yield\_group signifies Grouped interest rate into small medium and high of the previous application, While mostly falls in the XNA category, followed by middle, high, normal respectfully.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | | **Count of**  **NAME\_YIELD\_GROUP** | | |
| high | | 221900 |
| low\_action | | 58200 |
| low\_normal | | 202206 |
| middle | | 242328 |
| XNA | | 323941 |
| **Grand Total 1048575** | | | | |  |
| Yield\_Group  0  50000  100000  150000  200000  250000  300000  350000  high  low\_action  low\_normal  middle  XNA  Total | | | | | |

21. Mostly people are in the Business Type 2 domain followed by Self employed, mobile and medicine.

****

### **Univariate Analysis On Categorical Columns :**

### 1. The number of people who cleared the loan are both more in CASH LOANS as well as the REVOLVING LOANS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of**  **TARGET** | **Column Labels** |  |  |
| **Row Labels** | **0** | **1** | **Grand**  **Total** |
| Cash loans | 255011 | 23221 | 278232 |
| Revolving loans | 27675 | 1604 | 29279 |
| **Grand Total** | **282686** | **24825** | **307511** |



2. The number of female applicants is greater than male applicants in both defaulters and non defaulters list. Here, 1 stands for the defaulters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of TARGET** | **Column Labels** |  |  |  |
| **Row Labels** | **0** | **1** | **(blank)** | **Grand Total** |
| F | 145256 | 11921 |  | 157177 |
| M | 85042 | 9914 |  | 94956 |
| XNA | 4 |  |  | 4 |
| **Grand Total** | **230302** | **21835** |  | **252137** |

3. The ratio of non-defaulters owning a realty is greater than that of defaulters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of**  **TARGET** | **Column Labels** |  |  |  |
| **Row Labels** | **0** |  | **1** | **Grand**  **Total** |

N 86357 7842 94199

## Y 196329 16983 213312

|  |  |  |  |
| --- | --- | --- | --- |
| **Grand Total** | **282686** | **24825** | **307511** |

4. Ratio of non-defaulters owning a car is greater than those of non-defaulters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of**  **TARGET** | **Column Labels** |  |  |  |
| **Row Labels** | **0** |  | **1** | **Grand**  **Total** |

N 185675 17249 202924

## Y 97011 7576 104587

|  |  |  |  |
| --- | --- | --- | --- |
| **Grand Total** | **282686** | **24825** | **307511** |

5. Both cases seem to follow the same pattern where most applicants who are working have higher count in both defaulters and repayers list.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of TARGET** | **Column Labels** |  |  |
| **Row Labels** | **0** | **1** | **Grand**  **Total** |
| Businessman | 10 |  | 10 |
| Commercial associate | 66257 | 5360 | 71617 |
| Maternity leave | 3 | 2 | 5 |
| Pensioner | 52380 | 2982 | 55362 |
| State servant | 20454 | 1249 | 21703 |
| Student | 18 |  | 18 |
| Unemployed | 14 | 8 | 22 |
| Working | 143550 | 15224 | 158774 |
| **Grand Total** | **282686** | **24825** | **307511** |

6. People with education mostly payed the loans.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of TARGET** | **Column Labels** |  |  |
| **Row Labels** | **0** | **1** | **Grand**  **Total** |
| Academic degree | 161 | 3 | 164 |
| Higher education | 70854 | 4009 | 74863 |
| Incomplete higher | 9405 | 872 | 10277 |
| Lower secondary | 3399 | 417 | 3816 |
| Secondary / secondary special | 198867 | 19524 | 218391 |
| **Grand Total** | **282686** | **24825** | **307511** |

7. Both cases seem to follow the same pattern where most applicants are married.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of TARGET** | **Column Labels** |  |  |
| **Row Labels** | **0** | **1** | **Grand**  **Total** |
| Civil marriage | 26814 | 2961 | 29775 |
| Married | 181582 | 14850 | 196432 |
| Separated | 18150 | 1620 | 19770 |
| Single / not married | 40987 | 4457 | 45444 |
| Unknown | 2 |  | 2 |
| Widow | 15151 | 937 | 16088 |
| **Grand Total** | **282686** | **24825** | **307511** |

8. Both cases seem to follow the same pattern where most applicants no childern seem to repay as well as default, followed by having one child.

9. Both cases seem to follow the same pattern where most applicants have a house or an apartment are likely to the loan as well as default. However mostly repay the loans.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Count of TARGET** | | **Column Labels** |  |  |  |  | |
| **Row Labels** | | **0** |  |  | **1** | **Grand**  **Total** | |
| Co-op apartment | 1033 | 89 | | | 1122 |
| House / apartment Municipal | 251596 | 21272 | | | 272868 |
| apartment | 10228 | 955 | | | 11183 |

Office apartment 2445 172 2617

10. For the applicants who have mentioned their occupation, we find that Laborers are most likely to default followed by Sales staff.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Count of TARGET** | | **Column Labels** | |  |  |  | | |
| **Row Labels** | | **0** | |  | **1** | **Grand**  **Total** | | |
| Accountants | 9339 | | 474 | | 9813 | |
| Cleaning staff | 4206 | | 447 | | 4653 | |
| Cooking staff | 5325 | | 621 | | 5946 | |
| Core staff | 25832 | | 1738 | | 27570 | |
| Drivers | 16496 | | 2107 | | 18603 | |
| High skill tech staff | 10679 | | 701 | | 11380 | |
| HR staff | | 527 | 36 | | | 563 |
| IT staff | | 492 | 34 | | | 526 |
| Laborers | | 49348 | 5838 | | | 55186 |
| Low-skill Laborers | | 1734 | 359 | | | 2093 |
| Managers | | 20043 | 1328 | | | 21371 |
| Medicine staff | | 7965 | 572 | | | 8537 |
| Private service staff | | 2477 | 175 | | | 2652 |
| Realty agents | | 692 | 59 | | | 751 |
| Sales staff | | 29010 | 3092 | | | 32102 |
| Secretaries | | 1213 | 92 | | | 1305 |
| Security staff  Waiters/barmen | | 5999 | 722 | | | 6721 |
| staff | | 1196 | 152 | | | 1348 |
| (blank) | | 90113 | 6278 | | | 96391 |
| **Grand Total 282686 24825 307511** | | | | | | | | |  |

11. Both cases seem to follow the same pattern where most applicants with region ratings 1 seem to repay as well as default followed by reagion ratings as 3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Count of**  **TARGET** | **Column Labels** |  |  |  |  |
| **Row Labels** | **0** |  |  | **1** | **Grand**  **Total** |

1. 32513 1654 34167
2. 211314 18170 229484

### 3 38859 5001 43860

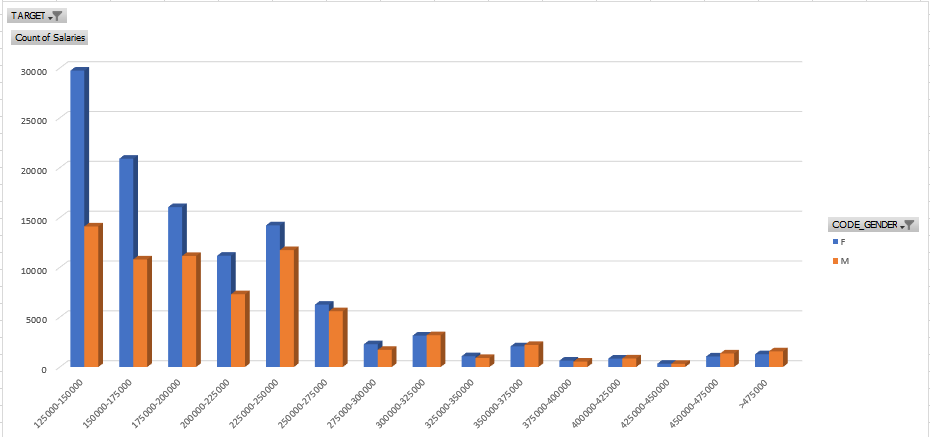
|  |  |  |  |
| --- | --- | --- | --- |
| **Grand Total** | **282686** | **24825** | **307511** |

12. Most of the people who repayed the loan as well as to default live in flats or ‘Not Mentioned’.

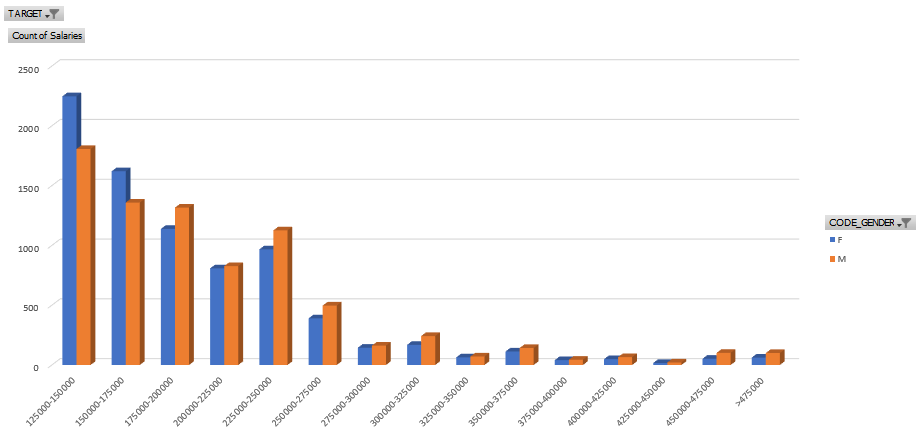
|  |  |  |  |
| --- | --- | --- | --- |
| **Count of TARGET** | **Column Labels** |  |  |
| **Row Labels** | **0** | **1** | **Grand Total** |
| block of flats | 140053 | 10450 | 150503 |
| specific housing | 1347 | 152 | 1499 |
| terraced house | 1109 | 103 | 1212 |
| (blank) | 140177 | 14120 | 154297 |
| **Grand Total** | **282686** | **24825** | **307511** |

13. Applicants who have less work experience in the range 0-5 years default more than applicants with more experience.

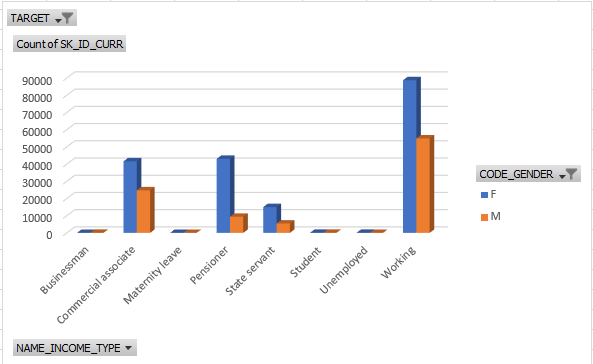
14. Mostly people who repayed the loan are in the range of 26-44, after which there is a slight dip and people who most likey to default were mostly 27.

15. This graph represents salary of people who repayed the loan, we can analyse that mostly people earm in the range of 125k-150k while least amount of people earn in the range of 425k-450k. It is also observed that the number of female is mostly greater than males.****

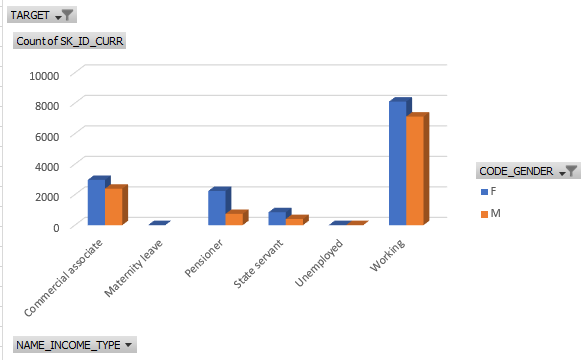
16. This graph represents salary of people who had trouble repaying the loan, we can analyse that here mostly people earm in the range of 125k-150k while least amount of people earn in the range of 425k-450k. It is also observed that the number of female is mostly greater than males.

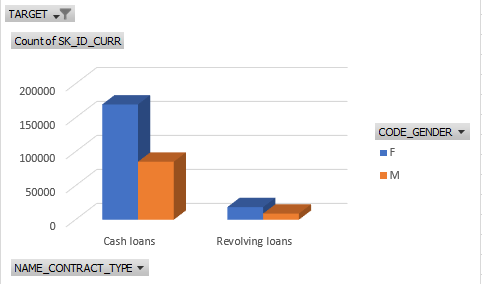
****

17. This graph represents the income type of people who repayed the loan on time. Mostly people are in the working category, followed by commercial associate and pensioner. The ratio of female category is much higher than that of the male catogery.

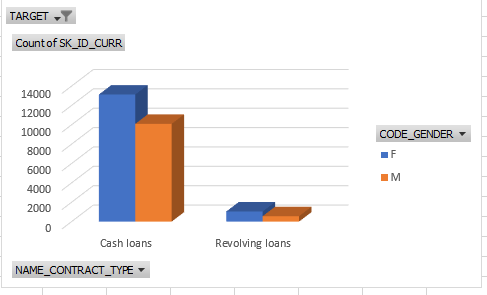
****

* 18. This graph represents the income type of people who had trouble repaying the loan on time. Mostly people are in the working category, followed by commercial associate, pensioner and State servant. The ratio of female category is much higher than that of the male catogery. Males have not been mentioned in the maternity leave category. Since there is no income type for ‘student’, “unemployed” and ‘Businessman’ , this suggests that they are not in the defaulter category.

****

19. For people who paid the loan on time, cash loans were taken in higher amount than that of revolving loans, also the ratio of females is much higher.****

20. For people who had trouble repaying the loan on time, cash loans were taken in higher amount than that of revolving loans, also the ratio of females to male is less here.

****

21. Organization entity type 3 is most likely to default followed by self-employed.

0

1000

2000

3000

4000

5000

6000

7000

Advertising

Business Entity Type 1

Cleaning

Electricity

Hotel

Industry: type 10

Industry: type 13

Industry: type 4

Industry: type 7

Insurance

Medicine

Other

Realtor

School

Self-employed

Trade: type 1

Trade: type 4

Trade: type 7

Transport: type 3

XNA

Organisation\_Type

1

22. With higher amount of annuity, more defaults are possible.

23. Defaulters decreases with increasing amount of the credit score.

24. We find that applicants in the Medium income group are the most to default than the applicants with very low or very high income.

24. Majority of the previous loans are either cash loans or consumer loans. And majority of the loan applications were mostly approved. Out of which consumer loan type, is the most approved one followed by the cash loan type. And, mostly cash loans were canceled or refused.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Count of NAME\_CONTRACT\_TYPE** | **Column Labels** |  |  |  |  |
| **Row Labels** | **Approved** | **Canceled** | **Refused** | **Unused offer** | **Grand Total** |
| Cash loans | 195339 | 167015 | 103197 | 310 | 465861 |
| Consumer loans | 395902 | 1009 | 47677 | 16461 | 461049 |
| Revolving loans | 61245 | 29002 | 31190 | 4 | 121441 |
| XNA |  | 205 | 19 |  | 224 |
| **Grand Total** | **652486** | **197231** | **182083** | **16775** | **1048575** |

**Bivariate Analysis:**

As we saw before as the price of the goods increases, the loan amount increases which results in increased annuity. Also from the plot we can see that AMT\_GOODS\_PRICE and AMT\_ANNUITY are correlated.

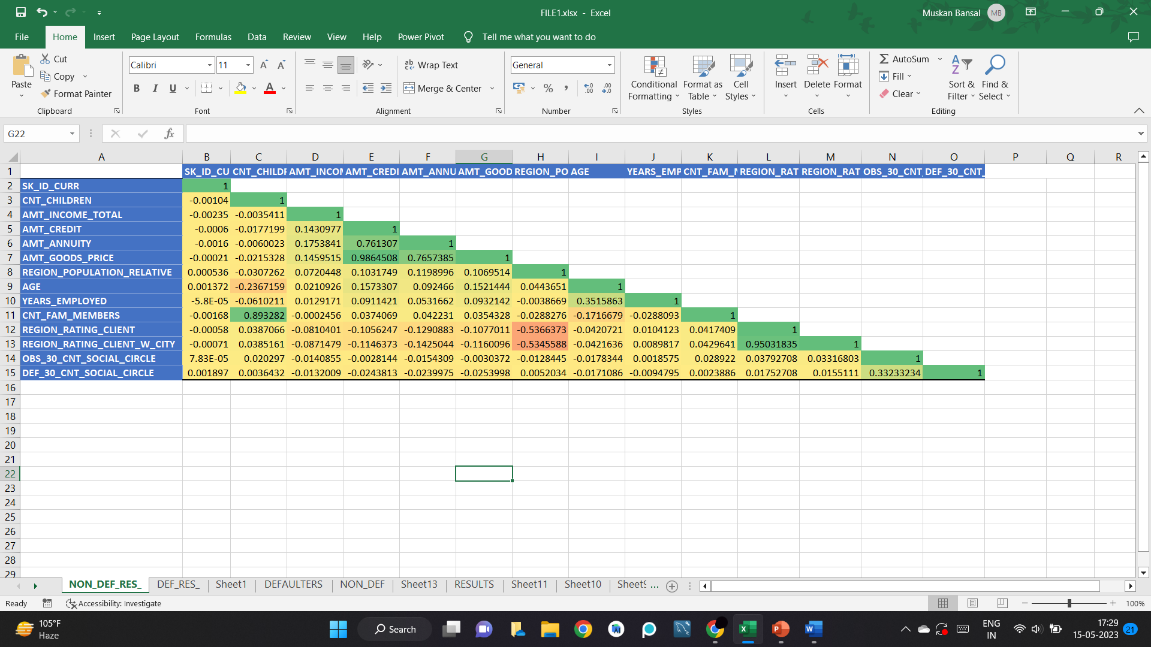
From the plot we can see that AMT\_CREDIT and AMT\_INCOME\_TOTAL are not correlated.

For Non-Defaulters:

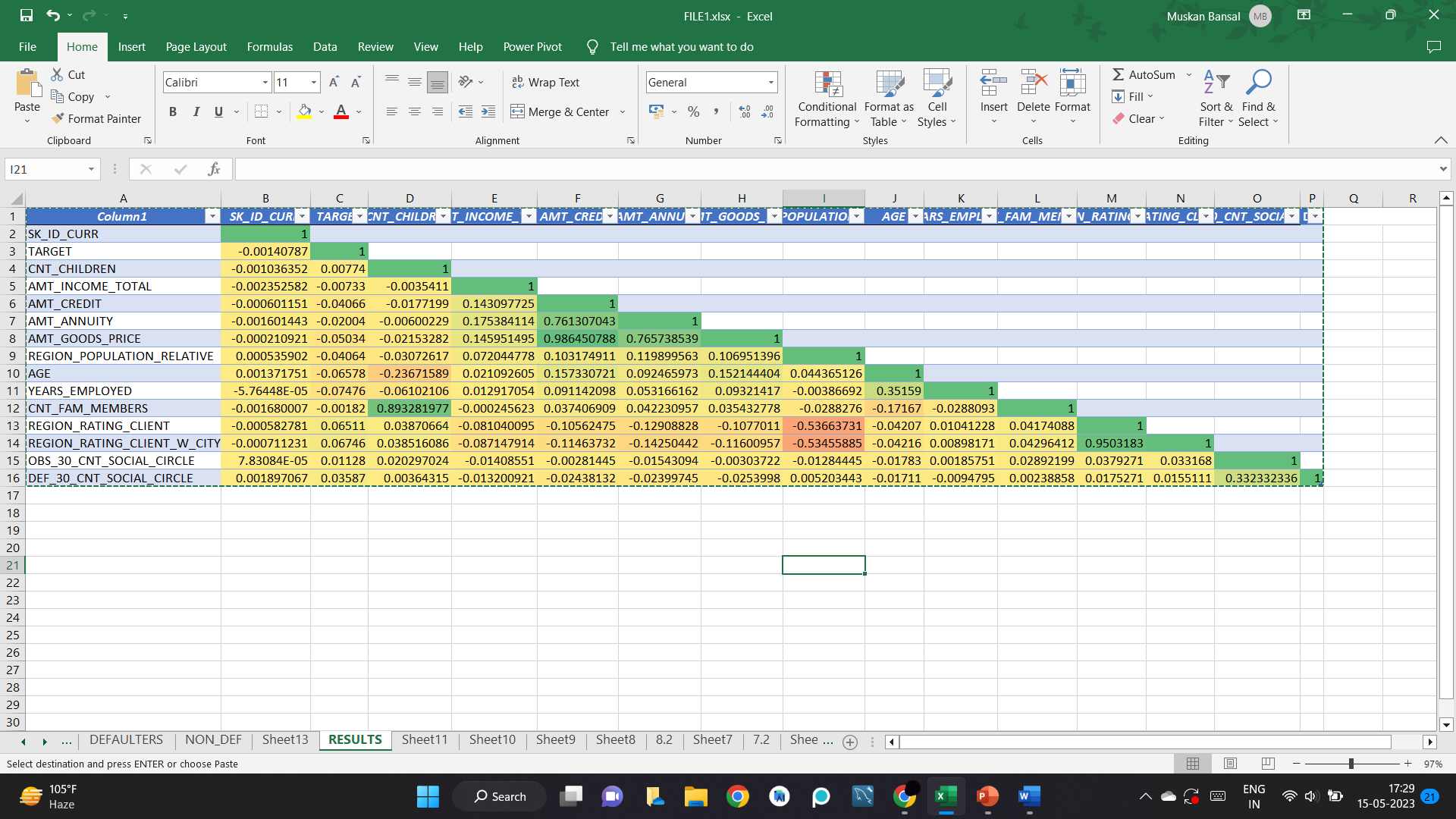
Number of children and number of family members are positively proportional, that means, with increase in the number of children, number of family count also increases.

AMT\_CREDIT and AMT\_ANNUITY are positively proportional that means, with increase in the annuity chances of greater credit score will happen.

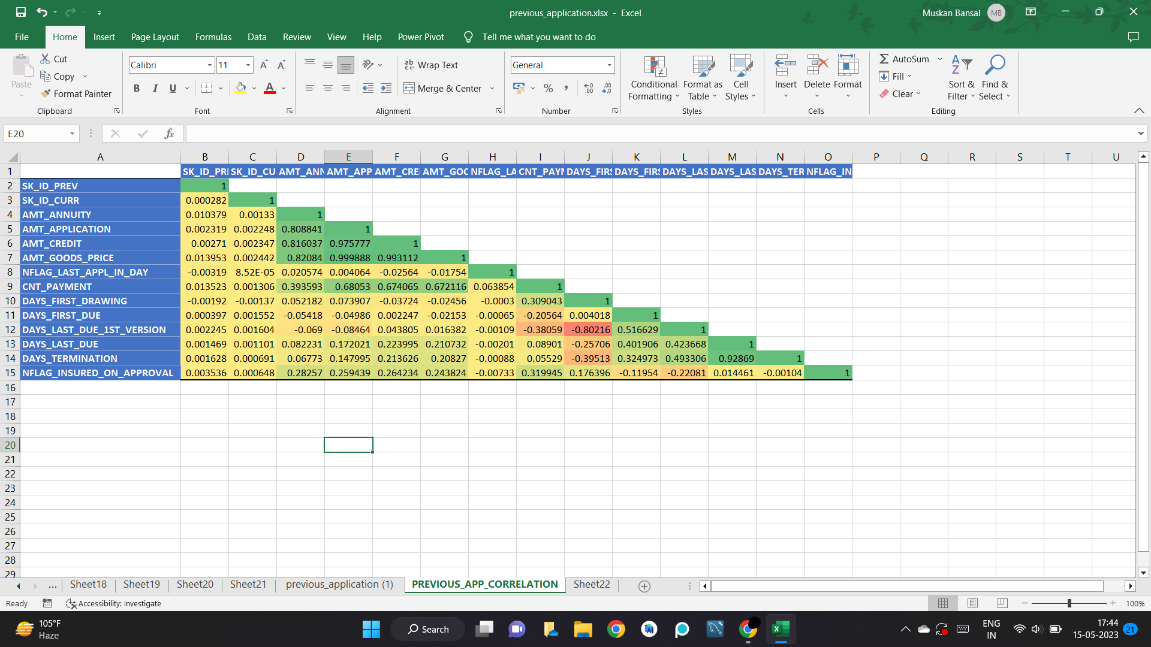
AMT\_CREDIT and AMT\_GOODS\_PRICE is also positively proportional that means, with increase in the price of the goods, the credit also increases.



For Defaulters:



From the PREVIOUS\_DATA, a lot of positive and negative corelations can be formed, from the below analysis you can see the following:



**Results:-**

Majority of the people who have applied for the loans are females and mostly have opted for cash loans, while male applicants with higher income have defaulted more. However, applicants belonging to higher income group were able to repay loans on time.

Banks should focus more on students, pensioners and businessmen and with housing type as co-op apartment since they repay the loans. It is also observed that people with ‘working’ income type have the most difficulty in repaying the loans so they can be avoided.

Applicants with greater work ex are credited higher loans.

Higher the price of the goods, higher will be the loan amount. Hence, AMT\_CREDIT and AMT\_GOODS\_PRICE are positively correlated.

With higher loan amount credited, higher will be the annuity. Therefore, AMT\_CREDIT and AMT\_ANNUITY are positively correlated.