CREDIT RISK ANALYSIS

```
import numpy as np

import pandas as pd

df1=pd.read_csv(r"C:\Users\radhi\OneDrive\Desktop\PythonProject\application_data.csv")

df1.describe()

<div>

<style scoped>

    .dataframe tbody tr th:only-of-type {

        vertical-align: middle;

    }


    .dataframe tbody tr th {

        vertical-align: top;

    }


    .dataframe thead th {

        text-align: right;

    }

</style>

<table border="1" class="dataframe">

  <thead>

    <tr style="text-align: right;">

      <th></th>

      <th>SK_ID_CURR</th>

      <th>TARGET</th>

      <th>CNT_CHILDREN</th>

      <th>AMT_INCOME_TOTAL</th>

      <th>AMT_CREDIT</th>

      <th>AMT_ANNUITY</th>

      <th>AMT_GOODS_PRICE</th>
```

```
      <th>REGION_POPULATION_RELATIVE</th>
      <th>DAYS_BIRTH</th>
      <th>DAYS_EMPLOYED</th>
      <th>...</th>
      <th>FLAG_DOCUMENT_18</th>
      <th>FLAG_DOCUMENT_19</th>
      <th>FLAG_DOCUMENT_20</th>
      <th>FLAG_DOCUMENT_21</th>
      <th>AMT_REQ_CREDIT_BUREAU_HOUR</th>
      <th>AMT_REQ_CREDIT_BUREAU_DAY</th>
      <th>AMT_REQ_CREDIT_BUREAU_WEEK</th>
      <th>AMT_REQ_CREDIT_BUREAU_MON</th>
      <th>AMT_REQ_CREDIT_BUREAU_QRT</th>
      <th>AMT_REQ_CREDIT_BUREAU_YEAR</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <th>count</th>
      <td>307511.000000</td>
      <td>307511.000000</td>
      <td>307511.000000</td>
      <td>3.075110e+05</td>
      <td>3.075110e+05</td>
      <td>307499.000000</td>
      <td>3.072330e+05</td>
      <td>307511.000000</td>
      <td>307511.000000</td>
      <td>307511.000000</td>
      <td>...</td>
      <td>307511.000000</td>
```

<td>307511.000000</td>

<td>307511.000000</td>

<td>307511.000000</td>

<td>265992.000000</td>

<td>265992.000000</td>

<td>265992.000000</td>

<td>265992.000000</td>

<td>265992.000000</td>

<td>265992.000000</td>

</tr>

<tr>

<th>mean</th>

<td>278180.518577</td>

<td>0.080729</td>

<td>0.417052</td>

<td>1.687979e+05</td>

<td>5.990260e+05</td>

<td>27108.573909</td>

<td>5.383962e+05</td>

<td>0.020868</td>

<td>-16036.995067</td>

<td>63815.045904</td>

<td>...</td>

<td>0.008130</td>

<td>0.000595</td>

<td>0.000507</td>

<td>0.000335</td>

<td>0.006402</td>

<td>0.007000</td>

<td>0.034362</td>

<td>0.267395</td>

```html
  <td>0.265474</td>
  <td>1.899974</td>
</tr>
<tr>
  <th>std</th>
  <td>102790.175348</td>
  <td>0.272419</td>
  <td>0.722121</td>
  <td>2.371231e+05</td>
  <td>4.024908e+05</td>
  <td>14493.737315</td>
  <td>3.694465e+05</td>
  <td>0.013831</td>
  <td>4363.988632</td>
  <td>141275.766519</td>
  <td>...</td>
  <td>0.089798</td>
  <td>0.024387</td>
  <td>0.022518</td>
  <td>0.018299</td>
  <td>0.083849</td>
  <td>0.110757</td>
  <td>0.204685</td>
  <td>0.916002</td>
  <td>0.794056</td>
  <td>1.869295</td>
</tr>
<tr>
  <th>min</th>
  <td>100002.000000</td>
  <td>0.000000</td>
```

```
  <td>0.000000</td>
  <td>2.565000e+04</td>
  <td>4.500000e+04</td>
  <td>1615.500000</td>
  <td>4.050000e+04</td>
  <td>0.000290</td>
  <td>-25229.000000</td>
  <td>-17912.000000</td>
  <td>...</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
</tr>
<tr>
  <th>25%</th>
  <td>189145.500000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>1.125000e+05</td>
  <td>2.700000e+05</td>
  <td>16524.000000</td>
  <td>2.385000e+05</td>
  <td>0.010006</td>
  <td>-19682.000000</td>
```

```html
  <td>-2760.000000</td>
  <td>...</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
</tr>
<tr>
  <th>50%</th>
  <td>278202.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>1.471500e+05</td>
  <td>5.135310e+05</td>
  <td>24903.000000</td>
  <td>4.500000e+05</td>
  <td>0.018850</td>
  <td>-15750.000000</td>
  <td>-1213.000000</td>
  <td>...</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
  <td>0.000000</td>
```

```
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>1.000000</td>
  </tr>
  <tr>
    <th>75%</th>
    <td>367142.500000</td>
    <td>0.000000</td>
    <td>1.000000</td>
    <td>2.025000e+05</td>
    <td>8.086500e+05</td>
    <td>34596.000000</td>
    <td>6.795000e+05</td>
    <td>0.028663</td>
    <td>-12413.000000</td>
    <td>-289.000000</td>
    <td>...</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>0.000000</td>
    <td>3.000000</td>
  </tr>
  <tr>
```

```
      <th>max</th>
      <td>456255.000000</td>
      <td>1.000000</td>
      <td>19.000000</td>
      <td>1.170000e+08</td>
      <td>4.050000e+06</td>
      <td>258025.500000</td>
      <td>4.050000e+06</td>
      <td>0.072508</td>
      <td>-7489.000000</td>
      <td>365243.000000</td>
      <td>...</td>
      <td>1.000000</td>
      <td>1.000000</td>
      <td>1.000000</td>
      <td>1.000000</td>
      <td>4.000000</td>
      <td>9.000000</td>
      <td>8.000000</td>
      <td>27.000000</td>
      <td>261.000000</td>
      <td>25.000000</td>
    </tr>
  </tbody>
</table>
<p>8 rows × 106 columns</p>
</div>
df1.isnull()
<div>
<style scoped>
    .dataframe tbody tr th:only-of-type {
```

```
        vertical-align: middle;
    }

    .dataframe tbody tr th {
        vertical-align: top;
    }

    .dataframe thead th {
        text-align: right;
    }
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_CURR</th>
      <th>TARGET</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>CODE_GENDER</th>
      <th>FLAG_OWN_CAR</th>
      <th>FLAG_OWN_REALTY</th>
      <th>CNT_CHILDREN</th>
      <th>AMT_INCOME_TOTAL</th>
      <th>AMT_CREDIT</th>
      <th>AMT_ANNUITY</th>
      <th>...</th>
      <th>FLAG_DOCUMENT_18</th>
      <th>FLAG_DOCUMENT_19</th>
      <th>FLAG_DOCUMENT_20</th>
      <th>FLAG_DOCUMENT_21</th>
      <th>AMT_REQ_CREDIT_BUREAU_HOUR</th>
```

```
    <th>AMT_REQ_CREDIT_BUREAU_DAY</th>
    <th>AMT_REQ_CREDIT_BUREAU_WEEK</th>
    <th>AMT_REQ_CREDIT_BUREAU_MON</th>
    <th>AMT_REQ_CREDIT_BUREAU_QRT</th>
    <th>AMT_REQ_CREDIT_BUREAU_YEAR</th>
  </tr>
</thead>
<tbody>
  <tr>
    <th>0</th>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>...</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
```

```
</tr>
<tr>
  <th>1</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>2</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>3</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
```

<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<th>4</th>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>...</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>

```
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>...</th>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
</tr>
<tr>
  <th>307506</th>
  <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
</tr>
<tr>
  <th>307507</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
</tr>
<tr>
  <th>307508</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>307509</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
```

```
   <tr>
     <th>307510</th>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>...</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
     <td>False</td>
   </tr>
  </tbody>
</table>
<p>307511 rows × 122 columns</p>
</div>
text_dict=df1.isnull().sum()
df1.isnull()
<div>
```

```
<style scoped>
    .dataframe tbody tr th:only-of-type {
        vertical-align: middle;
    }

    .dataframe tbody tr th {
        vertical-align: top;
    }

    .dataframe thead th {
        text-align: right;
    }
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_CURR</th>
      <th>TARGET</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>CODE_GENDER</th>
      <th>FLAG_OWN_CAR</th>
      <th>FLAG_OWN_REALTY</th>
      <th>CNT_CHILDREN</th>
      <th>AMT_INCOME_TOTAL</th>
      <th>AMT_CREDIT</th>
      <th>AMT_ANNUITY</th>
      <th>...</th>
      <th>FLAG_DOCUMENT_18</th>
      <th>FLAG_DOCUMENT_19</th>
      <th>FLAG_DOCUMENT_20</th>
```

```
      <th>FLAG_DOCUMENT_21</th>
      <th>AMT_REQ_CREDIT_BUREAU_HOUR</th>
      <th>AMT_REQ_CREDIT_BUREAU_DAY</th>
      <th>AMT_REQ_CREDIT_BUREAU_WEEK</th>
      <th>AMT_REQ_CREDIT_BUREAU_MON</th>
      <th>AMT_REQ_CREDIT_BUREAU_QRT</th>
      <th>AMT_REQ_CREDIT_BUREAU_YEAR</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <th>0</th>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>...</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>1</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>2</th>
  <td>False</td>
  <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>3</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
```

```
    <td>False</td>
    <td>...</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
  </tr>
  <tr>
    <th>4</th>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>...</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>...</th>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
</tr>
<tr>
```

```
    <th>307506</th>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>...</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
    <td>True</td>
  </tr>
  <tr>
    <th>307507</th>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
    <td>False</td>
```

```
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
  <td>True</td>
</tr>
<tr>
  <th>307508</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
```

  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
</tr>
<tr>
  <th>307509</th>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>...</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>
  <td>False</td>

```
      <td>False</td>
    </tr>
    <tr>
      <th>307510</th>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>...</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
      <td>False</td>
    </tr>
  </tbody>
</table>
<p>307511 rows × 122 columns</p>
</div>
#data cleaning,- REMOVING UNWANTED COLOUMNS IN THE FIRST DATA TABLE
```

#df1.drop(columns=['NAME_TYPE_SUITE','OCCUPATION_TYPE','CNT_FAM_MEMBERS','FLAG_WORK_PHONE'])

#df1

dfn=df1.drop(columns=['CNT_CHILDREN','NAME_TYPE_SUITE','OCCUPATION_TYPE','CNT_FAM_MEMBERS','FLAG_WORK_PHONE','FLAG_DOCUMENT_18','FLAG_DOCUMENT_19','FLAG_DOCUMENT_20','AMT_REQ_CREDIT_BUREAU_HOUR','AMT_REQ_CREDIT_BUREAU_DAY','AMT_REQ_CREDIT_BUREAU_WEEK','AMT_REQ_CREDIT_BUREAU_QRT','AMT_REQ_CREDIT_BUREAU_YEAR','FLAG_DOCUMENT_21','REGION_RATING_CLIENT_W_CITY','WEEKDAY_APPR_PROCESS_START','HOUR_APPR_PROCESS_START','APARTMENTS_AVG','BASEMENTAREA_AVG','YEARS_BEGINEXPLUATATION_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_BEGINEXPLUATATION_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','YEARS_BEGINEXPLUATATION_MEDI','YEARS_BUILD_MEDI','COMMONAREA_MEDI','ELEVATORS_MEDI','ENTRANCES_MEDI','FLOORSMAX_MEDI','FLOORSMIN_MEDI','LANDAREA_MEDI','LIVINGAPARTMENTS_MEDI','LIVINGAREA_MEDI','NONLIVINGAPARTMENTS_MEDI','NONLIVINGAREA_MEDI','FONDKAPREMONT_MODE','TOTALAREA_MODE','WALLSMATERIAL_MODE','DEF_30_CNT_SOCIAL_CIRCLE','OBS_60_CNT_SOCIAL_CIRCLE','FLAG_DOCUMENT_3','FLAG_DOCUMENT_4','FLAG_DOCUMENT_5','FLAG_DOCUMENT_6','FLAG_DOCUMENT_7','FLAG_DOCUMENT_8','FLAG_DOCUMENT_9','FLAG_DOCUMENT_10','FLAG_DOCUMENT_11','FLAG_DOCUMENT_12','FLAG_DOCUMENT_13','FLAG_DOCUMENT_14','FLAG_DOCUMENT_15','FLAG_DOCUMENT_16','FLAG_DOCUMENT_17','FLAG_DOCUMENT_18','FLAG_DOCUMENT_19','FLAG_DOCUMENT_20','FLAG_DOCUMENT_21'])

dfn

<div>

<style scoped>

.dataframe tbody tr th:only-of-type {

vertical-align: middle;

}

.dataframe tbody tr th {

vertical-align: top;

}

.dataframe thead th {

text-align: right;

}

```
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_CURR</th>
      <th>TARGET</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>CODE_GENDER</th>
      <th>FLAG_OWN_CAR</th>
      <th>FLAG_OWN_REALTY</th>
      <th>AMT_INCOME_TOTAL</th>
      <th>AMT_CREDIT</th>
      <th>AMT_ANNUITY</th>
      <th>AMT_GOODS_PRICE</th>
      <th>...</th>
      <th>EXT_SOURCE_1</th>
      <th>EXT_SOURCE_2</th>
      <th>EXT_SOURCE_3</th>
      <th>HOUSETYPE_MODE</th>
      <th>EMERGENCYSTATE_MODE</th>
      <th>OBS_30_CNT_SOCIAL_CIRCLE</th>
      <th>DEF_60_CNT_SOCIAL_CIRCLE</th>
      <th>DAYS_LAST_PHONE_CHANGE</th>
      <th>FLAG_DOCUMENT_2</th>
      <th>AMT_REQ_CREDIT_BUREAU_MON</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <th>0</th>
```

```
    <td>100002</td>
    <td>1</td>
    <td>Cash loans</td>
    <td>M</td>
    <td>N</td>
    <td>Y</td>
    <td>202500.0</td>
    <td>406597.5</td>
    <td>24700.5</td>
    <td>351000.0</td>
    <td>...</td>
    <td>0.083037</td>
    <td>0.262949</td>
    <td>0.139376</td>
    <td>block of flats</td>
    <td>No</td>
    <td>2.0</td>
    <td>2.0</td>
    <td>-1134.0</td>
    <td>0</td>
    <td>0.0</td>
  </tr>
  <tr>
    <th>1</th>
    <td>100003</td>
    <td>0</td>
    <td>Cash loans</td>
    <td>F</td>
    <td>N</td>
    <td>N</td>
    <td>270000.0</td>
```

<td>1293502.5</td>
<td>35698.5</td>
<td>1129500.0</td>
<td>...</td>
<td>0.311267</td>
<td>0.622246</td>
<td>NaN</td>
<td>block of flats</td>
<td>No</td>
<td>1.0</td>
<td>0.0</td>
<td>-828.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<th>2</th>
<td>100004</td>
<td>0</td>
<td>Revolving loans</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td>67500.0</td>
<td>135000.0</td>
<td>6750.0</td>
<td>135000.0</td>
<td>...</td>
<td>NaN</td>
<td>0.555912</td>
<td>0.729567</td>

<td>NaN</td>

<td>NaN</td>

<td>0.0</td>

<td>0.0</td>

<td>-815.0</td>

<td>0</td>

<td>0.0</td>

</tr>

<tr>

<th>3</th>

<td>100006</td>

<td>0</td>

<td>Cash loans</td>

<td>F</td>

<td>N</td>

<td>Y</td>

<td>135000.0</td>

<td>312682.5</td>

<td>29686.5</td>

<td>297000.0</td>

<td>...</td>

<td>NaN</td>

<td>0.650442</td>

<td>NaN</td>

<td>NaN</td>

<td>NaN</td>

<td>2.0</td>

<td>0.0</td>

<td>-617.0</td>

<td>0</td>

<td>NaN</td>

```
</tr>
<tr>
  <th>4</th>
  <td>100007</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>N</td>
  <td>Y</td>
  <td>121500.0</td>
  <td>513000.0</td>
  <td>21865.5</td>
  <td>513000.0</td>
  <td>...</td>
  <td>NaN</td>
  <td>0.322738</td>
  <td>NaN</td>
  <td>NaN</td>
  <td>NaN</td>
  <td>0.0</td>
  <td>0.0</td>
  <td>-1106.0</td>
  <td>0</td>
  <td>0.0</td>
</tr>
<tr>
  <th>...</th>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
```

```
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
</tr>
<tr>
  <th>307506</th>
  <td>456251</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>N</td>
  <td>N</td>
  <td>157500.0</td>
  <td>254700.0</td>
  <td>27558.0</td>
  <td>225000.0</td>
  <td>...</td>
```

```
  <td>0.145570</td>
  <td>0.681632</td>
  <td>NaN</td>
  <td>block of flats</td>
  <td>No</td>
  <td>0.0</td>
  <td>0.0</td>
  <td>-273.0</td>
  <td>0</td>
  <td>NaN</td>
</tr>
<tr>
  <th>307507</th>
  <td>456252</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>72000.0</td>
  <td>269550.0</td>
  <td>12001.5</td>
  <td>225000.0</td>
  <td>...</td>
  <td>NaN</td>
  <td>0.115992</td>
  <td>NaN</td>
  <td>block of flats</td>
  <td>No</td>
  <td>0.0</td>
  <td>0.0</td>
```

<td>0.0</td>

<td>0</td>

<td>NaN</td>

</tr>

<tr>

<th>307508</th>

<td>456253</td>

<td>0</td>

<td>Cash loans</td>

<td>F</td>

<td>N</td>

<td>Y</td>

<td>153000.0</td>

<td>677664.0</td>

<td>29979.0</td>

<td>585000.0</td>

<td>...</td>

<td>0.744026</td>

<td>0.535722</td>

<td>0.218859</td>

<td>block of flats</td>

<td>No</td>

<td>6.0</td>

<td>0.0</td>

<td>-1909.0</td>

<td>0</td>

<td>1.0</td>

</tr>

<tr>

<th>307509</th>

<td>456254</td>

<td>1</td>
<td>Cash loans</td>
<td>F</td>
<td>N</td>
<td>Y</td>
<td>171000.0</td>
<td>370107.0</td>
<td>20205.0</td>
<td>319500.0</td>
<td>...</td>
<td>NaN</td>
<td>0.514163</td>
<td>0.661024</td>
<td>block of flats</td>
<td>No</td>
<td>0.0</td>
<td>0.0</td>
<td>-322.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<th>307510</th>
<td>456255</td>
<td>0</td>
<td>Cash loans</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>157500.0</td>
<td>675000.0</td>

```
      <td>49117.5</td>
      <td>675000.0</td>
      <td>...</td>
      <td>0.734460</td>
      <td>0.708569</td>
      <td>0.113922</td>
      <td>block of flats</td>
      <td>No</td>
      <td>0.0</td>
      <td>0.0</td>
      <td>-787.0</td>
      <td>0</td>
      <td>2.0</td>
    </tr>
  </tbody>
</table>
<p>307511 rows × 43 columns</p>
</div>
#DATA CLEANING- chnaging the structure of the values in the column data
Dfn
<div>
<style scoped>
    .dataframe tbody tr th:only-of-type {
        vertical-align: middle;
    }

    .dataframe tbody tr th {
        vertical-align: top;
    }

    .dataframe thead th {
```

```
        text-align: right;
    }
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_CURR</th>
      <th>TARGET</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>CODE_GENDER</th>
      <th>FLAG_OWN_CAR</th>
      <th>FLAG_OWN_REALTY</th>
      <th>AMT_INCOME_TOTAL</th>
      <th>AMT_CREDIT</th>
      <th>AMT_ANNUITY</th>
      <th>AMT_GOODS_PRICE</th>
      <th>...</th>
      <th>EXT_SOURCE_1</th>
      <th>EXT_SOURCE_2</th>
      <th>EXT_SOURCE_3</th>
      <th>HOUSETYPE_MODE</th>
      <th>EMERGENCYSTATE_MODE</th>
      <th>OBS_30_CNT_SOCIAL_CIRCLE</th>
      <th>DEF_60_CNT_SOCIAL_CIRCLE</th>
      <th>DAYS_LAST_PHONE_CHANGE</th>
      <th>FLAG_DOCUMENT_2</th>
      <th>AMT_REQ_CREDIT_BUREAU_MON</th>
    </tr>
  </thead>
  <tbody>
```

```
<tr>
  <th>0</th>
  <td>100002</td>
  <td>1</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>N</td>
  <td>Y</td>
  <td>202500.0</td>
  <td>406597.5</td>
  <td>24700.5</td>
  <td>351000.0</td>
  <td>...</td>
  <td>0.083037</td>
  <td>0.262949</td>
  <td>0.139376</td>
  <td>block of flats</td>
  <td>No</td>
  <td>2.0</td>
  <td>2.0</td>
  <td>-1134.0</td>
  <td>0</td>
  <td>0.0</td>
</tr>
<tr>
  <th>1</th>
  <td>100003</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
```

<td>N</td>

<td>270000.0</td>

<td>1293502.5</td>

<td>35698.5</td>

<td>1129500.0</td>

<td>...</td>

<td>0.311267</td>

<td>0.622246</td>

<td>NaN</td>

<td>block of flats</td>

<td>No</td>

<td>1.0</td>

<td>0.0</td>

<td>-828.0</td>

<td>0</td>

<td>0.0</td>

</tr>

<tr>

<th>2</th>

<td>100004</td>

<td>0</td>

<td>Revolving loans</td>

<td>M</td>

<td>Y</td>

<td>Y</td>

<td>67500.0</td>

<td>135000.0</td>

<td>6750.0</td>

<td>135000.0</td>

<td>...</td>

<td>NaN</td>

```
  <td>0.555912</td>
  <td>0.729567</td>
  <td>NaN</td>
  <td>NaN</td>
  <td>0.0</td>
  <td>0.0</td>
  <td>-815.0</td>
  <td>0</td>
  <td>0.0</td>
</tr>
<tr>
  <th>3</th>
  <td>100006</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>135000.0</td>
  <td>312682.5</td>
  <td>29686.5</td>
  <td>297000.0</td>
  <td>...</td>
  <td>NaN</td>
  <td>0.650442</td>
  <td>NaN</td>
  <td>NaN</td>
  <td>NaN</td>
  <td>2.0</td>
  <td>0.0</td>
  <td>-617.0</td>
```

<td>0</td>

<td>NaN</td>

</tr>

<tr>

<th>4</th>

<td>100007</td>

<td>0</td>

<td>Cash loans</td>

<td>M</td>

<td>N</td>

<td>Y</td>

<td>121500.0</td>

<td>513000.0</td>

<td>21865.5</td>

<td>513000.0</td>

<td>...</td>

<td>NaN</td>

<td>0.322738</td>

<td>NaN</td>

<td>NaN</td>

<td>NaN</td>

<td>0.0</td>

<td>0.0</td>

<td>-1106.0</td>

<td>0</td>

<td>0.0</td>

</tr>

<tr>

<th>...</th>

<td>...</td>

<td>...</td>

```
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
    <td>...</td>
  </tr>
  <tr>
    <th>307506</th>
    <td>456251</td>
    <td>0</td>
    <td>Cash loans</td>
    <td>M</td>
    <td>N</td>
    <td>N</td>
    <td>157500.0</td>
    <td>254700.0</td>
    <td>27558.0</td>
```

```
  <td>225000.0</td>
  <td>...</td>
  <td>0.145570</td>
  <td>0.681632</td>
  <td>NaN</td>
  <td>block of flats</td>
  <td>No</td>
  <td>0.0</td>
  <td>0.0</td>
  <td>-273.0</td>
  <td>0</td>
  <td>NaN</td>
</tr>
<tr>
  <th>307507</th>
  <td>456252</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>72000.0</td>
  <td>269550.0</td>
  <td>12001.5</td>
  <td>225000.0</td>
  <td>...</td>
  <td>NaN</td>
  <td>0.115992</td>
  <td>NaN</td>
  <td>block of flats</td>
  <td>No</td>
```

```
  <td>0.0</td>
  <td>0.0</td>
  <td>0.0</td>
  <td>0</td>
  <td>NaN</td>
</tr>
<tr>
  <th>307508</th>
  <td>456253</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>153000.0</td>
  <td>677664.0</td>
  <td>29979.0</td>
  <td>585000.0</td>
  <td>...</td>
  <td>0.744026</td>
  <td>0.535722</td>
  <td>0.218859</td>
  <td>block of flats</td>
  <td>No</td>
  <td>6.0</td>
  <td>0.0</td>
  <td>-1909.0</td>
  <td>0</td>
  <td>1.0</td>
</tr>
<tr>
```

```
  <th>307509</th>
  <td>456254</td>
  <td>1</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>171000.0</td>
  <td>370107.0</td>
  <td>20205.0</td>
  <td>319500.0</td>
  <td>...</td>
  <td>NaN</td>
  <td>0.514163</td>
  <td>0.661024</td>
  <td>block of flats</td>
  <td>No</td>
  <td>0.0</td>
  <td>0.0</td>
  <td>-322.0</td>
  <td>0</td>
  <td>0.0</td>
</tr>
<tr>
  <th>307510</th>
  <td>456255</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>N</td>
```

```
      <td>157500.0</td>
      <td>675000.0</td>
      <td>49117.5</td>
      <td>675000.0</td>
      <td>...</td>
      <td>0.734460</td>
      <td>0.708569</td>
      <td>0.113922</td>
      <td>block of flats</td>
      <td>No</td>
      <td>0.0</td>
      <td>0.0</td>
      <td>-787.0</td>
      <td>0</td>
      <td>2.0</td>
    </tr>
  </tbody>
</table>
<p>307511 rows × 43 columns</p>
</div>
```

```
dfn['DAYS_BIRTH'].isnull().sum()
```

0

```
dfn['DAYS_BIRTH']=abs(dfn['DAYS_BIRTH'])#replacing negative values with absolute values
dfn['PERSON_AGE']=dfn.apply( lambda row : int(row.DAYS_BIRTH /365),axis=1)
dfn['PERSON_AGE'].isnull().sum()
```

0

```
dfn['YEARS_EMPLOYED']=dfn.apply( lambda row : int(row.DAYS_EMPLOYED/365),axis=1)
dfn['YEARS_EMPLOYED']=abs(dfn['YEARS_EMPLOYED'])
dfn['DAYS_LAST_PHONE_CHANGE']=abs(dfn['DAYS_LAST_PHONE_CHANGE'])
dfn['AMT_INCOME_TOTAL'].isnull().sum()
```

0

```
dfn.isnull().sum()
SK_ID_CURR                    0
TARGET                        0
NAME_CONTRACT_TYPE            0
CODE_GENDER                   0
FLAG_OWN_CAR                  0
FLAG_OWN_REALTY               0
AMT_INCOME_TOTAL              0
AMT_CREDIT                    0
AMT_ANNUITY                  12
AMT_GOODS_PRICE             278
NAME_INCOME_TYPE              0
NAME_EDUCATION_TYPE           0
NAME_FAMILY_STATUS            0
NAME_HOUSING_TYPE             0
REGION_POPULATION_RELATIVE    0
DAYS_BIRTH                    0
DAYS_EMPLOYED                 0
DAYS_REGISTRATION             0
DAYS_ID_PUBLISH               0
OWN_CAR_AGE              202929
FLAG_MOBIL                    0
FLAG_EMP_PHONE                0
FLAG_CONT_MOBILE              0
FLAG_PHONE                    0
FLAG_EMAIL                    0
REGION_RATING_CLIENT          0
REG_REGION_NOT_LIVE_REGION    0
REG_REGION_NOT_WORK_REGION    0
LIVE_REGION_NOT_WORK_REGION   0
REG_CITY_NOT_LIVE_CITY        0
```

```
REG_CITY_NOT_WORK_CITY            0
LIVE_CITY_NOT_WORK_CITY           0
ORGANIZATION_TYPE                 0
EXT_SOURCE_1                 173378
EXT_SOURCE_2                    660
EXT_SOURCE_3                  60965
HOUSETYPE_MODE               154297
EMERGENCYSTATE_MODE          145755
OBS_30_CNT_SOCIAL_CIRCLE       1021
DEF_60_CNT_SOCIAL_CIRCLE       1021
DAYS_LAST_PHONE_CHANGE            1
FLAG_DOCUMENT_2                   0
AMT_REQ_CREDIT_BUREAU_MON     41519
PERSON_AGE                        0
YEARS_EMPLOYED                    0
dtype: int64
```

```
dfn.dropna(axis=1,how='any',inplace=True)
dfn
```

```
<div>
<style scoped>
    .dataframe tbody tr th:only-of-type {
        vertical-align: middle;
    }

    .dataframe tbody tr th {
        vertical-align: top;
    }

    .dataframe thead th {
        text-align: right;
    }
```

```
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_CURR</th>
      <th>TARGET</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>CODE_GENDER</th>
      <th>FLAG_OWN_CAR</th>
      <th>FLAG_OWN_REALTY</th>
      <th>AMT_INCOME_TOTAL</th>
      <th>AMT_CREDIT</th>
      <th>NAME_INCOME_TYPE</th>
      <th>NAME_EDUCATION_TYPE</th>
      <th>...</th>
      <th>REG_REGION_NOT_LIVE_REGION</th>
      <th>REG_REGION_NOT_WORK_REGION</th>
      <th>LIVE_REGION_NOT_WORK_REGION</th>
      <th>REG_CITY_NOT_LIVE_CITY</th>
      <th>REG_CITY_NOT_WORK_CITY</th>
      <th>LIVE_CITY_NOT_WORK_CITY</th>
      <th>ORGANIZATION_TYPE</th>
      <th>FLAG_DOCUMENT_2</th>
      <th>PERSON_AGE</th>
      <th>YEARS_EMPLOYED</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <th>0</th>
```

```
  <td>100002</td>
  <td>1</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>N</td>
  <td>Y</td>
  <td>202500.0</td>
  <td>406597.5</td>
  <td>Working</td>
  <td>Secondary / secondary special</td>
  <td>...</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>Business Entity Type 3</td>
  <td>0</td>
  <td>25</td>
  <td>1</td>
</tr>
<tr>
  <th>1</th>
  <td>100003</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>N</td>
  <td>270000.0</td>
```

<td>1293502.5</td>

<td>State servant</td>

<td>Higher education</td>

<td>...</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>School</td>

<td>0</td>

<td>45</td>

<td>3</td>

</tr>

<tr>

<th>2</th>

<td>100004</td>

<td>0</td>

<td>Revolving loans</td>

<td>M</td>

<td>Y</td>

<td>Y</td>

<td>67500.0</td>

<td>135000.0</td>

<td>Working</td>

<td>Secondary / secondary special</td>

<td>...</td>

<td>0</td>

<td>0</td>

<td>0</td>

```
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>Government</td>
  <td>0</td>
  <td>52</td>
  <td>0</td>
</tr>
<tr>
  <th>3</th>
  <td>100006</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>135000.0</td>
  <td>312682.5</td>
  <td>Working</td>
  <td>Secondary / secondary special</td>
  <td>...</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>Business Entity Type 3</td>
  <td>0</td>
  <td>52</td>
  <td>8</td>
```

</tr>

<tr>

<th>4</th>

<td>100007</td>

<td>0</td>

<td>Cash loans</td>

<td>M</td>

<td>N</td>

<td>Y</td>

<td>121500.0</td>

<td>513000.0</td>

<td>Working</td>

<td>Secondary / secondary special</td>

<td>...</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>1</td>

<td>1</td>

<td>Religion</td>

<td>0</td>

<td>54</td>

<td>8</td>

</tr>

<tr>

<th>...</th>

<td>...</td>

<td>...</td>

<td>...</td>

<td>...</td>

```
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
</tr>
<tr>
  <th>307506</th>
  <td>456251</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>N</td>
  <td>N</td>
  <td>157500.0</td>
  <td>254700.0</td>
  <td>Working</td>
  <td>Secondary / secondary special</td>
  <td>...</td>
```

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>Services</td>

<td>0</td>

<td>25</td>

<td>0</td>

</tr>

<tr>

<th>307507</th>

<td>456252</td>

<td>0</td>

<td>Cash loans</td>

<td>F</td>

<td>N</td>

<td>Y</td>

<td>72000.0</td>

<td>269550.0</td>

<td>Pensioner</td>

<td>Secondary / secondary special</td>

<td>...</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>0</td>

<td>XNA</td>

```
  <td>0</td>
  <td>56</td>
  <td>1000</td>
</tr>
<tr>
  <th>307508</th>
  <td>456253</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>153000.0</td>
  <td>677664.0</td>
  <td>Working</td>
  <td>Higher education</td>
  <td>...</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>1</td>
  <td>1</td>
  <td>School</td>
  <td>0</td>
  <td>41</td>
  <td>21</td>
</tr>
<tr>
  <th>307509</th>
  <td>456254</td>
```

```
    <td>1</td>
    <td>Cash loans</td>
    <td>F</td>
    <td>N</td>
    <td>Y</td>
    <td>171000.0</td>
    <td>370107.0</td>
    <td>Commercial associate</td>
    <td>Secondary / secondary special</td>
    <td>...</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>1</td>
    <td>1</td>
    <td>0</td>
    <td>Business Entity Type 1</td>
    <td>0</td>
    <td>32</td>
    <td>13</td>
  </tr>
  <tr>
    <th>307510</th>
    <td>456255</td>
    <td>0</td>
    <td>Cash loans</td>
    <td>F</td>
    <td>N</td>
    <td>N</td>
    <td>157500.0</td>
    <td>675000.0</td>
```

```
      <td>Commercial associate</td>
      <td>Higher education</td>
      <td>...</td>
      <td>0</td>
      <td>0</td>
      <td>0</td>
      <td>0</td>
      <td>1</td>
      <td>1</td>
      <td>Business Entity Type 3</td>
      <td>0</td>
      <td>46</td>
      <td>3</td>
    </tr>
  </tbody>
</table>
<p>307511 rows × 33 columns</p>
</div>
import pandas as pd
df2=pd.read_csv(r"C:\Users\radhi\OneDrive\Desktop\PythonProject\previous_application.csv")
df2.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1670214 entries, 0 to 1670213
Data columns (total 37 columns):
 #   Column                       Non-Null Count    Dtype
---  ------                       --------------    -----
 0   SK_ID_PREV                   1670214 non-null  int64
 1   SK_ID_CURR                   1670214 non-null  int64
 2   NAME_CONTRACT_TYPE           1670214 non-null  object
 3   AMT_ANNUITY                  1297979 non-null  float64
 4   AMT_APPLICATION              1670214 non-null  float64
```

5 AMT_CREDIT 1670213 non-null float64

6 AMT_DOWN_PAYMENT 774370 non-null float64

7 AMT_GOODS_PRICE 1284699 non-null float64

8 WEEKDAY_APPR_PROCESS_START 1670214 non-null object

9 HOUR_APPR_PROCESS_START 1670214 non-null int64

10 FLAG_LAST_APPL_PER_CONTRACT 1670214 non-null object

11 NFLAG_LAST_APPL_IN_DAY 1670214 non-null int64

12 RATE_DOWN_PAYMENT 774370 non-null float64

13 RATE_INTEREST_PRIMARY 5951 non-null float64

14 RATE_INTEREST_PRIVILEGED 5951 non-null float64

15 NAME_CASH_LOAN_PURPOSE 1670214 non-null object

16 NAME_CONTRACT_STATUS 1670214 non-null object

17 DAYS_DECISION 1670214 non-null int64

18 NAME_PAYMENT_TYPE 1670214 non-null object

19 CODE_REJECT_REASON 1670214 non-null object

20 NAME_TYPE_SUITE 849809 non-null object

21 NAME_CLIENT_TYPE 1670214 non-null object

22 NAME_GOODS_CATEGORY 1670214 non-null object

23 NAME_PORTFOLIO 1670214 non-null object

24 NAME_PRODUCT_TYPE 1670214 non-null object

25 CHANNEL_TYPE 1670214 non-null object

26 SELLERPLACE_AREA 1670214 non-null int64

27 NAME_SELLER_INDUSTRY 1670214 non-null object

28 CNT_PAYMENT 1297984 non-null float64

29 NAME_YIELD_GROUP 1670214 non-null object

30 PRODUCT_COMBINATION 1669868 non-null object

31 DAYS_FIRST_DRAWING 997149 non-null float64

32 DAYS_FIRST_DUE 997149 non-null float64

33 DAYS_LAST_DUE_1ST_VERSION 997149 non-null float64

34 DAYS_LAST_DUE 997149 non-null float64

35 DAYS_TERMINATION 997149 non-null float64

```
 36  NFLAG_INSURED_ON_APPROVAL    997149 non-null   float64
dtypes: float64(15), int64(6), object(16)
memory usage: 471.5+ MB
df2.dropna(axis=1,how='any',inplace=True)
df2
df2n=df2.drop(columns=['HOUR_APPR_PROCESS_START','WEEKDAY_APPR_PROCESS_START',])
df2n
<div>
<style scoped>
    .dataframe tbody tr th:only-of-type {
        vertical-align: middle;
    }

    .dataframe tbody tr th {
        vertical-align: top;
    }

    .dataframe thead th {
        text-align: right;
    }
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_PREV</th>
      <th>SK_ID_CURR</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>AMT_APPLICATION</th>
      <th>FLAG_LAST_APPL_PER_CONTRACT</th>
      <th>NFLAG_LAST_APPL_IN_DAY</th>
```

```
    <th>NAME_CASH_LOAN_PURPOSE</th>
    <th>NAME_CONTRACT_STATUS</th>
    <th>DAYS_DECISION</th>
    <th>NAME_PAYMENT_TYPE</th>
    <th>CODE_REJECT_REASON</th>
    <th>NAME_CLIENT_TYPE</th>
    <th>NAME_GOODS_CATEGORY</th>
    <th>NAME_PORTFOLIO</th>
    <th>NAME_PRODUCT_TYPE</th>
    <th>CHANNEL_TYPE</th>
    <th>SELLERPLACE_AREA</th>
    <th>NAME_SELLER_INDUSTRY</th>
    <th>NAME_YIELD_GROUP</th>
  </tr>
</thead>
<tbody>
  <tr>
    <th>0</th>
    <td>2030495</td>
    <td>271877</td>
    <td>Consumer loans</td>
    <td>17145.0</td>
    <td>Y</td>
    <td>1</td>
    <td>XAP</td>
    <td>Approved</td>
    <td>-73</td>
    <td>Cash through the bank</td>
    <td>XAP</td>
    <td>Repeater</td>
    <td>Mobile</td>
```

```
  <td>POS</td>
  <td>XNA</td>
  <td>Country-wide</td>
  <td>35</td>
  <td>Connectivity</td>
  <td>middle</td>
</tr>
<tr>
  <th>1</th>
  <td>2802425</td>
  <td>108129</td>
  <td>Cash loans</td>
  <td>607500.0</td>
  <td>Y</td>
  <td>1</td>
  <td>XNA</td>
  <td>Approved</td>
  <td>-164</td>
  <td>XNA</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>x-sell</td>
  <td>Contact center</td>
  <td>-1</td>
  <td>XNA</td>
  <td>low_action</td>
</tr>
<tr>
  <th>2</th>
```

<td>2523466</td>

<td>122040</td>

<td>Cash loans</td>

<td>112500.0</td>

<td>Y</td>

<td>1</td>

<td>XNA</td>

<td>Approved</td>

<td>-301</td>

<td>Cash through the bank</td>

<td>XAP</td>

<td>Repeater</td>

<td>XNA</td>

<td>Cash</td>

<td>x-sell</td>

<td>Credit and cash offices</td>

<td>-1</td>

<td>XNA</td>

<td>high</td>

</tr>

<tr>

<th>3</th>

<td>2819243</td>

<td>176158</td>

<td>Cash loans</td>

<td>450000.0</td>

<td>Y</td>

<td>1</td>

<td>XNA</td>

<td>Approved</td>

<td>-512</td>

```
      <td>Cash through the bank</td>
      <td>XAP</td>
      <td>Repeater</td>
      <td>XNA</td>
      <td>Cash</td>
      <td>x-sell</td>
      <td>Credit and cash offices</td>
      <td>-1</td>
      <td>XNA</td>
      <td>middle</td>
    </tr>
    <tr>
      <th>4</th>
      <td>1784265</td>
      <td>202054</td>
      <td>Cash loans</td>
      <td>337500.0</td>
      <td>Y</td>
      <td>1</td>
      <td>Repairs</td>
      <td>Refused</td>
      <td>-781</td>
      <td>Cash through the bank</td>
      <td>HC</td>
      <td>Repeater</td>
      <td>XNA</td>
      <td>Cash</td>
      <td>walk-in</td>
      <td>Credit and cash offices</td>
      <td>-1</td>
      <td>XNA</td>
```

```
  <td>high</td>
</tr>
<tr>
  <th>...</th>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
</tr>
<tr>
  <th>1670209</th>
  <td>2300464</td>
  <td>352015</td>
  <td>Consumer loans</td>
  <td>267295.5</td>
  <td>Y</td>
```

```
  <td>1</td>
  <td>XAP</td>
  <td>Approved</td>
  <td>-544</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>Refreshed</td>
  <td>Furniture</td>
  <td>POS</td>
  <td>XNA</td>
  <td>Stone</td>
  <td>43</td>
  <td>Furniture</td>
  <td>low_normal</td>
</tr>
<tr>
  <th>1670210</th>
  <td>2357031</td>
  <td>334635</td>
  <td>Consumer loans</td>
  <td>87750.0</td>
  <td>Y</td>
  <td>1</td>
  <td>XAP</td>
  <td>Approved</td>
  <td>-1694</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>New</td>
  <td>Furniture</td>
  <td>POS</td>
```

```
  <td>XNA</td>
  <td>Stone</td>
  <td>43</td>
  <td>Furniture</td>
  <td>middle</td>
</tr>
<tr>
  <th>1670211</th>
  <td>2659632</td>
  <td>249544</td>
  <td>Consumer loans</td>
  <td>105237.0</td>
  <td>Y</td>
  <td>1</td>
  <td>XAP</td>
  <td>Approved</td>
  <td>-1488</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>Consumer Electronics</td>
  <td>POS</td>
  <td>XNA</td>
  <td>Country-wide</td>
  <td>1370</td>
  <td>Consumer electronics</td>
  <td>low_normal</td>
</tr>
<tr>
  <th>1670212</th>
  <td>2785582</td>
```

```
    <td>400317</td>
    <td>Cash loans</td>
    <td>180000.0</td>
    <td>Y</td>
    <td>1</td>
    <td>XNA</td>
    <td>Approved</td>
    <td>-1185</td>
    <td>Cash through the bank</td>
    <td>XAP</td>
    <td>Repeater</td>
    <td>XNA</td>
    <td>Cash</td>
    <td>x-sell</td>
    <td>AP+ (Cash loan)</td>
    <td>-1</td>
    <td>XNA</td>
    <td>low_normal</td>
  </tr>
  <tr>
    <th>1670213</th>
    <td>2418762</td>
    <td>261212</td>
    <td>Cash loans</td>
    <td>360000.0</td>
    <td>Y</td>
    <td>1</td>
    <td>XNA</td>
    <td>Approved</td>
    <td>-1193</td>
    <td>Cash through the bank</td>
```

<td>XAP</td>

<td>Repeater</td>

<td>XNA</td>

<td>Cash</td>

<td>x-sell</td>

<td>AP+ (Cash loan)</td>

<td>-1</td>

<td>XNA</td>

<td>middle</td>

</tr>

</tbody>

</table>

<p>1670214 rows × 19 columns</p>

</div>

df2n.dropna(axis=0,how='any',inplace=True)#removing all null values in the 'previous_application.csv' dataframe

df2n

<div>

<style scoped>

.dataframe tbody tr th:only-of-type {

vertical-align: middle;

}

.dataframe tbody tr th {

vertical-align: top;

}

.dataframe thead th {

text-align: right;

}

</style>

```
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_PREV</th>
      <th>SK_ID_CURR</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>AMT_APPLICATION</th>
      <th>FLAG_LAST_APPL_PER_CONTRACT</th>
      <th>NFLAG_LAST_APPL_IN_DAY</th>
      <th>NAME_CASH_LOAN_PURPOSE</th>
      <th>NAME_CONTRACT_STATUS</th>
      <th>DAYS_DECISION</th>
      <th>NAME_PAYMENT_TYPE</th>
      <th>CODE_REJECT_REASON</th>
      <th>NAME_CLIENT_TYPE</th>
      <th>NAME_GOODS_CATEGORY</th>
      <th>NAME_PORTFOLIO</th>
      <th>NAME_PRODUCT_TYPE</th>
      <th>CHANNEL_TYPE</th>
      <th>SELLERPLACE_AREA</th>
      <th>NAME_SELLER_INDUSTRY</th>
      <th>NAME_YIELD_GROUP</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <th>0</th>
      <td>2030495</td>
      <td>271877</td>
      <td>Consumer loans</td>
```

<td>17145.0</td>

<td>Y</td>

<td>1</td>

<td>XAP</td>

<td>Approved</td>

<td>-73</td>

<td>Cash through the bank</td>

<td>XAP</td>

<td>Repeater</td>

<td>Mobile</td>

<td>POS</td>

<td>XNA</td>

<td>Country-wide</td>

<td>35</td>

<td>Connectivity</td>

<td>middle</td>

</tr>

<tr>

<th>1</th>

<td>2802425</td>

<td>108129</td>

<td>Cash loans</td>

<td>607500.0</td>

<td>Y</td>

<td>1</td>

<td>XNA</td>

<td>Approved</td>

<td>-164</td>

<td>XNA</td>

<td>XAP</td>

<td>Repeater</td>

```
  <td>XNA</td>
  <td>Cash</td>
  <td>x-sell</td>
  <td>Contact center</td>
  <td>-1</td>
  <td>XNA</td>
  <td>low_action</td>
</tr>
<tr>
  <th>2</th>
  <td>2523466</td>
  <td>122040</td>
  <td>Cash loans</td>
  <td>112500.0</td>
  <td>Y</td>
  <td>1</td>
  <td>XNA</td>
  <td>Approved</td>
  <td>-301</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>x-sell</td>
  <td>Credit and cash offices</td>
  <td>-1</td>
  <td>XNA</td>
  <td>high</td>
</tr>
<tr>
```

<th>3</th>

<td>2819243</td>

<td>176158</td>

<td>Cash loans</td>

<td>450000.0</td>

<td>Y</td>

<td>1</td>

<td>XNA</td>

<td>Approved</td>

<td>-512</td>

<td>Cash through the bank</td>

<td>XAP</td>

<td>Repeater</td>

<td>XNA</td>

<td>Cash</td>

<td>x-sell</td>

<td>Credit and cash offices</td>

<td>-1</td>

<td>XNA</td>

<td>middle</td>

</tr>

<tr>

<th>4</th>

<td>1784265</td>

<td>202054</td>

<td>Cash loans</td>

<td>337500.0</td>

<td>Y</td>

<td>1</td>

<td>Repairs</td>

<td>Refused</td>

```
  <td>-781</td>
  <td>Cash through the bank</td>
  <td>HC</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>walk-in</td>
  <td>Credit and cash offices</td>
  <td>-1</td>
  <td>XNA</td>
  <td>high</td>
</tr>
<tr>
  <th>...</th>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
```

```
  <td>...</td>
  <td>...</td>
</tr>
<tr>
  <th>1670209</th>
  <td>2300464</td>
  <td>352015</td>
  <td>Consumer loans</td>
  <td>267295.5</td>
  <td>Y</td>
  <td>1</td>
  <td>XAP</td>
  <td>Approved</td>
  <td>-544</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>Refreshed</td>
  <td>Furniture</td>
  <td>POS</td>
  <td>XNA</td>
  <td>Stone</td>
  <td>43</td>
  <td>Furniture</td>
  <td>low_normal</td>
</tr>
<tr>
  <th>1670210</th>
  <td>2357031</td>
  <td>334635</td>
  <td>Consumer loans</td>
  <td>87750.0</td>
```

```
  <td>Y</td>
  <td>1</td>
  <td>XAP</td>
  <td>Approved</td>
  <td>-1694</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>New</td>
  <td>Furniture</td>
  <td>POS</td>
  <td>XNA</td>
  <td>Stone</td>
  <td>43</td>
  <td>Furniture</td>
  <td>middle</td>
</tr>
<tr>
  <th>1670211</th>
  <td>2659632</td>
  <td>249544</td>
  <td>Consumer loans</td>
  <td>105237.0</td>
  <td>Y</td>
  <td>1</td>
  <td>XAP</td>
  <td>Approved</td>
  <td>-1488</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>Consumer Electronics</td>
```

```
  <td>POS</td>
  <td>XNA</td>
  <td>Country-wide</td>
  <td>1370</td>
  <td>Consumer electronics</td>
  <td>low_normal</td>
</tr>
<tr>
  <th>1670212</th>
  <td>2785582</td>
  <td>400317</td>
  <td>Cash loans</td>
  <td>180000.0</td>
  <td>Y</td>
  <td>1</td>
  <td>XNA</td>
  <td>Approved</td>
  <td>-1185</td>
  <td>Cash through the bank</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>x-sell</td>
  <td>AP+ (Cash loan)</td>
  <td>-1</td>
  <td>XNA</td>
  <td>low_normal</td>
</tr>
<tr>
  <th>1670213</th>
```

```
      <td>2418762</td>
      <td>261212</td>
      <td>Cash loans</td>
      <td>360000.0</td>
      <td>Y</td>
      <td>1</td>
      <td>XNA</td>
      <td>Approved</td>
      <td>-1193</td>
      <td>Cash through the bank</td>
      <td>XAP</td>
      <td>Repeater</td>
      <td>XNA</td>
      <td>Cash</td>
      <td>x-sell</td>
      <td>AP+ (Cash loan)</td>
      <td>-1</td>
      <td>XNA</td>
      <td>middle</td>
    </tr>
  </tbody>
</table>
<p>1670214 rows × 19 columns</p>
</div>
```

```
df2n.rename(columns={'NAME_CONTRACT_TYPE':'PRE_NAME_CONTRACT_TYPE','AMT_CREDIT':'PRE_AMT_CREDIT'},inplace=True)

df2n

dfnew=pd.merge(dfn,df2n,on='SK_ID_CURR',how='left') #merged two dataframes

dfnew

for key in dfnew.columns[dfnew.isna().any()]:
    dfnew[key]=dfnew[key].replace(np.NaN,0)
```

dfnew.info()

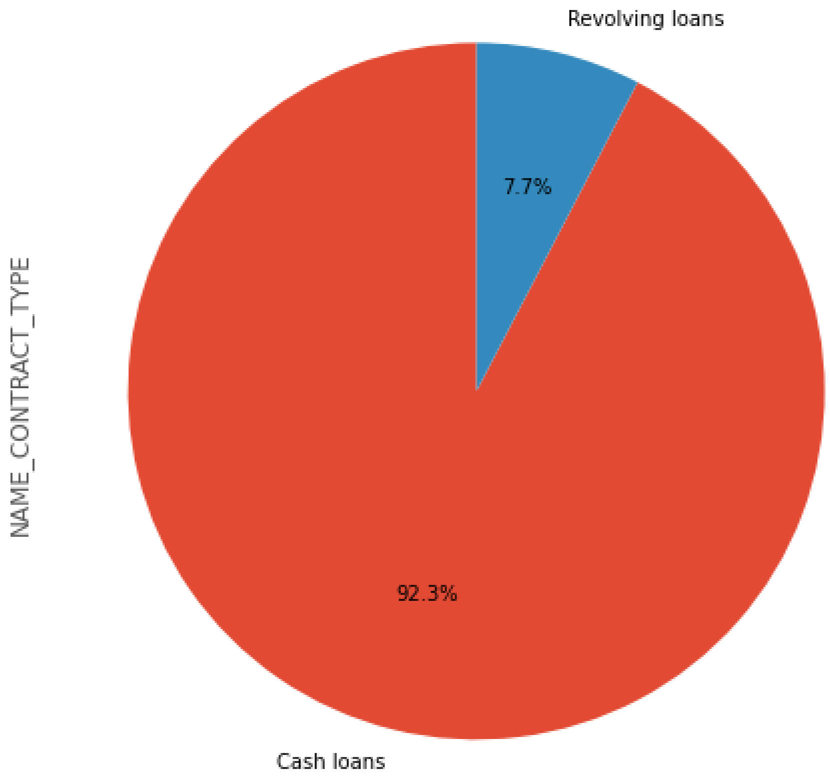
#analysis - COntract type distribution

import matplotlib.pyplot as plt

plt.figure(facecolor='white')

dfnew['NAME_CONTRACT_TYPE'].value_counts().plot(kind='pie',legend=False,autopct='%1.1f%%',figsize=(10,8),startangle=90)

plt.show()

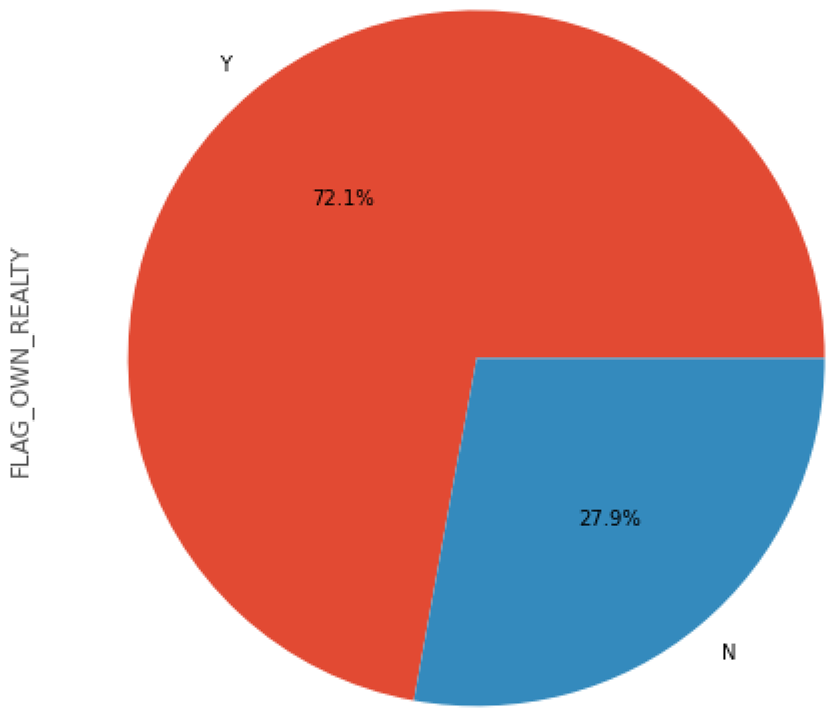


#piechart:property status

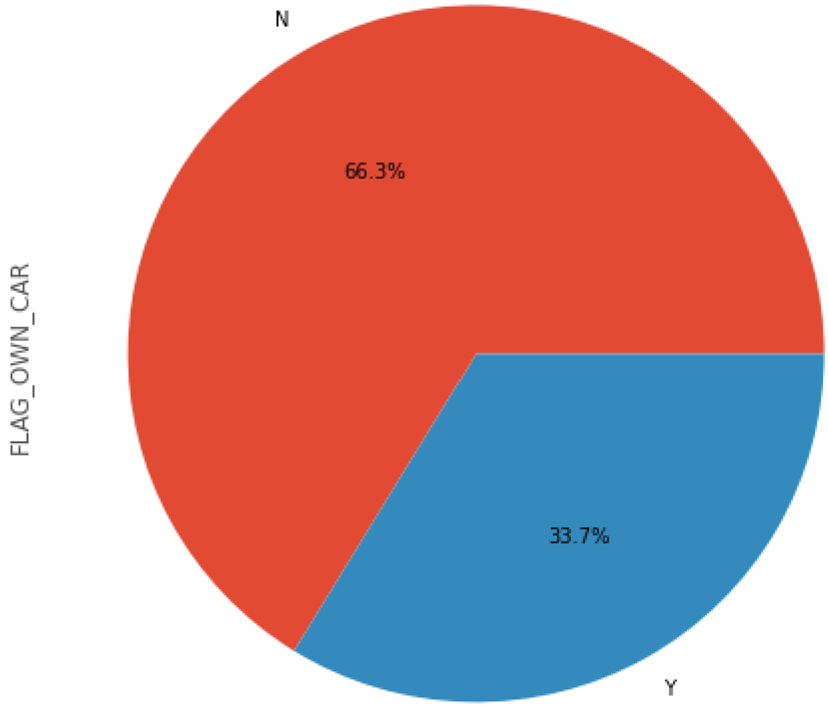
plt.figure(facecolor='white')

dfnew['FLAG_OWN_REALTY'].value_counts().plot(kind='pie', legend=False,autopct='%1.1f%%',figsize=(10,8),startangle=0)

plt.show()




```
#piechart:car status

plt.figure(facecolor='white')

dfnew['FLAG_OWN_CAR'].value_counts().plot(kind='pie',
legend=False,autopct='%1.1f%%',figsize=(10,8),startangle=0)

plt.show()
```



#dfnew['AMT_INCOME_TOTAL'].max()

#dfnew['AMT_INCOME_TOTAL'].min()

s=dfnew['AMT_INCOME_TOTAL'].astype(int)

dfnew['INCOME_RANGE']=pd.IntervalIndex.from_arrays(s,s+10000,000)

dfnew

<div>

<style scoped>

```
    .dataframe tbody tr th:only-of-type {
        vertical-align: middle;
    }

    .dataframe tbody tr th {
        vertical-align: top;
    }

    .dataframe thead th {
```

```
        text-align: right;
    }
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_CURR</th>
      <th>TARGET</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>CODE_GENDER</th>
      <th>FLAG_OWN_CAR</th>
      <th>FLAG_OWN_REALTY</th>
      <th>AMT_INCOME_TOTAL</th>
      <th>AMT_CREDIT</th>
      <th>NAME_INCOME_TYPE</th>
      <th>NAME_EDUCATION_TYPE</th>
      <th>...</th>
      <th>CODE_REJECT_REASON</th>
      <th>NAME_CLIENT_TYPE</th>
      <th>NAME_GOODS_CATEGORY</th>
      <th>NAME_PORTFOLIO</th>
      <th>NAME_PRODUCT_TYPE</th>
      <th>CHANNEL_TYPE</th>
      <th>SELLERPLACE_AREA</th>
      <th>NAME_SELLER_INDUSTRY</th>
      <th>NAME_YIELD_GROUP</th>
      <th>INCOME_RANGE</th>
    </tr>
  </thead>
  <tbody>
```

```html
<tr>
  <th>0</th>
  <td>100002</td>
  <td>1</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>N</td>
  <td>Y</td>
  <td>202500.0</td>
  <td>406597.5</td>
  <td>Working</td>
  <td>Secondary / secondary special</td>
  <td>...</td>
  <td>XAP</td>
  <td>New</td>
  <td>Vehicles</td>
  <td>POS</td>
  <td>XNA</td>
  <td>Stone</td>
  <td>500.0</td>
  <td>Auto technology</td>
  <td>low_normal</td>
  <td>(202500, 212500]</td>
</tr>
<tr>
  <th>1</th>
  <td>100003</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
```

```
  <td>N</td>
  <td>270000.0</td>
  <td>1293502.5</td>
  <td>State servant</td>
  <td>Higher education</td>
  <td>...</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>x-sell</td>
  <td>Credit and cash offices</td>
  <td>-1.0</td>
  <td>XNA</td>
  <td>low_normal</td>
  <td>(270000, 280000]</td>
</tr>
<tr>
  <th>2</th>
  <td>100003</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>N</td>
  <td>270000.0</td>
  <td>1293502.5</td>
  <td>State servant</td>
  <td>Higher education</td>
  <td>...</td>
  <td>XAP</td>
```

<td>Refreshed</td>
<td>Furniture</td>
<td>POS</td>
<td>XNA</td>
<td>Stone</td>
<td>1400.0</td>
<td>Furniture</td>
<td>middle</td>
<td>(270000, 280000]</td>
</tr>
<tr>
<th>3</th>
<td>100003</td>
<td>0</td>
<td>Cash loans</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td>270000.0</td>
<td>1293502.5</td>
<td>State servant</td>
<td>Higher education</td>
<td>...</td>
<td>XAP</td>
<td>Refreshed</td>
<td>Consumer Electronics</td>
<td>POS</td>
<td>XNA</td>
<td>Country-wide</td>
<td>200.0</td>
<td>Consumer electronics</td>

<td>middle</td>

<td>(270000, 280000]</td>

</tr>

<tr>

<th>4</th>

<td>100004</td>

<td>0</td>

<td>Revolving loans</td>

<td>M</td>

<td>Y</td>

<td>Y</td>

<td>67500.0</td>

<td>135000.0</td>

<td>Working</td>

<td>Secondary / secondary special</td>

<td>...</td>

<td>XAP</td>

<td>New</td>

<td>Mobile</td>

<td>POS</td>

<td>XNA</td>

<td>Regional / Local</td>

<td>30.0</td>

<td>Connectivity</td>

<td>middle</td>

<td>(67500, 77500]</td>

</tr>

<tr>

<th>...</th>

<td>...</td>

<td>...</td>

```
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
</tr>
<tr>
  <th>1430150</th>
  <td>456255</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>N</td>
  <td>157500.0</td>
  <td>675000.0</td>
  <td>Commercial associate</td>
```

```
  <td>Higher education</td>
  <td>...</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>x-sell</td>
  <td>Credit and cash offices</td>
  <td>-1.0</td>
  <td>XNA</td>
  <td>middle</td>
  <td>(157500, 167500]</td>
</tr>
<tr>
  <th>1430151</th>
  <td>456255</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>N</td>
  <td>157500.0</td>
  <td>675000.0</td>
  <td>Commercial associate</td>
  <td>Higher education</td>
  <td>...</td>
  <td>HC</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cards</td>
  <td>walk-in</td>
```

```
  <td>Country-wide</td>
  <td>20.0</td>
  <td>Connectivity</td>
  <td>XNA</td>
  <td>(157500, 167500]</td>
</tr>
<tr>
  <th>1430152</th>
  <td>456255</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>N</td>
  <td>157500.0</td>
  <td>675000.0</td>
  <td>Commercial associate</td>
  <td>Higher education</td>
  <td>...</td>
  <td>HC</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>walk-in</td>
  <td>Credit and cash offices</td>
  <td>-1.0</td>
  <td>XNA</td>
  <td>low_normal</td>
  <td>(157500, 167500]</td>
</tr>
<tr>
```

<th>1430153</th>

<td>456255</td>

<td>0</td>

<td>Cash loans</td>

<td>F</td>

<td>N</td>

<td>N</td>

<td>157500.0</td>

<td>675000.0</td>

<td>Commercial associate</td>

<td>Higher education</td>

<td>...</td>

<td>XAP</td>

<td>Repeater</td>

<td>XNA</td>

<td>Cash</td>

<td>x-sell</td>

<td>AP+ (Cash loan)</td>

<td>6.0</td>

<td>XNA</td>

<td>low_normal</td>

<td>(157500, 167500]</td>

</tr>

<tr>

<th>1430154</th>

<td>456255</td>

<td>0</td>

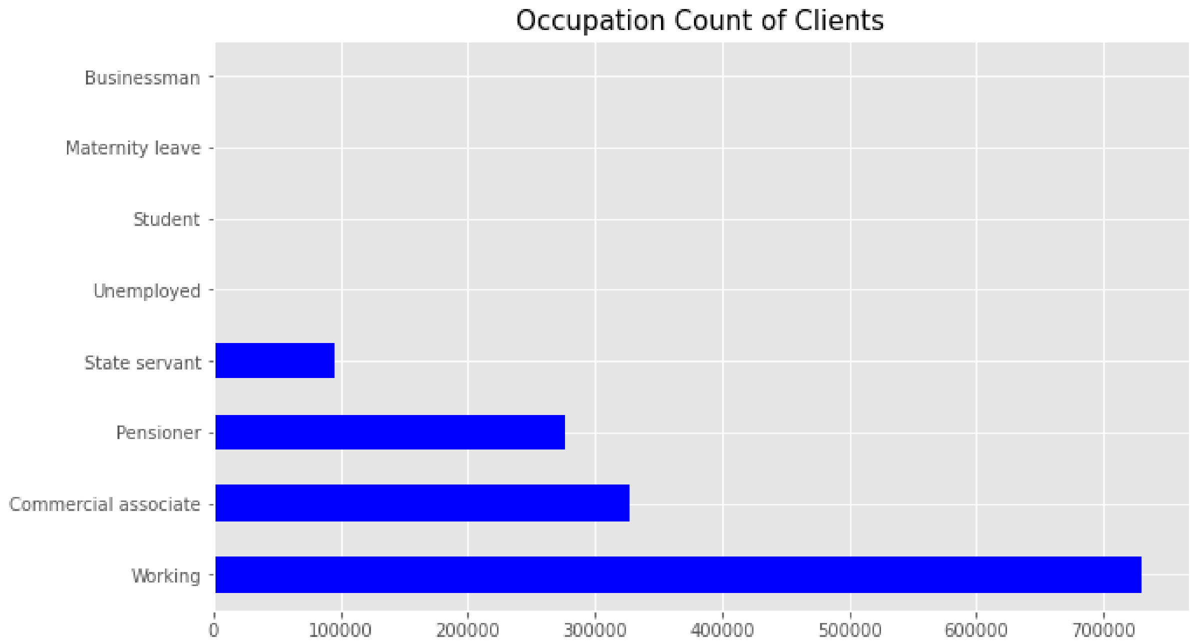
<td>Cash loans</td>

<td>F</td>

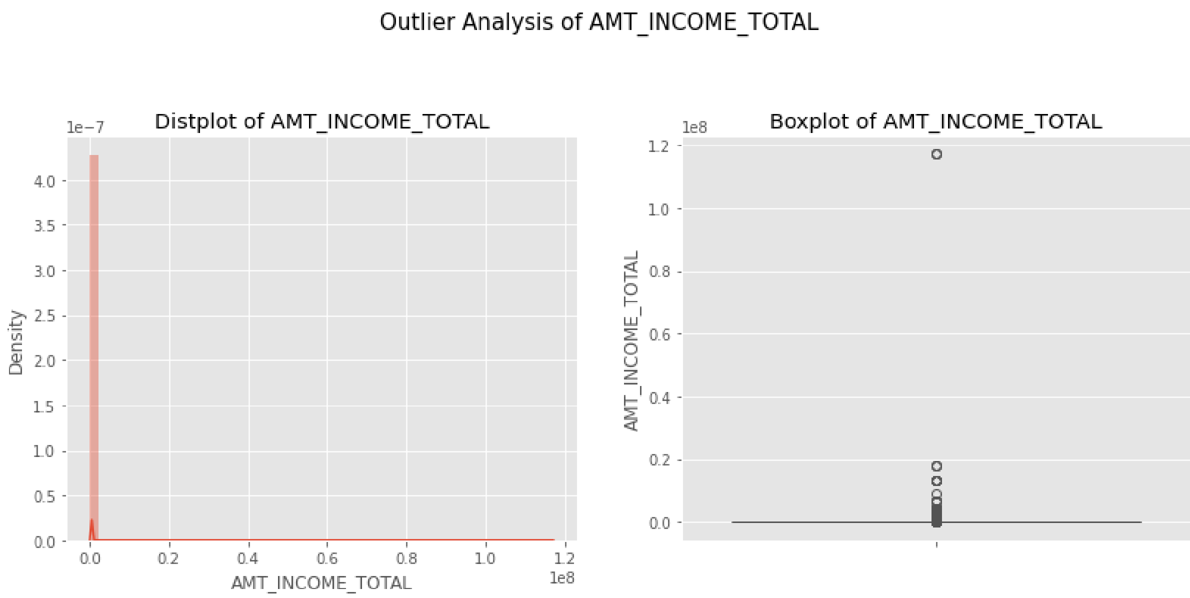
<td>N</td>

<td>N</td>

```
      <td>157500.0</td>
      <td>675000.0</td>
      <td>Commercial associate</td>
      <td>Higher education</td>
      <td>...</td>
      <td>XAP</td>
      <td>Repeater</td>
      <td>Computers</td>
      <td>POS</td>
      <td>XNA</td>
      <td>Country-wide</td>
      <td>20.0</td>
      <td>Connectivity</td>
      <td>high</td>
      <td>(157500, 167500]</td>
    </tr>
  </tbody>
</table>
<p>1430155 rows × 52 columns</p>
</div>
plt.figure(figsize=(10,6),facecolor='white')
dfnew['NAME_INCOME_TYPE'].value_counts().plot.barh(color='blue')
plt.title('Occupation Count of Clients',size=15)
plt.show()
```




```
!pip install seaborn

import seaborn as sns

#Outlier analysis of income

# create a function for outlier analysis

def outlier(column):
    plt.style.use('ggplot')
    plt.figure(figsize=(12,6))
    plt.subplot(1,2,1)
    sns.distplot(dfnew[column])
    plt.title('Distplot of'+' '+column)
    plt.subplot(1,2,2)
    sns.boxplot(y=dfnew[column])
    plt.title('Boxplot of'+' '+column)
    plt.suptitle('Outlier Analysis of'+' '+column ,size=15)
    plt.tight_layout(pad=3)
    plt.show()

#Analysis of AMT_INCOME_TOTAL

outlier('AMT_INCOME_TOTAL')

#got an understanding of existence of outliers
```



dfnew['AMT_INCOME_TOTAL'].describe() #statistical info

count   1.430155e+06

mean   1.736036e+05

std   1.983303e+05

min   2.565000e+04

25%   1.125000e+05

50%   1.575000e+05

75%   2.115000e+05

max   1.170000e+08

Name: AMT_INCOME_TOTAL, dtype: float64

dfnew['AMT_INCOME_TOTAL'].quantile([0.9,0.99,0.999,1.0]) #finding the percentile

0.900   270000.0

0.990   450000.0

0.999   900000.0

1.000   117000000.0

Name: AMT_INCOME_TOTAL, dtype: float64

so 99% of values are within 900000 range

#we need to check the data above 9lakh

dfnew[dfnew['AMT_INCOME_TOTAL']>900000]

<div>

<style scoped>

```
    .dataframe tbody tr th:only-of-type {
        vertical-align: middle;
    }

    .dataframe tbody tr th {
        vertical-align: top;
    }

    .dataframe thead th {
        text-align: right;
    }
</style>
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>SK_ID_CURR</th>
      <th>TARGET</th>
      <th>NAME_CONTRACT_TYPE</th>
      <th>CODE_GENDER</th>
      <th>FLAG_OWN_CAR</th>
      <th>FLAG_OWN_REALTY</th>
      <th>AMT_INCOME_TOTAL</th>
      <th>AMT_CREDIT</th>
      <th>NAME_INCOME_TYPE</th>
      <th>NAME_EDUCATION_TYPE</th>
      <th>...</th>
      <th>CODE_REJECT_REASON</th>
      <th>NAME_CLIENT_TYPE</th>
      <th>NAME_GOODS_CATEGORY</th>
      <th>NAME_PORTFOLIO</th>
```

```
    <th>NAME_PRODUCT_TYPE</th>
    <th>CHANNEL_TYPE</th>
    <th>SELLERPLACE_AREA</th>
    <th>NAME_SELLER_INDUSTRY</th>
    <th>NAME_YIELD_GROUP</th>
    <th>INCOME_RANGE</th>
  </tr>
</thead>
<tbody>
  <tr>
    <th>7390</th>
    <td>101769</td>
    <td>0</td>
    <td>Revolving loans</td>
    <td>M</td>
    <td>Y</td>
    <td>Y</td>
    <td>1080000.0</td>
    <td>180000.0</td>
    <td>Commercial associate</td>
    <td>Higher education</td>
    <td>...</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0.0</td>
    <td>0</td>
    <td>0</td>
```

```
  <td>(1080000, 1090000]</td>
</tr>
<tr>
  <th>8321</th>
  <td>102015</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>Y</td>
  <td>1935000.0</td>
  <td>269550.0</td>
  <td>Pensioner</td>
  <td>Secondary / secondary special</td>
  <td>...</td>
  <td>XAP</td>
  <td>New</td>
  <td>XNA</td>
  <td>Cash</td>
  <td>walk-in</td>
  <td>AP+ (Cash loan)</td>
  <td>50.0</td>
  <td>XNA</td>
  <td>low_normal</td>
  <td>(1935000, 1945000]</td>
</tr>
<tr>
  <th>16006</th>
  <td>103938</td>
  <td>0</td>
  <td>Cash loans</td>
```

```
  <td>F</td>
  <td>N</td>
  <td>N</td>
  <td>1350000.0</td>
  <td>2410380.0</td>
  <td>Commercial associate</td>
  <td>Higher education</td>
  <td>...</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0</td>
  <td>0.0</td>
  <td>0</td>
  <td>0</td>
  <td>(1350000, 1360000]</td>
</tr>
<tr>
  <th>21579</th>
  <td>105384</td>
  <td>0</td>
  <td>Revolving loans</td>
  <td>F</td>
  <td>Y</td>
  <td>Y</td>
  <td>1350000.0</td>
  <td>405000.0</td>
  <td>Commercial associate</td>
  <td>Higher education</td>
```

```
    <td>...</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0</td>
    <td>0.0</td>
    <td>0</td>
    <td>0</td>
    <td>(1350000, 1360000]</td>
  </tr>
  <tr>
    <th>26451</th>
    <td>106637</td>
    <td>0</td>
    <td>Cash loans</td>
    <td>M</td>
    <td>Y</td>
    <td>Y</td>
    <td>967500.0</td>
    <td>450000.0</td>
    <td>Commercial associate</td>
    <td>Higher education</td>
    <td>...</td>
    <td>XAP</td>
    <td>Repeater</td>
    <td>Mobile</td>
    <td>POS</td>
    <td>XNA</td>
    <td>Country-wide</td>
```

```
  <td>40.0</td>
  <td>Connectivity</td>
  <td>middle</td>
  <td>(967500, 977500]</td>
</tr>
<tr>
  <th>...</th>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
  <td>...</td>
</tr>
<tr>
  <th>1423829</th>
```

```
  <td>454746</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>Y</td>
  <td>Y</td>
  <td>949500.0</td>
  <td>735579.0</td>
  <td>Working</td>
  <td>Higher education</td>
  <td>...</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>XNA</td>
  <td>XNA</td>
  <td>Credit and cash offices</td>
  <td>-1.0</td>
  <td>XNA</td>
  <td>XNA</td>
  <td>(949500, 959500]</td>
</tr>
<tr>
  <th>1423830</th>
  <td>454746</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>Y</td>
  <td>Y</td>
  <td>949500.0</td>
```

```
  <td>735579.0</td>
  <td>Working</td>
  <td>Higher education</td>
  <td>...</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>XNA</td>
  <td>XNA</td>
  <td>XNA</td>
  <td>Credit and cash offices</td>
  <td>-1.0</td>
  <td>XNA</td>
  <td>XNA</td>
  <td>(949500, 959500]</td>
</tr>
<tr>
  <th>1423831</th>
  <td>454746</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>M</td>
  <td>Y</td>
  <td>Y</td>
  <td>949500.0</td>
  <td>735579.0</td>
  <td>Working</td>
  <td>Higher education</td>
  <td>...</td>
  <td>XAP</td>
  <td>Repeater</td>
  <td>Clothing and Accessories</td>
```

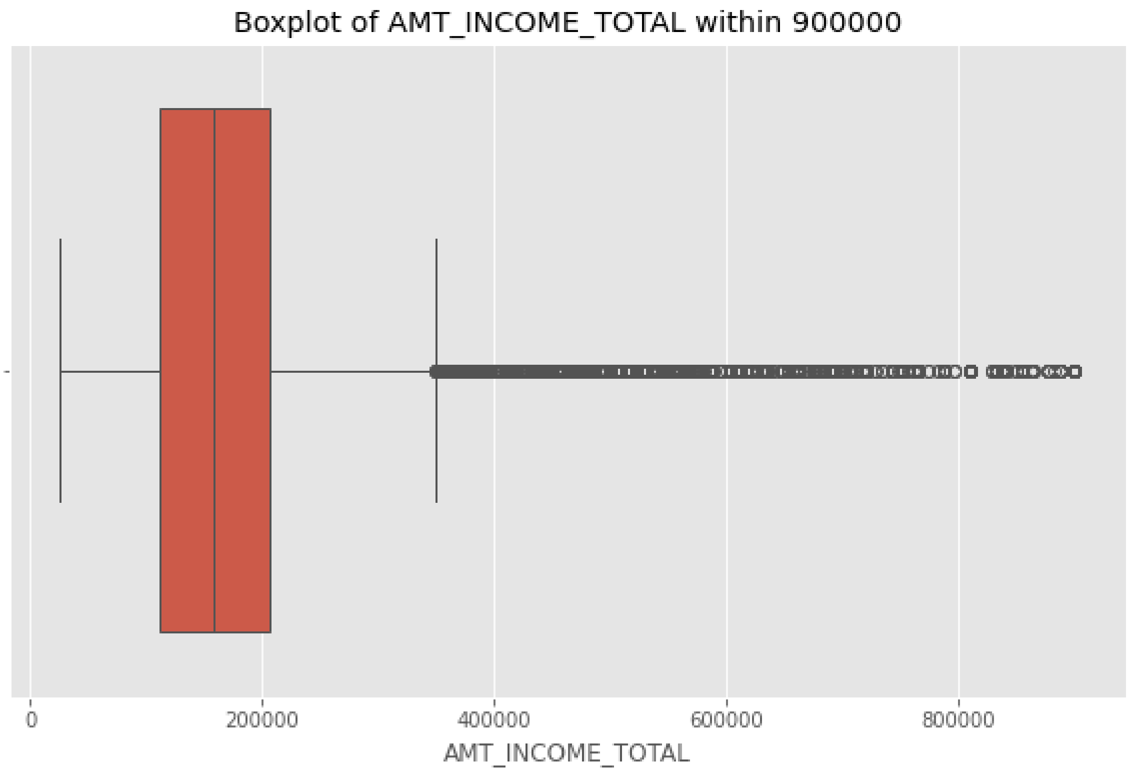
```
  <td>POS</td>
  <td>XNA</td>
  <td>Country-wide</td>
  <td>100.0</td>
  <td>Clothing</td>
  <td>middle</td>
  <td>(949500, 959500]</td>
</tr>
<tr>
  <th>1424300</th>
  <td>454864</td>
  <td>0</td>
  <td>Cash loans</td>
  <td>F</td>
  <td>N</td>
  <td>N</td>
  <td>936000.0</td>
  <td>1014493.5</td>
  <td>Commercial associate</td>
  <td>Higher education</td>
  <td>...</td>
  <td>XAP</td>
  <td>New</td>
  <td>XNA</td>
  <td>Cards</td>
  <td>walk-in</td>
  <td>Country-wide</td>
  <td>200.0</td>
  <td>Connectivity</td>
  <td>XNA</td>
  <td>(936000, 946000]</td>
```

```
    </tr>
    <tr>
      <th>1424301</th>
      <td>454864</td>
      <td>0</td>
      <td>Cash loans</td>
      <td>F</td>
      <td>N</td>
      <td>N</td>
      <td>936000.0</td>
      <td>1014493.5</td>
      <td>Commercial associate</td>
      <td>Higher education</td>
      <td>...</td>
      <td>XAP</td>
      <td>New</td>
      <td>Furniture</td>
      <td>POS</td>
      <td>XNA</td>
      <td>Regional / Local</td>
      <td>50.0</td>
      <td>Furniture</td>
      <td>middle</td>
      <td>(936000, 946000]</td>
    </tr>
  </tbody>
</table>
<p>882 rows × 52 columns</p>
</div>
#plot boxplot for below 9lakh cases
plt.figure(figsize=(10,6))
```

sns.boxplot(x=dfnew[dfnew['AMT_INCOME_TOTAL']<=900000]['AMT_INCOME_TOTAL'])

plt.title("Boxplot of AMT_INCOME_TOTAL within 900000")

plt.show()



#Most values lies between 1lakh and 2.5lakh.Also 99% of data lies below 9lakh. so we can consider anyvalue above 900000 are outliers

Analysis of AMT_CREDIT

dfnew.AMT_CREDIT.describe()

count   1.430155e+06

mean   5.893386e+05

std   3.874204e+05

min   4.500000e+04

25%   2.700000e+05

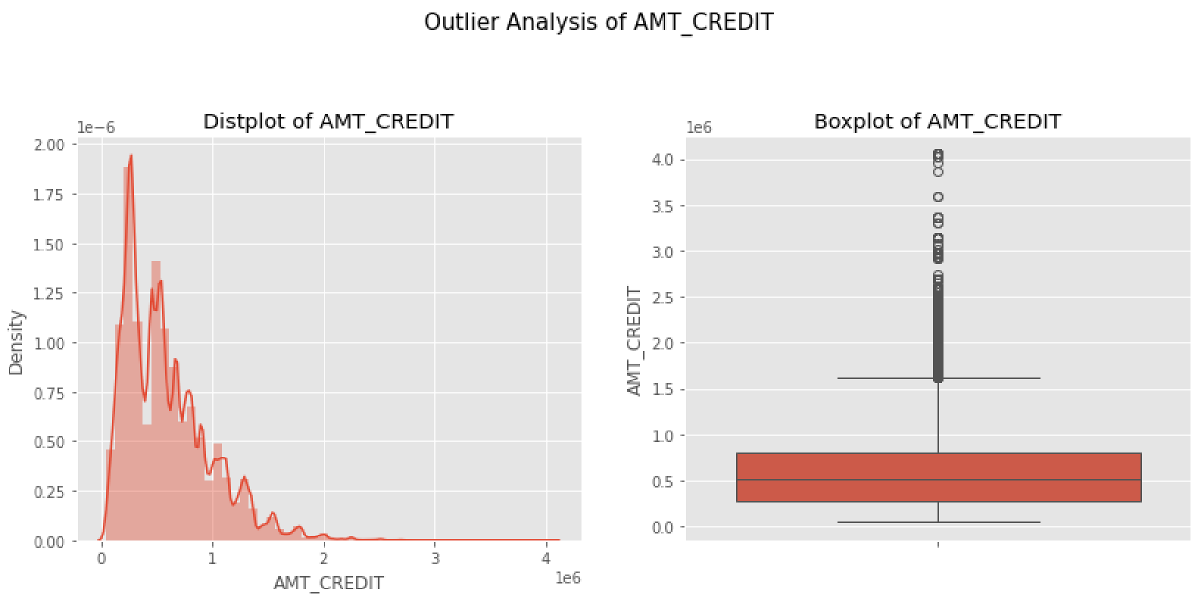
50%   5.084955e+05

75%   8.086500e+05

max   4.050000e+06

Name: AMT_CREDIT, dtype: float64

outlier('AMT_CREDIT')



#Most of the values are lying between 200000 and 800000. above 1.6M outliers are visible.

#Analysis of AGE

dfnew.PERSON_AGE.describe()

count 1.430155e+06

mean 4.419713e+01

std 1.190810e+01

min 2.000000e+01

25% 3.400000e+01

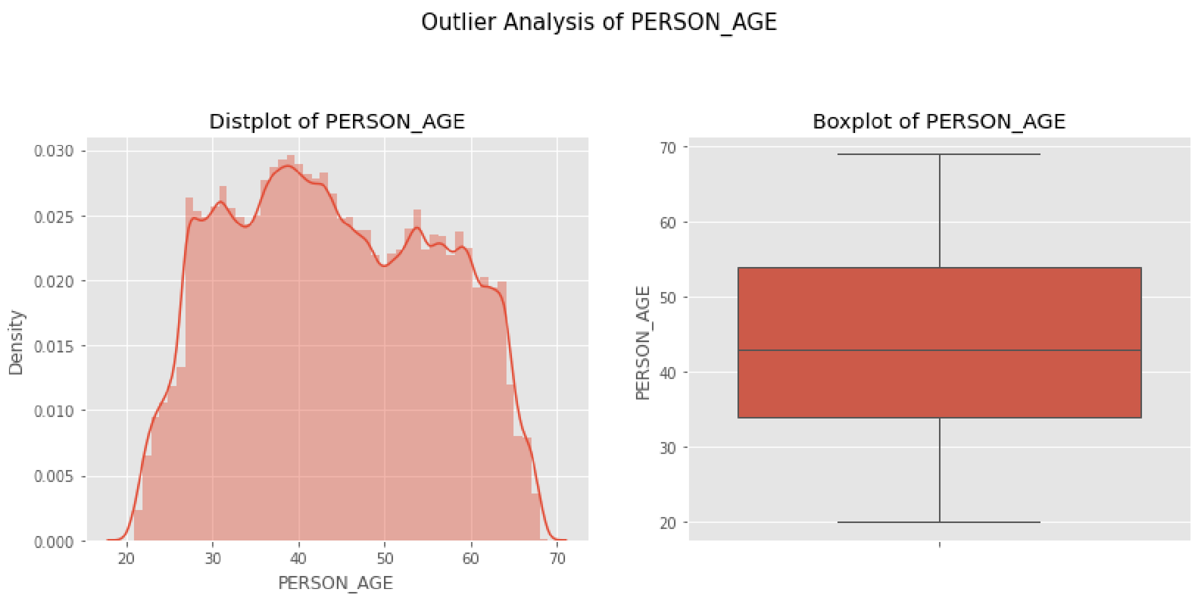
50% 4.300000e+01

75% 5.400000e+01

max 6.900000e+01

Name: PERSON_AGE, dtype: float64

outlier('PERSON_AGE')



#There are no outliers and mostof the applicants are in between 35 and 55 years of age

#Analysis of YEARS_EMPLOYED

dfnew.YEARS_EMPLOYED.describe()

count 1.430155e+06

mean 1.982652e+02

std 3.924261e+02

min 0.000000e+00

25% 2.000000e+00

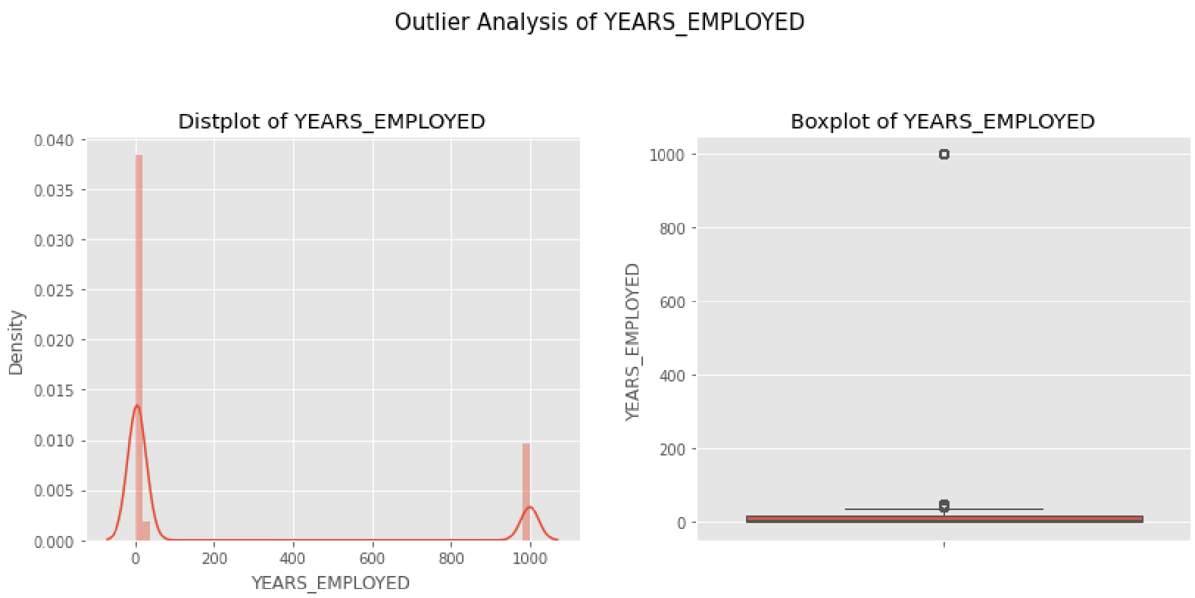
50% 6.000000e+00

75% 1.700000e+01

max 1.000000e+03

Name: YEARS_EMPLOYED, dtype: float64

outlier('YEARS_EMPLOYED')



#OBservation : there are outliers

#Check how many values lies above 600

dfnew[dfnew.YEARS_EMPLOYED > 600].shape

(276368, 52)

#there are values at 1000 yrs of employment which is not logical.lets ignore outliers

sns.boxplot(y=dfnew[dfnew['YEARS_EMPLOYED']< 600]['YEARS_EMPLOYED'])

plt.title("Boxplot of YEARS_EMPLOYED less than 600")

plt.show()

#Observation: Most of the values lies within 10 years of employment

#Categorical Analysis- Bucketing age data

dfnew['AGE_GROUP']=pd.cut(dfnew.PERSON_AGE,[0,20,30,40,50,60,999],labels=['0-20','20-30','30-40','40-50','50-60','60+'])

dfnew.AGE_GROUP.value_counts(normalize=True)*100

30-40   26.573903

40-50   24.624534

50-60   22.663348

20-30   15.316592

60+   10.821345

0-20   0.000280

Name: AGE_GROUP, dtype: float64

#create buckets for YEARS_EMPLOYED

dfnew['YEARS_OF_EMPLOYEMENT']=pd.cut(dfnew.YEARS_EMPLOYED,bins=[0,5,10,15,20,25,30,35,40,9999],labels=['0-5','5-10','10-15','15-20','20-25','25-30','30-35','35-40','40 & above'])

dfnew.YEARS_OF_EMPLOYEMENT.value_counts(normalize=True)*100

0-5   42.604137

5-10   21.048733

40 & above   20.994422

10-15   8.240442

15-20   3.424295

20-25   1.869285

25-30   0.995922

30-35   0.604613

35-40   0.218151

Name: YEARS_OF_EMPLOYEMENT, dtype: float64

#Create buckets for AMT_INCOME_TOTAL

dfnew['AMT_INCOME_RANGE']=pd.qcut(dfnew.AMT_INCOME_TOTAL,q=[0,0.2,0.5,.75,.95,1], labels = ['VERY_LOW','LOW','MEDIUM','HIGH','VERY_HIGH'])

dfnew.AMT_INCOME_RANGE.value_counts(normalize=True)*100

VERY_LOW   29.098524

LOW   27.065947

HIGH   19.993917

MEDIUM   19.161489

VERY_HIGH   4.680122

Name: AMT_INCOME_RANGE, dtype: float64

#Analysis of Target variable with Bar graph

plt.figure(figsize=(10,6))

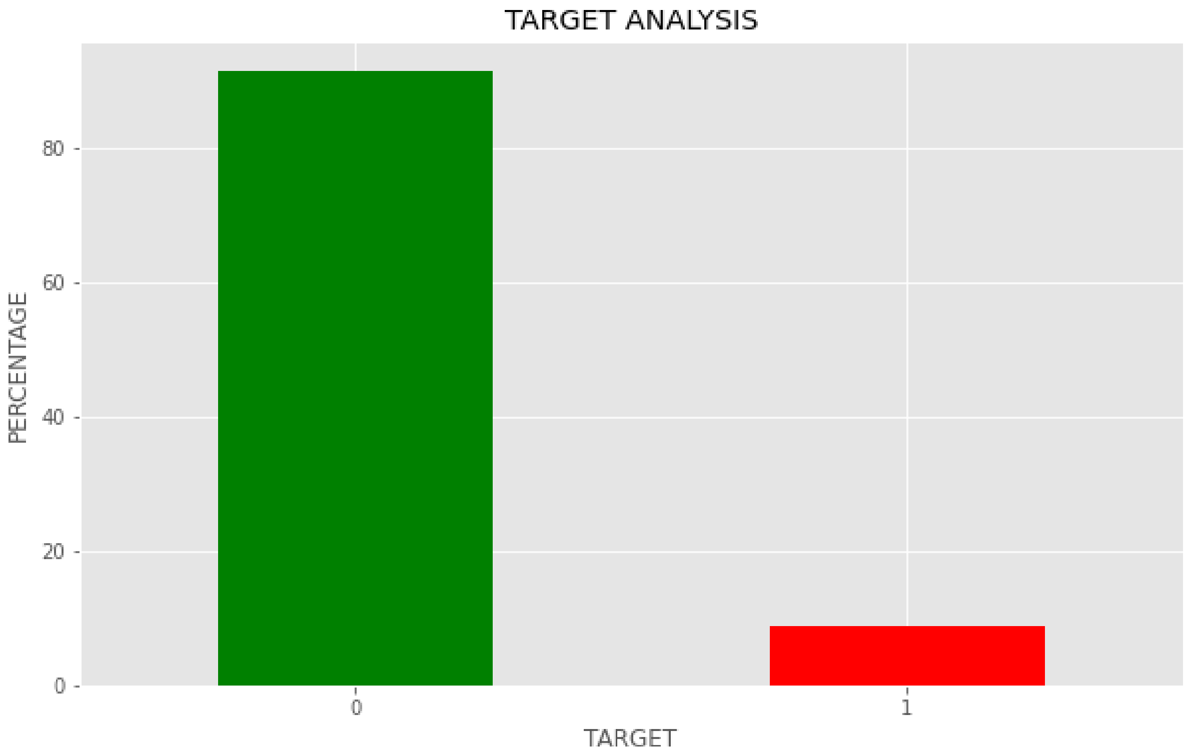
(dfnew.TARGET.value_counts(normalize=True)*100).plot.bar(title='TARGET ANALYSIS',color=['Green','Red'])

plt.xlabel('TARGET')

plt.ylabel('PERCENTAGE')

plt.xticks(rotation=0)

plt.show()



#from the chart it's clear that more than 80% of loan applicants are non defaulters and there are defaulters below 20%

#There is a data imbalance

For finding imbalace ratio split dataframe dfnew to 2,one with Target=1 and other with Target=0

#Create a new df with Target=1

TARGET_1=dfnew[dfnew.TARGET==1]

Create df for Target=0

TARGET_0=dfnew[dfnew.TARGET==0]

#check IMBALANCE

imbalace_ratio=len(TARGET_0)/len(TARGET_1)

imbalace_ratio

10.595224582455003

#the imbalance ratio is 10.59

#UNIVARIATE ANALYSIS

Define function for unsegmented univariate Analysis

def uni_var(column):

plt.figure(figsize=(13,6))

sns.countplot(data=dfnew,x=column)

plt.title("Univariate Analysis of"+ ' '+column)

plt.xticks(rotation=90)

Category_cols=['NAME_CONTRACT_TYPE', 'CODE_GENDER', 'FLAG_OWN_CAR', 'FLAG_OWN_REALTY',

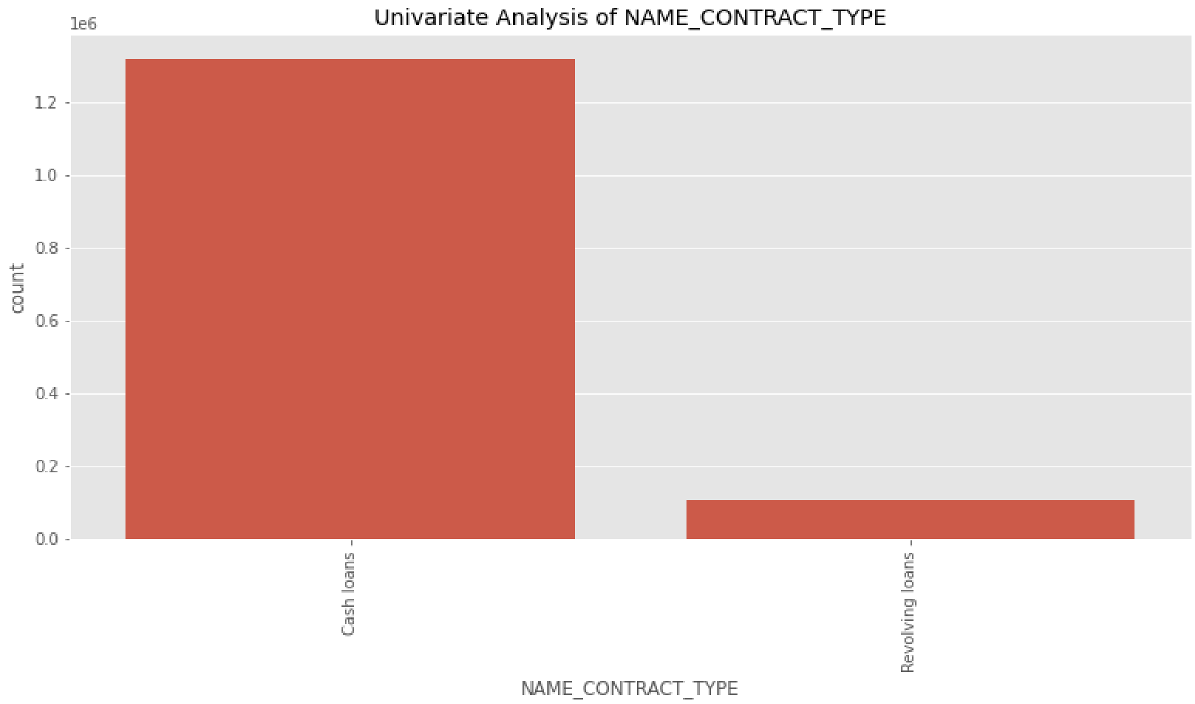
'NAME_INCOME_TYPE', 'NAME_EDUCATION_TYPE',

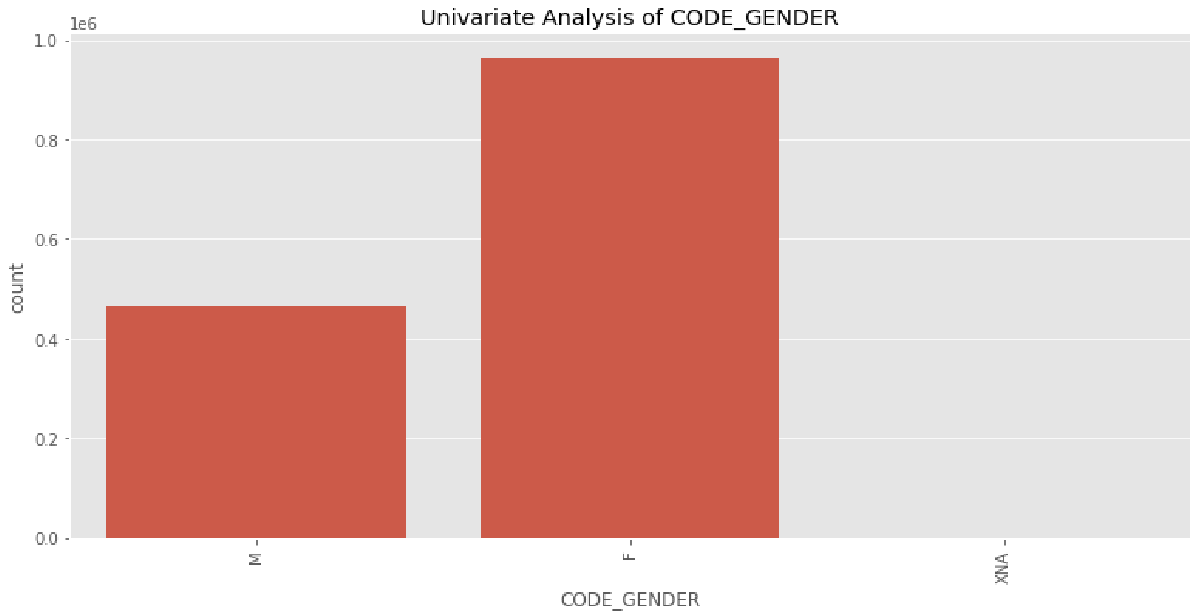
'NAME_FAMILY_STATUS', 'NAME_HOUSING_TYPE','ORGANIZATION_TYPE', 'AGE_GROUP',

'YEARS_OF_EMPLOYEMENT','AMT_INCOME_RANGE']

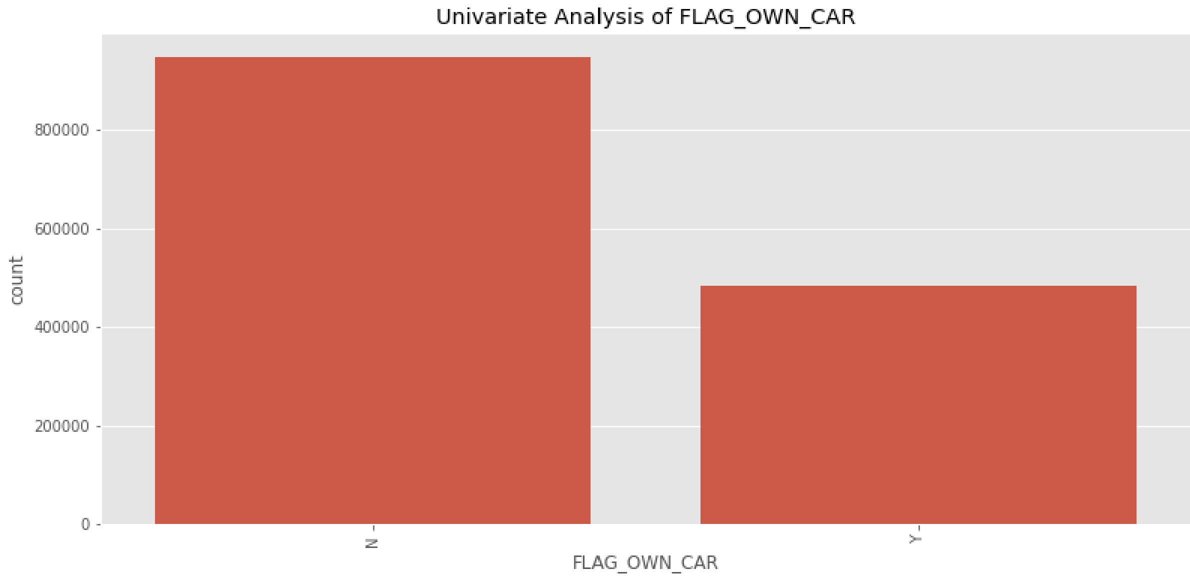
for i in Category_cols:

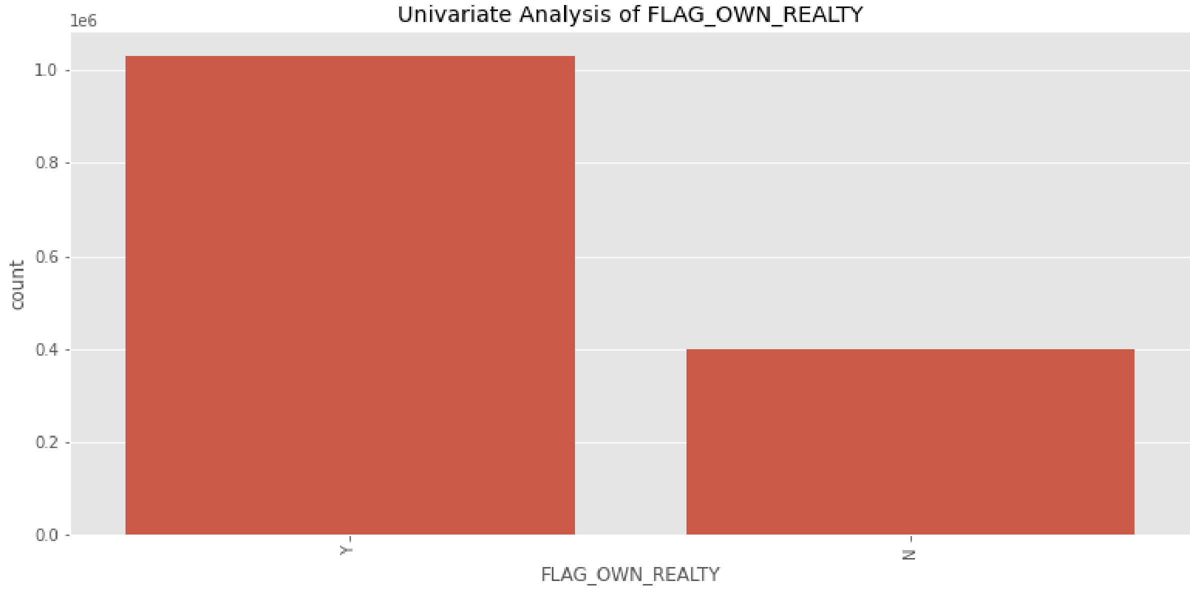
uni_var(i)

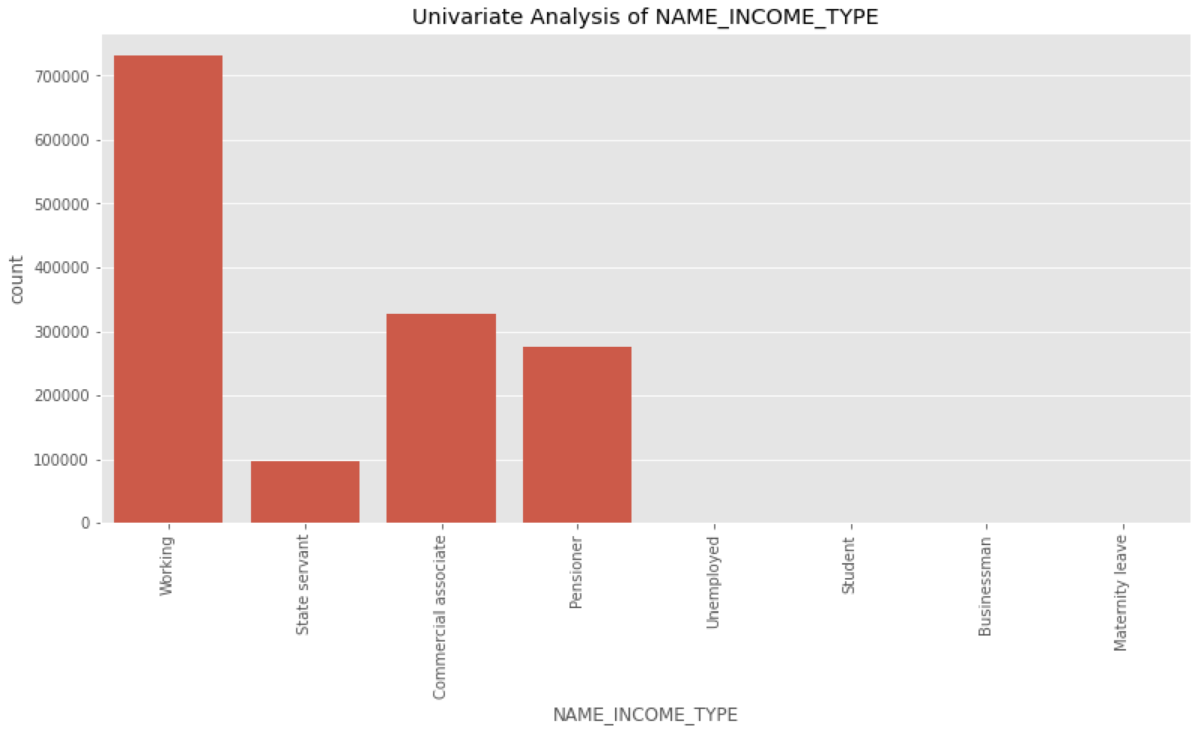


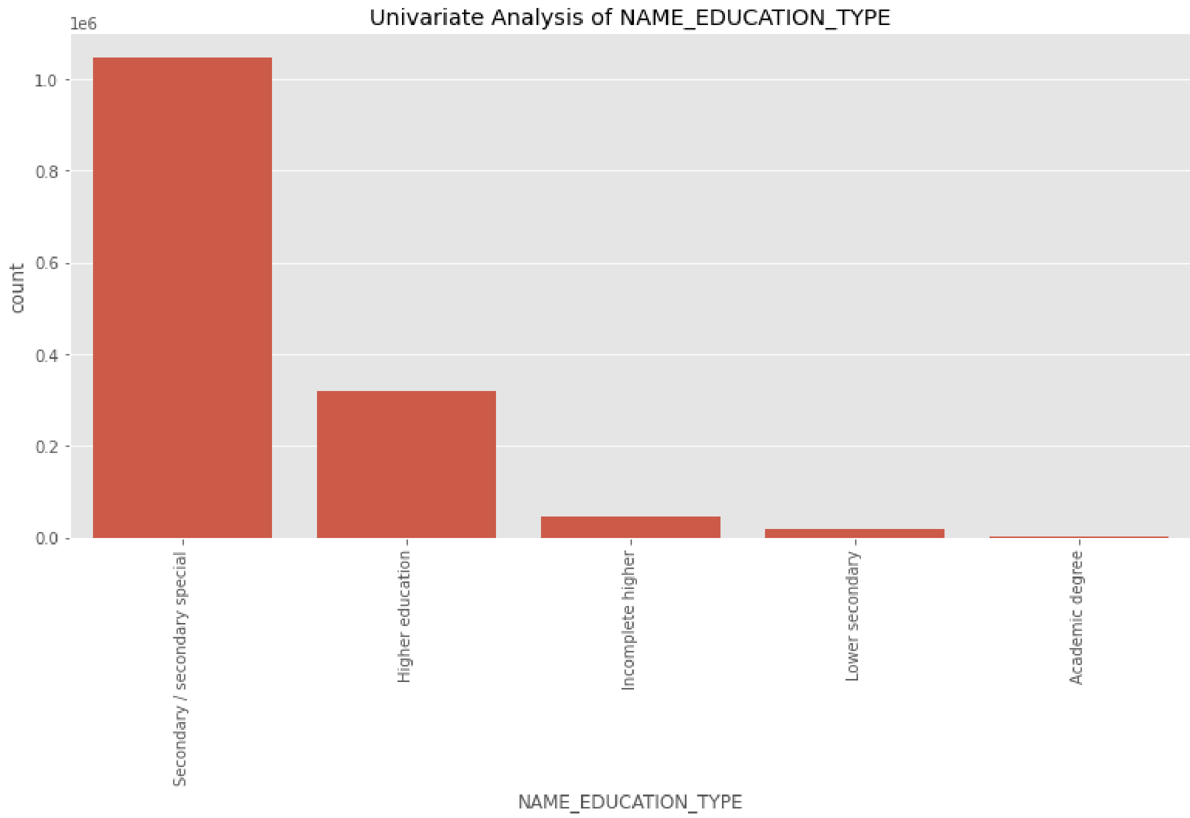


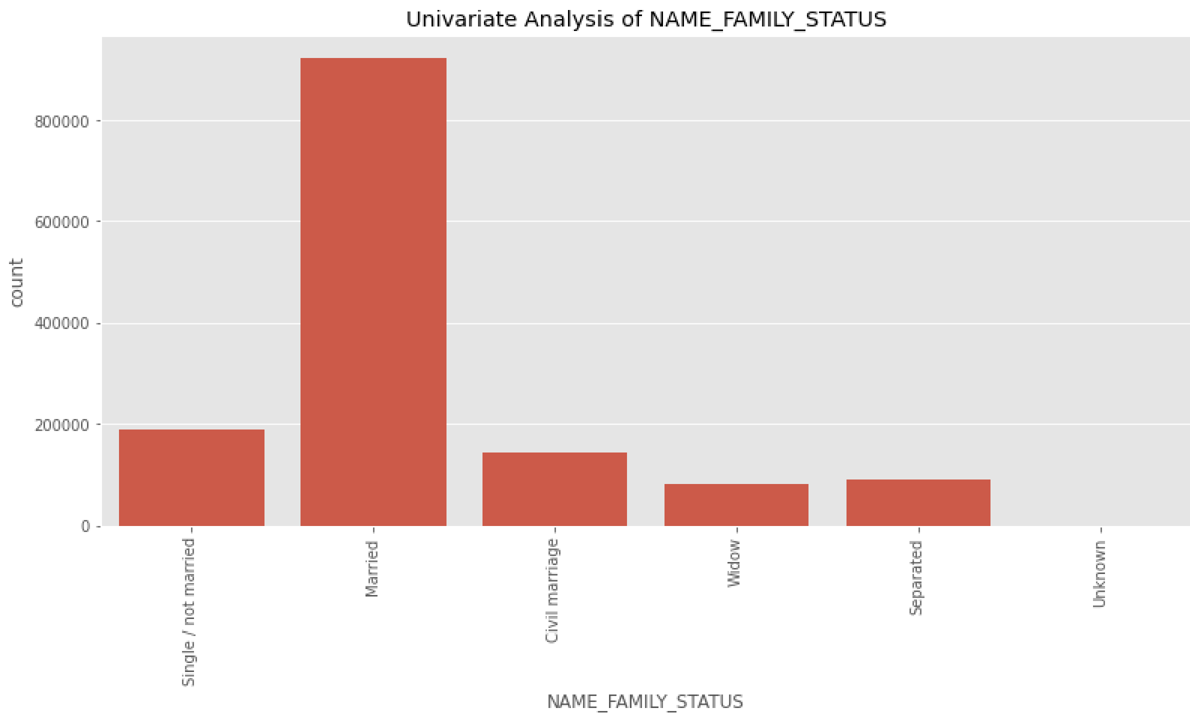
Univariate Analysis of CODE_GENDER
1e6
1.0
0.8
0.6
0.4
0.2
0.0
count
M
F
XNA
CODE_GENDER

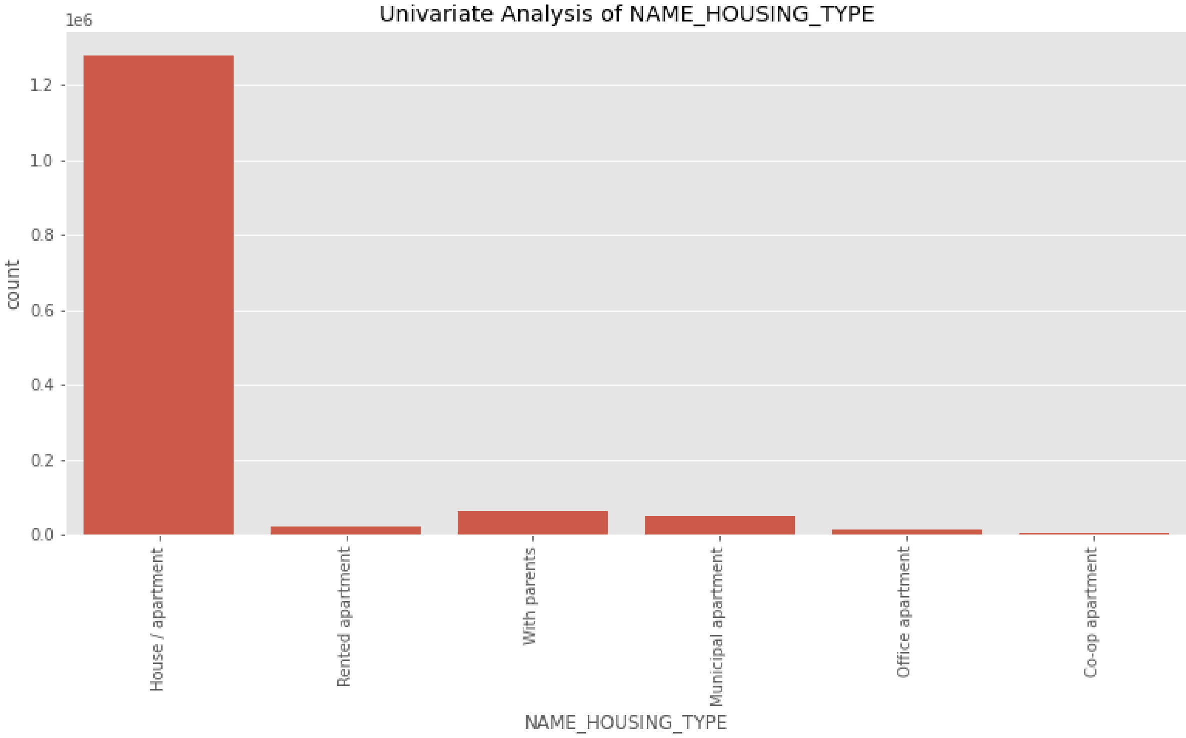

Univariate Analysis of FLAG_OWN_CAR
count
800000
600000
400000
200000
0
N
Y
FLAG_OWN_CAR

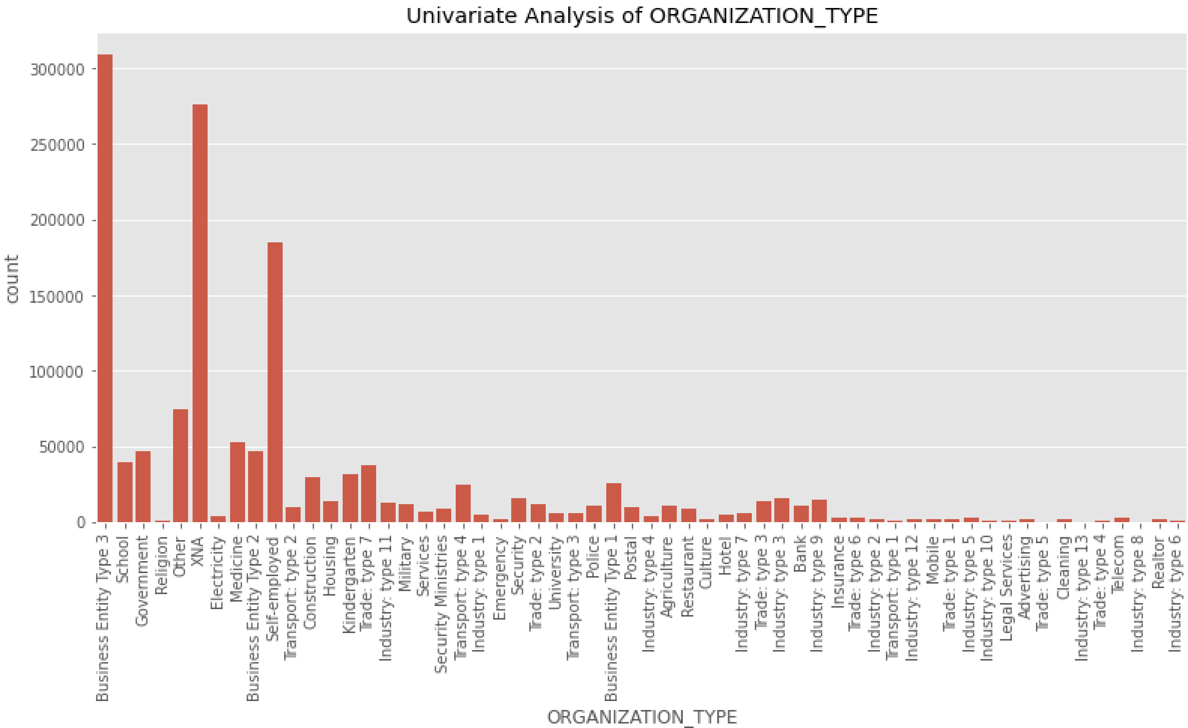

Univariate Analysis of FLAG_OWN_REALTY
1e6
count
1.0
0.8
0.6
0.4
0.2
0.0
Y
N
FLAG_OWN_REALTY

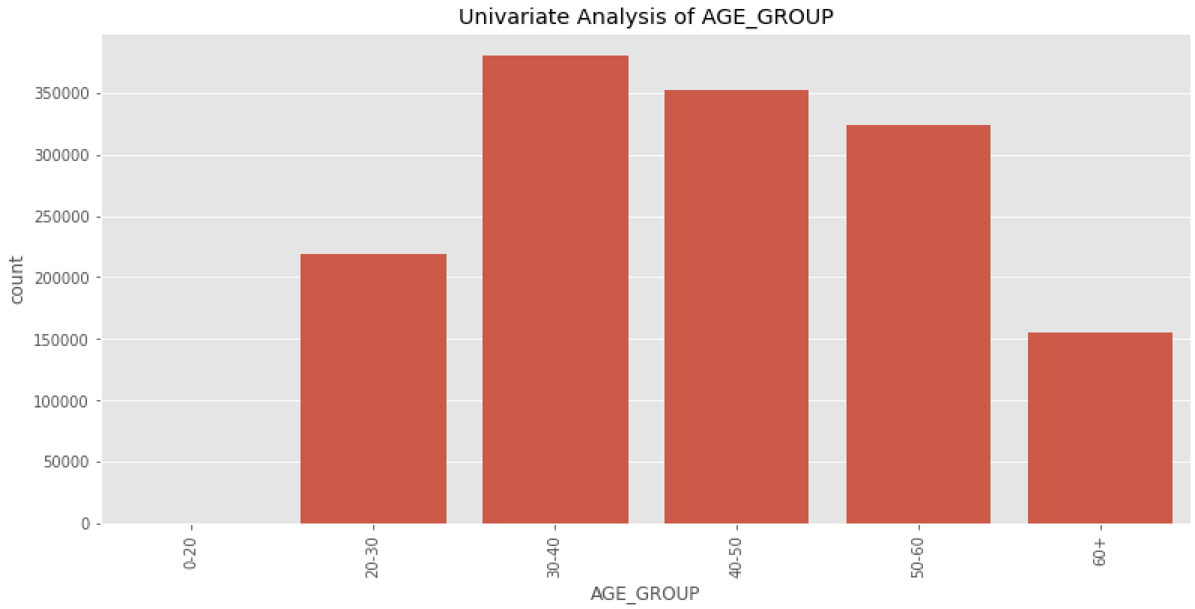

Univariate Analysis of NAME_INCOME_TYPE
count
700000
600000
500000
400000
300000
200000
100000
0
Working
State servant
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave
NAME_INCOME_TYPE

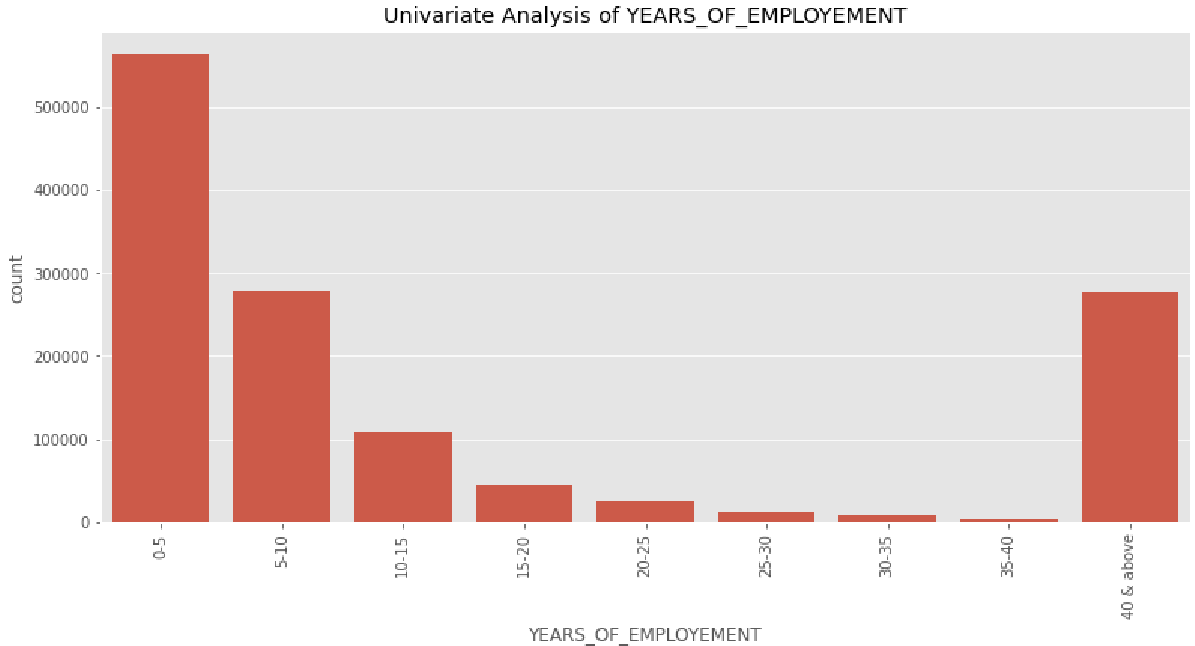

Univariate Analysis of NAME_EDUCATION_TYPE
1e6
1.0
0.8
0.6
0.4
0.2
0.0
count
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

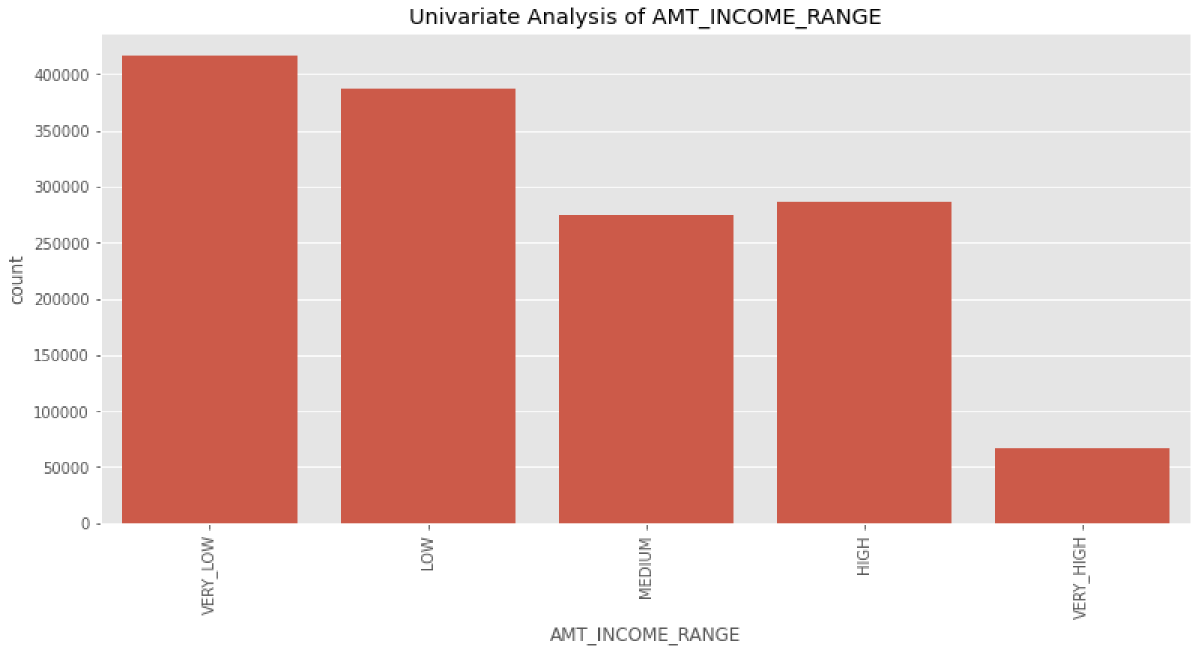

Univariate Analysis of NAME_FAMILY_STATUS
800000
600000
400000
200000
0
count
Single / not married
Married
Civil marriage
Widow
Separated
Unknown
NAME_FAMILY_STATUS


Univariate Analysis of NAME_HOUSING_TYPE
1e6
count
1.2
1.0
0.8
0.6
0.4
0.2
0.0
House / apartment
Rented apartment
With parents
Municipal apartment
Office apartment
Co-op apartment
NAME_HOUSING_TYPE


Univariate Analysis of ORGANIZATION_TYPE
count
300000
250000
200000
150000
100000
50000
0
Business Entity Type 3
School
Government
Religion
Other
XNA
Electricity
Medicine
Business Entity Type 2
Self-employed
Transport: type 2
Construction
Housing
Kindergarten
Trade: type 7
Industry: type 11
Military
Services
Security Ministries
Transport: type 4
Industry: type 1
Emergency
Security
Trade: type 2
University
Transport: type 3
Police
Business Entity Type 1
Postal
Industry: type 4
Agriculture
Restaurant
Culture
Hotel
Industry: type 7
Trade: type 3
Industry: type 3
Bank
Industry: type 9
Insurance
Trade: type 6
Industry: type 2
Transport: type 1
Industry: type 12
Mobile
Trade: type 1
Industry: type 5
Industry: type 10
Legal Services
Advertising
Trade: type 5
Cleaning
Industry: type 13
Trade: type 4
Telecom
Industry: type 8
Realtor
Industry: type 6
ORGANIZATION_TYPE


Univariate Analysis of AGE_GROUP
count
350000
300000
250000
200000
150000
100000
50000
0
0-20
20-30
30-40
40-50
50-60
60+
AGE_GROUP


Univariate Analysis of YEARS_OF_EMPLOYEMENT
count
500000
400000
300000
200000
100000
0
0-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40 & above
YEARS_OF_EMPLOYEMENT



Observations:

1.People rely more on cash loans than revolving loans

2. The applicants are more female than male

3. Most of the applicants doesn't own a car

4. But most of them own a property

5. Working class has highest percentage of loans

6. Majority of applicants with secondary education and people with high educatio have applied less

7. Most of the loan applicants are Married

8. Business entity needs more money

9. Majority of the applicants are in middle aged group

#10. People with less employment experience are more applied for loans

#11. Low and medium income people are more for loan application

```
#segmented univariate analysis

plt.style.use('default')

%matplotlib inline

plt.style.use('default')

%matplotlib inline

# define function for countplot

def univar_count(column):

    plt.figure(figsize=(13,6),facecolor='white')
```

plt.rcParams["axes.labelsize"]=12

plt.subplot(1,2,1)

sns.countplot(data=TARGET_0,x=column,order=sorted(TARGET_0[column].value_counts().index,reverse=True),palette='flare')

plt.title("Ontime Paying Clients")

plt.xticks(rotation=90)

plt.subplot(1,2,2)

sns.countplot(data=TARGET_1,x=column,order=sorted(TARGET_1[column].value_counts().index,reverse=True),palette='flare')

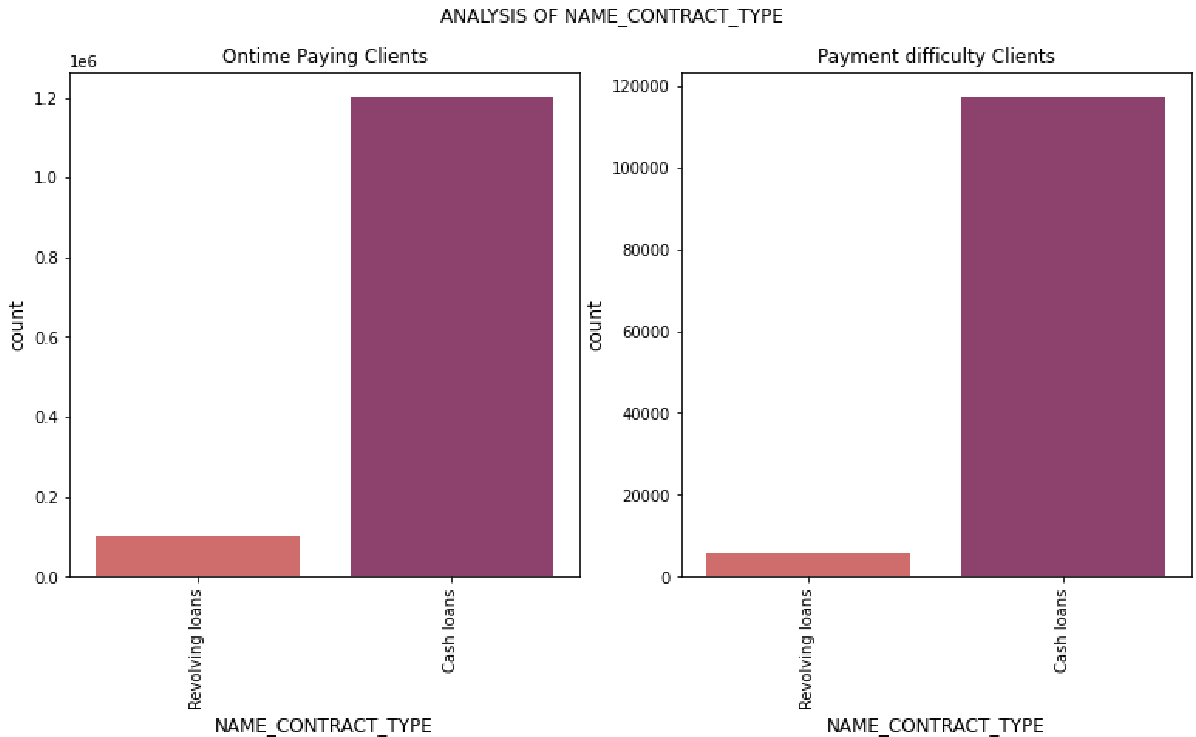
plt.title("Payment difficulty Clients")

plt.xticks(rotation=90)

plt.suptitle('ANALYSIS OF'+' '+ column)

#Analysis of NAME_CONTRACT_TYPE

univar_count('NAME_CONTRACT_TYPE')

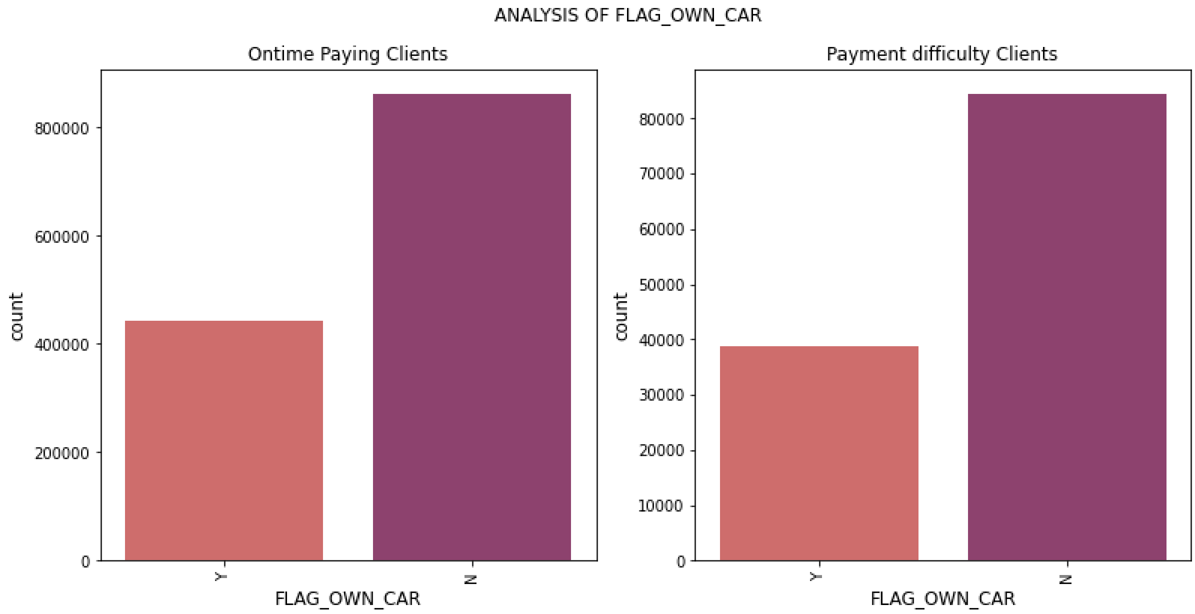


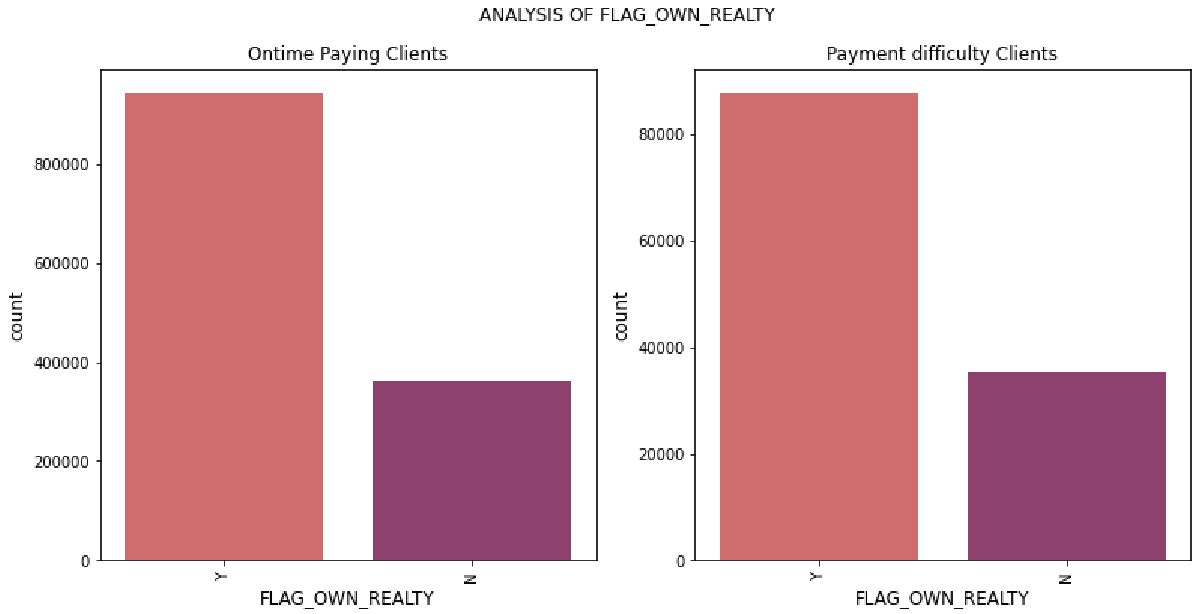
#observation: both ontime payers and default payers took cash loans than revolving loans

#Analysis of FLAG_OWN_CAR & FLAG_OWN_REALTY

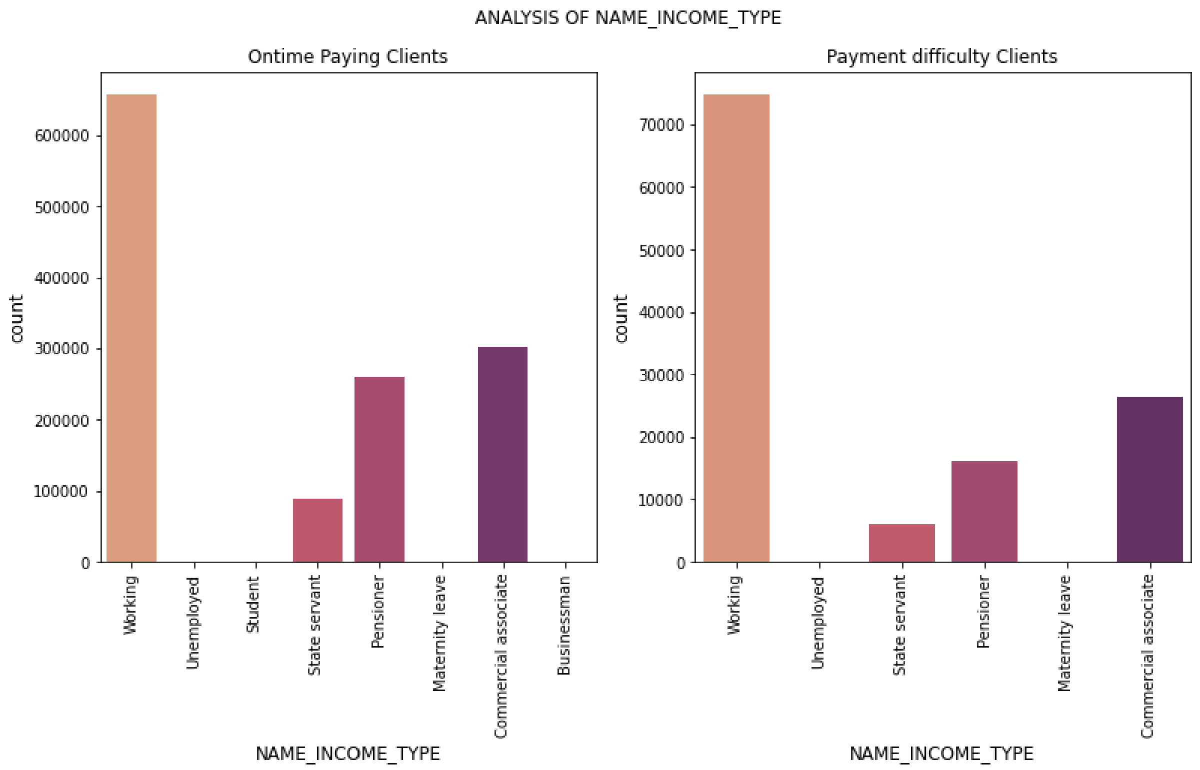
var=['FLAG_OWN_CAR','FLAG_OWN_REALTY']

```
for i in var:
    univar_count(i)
```






```
univar_count('NAME_INCOME_TYPE')
```



#Observation: Working Class holds high place in both categories

#Pensioners and Commercial associate tend to pay loans on-time

#Students and Businessman also have no payment difficulties so bank can target them in future

#Analysis of NAME_EDUCATION_TYPE

plt.figure(figsize=(13,6),facecolor='white')

plt.rcParams["axes.labelsize"]=12

plt.subplot(1,2,1)

TARGET_0['NAME_EDUCATION_TYPE'].value_counts().plot.bar(color=['Green','Yellow','Orange','Red',' Black'])

plt.title("Ontime Paying Clients")

plt.xticks(rotation=90)

plt.xlabel('NAME_EDUCATION_TYPE')

plt.ylabel('Count')

plt.subplot(1,2,2)

TARGET_1["NAME_EDUCATION_TYPE"].value_counts().plot.bar(color=['Green','Yellow','Orange','Red' ,'Black'])

plt.title("Payment difficulty Clients")

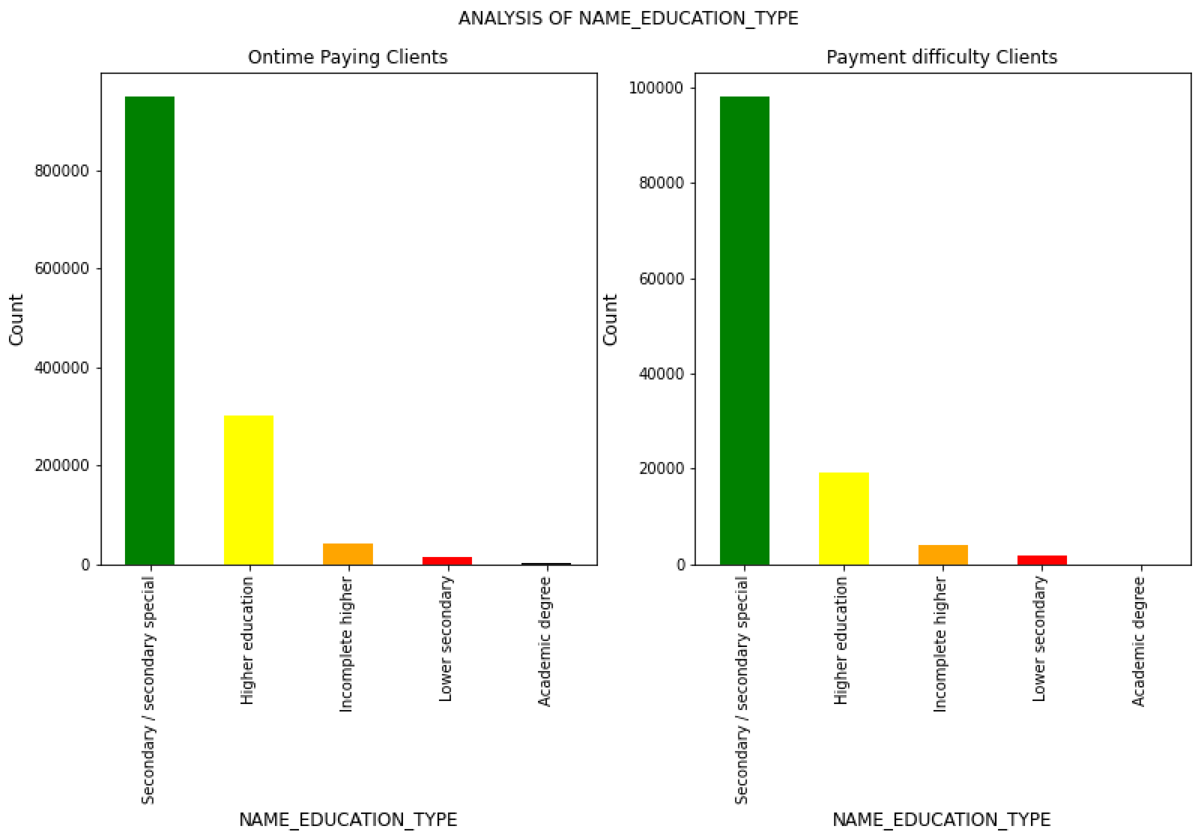
plt.xticks(rotation=90)

plt.xlabel('NAME_EDUCATION_TYPE')

plt.ylabel('Count')

plt.suptitle('ANALYSIS OF NAME_EDUCATION_TYPE')

plt.show()



#observations :

Secondary/secondary special is high in both categories

#Customers with Higher Education and Academic degree have higher ontime payments than defaulters and this can be because they are settled than others

#Analysis of NAME_FAMILY_STATUS

plt.figure(figsize=(13,6),facecolor='white')

plt.subplot(1,2,1)

TARGET_0['NAME_FAMILY_STATUS'].value_counts().plot.pie(autopct='%1.1f%%')

plt.title("Clients without Payment Difficulties")

plt.xticks(rotation=90)

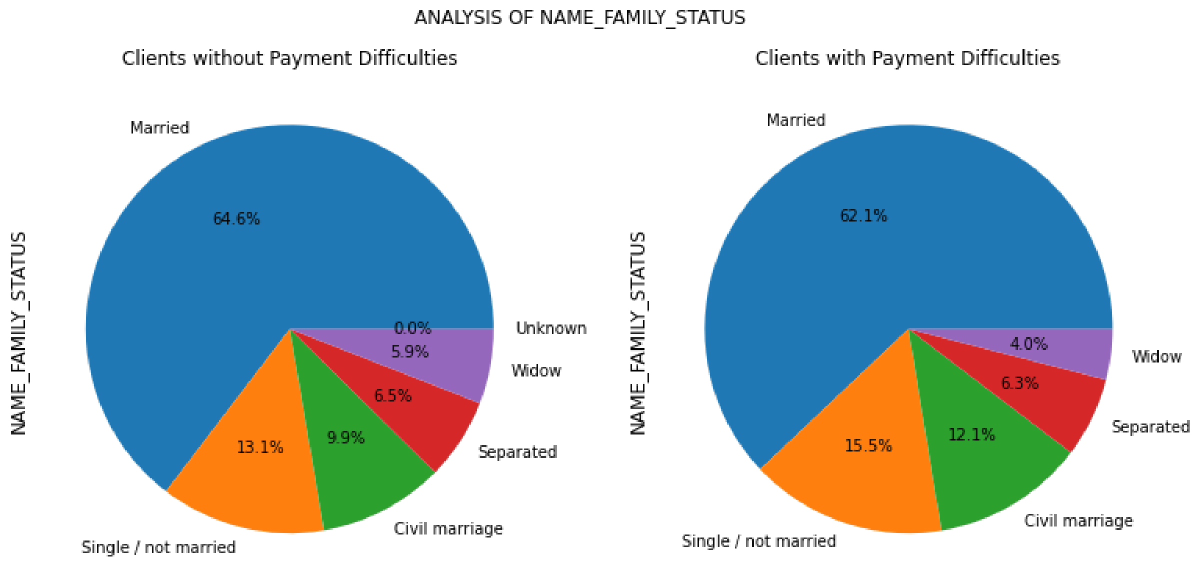
plt.subplot(1,2,2)

TARGET_1['NAME_FAMILY_STATUS'].value_counts().plot.pie(autopct='%1.1f%%')

plt.title("Clients with Payment Difficulties") plt.xticks(rotation=90)

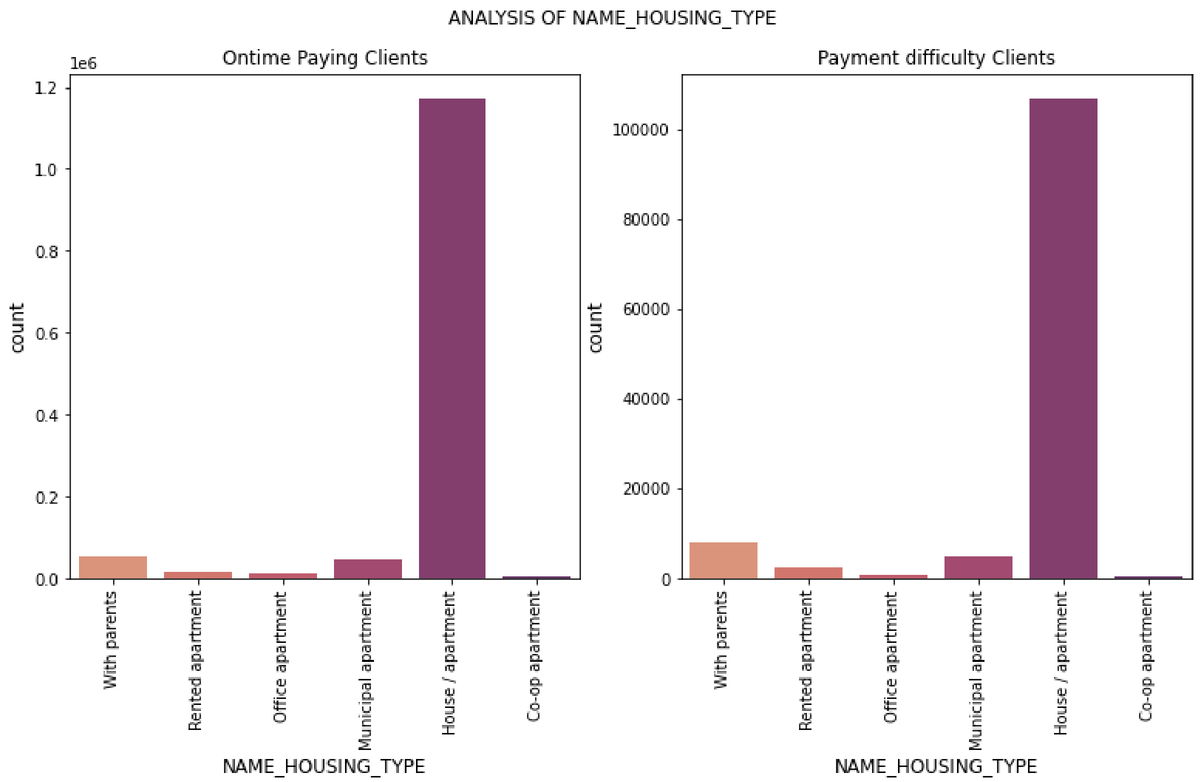
plt.suptitle('ANALYSIS OF NAME_FAMILY_STATUS')

plt.show()



#Analysis of NAME_HOUSING_TYPE

univar_count('NAME_HOUSING_TYPE')



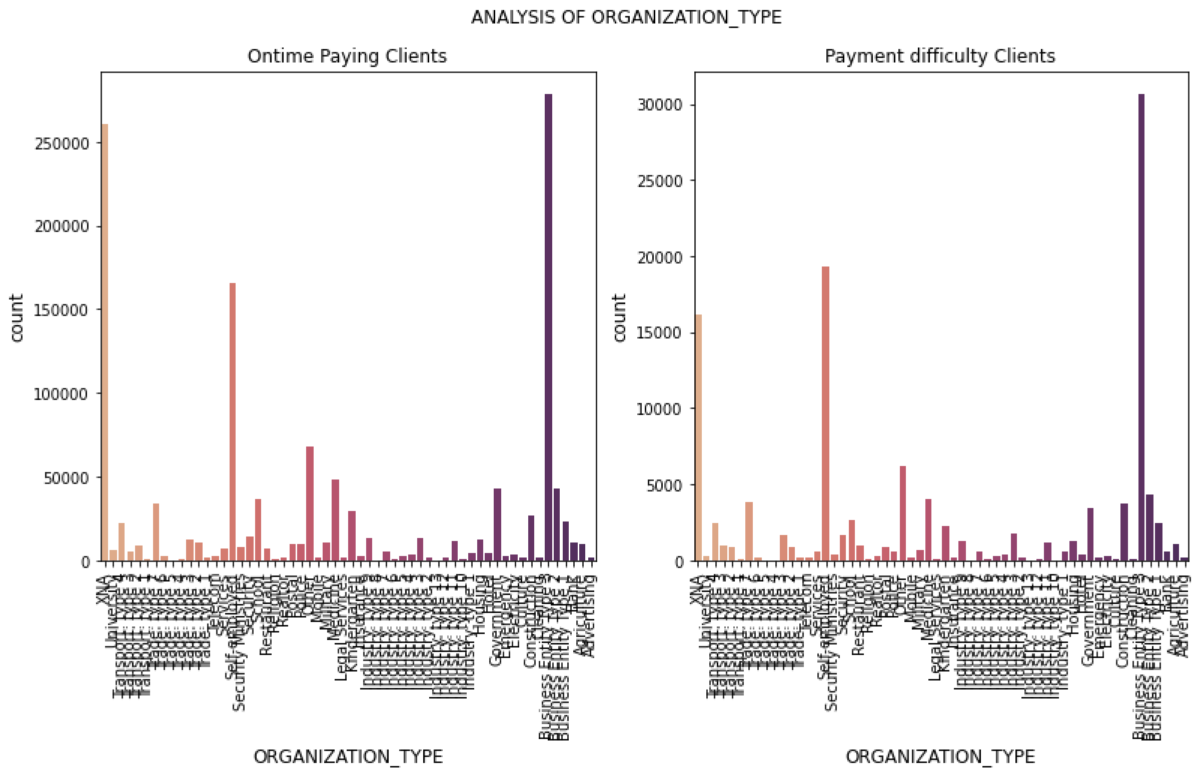
Observation:

Clients with house/Apartment is high in both categories.This can be because for without payment difficulties they already have house and liability is less and for defaulters it is be because they may have housing loan also.Also this is contrary to FLAG OWN REALTY analysis

#When compared both category people with Rented Apartments and with parents have more payment difficulties than other.It may be because they have more liabilites than others

#Analysis of ORGANIZATION_TYPE

univar_count('ORGANIZATION_TYPE')

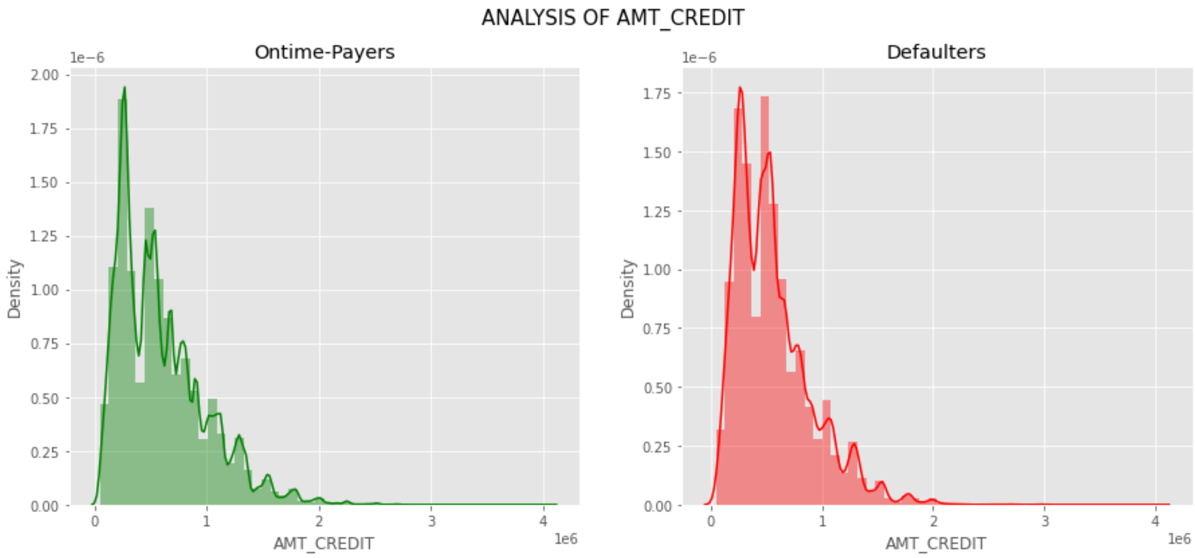


#Define function for easy analysis

```
def uni_num(column):
   plt.style.use('ggplot')
   plt.figure(figsize=(15,6))
   plt.subplot(1,2,1)
   sns.distplot(TARGET_0[column],color="green")
   plt.title('Ontime-Payers')
   plt.subplot(1,2,2)
   sns.distplot(TARGET_1[column],color="red")
   plt.title('Defaulters')
   plt.suptitle('ANALYSIS OF'+' '+ column,size=15)
   plt.show()
```

#Analysis of AMT_CREDIT

uni_num('AMT_CREDIT')



The graph shows the presence of outliers in both

#Approximately from 3 to 6 lakh there are more clients with difficulty in payments

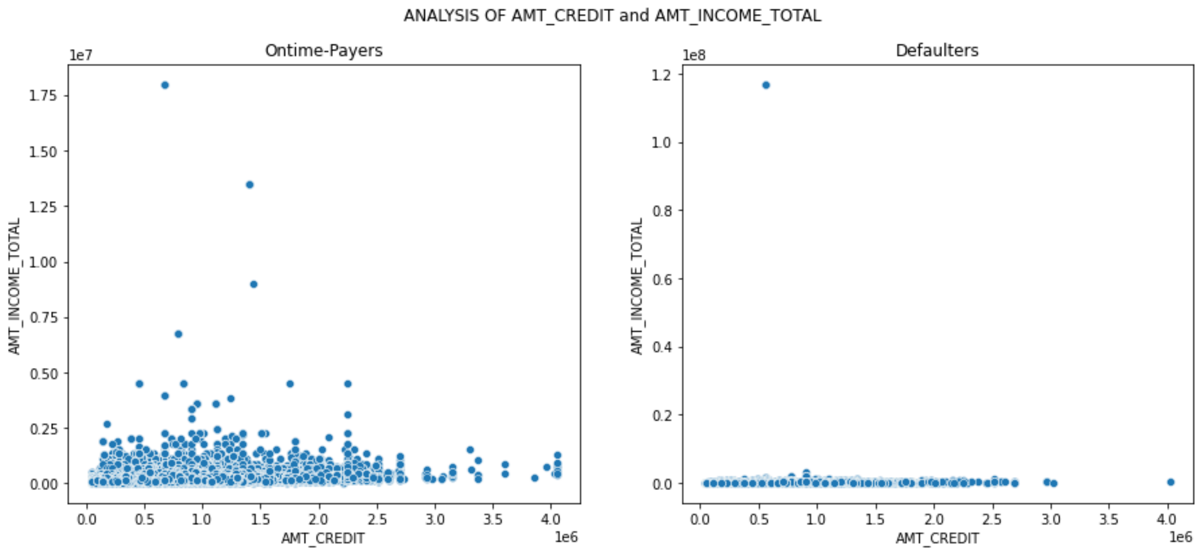
#BIVARIATE ANALYSIS

Define function for easy access for Analysis

```
plt.style.use('default')

%matplotlib inline

def scatter_plot(column1,column2):

    plt.figure(figsize=(15,6),facecolor='white')

    plt.subplot(1,2,1)

    sns.scatterplot(data=TARGET_0,x=column1,y=column2)

    plt.title('Ontime-Payers')

    plt.subplot(1,2,2)

    sns.scatterplot(data=TARGET_1,x=column1,y=column2)

    plt.title('Defaulters')

    plt.suptitle('ANALYSIS OF'+' '+ column1+' and '+column2)

    plt.show()
```

Anaiysis of AMT_CREDIT & AMT_INCOME_TOTAL

scatter_plot('AMT_CREDIT','AMT_INCOME_TOTAL')



#presence of outliers noticed

Plot AMT_CREDIT,AMT_INCOME_TOTAL

plt.figure(figsize=(15,8),facecolor='white')

plt.subplot(1,2,1)

sns.scatterplot(data=TARGET_0,x=TARGET_0[TARGET_0.AMT_INCOME_TOTAL < 337500].AMT_INCOME_TOTAL,y=TARGET_0[TARGET_0.AMT_CREDIT < 1620000].AMT_CREDIT)

plt.title('Ontime-Payers')

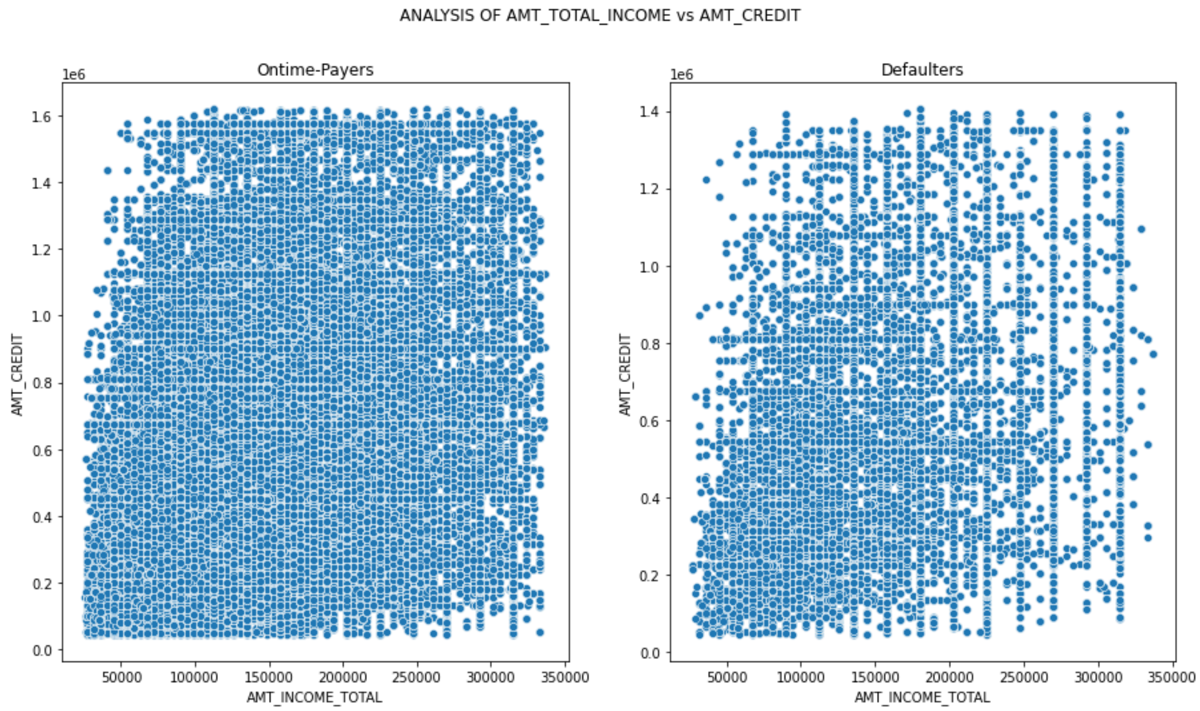
plt.subplot(1,2,2)

sns.scatterplot(data=TARGET_1,x=TARGET_1[TARGET_1.AMT_INCOME_TOTAL < 337500].AMT_INCOME_TOTAL,y=TARGET_1[TARGET_1.AMT_CREDIT < 1406688].AMT_CREDIT)

plt.title('Defaulters')

plt.suptitle('ANALYSIS OF AMT_TOTAL_INCOME vs AMT_CREDIT')

plt.show()



Observation:

#For Ontime payers its densely packed in all regions,but for Defaulters its densely packed in lower income-lower credit regions.

#So most of the defaulters are low income clients and bank should taken this to consideration

Multivariate Analysis

Analysing AMT_INCOME_TOTAL,AMT_CREDIT,PERSON_AGE,YEARS_EMPLOYED for TARGET_0

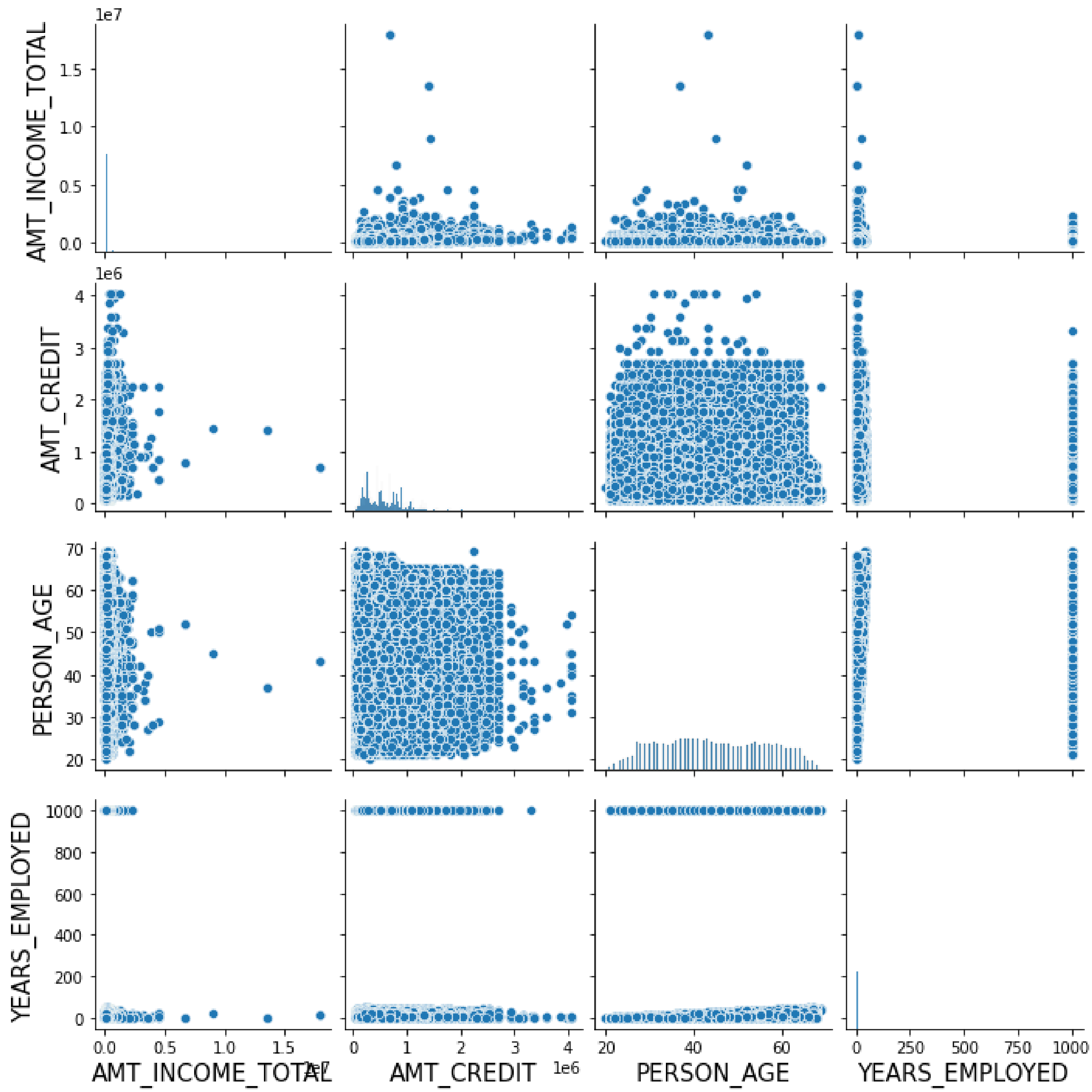
plt.figure(figsize=(21,6),facecolor='white')

plt.rc('xtick', labelsize=10)

plt.rc('ytick', labelsize=10)

sns.pairplot(TARGET_0[['AMT_INCOME_TOTAL','AMT_CREDIT','PERSON_AGE','YEARS_EMPLOYED']])

plt.show()



#Bivariate Analysis

Using function for easy acess of variables

def numcat_bivar(column1,column2):

    plt.figure(figsize=(21,6),facecolor='white')

    plt.rc('xtick', labelsize=12)

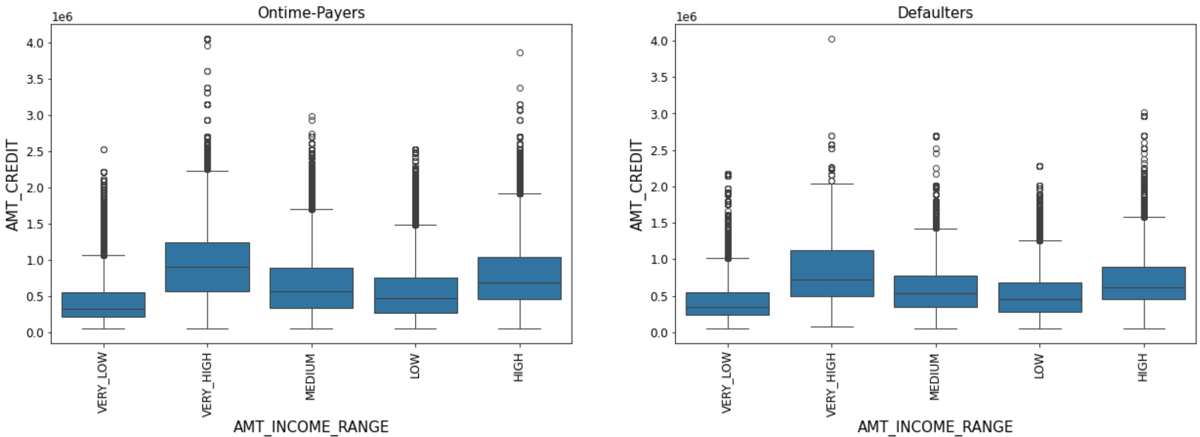
    plt.rc('ytick', labelsize=12)

    plt.rcParams['axes.labelsize']=15

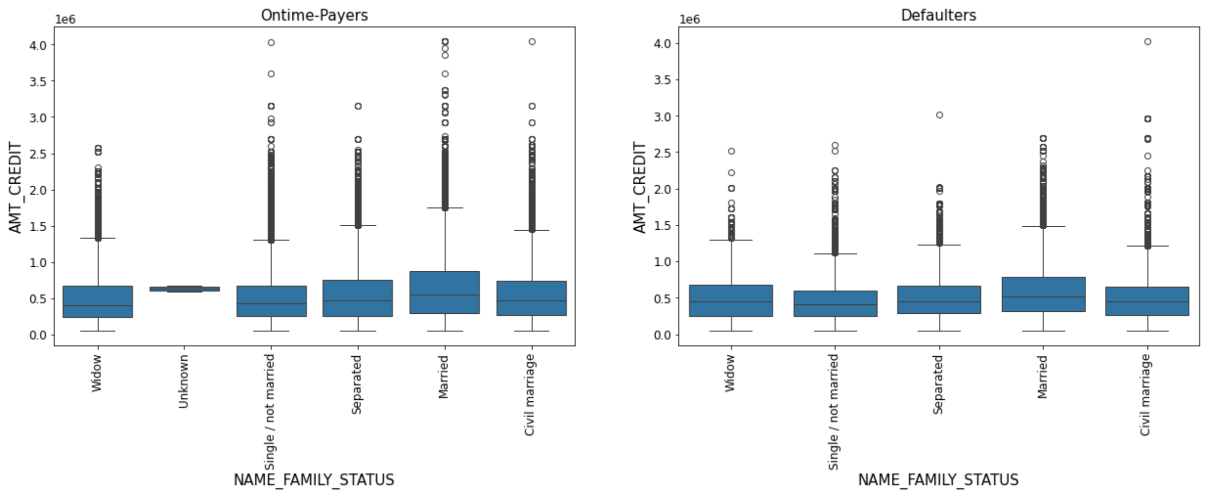
    plt.rcParams

    plt.subplot(1,2,1)

```
sns.boxplot(data=TARGET_0,x=column1,y=column2,order=sorted(TARGET_0[column1].value_counts(
).index,reverse=True))
    plt.title('Ontime-Payers',size=15)
    plt.xticks(rotation=90)
    plt.subplot(1,2,2)

sns.boxplot(data=TARGET_1,x=column1,y=column2,order=sorted(TARGET_1[column1].value_counts(
).index,reverse=True))
    plt.title('Defaulters',size=15)
    plt.xticks(rotation=90)
    plt.show()
# Analysis of NAME_EDUCATION_TYPE,AMT_CREDIT
#plt.figure(figsize=(5,8),facecolor='white')
numcat_bivar('AMT_INCOME_RANGE','AMT_CREDIT')
```

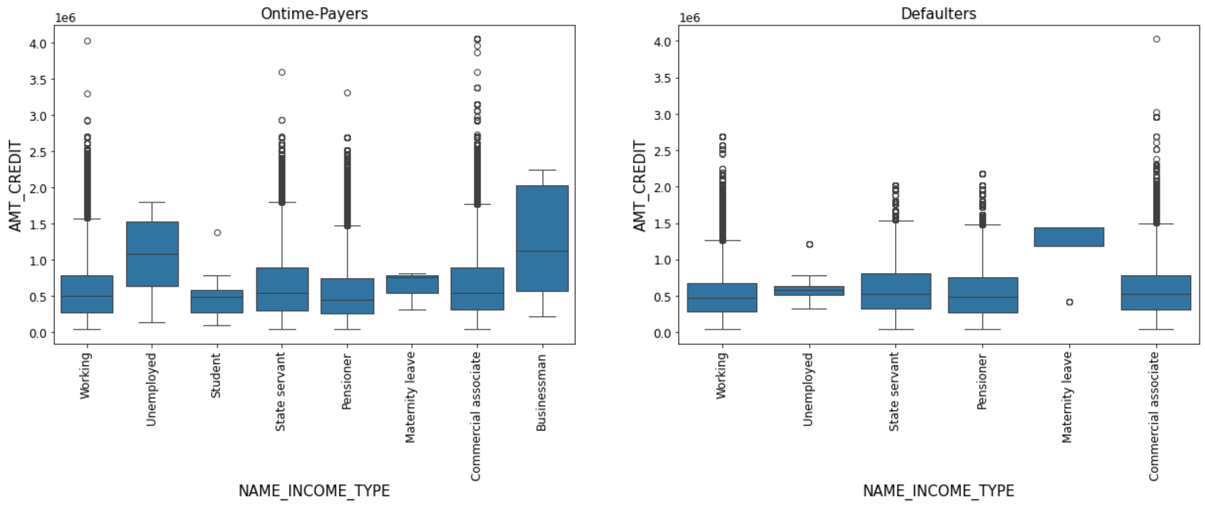



```
#High and Very_High got more credit amount than lower income categories and their presence is
high in Ontime payers category
numcat_bivar('NAME_FAMILY_STATUS','AMT_CREDIT')
```



#Married follows by Separated got higher credits has no payment difficulties

numcat_bivar('NAME_INCOME_TYPE','AMT_CREDIT')

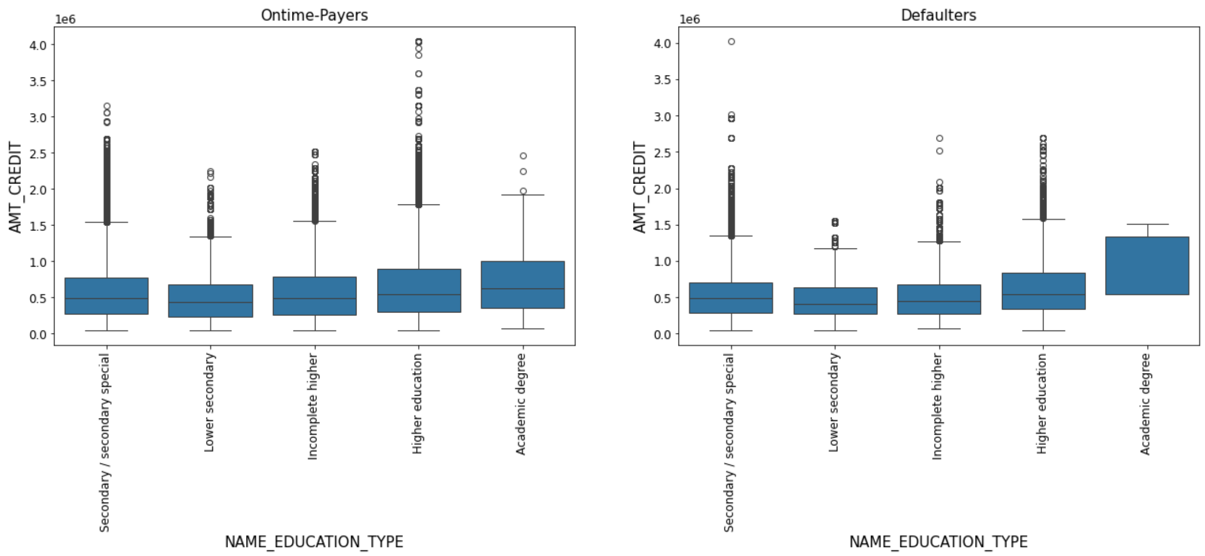


#Bussinessman who got high loan amount tends to pay it ontime

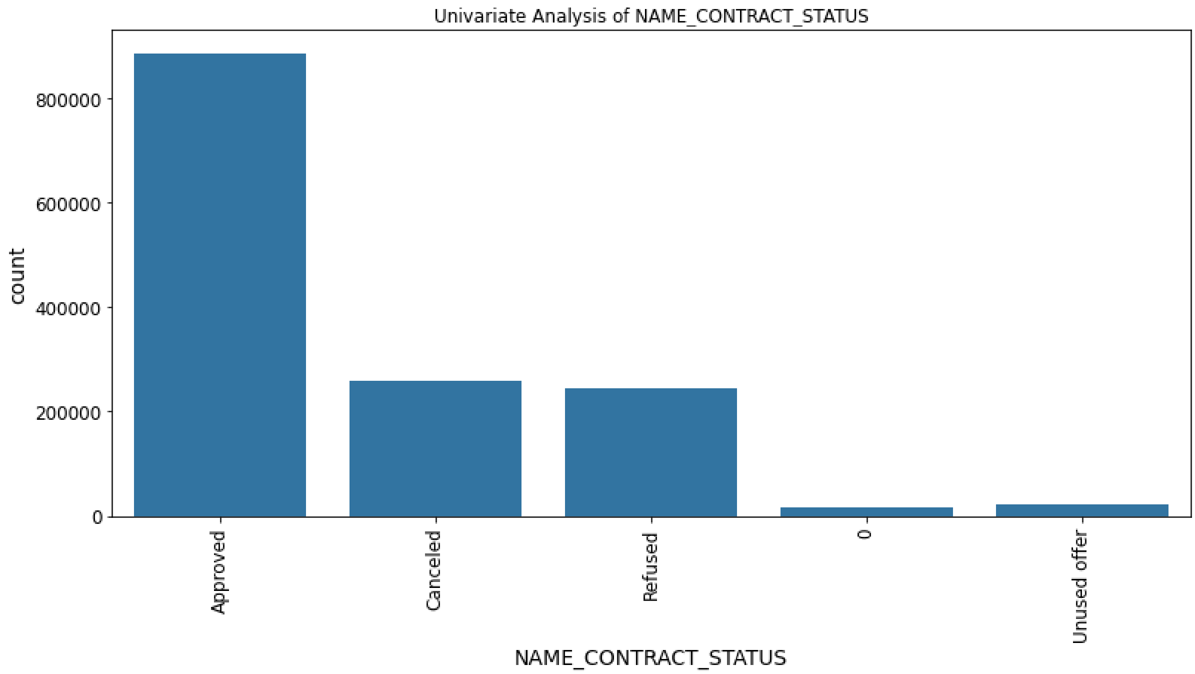
#Similarly Student eventhough credit amount is low tends to pay ontime

#But Maternity leave clients eventhough their representation is less in data tends to have more payement difficulty than ontime payment

numcat_bivar('NAME_EDUCATION_TYPE','AMT_CREDIT')




```
#OBSERVATIONS:Higher Education and Academic degree received more credits and their presence is more in Ontime paying category

#As Education level is low defaulting tendency is high

plt.style.use('default')

%matplotlib inline

# define function

def uni_var_cat(column):

    plt.figure(figsize=(13,6),facecolor='white')

    plt.rcParams['axes.labelsize']=14

    plt.rc('xtick', labelsize=12)

    plt.rc('ytick', labelsize=12)

    sns.countplot(data=dfnew,x=column)

    plt.title("Univariate Analysis of"+ ' '+column)

    plt.xticks(rotation=90)

    plt.show()

# Select categorical variables for analysis

categorical=['NAME_CASH_LOAN_PURPOSE', 'NAME_PAYMENT_TYPE', 'NAME_CLIENT_TYPE', 'NAME_GOODS_CATEGORY',

      'NAME_PORTFOLIO', 'NAME_PRODUCT_TYPE', 'CHANNEL_TYPE',

      'NAME_SELLER_INDUSTRY']

# Analysis of NAME_CONTRACT_STATUS

uni_var_cat('NAME_CONTRACT_STATUS')
```

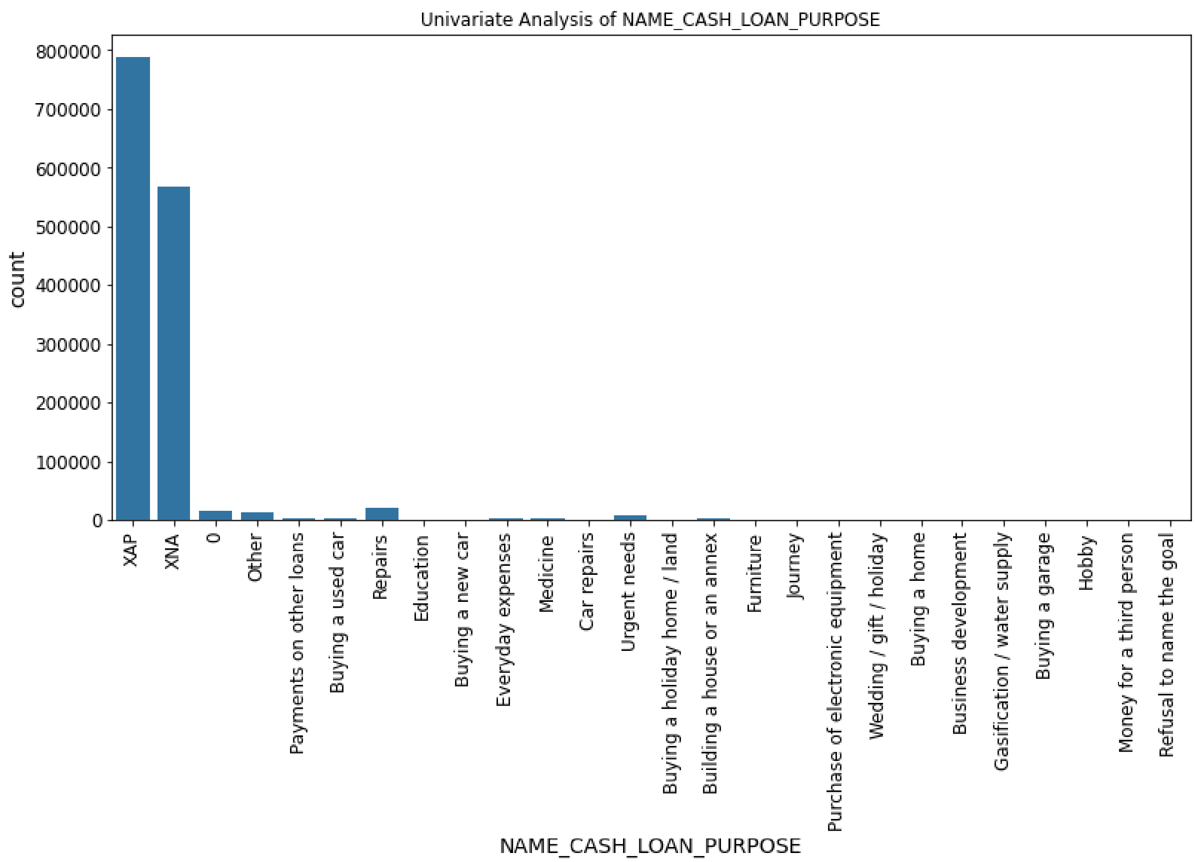


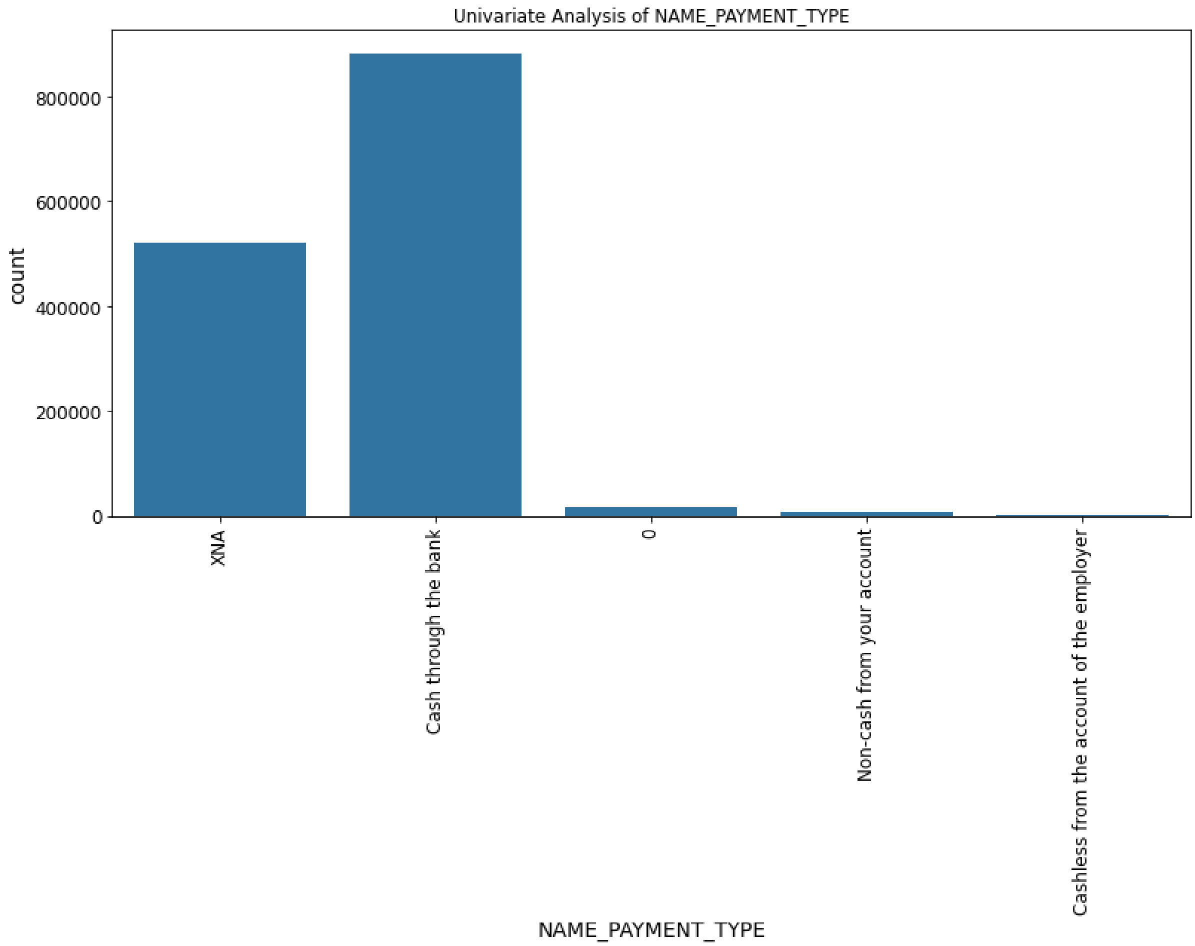
#most of the appplications are approved than canceled or refused

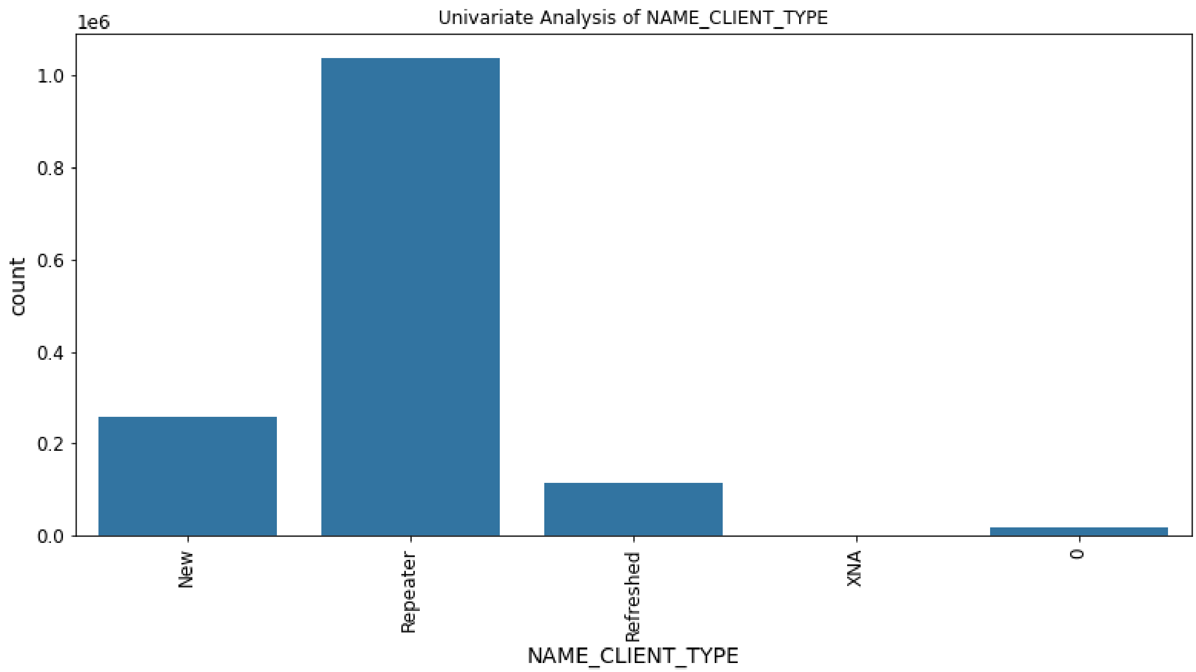
Analysis of other variables

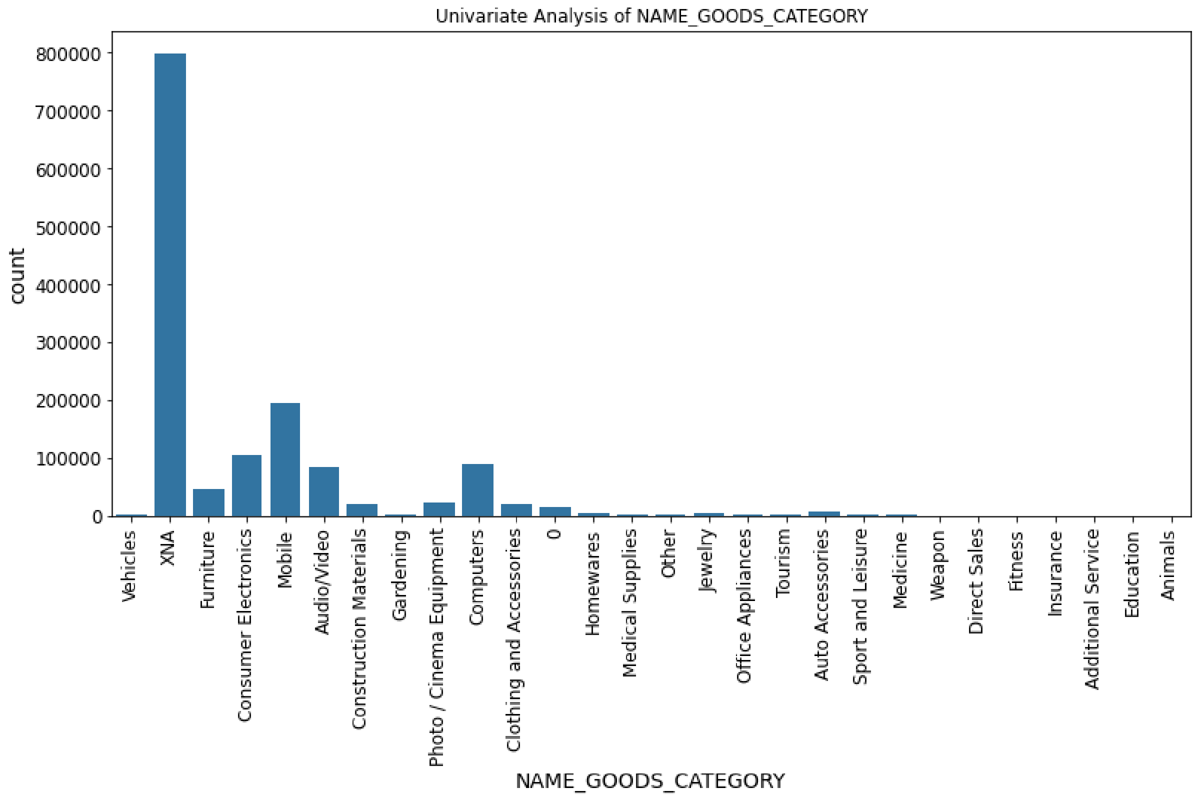
for i in categorical:

    uni_var_cat(i)

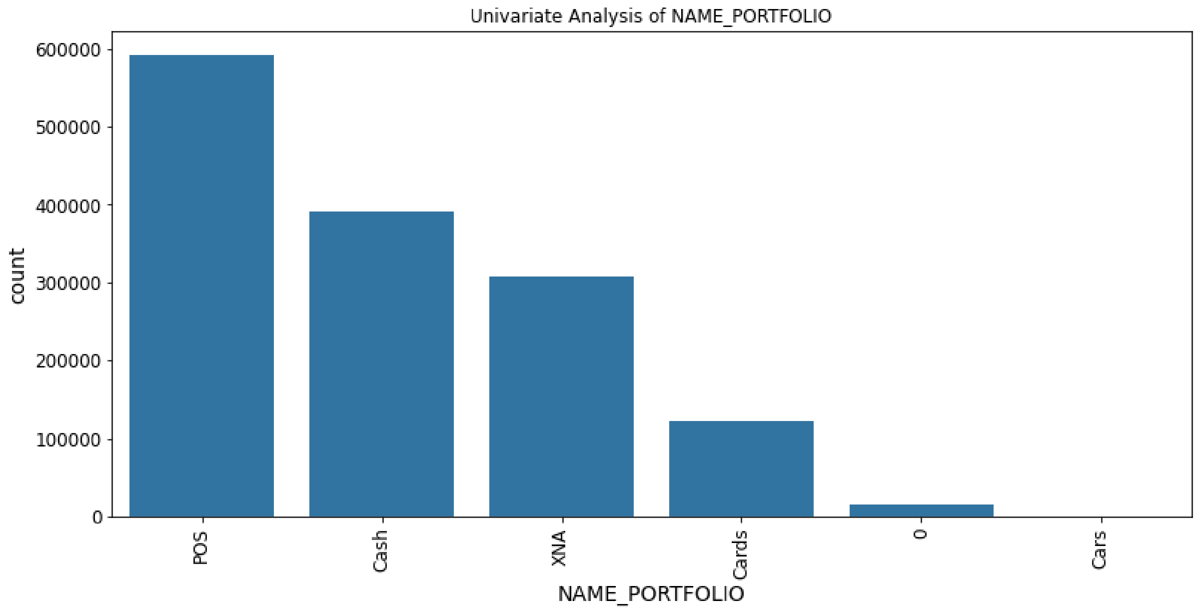


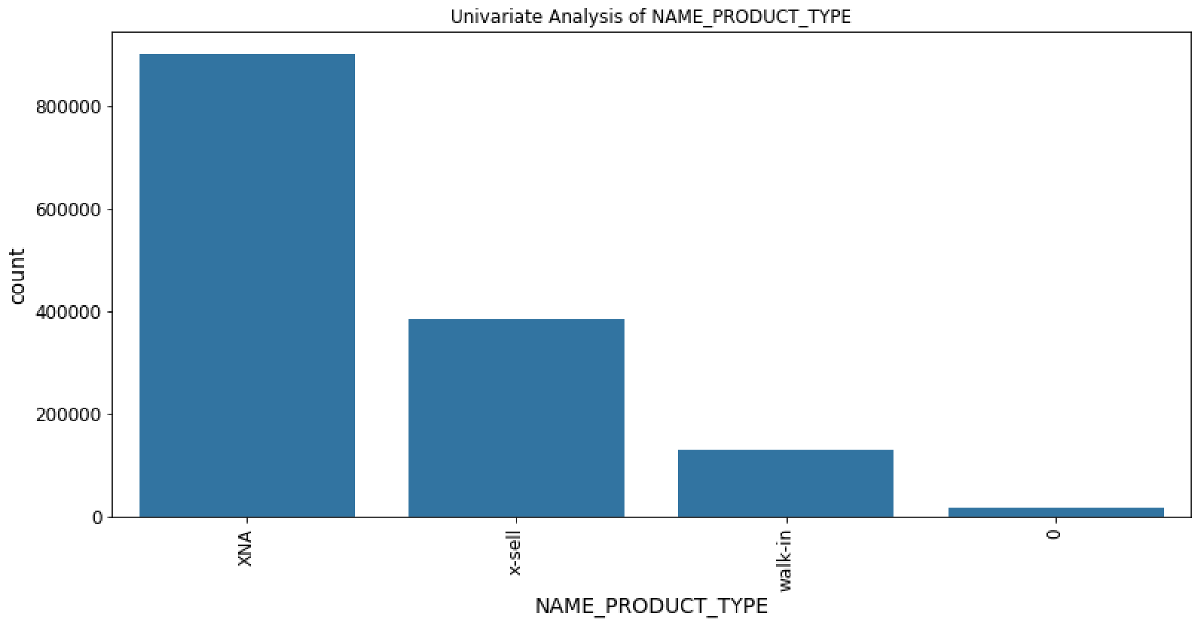

Univariate Analysis of NAME_PAYMENT_TYPE
count
800000
600000
400000
200000
0
XNA
Cash through the bank
0
Non-cash from your account
Cashless from the account of the employer
NAME_PAYMENT_TYPE

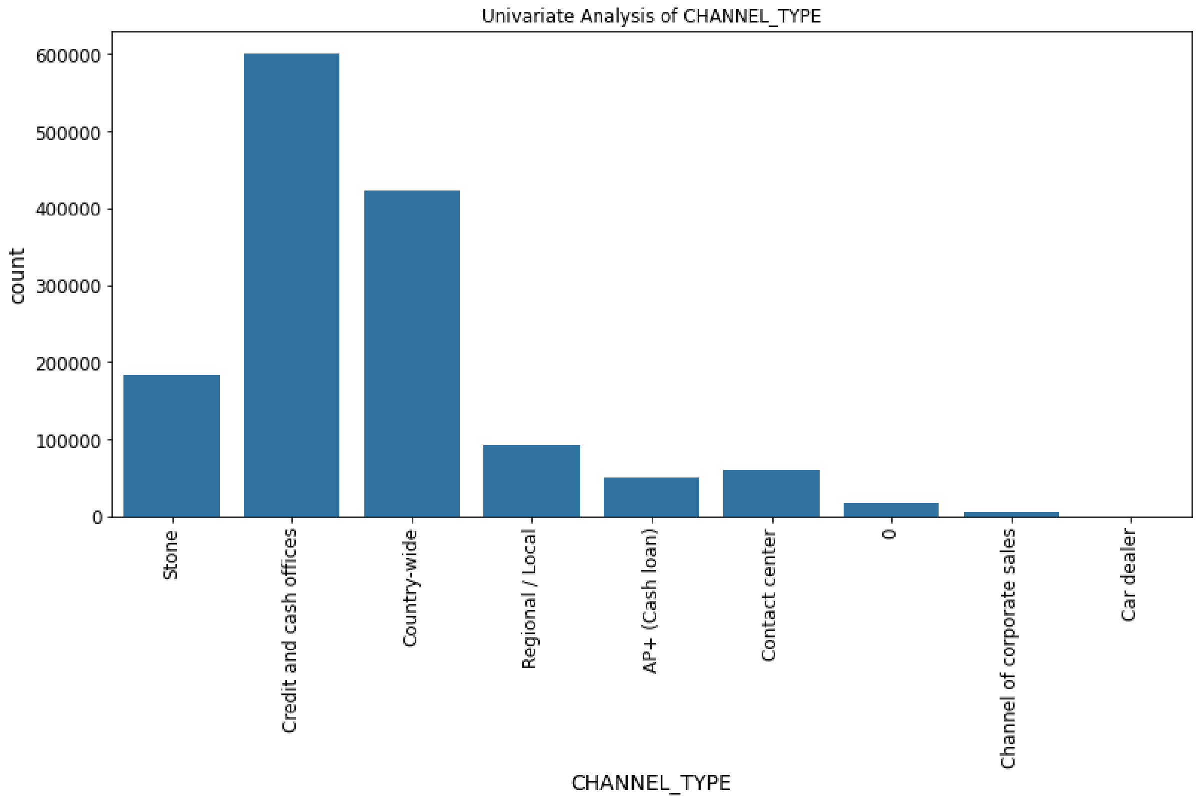

Univariate Analysis of NAME_CLIENT_TYPE
1e6
count
1.0
0.8
0.6
0.4
0.2
0.0
New
Repeater
Refreshed
XNA
0
NAME_CLIENT_TYPE

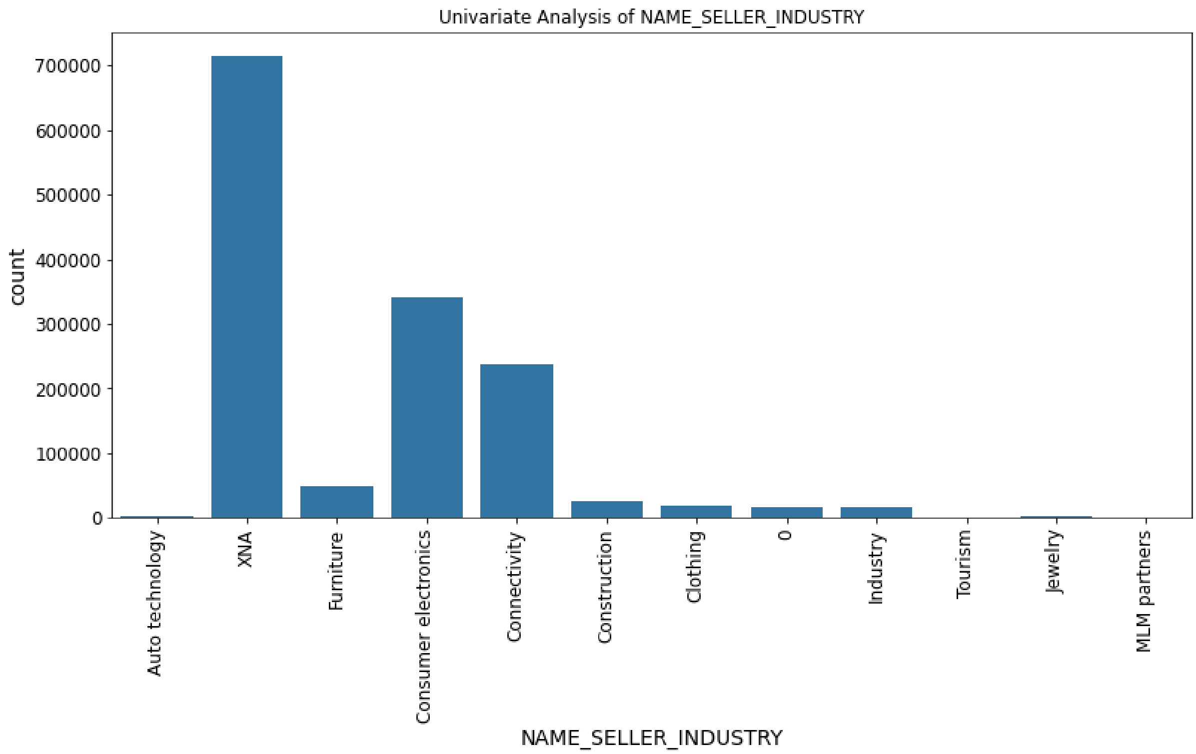


Univariate Analysis of NAME_GOODS_CATEGORY
count
800000
700000
600000
500000
400000
300000
200000
100000
0
Vehicles
XNA
Furniture
Consumer Electronics
Mobile
Audio/Video
Construction Materials
Gardening
Photo / Cinema Equipment
Computers
Clothing and Accessories
0
Homewares
Medical Supplies
Other
Jewelry
Office Appliances
Tourism
Auto Accessories
Sport and Leisure
Medicine
Weapon
Direct Sales
Fitness
Insurance
Additional Service
Education
Animals
NAME_GOODS_CATEGORY


Univariate Analysis of NAME_PORTFOLIO
count
0
100000
200000
300000
400000
500000
600000
POS
Cash
XNA
Cards
0
Cars
NAME_PORTFOLIO


Univariate Analysis of NAME_PRODUCT_TYPE
count
0
200000
400000
600000
800000
XNA
x-sell
walk-in
0
NAME_PRODUCT_TYPE





#Majority of the clients use loan for repairs,other and urgent needs.Rest of the parameters have negligible participation

#Clients done loan repayment through bank by cash

#Already existing client outnumbered New clients

#Majority of clients applied for Mobiles in previous Application(Ignoring xna)

#Most of the previous loan applications was for Point of sale followed by Cash and Cards

#Majority of the previuos application was cross shell than walkin.But most of the values was not filled properly

#Through Credit and cash offices and country-wide most of the previous clients were acquired

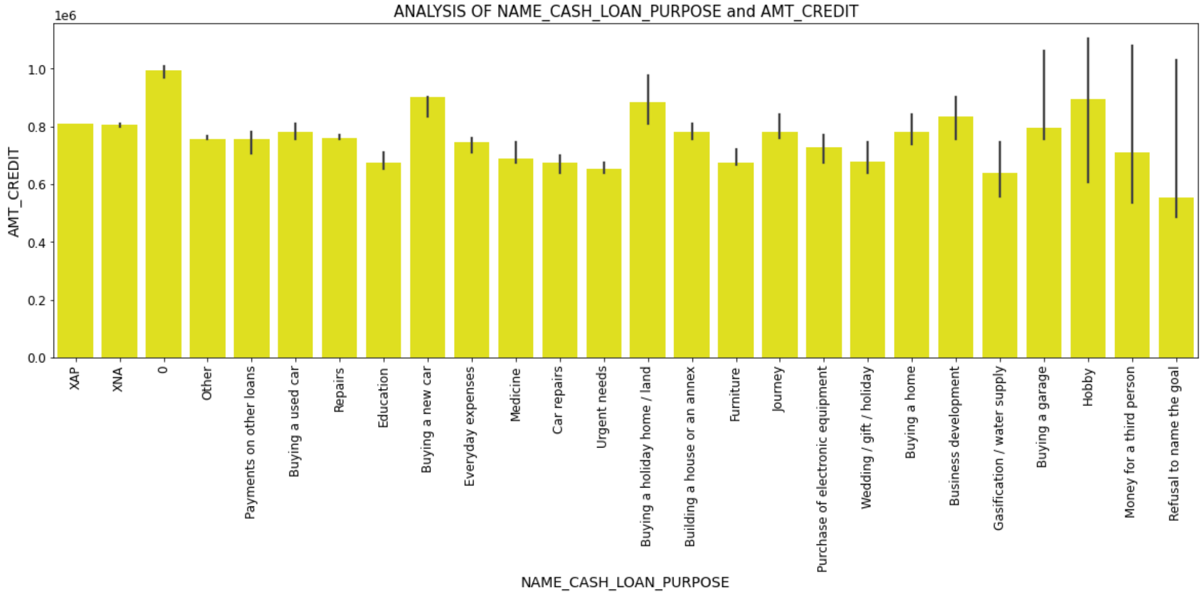
#consumer electronics and Connectivity are the good seller industries(ignoring XNA)

#Bivariate Analysis

```
def num_cat_bivar(column1,column2):

    plt.figure(figsize=(20,6),facecolor='white')

    plt.rcParams['axes.labelsize']=14

    plt.rc('xtick', labelsize=12)

    plt.rc('ytick', labelsize=12)

    sns.barplot(data=dfnew,x=column1,y=column2,estimator=lambda
x:np.quantile(x,0.75),color='yellow')

    plt.title('ANALYSIS OF'+' '+ column1+' and '+column2,size=15)

    plt.xticks(rotation=90)

    plt.show()
```

Analysis of NAME_CASH_LOAN_PURPOSE & AMT_CREDIT

num_cat_bivar('NAME_CASH_LOAN_PURPOSE','AMT_CREDIT')



#Buying a home,Buying a car,Buying a holidayhome/land got more credit than other categories.We know for these loans avail funds by providing your asset as collateral to the lender.So bank can promote this type of safer loans

Function definition

def cat_cat_new1(column1,column2):

plt.figure(figsize=(15,6),facecolor='white')

plt.rcParams['axes.labelsize']=14

plt.rc('xtick', labelsize=12)

plt.rc('ytick', labelsize=12)

sns.countplot(data=dfnew,x=column1,hue=column2,hue_order=sorted(dfnew[column2].value_counts().index,reverse=True))

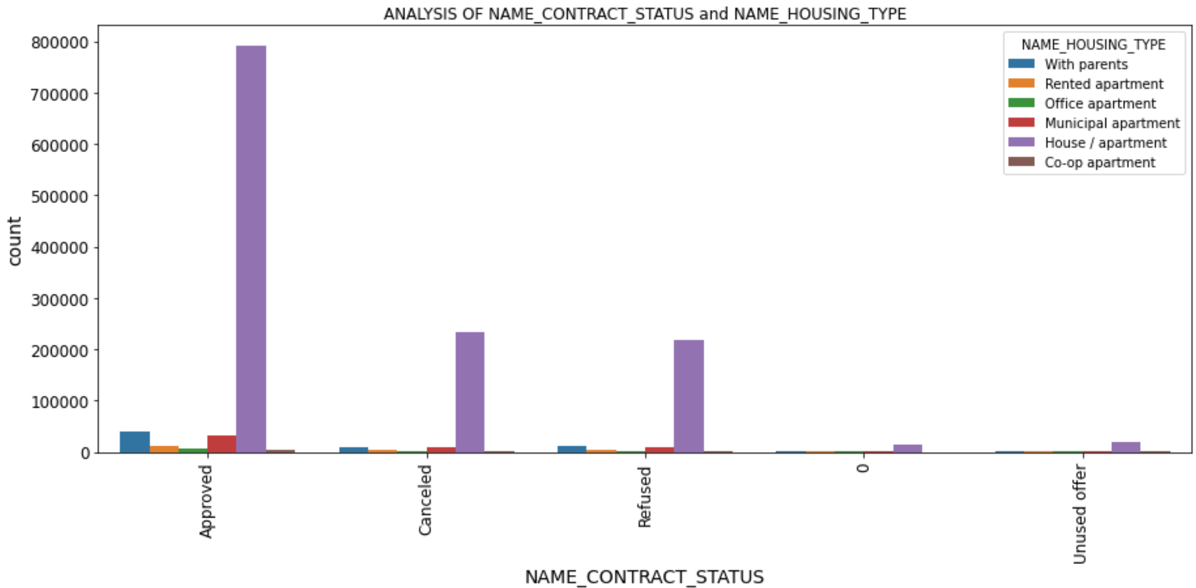
plt.title('ANALYSIS OF'+' '+ column1+' and '+column2)

plt.xticks(rotation=90)

plt.show()

Analysis of NAME_CONTRACT_STATUS & NAME_HOUSING_TYPE

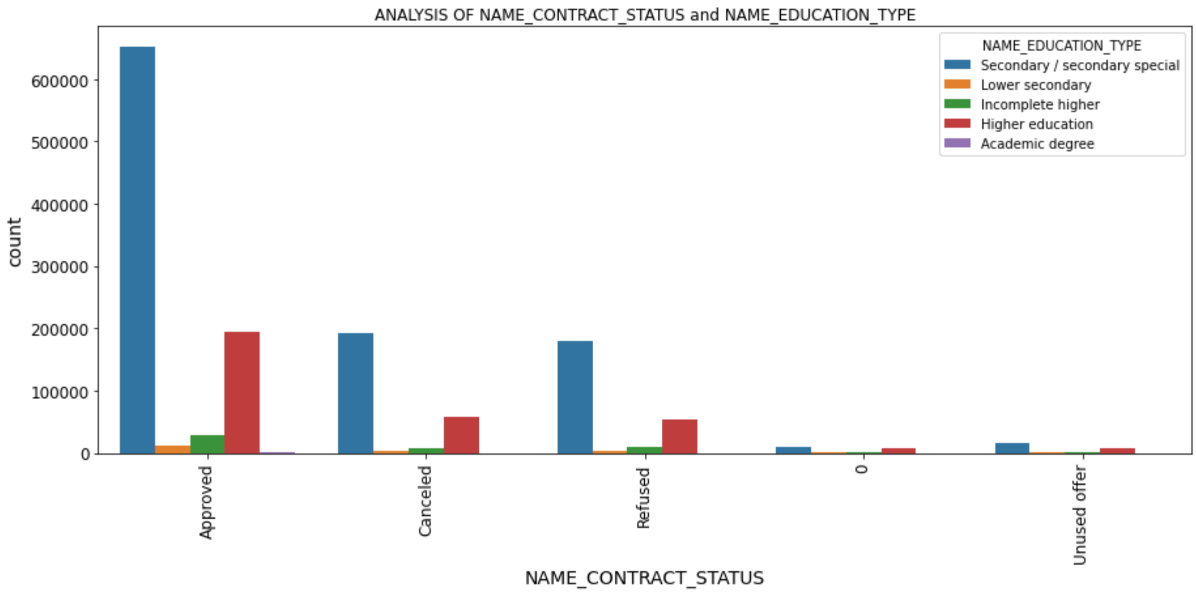
cat_cat_new1('NAME_CONTRACT_STATUS','NAME_HOUSING_TYPE')



#Clients have housing type as House/Apartment got more loan approvals than others

Analysis of NAME_CONTRACT_STATUS & NAME_EDUCATION_TYPE

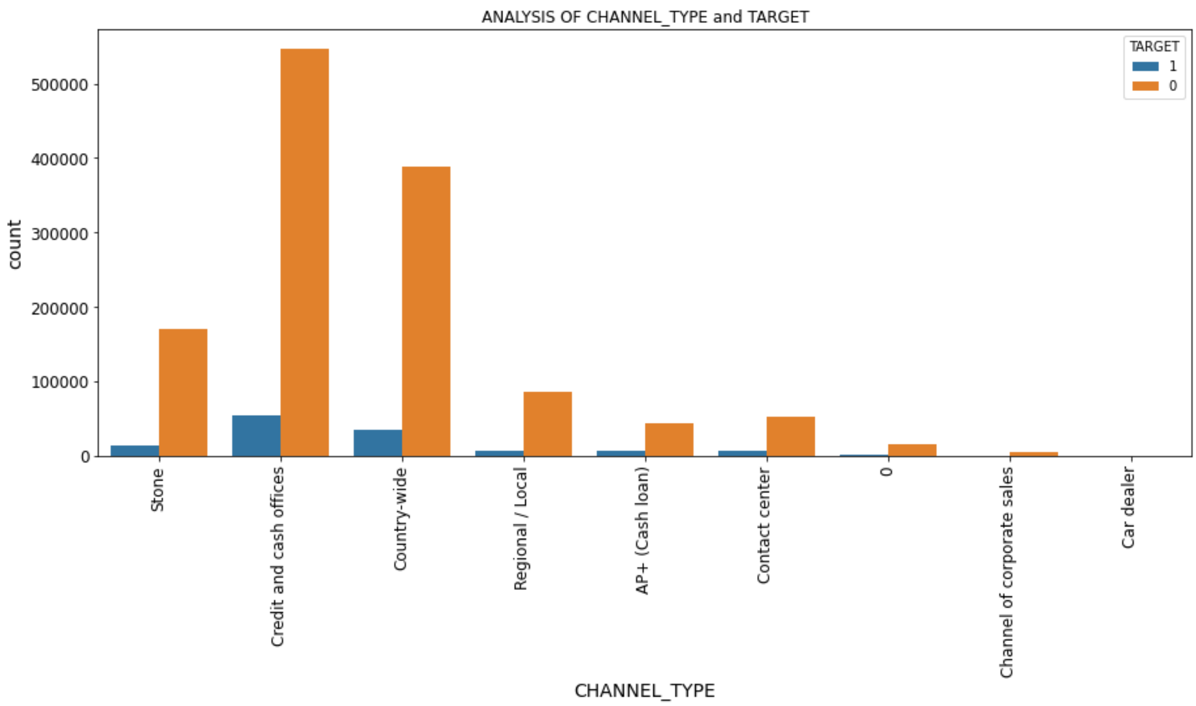
cat_cat_new1('NAME_CONTRACT_STATUS','NAME_EDUCATION_TYPE')



#Secondary/Secondary special has got higher loan approvals followed by Higher Education in previous application.But they are high in defaulters category also

Analysis of CHANNEL_TYPE & TARGET

cat_cat_new1('CHANNEL_TYPE','TARGET')

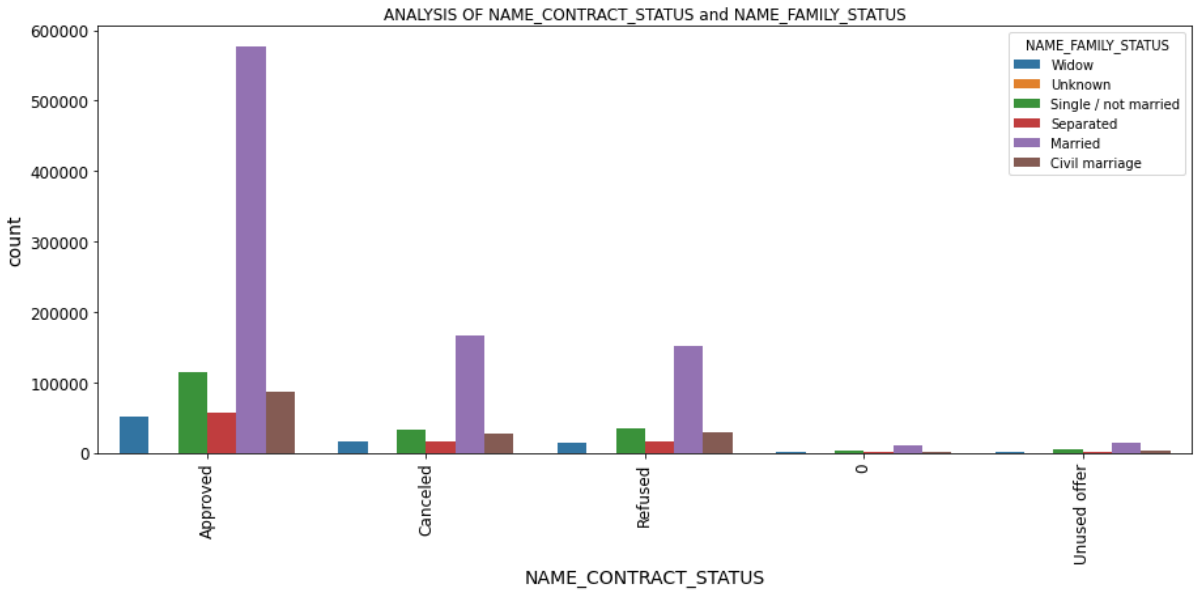


#Credit and cash offices has more clients without payement difficulties followes by country-wide

#Car dealer and Channel of corporate sales have clients who are defaulters

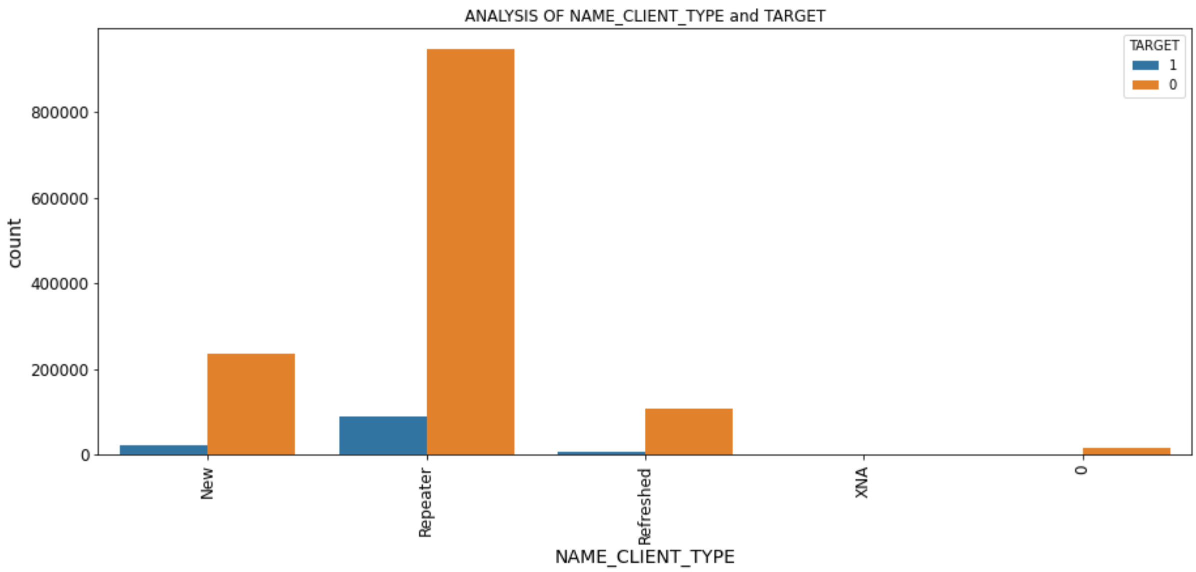
Analysis of NAME_CONTRACT_STATUS & NAME_FAMILY_STATUS

cat_cat_new1('NAME_CONTRACT_STATUS','NAME_FAMILY_STATUS')



Analysis of NAME_CLIENT_TYPE & TARGET

cat_cat_new1('NAME_CLIENT_TYPE','TARGET')



#Generally Repeaters or Old clients are ontime payers than New and Refreshed clients