

# Statistics and EDA - Module 09 – Credit EDA Case Study

***Author***: Anish Mahapatra, Karthik Premanand

***Email***: [anishmahapatra01@gmail.com](mailto:anishmahapatra01@gmail.com), karthikprem26@gmail.com

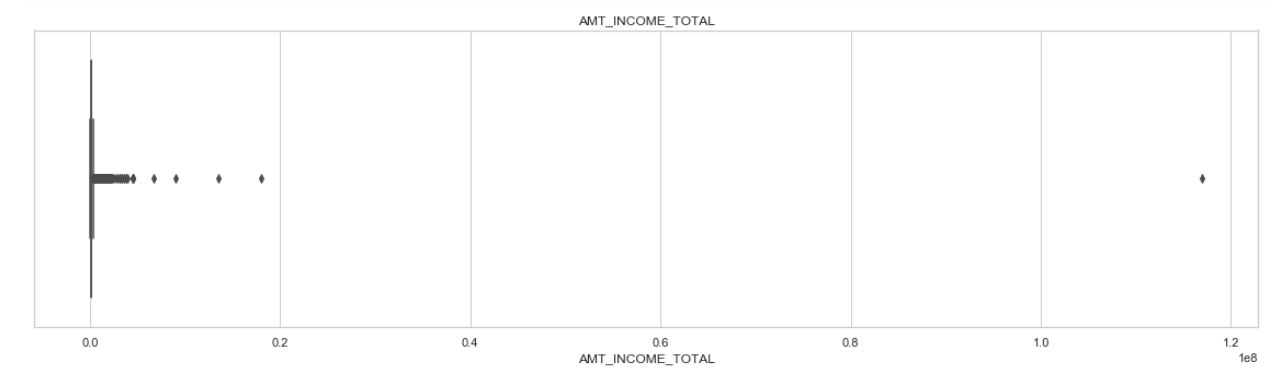
**Business Objectives**

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

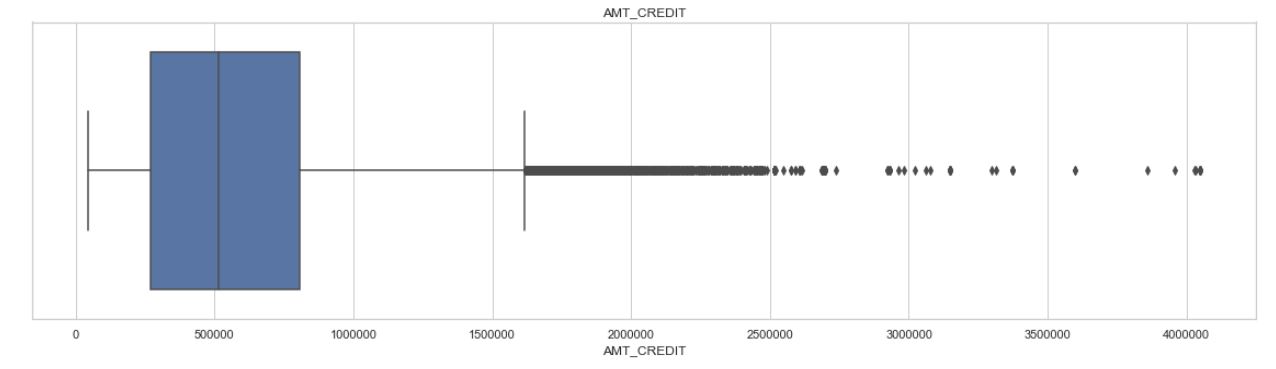
The conclusion from the graphs can be found as a part of the image in the interest of time. Kindly consider the same.

### Outlier Treatment Analysis

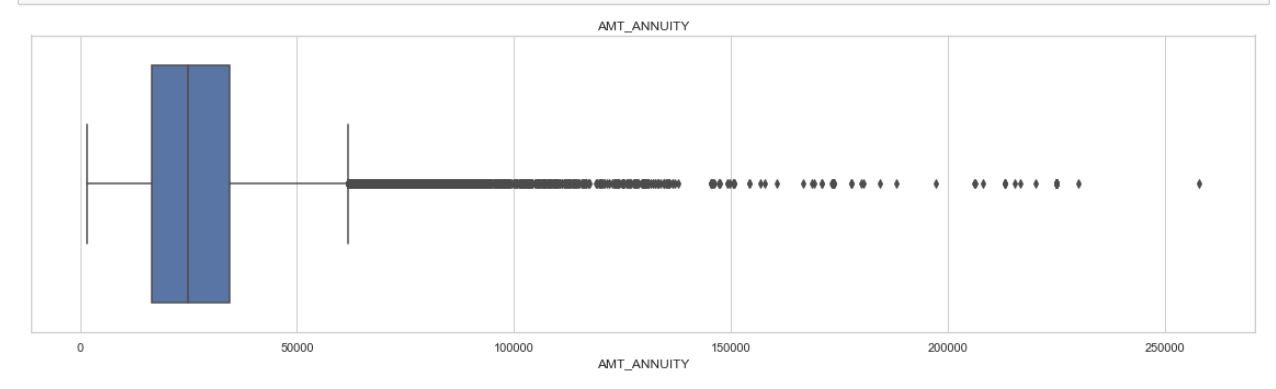
**Note:** The Boxplots below have been plotted with the standard whiskers of 1.5 x (IQR)



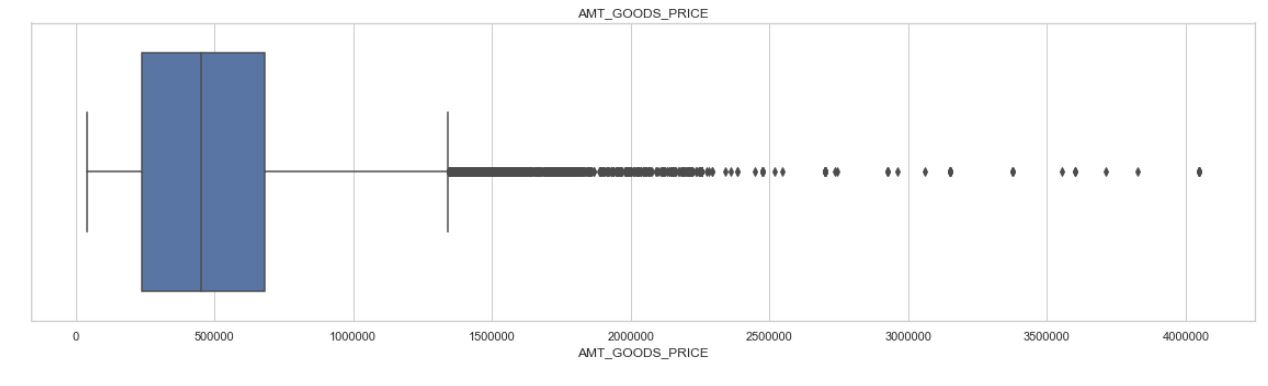
**Findings:** From the above plot of the income of the client above, we notice that quite a few people earn more than the average.



**Findings:** From the above plot, we see that most people get a loan between 3 to 8 lakhs and a maximum of about 7 lakhs. however, we notice that a considerable amount of people also get loans of upto 40 lakhs

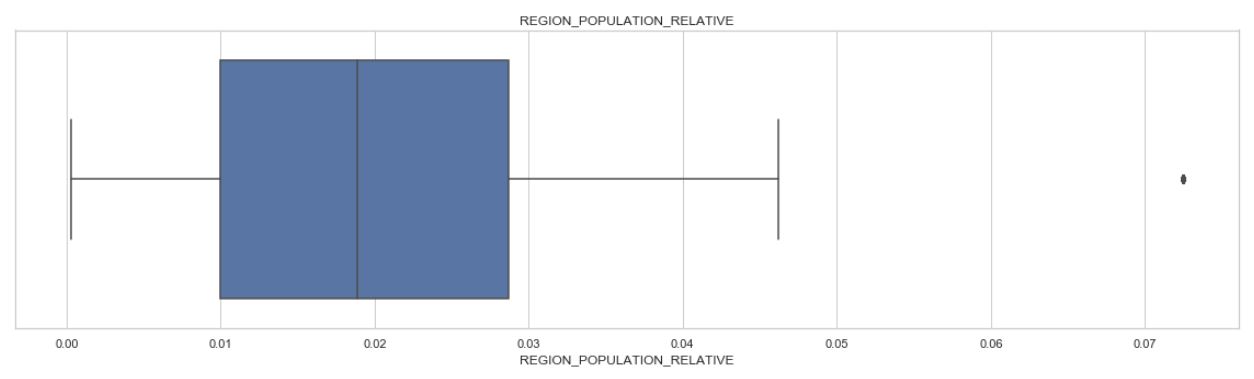
****

**Findings:** Due to annuity, loans, we see that history suggest that people hold around 65,000 worth of annuity. However, we see that this can go up significantly upto 20,00,000 and above



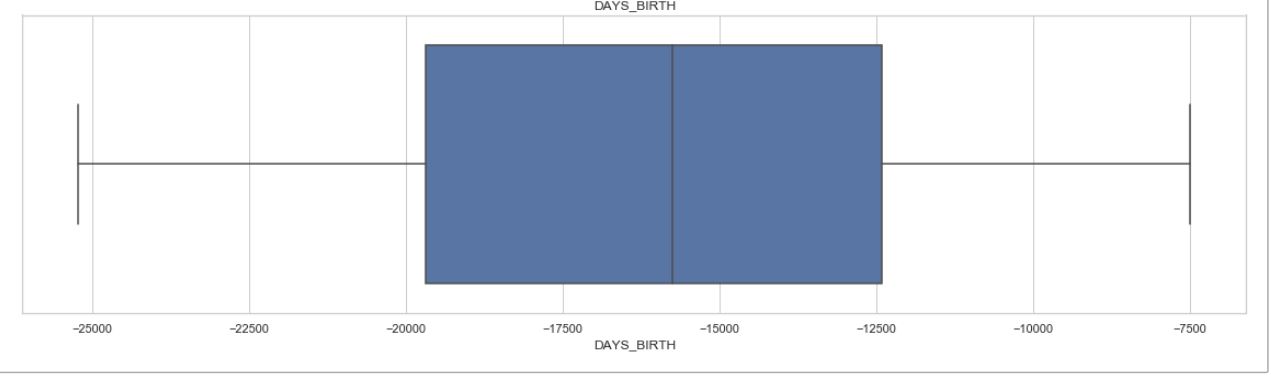
**Findings:** AMT\_GOODS\_PRICE is defined as: For consumer loans it is the price of the goods for which the loan is given.

We notice that it follows trends similar to AMT\_CREDIT with most of the spread lying within 13 lakhs and can go up to 40 lakhs.



**Findings:** REGION\_POPULATION\_RELATIVE - Normalized population of region where client lives (higher number means the client lives in more populated region)

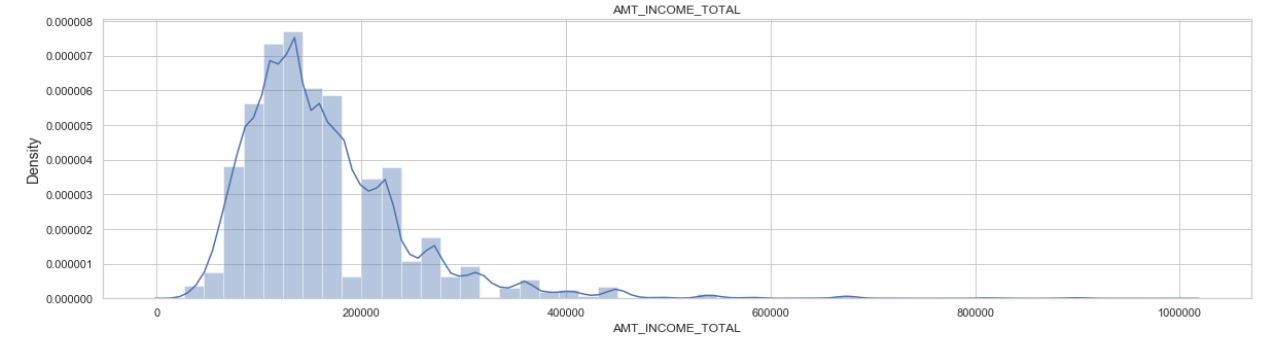
We notice that one client lives in a highly populated region, whereas the others mostly sat in regions that have 0.01 to 0.03 (approx) normalized population density



**Findings:** Based on the above plot, we notice that majority of the client are between 34 to 54 years of age with the youngest being 20 and the oldest being around 68

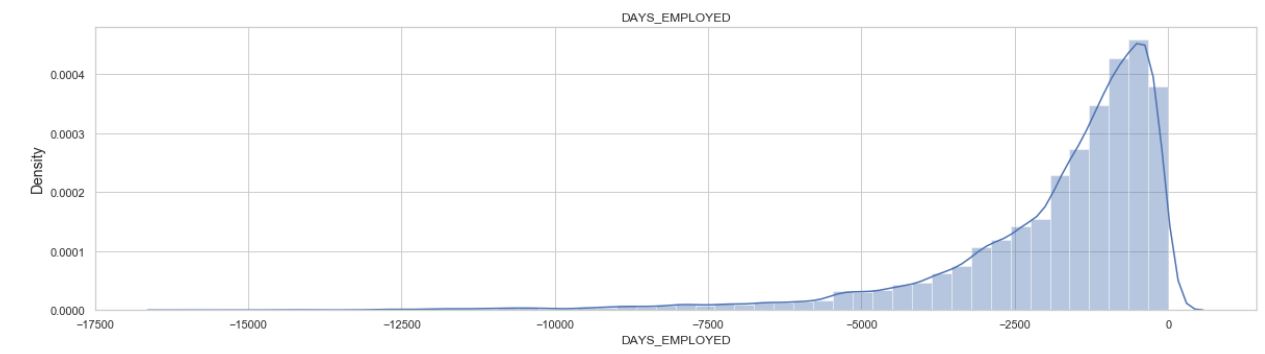
### Univariate Analysis on data where 'TARGET' = 1

#### Analysis of Client Income



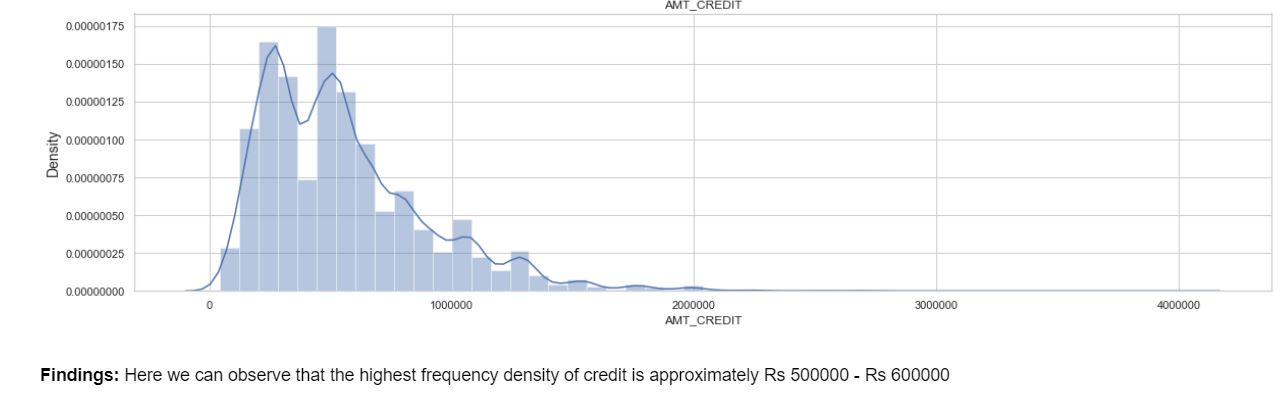
**Findings:** Here we can observe that the highest frequency of income amount for clients is between Rs. 100000 and Rs. 200000

#### Analysis of Days of Employment of the client

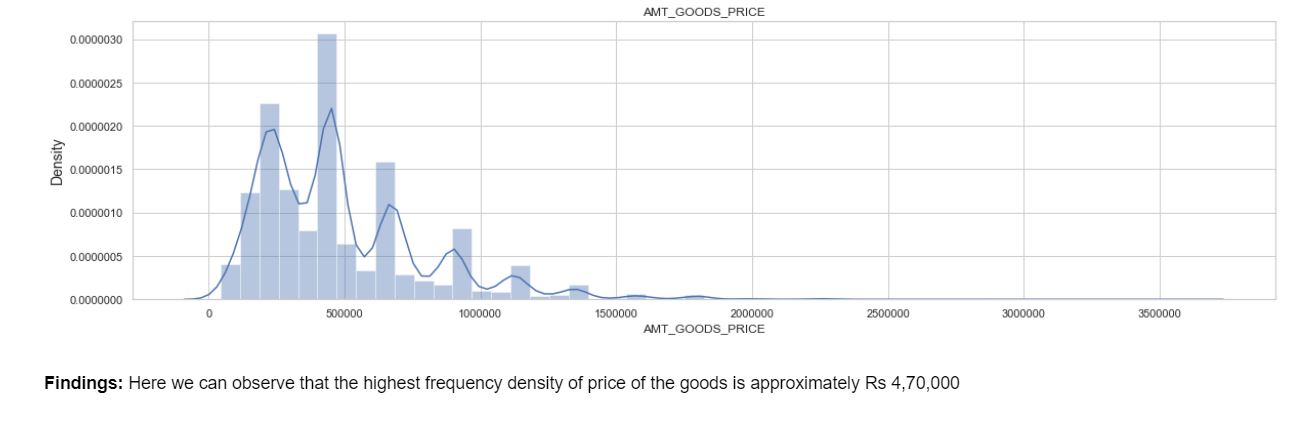


**Findings:** Here we can observe that the highest frequency of days that the clients have been employed is between 625 and 937 days i.e between 1.7 and 2.5 years

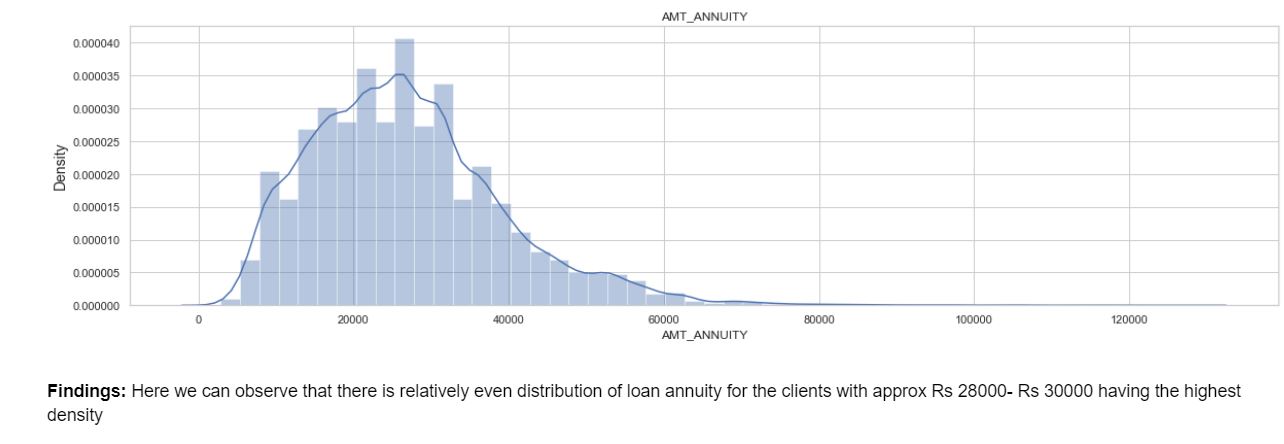
#### Analysis of Credit Amount of Loan



#### Analysis of Price of Loan Goods

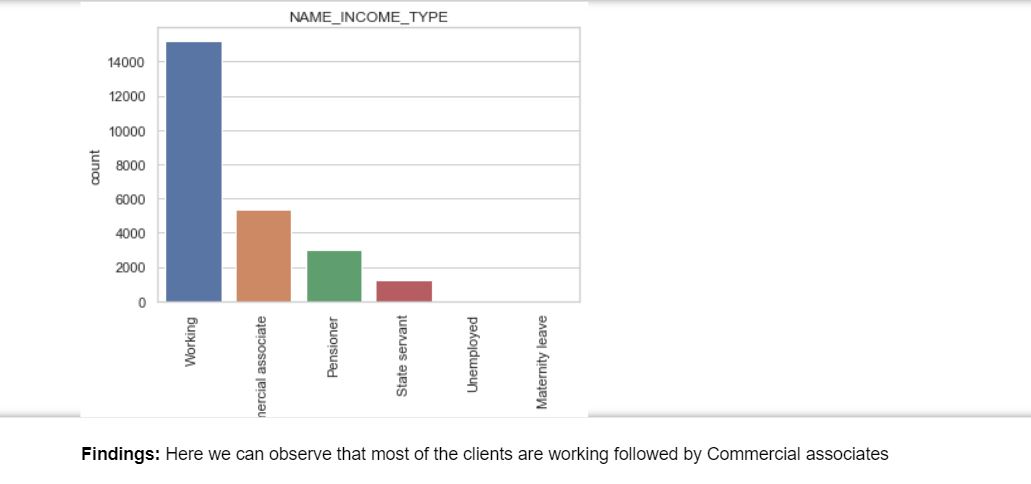


#### Analysis of Loan Annuity

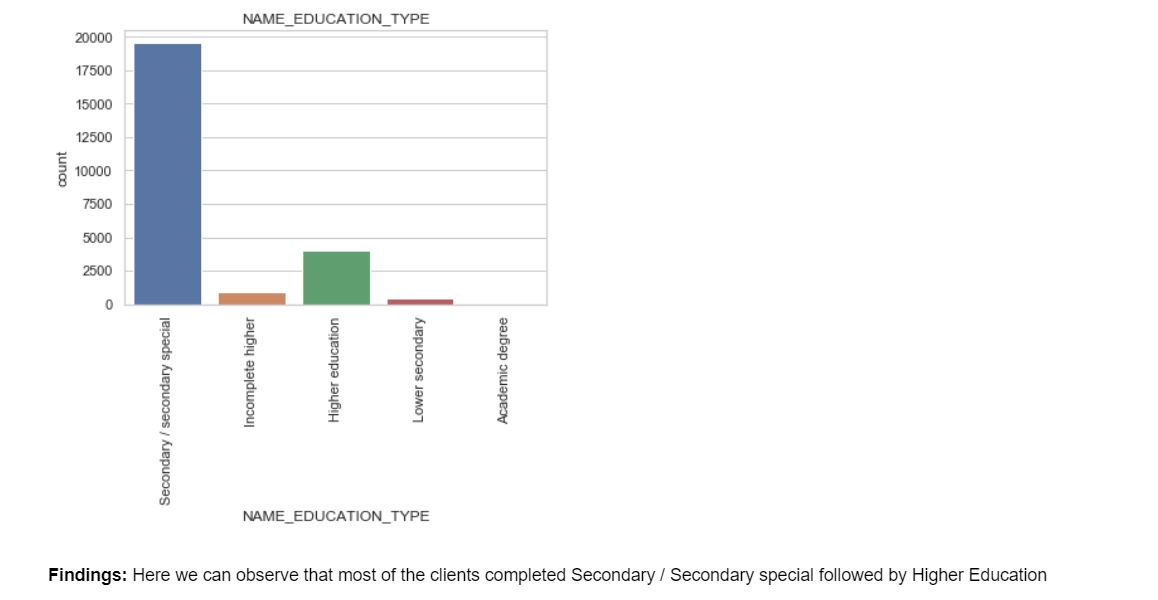


**Univariate Analysis for categorical variables**

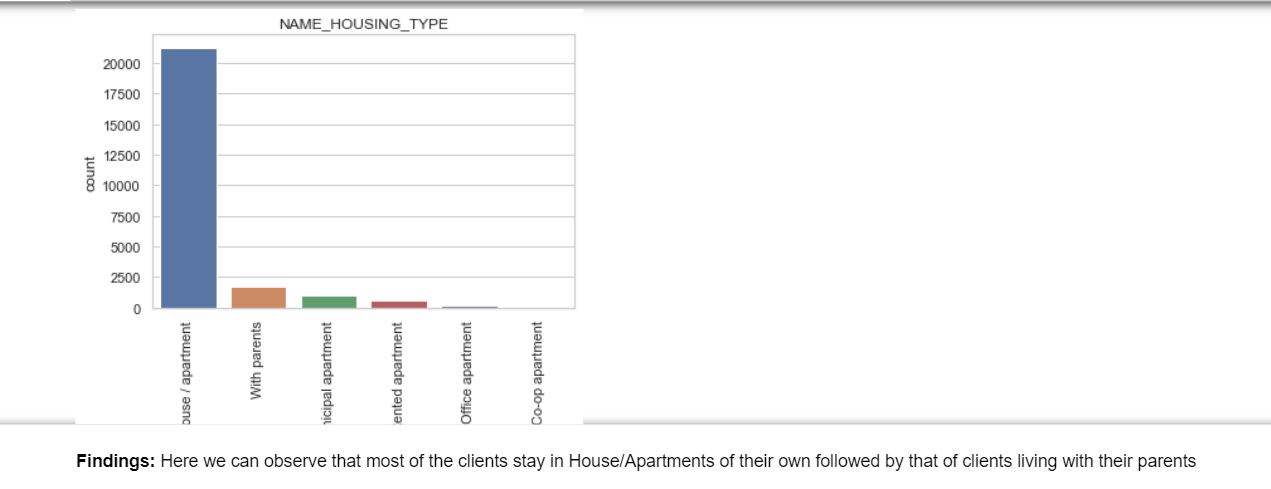
#### Analysis of Client Income Type



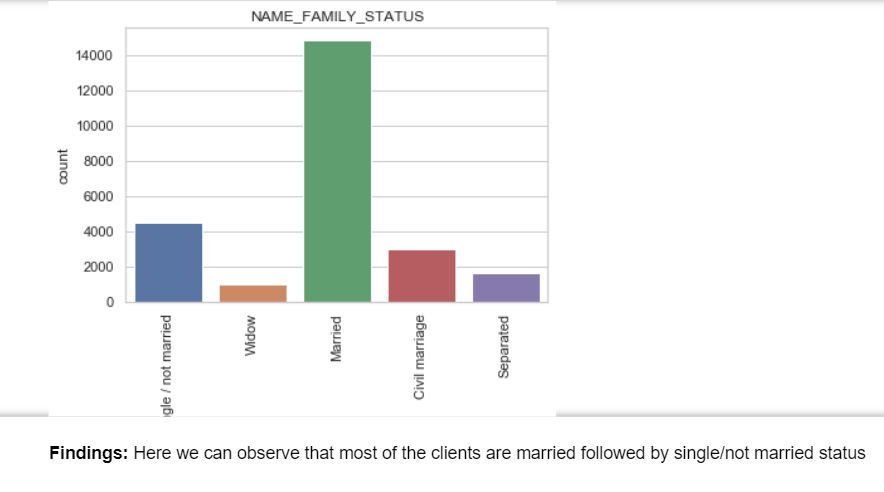
#### Analysis of Client Education Level



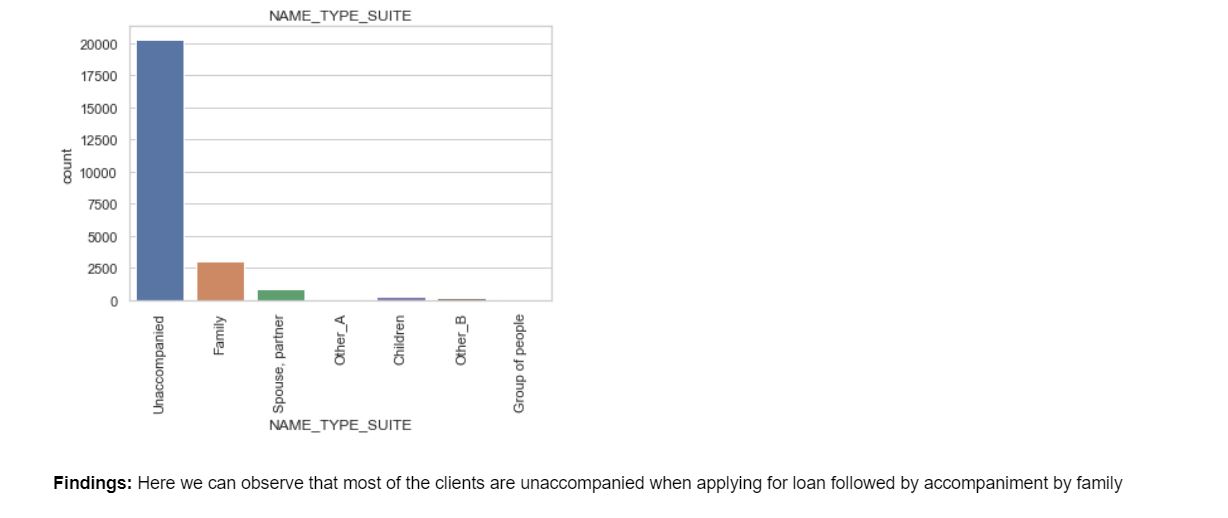
#### Analysis of Client Housing Type



#### Analysis of Client Family Status

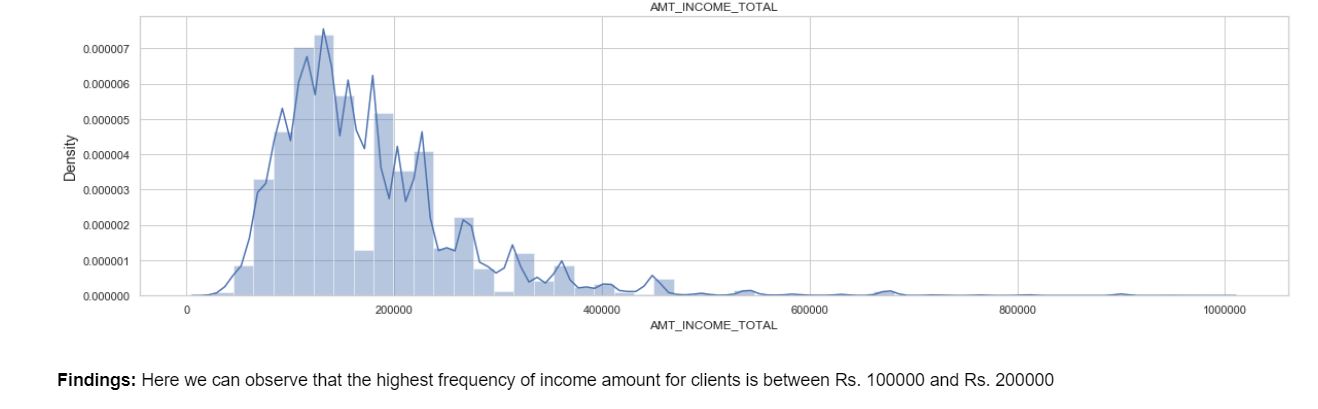


#### Analysis of Client Accompaniment status

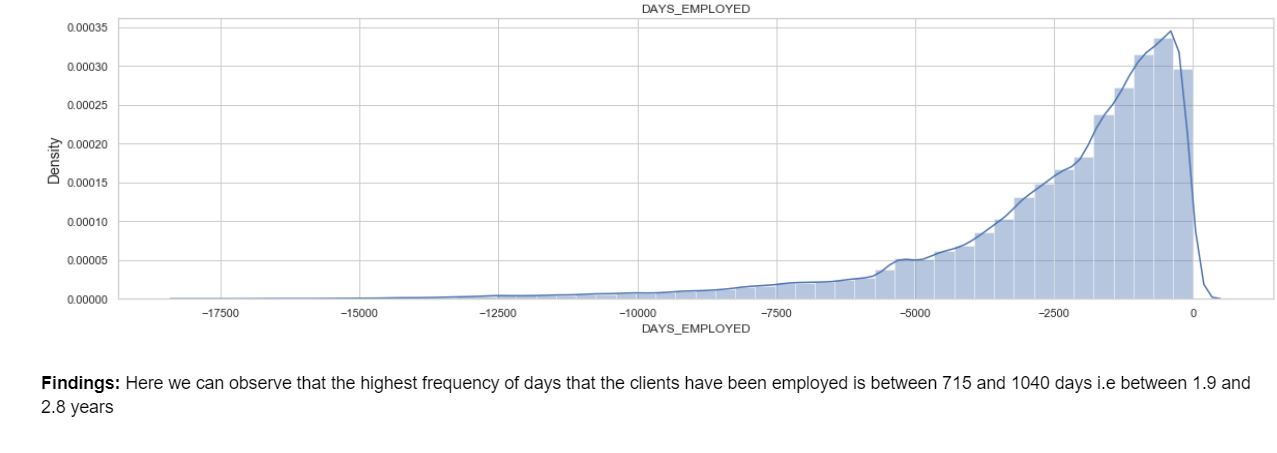


### Univariate Analysis where 'TARGET' = 0

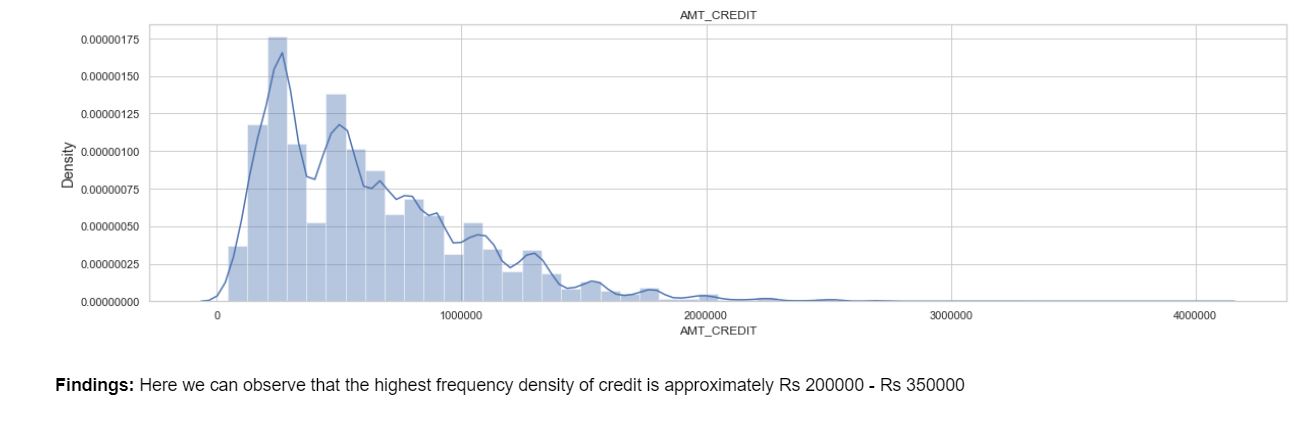
#### Analysis of Client Income



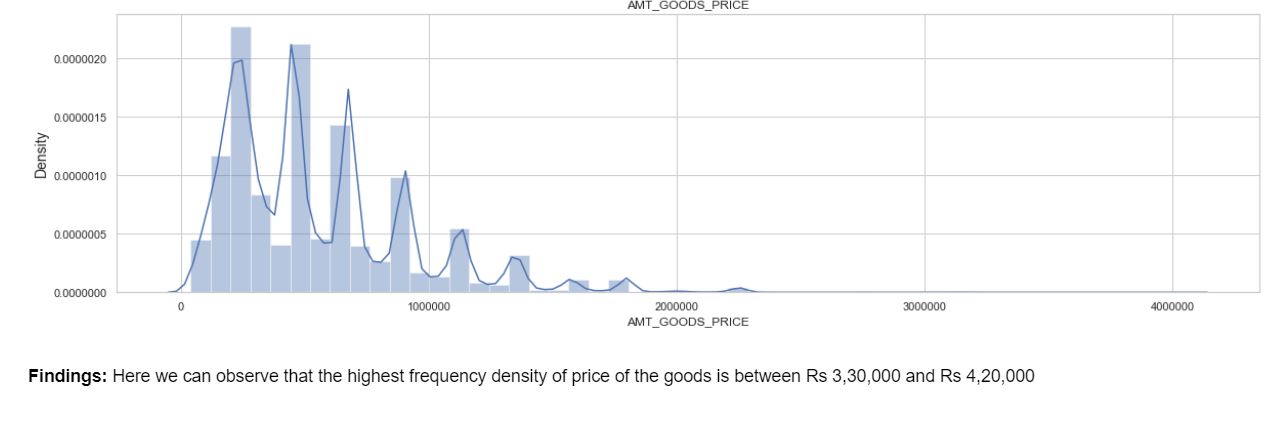
#### Analysis of Days of Employment of the client



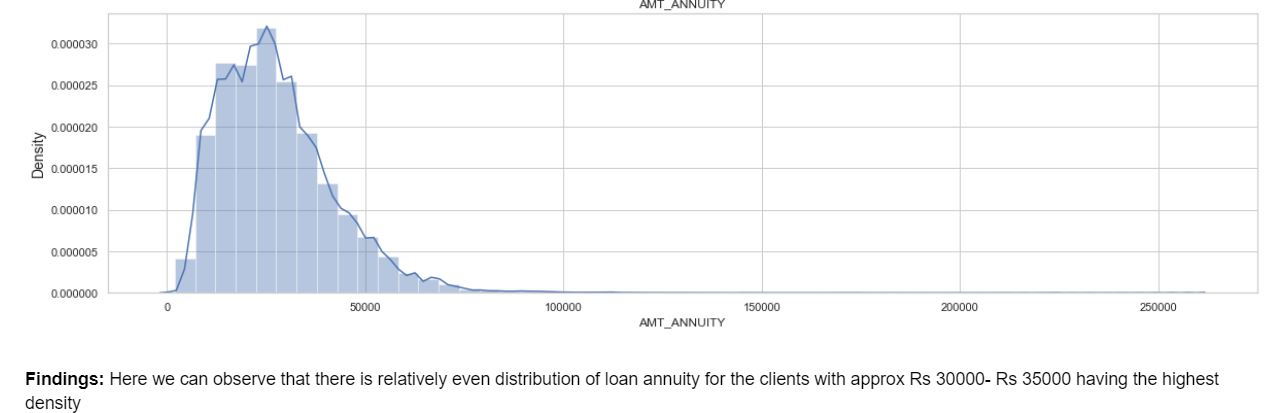
#### Analysis of Credit Amount of Loan



#### Analysis of Price of Loan Goods

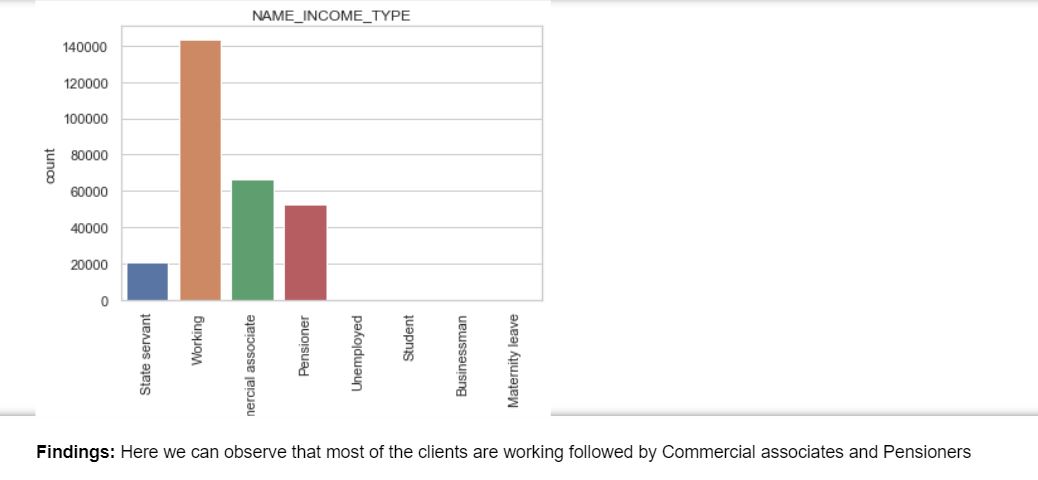


#### Analysis of Loan Annuity

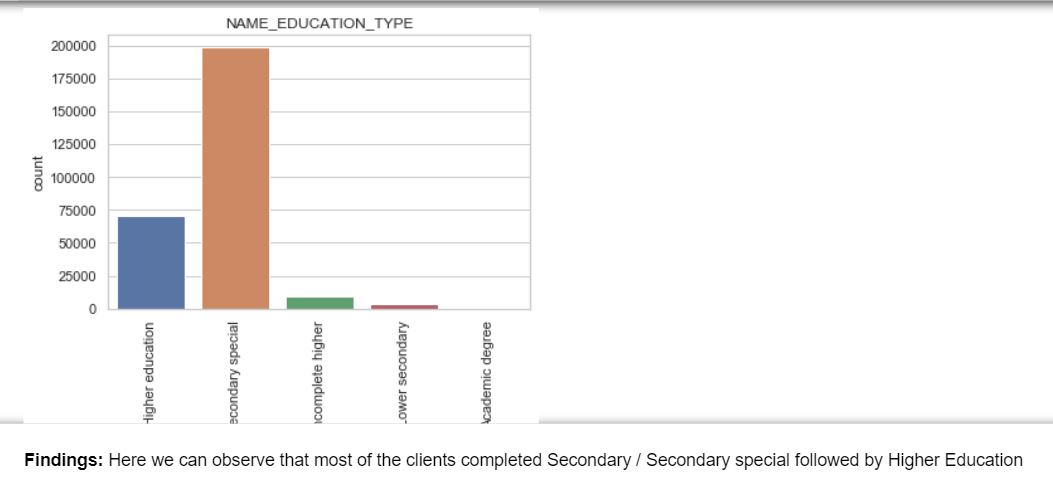


#### **Univariate Analysis for categorical variables**

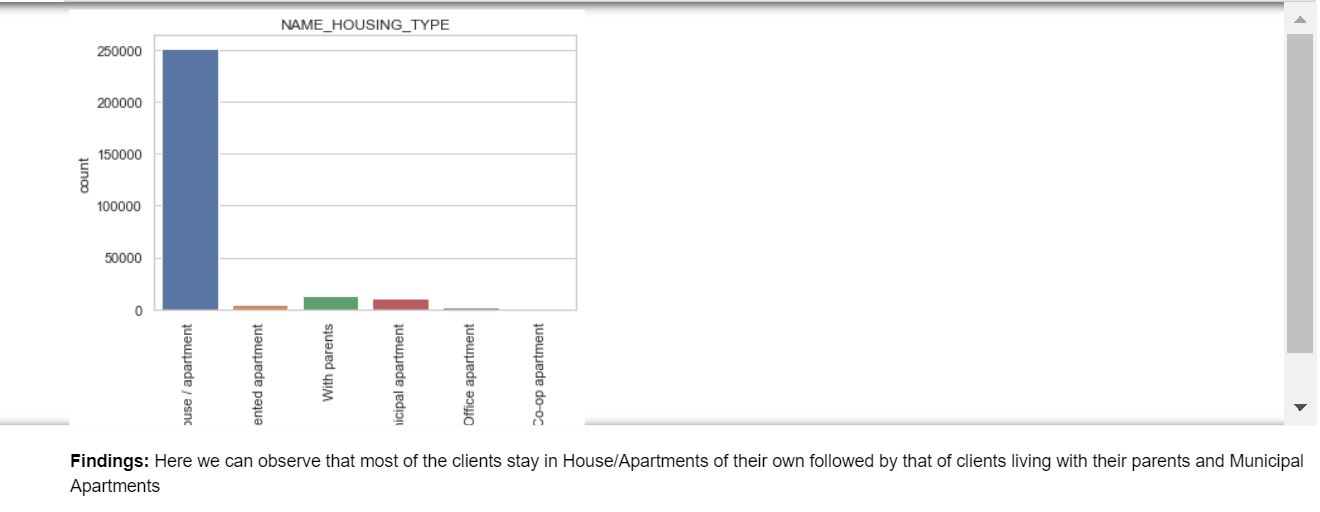
#### Analysis of Client Income Type



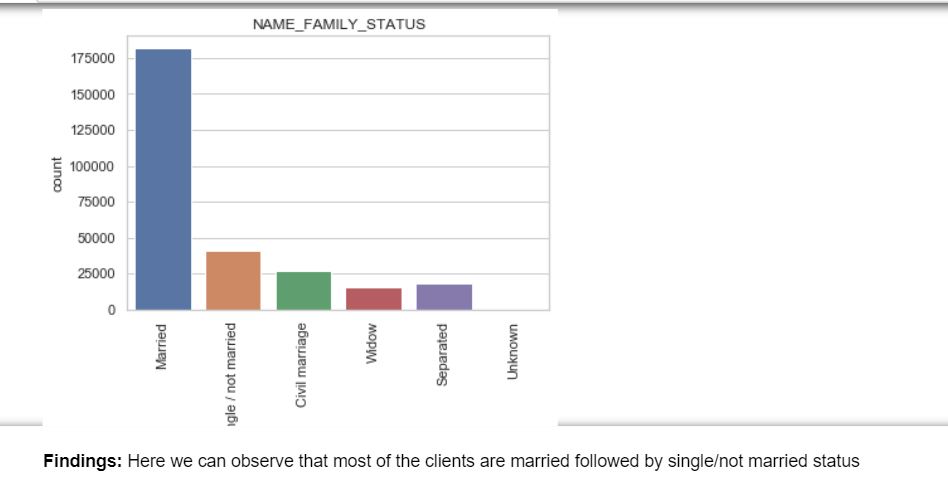
#### Analysis of Client Education Level



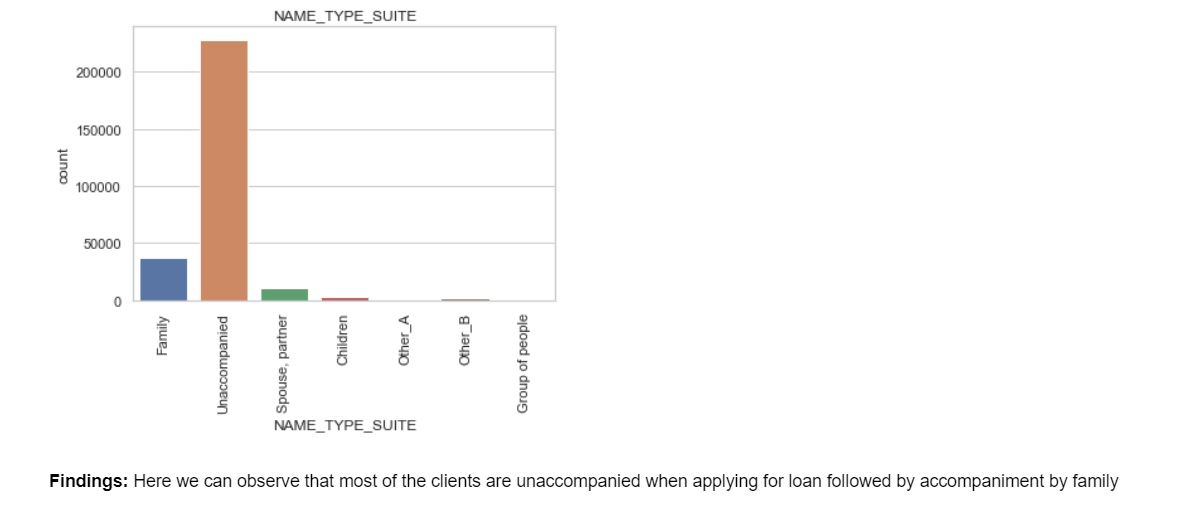
#### Analysis of Client Housing Type



#### Analysis of Client Family Status

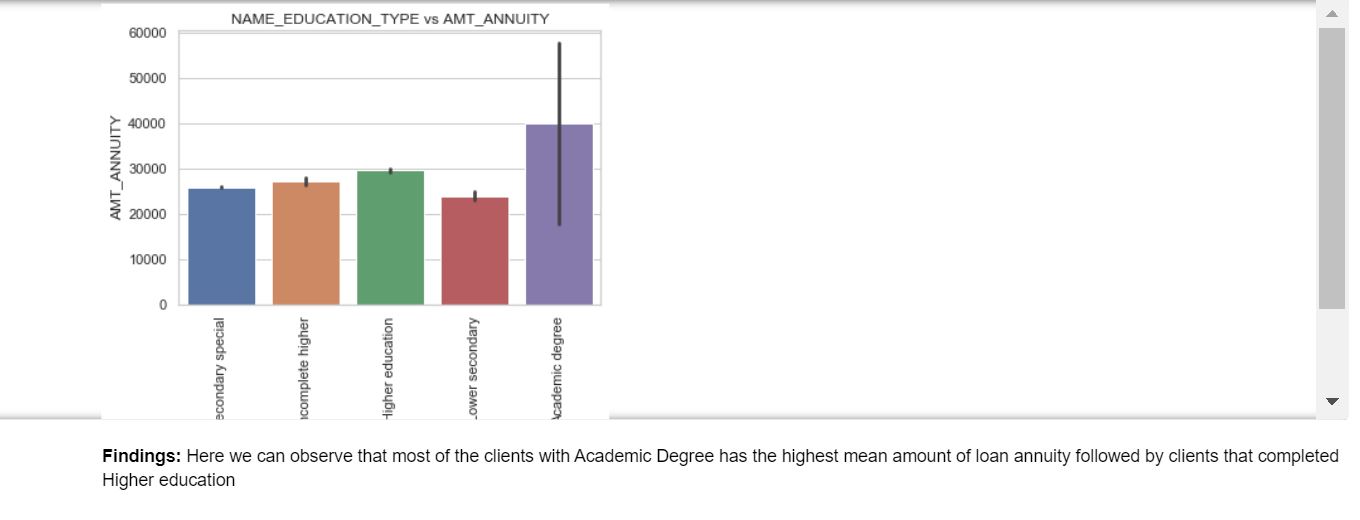


#### Analysis of Client Accompaniment status

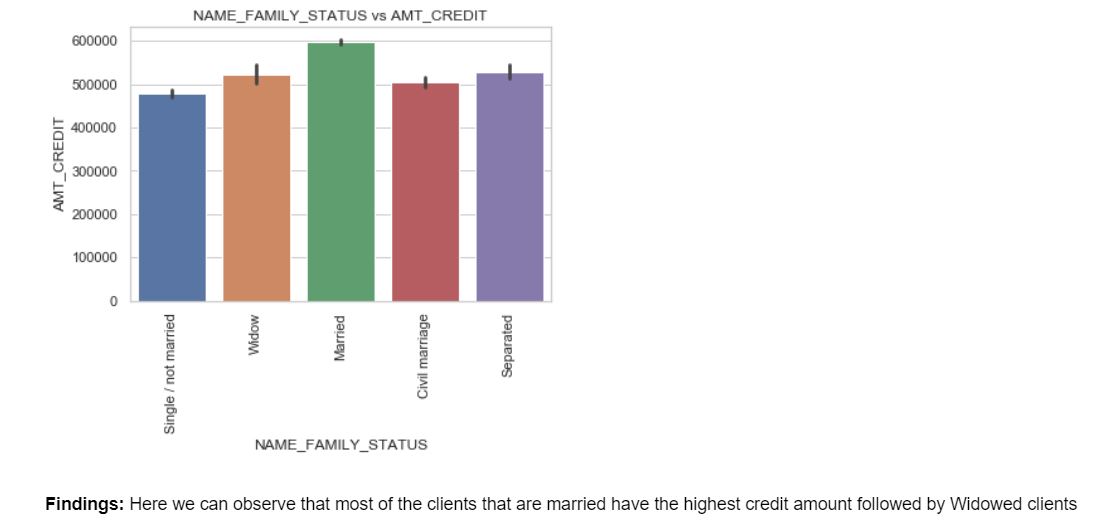


### Bivariate Analysis where 'TARGET' = 1

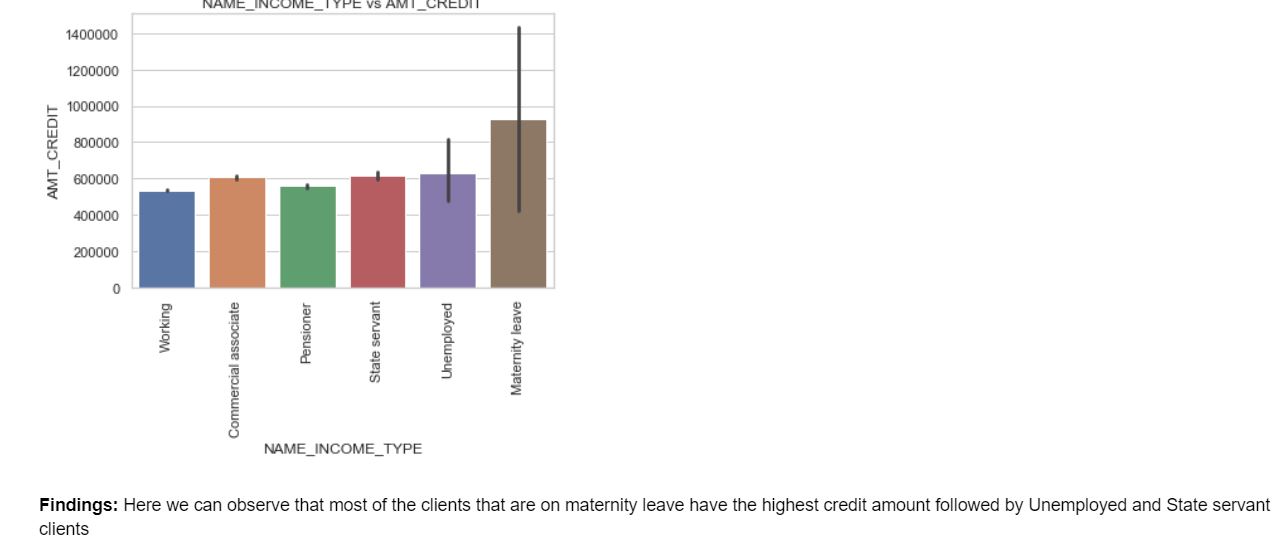
#### Analysis of Education Type versus Loan Annuity



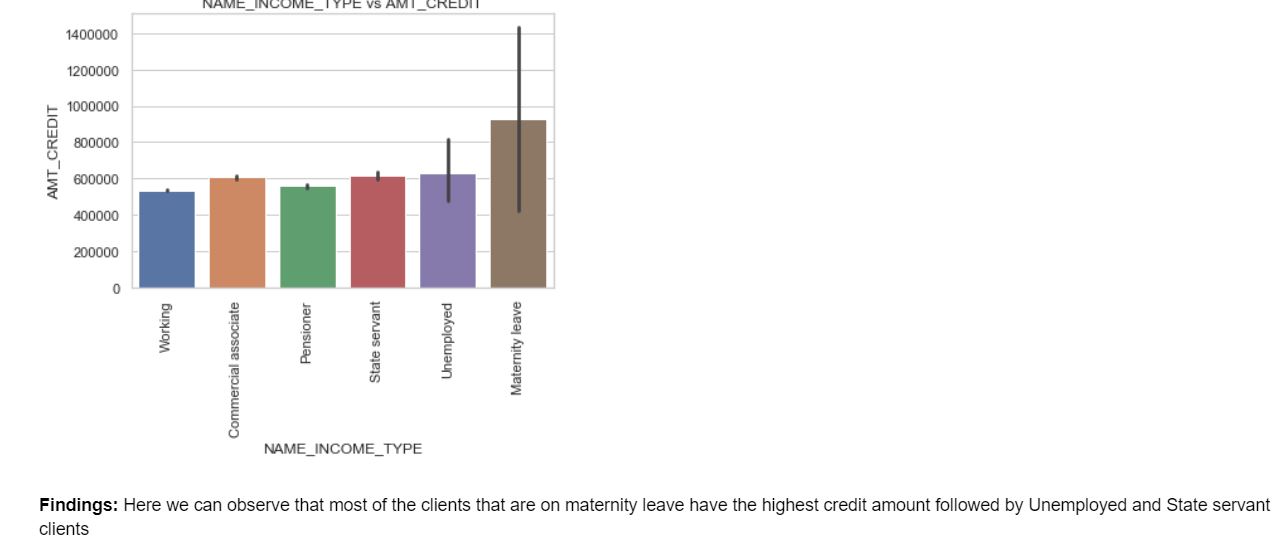
#### Analysis of Family Status versus Loan Credit Amount



#### Analysis of Income type versus Loan Credit Amount



#### Analysis of Housing status versus Loan Annuity

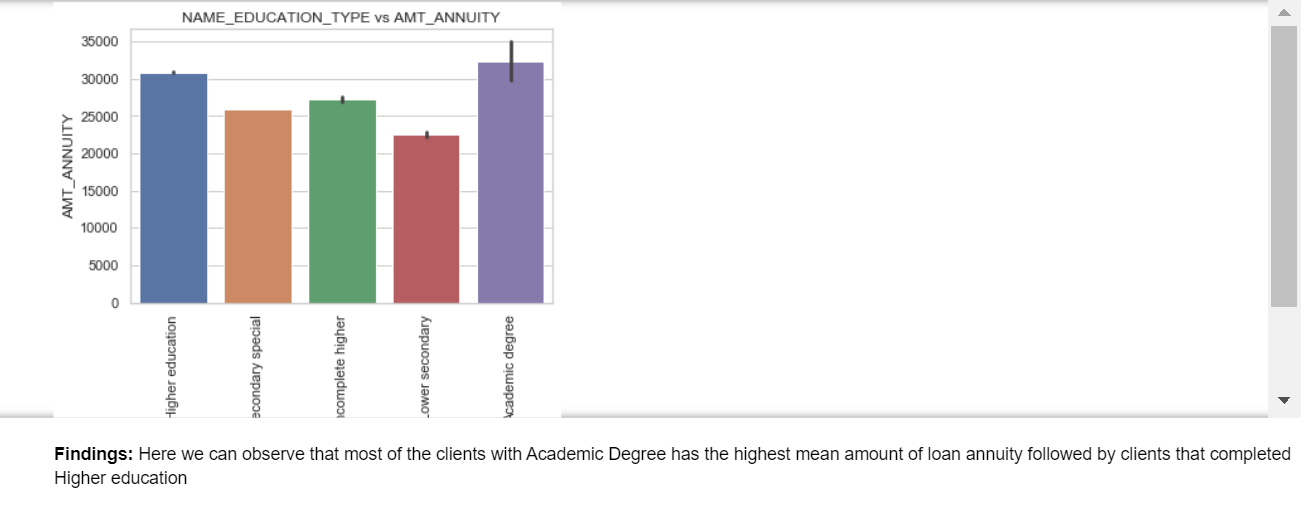


#### Analysis of Days employed versus Loan Credit Amount

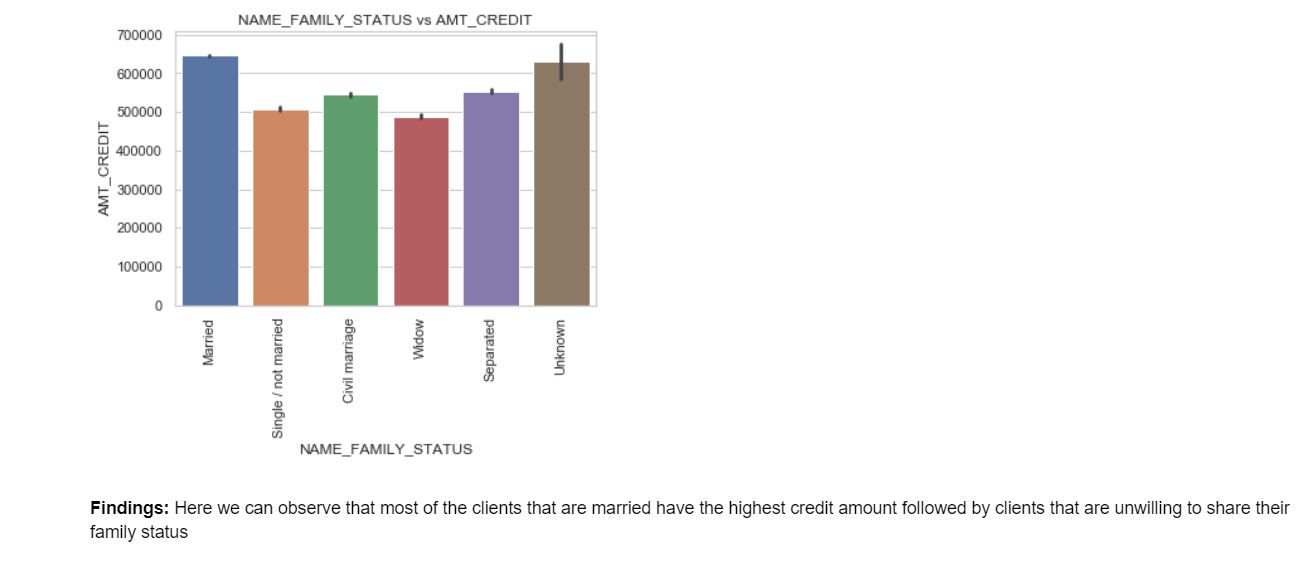


### Bivariate Analysis where 'TARGET' = 0

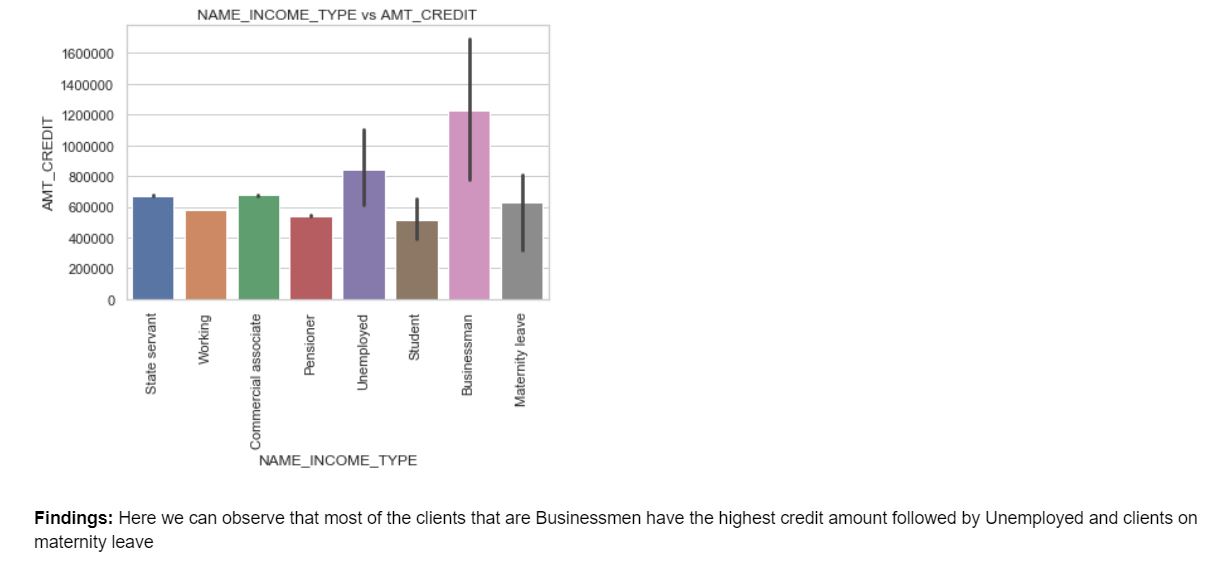
#### Analysis of Education Type versus Loan Annuity



#### Analysis of Family Status versus Loan Credit Amount



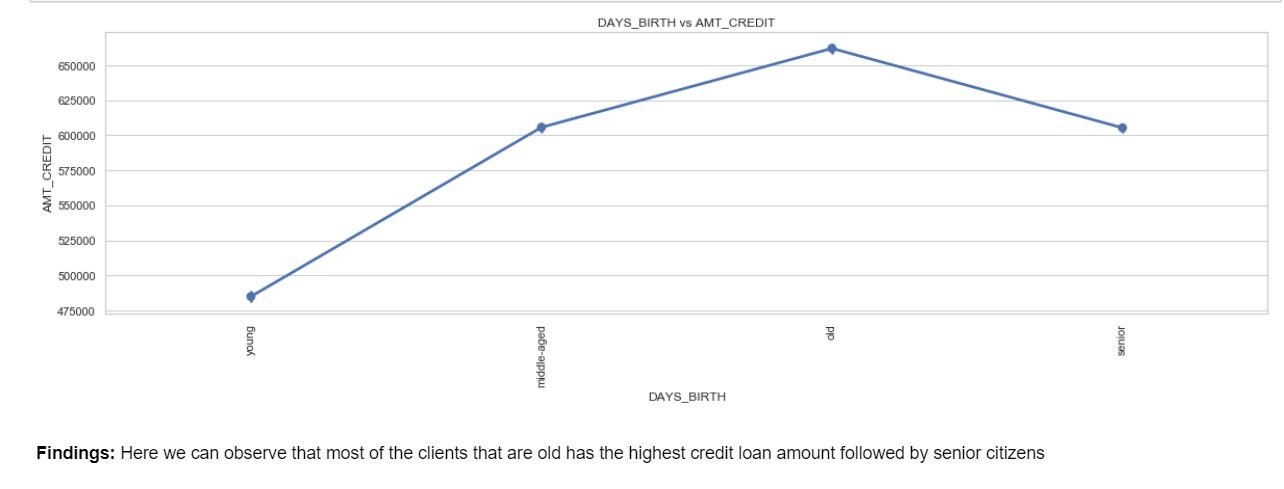
#### Analysis of Income type versus Loan Credit Amount



#### Analysis of Housing status versus Loan Annuity



#### Analysis of Days employed versus Loan Credit Amount

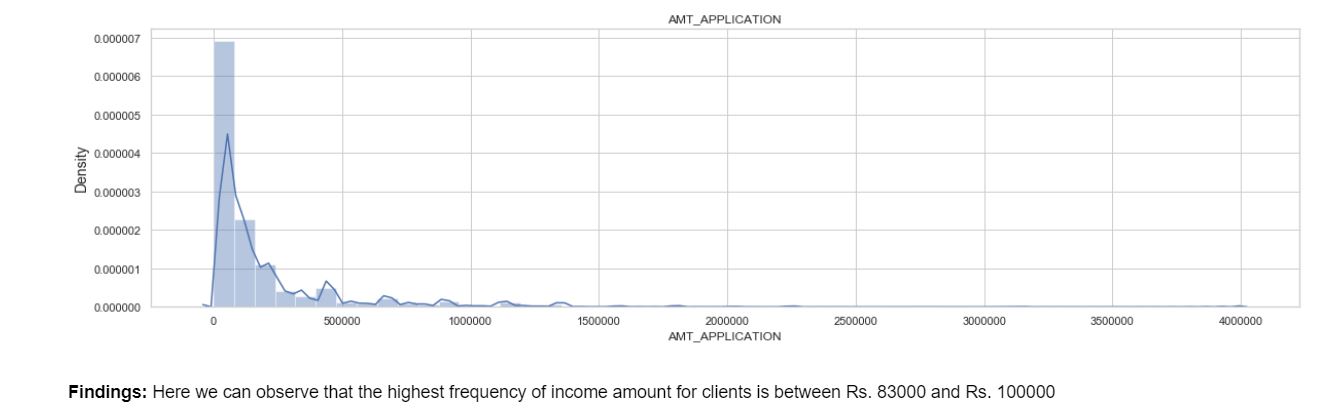


## EDA of Previous Application

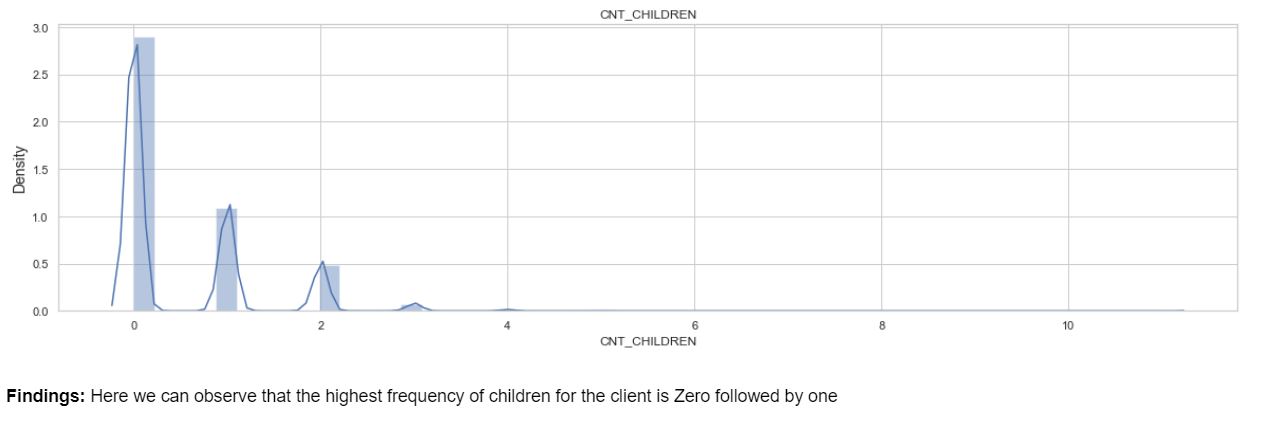
### Univariate Analysis on previous data where 'TARGET' = 1

#### Univariate Analysis for numerical variables

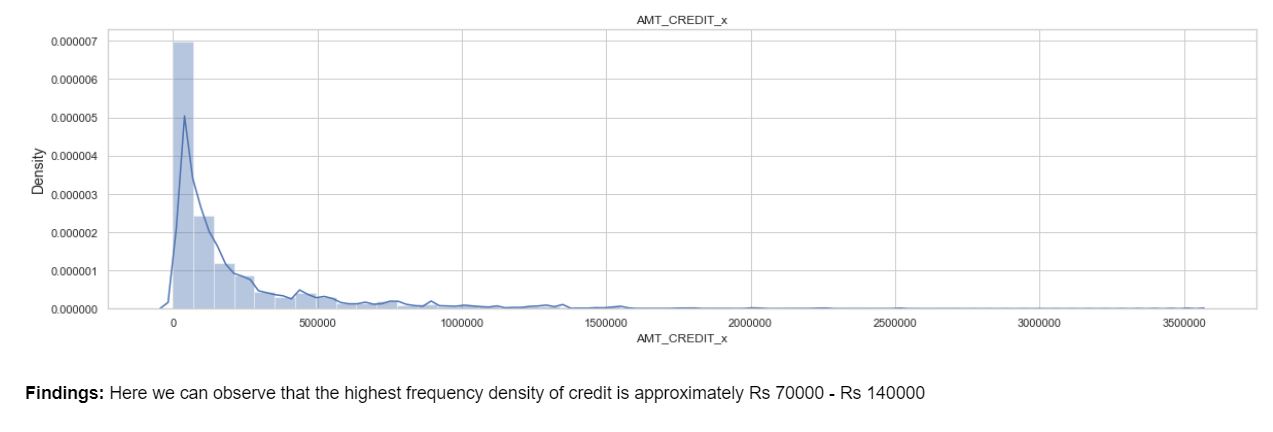
#### Analysis of Credit asked for the previous application



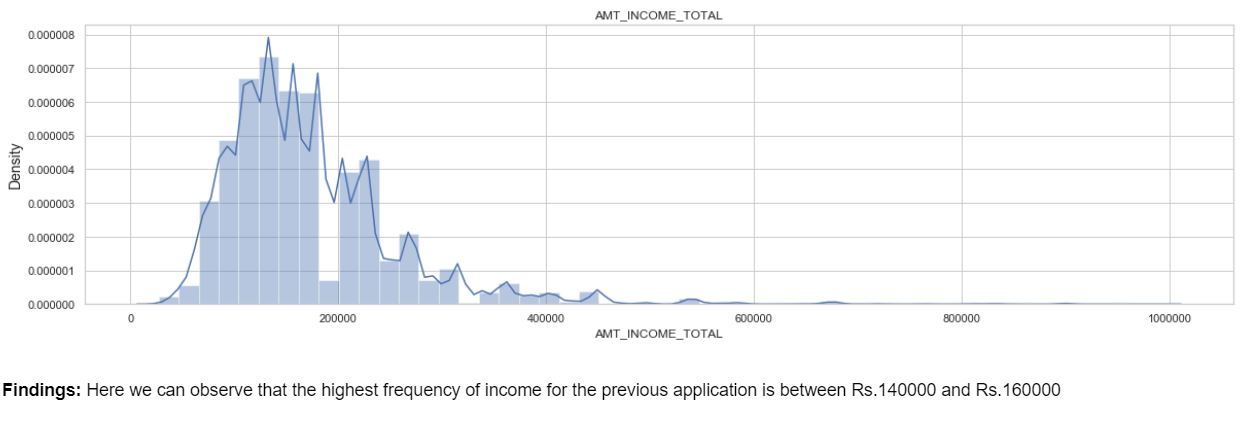
#### Analysis of Number of children for the client



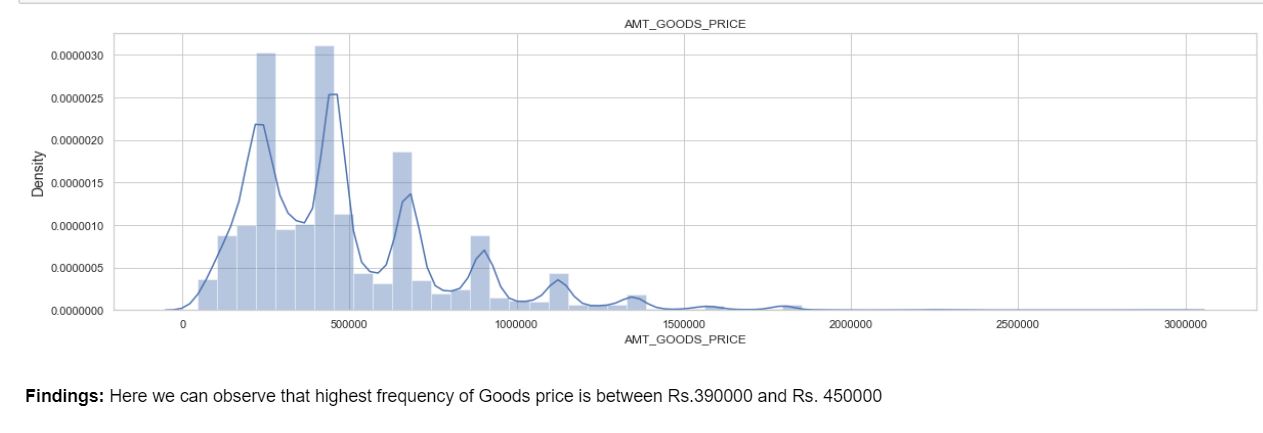
#### Analysis of Credit Amount of Loan for the previous application



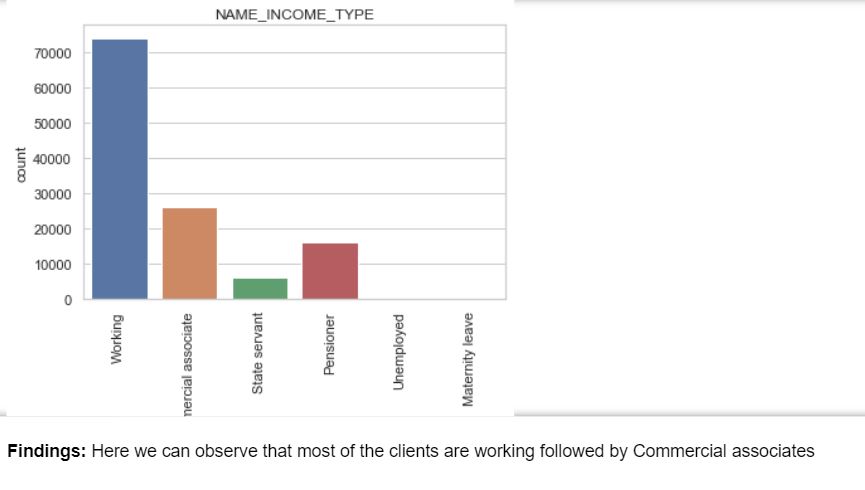
#### Analysis of Income of the previous application



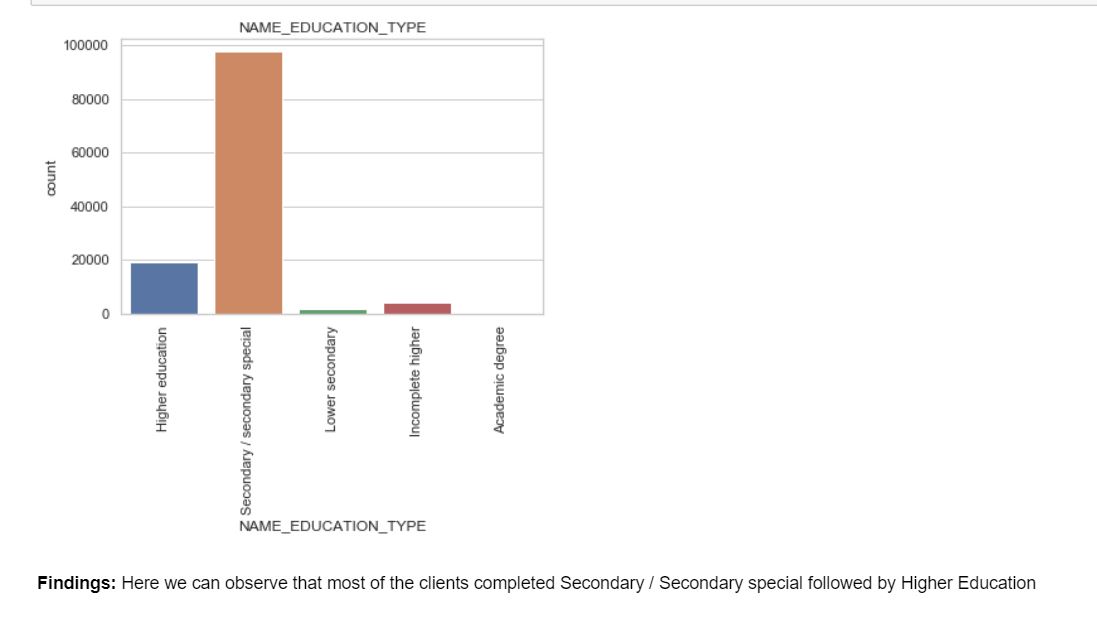
#### Analysis of Goods price of previous application



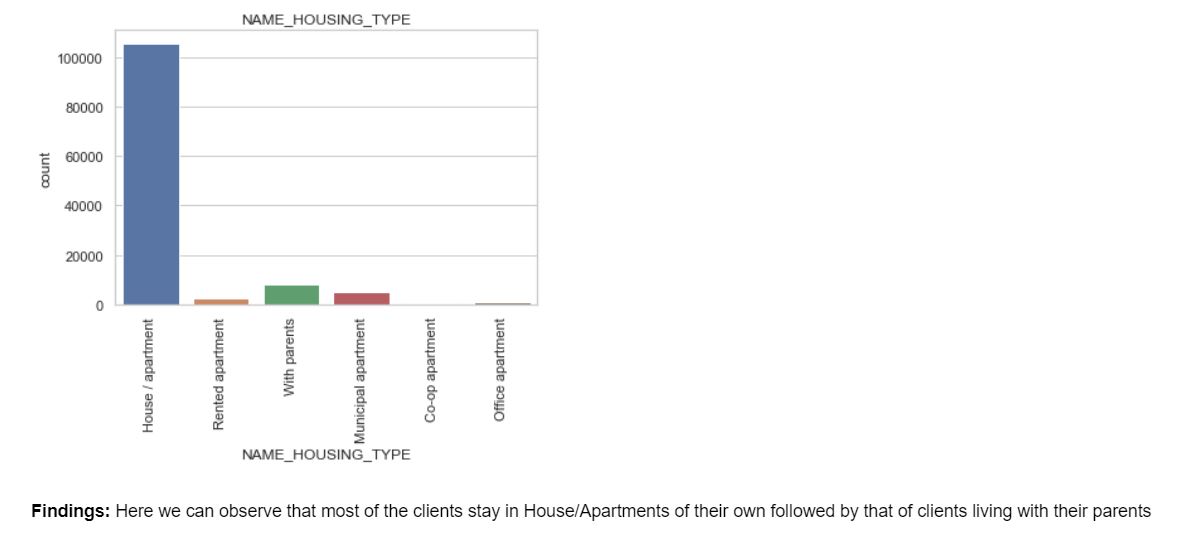
#### Univariate Analysis for categorical variables



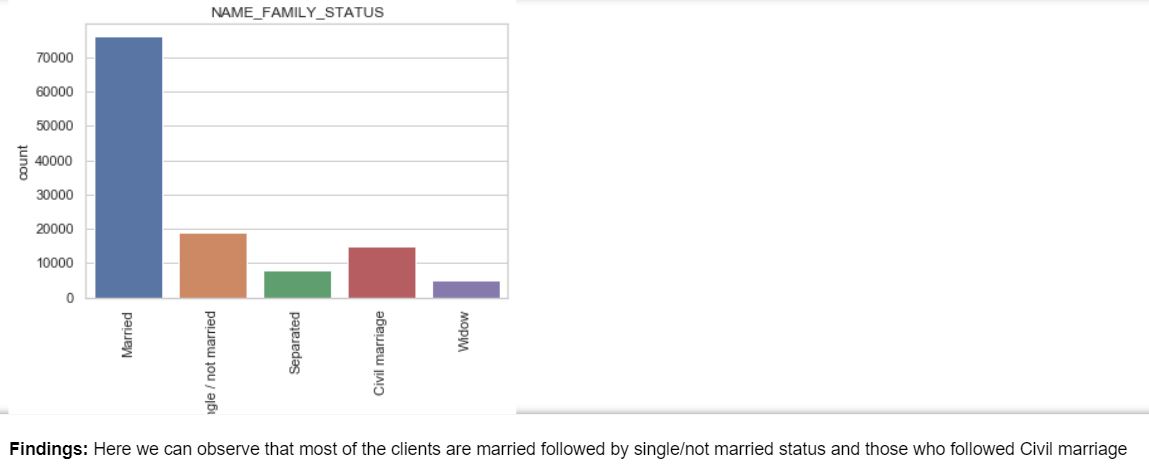
#### Analysis of Client Education Level



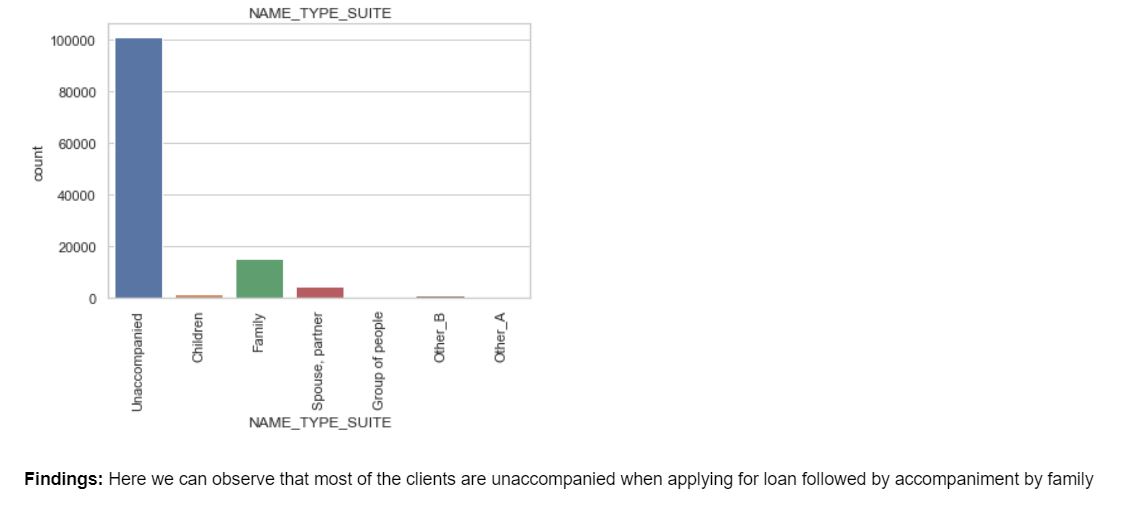
#### Analysis of Client Housing Type



#### Analysis of Client Family Status



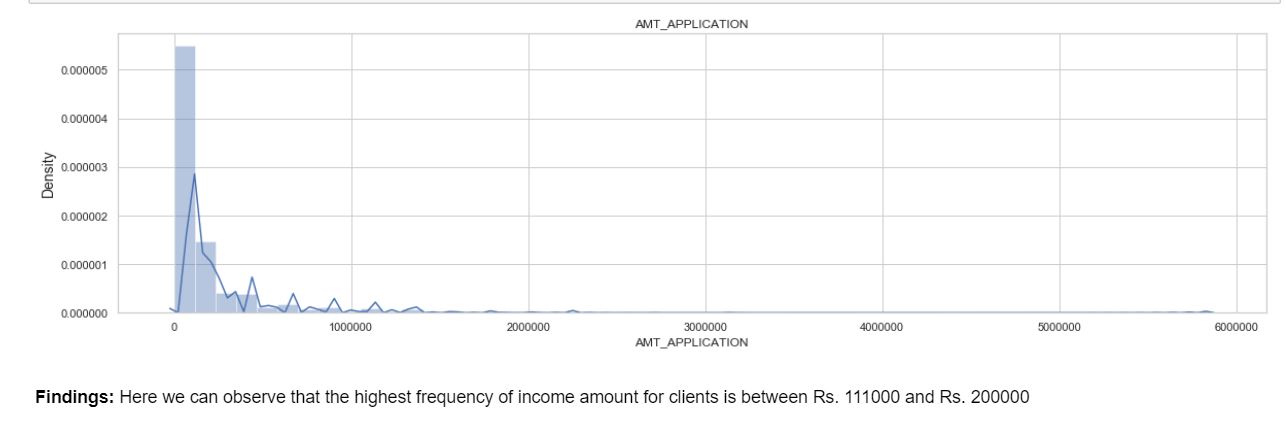
#### Analysis of Client Accompaniment status



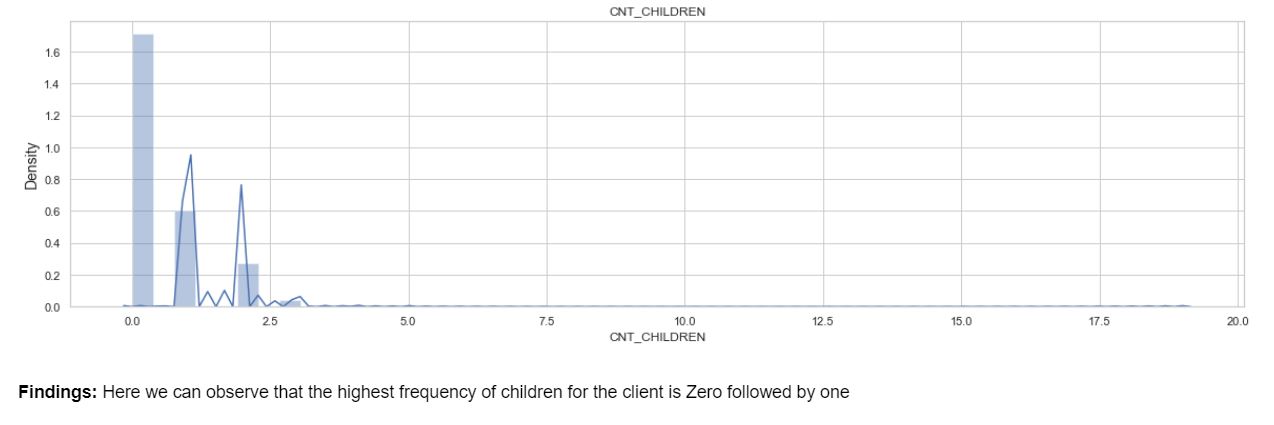
### Univariate Analysis on previous data where 'TARGET' = 0

#### Univariate Analysis for numerical variables

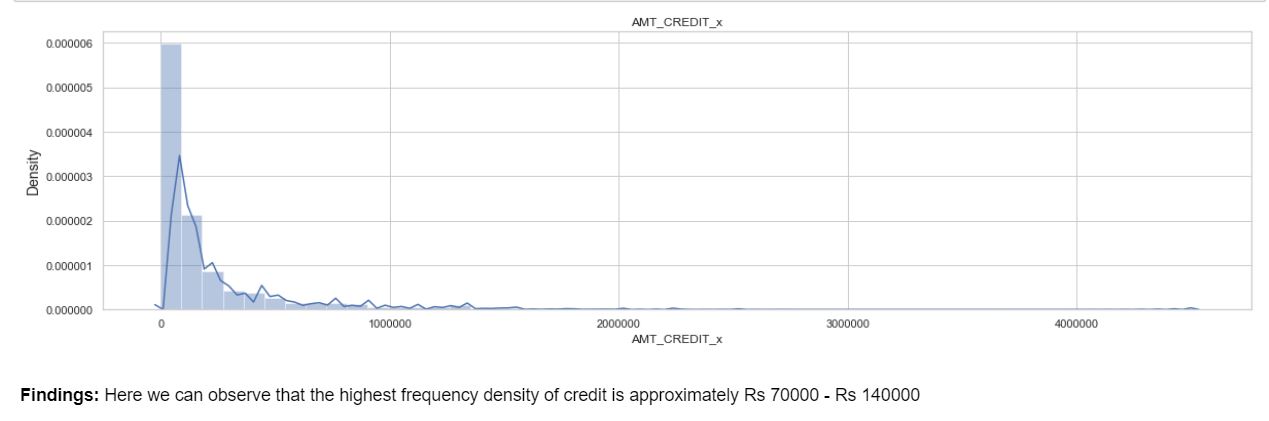
#### Analysis of Credit asked for the previous application



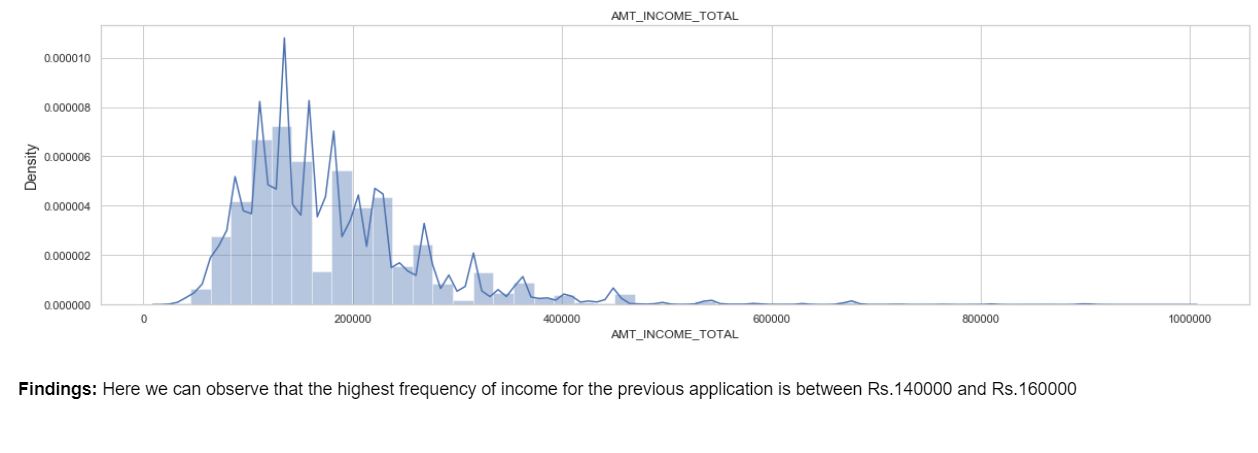
#### Analysis of Number of children for the client



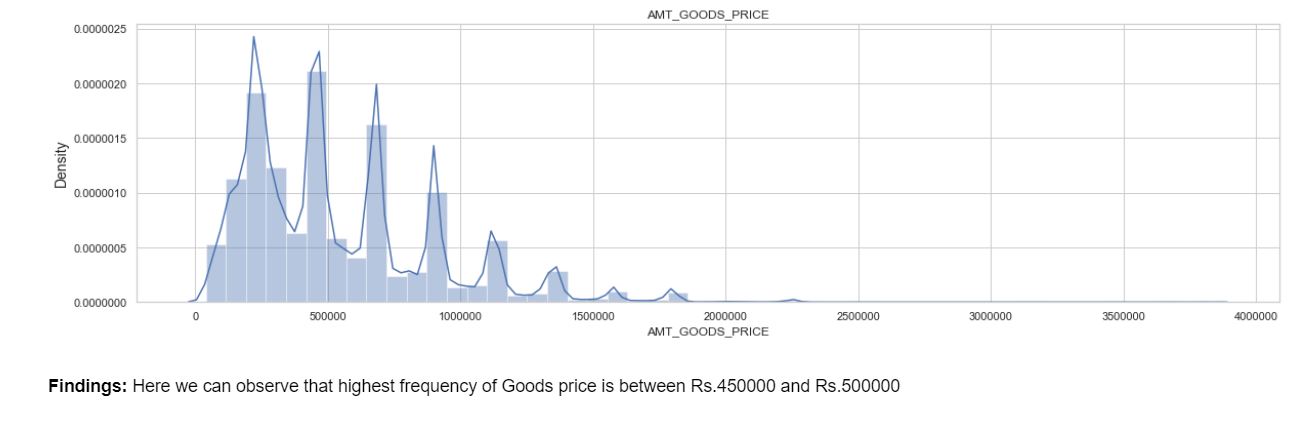
#### Analysis of Credit Amount of Loan for the previous application



#### Analysis of Income of the previous application

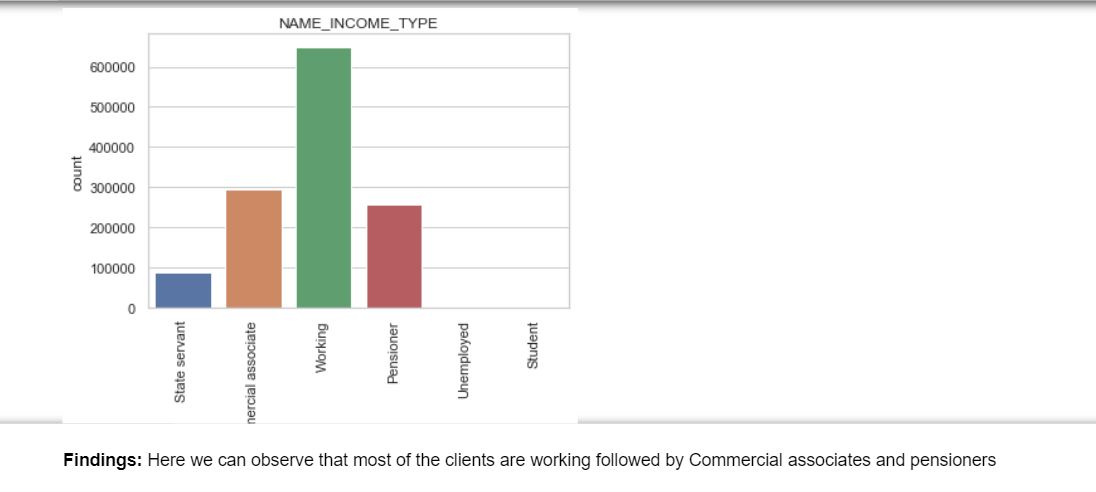


#### Analysis of Goods price of previous application

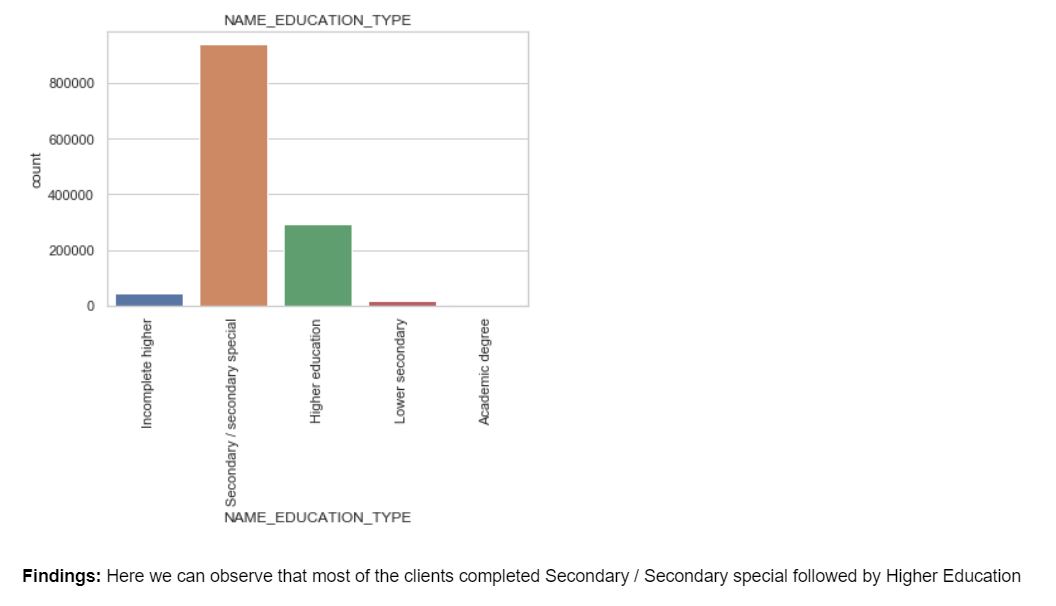


#### Univariate Analysis for categorical variables

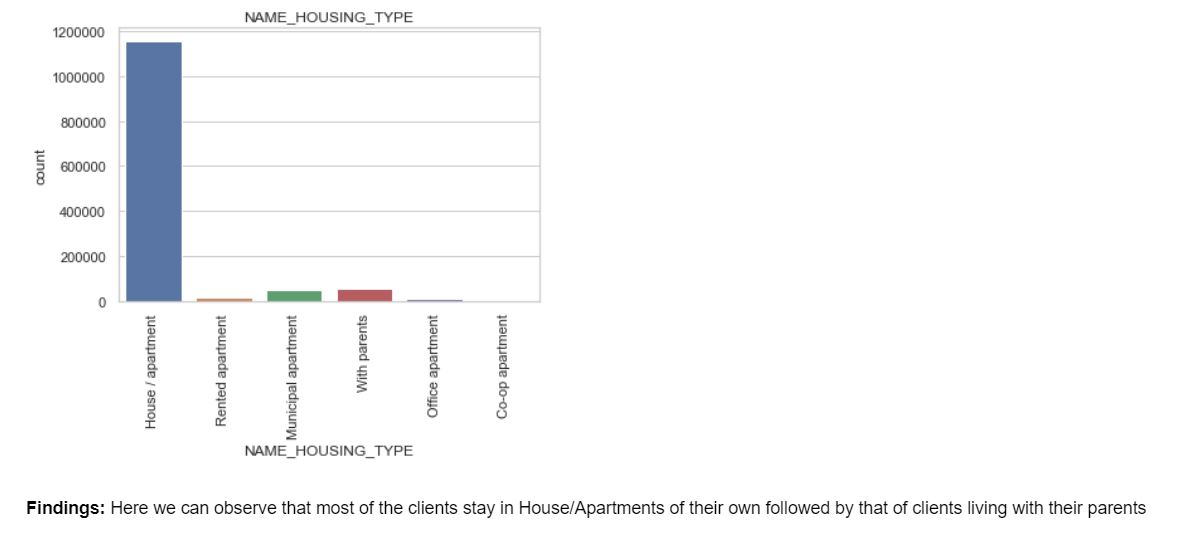
#### Analysis of Client Income Type



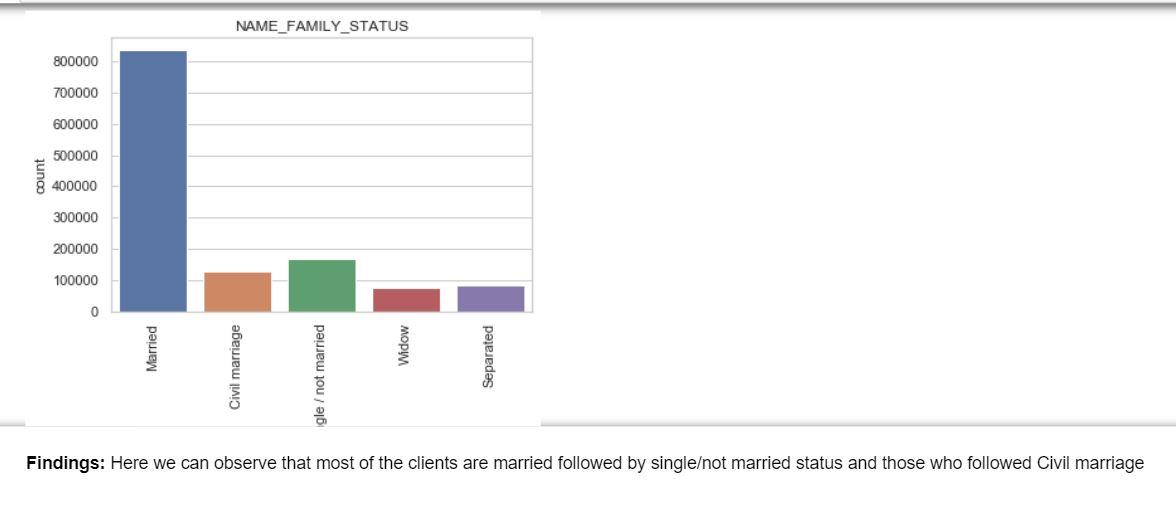
#### Analysis of Client Education Level



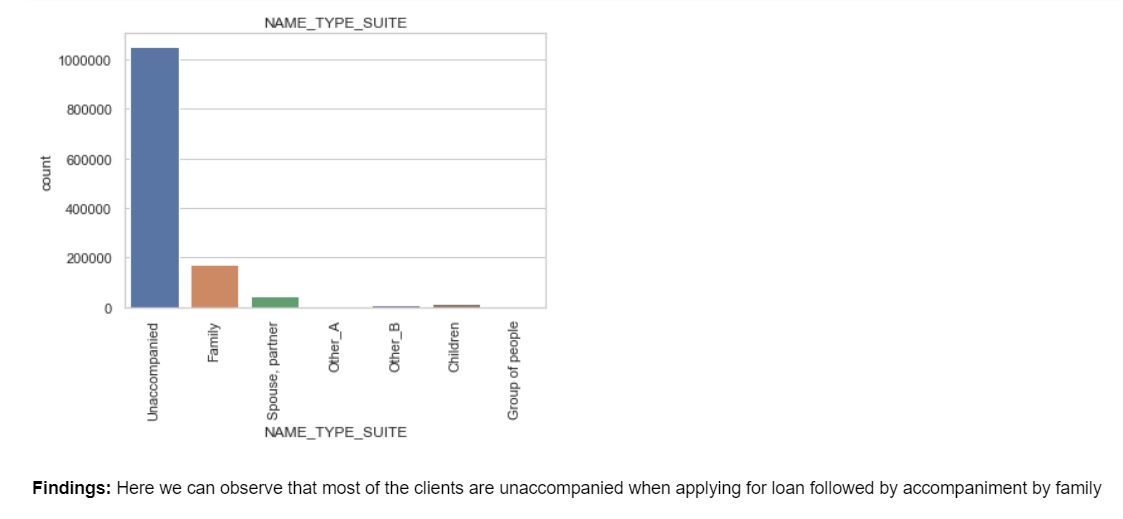
#### Analysis of Client Housing Type



#### Analysis of Client Family Status

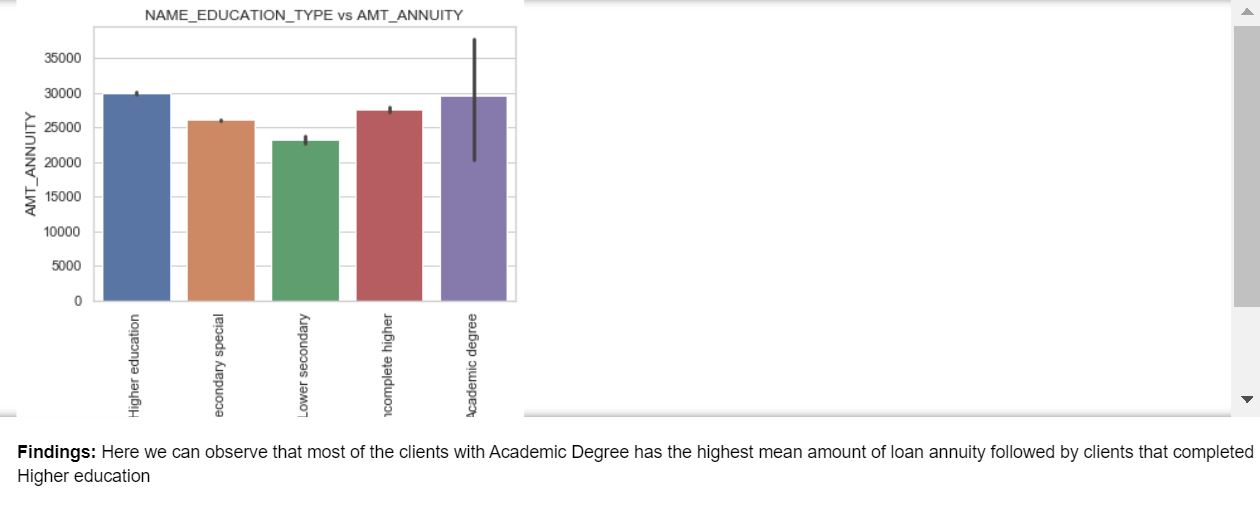


#### Analysis of Client Accompaniment status

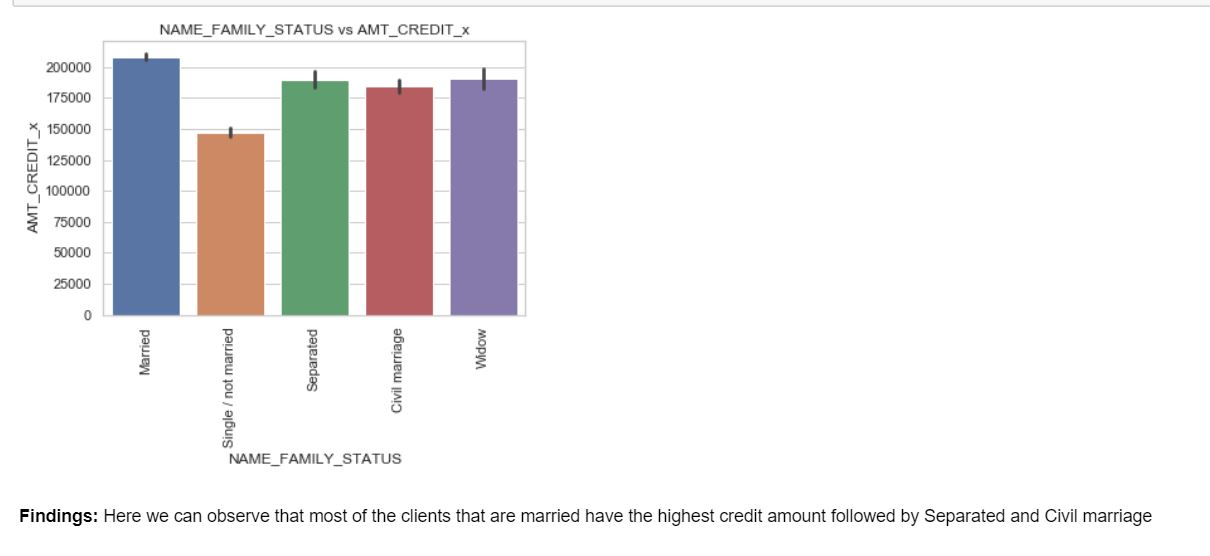


### Bivariate Analysis for previous application where 'TARGET' = 1

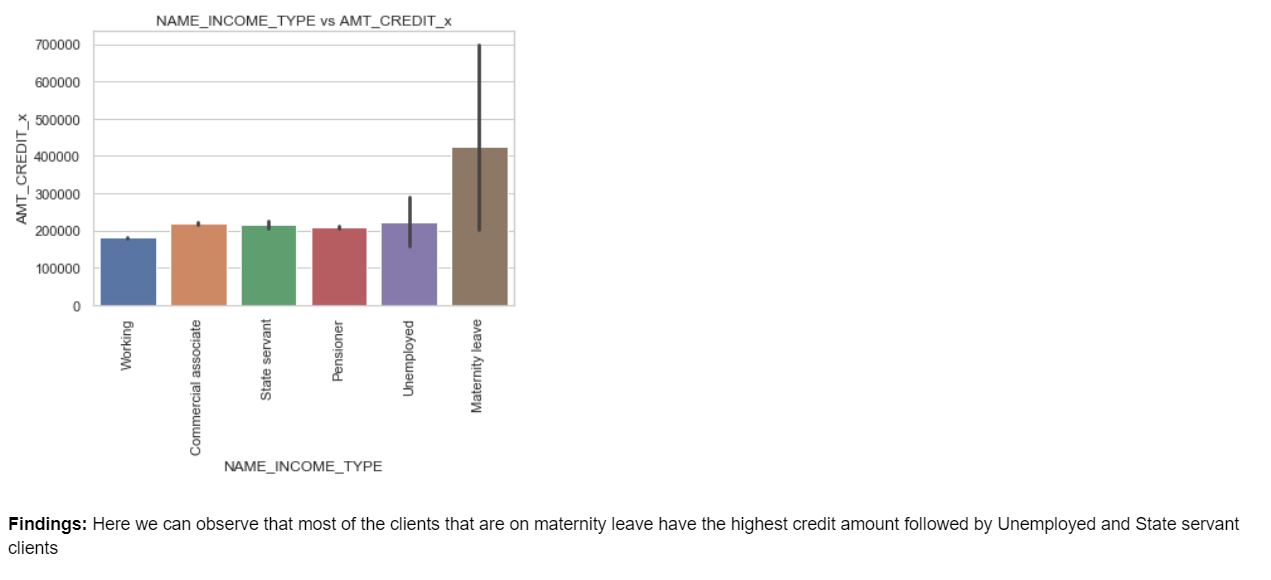
#### Analysis of Education Type versus Loan Annunity for previous application



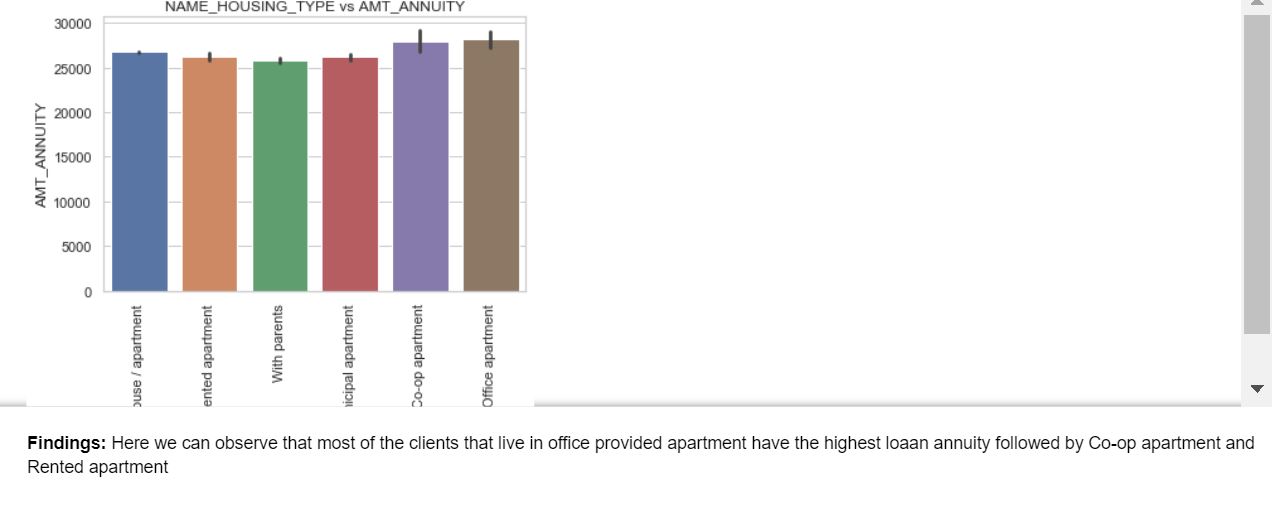
#### Analysis of Family Status versus Loan Credit Amount for previous application



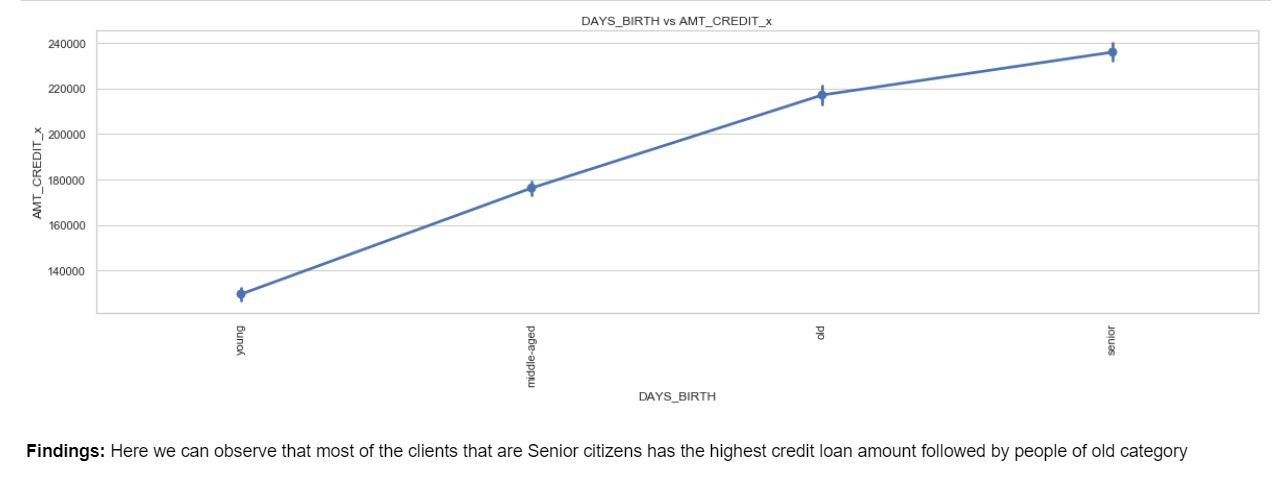
#### Analysis of Income type versus Loan Credit Amount



#### Analysis of Housing status versus Loan Annuity

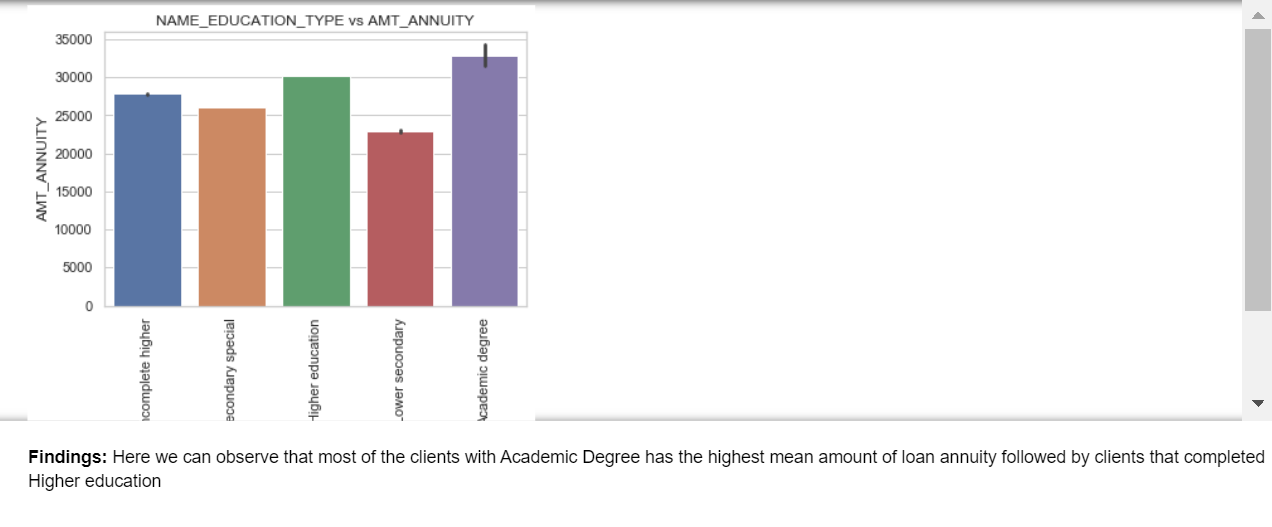


#### Analysis of Days employed versus Loan Credit Amount

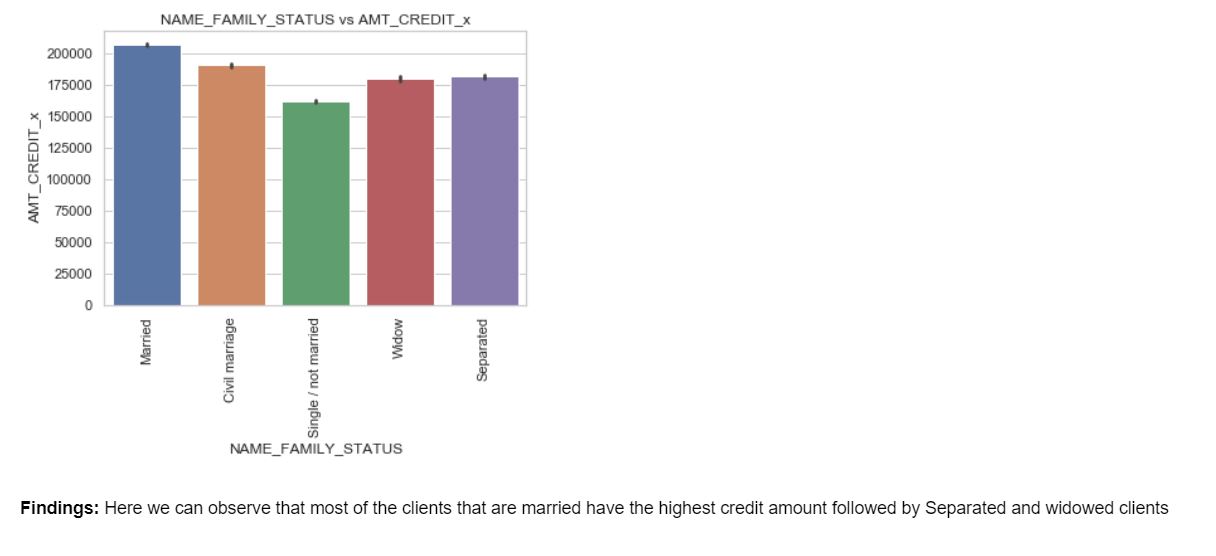


### Bivariate Analysis for previous application where 'TARGET' = 0

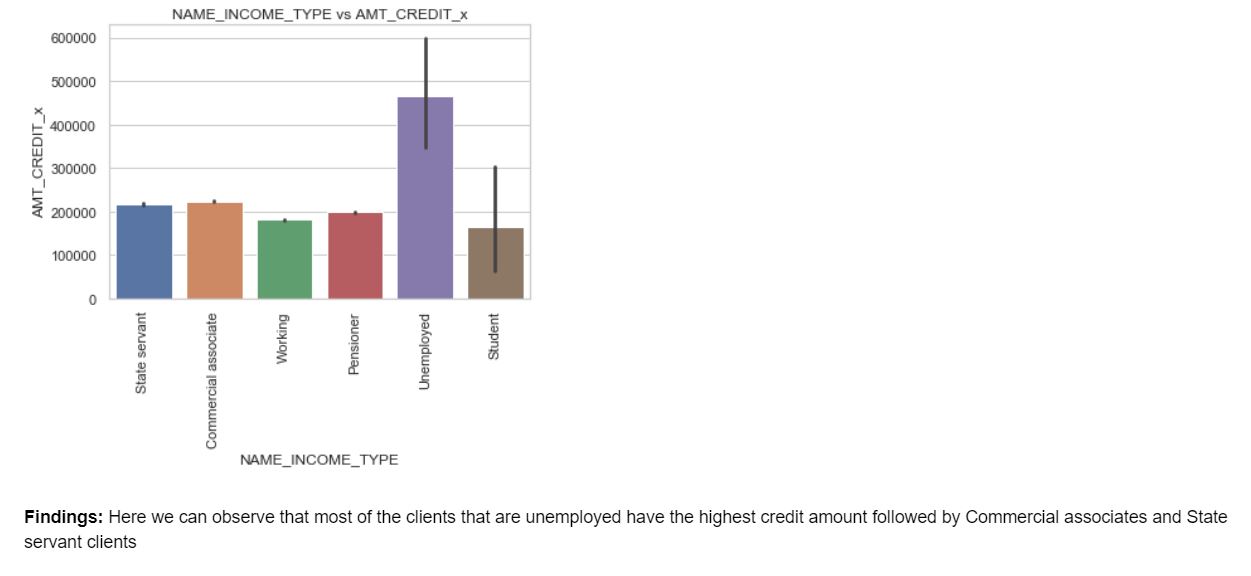
#### Analysis of Education Type versus Loan Annuity



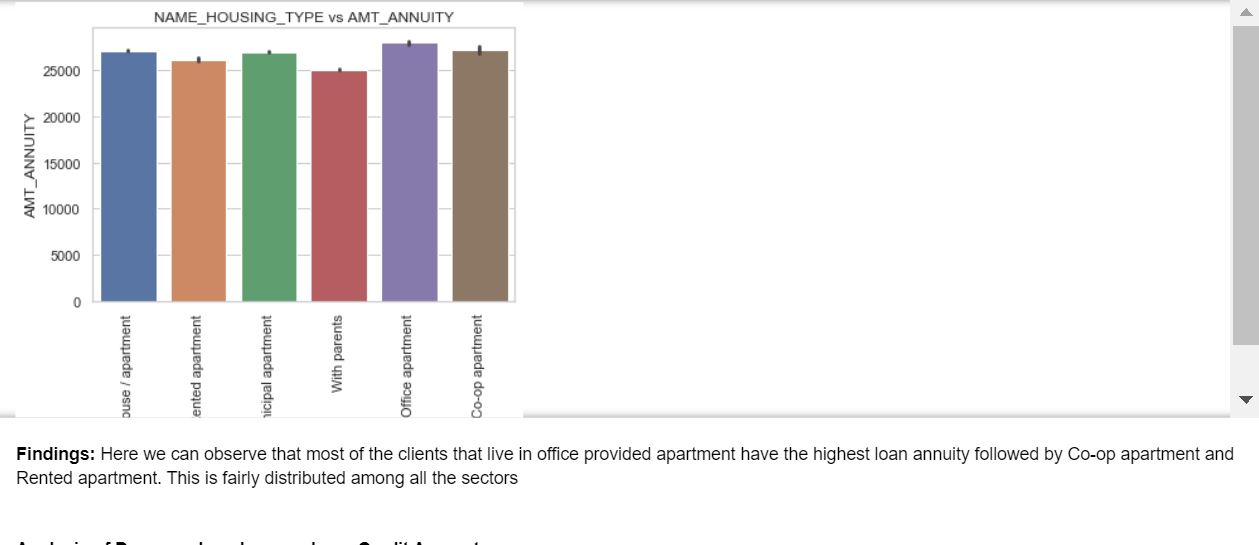
#### Analysis of Family Status versus Loan Credit Amount



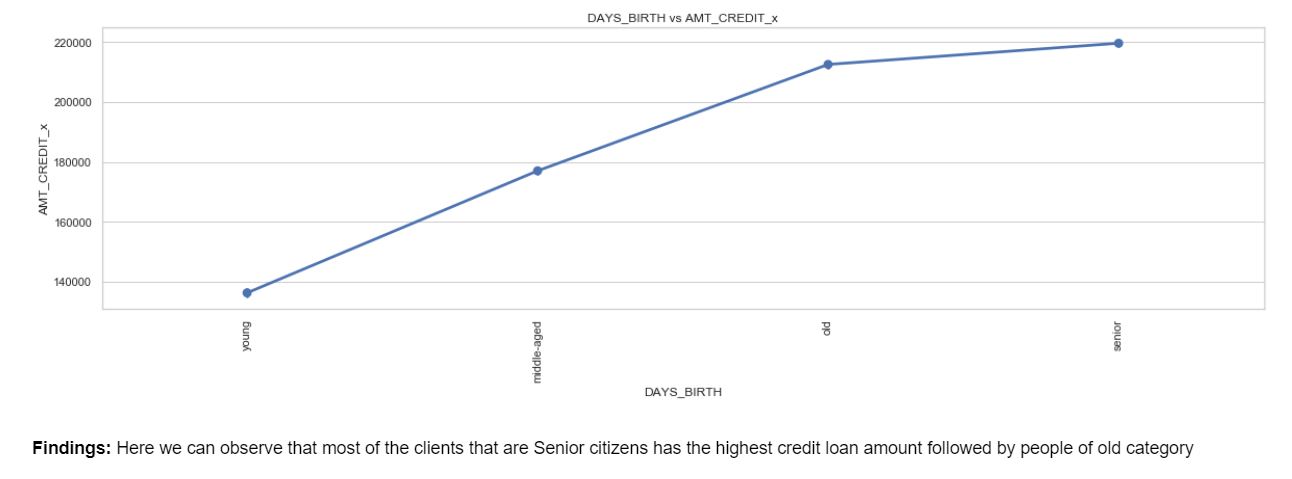
#### Analysis of Income type versus Loan Credit Amount



#### Analysis of Housing status versus Loan Annuity



#### Analysis of Days employed versus Loan Credit Amount



# Inferences From Exploratory Data Analysis

**1.** Overwhelming majority of credit applicants with risk involved are young working clients that earn 1-2 lakhs per month and have experience of 1.5 to 2 years, risk analysis should target this age group  
  
**2.** Majority of the applicants request for Credit between 5-6 lakhs for goods prices worth approximately 5 lakhs. These applicants are likely to default  
  
**3.** Most of the applicants that leads to credit default have loan annuity between Rs.25000-30000.  
  
**4.** Defaulters are applicants that completely higher secondary education, own a house and are married. We can conclude that experience of owning capital can still lead to loan defaults  
  
**5.** Applicants that are unaccompanied during application process overwhelming shows the possibility of defaulting. Risk analysis should be more focused among this group  
  
**6.** Applicants that are likely to pay the loan amount earn comparatively the same amount as those who default but these applicants are likely to have more job experience  
  
**7.** Applicants that pay the loans on time apply for lesser credit loan between Rs.200000 - Rs 350000 compared to the defaulters  
  
**8.** Credit loan amount of applicants in other cases are more inline with the price of the goods than that of defaulters  
  
**9.** Majority of applicants in non- defaulters group are educated, married and owns a house or an apartment  
  
**10.** Among the applicants having difficulty paying the loan amounts, a significant portion of those on maternity leaves have high loan annuity amounts  
  
**11.** Applicants that complete an academic degree and high annuity amount are more likely to default. This maybe due to pre-existing loans that the applicant may have paid off in the past.  
  
**12.** Senior Citizens and applicants of old-age group have higher loan amount and are more likely to default  
  
**13.** Married applicants generally have the highest load credit amount and are more likely to belong to non-defaulters group  
  
**14.** A large number of applicants of young age group have lower credit amount and show promising signs of being a non-defaulter. By targeting these sectors through marketing, higher revenue can be generated  
  
**15.** Comparing the previous application and current application data, most of the applicants apply for higher amount of loan in the current period and thus more likely to default  
  
**16.** There is no significant jump in the salary of defaulters between the current and previous application although they submit for relatively higher amount of loan. Applicants that apply for more loan without salary jump is a red flag  
  
**17.** Applicants with one or more children are more likely to pay the loan than applicants with no children. Defaulters also tend to not report the number of family members  
  
**18.** Comparing previous and current application, we can infer that the unemployed customers that recently procure a job are more likely to pay the loans on time  
  
**19.** Applicants that move from Rented apartments to their own house/apartments between applications are more likely to be a non-defaulter  
  
**20.** There is no different in trend for defaulters that belong to senior or old age group between current and previous applications

# Conclusion

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. We aim to understand the driving factors behind loan defaults for different types of applicants.  
Analysis was conducted on the prospective clients based on the their current as well as previous application data. Recommendations and insights were provided for the same.