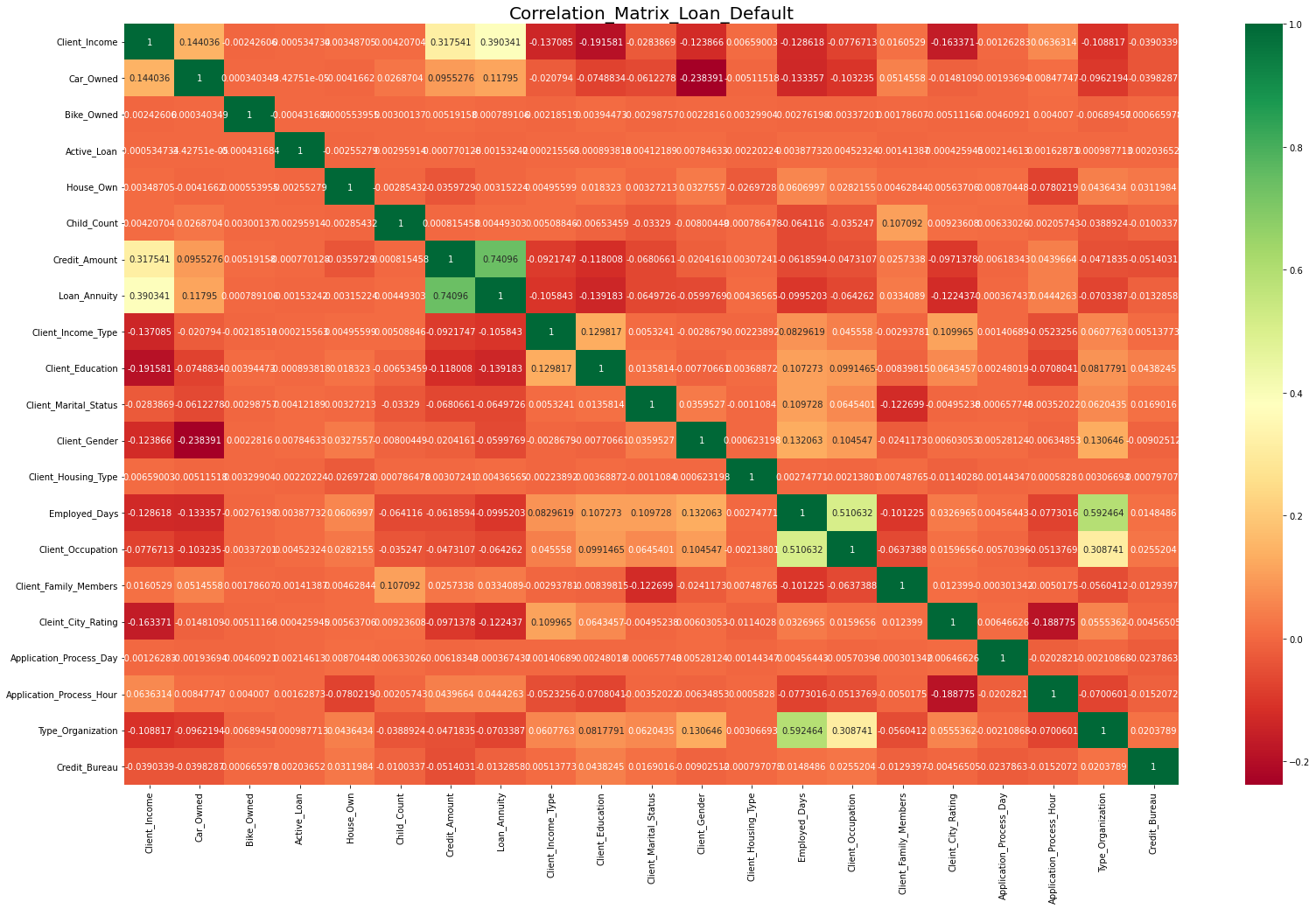
**Capstone Project: Loan Defaulter (By Jagesh Chauhan)**

**Loan Defaulter: Exploratory Data Analysis based on Prediction**

****

1. **Visual representation of the correlation by Heatmap**

**Based on the correlation, we can infer that below columns have high correlation –**

* ***Credit\_ Amount, Loan\_ Annuity, Employed\_ Days, Client\_ Occupation, Type\_ Organization***

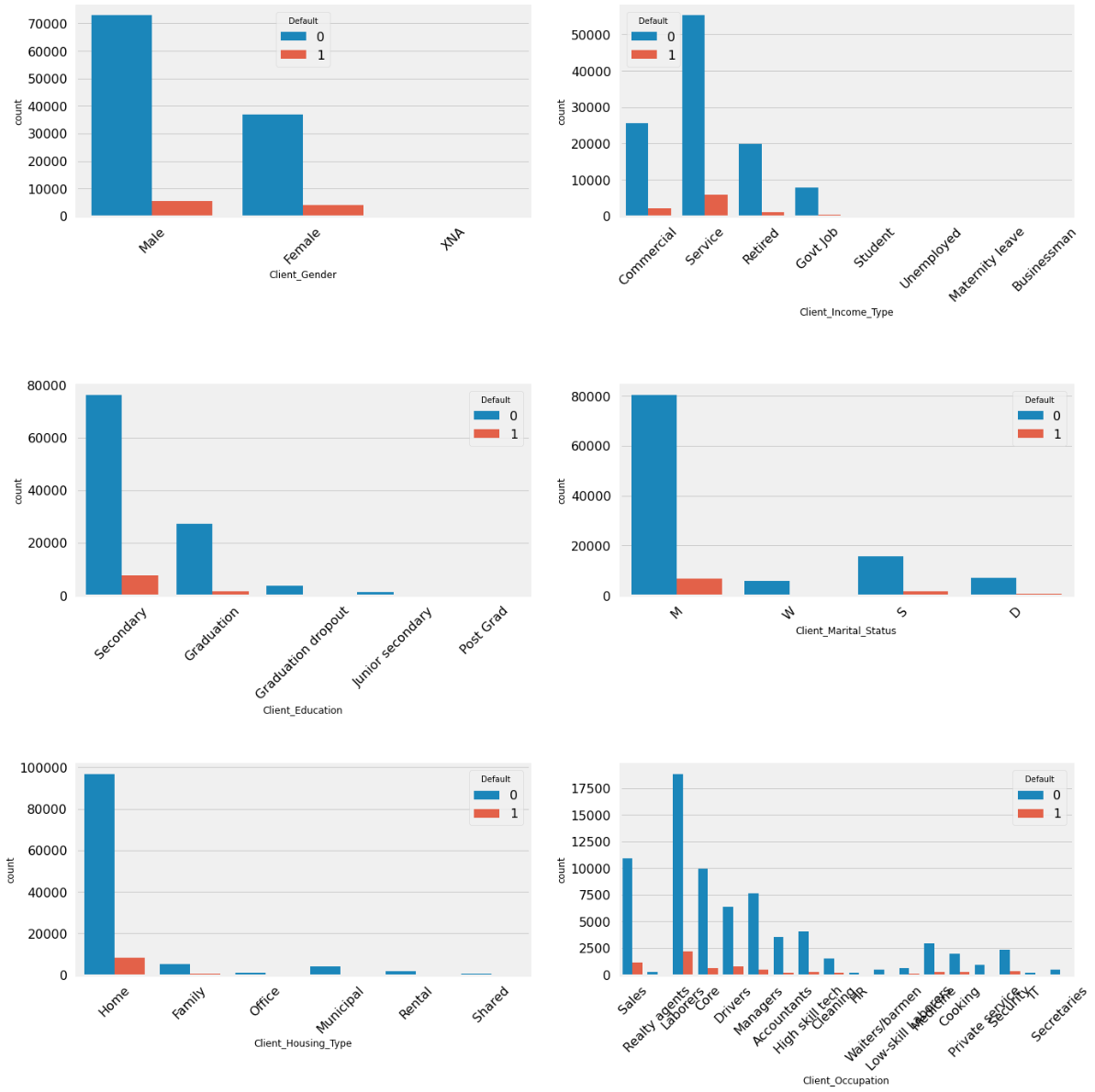
**Below are the columns have a high skewness bases on the correlation –**

* ***Client\_ Income, Client\_ Family\_ Members, Child\_ Count, Client\_ Housing\_ Type, Default***

**In the above image you can see that -1 in the Red highlighted that have a very lower rate correlation.**

**Rest miner negatives & miner positives in yellows that you can see in above image.**

### Analyses Categorical variables with respect to Target variable



##### ***From the above plot we can see that****,*

##### ***Female customers pay loan amount on time and banks can target more female customers for lending loan.***

##### ***Service customers can be targeted to lend loans as they have higher percentage of making payments on time.***

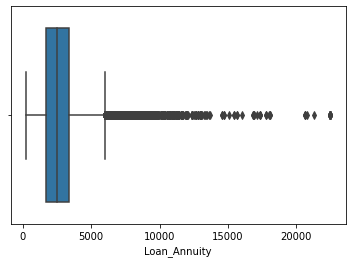
##### ***Customers with secondary education and Graduation are most likely to make payments when compared to customers with Post Grad.***

##### ***Married customers have paid loan amount on time when compared to widows.***[***¶***](https://www.kaggle.com/code/shrutidandagi/eda-bank-loan-default#Married-customers-have-paid-loan-amount-on-time-when-compared-to-widows.)

##### ***Customers owning House/family are most likely to make payments on time compared to those living in Rental and Shared.***

* ***Realty agents have high repayment percentage on time as bank will think to lending a big loan amount to them.***
* ***Laboure’s have high repayment percentage. Hence banks can think of lending small amount loans to them.***

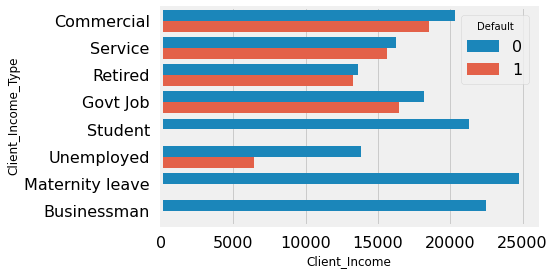
**Distribution of Loan \_ Annuity**

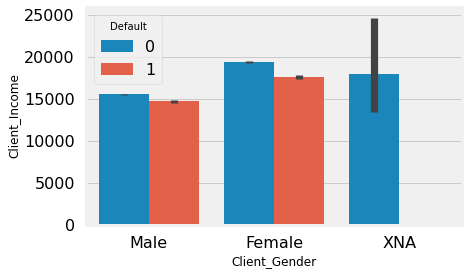
****

***We can infer from the above image that-***

* ***We take a look at Loan\_ Annuity column as we can see that there are outliers at 22500. But there is no much difference between the mean and median, i.e., mean is 2712.48 and median is 2499.75.***
* ***After the describe of Loan\_ Annuity columns we can impute the outliers with Median here.***

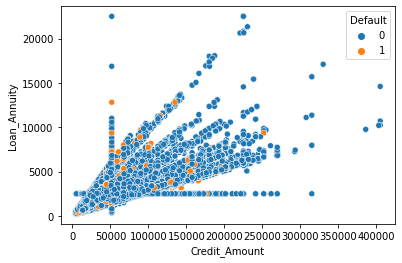
**Client\_ Income and Client\_ Income \_Type**

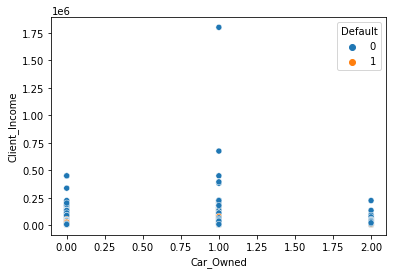
****

****

* ***Most of applicant are working.***
* ***Applicants on Maternity Leave, Businessman and Student have lowest percentage of Defaulters.***
* ***Commercial, Service, Retired and Govt job have a higher percentage of Defaulters.***
* ***Female applicant is more than the male applicant***
* ***Defaulter percentage is higher for female applicant.***

#### Insights from Numerical Bi- Variate Analysis

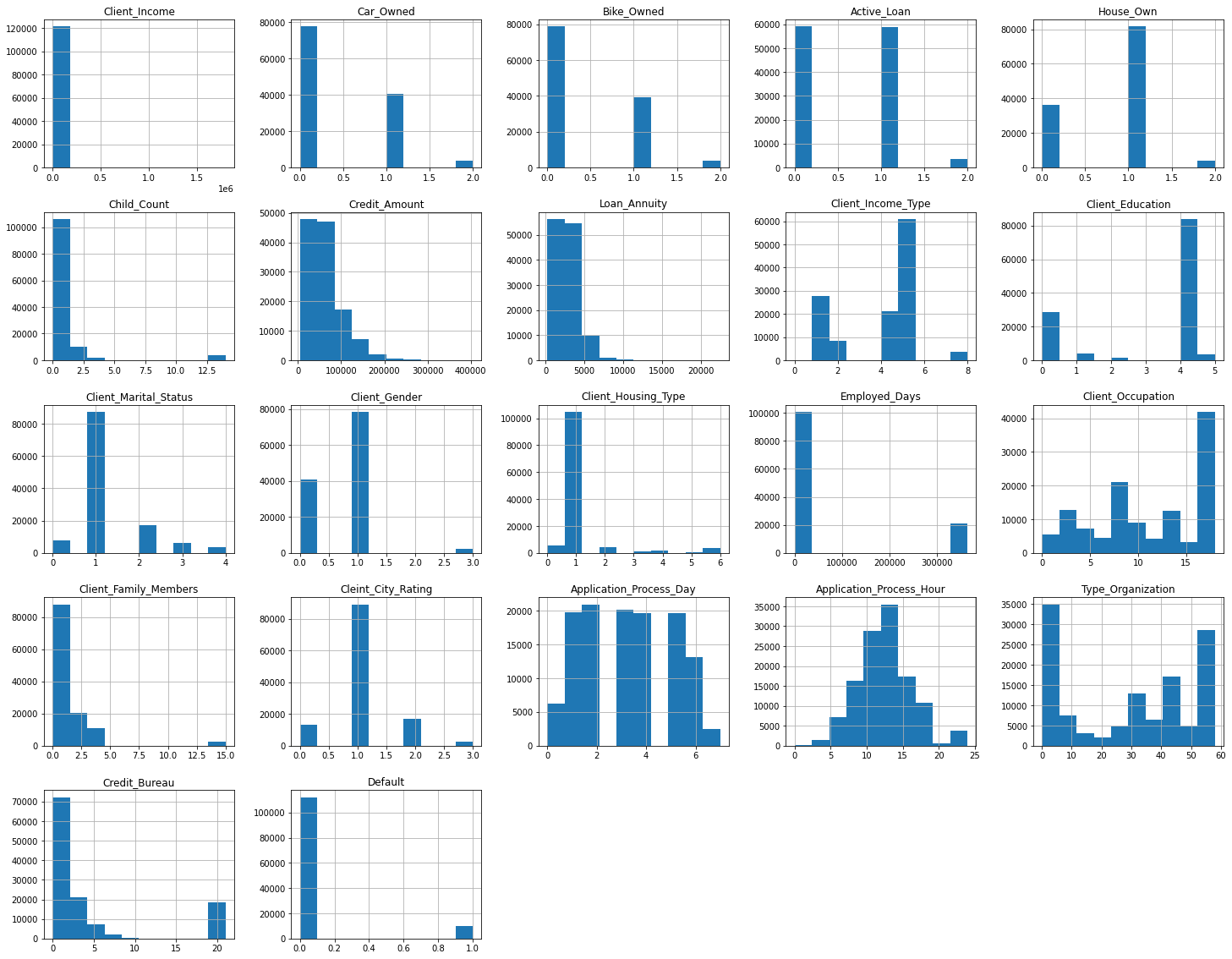
******

******

#### *Insights,*

* ***There are more re-payers’ rate when the loan annuity is more than 10k, that means applicants with 10k annuity will more likely to repay their loans, but their frequency is low.***
* ***Whose applicant’s income is 50k or more than 50k who has 1 car and more than 2 cars.***

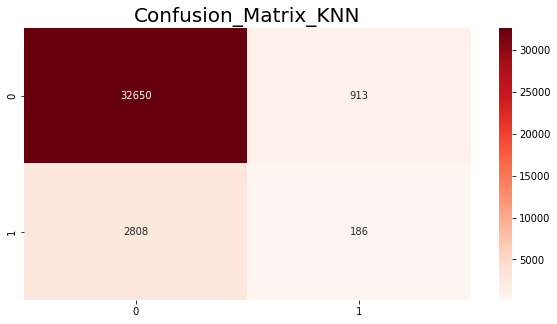
#### B. Univariate and Bivariate Analysis by using Histogram

****

***From the above Histogram plot we can say that following below points:***

* ***Whose income is less than Rs.40000. He/ She doesn’t have home and whose income is above Rs.40000. He/ She does have home.***
* ***Whose income is less than Rs.60000, his loan is active and whose income is above Rs.60000, his loan is not active.***
* ***Whose income is above Rs.40000 that amount is credited more than Rs. 2.5 lac. And whose income is less than Rs.20000 that amount is credited less than Rs.1 lac.***
* ***Whose income is more than Rs. 20000 that his/ her is not default and whose income is less than Rs.20000 that She/ He is default in term of paying loan.***
* ***Whose income is more than Rs.20000 as client city rating is \*2.0 and whose income is less than Rs.20000 as client city rating is \*1 – 3\*.***
* ***Whose income is more than Rs.20000. So, client family member count is 2 and whose income is less than Rs. 20000. So, client family member count is more than 2.***
* ***Whose income is less than Rs. 20000 as he/she is socially defaulting more than 1 times and whose income is more than Rs.20000 as he/she is not defaulting socially.***

1. **Visual representation of Logistic Regression and KNN by using Confusion Matrix**
2. **Visual representation of KNN-**

****

***We can infer below point from above image:***

* ***Top left quadrant= Actual value is [0] and predictive value is [0] that is True positive and the data calculated is 32650.***
* ***Bottom left quadrant = Actual value is [1] and predictive value is [0] that is False negative and the data calculated is 2808.***
* ***Top right quadrant = Actual value is [0] and predictive value is [1] that is False positive and the data calculated is 913.***
* ***Bottom right quadrant = Actual value is [1] and predictive value is [1] that is True negative and the data calculated is 186.***

1. **Visual representation of Logistic Regression-**

****

***We can infer below point from above image:***

* ***Top left quadrant= Actual value is [0] and predictive value is [0] that is True positive and the data calculated is 33563.***
* ***Bottom left quadrant = Actual value is [1] and predictive value is [0] that is False negative and the data calculated is 2994.***
* ***Top right quadrant = Actual value is [0] and predictive value is [1] that is False positive and the data calculated is 0.***
* ***Bottom right quadrant = Actual value is [1] and predictive value is [1] that is True negative and the data calculated is 0***

***Thanks,***

***Jagesh Chauhan***