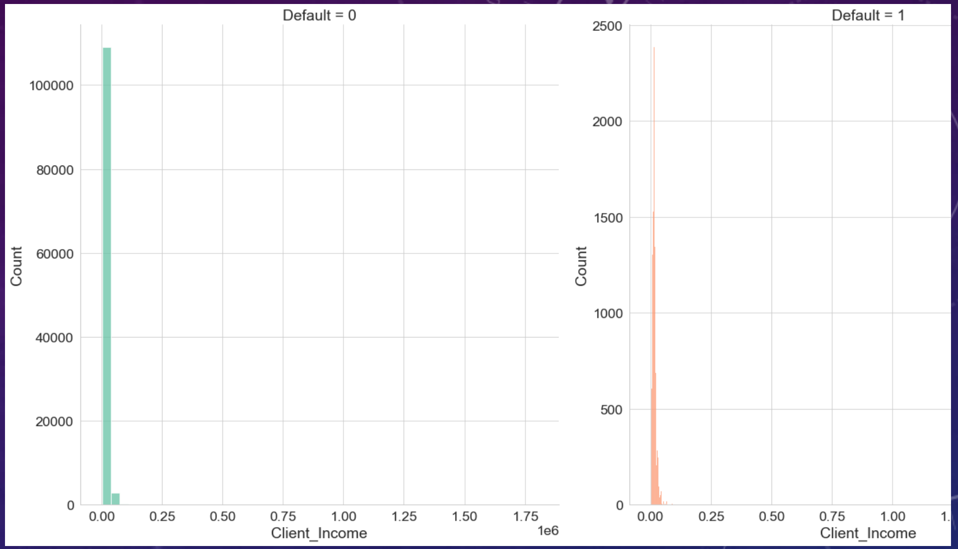
NBFI VEHICLE LOAN REPAYMENT ANALYSIS

INTRODUCTION

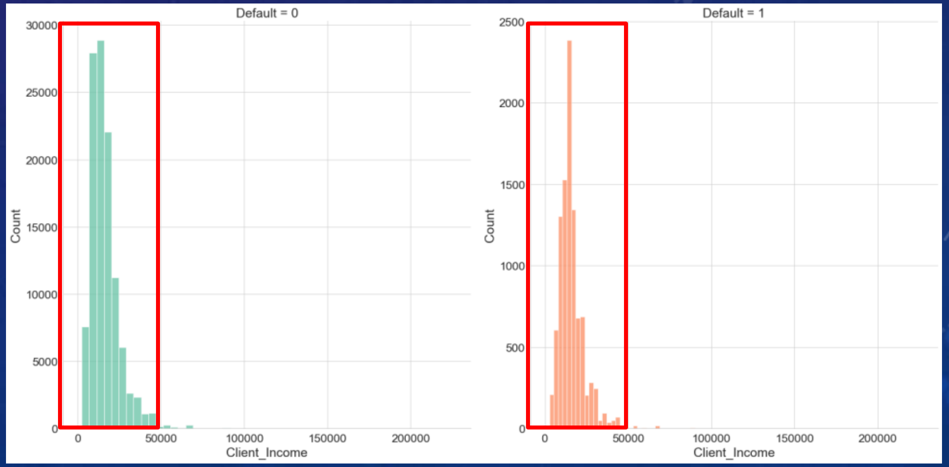
- Background:
  - A Non-Banking Financial Company (NBFC) is experiencing a decline in profitability due to a rise in defaults within the vehicle loan category. In response , the company wants to construct a prediction model by this dataset to predict the likelihood of a client defaulting on their vehicle loan payment.
- Objective:
  - Our objective for this project is to visualize pattern difference between the defaulted customers and non default customers on some of the variables to verify what variables can be kept and what can be removed for the further model-building process, in order to minimize noise during the model-building process
- Dataset:

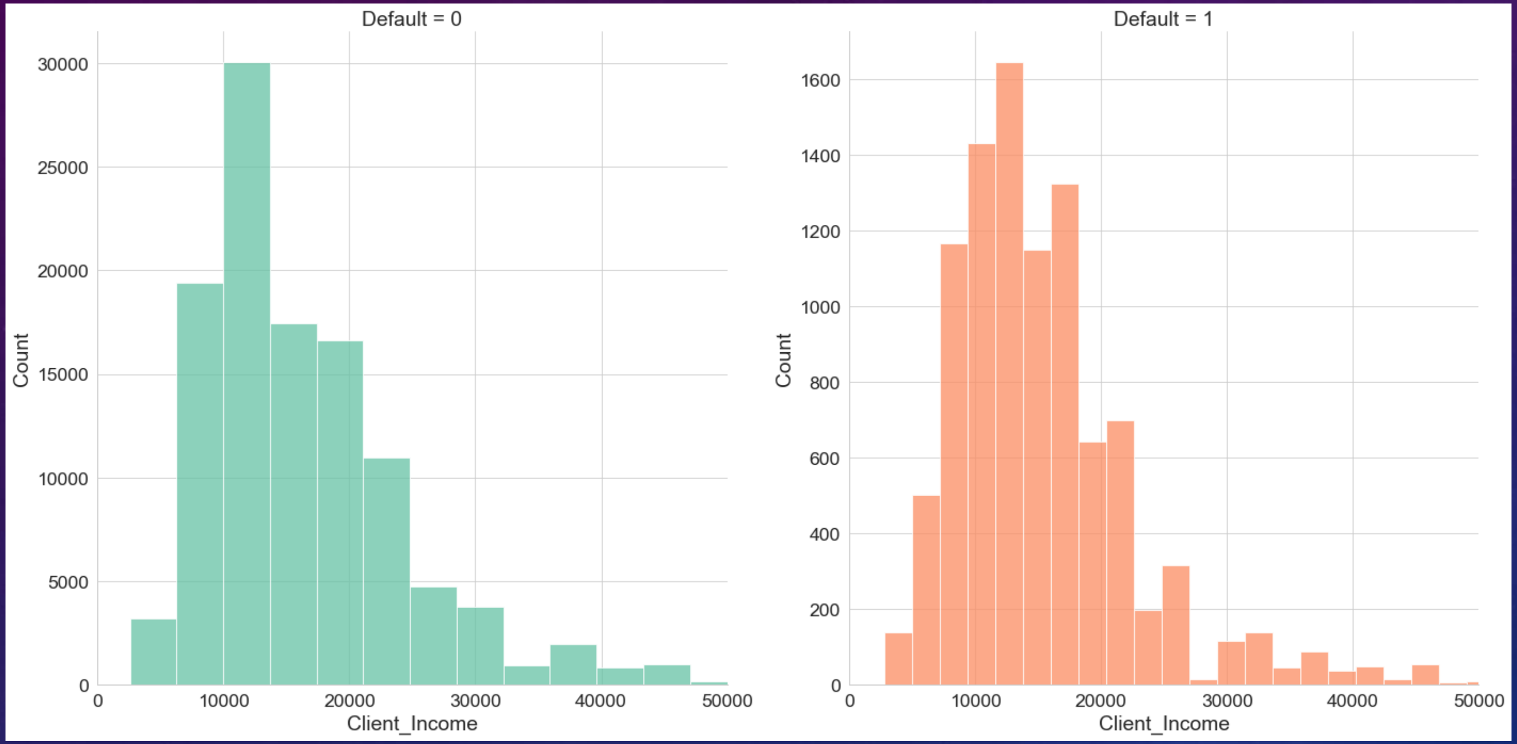
	Client_Income	Credit_Amount	Loan_Annuity	Client_Income_Type	Client_Education	Age_Days	Application_Process_Day	Application_Process_Hour	Score_Source_1	Score_Source_2	Score_Source_3	Car_Owned	Employed_Days	Registration_Days	Default
0	6750.0	61190.55	3416.85	Commercial	Secondary	13957.0	6.0	17.0	0.568066	0.478787	NaN	0.0	1062.0	6123.0	0
1	20250.0	15282.00	1826.55	Service	Graduation	14162.0	3.0	10.0	0.563360	0.215068	NaN	1.0	4129.0	7833.0	0
2	18000.0	59527.35	2788.20	Service	Graduation dropout	16790.0	4.0	NaN	NaN	0.552795	0.329655	0.0	5102.0	NaN	0
3	15750.0	53870.40	2295.45	Retired	Secondary	23195.0	2.0	15.0	NaN	0.135182	0.631355	0.0	NaN	NaN	0
4	33750.0	133988.40	3547.35	Commercial	Secondary	11366.0	3.0	NaN	0.508199	0.301182	0.355639	1.0	2977.0	5516.0	0

HISTOGRAM VISUALIZATION



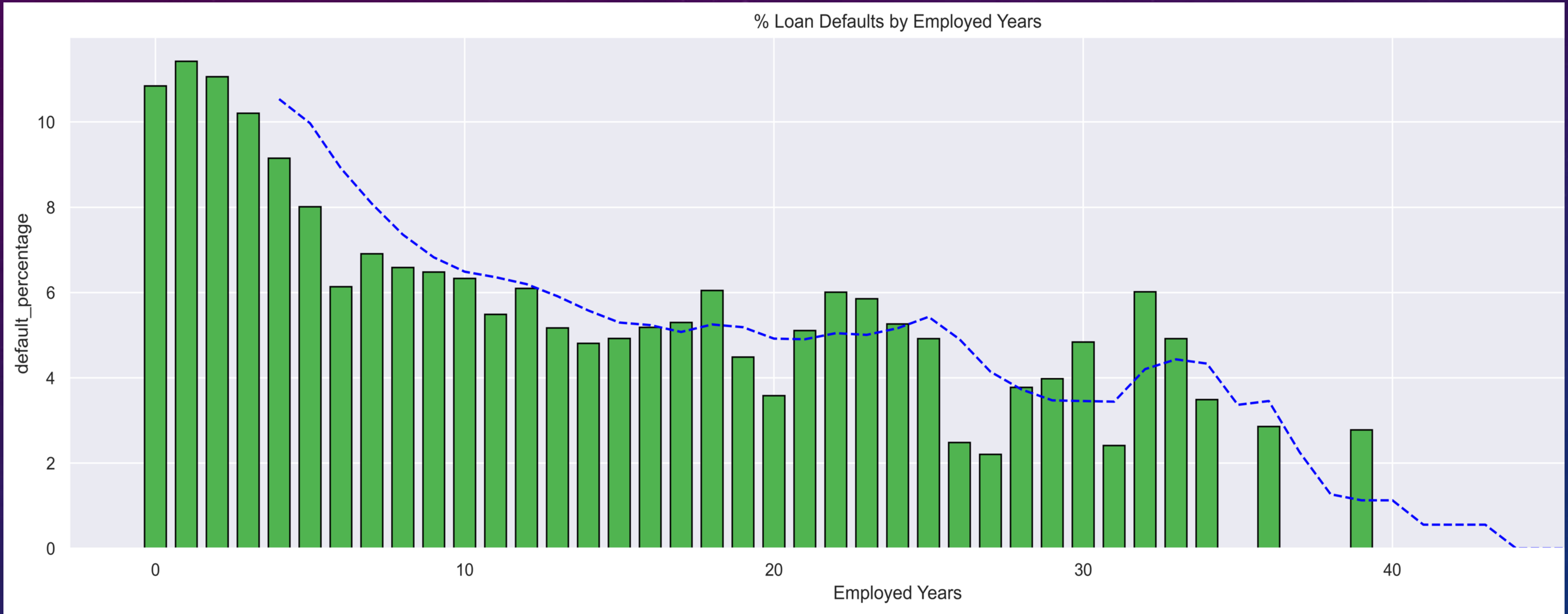
- $250K cut-off value to exclude outliers from index





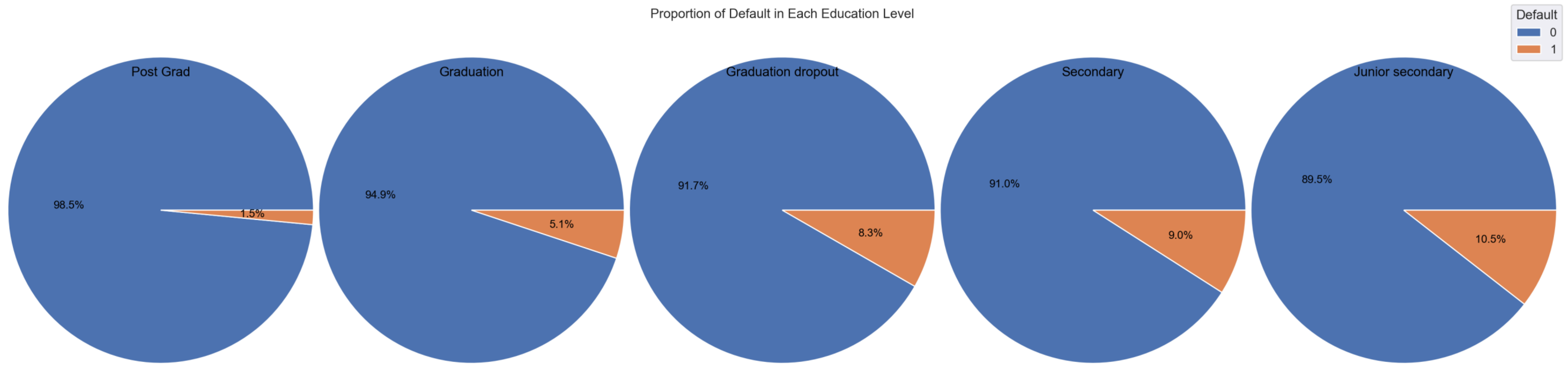
- These histograms involve client income
- Stagnant difference between defaulted and non defaulted clients

BAR CHART VISUALIZATION



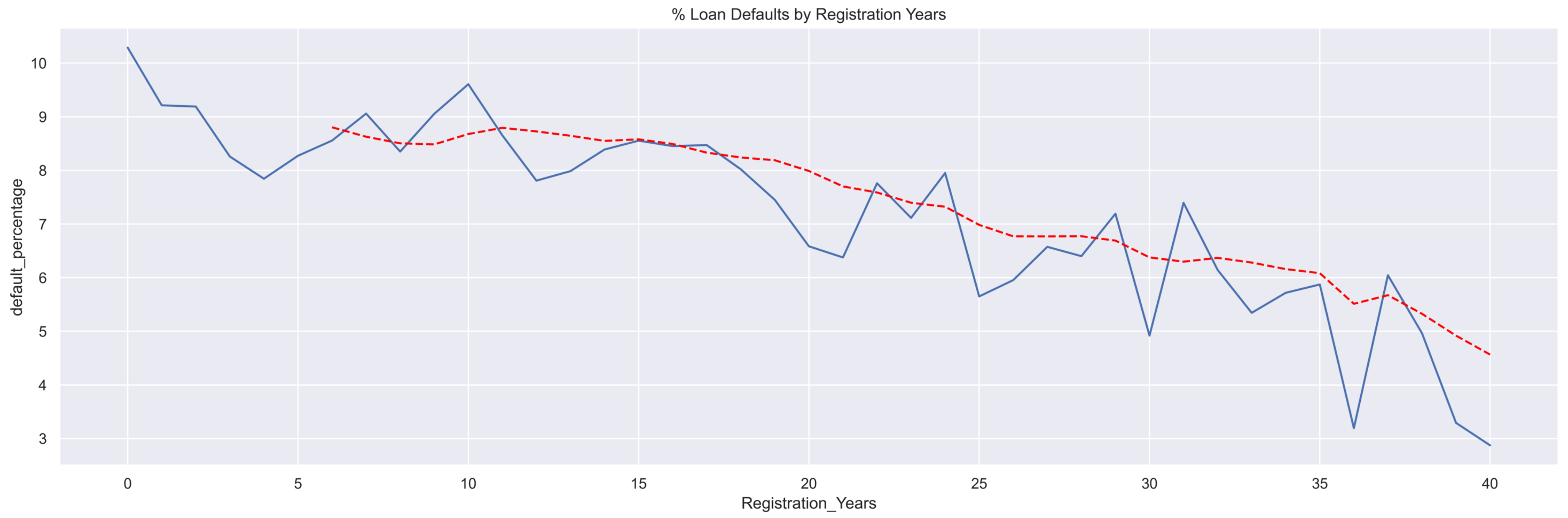
- The line chart on top of bar chart indicates moving average of default percentage
- Drastic loan default declines after employed # of years 6 and 25

PIE CHART



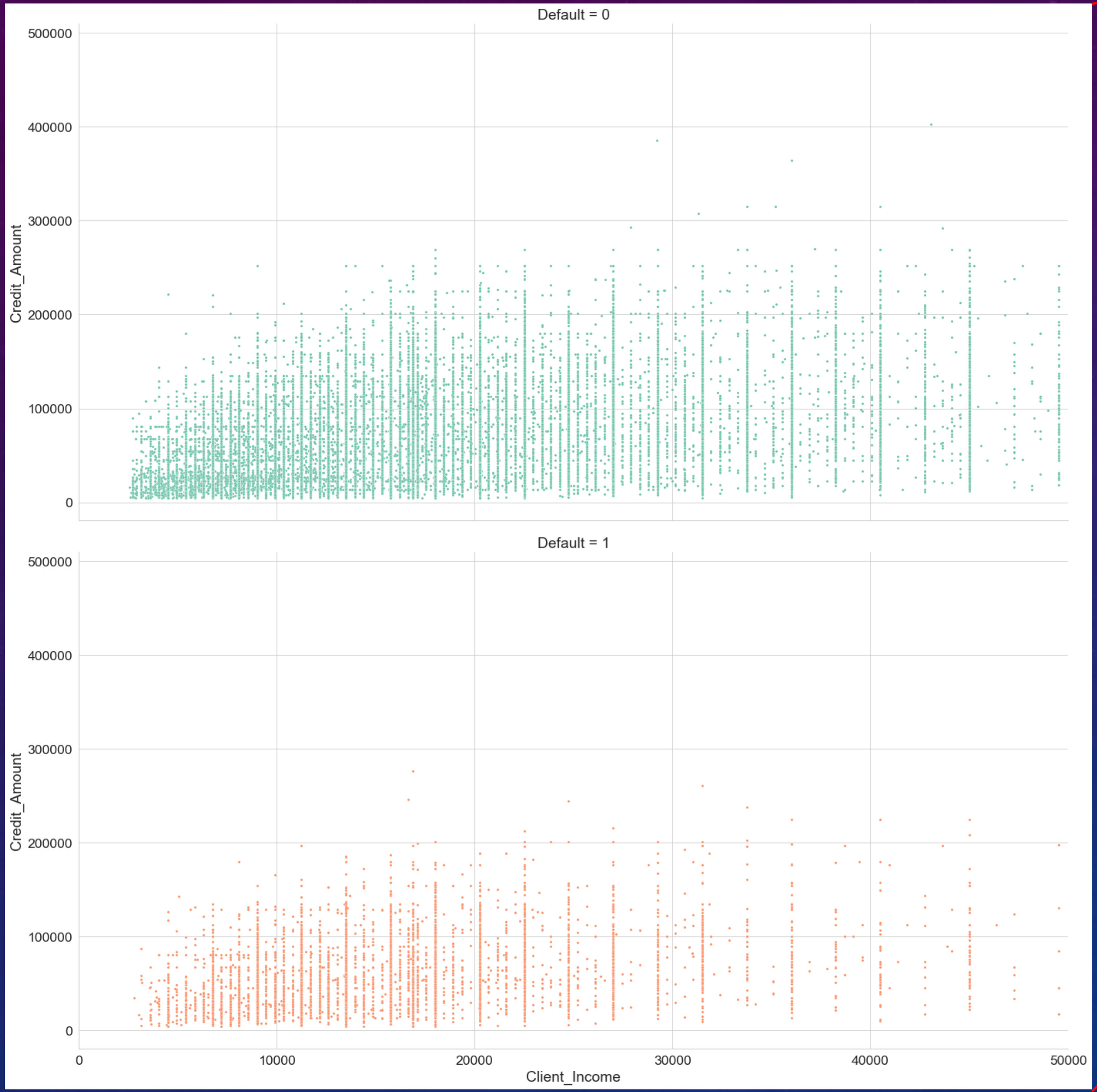
- The higher education level one has, the less likely they are to default on their loan
- This is evidenced by the percentages

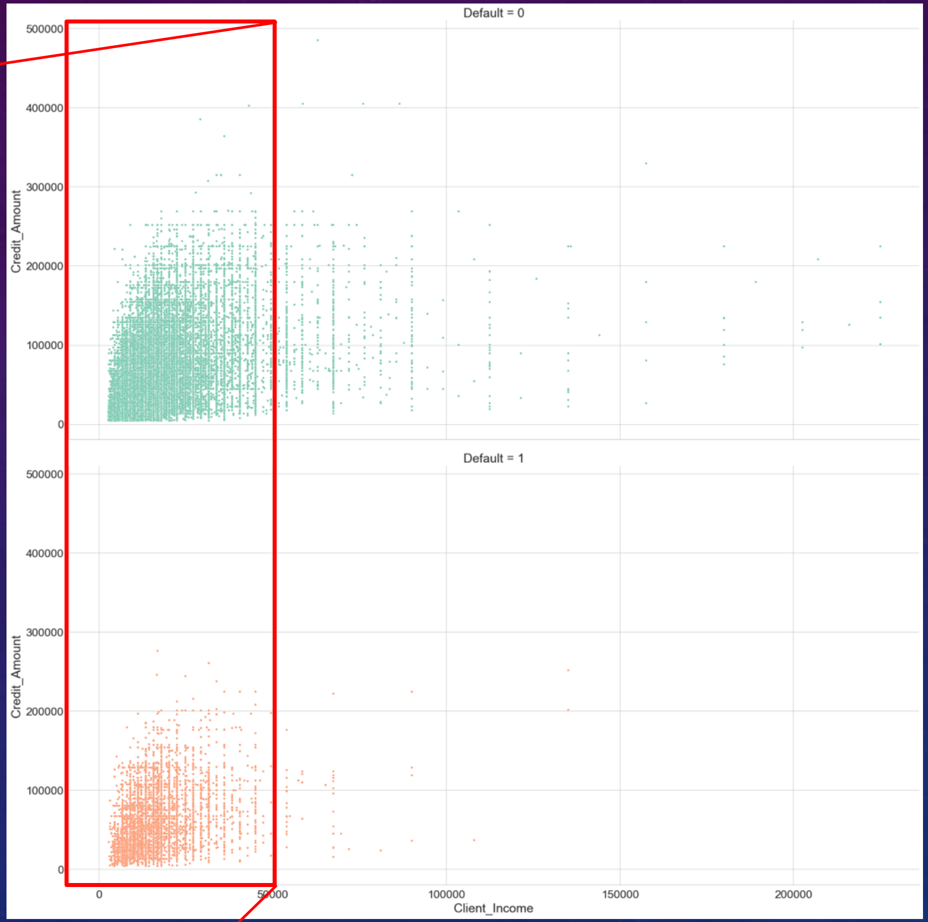
LINE PLOT



- Created new dataset (reg_yrs_distribution) to check if duration of registration had impact on defaults
- Inverse relationship between likelihood of defaulting and number of registration years
- Improved predictive power

SCATTERPLOT VISUALIZATION

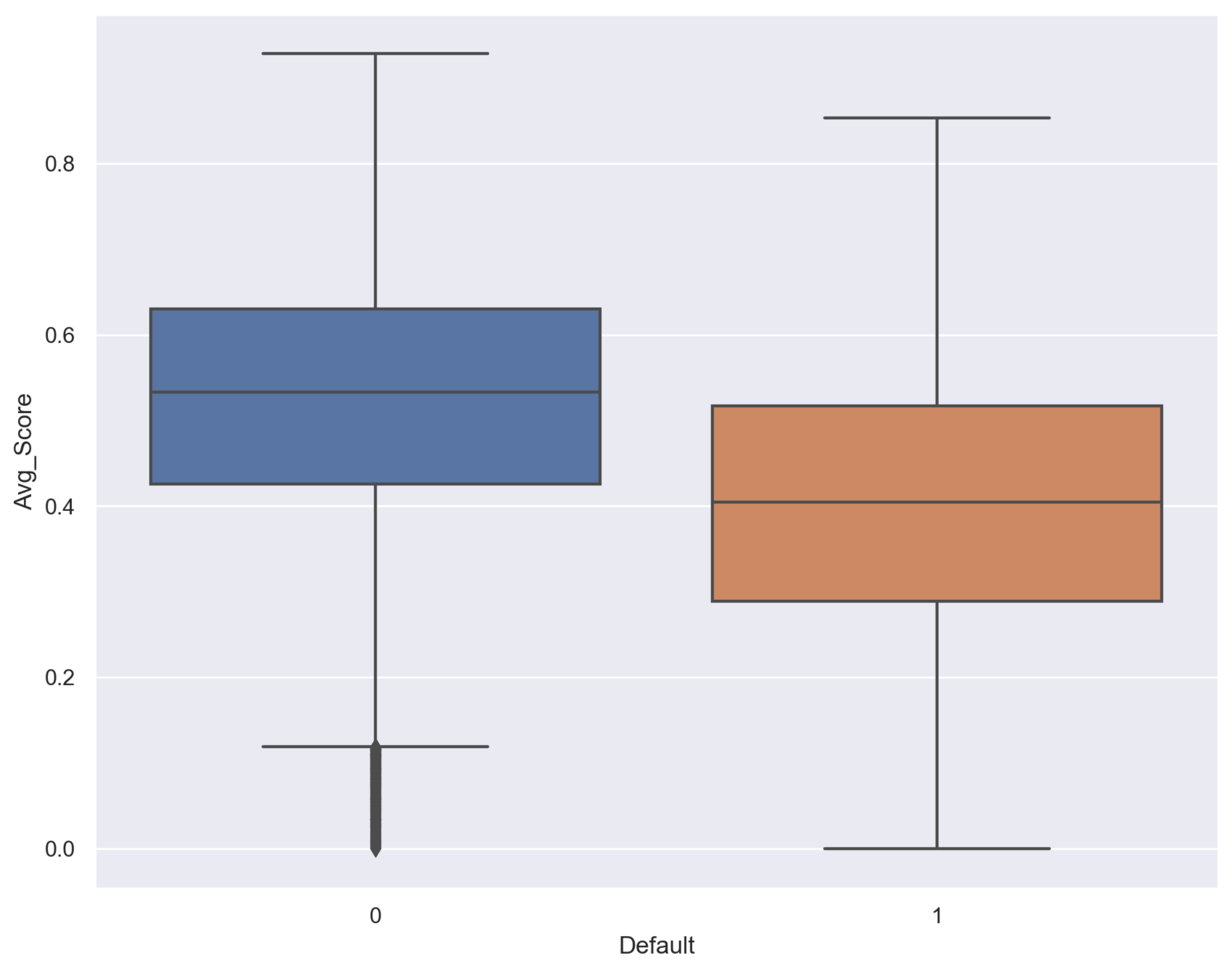




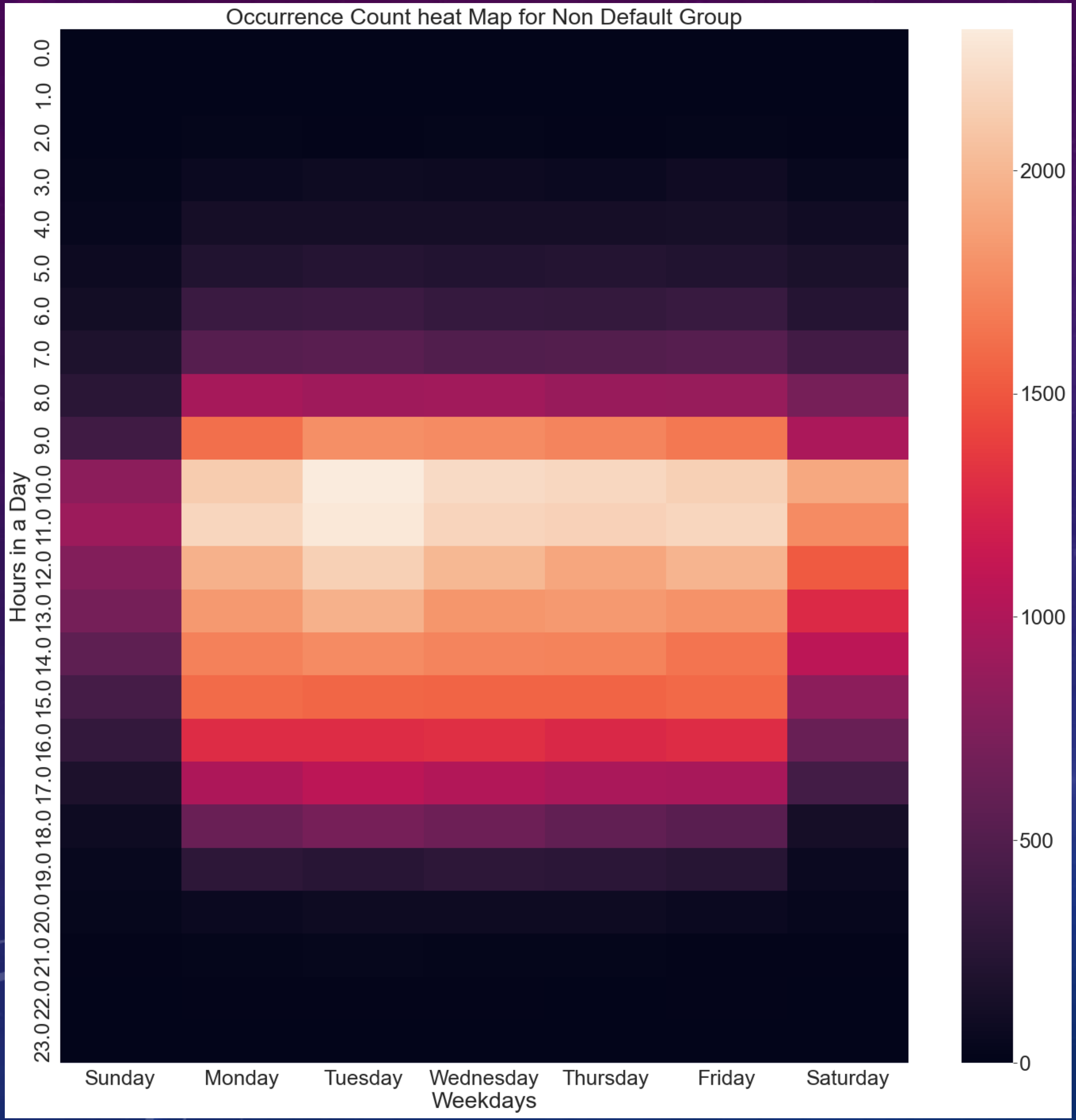
- Utilized python data visualization library's Seaborn to create these scatterplots
- Direct correlation between client income and credit amount

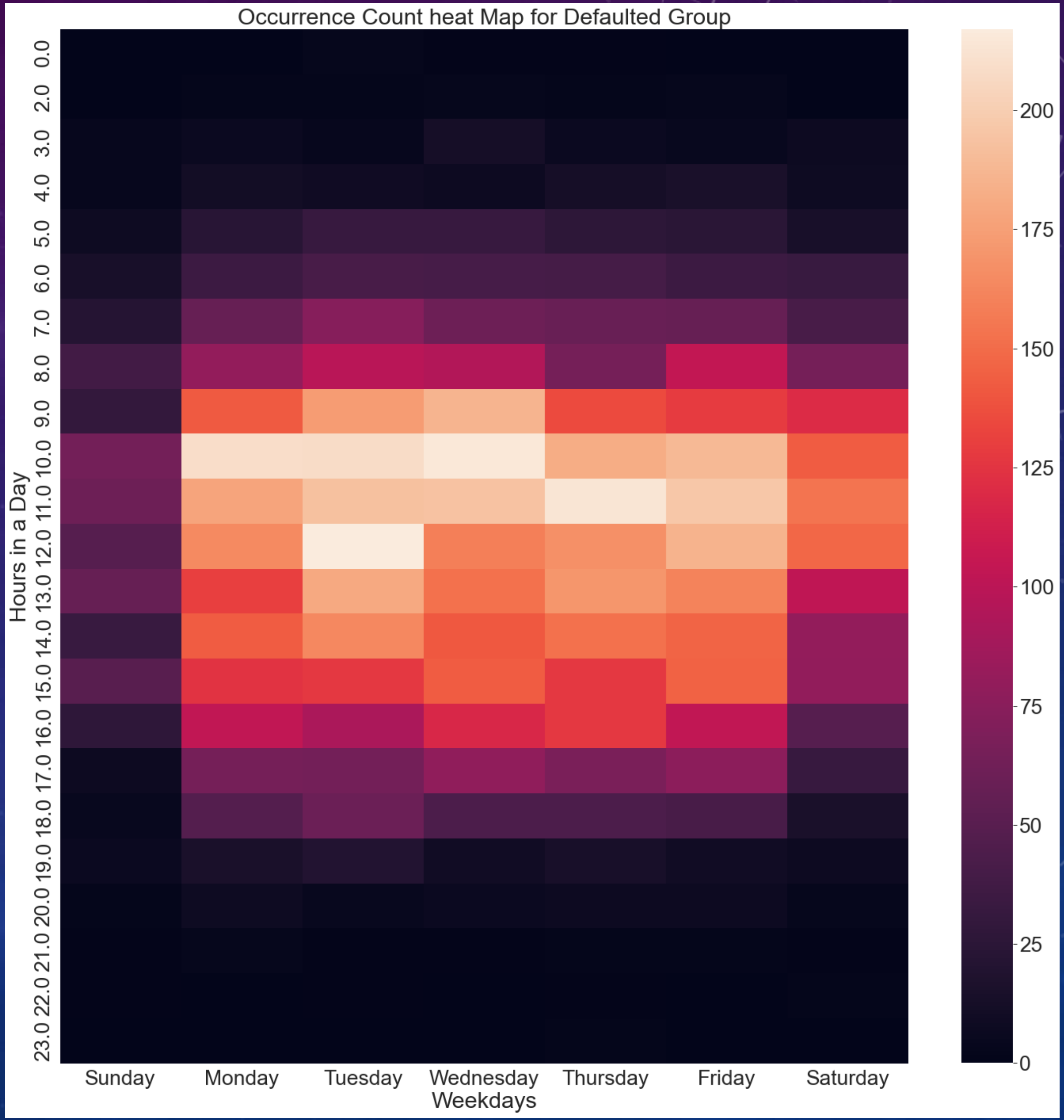
BOX PLOT

- The average credit score plays a huge role in customer defaults
- Regardless of which scale (min, max, median, etc), defaulted customers had lower credit score



HEAT MAP


Occurrence Count heat Map for Non Default Group
Hours in a Day
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
21.0
22.0
23.0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Weekdays
2000
1500
1000
500
0


Occurrence Count heat Map for Defaulted Group
Hours in a Day
0.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
21.0
22.0
23.0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Weekdays
200
175
150
125
100
75
50
25
0

CONCLUSION

Key variables include:

- Client_Income
- Client_Income_Type
- Credit_Amount
- Load_Annuity
- Client_Education
- Age_Date
- Employed_Days
- Registration_Days

Removed Variables:

- Application_Process_Day
- Application_Process_Hour
- Car_Owned

RECOMMENDATION

1. More detailed description in the Data Dictionary
2. Annual Percentage Rate (APR)
3. Vehicle Type or Specs