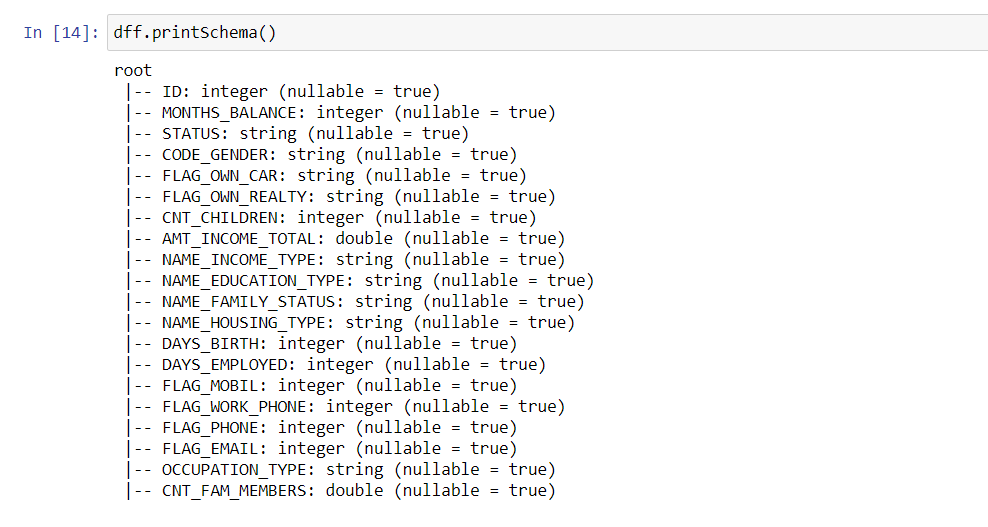
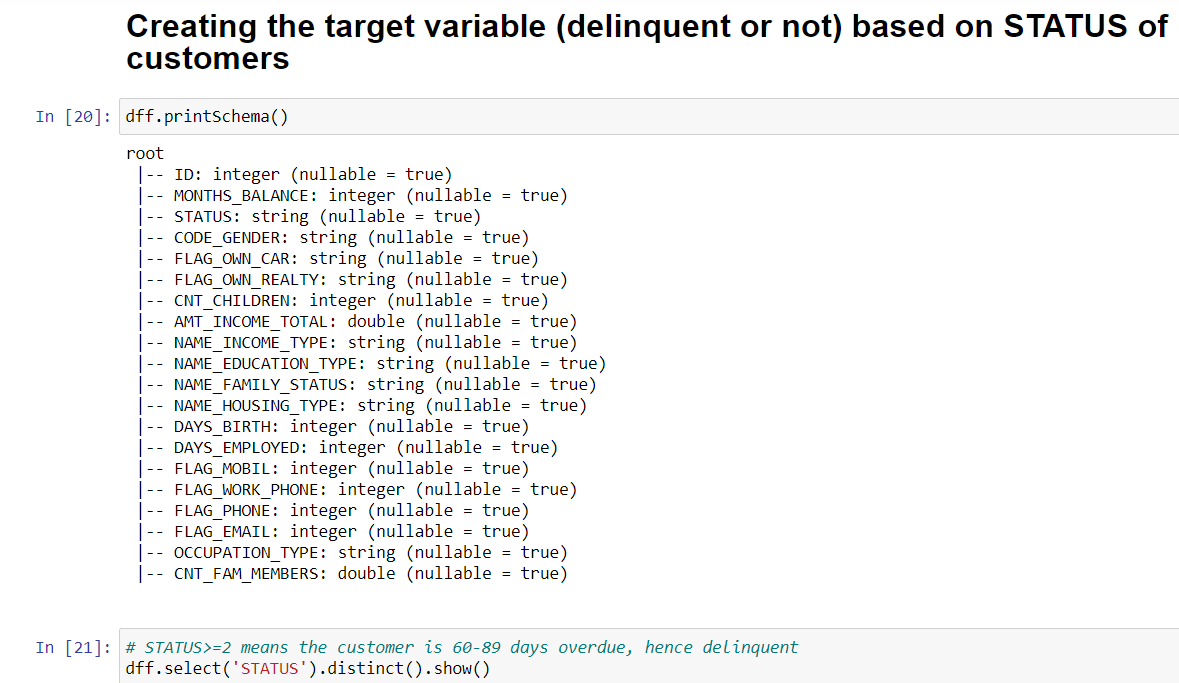
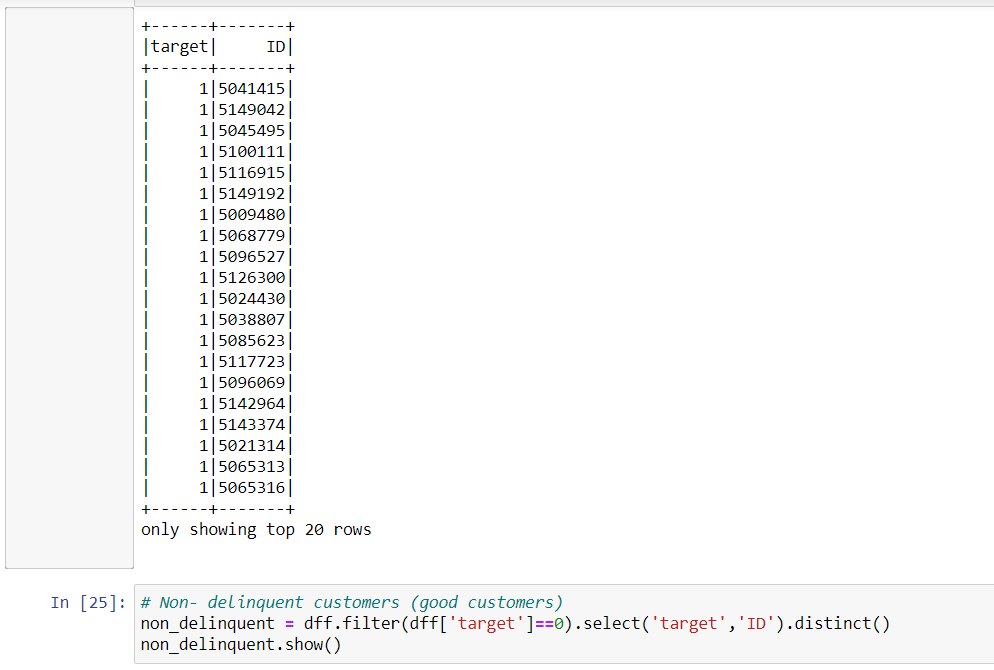
* **EDA**

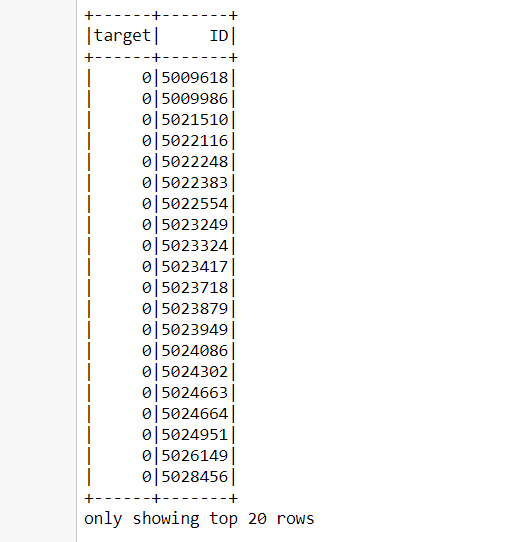
1. Schema of combined dataframe:



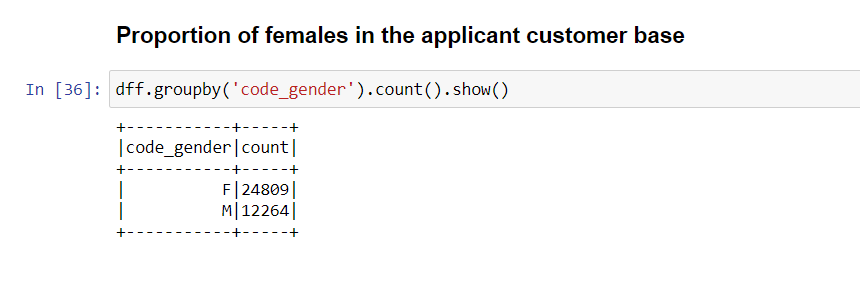








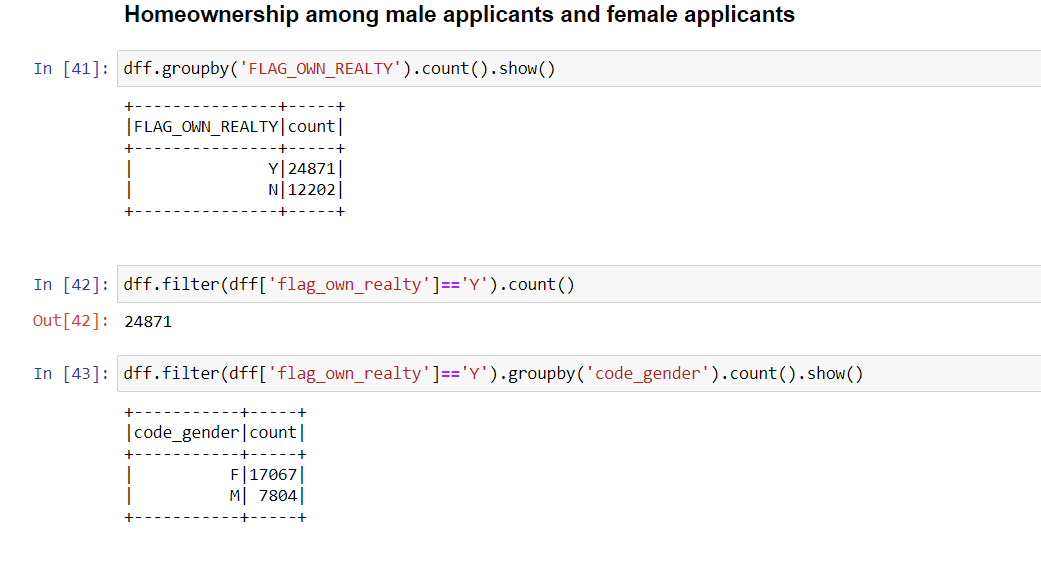
1. What is the proportion of females in the applicant customer base?

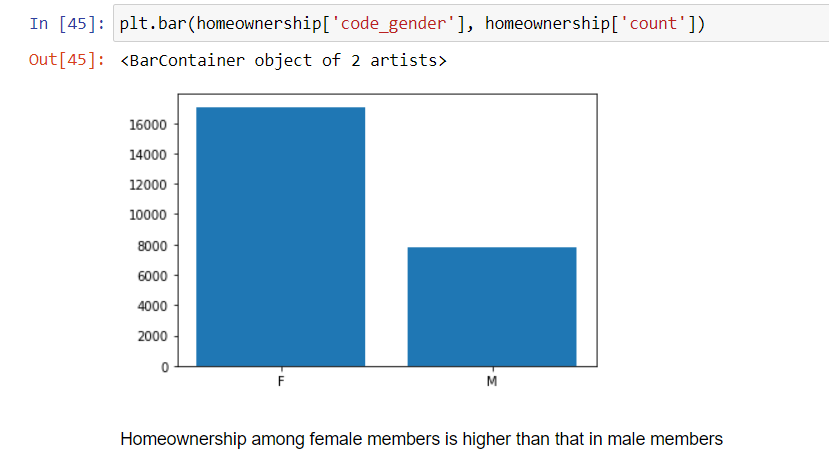






1. Is homeownership higher among male applicants or female applicants?

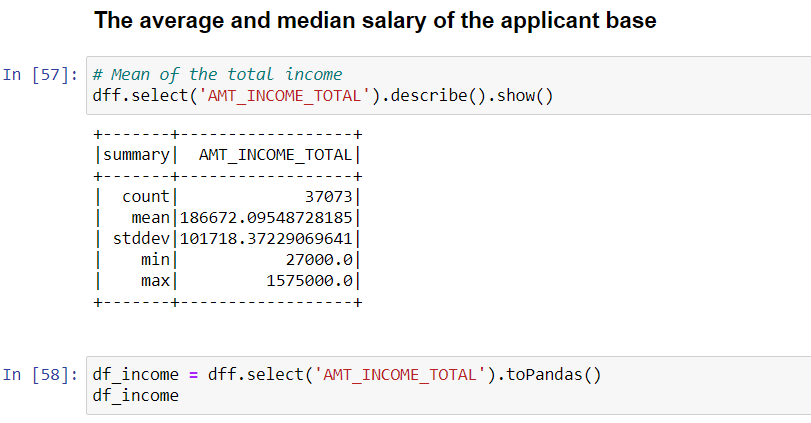


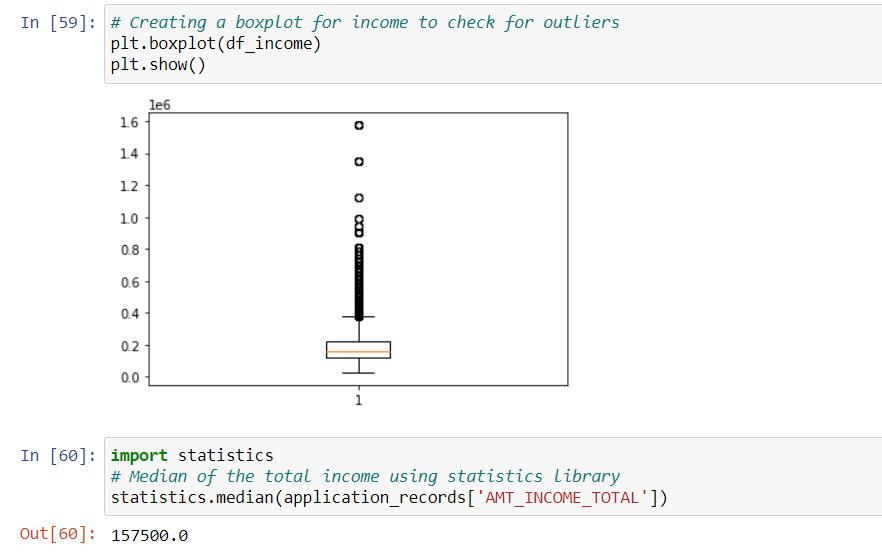


1. Is there any correlation between the customer's income level and education level?



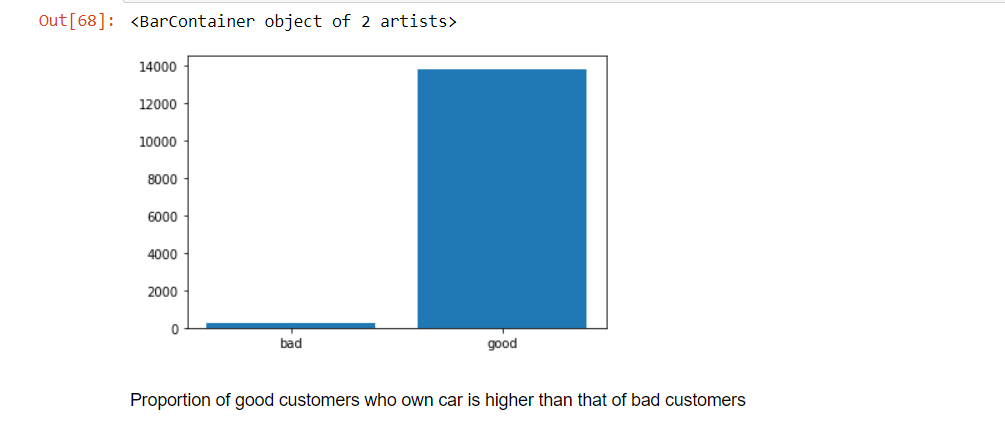
1. What is the average and median salary of the applicant base?



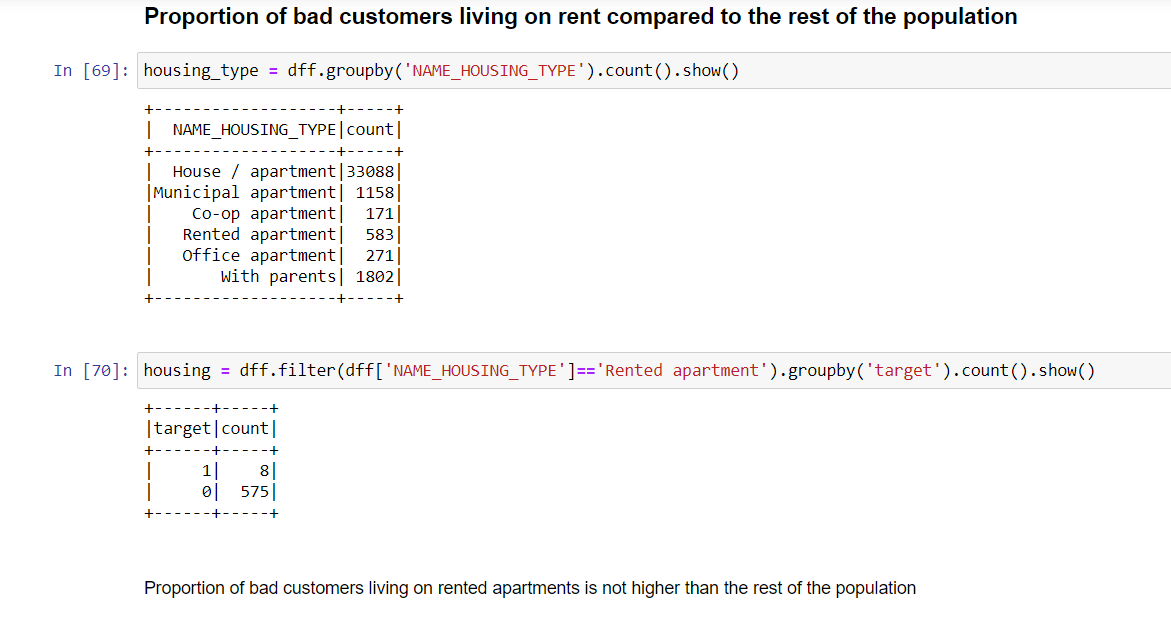


1. Is the proportion of bad customers higher for people who own cars?





1. Is the proportion of bad customers higher for those living on rent than the rest of the population?



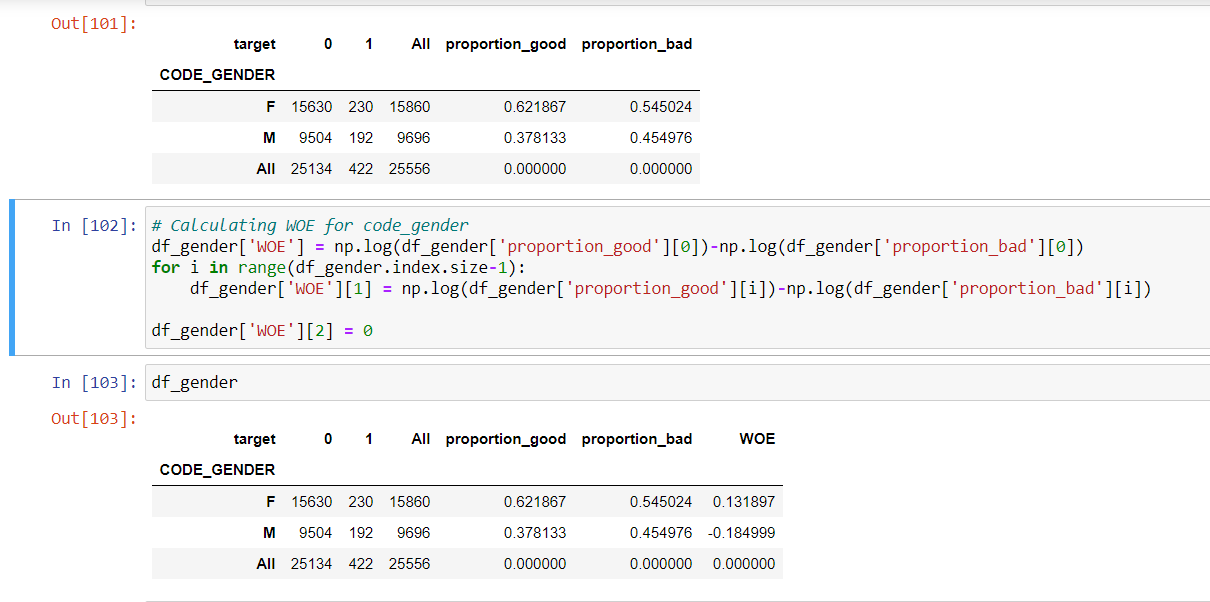
1. Is the proportion of bad customers higher for those who are single than married customers?

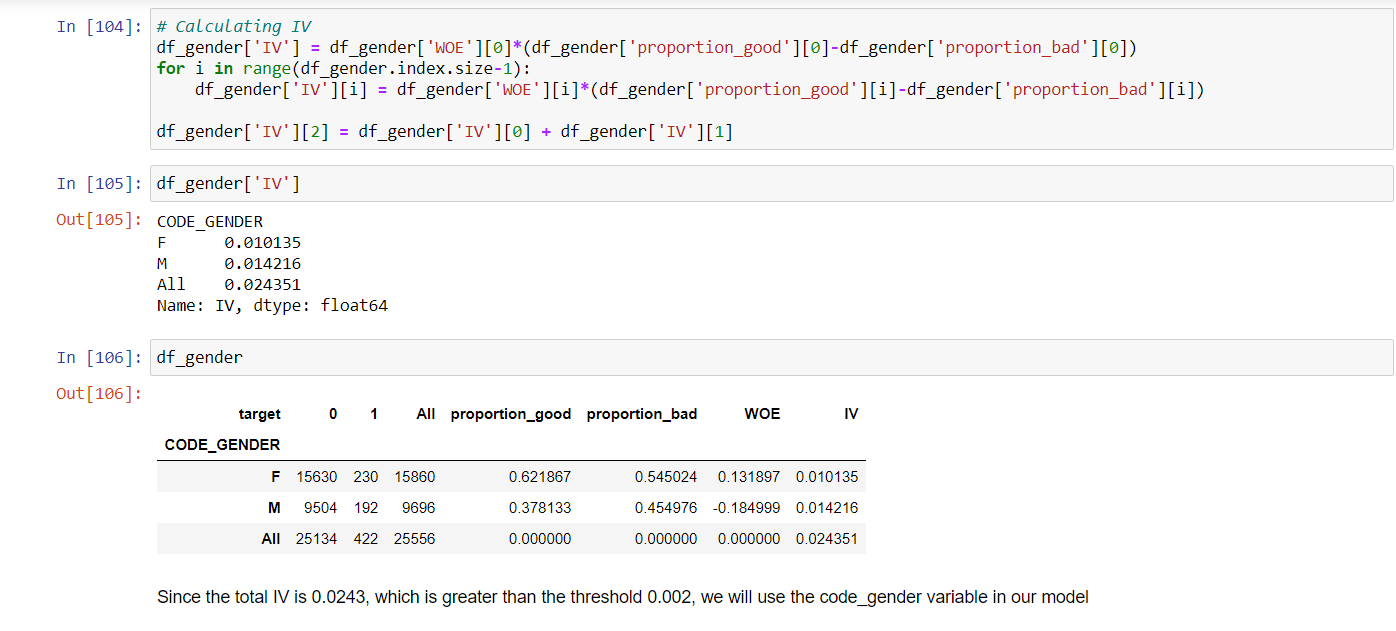


****

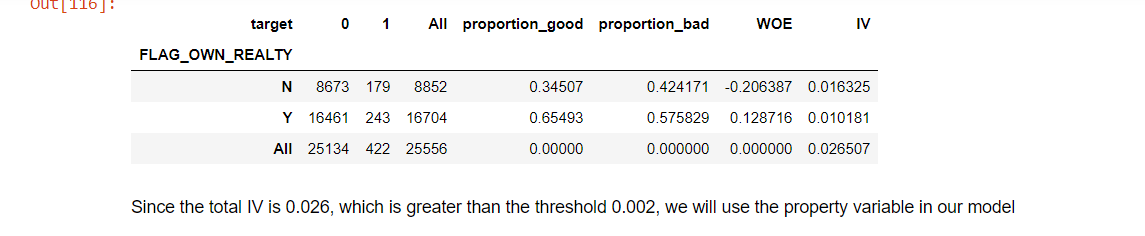
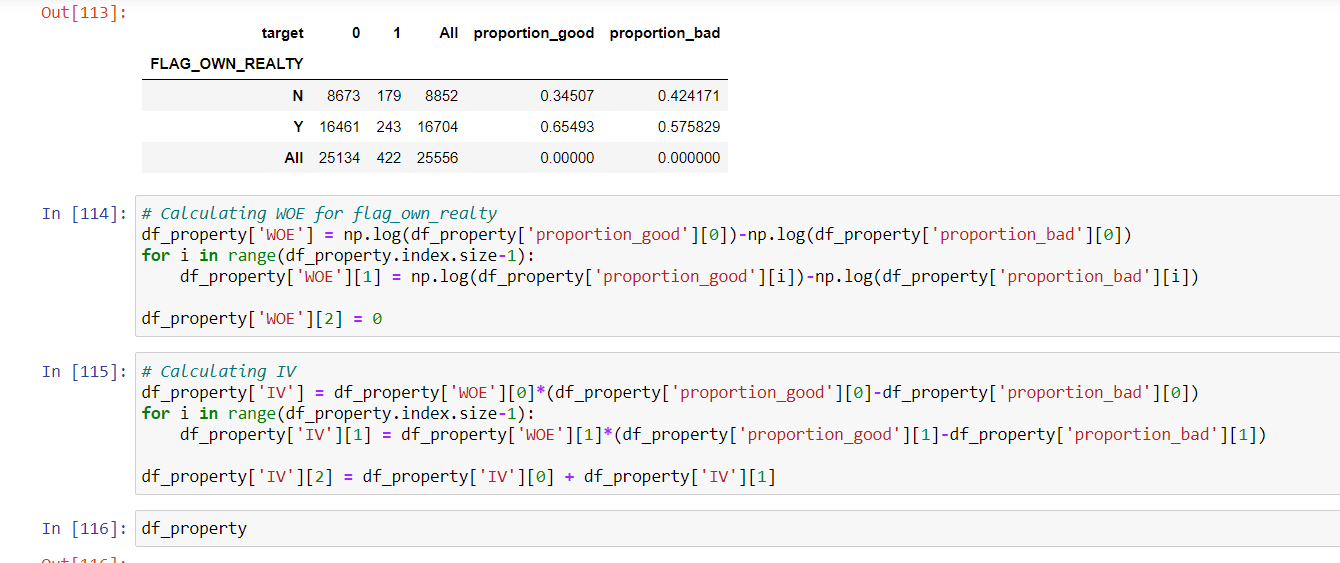
* **Variable Transformation: WOE and IV Calculation for Categorical variables**

1. Code\_gender

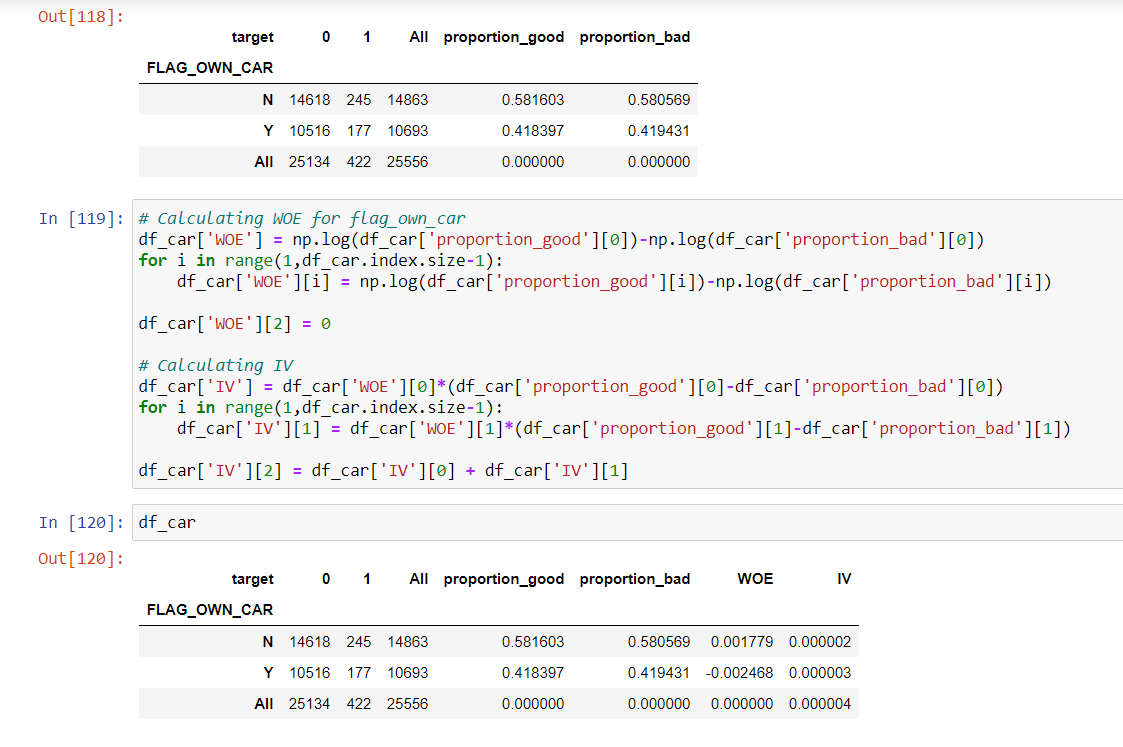




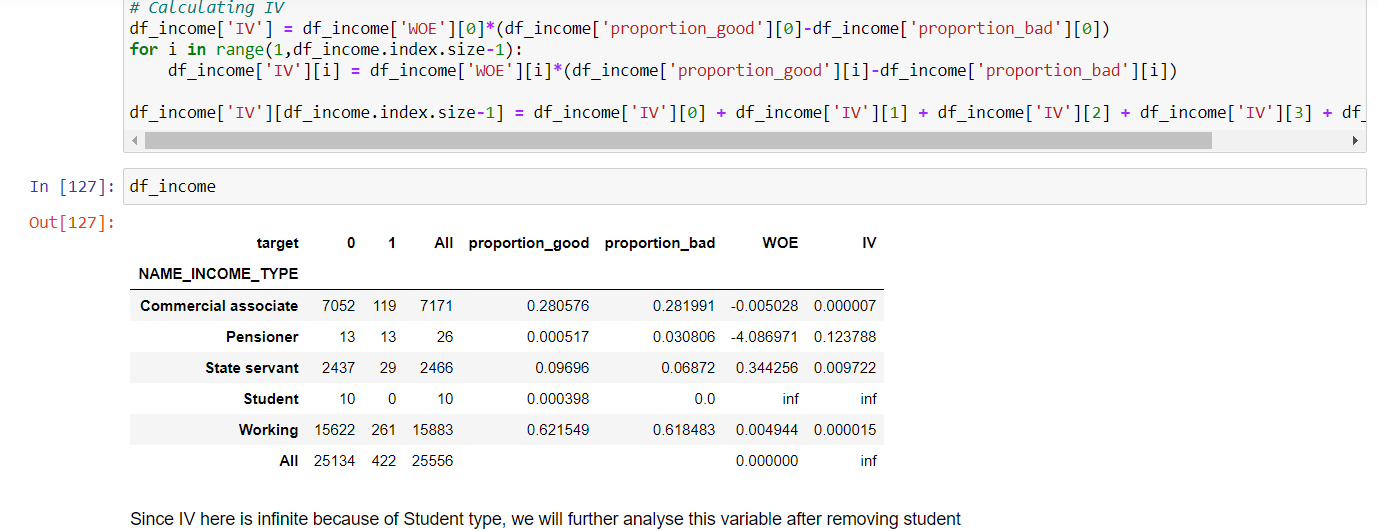
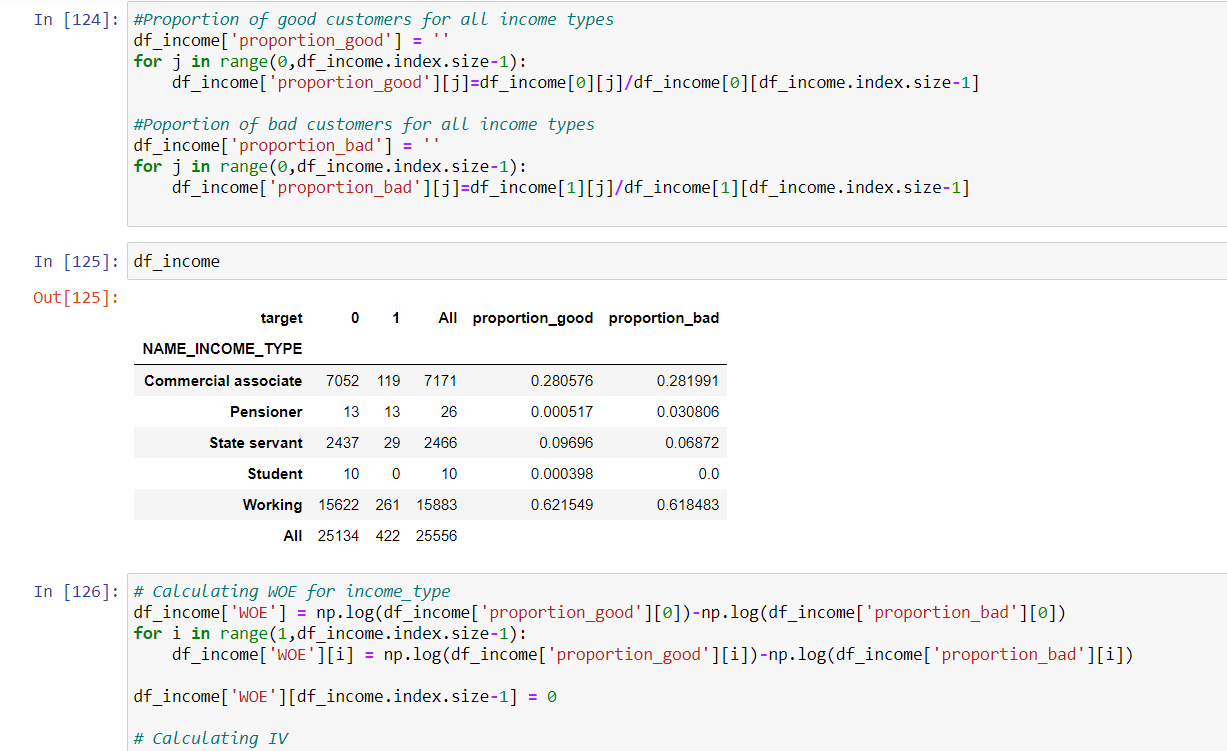
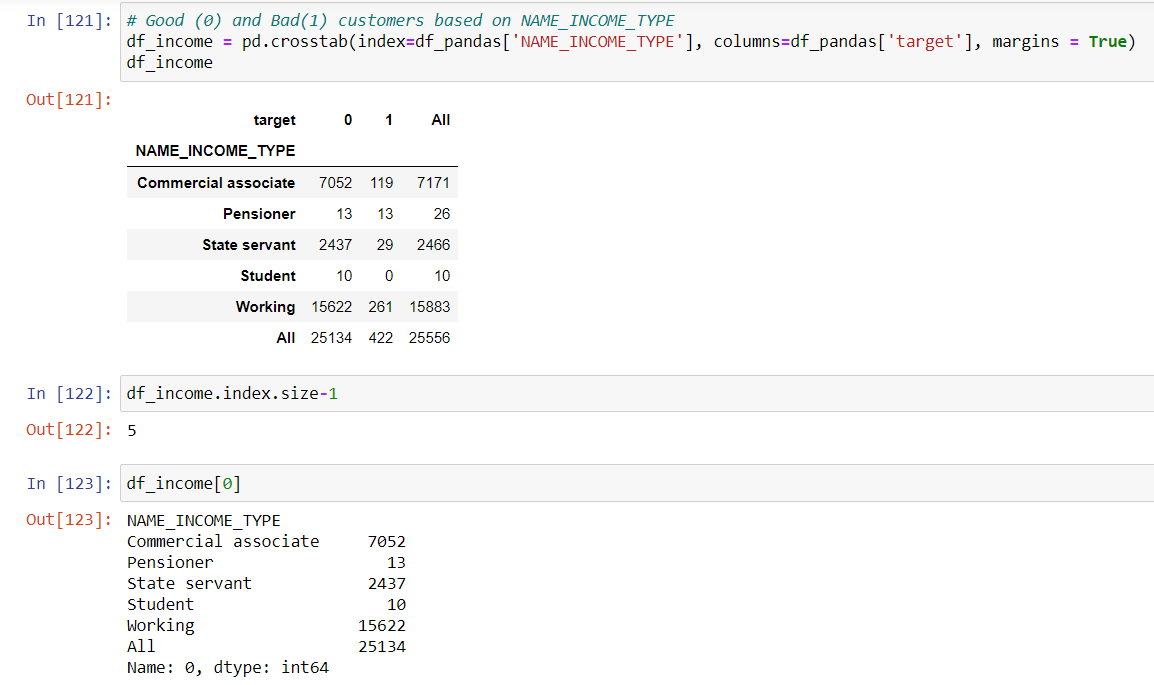
1. Flag\_own\_realty



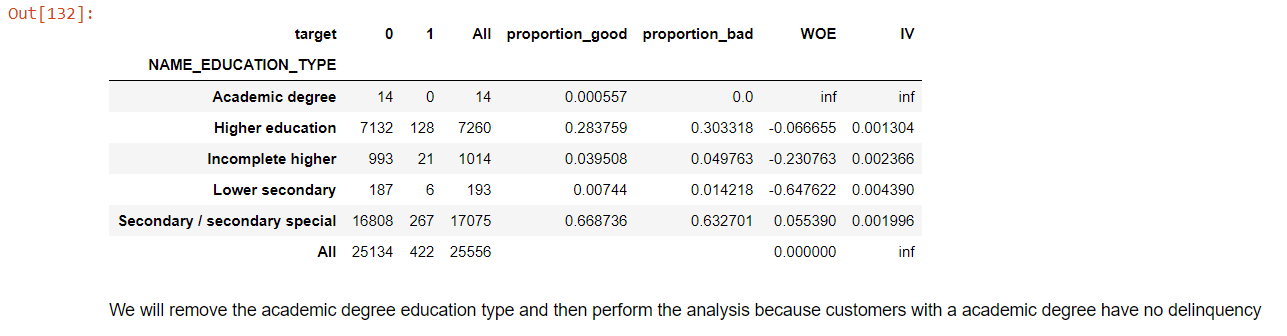
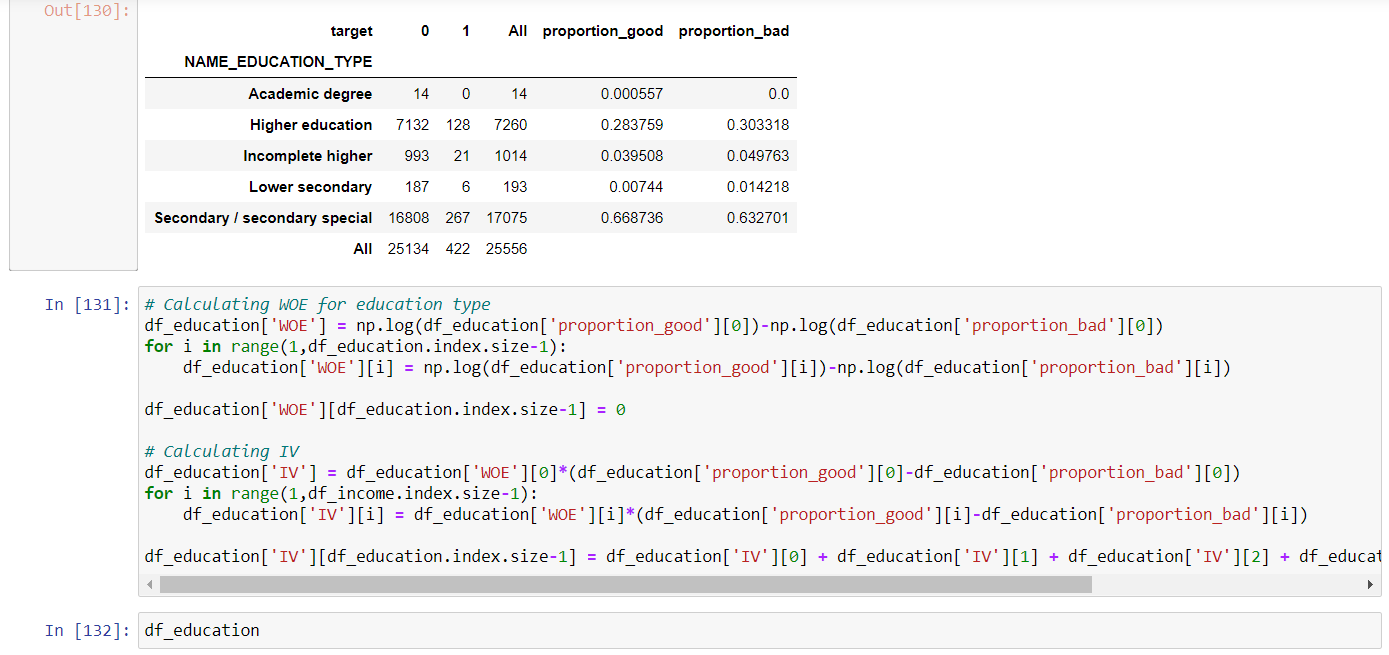
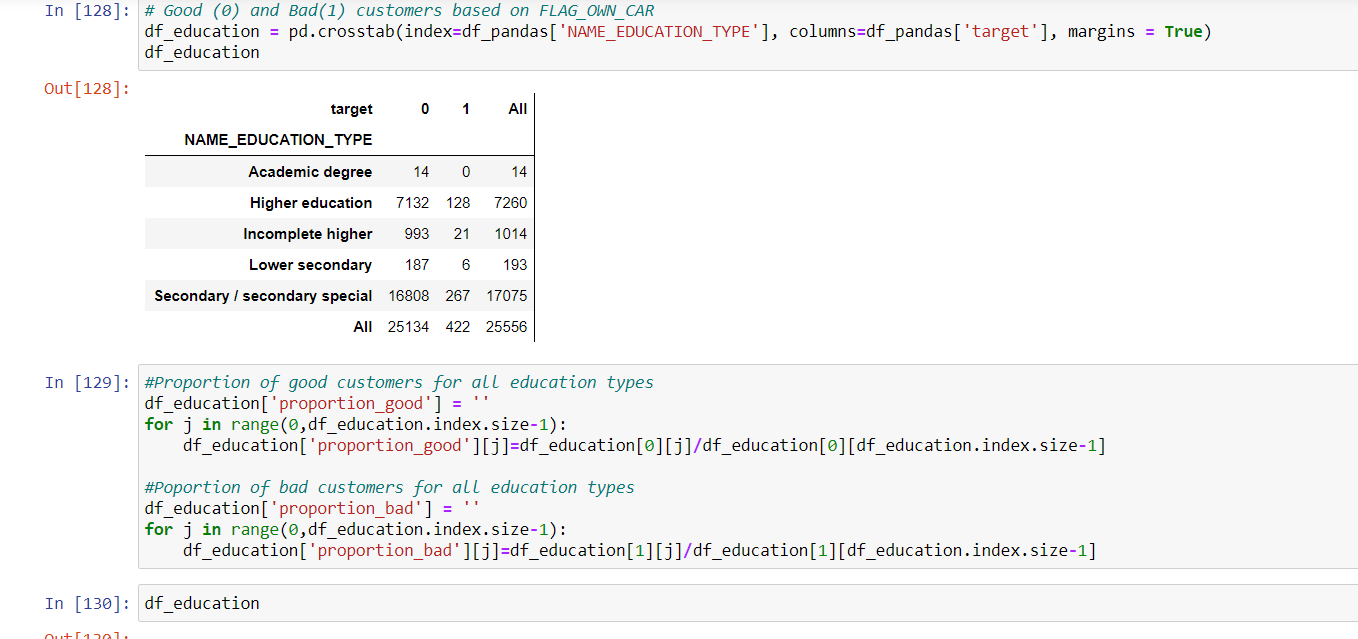
1. Flag\_own\_car



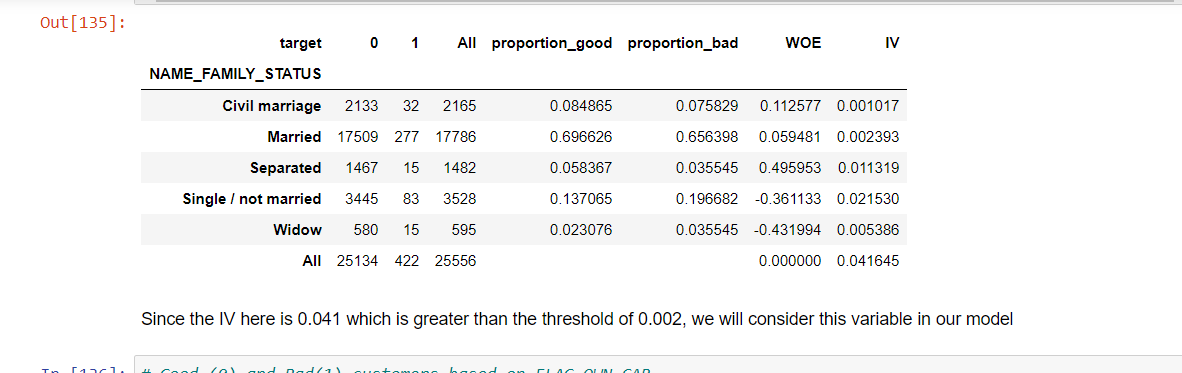
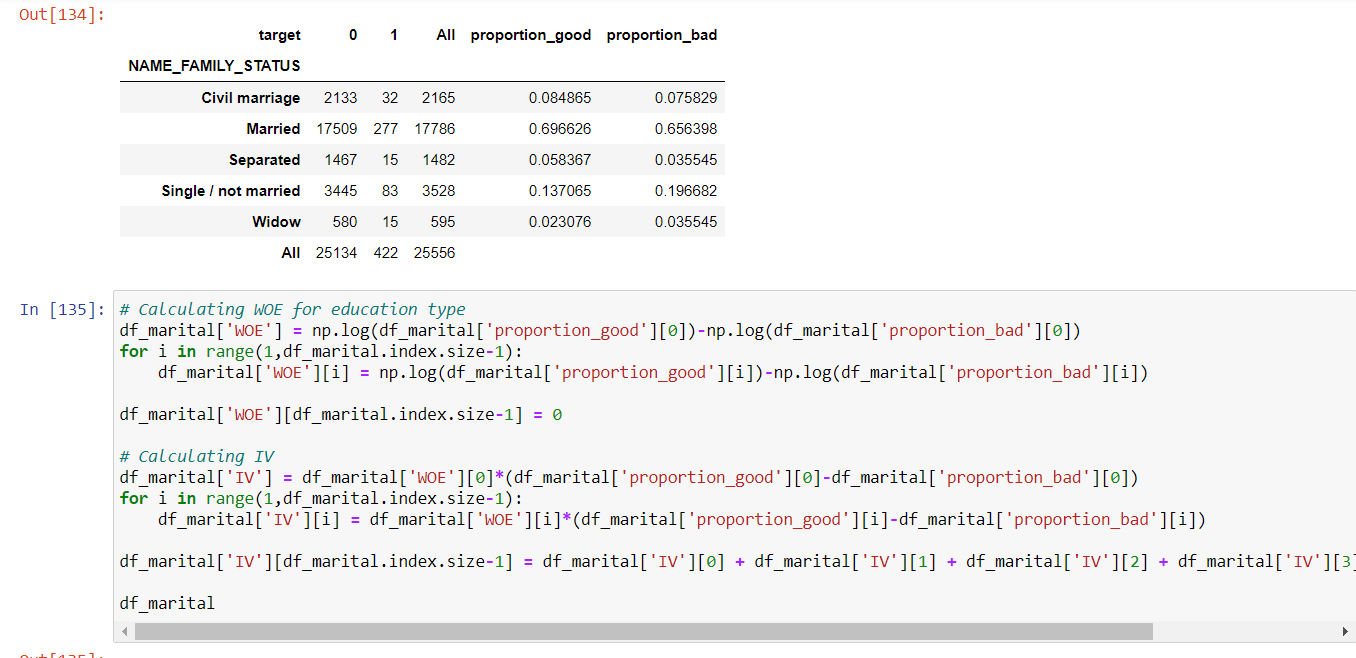
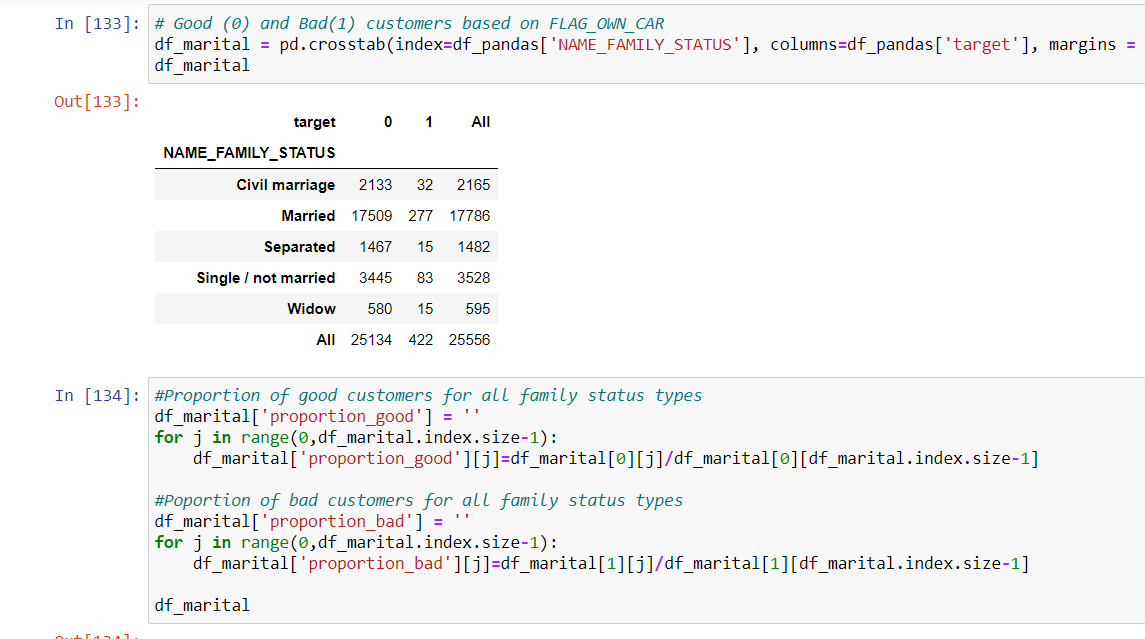
1. Name\_income\_type



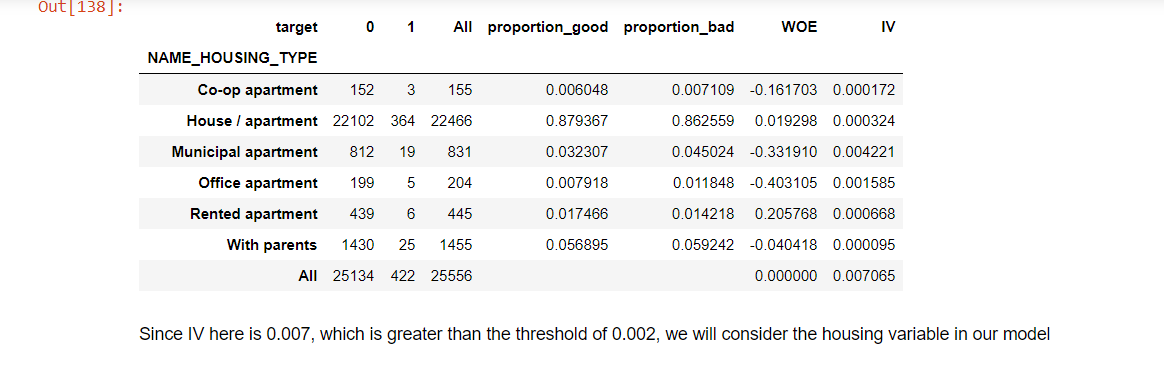
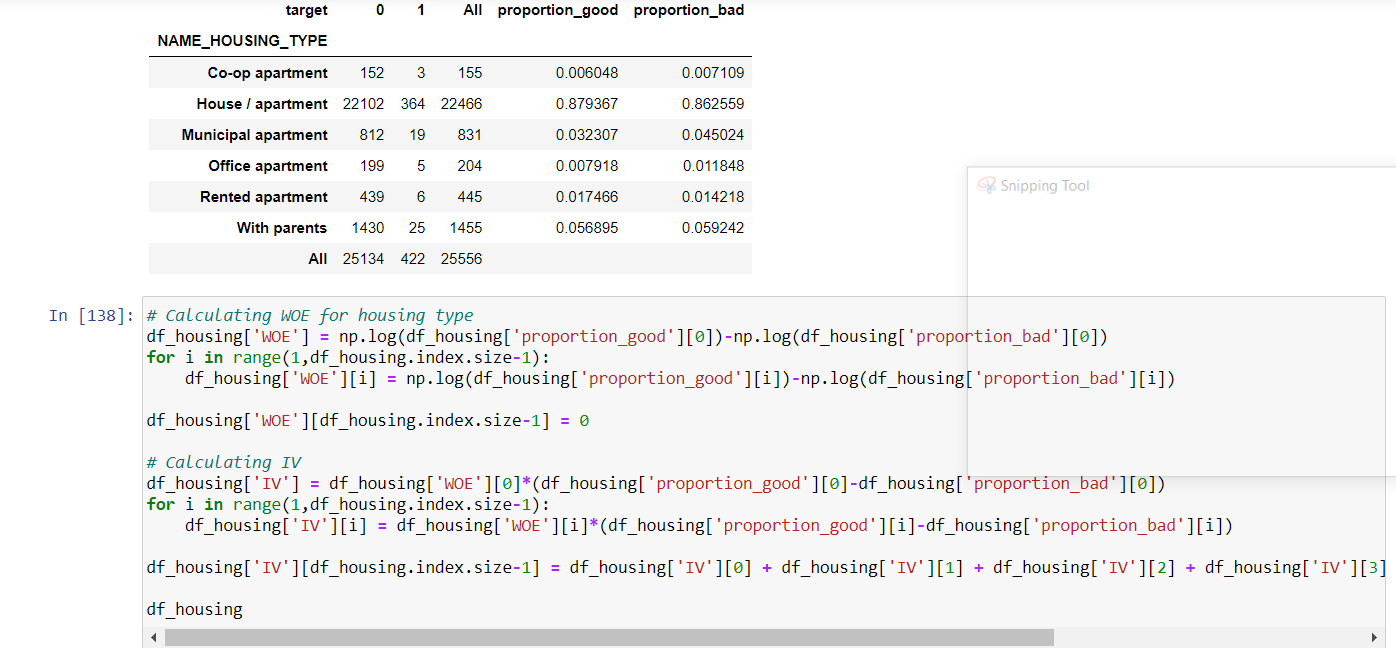
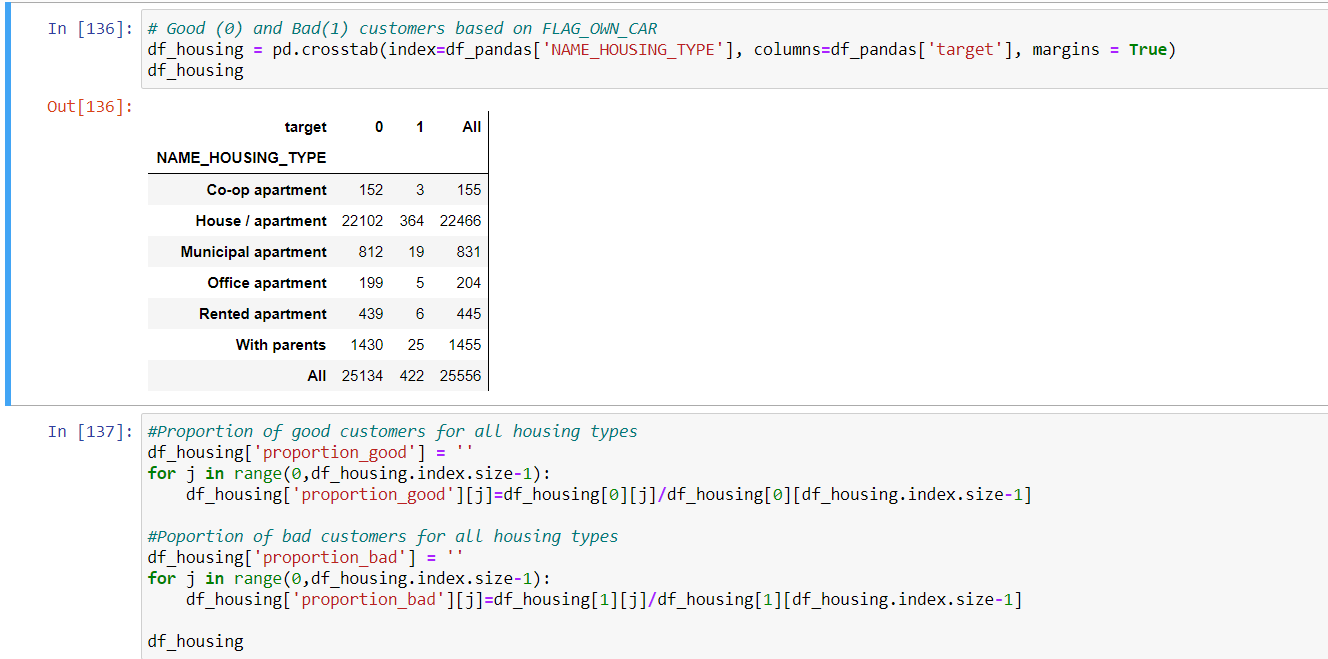
1. Education



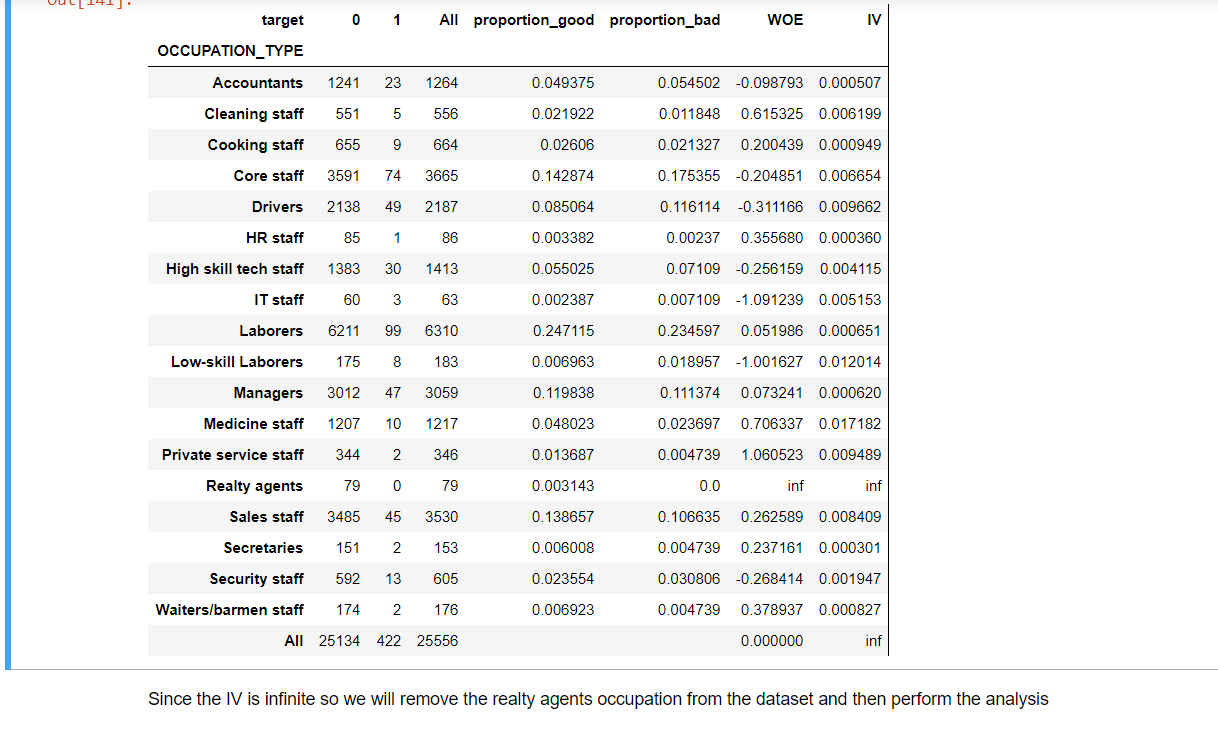
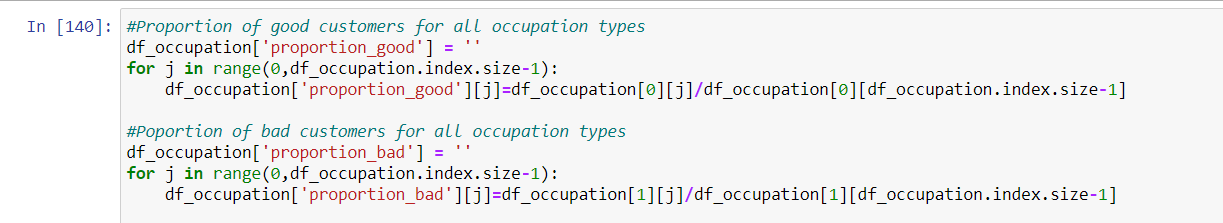
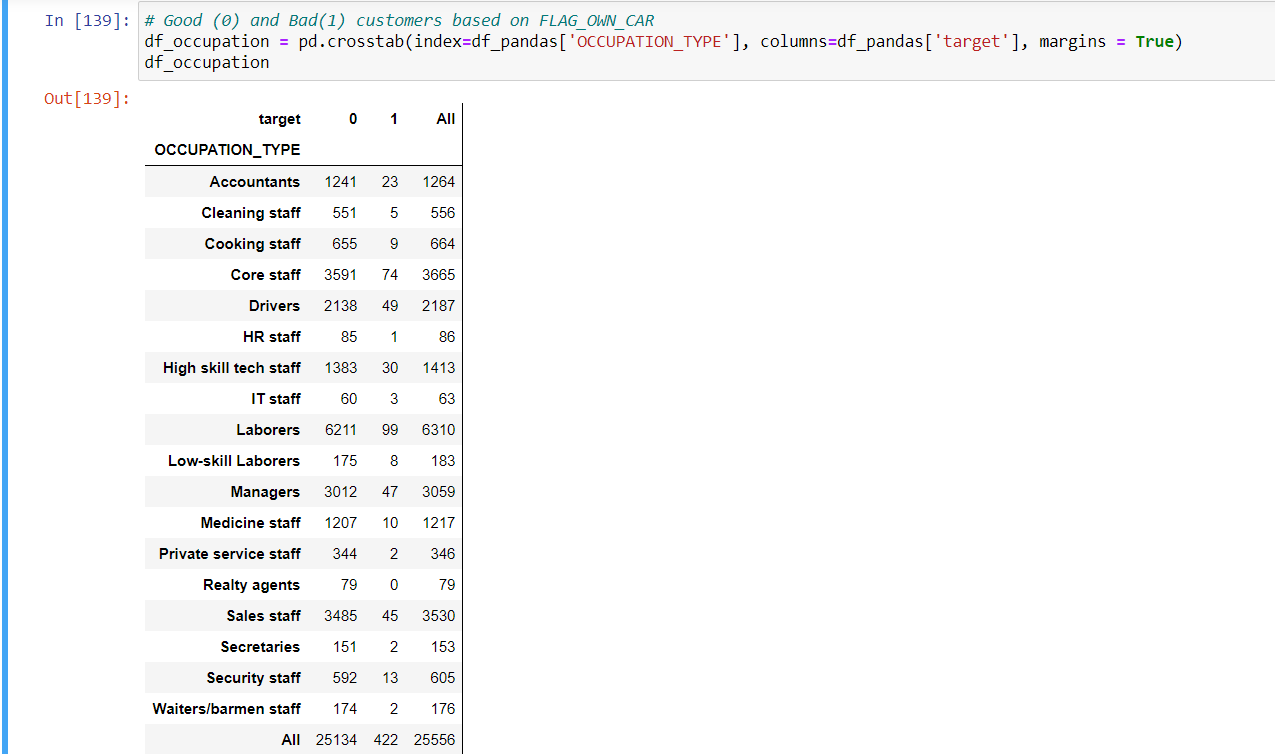
1. Name\_family\_status



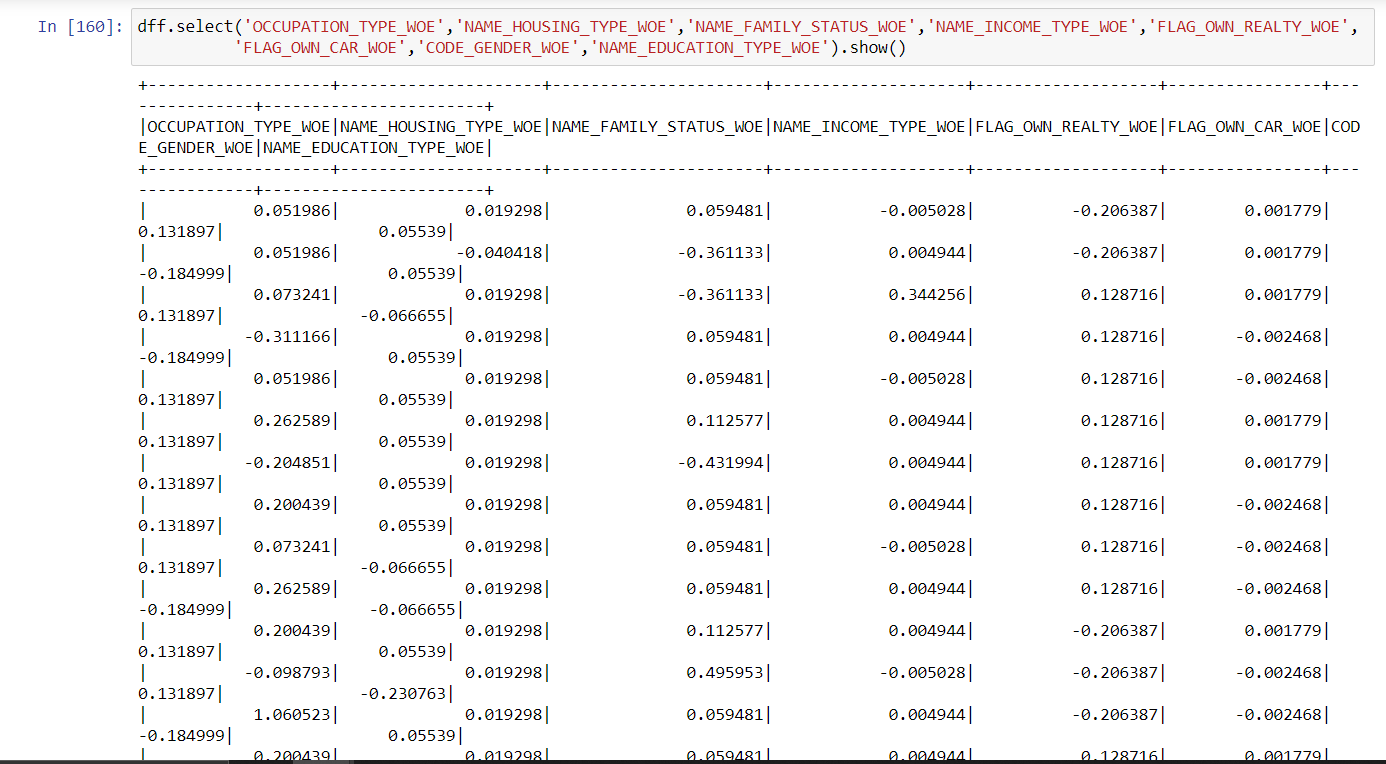
1. Housing



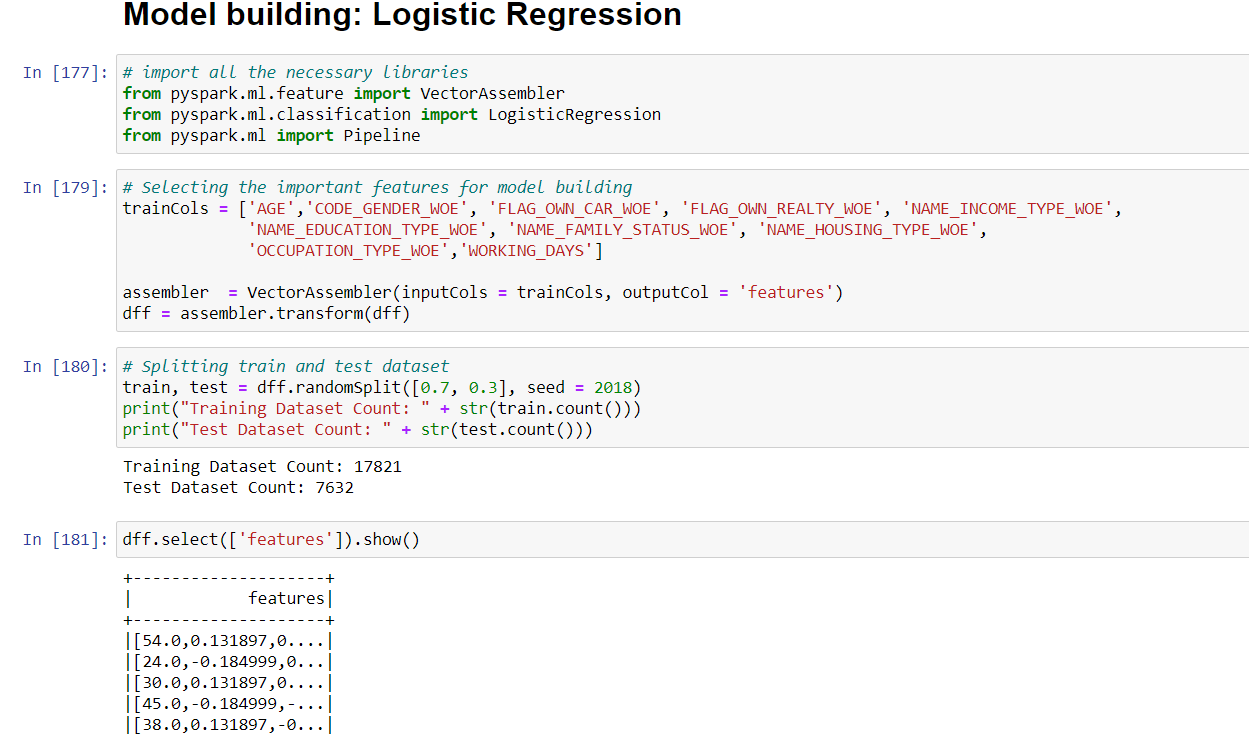
1. Occupation

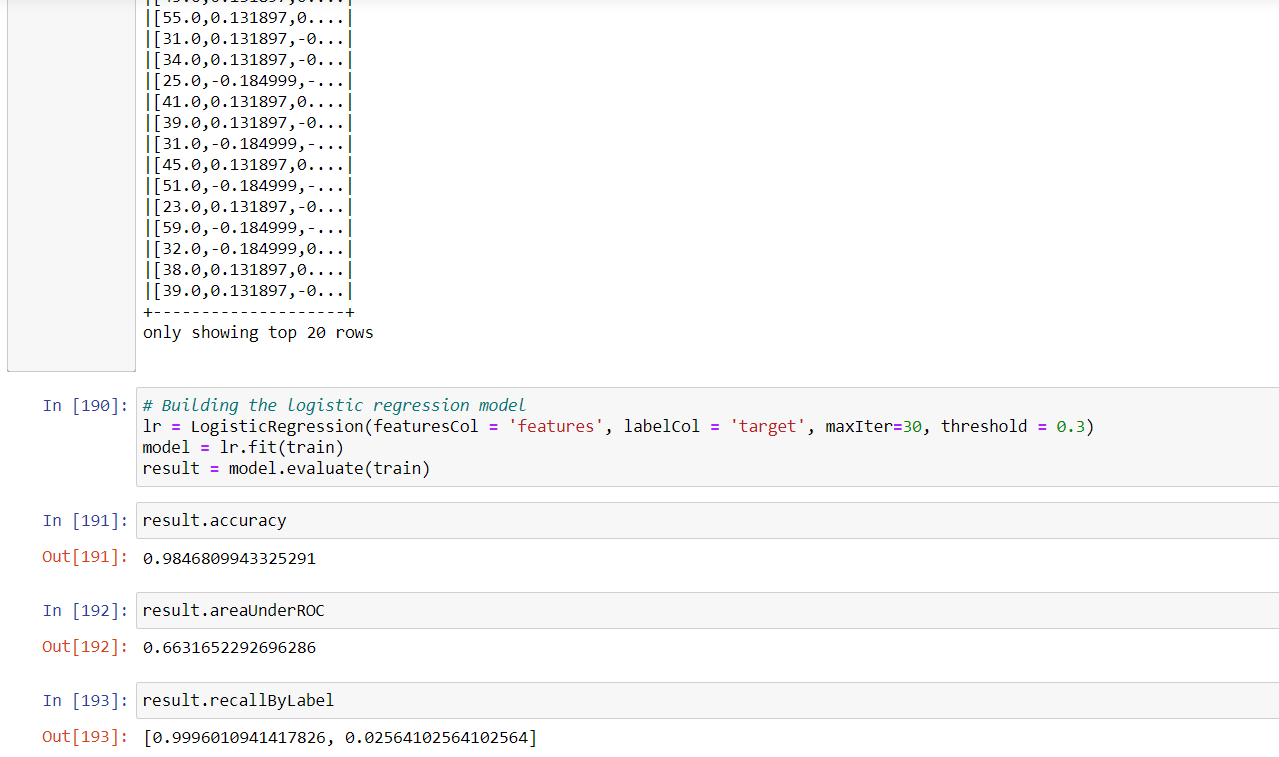


1. REPLACING ALL THE VALUES IN COLUMNS WITH WOE



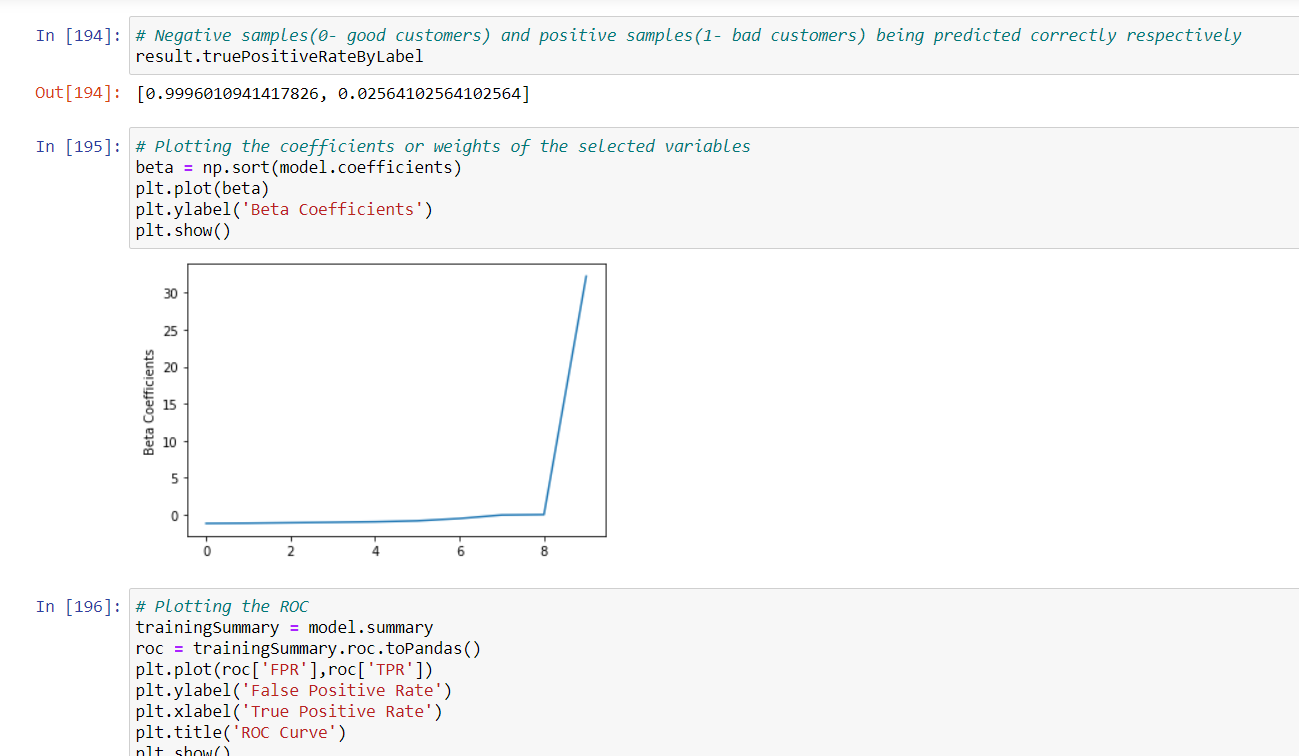
* **Model Building: Logistic Regression**



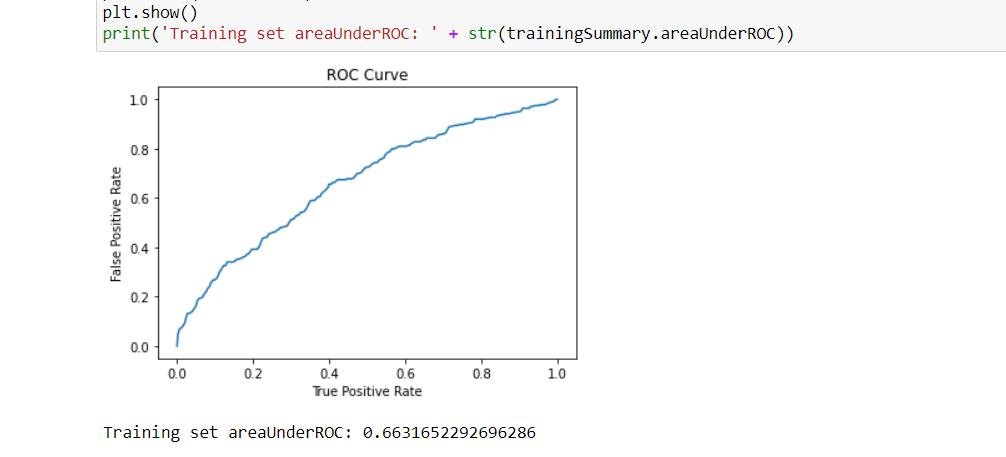


Model accuracy is 98.4%

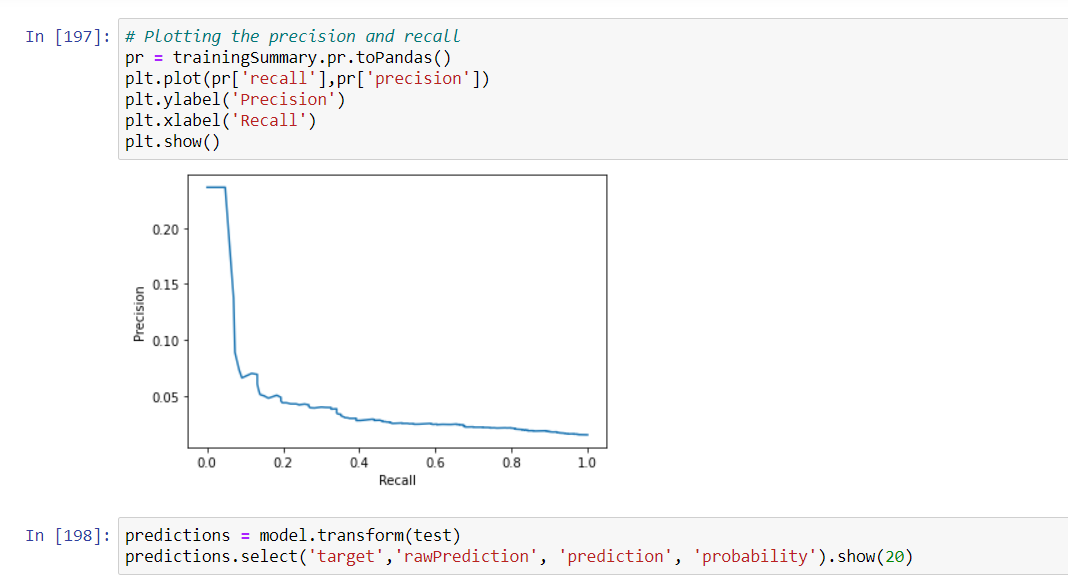
Area Under ROC is 66.3%



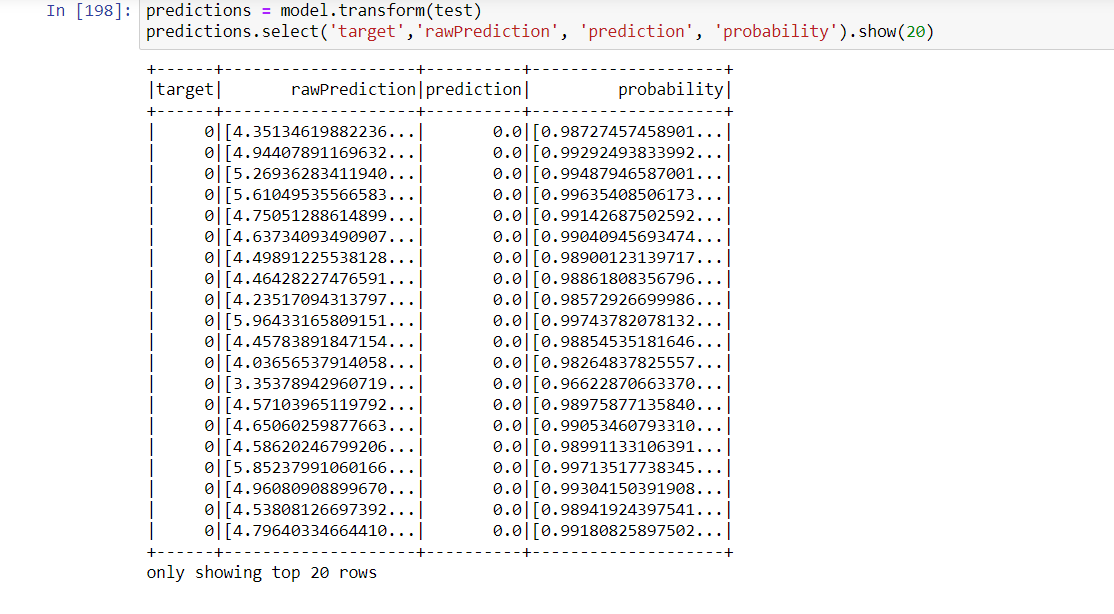
Beta coefficients

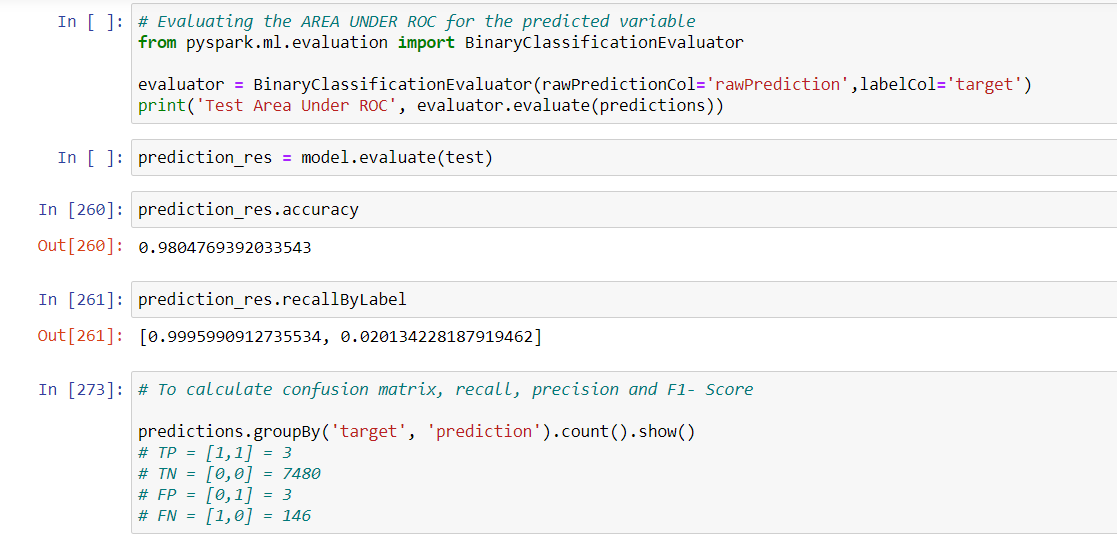


Roc curve

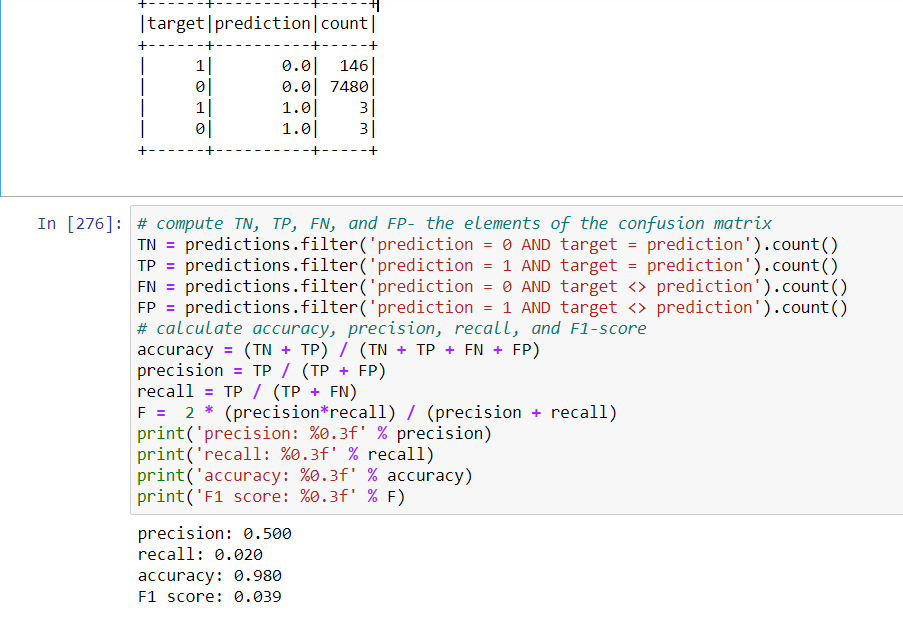


Precision VS Recall graph





Confusion Matrix, Precision, Recall, F1 Score



Precision is 50%

Recall is 2%

F1-Score is 3.9%

* FINE TUNING OF THE LOGISTIC REGRESSION MODEL

