**Problem Definition**

An increase in customer default rates negatively affects Credit One’s profitability, performance, and reliability since its business is approving customers for loans in the first place.

**Project Goal**

Investing:

* If/how Credit One can ensure that customers can/will pay their loans
* If/how Credit One can approve customers with high certainty

**Our Approach and Methodology:**

Performing a thorough exploratory data analysis to identify patterns in customers’ behavior. Visualizing trends and building a classification model using the Random Forest Algorithm to predict customers’ behavior based on the historical data we have been provided.

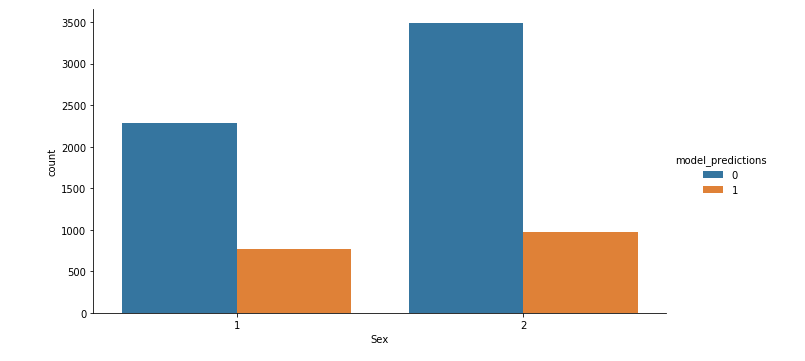
The confidence level of our model is 81%.

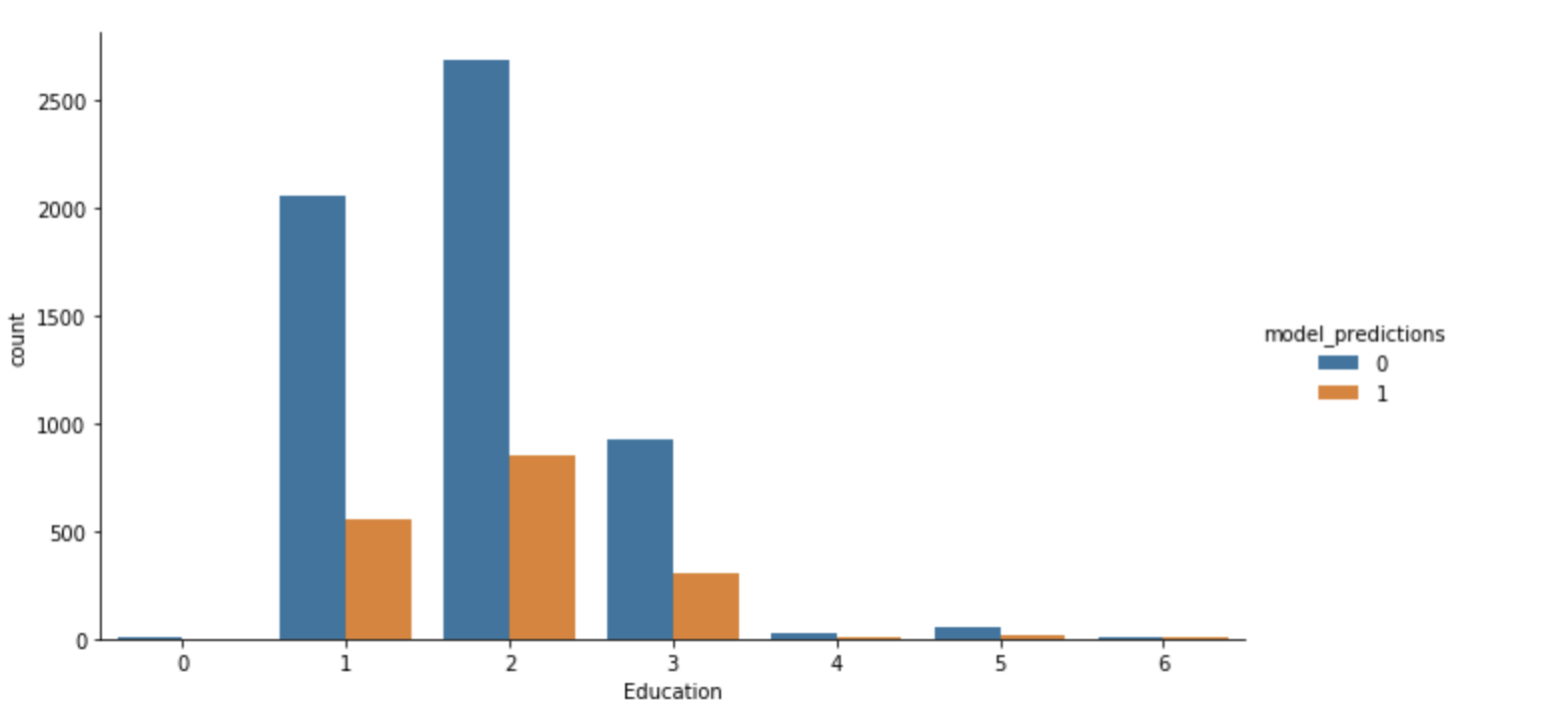
**Findings:**

* Early to mid-career customers are more likely to default their loans
* Customers with higher education levels (Education levels 4 to 6) can be considered low risk.
* Education level 2 (university level) is high risk in both genders.
* Singles are more likely to default their loans. Divorced customers are less likely to default.
* Overall, the age range of divorced customers and those with higher education levels is higher and there is a statistically significant positive relationship between age and loan payment. Older customers are more likely to pay their loans while yongers are more likely to default their loan.
* Number of female defaulters is higher than males.

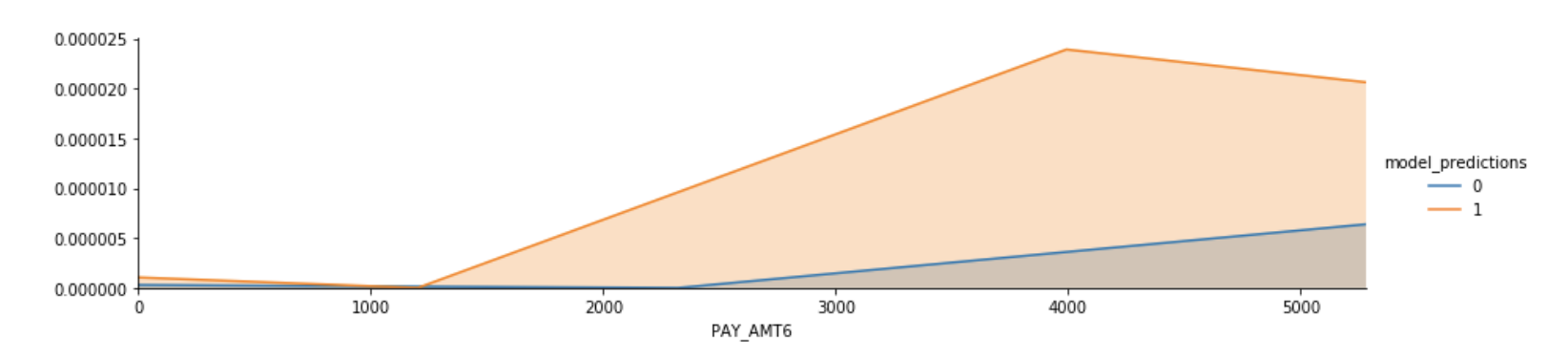
**Models Findings:**

About 23% of customers default their loan. Overal, the number of female customers is higher than males and the number of defaulters is also higher among females.





In general, customers with higher loan amount tend to default it more than those who get smaller loans.



**Conclusion:**

With the data we have, our study couldn’t generalize the specific demographics of customers who default their loans. However, we found a bell-curved payment pattern for defaulters while loyal customers seem to have more ups and downs in their payment patterns.

The model predicted that where education level is 2, customer can be considered ‘high risk’, since the number of defaulters are higher than other education groups.

I believe we can gain more insight from the historical data if we collect information about

customers’ income levels, credit scores, and employment status. With the information we have at the moment, our model can predict who default their loans, but we don’t have enough information to find a meaningful relationship between demographic attributions and customer behavior.