**Credit Default Analysis – Lessons Learned**

**Prepared for Internal Use ONLY**

**November 14, 2019**

**Credit One**

**S Womack**

**Objective:**

Perform Exploratory Data Analysis on Credit One client credit default data.

**Data Utilized:**

Data Source:

* Credit One credit default data
* 30,000 observations

Preprocessing:

* Discretized Age feature into 6 bins:

|  |  |
| --- | --- |
| **Age Bin** | **Count** |
| 18 - 25 | 3,871 |
| 26 - 35 | 12,938 |
| 36 - 45 | 8,522 |
| 46 - 55 | 3,825 |
| 56 – 67 | 799 |
| 68-100 | 45 |

* Combine Education ‘Other” categories:
* Raw data ‘Other’ categories: 0, 4, 5, 6
* New ‘Other’ category: 0
* Conversion of 0, 4, 5, 6 -> 0
* Features to be removed before modeling:
* ID: Id field
* AGE: replaced by AGE\_BIN

**Exploratory Data Analysis:**

Overall demographics

* Majority female
* Majority have higher education
* Close to even split between marital status
* Majority by age between 25 – 40, with peak between 25-32
* 22% in default status

Default demographics:

* More likely to default with lower credit limits
* Less likely to default between 25 – 35
* Males are more likely to default, except in the 55-67 age where women more likely
* High school and then University slightly higher percentage of default than Graduate
* Marriage only very slightly higher default rate than single

Default in relation to billing history:

* Lower credit limits are more likely to default
* More likely to default the longer the payment is delayed.
  + Significant increase in default after 2 delayed payments
  + Another significant increase in default after 6-7 delayed payments (especially going back to May and April at 100% default with 6 missed payments)
* The greater the difference between the billing amount versus the payment amount (high bill but low payment) they more likely to default.

**Recommendations:**

It is not clear what the default date is of the clients in the data file, or if these are just the clients that are in default as of the current month. This information would be important as activity only before default status would be most useful in predicting defaults.

Additionally, useful information to provide would be the default criteria. At which point is an account marked in default, at one delayed payment, at two, etc.?

**Lessons Learned:**

Visualizations can be very powerful and convey a lot of information, but it takes a lot of practice to know the right type of visualization to use for the data at hand. Otherwise it can be time consuming with much trial and error, to extract information.