CREDIT RISK ANALYSIS
&
APPLICATION SCORECARD

BY

<u>BFS 48 – Cohort Sep 2018</u>

S Dhivya *Simanta Dutta *Sushma B Achary *Theertendra Vadakattu

Overview of the Analytics

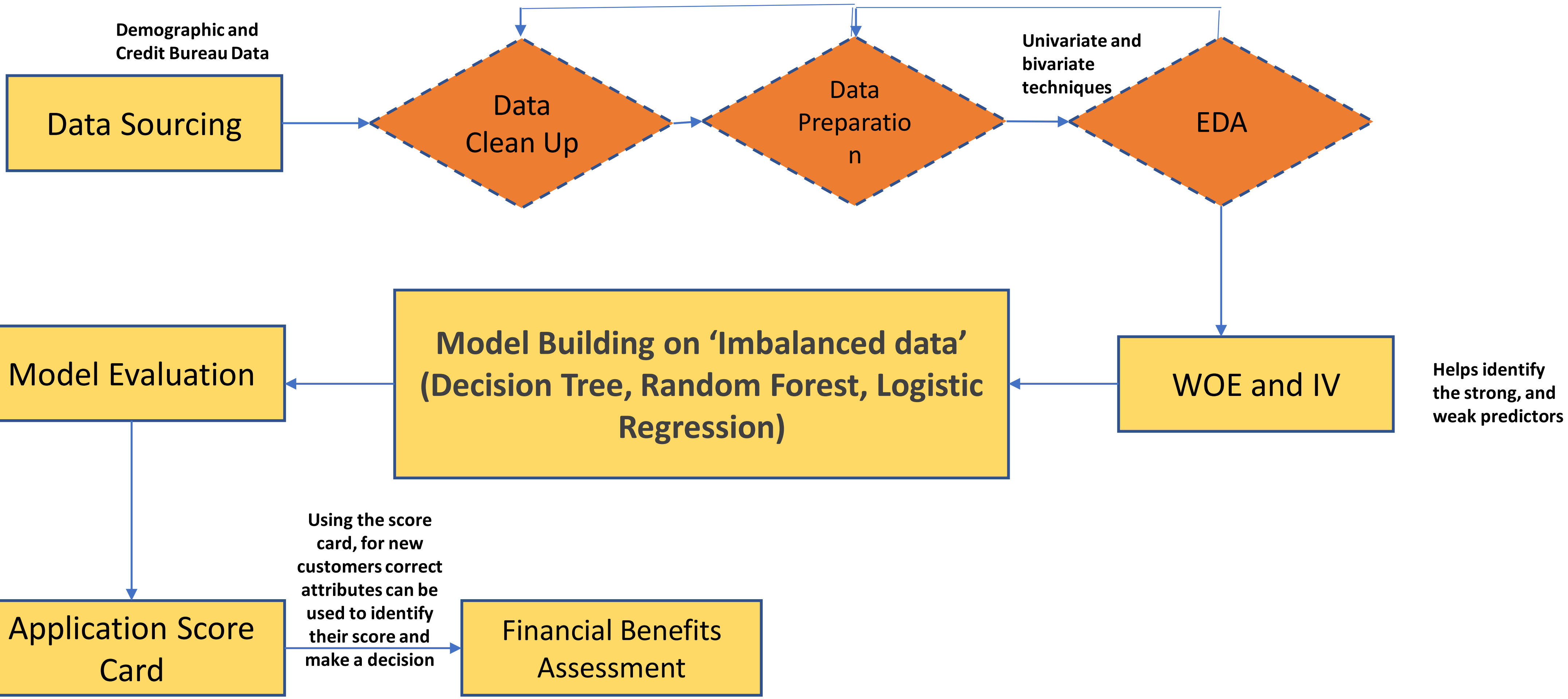
- CredX is a leading credit card provider who get innumerable credit card applications every year out of which approximately **4%** of them **default**, resulting in credit loss.
- The best way to mitigate the **credit risk** is to identify the **good** and **bad**, and help CredX acquire the right customers.
- We have categorised the data set broadly into ‘**Demographic’** and ‘**Credit Bureau’** data.
- The data thus acquired which is highly **‘IMBALANCED’** will go through various model building methodologies and a best modelling technique will then be evaluated.
- The evaluation would then be used, to generate an ‘**Application Scorecard’** using which we can help CredX identify their good v/s bad application requests.
- We will also help CredX with **identifying** the **‘factors’** that increase the **credit risk**, provide an assessment of the ‘**Financial Benefits’** with the **model** developed.

Risk Analytics

- To understand the credit worthiness of the borrower.
- To assign a risk rating to the borrower in lieu of the borrower and lending facilities proposed.
- Help in various financial analyses like ratio and trend analysis, projections and detailed analysis of cash flow, examine collateral and repayment sources, maintain credit history.
- To help predict the probability of a borrower defaulting on a debt and to predict the of losses in such scenarios.

Problem Solving Methodology

Flow chart



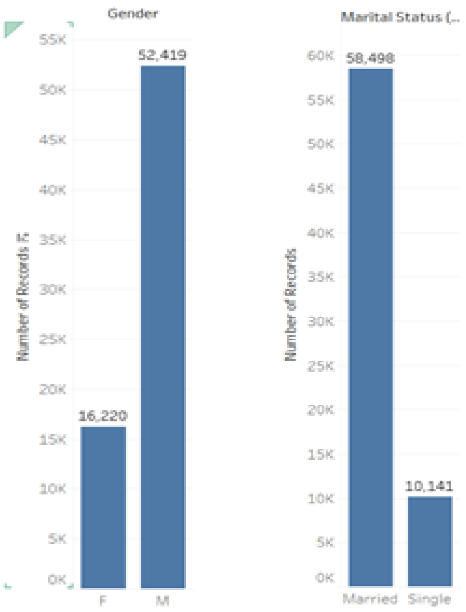
Data Understanding

Description	Demographic Data Information
Source	Past Records
Format	CSV
No. of rows	71298
Response Variable	Performance Tag

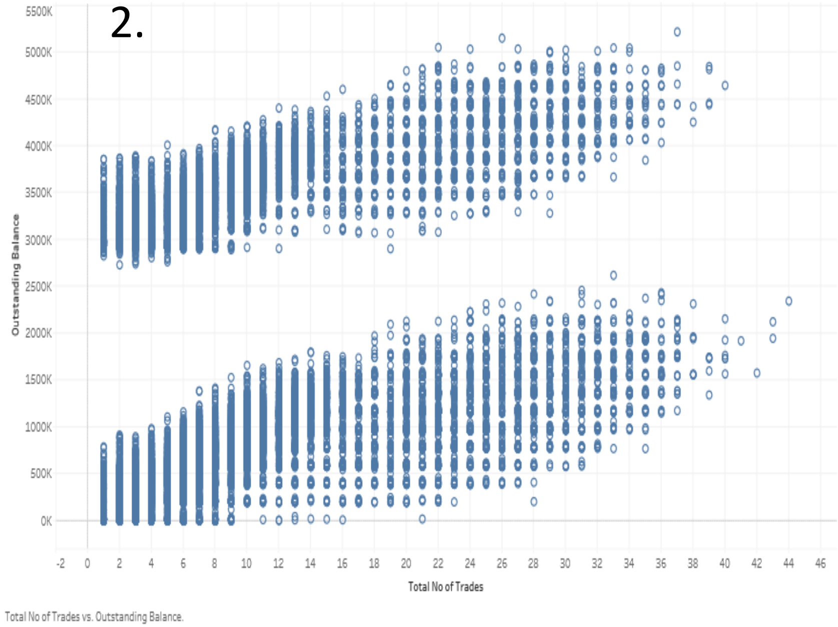
Description	Demographic Data Information
Source	Past Records
Format	CSV
No. of rows	71298
Response Variable	Performance Tag

Exploratory Data Analysis

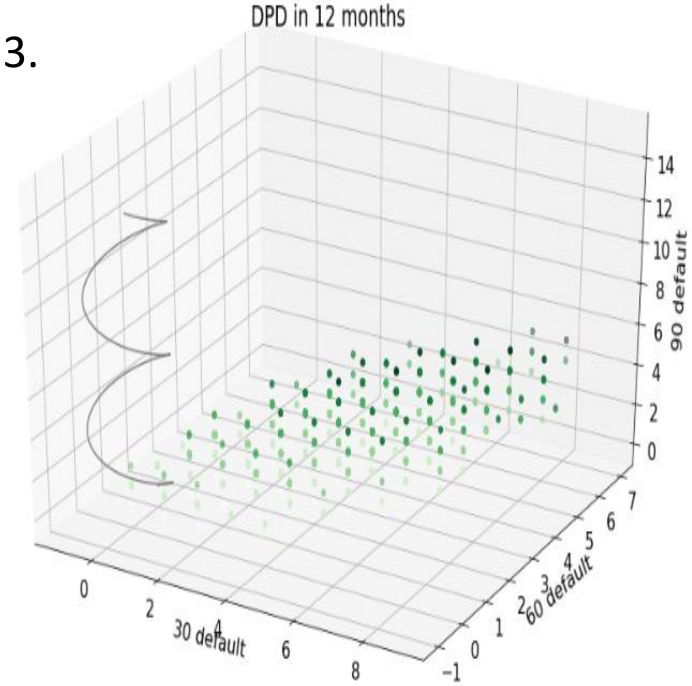
1.



2.



3.

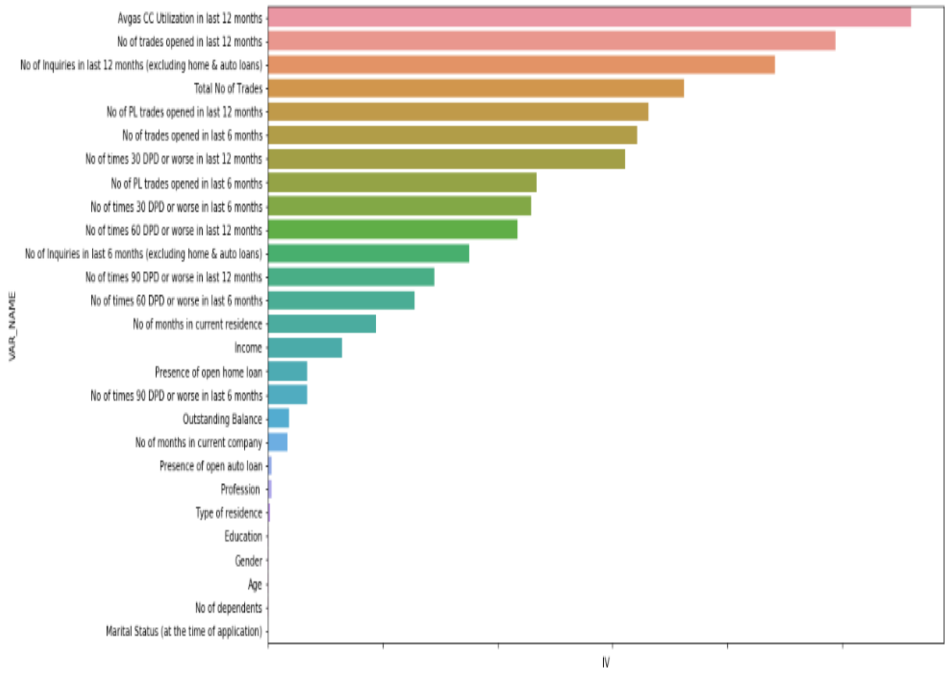


1. Univariate Analysis for Gender and Marital Status shows that Married and Male are placing more requests for Credit Card.
2. Bivariate Analysis for Total Trades and Outstanding Balance shows that as trades increase, outstanding balance also increases and then stays stable.
3. A 3D view of the 30, 60, 90 days default patterns within a year doesn’t significantly explain the defaulters.

WOE and IV Results

Five top most important variables as predictors of 'DEFAULT'

Predictor	Value	Predictive Power
Avgas CC Utilization in last 12 months	0.3	Strong
No of trades opened in last 12 months	0.25	Medium
No of Inquiries in last 12 months (excluding home and auto loans)	0.22	Medium
Total No of Trades	0.18	Medium
No of PL trades opened in last 12 months	0.16	Medium



Model Building and Evaluation

MODELS BUILT	EVALUATION OF MODELS
Supervised classification modelling methods used	
1. **Logistic Regression** : **66.6%**	AUC check being done
2. **Decision Tree** : **68.3%**	Hyperparameter tuning using GridSearch
3. Ensemble Model **Random Forest** : **69.4%**	Accuracy Score check after Hyperparameter tuning

Best Fit Model

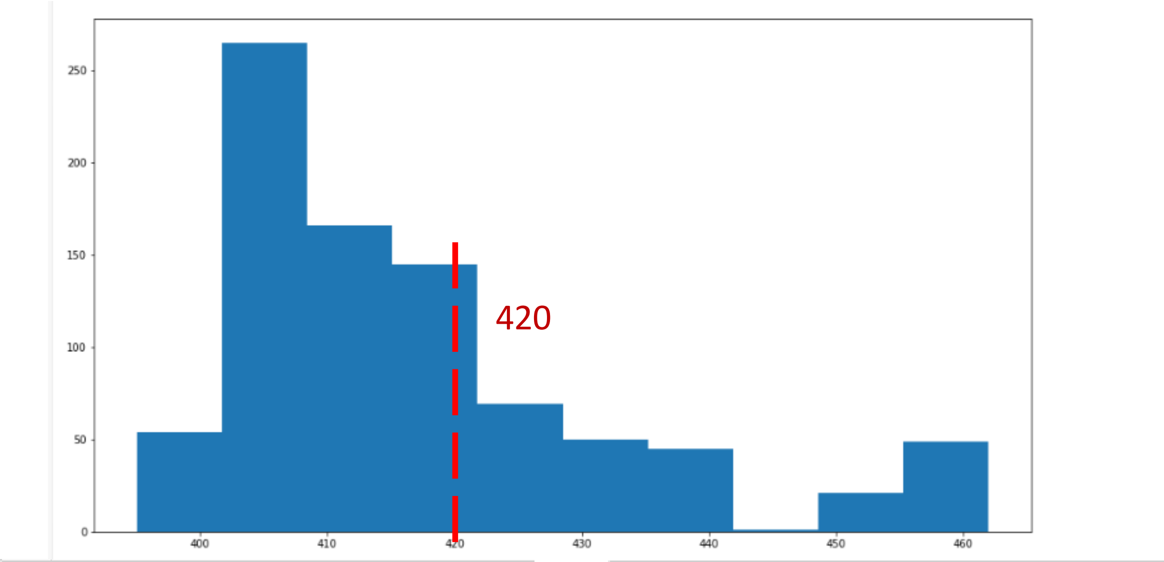
METRICS	SCORE
Model	**Random Forest**
Accuracy	95.78%
AUC (Area Under the Curve)	69.4% (70%)

Application Score Card

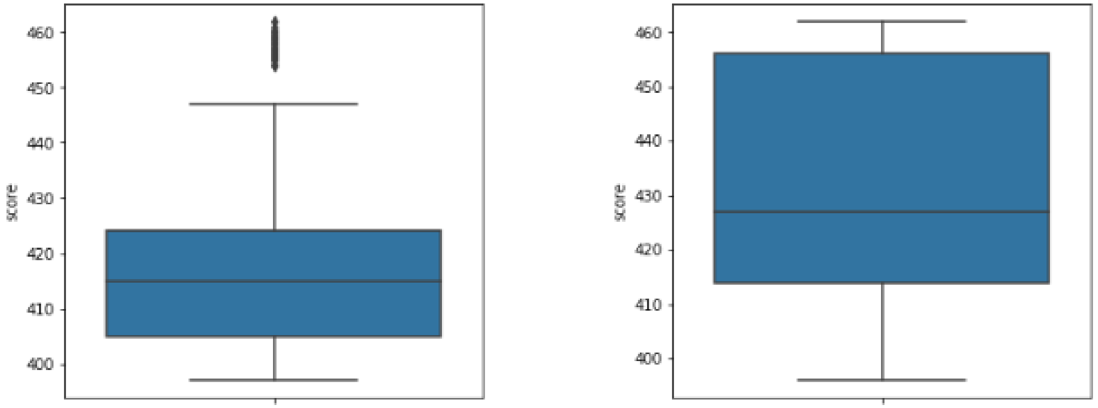
- **Automate** application decisions
- **Reduce** costs of manual underwriting
- Business can make accurate, quick and consistent fact based **decisions**.
- Provides **flexibility** in terms of credit risk strategy – cross selling, risk based pricing etc.
- An applicant's score is equivalent to '**log of predicted bad/good odds'**. A '**cut-off'** score is generated by the model and every application that a bank receives will be approved by the bank only if the applicant **scores above** the **cut-off score**, else application is reject.

Application Score Card Results

METRICS	SCORE
Cut Off Score	**420**
Prediction Probability	0.95
Std Deviation	15.82



- Going by the model ‘**accuracy’** and after the **cut off** score was applied, our model was able to identify **4** out of **1425** applicants as defaulter.
- This means a revenue gain of **99.65%** on the reject population



Rejected v/s approved Score Card

*Approved population (master loan data) is a population for which the application is accepted by bank

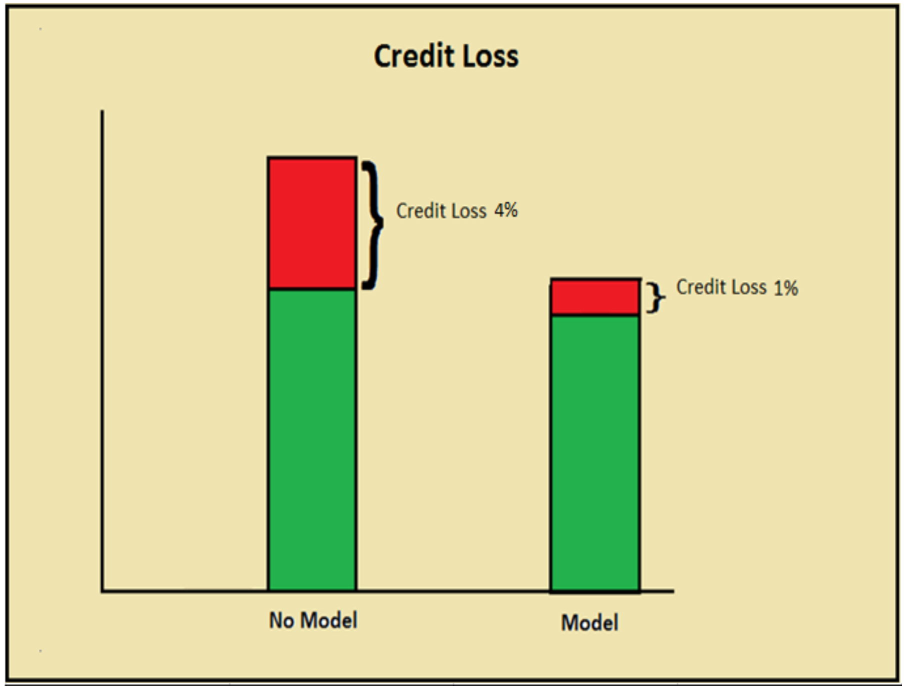
Financial Benefit Analysis.....1

The model reduces **'Credit loss'** from **4%** customers to **1%** customers (approved population). Therefore,

- Credit loss with no model = **4%**

- Credit loss with model = **1%**

- **Credit Loss prevented**: **3%**



Confusion Matrix		Actual Defaults	
		Good Customers(0)	Bad Customers(1)
Predicted Defaults	Good Customers(0)	47938	732
	Bad Customers (1)	18625	2206

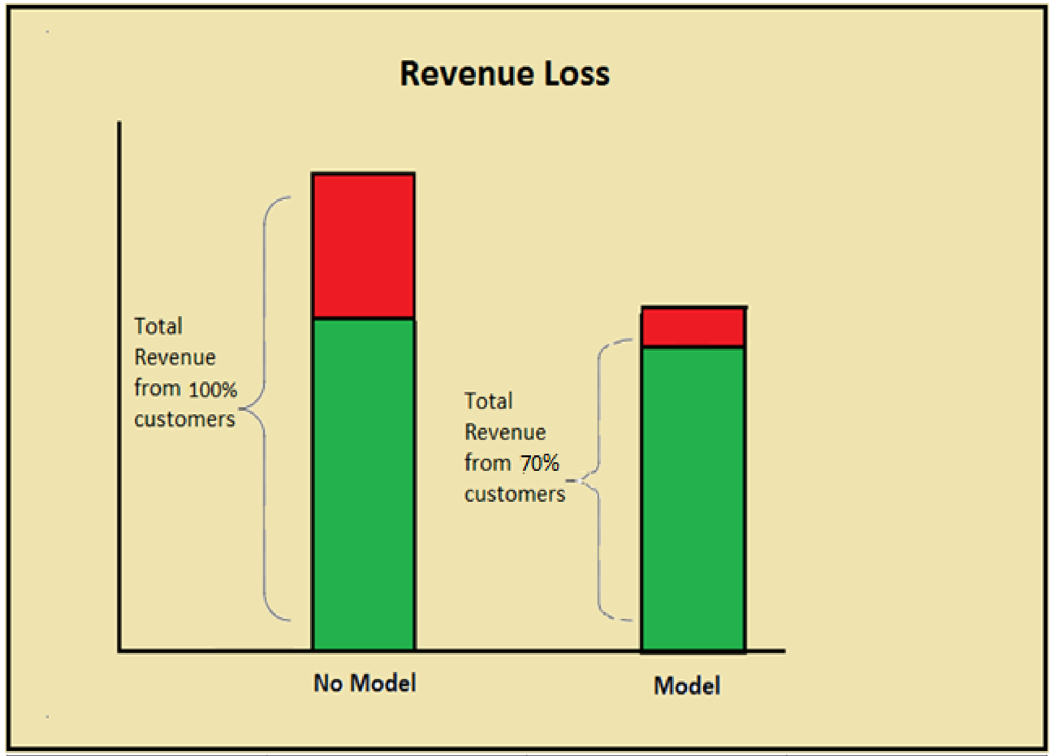
***When data balancing technique like SMOTE was adapted for model building the good bad customers and the financial analysis would have been this accurate.**

Cont….2

Revenue Loss*: Reduced to **30%** of revenue **(Auto-approval-Scorecard)**

- Revenue no model = **100%**
- Revenue with model = **70%**

Revenue Loss : **30%**



Confusion Matrix		Actual Defaults	
		Good Customers(0)	Bad Customers(1)
Predicted Defaults	Good Customers(0)	47938	732
	Bad Customers (1)	18625	2206

*The revenue loss actually occurred by wrongly identifying the good customers as "BAD"

THANK YOU