Data analysis: Banking application

Predicting loan repayment

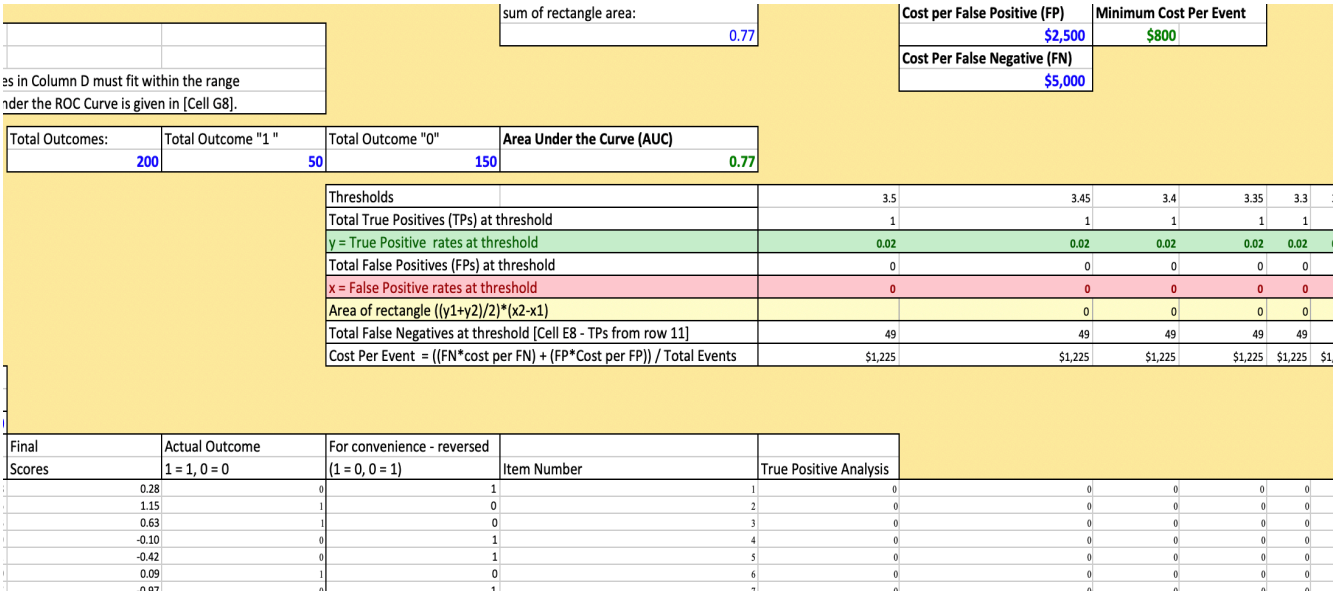
In this project a simple model was developed to lower the risk of default on credit cards by analyzing applicant's attributes such as age, years at current address, years at current employer, amount of debt, car debt and income using Excel.

- **The training and test datasets used in this project were located in AWS RDS (Check Ref [1] for the datasets source).**
- **MYSQL workbench was used to retrieve data into csv files from AWS RDS.**
- **AUC (Area under curve) of ROC curve was used to measure the model performance by using AUC calculator in Excel.**
- **The simple model was developed by trial and error and experimenting with various models, and comparison of AUCs of ROC curves.**

Simple model:

(-6*(raw cc debt + raw auto debt)/(raw income))
-(.3*(standardized age + standardized years at employer + standardized years at address)) – 2.9

This model has an AUC of .77 on the training set and 0.79 on the test set. The AUC was calculated in Excel using AUC calculator as shown below. Larger view of AUC calculator is uploaded in a separate file. This model is not the best one, however it provides robust performance on both training and test dataset.



			sum of rectangle area:			Cost per False Positive (FP)		Minimum Cost Per Event		
			0.77			$2,500		$800		
						Cost Per False Negative (FN)				
						$5,000				
es in Column D must fit within the range										
nder the ROC Curve is given in [Cell G8].										
Total Outcomes:	Total Outcome "1 "	Total Outcome "0"	Area Under the Curve (AUC)							
200	50	150	0.77							
		Thresholds		3.5	3.45	3.4	3.35	3.3		
		Total True Positives (TPs) at threshold		1	1	1	1	1		
		y = True Positive rates at threshold		0.02	0.02	0.02	0.02	0.02		
		Total False Positives (FPs) at threshold		0	0	0	0	0		
		x = False Positive rates at threshold		0	0	0	0	0		
		Area of rectangle ((y1+y2)/2)*(x2-x1)			0	0	0	0		
		Total False Negatives at threshold [Cell E8 - TPs from row 11]		49	49	49	49	49		
		Cost Per Event = ((FN*cost per FN) + (FP*Cost per FP)) / Total Events		$1,225	$1,225	$1,225	$1,225	$1,225	$1	

Final Scores	Actual Outcome 1 = 1, 0 = 0	For convenience - reversed (1 = 0, 0 = 1)	Item Number	True Positive Analysis					
0.28	0	1	1	0	0	0	0	0	
1.15	1	0	2	0	0	0	0	0	
0.63	1	0	3	0	0	0	0	0	
-0.10	0	1	4	0	0	0	0	0	
-0.42	0	1	5	0	0	0	0	0	
0.09	1	0	6	0	0	0	0	0	

Note: MySQL command to create the table in AWS RDS is available in a separate file.

Ref [1]: Mastering data analysis in Excel course, Duke University @ Coursera.