Mortgage Submissions - Analysis

Chris Baden

Nima Baradar

Dawn Bui

Curtis Hardrick

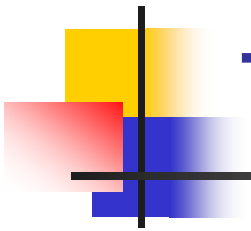
Subramanian Hariharan

Key findings

- *Our model indicates the loan approval process at XYZ Mortgage may be overly conservative.*
  - The Firm may be able to increase profits by adjusting the target approval rate.
- *Credit scores and length of loan appear to play an insignificant role in determining loan approval.*
  - The Firm may be able to decrease transaction costs.

Overview

- The Problem
- Data Characteristics
- Analysis
- Interpretation
- Recommendations
- Summary



The Problem

- XYZ Mortgage Company processes loan applications from borrowers
- Underwriting is done with software & rules determined by the company
- The Company wants to:
  - Analyze a sample of loan applications
  - Improve their decision making process

Data – Predictors & Response

- Predictors
  - Loan Amount
  - Loan term → Categorical
  - Interest Rate
  - Monthly Income
  - Credit Score
  - Outstanding Liability
- Response
  - Approve/Disapprove → Categorical/Binary
  - ~~Sell? (Not Considered)~~

Data - Source

- The data was obtained from a sample of submissions to the company over 1 month.
- Data required some normalization
  - Outliers
  - Invalid values
  - Duplicates
- Final sample (n = 237)

Data - Characteristics

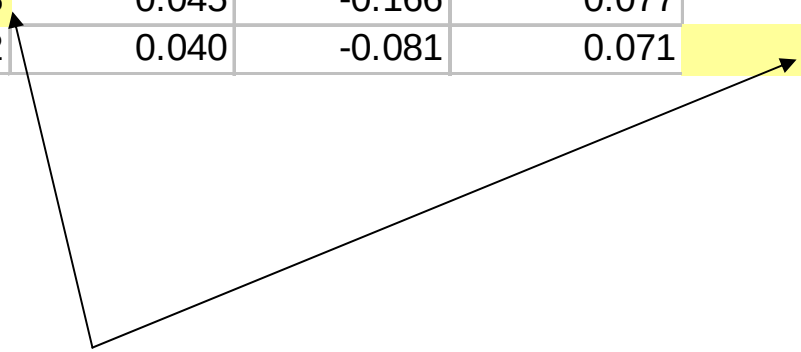
	Min	Max	Mean	Median
Amount	$14,000	$322,000	$140,000	$122,800
Term	10 yr	30 yr		
Rate	3.5%	8%	6%	6.125%
Income	$524	$18,739	$6046	$5552
Score	500	814	700	707
Liability	$632	$1.46M	$161,510	$114,971

Sample of the data

LOAN AMT	Term	Rate	Score	Income	Liability	Outcome
130000	360	7	743	2750	76889	NotApproved
87300	360	6.75	716	4565	66019	NotApproved
322000	360	6.5	624	5180	48846	NotApproved
153000	360	5.25	755	8750	179526	Approved
150000	240	5.75	707	5084	175471	Approved
247500	360	6.25	732	9000	174688	Approved

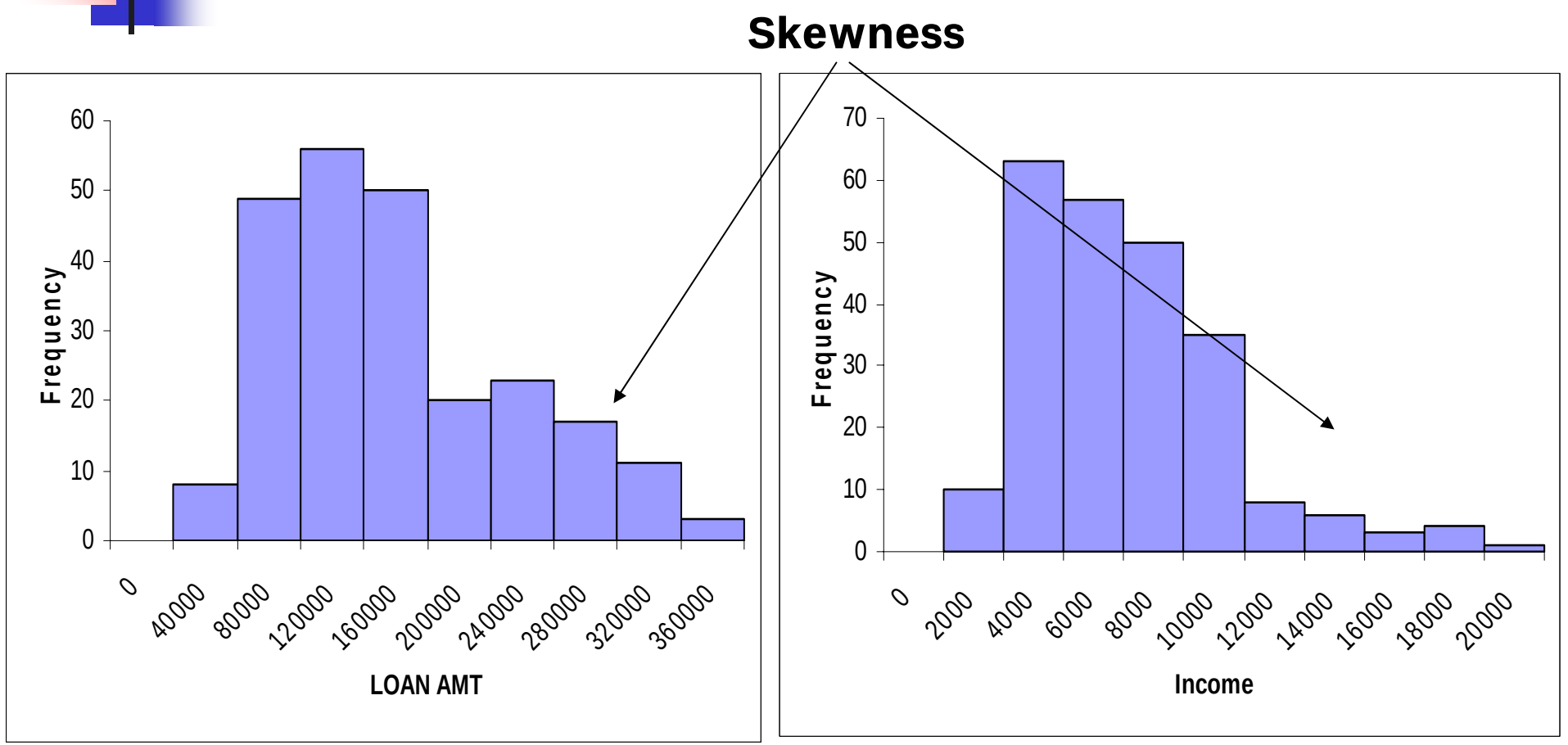
Data observations: Correlation Matrix

Table of correlations							
		Loan Amount	Term	Rate	Score	Income	Liability
	Loan Amount	1.000					
	Term	0.296	1.000				
	Rate	-0.210	0.190	1.000			
	Score	0.027	0.014	-0.095	1.000		
	Income	0.508	0.045	-0.166	0.077	1.000	
	Liability	0.232	0.040	-0.081	0.071	0.404	1.000

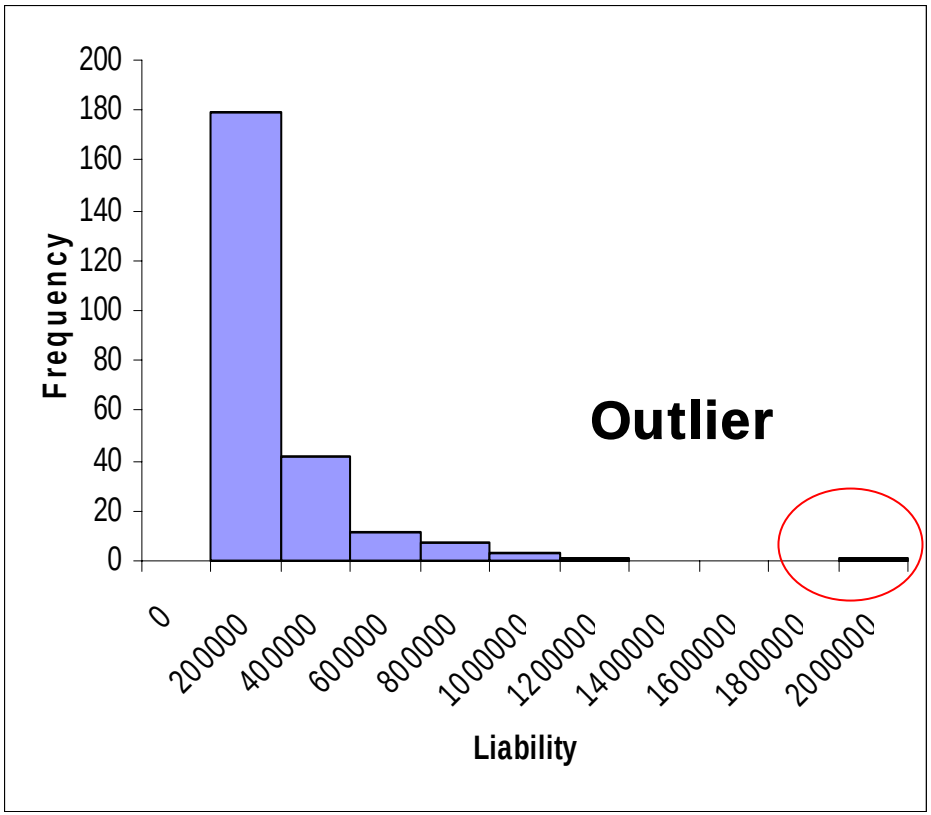


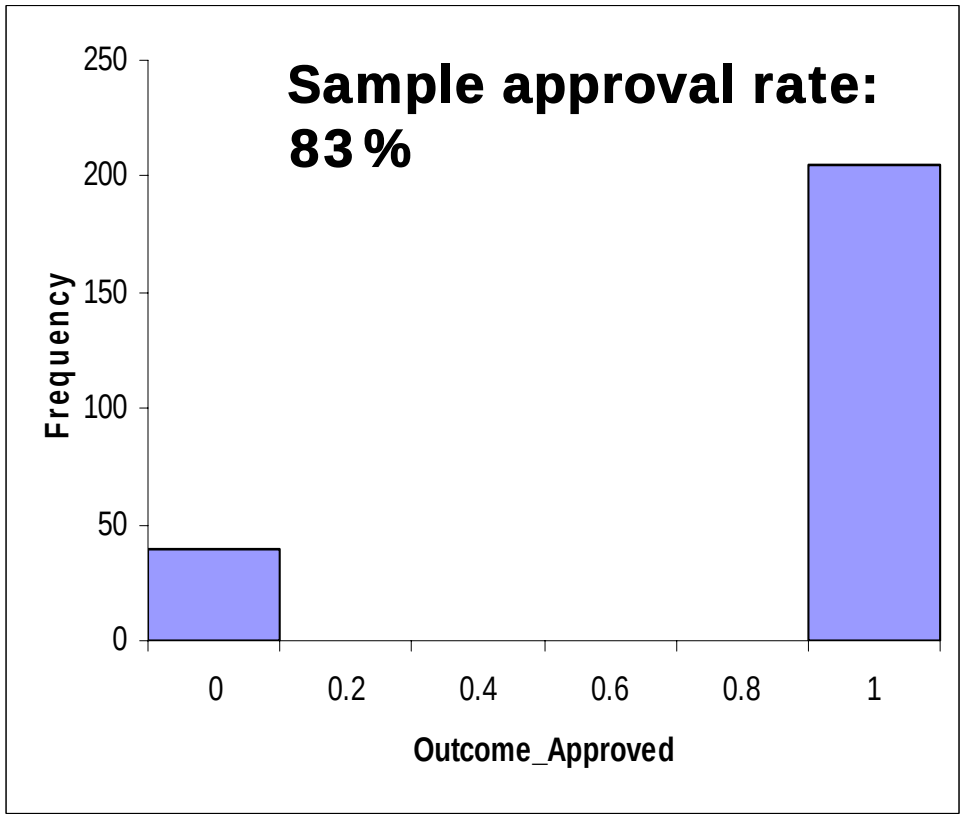
Correlations

Data observations: Loan Amount and Income



Data Observations: Liability and Approval Rate





Summary Statistics

Runs Test Results for Income		
	Number of obs	237
	Number above cutoff	106
	Number below cutoff	131
	Number of runs	2
	E(R)	118.181
	Stdev(R)	7.595
	Z-value	-15.297
	p-value (2-tailed)	0.000

Table of correlations								
		Loan Amount	Term	Rate	Score	Income	Liability	Outcome_Approved
	Loan Amount	1.000						
	Term	0.296	1.000					
	Rate	-0.210	0.190	1.000				
	Score	0.027	0.014	-0.095	1.000			
	Income	0.508	0.045	-0.166	0.077	1.000		
	Liability	0.232	0.040	-0.081	0.071	0.404	1.000	
	Outcome_Approved	-0.051	-0.085	-0.148	-0.032	0.149	-0.073	1.000

Data Discovery: Cluster Analysis

Cluster Centers

Cluster	Loan Amount	Term	Rate	Score	Income	Liability
Cluster-1	98094.74177	184.5280207	5.7099061	698.3395964	5820.85231	133090.635
Cluster-2	202983.7266	357.3134549	5.65067209	714.9403096	7106.115717	127508.6933
Cluster-3	97441.67526	356.842151	6.45210546	683.0210158	4204.941432	117237.0112
Cluster-4	231272.5361	351.8181823	5.76704564	729.2272189	11314.81459	524698.4483

Cluster Summary

Cluster	#Obs	Average distance in cluster
Cluster-1	53	1.554
Cluster-2	67	1.594
Cluster-3	95	1.425
Cluster-4	22	2.509
Overall	237	1.602

Cluster Analysis-- Interpretation

Clusters 1&3: Low loans, scores, incomes, liabilities

Cluster	Loan Amount	Term	Rate	Score	Income	Liability
Cluster-1	98094.74177	184.5280207		698.3395964	5820.85231	133090.635
Cluster-3	97441.67526	356.842151		683.0210158	4204.941432	117237.0112

Cluster Summary

Cluster	#Obs	Average distance in cluster
Cluster-1	53	1.554
Cluster-2	67	1.594
Cluster-3	95	1.425
Cluster-4	22	2.509
Overall	237	1.602

Cluster Analysis--Interpretation

Clusters 2&4: High loans, scores, incomes, liabilities

Cluster	Loan Amount	Term	Rate	Score	Income	Liability
Cluster-2	202983.7266	357.3134549		714.9403096	7106.115717	127508.6933
Cluster-4	231272.5361	351.8181823		729.2272189	11314.81459	524698.4483

Cluster Summary

Cluster	#Obs	Average distance in cluster
Cluster-1	53	1.554
Cluster-2	67	1.594
Cluster-3	95	1.425
Cluster-4	22	2.509
Overall	237	1.602

Data Analysis

- Methods considered
  - Logistic Regression (pursued)
    - Discriminate between the value of 6 predictors
    - Classify applicants into approved/disapproved
    - Predict future loan approvals
    - Reduce misclassifications
  - Discriminant Analysis (not pursued)
    - Actual proportion of approved loans unknown
    - Cost of misclassification unknown
    - Predictors do not appear normally distributed

Logistic Regression

- Followed an iterative process
- Started with all 6 predictors in the model
- Predictors with least impact were removed (based on p-values, and error %)
  - Loan Term
  - Credit Score

Logistic Regression – Results

Model with all six predictor variables

Regression coefficients								
		Coefficient	Std Err	Wald	p-value	Lower limit	Upper limit	Exp(Coeff)
	Constant	7.1683	2.7789	2.5795	0.0099	1.7216	12.6149	
	Loan Amount	0.0000	0.0000	-2.7072	0.0068	0.0000	0.0000	1.0000
	Term	-0.0003	0.0031	-0.0910	0.9275	-0.0064	0.0058	0.9997
	Rate	-0.7168	0.3118	-2.2989	0.0215	-1.3278	-0.1057	0.4883
	Score	-0.0018	0.0026	-0.6759	0.4991	-0.0068	0.0033	0.9982
	Income	0.0003	0.0001	3.3625	0.0008	0.0001	0.0005	1.0003
	Liability	0.0000	0.0000	-1.9276	0.0539	0.0000	0.0000	1.0000

Slice of the model with variable “term” removed

	p-value	0.0003						
Regression coefficients								
		Coefficient	Std Err	Wald	p-value	Lower limit	Upper limit	Exp(Coeff)
	Constant	7.149354	2.7669	2.5839	0.0098	1.7262	12.5725	
	Loan Amount	-0.000009	0.0000	-2.8777	0.0040	0.0000	0.0000	1.0000
	Rate	-0.726207	0.2945	-2.4657	0.0137	-1.3035	-0.1489	0.4837
	Score	-0.001767	0.0026	-0.6834	0.4944	-0.0068	0.0033	0.9982

Logistic Regression - Results

Best Fit Model

Regression coefficients								
		Coefficient	Std Err	Wald	p-value	Lower limit	Upper limit	Exp(Coeff)
	Constant	5.801298	1.9033	3.0480	0.0023	2.0708	9.5318	
	Loan Amount	-0.000009	0.0000	-2.8712	0.0041	0.0000	0.0000	1.0000
	Rate	-0.706254	0.2917	-2.4209	0.0155	-1.2780	-0.1345	0.4935
	Income	0.000316	0.0001	3.3863	0.0007	0.0001	0.0005	1.0003
	Liability	-0.000002	0.0000	-2.0187	0.0435	0.0000	0.0000	1.0000

Classification matrix		Predicted		
Actual		Disapproved	Approved	Pct Correct
	Disapproved	3	36	07.7%
	Approved	1	197	99.5%

Summary of overall classification results		
	Pct correct	84.4%
	Base	83.5%
	Improvement	05.1%

Model appears overly generous in predicting approvals that XYZ rejected.

Is XYZ too conservative?

Model is very accurate in predicting actual approvals

Recommendations

- Company may want to consider a less conservative approach to denying loans.
  - Costs of misclassification may be a significant driver—should be explored.
- Incorporating credit scores and loan terms into the approval/disapproval model may incur unnecessary transaction costs.
  - Score may be driven by misclassification costs.

Summary

- Loan Amount, Income, Liability and Rate are good predictors of approval/disapproval.

- Company may consider modifying its approval process to target the different borrower segments (as found from the cluster analysis)