DECISION 518Q Applied Probability and Statistics
Team Project Report - Final Submission
Truc Pham, Fengquan Sun, Devet Valecha, Yi Zhao & Tony Zitong Gao
August 22th, 2019

Research Question
Multidimensional Assessment of Home Mortgage Approval Rate in Washington State

Industry: Commercial Banking
Year of Data: 2016
Dataset: https://www.kaggle.com/washington-state-hmda
Description: https://cfpb.github.io/api/hmda/fields.html
Glossary: https://www.ffiec.gov/hmda/glossary.htm

Table of Contents

1. Executive Summary

The Home Mortgage Disclosure Act (HMDA) requires financial institutions to publish information about home loans upon origination. The law was enacted by congress in 1975 and was enforced by the Federal Reserve Board's Regulation C in 2011. This dataset discloses loan details of Washington State in 2016.

This report is aimed to predict the likeliness of applicants having their loans approved, by taking many factors into consideration, which could be classified as demographic, geographic, loan related, and property related information. On one hand, loan applicants can evaluate the chance of having their loans approved based on the customer-driven model. On the other hand, commercial banks can utilize the stepwise selection model to reduce default risks, while government could design the legislation accordingly.

With the infinite amount of factors that can influence the decision of an institution, 19 variables have been considered, spanning across the 4 categories mentioned above. In the later section, variables have been selected to construct various regression models based on 334363 records from HMDA Washington State, 2016.

2. Data Wrangling Process

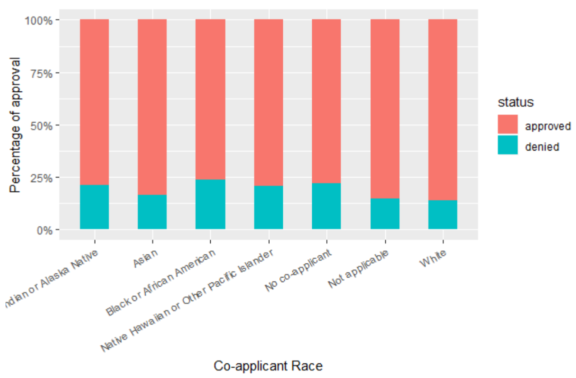
In order to prepare for the analysis, data has been wrangled transformed into a version that is easier for understanding and manipulation. Below are operations conducted on the data in sequential order:

1. Deleted 28 irrelevant variables that provide no value for solving the research question
2. Renamed every column variable for the ease of manipulation
3. Converted *"Minority_Population"* to percentage format, rounded to 2 decimal places
4. Filtered out irrelevant records that provide no value for solving the research question
    - Missing information of income from *"Applicant_Income"* column
    - Missing information of population from *"Tract_Population"* column
    - Missing records of *"Owner_Occupied_Units_Number"* column
5. Grouped remaining variables into 4 categories by key attributes
    a. Demographic Information -Including applicant's gender, race, income, etc.
    b. Property Information - Including property type
    c. Local Application Parameters - Including loan amount, loan type, loan purpose, etc.
    d. Geographic Information - Including tract population, county name, median county income, tract minority population etc.
6. Inserted *"Approval"* as an additional variable according to the categories listed by *"Loan_Status"* column

a. Approved = 1, contains categories including *"Application approved but not accepted"*, *"Loan originated"* & *"Loan purchased by the institution"*
b. Denied = 0, contains categories including *"Application denied by financial institution"*, *"Preapproval request approved but not accepted"* & *"Preapproval request denied by financial institution"*
c. Removed records contain categories of *"Application withdrawn by applicant"* & *"File closed for incompleteness"*

As mentioned in the Executive Summary, final dataset contains 19 variables and 334363 records with complete information on applicant's income and loan amount. Each row of the dataset represents a scenario of home loan application. Consequently, the relationship between the response variable *"Approval"* and independent variables are examined.

3. Data Exploration and Visualization



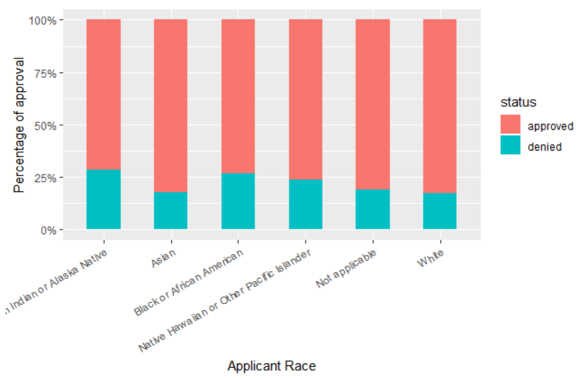


Exhibit 3.1: The above graphs imply insignificant discrimination based on ethnicity and race favoring Asians and White. The percentage of denial is higher for Indian/Asian Native and Black/African American.

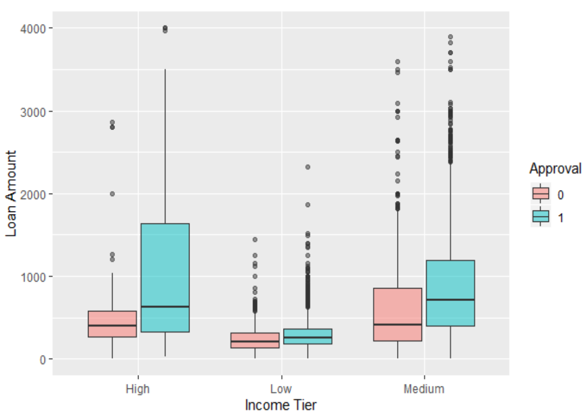


Exhibit 3.2: The annual income tiers are defined as follows: low (less than \$500, 000), medium (from \$500, 000 to less than \$2, 000, 000) and high (equal to or more than \$2, 000, 000). In general, high income borrowers face less restrictions on the amount of loans that could be granted, compared to the low income tier for which borrowers could only take out mini-sized home loans. This shows that income is likely to be a factor that influences the likelihood of loan application success.

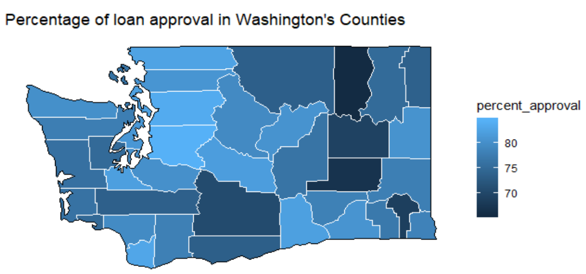


Exhibit 3.3: This graph shows that King County has the highest rate of approval (84.56%), while Ferry County holds the

lowest (65.61%). This is potentially due to the extreme variation in sample sizes between different counties.

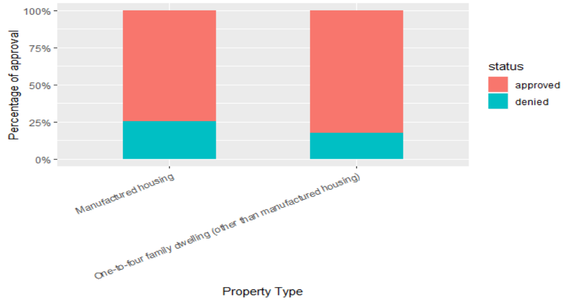


Exhibit 3.4: If people apply for manufactured housing, their applications will likely be rejected more times than one-to-four family dwelling. To be specific, the denial rate is 25.44% compared to 17.72%.

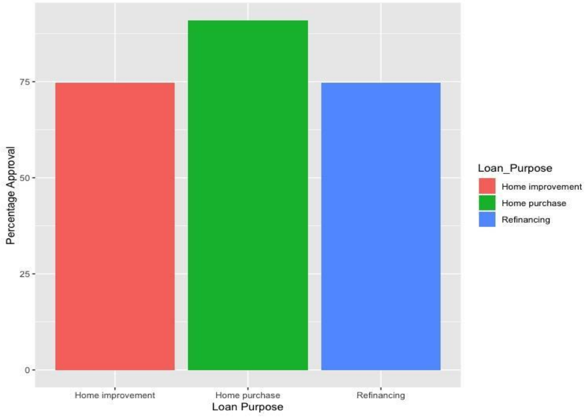


Exhibit 3.5: The graph suggests that people who applied for a loan for home purchasing (91.0%) had a greater chance of having their loans approved, compared to refinancing (74.7%), or home improvement (74.8%)

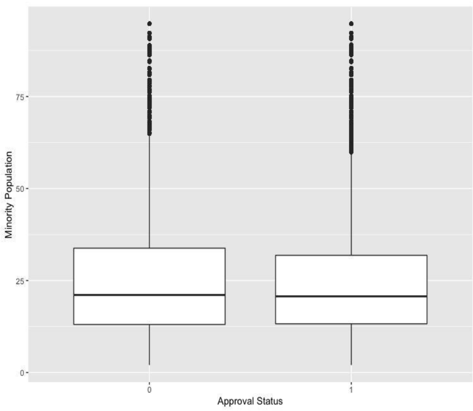


Exhibit 3.6: As per the graph, the minority population does have a slight effect on the approval status. In general, areas with lower minority population have a greater chance of approval.

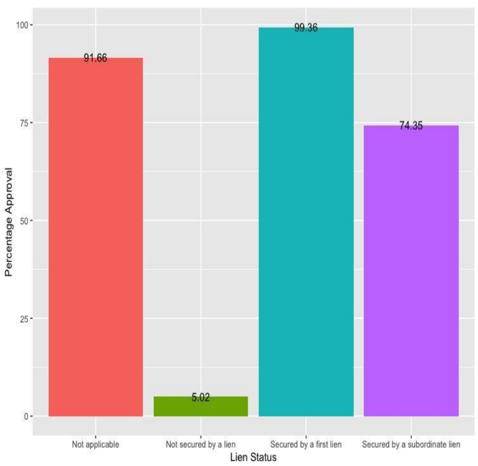
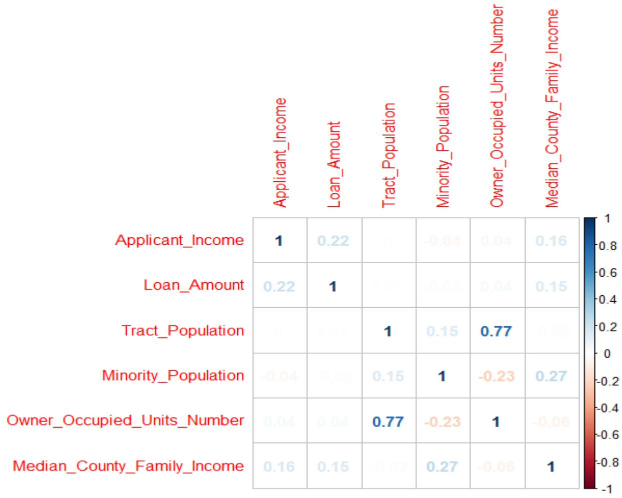


Exhibit 3.7: Evidently, the lien status was a major influencer in the application decision. Not having a lien to secure the loan reduced the chances to ~5% compared to ~74-99% for other categories.

4. Correlation Analysis

In order to gain a better understanding of the relationships between independent variables, correlation analysis is carried out using the PerformanceAnalytics package in R. Only continuous numerical variables are examined given their quantitative nature. The correlation matrix is drawn out for 6 variables: *"Applicant_Income"*, *"Loan_Amount"*, *"Tract_Population"*, *"Minority_Population"*, *"Owner_Occupied_Units_Number"*, *"Median_County_Family_Income"*.



The matrix shows that *"Tract_Population"* is highly correlated with *"Owner_Occupied_Units_Number"* given a correlation coefficient of 0.77. This logic is sound since a more condensed area would have more owners living in their owner property, one that they purchased with the home loan. In addition, *"Minority_Population"* is positively correlated with *"Tract_Population"* and *"Median_County_Family_Income"*, also negatively correlated with *"Owner_Occupied_Units_Number"*. This is possibly due to the fact that all 4 variables belong to the category of geographic information and hence relationships already exist between them.

Furthermore, *"Applicant_Income"* is positively correlated with *"Loan_Amount"* with the correlation coefficient equals to 0.22. Although not every borrower asks for a loan that matches with the maximum amount that they can borrow, the relationship presented demonstrates that loan originating institutions tend to award grant larger loan amounts to individuals with higher incomes. This finding matches with the expectation given that higher income earners are perceived as applicants with low default risks and high credits.

5a. Regression Analysis - Customer-driven Model

A customer-driven model is based on the outcomes of data visualization and built on variables that loan applicants can easily utilize to analyze the possibility of getting their loans approved. The base model shown below equips with an AIC of 302252 and R^2 of 0.04:

$$\hat{Approval} = Applicant\_Gender + Applicant\_Race + Applicant\_Income + County\_Name + Loan\_Amount + Property\_Type$$

It includes 6 variables in 4 categories, and it could be very helpful to loan applicants. This is because every variable in this model is easy to get and loan applicants can easily and intuitively assess their possibility of getting their loan applications approved. Key insights from modelling results include:

1. Demographic information - Male, Asian and While are more likely to get their loan approved. The higher the income, the higher the possibility of getting loans approved.
2. Property information - Manufactured housing are less likely to be approved than one-to-four family dwelling
3. Loan Application Parameters - The higher the loan amount, the higher the possibility of getting loan approved. However, this maybe because loan amount is positively correlated with applicant's income.
4. Geographic information - It is significant that the possibility of getting loans approved varies in different County.

5b. Regression Analysis - Intuitive Model

The intuitive model is mostly derived based on findings from data visualization as well as correlation analysis. Assuming that *"Log_{10}(Applicant_Income)"* and *"Log_{10}(Loan_Amount)"* are two major variables that directly influence the decision-making process of home loan approval, *"Minority_Population"* is added given its relevance observed from the correlation matrix. The negative coefficient from the result set implies that buying a house in a tract with greater minority population may reduce the probability of having the loan granted.

In addition, *"County_Name"* is not included in the intuitive model due to the suspicion that data related to county and tract may not be independent from each other. Given its significance as explained above, *"Tract_Population"* has been prioritised to represent geographic information in the intuitive regression model. The final model is shown below:

$$\widehat{Approval} = log(Applicant\_Income) + log(Loan\_Amount) + Loan\_Type + Loan\_Purpose + Minority\_Population + Applicant\_Gender + Applicant\_Race + Co\_Applicant\_Gender + Property\_Type$$

In hindsight, multiple variable iterations have been conducted to test the significance of other variables and the final intuitive model consists of 9 independent variables with log transformations performed on *"Applicant_Income"* and *"Loan_Amount"*. The result set also shows that there is a smaller probability of loan application acceptance associated with male applicants compared to female applicants. Although the estimated coefficient is not significant according to the p-value, out of all ethnicities, black or African American has the lowest probability to secure a home loan. In terms of the loan type, conventional loan is still the safer bet and has a higher probability of loan application acceptance compared to other loan types.

5c. Regression Analysis - AIC Stepwise Selection
Starting by selecting 16 variables from raw data. Using in-built function in R step() in both directions, some of the variables are omitted based on the AIC value. After excluding variables that are not statistically significant, the final model consists of 14 predictors with an AIC of 274828 and R^2 of 0.1273. After 19 iterations, the final model through automatic backwards selection is as below:

$$\widehat{Approval} = Applicant\_Gender + Applicant\_Race + Co\_Applicant\_Gender + Co\_Applicant\_Race + log(Applicant\_Income) + Property\_Type + County\_Name +$$

log(Loan_Amount) + Loan_Type + Loan_Purpose + Lien_Status + HOEPA_Status + Minority_Population + Owner_Occupied_Units_Number

The model has the lowest AIC compared to other regression models. Every variable are considered in this model step by step automatically through step() from both 'forward' and 'backward' directions.

This model includes 15 variables in 4 categories. It will be highly useful for commercial banks and governments since it contains every important aspects concerning whether the loans would get approved. Banks could utilize this model to revise loan decisions to reduce credit risk and avoid adverse selection. Government can supervise the loan market and guarantee the majority of loan approval procedure is standard, legal and undiscriminated. From the automatic backwards selection model, key findings include:

1. Demographic information: Males are less likely to get their loans approved. White applicants are most likely to get their loans approved, asian the next. No co-applicant's possibility of getting approval is less than co-applicants. The higher the income, the higher the possibility of getting loans approved.
2. Property information: Manufactured housing are less likely to be approved than one-to-four family dwelling
3. Loan application parametres: The higher the loan amount, the higher the possibility of getting loan approved, which may result from the positive correlation between income and loan amount. Conventional loan is most likely to be approved, while VA-guaranteed loan is least likely to be approved. Loan with the purpose of home purchasing is most likely to be approved, while loan with the purpose of refinancing is least likely to get approved
4. Geographic information: Minority population increases, possibility of getting approved decreases.The number of dwellings in the tract that are lived in by the owner increases, possibility of getting approved increases. There is a significant relationship between the possibility of getting approval and the County

6. Prediction & Empirical

First, predict() is utilized to predict the possibility of every applicant getting their loans approved. Accuracy rate equals expectation of getting loans approved over number of applications approved. The outcome is as follows:

	Customer-driven model	**Intuitive variable selection**	**Automatic backwards selection**
Accurate rate	0.8234759	0.837786	0.8383941

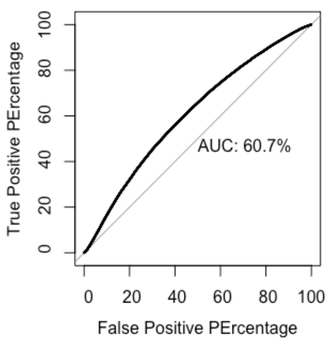
However, calculated accuracy rate only considers the accuracy extent when the loan is approved, which may well be biased (TD and FD are ignored).

	Approved (A)	**Denied (D)**
True Prediction (T)	TA	TD

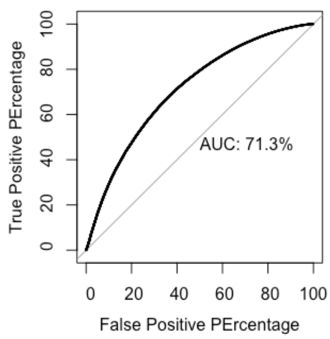
False Prediction (F)	FA	FD

Then, ROC and AUC are utilized to validate the logistic regression model though roc() function in R. ROC is receiver operating characteristics while AUC represents the area under the curve, which can be utilized as one of the most important evaluation metrics for checking any classification model's performance. Graphs presenting ROC and AUC are as below:

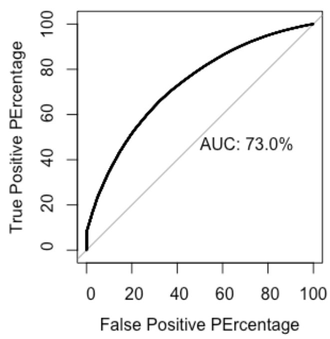
Customer-driven model



Intuitive variable selection



Automatic backwards selection



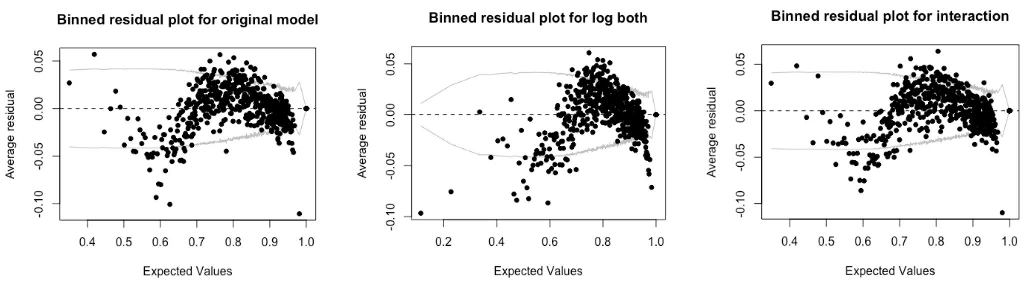
From the outcome of calculated accurate rate, ROC and AUC, the first customer-driven model is a simple model that could be utilized to predict customers' possibility of getting approved roughly with an AUC of 60.7% and calculated accurate rate of 82.35%. Intuitive model and automatic backwards selection model have a better capability of predicting with higher calculated accurate rate, ROC and AUC.

7. Risks and Other Considerations

Although the geographic graph in data visualization shows the loan approval rate in all counties displays significant density difference in different counties, the variable *"County_Name"* is not included in the intuitive selection model. Moreover, variable *"Minority_Population"* is related to specific tracts. This might seem inconsistent but since there is no specific tract name in the data set, it is not likely to map for tracts. Also, because a county consists of a number of tracts, the rate density of counties can be representative of the tract density.

In the prediction part, the accuracy rate equals the predicted number of loans being approved divided by the actual number of loans being approved. This only partly explains the model accuracy since the accuracy rate for loans not being approved is not accounted for. Therefore, we use the other analysis model which utilizes ROC graph and AUC percentage to represent the explanatory power of the model. This model takes into account accuracy rate for both approval and denial, which solves the problem in the accuracy rate calculation.

The residual distributions for numerical variables are examined. For practical reasons, only *"Loan_Amount"* and *"Applicant_Income"* are analysed. Specially, for logistic regression, instead of raw residuals, binned residual plot that divided the data into categories is utilized using "arm" package in R. The plots show the relationship between fitted/expected values and the average residual. Original and transformed model are fitted, the latter including logging income and loan amount, and adding interaction between income and loan amount. For all models, the relationship is displayed by a proximate cubic curve. Specifically, the grey lines surrounding the dots cover the data band that is within positive and negative two standard errors. Therefore, normally 95% of data is located within the area. Although data points are more concentrated under the model that logs both variables, it is not obvious which model is more consistent of the 95% distribution pattern, therefore not obvious which model is better. Moreover, similar to that for linear regression, because none of the graphs show a random distribution within the grey lines, all models still need further transformation. Again, for practical reasons, only income and loan amount variables have been transformed.



8. Summary

By analyzing data visualization, it is observed that demographic, geographic, property-related and loan-related information are all relevant to whether a loan could be approved or not. Based on the visualization, number of variables have been narrowed down from 47 to 19.

Logistic regression is applied in all three models, which takes the approval code (approved as 1 or not approved as 0) as dependent variable. Each model is built from the perspective of different users groups, namely they are customer-driven model, intuitive selection model and AIC stepwise selection model. The first model allows customers to utilize the information that is most readily accessible to them, plugging in their own circumstances and estimate their chances of being approved. However, customers should be aware that they should not only rely on the model because it has a relatively low McFadden R^2 and high AIC, indicating that the explanatory power is limited.

The second model is based on scientific intuition that derives from previous visualization and correlation analysis. Variables that are intuitively relevant to whether loans could be approved or not are included. Specifically, this model is based on tract information instead of county information due to drawbacks mentioned in Risks and Other Considerations and the fact that tract information is more significant.

The third model utilizes the stepwise selection function in R. The function automatically selects the model with the lowest AIC. This model has a higher McFadden R^2 than the customer-driven model, but the information required is difficult to collect for loan borrowers. Therefore, this model is more practical for loan issuers, such as commercial banks and loan origination institutions.

9. References

Jeff Webb, 2017, *Statistics and Predictive Analytics*. Available from: https://bookdown.org/jefftemplewebb/IS-6489/. [22 August 2019]

Jonathan Bartlett, 2014, *R squared in logistic regression*. Available from: https://thestatsgeek.com/2014/02/08/r-squared-in-logistic-regression/. [22 August 2019]

10. Appendix
Appendix 1 - Regression Model Output

Customer-driven Model

Variable	Coefficient	P value	Sig	Variable	Coefficient	P value	Sig
(Intercept)	-2.935793	< 2e-16	***	County_Name(Asotin County)	0.476361	0.000460	***
Applicant_Gender(Male)	-0.022018	0.040183	*	County_Name(Benton County)	0.600595	1.30e-08	***
Applicant_Race(Asian)	0.314568	4.43e-13	***	County_Name(Chelan County)	0.314235	0.004751	**
Applicant_Race(Black or African American)	-0.026779	0.579690		County_Name(Clallam County)	0.543821	1.51e-06	***
Applicant_Race(Native Hawaiian or Other Pacific Islander)	0.130566	0.033396	*	County_Name(Clark County)	0.585753	1.27e-08	***
Applicant_Race(White)	0.527156	< 2e-16	***	County_Name(Columbia County)	0.345429	0.206534	
log(Applicant_Income)	1.420558	< 2e-16	***	County_Name(Cowlitz County)	0.505925	2.88e-06	***
Property_Type(One-to-four family dwelling (other than manufactured housing))	0.084326	0.000562	***	County_Name(Douglas County)	0.487814	5.76e-05	***
log(Loan_Amount)	0.322141	< 2e-16		County_Name(Kitsap County)	0.422188	5.16e-05	***
County_Name(Ferry County)	-0.111774	0.575894		County_Name(Kittitas County)	0.438499	0.000177	***
County_Name(Franklin County)	0.444480	5.71e-05	***	County_Name(Klickitat County)	0.096962	0.477701	
County_Name(Garfield County)	-0.002325	0.993661		County_Name(Lewis County)	0.216600	0.051019	.
County_Name(Grant County)	0.334458	0.002943	**	County_Name(Lincoln County)	0.040749	0.789411	

County_Name(Grays Harbor County)	0.333906	0.003176	**	County_Name(Mason County)	0.221096	0.047620	*
County_Name(Island County)	0.522630	1.50e-06	***	County_Name(Okanogan County)	0.190285	0.130365	
County_Name(Jefferson County)	0.268490	0.027150	*	County_Name(Pacific County)	0.342051	0.009155	**
County_Name(King County)	0.461187	6.37e-06	***	County_Name(Spokane County)	0.575098	2.46e-08	***
County_Name(Pend Oreille County)	0.184409	0.194315		County_Name(Stevens County)	0.256282	0.031643	*
County_Name(Pierce County)	0.423120	3.60e-05	***	County_Name(Thurston County)	0.552444	1.26e-07	***
County_Name(San Juan County)	-0.037576	0.782584		County_Name(Wahkiakum County)	0.234257	0.239015	
County_Name(Skagit County)	0.464709	1.45e-05	***	County_Name(Walla Walla County)	0.527514	6.19e-06	***
County_Name(Skamania County)	0.306356	0.040679	*	County_Name(Whatcom County)	0.545437	2.20e-07	***
County_Name(Snohomish County)	0.540641	1.35e-07	***	County_Name(Yakima County)	0.087694	0.403606	
County_Name(Whitman County)	0.348456	0.00580	**				

Intuitive Model

Variable	**Coefficient**	**P value**	**Sig**	**Variable**	**Coefficient**	**P value**	**Sig**
(Intercept)	-2.2091810	< 2e-16	***	Applicant_RaceNot applicable	0.2112703	4.36e-06	***
log(Applicant_Income)	1.3285102	< 2e-16	***	Applicant_RaceWhite	0.5231667	< 2e-16	***
log(Loan_Amount)	0.2292458	< 2e-16	***	Loan_TypeFHA-insured	-0.4090203	< 2e-16	***
Minority_Population	-0.0024258	4.37e-15	***	Loan_TypeFSA/RHS-guaranteed	-0.3035178	7.29e-08	***
Applicant_GenderMale	-0.2109539	< 2e-16	***	Loan_TypeVA-guaranteed	-0.0968592	1.60e-08	***
Applicant_GenderNot applicable	0.0358015	0.2178		Loan_PurposeHome purchase	1.2222453	< 2e-16	***
Applicant_RaceAsian	0.2976211	3.79e-11	***	Loan_PurposeRefinancing	-0.1291009	3.43e-11	***
Applicant_RaceBlack or African American	-0.0157639	0.7533		Property_TypeOne-to-four family dwelling (other than manufactured housing)	0.3618292	< 2e-16	***
Applicant_RaceNative Hawaiian or Other Pacific Islander	0.1304783	0.0401	*	Co_Applicant_GenderMale	-0.3986484	< 2e-16	***

Co_Applicant_GenderNo co-applicant	-0.3896109	< 2e-16	***	Co_Applicant_GenderNot applicable	0.0583120	0.0434	*

AIC Stepwise Selection Model

Variable	Coefficient	P value	Sig	Variable	Coefficient	P value	Sig
(Intercept)	3.097e+01	0.978316		County_Name(Asotin County)	2.859e-01	0.047876	*
Applicant_Gender(Male)	-1.909e-01	< 2e-16	***	County_Name(Benton County)	4.052e-01	0.000307	***
Applicant_Gender(N/A)	-1.935e-03	0.950838		County_Name(Chelan County)	1.825e-01	0.122015	
Applicant_Race(Asian)	2.947e-01	4.35e-10	***	County_Name(Clallam County)	3.431e-01	0.004430	**
Applicant_Race(Balck or African American)	-1.557e-02	0.767853		County_Name(Clark County)	5.473e-01	6.22e-07	***
Applicant_Race(N/A)	1.326e-01	0.006754	**	County_Name(Columbia County)	-1.654e-01	0.564841	
Applicant_Race(Native Hawaiian or Other Pacific Islander)	1.390e-01	0.038420	*	County_Name(Cowlitz County)	2.856e-01	0.013316	*
Applicant_Race(White)	4.651e-01	< 2e-16	***	County_Name(Douglas County)	4.282e-01	0.000843	***
Co_Applicant_Gender(Male)	-3.844e-01	< 2e-16	***	County_Name(Ferry County)	-2.077e-01	0.329526	
Co_Applicant_Gender(No co-applicant)	-2.126e-01	0.003905	**	County_Name(Franklin County)	3.122e-01	0.007390	**
Co_Applicant_Gender(Not applicable)	-2.287e-01	6.72e-08	***	County_Name(Garfield County)	-2.968e-01	0.333319	
Co_Applicant_Race(Asian)	-7.269e-02	0.350964		County_Name(Grant County)	1.356e-01	0.254341	
Co_Applicant_Race(Black or African American)	-1.692e-01	0.062538	.	County_Name(Grays Harbor County)	-3.279e-02	0.786103	
Co_Applicant_Race(Native Hawaiian or Other Pacific Islander)	-8.025e-02	0.425886		County_Name(Island County)	2.964e-01	0.010639	*
Co_Applicant_Race(Not applicable)	2.345e-01	0.003486	**	County_Name(Jefferson County)	7.704e-02	0.552516	
Co_Applicant_Race(White)	2.306e-01	0.001709	**	County_Name(King County)	6.366e-01	4.22e-09	***
log(Applicant_Income)	1.454e+00	< 2e-16	***	County_Name(Kitsap County)	3.362e-01	0.002479	**
Property_Type(One-to-four family dwelling (other than manufactured housing))	4.094e-01	< 2e-16	***	County_Name(Kittitas County)	2.395e-01	0.054400	.
log(Loan_Amount)	-4.156e-01	< 2e-16	***	County_Name(Klickitat County)	-1.388e-01	0.340391	
Loan_Type(FHA-insurd)	-4.023e-01	< 2e-16	***	County_Name(Lewis County)	-2.930e-02	0.804705	

Loan_Type(FSA/RHS-guaranteed)	-3.263e-01	1.63e-08	***	County_Name(Lincoln County)	-3.017e-01	0.064279	.
Loan_Type(VA-guaranteed)	-2.656e-02	0.135524		County_Name(Mason County)	-1.144e-02	0.923477	
Loan_Purpose(Home purchase)	1.014e+00	< 2e-16	***	County_Name(Okanogan County)	7.593e-02	0.569104	
Loan_Purpose(Refinancing)	-3.757e-01	< 2e-16	***	County_Name(Pacific County)	-8.013e-04	0.995444	
Lien_Status(Not secured by a lien)	-1.734e+01	0.485989		County_Name(Pend Oreille County)	-1.638e-01	0.280790	
Lien_Status(Secured by a first lien)	-1.595e+01	0.521635		County_Name(Pierce County)	3.546e-01	0.001118	**
Lien_Status(Secured by a subordinate lien)	-1.661e+01	0.504425		County_Name(San Juan County)	-9.428e-02	0.514389	
HOEPA_Status(Not a HOEPA loan)	-1.644e+01	0.988486		County_Name(Skagit County)	2.857e-01	0.012206	*
Minority_Population	-4.250e-03	< 2e-16	***	County_Name(Skamania County)	2.063e-01	0.194269	
Owner_Occupied_Units_Number	5.097e-05	7.05e-07	***	County_Name(Snohomish County)	5.935e-01	5.24e-08	***
County_Name(Wahkiakum County)	-9.386e-02	0.659082		County_Name(Spokane County)	2.732e-01	0.013340	*
County_Name(Walla Walla County)	2.763e-01	0.025608	*	County_Name(Stevens County)	5.935e-03	0.962911	
County_Name(Whatcom County)	3.911e-01	0.000492	***	County_Name(Thurston County)	3.923e-01	0.000425	***
County_Name(Whitman County)	5.086e-02	0.704928					

Appendix 2 - Individual Contributions

We changed the main research question three times, as a result, detailed individual effort was not accurately recorded. Given the short turnaround time that we had for the final analysis, we worked together as a team and demonstrated dedication in delivering the piece of research under extreme pressure. In turn, each one of us has contributed equally in the final report and presentation. If a detailed individual contributions list is still required, we are happy to submit one for the sole purpose of evaluation. Thanks Natesh!