Credit Risk Analytics

Analysis of risk for Residential Mortgage and Equity

Group 2: Aditee Bhattarai,
Nam Dang, Zihan Huang, Jingxin Yao

AGENDA

INDUSTRY REVIEW

BUSINESS PROBLEM

DATA-DRIVEN ANALYSIS

Industry Review

Housing Bubble 2008

Source: https://medium.com/@maggiepolk/the-2008-mortgage-crisis-afdec4a92923

The 2008 Housing Bubble Burst (Magdoff & Yates, 2010)**:**

At 2006-2007:

- Unemployment rate at 4.4%
- Wages rising by 4.2%
- Dow Jones index hit all-time high
- Real estate market became core of the economy

At 2008 - 2009:

- Unemployment rate at 9.5% (actual rate could be at 16.5%)
- Housing prices fell by 9.5%
- Dow Jones index witnessed largest drop in intraday trading

Basel Capital Accord

BASEL Standard (Edward et al., 2010): Perceive failings of deregulation & prevent the credit losses

- **Shift the focus to data mining in credit risk management**

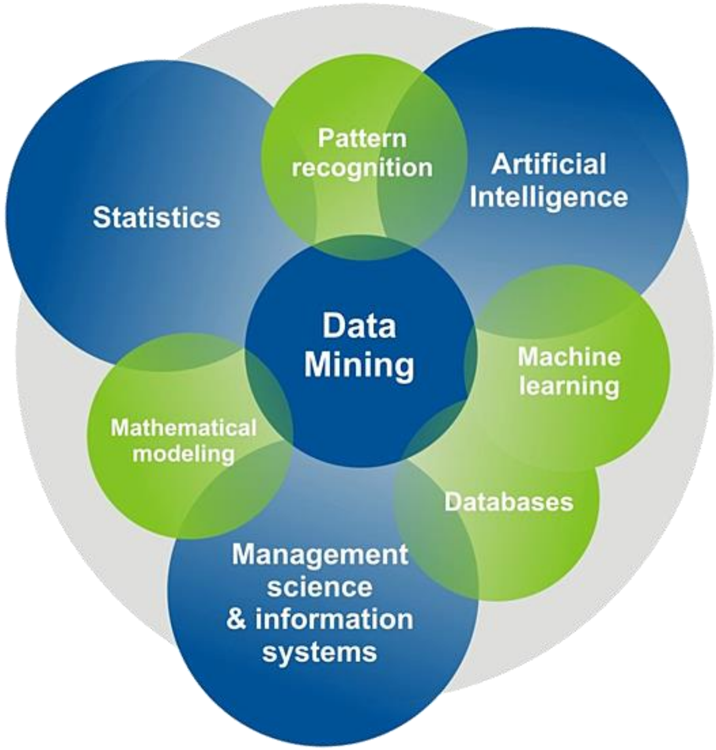
3 Key Estimation Parameters for Loan Portfolios

- PD: Probability of default in the next 12 months
- LGD: Loss given default
- EAD: Exposure at default

Main approaches to model LGD

- Ordinary Least Squares (OLS) Regression
- Two-Stage Approach
- Tobit Regression
- Censored Gamma Regression
- Zero-inflated Gamma Model

Data mining Techniques



Unsupervised Models

- Extract patterns that represent and describe distinct features of the data.

Supervised Models

- Use input variables to classify data or predict values for output variables.

Algorisms used in loan default prediction

- Classification and Regression Tree (Feldman et al., 2005)
- Logistic Regression (Butaru, Chen et al., 2006)
- Random Forest (Butaru, Chen et al., 2006)
- Neural Network (Atiya, 2001; Bahrammirzaee, 2010)
- Support Vector Machines (Pérez-Martín & Vaca， 2018）

Business Problem

Business Problem

- Analyze two loan datasets using analytics and modelling techniques

Mortgage Dataset

Home Equity Dataset

Analytics

Modelling

Goal

- Understand loan customers
- Predict customer creditworthiness

Data-Driven Analysis

MORTGAGE DATASET

Data Description

Original Dataset

- 622,489 Rows
- 21 Variables

Sample Dataset

- 8,146 Rows
- 27 Variables

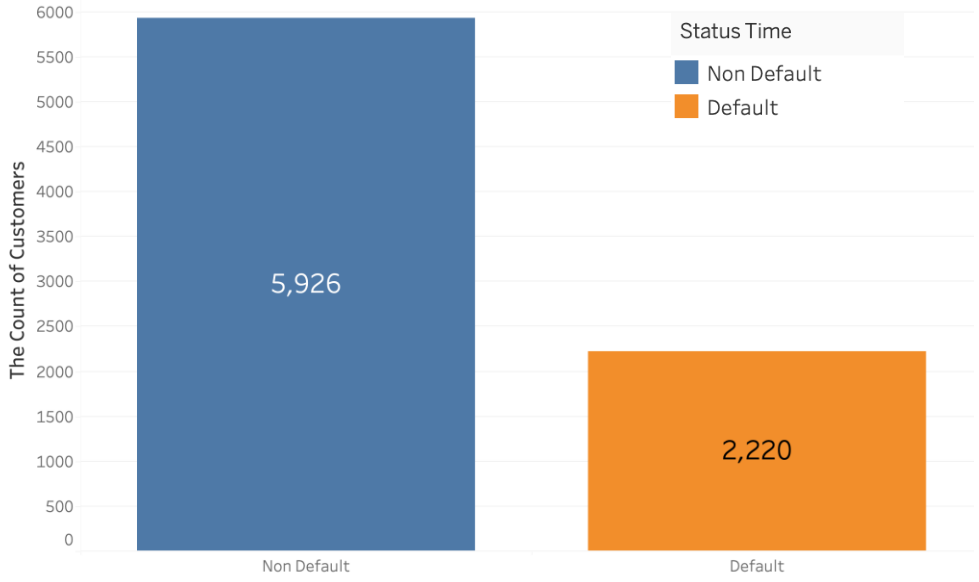
Data Aggregation

Data Description

- 5 nominal variables: real estate type, customer identity, mortgage status.
- 21 numerical variables: Time stamps, and values of mean and standard deviation for Loan-to-value ratio(LTV), GDP, unemployment rate, etc.
- Target variable- Mortgage Loan Status

  - 0: Non-Default

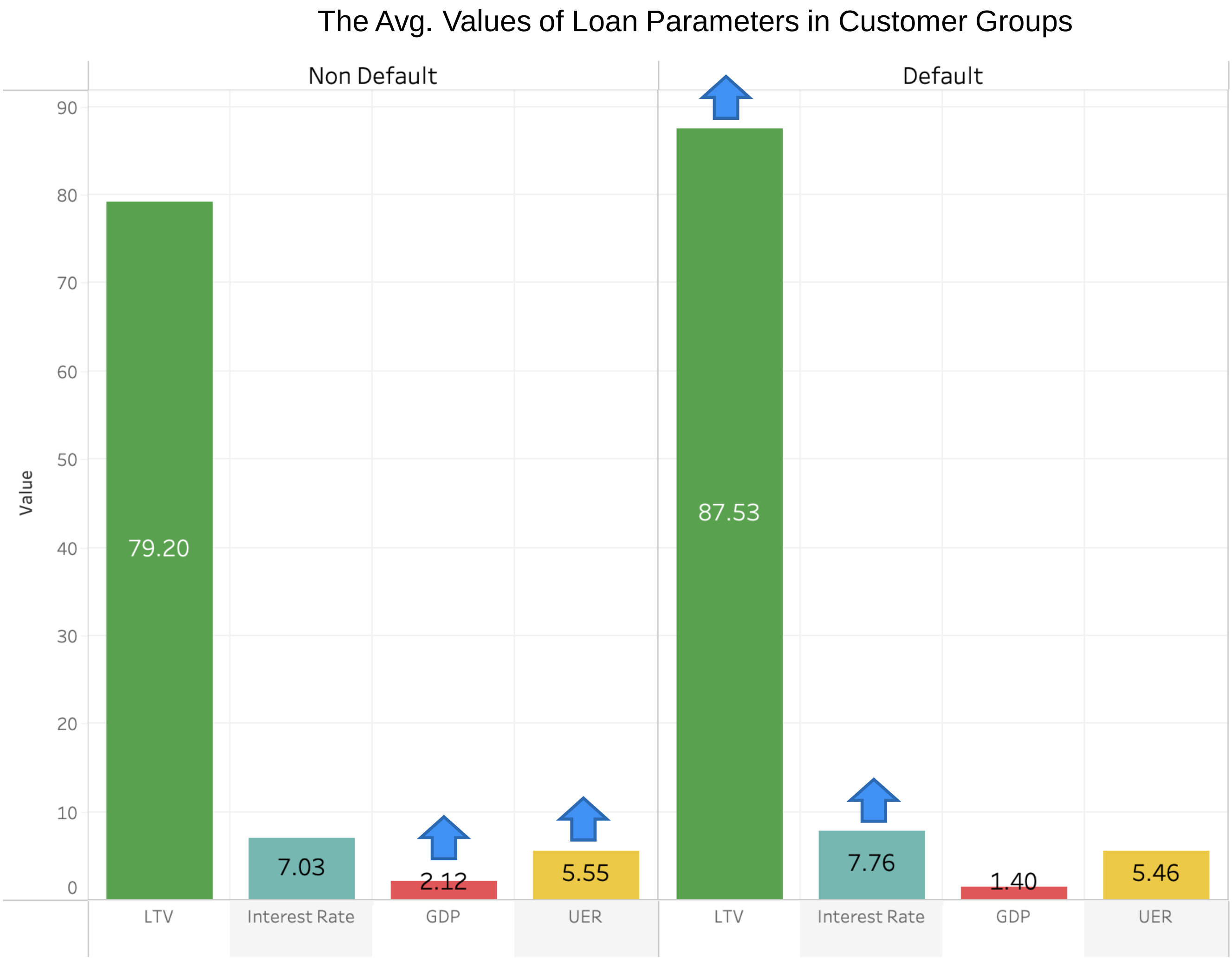
  - 1: Default



The Scale of Mortgage Loan Status

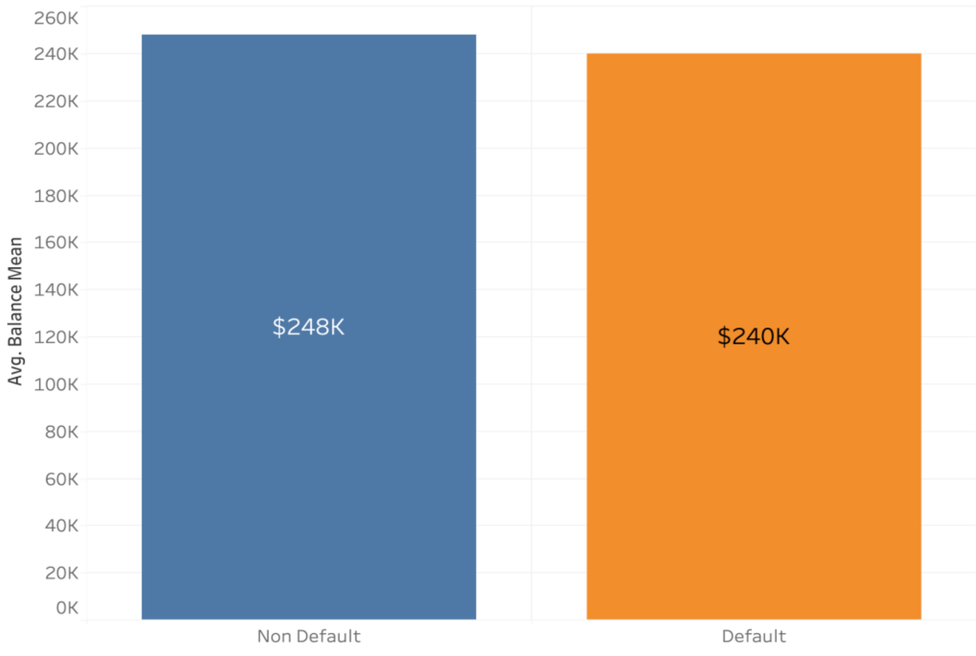
Data Exploration

- The bar chart shows the difference in average values between customer groups across all measurable parameters for the mortgage loan.
- Default customers have higher average values in both LTV and interest rate during the observation duration.

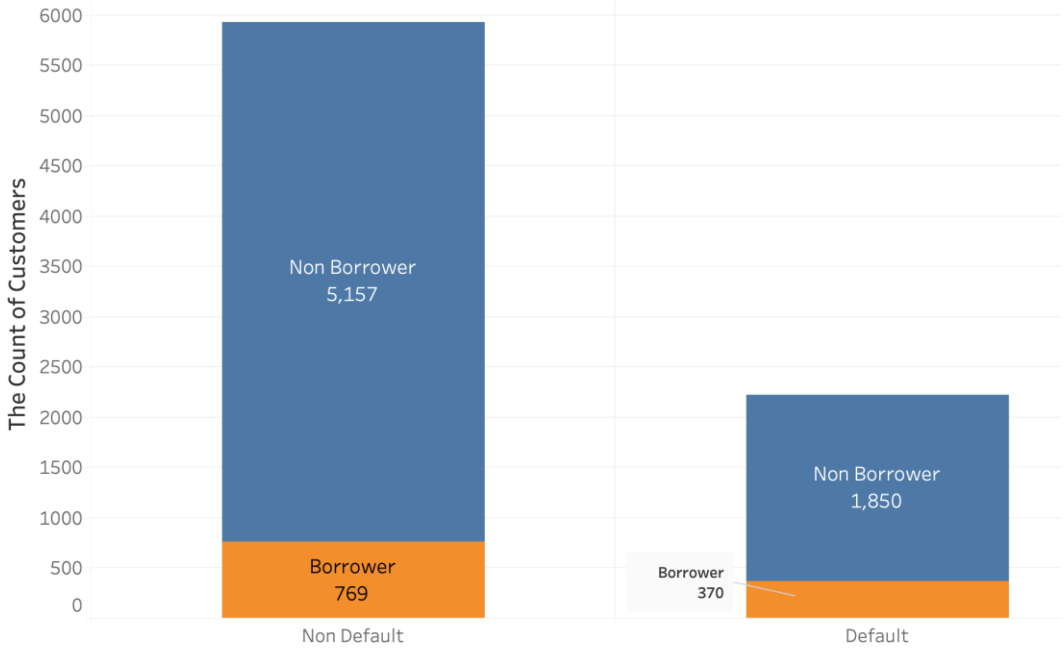


Data Exploration

The Average Outstanding Balance Amount for Customers

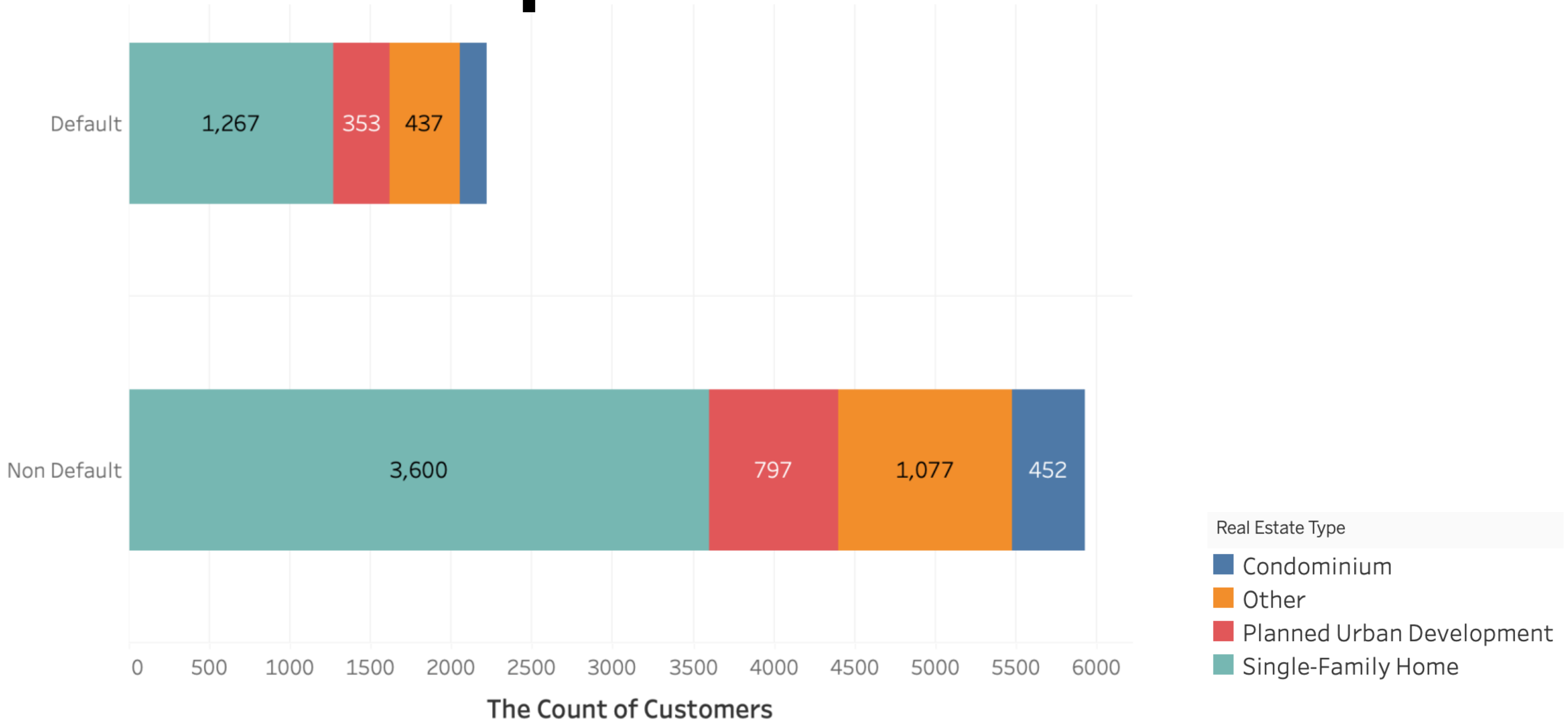


The Count of Investors among Customers



- Default customers have less outstanding balance amount on average.
- Most customers are not borrowers for each group.

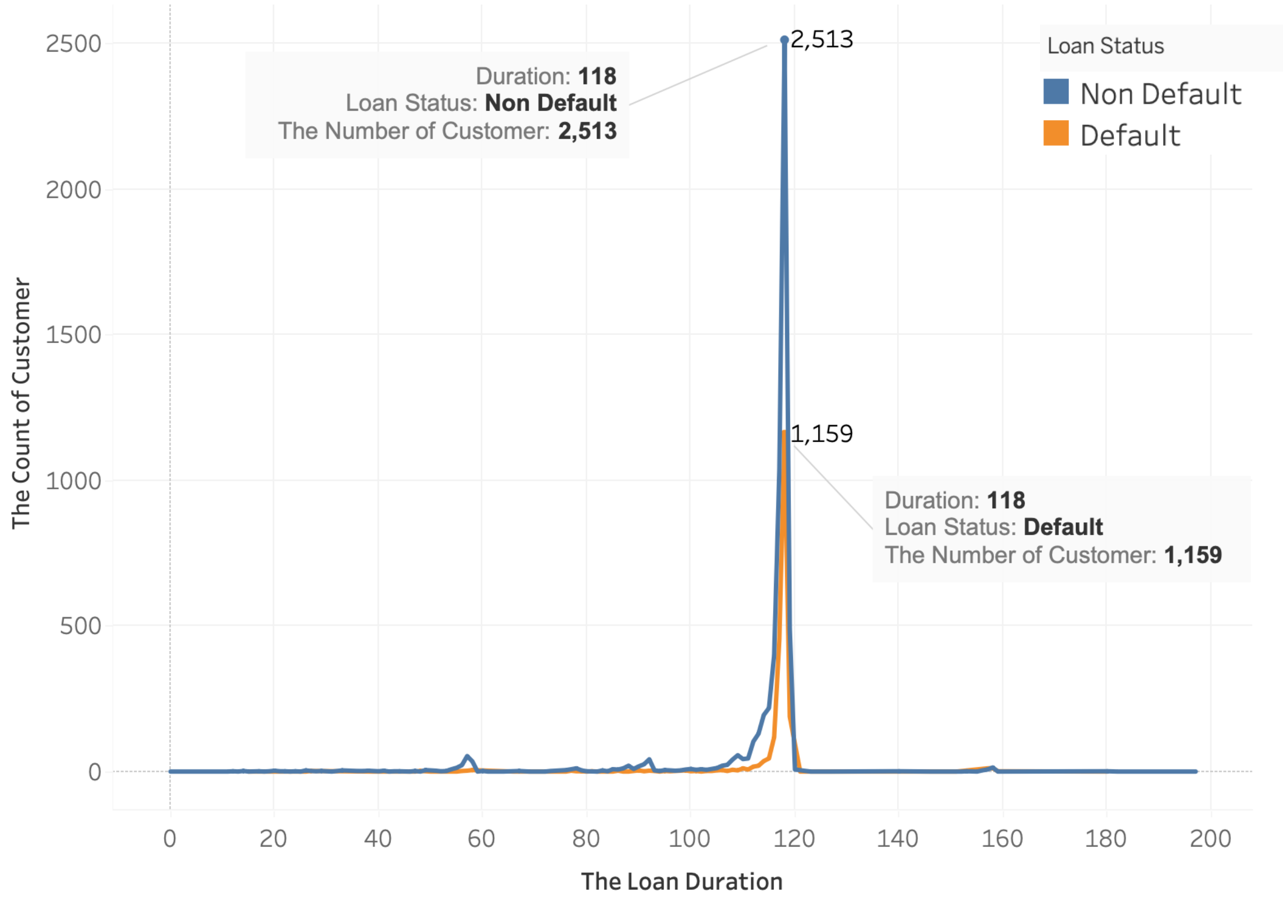
Data Exploration



The Mortgage Loan Status Distribution by Real Estate Type

- The distributions of all types of real estate have the same scales in two customer groups.
- Single-family home (SF) is the majority among all customers.

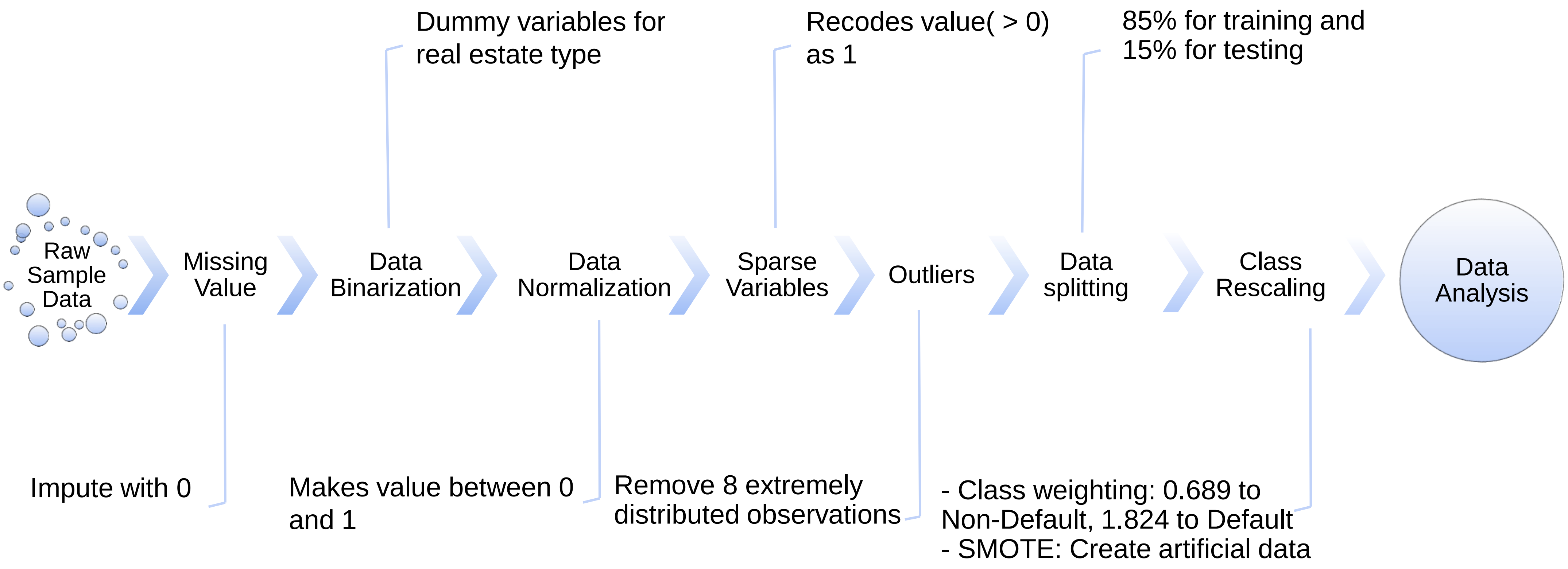
Data Exploration



The Line Chart of the Relation between Loan Duration and Status

- Observe in the loan duration: Maturity Time - Origin Time
- Loan Actions are typically manipulated at the loan duration of 110-120.

Data Preparation



Model Performance: Overall

Model	Training		Testing	
Parameters (Instruction)	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Decision Tree	0.8474	0.6552	0.7731	0.4947
SVM	**0.8406**	**0.5767**	**0.8264**	**0.5390**
kNN	0.8152	0.4906	0.7929	0.4189
Deep Learning	0.8388	0.6776	0.8133	0.6267

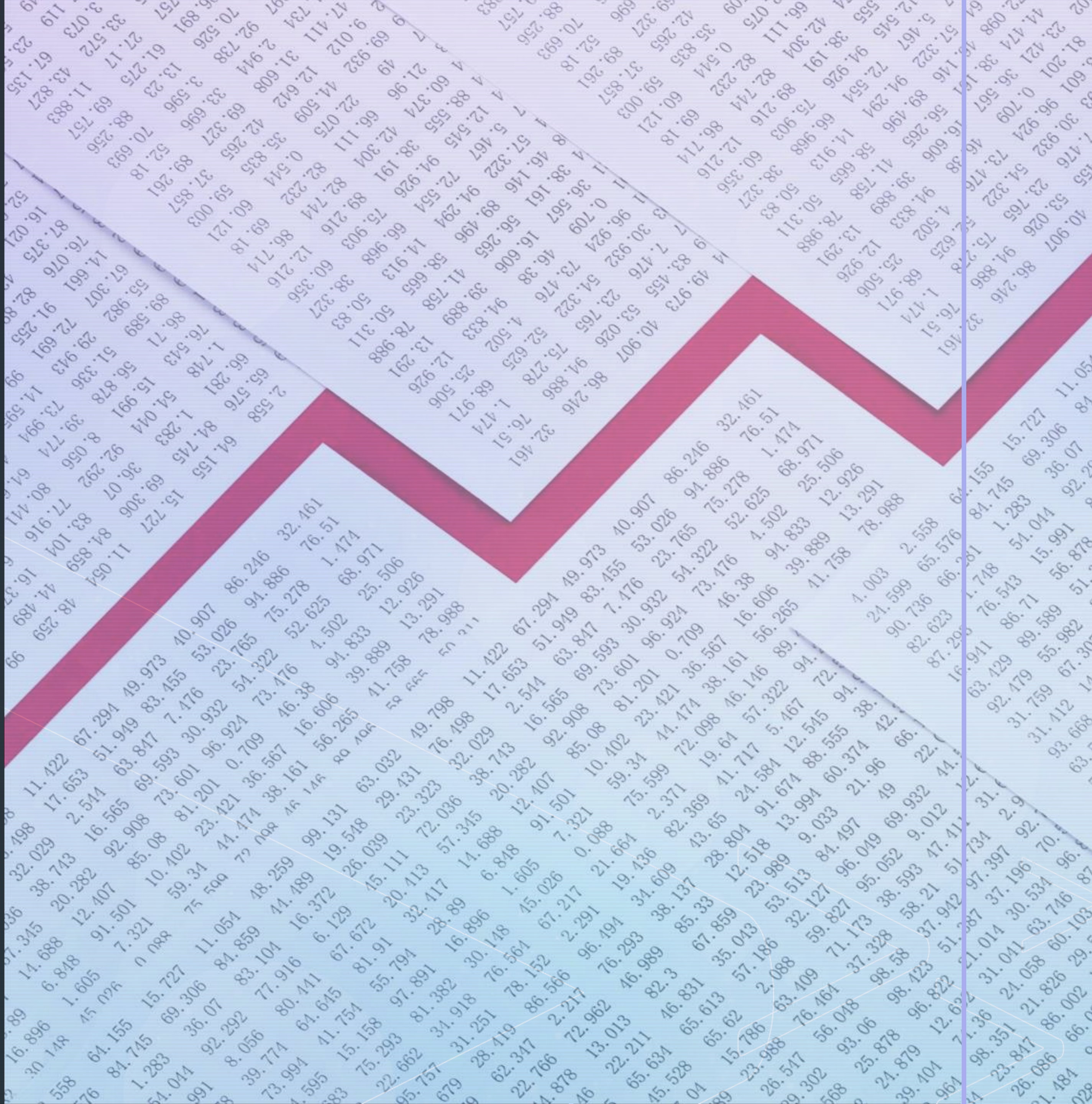
- SVM performed better in model accuracy.

Model Performance: Class Level

Model	Training					Testing				
Parameters (Instruction)	**Sensitivity**	**Specificity**	**Precision**	**Recall**	**F1 Value**	**Sensitivity**	**Specificity**	**Precision**	**Recall**	**F1 Value**
Decision Tree	0.9020	0.8267	0.6627	0.9020	0.7641	0.7934	0.7655	0.6627	0.7934	0.6567
SVM	0.6233	0.9226	0.7527	0.6233	0.6818	0.5988	0.9121	0.7194	0.5988	0.6536
kNN	0.5252	0.9238	0.7207	0.5252	0.6076	0.4625	0.9155	0.6725	0.4625	0.5480
Deep Learning	**0.8078**	**0.8698**	**0.8191**	**0.8698**	**0.8437**	**0.7770**	**0.8497**	**0.7916**	**0.8497**	**0.8196**

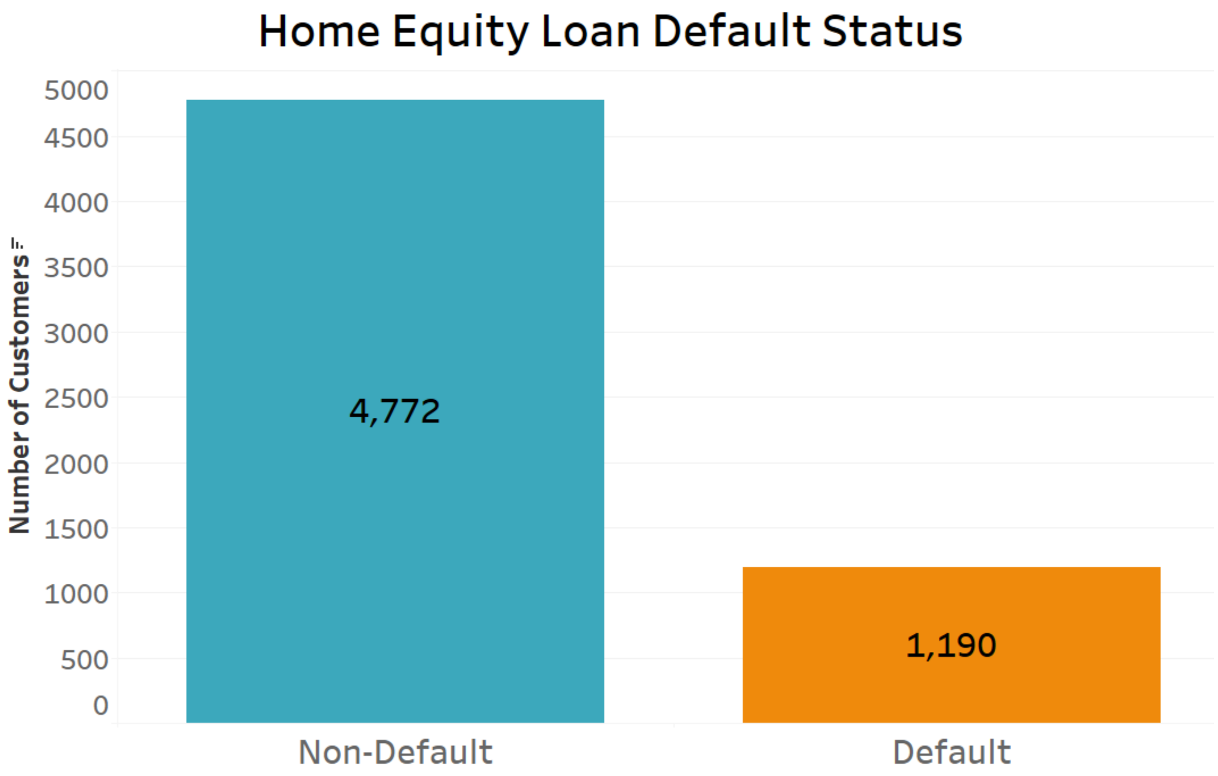
- Deep Learning performed better in correctly predicting mortgage default.

HOME EQUITY DATASET

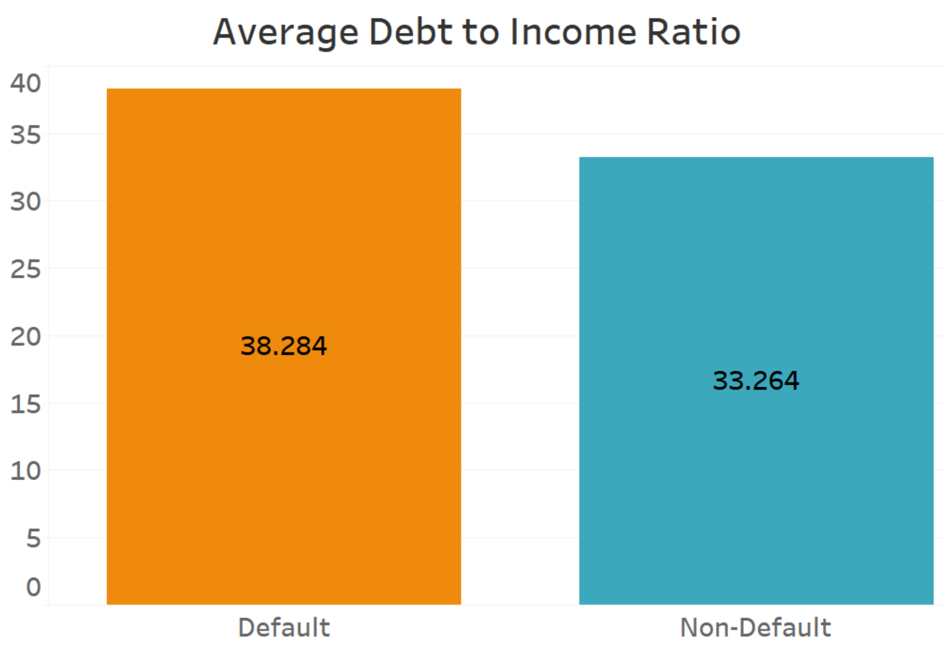


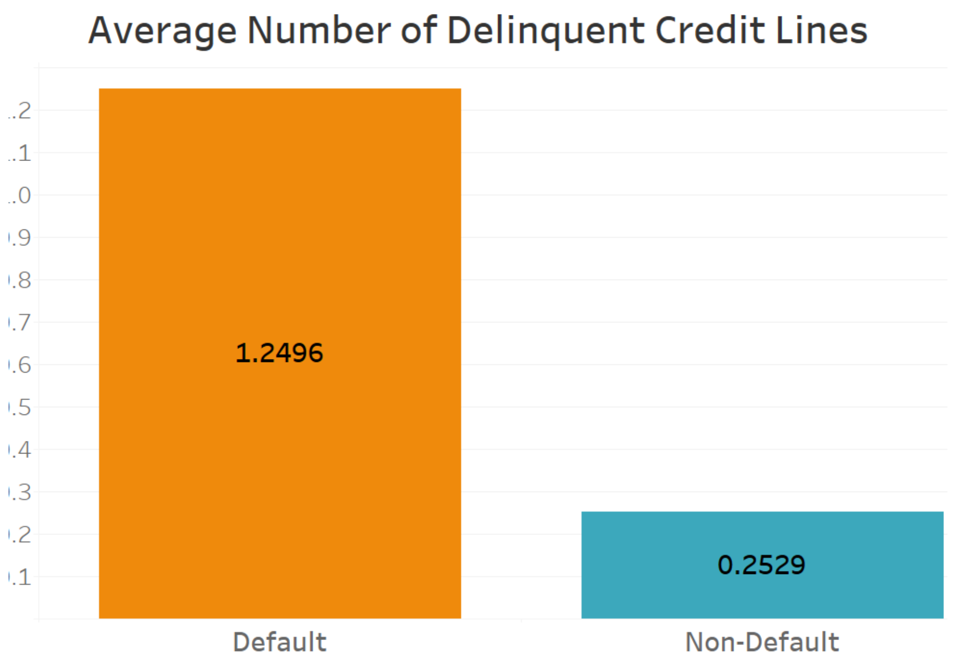
Data Description

- Dataset has 5964 rows and 14 columns.
- 2 categorical variables:
  Reason for Home Equity and Occupation
- 10 numerical variables:
  Including Loan Amount,
  Mortgage Balance, Home Value,
  Credit Line Age,
  Number of Delinquent Credit Line
- Target variable: Default
  0: No default
  1 : Default
- Class Imbalance exists in Target Variable



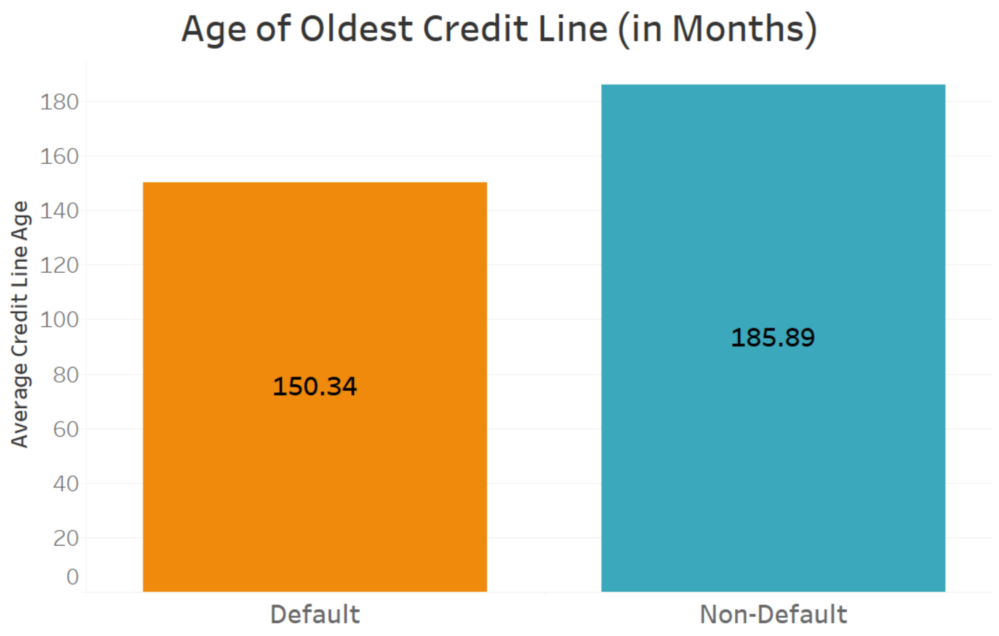
Data Exploration

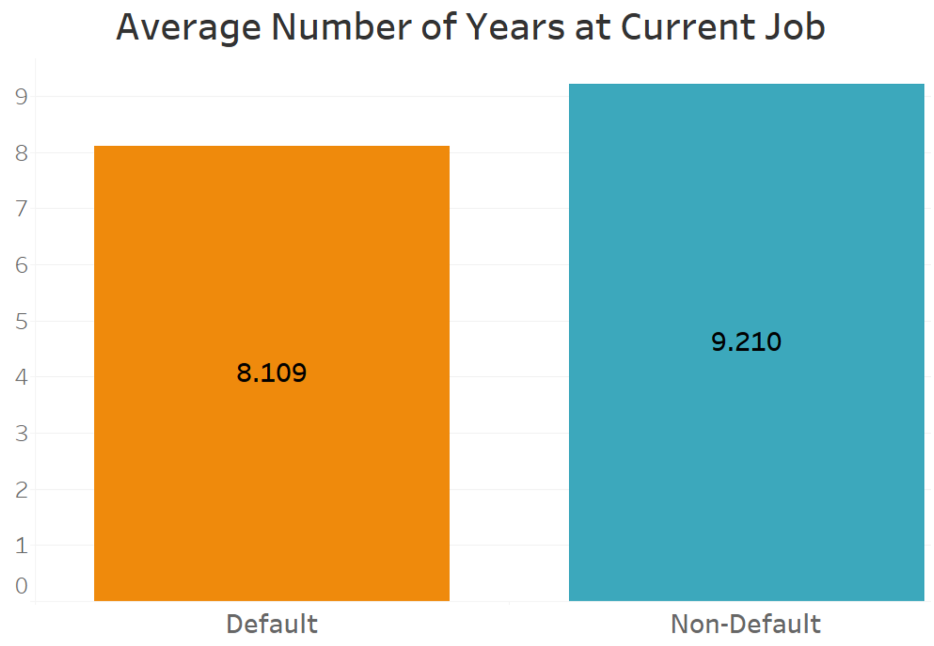




- Debt to income ratio is higher in defaulters.
- People with a higher number of delinquent credit lines have a higher rate of default.
- People with higher-than-average number of derogatory reports also have higher rate of default.

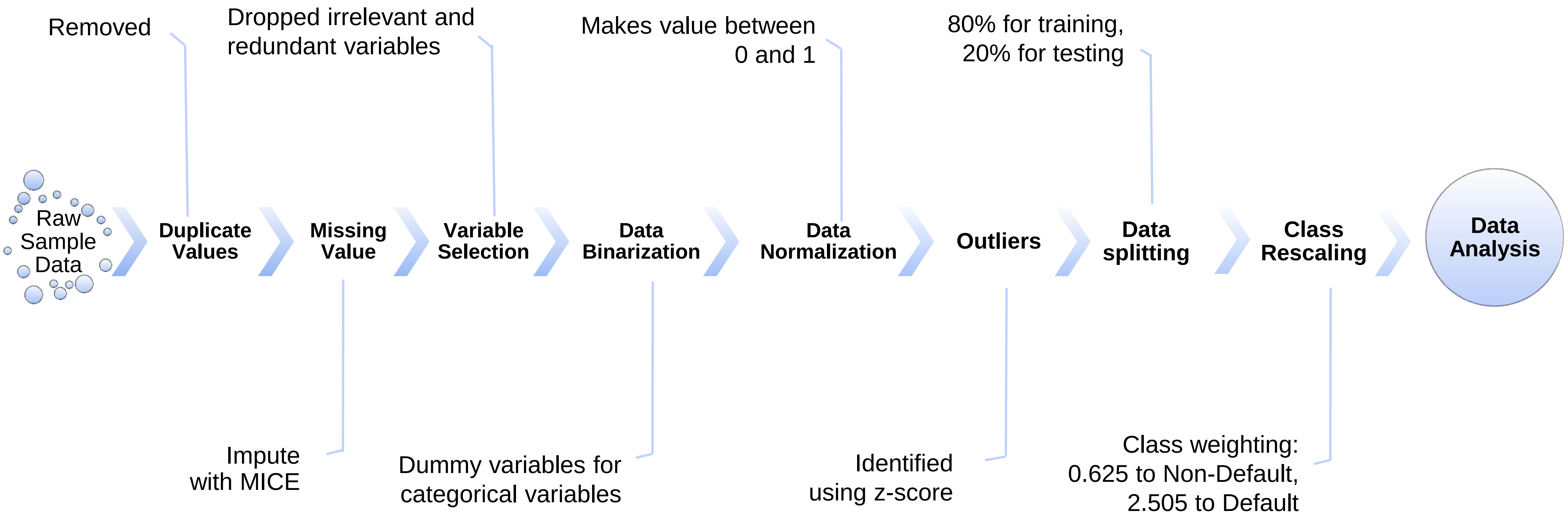
Data Exploration





- Credit Line Age is lower among people who default.
- Years on current job is lower among people who default.

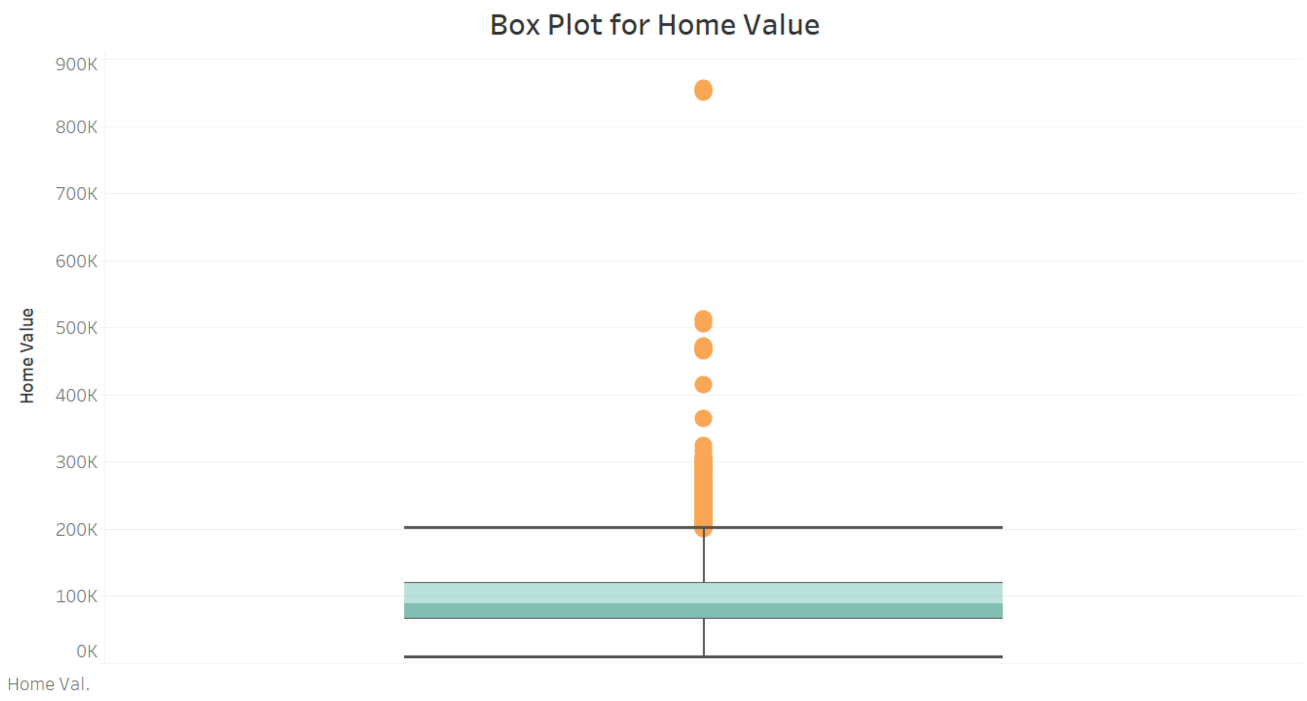
Data Preparation



Data Preparation

Outliers

- Most numerical variables also have outliers.
- Do these outliers make sense?
- Should consider models that are robust to outliers for analysis.



Data Preparation

- **Duplicate values**
  There are 2 duplicate entries within the dataset.
- **Missing values**
  There are 2620 entries with at least 1 missing value. That is 43% of the entire dataset.
- **Reasons:**
  Censored Data?
  Sensitive information?
- No non-random pattern in the Missing Values.
- Test for missing value at random: T-test to compare the means of cases with missing data versus not missing data on Loan Amount Variable.
- **Result of the test: MAR –** best approach is to use model-based imputation (Hair et al., 2019).

Model Performance: Overall

Model	Training (Tuned)		Testing (Tuned)	
Parameters (Instruction)	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
Decision Tree	0.783	0.566	0.793	0.416
SVM	**0.93**	**0.79**	**0.89**	**0.67**
KNN	0.829	0.224	0.865	0.469
Naïve Bayes	0.716	0.432	0.825	0.445

- SVM performs better in predicting Home Equity Loan Default.

Model Performance: Class Level

Model	Training					Testing				
Parameters	**Sensitivity**	**Specificity**	**Precision**	**Recall**	**F1**	**Sensitivity**	**Specificity**	**Precision**	**Recall**	**F1**
Decision Tree	0.72	0.845	0.823	0.720	0.768	0.623	0.836	0.486	0.623	0.546
SVM	**0.99**	**0.91**	**0.73**	**0.99**	**0.84**	**0.81**	**0.91**	**0.68**	**0.81**	**0.74**
KNN	0.157	0.997	0.933	0.157	0.269	0.387	0.984	0.862	0.387	0.534
Naïve Bayes	0.522	0.910	0.853	0.522	0.648	0.544	0.895	0.563	0.544	0.554

- SVM performs better in predicting Home Equity Loan Default.

References

Pérez-Martín*, A. Pérez-Torregrosa, M. Vaca. (2018), Big Data techniques to measure credit banking risk in home equity loans. Journal of business research, 2018, Vol.89, p.448-454 Retrieved from Big Data techniques to measure credit banking risk in home equity loans - ScienceDirect (drexel.edu)

Amadeo, K. (2022). When and why did the stock market crash in 2008? The Balance. Retrieved July 8, 2022, from https://www.thebalance.com/stock-market-crash-of-2008-3305535#:~:text=The%20stock%20market%20crash%20of%202008%20occurred%20on%20September%2029,largest%20point%20drop%20in%20history.

Chad Cowden, Frank J. Fabozzi, and Abdolreza Nazemi. (2019), Default Prediction of Commercial Real Estate Properties Using Machine Learning Techniques. Journal of portfolio management, 2019, Vol.45 (7), p.55-67. Retrieved from Default Prediction of Commercial Real Estate Properties Using Machine Learning Techniques - ProQuest

Christie, L. (2009, February 12). Home prices in Record Plunge. CNNMoney. Retrieved July 8, 2022, from https://money.cnn.com/2009/02/12/real_estate/Latest_median_prices/#:~:text=The%20median%20price%20for%20a,1.6%25%20between%202006%20and%202007

Edward N.C. Tong*, Christophe Mues & Lyn Thomas (2013). A zero-adjusted gamma model for mortgage loan loss given default, *International Journal of Forecasting*, 29 (4), 548-562. Retrieved from https://doi.org/10.1016/j.ijforecast.2013.03.003

Emad Azhar Ali, Syed; Sajjad Hussain Rizvi, Syed; Lai, Fong-Woon; Faizan Ali, Rao; Ali Jan, Ahmad (2021), Predicting Delinquency on Mortgage Loans: An Exhaustive Parametric Comparison of Machine Learning Techniques. International Journal of Industrial Engineering and Management, 2021, Vol.12 (Issue 1), p.1-13. Retrieved from Predicting Delinquency on Mortgage Loans: An Exhaustive Parametric Comparison of Machine Learning Techniques - ProQuest (drexel.edu)

Feibelman, A. (2022). BANKRUPTCY AND THE STATE. *Emory Bankruptcy Developments Journal, 38*(1), 1-50. Retrieved from http://ezproxy2.library.drexel.edu/login?url=https://www.proquest.com/scholarly-journals/bankruptcy-state/docview/2637689114/se-2?accountid=10559

Goolsbee, Austan D., and Alan B. Krueger. 2015. "A Retrospective Look at Rescuing and Restructuring General Motors and Chrysler." *Journal of Economic Perspectives*, 29 (2): 3-24.DOI: 10.1257/jep.29.2.3

Jackson, J. R., & Kaserman, D. L. (1980). Default risk on home mortgage loans: A test of competing hypotheses: ABSTRACT. *Journal of Risk and Insurance (Pre-1986), 47*(4), 678. Retrieved from http://ezproxy2.library.drexel.edu/login?url=https://www.proquest.com/scholarly-journals/default-risk-on-home-mortgage-loans-test/docview/235110404/se-2?accountid=10559

Kim, Dong-sup; Shin, Seungwoo, THE ECONOMIC EXPLAINABILITY OF MACHINE LEARNING AND STANDARD ECONOMETRIC MODELS-AN APPLICATION TO THE U.S. MORTGAGE DEFAULT RISK. International journal of strategic property management, 2021, Vol.25 (5), p.396-412. Retrieved from THE ECONOMIC EXPLAINABILITY OF MACHINE LEARNING AND STANDARD ECONOMETRIC MODELS-AN APPLICATION TO THE U.S. MORTGAGE DEFAULT RISK. - Document - Gale Academic OneFile (drexel.edu)

Magdoff, F., & Yates, M. D. (2010). ABCs of the economic crisis: What Working People Need To Know. Aakar Books. Retrieved from https://books.google.com/books?hl=en&lr=&id=LzIVCgAAQBAJ&oi=fnd&pg=PA9&dq=what+happens+after+economic+crisis&ots=Pd0VyZNGJW&sig=hnlfe1t32HVBmMj1hIpiN7UXPVY#v=onepage&q=what%20happens%20after%20economic%20crisis&f=false.

Mitrašević, M., & Bardarova, S. (2020). CAUSES OF NON-PAYMENT OF MORTGAGE LOANS: THEORETICAL AND PRACTICAL ASPECTS. *UTMS Journal of Economics, 11*(2), 138-150. Retrieved from http://ezproxy2.library.drexel.edu/login?url=https://www.proquest.com/scholarly-journals/causes-non-payment-mortgage-loans-theoretical/docview/2571154652/se-2

Song, G. (2022). Large US bank takeovers in 2008: Performance and implications. *Journal of Capital Markets Studies, 6*(1), 33-47. https://doi.org/10.1108/JCMS-06-2021-0021

Teng, Huei-Wen; Lee, Michael. (2019), Estimation Procedures of Using Five Alternative Machine Learning Methods for Predicting Credit Card Default. Review of Pacific basin financial markets and policies, 2019, Vol.22 (3), p.1950021. Retrieved from Estimation Procedures of Using Five Alternative Machine Learning Methods fo...: EBSCOhost (drexel.edu)

THE FINANCIAL CRISIS INQUIRY COMMISSION (2011), Final Report of the Causes of the Financial and Economic Crisis in the United States. Retrieved from https://cybercemetery.unt.edu/archive/fcic/20110310173545/http://c0182732.cdn1.cloudfiles.rackspacecloud.com/fcic_final_report_full.pdf

Credit Risk Analytics

Thank You

Group 2