

Unison Home Ownership Investors

Quantitative Financial Analyst – Research and Data

Mini Project 2021

Submitted by: Mohit Shukla

Date: 17th February 2021

Q1. SQL Database

Connect to the MySQL database which has a table of samples of the historical Freddie Mac singlefamily loan level data.

The following shows the connection details received after running the code.

```
<mysql.connector.connection.MySQLConnection object at 0x000001E00EF64520>
```

The following is a list of databases fetched from MySQL

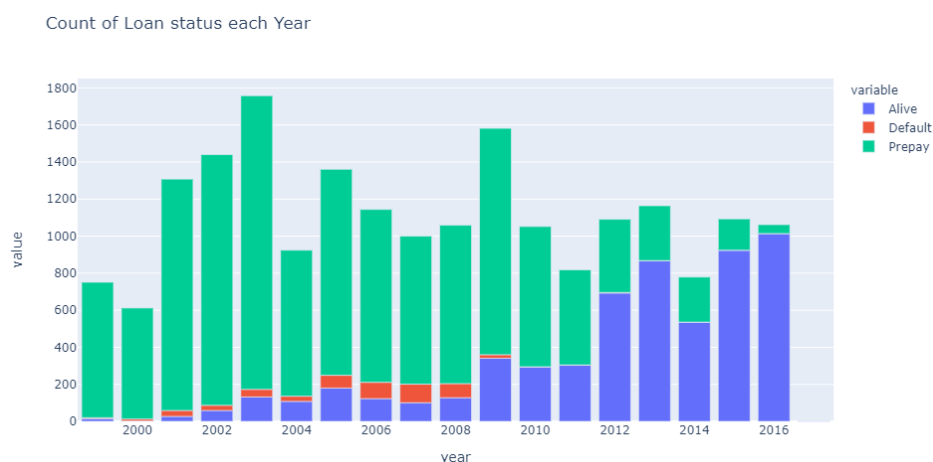
```
[('information_schema'), ('AgencyData'), ('mysql'), ('performance_schema'),]
```

After connecting with the database, we can now get the required sample data table.

Q2 Analyse Mortgage Statuses by first payment date

Analyze the breakdown of mortgage statuses (prepay, default, alive) for each origination cohort (group by year of first_pmt_date).

We first extract the year from first_pmt_date, then create a pivot table, grouping on the index 'year' and columns as different values of 'status'.



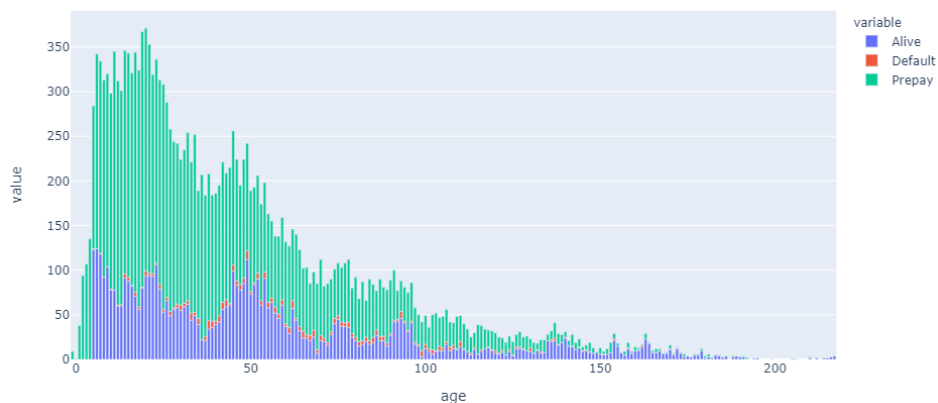
Comment on the change in mortgage behavior through the mortgage origination cycle.

We are likely to see more and more prepayments (or defaults) as the mortgages age, as consumers may see value in switching (or default if they are unable to pay the installments), as the economic conditions change. Thus, we see very few old loans alive, while most of the recent ones are still active.

Q3 Analyse Mortgage Statuses by age

Repeat the analysis above but instead of grouping the mortgages by first_pmt_date, group them by age.

Stacked Bar-Chart for Age



In addition, briefly explain how you may go about building prepayment and default models for residential mortgages. What data will be required if we want to build such models for Unison home equity agreements (UEA)?

To build a prepayment and/or default model, we need to consider which factors play a part in changing the status of a mortgage from 'Alive' to either 'Prepay' or 'Default'. The same can be achieved through logistic regression. We can also look at the CPR(t) - the conditional prepayment rate - to build the model dependent on these factors such as interest rate, fico score etc.

For Unison Home Equity Agreements (UEA), apart from the given factors, we also need the data for

- **Home prices** (for e.g. HPI - House Price Index), as higher prices also act as an incentive for borrowers to go for higher mortgage or selling the property.
- **Economic Indicators** like GDP growth rate, unemployment rate etc. A booming economy would lead to higher turnover in residential homes, as owners may want to expand or move to a better location/house.

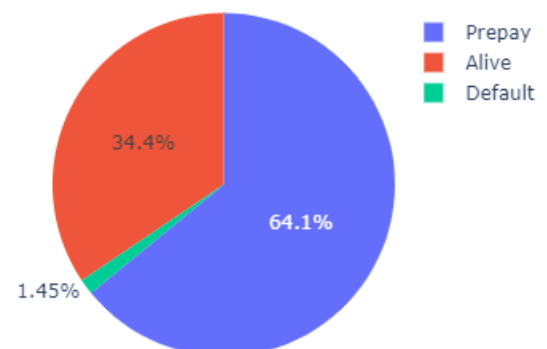
Q4. Unconditional Probabilities of Mortgage Behavior

Determine the unconditional probabilities of mortgage behavior based on the sample data and describe how you arrive at your results.

We are looking at the probability of a mortgage either defaulting, prepaying or staying alive in 5 years. Since these are the only 3 possibilities, the total of these probabilities must be 1. To look for unconditional probability, we need to make the following changes:

- **Remove all loans which are alive and have age less than 5 years (60 months)** - Since we do not know about the end state (at 5 years) of the loan which is alive at age 2 years, we need to remove such loans from the dataset.
- **Change status of all loans above 5 years to 'Alive'** - The loans marked 'Prepay' or 'Default' after 5 years would all be alive at the end of 5 years.

Probability of status within 5 years



How could prepayment or default behavior of the mortgage affect the Unison equity agreement (UEA) on the same home?

A prepayment may be refinanced due to lower rates or cash-out, which imply a rise in house price. In case of turnover as well, the borrower will also prepay and close the loan. All these scenarios are beneficial for Unison as the Equity value in the house goes up, which stands to benefit both Unison and UEA holder. Default on the other hand signifies an economic hardship, which may result in a distress sale or auction of property by lender. Such an event would negatively impact the value of house, and the return for Unison in the UEA.

Q5 Prepayment Model

Select one or more factors which you believe would affect the 'Prepay' response and build a simple model to predict prepayment or default rates

Let's build a model for Prepayment, in which case prepayment would signify 1 while Alive or Default would be 0 in the logistic regression. Next we create dummy variables for categorical columns. We can run an OLS Regression to test the t-stat for different factors, and then remove those where $|t\text{-stat}| < 1.96$, hence only considering those which are significant with 95% confidence interval.

OLS Regression Results						
=====						
Dep. Variable:	Y	R-squared:	0.314			
Model:	OLS	Adj. R-squared:	0.313			
Method:	Least Squares	F-statistic:	377.0			
Date:	Wed, 17 Feb 2021	Prob (F-statistic):	0.00			
Time:	00:53:32	Log-Likelihood:	-5051.6			
No. Observations:	13188	AIC:	1.014e+04			
Df Residuals:	13171	BIC:	1.026e+04			
Df Model:	16					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.7381	0.064	-11.491	0.000	-0.864	-0.612
age	-0.0031	8.13e-05	-38.664	0.000	-0.003	-0.003
credit_score	0.0009	6.21e-05	13.947	0.000	0.001	0.001
orig_cltv	-0.0016	0.000	-7.747	0.000	-0.002	-0.001
orig_dti	-0.0003	0.000	-1.078	0.281	-0.001	0.000
orig_ir	0.1652	0.003	58.509	0.000	0.160	0.171
num_borrowers	0.0612	0.006	9.545	0.000	0.049	0.074
first_time_ho_flag_Y	-0.0571	0.009	-6.069	0.000	-0.075	-0.039
occupancy_status_O	0.1210	0.018	6.709	0.000	0.086	0.156
occupancy_status_S	0.1943	0.025	7.757	0.000	0.145	0.243
prop_type_CP	-0.0059	0.057	-0.103	0.918	-0.118	0.107
prop_type_LH	0.0176	0.145	0.121	0.903	-0.268	0.303
prop_type_MH	-0.1392	0.040	-3.460	0.001	-0.218	-0.060
prop_type_PU	0.0170	0.014	1.238	0.216	-0.010	0.044
prop_type_SF	0.0139	0.012	1.154	0.248	-0.010	0.037
loan_purpose_N	0.0537	0.009	5.734	0.000	0.035	0.072
loan_purpose_P	-0.0805	0.009	-9.097	0.000	-0.098	-0.063
=====						
Omnibus:	869.205	Durbin-Watson:	1.994			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	750.724			
Skew:	-0.512	Prob(JB):	9.60e-164			
Kurtosis:	2.435	Cond. No.	3.49e+04			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.49e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Looking at the t-stats for above variables, we can discard prop_type and consider everything else in the model. This just indicates that property type does not significantly impact the decision to prepay a loan. The most significant factors, from the above, are interest rates, loan age and credit score, which are usually quoted as major factors for borrower behaviour. We can now run a logistic regression, the score of which will be displayed at the end. A high score denotes how well the model is performing.

Logistic Model Score: 0.83128601759175

The following confusion matrix shows the performance of the model vs the actual prepayments.

Actual 0s	1731	1466
Actual 1s	759	9232
	Predicted 0s	Predicted 1s

Comment on how your selected response (prepayment or default) might affect the market price of a mortgage and how you could devise an investment strategy around this model's output.

By building the model, we can predict the cashflows of the mortgage. These cashflows discounted back would then be used to arrive at the value of the mortgage. If the model is good, we can identify the mortgages (MBS/MSR) which are mispriced, earning an arbitrage in the process.

Q6 Why monitoring is essential for Unison

Why do you think we need to monitor, analyze and understand the mortgage industry in general and our customers' mortgages in particular?

By buying borrower's equity, Unison is a party to the performance of home equity. The mortgage may be a lead indicator, and the performance of mortgage and its behavior can tell us whether the investment is doing well or poorly and allow us to manage our risk better. Also, a good borrower is required for both mortgage as well as UEA, as a major risk is tied to the economic condition of the borrower. Analyzing these factors would improve our portfolio performance and mitigate the risk.

Name key metrics that you believe are good indicators of risk and explain how you think we can use those metrics in risk management, investment structuring, etc.

The major risk indicators are:

1. Housing prices - since we're buying home equity, this becomes the primary risk for the portfolio.
2. Economic conditions (GDP growth rate, unemployment rate, inflation/interest rate) - these macro factors can help us in determining whether the returns expected on the portfolio are good or bad.
3. Borrower characteristics (fico score, debt to income, loan to value) – These are the same characteristics used by the lenders to decide during the loan underwriting.