

We started this project by using the Bloomberg Terminal to download the loan and property level data for 2385 multifamily fixed-rate loans where the original loan balance was between 1,000,000 to 4,000,000. Upon completing the data cleaning and merging process to establish a one-to-one relationship between loans and properties, we recognized the need to streamline our dataset due to the fact that we had over 50 columns in the merged dataset. Thus, we decided to drop off the columns that had a significant amount of NA values, so we ultimately ended up with the following 10 columns: Loan Name, NOI(USD), Origination Date, City, Original Amortization, Cap Rate, DSCR, Original Bal(USD), Appr Val(USD), and Loan Status. As a result, after the cleaning, we reduced our data from 2385 loans to 423 loans

Looking at the summary statistics, we noticed that as NOI and Appraisal Value increases, there is a higher chance that a loan is prepaid as opposed to maturing / performing / being late. Additionally, we see that prepaid loans with penalty have the highest DSCR mean (3.092897) compared to the other loans. Also, loans that perform(w) have the highest mean original amortization (370.90909) and lowest mean original balance (1,892,579.6) while loans that simply perform have one of the lowest mean original amortization (319.65517). Loans that are prepaid with a penalty consistently have the highest original balance (3,400,000) compared to the other loans, especially as the standard deviation is the lowest.

To explore the relationships between the characteristics of the loans and properties and the performance of the loans, we initially ran a logistic regression as opposed to a linear regression because we treated each performance outcome as a categorical variable. And specifically, we used this regression to compare the DSCR, NOI(USD), Cap Rate, Days Since Origination, and Approximate Val(USD) of loans across all potential performance outcomes with loans of the "late" outcome. During our analysis, we identified notable relationships. For instance, we observed that as DSCR increases, the likelihood of a loan performing more poorly decreases, indicated by negative coefficients (-0.00941) for "perform(w)" and (-1.07602) for "prepaid(payload w/ penalty)." Consequently, we infer that higher DSCR values are associated with a lower probability of loan default. However, we did not extensively analyze this regression because we were aware of an overfitting concern. Given that our dataset contained only two loans categorized as "late," the analysis risked being skewed or unreliable due to the lack of sufficient data points for this particular outcome.

To further analyze the data, we produced visualizations with the ggplot2 library in R. For example, we created a bar graph of the Loan Status Counts by State. Based on the bar graph analysis, we have identified California, Florida, Maryland, and Texas as the strongest and most secure real estate markets. This conclusion is drawn from the significantly large percentages of loans that have performed within these states compared to other states. Additionally, we made a Boxplot of Days Since Origination by Loan Status and saw that perform(w) had the most median days since origination at around 4385 days, while prepaid (payload with YM) had the lowest median days since origination at approximately 4175 days. Furthermore, it is evident that loans that have performed typically have fewer days since origination than loans that haven't because

the box for those loans drops further down. Next, we looked at the Distribution of Days Since Origination by Loan Status. Not only do we see that there is a normal distribution, but we also see that the distribution centers around 4300 days as the median for loans that have performed, which confirms our findings from the Boxplot of Days Since Origination by Loan Status.

We then looked at the Distribution of Original Amortization by Loan Status and noticed that most loans have an original amortization of ~ 360 months. However, we did notice that some loans that have performed have an amortization of ~ 175 months and ~ 420 months. We also looked at the Distribution of Original Balance by Loan Status, and it is clear that loans perform better with a lower original balance ($< 2,500,000$). But to caveat, we noticed that the distribution is skewed to the right as there are more loans in our data with lower original balance.

After, we looked at the Distribution of NOI(USD) by Loan Status. From this histogram, we noticed that most properties had an NOI less than 500,000. And at the 500,000 mark, the number of properties dropped off, but we saw that there is a significant number of properties with loans that have performed with a NOI between 500,000 and 1,000,000. And loans with higher NOI values were mainly prepaid (payoff w/YM). We looked at the Distribution of Appraised Value by Loan Status next, and this distribution looked very similar to the distribution of NOI(USD) by Loan Status, which shows that NOI and Appraised Value are highly correlated. And after creating a Correlation Heatmap, we confirmed that NOI and Appraisal value have a very strong relationship. And to examine the relationship between DSCR and NOI, we created a Scatterplot of DSCR vs NOI. We noticed a highly positive correlation between DSCR and NOI, specifically when the DSCR is less than 2.5. However, the strength of this relationship significantly diminishes above 2.5, essentially becoming nonexistent, along with the probability of the loan performing.

Our analysis reveals several significant insights into the lending market across different states and loan performance indicators. Firstly, California, Texas, Florida, and Maryland emerge as the strongest markets in terms of loan performance and market activity. Among these, Maryland stands out with the highest percentage of loans that have performed well over time. Loans that perform (w) exhibit a much longer longevity compared to other loan statuses, especially prepaid loans which are labeled as "Payoff with YM." Interestingly, loans with an original amortization period around 175 months tend to demonstrate the best performance, albeit this observation is based on a limited sample size. Furthermore, loans originating with a smaller original balance often outperform their counterparts with higher initial amounts, which is most likely due to the fact that it is easier to pay off a smaller loan. Additionally, buildings with a higher NOI and Appraisal value predominantly decide to prepay their loan, while they rarely default or underperform. Lastly, a DSCR exceeding 2.5 serves as an indicator against loan performance, with such loans not performing nearly as often as buildings with a lower DSCR, and often opting to prepay the loans.