

group_eda_rmr

Data Overview

Table 1: Predictor / Outcome Statistics

Predictor / Outcome Variables	N = 2,341
Duration (years)	Mean: 7.61, StDev: 4.66, Median: 6.69
Rate of Amortization (%)	Mean: 0.02, StDev: 0.02, Median: 0.01
Debt to Asset (ratio)	Mean: 0.33, StDev: 0.17, Median: 0.33
Debt to Equity (ratio)	Mean: 1.30, StDev: 12.95, Median: 1.05
Interest Coverage (ratio)	Mean: 12.73, StDev: 35.19, Median: 8.04
Unknown	327
Debt Service Coverage (ratio)	Mean: 1.50, StDev: 3.05, Median: 0.60
Unknown	387
Cash Coverage (ratio)	Mean: 6.00, StDev: 11.10, Median: 1.98
Unknown	387
Current Ratio (ratio)	Mean: 1.36, StDev: 2.33, Median: 1.13
Unknown	350
Cash Ratio (ratio)	Mean: 0.37, StDev: 0.51, Median: 0.22
Unknown	350
Operating Profit Margin (ratio)	Mean: 0.72, StDev: 0.20, Median: 0.76
Unknown	374
Ebidta Margin (ratio)	Mean: 0.26, StDev: 0.20, Median: 0.26
Unknown	347
Sentiment Score	Mean: -45.56, StDev: 22.90, Median: -50.42
Unknown	17
Industry Sector	
Banking	Count: 209 %: (8.9%)
Basic Industry	Count: 70 %: (3.0%)
Brokerage/Asset Managers/Exchanges	Count: 27 %: (1.2%)
Capital Goods	Count: 233 %: (10.0%)
Communications	Count: 202 %: (8.6%)

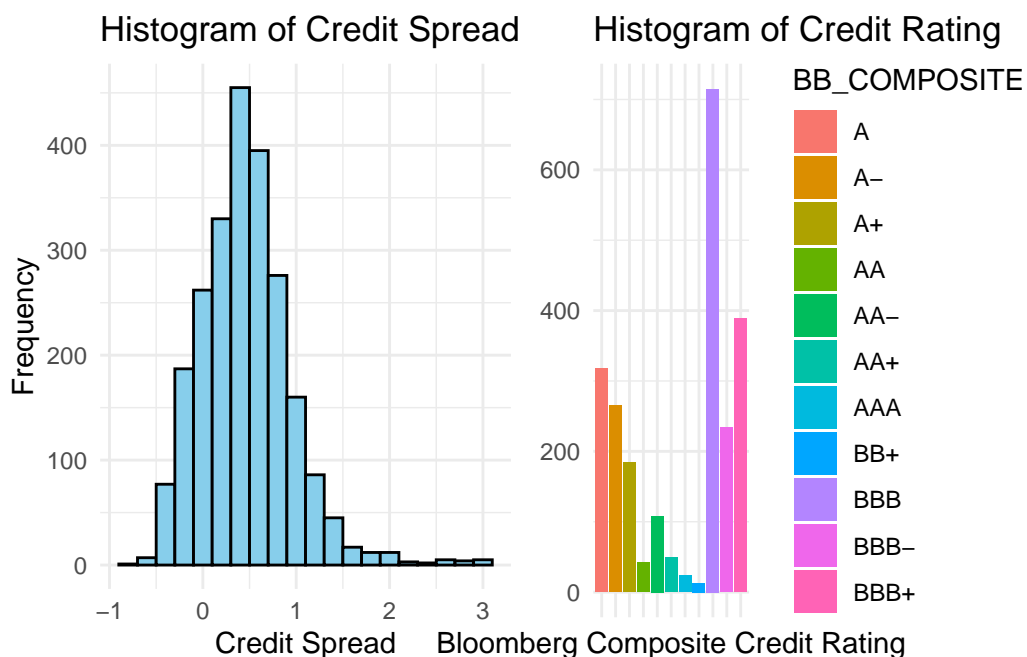
Predictor / Outcome Variables	N = 2,341
Consumer Cyclical	Count: 208 %: (8.9%)
Consumer Non-Cyclical	Count: 554 %: (23.7%)
Electric	Count: 83 %: (3.5%)
Energy	Count: 134 %: (5.7%)
Insurance	Count: 193 %: (8.2%)
Natural Gas	Count: 25 %: (1.1%)
Reits	Count: 25 %: (1.1%)
Technology	Count: 307 %: (13.1%)
Transportation	Count: 71 %: (3.0%)

Outcome Variables

The **Credit Spread** is a continuous outcome variable and represents the difference in yield or interest rate between a particular bond and a benchmark bond with similar characteristics but considered to be risk-free. For example, if a corporate bond yields 5% and a comparable risk-free government bond yields 3%, then the credit spread of the corporate bond is 2%. The distribution of credit risk in the data set with respect to the figure “x” below, appears to follow a normal distribution with a high positive skew. The mean Standard Deviation and Median of Credit Spread are as follows: 0.47, 0.49, 0.44.

The **Credit Rating** is an ordinal categorical variable, and represents the composite credit rating issued to each company by Bloomberg. When we look at the distribution of Credit Ratings in figure ‘X’, we notice that it appears more scattered and less structured compared to the distribution of Credit Spreads. To make sense of this, we’re considering combining certain categories of Credit Ratings. This could potentially help to reduce the variability and make the data clearer. We’ll discuss this further in the potential problem section. It’s worth noting that the ‘BBB’ category has a much larger number of observations in the data set compared to other categories. This makes it a significant portion (30.5%) of the data.

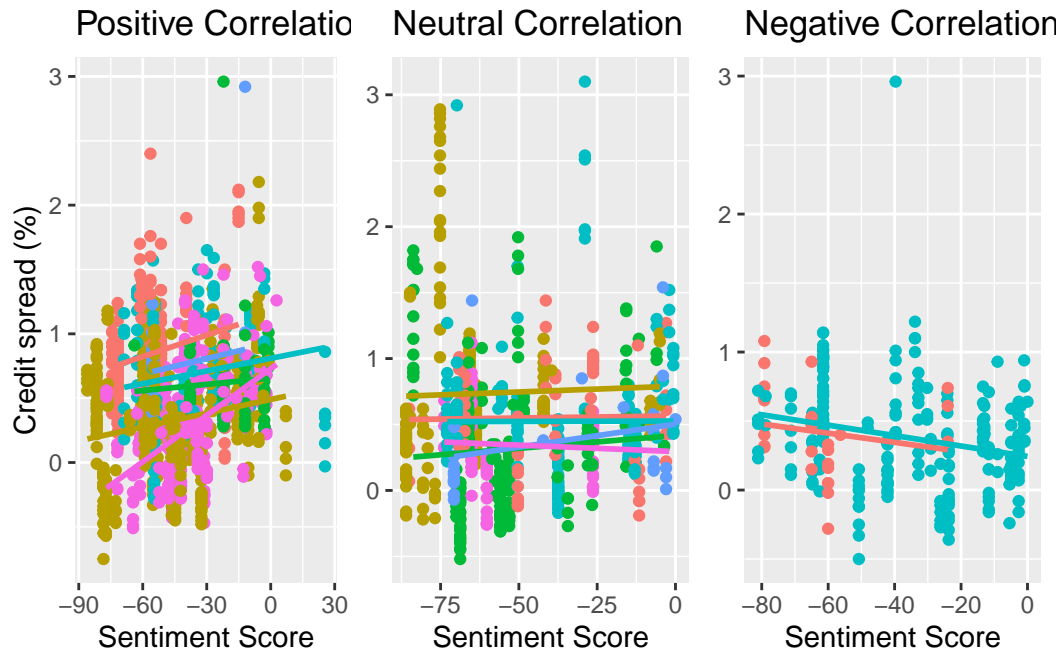
Figure 1.1, 1.2: Outcome Variables



Primary Relationships of Interest

The sentiment score exhibits a notable influence on the behavior of credit spreads, varying significantly across different sectors. For instance, in Figure 2, distinct sectors demonstrate varying correlations between Sentiment Score and Credit Spread. Some sectors display a positive correlation, while others show a negative or neutral association. Further examination revealed that sectors characterized by a more elastic demand, such as Technology and Banking, tend to exhibit a more pronounced positive correlation. Conversely, sectors with a more inelastic demand, like Basic Industry and Cyclical Consumer Industry, demonstrate a largely neutral response to the Sentiment Score.

Figure 2.1, 2.2 , 2.3: Sentiment Score Vs Credit Spread

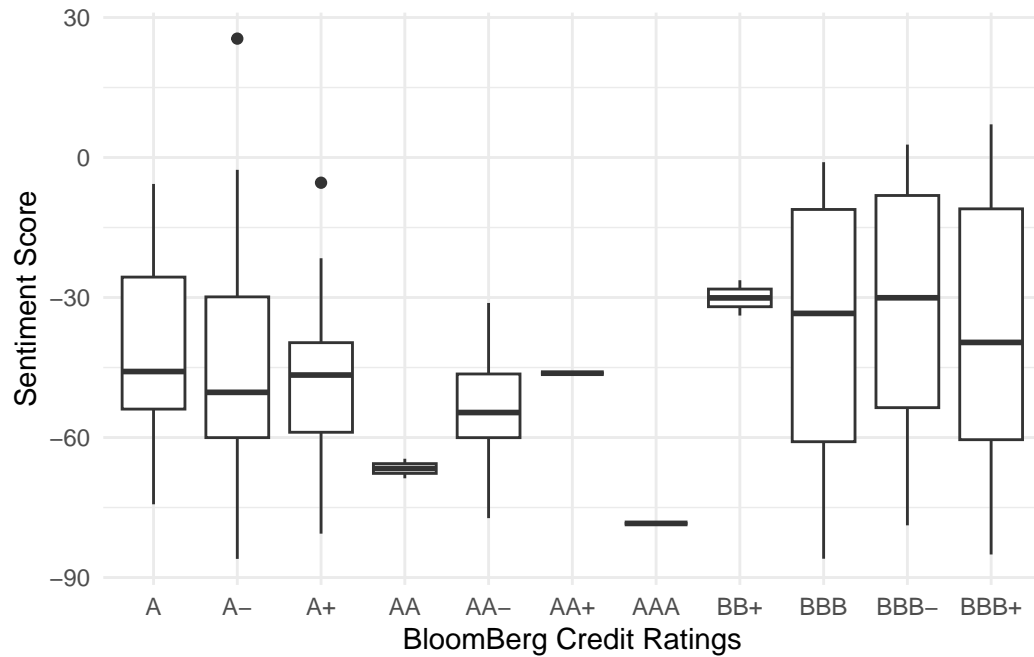


The data in Figure 3.1 suggests that companies with a credit rating of B (including BBB, BB+, etc.) exhibit a considerably broader interquartile range compared to those with an A rating. Additionally, all statistical measures in the box plot for A-rated companies are consistently lower when compared to their B-rated counterparts. *Figure 3.1: Sentiment Score Vs Credit Ratings*

```
most_common <- function(column) {
  frequencies <- table(column)
  most_common_category <- names(frequencies)[which.max(frequencies)]
  return(most_common_category)
}

bonds_agg <- bonds %>%
  group_by(ticker) %>%
  summarise(mean_value = mean(score), most_freq = most_common(BB_COMPOSITE))

ggplot(bonds_agg, aes(x = most_freq, y = mean_value)) +
  geom_boxplot() +
  labs(x = "BloomBerg Credit Ratings", y = "Sentiment Score") +
  theme_minimal()
```

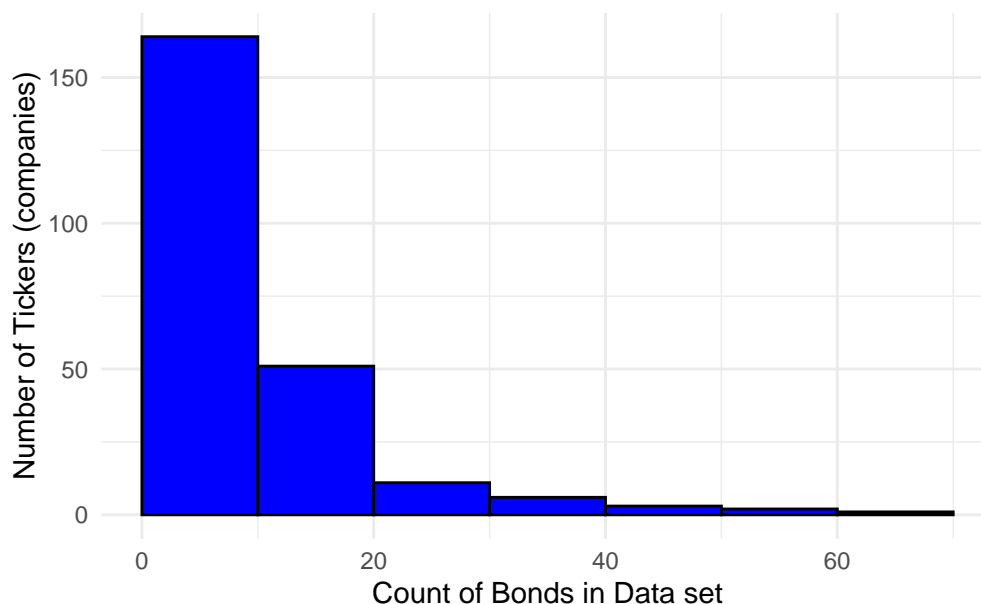


Other Characteristics

Other variables in the data set includes Other Technical Indicator such as Rate of Amortization, Ebidta Margin, Operating Profit Margin, etc.

Potential Challenges

Figure 3.1: Histogram of Number of Bonds per Ticker



One challenge our team may encounter relates to the varying frequencies at which each unique company's ticker appears in our raw data. As depicted in the figure above, most tickers have between 1 and 10 bonds in our data set. However, there are a few tickers at the higher end, with between 60 and 70 bonds. This imbalance leads to a data set where models trained on it might exhibit a bias towards companies with more bonds (observations). To address this concern, the team could implement strategies for handling the imbalanced data set. Some potential approaches could include, Weighted Sampling, Stratified Sampling, Re sampling Techniques, Etc.

Another potential challenge arises from the low counts in the categorical variables, specifically in Bloomberg Composite Rating (the outcome variable) and Industry Sector. In both cases, there are classes with limited representation. For instance, Banking and Brokerage/Asset Managers/Exchanges collectively account for only ~10% of the total data. One argument can be made that these industries share enough similarities to be collapsed into a single category. By consolidating certain categories, the team aims to mitigate the issue of low counts.

Once the adjustments are made, the team is confident that the resulting data set will be of sufficient size to effectively address the research questions.