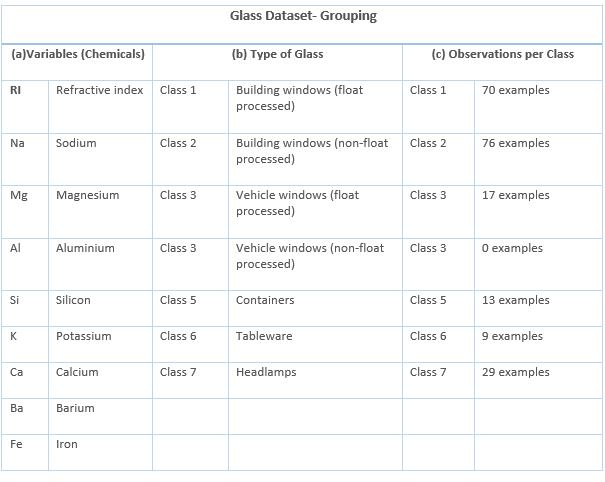
Data Analysis Report

Glass Identification Data Set

# Description of Dataset

The dataset describes that glass are categorised into one of six classes based on their chemical properties. The glass dataset's attributes are calculated by nine input variables, as shown in table 1(a). The weight percentage in corresponding oxide is used to measure the chemical compositions of glass. Table 1 (b) lists seven different kinds of glass and float glass refers to the process of creating the glass.

The dataset contains 214 observations, and table 1(c) lists the unequal number of observations in each class. For class 4 (non-float processed car windows), there are no examples included in the dataset. Classes 1-4 and 5-7 of the datasets are categorised as window glass and non-window glass, respectively. There are 51 examples of glass that is not a window and 163 examples of window glass. If the observations were limited to window glass only, another separation of the data would be between float processed glass and non-float processed glass. There is greater balance in this division.



# Verifying the Aspects of Data

To verify the different aspects dataset, histogram, boxplot pair scatter plot and density plot are used. To statistically analysis the relation of RI and glass type with predictor variables linear regression model is used.

## Identification of Outliers

To find the outliers, Boxplots are plotted, and the plots showed that there are outliers in all the variables expect magnesium. The K and Ba variables contain the most extreme outliers. Additionally displaying skewing in several variables, the Boxplots perform poorly when it comes to extreme outlier variables. However, no outliers were removed as it is not part of the report.

## Distribution and Relationship between Variables

It appears that the distributions of RI, Na, Al, and Si are relatively normal (symmetric) based on the histograms of the variables. The distributions of the remaining variables seem to be asymmetric. Strong correlations do not seem to be present for the majority of the variable pairings in the scatterplot matrix. The RI, Ca pair, which has an obviously positive association, and possibly the RI, Si pair, which exhibits a somewhat negative relationship, are the outliers.

## Linear regression Model

To identify the which predicator variables are more associated with refractive index and glass type simple linear regression models are used. Based on the linear mode results 'Na', 'Mg', 'K', 'Ca', and 'Ba' have p-values less than 0.05, indicating they are statistically significant predictors of refractive index RI. In case of glass type 'Al' seems to be statistically significant (p-value = 0.0354), indicating that 'Al' might be an important predictor for 'glass-type'. Other variables, such as 'Na', also show some significance (p-value = 0.0842), but not as strongly as 'Al'. In these results provide start point for finding trends abut simple linear model is not provide reasonable evidence which variable (chemical) will be strong indicator for reflective index and glass type.

# Visualization of Multi-Variate and Trends

This part explains the reasons for the plotting of the variables and the annotation of the plot's trends. part 4 offers more details on the trends.

## Oxides that Predict the Glass Type

Based on the findings of regression analysis, box plots were used to determine which two oxides could most accurately predict the kind of glass. The box plot results indicate that the two best oxides that are crucial in predicting the type of glass are sodium (Na) and aluminium (Al). The box plot of sodium and aluminium with different glass types is shown in figure 1 below. The different amounts of these two oxides will help in identifying the kind of glass for each of the seven categories of glass that are classified as window or non-window glass.

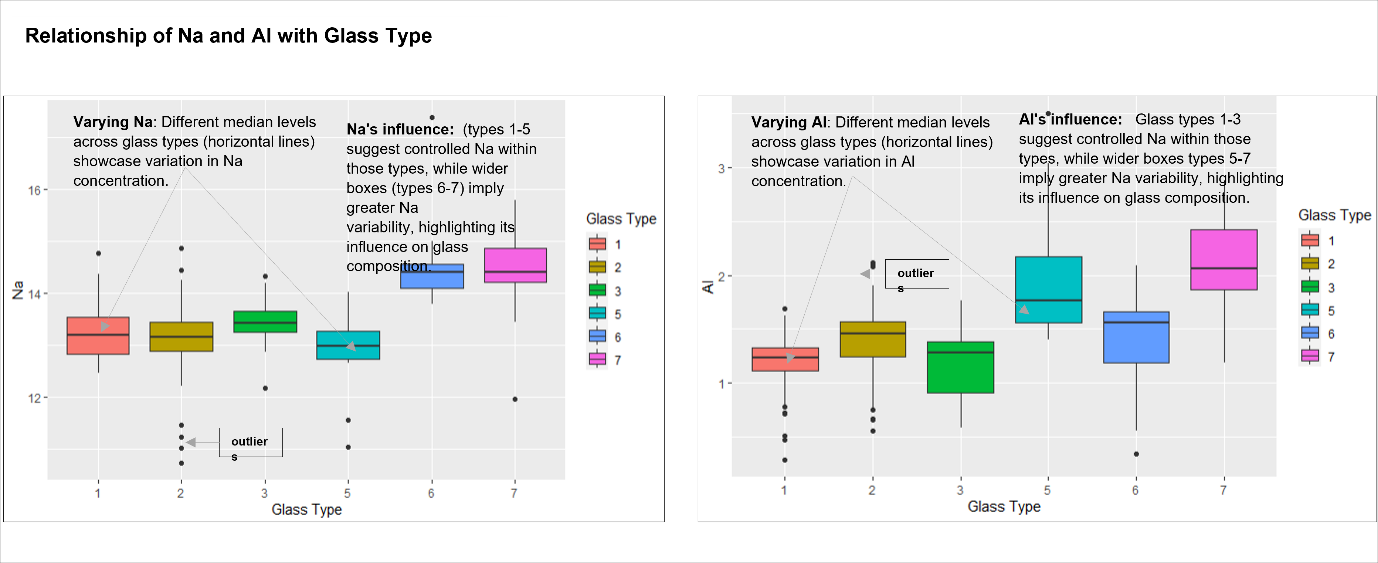


Figure : Sodium (Na) and Aluminium (Al) Oxides as Predicator for Glass Type( Windows and Non-Windows)

Relying less on the regression model, the average quantity of oxides is plotted against the type of glass to visualise the presence of other oxides in various glass types. This provides a clear image, as illustrated in figure 2, of which oxide types would be taken into consideration to predict the glass type. Section 4 paragraph one discussed the trends and optimum oxides to consider while predicting the types of glass.

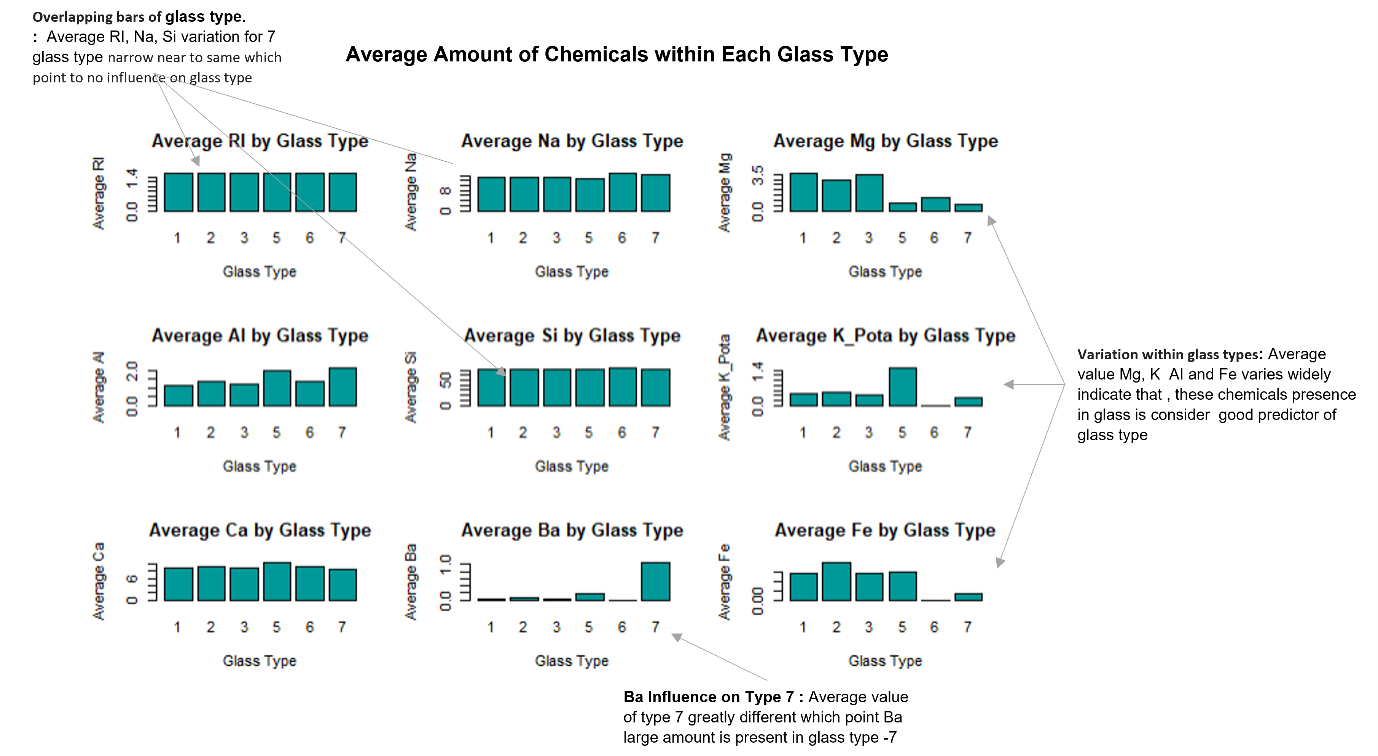


Figure : Oxides (chemicals) Average Amount Present in Glass Type(Windows and non-Windows)

## Oxides that Predict the Reflective Index (RI) of Glass

The Refractive Index (RI) of glass with different oxides is shown on a correlation heat map in order to determine which variable has the greatest positive and least negative relationship with the response variable. The glass type and iron (Fe) have the strongest refractive index association.

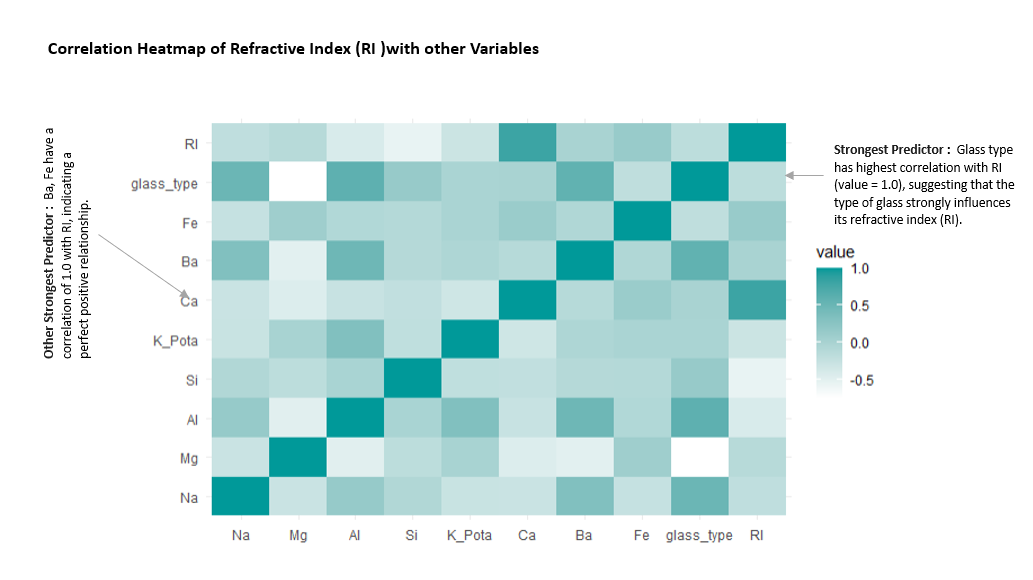


Figure : Correlation Matrix Heatmap to explain the correlation (Positive & Negative) of Refractive Index with other Oxides.

A grid scatter plot was created to better explore the impact of oxide presence on the refractive index of glass. The results indicate that the existence of additional oxides may have an impact on the refractive index of glass in addition to glass type.

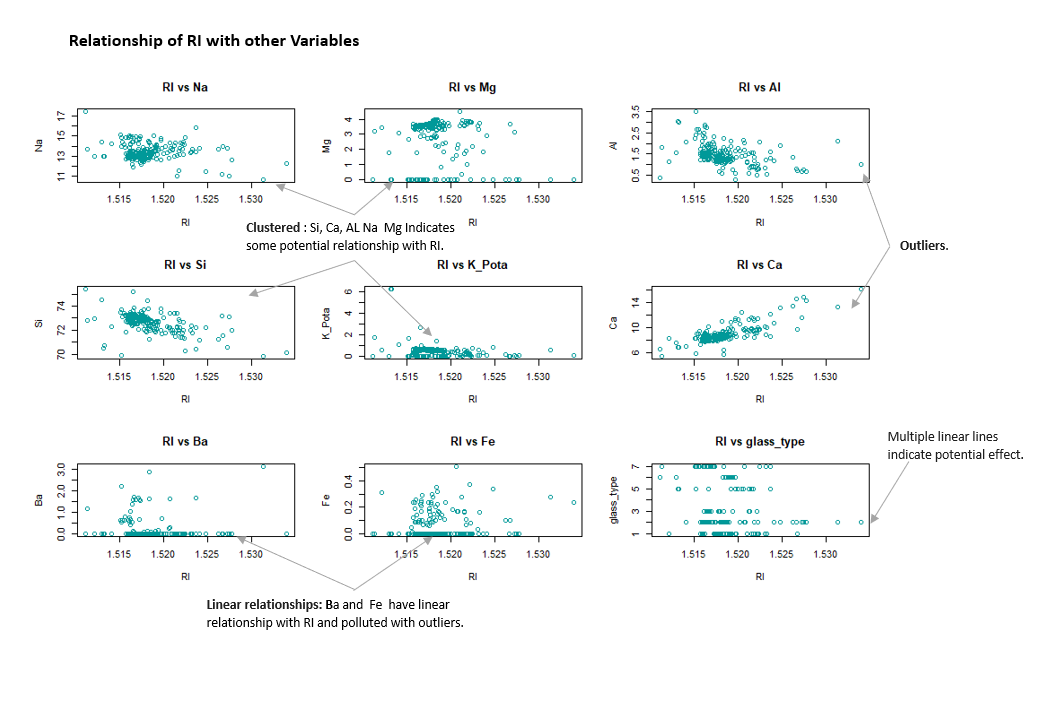


Figure : Scatter Plots of Oxides to Show the Spread for Refracive Index and presence of Outliers

Plotting the regression model result allows for a deeper exploration of the data. According to the regression model, the refractive index may be significantly predicted by calcium (Ca), magnesium (Mg), and sodium (Na). Nevertheless, patterns and the most important oxides as a refractive index predictor are discussed in of paragraph two.

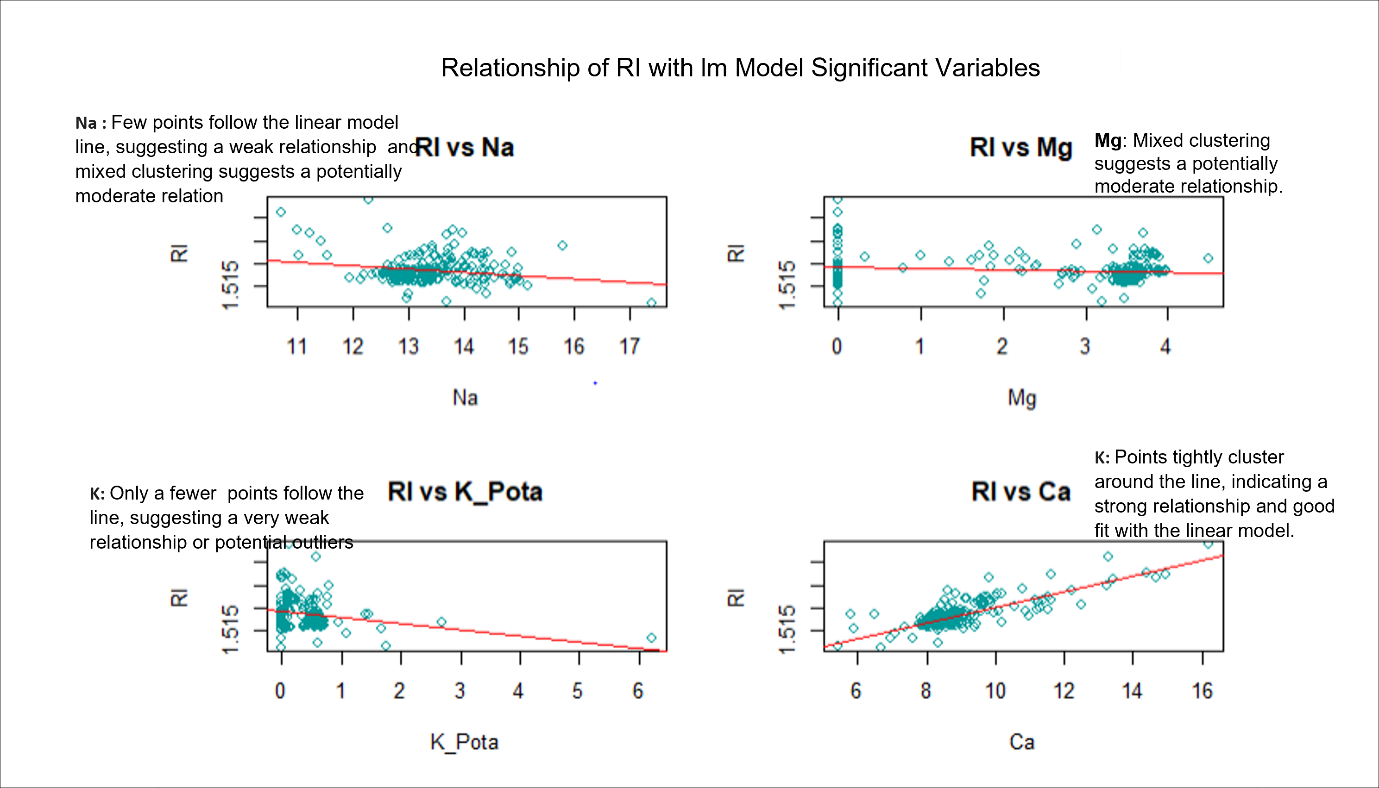


Figure : Most Significant Oxides to Predict the Refractive Index Based on Linear Regression Model Results

# Discussion

The defining role of Na in the various glass types is indicated by the varied medians and diverse spreads in Na concentration. The differences in the median Na concentration lines inside the boxes indicate that different varieties of glass have different usual Na concentrations. This variety implies that Na has a major role in differentiating these types of glass. Furthermore, the wider spreads of boxes for kinds 6 and 7 suggest higher fluctuation in Na content, the narrower boxes for types 1–5 suggest more constant Na concentrations within these categories. This implies that the concentration of Na in some glass varieties may be closely monitored or controlled, which contributes to their unique characteristics. But it is fair to conclude—after seeing the oxide bar graphs for each type of glass—that the presence of iron, potassium, magnesium, and aluminium in glass contributes to its type of identification. Determining the optimal oxides for glass type prediction required through additional statistical analysis and data training.

Three distinct plots are analysed to identify important refractive index (RI) predictors. The type of glass is the most significant variable among the other variables that positively linked with RI, with a perfect correlation value of 1.0. This indicates that the type of glass has a significant impact on the RI, which is supported by the grid plots' observed variations in RI for various glass types. Fe also shows favourable interaction with RI. Strong linear connections with RI are shown by the plots of Ba, Fe, and glass type, which are easily identifiable. However, the best fit line of linear mode says that Ca most significant indicator. Glass type and Ca followed by Ba can be considered as best predictor of refractive index.