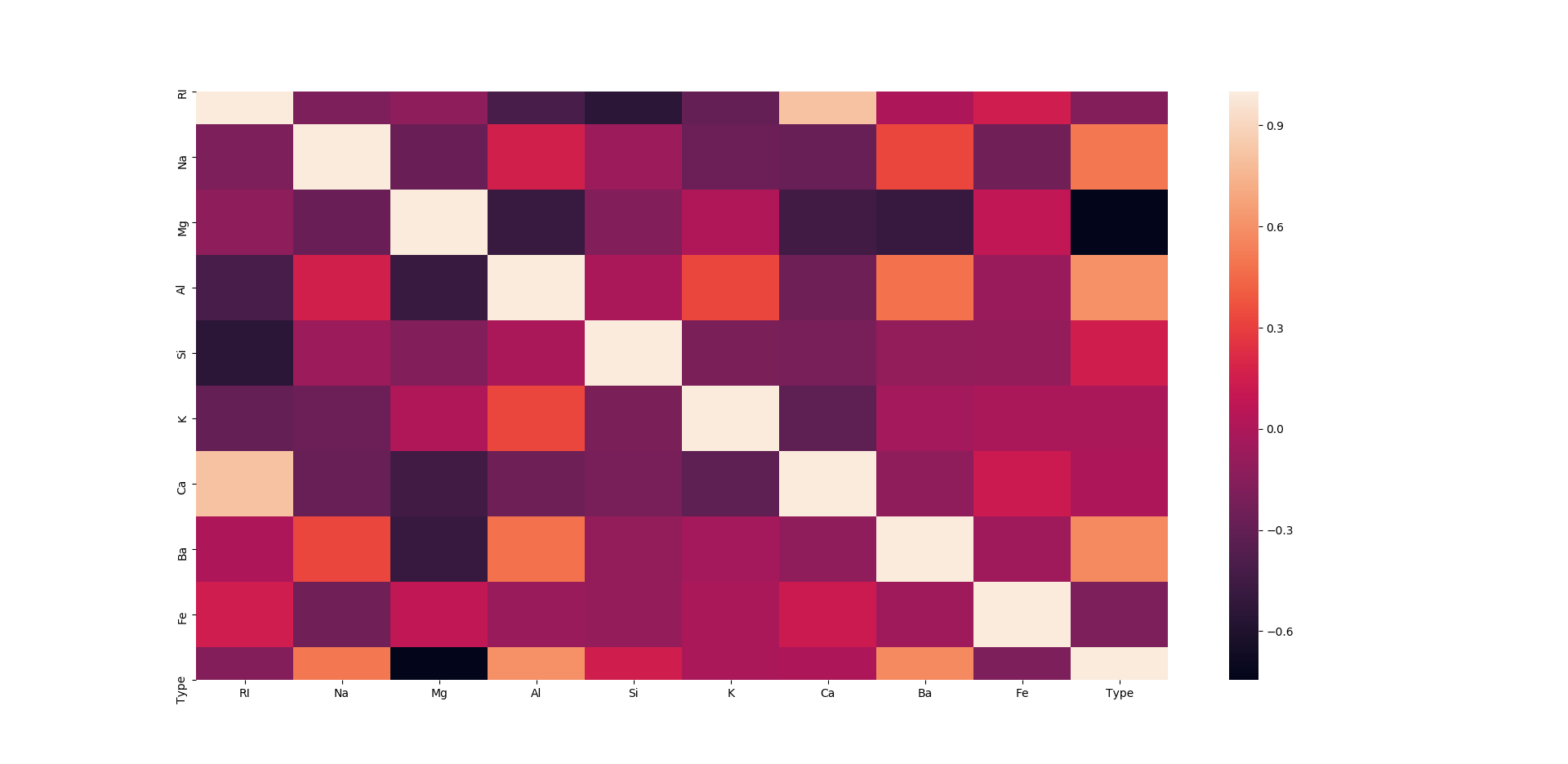
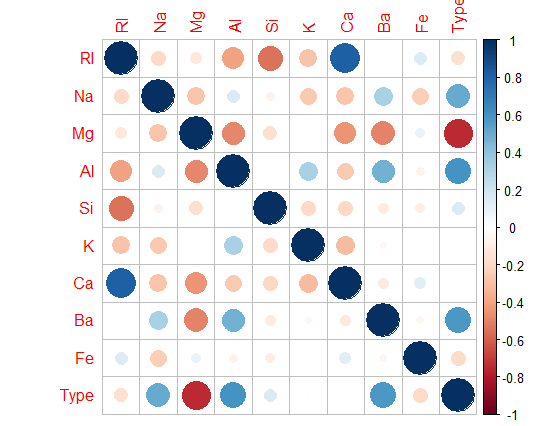
KNN CLASSIFICATION FOR GLASS DATA:

=>The distance metric is used to calculate its nearest neighbors (Euclidean, Manhattan).

=>Can solve classification (by determining the majority class of nearest neighbors) and regression problems (by determining the means of nearest neighbors).

=>If the majority of the nearest neighbors of the new data point belong to a certain class, the model classifies the new data point to that class.

DATA VISUALIZATION:

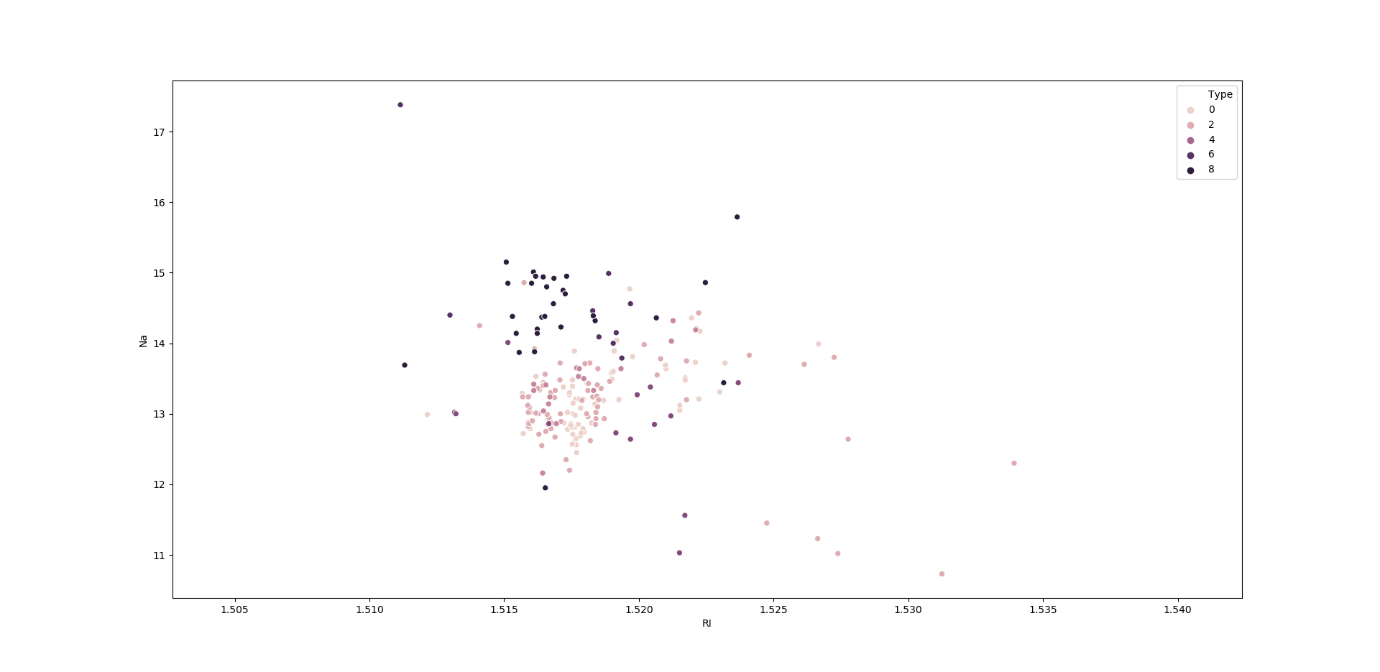


We can notice that Ca and K values don't affect Type that much.

Also, Ca and RI are highly correlated, this means using only RI is enough.

So, we can go ahead and drop Ca, and also K.

#### Scatter plot of two features, and pairwise plot:

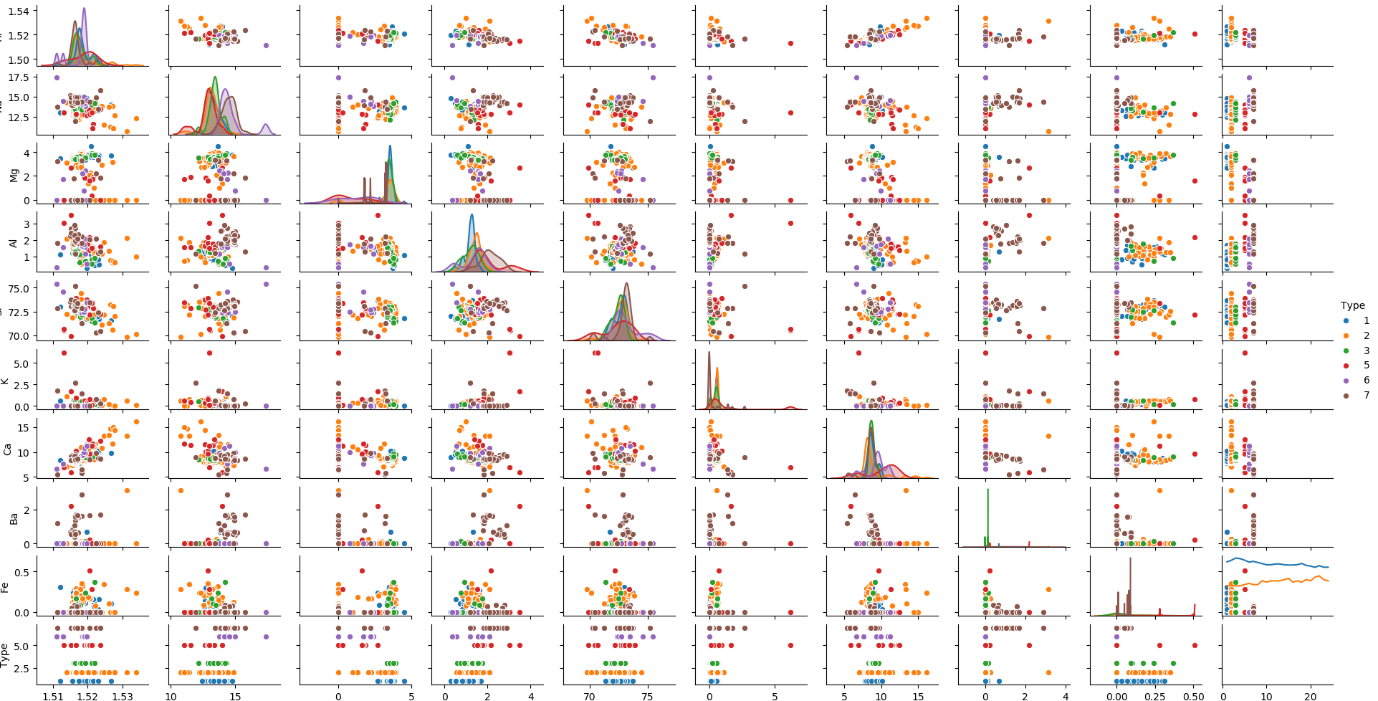


Suppose we consider only RI, and Na values for classification for glass type.

From the above plot, we first calculate the nearest neighbors from the new data point to be calculated.

If the majority of nearest neighbors belong to a particular class, say type 4, then we classify the data point as type 4.

But there are a lot more than two features based on which we can classify. So, let us take a look at pairwise plot to capture all the features.

PAIR PLOT:

he pair plot shows that the data is not linear and KNN can be applied to get nearest neighbors and classify the glass types.

Scaling is necessary for distance-based algorithms such as KNN. This is to avoid higher weightage being assigned to data with a higher magnitude.

Using standard scaler, we can scale down to unit variance.

Formula:

z = (x - u) / s

where x -> value, u -> mean, s -> standard deviation

## Applying KNN

Drop features that are not required

Use random state while splitting the data to ensure reproducibility and consistency

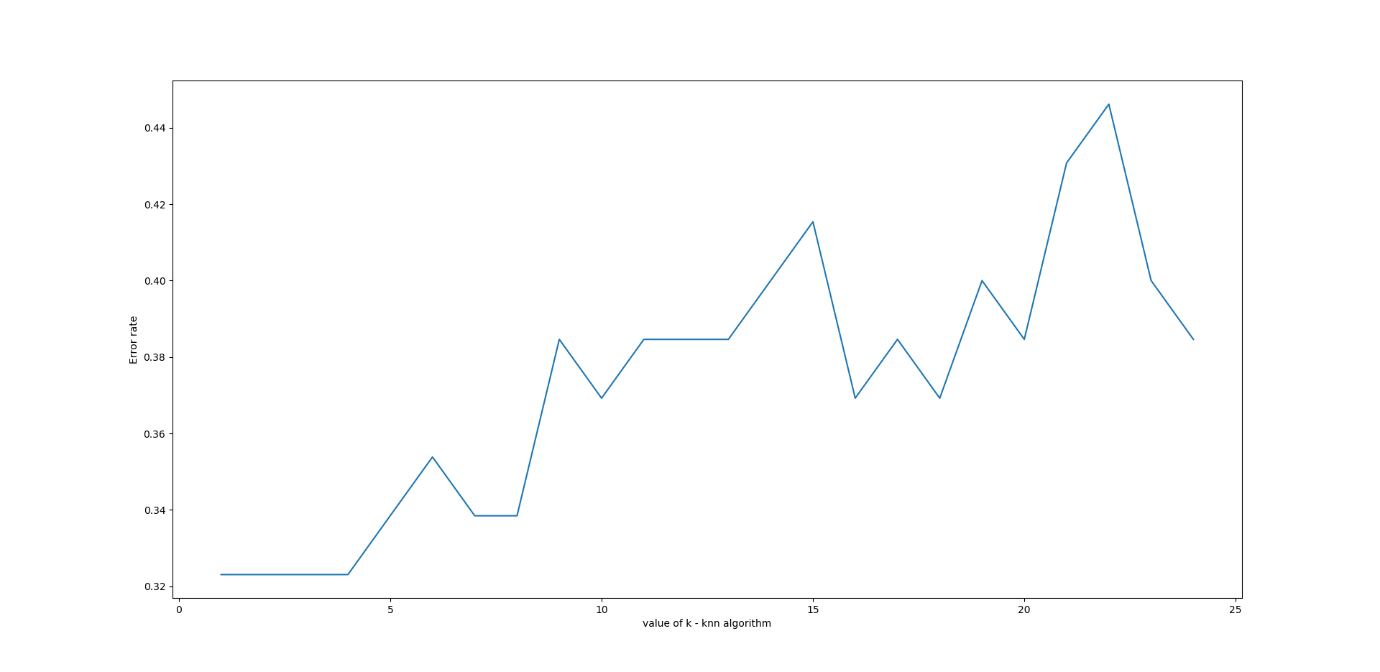
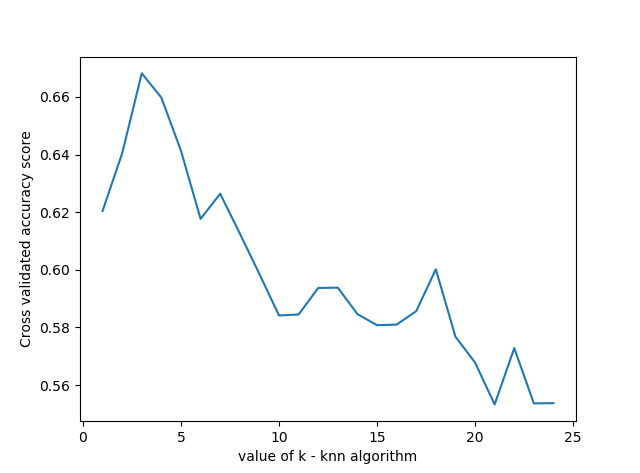
Experiment with distance metrics - Euclidean, Manhattan

By Building the model we got the accuracy score 73.84%

Find the best k value:

=>by plotting accuracy

=>by plotting error rate



we can see that k=4 produces the most accurate results.