Tests were ran to determine how much correlation there was between the volume of stocks (grouped by the industry) and the return. We then attempted to use these results to see if the stocks were increasing or decreasing on average with the hypothesis that a stock that increased over a 20 day period would be more profitable over one that did not increase over the same period.

After testing the correlation between several sets of variables, we did not find any linear correlations that stood out. The most significant observation from this method was found when we grouped the stocks by the industry. We noticed that there was a large variance in the volume of the stocks based on which industry they were in.

Through testing with linear regression, we were unable to get any reliable accuracy with our test data, but we did observe that a stock value against the industry group to visualize the variance in the industries.

Random Forest
- A random forest model was used against our given training data. The model took the industry of the stock as well as the volume and return rates for the past twenty days. Unfortunately due to memory constraints we were only able to train this model based on the first 50,000 rows (stocks).
- The predictions that we were able to generate using our random forest testing were 49.82% accurate to the given results.
- We concluded that this test was not an effective metric for stock prediction due to the accuracy being too low to produce any sort of profit.

Concluding Points
- Although reliable results could not be attained through the use of the random forest method, we concluded that it was the most accurate of our tested methods. All of our testing methods have been ruled out due to the absence of profitable results.
- After many tests, our team consistently found that our resulting accuracy was around 50%. All of these tests were as accurate as choosing randomly (a test that we did look into).
- Through comparison with the other scores of this competition, we observed that all teams were getting an accuracy within 52.04% of the actual result set.
- In the end, our team concluded that there is no sure way to predict the stock market. Some patterns can be observed to see which industries are more popular for investing, but observation of those patterns only yield a 2% window of profit.

We concluded that this test was not an effective metric for stock prediction due to the accuracy being too low to produce any sort of profit.

<table>
<thead>
<tr>
<th>Analysis Strategy</th>
<th>Relevant to Dataset</th>
<th>Viable Prediction Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>KNN</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Random Forest</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

The image above shows the error rates of the random forest and shows that as the forest grows (more trees), the error rate goes down.

The image above shows the correlation of the dates versus the residuals within this specific case. As the dates are randomized, these results are subject to change.

The image above shows the relevance of the industry groupings. We graphed the change in stock value against the industry group to visualize the variance in the industries.

The two images above show the correlation between the overall residual returns and stock volumes sold over the twenty-day period.