Abstract
Modern social media platforms have become influential forums for discussing and disseminating public opinion on topics like the financial markets. Previous literature has extensively demonstrated the impact of investor sentiment on stock price movements. This study seeks to further characterize the underlying structure of the relationship between time series data of equity price metrics and sentiment streams during the pandemic. We aggregated web scraped financial tweets and employed Twitter Sentiment Analysis techniques. We then constructed a network and employed spectral clustering to explore structures within the data. Then, to ascertain the predictive value of our sentiment metrics, we utilized deep learning techniques. We then constructed a network and employed spectral clustering to explore structures within the data. Then, to ascertain the predictive value of our sentiment metrics, we utilized deep learning techniques.

Models and Methods
Structure of Data
To understand the relationship between our Twitter sentiment time series and stock market movements we first conduct a χ² test, where we found statistically significant correlations. To visualize what companies are impacted by sentiment in similar ways, we construct an adjacency matrix using Pearson correlations. We then map this with a Gaussian kernel to network edges. Following Luxberg [4] we apply spectral clustering using the symmetrized Laplacian to this network. We map these clusters to economic sectors with an Adjusted Rand Index (ARI). We additionally consider a rolling window of time frames to reduce the impact of overarching market dynamics. Then, we analyze the entire stock time series in the same manner, applying Principal Components Analysis (PCA) to reduce the dimensionality of the data for computational efficiency.

Spectral Clustering
Structure of Data
Correlation
PS_d,B = \frac{\sum_i (S_d,i - \bar{S}_d)(S_B,i - \bar{S}_B)}{\sqrt{\sum_i (S_d,i - \bar{S}_d)^2} \sqrt{\sum_i (S_B,i - \bar{S}_B)^2}}

Feature Importance
Importance(f) = MSE_{\text{avg}} - MSE_{\text{featureless}}

Machine Learning Analysis
We utilized Long Short-Term Memory (LSTM) neural networks to assess whether Twitter sentiment can predict stock prices for heavily discussed companies. LSTMs are suitable for time series forecasting, long-term dependencies. We trained LSTM models with features from Twitter sentiment scores and traditional stock market metrics to forecast future stock prices.

Results and Analysis
Our sentiment analysis revealed significant variations in public sentiment towards different companies, correlating these with major events and stock price movements. We investigated “sentimental” stocks, where sentiment scores significantly predict price movements, by analyzing the top 50 stocks by tweet count using an LSTM model.

Conclusion
For the time period considered, we found significant correlations between investor sentiment and stock prices, particularly clustered together for tech stocks. The variable outcome of stock market predictions using machine learning suggest company-specific factors influence the effectiveness of sentiment as an indicator. These findings suggest that for sentimental stocks market observers can glean useful price signals.

Future Directions
Our research shows the significant impact of Twitter sentiment on stock prices. We plan to extend our study with a larger tweet dataset and include sentiment data from Reddit, TikTok, Instagram, and Mastodon. We also seek to pair real time stock and options data to better capture price movements due to sentiment. This could lead to the development of behavioral economic models capturing interplay of market forces and human psychology.

References

Analysis of impact of twitter sentiment on stock market dynamics using spectral clustering and deep learning
Rahul Gupta, Vagmin Viswanathan
Department of Mathematics, Dartmouth College