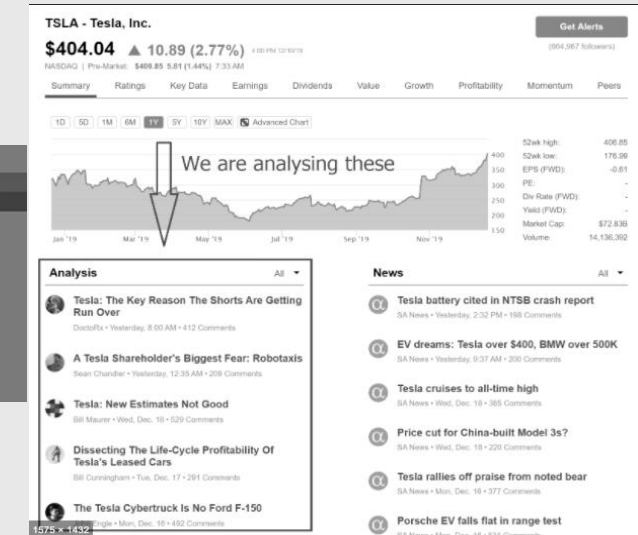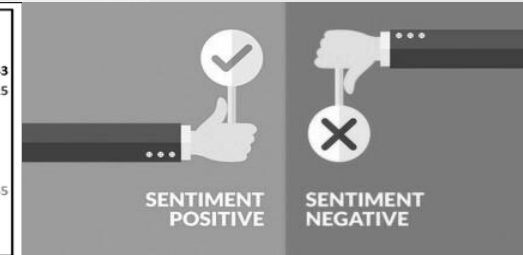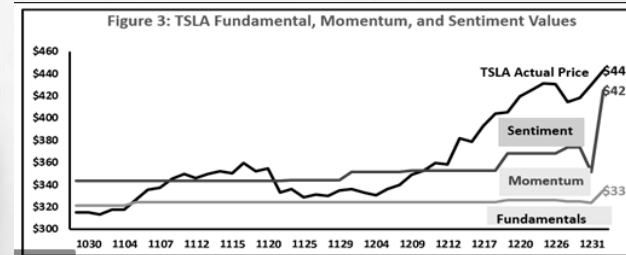# Project Overview:

Analyzing stock data in correlation to sentiments from different stock news sources and determining what (if there is any) impact it has on stock prices.

Project Team: Renae, Heena and Jatinder.

Figure 3: TSLA Fundamental, Momentum, and Sentiment Values

SENTIMENT POSITIVE   SENTIMENT NEGATIVE

We are analysing these

Stock Data:

Ticker: TSLA.

To limit the scope of project and create a working prototype, we decided to look at one stock initially. We picked TSLA as it seems to have more daily news and articles than many other stocks.

Stock prices source: yfinance.

We chose to use yfinance sdk to download historical stock prices as it was fairly easy to use and gave us all we needed in literally two lines of code.

Time period: year 2019 only.

We didn't want to include 2020 as year 2020 has seen some unprecedented times and it would probably go down as an anomaly in the world history.

date	TSLA_close
2019-01-02	310.119995
2019-01-03	300.359985
2019-01-04	317.690002

News/Articles Data:

Stock news sources: stocknewsapi.com; breakingapi.com

We downloaded the TSLA articles from stocknewsapi.com for 2019. However, stocknewsapi.com didn't give us the full articles. It provided news URLS, headlines and snippets though. It also provided it's own sentiment analysis score that we decided to use as a comparison to our own sentiment analysis. To get the full news articles, we wrote a custom function that would take the news url from stocknewsapi.com and pass it to breakingapi.com.

Dataframe created from data downloaded from Stocknewsapi:

	news_url	image_url	title	text	source_name	date	topics	sentiment	type	tickers
285	https://www.fool.com/investing/2019/11/16/what...	https://cdn.snapi.dev/images/v1/f/r/fredw22-1.jpg	What We Know About Tesla's Pickup Truck	The electric-car company will finally unveil t...	The Motley Fool	16 Nov 2019	[]	Neutral	Article	['TSLA']
286	https://www.fool.com/investing/2019/11/15/cons...	https://cdn.snapi.dev/images/v1/j/y/105797044-...	"Consumer Reports" Now Recommends Tesla's Mode...	The news comes as Tesla aims for record delive...	The Motley Fool	15 Nov 2019	[]	Positive	Article	['TSLA']
287	https://www.fool.com/investing/2019/11/15/3-st...	https://cdn.snapi.dev/images/v1/n/r/0101ee.jpg	3 Stocks That Soared This Earnings Season	Don't miss the stories behind these companies'...	The Motley Fool	15 Nov 2019	['earnings']	Positive	Article	['AAPL', 'DIS', 'TSLA']

Modified Dataframe after downloading full text from breakingapi.com:

	news_url	image_url	title	text	source_name	date	topics	sentiment	type	tickers	full_text
285	https://www.fool.com/investing/2019/11/16/what...	https://cdn.snapi.dev/images/v1/f/r/fredw22-1.jpg	What We Know About Tesla's Pickup Truck	The electric-car company will finally unveil t...	The Motley Fool	16 Nov 2019	[]	Neutral	Article	['TSLA']	As if electric-car maker Tesla (NASDAQ:TSLA) i...
286	https://www.fool.com/investing/2019/11/15/cons...	https://cdn.snapi.dev/images/v1/j/y/105797044-...	"Consumer Reports" Now Recommends Tesla's Mode...	The news comes as Tesla aims for record delive...	The Motley Fool	15 Nov 2019	[]	Positive	Article	['TSLA']	In a reversal of a previous decision not to re...
287	https://www.fool.com/investing/2019/11/15/3-st...	https://cdn.snapi.dev/images/v1/n/r/0101ee.jpg	3 Stocks That Soared This Earnings Season	Don't miss the stories behind these companies'...	The Motley Fool	15 Nov 2019	['earnings']	Positive	Article	['AAPL', 'DIS', 'TSLA']	As usual, earnings season brought with it its ...
											Tesla 's (TSLA...

Breakingapi allows to extract the full article by passing on the article link to its API. More info at
https://breakingapi.com/article-extraction-api

Data preparation, cleaning and processing:

Extracting sentiments from TSLA articles: custom code.

Sentiment analysis is a machine learning technique that detects polarity – positive or negative - within text. The models often focus on polarity but also on emotions e.g. angry, happy, sad, etc. or even intentions, interested or not interested. We only looked at polarity, a range of positive to negative, +1 to -1.

Because our focus is on stock market returns, we looked at articles that spoke to a particular stock – Tesla in our case – and decided to see if sentiments of news articles do, in fact, have a bearing on stock prices. We only looked at the articles of 2 sources, The Motley Fool and Zacks Investment Research for the year 2019, 289 articles. Full texts of all 289 articles were processed through 2 different Sentiment Analysis Models:

VADER *(Valence Aware Dictionary and **sEntiment** Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments in social media. This tool gives us 4 rating scores: Positive, Neutral, Negative, and Compound. Positive, Neutral, and Negative are given a ranking score within each document, with a Compound score for total polarity. The Compound score was used in our analysis.*

TextBlob *is a python library for processing textual data. This tool gives us 2 ratings scores: Polarity and Subjectivity. The Subjectivity scored whether the document was objectively or subjectively weighted, while the Polarity score rated the article based on positive or negative sentiment. The Polarity score was used in our analysis. Unlike VADER, input text needs to be preprocessed to tokens, because of this we ran this model twice, we wanted to see if this had any effect on the polarity and the predictive power of the sentiment:*

  a. *Lemmatizer breaks the word down to its root, dropping contractions, plurals, etc.*
  b. *PorterStemmer breaks the word down to its roots as does the Lemmatizer, however stemming is much stricter at stripping the words.*
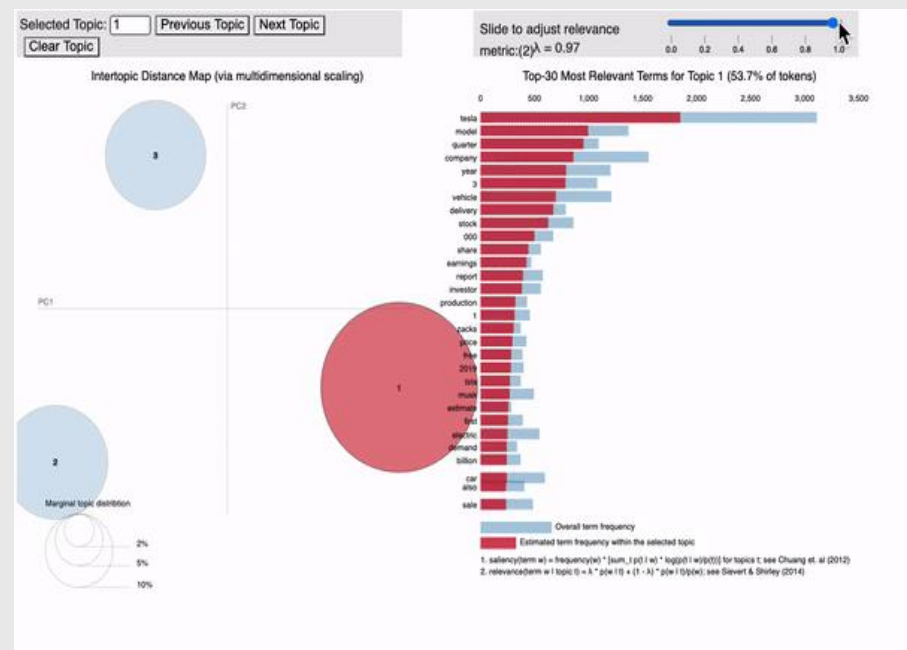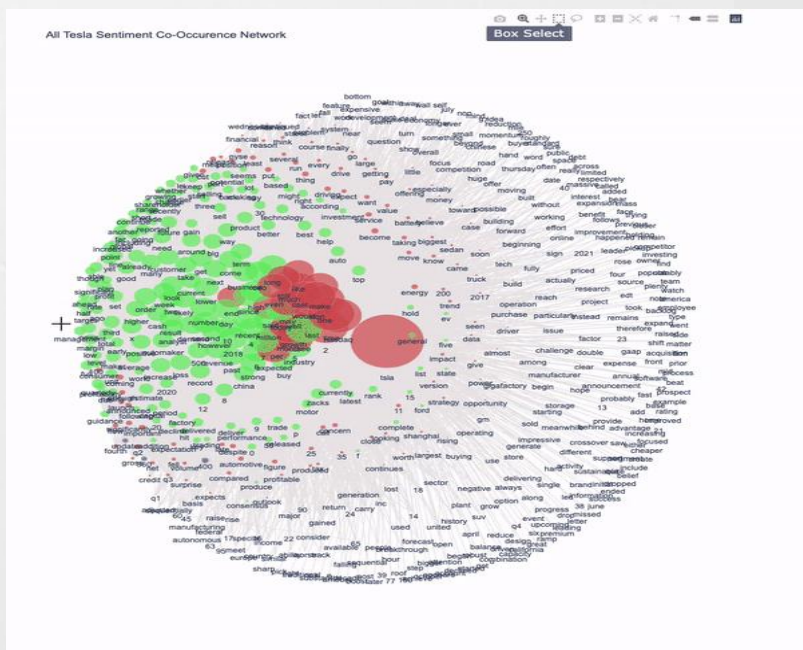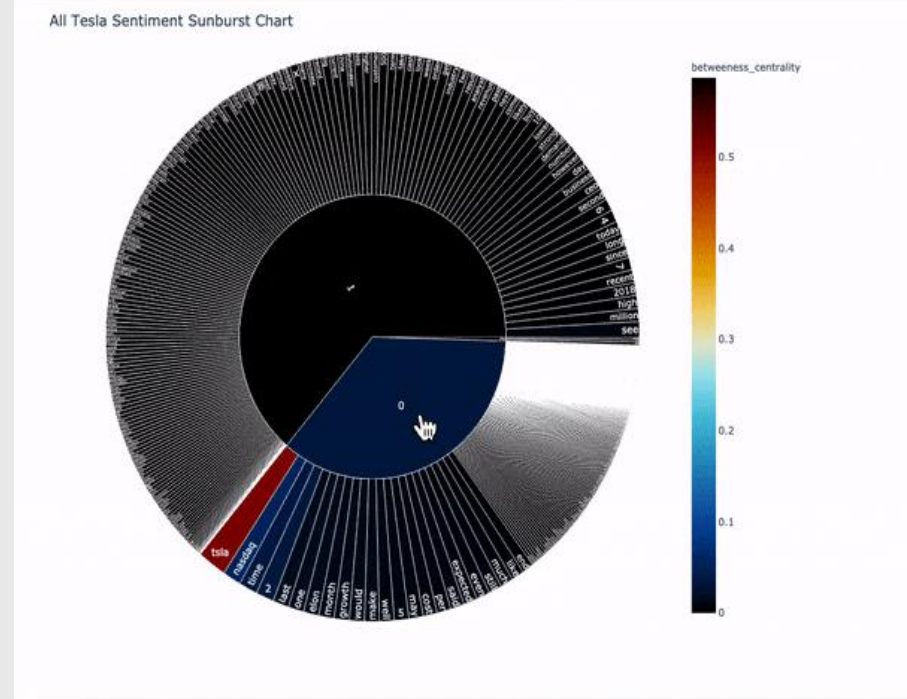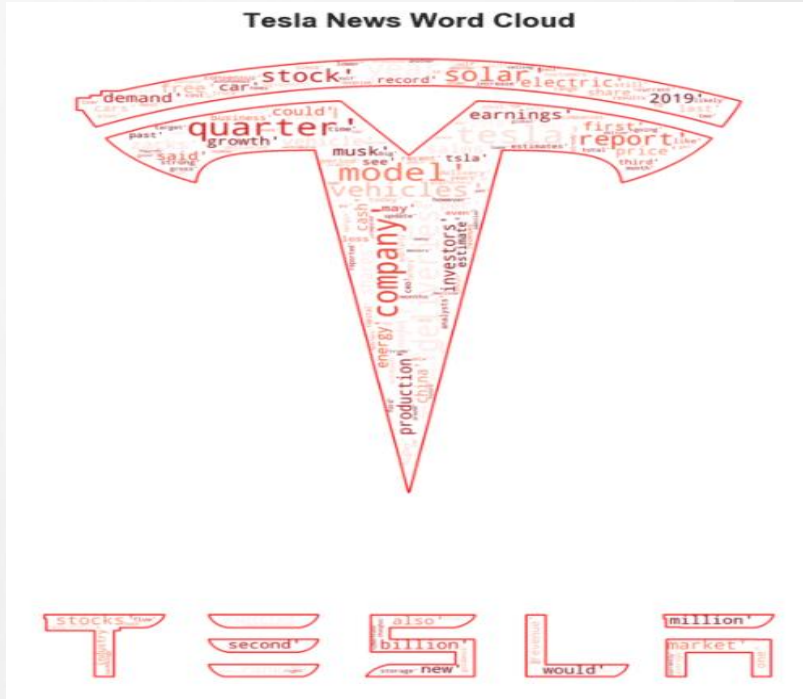
We also used a given polarity index from Stocknewsapi. We took this analysis at face value.
https://medium.com/@Intellica.AI/vader-ibm-watson-or-textblob-which-is-better-for-unsupervised-   sentiment-analysis-db4143a39445
https://towardsdatascience.com/sentiment-analysis-vader-or-textblob-ff25514ac540

Tesla News Word Cloud

All Tesla Sentiment Sunburst Chart

All Tesla Sentiment Co-Occurence Network

Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 1 (53.7% of tokens)

Merging stock prices and stock news sentiment data: custom code.

We created dataframes containing TSLA stock data, sentiment analysis from stocknewsapi and CSV files created by our custom code and joined them to create a final dataframe that was used for regression analysis.
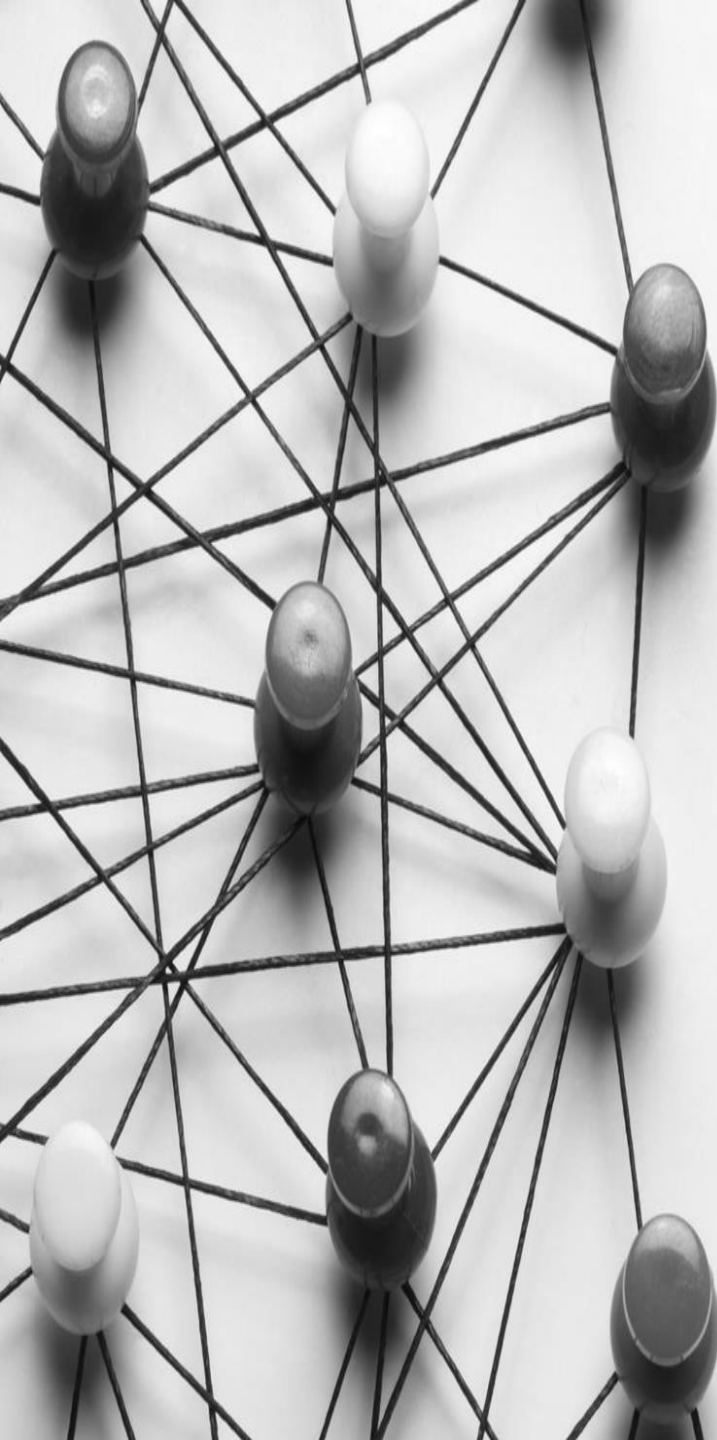
`Stocknewsapi sentiment translated to a numeric value:`

date	title	text	full_text	source_name	stocknews_sentiment	ticker
2019-12-31	Why the Stock Market Soured on Tesla and Centu...	The carmaker's CEO had some Boring news to rep...	Many of us are preparing for the new year and ...	The Motley Fool	-1	TSLA
2019-12-30	Tesla to Start Delivery of Model 3 Built at Ch...	The construction of Gigafactory3 in Shanghai i...	Tesla, Inc. (TSLA - Free Report) will start de...	Zacks Investment Research	0	TSLA
2019-12-30	Tesla (TSLA) Rolls Out First Model 3s Built in...	Tesla Motors (TSLA) remains near the peak of i...	Monday, December 30, 2019\nAdvance Trade in U...	Zacks Investment Research	1	TSLA
2019-12-29	How Many Vehicles Will Tesla Deliver in Q4?	The company is slated to report its quarterly ...	Sometime this week -- probably on Thursday or ...	The Motley Fool	0	TSLA
2019-12-28	Why Your 2020 Resolution Should Be to Buy More...	After a great 2019 capstone to the 2010s, tech...	After a decade of virtually no returns -- than...	The Motley Fool	1	TSLA
...	...	...	...	...	...	...
2019-11-16	What We Know About Tesla's Pickup Truck	The electric-car company will finally unveil t...	As if electric-car maker Tesla (NASDAQ:TSLA) i...	The Motley Fool	0	TSLA
2019-11-15	"Consumer Reports" Now Recommends Tesla's Mode...	The news comes as Tesla aims for record delive...	In a reversal of a previous decision not to re...	The Motley Fool	1	TSLA

`Added sentiments from our custom code to the above dataframe:`

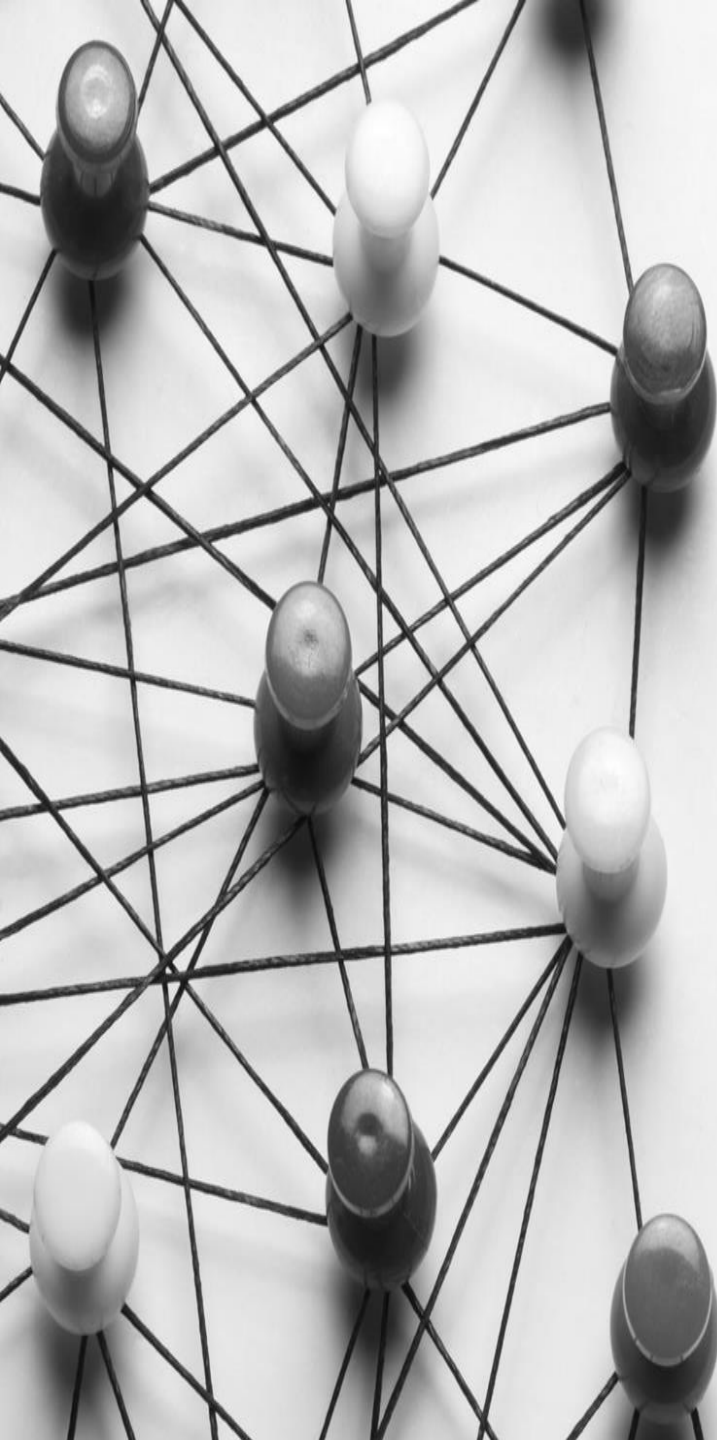| date	title	text	full_text	source_name	stocknews_sentiment	ticker	vader_sentiment	textblob_lemmatized_sentiments	textblob_stemmed_sentiments
2019-03-06	Did Audi Just Disrupt Tesla's Model Y?	Audi's e-tron Q4 Concept is aimed squarely at ...	Audi AG (OTC:AUDV.F) took the wraps off its la...	The Motley Fool	0	TSLA	0.9726	0.069364	0.041566
2019-03-08	Tesla (TSLA) Inks Deal With China Lenders for ...	Tesla (TSLA) is likely to benefit from a wholl...	Tesla, Inc. (TSLA - Free Report) has made the ...	Zacks Investment Research	1	TSLA	0.9928	0.056798	0.097043
2019-03-09	Is This the Final Nail in the Coffin for Tesla...	Elon Musk used to be the solar installer's big...	There were already signs that all was not well...	The Motley Fool	-1	TSLA	0.9965	0.127565	0.095899
2019-03-09	Tesla Has Some Problems	Reading between the lines of the Tesla news ci...	Join Motley Fool analysts Nick Sciple and John...	The Motley Fool	-1	TSLA	0.9999	0.115436	0.107687
2019-03-11	Stock Market Today: Tesla Plans Price Hikes, P...	Meanwhile, Boeing plummeted following a deadly...	Major stock market indexes climbed on Monday d...	The Motley Fool	-1	TSLA	-0.9235	0.063292	0.087270
...	...	...	...	...	...	...	...	...	...
2019-12-28	Why Your 2020 Resolution Should Be to Buy More...	After a great 2019 capstone to the 2010s, tech...	After a decade of virtually no returns -- than...	The Motley Fool	1	TSLA	0.9934	0.100860	0.057399
2019-12-29	How Many Vehicles Will Tesla Deliver in Q4?	The company is slated to report its quarterly ...	Sometime this week -- probably on Thursday or ...	The Motley Fool	0	TSLA	0.9880	0.119706	0.023636
2019-12-30	Tesla (TSLA) Rolls Out First Model 3s Built in...	Tesla Motors (TSLA) remains near the peak of i...	Monday, December 30, 2019\nAdvance Trade in U...	Zacks Investment Research	1	TSLA	-0.8271	0.078244	0.099321
2019-12-30	Tesla to Start Delivery of Model 3 Built at Ch...	The construction of Gigafactory3 in Shanghai i...	Tesla, Inc. (TSLA - Free Report) will start de...	Zacks Investment Research	0	TSLA	0.9961	0.028692	-0.029774
2019-12-31	Why the Stock Market Soured on Tesla and Centu...	The carmaker's CEO had some Boring news to rep...	Many of us are preparing for the new year and ...	The Motley Fool	-1	TSLA	0.9109	0.087333	0.112693

290 rows × 9 columns

Cleaned-up dataframe containing only the columns we needed:

date	ticker	stocknews_sentiment	vader_sentiment	textblob_lemmatized_sentiments	textblob_stemmed_sentiments
2019-03-06	TSLA	0.0	0.9726	0.069364	0.041566
2019-03-08	TSLA	1.0	0.9928	0.056798	0.097043
2019-03-09	TSLA	-1.0	0.9982	0.121500	0.101793
2019-03-11	TSLA	-1.0	-0.9235	0.063292	0.087270
2019-03-12	TSLA	-1.0	-0.9229	0.022327	-0.032525
...	...	...	...	...	...
2019-12-27	TSLA	0.5	0.9886	0.149190	0.081275
2019-12-28	TSLA	1.0	0.9934	0.100860	0.057399
2019-12-29	TSLA	0.0	0.9880	0.119706	0.023636

Merged Tesla stock closing prices for the respective dates:

date	ticker	stocknews_sentiment	vader_sentiment	textblob_lemmatized_sentiments	textblob_stemmed_sentiments	TSLA_close
2019-03-06	TSLA	0.0	0.9726	0.069364	0.041566	276.239990
2019-03-08	TSLA	1.0	0.9928	0.056798	0.097043	284.140015
2019-03-09	TSLA	-1.0	0.9982	0.121500	0.101793	NaN
2019-03-11	TSLA	-1.0	-0.9235	0.063292	0.087270	290.920013
2019-03-12	TSLA	-1.0	-0.9229	0.022327	-0.032525	283.359985
...	...	...	...	...	...	...
2019-12-27	TSLA	0.5	0.9886	0.149190	0.081275	430.380005
2019-12-28	TSLA	1.0	0.9934	0.100860	0.057399	NaN
2019-12-29	TSLA	0.0	0.9880	0.119706	0.023636	NaN
2019-12-30	TSLA	0.5	0.0845	0.053468	0.034774	414.700012
2019-12-31	TSLA	-1.0	0.9109	0.087333	0.112693	418.329987

***This was our Dataframe number 1 for regression analysis – grouped by date only**

Dataframe grouped by date and source:

date	ticker	source_name	stocknews_sentiment	vader_sentiment	textblob_lemmatized_sentiments	textblob_stemmed_sentiments
2019-03-06	TSLA	Motley	0.0	0.9726	0.069364	0.041566
2019-03-08	TSLA	Zacks	1.0	0.9928	0.056798	0.097043
2019-03-09	TSLA	Motley	-1.0	0.9982	0.121500	0.101793
2019-03-11	TSLA	Motley	-1.0	-0.9235	0.063292	0.087270
2019-03-12	TSLA	Motley	-1.0	-0.9229	0.022327	-0.032525
...	...	...	...	...	...	...
2019-12-27	TSLA	Zacks	1.0	0.9938	0.151768	0.110277
2019-12-28	TSLA	Motley	1.0	0.9934	0.100860	0.057399
2019-12-29	TSLA	Motley	0.0	0.9880	0.119706	0.023636
2019-12-30	TSLA	Zacks	0.5	0.0845	0.053468	0.034774
2019-12-31	TSLA	Motley	-1.0	0.9109	0.087333	0.112693

Source_name values transposed to columns:

date	ticker	TSLA_close	TSLA_close_shifted_by_1	zacks_stocknews_sentiment	zacks_vader_sentiment	zacks_textblob_lemmatized_sentiments	zacks_textblob_stemmed_sentiments	motley_stocknews_sentiment	motley_vader_sentiment	motley_textblob_lemmatized_sentiments	motley_textblob_stemmed_sentiments
2019-04-02	TSLA	285.880005	289.179993	-1.000000	0.996800	-0.069765	-0.154898	0.000000	-0.40370	0.077399	0.064030
2019-04-04	TSLA	267.779999	291.809998	-0.333333	-0.255867	0.025594	0.042549	-0.333333	0.11810	0.075108	0.117540
2019-04-11	TSLA	268.420013	276.059998	0.000000	0.610100	0.045378	0.041062	-1.000000	0.47315	0.039020	0.055446

This was our Dataframe number 2 for regression analysis – grouped by date and source.

Regression Analysis: custom code.

Sentiment Analysis Data Preparation
- News articles for Tesla were not available for every trading day
- Filled missing sentiment analysis with hysteresis prior to regression analysis

Monte Carlo Regression Analysis:
A monte carlo analysis was performed on 2 sentiment analysis data frames (All Sentiments and Sentiments by News Source) to determine which feature (sentiment analysis) has the best correlation to Telsa's daily returns. The analysis was performed using two different monte carlo functions which iterated over 20 days.
1) **Model 1: Lasso Regression**
    1) https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b
2) **Model 2: Lasso Regression using LogisticRegression**
    1) https://towardsdatascience.com/feature-selection-using-regularisation-a3678b71e499

➤ **Model 1 Monte Carlo Results- All Sentiments:**
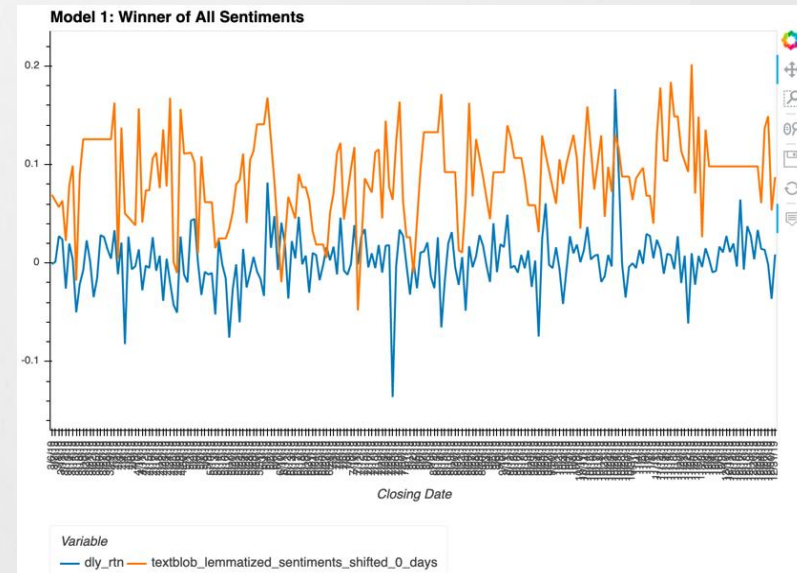
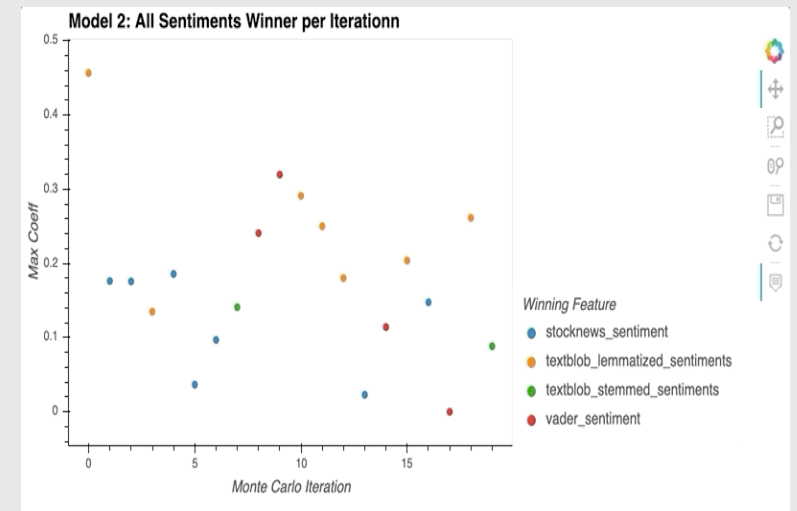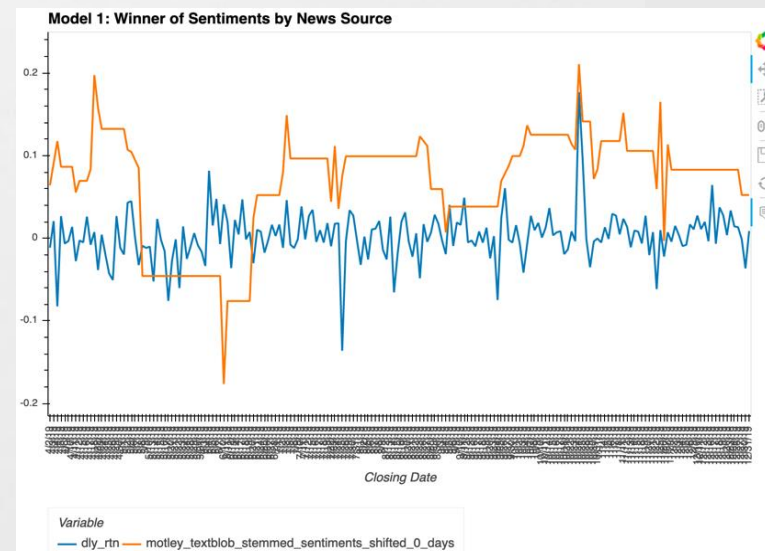*Model 1 Lasso Regression*   *Model 1 Lasso Regression using LogisticRegression*
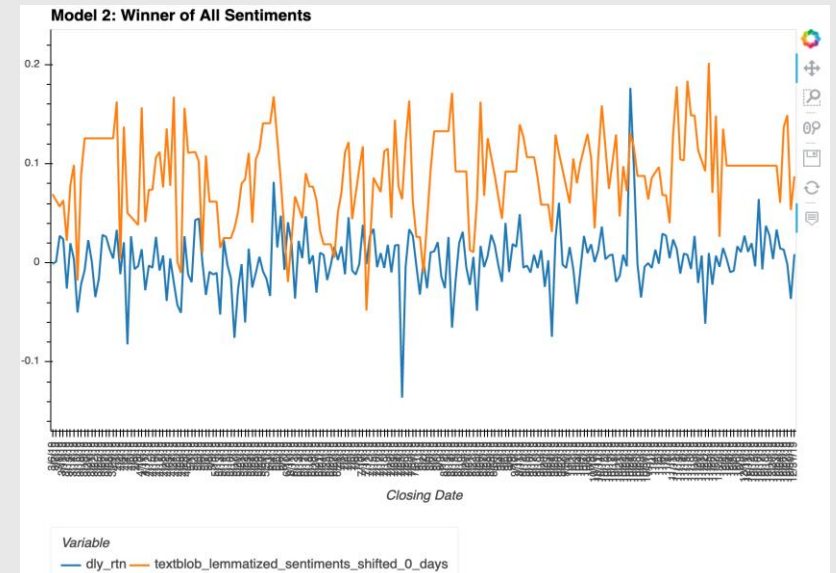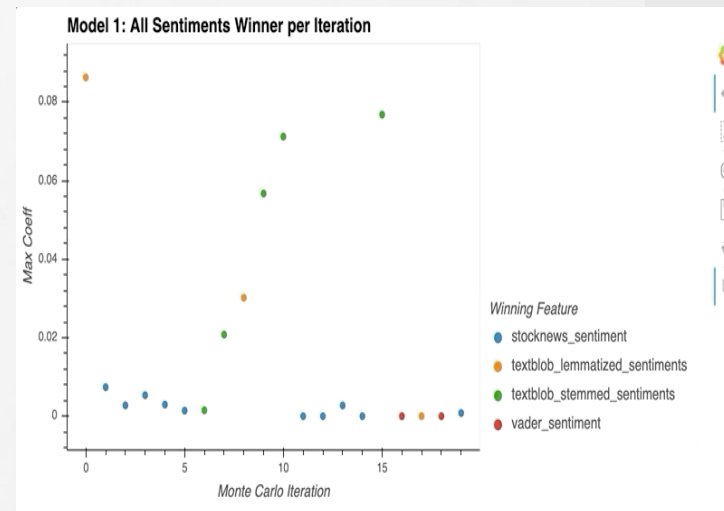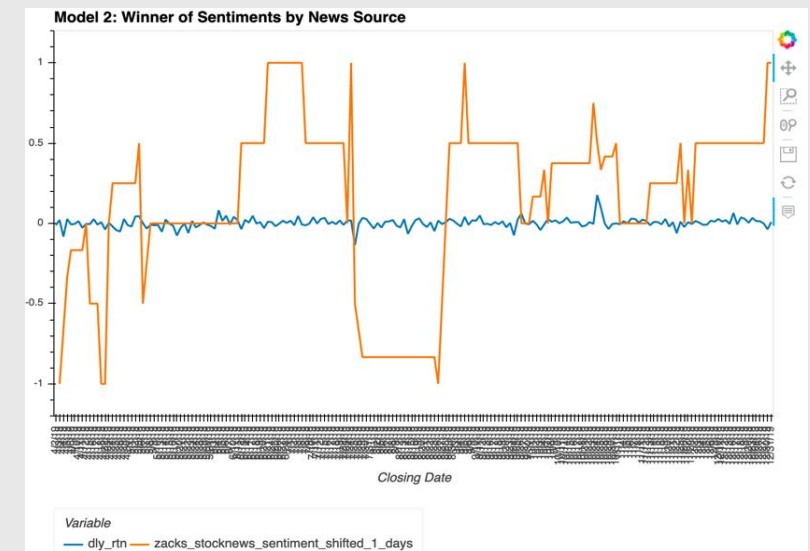
Model 1 Lasso Regression
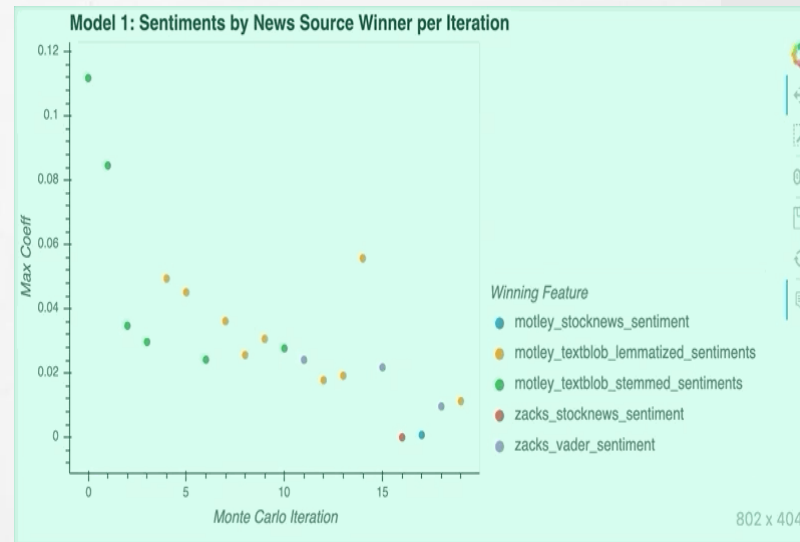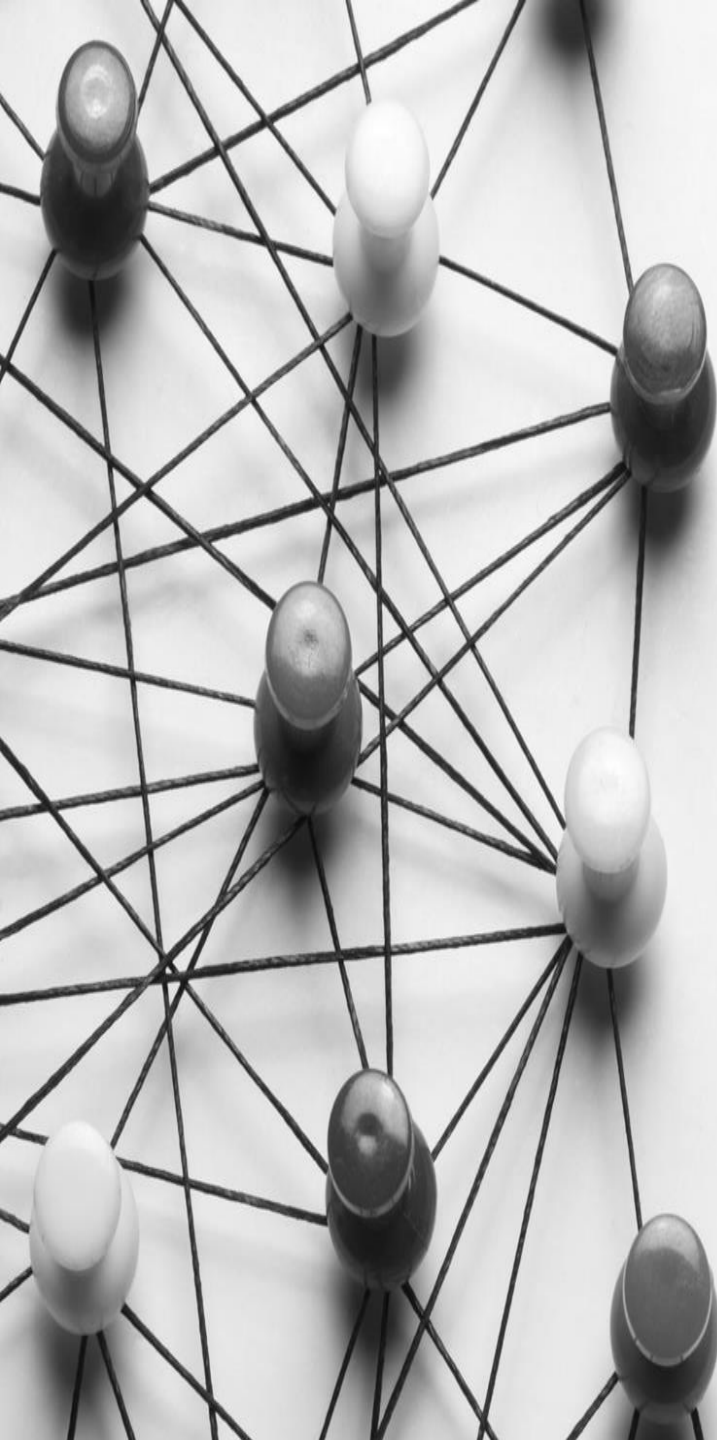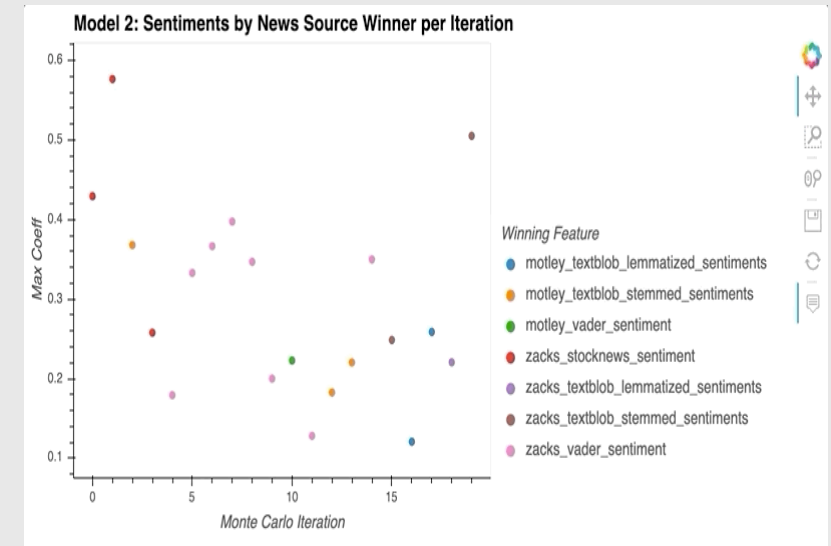
Model 1 Lasso Regression using LogisticRegression

> **Model 2 Monte Carlo Results- Sentiments by News Source:**

Model 2 Lasso Regression

Model 2 Lasso Regression using LogisticRegression

Model 2 Lasso Regression
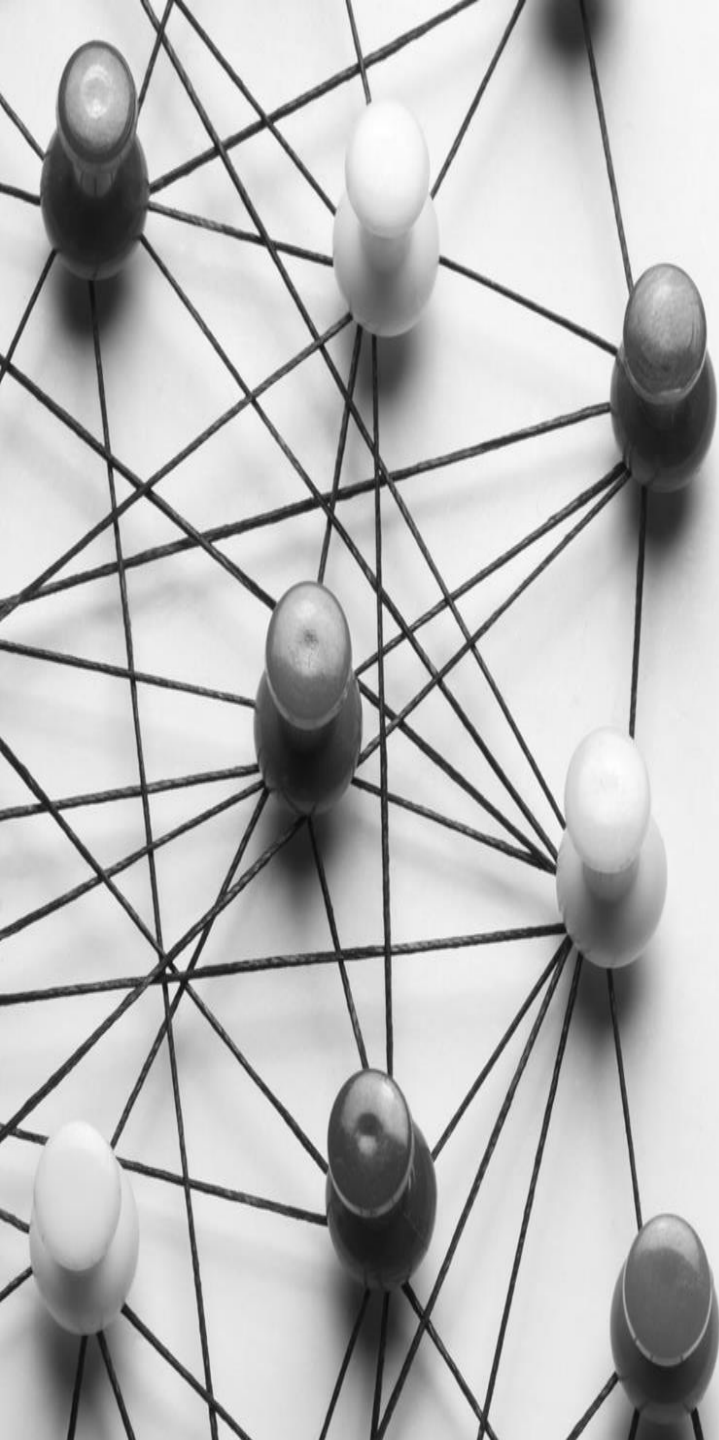

Model 2 Lasso Regression using LogisticRegression

Conclusion:

The charts above identify the winning feature (sentiment analysis) per iteration for each of the monte carlos performed. From analyzing the data, we confirmed that the overall winner shown on the hvplots isn't necessarily the feature that won the most times over the 20 day iteration. For the All Sentiment data frame, the textblob sentiment analysis has the most correlation to Tesla's stock returns. For the Sentiment by News Source data frame, Model 1 predicts that Motley Fool has the best correlation to Tesla's stock returns, whereas Model 2 predicts Zacks has the best correlation.

Enhancement Opportunities:
o   *Run analysis over a longer period of time (current analysis is only for 9 months.*
o   *Data mine for additional articles on stock for more sentiment analysis data.*
o   *Further customization of sentiment analysis on given article.*
    ▪   *Determine if article was written for Tesla vs Tesla being mentioned in the article.*
o   *Integrate volume of sales data and sell 'type' (industry vs retail) into regression analysis.*