

The Rise of Tesla: Analyzing the Effects of Sentiment on Stock Price (EDA)

Eustina Kim¹, Jennifer Lin², Ritesh Pendekanti², Diana Pham¹, Cassandra Tai¹, and Wanxin Xie²

¹Vivian Lew: Department of Statistics, UCLA, Los Angeles, CA 90095, USA

²Dave Zes: Department of Statistics, UCLA, Los Angeles, CA 90095, USA

November 23, 2020

1 Introduction

In the past few years, Tesla has revolutionized the automotive industry with their electric vehicles. Tesla is one of the fastest growing companies and has recently made its way into the S&P 500, one of the most reputable equity indices in the world. The company's stock price has been on an impressive bull run since mid-late 2019, with the world uncertain of when it will stop.



Figure 1: Tesla's stock price since 2015. Mid-late 2019 we can see Tesla begin its incredible bull run.

2 Project

For our project, we want to understand the driving force of Tesla Motor's emergence into today's top companies. Therefore, we came up with the idea to investigate public sentiment of Tesla, particularly from Reddit, Twitter, and Google. Sentiment in this context refers to a numeric score that indicates how positive or negative a body of text from the sources are on a given day. Being able to understand the relationship between these sentiment and stock prices would provide valuable insights for any type of investor.

3 Sentimental Calculation

To calculate the sentiment of a body of text, we used the popular Python library TextBlob. In specific, we utilized the senti-

ment polarity function which scores a sentence between -1 and 1, with -1 corresponding to most negative sentiment and 1 as the most positive sentiment. We took the average of each sentence's sentimental score to calculate the body of text's final sentiment. To ensure TextBlob worked as intended, we created Word Clouds of the most frequently used words for both positive and negative sentiment bodies of text.

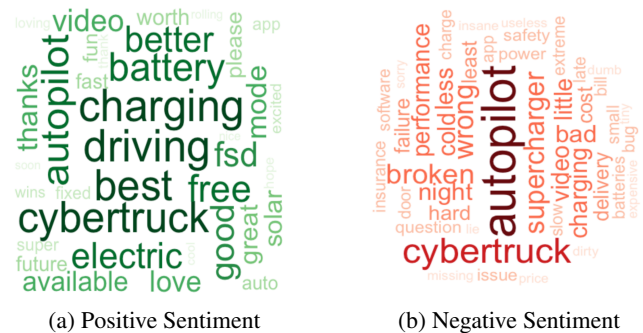


Figure 2: Word clouds of Reddit Posts.

We can see TextBlob works as intended because we can see positive sentiment words such as “love”, “super”, and “great” and negative sentiment words such as “broken”, “failure”, and “slow”.

4 Reddit EDA

Reddit is a social news website and forum that allows users to post articles and opinions on any topic. We are interested in Reddit sentiment because it is one of the most used sites ever created and has public opinion on everything happening in the world.

We used Reddit's public API to collect all posts and comments from Tesla related subreddits (i.e. r/TeslaMotors). Alternatively, we could have collected any post and comment that mentioned the word “Tesla”, however this method would result in many unrelated posts/comments appearing in our dataset. We prioritize quality over quantity of data, so we just collected directly from the Tesla related subreddits.

Reddit's API returned a large JSON, with each en-

try containing 85 attributes, all of which can be found at <https://pushshift.io/api-parameters/>. Since our project is focused on sentimental analysis, we reduced the dimensions by only saving the following attributes for our working dataset:

Attribute	Description
author	Author of Post/Comment
created_utc	UTC Timestamp of Post/Comment
id	Unique ID of Post/Comment
score	Sum of upvotes of Post/Comment
subreddit	Subreddit of Post/Comment
subreddit_id	Unique ID of Post/Comment
title	Text content (Post Only)
total_awards	Number of awards (Post Only)
sentiment	Sentiment of post or body of content

In total, we collected approximately 2,000 posts and over 200,000 comments from Reddit. From here, we wanted to visualize the relationship between daily Reddit sentiment and Tesla stock price. We recognized that some posts/comments are more valuable than others (higher Reddit score). Therefore, we create a new column called weighted sentiment that is simply the sentiment multiplied by the score. Then, we get the daily sentiment by calculating the average of weighted sentiment each day. This results in the following plot:

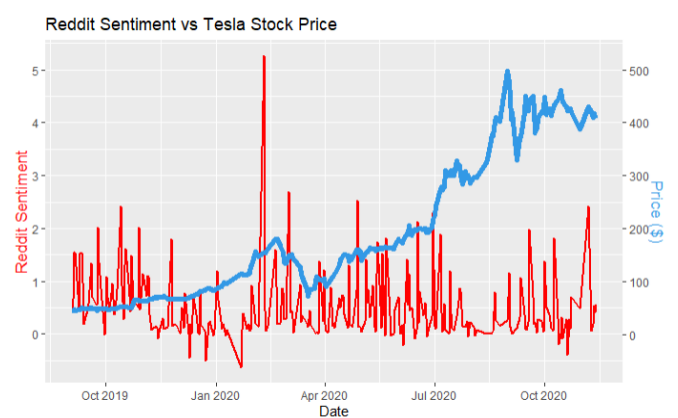


Figure 3: Weighted Reddit sentiment versus Tesla’s stock price.

One thing we found interesting in this graph is that before most of the bull runs, we see there is a significant spike in Reddit sentiment a few days before. This is a research topic we will likely explore.

5 Google EDA

For Google sentiment, we decided to use Google Trends data provided by their public API. Working with Google Trend data was much simpler because Google has their own scoring metric to indicate public sentiment. Each day, Google scores how popular a particular search term is between 0 to 100, with 100 correlating to highest traffic.

From here, we simply plot Google Trend data versus Tesla Stock price.

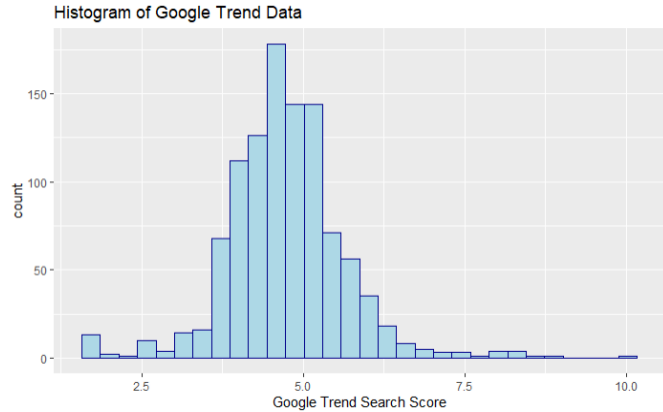


Figure 4: Normalized histogram of Google Trend data.

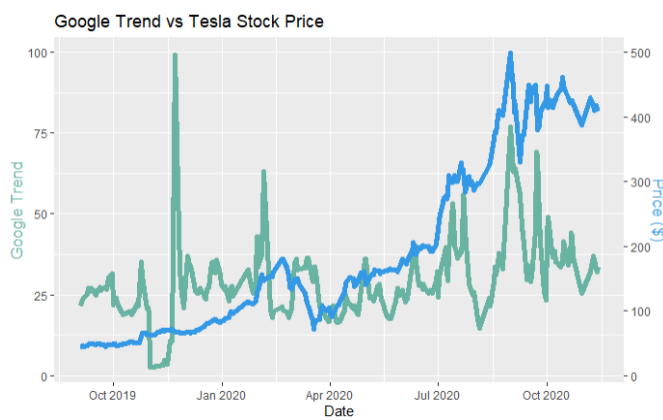


Figure 5: Google Trend versus Tesla’s stock price.

We notice that there is some pretty high correlation between stock price and Google Trends. This is another research topic we will likely explore. It is important to note that there is no sentiment score associated with Google Trends. This means we need to be careful in how we analyze it later because high amount of searches can be good or bad.

6 Potential Research Questions

1. Given weekly sentiment, can we predict if the future stock price will be bull or bear?
2. Given stock price, can we predict if Tesla google searches will go up or down?
3. What particular events (if any) triggered Tesla to begin and continue its bull run?
4. Will adding Twitter sentiment contribute in predicting future stock prices?

For question 4, we want to note that at the time of this report, we were not able to collect Twitter data. The Twitter API was complicated to work with because of API limit rates but we hope to add it to our final poster.