Sentiment and Time Series Analysis for

Amazon and Walmart

CIS 4680.01

Dr. Mehrdad Koohikamali

Roy Amador, Eric Arzoumanians, Sherleen Lee, Keyu Chen

Abstract

The Covid-19 pandemic has significantly impacted the world, including two of the world's largest retailers outside of China, Amazon and Walmart (Weise & Corkery, 2021). With Covid-19-related restrictions, consumers are changing the way they shop so companies adapt accordingly. We analyze their sentiments with sentiment analysis using VADER lexicon, which gives us compound scores representing the different kinds of sentiments that people have towards the two companies. To forecast, we used general forecasting and the Holt-Winters Exponential Smoothing model on stock prices obtained from Yahoo Finance.

Introduction

In examining the text data of Amazon and Walmart, we discover how the Covid-19 pandemic has changed the way people shop. Walmart is the world's largest brick-and-mortar retailer and has the fastest growing e-commerce site. Comparatively, Amazon, an online retailer, has been bigger than Walmart since the pandemic (Weise & Corkery, 2021). We want to know how people perceive them before and after the pandemic through news about stock prices. With sentiment analysis, we were able to predict the sentiments of the two companies using data from Finviz, a stock market and data site. We obtained news tables (text data) for October 2021 of both Amazon and Walmart. The news tables contain links to various news articles. Then, we used the VADER lexicon to analyze sentiments in the news. We then forecasted stock prices using stock prices data from Yahoo Finance from the past five years.

Background and Prior Research

We used multiple peer reviewed articles for some background on our chosen companies. To begin with, we discovered the corporate social responsibility that comes with being as big as Amazon and Walmart. The globalization of these companies has had a great effect on their corporate social responsibility because entry into more global markets causes an increase in influences from different political forces (Ertem-Eray, P3). Since corporate social responsibility pertains to a company's duty towards the communities they affect, the entry into new communities causes many changes in this aspect. In our next article, we studied how the pandemic generally affected companies with the magnitude of Amazon and Walmart. We found that "Amazon's sales were up 40 percent in the second fiscal quarter of 2020 alone" (2020, August) and "Walmart US sales increased by 10 percent during the pandemic" (2020, August). As we suspected, this confirmed the idea that these companies increased their performance during the pandemic. This gave us some definite facts to support our expected outcome and gave us a place to start. Next, we analyzed whether we could expect similar growth in countries outside the US since there were many unique methods by countries in terms of handling the pandemic. In this article it stated that, "Moreover, social distancing and staying home are further expected to drive customers towards online shopping after the COVID-19 pandemic. The e-commerce industry, however, may be impacted by volatile market demand and supply chain problems" (Memon, 2020). This highlighted that although there has been an increase in online shopping due to the pandemic, there were some obstacles that had to be overcome. The pandemic has created supply chain issues in various industries that many companies were unable to fulfill the increased demand regardless. Other issues were found in our next article including, "product availability; (ii) logistics and transportation disruptions; and (iii) consumer protection" (Alfonso et al., 2021). Transportation disruption affects our companies' ability to get resources needed for products and deliver the products to consumers. Our final article gave us most data

driven insights into how the pandemic affected e-commerce. This article also highlighted the "'catching-up' process in e-commerce growth among countries" (Alfonso, 2020). Countries that had below average e-commerce numbers pre-pandemic saw the biggest increases in activity during the pandemic. It was focused on Amazon and "it is shown both from the graph, and the correlation calculation from Gretl stock prices are going in line with the net sales of Amazon." (Işık, 2021). Amazon's net sales increased greatly due to customers' reluctance to go in stores during the pandemic, which led to an increase in Amazon's stock price since the two metrics were proven to be directly correlated. These articles gave us valuable insight into how the pandemic was affecting our companies and allowed us to continue our analysis from there.

Sentiment Analysis

Methodology and Dataset

We obtained the datasets from Finviz, a stock market and data website. The datasets contain headlines and links to news articles that mention the companies during the day. These articles are primarily published by reputable sources such as Bloomberg and Reuters. For simplicity, we are using the date, time, and headlines (title) in our sentiment analysis.

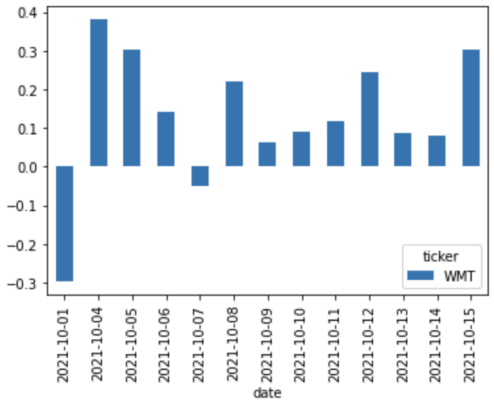
Firstly, we imported several libraries and all rows of the dataset. Then, we normalized the articles into a list (list_of_data). We proceeded with logistic regression and plotted the results on a line graph. We built a dataframe with headings 'ticker', 'date', 'time', and 'title'. We used the Sentiment Intensity Analyzer with VADER to check the compound scores for each ticker we assigned before, and created a bar chart for the compound scores for each date. Compound scores represent different sentiments. Compound scores over 0.05 are considered as the positive sentiment, between 0.05 and -0.05 are considered neutral, and scores less than -0.05 are negative sentiments. We applied the same methods to both notebooks.

Analysis and Results

Based on the compound scores of the sentiment analysis, we created bar charts for Amazon and Walmart. For Walmart, we observed a negative sentiment on October 1, 2021, but it is followed by a generally positive sentiment. Meanwhile, Amazon's analysis shows a positive sentiment, with all compound values scoring more than 0.05 for October 12 through October 16. Most scores appeared positive, indicating that people generally perceive the two retailers positively.

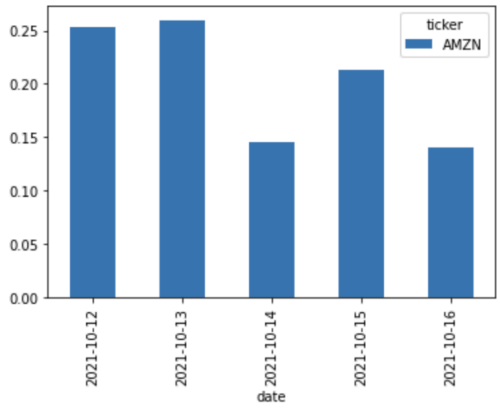
Walmart

```
[ ] mean_df.plot(kind='bar')
    plt.show()
```



Amazon

```
[ ] mean_df.plot(kind='bar')
    plt.show()
```



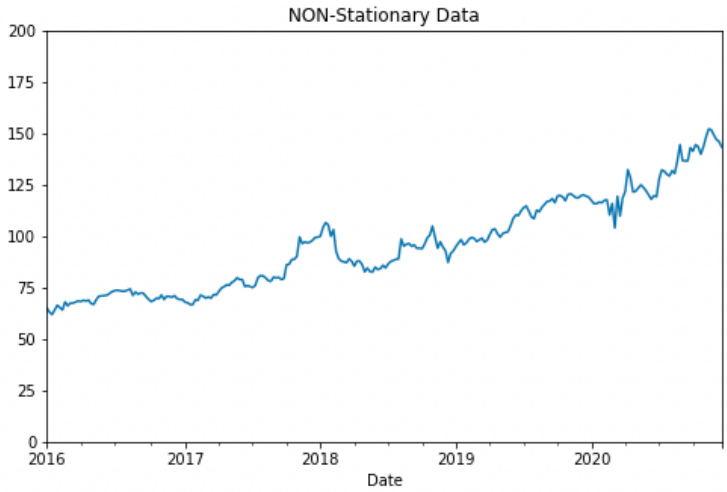
Time Series Analysis

Methodology and Dataset

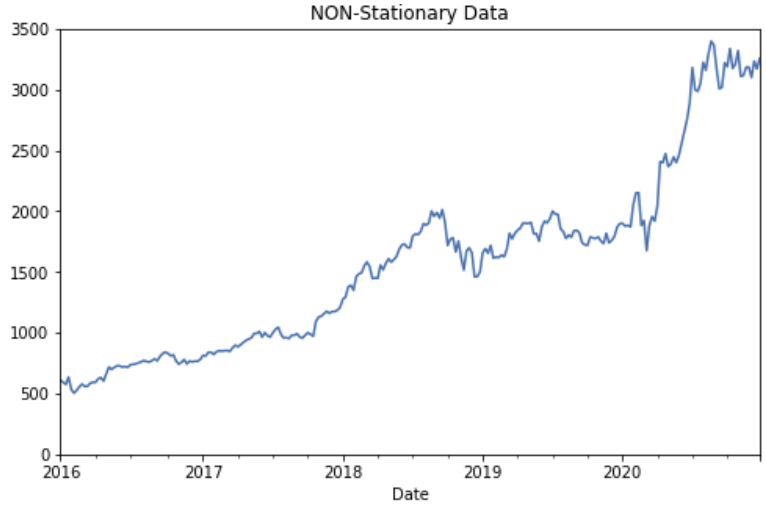
We downloaded the historical stock price of Amazon and Walmart from the past five years (2016 - 2021) from Yahoo Finance. The frequency of the time in the data is weekly, and we used it to predict the stock price of the next four weeks for the business.

We used the ARIMA model for prediction. First, we cleaned out the data to make sure there is no missing or duplicated one. And we checked if the data is stationarity based on our chart. We found our data to be non-stationarity. After cleaning, we created the ARIMA model with the order 5,1,0, and created a fitting model for fit in our data. Then we split our data into training and test dataset, and we used the model created from the last step to fit in the training data, and by using that, we got a new dataframe with the predicted stock price for the test dataset.

Walmart:



Amazon:

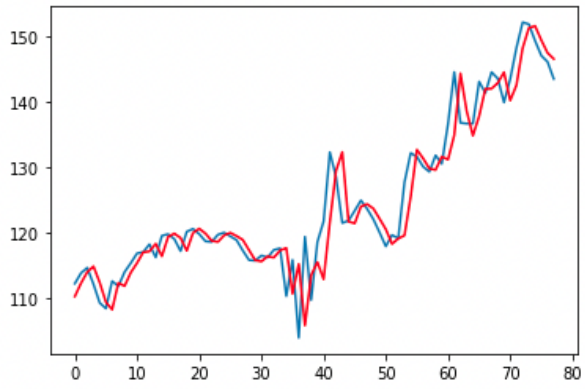


Analysis and Results

We created a chart and plotted the predicted stock price and the actual stock price, and we can see they are very similar; also, we evaluated and got the metrics of our model with a high R^2 score and a small root mean squared error (RMSE).

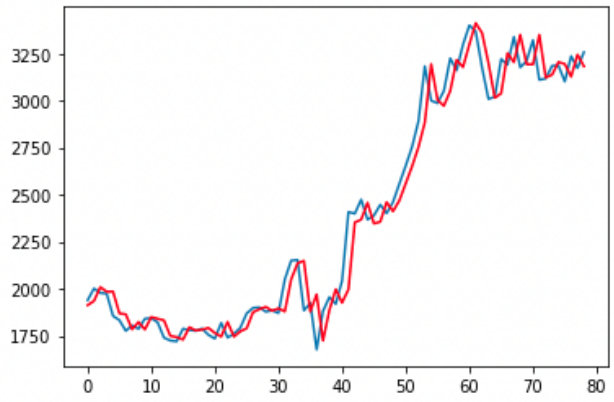
Walmart

```
RMSE:  4.047705662500223
R2:  0.883045736696225
```



Amazon

```
RMSE:  114.2978988070329
R2:  0.9644554229562872
```



This means that for Walmart, 88.30% of variance of the Stock Price is explained by our model, and for Amazon, 96.45% of variance of the Stock Price is explained by our model. RMSE of 4.05 for Walmart shows that there can be residuals of ±$4.05 around the best fitted value. In other words, our predicted Stock Price can be $4.05 away from the actual Stock Price. Similarly for Amazon, the RMSE 114.30 shows that there can be residuals of ±$114.30 around the best fitted value, and our predicted Stock Price can be $114.30 away from the actual Stock Price. Although the RMSE of 114.30 for Amazon seems to be large in comparison to the 4.05 that we got for Walmart, we must take into consideration that the Stock Price for Amazon is in the thousands and the Stock Price for Walmart is in the hundreds.

After testing the model, we predicted the following weeks' stock price in U.S. dollars for the two retailers and got final results as follows.

	Walmart	**Amazon**
2020-12-25	146.54	3182.51
2021-01-01	146.68	3190.35
2021-01-08	146.74	3211.21
2021-01-15	146.59	3211.24

Discussion

The headlines and the stock prices are strongly correlated. When stock prices drop, it is likely that negative sentiments are on the headlines, and vice versa. Our work can further be extended by looking at the news outlets that published the articles, and if a specific news outlet has a consistently positive, negative, or neutral sentiment on the company.

A limitation is that the dataset seems to only have news articles from established sources. Authors may not have the complete freedom to express their thoughts on the companies. Moreover, there are multiple platforms where people discuss a company and its stocks. We can use more sources in the future to reflect people's opinions more candidly. Additionally, the VADER lexicon is not designed to analyze financial jargon, but is attuned to sentiments expressed in social media. Hence, it may not be the best for analyzing news articles from official sources but is better for analyzing texts from sites like Twitter. Another limitation is that it seems like there is a limit on the amount of articles shown in Finviz, which explains the difference in the number of records in our analysis.

Looking ahead, we can compare our findings from the sentiment and time series analysis to further analyze their relationships. These models can give us continued insight into how these companies have transitioned out of the pandemic as many of the restrictions at the beginning of the pandemic are being lifted. This information and changes in sentiment can be compared to some of the past data and will give us information about the future for these companies. A possible shortcoming of our forecasting model is that it predicts stock prices of the companies on public holidays when the stock market is closed. This may cause anomalies in our data compared to what is actually happening. We could have remedied this with the use of exogenous variables with predetermined values that will keep the model accurate and more comparable to what is happening in reality.

Conclusion

In conclusion, we found mostly positive sentiments towards the two companies. Of course, due to the volatile nature of the pandemic period, there were some outlying days where the sentiment was negative. These days were few and far between due to the ability of both companies to adapt to changing conditions and keep their businesses profitable. The shift to e-commerce helped but, this increased demand and unpredictable pandemic conditions created a

new set of problems that needed to be handled in order to stay profitable. They made necessary adjustments while keeping their core principles intact which allowed for them to continue prospering through this time period. However, there are critical limitations in the dataset that can severely impact the accuracy of our results. Comparing the two companies before and after the pandemic means that we need headlines or titles dating from 2019 to 2021. Limitations in our sentiment analysis include the writers' freedom to express their true thoughts in their articles due to pressure from the company they work for and also shortcomings of Vader itself. Limitations are expected in such analysis but it is important to acknowledge them when coming to conclusions. Despite these shortcomings we were able to predict future stock prices and accurately analyze general sentiment towards each of the companies.

References

Weise, K., & Corkery, M. (2021, August 17). People Now Spend More at Amazon Than at Walmart. *The New York Times*. Retrieved October 12, 2021 from https://www.nytimes.com/2021/08/17/technology/amazon-walmart.html

Alfonso, V., Boar , C., Frost , J., Gambacorta, L., & Liu, J. (2021, January 12). *Online appendix - e-commerce in the pandemic and beyond*. BIS. Retrieved October 18, 2021, from https://www.bis.org/publ/bisbull36_appendix.pdf.

Ertem-Eray, T. (2020, October 16). *Addressing corporate social responsibility in corporations: a content analysis of Amazon's and Walmart's websites*. Emerald Insight . Retrieved October 17, 2021, from https://www-emerald-com.proxy.library.cpp.edu/insight/content/doi/10.1108/CCIJ-03-2020-0060/full/html.

Işık, S., İbiş, H., & Gulseven, O. (2021, January 16). *The impact of the COVID-19 pandemic on Amazon's business*. SSRN. Retrieved October 18, 2021, from https://poseidon01.ssrn.com/delivery.php?ID=056112118000093005127098118001091122118082063037061028081121126073113006086004094066101022100000122017012003111108079121093112061072021040019083119124029077067072098060085038118101001014071069071084092096124087084114005083122015076089002122122001016009&EXT=pdf&INDEX=TRUE.

Memon, S., & Kathar, G. N. (n.d.). *E-commerce and covid-19 pandemic in ... - excelpublication.com*. Excel Publication. Retrieved October 17, 2021, from http://excelpublication.com/wp-content/uploads/2021/08/Dr.-Ganesh-C.A-Sohel-Memon.pdf.

PANDEMIC PROFITEERS. Americans for Tax Fairness. (2020, August). Retrieved October 12, 2021, from https://americansfortaxfairness.org/wp-content/uploads/For-the-Many-ATF-CPD-NJ-Pandemic-Profiteers-Brief-August-2020-FINAL.pdf.