Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name of Variables** | | **Explanation** | | **Source** |
| **Date** | | Highlights the weekdays, strictly adhering to stock market operational days, spanning from April 30, 2013, to April 26, 2023. | | Yahoo Finance |
| **Open, High, Low, Close, Adj Close** | | These are conventional stock price indicators illustrating daily stock dynamics. | | Yahoo Finance |
| **Volume** | | Represents the quantity of shares exchanged on a particular day. | | Yahoo Finance |
| **Ticker** | | Identifiers for NASDAQ-listed firms, encapsulating a broad array of companies from diverse sectors. | | Yahoo Finance |
| **Sentiment\_Score** | | Quantifies the sentiment, derived as an average from all pertinent news articles associated with the respective company on a specific day. The score oscillates between -1 (denoting negative sentiment) and 1 (representing positive sentiment). | | RavenPack Analytics |
| **New\_Covid\_Cases** | | Chronicles the daily count of fresh COVID-19 cases reported in the US. | | Our World in Data |
| **MarketCap** | | Portrays the market capitalization of the corresponding company on the given date. | | Yahoo Finance |
| **PandemicPhase** | | A binary delineation demarcating the periods before and after the World Health Organization's official declaration of the pandemic on March 11, 2020. | | World Health Organization |
| **Volatility** | | Evaluated as the rolling standard deviation across the past 5 days of returns. | | Yahoo Finance |
| **Year, Month, Weekday** | | Time-oriented features extracted from the Date column. | | Yahoo Finance |
| **PositiveReturn** | | A binary flag signifying whether the stock registered a positive return on that day.If the value is set to one, it indicates that the stock registered a positive return on that particular day; otherwise, it suggests a non-positive return. | | Yahoo Finance |
| **Company name** | **Ticker** | |
| **Microsoft Corp** | MSFT | |
| **Apple Inc** | AAPL | |
| **NVIDIA Corp** | NVDA | |
| **…** | … | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| **Open** | 228099.0 | 135.13 | 219.70 | 0.70 | 38.23 | 71.41 | 149.69 | 2697.75 |
| **High** | 228099.0 | 136.92 | 222.53 | 0.71 | 38.66 | 72.30 | 151.75 | 2721.85 |
| **Low** | 228099.0 | 133.29 | 216.77 | 0.65 | 37.74 | 70.54 | 147.60 | 2687.81 |
| **Close** | 228099.0 | 135.15 | 219.67 | 0.70 | 38.21 | 71.46 | 149.69 | 2703.26 |
| **Adj Close** | 228099.0 | 131.23 | 220.08 | 0.70 | 34.34 | 65.79 | 144.80 | 2703.26 |
| **Volume** | 228099.0 | 11554842.24 | 28560594.65 | 0.00 | 1472750.00 | 3154700.00 | 8661400.00 | 1065523200.00 |
| **Sentiment\_Score** | 228099.0 | 0.67 | 2.38 | -96.88 | 0.00 | 0.08 | 0.38 | 66.62 |
| **New\_Covid\_Cases** | 228099.0 | 28569.96 | 83851.28 | 0.00 | 0.00 | 0.00 | 20673.00 | 1265520.00 |
| **MarketCap** | 228099.0 | 214808803513.93 | 478458310487.78 | 16346392576.00 | 39679574016.00 | 62220886016.00 | 131073245184.00 | 2828020482048.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Max AUC** | **Min AUC** | **Mean AUC** | **Median AUC** |
| **Random Forest** | 0.8119747899159660 | 0.5270964691046660 | 0.6773262130929950 | 0.677745245825603 |
| **XGBoost** | 0.9107142857142860 | 0.5805366483897400 | 0.7526194751681700 | 0.7519644741193330 |
| **LightGBM** | 0.8396358543417370 | 0.4333333333333330 | 0.6481785020287460 | 0.6392358095839580 |
| **CatBoost** | 0.8416610132054340 | 0.470279086625589 | 0.6209455609252810 | 0.6113474506838090 |

## 7.3 Cluster Analysis without Big Companies

In this section, because the previous clustering included big companies, Cluster 2

(Microsoft Corp, Meta Platforms Inc, Alphabet Inc), Cluster 3 (Apple Inc) and Cluster 4 (NVIDIA Corp, Amazon.com Inc, Tesla Inc), they are excluded to delve deeper into the rest of the companies. Likewise, through a KMeans clustering approach, I have segmented the companies into four distinct groups.

**A graph with colored crosses

Description automatically generated with medium confidence**

Figure 11 KMeans clustering of firm (Exclusing Outliers)

Cluster 0 (21 firms):

Companies in this cluster include:

* PDD Holdings Inc ADR
* Booking Holdings Inc
* Illumina Inc
* ... and 18 others.

Cluster 0 predominantly comprises technology and internet-based companies. These firms, despite their differences in market capitalization, exhibit similar sentiment scores, signifying a shared news sentiment trend. Notably, Cluster 0 tends to have sentiment scores higher than those in Cluster 3.

Cluster 1 (2 firms):

Companies in this cluster:

* Cisco Systems Inc
* Comcast Corp

Cluster 1 is minimal, consisting of two major tech companies. Both are renowned for their vast consumer and enterprise networks. Their influential positions in the tech industry might be the reason they share similar market sentiments.

Cluster 2 (6 firms):

Companies in this cluster include:

* ASML Holding NV
* Adobe Inc
* Broadcom Inc
* ... and 3 others.

Primarily involved in technology, software, and hardware development, these firms' shared industry focus might explain their grouping in terms of market capitalization and sentiment score.

Cluster 3 (62 firms):

Companies in this cluster include:

* Lucid Group Inc
* Atlassian Corp
* Constellation Energy Corp
* ... and 59 others.

Cluster 3, the most extensive cluster, encompasses a mix of companies from various sectors, including technology, energy, finance, and more. Despite the diversity, this cluster has the lowest sentiment scores among all clusters, suggesting specific underlying factors influencing their media portrayal.

Cluster 1

['PDD Holdings Inc ADR', 'Booking Holdings Inc', 'Intuitive Surgical Inc', 'Lam Research Corp', 'Advanced Micro Devices Inc', 'Applied Materials Inc', 'Automatic Data Processing Inc', 'Analog Devices Inc', 'Intuit Inc', 'T-Mobile US Inc', 'Regeneron Pharmaceuticals Inc', 'Gilead Sciences Inc', 'Airbnb Inc', 'Texas Instruments Inc', 'Mondelez International Inc', 'PayPal Holdings Inc', 'Amgen Inc', 'QUALCOMM Inc', 'Starbucks Corp', 'Intel Corp', 'Honeywell International Inc'],

['Cisco Systems Inc', 'Comcast Corp'],

['ASML Holding NV', 'Adobe Inc', 'Broadcom Inc', 'Costco Wholesale Corp', 'Netflix Inc', 'AstraZeneca PLC ADR'],

['Lucid Group Inc', 'Atlassian Corp', 'Constellation Energy Corp', 'Old Dominion Freight Line Inc', 'Workday Inc', 'MercadoLibre Inc', 'Enphase Energy Inc', 'GLOBALFOUNDRIES Inc', 'American Electric Power Co Inc', 'Diamondback Energy Inc', 'Fastenal Co', 'ON Semiconductor Corp', 'Copart Inc', 'Datadog Inc', 'KLA Corp', 'Cadence Design Systems Inc', 'IDEXX Laboratories Inc', 'Marvell Technology Inc', 'Micron Technology Inc', 'Align Technology Inc', 'ANSYS Inc', 'Seagen Inc', 'Cintas Corp', 'Monster Beverage Corp', 'Synopsys Inc', 'CoStar Group Inc', 'Dexcom Inc', 'Warner Bros Discovery Inc', 'Vertex Pharmaceuticals Inc', 'Exelon Corp', 'PACCAR Inc', 'Zscaler Inc', 'Verisk Analytics Inc', 'CSX Corp', 'Sirius XM Holdings Inc', 'Dollar Tree Inc', 'Keurig Dr Pepper Inc', "O'Reilly Automotive Inc", 'Cognizant Technology Solutions Corp', 'Lululemon Athletica Inc', 'Crowdstrike Holdings Inc', 'Ross Stores Inc', 'Microchip Technology Inc', 'Palo Alto Networks Inc', 'Electronic Arts Inc', 'Autodesk Inc', 'Xcel Energy Inc', 'JD.com Inc ADR', 'Fortinet Inc', 'Paychex Inc', 'Charter Communications Inc', 'GE HealthCare Technologies Inc', 'Marriott International Inc/MD', 'Illumina Inc', 'Baker Hughes Co', 'NXP Semiconductors NV', 'Walgreens Boots Alliance Inc', 'Kraft Heinz Co/The', 'Biogen Inc', 'Zoom Video Communications Inc', 'eBay Inc', 'Moderna Inc']