



Data processing And Analytics

Spark Stream Report

Realised By:

Adnan EL MOUTTAKI
Saad BELOUAD
Anouar ZAHRAN

Supervised By:

Pr. Mauri Andrea

Report on Real-Time Stock Market Analysis Using Spark Streaming

Introduction

The objective of this project was to leverage Spark Streaming to conduct real-time analysis of stock market data.

The project aimed to create a dynamic dashboard that could provide investors with timely insights into stock market trends and asset values.

Data Source

The dataset used for this project was derived from **stock.csv**, which was modified to focus on two key columns: the stock **name** and its **value**.

Task Details and Methodology

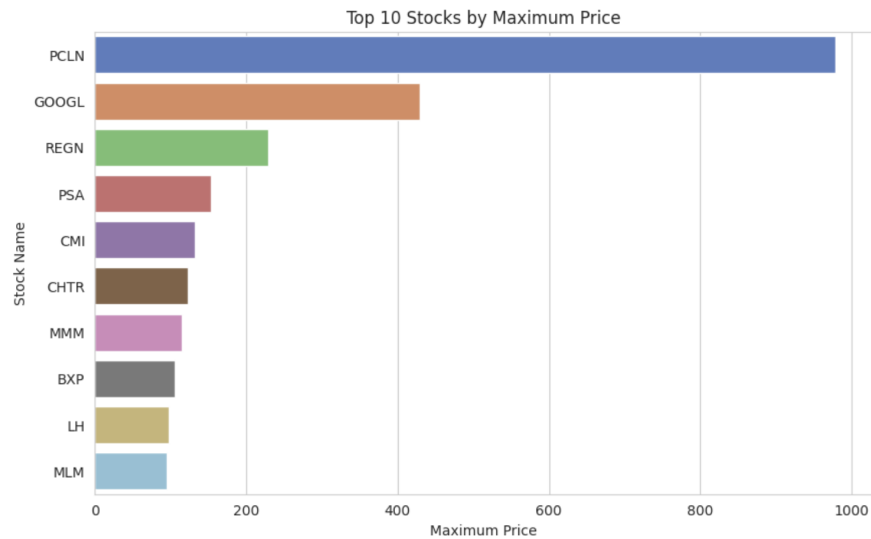
Task 1: Identifying the Most Valuable Stocks

Objective: Determine the top N stocks with the highest value within each time window.

Methodology: Applied a window operation to segment the data into fixed intervals, grouping by stock name and calculating the maximum stock price.

Results: Produced a ranked list of stocks by maximum price within each window, with a bar plot visualisation for the top 10 stocks in 80-second windows.

| window | name | max_price | rank |
|----------------------|-------|-----------|------|
| [2023-11-06 09:28... | PCLN | 979.23 | 1 |
| [2023-11-06 09:28... | GOOGL | 429.2839 | 2 |
| [2023-11-06 09:28... | REGN | 229.55 | 3 |
| [2023-11-06 09:28... | PSA | 153.06 | 4 |
| [2023-11-06 09:28... | CMI | 131.8 | 5 |
| [2023-11-06 09:28... | CHTR | 123.31 | 6 |
| [2023-11-06 09:28... | MMM | 115.63 | 7 |
| [2023-11-06 09:28... | BXP | 106.39 | 8 |
| [2023-11-06 09:28... | LH | 97.75 | 9 |
| [2023-11-06 09:28... | MLM | 95.67 | 10 |



Task 2: Tracking Stocks That Lost Value

Objective: Identify stocks that decreased in value from one window to the next.

Methodology: Compared stock values between consecutive windows to flag those with a decrease in price.

Results: Generated a list of stocks that lost value, potentially indicating a downward trend.

| window | name | current_max_price | previous_max_price | price_difference |
|----------------------|------|-------------------|--------------------|---------------------|
| [2023-11-06 09:28... | COP | 60.67 | 61.51 | -0.8399999999999963 |
| [2023-11-06 09:28... | NEE | 79.18 | 80.39 | -1.2099999999999937 |
| [2023-11-06 09:28... | ITW | 69.27 | 69.74 | -0.4699999999999986 |

Task 3: Finding Stocks with the Greatest Gains

Objective: Pinpoint stocks that gained the most value between windows.

Methodology: Calculated the difference in stock prices between windows to identify the highest positive changes.

Results: Highlighted stocks with significant gains, offering insights into strong market performers.

| window | name | current_max_price | previous_max_price | price_difference |
|----------------------|------|-------------------|--------------------|--------------------|
| [2023-11-06 09:29... | URI | 66.69 | 65.32 | 1.3700000000000045 |

Task 4: Implementing a Control for Excessive Value Loss

Objective: Monitor and alert if a stock's value falls below a set threshold within a time frame.

Methodology: Established a threshold for excessive loss and flagged stocks exceeding this limit.

For this project, the threshold was set at a maximum allowed loss of -5%.

Results: Provided a control mechanism for risk management by signalling potential critical value drops.

This alert system is crucial for investors to make timely decisions to mitigate losses.

| window | name | current_max_price | previous_max_price | price_difference | excessive_loss |
|----------------------|------|-------------------|--------------------|---------------------|----------------|
| [2023-11-06 09:53... | AIZ | 68.43 | 79.13 | -10.699999999999989 | true |
| [2023-11-06 09:53... | STZ | 137.33 | 146.76 | -9.429999999999978 | true |

Task 5: Calculating Personal Asset Fluctuations

Objective: Assess how the value of a personal portfolio changes with market fluctuations.

Methodology: Joined a personal stock portfolio dataframe, which contains a sample of stock names, with real-time market data to compute total asset value.

Results: Offers a real-time view of the portfolio's worth, enabling timely investment decisions.

| name | price | timestamp | name | amount | market_value |
|------|----------|---------------------|------|--------|--------------------|
| RE | 127.16 | 2023-11-06 09:27:12 | RE | 64 | 8138.24 |
| ACN | 75.56 | 2023-11-06 09:27:09 | ACN | 25 | 1889.0 |
| ARE | 70.97 | 2023-11-06 09:27:06 | ARE | 66 | 4684.0199999999995 |
| KMB | 89.85 | 2023-11-06 09:26:58 | KMB | 8 | 718.8 |
| ISRG | 191.9256 | 2023-11-06 09:26:56 | ISRG | 5 | 959.628 |
| LH | 89.64 | 2023-11-06 09:27:00 | LH | 86 | 7709.04 |
| CMI | 119.32 | 2023-11-06 09:27:00 | CMI | 56 | 6681.92 |
| PXD | 122.17 | 2023-11-06 09:27:06 | PXD | 40 | 4886.8 |
| PXD | 124.88 | 2023-11-06 09:27:10 | PXD | 40 | 4995.2 |
| VTR | 71.95 | 2023-11-06 09:27:09 | VTR | 70 | 5036.5 |
| AZO | 382.14 | 2023-11-06 09:26:58 | AZO | 14 | 5349.96 |

User Guide

To replicate the analysis presented in this report, follow these steps:

1. Run the `kafka_project` notebook to initiate the data streaming process.
2. Modify the `construct_stock` function to match the schema of your modified `stock.csv` file. Ensure the row indices match the columns for stock name and price.

The function should look like this:

```
```python
import time
from datetime import datetime

def construct_stock(row):
 time_stamp = time.time()
 date_time = datetime.fromtimestamp(time_stamp)
 str_date_time = date_time.strftime("%Y-%m-%dT%H:%M:%SZ")
 #"%d-%m-%Y, %H:%M:%S"
 stock = {"name": row[1],
 "price": float(row[0]),
 "timestamp": str_date_time
 }
 return stock
```
```

3. If you encounter an error due to file size limitations, compress the original file into a ZIP archive and then decompress it within the notebook environment.
4. Once the Kafka producer is running and streaming data, open the `project_template` notebook.
5. Execute the cells in the `project_template` notebook in order. Each cell contains comments explaining the purpose and function of the code within.
6. If any issues arise during execution, please reach out to one of the authors of this report for assistance.

Conclusion

The project successfully demonstrated the use of Spark Streaming to analyse stock market data in real-time.

The tasks were executed to provide valuable insights into stock performance and portfolio management, showcasing the potential of big data technologies in financial analytics.