****

**Analysis of the Various companies using the Real time data**

Name: Sohan Singh Rawat

Reg no: CU22MSD004A

School of Mathematics and Natural Science

Github link of the Assignment:

Submitted to Prof. Sarath B Sribhasyam

**Tools and Software Used in this Assignment**

* Finn hub
* Python
* Kafka
* Zookeeper
* Postgres
* Superset

**Work flow of the Assignment**

Consumer

Producer

Defining the topic

Inserting to the Database

Data collection

Dashboard

In this Assignment real time streamline data need to collected. So for that Finn hub is used. The Data that is collected is about the Insider Transactions.  
It contains the following things

**change**

Number of share changed from the last period. A positive value suggests a BUY transaction. A negative value suggests a SELL transaction.

**filingDate**

The data on which it was filled.

**name**

Insider's name or the name of the person who bought the specific share

**share**

Number of shares held after the transaction.

**symbol**

Symbol of the company or the name of the companies

**transactionCode**

Transaction code

**transactionDate**

Data of the transaction

**transactionPrice**

Average transaction price.

Credit:finnhub

About the Data:

The Data is basically about the 6 companies Tesla, Amazon, Apple, Google, Intel, Microsoft. The date range is from 2023 to 2024. The Looks something like this:  
Data:

<https://drive.google.com/file/d/1paj-oToxUXWqUh_kEM0VfstXe5DGynRs/view?usp=sharing>

After getting the continuous data from the Finn hub. Kafka comes into play and topic is being to store it the name of the topic is Project2 with 1 partition.

Then the Data that is coming is pushed to the Producer which reads the data and send it to the consumer. Consumer reads the data that is sent by the Producer. After reaching to the Consumer, it is sent to the Postgres.

In the Postgres a database is created that is ‘project’ and simultaneously a table is create ‘tt’ to store the data. The Data is pushed from the Consumer to the Database and then to the table.

After getting the Data it is sent it to the Superset where various kpi is defined and dashboard is created for the better understanding through the visualization.

Dashboard: <https://drive.google.com/file/d/1Avkv4610Pa85T4selorFRLyOn9CPNgHz/view?usp=sharing>

KPI

1. Companies Share : It basically the sum of the share available in the data

And the chart used to visualize it is pie chart.

The Querry used to the output.

SELECT symbol AS symbol,

count(records\_count) AS "COUNT(records\_count)"

FROM

(SELECT symbol,

COUNT(\*) AS records\_count

FROM tt

GROUP BY symbol) AS virtual\_table

GROUP BY symbol

ORDER BY "COUNT(records\_count)" DESC

LIMIT 5000;

Mathematically: Count(si)=Number of occurrences of symbol siCount(si​)=Number of occurrences of symbol si

1. Companies Data: Sum of the data available in the dataset to get the insight of the Companies record. And the chart used for it the table.

Output: Amazon and Tesla high numbers

SELECT symbol AS symbol,

records\_count AS records\_count

FROM

(SELECT symbol,

COUNT(\*) AS records\_count

FROM tt

GROUP BY symbol) AS virtual\_table

LIMIT 1000;

1. Transaction of the Share: Here I have calculated the number of share and volume. To get the sum of the transaction of the volume. The chart that is used bar on the y axis it is the share and x axis it the companies available.

Querry for it is:

Output: Amazon has the highest number

SELECT symbol AS symbol,

sum(total\_transaction\_volume) AS "SUM(total\_transaction\_volume)"

FROM

(SELECT symbol,

SUM(ABS(share)) AS total\_transaction\_volume

FROM tt

GROUP BY symbol) AS virtual\_table

GROUP BY symbol

ORDER BY "SUM(total\_transaction\_volume)" DESC

LIMIT 1000;

1. Price analysis with outlier companies: For this I have calculate the Average Transaction done mythe people. Chart that is used box plot and it will help in getting the outliers, on the y axis max transaction price and on the x axis details of the companies.

SELECT share AS share,

symbol AS symbol,

AVG(transactionprice) AS "AVG(transactionprice)"

FROM public.tt

GROUP BY share,

symbol

LIMIT 5000;

1. Trend Analysis over a Month: In this I have taken transaction count and month with fill date to get the appropriate result. I have taken the Max of the transaction count. On x axis I have companies.

Querry

SELECT symbol AS symbol,

max(transaction\_count) AS "MAX(transaction\_count)"

FROM

(SELECT symbol,

DATE\_TRUNC('month', filingdate) AS month,

COUNT(\*) AS transaction\_count

FROM tt

GROUP BY symbol,

month

ORDER BY symbol,

month) AS virtual\_table

GROUP BY symbol

ORDER BY "MAX(transaction\_count)" DESC

LIMIT 5000;

1. Market Analysis: For this I have taken the companies, transaction code and volume. In the transaction code I have some letter and it means this

S - Sell: This indicates that the transaction is a sell order, meaning the investor is selling their shares of a particular security.

M - Market Order: This indicates a market order, where the investor wants to buy or sell a security at the current market price. Market orders prioritize speed of execution over price.

A - Ask Price: This might represent the ask price, which is the price at which a seller is willing to sell a security. It can also represent an 'Ask' order, where an investor is willing to buy at a specific price.

F - Fill or Filled Order: This usually indicates that the order has been filled, meaning the buy or sell transaction has been completed.

G - Good Till Canceled (GTC) Order: This type of order remains active until it is either filled or canceled by the investor. It stays in the system until it achieves its objective or the investor decides to cancel it.

C - Cancel Order: This indicates that the investor has canceled their order before it was executed. It could be a buy or sell order that was placed but later withdrawn.

Chart that is used Gauge and I have added the transaction volume

Querry:

SELECT symbol AS symbol,

transactioncode AS transactioncode,

sum(total\_transaction\_volume) AS "SUM(total\_transaction\_volume)"

FROM

(SELECT symbol,

transactioncode,

SUM(ABS(share)) AS total\_transaction\_volume

FROM tt

GROUP BY symbol,

transactioncode

ORDER BY symbol,

transactioncode) AS virtual\_table

GROUP BY symbol,

transactioncode

LIMIT 1000;

1. Number of operation: I have taken the transaction count and added based on the companies available on the data.

Chart used in the pie chart

Querry:

SELECT transactioncode AS transactioncode,

sum(transaction\_count) AS "SUM(transaction\_count)"

FROM

(SELECT transactioncode,

COUNT(\*) AS transaction\_count

FROM tt

GROUP BY transactioncode) AS virtual\_table

GROUP BY transactioncode

ORDER BY "SUM(transaction\_count)" DESC

LIMIT 100;