

Stock Exchange Data Analysis

Objective: To use hive features for data engineering or analysis and sharing the actionable insights

Problem Statement:

NewYork stock exchange data of seven years, between 2010 to 2016, is captured for 500+ listed companies. The data set comprises of intra-day prices and volume traded for each listed company. The data serves both for machine learning and exploratory analysis projects, to automate the trading process and to predict the next trading-day winners or losers.. The scope of this project is limited to exploratory data analysis.

Domain: BFSI

Analysis to be done: Exploratory analysis to understand how MoM or YoY companies from different sectors or industries and states have progressed in a period of 7 years

Content: This data set contains prices.csv and securities.csv files having the following features:

Prices.csv:

1. Date: Trading date
2. Symbol: Ticker code or listed company code on NY stock exchange
3. Open: Intra-day opening price for each listed company
4. Close: Intra-day closing price for each listed company
5. Low: Intra-day lowest price for each listed company
6. High: Intra-day highest price for each listed company
7. Volume: Number of shares traded per day per company

Securities.csv:

1. Ticker_Symbol: Country to which the customer belongs
2. Security: Legal name of the listed company
3. Sector: Business vertical of the listed company
4. Sub_Industry: Business domain of the listed company within a Sector.
5. Headquarter: Headquarters address

Steps to perform:

1) Create a data base in Hive (DATABASE NAME: BDHS_PROJECT)

- i. Stock_prices
- ii. Stock_companies

Check the TABLE description: STOCK_PRICES

Column Name

Datatype

Trading_date	Date
Symbol	String
Open	double
Close	double
Low	double
High	double
Volume	int

TABLE: STOCK_COMPANIES

Column Name	Datatype
Symbol	String
Company_name	String
Sector	String
Sub_industry	String
Headquarter	String

2) Create a new hive table with the following fields by joining the above two hive tables. Please use appropriate Hive built-in functions for columns (a,b,e and h to l).

- Trading_year: Should contain YYYY for each record
- Trading_month: Should contain MM or MMM for each record
- Symbol: Ticker code
- CompanyName: Legal name of the listed company
- State: State to be extracted from headquarters value.
- Sector: Business vertical of the listed company
- Sub_Industry: Business domain of the listed company within a sector
- Open: Average of intra-day opening price by month and year for each listed company
- Close: Average of intra-day closing price by month and year for each listed company
- Low: Average of intra-day lowest price by month and year for each listed company
- High: Average of intra-day highest price by month and year for each listed company
- Volume: Average of number of shares traded by month and year for each listed company

DATA ANALYSIS USING HIVE

- 3) Find the top five companies that are good for investment
- 4) Show the best-growing industry by each state, having at least two or more industries mapped.
- 5) For each sector find the following.

- Worst year
- b. Best year
- c. Stable year