

1. Pendahuluan

Menurut Economatica, sebuah perusahaan yang mengkhususkan diri dalam pasar saham Amerika Latin, pasar saham Brasil, yang diatur oleh Brasil, Bolsa, Balcão (B3), mempertukarkan BRL sekitar 25,9 miliar per hari pada paruh pertama tahun 2020, selama epidemi virus corona. Pada periode yang sama terjadi pertumbuhan 18% dalam jumlah investor Brasil, dengan total sekitar 2,6 juta investor aktif. Oleh karena itu, pasar keuangan menggerakkan sejumlah besar nilai sehingga menghasilkan informasi dan data setiap hari. Data ini menunjukkan pergerakan saham, harga masing-masing, nilai tukar dolar dan sebagainya.

2. Deskripsi Dataset

a. Nama dan Sumber Dataset

Nama Dataset: Brazilian Stock Market

Sumber: www.kaggle.com/datasets/leomauro/brazilian-stock-market-data-warehouse

b. Struktur Data (Skema dan Format)

Dataset ini menggunakan Fact Constellation Schema (Galaxy Schema), dengan detail:

1) Tabel fakta:

- factStocks: Menyimpan transaksi saham harian
- factCoins: Menyimpan nilai tukar mata uang harian

2) Tabel dimensi:

- dimCompany: Menyimpan informasi perusahaan terkait
- dimCoin: Menyimpan informasi mata uang terkait
- dimTime: Menyimpan informasi waktu transaksi

Dataset ini dikemas dalam format SQL dengan detail setiap tabel sebagai berikut:

1) Tabel "dimCompany" (Dimensi Perusahaan)

- keyCompany: ID perusahaan
- stockCodeCompany: Kode saham perusahaan
- nameCompany: Nama perusahaan
- sectorCodeCompany: Kode sektor perusahaan
- sectorCompany: Nama sektor perusahaan
- segmentCompany: Nama segmen perusahaan
- startedAt: Tanggal mulai aktif perusahaan
- endedAt: Tanggal berhenti aktif (bisa null)
- isActive: Status perusahaan (aktif atau tidak)

2) Tabel "dimCoin" (Dimensi Mata Uang)

- keyCoin: ID mata uang
- abbrevCoin: Singkatan mata uang

- nameCoin: Nama mata uang
 - symbolCoin: Simbol mata uang
- 3) Tabel "dimTime" (Dimensi Waktu)
- keyTime: ID waktu
 - datetime: Format tanggal-waktu lengkap
 - dayTime: Hari dalam bulan
 - dayWeekTime: Hari dalam minggu (angka)
 - dayWeekAbbrevTime: Singkatan nama hari (misalnya: Sen, Sel, Rab)
 - dayWeekCompleteTime: Nama hari (misalnya: Senin, Selasa)
 - monthTime: Bulan dalam angka
 - monthAbbrevTime: Singkatan nama bulan (misalnya: Jan, Feb)
 - monthCompleteTime: Nama lengkap bulan (misalnya: Januari, Februari)
 - bimonthTime: Periode dua bulanan
 - quarterTime: Kuartal dalam tahun
 - semesterTime: Semester dalam tahun
 - yearTime: Tahun dalam angka
- 4) Tabel "factCoins" (Fakta Mata Uang)
- keyTime: ID waktu
 - keyCoin: ID mata uang
 - valueCoin: Nilai tukar mata uang
- 5) Tabel "factStocks" (Fakta Saham)
- keyTime: ID waktu
 - keyCompany: ID perusahaan
 - openValueStock: Harga pembukaan saham
 - closeValueStock: Harga penutupan saham
 - highValueStock: Harga tertinggi dalam sehari
 - lowValueStock: Harga terendah dalam sehari
 - quantityStock: Volume perdagangan saham

c. Volume Data (Jumlah Record)

Dataset ini mencakup data harian dari tahun 1994 hingga 2020 dengan jumlah record:

- Dimensi Coin: 2 baris
- Dimensi Company: 607 baris
- Dimensi Time: 9680 baris
- Fakta Coin: 13300 baris
- Fakta Stock: 680150 baris