ECE 356 Project

Stock Market Data

This document is a description of details relevant to those doing the course project using the **Stock Market** datasets. It provides details about the dataset, as well as suggestions pertaining to the client application, the entity-relationship design, and the data-mining exercise.

Data Source

The main source of data for those doing a project in the stock-market domain is:

`https://www.kaggle.com/qks1lver/amex-nyse-nasdaq-stock-histories`

It contains daily stock history data for several thousand stocks. This is required in your database, but not sufficient as each CSV contains just seven distinct attributes. There are three other sources:

`https://www.kaggle.com/proselotis/financial-ipo-data`
`https://www.kaggle.com/cnic92/200-financial-indicators-of-us-stocks-20142018`
`https://www.kaggle.com/miguelaenlle/massive-stock-news-analysis-db-for-nlpbacktests`

You are not expected to use all three, nor even necessarily to use all attributes within any given one of these three datasets. However, if your final database has only a few relations with only a couple of dozen distinct attributes in total, you have probably done a mediocre project. All CSV files from the above sources are available in "`/var/lib/mysql-files/18-Stocks/`" using exactly the names as on the Kaggle site.

In all instances, you should look at any of the relevant datasets on Kaggle to determine what the different attributes are within the CSV.

Client Application

As noted in the main project document, there will be little additional to add to the generic client-application requirements listed there. If you want a sense of what a client application for the stock-market domain should do, you should think about the potential users of such a database. Likely users include stock analysts who wish to look up the history of stock prices, and examine that data in the light of things such as analyst recommendations and IPO information. Any such database should allow the addition of stock-data as such information becomes available. It is also likely that such a database would be used by a group of people and they would want to add comments on any given stock.

Entity-Relationship Design

Per the main project document, you will need to determine an appropriate ER design for your dataset. There have been no prior projects done using this data. However, there are certain entity sets that seem likely to be needed, such as “Stock”, “TradingData” (it is likely worthwhile to distinguish market prices from a particular stock, which is simply the company name, symbol, and other pertinent facts about the company), “IPOInformation” (if you are using that data), “AnalystInformation”, and so forth. If you have difficulty in thinking about different relevant entity sets for this domain you should consult with your designated instruction-team member.

Data-Mining Investigation

For stock-market data there are numerous possible data mining-exercises that are worth considering. One of the most useful ones is expressed in the question, “What factors (attributes) determine the likelihood of a stock price rise in excess of the market average?” Other questions of relevance include, “What factors determine an overpriced stock?” and “Categorize (cluster) stocks based on market attributes rather than the usual industry-segment categorization”

If you have difficulty thinking about an appropriate data-mining exercise, you should consult with your designated instruction-team member. If you think you have a good idea for a data-mining exercise, it is probably worthwhile checking with your designated instruction-team member to confirm that it is of appropriate scope.