

Assignment Description

You will use SQL Server (database and analysis services) or MySQL + Pentaho:

a) [40%] A large data set will be found, which will be cleaned and inserted into a data warehouse designed by the groups (2 students per group). Then a data cube and various metrics should be created.

b) [20%] A visualization tool (Tableau or Power BI) will be used to create various instances of data visualization.

c) [40%] The warehouse data will be used for some mining operations, such as categorization, correlation rules, clustering, etc. Use trading system methods and models or an open-source tool. You will implement at least two models.

To find datasets, you can search the internet for relevant ones websites (eg <https://www.kaggle.com/>, <http://archive.ics.uci.edu/ml/>, <http://www.kdnuggets.com/datasets/index.html>)