BSc (Hons) in Information Technology
Specialising in Data Science

IT3021 – Data Warehousing and Business Intelligence
Year 3

Assignment 1: Guide to Find Data

Semester 1, 2021

Finding your data set

The whole point of using different sorts of data sets is to encourage students to experience different scenarios rather than everybody ending up working on the same data set.

If you try searching "sample transaction data sets" or "sample data warehouse data sets", it's highly unlikely you come up with a data set. Instead think of a scenario first:
Ex: Call records of a telecommunication company
    Insurance claims data
    Inventory data

Then you can find sample data. Maybe not all the data you can find in a single file:
Ex; Insurance claims data scenarios -
        You will be able to find a list of insurance claims in one file.
        Then you can get a list of customer data from a separate file.
        Get set of insurance products details from another file.
        Locations from another. Likewise you will have to build your data set.

Following are some links for you to try to get some data too:

https://www.kaggle.com/datasets

https://community.tableau.com/docs/DOC-1236

http://eforexcel.com/wp/downloads-18-sample-csv-files-data-sets-for-testing-sales/

http://nodobo.com/release.html

http://www.baseball-databank.org/

http://seanlahman.com/baseball-archive/statistics/

https://busintblog.wordpress.com/2014/06/27/publically-avaliable-data-sets-useful-to-practice-etl/

http://www.altanova.pl/blog/best-free-sample-sql-data-set-anyone/

https://www.jenunderwood.com/2016/01/14/my-favorite-public-data-sources/

https://github.com/oracle/data-warehouse-etl-offload-samples

https://www.toolsverse.com/etl-framework/etl-examples/index.html

https://data.world/datasets/ecommerce

https://support.spatialkey.com/spatialkey-sample-csv-data/

https://www.lib.ncsu.edu/teaching-and-learning-datasets

https://www.stats.govt.nz/large-datasets/csv-files-for-download/

https://www.sqlchick.com/entries/2011/10/3/places-to-download-sample-data.html