Modern Data Management & Business Intelligence
Assignment #2 – Due Date: December 4th, 2017 (presentation)

You are going to use SQL Server Database, SQL Server Analysis Services and Power BI or Tableau for this project (alternatively you may use MySQL/Pentaho/Tableau). You are going to design and develop a data warehouse, build one or more data cubes on top of it, develop some OLAP reports and visualize your results. You are going to present your project in class (15' each group). This should be in the form of a business case. This includes:

- business goals, description of the problem/domain
- description of data sources, where did you find the datasets
- design of the data warehouse, cubes, etc
- import/cleaning/transformation challenges and what did you do
- examples of OLAP queries, reports, etc.
- visualization examples

Try to make it as a story – you are the story teller!

1. Find a dataset in the web that seems attractive and interesting to you. Possible links:

   https://github.com/caesar0301/awesome-public-datasets
   http://www.kdnuggets.com/datasets/index.html
   https://catalog.data.gov/dataset?tags=data-warehouse

   or, search google for "datasets for data warehousing / data mining / OLAP / etc."

2. Understand the facts and the dimensions of the application. Define a star/snowflake schema in your database SQLServer (or MySQL). Populate the fact and the dimension tables from the dataset you found - for example by using the import task in your database server. You may have to clean, transform the dataset, manually define dimension tables or insert values.

3. Use SQL Server Analysis Services (or Pentaho) to define a multi-dimensional model (a cube) over your schema. Play with the reporting capabilities of your tool and show some OLAP reports (drill down/roll up, pivoting, ranking, etc.)

4. Install Tableau and using your database schema, show OLAP examples and visualize these - or whatever else you consider interesting. Better (and more interesting/interactive/etc) visualizations mean better grade ☺

The deliverable (aside the presentation) should be a document (.doc or .pdf) describing in detail each of the above steps - with a lot of screenshots. (a) what kind of application you are targeting, what dataset you used, where did you find it, what problems you are trying to solve, what analysis you want to do, (b) description of the relational design of your fact and dimension tables, import methods, cleaning/transformation procedures, (c) what cube you have built on top of your schema, dimensions, measures, calculated - if any - measures; description (in English) of OLAP reports and screenshots, and (d) visualizations of these reports and description of the visualization, how it was produced, etc.

Instructions on downloading systems and tools:

SQL Server (I am using SQL Server 2012 but you can use any version >= 2008):

Download SQL Server Developer's Edition from Microsoft's Imagine Software.

Tableau*:

Follow the instructions announced earlier in the semester

*Tableau's [data visualization software](#) is provided through the Tableau for Teaching program