

IICT6011 - Business Intelligence & Reporting

Assignment 02 [Home]

4th April 2017

General Guidelines

- This is a **home** assignment and a total of **9 weeks** has been allocated.
- This assignment carries a total of **56%** from the final module mark.
- You are requested to print this document and present all your content in an accompanying **CD**.
- Your documentation is to be submitted on **TurnItIn** which details have been provided on the module Moodle page.
- Copying is strictly prohibited and any students caught will be subject to the respective **MCAST Disciplinary Procedures**.

Business Understanding

You are being provided with a dataset for which you are to analyse this data, clean it, design and implement an OLAP cube, perform an ETL process and finally prepare a data mining report of your choice. You are free to use the tools of your choice, yet the assignment is completely possible with Microsoft SQL Server and R-Project alone. For your documentation it is strongly recommended that you use \LaTeX .

The only restrictions being imposed are the adoption of the CRISP-DM and the use of a **GIT** repository, even though you are free to choose between GitHub or BitBucket.

A. GIT

Your first task is to create a GIT repository. Add your lecturer to the repository list of users. Make sure to commit your work regularly.

B. DATASET

You are being provided with an OLTP database for the sale of items across different stores for the years 1997 and 1998. The file **oltp - erd.png** contains the logical ERD, whilst the **oltp - setup.sql** file creates the schema and loads the data. You must create a database labelled **iiict6011a02** and run the setup script. Please be patient since there are 251,395 individual items sold.

C. LITERATURE

You are being provided with a number of academic papers documenting research on the subject of this dataset. Review said literature with any other that you find, then document in the literature review section. Details about the report is provided for you in section G.

D. PRELIMINARY ANALYSIS & DATA CLEANING

Your next task is to go through the actual data and perform some preliminary analysis and identify any data cleaning issues. You can use any tool of your choice for the analysis such as: Excel; SQL; or R. Either way, you are expected to document your process and key aspects in your report. Please note that some attributes have only 1 unique value (so you can delete), other attributes need to be computed (so need to be generated in next section).

E. OLAP CUBE

You are requested to create at least one cube, yet you are free to create multiple such as: product sales; order total; client spending; store profit. The subsequent reports that you will generate are based on the cube you design, so do consider different designs and allow yourself time to experiment.

F. DATA ANALYSIS

Prepare your dataset for analysis and use the research documented in your literature review as inspiration to create reports of your choice. You can consider clustering, decision trees, correlation analysis, basket analysis or similar reports. Be sure to identify your research question, which is heavily based on the cube(s) you design, such as:

- Investigating buying patterns across different neighbourhoods
- A temporal multi-store market basket analysis
- Investigating buying patterns across different social classes
- Investigating temporal multi-store sales trends

N.B. At some point you need to consider normalising data, creating appropriate buckets (based on analysis such as for annual income) prior to generating any reports.

G. REPORT

A 5-10 page IEEE conference styled report is expected for this project with the following sections:

- **Title** This should include a short name for your research, your full name as the author, the institute information, your MCAST email and the date of the report.
- **Abstract** A brief (100-word approx.) explanation of what this research is all about.
- **Keywords** A few key terms about the techniques used in this research.
- **Introduction** A short overview of the dataset and motivation behind this research. A general overview of the techniques used. An overview of how the report is structured.
- **Literature Review** A review of some of the current research published on the provided dataset.
- **Research Methodology** A review of the approach taken (CRISP-DM).
- **Data Gathered** A review of the data analysis process performed on the OLAP data.
- **Data Analysis** A discussion of the generated reports from the exported data, model evaluation.
- **Conclusion** Some concluding points, highlighting achievements of this research and its limitations. Do provide any recommendations for future research.

H. INTERVIEW

After submitting your work a short interview will be held with your lecturer where you will have an opportunity to present your work. You are free to use any media of your choice (Presentation, Video, etc.). A video demonstration is not expected for this assignment, yet greatly welcomed. Kindly bring a printed copy of your report. A few questions will follow your demonstration. A casual smart dress code is expected.

Grading Criteria

Literature Review - KU1.2	
Document current research (at least 3 papers) on the provided dataset.	_ / 5
Research Methodology - AA2.2	
Document the CRISP-DM methodology as adopted in your project.	
Document and illustrate a 3D OLAP cube.	_ / 3
Implement the cube.	_ / 4
Research Methodology - AA4.3	
Document and illustrate the OLAP cube ERD.	_ / 3
Implement the cube.	_ / 4
Research Methodology - KU3.1	
Document the data cleaning process prior to migration.	_ / 2
Provide the code for data cleaning	_ / 3
Research Methodology - AA3.1	
Document the ETL process to the OLAP cube.	_ / 2
Provide the SQL code used for migration.	_ / 5
Data Gathering - KU3.2	
Document the data analysis process performed on the OLAP data.	_ / 3
Provide the code used.	_ / 2
Data Analysis - SE4.1	
Document the data analysis process applied on the OLAP data.	_ / 2
Document and assess generated reports. Provide any code used.	_ / 8
Final Report - SE4.2	
Include an abstract, introduction and conclusion in your report.	_ / 3
Make proper use of IEEE conference template including referencing.	_ / 7
Total	_ / 56