**Data Warehousing**

**Data Warehousing – Select Project Subject**

**Date Due (See Canvas)**

The focus of this course will be the creation of a data warehouse prototype.

We live in a world where data abounds. Much of it is useless until converted into some form of meaningful information. That is the primary goal of any data warehouse – to organise data from one or more sources into coherent information for the stakeholders.

The first task of your group is to identify a subject for your data warehouse project. You may choose to get your data from one of the suggested data sources or choose another source in a topic that interests you.

# Data

If you choose to use your own sourced data set, the following conditions must be met.

Your data warehouse fact table will require a minimum of 50,000 rows, so it will not be possible to completely manufacture this data while maintaining a realistic range of values.

Your data warehouse must contain at least 4 realistic dimension tables. One of these must be a time dimension. Ensure your data contains a time variance large enough to be usable. The time span must contain current time frames (within the last 2 years). I am willing to remove this restriction in exceptional cases (e.g. reporting on troop movements during WWII)

The following topics are not permitted:

* Prepared datasets such as the AdventureWorks databases on the Microsoft website.
* Airline flight delays and/or on-time performance
* Crime incident data from Chicago or another major city

**Business Case**

There must be a business case for why you are building your data warehouse. “Business” does not necessarily mean you want to make money. A business case simply means that there is a goal for your data warehouse, and you want to find out something by compiling the data.

Try to be creative in the way you approach the topic. Think of novel ways you could use the data to solve a business problem.

# Initial Data Model

In order to verify that your data can support a data warehouse, a draft of the data model must be developed as part of the proposal. The goal of this model is to identify the Fact Table Measures and the Dimensions and ensure the data source supports this model. The model should include at least one non-date hierarchy.

In the past, teams have realised too late that their data was not robust enough to complete the project, adding complexity and effort to the project.

# Submission Instructions

Submit via Canvas.

There only needs to be one submission per group. I will review your topic briefly and send you feedback if I think your topic could present any issues. Include the following information:

1. Group number of your group.
2. A brief statement (i.e. one or two sentences) of the business case.
3. A link (URL) to the data source chosen.

4. A data model diagram in either Visio format (.vsd), an image created in SQL server, or a PDF from another source.