CT Academy | Module 2 Acquiring & Exploring Data

Overview

This module focuses on the foundational practices of acquiring, sourcing, and exploring data. Participants will examine how data enters their organizations, assess documentation practices, and begin the exploration process using simple tools like Excel and Power BI. The module looks at what we can, and more importantly cannot do with different sets of data.

The exercises for this module are divided into two sections, the first is a set of reflection questions that we will use as the basis for discussion in the data lab. The second section will be an active exercise for you to develop and embed some of the concepts that you have learned by following the content.

This document also outlines the activities in the associated Data Lab.

Learning Objectives

- Understand different methods of acquiring data, for example studies and surveys
- Recognize the danger of bias by looking at different sampling methods
- Review approaches to data exploration to fully understand the capabilities and limitations of your dataset

Data Lab

The data lab for Module 2 will have the following format:

1. Review and discussion of the answers to the reflection questions for Module 2 with some real-life examples provided by the facilitators
2. Review what data acquisition techniques different departments (aside from SAP and PeopleSoft data)
3. Discuss why mapping data is a critical activity prior to data modeling
4. Follow the instructors as they create a Mastery Level 3 data model and discuss how it might need to be adapted for different organizational needs

CT Academy | Module 2 Acquiring & Exploring Data

Reflection Exercises

Exercise 1:

In the learning resources for this module, we learned about *acquiring data*. List out a number of ways that you acquire data for your reports or analysis. For example, you might run an SAP TCode, a query from Cognos or access a Power BI data model or dataflow.

Exercise 2:

In the learning resources for this module, we learned about *data exploration*. List out the ways that you investigate your data including the tools that you use. For example, you might use excel to plot actuals against budget over time to see what the trend is.

Exercise 5:

In the learning resources for this module, we were introduced to the concept of *sampling*. Identify datasets that might use sampling to get their data. For example, you might use data from different employee surveys as part of a dashboard.

CT Academy | Module 2 Acquiring & Exploring Data

Active Exercise

As part of this course you have been given an excel file (GoC Financial Sample Data CLEAN.xlsx) with several tabs each containing different types of data.

In the learning resources for this module, we learned about *mapping data*. In this we want you to identify columns of data from different tables that you can use to "map" the data (primary keys). For example, you might have two tables in excel, both of which contain a cost or fund centre.

In the learning resources for this module, we also learned about *data modeling*. We want you to at least understand the concept of data modelling and at higher mastery levels actually create a data model yourself!

NOTE: Mastery levels are not exclusive, you might need (or want) to do Mastery Level 1 *AND* 2 *AND* 3 to get to the result that you need.

To perform the exercises first open the Excel file, then click on the "Transactions" tab and perform the activity underneath. Then repeat for the "Budget" tab.

	Mastery Level Examples		
Activity	**Mastery Level 1**	**Mastery Level 2**	**Mastery Level 3**
Identify which columns from different tables are common so we can map and then model the data	Highlight the column titles in colours that match, for example highlight the "Cost Centre" column title in the "Transactions" tab in (for example) yellow, then go to the "Cost Centre" tab and highlight the "Cost Centre" column title again in yellow to show it matches	Go to the Mastery Level 3 table you created in Module 1 and add in two additional columns that identify the tab and column that you can map to. For example, in your "Fund Code" row you should note that the matching column you will use as a primary key comes from the "Fund" tab and in that tab the "Fund Code" column. If you wanted you could also create a hyperlink as well!	Create and Entity Relationship (ER) Diagram in a program like Visio (Power Point is ok) or using a scripting language like DBML. If you don't know what this is don't worry, we will review in the Data Lab!

	Mastery Level Examples		
Activity	**Mastery Level 1**	**Mastery Level 2**	**Mastery Level 3**
Create a data model / relate the data together programmatically	In the transactions or budget tab in the excel sheet use a XLOOKUP (NOT VLOOKUP, will explain in the lab!) to copy in additional columns of interest, for example looking up the "Cost Centre Desc EN" values in the Cost Centre table and copying them to a new column in either the Transactions or Budget table	Using a program like Power BI (in which you are able to create a data model) link all of the tables together using primary keys you identified in the mapping exercise above.	Using a program like Power BI (in which you are able to create a data model) link all of the tables together using primary keys you identified in the mapping exercise above. In addition, identify the tables as transactional or reference and add in additional useful data (e.g., in a Date reference table add in fiscal year details)