



Institute for Management & Innovation
UNIVERSITY OF TORONTO
MISSISSAUGA

IMix Course Overview

Title: Data Cleaning and Manipulation

Length: 1 day

Rationale/purpose of the course:

- What does the course/module seek to achieve?
 - This course will help participants understand the importance of clean and accurate data.

- Why is the course important?
 - In order to be useful for data mining purposes, the databases need to undergo preprocessing and depending on the data set, this alone can account for 10–60% of all the time and effort for the entire data mining process. Therefore, in this session, we will introduce a very powerful Python tool, pandas, which has many features and capabilities to manipulate large data sets. No programming knowledge is assumed because we will start with a short introduction to Python.

- Who should attend?
 - Managers, data analysts, and others who need to keep abreast of the latest methods for enhancing return on investment.

Instructor: Gerhard Trippen, Associate Professor, Operations Management

Issues/Topics to be covered:

- How do we clean raw data, filter out fields that are obsolete or redundant, deal with missing data and outliers, and perform transformation on certain variables?

Modes of instruction:

- Discussion
- Interactive Sessions including Coding

*This course is a requirement for the Certificate in Data Analytics