Week 1: Data Management

1. What is Data Management?

* The process of collecting, organizing, protecting and storing an organization’s data so it can be analyzed for business decisions.

1. Database Management Systems

**Database Management System (DBMS)**

* Allows users to store, retrieve, and manipulate data in a database.
* Provide tools that enable users to create, update, and delete data in the database.
* Includes data security, data integrity, and data recovery features.

**Relational Database Management System (RDBMS)**

* Relational databases organize data into tables with rows and columns.
* Records in different tables can be connected using primary and foreign keys.
* Built around the SQL programming language.

1. Data Lake vs Data Warehouse

Data Lake: A repository that stores vast pools of raw data for use in predictive modelling, machine learning and other advanced analytics applications.

Data Warehouse: A repository (storage space) for business data where structured and filtered data that has already been cleansed and processed.

1. Data Discovery

* The process of applying advanced analytics to data to detect informative patterns.

1. Data Cleaning

* The process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

1. Data Integration

* The process of combining data from multiple sources to provide a unified, single view of data.

1. Data Labelling

* The process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to that data to specify its context.

1. Data Lineage

* The process of understanding, recording and visualizing data as it flows from data sources to consumption.