**Importing Data Sets**

In this module, you will learn how to understand data and learn about how to use the libraries in Python to help you import data from multiple sources. You will then learn how to perform some basic tasks to start exploring and analyzing the imported data set.

**Learning Objectives**

* Access databases using Python database APIs
* Analyze Python data using a dataset
* Identify three Python libraries and describe their uses
* Read data using Python's Pandas package
* Demonstrate how to import and export data in Python

# **Importing Data Sets**

## **Course Introduction**

* **Introduction to Data Science with Python**:
  + Importance of Python in data science.
  + Overview of libraries like NumPy, Pandas, and scikit-learn.
* **Module Breakdown**:
  + **Module 1**: Understanding dataset characteristics, importing data, and starting analysis.
  + **Module 2**: Data wrangling, preprocessing, handling missing values, formatting, and normalization.
  + **Module 3**: Exploratory data analysis (EDA), descriptive statistics, GroupBy, and correlation.
  + **Module 4**: Linear regression, model evaluation, polynomial regression, and decision-making.
  + **Module 5**: Model evaluation and refinement, overfitting, underfitting, and grid search.
* **Hands-On Project**: Application of skills learned using a real-world dataset.

## **Understanding the Data**

* **Dataset Overview**:
  + The dataset is an open dataset by Jeffrey C Schlimmer in CSV format.
  + Each line represents a row in the dataset.
* **First Row**:
  + The first row is not a header; it contains data.
* **Attributes**:
  + There are **26 columns** in total.
  + **Symboling**: Indicates the insurance risk level of a car (values range from -3 to +3).
  + **Normalized Losses**: Represents the average loss payment per insured vehicle per year, normalized for car size (values range from 65 to 256).
  + The **26th attribute** is **price**, which is the target value we want to predict.
* **Goal**:
  + The objective is to predict the price based on other car features.
* **Note**:
  + The dataset is from **1985**, so car prices may seem low.