Date: December 30 2022

# Introduction to Spreadsheets for Data Analysis

Course Overview:

**Module 1**

In this module, we will learn about the:

* basics of spreadsheets
* spreadsheet terminology
* The interface – Navigating Around a Spreadsheet

**Module 2**

* Selecting, Entering and Editing Data
* Copying and Auto-Filling Data
* Formatting Data
* Using Functions and Formulas

**Module 3**

* Cleaning and Wrangling Data Using a Spreadsheet
* The Fundamentals of Data Quality and Data Privacy
* Removing Duplicated and Inaccurate Data
* Removing Empty Rows
* Removing Data Inconsistencies and Whitespaces
* Using the ‘Flash Fill’ and ‘Text to Columns’ Features

**Module 4**

* Analyzing Data Using Spreadsheets
* Filtering and Sorting Data
* Using Common Data Analysis Function
* Creating and Using Pivot Tables
* Creating and Using Slicers and Timelines

At the end of the course,

* Understand how spreadsheets can be used as a data analysis tool.
* Understand when to use spreadsheets as a data analysis tool and their limitations.
* Create a spreadsheet and explain its basic functionality
* Perform data wrangling and data cleaning tasks using Excel.
* Analyze data using filter, sort, and pivot table features within Excel spreadsheets.
* Perform some intermediate level data wrangling and data analysis tasks to address a business scenario.

## How a Data Analyst Uses Spreadsheets?

As a Data Analyst, we can use spreadsheets as a tool for our data analysis tasks, including

* collecting and harvesting data from one or more distributed and different sources.
* Cleaning data to remove duplicates, inaccuracies, errors and resolve missing values to improve the quality of the data.
* Analyzing data by filtering, sorting and interpreting it to determine what useful information can be gleaned from it.
* And visualizing data, to help us tell a story about our data analysis findings to key business stakeholders and any other interested parties within our organizations.

## Spreadsheet Basics:

* Workbooks are the highest-level component in Excel and are represented as a .XLSX file.
* Workbook consists of one or more worksheets each of which is represented by a tab in Excel.

Getting Started using Spreadsheets:

* Viewing, Entering and Editing Data:

Viewing features in Excel, enter some data and edit data.

* Copying, Fulling and Formatting Cells and Data

Entered and edited some data., how to move, copy and fill data, format cells

The Basics of Formulas:

* Basic calculations (SUM)
* Selecting ranges in formulas
* How to copy formulas

The common functions used by Data Analysts:

* Some statistical functions [Formulas tab at the top bar]

Referencing Data in Formulas:

* The difference between relative, absolute, and mixed references
* How to use relative, absolute, and mixed references in formulas
* Formula errors in Excel

**Practice Quiz**:

1. How can you zoom to a specific area of data in an Excel spreadsheet?

* The Zoom to Selection button will zoom in to a specific area of your data.

1. What do you use the AutoFill feature for?

* AutoFill is used to fill cells automatically for you, when the data is identified to be in a sequential series or pattern.

1. What is one of the key components of a typical formula?

* While you can use division and percentages in your calculations in a formula, of these options, only **Reference** is a key component of a typical formula

1. What happens when you use the median calculation but select an even number of values in a range?

* Returns middle figure between the two middle values in the selected range.
* The median is a midway point between a range of values, so if you select an odd number of values, it uses the middle value, but when you select an even number of values, the median is the midway point between the two middle values.

1. What are Excel cell references by default?

* By default, in Excel, cell references are always relative, and you must manually configure a cell reference to be absolute or mixed.

1. When creating formulas, what is a mixed reference?

* A mixed reference has at least one relative cell reference and at least one absolute cell reference.