Introduction to Data Science

Renato R. Maaliw III, *DIT*

Professor 1, College of Engineering

Southern Luzon State University

Lucban, Quezon, Philippines

"To **understand** data is to possess the lens through which the world reveals its patterns, behaviors, and **truths** in the language of numbers."

Why learn data science?

In today's digital age, data is the new currency.

It powers every industry, from healthcare to finance, from entertainment to technology.

The world generates an overwhelming amount of data every second, and hidden within these vast streams of information are patterns, insights, and trends that have the potential to transform our understanding of the world

"When you study data science, you aren't just learning how to analyze numbers—you're learning **how to decode** the universe."

Data is the language in which the world speaks to us, and by mastering it, you gain the power to uncover the truths that lie beneath the surface of everyday phenomena.

Imagine being able to predict disease outbreaks, optimize business strategies, personalize learning experiences, or even contribute to solving some of the **world's most pressing challenges**, like climate change or poverty.

But the power of data science goes beyond just making predictions or finding efficiencies. It's about **storytelling**—telling stories with numbers that **can influence** decisions, inspire change, and guide the future.

Moreover, data science is a field that is in high demand. Companies and organizations around the globe are eager to find talented individuals who can help them navigate the complexities of a data-driven world.

Remember, data is **more than just numbers**—it's the key to unlocking a deeper understanding of the world.

By mastering data science, you are not just preparing for a job; you are **preparing to change the world**, one dataset at a time.

Data Science is a multidisciplinary field that combines statistical analysis, computer science, and domain-specific knowledge to extract meaningful insights from data.

It involves various techniques to **process**, **analyze**, and **interpret** large volumes of structured and unstructured data.

By leveraging algorithms, **machine learning**, and **statistical models**, you can uncover patterns and trends that inform decision-making.

In today's data-driven world, your ability to harness data effectively will be your competitive advantage

CPE15 – Programming for Data Science
CPE21 – Machine Learning 1
CPE28 – Machine Learning 2

Week 1 (Preliminaries)

- **Data science overview**
- **IDE (Setup, offline and online)**
- **Jupyter Notebook overview**
- **Python Basics – Part 1**
  a. data types, variables, list, dictionary, boolean, tuple, sets
  b. comparison operators
  c. logical operators
  d. if constructs
  e. loops

Week 2 (Numpy)

- Array manipulations: reshaping, flattening, concatenation, splitting, etc.
- Universal functions and aggregation
- Working with multidimensional arrays
- Linear Algebra operations with Numpy

Week 3 (Introduction to Pandas)

- Series & DataFrames
- Data Loading, Storage, and File Formats
- Indexing and Selecting Data
- Handling Missing Data

Week 4 (Data Manipulation with Pandas)

- DataFrame Operations: Merging, Joining, Concatenating, and Resphaping
- GroupBy Operations and Aggregations
- Functions: Mapping, Applying, and Vectorizing
- Handling Time Series Data with Pandas

Week 5 (Data Cleaning & Preprocessing)

- Data Cleaning: Duplicates, Inconsistencies, and Outliers
- Data Transformation: Scaling, Normalization, and Encoding
- Working with Text Data in Pandas
- Handling Missing Data for DataFrames

Week 6 (Introduction to Data Wrangling)

- **Data Wrangling Fundamentals**
- **Data Inspection and Profiling**
- **Common Data Wrangling Tasks and Challenges**
- **Introduction to Regular Expressions for Data Cleaning**

Week 7 (Data Wrangling Techniques)

- **Imputation Techniques**
- **Combining and Reshaping Datasets**
- **Outlier Treatment**
- **Feature Engineering for Data Science**

Week 8 (Data Extraction)

- **Extracting Data using Native Python**
- **Extraction using SQL**

Week 9 (Web Scraping)

- Introduction to Web Scraping
- Tools: BeautifulSoup, etc.
- Parsing HTML and XML data
- Ethical Considerations and Legal Issues in Web Scraping

Week 10 (Exploratory Data Analysis)

- Statistics and Distributions
- Patterns, Anomalies, and Relationships in Data
- Correlational Analysis

Week 11 (Data Visualizations – Part 1)

- **Overview of Matplotlib**
- **Plots: Line plots, Bar plots, Histograms, etc.**
- **Plot Cusomizations**

Week 12 (Data Visualizations – Part 2)

- Subplots & Grids
- More Customizations
- 3D Plotting

Week 13 (Statistical Data Visualization)

- Overview of Seaborn
- Distribution Plots: KDE Plots, Box Plots
- Relational Plots: Scatter Plots, Line Plots, Pair Plots
- Categorical Plots: Bar Plots, Count Plots, Violin Plots

Week 14 (GeoSpatial Data Visualization)

- Overview of GeoSpatial Data and GeoPandas
- Shapefiles and GeoJSON
- Plotting Geospatial with GeoPandas and Folium

Week 15 (Interactive Data Visualizations)

- Dashboards and Reports
- Best Practices for Effective Data Visualization
- End-to-End Data Science Pipeline Project

Thank you very much for listening.