Course Syllabus: Data Analysis Using Python

College: Mella Technology Services LLC
Address: 962 Wayne Ave #902
Silver Spring, MD
Greenbelt, Maryland, 20910

Authors: **Tesfay Yemane Tesfu** And **Addisu Alemneh Tesfaye**

June 17, 2025

Course Information

- **Course Title:** Introduction to Data Analysis Using Python
- **Course Duration:** 6-12 Weeks
- **Prerequisites:** Any basic programming knowledge
- **Course Level:** Beginner to Intermediate
- **Instructor:** Tesfay Yemane Tesfu
- **Contact:** tesfayphysics@gmail.com

Course Description

This course introduces the fundamentals of data analysis using Python. Students will learn how to manipulate, analyze, and visualize data using popular Python libraries such as **Pandas**, **NumPy**, **Matplotlib**, and **Seaborn**. By the end of the course, students will be able to perform exploratory data analysis (EDA), clean datasets, and generate insights from real-world data.

Key Focus Areas:

- Python programming environment fundamentals (lambdas, CSV manipulation, NumPy)
- Data manipulation/cleaning with pandas (Series/DataFrame as core structures)
- Effective use of functions like `groupby`, `merge`, and pivot tables
- Running basic inferential statistical analyses on cleaned tabular data

By the end of the course, students will be able to perform exploratory data analysis (EDA), clean datasets, and generate insights from real-world data.

Participation Strategies

- Engaged learning looks different for everybody. In this course, we hope you will:
  - Define your own measures of success
  - Engage with material in ways that suit your needs
  - Reflect on your unique skills, needs, and aspirations
- While time estimates are provided, you are empowered to:
  - Engage with content at your preferred pace
  - Focus on areas aligning with your goals
  - Utilize diverse learning approaches
- We celebrate diverse engagement methods and encourage you to find what works best for you.

Course Objectives

By the end of this course, students will be able to:

1. Understand the basics of Python for data analysis.
2. Use Pandas for data manipulation and cleaning.
3. Perform statistical analysis using NumPy and SciPy.
4. Create meaningful visualizations using Matplotlib and Seaborn.
5. Conduct exploratory data analysis (EDA) on real datasets.
6. Work with real-world datasets (CSV, Excel, JSON).

Course Outline

Week 1: Introduction to Python for Data Analysis

- Python basics (variables, loops, functions)
- Introduction to Jupyter Notebook
- Essential Python libraries (Pandas, NumPy, Matplotlib)
- Loading and inspecting datasets

Week 2: Python Data Structures for Data Analysis

- **Lists**: Creation, indexing, slicing, methods
- **Tuples**: Immutable sequences, packing/unpacking
- **Dictionaries**: Key-value pairs, methods, use cases
- **Sets**: Unique elements, set operations
- Choosing appropriate data structures
- Practical applications in data analysis

Week 3: Data Manipulation with Pandas

- Pandas DataFrames and Series
- Indexing, filtering, and sorting data
- Handling missing data
- Aggregating data (groupby, pivot tables)

Week 4: Data Cleaning & Preprocessing

- Removing duplicates and outliers
- Data transformation (normalization, encoding)
- Merging and joining datasets
- Working with dates and times

Week 5: Exploratory Data Analysis (EDA)

- Descriptive statistics (mean, median, variance)
- Correlation and covariance
- Univariate and multivariate analysis
- Hands-on EDA project

Week 6: Data Visualization

- Introduction to Matplotlib (line plots, bar charts, histograms)
- Advanced visualizations with Seaborn (heatmaps, box plots, pair plots)
- Customizing plots (labels, legends, styles)
- Storytelling with data

Week 7 - Week 10: Capstone Project & Real-World Applications

- Working with a real world dataset
- End-to-end data analysis project
- Presenting findings
- Course wrap-up

Week 11 – Week 12: Launch of Top Selected Projects

- Evaluate and compare all final trainee projects
- Identify and prioritize the top three projects
- Refine and improve the selected projects
- Officially launch the top projects on the Mella Technology platform

Assessment & Grading

- ***All assessments apply in their designated weeks***
- ***Final project evaluations occur in the last week***
- **Quizzes & Exercises (25%)** – Weekly coding exercises
- **Assignments (25%)** – Hands-on data analysis tasks
- **Final Project (50%)** – Complete EDA on a provided dataset

Recommended Resources

- **Books:**
  - *Data Science from Scratch* by Joel Grus
  - *Python for Data Analysis* by Wes McKinney
- **Online:**
  - **Support Code:** Available at (Tesfay-Tesfu)
  - Pandas Documentation (Pandas)
  - Kaggle for datasets and competitions (Kaggle)

Tools & Software

- Python Version greater than 3.7 will be convenient
- Better to install Jupyter Notebook or you can use free Google Colab
- Most useful Libraries: NumPy, Pandas, Matplotlib, Seaborn, and Basemab toolkits

Other Advanced Data-Analysis Packages

Data Analysis Using Complicated Data Formats:

- netcdf files
- Cmn files
- Fits files

Mapping and 3D simulations using:

- Basemap toolkits; Used to simulate various interesting simulations:
  - **Flow Maps** for trade routes/capital flows
  - **3D Economic Landscapes** using elevation for metrics
  - **Time-Series Animations** of economic changes
  - **Interactive Web Visualizations** using Folium
  - **Regional economic disparity analysis**
  - **Stock market performance by geographic region**
  - **GDP growth mapping by country/region**
  - **Mapping disease spread patterns**
  - **Visualizing voting patterns by geographic region**
  - **Weather Analysis (rain, hurricane, storm, wind, etc)**

Python Integration with Scientific Writing Platforms:

- LaTeX
- Kile
- Other scientific writing platforms

Course Topics Include in-person training :

- Advanced Python techniques for scientific writing
- Real-time data scraping methods
- Real-time data analysis and streaming
- Secure and encrypted data analysis approaches
- Interactive analysis
- Wavelet Analysis
- Python-based SQL, Pandas, and Matplotlib analysis
- Basic Practice on how to Develop Python Libraries and launch in PyPI