

Course Title: Data Analysis Course Syllabus for Beginners and Intermediate

Course Description: This course provides an introduction to data analysis for beginners and intermediate learners. Students will learn fundamental concepts, techniques, and tools for collecting, cleaning, exploring, and visualizing data. The course covers both theoretical foundations and practical applications of data analysis using Python and popular libraries such as pandas, NumPy, and Matplotlib.

Prerequisites: Basic knowledge of programming concepts and familiarity with Python programming language. No prior experience in data analysis is required.

Course Objectives:

1. Understand the fundamentals of data analysis and its importance in decision-making processes.
2. Learn how to collect, clean, and preprocess data for analysis.
3. Gain proficiency in using Python libraries such as pandas and NumPy for data manipulation and analysis.
4. Develop skills in data visualization using Matplotlib and other visualization libraries.
5. Explore exploratory data analysis (EDA) techniques and statistical methods for deriving insights from data.

Course Outline:**Module 1: Introduction to Data Analysis**

- Overview of data analysis and its applications
- Introduction to Python for data analysis
- Setting up Python environment and Jupyter Notebook

Module 2: Data Collection and Cleaning

- Data types and formats
- Data collection methods (web scraping, APIs, databases)
- Data cleaning techniques (handling missing values, duplicates, outliers)

Module 3: Introduction to Pandas

- Introduction to pandas library
- Series and DataFrame objects
- Data manipulation and transformation with pandas

Module 4: Data Visualization with Matplotlib

- Introduction to Matplotlib library
- Basic plots (line plots, bar plots, scatter plots)
- Customizing plots and adding annotations

Module 5: Exploratory Data Analysis (EDA)

- Understanding the structure and distribution of data
- Summary statistics and descriptive analysis
- Correlation analysis and heatmap visualization

Module 6: Data Analysis with NumPy

- Introduction to NumPy library
- Array creation and manipulation
- Mathematical operations and statistical functions

Module 7: Advanced Data Analysis Techniques

- Grouping and aggregation with pandas
- Time series analysis and visualization
- Introduction to machine learning for data analysis

Module 8: Data Visualization with Seaborn

- Introduction to Seaborn library for statistical data visualization
- Advanced plots (box plots, violin plots, pair plots)
- Visualizing relationships and patterns in data

Module 9: Introduction to Statistical Analysis

- Basic concepts of statistical analysis
- Hypothesis testing and p-values
- Regression analysis and model fitting

Module 10: Real-world Data Analysis Projects

- Working on real-world datasets
- Applying data analysis techniques learned throughout the course
- Presenting findings and insights from data analysis projects

Module 11: Intermediate Data Analysis

- Advanced data cleaning techniques
- Handling large datasets with pandas and Dask
- Advanced statistical analysis and machine learning concepts

Module 12: Capstone Project

- Developing a comprehensive data analysis project
- Identifying a research question or problem statement
- Collecting, cleaning, analyzing, and visualizing data to derive insights

Assessment:

- Weekly assignments to reinforce learning concepts.
- Midterm project: Performing exploratory data analysis on a given dataset.
- Final project: Developing a complete data analysis project from data collection to visualization and presenting findings.

Textbook: "Python for Data Analysis" by Wes McKinney

Additional Resources:

- Online tutorials and documentation (pandas documentation, NumPy documentation, Matplotlib documentation, etc.).
- Supplemental readings and materials provided by the instructor.

Grading:

- Assignments: 30%
- Midterm Project: 20%
- Final Project: 40%
- Participation and Attendance: 10%

Attendance Policy: Regular attendance is expected. Students are allowed a maximum of three unexcused absences. Excessive absences may result in a reduction of the final grade.

Office Hours: Instructor office hours will be held twice a week for additional help and clarification.