Government of Pakistan

National Vocational and Technical Training Commission

Prime Minister Youth Skills Development Program

"Skills for All"

Course Contents / Lesson Plan
Course Title: Data Analysis using Python
Duration: 2 Months

Trainer Name	
Author Name	**Dr. Fawad Salam Khan (Air University, Islamabad)** **Muhammad Nasir Khan (DACUM Facilitator, Ex-DD VT, SS&C Wing, NAVTTC, Islamabad**
Course Title	**Data Analysis using Python**
Objectives and Expectations	**Course Objectives** By the end of the "Data Analysis using Python" course, students will: 1. **Understand Python Programming**: Gain proficiency in Python programming language, focusing on data analysis tasks, including working with essential libraries such as NumPy, Pandas, Matplotlib, and Seaborn. 2. **Master Data Manipulation**: Develop the skills to efficiently manipulate data using Python, including importing, cleaning, preprocessing, and transforming data to prepare it for analysis. 3. **Implement Data Cleaning and Preprocessing**: Learn various techniques for data cleaning, including handling missing values, normalizing data, encoding categorical variables, and performing feature engineering. 4. **Create Effective Data Visualizations**: Be able to create both basic and advanced data visualizations using Matplotlib, Seaborn, and Plotly, making complex data more accessible and understandable through visual representation. 5. **Conduct Exploratory Data Analysis (EDA)**: Perform comprehensive exploratory data analysis to uncover patterns, correlations, and insights within datasets, laying the groundwork for more advanced statistical analysis or machine learning. 6. **Develop Interactive Dashboards**: Gain the ability to create interactive data visualizations and dashboards using Plotly and Dash, enhancing the presentation and exploration of data in a user-friendly manner. 7. **Apply Knowledge to Real-World Problems**: Synthesize all the skills learned in the course to conduct a complete analysis on a real-world dataset, from data cleaning to visualization and interpretation, culminating in a capstone project. 8. **Communicate Findings Effectively**: Learn to present data analysis results through well-structured reports, incorporating visualizations and statistical summaries that effectively communicate key insights. **Course Expectations** Students enrolled in the course are expected to: 1. **Engage Actively in Learning**: Attend lectures, participate in discussions, and engage with lab sessions to fully grasp the course content. 2. **Complete Assignments Promptly**: Submit weekly assignments on time, ensuring that they demonstrate an understanding of the week's material and practical application skills. 3. **Work Independently and Collaboratively**: While individual assignments will assess personal understanding, students should also engage in peer

	discussions and group activities when applicable, to enhance learning through collaboration. 4. **Practice Consistently**: Regularly practice coding and data analysis outside of class hours, utilizing provided datasets and recommended resources to reinforce learning. 5. **Ask Questions and Seek Help**: Be proactive in seeking clarification on challenging topics, either through discussion forums, during lab sessions, or directly from the instructor. 6. **Apply Critical Thinking**: Approach data analysis tasks with a critical mindset, questioning assumptions, considering alternative methods, and validating results to ensure accuracy and reliability. 7. **Adhere to Ethical Standards**: Uphold ethical standards in data analysis, including respecting data privacy, acknowledging sources, and presenting analysis results honestly and transparently. 8. **Complete the Capstone Project**: Dedicate sufficient time and effort to the capstone project, which is a significant portion of the final grade, ensuring it reflects a comprehensive understanding of the course material.
Entry-level of trainees	**Prerequisites:** • **Basic Computer Literacy**: Students should be comfortable using a computer, navigating software, and managing files. • **Basic Understanding of Mathematics**: A foundational knowledge of basic mathematical concepts, such as algebra and statistics, is beneficial but not mandatory. • **No Prior Programming Experience Required**: This course is designed to accommodate those who are new to programming and Python, though individuals with some programming experience may find it easier to grasp the initial concepts. **Target Audience:** • Individuals interested in learning data analysis, including students, professionals, and enthusiasts from various fields. • Beginners who want to start their journey in data science and Python programming. • Intermediate learners who want to enhance their data manipulation and analysis skills using Python.

Learning Outcomes of the course	**Learning Outcomes**
	
	By the end of the "Data Analysis using Python" course, students will be able to:
	
	1. **Understand and Utilize Python for Data Analysis**:
	  o Develop a strong foundation in Python programming, including the use of key libraries such as NumPy, Pandas, Matplotlib, and Seaborn.
	  o Write Python scripts to perform various data analysis tasks.
	2. **Efficiently Manipulate and Process Data**:
	  o Import, clean, preprocess, and transform datasets using Python.
	  o Handle missing data, perform data normalization and standardization, and apply feature engineering techniques.

	3. **Create and Interpret Data Visualizations**: o Produce a variety of data visualizations (e.g., line plots, bar plots, histograms, heatmaps) using Matplotlib and Seaborn. o Use Plotly to create interactive visualizations and Dash to develop data dashboards. 4. **Perform Comprehensive Exploratory Data Analysis (EDA)**: o Conduct exploratory data analysis to identify patterns, correlations, and key insights within datasets. o Apply statistical techniques such as descriptive statistics and hypothesis testing during the EDA process. 5. **Develop and Present Data-Driven Insights**: o Integrate data cleaning, preprocessing, visualization, and analysis skills to analyze real-world datasets. o Communicate data-driven insights through well-structured reports and presentations that incorporate visualizations and statistical summaries. 6. **Build Interactive Data Dashboards**: o Create and deploy interactive dashboards using Dash, enabling users to explore data dynamically. 7. **Apply Python Skills to Real-World Projects**: o Complete a capstone project that involves cleaning, analyzing, and visualizing a real-world dataset, demonstrating the ability to apply learned skills in a practical context. 8. **Work Independently on Data Analysis Projects**: o Develop the confidence and competence to undertake independent data analysis projects, from data acquisition to presentation of findings.
Course Execution Plan	**Course Duration**: 8 Weeks (2 Months) **Course Level**: Beginner to Intermediate **Total Hours**: 40 Hours (5 Hours per Week) **Delivery Mode**: Lectures, Hands-on Lab Sessions, and Assignments
Companies offering jobs in the respective trade	**1. Software Houses and IT Companies** • **NetSol Technologies**: A leading IT company offering services in software development, data analysis, and IT consulting. They frequently hire data analysts, Python developers, and data scientists. • **Systems Limited**: A well-known IT services company providing solutions in data analytics, business intelligence, and software development. • **10Pearls**: A global technology company with a significant presence in Pakistan, focusing on digital transformation, including data analytics and AI solutions. • **Afiniti**: A pioneer in AI and big data, offering opportunities in data analysis and data science.

Data Analysis using Python

2. Telecommunication Companies

- **Telenor Pakistan**: A major telecom operator that hires data analysts and business intelligence professionals to analyze customer data and improve service delivery.
- **Jazz (Mobilink)**: One of Pakistan's largest telecom companies, offering roles in data analysis, customer insights, and data-driven decision-making.
- **Zong**: A leading telecom provider that uses data analytics to enhance customer experiences and optimize operations.

3. Financial Services and Banks

- **Habib Bank Limited (HBL)**: Pakistan's largest bank, often recruiting data analysts and financial analysts to support their data-driven strategies.
- **United Bank Limited (UBL)**: A major bank in Pakistan that leverages data analytics for risk management, customer insights, and financial modeling.
- **Meezan Bank**: Pakistan's leading Islamic bank, offering opportunities in data analysis, especially in the areas of financial performance and customer behavior analysis.

4. E-Commerce and Retail

- **Daraz.pk**: The largest online marketplace in Pakistan, frequently hiring data analysts and business intelligence professionals to enhance their e-commerce platform.
- **Foodpanda Pakistan**: A prominent food delivery service that relies heavily on data analytics to optimize operations, marketing strategies, and customer experience.
- **Careem**: A ride-hailing service that uses data to improve its operations, customer satisfaction, and service delivery.

5. FMCG Companies

- **Unilever Pakistan**: A global FMCG giant with a significant presence in Pakistan, often hiring data analysts to support market research, sales forecasting, and supply chain optimization.
- **Nestlé Pakistan**: A major player in the FMCG sector, offering roles in data analysis for market research, product development, and operational efficiency.

6. Consulting and Analytics Firms

- **KPMG Taseer Hadi & Co.**: A global professional services firm offering audit, tax, and advisory services, including data analytics roles.
- **PwC Pakistan**: Part of the global PwC network, this firm offers opportunities in data analytics, financial modeling, and business intelligence.
- **Arbisoft**: A technology consulting firm that offers data analytics services, often recruiting data scientists and analysts.

7. Tech Startups

- **Airlift Technologies**: A tech startup focusing on logistics and transportation, leveraging data analytics to optimize operations and customer experience.

	• **Bykea**: A local ride-hailing and delivery startup that uses data analysis to enhance service efficiency and customer satisfaction. • **Bazaar Technologies**: A B2B e-commerce platform for small businesses in Pakistan, relying on data analytics for market insights and operational decision-making. **8. Healthcare and Pharmaceutical Companies** • **Siemens Healthineers**: A global healthcare company with a strong focus on data-driven healthcare solutions, including opportunities in data analytics. • **GlaxoSmithKline (GSK) Pakistan**: A leading pharmaceutical company that uses data analysis for market research, product development, and supply chain management.
Job Opportunities	**1. Python Developer (with a focus on Data Analysis)** • **Role**: Develop and maintain Python scripts and applications that automate data collection, processing, and analysis tasks. • **Skills Utilized**: Python programming, data manipulation, automation scripts, and integration of data analysis libraries. **2. Research Analyst** • **Role**: Perform data-driven research, analyze datasets to support academic or industry research projects, and present findings. • **Skills Utilized**: Data analysis, statistical testing, data visualization, and reporting. **3. Financial Analyst** • **Role**: Analyze financial data, create financial models, and provide insights into market trends, investment opportunities, and risk management. • **Skills Utilized**: Python for financial data analysis, data cleaning, and advanced data visualization. **4. Marketing Analyst** • **Role**: Analyze marketing data, including customer behavior, sales trends, and campaign effectiveness, to improve marketing strategies. • **Skills Utilized**: Data analysis, segmentation, trend analysis, and visualization using Python. **5. Operations Analyst** • **Role**: Analyze operational data to improve efficiency, optimize processes, and reduce costs within an organization. • **Skills Utilized**: Data manipulation, performance metrics analysis, and reporting using Python.

Data Analysis using Python

	6. Data Visualization Specialist • **Role**: Focus on creating effective and visually appealing data visualizations and dashboards to present complex data insights in an understandable way. • **Skills Utilized**: Matplotlib, Seaborn, Plotly, and Dash for creating visualizations and dashboards. **7. Entry-Level Machine Learning Engineer** • **Role**: Work with data scientists to prepare data for machine learning models, perform EDA, and assist in developing basic machine learning algorithms. • **Skills Utilized**: Python programming, data preprocessing, and basic understanding of machine learning workflows.
No of Students	25
Learning Place	Classroom / Lab
Instructional Resources	1. **"Python for Data Analysis" by Wes McKinney** o A comprehensive guide to using Python libraries like Pandas and NumPy for data manipulation and analysis. o Amazon 2. **"Automate the Boring Stuff with Python" by Al Sweigart** o Excellent for beginners, this book covers practical Python applications, including data manipulation tasks. o Automate the Boring Stuff 3. **"Practical Statistics for Data Scientists" by Peter Bruce and Andrew Bruce** o Focuses on essential statistical concepts using Python and R for data analysis. o O'Reilly 4. **"Hands-On Data Analysis with Pandas" by Stefanie Molin** o Detailed examples and exercises using Pandas for data manipulation, cleaning, and analysis. o Packt **Tutorial Websites** 1. **Kaggle** o Offers Python tutorials, datasets for practice, and a community for data science enthusiasts. Great for practical, hands-on learning. o Kaggle 2. **Real Python** o A comprehensive resource for learning Python, including tutorials on data analysis, web scraping, and data visualization. o Real Python 3. **W3Schools Python Tutorial** o Beginner-friendly tutorials that cover Python programming basics and data analysis topics. o W3Schools 4. **Towards Data Science** o A popular Medium publication with articles and tutorials on Python, data analysis, machine learning, and more. o Towards Data Science

	5. **GeeksforGeeks: Python Programming** o Provides a wide range of Python tutorials, from basic to advanced, with a focus on data structures, algorithms, and data science. o GeeksforGeeks

MODULES

Scheduled Weeks	Module Title	Learning Units	Home Assignment
Week 1	**1.1 Introduce Python for Data Analysis**	**1.1 Introduce Python for Data Analysis** • Overview of Python programming language • Setup Python environment (Anaconda, Jupyter Notebook) • Introduction to Python libraries: NumPy, Pandas, Matplotlib, Seaborn • Interpret Python syntax and operations (variables, data types, loops, functions) **1.2 Working with Data in Python** • Introduce to data structures: Lists, Tuples, Dictionaries, and Sets • Introduction to NumPy: Arrays, array operations, and basic mathematical functions • Introduction to Pandas: Series and DataFrames • Importing and exporting data (CSV, Excel, JSON) **Lab Session:** • Setting up Jupyter Notebook for data analysis Basic data manipulation with NumPy and Pandas	**Assignment 1:** • Basic data manipulation using Pandas: Creating, reading, and writing Data Frames

Week 2	**Working with Data in Python**	**3.1 Data Cleaning Techniques** • Handling missing data (drop, fillna, interpolation) • Data normalization and standardization • Detecting and treating outliers • Data transformation (log, square root, etc.) **3.2 Data Preprocessing** • Deal with categorical data: Encoding techniques (One-Hot, Label Encoding) • Handling date and time data • Data binning and discretization • Feature selection and extraction **Lab Session:** • Practical data cleaning and preprocessing exercises using Pandas	**Assignment 2:** • Create a Python script that reads a data file, processes the data, and outputs results

Week 3	**Advanced- Data Cleaning and Feature Engineering**	**3.1 Data Cleaning Techniques** • Handle missing data (drop, fillna, interpolation) • Data normalization and standardization • Detect and treat outliers • Data transformation (log, square root, etc.) **3.2 Data Preprocessing** • Deal with categorical data: Encoding techniques (One-Hot, Label Encoding) • Handling date and time data • Data binning and discretization • Feature selection and extraction **Lab Session:** • Practical data cleaning and preprocessing exercises using Pandas	**Assignment 3:** • Cleaning and preprocessing a dataset (handling missing data, encoding categorical variables, normalizing features)

Week 4	**Advanced Data Cleaning and Feature Engineering**	**4.1 Advanced Feature Engineering** • Creating new features from existing data • Feature scaling and polynomial features • Interaction features **4.2 Handling Large Datasets** • Working with large datasets in Pandas • Optimizing memory usage and performance • Introduction to Disk for handling large-scale data **Lab Session:** • Feature engineering and handling large datasets	**Assignment 4:** • Engineer new features for a dataset and analyze their impact on data analysis

Week 5	**Data Visualization**	**5.1 Introduction to Data Visualization** • Importance of data visualization in data analysis • Overview of Matplotlib and Seaborn libraries **5.2 Basic Plotting with Matplotlib** • Create simple plots: Line plot, Bar plot, Histogram • Customizing plots: Titles, labels, legends, and colors • Subplots and figure layouts **5.3 Advanced Data Visualization with Seaborn** • Create advanced plots: Heatmaps, Box plots, Pair plots • Visualizing distributions and correlations • Plot aesthetics and customization **Lab Session:** • Hands-on practice with Matplotlib and Seaborn for data visualization	**Assignment 5:** • Creating visualizations to analyze trends, distributions, and relationships in a given dataset

Week 6	**Interactive Visualizations and Dashboarding**	**6.1 Interactive Visualizations** • Introduction to Plotly for interactive visualizations • Create interactive plots: Scatter plots, Line charts, and more • Customizing interactive visualizations **6.2 Dashboarding with Dash** • Introduction to Dash for creating dashboards • Build a simple data dashboard • Deploying dashboards for data analysis **Lab Session:** • Create a interactive visualizations and dashboards	**Assignment 6:** • Develop a simple dashboard to visualize a dataset interactively
Week 7	**Exploratory Data Analysis (EDA)**	**7.1 Introduction to Exploratory Data Analysis (EDA)** • Understand the importance of EDA • Steps in performing EDA • Identifying patterns, correlations, and insights from data **7.2 EDA Techniques and Best Practices** • Descriptive statistics: Mean, median, mode, standard deviation • Correlation analysis and covariance • Hypothesis testing basics • Identifying and interpreting trends **Lab Session:** • Performing a complete EDA process on a sample dataset	**Assignment 7:** • Conduct an EDA on a provided dataset and report the findings

Week 8	**Capstone Project and Case Study**	**8.1 Case Study: Real-World Data Analysis** • Applying data cleaning, preprocessing, visualization, and EDA on a real-world dataset • Report writing: Presenting findings and insights through visualizations and statistical summaries **8.2 Capstone Project Development** • Students work on a capstone project to apply all learned concepts • Guidance on project structuring and report writing **Lab Session:** • Capstone project development with instructor guidance	**Final Project Submission:** • Submission of the final project report including data cleaning, preprocessing, visualization, and EDA findings.

Practical Tasks:

	Task	Description	Week
1	**Basic data manipulation using Pandas: Creating, reading, and writing Data Frames**	● The goal of this task is to practice basic data manipulation using the Pandas library in Python. You will learn how to create, read, and write Data Frames, which are essential for handling and analyzing structured data. This exercise will help you build foundational skills in working with data in Python.	Week 1
2	**Create a Python script that reads a data file, processes the data, and outputs results**	● The task is to develop a Python script that reads a data file, processes the data through various stages of cleaning and transformation, and outputs the results in a specified format. This exercise is designed to enhance your skills in Python programming, data manipulation, and exploratory data analysis.	Week 2
3	**Practical data cleaning and preprocessing exercises using Pandas**	● The aim of this task is to perform practical data cleaning and preprocessing exercises using the Pandas library in Python. This exercise will help you develop essential skills in preparing raw data for analysis, ensuring that the data is clean, consistent, and ready for further exploration or modeling.	Week 3
4	**Engineer new features for a dataset and analyse their impact on data analysis**	● The goal of this task is to engineer new features for an existing dataset and analyze how these new features impact the overall data analysis. Feature engineering is a critical step in data preprocessing that can significantly enhance the predictive power of your models, and the insights derived from your data.	Week 4
5	**Creating visualizations to analyze trends, distributions, and relationships in a given dataset**	● The objective of this task is to create visualizations that effectively analyze and present trends, distributions, and relationships within a given dataset. Visualizations are a powerful tool for uncovering insights and communicating findings in an intuitive	Week 5

		and impactful way.	
6	**Develop a simple dashboard to visualize a dataset interactively**	● Create an interactive dashboard to visualize a dataset involves several steps, from data preparation to designing the dashboard itself. Below is a high-level outline of how you can develop a simple interactive dashboard using Python	Week 6
7	**Conduct an EDA on a provided dataset and report the findings**	● The objective of this Exploratory Data Analysis (EDA) is to understand the underlying patterns, distributions, and relationships within the provided dataset. EDA will help identify any anomalies, trends, or insights that could inform subsequent data processing and model-building phases.	Week 7
8	**Submit of the final project report including data cleaning, preprocessing, visualization, and EDA findings**	● The purpose of this task is to perform a comprehensive Exploratory Data Analysis (EDA) on the provided dataset. This process aims to uncover underlying patterns, relationships, and anomalies within the data, which will be crucial for informing subsequent stages of data processing and model development..	Week 8

Workplace/Institute Ethics Guide

Work ethic is a standard of conduct and values for job performance. The modern definition of what constitutes good work ethics often varies. Different businesses have different expectations. Work ethic is a belief that hard work and diligence have a moral benefit and an inherent ability, virtue, or value to strengthen character and individual abilities. It is a set of values-centered on the importance of work and manifested by determination or desire to work hard.

The following ten work ethics are defined as essential for student success:

1. **Attendance:**
   Be at work every day possible, plan your absences don't abuse leave time. Be punctual every day.
2. **Character:**
   Honesty is the single most important factor having a direct bearing on the final success of an individual, corporation, or product. Complete assigned tasks correctly and promptly. Look to improve your skills.

Data Analysis using Python

3. **Team Work:**

   The ability to get along with others including those you don't necessarily like. The ability to carry your weight and help others who are struggling. Recognize when to speak up with an idea and when to compromise by blend ideas together.

4. **Appearance:**

   Dress for success set your best foot forward, personal hygiene, good manner, remember that the first impression of who you are can last a lifetime

5. **Attitude:**

   Listen to suggestions and be positive, accept responsibility. If you make a mistake, admit it. Values workplace safety rules and precautions for personal and co-worker safety. Avoids unnecessary risks. Willing to learn new processes, systems, and procedures in light of changing responsibilities.

6. **Productivity:**

   Do the work correctly, quality and timelines are prized. Get along with fellows, cooperation is the key to productivity. Help out whenever asked, do extra without being asked. Take pride in your work, do things the best you know-how. Eagerly focuses energy on accomplishing tasks, also referred to as demonstrating ownership. Takes pride in work.

7. **Organizational Skills:**

   Make an effort to improve, learn ways to better yourself. Time management; utilize time and resources to get the most out of both. Take an appropriate approach to social interactions at work. Maintains focus on work responsibilities.

8. **Communication:**

   Written communication, being able to correctly write reports and memos. Verbal communications, being able to communicate one on one or to a group.

9. **Cooperation:**

   Follow institute rules and regulations, learn and follow expectations. Get along with fellows, cooperation is the key to productivity. Able to welcome and adapt to changing work situations and the application of new or different skills.

10. **Respect:**

    Work hard, work to the best of your ability. Carry out orders, do what's asked the first time. Show respect, accept, and acknowledge an individual's talents and knowledge. Respects diversity in the workplace, including showing due respect for different perspectives, opinions, and suggestions.