DATA SCIENCE COURSE SYLLABUS

Introduction to Python
- Introduction to python
- Environment setup & start programming
- Python Conditional Statements, Loops and File Handling
- Core Objects and Advanced Data Structures; Functions and Lambdas
- The Object Oriented Side of it

Module 1: Introduction to Data Analysis with Python
- What is data analysis?
- Overview of Python and its data analysis libraries (NumPy, pandas, Matplotlib, Seaborn)
- Setting up your Python environment

Module 2: Data Preprocessing and Cleaning
- Importing data from various sources (CSV, Excel, SQL)
- Exploring and understanding the dataset
- Handling missing data: imputation techniques
- Dealing with outliers and anomalies
- Data transformation: normalization, standardization
- Data integration and manipulation using pandas

Module 3: Exploratory Data Analysis (EDA)
- Descriptive statistics: mean, median, mode, variance, etc.
- Histograms, box plots, scatter plots
- Correlation analysis and heatmaps
- Univariate and bivariate analysis
- Data visualization using Matplotlib and Seaborn

Module 4: Data Visualization
- Advanced data visualization techniques: bar plots, line plots, pie charts, etc.
- Interactive visualizations using Plotly
- Geospatial visualization
- Effective data storytelling and communication

Module 5: Statistical Analysis
- Sampling techniques and the Central Limit Theorem
- Hypothesis testing: t-tests, chi-square tests, ANOVA
- Confidence intervals and p-values

- Interpreting statistical results

Module 6: Machine Learning Algorithms for Data Analysis
- Introduction to machine learning
- Feature engineering and selection
- Linear regression: simple and multiple regression
- Logistic regression for classification
- Decision trees and random forests
- Model evaluation metrics: R-squared, MAE, RMSE, accuracy, precision, recall, F1-score
- Model assumptions and diagnostics

Module 7: Time Series Analysis
- Introduction to time series data
- Time series components: trend, seasonality, noise
- Decomposition techniques
- Time series forecasting methods: moving average, ARIMA, exponential smoothing
- Implementing time series analysis in Python

Module 8: Advanced Topics
- Dimensionality reduction techniques (PCA, t-SNE)
- Clustering algorithms (K-means, hierarchical clustering)
- Advanced statistical techniques (non-parametric tests, ANCOVA)

Module 9: Real-World Projects and Case Studies
- Applying data analysis concepts to real datasets
- Solving data analysis challenges and problems
- Creating a portfolio of data analysis projects