Applied Data Science With Python

Module 1: Foundations of Data Science

- Introduction to Data Science: Concepts, scope, relevance in digital ecosystems

- Growth of data-centric solutions in industry and research

- Data Science Workflow: Data acquisition → preprocessing → exploratory study → model building → results interpretation

- Key Professional Roles: Data Architect, ML Engineer, Business Data Analyst, Data Visualization Specialist

- Industry Use Cases: Smart retail analytics, digital health monitoring, intelligent credit scoring

- **Learning Task:** Discussion on "How Data Science influences automated decision-making in modern businesses."

Module 2: Python Programming for Data Manipulation

- Core Python constructs essential for MCA students

- Python Libraries for Data Intelligence: NumPy for computation, Pandas for structured data handling, Matplotlib & Seaborn for visual analytics

- Working with DataFrames: Importing, parsing, cleaning, transforming, labelling

- Managing missing entries, noisy values, and data inconsistencies

- Data operations: Filtering, merging, aggregation, encoding, type casting

- **Hands-on Exercise:** Implement data wrangling on a dataset such as **online food delivery orders** or **ride-sharing logs**

Module 3: Data Exploration and Insight Extraction(EDA)

- Understanding data attributes: Categorical vs Numerical

- Descriptive analysis, correlation study, outlier exploration, and feature interactions

- Visual Analytics: Histograms, KDE plots, scatter diagrams, comparative bar charts, correlation heatmaps

- Extracting hidden trends and behaviour patterns

- **Hands-on Project:** Conduct **EDA** on datasets such as **student academic performance**, **driver trip patterns**, or **product review logs**

Module 4: Statistical Computing for Data Analysis

- Essentials of Statistical Thinking: Mean, median, mode, dispersion, skewness

- Sampling strategies and inferential foundations

- Probability Distributions: Gaussian, Poisson, and categorical models

- Hypothesis building and testing: t-tests, chi-square, and ANOVA concepts

- Confidence intervals, significance values, and interpretation

- **Lab Task:** Perform analytics on datasets like **LMS login activity** or **hardware performance metrics** using SciPy and Python's statistics modules

---

Module 5: Machine Learning Algorithms & Predictive Modelling

- Role of Machine Learning in Intelligent Systems

- Data preparation, feature construction, encoding strategies

- Supervised Learning Techniques: Linear Regression, Logistic Regression, Decision Trees

- Unsupervised Learning: Clustering models such as K-Means

- Model Evaluation: RMSE, accuracy, recall, and confusion matrix interpretation

- **Hands-on:** Develop ML solutions for **smart building energy prediction**, **e-commerce return fraud detection**, or **telecom churn forecasting** using Scikit-Learn

---

Module 6: Visual Narratives & Analytical Storytelling

- Principles of effective information design

- Building interactive visual analytics through Plotly, Power BI, and dashboards

- Structuring insights for managerial interpretation

- Data storytelling methods for strategic alignment and decision clarity

- **Hands-on:** Create a dashboard for **IoT-based smart parking occupancy**, **student engagement insights**, or **App rating sentiment visualisation**

Module 7: Data Platforms, Tools & Collaborative Ecosystems

- Introduction to cloud-based data science platforms

- Working environment setup using **Google Colab** / **Azure ML Studio**

- Accessing and integrating public repositories like Kaggle, UCI Machine Learning Repository

- Team collaboration and code versioning using GitHub

- **Activity:** Configure and deploy a mini data project on a cloud notebook environment

---

Capstone Project (Mini Project)

Objective: Implement an end-to-end Data Science solution addressing a real-world challenge

Possible Themes:
 Choose **ONE** from the following ideas:

1 **Smart Campus Resource Optimizer** – Predict utilisation of labs/classrooms using device logs
2 **Wearable Health Insights** – Analyse sleep/step/calorie patterns for advisory analytics
3 **Transit Demand Predictor** – Forecast metro/bus crowding patterns
4 **EdTech Learning Recommender** – Suggest study materials based on LMS usage data