GSP Focus

Tenure: 3rd Year Students; 4 Semesters

# Cohort 1 (Semester 5): Introduction to Data Science and Python Programming (84 Hours)

Total Days: 12 days | Hours per Day: 7 hours

Course Outline:

1. Introduction to Data Science and Data Analytics (1 day)
   * Overview of data science and analytics
   * Real-world applications and industry trends
   * Understanding the data science workflow
2. Computational Thinking (1 day)
   * Problem-solving strategies
   * Algorithmic thinking and logic
   * Decomposition and pattern recognition
3. Excel for Data Analysis (3 days)
   * Day 1:
     + Excel basics refresher
     + Data entry, formatting, and basic formulas
   * Day 2:
     + Advanced functions and formulas
     + Data sorting, filtering, and conditional formatting
   * Day 3:
     + PivotTables and PivotCharts
     + Data visualization techniques
     + Introduction to Excel macros
4. Introduction to Python Programming (6 days)
   * Day 1:
     + Python setup and environment
     + Basic syntax, variables, and data types
   * Day 2:
     + Control structures: loops and conditionals
     + Functions and scope
   * Day 3:
     + Data structures: lists, tuples, dictionaries, and sets
   * Day 4:
     + File handling and I/O operations
     + Error handling and exceptions
   * Day 5:
     + Modules and packages
     + Working with external libraries
   * Day 6:
     + Introduction to NumPy and pandas for data manipulation
5. Mini Project (1 day)
   * Project Description:
     + Develop a basic data analysis report using Excel and Python.
     + Tasks include data cleaning, basic analysis, and simple visualization.
   * Activities:
     + Project work (5 hours)
     + Presentation and feedback session (2 hours)

# Cohort 2 (Semester 6): Data Analysis and Visualization with Python (84 Hours)

Total Days: 12 days | Hours per Day: 7 hours

Course Outline:

1. Advanced Python for Data Analysis (3 days)
   * Day 1:
     + Object-oriented programming concepts
     + Classes, objects, and methods
   * Day 2:
     + Advanced data structures
     + List comprehensions and generators
   * Day 3:
     + Working with dates and times
     + Regular expressions
     + Introduction to functional programming
2. Data Manipulation with Pandas and NumPy (4 days)
   * Day 1:
     + Pandas Series and DataFrames
     + Importing and exporting data (CSV, Excel, JSON)
   * Day 2:
     + Data cleaning and preprocessing
     + Handling missing data and duplicates
   * Day 3:
     + Data transformation and normalization
     + Grouping, aggregating, and pivoting data
   * Day 4:
     + Merging, joining, and concatenating datasets
     + Time-series data analysis
3. Data Visualization with Matplotlib and Seaborn (3 days)
   * Day 1:
     + Basic plots with Matplotlib (line, bar, scatter)
     + Customizing plots (titles, labels, annotations)
   * Day 2:
     + Advanced visualizations with Seaborn (heatmaps, pair plots)
     + Styling and themes
   * Day 3:
     + Interactive visualizations with Plotly
     + Creating dashboards and reports
4. Introduction to Databases and SQL (1 day)
   * Relational database concepts
   * Basic SQL queries (SELECT, WHERE, ORDER BY)
   * Joining tables and subqueries
   * Integrating SQL with Python using SQLite and SQLAlchemy
5. Intensive Mini Project (1 day)
   * Project Description:
     + Perform an end-to-end data analysis project using Python.
     + Tasks include data extraction (possibly from databases), cleaning, analysis, and advanced visualization.
   * Activities:
     + Project work (5 hours)
     + Presentation and peer review (2 hours)
6. Additional Practice and Revision (Optional sessions spread over the cohort)
   * Hands-on exercises and case studies
   * Q&A sessions and doubt clearing

# Cohort 3 (Semester 7): Machine Learning Fundamentals and Applications (105 Hours)

Total Days: 15 days | Hours per Day: 7 hours

Course Outline:

1. Statistics for Data Science (3 days)
   * Day 1:
     + Descriptive statistics (mean, median, mode)
     + Measures of dispersion (variance, standard deviation)
   * Day 2:
     + Probability concepts
     + Probability distributions (normal, binomial, Poisson)
   * Day 3:
     + Hypothesis testing
     + Confidence intervals and p-values
     + Correlation and regression analysis
2. Machine Learning Fundamentals (3 days)
   * Day 1:
     + Introduction to machine learning concepts
     + Supervised vs. unsupervised learning
     + Machine learning pipeline
   * Day 2:
     + Data preprocessing for machine learning
     + Feature scaling and encoding
     + Handling imbalanced datasets
   * Day 3:
     + Model evaluation metrics
     + Cross-validation techniques
     + Bias-variance trade-off
3. Supervised Learning Techniques (4 days)
   * Day 1:
     + Regression algorithms (Linear Regression, Polynomial Regression)
     + Model implementation and interpretation
   * Day 2:
     + Classification algorithms (Logistic Regression, Naive Bayes)
     + Decision Trees and Random Forests
   * Day 3:
     + Support Vector Machines
     + K-Nearest Neighbors
   * Day 4:
     + Ensemble methods (Bagging, Boosting)
     + Gradient Boosting Machines (XGBoost, LightGBM)
4. Unsupervised Learning Techniques (3 days)
   * Day 1:
     + Clustering algorithms (K-Means, Hierarchical Clustering)
     + Evaluating clustering performance
   * Day 2:
     + Density-based clustering (DBSCAN)
     + Applications of clustering in real-world scenarios
   * Day 3:
     + Dimensionality reduction techniques (PCA, t-SNE)
     + Anomaly detection methods
5. Model Optimization and Deployment (1 day)
   * Hyperparameter tuning (Grid Search, Random Search)
   * Preventing overfitting and underfitting
   * Introduction to model deployment concepts
6. Mini Project (1 day)
   * Project Description:
     + Develop a machine learning model to solve a complex problem.
     + Tasks include data preprocessing, model selection, training, evaluation, and basic deployment strategy.
   * Activities:
     + Project work (5 hours)
     + Presentation and feedback session (2 hours)
7. Additional Practice and Case Studies (Optional sessions spread over the cohort)
   * Real-world datasets and problem-solving
   * Group discussions and collaborative learning

# Cohort 4 (Semester 8): Advanced Topics and Capstone Project (105 Hours)

Total Days: 15 days | Hours per Day: 7 hours

Course Outline:

1. Advanced Machine Learning Techniques (4 days)
   * Day 1:
     + Advanced ensemble methods (Stacking, Blending)
     + Model interpretation tools (SHAP values, LIME)
   * Day 2:
     + Time series analysis and forecasting
     + ARIMA models and Prophet library
   * Day 3:
     + Recommender systems
     + Collaborative filtering and content-based filtering
   * Day 4:
     + Natural Language Processing (NLP) basics
     + Text preprocessing and sentiment analysis
2. Introduction to Deep Learning (3 days)
   * Day 1:
     + Neural network architectures
     + Activation functions and backpropagation
   * Day 2:
     + Convolutional Neural Networks (CNNs) for image processing
     + Implementing CNNs with TensorFlow/Keras
   * Day 3:
     + Recurrent Neural Networks (RNNs) for sequential data
     + Introduction to LSTMs and GRUs
3. Big Data Tools and Technologies (3 days)
   * Day 1:
     + Big data concepts and challenges
     + Hadoop ecosystem overview
   * Day 2:
     + Working with Apache Spark for big data processing
     + Spark DataFrames and Spark SQL
   * Day 3:
     + Distributed machine learning with Spark MLlib
     + Real-time data processing with Spark Streaming
4. Data Engineering and Cloud Computing (2 days)
   * Day 1:
     + Building data pipelines
     + ETL (Extract, Transform, Load) processes
     + Working with APIs and web scraping
   * Day 2:
     + Introduction to cloud platforms (AWS, Azure, GCP)
     + Deploying machine learning models in the cloud
     + Serverless computing and containerization (Docker)
5. Data Ethics and Legal Considerations (1 day)
   * Data privacy laws (GDPR, CCPA)
   * Ethical considerations in data science
   * Bias and fairness in machine learning models
6. Data Storytelling and Communication (1 day)
   * Principles of effective data visualization
   * Crafting a narrative with data
   * Advanced presentation skills and tools
7. Capstone Project (1 day for project kickoff, 1 day for presentations)
   * Day 1: Project Kickoff
     + Project planning and proposal
     + Team formation (if applicable)
     + Defining project scope and objectives
     + Project Work (to be conducted throughout the cohort):
     + Applying advanced analytics to a complex, real-world problem
     + Incorporating machine learning, deep learning, and big data tools
     + Emphasis on end-to-end solution development
   * Day 2: Final Presentations
     + Project presentations to a panel of instructors and peers
     + Q&A session and feedback
     + Discussion of challenges, solutions, and learnings
8. Additional Support and Consultation (Available throughout the cohort)
   * Instructor office hours
   * Peer collaboration sessions
   * Access to additional learning resources