Milestone	Module	Sub-Topics
Prerequisites	**Prerequisites**	• Programming Basics: Understand basic programming concepts and be proficient in Python • Mathematics Foundation: Have a grasp of algebra, calculus, statistics, and linear algebra. • Data Handling Skills: Be comfortable working with different data formats and manipulating data using libraries like Pandas • Data Visualization: Learn to create effective visualizations using tools like Matplotlib and Seaborn • SQL Knowledge: Familiarity with SQL for data retrieval and manipulation from relational databases. • Understanding Python Data Structures: Know essential data structures like lists, arrays, dictionaries, and data frames
Introduction	**Introduction to Data Science**	• Titanic Case Study • What is Data Science and scope and impact • Explain the role of a Data Scientist, the skills required and industry demand and pay. • Share Data Science Roadmap and explain its component such as Python Programming for Datascience, Analytics, AI, ML, Prompt Engineering (Generative AI) • Essential Data Science tools and techniques • Lifecycle of Data Science Projects • Showcase examples of Data Science applications, their functionalities, and discuss some sample projects and resourcess. • Highlight career opportunities and competitive advantages of being a Data Scientists in the job market • Share success stories of Data Scientist and their impacts on society.
Data Wrangling with SQL	**SQL Fundamentals**	• Fundamentals of Structured Query Language • DDL, DML, DQL, TCL • SQL Functions • SQL Clauses and Statements for fetching Data
	Advanced SQL Queries	• SQL Joins • Nested Queries • Python with SQL
Python Programming for Data Science	**Python basics: Data types, functions, and control structures**	Python Data types, Variables, Python Operators
		Conditional Statements, Loops
		Python Flow Control
		User-defined Functions, Function Arguments, Lambda Functions
		Recap of Python Data Structures
Libraries for Data Science Numpy & Pandas	**Data Handling with NumPy**	NumPy: Introduction, NumPy Array, Creating NumPy Array, Array Attributes, Array Methods, Array Indexing, Slicing Arrays,
		NumPy: Array Operations, Saving and Loading Arrays
		CRUD Operations
	Data Manipulation with pandas	Pandas: Introduction, Pandas Series, Creating Pandas Series, Accessing Series Elements, Filtering a Series
		Arithmetic Operations, Series Ranking and Sorting, Checking Null Values, Concatenate a Series
		Data Frame Manipulation: Introduction, Dataframe Creation, Reading Data from Various Files, Understanding Data
		Accessing Dataframe Elements using Indexing Dataframe Sorting, Ranking in Dataframe
		Dataframe Concatenation, Dataframe Joins, Dataframe Merge, Reshaping Dataframe
		Pivot Tables, Cross Tables, Dataframe Operations, Checking Duplicates
Exploratory Data Analysis	**Data Visualization**	Visualisation I: Visualisation using Matplotlib, Plot Styles & Settings, Line Plot, Multiline Plot
		Matplotlib Subplots, Histogram, Boxplot, Pie Chart, Scatter Plot.
		Visualisation II: Visualisation using Seaborn, Strip Plot, Distribution Plot, Joint Plot, Violin Plot, Swarm Plot, Pair Plot, Count Plot, Heatmap
		Regression plots, categorical plots, area plots, etc, with Python seaborn.
	Data Preprocessing	EDA: Summary Statistics, Missing Value Treatment, Outliers Analysis and Handling, Dataframe Analysis using Groupby and Advanced Data Explorations, Normalization, standardization, etc.
		Statistics: Introduction to Statistics, Random Variables, Descriptive Statistics, Measure of Central Tendency, Measure of Dispersion, Skewness and Kurtosis, Covariance and Correlation
Generative AI	**Introduction of Generative AI**	GPT – What are generative pre-trained models (GPT), how does a GPT work?, real life examples of GPT, etc. LLM – NLP and Language models, what are LLMs, how does a LLM work, applications of LLM, etc.

15 Project Statements	1. Stock Price Prediction: Develop a model to predict stock prices using historical data and various machine learning algorithms. Implement the model to provide real-time predictions based on the latest market data.
	2. Sentiment Analysis for Social Media: Create a system to analyze the sentiment of social media posts or tweets in real-time. This could involve classifying posts as positive, negative, or neutral and tracking trends over time.
	3. Predictive Maintenance for Equipment: Build a predictive maintenance system that analyzes sensor data from machinery or equipment in real-time to predict when maintenance is required, helping to prevent unexpected downtime.
	4. Fraud Detection in Financial Transactions: Develop an algorithm to detect fraudulent transactions in real-time by analyzing patterns and anomalies in financial data streams.
	5. Real-Time Traffic Prediction: Use historical traffic data and real-time sensor data to predict traffic congestion and suggest alternate routes for drivers in real-time.
	6. Energy Consumption Forecasting: Build a model to forecast energy consumption for buildings or regions based on historical usage data and external factors like weather conditions. Provide real-time predictions to optimize energy usage.
	7. Health Monitoring and Early Warning System: Develop a system to monitor vital signs and health data in real-time, detecting anomalies and providing early warnings for health issues.
	8. Supply Chain Optimization: Analyze real-time data from the supply chain, including inventory levels, production schedules, and transportation routes, to optimize logistics and reduce costs.
	9. Customer Churn Prediction: Predict customer churn in real-time by analyzing customer behavior and engagement metrics, allowing businesses to take proactive measures to retain customers.
	10. Dynamic Pricing Optimization: Implement a dynamic pricing system that adjusts prices in real-time based on factors like demand, competitor prices, and inventory levels.
	11. Weather Forecasting and Disaster Prediction: Develop models to forecast weather patterns and predict natural disasters like hurricanes or earthquakes in real-time, helping authorities and individuals to take preventive measures.
	12. Smart Home Automation: Create a system that uses sensors and data analysis to automate home devices like thermostats, lights, and security systems in real-time based on user preferences and environmental conditions.
	13. Recommendation System for Online Content: Build a recommendation engine that suggests personalized content or products to users in real-time based on their preferences, browsing history, and behavior.
	14. Predictive Hiring Analytics: Develop a model to predict the success and fit of candidates in real-time based on their resumes, interview performance, and other relevant data, aiding in hiring decisions.
	15. Healthcare Fraud Detection: Build a system to detect healthcare fraud in real-time by analyzing medical claims data and identifying suspicious patterns or anomalies.