

## 7-Month Data Science Curriculum

---

### Week 1: Orientation

- **Overview of Data Science**
    - Introduction to Data Science: Definitions, applications, and impact
    - How to Learn: Peer-to-peer learning, hands-on approach, and resource utilization
    - Career Opportunities: Roles, career paths, and industry applications
  - **Hands-On Activities:**
    - Icebreaker activities to understand learning goals
    - Introduction to learning platforms and tools
- 

### Month 1: Databases and SQL

- **Week 1: Database Basics**
    - Introduction to Database Concepts: Types of databases, database design, and relational databases
    - **SQL Fundamentals:**
      - Data Definition Language (DDL)
      - Data Manipulation Language (DML)
      - Basic queries, joins, and aggregations
  - **Week 2: NoSQL Databases**
    - Introduction to NoSQL: Types (document-based, key-value, column-family, graph)
    - Basic CRUD operations in NoSQL databases
  - **Week 3: SQL Project**
    - Hands-on SQL exercises: Creating and querying databases
    - Mini-project: Design and implement a small database with SQL queries
  - **Week 4: NoSQL Project**
    - Hands-on exercises with NoSQL databases
    - Mini-project: Implement and query a NoSQL database
- 

### Month 2: Python for Data Science

- **Unit 1: What is Data Science?**
  - Introduction to Data Science: Overview, tools, and methodologies
- **Unit 2: Python for Data Science**
  - Python Basics: Variables, data types, control structures
  - Using Google Colaboratory for Python programming

- **Unit 3: The NumPy Module**
    - Array operations: Creating, manipulating, and performing operations on arrays
    - Hands-on exercises with NumPy arrays
  - **Unit 4: Applied Statistics in Python**
    - Basic statistical concepts: Mean, median, mode, variance, standard deviation
    - Using Python for statistical analysis
  - **Unit 5: The Pandas Module**
    - DataFrames and Series: Creation, manipulation, and analysis
    - Data cleaning and transformation using Pandas
  - **Unit 6: Visualization**
    - Data visualization techniques using Matplotlib and Seaborn
    - Creating and interpreting various types of plots
  - **Unit 7: Data Mining I – Supervised Learning**
    - Introduction to supervised learning: Concepts and techniques
    - Implementing classification and regression algorithms using scikit-learn
  - **Unit 8: Data Mining II – Clustering Techniques**
    - Introduction to clustering: K-Means, hierarchical clustering
    - Hands-on exercises with clustering techniques
  - **Unit 9: Data Mining III - Statistical Modeling**
    - Statistical modeling concepts: Linear regression, logistic regression
    - Building and evaluating models with scikit-learn and statsmodels
  - **Unit 10: Time Series Analysis**
    - Introduction to time series data: Components and forecasting methods
    - Time series modeling using statsmodels: AR, MA, ARIMA models
  - **Week 4: Python Project**
    - Implement and test Python programs using Google Colaboratory
    - Comprehensive project applying all learned Python and data science techniques
- 

### Month 3: Mathematics for Data Science

- **Unit 1: Descriptive Statistics**
  - Measures of central tendency and dispersion
  - Data distribution analysis: Histograms, box plots
- **Unit 2: Probability Distributions**
  - Common distributions: Normal, binomial, Poisson
  - Application of distributions in data science
- **Unit 3: Matrix Algebra**
  - Basics of matrix operations: Addition, multiplication, inversion
  - Application of matrix algebra in data science: Linear regression, PCA
- **Week 4: Mathematics Project**
  - Hands-on exercises applying descriptive statistics, distributions, and matrix algebra
  - Mini-project: Analyzing a dataset using mathematical concepts

---

## Month 4: Advanced Data Science Techniques

- **Unit 1: Advanced Supervised Learning**
    - Advanced techniques in supervised learning: Decision Trees, Random Forests, Gradient Boosting
    - Model tuning and validation
  - **Unit 2: Advanced Unsupervised Learning**
    - Techniques: Principal Component Analysis (PCA), t-SNE
    - Dimensionality reduction and feature extraction
  - **Unit 3: Big Data Technologies**
    - Introduction to PySpark and its ecosystem
    - Hands-on exercises with PySpark for big data processing
  - **Unit 4: Deep Learning Basics**
    - Introduction to neural networks and deep learning
    - Building and training simple neural networks using TensorFlow or PyTorch
  - **Week 4: Advanced Techniques Project**
    - Implementing advanced supervised and unsupervised learning techniques
    - PySpark and deep learning project
- 

## Month 5: Real-World Data Science Applications

- **Unit 1: Sentiment Analysis**
    - Techniques and tools for analyzing sentiment in text data
    - Hands-on exercises with sentiment analysis using Python
  - **Unit 2: Time Series Forecasting**
    - Advanced time series analysis techniques
    - Implementing forecasting models and evaluating performance
  - **Unit 3: Feature Engineering**
    - Techniques for feature extraction and transformation
    - Improving model performance through feature engineering
  - **Week 4: Applications Project**
    - Comprehensive project involving sentiment analysis, time series forecasting, and feature engineering
    - Presentations and peer reviews
- 

## Month 6: Capstone Project Preparation

- **Week 1-2: Project Planning**
  - Defining project scope, objectives, and deliverables
  - Data collection and preprocessing

- **Week 3-4: Project Implementation**
    - Developing and implementing the project
    - Regular check-ins and progress reviews
- 

## **Month 7: Capstone Project Execution and Presentation**

- **Week 1-2: Finalizing the Project**
  - Completing project development and testing
  - Preparing project report and presentation
- **Week 3: Presentation Preparation**
  - Creating and refining presentation slides
  - Practice presentation skills and receive feedback
- **Week 4: Capstone Project Presentation**
  - Final presentations to peers and instructors
  - Receiving feedback and discussing potential improvements