**Data Science Topic-wise roadmap**

**1. Foundations**

**Estimated Time: 2-3 Weeks**

1. **Mathematics Basics**
   * Mean, median, mode, variance, standard deviation.
   * Probability: Basics, distributions (normal, binomial).
   * Linear algebra: Vectors and matrices (basic operations).
   * Basic calculus: Derivatives and gradients (as needed for ML).
2. **Programming Basics (Python)**
   * Variables, data types, conditionals, loops, and functions.
   * File handling and working with libraries.

**2. Data Manipulation & Cleaning**

**Estimated Time: 3-4 Weeks**

1. **Data Handling with Python**
   * Learn Pandas: Loading data, dataframes, filtering, and grouping.
   * NumPy: Arrays, indexing, broadcasting.
2. **Data Cleaning**
   * Handling missing values.
   * Data transformation (normalization, encoding categorical data).

**Resources:**

* Kaggle tutorials: Pandas and NumPy.

**3. Exploratory Data Analysis (EDA) & Visualization**

**Estimated Time: 2-3 Weeks**

1. **Exploratory Data Analysis**
   * Descriptive statistics and summary statistics.
   * Identifying trends, patterns, and outliers.
2. **Data Visualization**
   * Python libraries: Matplotlib, Seaborn.
   * Interactive dashboards: Tableau/Power BI (optional).

**4. Database Management with SQL**

**Estimated Time: 2-3 Weeks**

1. **Basics of SQL**
   * SELECT, WHERE, GROUP BY, ORDER BY.
   * Joins: INNER, LEFT, RIGHT, FULL.
   * Subqueries and CTEs.
2. **Advanced SQL Concepts**
   * Window functions.
   * Indexing and optimization basics.

**Resources:**

* SQL Zoo or Mode Analytics SQL tutorials.

**5. Statistics and Analytics**

**Estimated Time: 4-5 Weeks**

1. **Inferential Statistics**
   * Hypothesis testing: t-tests, chi-square tests.
   * Confidence intervals and p-values.
2. **Regression Analysis**
   * Simple and multiple linear regression.
   * Logistic regression for classification problems.

**Resources:**

* *Statistics for Data Science* by Khan Academy.
* *An Introduction to Statistical Learning* (ISLR).

**6. Machine Learning Basics**

**Estimated Time: 6-8 Weeks**

1. **Supervised Learning**
   * Linear regression, logistic regression.
   * Decision Trees, Random Forests, Gradient Boosting (XGBoost).
2. **Unsupervised Learning**
   * Clustering: K-Means, DBSCAN.
   * Dimensionality reduction: PCA.
3. **Model Evaluation**
   * Metrics: Accuracy, precision, recall, F1-score, RMSE.
   * Cross-validation and hyperparameter tuning.

**Resources:**

* Google’s Machine Learning Crash Course.

**7. Big Data and Data Engineering (Optional)**

**Estimated Time: 6-8 Weeks**

1. **Big Data Basics**
   * Introduction to Hadoop and Spark.
   * Working with large datasets.
2. **NoSQL Databases**
   * MongoDB basics.

**Resources:**

* MongoDB University tutorials.

**8. Cloud and Deployment (Optional Advanced)**

**Estimated Time: 4 Weeks**

1. **Cloud Basics**
   * AWS, Azure, or Google Cloud for data storage and analysis.
2. **Model Deployment**
   * Flask or Django for building APIs.
   * Containerization using Docker.

**Resources:**

* AWS free-tier tutorials.
* Flask documentation.

**9. Specialized Topics (Optional Advanced)**

**Estimated Time: 4-6 Weeks**

1. **Deep Learning**
   * Basics of neural networks using TensorFlow or PyTorch.
2. **Natural Language Processing (NLP)**
   * Tokenization, sentiment analysis, embeddings (e.g., Word2Vec, BERT).

**Resources:**

* Deep Learning Specialization on Coursera by Andrew Ng.

**10. Projects and Portfolio**

**Continuous**

1. Build portfolio projects:
   * EDA on public datasets (Kaggle, UCI).
   * Build machine learning models and deploy them.
   * Create dashboards (e.g., Tableau, Power-BI).
2. Showcase on GitHub, LinkedIn, and personal blogs.