**Data Scientist Roadmap**

**|**

**|-- 1. Basic Foundations**

**|   |-- a. Mathematics**

**|   |   |-- i. Linear Algebra**

**|   |   |-- ii. Calculus**

**|   |   |-- iii. Probability**

**|   |   -- iv. Statistics**

**|   |**

**|   |-- b. Programming**

**|   |   |-- i. Python**

**|   |   |   |-- 1. Syntax and Basic Concepts**

**|   |   |   |-- 2. Data Structures**

**|   |   |   |-- 3. Control Structures**

**|   |   |   |-- 4. Functions**

**|   |   |   -- 5. Object-Oriented Programming**

**|   |   |**

**|   |   -- ii. R (optional, based on preference)**

**|   |**

**|   |-- c. Data Manipulation**

**|   |   |-- i. Numpy (Python)**

**|   |   |-- ii. Pandas (Python)**

**|   |   -- iii. Dplyr (R)**

**|   |**

**|   -- d. Data Visualization**

**|       |-- i. Matplotlib (Python)**

**|       |-- ii. Seaborn (Python)**

**|       -- iii. ggplot2 (R)**

**|**

**|-- 2. Data Exploration and Preprocessing**

**|   |-- a. Exploratory Data Analysis (EDA)**

**|   |-- b. Feature Engineering**

**|   |-- c. Data Cleaning**

**|   |-- d. Handling Missing Data**

**|   -- e. Data Scaling and Normalization**

**|**

**|-- 3. Machine Learning**

**|   |-- a. Supervised Learning**

**|   |   |-- i. Regression**

**|   |   |   |-- 1. Linear Regression**

**|   |   |   -- 2. Polynomial Regression**

**|   |   |**

**|   |   -- ii. Classification**

**|   |       |-- 1. Logistic Regression**

**|   |       |-- 2. k-Nearest Neighbors**

**|   |       |-- 3. Support Vector Machines**

**|   |       |-- 4. Decision Trees**

**|   |       -- 5. Random Forest**

**|   |**

**|   |-- b. Unsupervised Learning**

**|   |   |-- i. Clustering**

**|   |   |   |-- 1. K-means**

**|   |   |   |-- 2. DBSCAN**

**|   |   |   -- 3. Hierarchical Clustering**

**|   |   |**

**|   |   -- ii. Dimensionality Reduction**

**|   |       |-- 1. Principal Component Analysis (PCA)**

**|   |       |-- 2. t-Distributed Stochastic Neighbor Embedding (t-SNE)**

**|   |       -- 3. Linear Discriminant Analysis (LDA)**

**|   |**

**|   |-- c. Reinforcement Learning**

**|   |-- d. Model Evaluation and Validation**

**|   |   |-- i. Cross-validation**

**|   |   |-- ii. Hyperparameter Tuning**

**|   |   -- iii. Model Selection**

**|   |**

**|   -- e. ML Libraries and Frameworks**

**|       |-- i. Scikit-learn (Python)**

**|       |-- ii. TensorFlow (Python)**

**|       |-- iii. Keras (Python)**

**|       -- iv. PyTorch (Python)**

**|**

**|-- 4. Deep Learning**

**|   |-- a. Neural Networks**

**|   |   |-- i. Perceptron**

**|   |   -- ii. Multi-Layer Perceptron**

**|   |**

**|   |-- b. Convolutional Neural Networks (CNNs)**

**|   |   |-- i. Image Classification**

**|   |   |-- ii. Object Detection**

**|   |   -- iii. Image Segmentation**

**|   |**

**|   |-- c. Recurrent Neural Networks (RNNs)**

**|   |   |-- i. Sequence-to-Sequence Models**

**|   |   |-- ii. Text Classification**

**|   |   -- iii. Sentiment Analysis**

**|   |**

**|   |-- d. Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU)**

**|   |   |-- i. Time Series Forecasting**

**|   |   -- ii. Language Modeling**

**|   |**

**|   -- e. Generative Adversarial Networks (GANs)**

**|       |-- i. Image Synthesis**

**|       |-- ii. Style Transfer**

**|       -- iii. Data Augmentation**

**|**

**|-- 5. Big Data Technologies**

**|   |-- a. Hadoop**

**|   |   |-- i. HDFS**

**|   |   -- ii. MapReduce**

**|   |**

**|   |-- b. Spark**

**|   |   |-- i. RDDs**

**|   |   |-- ii. DataFrames**

**|   |   -- iii. MLlib**

**|   |**

**|   -- c. NoSQL Databases**

**|       |-- i. MongoDB**

**|       |-- ii. Cassandra**

**|       |-- iii. HBase**

**|       -- iv. Couchbase**

**|**

**|-- 6. Data Visualization and Reporting**

**|   |-- a. Dashboarding Tools**

**|   |   |-- i. Tableau**

**|   |   |-- ii. Power BI**

**|   |   |-- iii. Dash (Python)**

**|   |   -- iv. Shiny (R)**

**|   |**

**|   |-- b. Storytelling with Data**

**|   -- c. Effective Communication**

**|**

**|-- 7. Domain Knowledge and Soft Skills**

**|   |-- a. Industry-specific Knowledge**

**|   |-- b. Problem-solving**

**|   |-- c. Communication Skills**

**|   |-- d. Time Management**

**|   -- e. Teamwork**

**|**

**-- 8. Staying Updated and Continuous Learning**

**|-- a. Online Courses**

**|-- b. Books and Research Papers**

**|-- c. Blogs and Podcasts**

**|-- d. Conferences and Workshops**

**`-- e. Networking and Community Engagement**

**All the best 👍👍**