🚀 Guide2Code - Data Science & Analytics Roadmap 📊

🔰 Beginner Level - Foundations of Data & Analytics

Required Programming Languages:

- **Python** 🐍
- **SQL** 💾
- **Excel/Google Sheets** 📈

Required Skills:

- **Data Cleaning & Preprocessing** 🧹
- **Basic Statistics** 📊
- **Data Visualization** 🎨
- **Exploratory Data Analysis (EDA)** 🔍

Learn the Fundamentals:

- **Python Basics**: Learn Python syntax, data types, and libraries like Pandas and NumPy for data manipulation.
- **SQL Basics**: Understand how to query databases using SQL to retrieve, insert, update, and delete data.
- **Statistics Fundamentals**: Learn basic statistical concepts such as mean, median, mode, standard deviation, and probability.
- **Data Visualization**: Learn how to visualize data using libraries like Matplotlib, Seaborn, and tools like Excel/Google Sheets.
- **Data Cleaning**: Understand how to clean and preprocess raw data by handling missing values, duplicates, and inconsistent formats.

Beginner Projects 🚀:

1. **Basic Data Cleaning Project** - Clean a real-world dataset by handling missing values and outliers.
2. **Exploratory Data Analysis (EDA)** - Perform EDA on a dataset and summarize findings using graphs and statistical methods.
3. **Data Visualization Dashboard** - Create a simple dashboard using Matplotlib or Seaborn to visualize trends in data.

4. **SQL Queries** - Write SQL queries to analyze data from a database (e.g., SELECT, JOIN, GROUP BY).
5. **Excel Data Analysis** - Analyze and visualize data using Excel/Google Sheets.

🔄 Intermediate Level - Expanding Data & Analytics Skills

Required Programming Languages:

- **Python (Advanced)** 🐍
- **R** 📊
- **SQL (Advanced)** 💾

Required Skills:

- **Machine Learning Fundamentals** 🤖
- **Advanced Data Visualization** 📊
- **Data Wrangling** 🧹
- **Big Data Tools (Hadoop, Spark)** 💡

Expanding Your Knowledge:

- **Advanced Python**: Learn advanced Python libraries like Scikit-learn for machine learning, and SQLAlchemy for database interaction.
- **R Programming**: Learn the basics of R for statistical analysis and data visualization.
- **Machine Learning Basics**: Understand the core concepts of supervised and unsupervised learning, algorithms like linear regression, decision trees, and clustering.
- **Big Data Technologies**: Get familiar with distributed computing and big data tools like Hadoop and Apache Spark for handling large datasets.
- **Data Wrangling**: Learn how to transform and reshape data for analysis using advanced techniques and libraries (e.g., Pandas, dplyr in R).

Intermediate Projects 🔄:

1. **Data Wrangling Project** - Work with messy, unstructured data and clean it using Python (Pandas).
2. **Predictive Model** - Build a basic machine learning model to predict an outcome using Scikit-learn (e.g., predicting house prices).

3. **Interactive Dashboard with Plotly** - Create a data visualization dashboard with interactive charts.
4. **SQL Data Aggregation** - Perform complex SQL queries involving aggregation, subqueries, and joins.
5. **Big Data Analysis with Spark** - Analyze large datasets using Apache Spark.

🔝 Advanced Level - Mastering Data & Analytics

Required Programming Languages:

- **Python (Advanced)** 🐍
- **R** 📊
- **Scala/Java (for Spark)** 💻

Required Skills:

- **Deep Learning** 🧠
- **Big Data & Distributed Systems** 🌐
- **Data Engineering** 🔧
- **Data-Driven Decision Making** 📊
- **Cloud Data Platforms (AWS, GCP, Azure)** ☁️

Deep Dive Into Advanced Topics:

- **Deep Learning**: Learn advanced techniques such as neural networks, CNNs, RNNs, and reinforcement learning.
- **Big Data Systems**: Work with tools like Apache Hadoop, Spark, and Kafka to process and analyze large-scale datasets.
- **Data Engineering**: Get into the technical side of building data pipelines, ETL processes, and managing data storage solutions.
- **Data-Driven Decision Making**: Understand how to leverage analytics to drive business decisions.
- **Cloud Platforms**: Work with cloud data platforms like AWS, GCP, and Azure to scale data processing and analysis.

Advanced Projects 🌟:

1. **Deep Learning Model** - Build and deploy a deep learning model using TensorFlow or PyTorch (e.g., image classification).

2. **Big Data Analysis with Hadoop/Spark** - Analyze large datasets using Hadoop or Spark for batch and stream processing.
3. **Data Pipeline in Cloud** - Design and implement a data pipeline for real-time analytics in a cloud platform (AWS/GCP).
4. **Customer Segmentation Using ML** - Use unsupervised learning techniques like clustering to segment customers for marketing purposes.
5. **Business Intelligence Dashboard** - Build a comprehensive dashboard for analyzing business KPIs using Power BI or Tableau.

Thank You for Visiting Guide2Code!

"Turn raw data into actionable insights and drive decisions with confidence!"