Data Scientist Roadmap Checklist

# Math & Statistics

* ☐ Basic statistics: mean, median, mode, standard deviation
* ☐ Probability theory & distributions
* ☐ Hypothesis testing
* ☐ Linear algebra (vectors, matrices)
* ☐ Calculus basics (optional at first)

# Python Programming

* ☐ Python basics: variables, loops, functions, classes
* ☐ Data structures: lists, dictionaries, sets, tuples
* ☐ Numpy & Pandas for data manipulation
* ☐ Jupyter Notebooks

# Data Analysis & Visualization

* ☐ Data cleaning (missing values, duplicates, outliers)
* ☐ Data transformation (grouping, merging, filtering)
* ☐ Matplotlib and Seaborn for charts and plots

# SQL & Databases

* ☐ Basics of relational databases
* ☐ SQL: SELECT, JOIN, GROUP BY, WHERE, etc.
* ☐ Practice queries on real datasets

# Machine Learning

* ☐ Supervised vs. unsupervised learning
* ☐ Regression and classification algorithms
* ☐ Clustering algorithms
* ☐ Model evaluation metrics
* ☐ Scikit-learn library

# Projects & Portfolio

* ☐ Clean and analyze real-world datasets
* ☐ Build 3–5 end-to-end projects
* ☐ Publish projects on GitHub
* ☐ Write blog posts or documentation

# Model Deployment & Cloud

* ☐ Build web apps using Flask or Streamlit
* ☐ Deploy models using Heroku or Render
* ☐ Basics of AWS, GCP, or Azure

# Job Preparation

* ☐ Create strong resume and LinkedIn profile
* ☐ Practice SQL and ML interview questions
* ☐ Apply for internships or junior roles