Roadmap to Become a Data Scientist

A Data Scientist doesn't just "build models" they **understand data deeply, solve business problems, and communicate insights effectively.**

Data Scientist = Mathematics + Statistics + Programming + Machine Learning + Business Intelligence + Communication

🔥 **What Does a Data Scientist Do?**

- Collect, clean & explore large datasets
- Build predictive ML models & run experiments
- Visualize insights using dashboards/plots
- Collaborate with domain teams & decision-makers
- Turn data into revenue-saving/business-growth actions

A Data Scientist must know **Data Analysis → Machine Learning → Deployment → Reporting**

Step 1: Mathematics, Statistics & Probability

Why? → This is the **brain** of Data Science & ML.

Learn Concepts:

A. Statistics

- Descriptive Statistics: Mean, Median, Mode, Variance, IQR, Skewness, Kurtosis
- Inferential Statistics: Hypothesis Testing, p-value, Z-test, T-test, ANOVA
- Sampling Methods: Random, Stratified, Systematic
- Correlation vs Causation
- A/B Testing

B. Probability

- Conditional Probability
- Bayes' Theorem
- Random Variables (Discrete/Continuous)
- Law of Large Numbers, Central Limit Theorem

C. Mathematics

- Linear Algebra → Matrices, Vectors, Eigenvalues, Singular Value Decomposition
- Calculus → Derivatives, Integrals, Gradient, Optimization
- Combinatorics → Permutations/Combinations
- ✓ Outcome → You understand how ML algorithms work mathematically.

Step 2: Programming for Data Science (Python)

Core Concepts to Learn:

- Data Types, Loops, Functions
- OOP (Classes, Inheritance, Polymorphism)
- File Handling, Error Handling
- Working with APIs & JSON

Libraries Examples

Library	**Use**
NumPy	Numerical operations
Pandas / Polars	Data manipulation, cleaning
Matplotlib, Seaborn, Plotly	Visualization
Scipy	Statistical operations
Regex	Data cleaning tasks
Requests/BeautifulSoup/Scrapy	Web scraping

- Optional (R Programming): **ggplot2, tidyverse, dplyr, R Shiny**
  - ✓ Outcome → You can analyze datasets & write scripts confidently.

📍 **Step 3: SQL + Databases (Parallel with Python)**

Must-have skill for BD & Abroad.

Learn:

- SELECT, WHERE, GROUP BY, HAVING
- JOINS (INNER, LEFT, RIGHT, FULL)
- CTE, Subqueries
- Window Functions (RANK, ROW_NUMBER, LAG, LEAD)
- Indexing & Query Optimization
- Database Concepts → OLTP/OLAP, ERD, Normalization

Databases to Practice:

- **MySQL / PostgreSQL (Primary)**
- **SQL Server, Snowflake, BigQuery (Global focus)**

✓ Outcome → You can pull/analyze data from relational DBs like an analyst.

Step 4: Exploratory Data Analysis (EDA)

- Outlier detection
- Missing value handling
- Feature selection
- Univariate/Bivariate/Multivariate Analysis
- Correlation Heatmaps, Pairplots
- ✓ Outcome → You generate insights from raw data.

Step 5: Machine Learning

Supervised Learning

- Regression → Linear, Ridge, Lasso, ElasticNet
- Classification → Logistic, SVM, KNN, Decision Tree, Naive Bayes

- Ensemble Methods → Random Forest, XGBoost, LightGBM, CatBoost

Unsupervised Learning

- Clustering → K-Means, DBSCAN
- Dimensionality Reduction → PCA, t-SNE

Time Series

- Moving Average, ACF/PACF
- ARIMA, SARIMA, Prophet

Model Skills

- Feature Engineering
- Train/Test Split, Cross Validation
- Hyperparameter Tuning
- Evaluation Metrics → Accuracy, F1, ROC-AUC, RMSE, R2
- ✓ Outcome → You can build ML models end to end.

Step 6: Deep Learning & AI

Neural Networks

- Perceptron, Feedforward, Backpropagation
- Activation Functions (ReLU, Sigmoid, Softmax)

Architectures

Model	**Use Case**
ANN	General prediction tasks
CNN	Images
RNN/LSTM/GRU	Time series, sequences
Transformers	NLP & LLMs

Frameworks

- TensorFlow / Keras / PyTorch

Projects

- Image Classifier
- Sentiment Analysis (NLP)
- Chatbot
- Transfer Learning with ResNet/BERT

✓ Outcome → You enter the AI side of Data Science.

Step 7: Data Visualization, BI & Storytelling

Tools:

- Power BI
- Tableau
- Looker Studio/Metabase
- Excel/Google Sheets (Must!)

Skills:

- KPIs, Dashboards, DAX
- Data storytelling
- Business understanding

✓ Outcome → You communicate insights visually like an analyst.

📍 **Step 8: Deployment + MLOps Basics (Optional but Valuable)**

- Flask / FastAPI
- Streamlit Dashboard Apps
- Docker Basics
- Git/GitHub+ CI/CD

- AWS/GCP/Azure (Basic: S3, EC2, Lambda)

✓ Outcome → You deploy ML apps for real use.

Step 9: Project Portfolio [examples]

Must build **hands-on** project experience.

ML Projects

- House Price Prediction
- Customer Churn
- Recommendation System
- Medical Diagnosis Classifier

Deep Learning

- Face Recognition
- Emotion Detection
- Chatbot / LLM Mini App

SQL

- Sales Analysis Dashboard
- Retail DB Query Tasks

BI Visuals

- E-commerce BI Dashboard
- Finance KPI Board

✓ Target → Minimum **10–15 portfolio projects**.

🌎 **Bangladesh vs Global Job Focus**

Skill	BD Priority	Global Priority
Python + SQL	Must	Must (Strong)

Skill	BD Priority	Global Priority
Statistics & ML	Must	Must (Advanced)
Power BI/Tableau	High	Medium
MLOps/Cloud	Bonus	Often Required
Communication	Important	Critical
Portfolio	Bonus	Mandatory

Final Advice ✨

- Learn consistently
- Build real projects
- Post progress publicly
- Network with professionals
- Apply without fear

Thanks for Reading!

🌱 *Made with love to help beginners grow in Data Science*

👩🏿 **Author:** Mahfuzatul Bushra