

Data Scientist **Core Skills** and required **Tools Skills**



Python

Topics

- ✓ Variables, Numbers, Strings
- ✓ Lists, Dictionaries, Tuples
- ✓ If condition, for loop
- ✓ Functions, modules
- ✓ Read, write files
- ✓ Exception handling
- ✓ Classes, Objects

Pandas, Data Visualization (Matplotlib or Seaborn)

Numpy

Pandas

Matplotlib or Seaborn

- ✓ Do not learn both or One
- ✓ Matplotlib and Seaborn are libraries for data visualization and exploration

Statistics and Math for Data Science

- ✓ Finish this excellent Khan academy course on statistics and probability
Course link: <https://www.khanacademy.org/math/statistics-probability>

Machine Learning

Topics

- ✓ Feature engineering
- ✓ Regression
- ✓ Classification
- ✓ Clustering

Project Management

- ✓ Scrum: Azure ADO
- ✓ Tools: JIRA, Notion

Database: SQL and MongoDB - No SQL

SQL

Topics

- ✓ Basics of relational databases
- ✓ Basic Queries: SELECT, WHERE LIKE, DISTINCT, BETWEEN, GROUP BY, ORDER BY
- ✓ Advanced Queries: CTE, Sub queries, Window Functions
- ✓ Joins: Left, Right, Inner, Full
- ✓ Stored procedures and functions
- ✓ No need to learn database creation, indexes, triggers etc. as those things are rarely used by data scientists

Machine Learning Projects with Deployment

Flask Web Framework

You need to finish two end to end ML projects. One on Regression, the other on Classification

e.g. Regression Project: Bangalore property price prediction

Project covers following

- ✓ Data cleaning
- ✓ Feature engineering
- ✓ Model building and hyper parameter tuning
- ✓ Write flask server as a web backend
- ✓ Building website for price prediction
- ✓ Deployment to AWS/GCP/Other providers

Advance Excel

BI Tool (Power BI or Tableau)

Power BI

Tableau

- ✓ Should I learn Power BI or Tableau?
- ✓ If someone asks me to pick between Power BI and Tableau, I always suggest Power BI as it is growing in popularity as compared to Tableau.
- ✓ This Gartner research shows Power BI is leading a BI game:

<https://info.microsoft.com/ww-landing-2022-gartner-mq-report-on-bi-andanalytics-platforms.html?LCID=EN-US>

Deep Learning

- ✓ Natural Language Processing (NLP)
- ✓ ANN
- ✓ CNN

Computer Vision

- ✓ Computer vision is a vast field where one can use OpenCV, PyTorch, Tensorflow etc for deep learning approaches for computer vision as well. You can find many resources online on this. I do not have a specific recommendation for this

Tips of effective learning

Spend less time in consuming information, more time in

- ✓ Digesting
- ✓ Implementing
- ✓ Sharing

Advanced Topics

Cloud ML Platforms

- ✓ Big cloud service providers such as **AWS, Azure, Google Cloud** have their own ML offering such as **Amazon** in case of **AWS**. As a fresher it is ok if you are not familiar with these cloud platforms but once you have some experience it is good to have experience and know-how of at least one **cloud ML platform**.
- ✓ Perform EDA (Exploratory data analysis on at least 3 datasets on <https://www.kaggle.com/>)

Core/Soft Skills

