Data Science Roadmap for Beginners

Following is the roadmap to learn **Data Science** skills for a total beginner (**no coding** or **computer science background** needed). It includes FREE learning resources for technical skills (or tool skills) and soft (or core) skills 🏂

Total Duration: **6 Months**

3 hours in Tool Skills + **1 hour** in Core Skills = **4 hours** study Every Day

Week 1 and 2: Python 🐍

- Topics
    - Variables, Numbers, Strings
    - Lists, Dictionaries, Tuples
    - If condition, for loop
    - Functions, modules

- o Read, write files
- o Exception handling
- o Classes, Objects
- Learning Resources
  - o Track A (Free)
    - ▪ Codebasics python tutorials (first 16) - https://bit.ly/3X6CCC7
    - ▪ Codebasics python HINDI tutorials - https://bit.ly/3vmXrgw
  - o Track B (Paid with minor fees)
    - ▪ Python course: https://codebasics.io/courses/python-for-beginner-and-intermediate-learners

- **Assignment**

  - ☐ Track A: Finish all these exercises: https://bit.ly/3k1mof5
  - ☐ Track B: Finish exercises and quizzes for relevant topics
  - ☐ Create a professional looking LinkedIn profile
    - ▪ Have a clear profile picture, banner image (data scientist loading…)
    - ▪ Add tags such as: Open to work etc.

Week 3, 4: Pandas, Data Visualization (matplotlib or seaborn) 📊

- **Tech Skills**
  - o **numpy**
    - ▪ numpy YouTube playlist: https://bit.ly/3GTppa8
  - o **pandas**
    - ▪ pandas YouTube playlist (**first 10 videos only**): https://bit.ly/3vPJWpX
  - o **matplotlib or seaborn**
    - ▪ Do not learn both
    - ▪ matplotlib and seaborn are libraries for data visualization and exploration
    - ▪ matplotlib YouTube playlist: https://bit.ly/3k55egu
- **Core/Soft Skills**
  - o **Linkedin**

- - - Start following prominent data science, analytics influencers
      - Daliana Liu: https://www.linkedin.com/in/dalianaliu/
      - Hemanand Vadivel: https://www.linkedin.com/in/hemvad/
    - Increase engagement
      - Start commenting meaningfully on data science and career related posts
      - Helps network with others working in the industry build connections
      - Learning and brainstorming opportunity
    - Remember *online presence is a new form of resume*
  - **Business Fundamentals - Soft Skill**
    - Learn business concepts from ThinkSchool and other YT Case Studies
    - Example: How Amul beat competition: https://youtu.be/nnwqtZiYMxQ
  - **Discord**
    - Start asking questions and get help from the community. This post shows how to ask questions the right way: https://bit.ly/3I70EbI
    - Join codebasics discord server: https://discord.gg/r42Kbuk

- **Assignment**

  ☐ Write meaningful comments on at least **10 data science related LinkedIn posts**

  ☐ Note down your key learnings from **3 case studies** on ThinkSchool and share with your friend

Week 5, 6, 7, 8: 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
- When you are doing khan academy course, you can use stat quest YouTube channel to clear your doubts
- Complete math and statistics for data science YouTube playlist with Python code (Khan academy course doesn't have Python code)
  - Playlist link: https://bit.ly/3QrSXis

- **Assignment**
  ☐ Finish all exercises in that playlist: https://bit.ly/3QrSXis

☐ Perform EDA (Exploratory data analysis on **at least 3 datasets** on https://www.kaggle.com/

Week 9, 10, 11, 12: Machine Learning ⚒️

- **Machine Learning**
  - Topics
    - Feature engineering
    - Regression
    - Classification
    - Clustering
  - Learning Resources
    - YouTube playlist (more than 2 million views): https://bit.ly/3io5qqX
      - First 21 videos
    - Feature engineering playlist: https://bit.ly/3IFa3Yf
- **Core/Soft Skills**
  - **Project Management**
    - Scrum: https://scrumtrainingseries.com/
    - Kanban: https://youtu.be/jf0tlbt9lx0
    - Tools: JIRA, Notion
- **Assignment**
  - ☐ Complete all exercises in ML playlist: https://bit.ly/3io5qqX
  - ☐ Work on **2 Kaggle ML notebooks**
  - ☐ Write **2 LinkedIn posts** on whatever you have learnt in ML
  - ☐ Discord: Help people with **at least 10 answers**

Week 13, 14, 15: Machine Learning Projects with Deployment 🎯

- You need to finish **two** end to end ML projects. One on **Regression**, the other on **Classification**
- Regression Project: Bangalore property price prediction
  - YouTube playlist link: https://bit.ly/3ivycWr

- o 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
- Classification Project: Sports celebrity image classification
  - o YouTube playlist link: https://bit.ly/3ioaMSU
  - o Project covers following
    - Data collection and data cleaning
    - Feature engineering and model training
    - Flask server as a web backend
    - Building website and deployment
- **Assignment**
  - o In above two projects make following changes
    - ☐ Use **FastAPI** instead of **flask**. FastAPI tutorial: https://youtu.be/Wr1JjhTt1Xg
    - ☐ **Regression project**: Instead of property prediction, take any other project of your interest from Kaggle for regression
    - ☐ **Classification project**: Instead of sports celebrity classification, take any other project of your interest from Kaggle for classification and build end to end solution along with **deployment to AWS or Azure**

Week 16, 17: SQL 👝

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

- Learning Resources
  - Track A
    - Khan academy: https://bit.ly/3WFku20
    - https://www.w3schools.com/sql/
    - https://sqlbolt.com/
  - Track B
    - SQL course for data professionals: https://codebasics.io/courses/sql-beginner-to-advanced-for-data-professionals

- **Core/Soft Skills**
  - Presentation skills
    - Death by PowerPoint: https://youtu.be/Iwpi1Lm6dFo

- **Assignment**
  - ☐ Participate in resume project challenge on https://codebasics.io/
    - These challenges help you improve technical skills, soft skills and business understanding
    - Link: https://codebasics.io/event/codebasics-resume-project-challenge
  - ☐ Make a LinkedIn post with a submission of your resume project challenge
    - Sample post: https://bit.ly/3GxGaq1
    - Codebasics is promoting winning entries to employers. This way you can get interview calls. We do this in two ways
      - We have a database of employers hiring for data analyst positions. We send first 10 or 20 profiles based on their performance
      - LinkedIn post by Dhaval (who has more than 100k followers and some of them are HR managers, data analytics senior managers): https://bit.ly/3jnni5c

Week 18, 19, 20: BI Tool (Power BI or Tableau) ☑

- **Power BI**
  - Track A (Free)
    - Sales insights Power BI project: https://bit.ly/3C1WKgA
    - Personal finance project:
      project: https://www.youtube.com/watch?v=pqSoCa2NGj4
    - HR data analytics project: https://bit.ly/3C7cw9P

- o Track B (Paid for affordable fees)
    - Sales insights project tutorial playlist: https://bit.ly/3C1WKgA
    - My Power BI course that can make you job ready: https://codebasics.io/courses/power-bi-data-analysis-with-end-to-end-project

- **Tableau**
    - o Codebasics sales insights project: http://bit.ly/3YQSBFV
    - o HINDI codebasics sales insights project: https://bit.ly/3hZXUCb

- Should I learn Power BI or Tableau?
    - o If someone asks me to pick between Power BI and Tableau, I always suggest <mark>Power BI</mark> as it is growing in popularity as compared to Tableau.
    - o This Gartner research shows Power BI is leading a BI game: https://info.microsoft.com/ww-landing-2022-gartner-mq-report-on-bi-and-analytics-platforms.html?LCID=EN-US

- **Assignment**
    - ☐ Participate in one resume project challenge
        - These challenges help you improve technical skills, soft skills and business understanding
    - ☐ Make a LinkedIn post with video presentation
        - Example post: https://bit.ly/3WMTgGK (Naveen S)
    - ☐ Discord server participation

Week 21, 22, 23, 24: Deep Learning 🤖

- Deep Learning YouTube playlist: https://bit.ly/3vOZ3zV
- End to end potato disease classification project: https://bit.ly/3QzkVJi

- **Assignment**
    - ☐ Instead of potato plant images use tomato plant images or some other image classification dataset
    - ☐ Deploy to Azure instead of GCP

☐ Create a presentation as if you are presenting to stakeholders and upload video presentation on LinkedIn

Week 25 onwards…. 😀 😀 😀

- More projects 🎥
- Online brand building through LinkedIn, Kaggle, Discord, Opensource contribution 👨‍👩‍👧
- Resume and interview preparation 📃
  - Resume prep video: https://www.youtube.com/watch?v=buQSI8NLOMw
- Job application and Success 🚀

Tips of effective learning 💧

- **Spend less time in consuming information, more time in**
  - Digesting
  - Implementing
  - Sharing
- **Group learning**
  - Use **partner-and-group-finder** channel on codebasics discord server for group study and hold each other accountable for the progress of your study plan. Here is the discord server link: https://discord.gg/r42Kbuk

Inspirational Stories 😄

- Career transition stories: https://bit.ly/3PUZ4f3

Advanced Topics 🔭

- **ML Ops. What is it and how can I learn it?**

- o This post has necessary information: https://bit.ly/3X220Jk
- **Cloud ML Platforms**
  - o Big cloud service providers such as AWS, Azure, Google Cloud have their own ML offering such as Amazon Sagemaker 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.
- **Natural Language Processing (NLP)**
  - o NLP YouTube playlist: https://bit.ly/3XnjfEZ
- **Computer Vision**
  - o 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