Complete Roadmap to become a Data Scientist by CodeWithHarry

As a data scientist you will collect, analyze, process, and model data to gain valuable insights!
Why become a data scientist in India

- The average data scientist salary is 956K/Yr (Source: Glassdoor)
- 62% expected increase in demand (Source: TOI)
- Internet = Users = Data = DS Opportunity = Data Science Jobs in India
- Google Microsoft Flipkart and Amazon on constant search for talented Data Scientists

The RoadMap

In order to become a data scientist, one needs to learn the following things. I have added the resource links of all these things in this PDF

1. Mathematics & Statistics

Why not Programming? - Because understanding maths takes practice and you have to be good with numbers and logic to be able to code something

- ➔ Learn about basic statistics from a book like hines.(PDF is available online!)
- ➔ Learn what dy/dx actually means!
- ➔ Learn about Optimization and gradient descent. This is a good playlist to learn about basics of gradient descent
- ➔ Learn to plot simple functions in excel itself! You don't need a hammer to kill mosquitoes
- ➔ Learn about basic probability distributions with a bit more emphasis on normal distribution. Hines is a good book for learning these basics.
- ➔ Mathematics for Machine Learning: Download this book

2. Programming

Can be any programming language but I recommend python due to its popularity. Most of the companies ask this as a necessary requirement.

- ➔ Learn about Python Basics from here
- ➔ Learn about Numpy from CodeWithHarry one video
- ➔ Learn about Pandas from CodeWithHarry one video
- ➔ Learn about Matplotlib/Seaborn
- ➔ Learn about what time complexity of algorithms is (just the understanding)!
- ➔ Learn about storing data to the Database

3. Big Data and External Data Visualization Tools (Optional)

Data visualization and big data can be handled in python but some of the external tools are made just for the task in hand.Once you have started to use python to solve few of the data science problems, you need to look into these tools to understand what they have to offer on the table. These tools include:

- ➔ Tableau
- ➔ Excel (& VBA)
- ➔ Hadoop
- ➔ AWS Offerings

Note that learning these tools at the expert level might take some time!

4. Machine learning and Deep Learning

Once you have mastered data wrangling with python, you need to work on your Machine learning concepts. Start with this Google's free course which is made just for people getting started with machine learning. The best thing about this course is that it is free and it has just the topics you need to know to work in the industry!

- ➔ Learn Sklearn
- ➔ Learn to build a neural network in tensorflow
- ➔ Learn to use tensorflow_hub
- ➔ Learn how to use the tensorboard

Another amazing course for Data Science is from Kaggle. This one is free as well and is a must checkout for all the Data Science Enthusiasts! This one from coursera is free and amazing too and I am personally doing it!

5. Linux & Version Control

Apart from learning Python and Mathematics, you need to know how to manage and collaborate with others on the software you are creating.

- ➔ Learn Linux basics from CodeWithHarry Linux in one Video
- ➔ Learn Git from CodeWithHarry Git in One Video
- ➔ Keep your eye on this papers with code website
- ➔ Keep your eye on Github trending repositories
- ➔ Read Data Science news from sources like this
- ➔ Learn to scrape data from websites from here

6. Other Points to Remember

- ➔ Computer science degree matters but if it's too late for you to acquire a CS degree, there is nothing you can do about it so nothing to worry about!
- ➔ Domain specific knowledge helps in getting a job

- ➔ Good communication skills and spoken english helps!
- ➔ Staying ahead of the competition helps

I wish you all the best in learning data science. Hope all this information helps