Data Analyst to Data Scientist

8 Week-Roadmap

WEEK 1

Python Programming for Data Science

DAY 1-3

Basics of Python; Variables, Data Types, and Operations.

Practice:

Solve 5 problems on basic operations and data types on HackerRank.

DAY 4-5

Control Structures; Loops and Conditional Statements.

Practice:

Create a Python script to filter even numbers from a list and use conditional statements to categorize them.

DAY 6-7

Functions and Modules; Writing Reusable Code.

Practice:

Write a function to calculate the factorial of a number and another to check if a number is prime.

WEEK 2

Advanced Python and Introduction to Pandas

DAY 1-2

Advanced Data Structures; Lists, Dictionaries, Sets, and Tuples.

Practice:

Implement a dictionary-based phonebook application.

DAY 3-4

Introduction to Pandas; Series and DataFrames.

Practice:

Load a CSV file using Pandas and perform basic data exploration.

DAY 5-7

Data Manipulation with Pandas.

Practice:

Perform data cleaning on a dataset: handle missing values, duplicate data, and filter rows/columns.

WEEK 3

Exploratory Data Analysis (EDA)

DAY 1-2

Visualization with Matplotlib and Seaborn.

Practice:

Create a histogram of a dataset's numerical feature and a bar plot of a categorical feature.

DAY 3-4

Statistical Foundations of EDA.

Practice:

Calculate mean, median, mode, variance, and standard deviation of a dataset.

DAY 5-7

Practical EDA on Real Datasets.

Practice:

Conduct a full EDA on a dataset: identify outliers, perform hypothesis testing, and generate insights.

WEEK 4

Introduction to Machine Learning

DAY 1-2

Overview of Machine Learning; Types of ML.

Practice:

Classify problems into regression, classification, or clustering.

DAY 3-4

Linear Regression.

Practice:

Implement a simple linear regression model on a dataset.

DAY 5-7

Logistic Regression and Decision Trees.

Practice:

Build a logistic regression model and a decision tree to classify binary outcomes in a dataset

WEEK 5

Intermediate Machine Learning

DAY 1-3

Ensemble Methods; Random Forests and Gradient Boosting.

Practice:

Compare the performance of a decision tree, random forest, and gradient-boosted model on the same dataset.

DAY 4-5

Clustering Techniques; K-Means and Hierarchical Clustering.

Practice:

Implement a simple linear regression model on a dataset.

DAY 6-7

Dimensionality Reduction; PCA.

Practice:

Apply PCA on a high-dimensional dataset and visualize the results.

WEEK 6

Advanced Topics in Machine Learning

DAY 1-3

Introduction to Neural Networks and Deep Learning.

Practice:

Use TensorFlow or PyTorch to build a basic neural network for a classification problem.

DAY 4-7

Natural Language Processing (NLP) Basics.

Practice:

Implement a simple bag-of-words model and perform sentiment analysis on text data.

WEEK 7

SQL for Data Science

DAY 1-2

Advanced SQL Queries; Subqueries and Window Functions.

Practice:

Write SQL queries to perform analytical functions over partitioned data.

DAY 3-4

Data Warehousing Concepts.

Practice:

Design a simple star schema for a retail database.

DAY 5-7

ETL Processes; Introduction to Data Pipelines.

Practice:

Create an SQL script to transform and load data from one table to another.

WEEK 8

Real-World Projects and Portfolio Building

DAY 1-3

Complete a Capstone Project.

Practice:

Choose a problem statement and work on a project from scratch using datasets from Kaggle or GitHub.

DAY 4-5

Document Your Project.

Practice:

Write a detailed report of your project, including the problem statement, your approach, methodologies used, and insights gained.

DAY 6-7

Portfolio and Resume Building.

Practice:

Compile your projects and analyses into a professional portfolio. Update your resume to highlight your newly acquired data science skills and projects.

Conclusion

The progress of everyone will vary, so KEEP IN MIND THE FOLLOWING:

→ Individuals with a strong background in analytics, programming, or a related field may progress faster than those with less relevant experience.

→ The amount of time dedicated weekly to learning and practicing these skills plays a crucial role. The guide assumes a full-time commitment, which might not be feasible for everyone.

→ Everyone has a unique learning pace. Some may grasp new concepts quickly, while others may need more time to fully understand and apply them.

→ Depth of Knowledge: While this guide covers essential topics, mastery in data science often requires going beyond the basics, especially for highly specialized roles or industries.

→ Access to Resources: Availability of learning resources, mentorship, and hands-on projects can influence the speed and effectiveness of the transition.

Happy Learning