

PYTHON FOR DATA SCIENCE

A Complete Beginners Roadmap
Libraries • Weekly Plan • Projects • Cheat Sheet

16-WEEK STEP-BY-STEP LEARNING PLAN

Kaise Shuru Karein Python for Data Science?

Python data science ki sabse popular language hai — beginner-friendly syntax, huge community, aur powerful libraries. Aapko koi prior coding background ki zaroorat nahi hai. Ye guide aapko ek clear, week-by-week roadmap deti hai — zero se lekar apna pehla project banane tak.

Is Guide Mein Kya Milega:

- 16-week structured roadmap — koi confusion nahi, sirf clear steps
- Python + Pandas + NumPy quick-reference cheat sheet
- Data visualization basics — Matplotlib & Seaborn
- 5 beginner-friendly project ideas with real datasets
- Common mistakes jo beginners avoid karni chahiye

16-Week Roadmap — Overview

Weeks	Focus Area	Key Topics
1–3	Python Basics	Variables, loops, functions, data structures
4–6	Data Handling	NumPy arrays, Pandas DataFrames
7–8	Visualization	Matplotlib, Seaborn, chart types
9–10	Statistics	Mean, median, distributions, correlation
11	SQL + Python	Connecting databases, querying data
12–14	Intro to ML	Scikit-learn, regression, classification
15–16	Projects	Portfolio-ready end-to-end projects

Reality Check: Ye timeline consistent 1-2 hours/day practice assume karta hai. Speed se zyada consistency important hai — slow bhi chalega, ruko mat.

Phase 1: Python Basics (Weeks 1–3)

Topics to Cover

- Variables, data types (int, str, list, dict)
- Conditional statements — if/else
- Loops — for, while
- Functions — defining and calling
- Basic file handling

```
name = "Data Science"
for letter in name:
    print(letter)
```

Recommended Resources

- freeCodeCamp — Python for Beginners (YouTube, free)
- Python official documentation — docs.python.org
- Practice: HackerRank Python track (30 days of code)

Phase 2: Data Handling (Weeks 4–6)

NumPy — Numerical Computing

- Arrays create karna, indexing, slicing
- Mathematical operations on arrays
- Array reshaping aur aggregation

```
import numpy as np
arr = np.array([1, 2, 3, 4])
print(arr.mean())
```

Pandas — Data Manipulation

- DataFrames aur Series samajhna
- CSV/Excel files read karna
- Filtering, sorting, grouping data
- Missing values handle karna

```
import pandas as pd
df = pd.read_csv("data.csv")
print(df.head())
print(df.groupby("category").mean())
```

Phase 3: Visualization & Stats (Weeks 7–10)

Matplotlib & Seaborn

- Line charts, bar charts, histograms banana
- Seaborn se advanced statistical plots
- Chart labels, titles, aur styling

```
import matplotlib.pyplot as plt
plt.plot(df["date"], df["sales"])
plt.title("Sales Trend")
plt.show()
```

Statistics Fundamentals

- Mean, median, mode, standard deviation
- Probability distributions (normal, binomial)
- Correlation aur hypothesis testing basics

Phase 4: Intro to ML (Weeks 11–14)

Scikit-learn Basics

- Train-test split samajhna
- Linear regression — first model banayein
- Classification — logistic regression, decision trees
- Model evaluation — accuracy, precision, recall

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

5 Beginner Project Ideas (Weeks 15–16)

1. Sales Data Analysis

Ek retail dataset lekar sales trends, top products, aur regional performance analyze karein. (Dataset: Kaggle 'Superstore Sales')

2. Movie Ratings Dashboard

IMDb ya TMDb dataset use karke top-rated movies, genre trends visualize karein.

3. COVID-19 Data Tracker

Public COVID datasets se case trends, recovery rates plot karein.

4. Titanic Survival Prediction

Classic beginner ML project — Scikit-learn se survival predict karein. (Kaggle 'Titanic' dataset)

5. Personal Expense Analyzer

Apna khud ka spending data CSV mein daalkar categories, monthly trends analyze karein.

Quick-Reference Cheat Sheet

Task	Code
Read CSV	<code>pd.read_csv('file.csv')</code>
View first rows	<code>df.head()</code>
Column info	<code>df.info()</code>
Summary stats	<code>df.describe()</code>
Filter rows	<code>df[df['col'] > 10]</code>
Group + aggregate	<code>df.groupby('col').mean()</code>
Handle missing	<code>df.dropna() / df.fillna(0)</code>
Plot chart	<code>df['col'].plot(kind='bar')</code>

Common Mistakes Beginners Make

✕

Sirf tutorials dekhna

Hands-on practice ke bina sirf videos dekhna — coding seekhne ka best tareeka khud likhna hai, dekhna nahi.

✕

Ek saath sab kuch seekhne ki koshish

Python + ML + Deep Learning + Cloud sab ek saath — overwhelming ho jaata hai. Roadmap follow karo, step by step.

✕

Projects skip karna

Sirf theory se job nahi milti — recruiters portfolio dekhte hain. Har phase ke baad chhota project zaroor banao.

✕

Documentation ignore karna

Official docs padhna seekhna zaroori hai — ye best aur most reliable resource hai.

✕

Consistency na rakhna

1 din 5 hours, phir 2 hafte kuch nahi — ye kaam nahi karta. Daily 30-60 min consistent practice better hai.

Ready to Start Coding?

Ab aapke paas ek complete roadmap hai — Python basics se lekar apna pehla ML model banane tak. Consistency rakhein, roz thoda practice karein, aur projects banate rahein. Data science ka safar ek marathon hai, sprint nahi.

Thank you for downloading this guide!