7 Days to Learn 80% of Data Analysis with Python

This plan is designed to cover the most essential 80% of data analysis skills using Python, based on the Pareto Principle. Each day focuses on hands-on learning with real datasets using core tools: Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn.

Day 1: Python Basics + Pandas Intro

Goal: Load and explore datasets efficiently.

Topics: - Python recap: lists, dictionaries, loops, functions - Installing libraries: pandas, numpy, matplotlib - Reading data: `pd.read_csv()` - Exploring: `.head()`, `.info()`, `.describe()` - DataFrame selection with `.loc[]`, `.iloc[]`

Practice: Load a CSV and display the first 10 rows.

Day 2: Data Cleaning with Pandas

Goal: Make messy data usable.

Topics: - Checking for missing values: `.isnull().sum()` - Dropping or filling missing values: `.dropna()`, `.fillna()` - Converting data types: `.astype()` - Removing duplicates: `.drop_duplicates()`

Practice: Clean a dataset (like automobile or titanic).

Day 3: Data Manipulation

Goal: Filter, sort, and aggregate data.

Topics: - Filtering with conditions: `df[df['column'] > value]` - Sorting: `df.sort_values()` - Creating new columns - Grouping: `.groupby()` - Aggregations: `.sum()`, `.mean()`, `.count()`

Practice: Group data by a category and calculate average price.

Day 4: Data Visualization

Goal: Reveal insights visually.

Topics: - Line, bar, histogram, scatter plots using `matplotlib` - Plot styling: labels, titles, grid - Seaborn: `sns.histplot()`, `sns.boxplot()`, `sns.heatmap()`

Practice: Visualize numeric columns (e.g. price distribution, correlation).

Day 5: Descriptive Stats & Correlation

Goal: Understand data distributions and relationships.

Topics: - Summary statistics: mean, median, mode, std, var - Correlation matrix: `.corr()` - Visualizing correlations with heatmaps

Practice: Correlation between features like engine-size and price.

Day 6: Linear Regression with Scikit-Learn

Goal: Model and predict outcomes.

Topics: - Simple & Multiple Linear Regression - Model training with `LinearRegression().fit()` - Predicting with `.predict()` - Evaluating model: R^2, MSE

Practice: Predict car price or sales revenue.

Day 7: Mini Project & Wrap-Up

Goal: Build and showcase a complete project.

Project Steps: - Load and clean data - Explore & visualize patterns - Apply regression - Share insights (Jupyter Notebook, PDF, or blog)

Ideas: - Netflix ratings analysis - IPL data insights - E-commerce sales trends

---

Tools You'll Use: - `pandas`, `numpy`, `matplotlib`, `seaborn`, `sklearn`

By Day 7, you'll be comfortable with the **core 80% of data analysis tasks** used in real-world projects and interviews.