**DATA SCIENCE ROADMAP**

**DAY 1-4 NUMPY**

DAY 1- 2

1. Introduction to NumPy
   * Overview and installation
   * Importing NumPy
   * Creating NumPy arrays
     + From lists
     + Using functions like arange, linspace, zeros, ones, empty
2. Basic Array Operations
   * Array attributes (shape, size, ndim, dtype)
   * Basic indexing and slicing
   * Reshaping arrays
   * Array concatenation and splitting
3. Array Operations and Mathematics
   * Element-wise operations
   * Basic mathematical functions (sum, mean, std, min, max, etc.)
   * Broadcasting rules
   * Vectorized operations and performance benefits
4. Statistical Functions
   * Basic statistical functions (mean, median, std, var, etc.)
   * Random sampling (random module, random functions)
5. Advanced Indexing
   * Boolean indexing
   * Fancy indexing
   * Indexing with np.ix\_
6. Array Manipulation
   * Transposing arrays
   * Changing array shapes
   * Adding and removing elements
7. Linear Algebra
   * Dot product and matrix multiplication
   * Eigenvalues and eigenvectors
   * Inversion, decomposition, and solving linear equations

Day 3-4: Advanced Features and Practical Applications

1. Advanced Functions
   * Universal functions (ufuncs)
   * Applying custom functions along array axes (apply\_along\_axis)
   * Vectorize functions using np.vectorize
2. Broadcasting and Memory Layout
   * In-depth understanding of broadcasting
   * Memory layout of arrays (C-order vs. F-order)

Midday Session (2 hours)

1. File I/O
   * Saving and loading data (np.save, np.load, np.savetxt, np.loadtxt)
   * Handling large datasets with memmap
2. Integration with Other Libraries
   * Pandas and NumPy
   * SciPy and NumPy
3. Performance Optimization
   * Profiling NumPy code
   * Optimizing performance with Cython, Numba
   * Parallelizing computations

**DAY 5-8 PANDAS**

**Day 5-6: Core Concepts and Basic Operations**

1. **Introduction to Pandas**
   * Overview and installation
   * Importing pandas
   * Understanding Series and DataFrame
   * Creating Series and DataFrames from different sources (lists, dictionaries, NumPy arrays, CSV files)
2. **Basic DataFrame Operations**
   * Viewing data (head, tail, info, describe)
   * DataFrame attributes (shape, columns, index, dtypes)
   * Accessing data (loc, iloc, at, iat)
   * Indexing and selecting data
3. **Data Manipulation**
   * Adding and dropping columns/rows
   * Renaming columns and indices
   * Handling missing data (isnull, dropna, fillna)
   * Changing data types (astype)
4. **Data Cleaning and Preparation**
   * Detecting and filtering outliers
   * Handling duplicates (duplicated, drop\_duplicates)
   * String operations (str methods)
   * Date and time manipulation (pd.to\_datetime, dt accessor)
5. **Data Aggregation and Grouping**
   * GroupBy operations (groupby, aggregate, transform)
   * Pivot tables (pivot\_table)
   * Crosstabulation (crosstab)
   * Applying functions to groups
6. **Merging and Joining DataFrames**
   * Concatenating DataFrames (concat)
   * Merging DataFrames (merge)
   * Joining DataFrames (join)

**Day 7-8: Advanced Features and Practical Applications**

1. **Advanced Data Analysis**
   * Window operations (rolling, expanding)
   * Time series analysis
   * Resampling data (resample)
   * Hierarchical indexing (MultiIndex)
2. **Visualization with Pandas**
   * Basic plotting (plot)
   * Customizing plots (plot options, subplots)
   * Integration with Matplotlib and Seaborn
3. **File I/O Operations**
   * Reading and writing data (CSV, Excel, JSON, SQL)
   * Working with large datasets (chunking, memory optimization)
   * Performance tips (vectorization, apply vs. vectorized operations)
4. **Advanced DataFrame Manipulations**
   * Reshaping data (melt, stack, unstack)
   * Pivoting DataFrames (pivot)
   * Combining data (combine\_first, update)
5. **Integration with Other Libraries**
   * Using pandas with NumPy
   * Pandas and SciPy
   * Pandas and SQL databases
   * Pandas in data pipelines (dask, vaex)

**Study Tips**

* **Practice Coding**: Apply what you learn through small exercises and problems.
* **Use Documentation**: Pandas documentation is comprehensive. Use it to understand functions and their applications.
* **Hands-On Projects**: Implement small projects or case studies to reinforce your learning.
* **Breaks and Reviews**: Take short breaks between sessions to avoid burnout and review what you’ve learned periodically.

**DAY 9-11 EDA**

**Day 9-10 Data Sourcing and Data Cleaning**

**Day 11 Univariate, Bivariate and Multivariate analysis**

**DAY 11-12 MATPLOTLIB**

**Day 11: Basics and Core Concepts (2 hours)**

1. **Introduction to Matplotlib**
   * Overview and installation
   * Importing Matplotlib
   * Basic structure of a Matplotlib plot
   * Using pyplot vs. object-oriented API
2. **Creating Basic Plots**
   * Line plots
     + Creating a simple line plot
     + Customizing line styles (color, width, style)
   * Scatter plots
     + Basic scatter plot
     + Customizing markers (size, color, shape)
3. **Plot Customization**
   * Titles and labels
     + Adding and customizing titles
     + Adding and customizing axis labels
   * Legends
     + Adding a legend
     + Customizing legend location and appearance

**Day 12: Advanced Features and Practical Applications (2 hours)**

1. **Advanced Plot Types**
   * Bar plots
     + Vertical and horizontal bar plots
     + Stacked bar plots
   * Histograms
     + Creating histograms
     + Customizing bins and density
   * Box plots
     + Creating and customizing box plots
   * Pie charts
     + Creating and customizing pie charts
2. **Advanced Customization**
   * Subplots
     + Creating multiple subplots using subplots and subplot
     + Adjusting subplot layout
   * Styles and themes
     + Applying built-in styles
     + Creating and using custom styles
   * Annotations
     + Adding annotations to plots
     + Customizing annotation appearance

**DAY 13-14 SEABORN**

**Day 13: Basics and Core Plotting**

**Session 1 (2 hours)**

1. **Introduction to Seaborn**
   * Overview and installation
   * Importing Seaborn
   * Differences between Seaborn and Matplotlib
2. **Basic Plotting with Seaborn**
   * Creating common plots:
     + Scatter plots
     + Line plots
     + Histograms
     + Bar plots
   * Customizing aesthetics:
     + Color palettes
     + Markers and line styles
     + Plot styles

**Day 14: Advanced Features and Specialized Plots**

**Session 2 (2 hours)**

1. **Advanced Plotting Techniques**
   * Pair plots for exploring relationships in datasets
   * Heatmaps for visualizing correlation matrices and categorical data
   * Joint plots for combining univariate and bivariate plots
2. **Specialized Plots**
   * Violin plots for visualizing distributions
   * Box plots for showing quartiles and outliers
   * Swarm plots for combining categorical and quantitative data

**DAY 15-18 INTERVIEW QUESTIONS AND CASE STUDIES**

Day 15-16 Interview Question on Visualization, EDA and NumPy and Pandas

Day 17- 18 Case Studies on EDA and visualization **Case Studies and Projects**

* + Real-world data analysis examples
  + Financial data analysis
  + Web scraping and data extraction
  + Data cleaning and preprocessing pipelines