Data Science Course Roadmap (EDA Focused)

1. Introduction to Data Science

We will start by understanding what Data Science actually is, why it matters in the real world, and how it connects with fields like AI, Machine Learning, and Data Analytics. You will also learn about the different roles in Data Science, the lifecycle of a project, and how Exploratory Data Analysis (EDA) is used in practice.

2. Understanding Data Types and Formats

Next, we will explore the different kinds of data you will work with, from structured tables to unstructured text. You will learn about numerical and categorical data, time series, and common file formats like CSV, Excel, JSON, and SQL. We will also cover metadata and how to document your data.

3. Python Essentials for Data Science

Before diving deep into analysis, we will make sure you are comfortable with Python basics. You will learn about variables, data types, lists, dictionaries, loops, functions, and working with Python libraries. We will also introduce basic error handling and list comprehensions.

4. Setting Up Python for Data Science

We will guide you through installing Python with Miniconda, setting up VS Code, and installing essential libraries. You will also learn how to create clean project environments and organize your work professionally.

5. Working with Pandas

Pandas is a must-have library for data manipulation. You will learn how to read and write data, select and filter information, handle missing values, group and aggregate data, merge datasets, and create pivot tables. We will also explore useful Pandas display settings.

6. Numpy for Numerical Computation

Here we will work with Numpy arrays to perform fast and efficient numerical operations. You will learn array creation, slicing, reshaping, broadcasting, statistical calculations, and even basic linear algebra.

7. Data Cleaning and Preprocessing

Clean data is the foundation of good analysis. You will learn how to find and fix missing values, remove duplicates, detect outliers, encode categorical data, scale features, parse dates, and handle messy text formatting.

8. Exploratory Data Analysis (EDA)

This is the heart of the course. We will explore data through statistics, patterns, and visualizations. You will learn univariate, bivariate, and multivariate analysis, calculate central tendency and dispersion, check correlations, and identify trends.

9. Data Visualization with Matplotlib

We will create clear, professional charts using Matplotlib, including line, bar, scatter, and histogram plots. You will learn to customize your plots, add annotations, and save them for reports.

10. Data Visualization with Seaborn

```
Seaborn makes beautiful statistical plots easy to create. You will learn to build
distribution plots, box plots, violin plots, heatmaps, and pair plots.
```

11. Interactive Visualization with Plotly

```
We will move into interactive visualizations where you can zoom, filter, and hover
for more information. You will learn to create line charts, bar charts, pie
charts, and dashboards.
```

12. Data Profiling Overview

```
Here you will discover automated profiling tools like ydata-profiling that give
you a quick, detailed overview of your dataset, including missing values,
distributions, and correlations.
```

13. EDA Project 1 - Titanic Dataset Analysis

```
Apply everything you have learned to a real dataset by cleaning, exploring, and
visualizing the Titanic dataset, followed by extracting insights.
```

14. EDA Project 2 - Google Play Store App Analysis

```
Work with a business-focused dataset to analyze app ratings, installs, and
categories. Visualize key trends and find correlations.
```

15. EDA Project 3 - Pakistan Population Dataset

```
Analyze time-series population trends, compare regions, and create interactive
visualizations to derive meaningful conclusions.
```

16. Final Review and What's Next

Wrap up by reviewing the EDA process, learning tips for real-world projects, exploring the basics of Machine Learning, and planning your next steps in Data Science.