Exploratory Data Analysis

Complete **curriculum of Exploratory Data Analysis (EDA)** from **basic to advanced**, suitable for students or beginners aiming for professional-level data analysis skills using Python:

Basic EDA

1. Introduction to EDA

- What is EDA?
- Why EDA is important
- Types of Data (Numerical, Categorical, Text, etc.)

2. Data Collection & Loading

- CSV, Excel, SQL, APIs
- Tools: `pandas`, `numpy`

3. Data Inspection

- `.head()`, `.tail()`, `.info()`, `.describe()`
- Checking data types
- Shape of dataset
- Identifying columns and data summary

4. Handling Missing Values

- Detecting missing data (`.isnull().sum()`)
- Dropping vs Imputing missing values
- Techniques: mean, median, mode, interpolation

5. Handling Duplicates

- `.duplicated()`, `.drop_duplicates()`

Intermediate EDA

6. Data Cleaning

- Renaming columns
- Correcting data types
- Removing outliers (IQR method, z-score)

7. Data Transformation

- Scaling: Min-Max, StandardScaler
- Encoding: Label Encoding, One-Hot Encoding
- Binning (converting continuous → categorical)

8. Univariate Analysis

- Categorical: Count plots, bar charts
- Numerical: Histograms, boxplots, KDE plots
- Tools: `matplotlib`, `seaborn`

9. Bivariate Analysis

- Numerical vs Numerical: Scatter plot, correlation heatmap
- Categorical vs Numerical: Boxplot, violin plot
- Categorical vs Categorical: Crosstab, stacked bar chart

10. Feature Engineering

- Creating new features
- Date/time features
- Groupby statistics

Advanced EDA

11. Multivariate Analysis

- Pair plots
- Grouped box plots
- Heatmaps & correlation matrix
- Pivot tables

12. Outlier Detection Techniques

- Boxplot
- Z-score
- IQR Method
- Isolation Forest (basic idea)

13. Dimensionality Reduction (Intro)

- PCA for visualization (2D/3D)

- t-SNE (for complex datasets)

14. EDA for Text Data

- Word frequency, word clouds
- Text length, stop words, n-grams
- Tools: `nltk`, `wordcloud`, `sklearn`

15. EDA for Time-Series

- Datetime parsing
- Time-based grouping (daily, weekly, monthly)
- Line plots, seasonality, trend detection

Expert/Project Level

16. EDA Best Practices

- Making EDA reproducible
- Writing data profiling reports (e.g., using `pandas_profiling`, `sweetviz`, or `ydata-profiling`)
- Business context + data insights

17. Automated EDA Tools

- `pandas_profiling` / `ydata-profiling`
- `sweetviz`
- `dtale`
- `AutoViz`

18. Interactive Dashboards (Optional)

- `Plotly`, `Dash`, `Streamlit` for visual EDA