

Mastering **Exploratory Data Analysis (EDA)** using Python is a valuable skill for any data analyst, engineer, or scientist. Here's a structured path to help you go from beginner to advanced.

---

## ✓ 1. Fundamentals to Know First

Before diving into EDA, make sure you're comfortable with:

- **Python basics:** loops, functions, conditionals, list/dict/set
- **Numpy:** for numerical operations
- **Pandas:** for data manipulation
- **Matplotlib / Seaborn:** for visualizations

If you're rusty on any of these, I can provide crash guides or exercises.

---

## 📦 2. Essential Libraries for EDA

Library	Purpose
pandas	Data cleaning and manipulation
numpy	Numerical operations
matplotlib	Basic plotting
seaborn	Statistical visualization
plotly	Interactive plots
missingno	Visualizing missing data
scipy.stats	Statistical analysis

---

## 📊 3. EDA Process Breakdown

### 🔧 Step 1: Load and Inspect Data

```
python
CopyEdit
import pandas as pd

df = pd.read_csv('your_file.csv')
df.head()
df.info()
df.describe()
df.columns
df.shape
```

### 🔍 Step 2: Understand the Data

- Identify data types (`df.dtypes`)
- Value counts: `df['column'].value_counts()`
- Check for duplicates: `df.duplicated().sum()`

### 💡 Step 3: Handle Missing Data

```
python
CopyEdit
df.isnull().sum()
df.dropna() # or df.fillna()
```

Use `missingno` for visualizing:

```
python
```

```
CopyEdit
import missingno as msno
msno.matrix(df)
```

#### **Step 4: Summary Stats and Distribution**

```
python
CopyEdit
df.describe()
df['column'].hist()
df['column'].value_counts().plot(kind='bar')
```

#### **Step 5: Data Visualization**

```
python
CopyEdit
import seaborn as sns
import matplotlib.pyplot as plt

sns.histplot(df['column'])
sns.boxplot(x='column', data=df)
sns.heatmap(df.corr(), annot=True)
sns.pairplot(df)
```

#### **Step 6: Correlation & Patterns**

```
python
CopyEdit
df.corr() # Pearson correlation
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
```