# **Exploratory Data Analysis & Data Visualization in Python**

Communicating insights effectively through data visualization





## What is EDA?

1

**EDA** stands for Exploratory Data Analysis.

2

It's a crucial first step in data analysis.

#### **EDA** helps in:

- Understanding data structure
- Detecting outliers
- Validating assumptions
- Identifying relationships and patterns

3



## **Loading and Inspecting Data**



#### **Python Code**

import pandas as pd

# Load dataset

df = pd.read\_csv("data.csv")



#### **Inspection Commands**

# Display first few rows

print(df.head())

# Get data types and summary

print(df.info())

print(df.describe())





## **Checking and Converting Data Data Types**

# Check data types

print(df.dtypes)

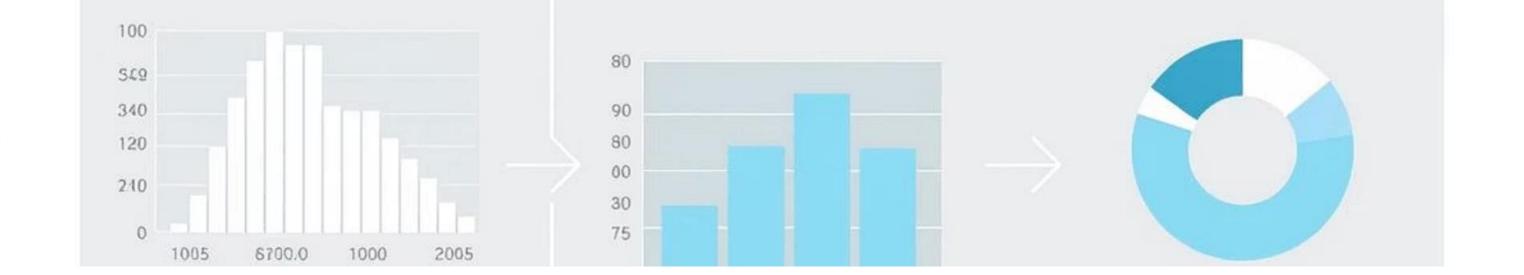
# Convert data type of a column

df["column\_name"] =

df["column\_name"].astype("int")



```
celesseflartivald.cnate_tape: far
comel: ietloectifietesr inlight a
ctetenlfourcremption, techiled:
 erribd: inn, all essectatiive. saw
ctley: The Motlorsfalel, deff - for )
ctivlengahaerlock, prop;
m, tab:, rab)
 avclinec Wontt leal((0)
 ijiock: corertisisters: (lfcrccaller
 r 1ff:2: go mitonlactenwall). (Intreman
 yblal: inacide seedulel, fosciabli; Meteril(wheellook
 rrihale: use lall>
sctec: ("2026 tatec) hari, sencractiubill, t
   Tutkscnoday reelcipstriess, that fil.dal
stesciiveal adischouly, ifer (coarsfuntions
 col: rpect(etagesaif/ffpselenderred
mirstay: aal, teerriseefiljyussa(Ilhanta
statale to evertyh instistiatall, met
piefe:Tytertabc restund)
                Hones dishrwill: (Or
    Mcloahat; (ame)
```



## **Univariate Analysis**

1 Frequency Distribution

```
# Frequency distribution
print(df["category_column"].value_counts())
```

2 Histogram for Numerical Data

```
import matplotlib.pyplot as plt
plt.hist(df["numeric_column"], bins=10,
color='skyblue', edgecolor='black')
plt.title("Histogram of Numeric Column")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```



## 105 15 44 20 0 011 128 204 109 173 188 200 104 295 1233 228 235 145 146

## **Bivariate Analysis**



#### **Scatter Plot**

```
# Scatter plot
plt.scatter(df["feature1"], df["feature2"], alpha=0.5)
plt.xlabel("Feature1")
plt.ylabel("Feature2")
plt.title("Feature1 vs Feature2")
plt.show()
```



## **Data Visualization in Python**



Data visualization helps in communicating insights effectively.



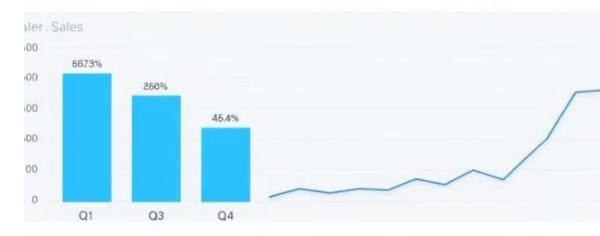
It translates complex data into clear visual stories.



#### erming by

#### erttner Nasies











arket Competitors		
10	Irayles	156%
18	Tranffis	350%
000	Tardars	553%
240	Memarts	307%



## **Matplotlib Basics**

#### **Code Example**

```
import matplotlib.pyplot as plt
# Line plot example
plt.plot(df["date_column"], df["value_column"])
plt.xlabel("Date")
plt.ylabel("Value")
plt.title("Line Plot of Value Over Time")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```







### **Seaborn for Enhanced Plots**

**Boxplot** import seaborn as sns # Boxplot Heatmap sns.boxplot(x="category ", y="value", data=df, import seaborn as sns palette="Set3") # Heatmap for plt.title("Boxplot of correlation Values by Category") sns.heatmap(df.corr(), plt.show() annot=True, cmap="coolwarm") plt.title("Correlation Matrix Heatmap") plt.show()





## **Pair Plots**

```
sns.pairplot(df, hue="target_column")
plt.show()
```



## **Customizing Plots**

1

```
sns.set_style("whitegrid")

sns.histplot(df["numeric_column"], bins=20, color="coral", kde=True)

plt.title("Enhanced Histogram with Seaborn Style")

plt.show()
```



## **Key Takeaways**

#### **Essential Step**

✓ EDA is essential for understanding data.

#### **Effective Communication**

✓ Data visualization conveys insights effectively.

#### **Tools of the Trade**

✓ Use Pandas, Matplotlib, and Seaborn for your EDA tasks.





## **Thank You!**

- Feel free to explore the GitHub repo for the project's complete code and code and documentation.
- <u>GitHub Link Placeholder</u>

