Exploratory Data Analysis & Data Visualization in Python

1 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.

3

EDA helps in:

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

Awais Manzoor

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())
```

Awais Manzoor

Checking and Converting Data Data Types

Check data types

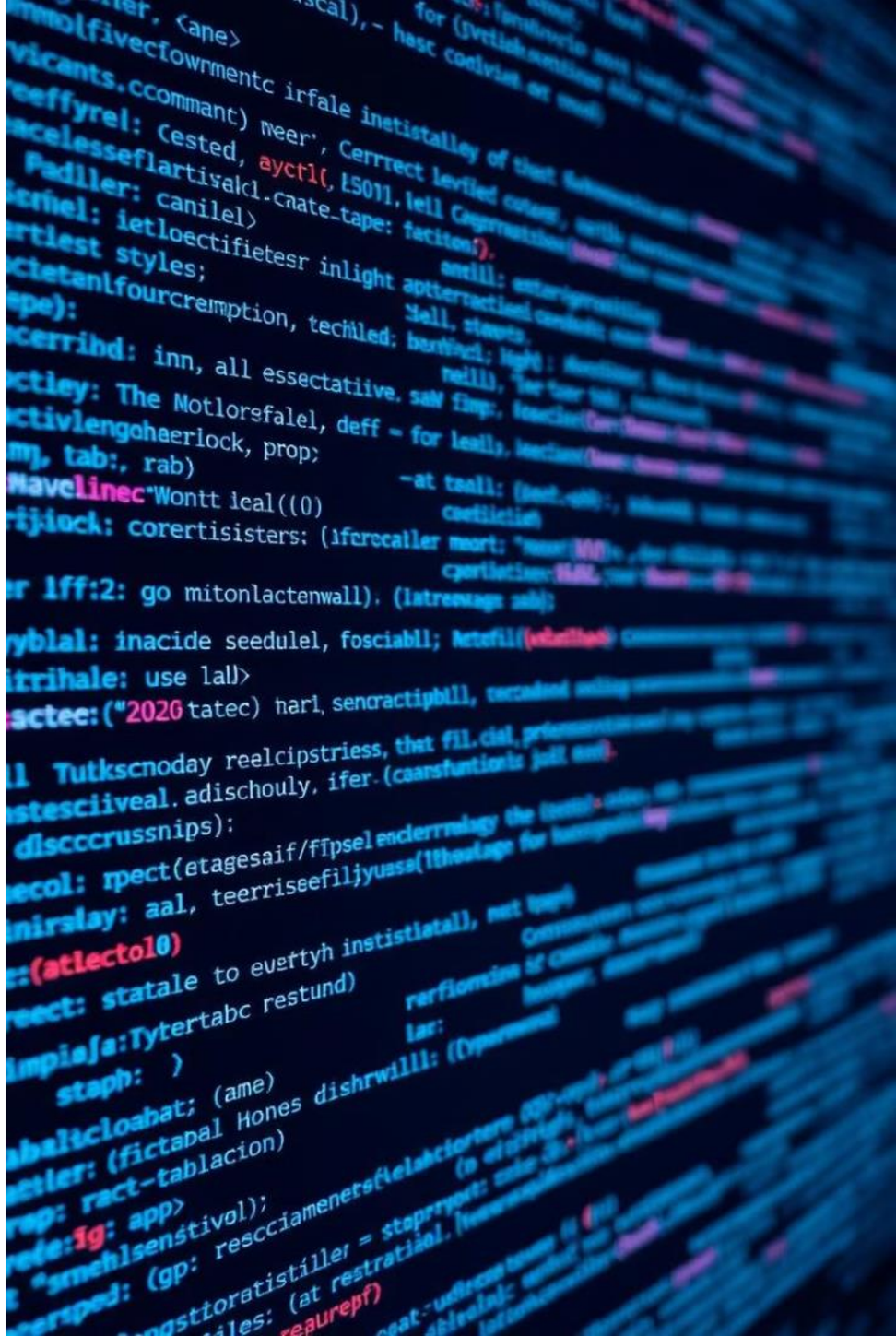
print(df.dtypes)

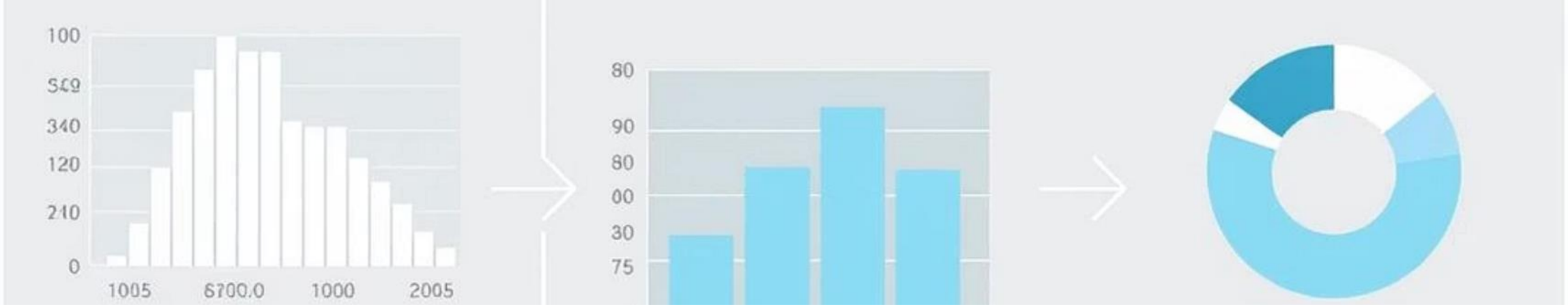
Convert data type of a column

df["column_name"] =

df["column_name"].astype("int")

Awais Manzoor





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()
```

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()
```

Awais Manzoor

Data Visualization in Python

- Data visualization helps in communicating insights effectively.
- It translates complex data into clear visual stories.

Awais Manzoor

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()
```

Awais Manzoor

Seaborn for Enhanced Plots

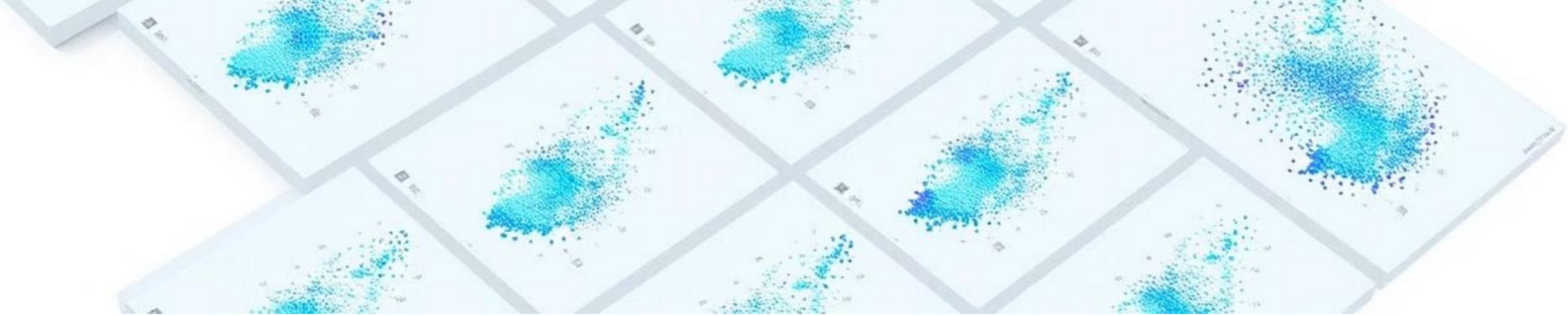
Boxplot — 1

```
import seaborn as sns
# Boxplot
sns.boxplot(x="category
", y="value", data=df,
palette="Set3")
plt.title("Boxplot of
Values by Category")
plt.show()
```

2 — **Heatmap**

```
import seaborn as sns
# Heatmap for
correlation
sns.heatmap(df.corr(),
annot=True,
cmap="coolwarm")
plt.title("Correlation
Matrix Heatmap")
plt.show()
```

Awais Manzoor



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()
```

Awais Manzoor

Key Takeaways

Essential Step

✓ EDA is essential for understanding data. data.

Effective Communication

✓ Data visualization conveys insights effectively.

Tools of the Trade

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

Thank You!

- 🖇️ Let's connect!
- 👉 Feel free to explore the GitHub repo for the project's complete code and code and documentation.
- GitHub Link Placeholder

Awais Manzoor