Python Notes for Data Analysts

1. What is Python?

Python is a **high-level, interpreted programming language** used for **data analysis, machine learning, web development, and automation**. It is widely used due to its **simplicity and powerful libraries**.

---

2. Basic Python Syntax

Concept	Example
Variables	x = 10; name = 'Alice'
Data Types	int, float, str, bool, list, tuple, dict, set
Conditional Statements	if, elif, else
Loops	for, while
Functions	def my_function(): return

Example:

```
name = "Alice"
print(f"Hello, {name}!")
```

---

3. Data Structures in Python

3.1 Lists

```
my_list = [1, 2, 3, 4]
my_list.append(5)
print(my_list)
```

3.2 Tuples

```
my_tuple = (1, 2, 3)
print(my_tuple[0])
```

3.3 Dictionaries

```
my_dict = {"name": "Alice", "age": 25}
print(my_dict["name"])
```

---

4. Working with Libraries

Library	Purpose
numpy	Numerical computing
pandas	Data manipulation
matplotlib	Data visualization
seaborn	Statistical plots
sklearn	Machine learning

Example:

```
import pandas as pd


data = {"Name": ["Alice", "Bob"], "Age": [25, 30]}
df = pd.DataFrame(data)
print(df)
```

---

5. File Handling

5.1 Reading a File

```
with open("file.txt", "r") as f:
    content = f.read()
    print(content)
```

5.2 Writing to a File

```
with open("file.txt", "w") as f:
    f.write("Hello, world!")
```

---

6. Data Analysis with Pandas

6.1 Reading Data

```
df = pd.read_csv("data.csv")
print(df.head())
```

6.2 Data Cleaning

```
df.dropna(inplace=True)  # Remove missing values
```

6.3 Data Aggregation

```
print(df.groupby("Category")["Sales"].sum())
```

---

7. Data Visualization

7.1 Matplotlib

```
import matplotlib.pyplot as plt
plt.plot([1, 2, 3], [4, 5, 6])
plt.show()
```

7.2 Seaborn

```
import seaborn as sns
sns.histplot(df["Age"])
```

---

8. SQL Integration with Python

```
import sqlite3

conn = sqlite3.connect("database.db")
cursor = conn.cursor()
cursor.execute("SELECT * FROM Customers")
print(cursor.fetchall())
```

---

9. Machine Learning Basics

9.1 Linear Regression

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

9.2 Classification

```
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier()
model.fit(X_train, y_train)
```

---

10. Web Scraping

```
import requests
from bs4 import BeautifulSoup

url = "https://example.com"
response = requests.get(url)
soup = BeautifulSoup(response.text, "html.parser")
print(soup.title.text)
```

---

11. Automation with Python

11.1 Automating Email Sending

```
import smtplib
server = smtplib.SMTP('smtp.gmail.com', 587)
server.starttls()
server.login('your_email@gmail.com', 'password')
server.sendmail('from_email', 'to_email', 'Subject: Test\n\nHello!')
server.quit()
```

📌 **Summary Table for Data Analysts**

Topic	Key Python Concept
Basic Syntax	Variables, Data Types, Loops, Functions
Data Structures	Lists, Tuples, Dictionaries, Sets
File Handling	open(), read(), write()
Data Manipulation	pandas for CSV, DataFrames
Visualization	matplotlib, seaborn
SQL & Python	sqlite3 for database operations
Machine Learning	sklearn, LinearRegression, DecisionTreeClassifier
Automation	requests, smtplib for web & email automation

Final Notes

- Python is **essential for Data Analysts** to **clean, analyze, and visualize data**.
- **Practice with real datasets** to improve skills.
- **Learn Pandas, Matplotlib, Seaborn, and Sklearn** for advanced analysis.
