DMV PRACTICAL FILE – DETAILED

Practical 1 – Data Loading, Storage and File Formats

Aim: To load and analyze sales data from CSV, Excel, and JSON file formats.
Theory: Data comes in various formats like CSV, Excel, and JSON. Pandas provides read_csv(), read_excel(), and read_json() to load data. Cleaning includes removing duplicates and missing values.
Code:

```
import pandas as pd
csv_data = pd.read_csv('sales.csv')
excel_data = pd.read_excel('sales.xlsx')
json_data = pd.read_json('sales.json')
combined = pd.concat([csv_data, excel_data, json_data])
combined.drop_duplicates(inplace=True)
combined.fillna(method='ffill', inplace=True)
print(combined.head())
```

Conclusion: Data from 3 different formats was loaded, cleaned, merged and analyzed.
Viva Q&A;:
1. Q: What is CSV? A: A comma-separated values file storing tabular data.
2. Q: How to load Excel file in pandas? A: Using read_excel().
3. Q: What is data cleaning? A: Removing missing, wrong, or duplicate data.
4. Q: What does concat() do? A: Combines multiple datasets.
5. Q: Why remove duplicates? A: To avoid repeated entries in analysis.

Practical 2 – Interacting with Web APIs

Aim: To fetch weather data using OpenWeatherMap API.

Theory: APIs allow apps to communicate. Weather API returns JSON data. JSON can be parsed and analyzed in Python.

Code:

```
import requests
api='your_api_key'
url=f"https://api.openweathermap.org/data/2.5/weather?q=Pune&appid;={api}"
data=requests.get(url).json()
print(data['main'])
```

Conclusion: Live weather info fetched from API.

Viva Q&A;:

1. What is API? - A system to connect applications.
2. What is JSON? - A text-based data format.
3. What library is used? - requests.
4. Why API key? - Authentication.
5. What is HTTP request? - A call made to a server.

Practical 3 – Data Cleaning and Preparation

Aim: Clean customer churn dataset.

Theory: Raw data contains missing values, duplicates, wrong types. Cleaning is required before analysis.

Steps:

- load CSV
- remove duplicates
- fill missing values
- convert data types
- handle outliers

Viva Q&A;:

1. What is data preprocessing?
2. Why remove duplicates?
3. What are outliers?
4. What is missing value treatment?
5. Why data types are important?

Practical 4 – Data Wrangling

Aim: Wrangle and transform real estate dataset.

Theory: Data wrangling means converting raw data into usable form:

- rename columns
- encode categorical data
- filter rows
- merge datasets

Viva Q&A; included.

Practical 5 – Data Visualization using matplotlib

Aim: Visualize AQI and pollutant trends.

Theory: Visualization helps understand patterns using charts like line plot, bar plot, scatter plot.

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('City_Air_Quality.csv')
plt.plot(df['Date'], df['AQI'])
plt.show()
```

Viva Q&A; included.

Practical 6 – Data Aggregation

Aim: Analyze sales by region using aggregation.
Theory: Groupby() helps summarize data.
Code:
df.groupby('Region')['Sales'].sum()
Viva Q&A; included.

Practical 7 – Time Series Data Analysis

Aim: Analyze stock price trends.

Theory: Time series uses dates as index. Trend, seasonality and forecasting are major parts.

Code:

```
df['Date']=pd.to_datetime(df['Date'])
plt.plot(df['Date'], df['Close'])
```

Viva Q&A; included.