EXPERIMENT NO: 1(c)

Data visualization

Conduct an experiment to differentiate Structured, Unstructured, and Semi-Structured data

Aim:

To create sample datasets for structured, unstructured, and semi-structured data and explain their characteristics using Python.

Algorithm:

1. Import required libraries (pandas, json).
2. Create a structured dataset using a DataFrame.
3. Represent unstructured data using plain text.
4. Represent semi-structured data using a JSON format.
5. Display all datasets and analyze their structure and organization.

Program:

UMAR FAROOK K
240701573

[5]:
```
import pandas as pd
structured_data = {
    'Employee_ID': [101, 102, 103],
    'Name': ['Alice', 'Bob', 'Charlie'],
    'Department': ['Data Science', 'IT', 'HR'],
    'Salary': [85000, 78000, 62000]
}
structured_df = pd.DataFrame(structured_data)
structured_df
semi_structured_data = [
    {'Name': 'Alice', 'Skills': ['Python', 'ML', 'SQL']},
    {'Name': 'Bob', 'Skills': ['Java', 'AWS'], 'Experience': 3},
    {'Name': 'Charlie', 'Skills': ['Excel', 'Recruitment']}
]
semi_structured_df = pd.DataFrame(semi_structured_data)
semi_structured_df
unstructured_data = [
    "Alice completed the Data Science project successfully.",
    "Team meeting notes: discuss new hiring process.",
    "Image_01.jpg - Company event photo."
]
unstructured_df = pd.DataFrame(unstructured_data, columns=['Unstructured_Data'])
unstructured_df
summary = {
    'Type': ['Structured', 'Semi-Structured', 'Unstructured'],
    'Example Format': ['CSV / SQL Table', 'JSON / XML', 'Text / Image / Audio'],
    'Schema': ['Fixed', 'Flexible', 'No Schema'],
    'Ease of Analysis': ['Easy', 'Moderate', 'Complex']
}
```

```
    "Image_01.jpg - Company event photo."
]
unstructured_df = pd.DataFrame(unstructured_data, columns=['Unstructured_Data'])
unstructured_df
summary = {
    'Type': ['Structured', 'Semi-Structured', 'Unstructured'],
    'Example Format': ['CSV / SQL Table', 'JSON / XML', 'Text / Image / Audio'],
    'Schema': ['Fixed', 'Flexible', 'No Schema'],
    'Ease of Analysis': ['Easy', 'Moderate', 'Complex']
}
summary_df = pd.DataFrame(summary)
summary_df
```

[5]:

	Type	Example Format	Schema	Ease of Analysis
0	Structured	CSV / SQL Table	Fixed	Easy
1	Semi-Structured	JSON / XML	Flexible	Moderate
2	Unstructured	Text / Image / Audio	No Schema	Complex

Result:

Thus, the Python code to perform the differentiation of structured, unstructured, and semi-structured data is successfully executed.