##Pandas--->> It is a powerful and widely-used open-source library in Python for data manipulation and analysis. It provides data structures and functions needed to work with structured data seamlessly and efficiently. Here are some key features and components of the Pandas library:

Key Features of Pandas

1. **Data Structures**:
   - **Series**: A one-dimensional array-like object containing a sequence of values. It is similar to a list or a one-dimensional NumPy array but with labeled axes (index).
   - **DataFrame**: A two-dimensional table of data with labeled axes (rows and columns). It can be thought of as a collection of Series objects sharing the same index.
2. **Data Alignment**: Automatic alignment of data for arithmetic operations on Series and DataFrame objects, making it easy to manage missing data.
3. **Data Cleaning**: Functions for handling missing data, such as filling, replacing, and dropping missing values.
4. **Data Transformation**: Tools for reshaping, pivoting, and transforming data, including functions for merging, joining, and concatenating DataFrame objects.
5. **Data Aggregation and Grouping**: Powerful group-by functionality to split data into groups, apply functions to each group, and combine the results.
6. **Time Series Handling**: Specialized tools for working with time series data, including date range generation and frequency conversion.
7. **Input and Output**: Functions for reading and writing data to and from various file formats, including CSV, Excel, SQL databases, and more.
8. **Visualization**: Integration with plotting libraries like Matplotlib to provide basic data visualization capabilities directly from DataFrame and Series objects.

Import pandas

##Panda---> Various functions----->>>

```
import numpy as np
import pandas as pd

l1 = [1, 2, 3, 4, 5, 6]
labels = ['a', 'b', 'c', 'd', 'e', 'f']
d1 = {"A":10, "B":20, "C":30, "D":40, "E":50}
s1 = pd.Series(l1)
s1
```

```
0    1
1    2
2    3
3    4
4    5
5    6
dtype: int64
```

```
s1[4]
```

```
5
```

```
s3 = pd.Series(data=l1, index=labels)
s3
```

```
a    1
b    2
c    3
d    4
e    5
f    6
dtype: int64
```

```
pd.Series(d1)
```

```
A    10
B    20
C    30
D    40
E    50
dtype: int64
```

```
arr = np.random.randint(low=1, high=100, size=(5, 6))
arr
```

```
array([[32, 47, 74,  4, 76, 87],
       [60, 54, 52, 57, 18, 16],
       [50, 32, 60, 73, 76,  3],
       [ 3, 86, 36, 53, 32, 60],
       [14, 98, 29, 82, 65, 30]])
```

```
pd.DataFrame(arr)
```


```
{"summary":"{\n  \"name\": \"pd\",\n  \"rows\": 5,\n  \"fields\": [\n
{\n      \"column\": 0,\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 23,\n        \"min\": 3,\n
\"max\": 60,\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          60,\n          14,\n          50\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": 1,\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 27,\n        \"min\": 32,\n
```

\"max\": 98,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   54,\n   98,\n   32\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": 2,\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 18,\n   \"min\": 29,\n
\"max\": 74,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   52,\n   29,\n   60\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": 3,\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 30,\n   \"min\": 4,\n
\"max\": 82,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   57,\n   82,\n   73\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": 4,\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 26,\n   \"min\": 18,\n
\"max\": 76,\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   18,\n   65,\n   76\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": 5,\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 34,\n   \"min\": 3,\n
\"max\": 87,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   16,\n   30,\n   3\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe"}

```
df = pd.DataFrame(arr, index=["A", "B", "C", "D", "E"], columns=["U",
"V", "W", "X", "Y", "Z"])
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 5,\n \"fields\": [\n
{\n   \"column\": \"U\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 23,\n   \"min\": 3,\n
\"max\": 60,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   60,\n   14,\n   50\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"V\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 27,\n   \"min\": 32,\n
\"max\": 98,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   54,\n   98,\n   32\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"W\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 18,\n   \"min\": 29,\n
\"max\": 74,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   52,\n   29,\n   60\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"X\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 30,\n   \"min\": 4,\n
\"max\": 82,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   57,\n   82,\n   73\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\

```
n    },\n    {\n      \"column\": \"Y\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 26,\n        \"min\": 18,\n
\"max\": 76,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          18,\n          65,\n          76\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Z\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 34,\n        \"min\": 3,\n
\"max\": 87,\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          16,\n          30,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


```
type(df)
```

```
pandas.core.frame.DataFrame
```

```
df[["X", "Z", "V"]]
```


```
{"summary":"{\n  \"name\": \"df[[\\\"X\\\", \\\"Z\\\", \\\"V\\\"]]\",\
n  \"rows\": 5,\n  \"fields\": [\n    {\n      \"column\": \"X\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
30,\n        \"min\": 4,\n        \"max\": 82,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          57,\n
82,\n          73\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Z\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 34,\n        \"min\": 3,\n        \"max\": 87,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          16,\n
30,\n          3\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"V\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 27,\n        \"min\": 32,\n        \"max\": 98,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          54,\n
98,\n          32\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```


```
df.loc["C"]
```

```
U    50
V    32
W    60
X    73
Y    76
Z     3
Name: C, dtype: int64
```

```
df.loc[["A", "B", "E"]]
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 3,\n  \"fields\": [\n
{\n      \"column\": \"U\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 23,\n        \"min\": 14,\n
```

\"max\": 60,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   32,\n   60,\n   14\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"V\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 27,\n   \"min\": 47,\n
\"max\": 98,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   47,\n   54,\n   98\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"W\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 22,\n   \"min\": 29,\n
\"max\": 74,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   74,\n   52,\n   29\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"X\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 39,\n   \"min\": 4,\n
\"max\": 82,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   4,\n   57,\n   82\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Y\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 30,\n   \"min\": 18,\n
\"max\": 76,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   76,\n   18,\n   65\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Z\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 37,\n   \"min\": 16,\n
\"max\": 87,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   87,\n   16,\n   30\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe"}

```
df.iloc[2]
```

```
U    50
V    32
W    60
X    73
Y    76
Z     3
Name: C, dtype: int64
```

```
df['New'] = [10, 20, 30, 40, 50]
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 5,\n \"fields\": [\n
{\n   \"column\": \"U\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 23,\n   \"min\": 3,\n
\"max\": 60,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   60,\n   14,\n   50\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"V\",\n   \"properties\": {\n

\"dtype\": \"number\",\n   \"std\": 27,\n   \"min\": 32,\n
\"max\": 98,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   54,\n   98,\n   32\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"W\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 18,\n   \"min\": 29,\n
\"max\": 74,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   52,\n   29,\n   60\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"X\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 30,\n   \"min\": 4,\n
\"max\": 82,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   57,\n   82,\n   73\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Y\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 26,\n   \"min\": 18,\n
\"max\": 76,\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   18,\n   65,\n   76\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Z\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 34,\n   \"min\": 3,\n
\"max\": 87,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   16,\n   30,\n   3\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"New\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 15,\n   \"min\": 10,\n
\"max\": 50,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   20,\n   50,\n   30\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
df.drop('New', axis=1)
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 5,\n \"fields\": [\n
{\n   \"column\": \"U\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 23,\n   \"min\": 3,\n
\"max\": 60,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   60,\n   14,\n   50\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"V\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 27,\n   \"min\": 32,\n
\"max\": 98,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   54,\n   98,\n   32\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"W\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 18,\n   \"min\": 29,\n
\"max\": 74,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   52,\n   29,\n   60\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"X\",\n   \"properties\": {\n

\"dtype\": \"number\",\n   \"std\": 30,\n   \"min\": 4,\n
\"max\": 82,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   57,\n   82,\n   73\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Y\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 26,\n   \"min\": 18,\n
\"max\": 76,\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   18,\n   65,\n   76\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Z\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 34,\n   \"min\": 3,\n
\"max\": 87,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   16,\n   30,\n   3\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe"}

```
df['X'] % 2 == 0
```

```
A     True
B    False
C    False
D    False
E     True
Name: X, dtype: bool
```

```
df[df['X'] % 2 == 0]
```

{"summary":"{\n \"name\": \"df[df['X'] % 2 == 0]\",\n \"rows\": 2,\n
\"fields\": [\n   {\n   \"column\": \"U\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\": 12,\n
\"min\": 14,\n   \"max\": 32,\n   \"num_unique_values\": 2,\
n   \"samples\": [\n   14,\n   32\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"V\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 36,\n   \"min\": 47,\n
\"max\": 98,\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   98,\n   47\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"W\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 31,\n   \"min\": 29,\n
\"max\": 74,\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   29,\n   74\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"X\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 55,\n   \"min\": 4,\n
\"max\": 82,\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   82,\n   4\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"Y\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 7,\n   \"min\": 65,\n

\"max\": 76,\n   \"num_unique_values\": 2,\n   \"samples\": [\n   65,\n   76\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Z\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 40,\n   \"min\": 30,\n   \"max\": 87,\n   \"num_unique_values\": 2,\n   \"samples\": [\n   30,\n   87\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"New\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 28,\n   \"min\": 10,\n   \"max\": 50,\n   \"num_unique_values\": 2,\n   \"samples\": [\n   50,\n   10\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
df[df['X'] % 2 == 0]['Y']
```

```
A    76
E    65
Name: Y, dtype: int64
```

```
df.reset_index()
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 5,\n \"fields\": [\n   {\n   \"column\": \"index\",\n   \"properties\": {\n   \"dtype\": \"string\",\n   \"num_unique_values\": 5,\n   \"samples\": [\n   \"B\",\n   \"E\",\n   \"C\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"U\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 23,\n   \"min\": 3,\n   \"max\": 60,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   60,\n   14,\n   50\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"V\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 27,\n   \"min\": 32,\n   \"max\": 98,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   54,\n   98,\n   32\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"W\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 18,\n   \"min\": 29,\n   \"max\": 74,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   52,\n   29,\n   60\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"X\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 30,\n   \"min\": 4,\n   \"max\": 82,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   57,\n   82,\n   73\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Y\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 26,\n   \"min\": 18,\n

\"max\": 76,\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   18,\n   65,\n   76\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Z\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 34,\n   \"min\": 3,\n
\"max\": 87,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   16,\n   30,\n   3\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"New\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 15,\n   \"min\": 10,\n
\"max\": 50,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   20,\n   50,\n   30\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe"}

```
df['States'] = "PB RJ DL CHD J&K".split()
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 5,\n \"fields\": [\n
{\n   \"column\": \"U\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 23,\n   \"min\": 3,\n
\"max\": 60,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   60,\n   14,\n   50\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"V\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 27,\n   \"min\": 32,\n
\"max\": 98,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   54,\n   98,\n   32\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"W\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 18,\n   \"min\": 29,\n
\"max\": 74,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   52,\n   29,\n   60\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"X\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 30,\n   \"min\": 4,\n
\"max\": 82,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   57,\n   82,\n   73\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Y\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 26,\n   \"min\": 18,\n
\"max\": 76,\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   18,\n   65,\n   76\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Z\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 34,\n   \"min\": 3,\n
\"max\": 87,\n   \"num_unique_values\": 5,\n   \"samples\":
[\n   16,\n   30,\n   3\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"New\",\n   \"properties\": {\n

```
\"dtype\": \"number\",\n        \"std\": 15,\n        \"min\": 10,\n
\"max\": 50,\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          20,\n          50,\n          30\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"States\",\n      \"properties\":
{\n        \"dtype\": \"string\",\n        \"num_unique_values\": 5,\n
\"samples\": [\n          \"RJ\",\n          \"J&K\",\n
\"DL\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```


```
df.set_index('States')
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 5,\n  \"fields\": [\n
{\n      \"column\": \"States\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 5,\n
\"samples\": [\n          \"RJ\",\n          \"J&K\",\n
\"DL\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"U\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 23,\n        \"min\": 3,\n        \"max\": 60,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          60,\n
14,\n          50\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"V\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 27,\n        \"min\": 32,\n        \"max\": 98,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          54,\n
98,\n          32\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"W\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 18,\n        \"min\": 29,\n        \"max\": 74,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          52,\n
29,\n          60\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"X\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 30,\n        \"min\": 4,\n        \"max\": 82,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          57,\n
82,\n          73\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Y\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 26,\n        \"min\": 18,\n        \"max\": 76,\n
\"num_unique_values\": 4,\n        \"samples\": [\n          18,\n
65,\n          76\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Z\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 34,\n        \"min\": 3,\n        \"max\": 87,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          16,\n
30,\n          3\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"New\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
```

\"std\": 15,\n   \"min\": 10,\n   \"max\": 50,\n
\"num_unique_values\": 5,\n   \"samples\": [\n   20,\n
50,\n   30\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
d = {"A":[1, 2, 3, np.nan],
    "B":[5, np.nan, np.nan, np.nan],
    "C":[10, 20, 30, 40],
    "D":[np.nan, np.nan, np.nan, np.nan]}

df = pd.DataFrame(d)
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 4,\n \"fields\": [\n
{\n   \"column\": \"A\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 1.0,\n   \"min\": 1.0,\n
\"max\": 3.0,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   1.0,\n   2.0,\n   3.0\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"B\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": null,\n   \"min\": 5.0,\
n   \"max\": 5.0,\n   \"num_unique_values\": 1,\n
\"samples\": [\n   5.0\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"C\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 12,\n   \"min\": 10,\n
\"max\": 40,\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   20\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"D\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": null,\n   \"min\": null,\n   \"max\": null,\n
\"num_unique_values\": 0,\n   \"samples\": [],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
df.isnull()
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 4,\n \"fields\": [\n
{\n   \"column\": \"A\",\n   \"properties\": {\n
\"dtype\": \"boolean\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   true,\n   false\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"B\",\n   \"properties\": {\n
\"dtype\": \"boolean\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   true,\n   false\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"C\",\n   \"properties\": {\n
\"dtype\": \"boolean\",\n   \"num_unique_values\": 1,\n
\"samples\": [\n   false\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\

```
n    },\n    {\n      \"column\": \"D\",\n      \"properties\": {\n
\"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          true\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe"}
```


```
df.isnull().sum()
```

```
A    1
B    3
C    0
D    4
dtype: int64
```

```
df.dropna(axis=1, thresh=2)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"A\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 1.0,\n        \"min\": 1.0,\n
\"max\": 3.0,\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          1.0,\n          2.0,\n          3.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"C\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 12,\n        \"min\": 10,\n
\"max\": 40,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          20,\n          40,\n          10\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```


```
df.fillna("FILL")
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"A\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 4,\n
\"samples\": [\n          2.0,\n          \"FILL\",\n          1.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"B\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"FILL\",\n          5.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"C\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 12,\n        \"min\": 10,\n
\"max\": 40,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          20,\n          40\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"D\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 1,\n        \"samples\":
[\n          \"FILL\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```

```
d = {"Company":["FB", "GOOGLE", "MICROSOFT", "FB", "GOOGLE", "FB",
"MICROSOFT", "FB"],
    "Employee":["Sam", "Rachel", "Maddy", "Joe", "Srishti", "Shivay",
"Pushpa", "Kirti"],
    "Sales":[1000, 500, 550, 2000, 890, 500, 350, 350]}

df = pd.DataFrame(d)
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"Company\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 3,\n
\"samples\": [\n          \"FB\",\n          \"GOOGLE\",\n
\"MICROSOFT\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Employee\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 8,\n        \"samples\":
[\n          \"Rachel\",\n          \"Shivay\",\n          \"Sam\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Sales\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 551,\n
\"min\": 350,\n        \"max\": 2000,\n        \"num_unique_values\":
6,\n        \"samples\": [\n          1000,\n          500,\n
350\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```

```
df.min()
```

```
Company      FB
Employee    Joe
Sales       350
dtype: object
```

```
grouped_df = df.groupby('Company')
grouped_df
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at
0x782090c79c90>
```

```
grouped_df.describe()
```

```
{"summary":"{\n  \"name\": \"grouped_df\",\n  \"rows\": 3,\n
\"fields\": [\n    {\n      \"column\": [\n        \"Company\",\n
\"\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          \"FB\",\n          \"GOOGLE\",\n          \"MICROSOFT\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"count\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 1.1547005383792515,\n        \"min\":
```

```
2.0,\n        \"max\": 4.0,\n        \"num_unique_values\": 2,\n
\"samples\": [\n          2.0,\n          4.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"mean\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 256.3323038557567,\n        \"min\":
450.0,\n        \"max\": 962.5,\n        \"num_unique_values\": 3,\n
\"samples\": [\n          962.5,\n          695.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"std\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 317.12156582082565,\n        \"min\":
141.4213562373095,\n        \"max\": 745.4025757937787,\n
\"num_unique_values\": 3,\n        \"samples\": [\n
745.4025757937787,\n          275.77164466275354\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"min\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 86.60254037844386,\n        \"min\":
350.0,\n        \"max\": 500.0,\n        \"num_unique_values\": 2,\n
\"samples\": [\n          500.0,\n          350.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"25%\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 100.943466025956,\n        \"min\":
400.0,\n        \"max\": 597.5,\n        \"num_unique_values\": 3,\n
\"samples\": [\n          462.5,\n          597.5\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"50%\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 159.7132847741018,\n        \"min\":
450.0,\n        \"max\": 750.0,\n        \"num_unique_values\": 3,\n
\"samples\": [\n          750.0,\n          695.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"75%\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 378.0128966053936,\n        \"min\":
500.0,\n        \"max\": 1250.0,\n        \"num_unique_values\": 3,\n
\"samples\": [\n          1250.0,\n          792.5\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": [\n        \"Sales\",\n
\"max\"\n      ],\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 758.3095234357362,\n        \"min\":
550.0,\n        \"max\": 2000.0,\n        \"num_unique_values\": 3,\n
\"samples\": [\n          2000.0,\n          890.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```


```
df["Sales"]=df["Sales"].apply(lambda x: x+100)
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 8,\n \"fields\": [\n{\n   \"column\": \"Company\",\n   \"properties\": {\n\"dtype\": \"category\",\n   \"num_unique_values\": 3,\n\"samples\": [\n   \"FB\",\n   \"GOOGLE\",\n\"MICROSOFT\"\n   ],\n   \"semantic_type\": \"\",\n\"description\": \"\"\n   }\n   },\n   {\n   \"column\":\"Employee\",\n   \"properties\": {\n   \"dtype\":\"string\",\n   \"num_unique_values\": 8,\n   \"samples\":[\n   \"Rachel\",\n   \"Shivay\",\n   \"Sam\"\n],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n}\n   },\n   {\n   \"column\": \"Sales\",\n   \"properties\":{\n   \"dtype\": \"number\",\n   \"std\": 551,\n\"min\": 450,\n   \"max\": 2100,\n   \"num_unique_values\":6,\n   \"samples\": [\n   1100,\n   600,\n450\n   ],\n   \"semantic_type\": \"\",\n\"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
new_employee=pd.DataFrame({"Company":["Google"],"Employee":
["Jacky"],"Sales":[1230]})
new_employee
```

{"summary":"{\n \"name\": \"new_employee\",\n \"rows\": 1,\n\"fields\": [\n   {\n   \"column\": \"Company\",\n\"properties\": {\n   \"dtype\": \"string\",\n\"num_unique_values\": 1,\n   \"samples\": [\n\"Google\"\n   ],\n   \"semantic_type\": \"\",\n\"description\": \"\"\n   }\n   },\n   {\n   \"column\":\"Employee\",\n   \"properties\": {\n   \"dtype\":\"string\",\n   \"num_unique_values\": 1,\n   \"samples\":[\n   \"Jacky\"\n   ],\n   \"semantic_type\": \"\",\n\"description\": \"\"\n   }\n   },\n   {\n   \"column\":\"Sales\",\n   \"properties\": {\n   \"dtype\": \"number\",\n\"std\": null,\n   \"min\": 1230,\n   \"max\": 1230,\n\"num_unique_values\": 1,\n   \"samples\": [\n   1230\n],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n}\n   }\n]\n}","type":"dataframe","variable_name":"new_employee"}

```
df=pd.concat([df,new_employee])
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 9,\n \"fields\": [\n{\n   \"column\": \"Company\",\n   \"properties\": {\n\"dtype\": \"category\",\n   \"num_unique_values\": 4,\n\"samples\": [\n   \"GOOGLE\",\n   \"Google\",\n\"FB\"\n   ],\n   \"semantic_type\": \"\",\n\"description\": \"\"\n   }\n   },\n   {\n   \"column\":\"Employee\",\n   \"properties\": {\n   \"dtype\":\"string\",\n   \"num_unique_values\": 9,\n   \"samples\":[\n   \"Kirti\",\n   \"Rachel\",\n   \"Shivay\"\n

```
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Sales\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 529,\n
\"min\": 450,\n        \"max\": 2100,\n        \"num_unique_values\":
7,\n        \"samples\": [\n          1100,\n          600,\n
450\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```


```
df[df['Company'] == 'MICROSOFT'].index
```

```
Index([2, 6], dtype='int64')
```