Lab Assignment 24

Student Name: Chauhan Vandana Ramdayal

Student Id: AF0411629

Topic: Panda series

Pandas library

Pandas is an open-source data manipulation and analysis library for the Python programming language that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series and for working with structured data, making it one of the most popular and widely used libraries for data analysis and manipulation in the Python ecosystem.

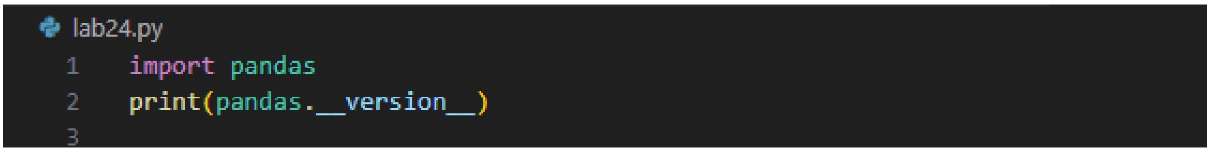
This library is built on top of the NumPy library of Python. Pandas is fast and it has high performance & productivity for users. Key features and components of the Pandas library include:

Key features and components of the Pandas library include:

- **DataFrame**: Two-dimensional table, similar to a spreadsheet/SQL table, with columns supporting different data types.
- **Series**: One-dimensional array-like structure; used to represent a single column of data with an index.
- **Data Input/Output**: Functions to read and write data from/to formats like CSV, Excel, SQL databases, etc.
- **Data Cleaning**: Tools to handle missing data, remove duplicates, and prepare data for analysis.
- **Data Transformation**: Supports filtering, merging, reshaping, and other preprocessing tasks.
- **Grouping & Aggregation**: Group data by one or more columns and perform summary/aggregation operations.
- **Indexing & Selection**: Powerful label-based and integer-based indexing for accessing specific parts of the data.
- **Data Visualization**: Integrates with Matplotlib and Seaborn for creating plots and visualizations.
- **Performance**: Optimized for handling large datasets efficiently using vectorized operations.
- **Time Series Support**: Strong functionality for working with time series data, ideal for financial/temporal analysis.

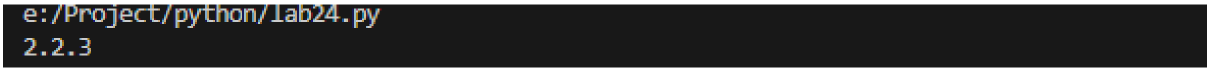
Q1. Write a program to check python version.

Code:



```python
import pandas
print(pandas.__version__)
```

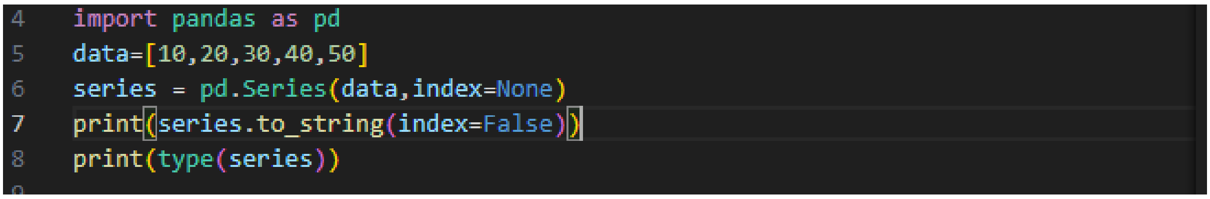
Output:



```
e:/Project/python/lab24.py
2.2.3
```

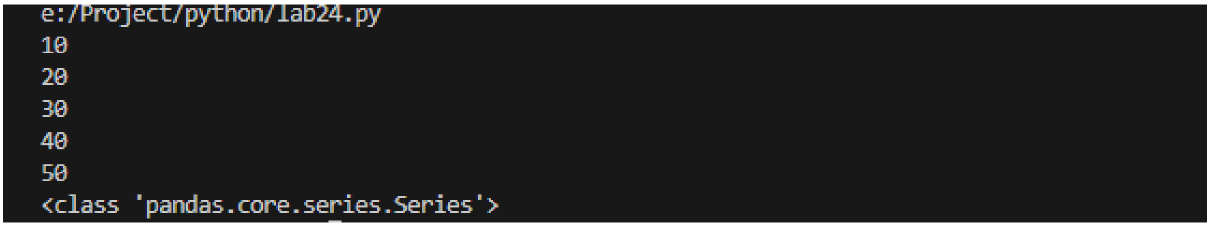
Q2.Create Pandas series [10,20,30,40,50] without index

Code:



```python
import pandas as pd
data=[10,20,30,40,50]
series = pd.Series(data,index=None)
print(series.to_string(index=False))
print(type(series))
```

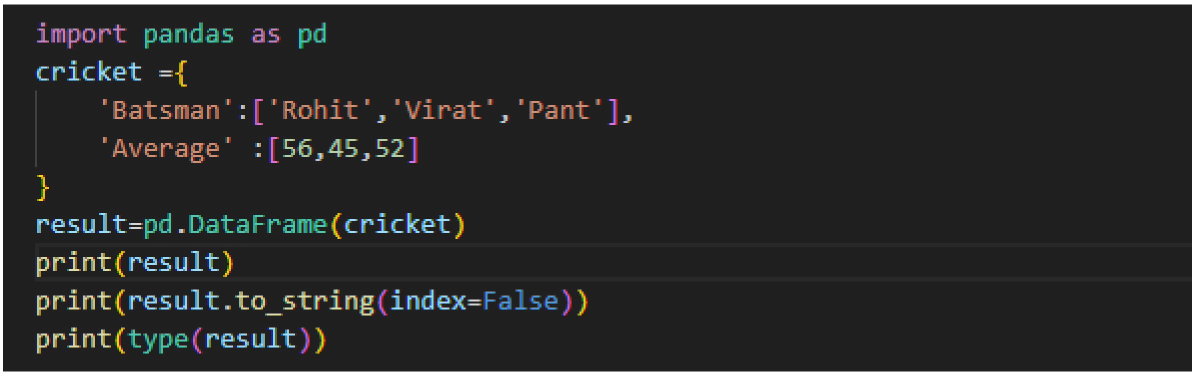
Output:



```
e:/Project/python/lab24.py
10
20
30
40
50
<class 'pandas.core.series.Series'>
```

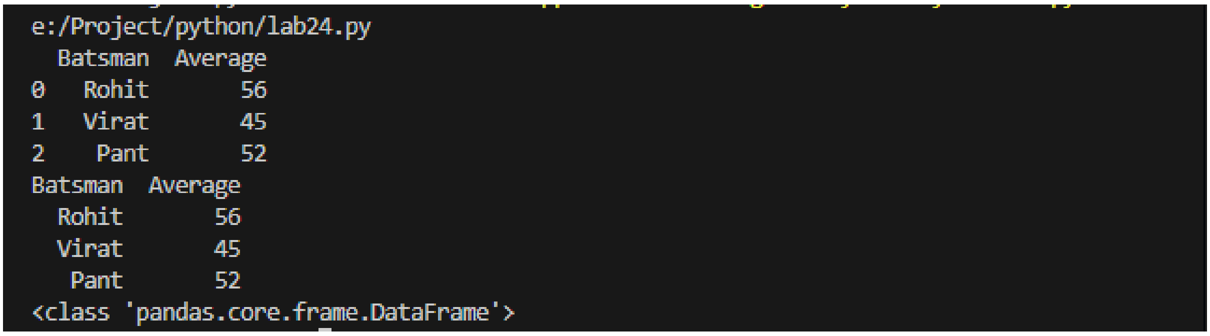
Q3.Create Pandas DataFrame example

Code:



```python
import pandas as pd
cricket ={
    'Batsman':['Rohit','Virat','Pant'],
    'Average' :[56,45,52]
}
result=pd.DataFrame(cricket)
print(result)
print(result.to_string(index=False))
print(type(result))
```

Output:

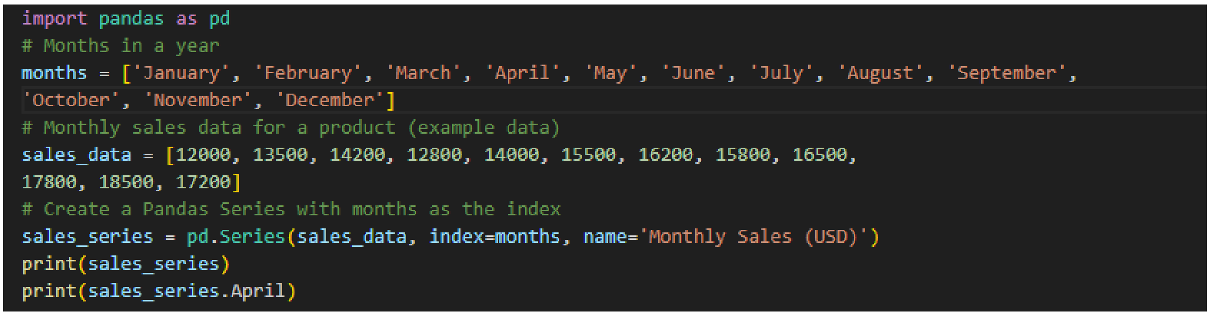


```
e:/Project/python/lab24.py
  Batsman  Average
0   Rohit       56
1   Virat       45
2    Pant       52
Batsman  Average
  Rohit       56
  Virat       45
   Pant       52
<class 'pandas.core.frame.DataFrame'>
```

Q4. **Example: Monthly Sales Data Imagine you are a sales manager for a retail company, and you want to analyze the monthly sales performance of a particular product in a given**

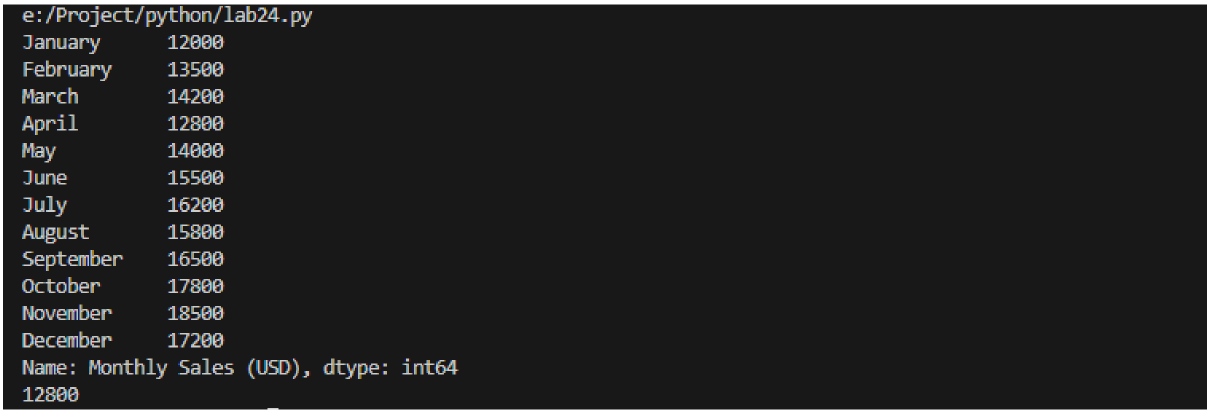
year. You have recorded the monthly sales figures for that product, and you want to represent this data using a Pandas Series.

Code:



```
import pandas as pd
# Months in a year
months = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September',
'October', 'November', 'December']
# Monthly sales data for a product (example data)
sales_data = [12000, 13500, 14200, 12800, 14000, 15500, 16200, 15800, 16500,
17800, 18500, 17200]
# Create a Pandas Series with months as the index
sales_series = pd.Series(sales_data, index=months, name='Monthly Sales (USD)')
print(sales_series)
print(sales_series.April)
```

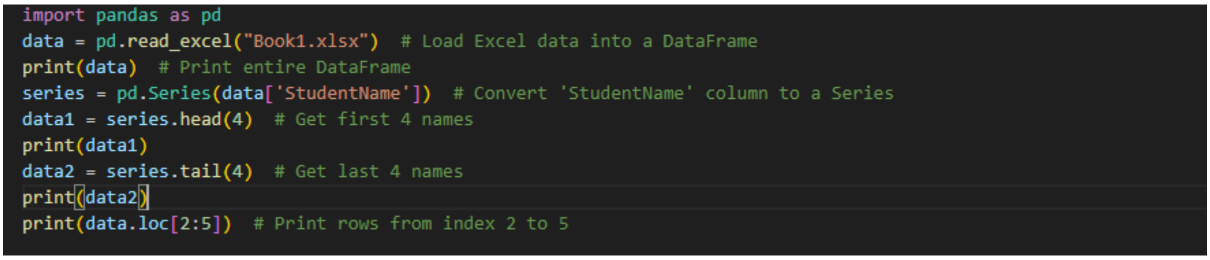
Output:



```
e:/Project/python/lab24.py
January      12000
February     13500
March        14200
April        12800
May          14000
June         15500
July         16200
August       15800
September    16500
October      17800
November     18500
December     17200
Name: Monthly Sales (USD), dtype: int64
12800
```

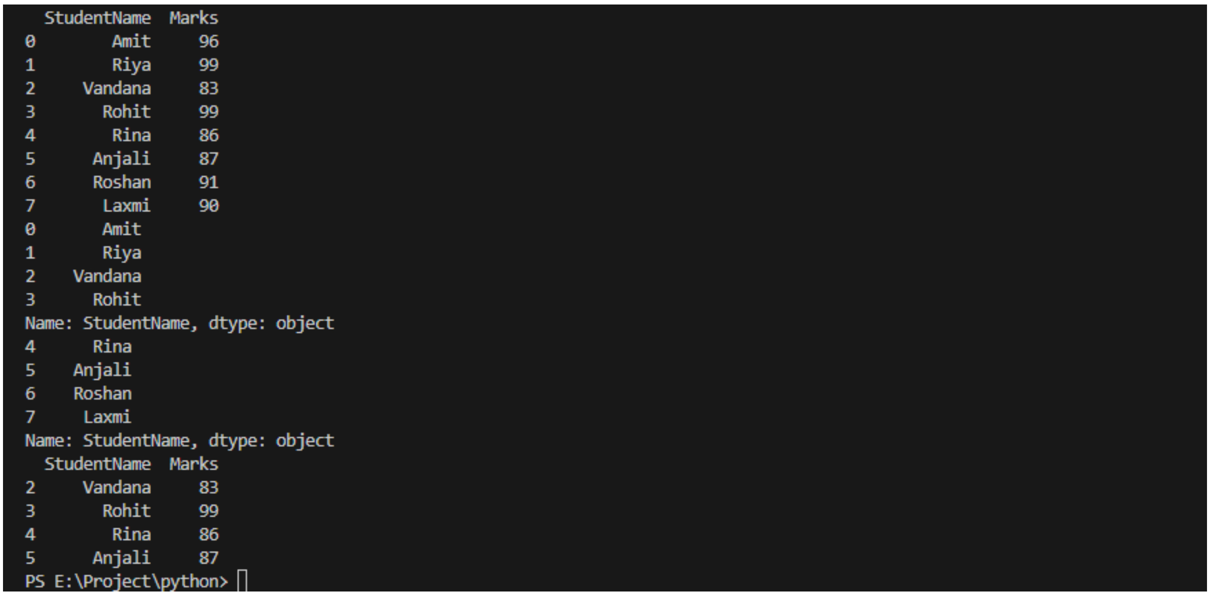
Example

Code:



```
import pandas as pd
data = pd.read_excel("Book1.xlsx")  # Load Excel data into a DataFrame
print(data)  # Print entire DataFrame
series = pd.Series(data['StudentName'])  # Convert 'StudentName' column to a Series
data1 = series.head(4)  # Get first 4 names
print(data1)
data2 = series.tail(4)  # Get last 4 names
print(data2)
print(data.loc[2:5])  # Print rows from index 2 to 5
```

Output:



```
  StudentName  Marks
0        Amit     96
1        Riya     99
2     Vandana     83
3       Rohit     99
4        Rina     86
5      Anjali     87
6      Roshan     91
7       Laxmi     90
0       Amit
1       Riya
2    Vandana
3      Rohit
Name: StudentName, dtype: object
4      Rina
5    Anjali
6    Roshan
7     Laxmi
Name: StudentName, dtype: object
  StudentName  Marks
2     Vandana     83
3       Rohit     99
4        Rina     86
5      Anjali     87
PS E:\Project\python> 
```