rutvik balar

PANDAS Creating data frames

```
import pandas as pd  #importing panda
```

creating data frame

```
# Calling DataFrame constructor
df = pd.DataFrame()

print(df)
```

```
Empty DataFrame
Columns: []
Index: []
```

reading files

```
df.to_csv('data.csv')  #file created
df = pd.read_csv('data.csv', index_col=0) #reading file
df
```

slicing and manipulation

```
# Initializing the nested list with Data set
player_list = [['Dhoni', 36, 75, 5428000],
               ['A.B.D', 38, 74, 3428000],
               ['Kholi', 31, 70, 8428000],
               ['Smith', 34, 80, 4428000],
               ['Gayle', 40, 100, 4528000],
               ['Root', 33, 72, 7028000],
               ['Peterson', 42, 85, 2528000]]

# creating a pandas dataframe
df = pd.DataFrame(player_list, columns=['Name', 'Age', 'Weight', 'Salary'])

# data frame before slicing
df
```

	Name	Age	Weight	Salary
0	Dhoni	36	75	5428000
1	A.B.D	38	74	3428000
2	Kholi	31	70	8428000
3	Smith	34	80	4428000
4	Gayle	40	100	4528000
5	Root	33	72	7028000
6	Peterson	42	85	2528000

```
# Slicing rows in data frame
df1 = df.iloc[0:4]

# data frame after slicing
df1
```

	Name	Age	Weight	Salary
0	Dhoni	36	75	5428000
1	A.B.D	38	74	3428000
2	Kholi	31	70	8428000
3	Smith	34	80	4428000

manipulation

```
# Sorting by column 'Weight'
df.sort_values(by=['Weight'])
```

	Name	Age	Weight	Salary
2	Kholi	31	70	8428000
5	Root	33	72	7028000
1	A.B.D	38	74	3428000
0	Dhoni	36	75	5428000
3	Smith	34	80	4428000
6	Peterson	42	85	2528000
4	Gayle	40	100	4528000

Exporting data to files

```
df.to_csv("record.csv")
```

Columns and row manipulations with loops

```
for col_name, data in df.items():
    print("col_name:",col_name, "\ndata:",data)
```

```
col_name: Name
data: 0       Dhoni
1       A.B.D
2       Kholi
3       Smith
4       Gayle
5        Root
6    Peterson
Name: Name, dtype: object
col_name: Age
data: 0    36
1    38
2    31
3    34
4    40
5    33
6    42
Name: Age, dtype: int64
col_name: Weight
data: 0     75
1     74
2     70
3     80
4    100
5     72
6     85
Name: Weight, dtype: int64
col_name: Salary
data: 0    5428000
1    3428000
2    8428000
3    4428000
4    4528000
5    7028000
6    2528000
Name: Salary, dtype: int64
```

Use pandas for masking data and reading if in Boolean format.

```
df.Weight>75
```

```
0    False
1    False
2    False
3     True
4     True
5    False
6     True
Name: Weight, dtype: bool
```

mask

```
df.mask(df.Weight > 75, 0)
```

	Name	Age	Weight	Salary
0	Dhoni	36	75	5428000
1	A.B.D	38	74	3428000
2	Kholi	31	70	8428000
3	0	0	0	0
4	0	0	0	0
5	Root	33	72	7028000
6	0	0	0	0