Name: Abhishek Mali
Roll No.: 213070
PRN No.: 22320024

Tutorial No.1

1. Perform the following operations using Python on suitable data sets, read data from different formats(like csv, xls),indexing and selecting data, sort data, describe attributes of data, checking data types of each column, counting unique values of data, format of each column, converting variable data type (e.g. from long to short, vice versa), identifying missing values and fill in the missing values.

```
[63]: import pandas as pd
      import numpy as np

[77]: dfnpd.read_csv("nba.csv")
```

Read Csv File

```
[78]: df
```

[78]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25	2-Jun	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99.0	SF	25	6-Jun	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30.0	SG	27	5-Jun	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28.0	SG	22	5-Jun	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8.0	PF	29	10-Jun	231	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
452	Trey Lyles	Utah Jazz	41.0	PF	20	10-Jun	234	Kentucky	2239800.0
453	Shelvin Mack	Utah Jazz	8.0	PG	26	3-Jun	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25.0	PG	24	1-Jun	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21.0	C	26	3-Jul	256	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24.0	C	26	Jul-00	231	Kansas	947276.0

457 rows × 9 columns

Select column Here

```
[51]: coludf["Position"]

[54]: colu

[54]: 0      PG
      1      SF
      2      SG
      3      SG
      4      PF
            ...
      452    PF
      453    PG
      454    PG
      455     C
      456     C
      Name: Position, Length: 457, dtype: object
```

Toggle output scrolling

		Utah Jazz	2239800.0	PF	41.0	Trey Lyles	10-Jun	Kentucky	20
453	203	Utah Jazz	2433333.0	PG	8.0	Shelvin Mack	3-Jun	Butler	26
454	179	Utah Jazz	900000.0	PG	25.0	Raul Neto	1-Jun	NaN	24
455	256	Utah Jazz	2900000.0	C	21.0	Tibor Pleiss	3-Jul	NaN	26
456	231	Utah Jazz	947276.0	C	24.0	Jeff Withey	Jul-00	Kansas	26

457 rows × 9 columns

Describe

```
df.describe()
```

	Number	Age	Weight	Salary
count	455.000000	457.000000	457.000000	4.450000e+02
mean	17.669132	26.938731	221.522976	4.853190e+06
std	15.945315	4.404016	26.368343	5.230409e+06
min	0.000000	19.000000	161.000000	3.088800e+04
25%	5.000000	24.000000	200.000000	1.074169e+06
50%	13.000000	26.000000	220.000000	2.841960e+06
75%	25.000000	30.000000	240.000000	6.500000e+06
max	99.000000	40.000000	307.000000	2.500000e+07

Row index

[51]:
```
rowe=df.loc[1]
```

[53]:
```
rowe
```

[53]:
```
Name            Jae Crowder
Team         Boston Celtics
Number                 99.0
Position                 SF
Age                      25
Height                6-Jun
Weight                  235
College           Marquette
Salary            6796117.0
Name: 1, dtype: object
```

Sorting Values

[54]:
```
df.sort_index(axis=1,ascending=False)#Sort the in  Descendi axis =1 mean column#ascending false mean descene
```

[54]:

	Weight	Team	Salary	Position	Number	Name	Height	College	Age
0	180	Boston Celtics	7730337.0	PG	0.0	Avery Bradley	2-Jun	Texas	25
1	235	Boston Celtics	6796117.0	SF	99.0	Jae Crowder	6-Jun	Marquette	25
2	205	Boston Celtics	NaN	SG	30.0	John Holland	5-Jun	Boston University	27
3	185	Boston Celtics	1148640.0	SG	28.0	R.J. Hunter	5-Jun	Georgia State	22
4	231	Boston Celtics	5000000.0	PF	8.0	Jonas Jerebko	10-Jun	NaN	29

Name: Abhishek Mali
Roll No.: 213070
PRN No.: 22320024

Checking The DataType

```
[21]: df.dtypes #Checking The Datatype
```

```
[21]: Name         object
      Team         object
      Number      float64
      Position     object
      Age           int64
      Height       object
      Weight        int64
      College      object
      Salary      float64
      dtype: object
```

Unique Values

```
[59]: df.nunique()#Unique Value
```

```
[59]: Name        457
```

Checking The DataType

[21] df.dtypes #Checking The Datatype

[21] Name object
Team object
Number float64
Position object
Age int64
Height object
Weight int64
College object
Salary float64
dtype: object

Unique Values

[59] df.nunique()#Unique Value

[59] Name 457
Team 30
Number 53
Position 5
Age 22
Height 18
Weight 87
College 118
Salary 308
dtype: int64

Name: Abhishek Mali
Roll No.: 213070
PRN No.: 22320024

Format Of The Column

```
[60]: df.info()#Format of the column

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 457 entries, 0 to 456
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      457 non-null    object
 1   Team      457 non-null    object
 2   Number    455 non-null    float64
 3   Position  457 non-null    object
 4   Age       457 non-null    int64
 5   Height    457 non-null    object
 6   Weight    457 non-null    int64
 7   College   373 non-null    object
 8   Salary    445 non-null    float64
dtypes: float64(2), int64(2), object(5)
memory usage: 32.3+ KB
```

```
[61]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 457 entries, 0 to 456
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      457 non-null    object
 1   Team      457 non-null    object
 2   Number    455 non-null    float64
 3   Position  457 non-null    object
 4   Age       457 non-null    int64
 5   Height    457 non-null    object
 6   Weight    457 non-null    int64
```

```
 7   College   373 non-null    object
 8   Salary    445 non-null    float64
dtypes: float64(2), int64(2), object(5)
memory usage: 32.3+ KB
```

Converting Variable DataType

```
[77]: var_change=df["Weight"].df["Weight"].astype("float")

[45]: var_change

[45]: 0      180.0
      1      235.0
      2      205.0
      3      185.0
      4      231.0
             ...
      452    234.0
      453    203.0
      454    179.0
      455    256.0
      456    231.0
      Name: Weight, Length: 457, dtype: float64
```

isNull Fun

#It show missing value

```
[68]: df.isnull().sum()

[68]: Name        0
      Team        0
      Number      2
      Position    0
      Age         0
      Height      0
      Weight      0
      College    84
      Salary     12
      dtype: int64

[ ]: ###
```