```python
import pandas as pd

a = pd.read_csv("/content/nces330_20.csv")

a.shape
# will return the dimension of the data set whether it is 2D, 3D,etc.
```

```
(3548, 6)
```

```python
a
```

```
      Year    State                Type  Length       Expense  Value
0     2013  Alabama             Private  4-year  Fees/Tuition  13983
1     2013  Alabama             Private  4-year    Room/Board   8503
2     2013  Alabama      Public In-State  2-year  Fees/Tuition   4048
3     2013  Alabama      Public In-State  4-year  Fees/Tuition   8073
4     2013  Alabama      Public In-State  4-year    Room/Board   8473
...    ...      ...                 ...     ...           ...    ...
3543  2021  Wyoming      Public In-State  2-year  Fees/Tuition   3987
3544  2021  Wyoming      Public In-State  4-year    Room/Board   9799
3545  2021  Wyoming  Public Out-of-State  2-year  Fees/Tuition   9820
3546  2021  Wyoming  Public Out-of-State  4-year  Fees/Tuition  14710
3547  2021  Wyoming  Public Out-of-State  4-year    Room/Board   9799

[3548 rows x 6 columns]
```

```python
a['Length'].unique()
# will return the unique values of the column 'Length'
```

```
array(['4-year', '2-year'], dtype=object)
```

```python
a.describe()
# will return Statistical Information of dataset
```

```
              Year         Value
count  3548.000000   3548.000000
mean   2016.923337  13027.720124
std       2.553910   8734.568645
min    2013.000000   1225.000000
25%    2015.000000   7756.750000
50%    2017.000000  10203.500000
75%    2019.000000  14830.750000
max    2021.000000  49152.000000
```

```python
a.info()
# will return information about no. of columns,column data types,etc
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3548 entries, 0 to 3547
Data columns (total 6 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
```

```
 0   Year     3548 non-null   int64
 1   State    3548 non-null   object
 2   Type     3548 non-null   object
 3   Length   3548 non-null   object
 4   Expense  3548 non-null   object
 5   Value    3548 non-null   int64
dtypes: int64(2), object(4)
memory usage: 166.4+ KB
```

```
a.columns # will return names of the columns
a.columns[0:2] # will return the name of the column form 0 to 2.
```

```
Index(['Year', 'State'], dtype='object')
```

```
a[['Year','State'] ]
```

```
      Year    State
0     2013  Alabama
1     2013  Alabama
2     2013  Alabama
3     2013  Alabama
4     2013  Alabama
...    ...      ...
3543  2021  Wyoming
3544  2021  Wyoming
3545  2021  Wyoming
3546  2021  Wyoming
3547  2021  Wyoming

[3548 rows x 2 columns]
```

```
list1 = ['Year','Value']
a[list1]
```

```
      Year  Value
0     2013  13983
1     2013   8503
2     2013   4048
3     2013   8073
4     2013   8473
...    ...    ...
3543  2021   3987
3544  2021   9799
3545  2021   9820
3546  2021  14710
3547  2021   9799

[3548 rows x 2 columns]
```

```
a.loc[0:2] # will return the row form 0 to 2 i.e 3 rows
```

```
   Year    State            Type  Length       Expense  Value
0  2013  Alabama         Private  4-year  Fees/Tuition  13983
1  2013  Alabama         Private  4-year    Room/Board   8503
2  2013  Alabama  Public In-State  2-year  Fees/Tuition   4048
```

```
a.iloc[0:2]
```

```
   Year    State     Type  Length       Expense  Value
0  2013  Alabama  Private  4-year  Fees/Tuition  13983
1  2013  Alabama  Private  4-year    Room/Board   8503
```

```
a.iloc[0:4,0:2]
# will display rows from 0 to 4  and columns from 0 to 2.
```

```
   Year    State
0  2013  Alabama
1  2013  Alabama
2  2013  Alabama
3  2013  Alabama
```

```
a.iloc[:,0:2]
# will display all rows and columns from 0 to 2
```

```
      Year    State
0     2013  Alabama
1     2013  Alabama
2     2013  Alabama
3     2013  Alabama
4     2013  Alabama
...    ...      ...
3543  2021  Wyoming
3544  2021  Wyoming
3545  2021  Wyoming
3546  2021  Wyoming
3547  2021  Wyoming

[3548 rows x 2 columns]
```

```
a.head(10)
# Display first 10 values
```

```
   Year    State               Type  Length       Expense  Value
0  2013  Alabama            Private  4-year  Fees/Tuition  13983
1  2013  Alabama            Private  4-year    Room/Board   8503
2  2013  Alabama     Public In-State  2-year  Fees/Tuition   4048
3  2013  Alabama     Public In-State  4-year  Fees/Tuition   8073
4  2013  Alabama     Public In-State  4-year    Room/Board   8473
5  2013  Alabama  Public Out-of-State  2-year  Fees/Tuition   7736
6  2013  Alabama  Public Out-of-State  4-year  Fees/Tuition  20380
7  2013  Alabama  Public Out-of-State  4-year    Room/Board   8473
8  2013   Alaska            Private  4-year  Fees/Tuition  21496
9  2013   Alaska            Private  4-year    Room/Board   8923
```

```
a.tail(10)
# Will retuen last 10 values
```

```
      Year      State                  Type  Length       Expense
Value
3538  2021  Wisconsin      Public In-State  2-year  Fees/Tuition
4534
3539  2021  Wisconsin      Public In-State  4-year    Room/Board
9093
3540  2021  Wisconsin  Public Out-of-State  2-year  Fees/Tuition
6552
3541  2021  Wisconsin  Public Out-of-State  4-year  Fees/Tuition
26970
3542  2021  Wisconsin  Public Out-of-State  4-year    Room/Board
9093
3543  2021    Wyoming      Public In-State  2-year  Fees/Tuition
3987
3544  2021    Wyoming      Public In-State  4-year    Room/Board
9799
3545  2021    Wyoming  Public Out-of-State  2-year  Fees/Tuition
9820
3546  2021    Wyoming  Public Out-of-State  4-year  Fees/Tuition
14710
3547  2021    Wyoming  Public Out-of-State  4-year    Room/Board
9799
```

```
a.head()
# will return first 5 values
```

```
   Year    State             Type  Length       Expense  Value
0  2013  Alabama          Private  4-year  Fees/Tuition  13983
1  2013  Alabama          Private  4-year    Room/Board   8503
2  2013  Alabama  Public In-State  2-year  Fees/Tuition   4048
3  2013  Alabama  Public In-State  4-year  Fees/Tuition   8073
4  2013  Alabama  Public In-State  4-year    Room/Board   8473
```

```
a2 = a.iloc[0:10,0:4]
a2
#Here, we have created new data frame and added first 10 entries of
data frame a
```

```
   Year    State                 Type  Length
0  2013  Alabama              Private  4-year
1  2013  Alabama              Private  4-year
2  2013  Alabama      Public In-State  2-year
3  2013  Alabama      Public In-State  4-year
4  2013  Alabama      Public In-State  4-year
5  2013  Alabama  Public Out-of-State  2-year
6  2013  Alabama  Public Out-of-State  4-year
7  2013  Alabama  Public Out-of-State  4-year
```

```
8  2013   Alaska              Private  4-year
9  2013   Alaska              Private  4-year
```

```
a2['Stud_Id']=5 # We add new column.
a2
```

```
   Year    State                Type  Length  Stud_Id
0  2013  Alabama              Private  4-year        5
1  2013  Alabama              Private  4-year        5
2  2013  Alabama       Public In-State  2-year        5
3  2013  Alabama       Public In-State  4-year        5
4  2013  Alabama       Public In-State  4-year        5
5  2013  Alabama  Public Out-of-State  2-year        5
6  2013  Alabama  Public Out-of-State  4-year        5
7  2013  Alabama  Public Out-of-State  4-year        5
8  2013   Alaska              Private  4-year        5
9  2013   Alaska              Private  4-year        5
```

```
a2['Stud_Id']=a['Year']
a2
# Here, we updated values of 'Stud_Id' by the values of column 'Year'
from data frame 'a'
```

```
   Year    State                Type  Length  Stud_Id
0  2013  Alabama              Private  4-year     2013
1  2013  Alabama              Private  4-year     2013
2  2013  Alabama       Public In-State  2-year     2013
3  2013  Alabama       Public In-State  4-year     2013
4  2013  Alabama       Public In-State  4-year     2013
5  2013  Alabama  Public Out-of-State  2-year     2013
6  2013  Alabama  Public Out-of-State  4-year     2013
7  2013  Alabama  Public Out-of-State  4-year     2013
8  2013   Alaska              Private  4-year     2013
9  2013   Alaska              Private  4-year     2013
```

```
a2.drop(['Stud_Id'],axis=1)
# Here, we drop the column 'Stud_Id'.
```

```
   Year    State                Type  Length
0  2013  Alabama              Private  4-year
1  2013  Alabama              Private  4-year
2  2013  Alabama       Public In-State  2-year
3  2013  Alabama       Public In-State  4-year
4  2013  Alabama       Public In-State  4-year
5  2013  Alabama  Public Out-of-State  2-year
6  2013  Alabama  Public Out-of-State  4-year
7  2013  Alabama  Public Out-of-State  4-year
8  2013   Alaska              Private  4-year
9  2013   Alaska              Private  4-year
```

```
a['New_Length'] = a['Length'].map({'4-year':4, '2-year':2}) # will
replace the value of '4-year' by 4 and values of '2-year' by 2 and add
```

```
it in new column 'New_Length
a
```

```
      Year    State                 Type  Length       Expense  Value
\
0     2013  Alabama              Private  4-year  Fees/Tuition  13983

1     2013  Alabama              Private  4-year    Room/Board   8503

2     2013  Alabama       Public In-State  2-year  Fees/Tuition   4048

3     2013  Alabama       Public In-State  4-year  Fees/Tuition   8073

4     2013  Alabama       Public In-State  4-year    Room/Board   8473

...    ...      ...                   ...     ...           ...    ...

3543  2021  Wyoming       Public In-State  2-year  Fees/Tuition   3987

3544  2021  Wyoming       Public In-State  4-year    Room/Board   9799

3545  2021  Wyoming  Public Out-of-State  2-year  Fees/Tuition   9820

3546  2021  Wyoming  Public Out-of-State  4-year  Fees/Tuition  14710

3547  2021  Wyoming  Public Out-of-State  4-year    Room/Board   9799


      New_Length
0              4
1              4
2              2
3              4
4              4
...          ...
3543           2
3544           4
3545           2
3546           4
3547           4

[3548 rows x 7 columns]
```

```
a['Year'].dtype # will return the type of data
```

```
dtype('int64')
```

```
a['Value']<10000 # will check the condition is  true or false
```

```
0       False
1        True
2        True
3        True
4        True
        ...
3543     True
3544     True
3545     True
3546    False
3547     True
Name: Value, Length: 3548, dtype: bool
```

```
a[a['Value']<10000] # will return all the record were it is true
```

```
      Year      State                 Type  Length       Expense
Value  \
1     2013    Alabama              Private  4-year    Room/Board
8503
2     2013    Alabama      Public In-State  2-year  Fees/Tuition
4048
3     2013    Alabama      Public In-State  4-year  Fees/Tuition
8073
4     2013    Alabama      Public In-State  4-year    Room/Board
8473
5     2013    Alabama  Public Out-of-State  2-year  Fees/Tuition
7736
...    ...        ...                  ...     ...           ...    ..
.
3542  2021  Wisconsin  Public Out-of-State  4-year    Room/Board
9093
3543  2021    Wyoming      Public In-State  2-year  Fees/Tuition
3987
3544  2021    Wyoming      Public In-State  4-year    Room/Board
9799
3545  2021    Wyoming  Public Out-of-State  2-year  Fees/Tuition
9820
3547  2021    Wyoming  Public Out-of-State  4-year    Room/Board
9799

      New_Length
1              4
2              2
3              4
4              4
5              2
...          ...
3542           4
3543           2
3544           4
```

```
3545           2
3547           4

[1695 rows x 7 columns]
```

```
a[(a['Value']<10000) | (a['Value']>5000)] # 'OR' condition
```

```
      Year    State                 Type  Length       Expense  Value
\
0     2013  Alabama              Private  4-year  Fees/Tuition  13983

1     2013  Alabama              Private  4-year    Room/Board   8503

2     2013  Alabama       Public In-State  2-year  Fees/Tuition   4048

3     2013  Alabama       Public In-State  4-year  Fees/Tuition   8073

4     2013  Alabama       Public In-State  4-year    Room/Board   8473

...    ...      ...                   ...     ...           ...    ...

3543  2021  Wyoming       Public In-State  2-year  Fees/Tuition   3987

3544  2021  Wyoming       Public In-State  4-year    Room/Board   9799

3545  2021  Wyoming  Public Out-of-State  2-year  Fees/Tuition   9820

3546  2021  Wyoming  Public Out-of-State  4-year  Fees/Tuition  14710

3547  2021  Wyoming  Public Out-of-State  4-year    Room/Board   9799


      New_Length
0              4
1              4
2              2
3              4
4              4
...          ...
3543           2
3544           4
3545           2
3546           4
3547           4

[3548 rows x 7 columns]
```

```
a[(a['Value']<10000) & (a['Value']>5000)] # 'AND' condition
```

```
      Year      State                 Type  Length       Expense
Value  \
1     2013    Alabama              Private  4-year    Room/Board
8503
3     2013    Alabama      Public In-State  4-year  Fees/Tuition
8073
4     2013    Alabama      Public In-State  4-year    Room/Board
8473
5     2013    Alabama  Public Out-of-State  2-year  Fees/Tuition
7736
7     2013    Alabama  Public Out-of-State  4-year    Room/Board
8473
...    ...        ...                  ...     ...           ...    ..
.
3540  2021  Wisconsin  Public Out-of-State  2-year  Fees/Tuition
6552
3542  2021  Wisconsin  Public Out-of-State  4-year    Room/Board
9093
3544  2021    Wyoming      Public In-State  4-year    Room/Board
9799
3545  2021    Wyoming  Public Out-of-State  2-year  Fees/Tuition
9820
3547  2021    Wyoming  Public Out-of-State  4-year    Room/Board
9799

      New_Length
1              4
3              4
4              4
5              2
7              4
...          ...
3540           2
3542           4
3544           4
3545           2
3547           4

[1255 rows x 7 columns]
```

```
def doublevalue(x):
  return(x*2)
a['New_value']= a['Value'].apply(doublevalue)
a[['Value', 'New_value']]
```

```
      Value  New_value
0     13983      27966
1      8503      17006
2      4048       8096
3      8073      16146
```

```
4      8473      16946
...     ...        ...
3543   3987       7974
3544   9799      19598
3545   9820      19640
3546  14710      29420
3547   9799      19598

[3548 rows x 2 columns]
```

```
# For inline function
# def doublevalue(x):
#   return(x*2)
a['New_value']= a['Value'].apply(lambda x:x*2)
a[['Value', 'New_value']]
```

```
      Value  New_value
0     13983      27966
1      8503      17006
2      4048       8096
3      8073      16146
4      8473      16946
...     ...        ...
3543   3987       7974
3544   9799      19598
3545   9820      19640
3546  14710      29420
3547   9799      19598

[3548 rows x 2 columns]
```

```
a.isnull() # to find null values
```

```
       Year  State   Type  Length  Expense  Value  New_Length
New_value
0     False  False  False   False    False  False       False
False
1     False  False  False   False    False  False       False
False
2     False  False  False   False    False  False       False
False
3     False  False  False   False    False  False       False
False
4     False  False  False   False    False  False       False
False
...     ...    ...    ...     ...      ...    ...         ...        .
..
3543  False  False  False   False    False  False       False
False
3544  False  False  False   False    False  False       False
False
```

```
3545  False  False  False   False    False  False       False
False
3546  False  False  False   False    False  False       False
False
3547  False  False  False   False    False  False       False
False

[3548 rows x 8 columns]
```

```
a.isnull().sum() # to count the null values
```

```
Year          0
State         0
Type          0
Length        0
Expense       0
Value         0
New_Length    0
New_value     0
dtype: int64
```

```
name = "12"
type(name)
```

```
str
```

```
# convert string into integer
name = int(name)
type(name)
```

```
int
```