In [1]:
```
import pandas as pd
data = [1,2,3,4,5]
```

In [2]:
```
series = pd.Series(data, index = ['a','b','c','d','e'])
print(series)
```

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

In [3]:
```
dic = {'Fruits':['apple', 'banana', 'mango'], 'count':['10', '20', '30']}
print(pd.DataFrame(dic))
```

```
   Fruits count
0   apple    10
1  banana    20
2   mango    30
```

In [4]:
```
print(pd.DataFrame({'Name':['Ram','Shyam'], 'Salary':['2222', '232323']}, index = ['a','b']))
```

```
    Name  Salary
a    Ram    2222
b  Shyam  232323
```

In [5]:
```
import numpy as np
num = np.array([['Ram','Shyam'],['30000', '35000']])
print(pd.DataFrame({'Name':num[0],'Salary':num[1]}))
```

```
    Name Salary
0    Ram  30000
1  Shyam  35000
```

Merge opreations

In [6]:
```
Player = ['Player1', 'Player2', 'Plaer3']
Points = [2,3,4]
Title = ['Game1', 'Game2', 'Game3']
df1 = pd.DataFrame({'Player':Player, 'Points':Points, 'Title':Title})
df1
```

Out[6]:

	Player	Points	Title
0	Player1	2	Game1
1	Player2	3	Game2
2	Plaer3	4	Game3

In [7]:
```
Player = ['Player1', 'Player5', 'Player6']
Power = ['Punch', 'Kick', 'Elbow']
Title = ['Game1', 'Game5', 'Game6']
df2 = pd.DataFrame({'Player':Player, 'Power':Power, 'Title': Title})
df2
```

Out[7]:

	Player	Power	Title
0	Player1	Punch	Game1
1	Player5	Kick	Game5
2	Player6	Elbow	Game6

In [8]:
```
df1
```

Out[8]:

	Player	Points	Title
0	Player1	2	Game1
1	Player2	3	Game2
2	Plaer3	4	Game3

In [9]:
```
df2
```

Out[9]:

	Player	Power	Title
0	Player1	Punch	Game1
1	Player5	Kick	Game5
2	Player6	Elbow	Game6

In [10]:
```
#Inner merge
df1.merge(df2, on='Title', how='inner')
```

Out[10]:

	Player_x	Points	Title	Player_y	Power
0	Player1	2	Game1	Player1	Punch

In [11]:
```
#Right merge
df1.merge(df2, on='Player', how="right")
```

Out[11]:

	Player	Points	Title_x	Power	Title_y
0	Player1	2.0	Game1	Punch	Game1
1	Player5	NaN	NaN	Kick	Game5
2	Player6	NaN	NaN	Elbow	Game6

In [12]:
```
#Left merge
df1.merge(df2, on='Player',how='left')
```

Out[12]:

	Player	Points	Title_x	Power	Title_y
0	Player1	2	Game1	Punch	Game1
1	Player2	3	Game2	NaN	NaN
2	Plaer3	4	Game3	NaN	NaN

In [13]:
```
#Outer merge
df1.merge(df2, on='Player', how='outer')
```

Out[13]:

	Player	Points	Title_x	Power	Title_y
0	Player1	2.0	Game1	Punch	Game1
1	Player2	3.0	Game2	NaN	NaN
2	Plaer3	4.0	Game3	NaN	NaN
3	Player5	NaN	NaN	Kick	Game5
4	Player6	NaN	NaN	Elbow	Game6

In case of Join the two dataframes are join in bases of Index name and In case of Merge the two dataframes are Merge in bases of attributes name

Join Opreations

In [14]:
```
Player1 = ['Player1', 'Player2', 'Plaer3']
Points1 = [2,3,4]
Title1 = ['Game1', 'Game2', 'Game3']
df3 = pd.DataFrame({'Player':Player, 'Points':Points1, 'Title':Title}, index = ['L1', 'L2', 'L3'])
df3
```

Out[14]:

	Player	Points	Title
L1	Player1	2	Game1
L2	Player5	3	Game5
L3	Player6	4	Game6

In [15]:
```
Player1 = ['Player1', 'Player5', 'Player6']
Power1 = ['Punch', 'Kick', 'Elbow']
Title1 = ['Game1', 'Game5', 'Game6']
df4 = pd.DataFrame({'Player1':Player, 'Power1':Power, 'Title1': Title}, index = ['L2', 'L3', 'L4'])
df4
```

Out[15]:

	Player1	Power1	Title1
L2	Player1	Punch	Game1
L3	Player5	Kick	Game5
L4	Player6	Elbow	Game6

In [16]:
```
#Inner Join
df3.join(df4, how='inner')
```

Out[16]:

	Player	Points	Title	Player1	Power1	Title1
L2	Player5	3	Game5	Player1	Punch	Game1
L3	Player6	4	Game6	Player5	Kick	Game5

In [17]:
```
df3.join(df4)
```

Out[17]:

	Player	Points	Title	Player1	Power1	Title1
L1	Player1	2	Game1	NaN	NaN	NaN
L2	Player5	3	Game5	Player1	Punch	Game1
L3	Player6	4	Game6	Player5	Kick	Game5

In [18]:
```
df3.join(df4, how='outer')
```

Out[18]:

	Player	Points	Title	Player1	Power1	Title1
L1	Player1	2.0	Game1	NaN	NaN	NaN
L2	Player5	3.0	Game5	Player1	Punch	Game1
L3	Player6	4.0	Game6	Player5	Kick	Game5
L4	NaN	NaN	NaN	Player6	Elbow	Game6

In [19]:
```
df3.join(df4, how='right')
```

Out[19]:

	Player	Points	Title	Player1	Power1	Title1
L2	Player5	3.0	Game5	Player1	Punch	Game1
L3	Player6	4.0	Game6	Player5	Kick	Game5
L4	NaN	NaN	NaN	Player6	Elbow	Game6

In [20]:
```
df3.join(df4,how='left')
```

Out[20]:

	Player	Points	Title	Player1	Power1	Title1
L1	Player1	2	Game1	NaN	NaN	NaN
L2	Player5	3	Game5	Player1	Punch	Game1
L3	Player6	4	Game6	Player5	Kick	Game5

Concatinate Opreations

In [21]:
```
pd.concat([df3,df4])
```

Out[21]:

	Player	Points	Title	Player1	Power1	Title1
L1	Player1	2.0	Game1	NaN	NaN	NaN
L2	Player5	3.0	Game5	NaN	NaN	NaN
L3	Player6	4.0	Game6	NaN	NaN	NaN
L2	NaN	NaN	NaN	Player1	Punch	Game1
L3	NaN	NaN	NaN	Player5	Kick	Game5
L4	NaN	NaN	NaN	Player6	Elbow	Game6

In [22]:
```
pd.concat([df1,df2])
```

Out[22]:

	Player	Points	Title	Power
0	Player1	2.0	Game1	NaN
1	Player2	3.0	Game2	NaN
2	Plaer3	4.0	Game3	NaN
0	Player1	NaN	Game1	Punch
1	Player5	NaN	Game5	Kick
2	Player6	NaN	Game6	Elbow

In [23]:
```
pd.concat([df1,df2,df3,df4])
```

Out[23]:

	Player	Points	Title	Power	Player1	Power1	Title1
0	Player1	2.0	Game1	NaN	NaN	NaN	NaN
1	Player2	3.0	Game2	NaN	NaN	NaN	NaN
2	Plaer3	4.0	Game3	NaN	NaN	NaN	NaN
0	Player1	NaN	Game1	Punch	NaN	NaN	NaN
1	Player5	NaN	Game5	Kick	NaN	NaN	NaN
2	Player6	NaN	Game6	Elbow	NaN	NaN	NaN
L1	Player1	2.0	Game1	NaN	NaN	NaN	NaN
L2	Player5	3.0	Game5	NaN	NaN	NaN	NaN
L3	Player6	4.0	Game6	NaN	NaN	NaN	NaN
L2	NaN	NaN	NaN	NaN	Player1	Punch	Game1
L3	NaN	NaN	NaN	NaN	Player5	Kick	Game5
L4	NaN	NaN	NaN	NaN	Player6	Elbow	Game6

Import the dataframe

In [24]:
```
train = pd.read_csv("C:\\Users\\Utkarsh Prajapati\\Downloads\\forest-cover-type-kernels-only\\train.csv")
train.head(10)
```

Out[24]:

	Id	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am	H
0	1	2596	51	3	258	0	510	221	
1	2	2590	56	2	212	-6	390	220	
2	3	2804	139	9	268	65	3180	234	
3	4	2785	155	18	242	118	3090	238	
4	5	2595	45	2	153	-1	391	220	
5	6	2579	132	6	300	-15	67	230	
6	7	2606	45	7	270	5	633	222	
7	8	2605	49	4	234	7	573	222	
8	9	2617	45	9	240	56	666	223	
9	10	2612	59	10	247	11	636	228	

10 rows × 56 columns

In [25]:
```
train.size
```

Out[25]:
```
846720
```

In [26]:
```
train.shape
```

Out[26]:
```
(15120, 56)
```

In [27]:
```
type(train)
```

Out[27]:
```
pandas.core.frame.DataFrame
```

In [28]:
```
train.tail(10)
```

Out[28]:

	Id	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade
15110	15111	2508	33	26	67	1	644	
15111	15112	2610	59	17	60	10	674	
15112	15113	2600	38	25	124	0	589	
15113	15114	2688	104	15	443	10	805	
15114	15115	2670	108	12	624	24	730	
15115	15116	2607	243	23	258	7	660	
15116	15117	2603	121	19	633	195	618	
15117	15118	2492	134	25	365	117	335	
15118	15119	2487	167	28	218	101	242	
15119	15120	2475	197	34	319	78	270	

10 rows × 56 columns

In [29]:
```
#Summary of the data
train.info(null_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15120 entries, 0 to 15119
Data columns (total 56 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Id                                  15120 non-null  int64
 1   Elevation                           15120 non-null  int64
 2   Aspect                              15120 non-null  int64
 3   Slope                               15120 non-null  int64
 4   Horizontal_Distance_To_Hydrology    15120 non-null  int64
 5   Vertical_Distance_To_Hydrology      15120 non-null  int64
 6   Horizontal_Distance_To_Roadways     15120 non-null  int64
 7   Hillshade_9am                       15120 non-null  int64
 8   Hillshade_Noon                      15120 non-null  int64
 9   Hillshade_3pm                       15120 non-null  int64
 10  Horizontal_Distance_To_Fire_Points  15120 non-null  int64
 11  Wilderness_Area1                    15120 non-null  int64
 12  Wilderness_Area2                    15120 non-null  int64
 13  Wilderness_Area3                    15120 non-null  int64
 14  Wilderness_Area4                    15120 non-null  int64
 15  Soil_Type1                          15120 non-null  int64
 16  Soil_Type2                          15120 non-null  int64
 17  Soil_Type3                          15120 non-null  int64
 18  Soil_Type4                          15120 non-null  int64
 19  Soil_Type5                          15120 non-null  int64
 20  Soil_Type6                          15120 non-null  int64
 21  Soil_Type7                          15120 non-null  int64
 22  Soil_Type8                          15120 non-null  int64
 23  Soil_Type9                          15120 non-null  int64
 24  Soil_Type10                         15120 non-null  int64
 25  Soil_Type11                         15120 non-null  int64
 26  Soil_Type12                         15120 non-null  int64
 27  Soil_Type13                         15120 non-null  int64
 28  Soil_Type14                         15120 non-null  int64
 29  Soil_Type15                         15120 non-null  int64
 30  Soil_Type16                         15120 non-null  int64
 31  Soil_Type17                         15120 non-null  int64
 32  Soil_Type18                         15120 non-null  int64
 33  Soil_Type19                         15120 non-null  int64
 34  Soil_Type20                         15120 non-null  int64
 35  Soil_Type21                         15120 non-null  int64
 36  Soil_Type22                         15120 non-null  int64
 37  Soil_Type23                         15120 non-null  int64
 38  Soil_Type24                         15120 non-null  int64
 39  Soil_Type25                         15120 non-null  int64
 40  Soil_Type26                         15120 non-null  int64
 41  Soil_Type27                         15120 non-null  int64
 42  Soil_Type28                         15120 non-null  int64
 43  Soil_Type29                         15120 non-null  int64
 44  Soil_Type30                         15120 non-null  int64
 45  Soil_Type31                         15120 non-null  int64
 46  Soil_Type32                         15120 non-null  int64
 47  Soil_Type33                         15120 non-null  int64
 48  Soil_Type34                         15120 non-null  int64
 49  Soil_Type35                         15120 non-null  int64
 50  Soil_Type36                         15120 non-null  int64
 51  Soil_Type37                         15120 non-null  int64
 52  Soil_Type38                         15120 non-null  int64
 53  Soil_Type39                         15120 non-null  int64
 54  Soil_Type40                         15120 non-null  int64
 55  Cover_Type                          15120 non-null  int64
dtypes: int64(56)
memory usage: 6.5 MB
```

In [30]:
```
train.mean()
```

Out[30]:
```
Id                                  7560.500000
Elevation                           2749.322553
Aspect                               156.676653
Slope                                 16.501587
Horizontal_Distance_To_Hydrology      227.195701
Vertical_Distance_To_Hydrology         51.076521
Horizontal_Distance_To_Roadways      1714.023214
Hillshade_9am                         212.704299
Hillshade_Noon                        218.965608
Hillshade_3pm                         135.091997
Horizontal_Distance_To_Fire_Points   1511.147288
Wilderness_Area1                        0.237897
Wilderness_Area2                        0.033003
Wilderness_Area3                        0.419907
Wilderness_Area4                        0.309193
Soil_Type1                              0.023479
Soil_Type2                              0.041204
Soil_Type3                              0.063624
Soil_Type4                              0.055754
Soil_Type5                              0.010913
Soil_Type6                              0.042989
Soil_Type7                              0.000000
Soil_Type8                              0.000066
Soil_Type9                              0.000661
Soil_Type10                             0.141667
Soil_Type11                             0.026852
Soil_Type12                             0.015013
Soil_Type13                             0.031481
Soil_Type14                             0.011177
Soil_Type15                             0.000000
Soil_Type16                             0.007540
Soil_Type17                             0.040476
Soil_Type18                             0.003968
Soil_Type19                             0.003042
Soil_Type20                             0.009193
Soil_Type21                             0.001058
Soil_Type22                             0.022817
Soil_Type23                             0.050066
Soil_Type24                             0.016997
Soil_Type25                             0.000066
Soil_Type26                             0.003571
Soil_Type27                             0.000992
Soil_Type28                             0.000595
Soil_Type29                             0.085384
Soil_Type30                             0.047950
Soil_Type31                             0.021958
Soil_Type32                             0.045635
Soil_Type33                             0.040741
Soil_Type34                             0.001455
Soil_Type35                             0.006746
Soil_Type36                             0.000661
Soil_Type37                             0.002249
Soil_Type38                             0.048148
Soil_Type39                             0.043452
Soil_Type40                             0.030357
Cover_Type                              4.000000
dtype: float64
```

In [31]:
```
train.median()
```

Out[31]:
```
Id                                  7560.5
Elevation                           2752.0
Aspect                               126.0
Slope                                 15.0
Horizontal_Distance_To_Hydrology     180.0
Vertical_Distance_To_Hydrology        32.0
Horizontal_Distance_To_Roadways     1316.0
Hillshade_9am                        220.0
Hillshade_Noon                       223.0
Hillshade_3pm                        138.0
Horizontal_Distance_To_Fire_Points  1256.0
Wilderness_Area1                       0.0
Wilderness_Area2                       0.0
Wilderness_Area3                       0.0
Wilderness_Area4                       0.0
Soil_Type1                             0.0
Soil_Type2                             0.0
Soil_Type3                             0.0
Soil_Type4                             0.0
Soil_Type5                             0.0
Soil_Type6                             0.0
Soil_Type7                             0.0
Soil_Type8                             0.0
Soil_Type9                             0.0
Soil_Type10                            0.0
Soil_Type11                            0.0
Soil_Type12                            0.0
Soil_Type13                            0.0
Soil_Type14                            0.0
Soil_Type15                            0.0
Soil_Type16                            0.0
Soil_Type17                            0.0
Soil_Type18                            0.0
Soil_Type19                            0.0
Soil_Type20                            0.0
Soil_Type21                            0.0
Soil_Type22                            0.0
Soil_Type23                            0.0
Soil_Type24                            0.0
Soil_Type25                            0.0
Soil_Type26                            0.0
Soil_Type27                            0.0
Soil_Type28                            0.0
Soil_Type29                            0.0
Soil_Type30                            0.0
Soil_Type31                            0.0
Soil_Type32                            0.0
Soil_Type33                            0.0
Soil_Type34                            0.0
Soil_Type35                            0.0
Soil_Type36                            0.0
Soil_Type37                            0.0
Soil_Type38                            0.0
Soil_Type39                            0.0
Soil_Type40                            0.0
Cover_Type                             4.0
dtype: float64
```

In [32]:
```
train.std()
```

Out[32]:
```
Id                                  4364.912370
Elevation                            417.678187
Aspect                               110.085801
Slope                                  8.453927
Horizontal_Distance_To_Hydrology     210.075296
Vertical_Distance_To_Hydrology        61.239406
Horizontal_Distance_To_Roadways     1325.066358
Hillshade_9am                         30.561287
Hillshade_Noon                        22.801966
Hillshade_3pm                         45.895189
Horizontal_Distance_To_Fire_Points  1099.936493
Wilderness_Area1                       0.425810
Wilderness_Area2                       0.178649
Wilderness_Area3                       0.493560
Wilderness_Area4                       0.462176
Soil_Type1                             0.151424
Soil_Type2                             0.198768
Soil_Type3                             0.244091
Soil_Type4                             0.229454
Soil_Type5                             0.103896
Soil_Type6                             0.202840
Soil_Type7                             0.000000
Soil_Type8                             0.008133
Soil_Type9                             0.025710
Soil_Type10                            0.348719
Soil_Type11                            0.161656
Soil_Type12                            0.121609
Soil_Type13                            0.174621
Soil_Type14                            0.105133
Soil_Type15                            0.000000
Soil_Type16                            0.086506
Soil_Type17                            0.197080
Soil_Type18                            0.062871
Soil_Type19                            0.055075
Soil_Type20                            0.095442
Soil_Type21                            0.032514
Soil_Type22                            0.149326
Soil_Type23                            0.218089
Soil_Type24                            0.129265
Soil_Type25                            0.008133
Soil_Type26                            0.059657
Soil_Type27                            0.031482
Soil_Type28                            0.024391
Soil_Type29                            0.279461
Soil_Type30                            0.213667
Soil_Type31                            0.146550
Soil_Type32                            0.208699
Soil_Type33                            0.197696
Soil_Type34                            0.038118
Soil_Type35                            0.081859
Soil_Type36                            0.025710
Soil_Type37                            0.047368
Soil_Type38                            0.214086
Soil_Type39                            0.203880
Soil_Type40                            0.171574
Cover_Type                             2.000066
dtype: float64
```

In [33]:
```
train.max()
```

Out[33]:
```
Id                                  15120
Elevation                            3849
Aspect                                360
Slope                                  52
Horizontal_Distance_To_Hydrology     1343
Vertical_Distance_To_Hydrology        554
Horizontal_Distance_To_Roadways      6890
Hillshade_9am                         254
Hillshade_Noon                        254
Hillshade_3pm                         248
Horizontal_Distance_To_Fire_Points   6993
Wilderness_Area1                        1
Wilderness_Area2                        1
Wilderness_Area3                        1
Wilderness_Area4                        1
Soil_Type1                              1
Soil_Type2                              1
Soil_Type3                              1
Soil_Type4                              1
Soil_Type5                              1
Soil_Type6                              1
Soil_Type7                              0
Soil_Type8                              1
Soil_Type9                              1
Soil_Type10                             1
Soil_Type11                             1
Soil_Type12                             1
Soil_Type13                             1
Soil_Type14                             1
Soil_Type15                             0
Soil_Type16                             1
Soil_Type17                             1
Soil_Type18                             1
Soil_Type19                             1
Soil_Type20                             1
Soil_Type21                             1
Soil_Type22                             1
Soil_Type23                             1
Soil_Type24                             1
Soil_Type25                             1
Soil_Type26                             1
Soil_Type27                             1
Soil_Type28                             1
Soil_Type29                             1
Soil_Type30                             1
Soil_Type31                             1
Soil_Type32                             1
Soil_Type33                             1
Soil_Type34                             1
Soil_Type35                             1
Soil_Type36                             1
Soil_Type37                             1
Soil_Type38                             1
Soil_Type39                             1
Soil_Type40                             1
Cover_Type                              7
dtype: int64
```

In [34]:
```
train.min()
```

Out[34]:
```
Id                                   1
Elevation                         1863
Aspect                               0
Slope                                0
Horizontal_Distance_To_Hydrology     0
Vertical_Distance_To_Hydrology    -146
Horizontal_Distance_To_Roadways      0
Hillshade_9am                        0
Hillshade_Noon                      99
Hillshade_3pm                        0
Horizontal_Distance_To_Fire_Points   0
Wilderness_Area1                     0
Wilderness_Area2                     0
Wilderness_Area3                     0
Wilderness_Area4                     0
Soil_Type1                           0
Soil_Type2                           0
Soil_Type3                           0
Soil_Type4                           0
Soil_Type5                           0
Soil_Type6                           0
Soil_Type7                           0
Soil_Type8                           0
Soil_Type9                           0
Soil_Type10                          0
Soil_Type11                          0
Soil_Type12                          0
Soil_Type13                          0
Soil_Type14                          0
Soil_Type15                          0
Soil_Type16                          0
Soil_Type17                          0
Soil_Type18                          0
Soil_Type19                          0
Soil_Type20                          0
Soil_Type21                          0
Soil_Type22                          0
Soil_Type23                          0
Soil_Type24                          0
Soil_Type25                          0
Soil_Type26                          0
Soil_Type27                          0
Soil_Type28                          0
Soil_Type29                          0
Soil_Type30                          0
Soil_Type31                          0
Soil_Type32                          0
Soil_Type33                          0
Soil_Type34                          0
Soil_Type35                          0
Soil_Type36                          0
Soil_Type37                          0
Soil_Type38                          0
Soil_Type39                          0
Soil_Type40                          0
Cover_Type                           1
dtype: int64
```

In [35]:
```
train.count()
```

Out[35]:
```
Id                                  15120
Elevation                           15120
Aspect                              15120
Slope                               15120
Horizontal_Distance_To_Hydrology    15120
Vertical_Distance_To_Hydrology      15120
Horizontal_Distance_To_Roadways     15120
Hillshade_9am                       15120
Hillshade_Noon                      15120
Hillshade_3pm                       15120
Horizontal_Distance_To_Fire_Points  15120
Wilderness_Area1                    15120
Wilderness_Area2                    15120
Wilderness_Area3                    15120
Wilderness_Area4                    15120
Soil_Type1                          15120
Soil_Type2                          15120
Soil_Type3                          15120
Soil_Type4                          15120
Soil_Type5                          15120
Soil_Type6                          15120
Soil_Type7                          15120
Soil_Type8                          15120
Soil_Type9                          15120
Soil_Type10                         15120
Soil_Type11                         15120
Soil_Type12                         15120
Soil_Type13                         15120
Soil_Type14                         15120
Soil_Type15                         15120
Soil_Type16                         15120
Soil_Type17                         15120
Soil_Type18                         15120
Soil_Type19                         15120
Soil_Type20                         15120
Soil_Type21                         15120
Soil_Type22                         15120
Soil_Type23                         15120
Soil_Type24                         15120
Soil_Type25                         15120
Soil_Type26                         15120
Soil_Type27                         15120
Soil_Type28                         15120
Soil_Type29                         15120
Soil_Type30                         15120
Soil_Type31                         15120
Soil_Type32                         15120
Soil_Type33                         15120
Soil_Type34                         15120
Soil_Type35                         15120
Soil_Type36                         15120
Soil_Type37                         15120
Soil_Type38                         15120
Soil_Type39                         15120
Soil_Type40                         15120
Cover_Type                          15120
dtype: int64
```

In [36]:
```
train.describe()
```

Out[36]:

	Id	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_T
count	15120.00000	15120.000000	15120.000000	15120.000000	15120.000000	15120.000000	
mean	7560.50000	2749.322553	156.676653	16.501587	227.195701	51.076521	
std	4364.91237	417.678187	110.085801	8.453927	210.075296	61.239406	
min	1.00000	1863.000000	0.000000	0.000000	0.000000	-146.000000	
25%	3780.75000	2376.000000	65.000000	10.000000	67.000000	5.000000	
50%	7560.50000	2752.000000	126.000000	15.000000	180.000000	32.000000	
75%	11340.25000	3104.000000	261.000000	22.000000	330.000000	79.000000	
max	15120.00000	3849.000000	360.000000	52.000000	1343.000000	554.000000	

8 rows × 56 columns

Cleaning the dataset

In [37]:
```
#Fill the missing value
#train.aspect = train.aspect.fillna(train.aspect.mean())
```

In [38]:
```
#Drop unwanted coloum
train = train.drop(columns=['Id'])
train.head()
```

Out[38]:

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am	Hillsl
0	2596	51	3	258	0	510	221	
1	2590	56	2	212	-6	390	220	
2	2804	139	9	268	65	3180	234	
3	2785	155	18	242	118	3090	238	
4	2595	45	2	153	-1	391	220	

5 rows × 55 columns

In [41]:
```
train.columns
```

Out[41]:
```
Index(['Elevation', 'Aspect', 'Slope', 'Horizontal_Distance_To_Hydrology',
       'Vertical_Distance_To_Hydrology', 'Horizontal_Distance_To_Roadways',
       'Hillshade_9am', 'Hillshade_Noon', 'Hillshade_3pm',
       'Horizontal_Distance_To_Fire_Points', 'Wilderness_Area1',
       'Wilderness_Area2', 'Wilderness_Area3', 'Wilderness_Area4',
       'Soil_Type1', 'Soil_Type2', 'Soil_Type3', 'Soil_Type4', 'Soil_Type5',
       'Soil_Type6', 'Soil_Type7', 'Soil_Type8', 'Soil_Type9', 'Soil_Type10',
       'Soil_Type11', 'Soil_Type12', 'Soil_Type13', 'Soil_Type14',
       'Soil_Type15', 'Soil_Type16', 'Soil_Type17', 'Soil_Type18',
       'Soil_Type19', 'Soil_Type20', 'Soil_Type21', 'Soil_Type22',
       'Soil_Type23', 'Soil_Type24', 'Soil_Type25', 'Soil_Type26',
       'Soil_Type27', 'Soil_Type28', 'Soil_Type29', 'Soil_Type30',
       'Soil_Type31', 'Soil_Type32', 'Soil_Type33', 'Soil_Type34',
       'Soil_Type35', 'Soil_Type36', 'Soil_Type37', 'Soil_Type38',
       'Soil_Type39', 'Soil_Type40', 'Cover_Type'],
      dtype='object')
```

In [43]:
```
#Find correlation matric
corr = train[['Elevation', 'Aspect', 'Slope', 'Horizontal_Distance_To_Hydrology',
       'Vertical_Distance_To_Hydrology', 'Horizontal_Distance_To_Roadways',
       'Hillshade_9am', 'Hillshade_Noon', 'Hillshade_3pm',
       'Horizontal_Distance_To_Fire_Points', 'Wilderness_Area1',
       'Wilderness_Area2', 'Wilderness_Area3', 'Wilderness_Area4']].corr()
corr
```

Out[43]:

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distan
Elevation	1.000000	-0.011096	-0.312640	0.412712	0.122092	
Aspect	-0.011096	1.000000	0.028148	0.040732	0.056412	
Slope	-0.312640	0.028148	1.000000	-0.055976	0.265314	
Horizontal_Distance_To_Hydrology	0.412712	0.040732	-0.055976	1.000000	0.652142	
Vertical_Distance_To_Hydrology	0.122092	0.056412	0.265314	0.652142	1.000000	
Horizontal_Distance_To_Roadways	0.578659	0.066184	-0.277049	0.203397	0.011555	
Hillshade_9am	0.097900	-0.593997	-0.200072	-0.033803	-0.095930	
Hillshade_Noon	0.215782	0.324912	-0.612613	0.080047	-0.132948	
Hillshade_3pm	0.089518	0.635022	-0.326887	0.080833	-0.035559	
Horizontal_Distance_To_Fire_Points	0.443563	-0.052169	-0.239527	0.158817	-0.015048	
Wilderness_Area1	0.330417	-0.131262	-0.152820	-0.009402	-0.117835	
Wilderness_Area2	0.261729	0.028238	-0.065923	0.087484	0.017108	
Wilderness_Area3	0.354025	0.032578	-0.113033	0.200532	0.069884	
Wilderness_Area4	-0.783651	0.075228	0.286985	-0.239303	0.027321	

In [50]:
```
#Changing the type of data
train.Hillshade_3pm = train.Hillshade_3pm.astype(float)
train.Hillshade_3pm.dtype
```

Out[50]:
```
dtype('float64')
```

In [58]:
```
#View particular column
#iloc <- Index Location
train.iloc[0:5,4]
```

Out[58]:
```
0      0
1     -6
2     65
3    118
4     -1
Name: Vertical_Distance_To_Hydrology, dtype: int64
```

In [59]:
```
train.iloc[:,:]
```

Out[59]:

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am	I
0	2596	51	3	258	0	510	221	
1	2590	56	2	212	-6	390	220	
2	2804	139	9	268	65	3180	234	
3	2785	155	18	242	118	3090	238	
4	2595	45	2	153	-1	391	220	
...	...	...	...	...	...	...	...	
15115	2607	243	23	258	7	660	170	
15116	2603	121	19	633	195	618	249	
15117	2492	134	25	365	117	335	250	
15118	2487	167	28	218	101	242	229	
15119	2475	197	34	319	78	270	189	

15120 rows × 55 columns

In [61]:
```
#Row starting from number 6 and column startinf from 4
train.iloc[6:,4:]
```

Out[61]:

	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am	Hillshade_Noon	Hillshade_3pm	Horizontal_Distance_To_Fire_P
6	5	633	222	225	138.0	
7	7	573	222	230	144.0	
8	56	666	223	221	133.0	
9	11	636	228	219	124.0	
10	51	735	218	243	161.0	
...	...	...	...	...	...	
15115	7	660	170	251	214.0	
15116	195	618	249	221	91.0	
15117	117	335	250	220	83.0	
15118	101	242	229	237	119.0	
15119	78	270	189	244	164.0	

15114 rows × 51 columns

In [70]:
```
#Double up the data
f = lambda x: x*2
train['Vertical_Distance_To_Hydrology'] = train['Vertical_Distance_To_Hydrology'].apply(f)
```

In [77]:
```
train.loc[:,"Vertical_Distance_To_Hydrology"]
```

Out[77]:
```
0          0
1        -12
2        130
3        236
4         -2
        ...
15115     14
15116    390
15117    234
15118    202
15119    156
Name: Vertical_Distance_To_Hydrology, Length: 15120, dtype: int64
```

Split the dataset into train and testing

In [78]:
```
train.shape
```

Out[78]:
```
(15120, 55)
```

In [80]:
```
train.tail()
```

Out[80]:

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am	H
15115	2607	243	23	258	14	660	170	
15116	2603	121	19	633	390	618	249	
15117	2492	134	25	365	234	335	250	
15118	2487	167	28	218	202	242	229	
15119	2475	197	34	319	156	270	189	

5 rows × 55 columns

In [88]:
```
x = train.iloc[:,:54]
```

In [90]:
```
y = train.iloc[:,54]
```

In [92]:
```
print(x.head())
print(y.head())
```

```
   Elevation  Aspect  Slope  Horizontal_Distance_To_Hydrology  \
0       2596      51      3                               258
1       2590      56      2                               212
2       2804     139      9                               268
3       2785     155     18                               242
4       2595      45      2                               153

   Vertical_Distance_To_Hydrology  Horizontal_Distance_To_Roadways  \
0                               0                              510
1                             -12                              390
2                             130                             3180
3                             236                             3090
4                              -2                              391

   Hillshade_9am  Hillshade_Noon  Hillshade_3pm  \
0            221             232          148.0
1            220             235          151.0
2            234             238          135.0
3            238             238          122.0
4            220             234          150.0

   Horizontal_Distance_To_Fire_Points  ...  Soil_Type31  Soil_Type32  \
0                                6279  ...            0            0
1                                6225  ...            0            0
2                                6121  ...            0            0
3                                6211  ...            0            0
4                                6172  ...            0            0

   Soil_Type33  Soil_Type34  Soil_Type35  Soil_Type36  Soil_Type37  \
0            0            0            0            0            0
1            0            0            0            0            0
2            0            0            0            0            0
3            0            0            0            0            0
4            0            0            0            0            0

   Soil_Type38  Soil_Type39  Soil_Type40
0            0            0            0
1            0            0            0
2            0            0            0
3            0            0            0
4            0            0            0

[5 rows x 54 columns]
0    5
1    5
2    2
3    2
4    5
Name: Cover_Type, dtype: int64
```

In [98]:
```
print(x.shape)
print(y.shape)
```

```
(15120, 54)
(15120,)
```

In [99]:
```
from sklearn.model_selection import train_test_split
```

In [113]:
```
x_train, x_test, y_train, y_test = train_test_split(x,y, train_size = 0.70, random_state = 10)
```

In [114]:
```
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(10584, 54)
(4536, 54)
(10584,)
(4536,)
```

In [115]:
```
x_train.head(2)
```

Out[115]:

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am
10719	2577	21	4	0	0	819	217
11445	3024	357	5	663	104	5646	213

2 rows × 54 columns

In [116]:
```
x_test.head(2)
```

Out[116]:

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am
11813	2790	62	20	228	194	234	233
940	2749	169	4	30	-4	1895	222

2 rows × 54 columns

In [117]:
```
y_train.head(2)
```

Out[117]:
```
10719    6
11445    2
Name: Cover_Type, dtype: int64
```

In [118]:
```
y_test.head(2)
```

Out[118]:
```
11813    5
940      2
Name: Cover_Type, dtype: int64
```

In [119]:
```
from sklearn.linear_model import LinearRegression
clf = LinearRegression()
```

In [120]:
```
clf.fit(x_train, y_train)
```

Out[120]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [121]:
```
clf.predict(x_test)
```

Out[121]:
```
array([3.15192589, 1.95319384, 6.29086086, ..., 2.00057448, 2.52889644,
       3.71249324])
```

In [123]:
```
y_test
```

Out[123]:
```
11813    5
940      2
10527    7
450      2
13227    4
        ..
14447    7
9514     7
258      2
14317    2
850      1
Name: Cover_Type, Length: 4536, dtype: int64
```

In [125]:
```
clf.score(x_test, y_test)
```

Out[125]:
```
0.40340958385732684
```

Accuracy is 40.3%