

## Data Science and Machine Learning Internship (HDL Technologies)

**Name:** *B. Deva SriRama Sai Ganesh*

**Email:** [bandarusaiganesh2003@gmail.com](mailto:bandarusaiganesh2003@gmail.com)

### WEEK - 1 Assignment

#### Solutions:

1 Ans:

```
In [9]: import pandas as pd

d = {'col1': [10, 20, 30],
      'col2': [40, 50, 60],
      'col3': [70, 80, 90],
      'col4': [100, 110, 120]}

df = pd.DataFrame(data = d)
print(df)
print("\n Table created using DataFrame.")

count_row = df.shape[0]
print("\n Displaying the number of rows: ", count_row)

count_col = df.shape[1]
print("\n Displaying the number of columns: ", count_col)
```

	col1	col2	col3	col4
0	10	40	70	100
1	20	50	80	110
2	30	60	90	120

Table created using DataFrame.

Displaying the number of rows: 3

Displaying the number of columns: 4

2 Ans:

```
In [20]: import pandas as obj
House_data = obj.read_csv('C:\\Users\\banda\\Downloads\\data.csv')

print("\n a) Printing the dataset (House_data): \n")
print(House_data)

print("\n b) Printing only the first 5 rows in the data using head(): \n")
print(House_data.head())

print("\n c) Removing the NULL values in the data: \n")
House_data.dropna(axis = 0, inplace = True)
print(House_data)

print("\n d) Analyzing the data using describe() \n")
print(House_data.describe())
```

a) Printing the dataset (House\_data):

	Price	Area	Location	No. of Bedrooms	Gymnasium	\
0	6968000	1340	Nizampet	2.0	1.0	
1	29000000	3498	Hitech City	4.0	1.0	
2	6590000	1318	Manikonda	2.0	1.0	
3	5739000	1295	Alwal	3.0	0.0	
4	5679000	1145	Kukatpally	2.0	0.0	
5	6099000	1230	Nizampet	2.0	0.0	
6	7000000	1350	Manikonda	2.0	0.0	
7	4198000	1400	Gachibowli	3.0	1.0	
8	9900000	2240	Tellapur	NaN	NaN	
9	9800000	1520	Gachibowli	2.0	0.0	
10	8903000	1535	Kokapet	3.0	1.0	
11	6300000	1150	Hyder Nagar	2.0	0.0	
12	6300000	1285	Mehdipatnam	2.0	0.0	
13	9000000	1897	Tellapur	3.0	1.0	
14	11100000	1725	Narsingi	3.0	1.0	
15	11500000	XYZ	Khajaguda Nanakramguda Road	3.0	NaN	
16	10900000	1592	Kukatpally	3.0	1.0	
17	8625000	1150	Madhapur	2.0	0.0	
18	9900000	2240	Tellapur	3.0	1.0	

	SwimmingPool	JoggingTrack	RainWaterHarvesting	IndoorGames
0	1.0	1.0	1.0	1.0
1	1.0	1.0	1.0	1.0
2	0.0	0.0	0.0	1.0
3	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0
7	1.0	0.0	0.0	0.0
8	NaN	0.0	0.0	1.0
9	0.0	0.0	0.0	0.0
10	1.0	1.0	1.0	1.0
11	0.0	0.0	0.0	0.0
12	0.0	NaN	NaN	NaN
13	1.0	1.0	1.0	1.0
14	1.0	1.0	0.0	1.0
15	NaN	0.0	0.0	0.0
16	1.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0
18	1.0	0.0	0.0	1.0

b) Printing only the first 5 rows in the data using head():

	Price	Area	Location	No. of Bedrooms	Gymnasium	SwimmingPool	\
0	6968000	1340	Nizampet	2.0	1.0	1.0	
1	29000000	3498	Hitech City	4.0	1.0	1.0	
2	6590000	1318	Manikonda	2.0	1.0	0.0	
3	5739000	1295	Alwal	3.0	0.0	0.0	
4	5679000	1145	Kukatpally	2.0	0.0	0.0	

  

	JoggingTrack	RainWaterHarvesting	IndoorGames
0	1.0	1.0	1.0
1	1.0	1.0	1.0
2	0.0	0.0	1.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0

c) Removing the NULL values in the data:

	Price	Area	Location	No. of Bedrooms	Gymnasium	SwimmingPool	\
0	6968000	1340	Nizampet	2.0	1.0	1.0	
1	29000000	3498	Hitech City	4.0	1.0	1.0	
2	6590000	1318	Manikonda	2.0	1.0	0.0	
3	5739000	1295	Alwal	3.0	0.0	0.0	
4	5679000	1145	Kukatpally	2.0	0.0	0.0	
5	6099000	1230	Nizampet	2.0	0.0	0.0	
6	7000000	1350	Manikonda	2.0	0.0	0.0	
7	4198000	1400	Gachibowli	3.0	1.0	1.0	
9	9800000	1520	Gachibowli	2.0	0.0	0.0	
10	8903000	1535	Kokapet	3.0	1.0	1.0	
11	6300000	1150	Hyder Nagar	2.0	0.0	0.0	
13	9000000	1897	Tellapur	3.0	1.0	1.0	
14	11100000	1725	Narsingi	3.0	1.0	1.0	
16	10900000	1592	Kukatpally	3.0	1.0	1.0	
17	8625000	1150	Madhapur	2.0	0.0	0.0	
18	9900000	2240	Tellapur	3.0	1.0	1.0	

  

	JoggingTrack	RainWaterHarvesting	IndoorGames
0	1.0	1.0	1.0
1	1.0	1.0	1.0
2	0.0	0.0	1.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0
5	0.0	0.0	0.0
6	0.0	0.0	0.0
7	0.0	0.0	0.0
9	0.0	0.0	0.0
10	1.0	1.0	1.0
11	0.0	0.0	0.0
13	1.0	1.0	1.0
14	1.0	0.0	1.0
16	0.0	0.0	0.0
17	0.0	0.0	0.0
18	0.0	0.0	1.0



d) Analyzing the data using describe()

	Price	No. of Bedrooms	Gymnasium	SwimmingPool	JoggingTrack \
count	1.600000e+01	16.000000	16.000000	16.000000	16.000000
mean	9.112562e+06	2.562500	0.562500	0.500000	0.312500
std	5.679100e+06	0.629153	0.512348	0.516398	0.478714
min	4.198000e+06	2.000000	0.000000	0.000000	0.000000
25%	6.249750e+06	2.000000	0.000000	0.000000	0.000000
50%	7.812500e+06	2.500000	1.000000	0.500000	0.000000
75%	9.825000e+06	3.000000	1.000000	1.000000	1.000000
max	2.900000e+07	4.000000	1.000000	1.000000	1.000000

	RainWaterHarvesting	IndoorGames
count	16.000000	16.000000
mean	0.250000	0.437500
std	0.447214	0.512348
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.250000	1.000000
max	1.000000	1.000000

---