```
"""Name: Akshara.D

Registration Number: 21BAI1399

Email ID: akshara.d2021@vitstudent.ac.in"""

#Drive link: https://colab.research.google.com/drive/1XNC51HT_aV96oKk_sF0uhDoMdWFHlTen?usp=sharing
#Git hub: https://github.com/Akshara-Daram/DATA-SCIENCE-1/blob/main/Assignment_1_Data_Science%20(2).ipynb


     'Name: Akshara.D\n\nRegistration Number: 21BAI1399\n\nEmail ID: akshara.d2021@vitstudent.ac.in'


#Task 1 - Create a pandas dataframe (Dataframe name as 'df') with numpy random values (4 features and 4 observations)


import pandas as pd
import numpy as np

np.random.seed(10)

data = {
    'CustomerID': np.random.randint(10000, 99999, 4),
    'ProductID': np.random.randint(1000, 9999, 4),
    'OrderID': np.random.randint(100000, 999999, 4),
    'Quantity': np.random.randint(1, 10, 4),
}

df = pd.DataFrame(data)

print(df)
```

```
   CustomerID  ProductID  OrderID  Quantity
0       93209       5829   940036         2
1       60496       2520   280463         9
2       19372       7400   675883         5
3       20201       6648   922616         2
```

```
#Task 2 - Rename the task 1 'df' dataframe column names to 'Random value 1' , 'Random value 2' , 'Random value 3' , 'Random value 4'

df.rename(columns={'CustomerID':'Random value 1','ProductID':'Random value 2','OrderID':'Random value 3','Quantity':'Random value 4'},inplace
df
```

	Random value 1	Random value 2	Random value 3	Random value 4
0	93209	5829	940036	2
1	60496	2520	280463	9
2	19372	7400	675883	5
3	20201	6648	922616	2

```
# Task 3 - Find the descriptive statistics of the 'df' dataframe

statistics = df.describe()
print(statistics)
```

```
       Random value 1  Random value 2  Random value 3  Random value 4
count        4.000000         4.00000        4.000000        4.000000
mean     48319.500000      5599.25000   704749.500000        4.500000
std      35552.506527      2150.74737   307504.936442        3.316625
min      19372.000000      2520.00000   280463.000000        2.000000
25%      19993.750000      5001.75000   577028.000000        2.000000
50%      40348.500000      6238.50000   799249.500000        3.500000
75%      68674.250000      6836.00000   926971.000000        6.000000
max      93209.000000      7400.00000   940036.000000        9.000000
```

```
#Task 4 - Check for the null values in 'df' and find the data type of the columns

null_values = df.isnull().sum()
print(null_values)

null_values = df.isnull().any()
print(null_values)
```

```
#Data types of the columns
data_types = df.dtypes
print("\nData Types:")
print(data_types)
```

```
Random value 1    0
Random value 2    0
Random value 3    0
Random value 4    0
dtype: int64
Random value 1    False
Random value 2    False
Random value 3    False
Random value 4    False
dtype: bool

Data Types:
Random value 1    int64
Random value 2    int64
Random value 3    int64
Random value 4    int64
dtype: object
```

Double-click (or enter) to edit

```
#Task 5 - Display the 'Random value 2' and 'Random value 3' columns with location method and index location method

# Location method
columns_loc = df.loc[:, ['Random value 2', 'Random value 3']]
print("Columns using .loc[]:")
print(columns_loc)

# Index location method
columns_iloc = df.iloc[:, [1, 2]]
print("\nColumns using .iloc[]:")
print(columns_iloc)
```

```
Columns using .loc[]:
   Random value 2  Random value 3
0            5829          940036
1            2520          280463
2            7400          675883
3            6648          922616

Columns using .iloc[]:
   Random value 2  Random value 3
0            5829          940036
1            2520          280463
2            7400          675883
3            6648          922616
```