

```
In [1]: import pandas as pd      #importing pandas
```

```
In [2]: data=pd.read_csv("/home/placemnet/YUVA/customer_details.csv")      #reading the customer details file
data
```

Out[2]:

	customer_id	sex	customer_age	tenure
0	9798859	Male	44.0	93
1	11413563	Male	36.0	65
2	818195	Male	35.0	129
3	12049009	Male	33.0	58
4	10083045	Male	42.0	88
...
19995	12557307	Male	41.0	52
19996	12595961	Male	29.0	52
19997	12520991	Male	35.0	52
19998	12612719	Male	39.0	52
19999	12572063	Male	28.0	52

20000 rows × 4 columns

```
In [3]: data1=pd.read_csv("/home/placemnet/YUVA/basket_details.csv")    #reading the basket details file
data1.describe()
```

Out[3]:

	customer_id	product_id	basket_count
count	1.500000e+04	1.500000e+04	15000.000000
mean	1.808567e+07	3.269771e+07	2.153733
std	1.233000e+07	1.629455e+07	0.517929
min	4.784000e+03	4.939000e+04	2.000000
25%	8.659327e+06	3.137412e+07	2.000000
50%	1.520775e+07	3.694759e+07	2.000000
75%	2.663904e+07	4.502408e+07	2.000000
max	4.460824e+07	5.579097e+07	10.000000

```
In [4]: data.groupby(['customer_id']).count()      #to count the number of values in customer id column in data
```

Out[4]:

	sex	customer_age	tenure
customer_id			
2093	1	1	1
12817	1	1	1
14309	1	1	1
15155	1	1	1
23205	1	1	1
...
44392831	1	1	1
44401175	1	1	1
44431821	1	1	1
44621778	1	1	1
44625658	1	1	1

20000 rows × 3 columns

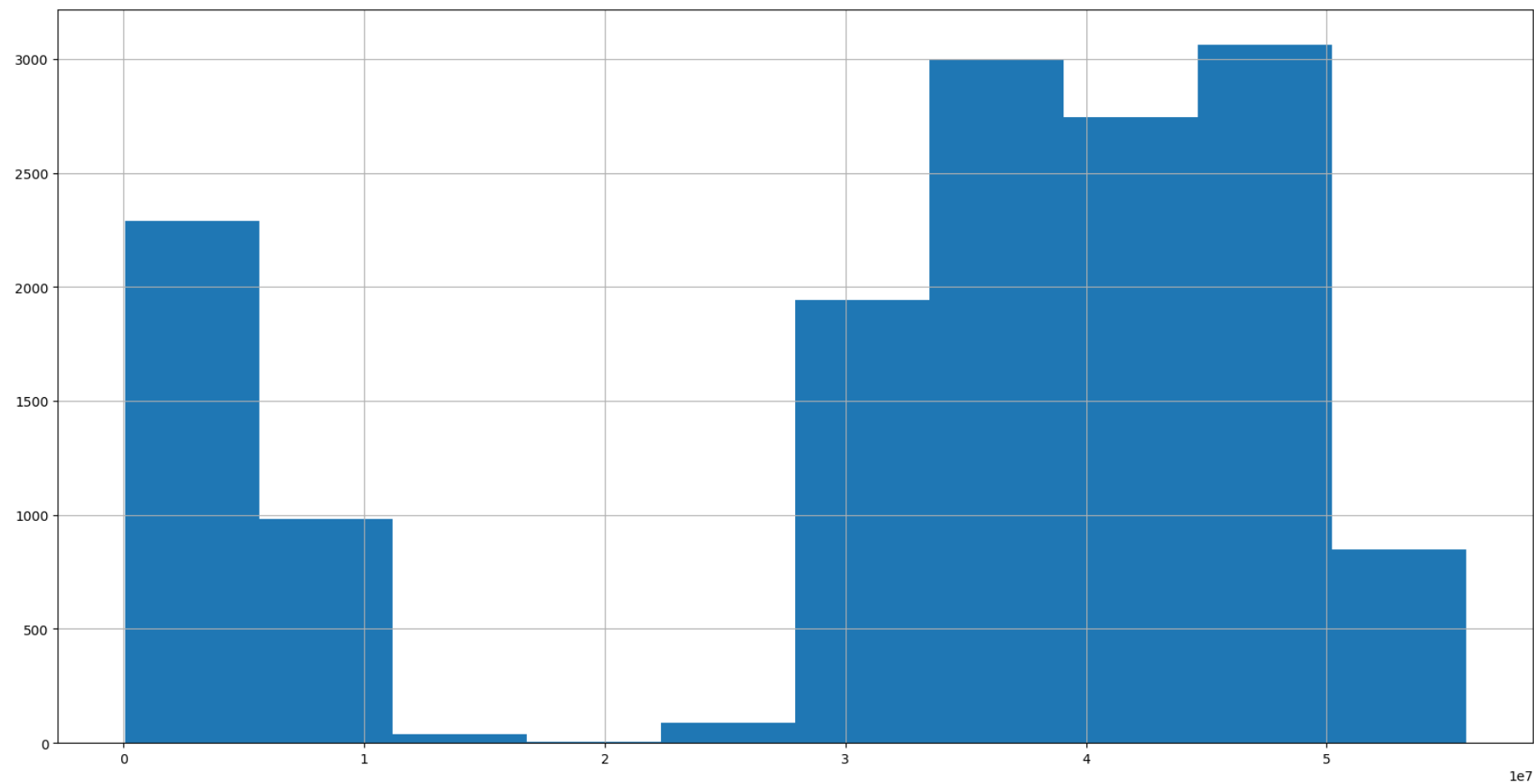
```
In [5]: data1.groupby(['customer_id']).count()    #to count the number of values in customer id column in data1
```

Out[5]:

	product_id	basket_date	basket_count
customer_id			
4784	1	1	1
8314	2	2	2
8857	1	1	1
9273	1	1	1
11172	1	1	1
...
44460516	1	1	1
44461180	1	1	1
44473609	1	1	1
44486815	1	1	1
44608245	1	1	1

13871 rows × 3 columns

```
In [6]: import matplotlib.pyplot as plt  
data1['product_id'].hist(figsize=(20,10))    #  
plt.show()
```



```
In [7]: y=pd.merge(data,data1,on="customer_id")    #merging the two data frames using a column name
```

In [8]:

y

Out[8]:

	customer_id	sex	customer_age	tenure	product_id	basket_date	basket_count
0	9500953	Male	55.0	96	3446783	2019-06-10	3
1	851739	Male	40.0	129	32920704	2019-06-19	2
2	9654043	Male	37.0	95	51307669	2019-06-08	2
3	4912369	Male	36.0	114	33923115	2019-05-20	2
4	9875271	Male	34.0	92	31586037	2019-06-06	2
...
67	13278573	Male	28.0	47	4488682	2019-05-26	2
68	12901520	Female	40.0	50	38610580	2019-05-28	3
69	12737235	Male	39.0	51	32933848	2019-05-21	2
70	12737235	Male	39.0	51	46373374	2019-05-21	3
71	12574807	Male	33.0	52	32056122	2019-05-25	2

72 rows × 7 columns

In [9]: data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False) *#arranging in descending*

Out[9]:

product_id	
43524799	69
31516269	59
39833031	50
46130148	36
34913531	28
...	..
34003520	2
34003697	2
34004660	2
34013459	2
55790974	2

Name: basket_count, Length: 13161, dtype: int64

```
In [10]: data1.groupby(['product_id'])['basket_count'].sum().unique()
```

```
Out[10]: array([ 2,  3,  7,  8,  4,  5,  6,  9, 14, 15, 10, 20, 27, 12, 23, 13, 11,  
                19, 59, 22, 18, 28, 17, 21, 16, 50, 69, 36])
```

```
In [11]: y.groupby(['customer_age']).count()
```

```
Out[11]:
```

	customer_id	sex	tenure	product_id	basket_date	basket_count
customer_age						
5.0	1	1	1	1	1	1
22.0	2	2	2	2	2	2
23.0	1	1	1	1	1	1
24.0	2	2	2	2	2	2
25.0	2	2	2	2	2	2
26.0	1	1	1	1	1	1
27.0	4	4	4	4	4	4
28.0	3	3	3	3	3	3
29.0	6	6	6	6	6	6
30.0	3	3	3	3	3	3
32.0	4	4	4	4	4	4
33.0	2	2	2	2	2	2
34.0	3	3	3	3	3	3
35.0	2	2	2	2	2	2
36.0	4	4	4	4	4	4
37.0	2	2	2	2	2	2
39.0	3	3	3	3	3	3
40.0	5	5	5	5	5	5
41.0	1	1	1	1	1	1
42.0	2	2	2	2	2	2
43.0	3	3	3	3	3	3
45.0	1	1	1	1	1	1

	customer_id	sex	tenure	product_id	basket_date	basket_count
customer_age						
46.0	1	1	1	1	1	1
51.0	3	3	3	3	3	3
55.0	1	1	1	1	1	1
57.0	2	2	2	2	2	2
61.0	1	1	1	1	1	1
67.0	2	2	2	2	2	2