

```
In [2]: import pandas as pd
```

```
In [3]: data=pd.read_csv("/home/placement/Downloads/customer_details.csv")#read csv file
```

```
In [4]: data.describe()#describe data
```

Out[4]:

	customer_id	customer_age	tenure
count	2.000000e+04	20000.000000	20000.000000
mean	1.760040e+07	262.222550	44.396800
std	8.679505e+06	604.321589	31.998376
min	2.093000e+03	-34.000000	4.000000
25%	1.188115e+07	29.000000	21.000000
50%	1.560912e+07	38.000000	35.000000
75%	2.228484e+07	123.000000	60.000000
max	4.462566e+07	2022.000000	133.000000

```
In [5]: data1=pd.read_csv("/home/placement/Downloads/basket_details.csv")#read another csv file
```

```
In [6]: data1.describe()#describe data
```

Out[6]:

	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 [7]: data.tail()#to get bottom rows
```

Out[7]:

	customer_id	sex	customer_age	tenure
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

In [8]: data

Out[8]:

	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 [9]: data1.tail()*#to get bottom rows*

Out[9]:

	customer_id	product_id	basket_date	basket_count
14995	8336862	50977318	2019-05-26	2
14996	9500785	43862061	2019-05-26	2
14997	22787344	6041664	2019-05-26	2
14998	8221263	3597369	2019-05-26	2
14999	4912577	46646893	2019-05-26	2

```
In [10]: data1
```

```
Out[10]:
```

	customer_id	product_id	basket_date	basket_count
0	42366585	41475073	2019-06-19	2
1	35956841	43279538	2019-06-19	2
2	26139578	31715598	2019-06-19	3
3	3262253	47880260	2019-06-19	2
4	20056678	44747002	2019-06-19	2
...
14995	8336862	50977318	2019-05-26	2
14996	9500785	43862061	2019-05-26	2
14997	22787344	6041664	2019-05-26	2
14998	8221263	3597369	2019-05-26	2
14999	4912577	46646893	2019-05-26	2

15000 rows × 4 columns

```
In [11]: data1.groupby(['customer_id']).count()#determinbe the unique ids of a customer
```

```
Out[11]:
```

	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 [12]: data.groupby(['customer_id']).count()#determinbe the n ids of a customer
```

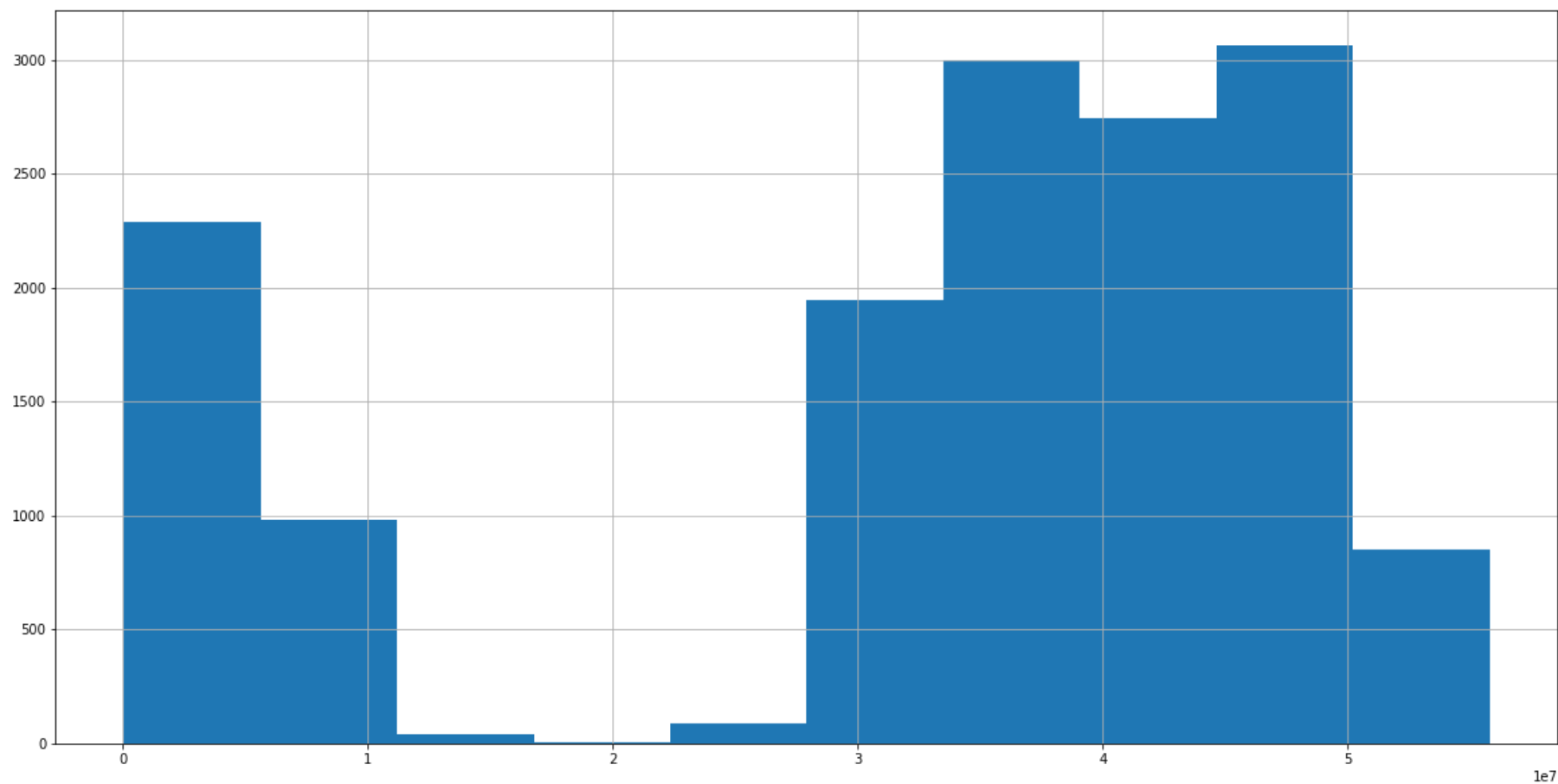
```
Out[12]:
```

	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 [13]: data1['product_id'].hist(figsize=(20,10))#creating a histogram for data1
```

Out[13]: <AxesSubplot:>



```
In [14]: data2=pd.merge(data,data1,on="customer_id")#merge the datas into a single column
```

```
In [15]: data2
```

```
Out[15]:
```

	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 [16]: data2.describe()#describe the data
```

```
Out[16]:
```

	customer_id	customer_age	tenure	product_id	basket_count
count	7.200000e+01	72.000000	72.000000	7.200000e+01	72.000000
mean	1.554364e+07	68.458333	56.180556	3.140376e+07	2.152778
std	9.961282e+06	234.574289	38.948621	1.616160e+07	0.362298
min	3.809750e+05	5.000000	4.000000	8.287500e+04	2.000000
25%	1.026443e+07	29.000000	24.750000	2.980404e+07	2.000000
50%	1.352736e+07	35.500000	45.500000	3.498005e+07	2.000000
75%	2.037478e+07	43.000000	83.750000	4.359420e+07	2.000000
max	4.328080e+07	2022.000000	130.000000	5.130767e+07	3.000000

```
In [17]: data2.customer_id.unique()#determinbe the unique ids of a customer
```

```
Out[17]: array([ 9500953,   851739,   9654043,   4912369,   9875271,  11737579,
  10619833,   4193819,   4897641,   4643359,   380975,  11623549,
  11724853,  12410433,  10394153,    537173,  11440499,  10439331,
  10629563,   4257099,  11346069,   8508353,   9700145,  10814041,
    9804585,   4238087,  11665521,   1030589,  11072047,  43280797,
  41790413,  39814593,  36623391,  34677755,  29144255,  27081691,
  25055107,  25567283,  23179191,  22524187,  21765975,  21142247,
  20789769,  20236456,  20174063,  17909829,  18256077,  17830393,
  16944627,  16398473,  16029475,  15436141,  15570891,  15192667,
  15067633,  14966315,  15141119,  14248059,  14053193,  13776147,
  13278573,  12901520,  12737235,  12574807])
```

```
In [18]: data1.head()#display top 5 rows default
```

```
Out[18]:
```

	customer_id	product_id	basket_date	basket_count
0	42366585	41475073	2019-06-19	2
1	35956841	43279538	2019-06-19	2
2	26139578	31715598	2019-06-19	3
3	3262253	47880260	2019-06-19	2
4	20056678	44747002	2019-06-19	2

```
In [19]: data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)#true=ascending order
```

```
Out[19]: product_id
49390      2
42094163   2
42102274   2
42110403   2
42110580   2
...
34913531  28
46130148  36
39833031  50
31516269  59
43524799  69
Name: basket_count, Length: 13161, dtype: int64
```

```
In [20]: data2.groupby(['customer_age']).count()#determine customer age with different ages
```

Out[20]:

	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
46.0	1	1	1	1	1	1

	customer_id	sex	tenure	product_id	basket_date	basket_count
customer_age						
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
123.0	4	4	4	4	4	4
2022.0	1	1	1	1	1	1

```
In [21]: data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)#false=descending order
```

```
Out[21]: 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 [22]: cor_mat=data1.corr()#to detrmine correlation  
cor_mat
```

Out[22]:

	customer_id	product_id	basket_count
customer_id	1.000000	0.001937	0.058235
product_id	0.001937	1.000000	-0.006407
basket_count	0.058235	-0.006407	1.000000

```
In [23]: import seaborn as sns#to view the correlation using graph  
sns.heatmap(cor_mat,vmax=1,vmin=-1,annot=True,linewidth=.5,cmap='bwr')
```

Out[23]: <AxesSubplot:>



In []: