In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sb
```

In [2]:
```
data=pd.read_csv("/home/placement/Downloads/basket_details.csv")#Reading the baskect data
data
```

Out[2]:

	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 [3]:
```
data1=pd.read_csv("/home/placement/Downloads/customer_details.csv")#Reading the customer data
data1
```

Out[3]:

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

Out[4]:

	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 [5]:
```
data1.describe()
```

Out[5]:

	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 [6]:
```
data.tail()
```

Out[6]:

	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 [7]:
```
data1['sex']=data1['sex'].map({'Male':1,'Female':2})#Converting strings into integers
data1
```

Out[7]:

	customer_id	sex	customer_age	tenure
0	9798859	1.0	44.0	93
1	11413563	1.0	36.0	65
2	818195	1.0	35.0	129
3	12049009	1.0	33.0	58
4	10083045	1.0	42.0	88
...	...	...	...	...
19995	12557307	1.0	41.0	52
19996	12595961	1.0	29.0	52
19997	12520991	1.0	35.0	52
19998	12612719	1.0	39.0	52
19999	12572063	1.0	28.0	52

20000 rows × 4 columns

In [8]:
```
a=data1.loc[data1.sex==2]
a
```

Out[8]:

	customer_id	sex	customer_age	tenure
16	831271	2.0	38.0	129
18	11350661	2.0	24.0	66
23	11328737	2.0	41.0	66
28	12417929	2.0	35.0	54
32	10189011	2.0	39.0	86
...	...	...	...	...
19973	12623079	2.0	49.0	52
19977	12606531	2.0	36.0	52
19986	12560981	2.0	46.0	52
19987	12525219	2.0	40.0	52
19990	12595849	2.0	27.0	52

4669 rows × 4 columns

In [9]:
```
data.customer_id.unique()#unique values
```

Out[9]:
```
array([42366585, 35956841, 26139578, ...,  8336862,  9500785, 22787344])
```

In [10]:
```
data.groupby(['customer_id']).count()#grouping the basket detials file with respect to customer_id
```

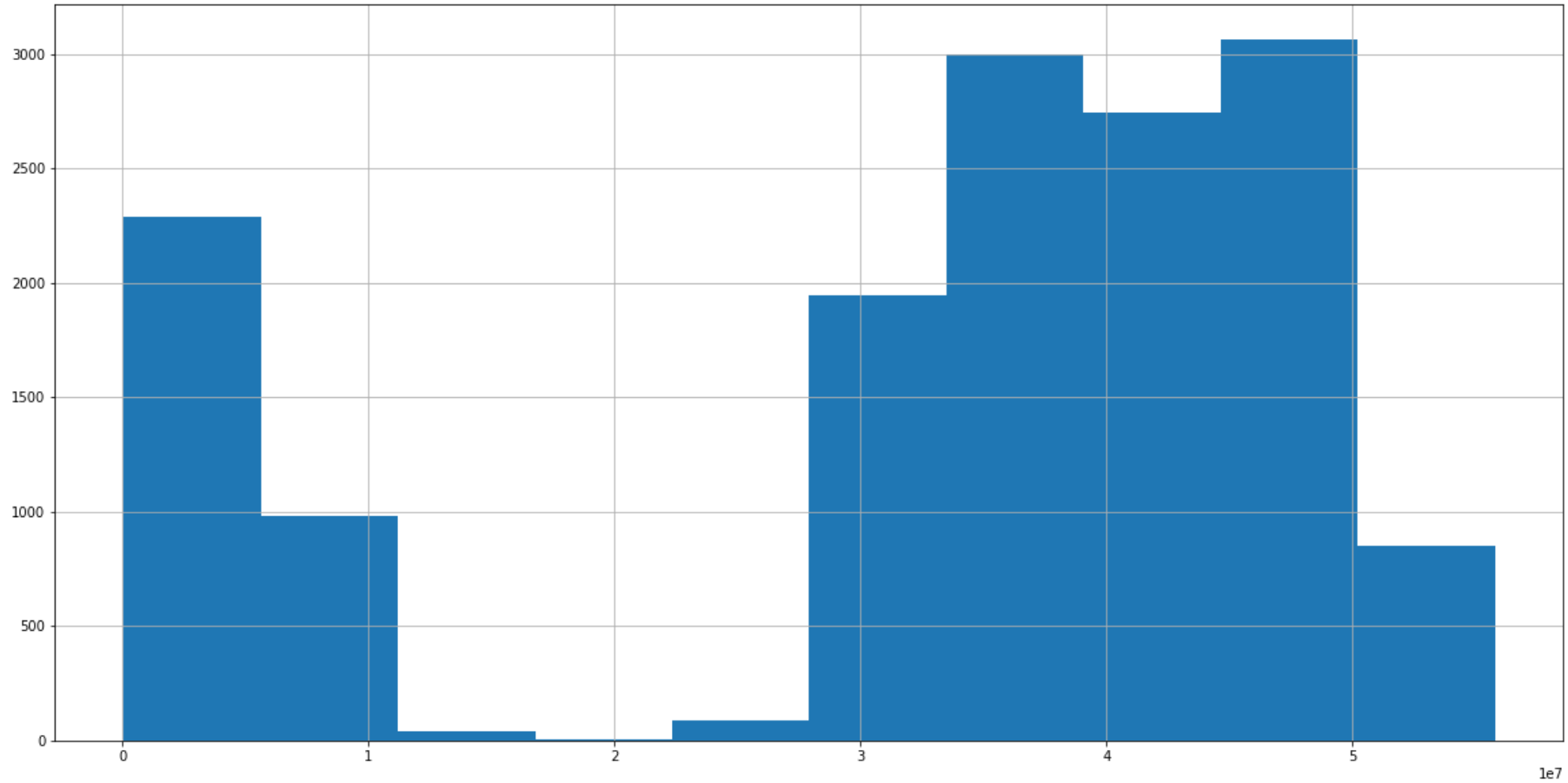
Out[10]:

customer_id	product_id	basket_date	basket_count
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 [11]:
```
data['product_id'].hist(figsize=(20,10))#Histograph to the data
```

Out[11]: <Axes: >



In [12]:
```
cor=data1.corr()
sb.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidth=5,cmap='bwr')#Correlating graph to the data
```

Out[12]: <Axes: >

In [ ]:
```

```

In [13]:
```
test=pd.merge(data,data1, on = "customer_id")#mergeing customer and basket data
```

In [14]:
```
test
```

Out[14]:

	customer_id	product_id	basket_date	basket_count	sex	customer_age	tenure
0	4897641	34525548	2019-06-15	2	1.0	40.0	114
1	11623549	50394038	2019-06-18	2	1.0	30.0	63
2	11665521	41476812	2019-06-15	2	2.0	51.0	62
3	4193819	6455162	2019-06-15	2	1.0	42.0	117
4	1030589	38578121	2019-05-26	2	1.0	45.0	127
...	...	...	...	...	...	...	...
67	12574807	32056122	2019-05-25	2	1.0	33.0	52
68	15192667	31272089	2019-05-24	2	1.0	46.0	37
69	14248059	48790153	2019-05-21	2	1.0	29.0	41
70	10629563	47864502	2019-06-01	2	1.0	29.0	76
71	11737579	46626448	2019-05-27	2	1.0	35.0	61

72 rows × 7 columns

In [15]:
```
(['product_id'])['basket_count'].sum().sort_values(ascending=False)#Grouping data and arranging in descendin
```

Out[15]:
```
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 [16]:
```
data.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)
```

Out[16]:
```
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 [ ]:
```

```