

```
In [2]: import pandas as pd
data=pd.read_csv("/home/placement/Downloads/customer_details.csv")
data1=pd.read_csv("/home/placement/Downloads/basket_details.csv")
```

```
In [3]: data.describe()
```

Out[3]:

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

```
Out[4]:
```

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

```
In [5]: data.tail()
```

```
Out[5]:
```

	customer_id	sex	customer_age	tenure
<b>19995</b>	12557307	Male	41.0	52
<b>19996</b>	12595961	Male	29.0	52
<b>19997</b>	12520991	Male	35.0	52
<b>19998</b>	12612719	Male	39.0	52
<b>19999</b>	12572063	Male	28.0	52

```
In [6]: data1.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.groupby(['customer_id']).count()
```

```
Out[7]:
```

	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 [8]: data.groupby(['customer_id']).count()
```

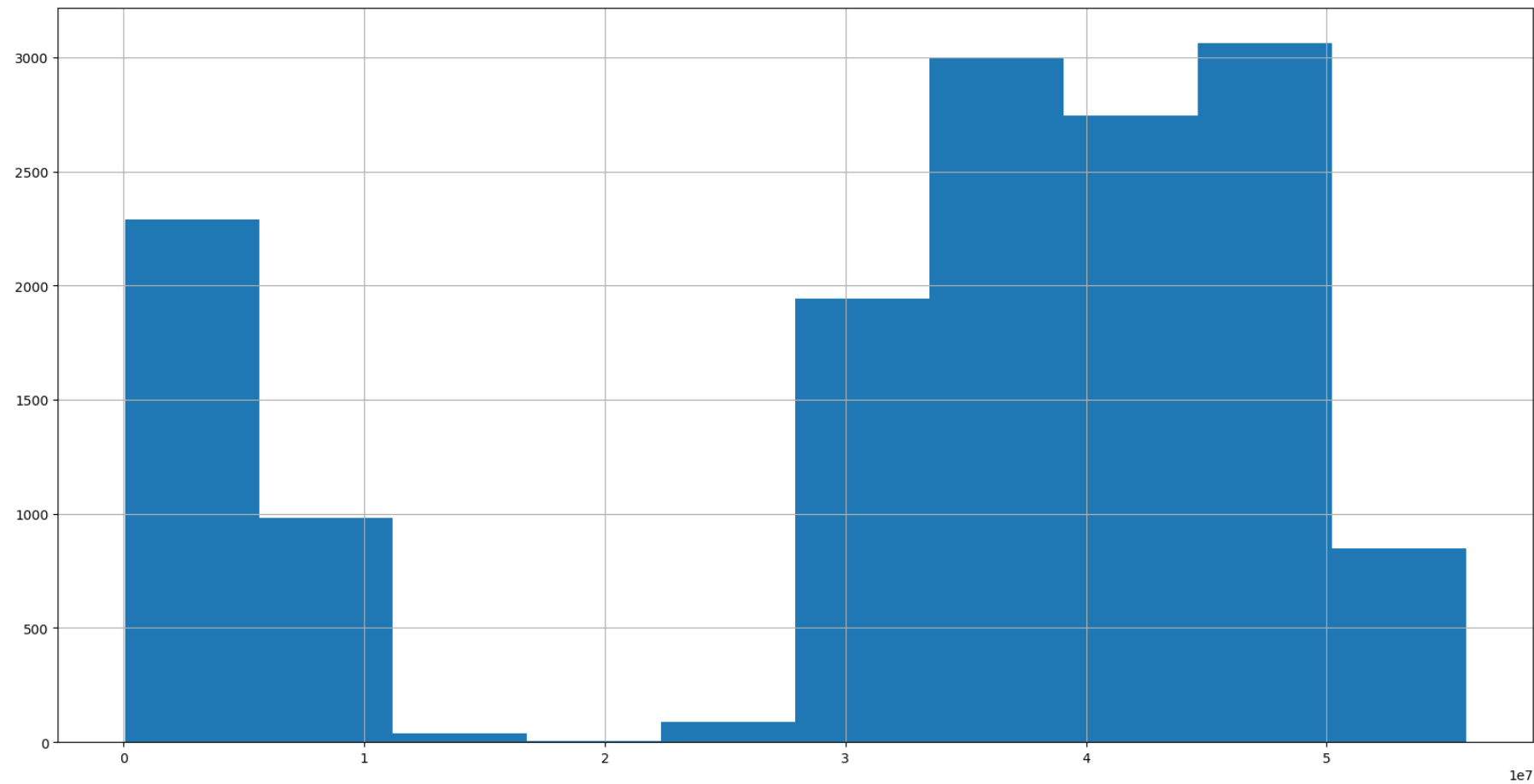
```
Out[8]:
```

	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 [9]: data1['product_id'].hist(figsize=(20,10))
```

```
Out[9]: <Axes: >
```



```
In [10]: !pip install seaborn
```

```
Requirement already satisfied: seaborn in ./anaconda3/lib/python3.10/site-packages (0.12.2)  
Requirement already satisfied: pandas>=0.25 in ./anaconda3/lib/python3.10/site-packages (from seaborn) (1.5.3)  
Requirement already satisfied: numpy!=1.24.0,>=1.17 in ./anaconda3/lib/python3.10/site-packages (from seaborn) (1.23.5)  
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in ./anaconda3/lib/python3.10/site-packages (from seaborn) (3.7.0)  
Requirement already satisfied: contourpy>=1.0.1 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)  
Requirement already satisfied: fonttools>=4.22.0 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)  
Requirement already satisfied: pyparsing>=2.3.1 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)  
Requirement already satisfied: cycler>=0.10 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)  
Requirement already satisfied: kiwisolver>=1.0.1 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)  
Requirement already satisfied: python-dateutil>=2.7 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)  
Requirement already satisfied: pillow>=6.2.0 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)  
Requirement already satisfied: packaging>=20.0 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (22.0)  
Requirement already satisfied: pytz>=2020.1 in ./anaconda3/lib/python3.10/site-packages (from pandas>=0.25->seaborn) (2022.7)  
Requirement already satisfied: six>=1.5 in ./anaconda3/lib/python3.10/site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
```

```
In [11]: test=pd.merge(data,data1,on="customer_id")
```

In [12]: test

Out[12]:

	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 [13]: `test.describe()`

Out[13]:

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

In [14]: `data1.head()`

Out[14]:

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



```
In [15]: data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
```

```
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]: data1.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 [17]: test.groupby(['customer_age']).count()
```

Out[17]:

	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

```
In [20]: import warnings
warnings.filterwarnings('ignore')
```

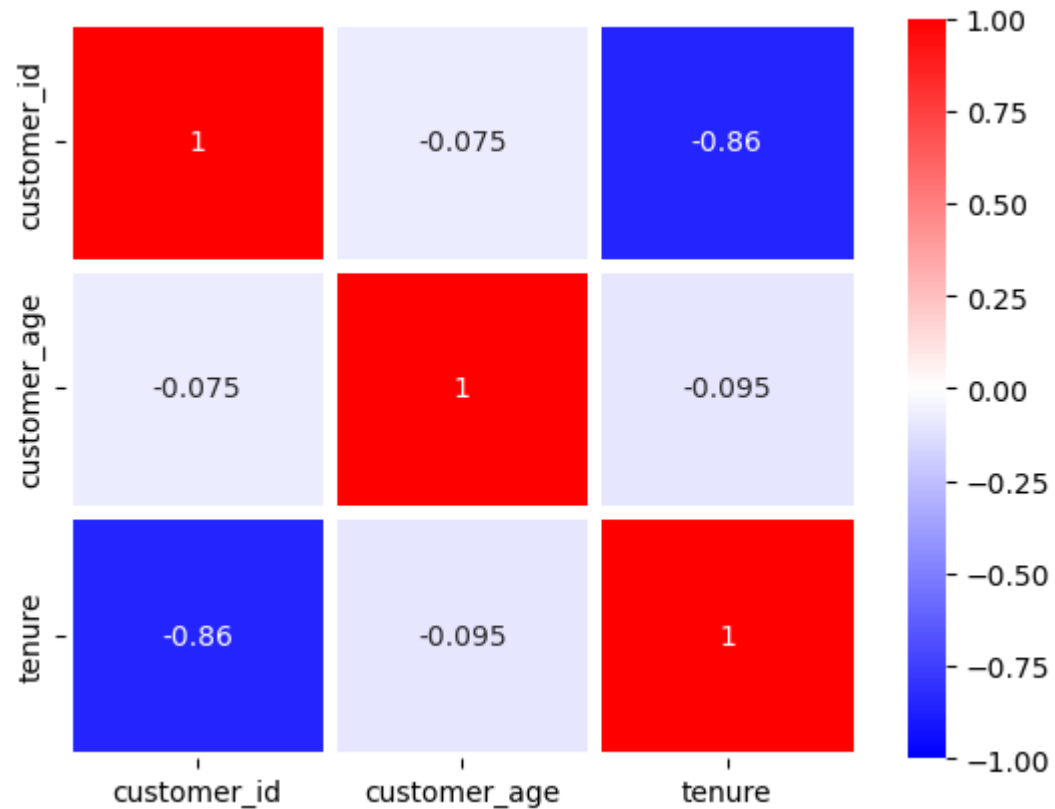
```
In [21]: cor=data.corr()
cor
```

Out[21]:

	customer_id	customer_age	tenure
customer_id	1.000000	-0.075467	-0.855410
customer_age	-0.075467	1.000000	-0.095013
tenure	-0.855410	-0.095013	1.000000

```
In [22]: import seaborn as sns
sns.heatmap(cor, vmax=1, vmin=-1, annot=True, linewidth=5, cmap='bwr')
```

Out[22]: <Axes: >



In [ ]: