In [8]:

```
import pandas as pd
```

In [9]:

```
import numpy as np
```

In [10]:

```
data1=pd.read_csv('/home/palcement/Downloads/basket_details.csv')
tharun=pd.read_csv('/home/palcement/Downloads/customer_details.csv')
```

In [11]:

```
tharun.head()
```

Out[11]:

	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

In [5]:

```
data1.head()
```

Out[5]:

	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 [6]:

```
list(data1)
```

Out[6]:

```
['customer_id', 'product_id', 'basket_date', 'basket_count']
```

In [12]:

```
data1.describe()
tharun.describe()
```

Out[12]:

	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 [13]:

```
data1.tail()
tharun.tail()
```

Out[13]:

	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 [14]:

```
tharun.tail()
```

Out[14]:

	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 [15]:

```
data1.tail()
```

Out[15]:

	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 [16]:

```
tharun.head()
data1.head()
```

Out[16]:

	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 [17]:

```
tharun.head()
```

Out[17]:

	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

In [18]:

```
data1.groupby(['customer_id']).count()
```

Out[18]:

	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

In [19]:

```
tharun.groupby(['customer_id']).count()
```

Out[19]:

	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 [20]:

```
data1.groupby(['product_id']).count()
```

Out[20]:

	customer_id	basket_date	basket_count
product_id			
49390	1	1	1
52798	1	1	1
53091	1	1	1
53093	1	1	1
53238	3	3	3
...	...	...	...
55445659	1	1	1
55464635	1	1	1
55521098	1	1	1
55578837	1	1	1
55790974	1	1	1

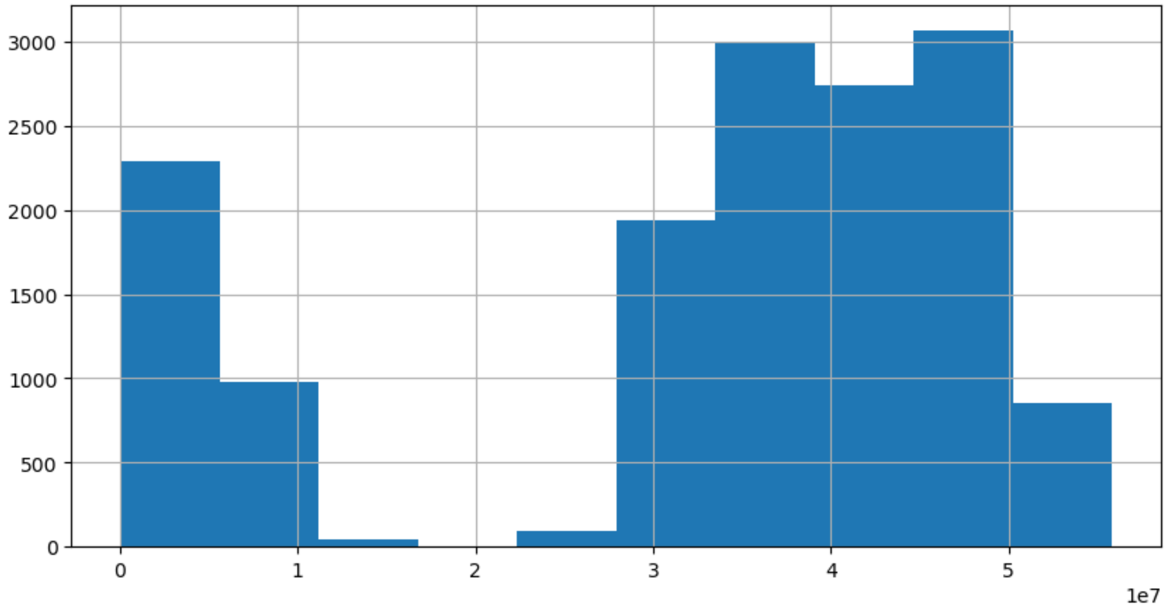
13161 rows × 3 columns

In [24]:

```
data1['product_id'].hist(figsize=(10,5))
```

Out[24]:

```
<Axes: >
```



In [27]:

```
pip install seaborn
```

```
Requirement already satisfied: seaborn in ./anaconda3/lib/python3.10/s
ite-packages (0.12.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in ./anaconda3/
lib/python3.10/site-packages (from seaborn) (3.7.0)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in ./anaconda3/li
b/python3.10/site-packages (from seaborn) (1.23.5)
Requirement already satisfied: pandas>=0.25 in ./anaconda3/lib/python
3.10/site-packages (from seaborn) (1.5.3)
Requirement already satisfied: python-dateutil>=2.7 in ./anaconda3/li
b/python3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.
8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in ./anaconda3/lib/py
thon3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: pyparsing>=2.3.1 in ./anaconda3/lib/pyt
hon3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: fonttools>=4.22.0 in ./anaconda3/lib/py
thon3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.
0)
Requirement already satisfied: contourpy>=1.0.1 in ./anaconda3/lib/pyt
hon3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: packaging>=20.0 in ./anaconda3/lib/pyth
on3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (22.0)
Requirement already satisfied: cycler>=0.10 in ./anaconda3/lib/python
3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in ./anaconda3/lib/python
3.10/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: pytz>=2020.1 in ./anaconda3/lib/python
3.10/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->sea
born) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [28]:

```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in ./anaconda3/lib/python3.1
0/site-packages (3.7.0)
Requirement already satisfied: pyparsing>=2.3.1 in ./anaconda3/lib/pyt
hon3.10/site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: contourpy>=1.0.1 in ./anaconda3/lib/pyt
hon3.10/site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: python-dateutil>=2.7 in ./anaconda3/li
b/python3.10/site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: fonttools>=4.22.0 in ./anaconda3/lib/py
thon3.10/site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: numpy>=1.20 in ./anaconda3/lib/python3.
10/site-packages (from matplotlib) (1.23.5)
Requirement already satisfied: cycler>=0.10 in ./anaconda3/lib/python
3.10/site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in ./anaconda3/lib/python
3.10/site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: packaging>=20.0 in ./anaconda3/lib/pyth
on3.10/site-packages (from matplotlib) (22.0)
Requirement already satisfied: kiwisolver>=1.0.1 in ./anaconda3/lib/py
thon3.10/site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: six>=1.5 in ./anaconda3/lib/python3.10/
site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [29]:

```
test=pd.merge(data1,tharun, on = 'customer_id')
```

In [30]:

```
test
```

Out[30]:

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

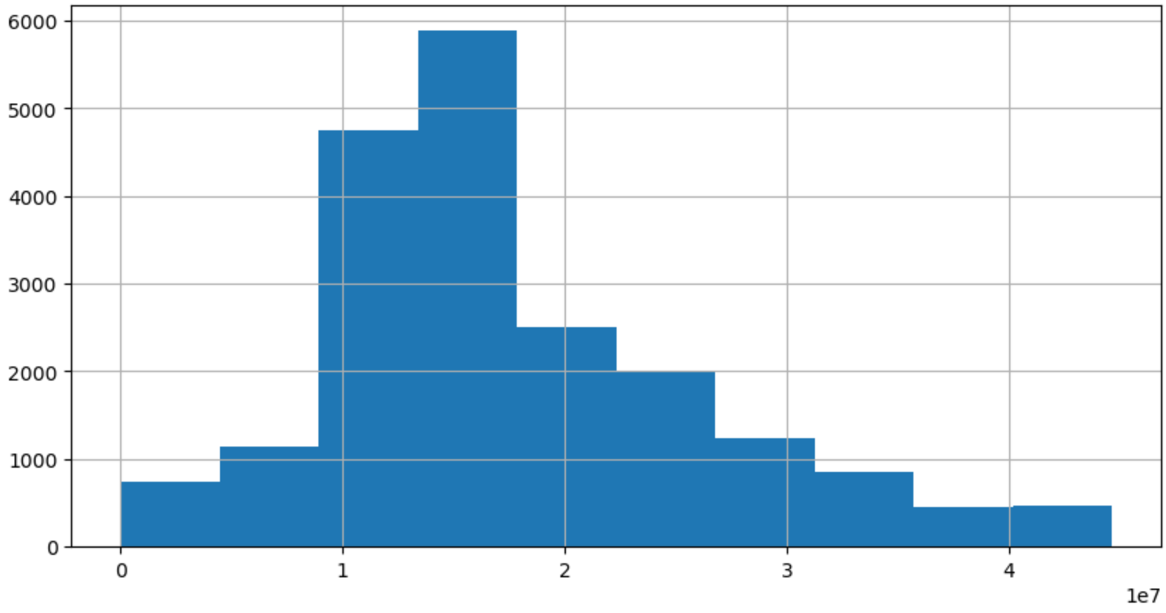
72 rows × 7 columns

In [32]:

```
tharun['customer_id'].hist(figsize=(10,5))
```

Out[32]:

```
<Axes: >
```



In [33]:

```
test.describe()
```

Out[33]:

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

In [35]:

```
data1.head()
```

Out[35]:

	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 [37]:

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

Out[37]:

```
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 [38]:

```
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)
```

Out[38]:

```
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 [42]:

```
test.groupby(['customer_age']).count()
```

Out[42]:

	customer_id	product_id	basket_date	basket_count	sex	tenure
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
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 [43]:

```
cor=data1.corr()
cor
```

```
/tmp/ipykernel_9186/870474124.py:1: FutureWarning: The default value o
f numeric_only in DataFrame.corr is deprecated. In a future version, i
t will default to False. Select only valid columns or specify the valu
e of numeric_only to silence this warning.
  cor=data1.corr()
```

Out[43]:

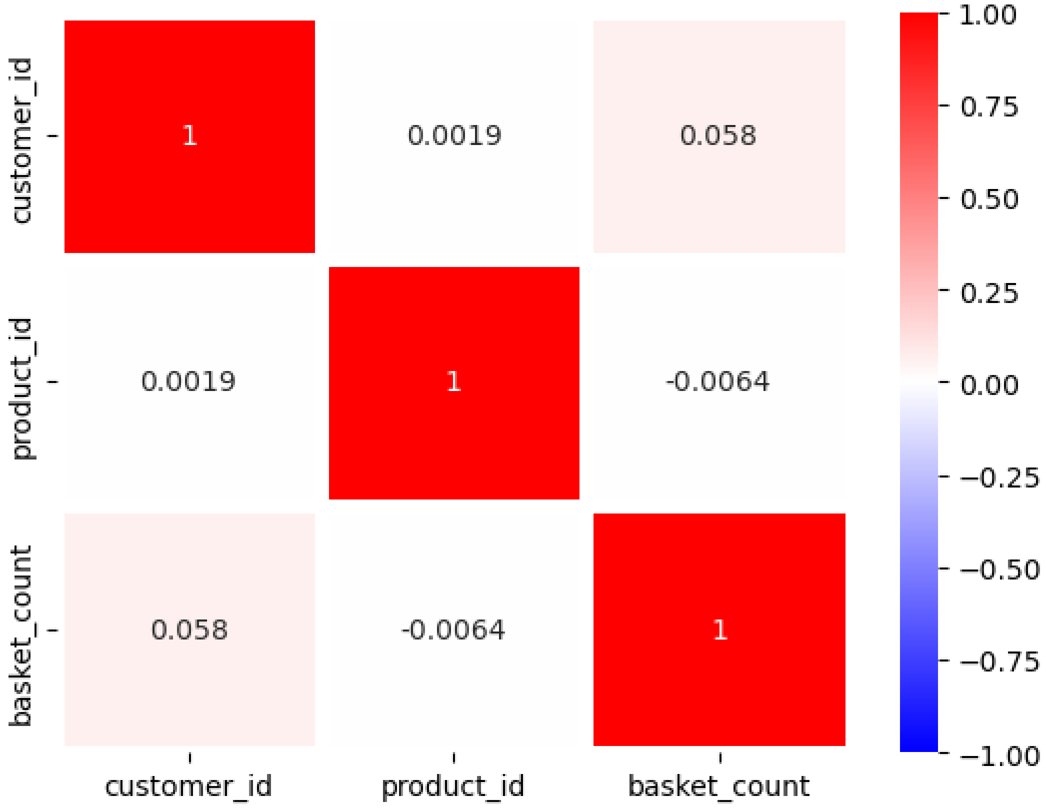
	customer_id	product_id	basket_count
customer_id	1.000000	0.001937	0.058235
product_id	0.001937	1.000000	-0.006407
basket_count	0.058235	-0.006407	1.000000

In [47]:

```
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='bwr')
```

Out[47]:

```
<Axes: >
```



In [48]:

```
cor=tharun.corr()
cor
```

```
/tmp/ipykernel_9186/1326857595.py:1: FutureWarning: The default value
of numeric_only in DataFrame.corr is deprecated. In a future version,
it will default to False. Select only valid columns or specify the val
ue of numeric_only to silence this warning.
  cor=tharun.corr()
```

Out[48]:

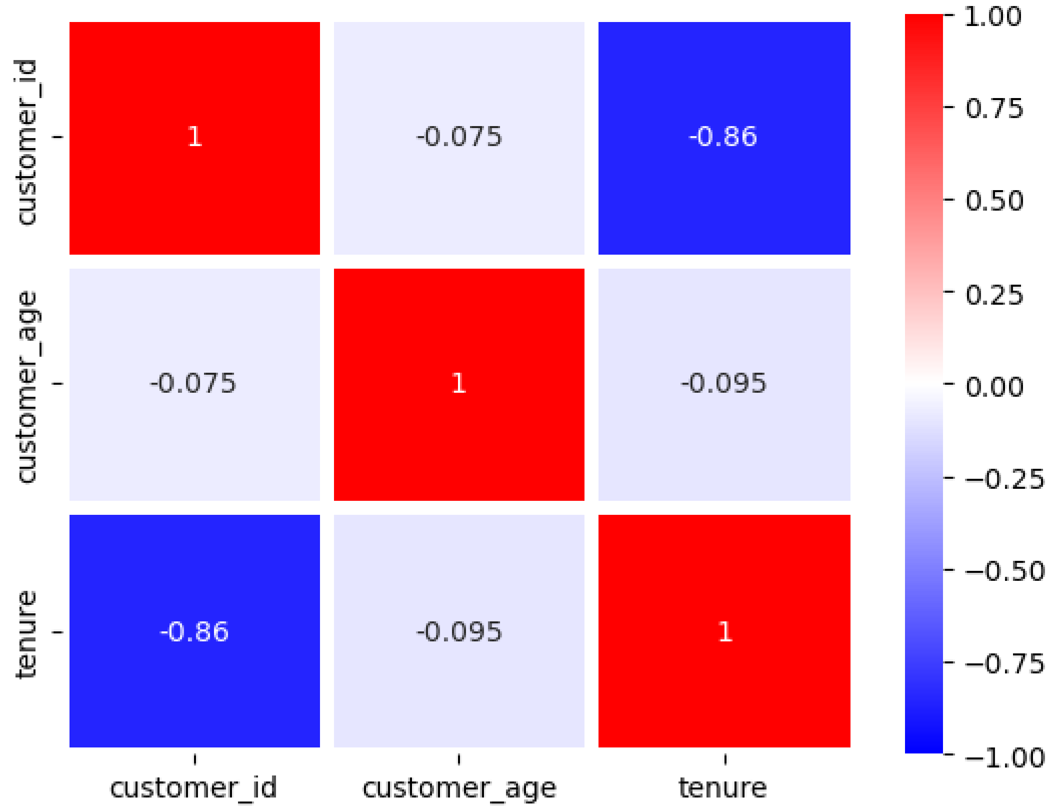
	customer_id	customer_age	tenure
customer_id	1.000000	-0.075467	-0.855410
customer_age	-0.075467	1.000000	-0.095013
tenure	-0.855410	-0.095013	1.000000

In [49]:

```
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='bwr')
```

Out[49]:

```
<Axes: >
```



In [50]:

```
cor=test.corr()
cor
```

```
/tmp/ipykernel_9186/2206162927.py:1: FutureWarning: The default value
of numeric_only in DataFrame.corr is deprecated. In a future version,
it will default to False. Select only valid columns or specify the val
ue of numeric_only to silence this warning.
  cor=test.corr()
```

Out[50]:

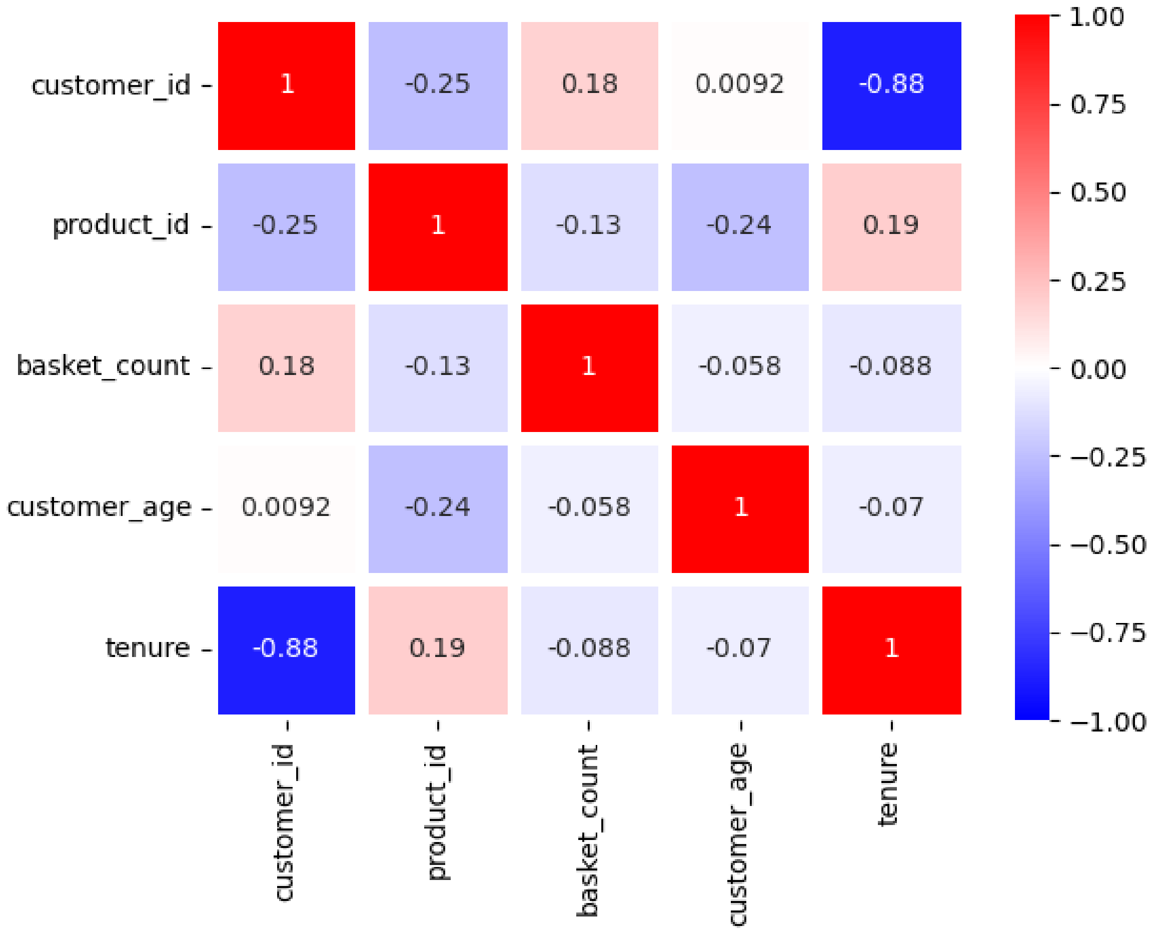
	customer_id	product_id	basket_count	customer_age	tenure
customer_id	1.000000	-0.252572	0.179558	0.009194	-0.882379
product_id	-0.252572	1.000000	-0.125352	-0.243038	0.190134
basket_count	0.179558	-0.125352	1.000000	-0.058177	-0.087821
customer_age	0.009194	-0.243038	-0.058177	1.000000	-0.069814
tenure	-0.882379	0.190134	-0.087821	-0.069814	1.000000

In [51]:

```
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='bwr')
```

Out[51]:

```
<Axes: >
```


customer_id
product_id
basket_count
customer_age
tenure
customer_id
product_id
basket_count
customer_age
tenure