#16/june/23 #saimohan

#basket and customer csv files

both files we can read and analize

In [59]:
```
import pandas as pd
```

In [60]:
```
!pip3 install seaborn
```

```
Requirement already satisfied: cycler>=0.10 in /home/placement/anaconda3/lib/python3.10/site-packages (fr
om matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in /home/placement/anaconda3/lib/python3.10/site-package
s (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: pyparsing>=2.3.1 in /home/placement/anaconda3/lib/python3.10/site-packages
(from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: packaging>=20.0 in /home/placement/anaconda3/lib/python3.10/site-packages
(from matplotlib!=3.6.1,>=3.1->seaborn) (22.0)
Requirement already satisfied: pillow>=6.2.0 in /home/placement/anaconda3/lib/python3.10/site-packages (f
rom matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: contourpy>=1.0.1 in /home/placement/anaconda3/lib/python3.10/site-packages
(from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: kiwisolver>=1.0.1 in /home/placement/anaconda3/lib/python3.10/site-package
s (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: python-dateutil>=2.7 in /home/placement/anaconda3/lib/python3.10/site-pack
ages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /home/placement/anaconda3/lib/python3.10/site-packages (fr
om pandas>=0.25->seaborn) (2022.7)
Requirement already satisfied: six>=1.5 in /home/placement/anaconda3/lib/python3.10/site-packages (from p
```

In [61]:
```
data=pd.read_csv("/home/placement/Desktop/saimohan/csv files/customer_details.csv")
data1=pd.read_csv("/home/placement/Desktop/saimohan/csv files/basket_details.csv")
```

In [62]:
```
data1.describe()
```

Out[62]:

| | 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 [63]:
```
data.describe()
```

Out[63]:

| | 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 [64]:
```
data.shape
```

Out[64]:
```
(20000, 4)
```

In [65]:
```
data1.shape
```

Out[65]:
```
(15000, 4)
```

In [66]:
```
data.tail
```

Out[66]:
```
<bound method NDFrame.tail of        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 x 4 columns]>
```

In [67]:
```
data1.tail
```

Out[67]:
```
<bound method NDFrame.tail of        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 x 4 columns]>
```

In [68]:
```
data1.groupby(['customer_id']).count()
```

Out[68]:

| 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 [69]:
```
data.groupby(['customer_id']).count()
```

Out[69]:

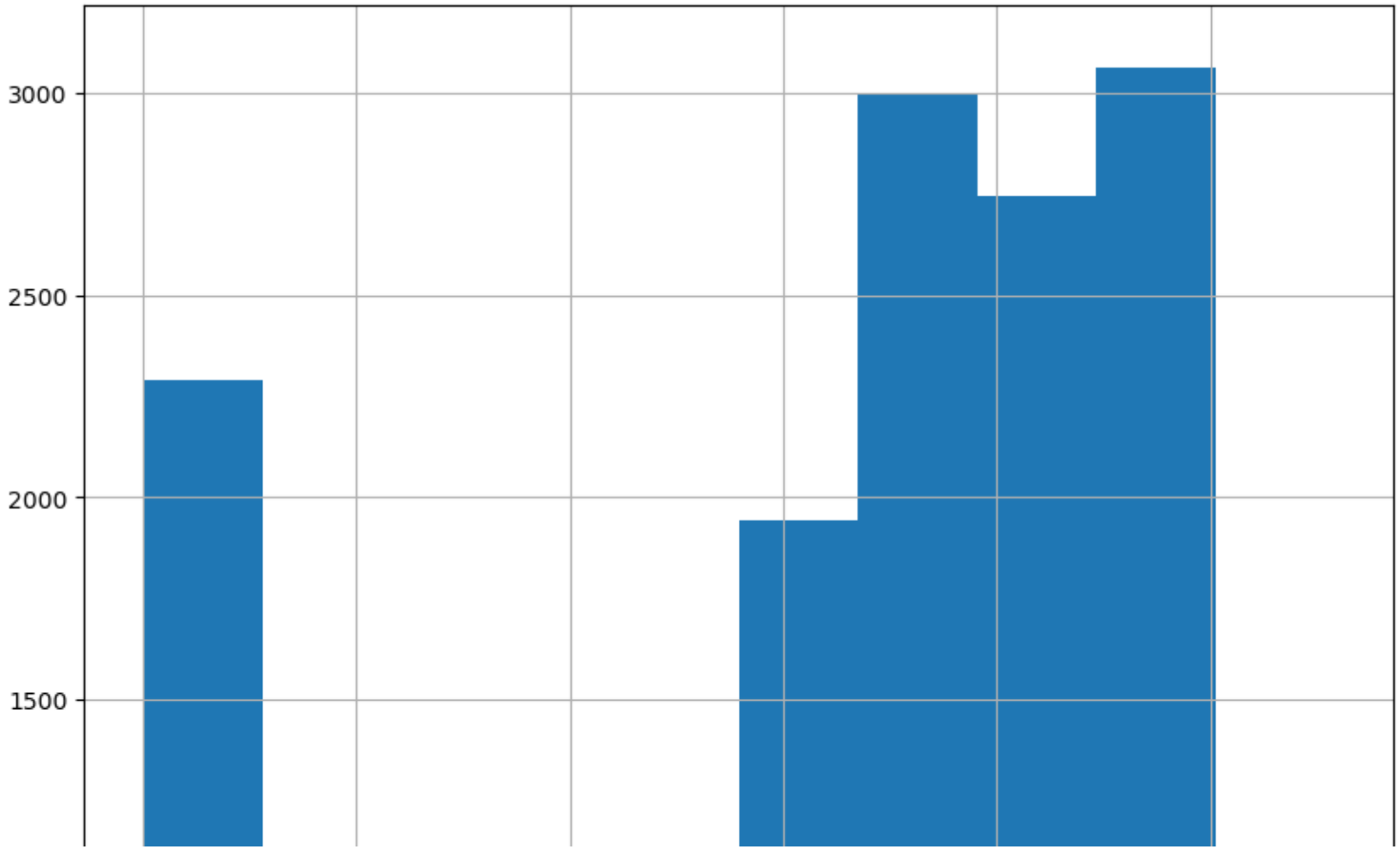
| customer_id | sex | customer_age | tenure |
|---|---|---|---|
| **2093** | 1 | 1 | 1 |
| **12817** | 1 | 1 | 1 |
| **14309** | 1 | 1 | 1 |
| **15155** | 1 | 1 | 1 |
| **23205** | 1 | 1 | 1 |
| **...** | ... | ... | ... |
| **44392831** | 1 | 1 | 1 |
| **44401175** | 1 | 1 | 1 |
| **44431821** | 1 |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

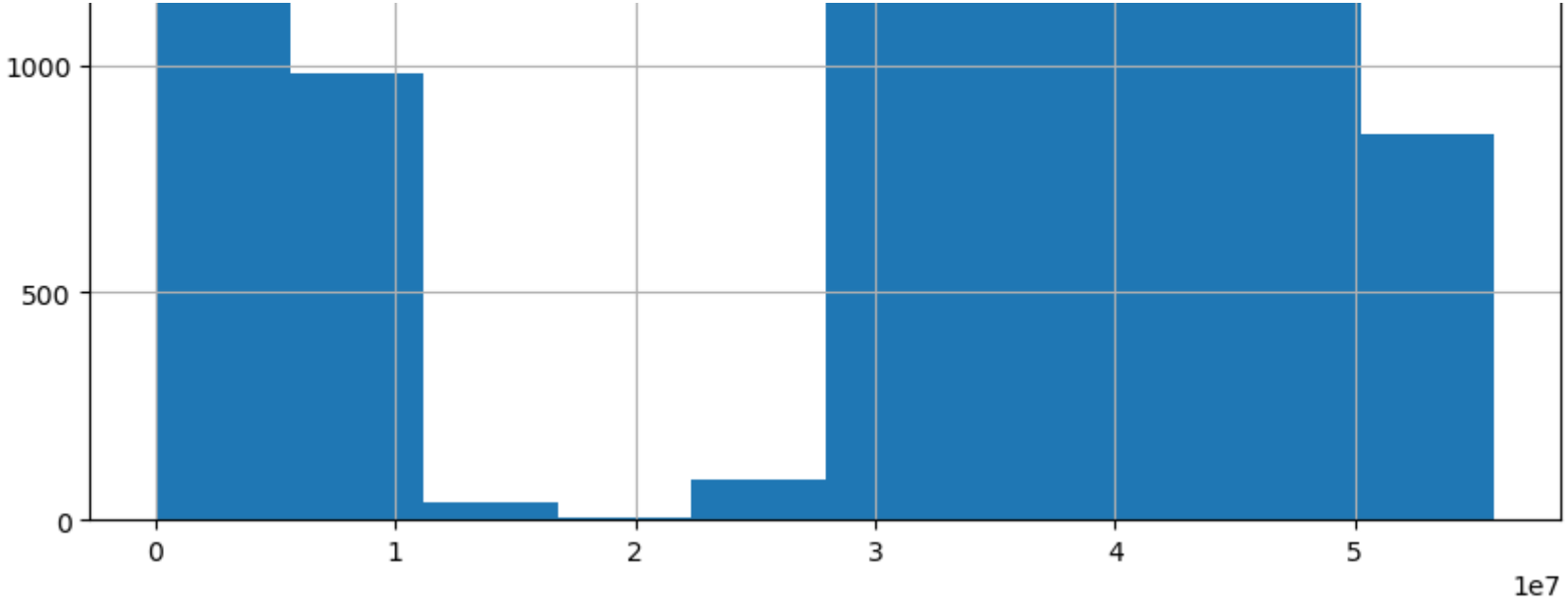
20000 rows × 3 columns

In [70]:
```
data1['product_id'].hist(figsize=(10,10))
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[70], line 2
      1 data1['product_id'].hist(figsize=(10,10))
----> 2 plt.show()

NameError: name 'plt' is not defined
```






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

In [ ]:
```
test
```

In [ ]:
```
test.head
```

In [ ]:
```
test.describe()
```

In [ ]:
```
test.customer_id.unique()
```

In [ ]:
```
data1.head
```

In [73]:
```
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
#true means ascending order
#false means decending order
```

Out[73]:
```
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 [74]:
```
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)
#true means ascending order
#false means decending order
```

Out[74]:
```
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 [ ]:
```
#true means ascending order
#false means decending order
```

In [71]:
```
test.groupby(['customer_age']).count()
```

Out[71]:

| customer_age | customer_id | sex | tenure | product_id | basket_date | basket_count |
|---|---|---|---|---|---|---|
| **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_age | customer_id | sex | tenure | product_id | basket_date | basket_count |
|---|---|---|---|---|---|---|
| **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 [75]:
```
#-------------------------------------------------------------
import seaborn as sns
```

In [ ]:
```
```

In [76]:
```
cor=data1.corr()
cor
```

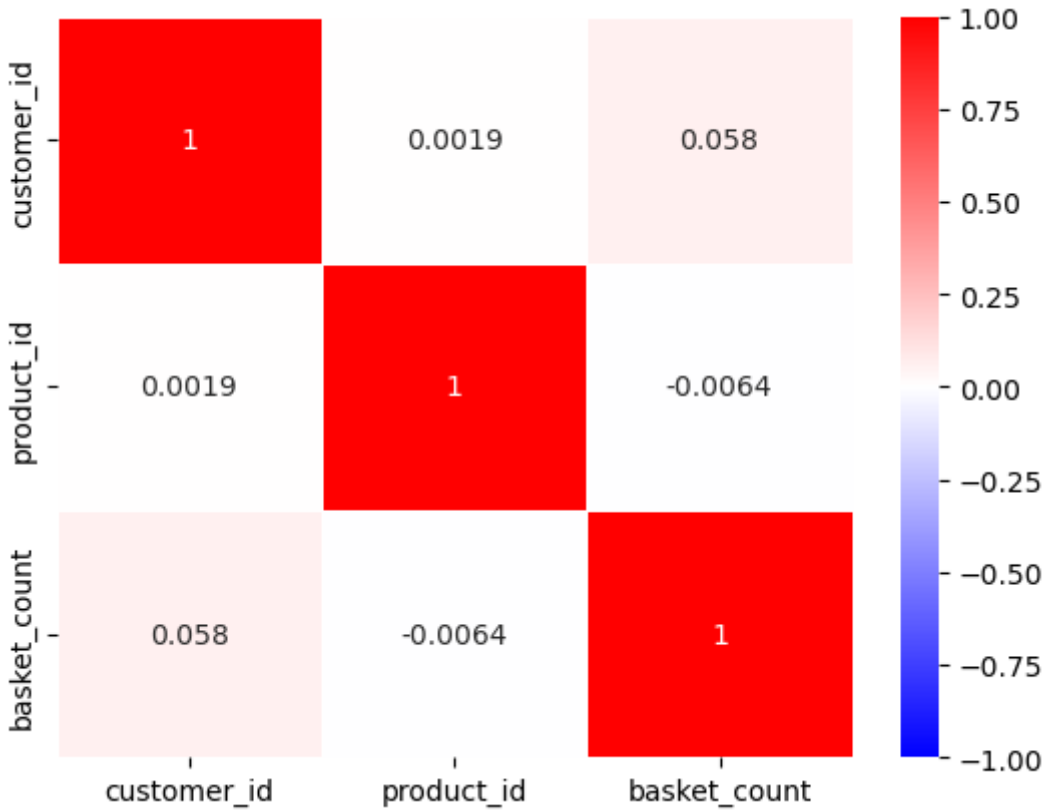
```
/tmp/ipykernel_4415/870474124.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is d
eprecated. In a future version, it will default to False. Select only valid columns or specify the value of
numeric_only to silence this warning.
  cor=data1.corr()
```

Out[76]:

| | 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 [79]:
```
#-------------------------------------------------------------
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=.5,cmap='bwr')
```

Out[79]: <Axes: >




In [ ]: