

In [13]:

```
import pandas as pd
b=pd.read_csv('/home/placement/Desktop/college training/basket_details.csv')
```

In [14]:

```
import seaborn as se #importing desborn
```

In [15]:

```
!pip3 install seaborn #to install the seaborn
```

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

```
In [16]: b.describe()
```

```
Out[16]:
```

	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 [28]: b.head(10) #to describe the head
```

```
Out[28]:
```

	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
5	32037116	33739394	2019-06-19	2
6	17565651	46000191	2019-06-19	2
7	42079380	46881033	2019-06-19	2
8	25533477	44752779	2019-06-19	2
9	10385144	41882886	2019-06-19	2

## # CUSTOMER DETAILS

```
In [18]: import pandas as pd  
c=pd.read_csv('/home/placement/Desktop/college training/customer_details.csv')
```

```
In [19]: c.describe()
```

Out[19]:

	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 [29]: `c.head(10)`

Out[29]:

	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
5	11248447	Male	37.0	68
6	819721	Male	46.0	129
7	4713723	Male	35.0	115
8	11141669	Male	36.0	69
9	10844015	Male	37.0	73

```
In [21]: d=c.loc[(c.sex=='Female')]  
d
```

Out[21]:

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

4669 rows × 4 columns

```
In [22]: s=c.groupby(["customer_id"]).count()  
s
```

#groupby

Out[22]:

	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 [23]: d=s.sort_values('customer_age') | s.sort_values('customer_id')
d
```

Out[23]:

	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 [24]: b['product_id'].hist(figsize=(20,10))    # to plot the histographic graph  
plt.show()
```

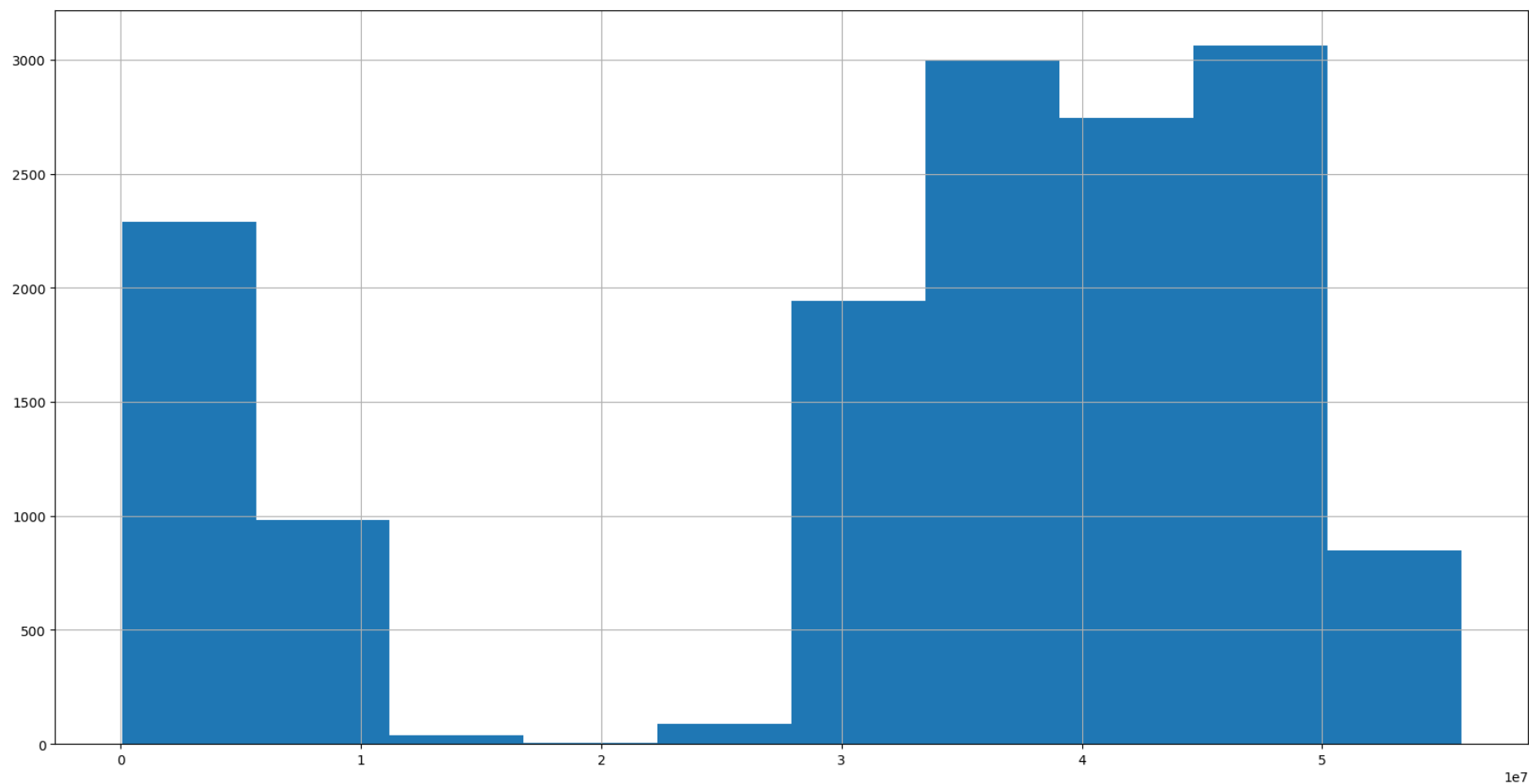
-----  
NameError

Traceback (most recent call last)

Cell In[24], line 2

```
1 b['product_id'].hist(figsize=(20,10))    # to plot the histographic graph  
----> 2 plt.show()
```

NameError: name 'plt' is not defined





```
In [25]: test=pd.merge(c,b,on ='customer_id')           # merge the two files
```

```
In [26]: test.describe()
```

Out[26]:

	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 [27]: b.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
#b.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)
```

```
Out[27]: 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
```