In [17]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
# importing the required files to ANALYZE THE DATA
```

In [18]:
```
data=pd.read_csv("/home/placement/Desktop/nio/basket_details.csv")# reading the data from file 1
```

In [19]:
```
data1=pd.read_csv("/home/placement/Desktop/nio/customer_details.csv")# reding the dtaa from file 2
```

In [20]:
```
data.describe()# describing the data from file 1
```

Out[20]:

	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 [21]:
```
data1.describe()# describing the data from file 2
```

Out[21]:

	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 [22]:
```
list(data)# it tells us about the columns the file had
```

Out[22]:
```
['customer_id', 'product_id', 'basket_date', 'basket_count']
```

In [23]:
```
list(data1)
```

Out[23]:
```
['customer_id', 'sex', 'customer_age', 'tenure']
```

In [24]:
```
data1.shape #just giv efirm number about rows and cols
```

Out[24]:
```
(20000, 4)
```

In [25]:
```
data.shape
```

Out[25]:
```
(15000, 4)
```

In [26]:
```
data#DESCRIBE THE TOTAL DATA
```

Out[26]:

	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 [27]:
```
data1
```

Out[27]:

	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 [28]:
```
import pandas as pd
import numpy as np
data1.groupby(['customer_id']).count()
```

Out[28]:

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

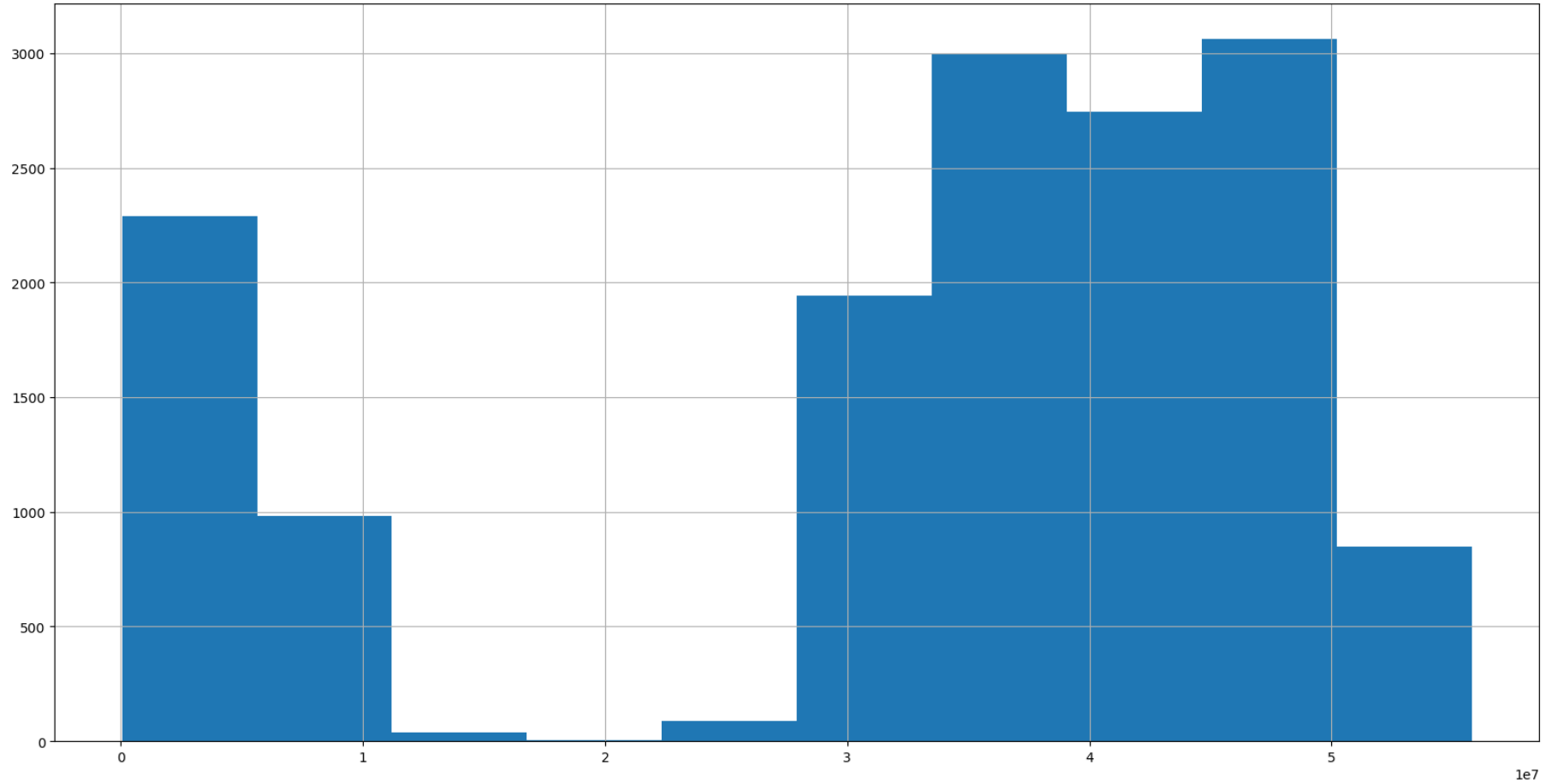
In [29]:
```
data.groupby(['customer_id']).count() #
```

Out[29]:

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 [47]:
```
data['product_id'].hist(figsize=(20,10))
plt.show()# ploting the histogram
```



In [ ]:

In [31]:
```
test=pd.merge(data,data1,on='customer_id')
```

In [32]:
```
test
```

Out[32]:

	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 [33]:
```
test.shape
```

Out[33]:
```
(72, 7)
```

In [34]:
```
dat2=test.loc[(test.sex=='Female')]
```

In [35]:
```
dat2
```

Out[35]:

	customer_id	product_id	basket_date	basket_count	sex	customer_age	tenure
2	11665521	41476812	2019-06-15	2	Female	51.0	62
9	10619833	43524799	2019-05-22	3	Female	32.0	76
10	21765975	6736398	2019-05-22	2	Female	67.0	22
11	21765975	47732556	2019-05-25	2	Female	67.0	22
17	4238087	3213654	2019-06-09	3	Female	43.0	117
28	25055107	39727770	2019-06-04	2	Female	23.0	16
36	29144255	7163984	2019-06-16	2	Female	29.0	13
38	36623391	32252271	2019-06-13	2	Female	41.0	7
42	12901520	38610580	2019-05-28	3	Female	40.0	50
46	15141119	4033598	2019-05-20	2	Female	123.0	38
47	41790413	82875	2019-05-31	2	Female	28.0	4
51	18256077	34222951	2019-05-24	2	Female	42.0	27
53	43280797	35722328	2019-06-10	3	Female	24.0	4
64	34677755	35058279	2019-05-23	3	Female	29.0	8

In [36]:
```
dat2.shape
```

Out[36]: (14, 7)

In [37]:
```
test.describe()
```

Out[37]:

	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 [38]:
```
dat5=dat2.loc[(dat2.customer_age<=28)]
```

In [39]:
```
dat5
```

Out[39]:

	customer_id	product_id	basket_date	basket_count	sex	customer_age	tenure
28	25055107	39727770	2019-06-04	2	Female	23.0	16
47	41790413	82875	2019-05-31	2	Female	28.0	4
53	43280797	35722328	2019-06-10	3	Female	24.0	4

In [40]:
```
test.customer_id.unique()
```

Out[40]:
```
array([ 4897641, 11623549, 11665521,  4193819,  1030589, 20236456,
       15436141, 10394153, 10619833, 21765975, 16029475, 12737235,
       21142247, 15067633,  4238087, 17909829, 11346069, 25567283,
         380975,  4257099, 11440499, 20174063,   537173, 25055107,
       39814593,  9654043, 16398473, 11724853,  4643359,  9700145,
       29144255, 14053193, 36623391, 22524187,  8508353, 12901520,
       20789769, 16944627, 23179191, 15141119, 41790413, 27081691,
        9804585, 18256077,  4912369, 43280797,  9500953, 12410433,
        9875271,   851739, 10439331, 13776147, 11072047, 15570891,
       14966315, 10814041, 34677755, 17830393, 13278573, 12574807,
       15192667, 14248059, 10629563, 11737579])
```

In [41]:
```
data.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
```

Out[41]:
```
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 [42]:
```
data.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)
```

Out[42]:
```
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 [43]:
```
test.groupby(['customer_age']).count()
```

Out[43]:

customer_age	customer_id	product_id	basket_date	basket_count	sex	tenure
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	product_id	basket_date	basket_count	sex	tenure
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 [44]:
```
test.groupby(['customer_age']).count()
```

Out[44]:

customer_age	customer_id	product_id	basket_date	basket_count	sex	tenure
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	product_id	basket_date	basket_count	sex	tenure
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 [ ]:

In [ ]:

In [ ]: