In [1]:
```python
import pandas as pd
```

In [6]:
```python
data=pd.read_csv("/home/placement/Desktop/csv/customer_details.csv")
```

In [7]:
```python
data1=pd.read_csv("/home/placement/Desktop/csv/basket_details.csv")
```

In [11]:
```python
data.describe()
```

Out[11]:

	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 [9]: `data1.describe()`

Out[9]:

	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 [12]: `data.tail()`

Out[12]:

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

Out[15]:

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 [16]:
```python
data1.groupby(['customer_id']).count()
```

Out[16]:

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 [19]:
```python
data1['product_id'].hist(figsize=(20,10))
```

Out[19]: <Axes: >
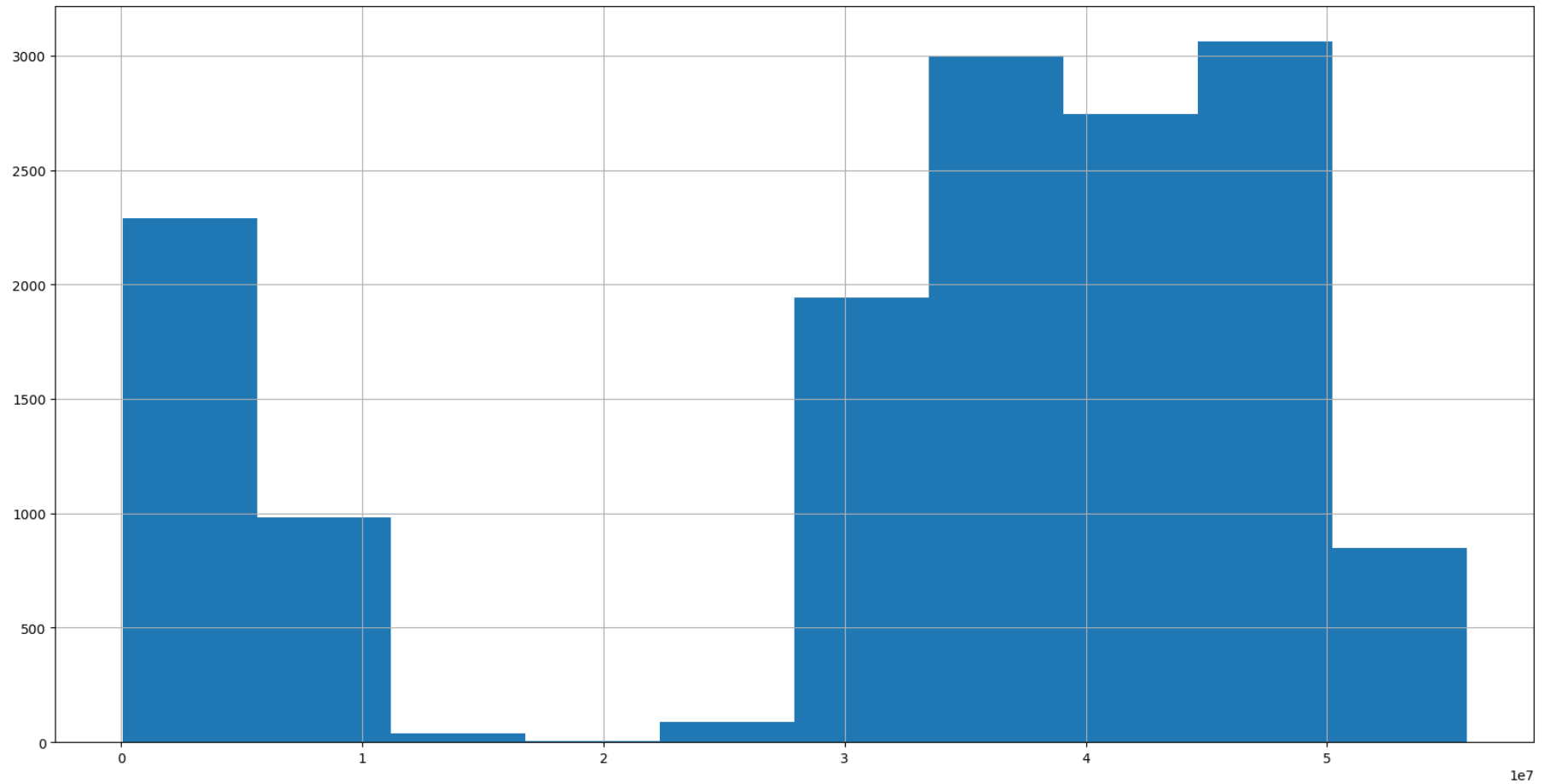
test=pd.merge(data,data1, on = 'customer_id)'

```
In [24]: test=pd.merge(data,data1, on = "customer_id")
         test
```

Out[24]:

	customer_id	sex	customer_age	tenure	product_id	basket_date	basket_count
0	380975	1	1	1	46524972	2019-05-27	2
1	380975	1	1	1	39671495	2019-05-27	2
2	537173	1	1	1	25795609	2019-06-09	2
3	537173	1	1	1	49217445	2019-06-09	2
4	851739	1	1	1	32920704	2019-06-19	2
...	...	...	...	...	...	...	...
67	36623391	1	1	1	32252271	2019-06-13	2
68	39814593	1	1	1	6616058	2019-06-08	3
69	39814593	1	1	1	34219300	2019-05-24	2
70	41790413	1	1	1	82875	2019-05-31	2
71	43280797	1	1	1	35722328	2019-06-10	3

72 rows × 7 columns

In [25]: `test.head()`

Out[25]:

	customer_id	sex	customer_age	tenure	product_id	basket_date	basket_count
0	380975	1	1	1	46524972	2019-05-27	2
1	380975	1	1	1	39671495	2019-05-27	2
2	537173	1	1	1	25795609	2019-06-09	2
3	537173	1	1	1	49217445	2019-06-09	2
4	851739	1	1	1	32920704	2019-06-19	2

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

Out[26]:

	customer_id	sex	customer_age	tenure	product_id	basket_count
count	7.200000e+01	72.0	72.0	72.0	7.200000e+01	72.000000
mean	1.554364e+07	1.0	1.0	1.0	3.140376e+07	2.152778
std	9.961282e+06	0.0	0.0	0.0	1.616160e+07	0.362298
min	3.809750e+05	1.0	1.0	1.0	8.287500e+04	2.000000
25%	1.026443e+07	1.0	1.0	1.0	2.980404e+07	2.000000
50%	1.352736e+07	1.0	1.0	1.0	3.498005e+07	2.000000
75%	2.037478e+07	1.0	1.0	1.0	4.359420e+07	2.000000
max	4.328080e+07	1.0	1.0	1.0	5.130767e+07	3.000000

In [27]: `test.customer_id.unique()`

Out[27]:
```
array([  380975,    537173,    851739,   1030589,   4193819,   4238087,
         4257099,   4643359,   4897641,   4912369,   8508353,   9500953,
         9654043,   9700145,   9804585,   9875271,  10394153,  10439331,
        10619833,  10629563,  10814041,  11072047,  11346069,  11440499,
        11623549,  11665521,  11724853,  11737579,  12410433,  12574807,
        12737235,  12901520,  13278573,  13776147,  14053193,  14248059,
        14966315,  15067633,  15141119,  15192667,  15436141,  15570891,
        16029475,  16398473,  16944627,  17830393,  17909829,  18256077,
        20174063,  20236456,  20789769,  21142247,  21765975,  22524187,
        23179191,  25055107,  25567283,  27081691,  29144255,  34677755,
        36623391,  39814593,  41790413,  43280797])
```

In [28]: `data1.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

In [30]:
```python
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
```

Out[30]:
```
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 [31]:
```python
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)
#false=decending order true=ascending order
```

Out[31]:
```
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 [37]:
```python
test.groupby(['customer_age']).count()
```

Out[37]:

customer_age	customer_id	sex	tenure	product_id	basket_date	basket_count
1	72	72	72	72	72	72

In [ ]: