In [16]:
```
import pandas as pd
```

In [17]:
```
data1=pd.read_csv("basket_details.csv")
```

In [18]:
```
data=pd.read_csv("customer_details.csv")
```

In [19]:
```
data.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 [20]:
```
data1.describe()
```

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]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   customer_id   20000 non-null  int64  
 1   sex           20000 non-null  object 
 2   customer_age  20000 non-null  float64
 3   tenure        20000 non-null  int64  
dtypes: float64(1), int64(2), object(1)
memory usage: 625.1+ KB
```

In [22]:
```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   customer_id   15000 non-null  int64 
 1   product_id    15000 non-null  int64 
 2   basket_date   15000 non-null  object
 3   basket_count  15000 non-null  int64 
dtypes: int64(3), object(1)
memory usage: 468.9+ KB
```

In [23]:
```
data.groupby(['customer_id']).count()
```

Out[23]:

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 [24]:
```
data1.groupby(['customer_id']).count()
```

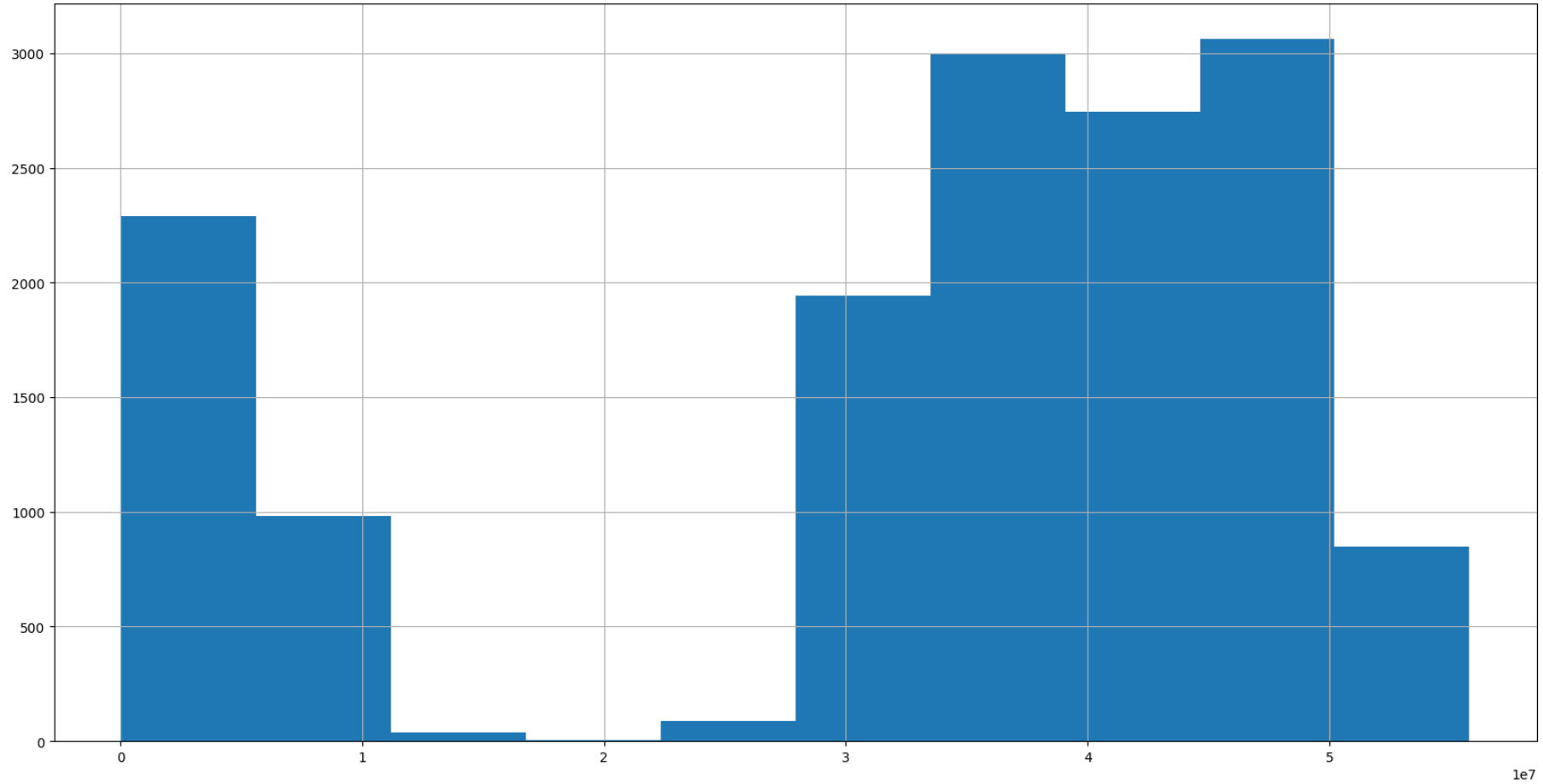
Out[24]:

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 [25]:
```
data1['product_id'].hist(figsize=(20,10))
```

Out[25]: <Axes: >



In [27]:
```
!pip3 install seaborn
```

```
Collecting seaborn
  Downloading seaborn-0.12.2-py3-none-any.whl (293 kB)
     |████████████████████████████████| 293 kB 2.3 MB/s eta 0:00:01
Requirement already satisfied: pandas>=0.25 in /home/placement/.local/lib/python3.8/site-packages (from sea
born) (2.0.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in /home/placement/.local/lib/python3.8/site-packages
(from seaborn) (1.24.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in /home/placement/.local/lib/python3.8/site-package
s (from seaborn) (3.7.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /home/placement/.local/lib/python3.8/site-packages
(from pandas>=0.25->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /home/placement/.local/lib/python3.8/site-packages (from pan
das>=0.25->seaborn) (2023.3)
Requirement already satisfied: tzdata>=2022.1 in /home/placement/.local/lib/python3.8/site-packages (from p
andas>=0.25->seaborn) (2023.3)
Requirement already satisfied: importlib-resources>=3.2.0; python_version < "3.10" in /home/placement/.loca
l/lib/python3.8/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (5.12.0)
Requirement already satisfied: fonttools>=4.22.0 in /home/placement/.local/lib/python3.8/site-packages (fro
m matplotlib!=3.6.1,>=3.1->seaborn) (4.40.0)
Requirement already satisfied: packaging>=20.0 in /home/placement/.local/lib/python3.8/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pyparsing>=2.3.1 in /home/placement/.local/lib/python3.8/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: contourpy>=1.0.1 in /home/placement/.local/lib/python3.8/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (1.1.0)
Requirement already satisfied: cycler>=0.10 in /home/placement/.local/lib/python3.8/site-packages (from mat
plotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in /usr/lib/python3/dist-packages (from matplotlib!=3.6.1,>=3.
1->seaborn) (7.0.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /home/placement/.local/lib/python3.8/site-packages (fro
m matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: six>=1.5 in /usr/lib/python3/dist-packages (from python-dateutil>=2.8.2->pan
das>=0.25->seaborn) (1.14.0)
Requirement already satisfied: zipp>=3.1.0; python_version < "3.10" in /home/placement/.local/lib/python3.
8/site-packages (from importlib-resources>=3.2.0; python_version < "3.10"->matplotlib!=3.6.1,>=3.1->seabor
n) (3.15.0)
Installing collected packages: seaborn
Successfully installed seaborn-0.12.2
```

In [28]:
```
import pandas as pd
```

In [29]:
```
import numpy as np
```

In [30]:
```
import matplotlib.pyplot as plt
```

In [31]:
```
import seaborn as sns
```

In [32]:
```
test=pd.merge(data,data1,on='customer_id')
#test=pd.merge(data,data1)
```

In [33]:
```
test
```

Out[33]:

	customer_id	sex	customer_age	tenure	product_id	basket_date	basket_count
0	9500953	Male	55.0	96	3446783	2019-06-10	3
1	851739	Male	40.0	129	32920704	2019-06-19	2
2	9654043	Male	37.0	95	51307669	2019-06-08	2
3	4912369	Male	36.0	114	33923115	2019-05-20	2
4	9875271	Male	34.0	92	31586037	2019-06-06	2
...	...	...	...	...	...	...	...
67	13278573	Male	28.0	47	4488682	2019-05-26	2
68	12901520	Female	40.0	50	38610580	2019-05-28	3
69	12737235	Male	39.0	51	32933848	2019-05-21	2
70	12737235	Male	39.0	51	46373374	2019-05-21	3
71	12574807	Male	33.0	52	32056122	2019-05-25	2

72 rows × 7 columns

In [35]:
```
data1
```

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
...	...	...	...	...
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 [36]:
```
test.describe()
```

Out[36]:

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

In [37]:
```
test.customer_id.unique()
```

Out[37]:
```
array([ 9500953,   851739,  9654043,  4912369,  9875271, 11737579,
       10619833,  4193819,  4897641,  4643359,   380975, 11623549,
       11724853, 12410433, 10394153,   537173, 11440499, 10439331,
       10629563,  4257099, 11346069,  8508353,  9700145, 10814041,
        9804585,  4238087, 11665521,  1030589, 11072047, 43280797,
       41790413, 39814593, 36623391, 34677755, 29144255, 27081691,
       25055107, 25567283, 23179191, 22524187, 21765975, 21142247,
       20789769, 20236456, 20174063, 17909829, 18256077, 17830393,
       16944627, 16398473, 16029475, 15436141, 15570891, 15192667,
       15067633, 14966315, 15141119, 14248059, 14053193, 13776147,
       13278573, 12901520, 12737235, 12574807])
```

In [39]:
```
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
```

Out[39]:
```
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 [40]:
```
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=True)
```

Out[40]:
```
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 [48]:
```
test.groupby('customer_age').count()
```

Out[48]:

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 [46]:
```
test.head()
```

Out[46]:

	customer_id	sex	customer_age	tenure	product_id	basket_date	basket_count
0	9500953	Male	55.0	96	3446783	2019-06-10	3
1	851739	Male	40.0	129	32920704	2019-06-19	2
2	9654043	Male	37.0	95	51307669	2019-06-08	2
3	4912369	Male	36.0	114	33923115	2019-05-20	2
4	9875271	Male	34.0	92	31586037	2019-06-06	2

In [42]:
```
import seaborn as sns
```