In [9]: `import pandas as pd`

In [10]: `data=pd.read_csv("/home/placement/Downloads/customer_details.csv")`

In [11]: `data1=pd.read_csv("/home/placement/Downloads/basket_details.csv")`

In [14]: `data+data1`

Out[14]:

	basket_count	basket_date	customer_age	customer_id	product_id	sex	tenure
0	NaN	NaN	NaN	52165444.0	NaN	NaN	NaN
1	NaN	NaN	NaN	47370404.0	NaN	NaN	NaN
2	NaN	NaN	NaN	26957773.0	NaN	NaN	NaN
3	NaN	NaN	NaN	15311262.0	NaN	NaN	NaN
4	NaN	NaN	NaN	30139723.0	NaN	NaN	NaN
...	...	...	...	...	...	...	...
19995	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19996	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19997	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19998	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19999	NaN	NaN	NaN	NaN	NaN	NaN	NaN

20000 rows × 7 columns

In [15]: `data`

Out[15]:

	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 [16]: `data1`

Out[16]:

	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 [19]: `data*data1`

Out[19]:

	basket_count	basket_date	customer_age	customer_id	product_id	sex	tenure
0	NaN	NaN	NaN	4.151442e+14	NaN	NaN	NaN
1	NaN	NaN	NaN	4.103957e+14	NaN	NaN	NaN
2	NaN	NaN	NaN	2.138727e+13	NaN	NaN	NaN
3	NaN	NaN	NaN	3.930692e+13	NaN	NaN	NaN
4	NaN	NaN	NaN	2.022324e+14	NaN	NaN	NaN
...	...	...	...	...	...	...	...
19995	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19996	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19997	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19998	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19999	NaN	NaN	NaN	NaN	NaN	NaN	NaN

20000 rows × 7 columns

In [20]: `data.tail()`

Out[20]:

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

Out[21]:

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

Out[22]:

	product_id	basket_date	basket_count
customer_id			
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 [23]: `data1.tail()`

Out[23]:

	customer_id	product_id	basket_date	basket_count
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

In [24]: `data1['product_id'].hist(figsize=(20,10))`

Out[24]: `<Axes: >`
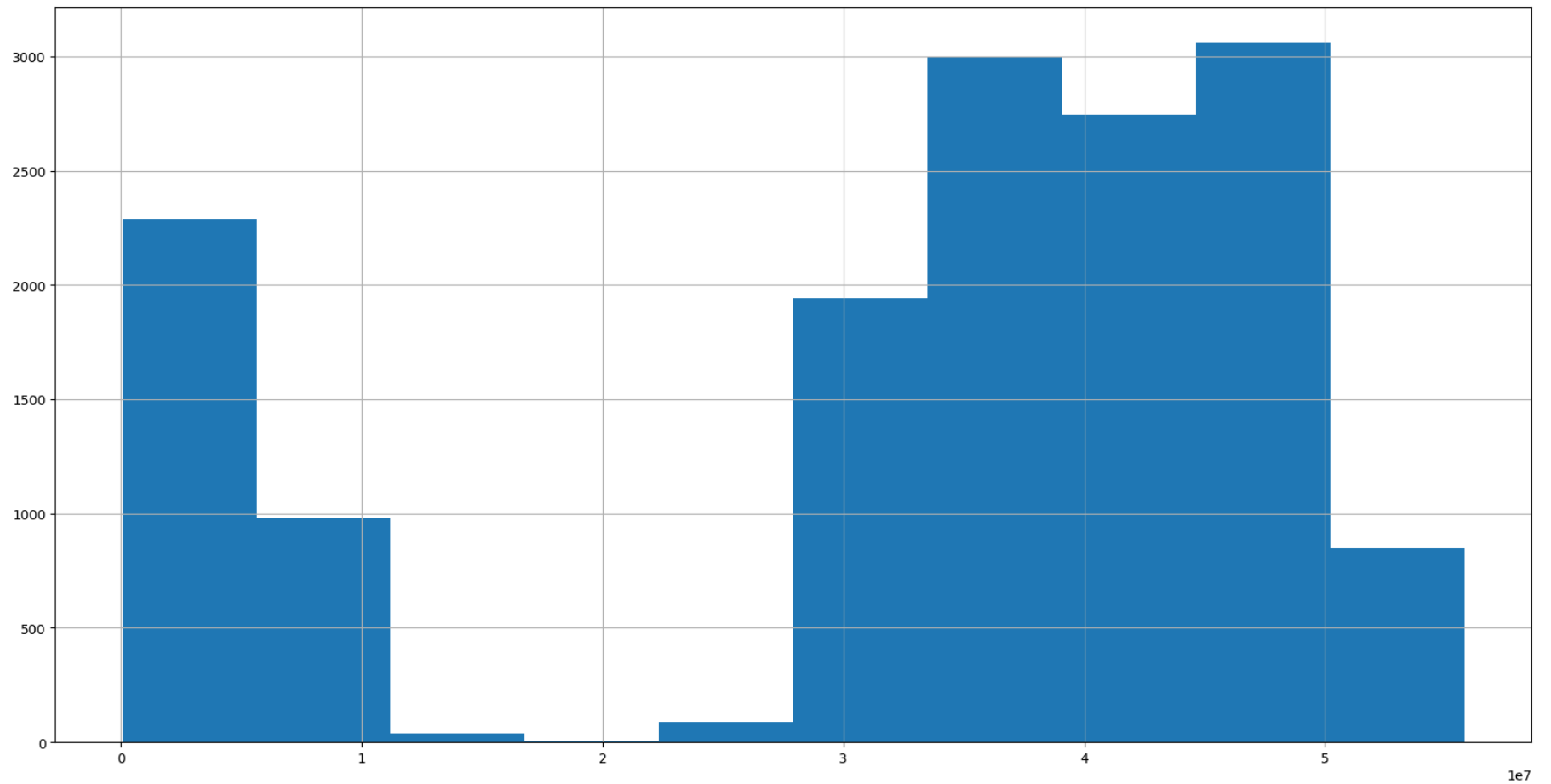
In [25]:
```python
test=pd.merge(data, data1, on = "customer_id")
test
```

Out[25]:

	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 [26]: `test.describe()`

Out[26]:

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

Out[27]:
```
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 [ ]: