In [1]:
```
import pandas as r
#reading the file into the jupyter
d=r.read_csv("/home/placement/Downloads/customer_details.csv")
```

In [2]:
```
#This command is to describe the data present in the DataFrame in statistically
d.describe()
```

Out[2]:

	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 [3]:
```
'''groupby count method is used to count the values in each group by
ignoring the missing values or NaN values in the data frame.'''
d.groupby(['customer_id']).count()
```

Out[3]:

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 [4]:
```
#reading the file into the jupyter
d1=r.read_csv("/home/placement/Downloads/basket_details.csv")
d1.describe()
```

Out[4]:

	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 [5]:
```
d1.groupby(['customer_id']).count()
```

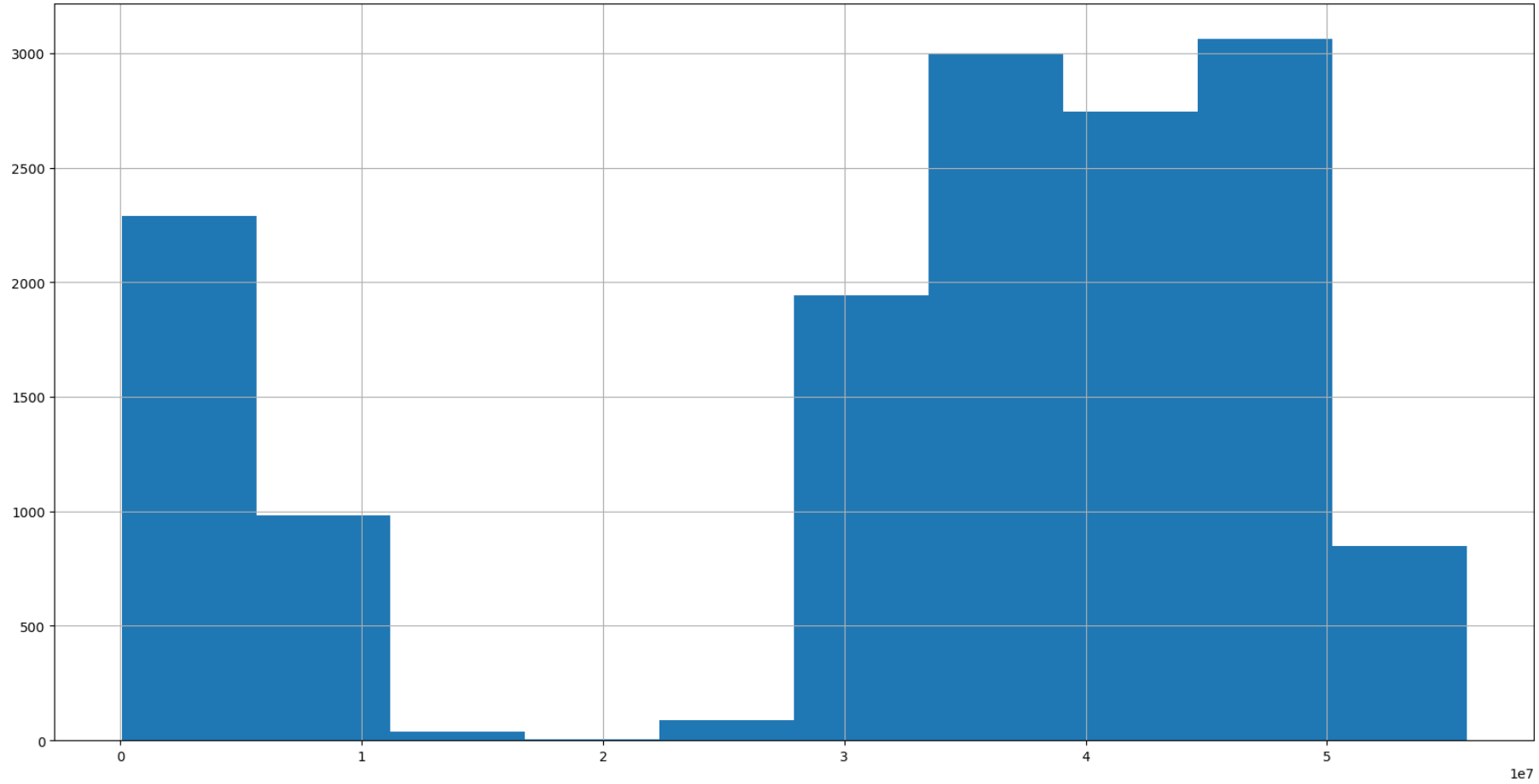
Out[5]:

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 [6]:
```
d1['product_id'].hist(figsize=(20,10))
```

Out[6]: <Axes: >



In [7]:
```
#Merging the two data frames using a column name
test=r.merge(d,d1,on="customer_id")
test
```

Out[7]:

	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 [8]:
```
#grouping the data and summing the basket count and arranging in descending order
d1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
```

Out[8]:
```
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 [9]:
```
'''groupby count method is used to count the values in each group by
ignoring the missing values'''
d.groupby(['customer_age']).count()
```

Out[9]:

customer_age	customer_id	sex	tenure
-34.0	1	1	1
3.0	2	2	2
4.0	1	1	1
5.0	710	710	710
6.0	1	1	1
...	...	...	...
127.0	1	1	1
130.0	1	1	1
139.0	1	1	1
149.0	1	1	1
2022.0	2102	2102	2102

93 rows × 3 columns