In [2]:
```
import numpy as np
```

In [3]:
```
import pandas as pd
```

In [4]:
```
data=pd.read_csv("/home/placement/Downloads/customer_details.csv")
data1=pd.read_csv("/home/placement/Downloads/basket_details.csv")
```

In [5]:
```
data.head()
```

Out[5]:

	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

In [6]:
```
data1.head()
```

Out[6]:

	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 [7]:
```
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 [8]:
```
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 [9]:
```
data.describe()
```

Out[9]:

	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 [10]:
```
data1.describe()
```

Out[10]:

	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 [11]:
```
data.tail()
```

Out[11]:

	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 [12]:
```
data.groupby(['customer_id']).count()
```

Out[12]:

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 [13]:
```
data.groupby(['sex']).count()
```

Out[13]:

sex	customer_id	customer_age	tenure
Female	4669	4669	4669
Male	15322	15322	15322
UNKNOWN	1	1	1
kvkktalepsilindi	8	8	8

In [14]:
```
data.groupby(['customer_age']).count()
```

Out[14]:

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

In [15]:
```
data1.groupby(['customer_id']).count()
```

Out[15]:

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

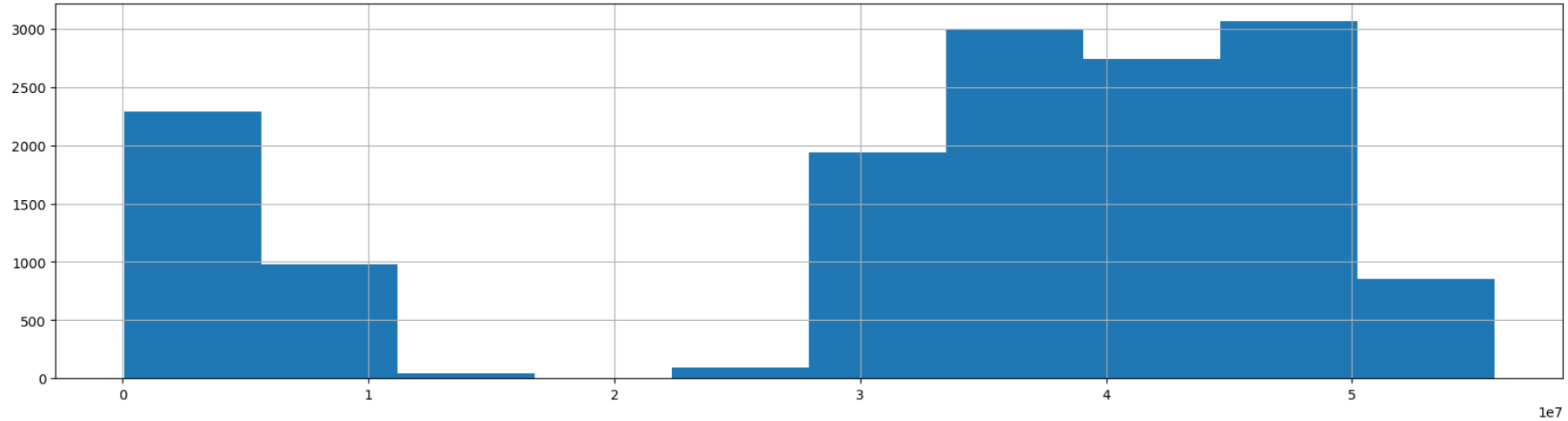
Out[16]:

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 [17]:
```
data1['product_id'].hist(figsize=(20,5))
```

Out[17]: <Axes: >



In [18]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in ./anaconda3/lib/python3.10/site-packages (0.12.2)
Requirement already satisfied: pandas>=0.25 in ./anaconda3/lib/python3.10/site-packages (from seaborn) (1.
5.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in ./anaconda3/lib/python3.10/site-packages (from se
aborn) (3.7.0)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in ./anaconda3/lib/python3.10/site-packages (from seabo
rn) (1.23.5)
Requirement already satisfied: contourpy>=1.0.1 in ./anaconda3/lib/python3.10/site-packages (from matplotli
b!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: fonttools>=4.22.0 in ./anaconda3/lib/python3.10/site-packages (from matplotl
ib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: packaging>=20.0 in ./anaconda3/lib/python3.10/site-packages (from matplotli
b!=3.6.1,>=3.1->seaborn) (22.0)
Requirement already satisfied: python-dateutil>=2.7 in ./anaconda3/lib/python3.10/site-packages (from matpl
otlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in ./anaconda3/lib/python3.10/site-packages (from matplotl
ib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: pillow>=6.2.0 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=
3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: cycler>=0.10 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=
3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: pyparsing>=2.3.1 in ./anaconda3/lib/python3.10/site-packages (from matplotli
b!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: pytz>=2020.1 in ./anaconda3/lib/python3.10/site-packages (from pandas>=0.25-
>seaborn) (2022.7)
Requirement already satisfied: six>=1.5 in ./anaconda3/lib/python3.10/site-packages (from python-dateutil>=
2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:
```

```