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

In [2]:
```
data=pd.read_csv("/home/placement/Downloads/customer_details.csv")
```

In [3]:
```
data1=pd.read_csv("/home/placement/Downloads/basket_details.csv")
```

In [4]:
```
data+data1
```

Out[4]:

| | 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 [5]:
```
data
```

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 |
| **...** | ... | ... | ... | ... |
| **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 [6]:
```
data1
```

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 |
| **...** | ... | ... | ... | ... |
| **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 [7]:
```
data*data1
```

Out[7]:

| | 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 [8]:
```
data-data1
```

Out[8]:

| | basket_count | basket_date | customer_age | customer_id | product_id | sex | tenure |
|---|---|---|---|---|---|---|---|
| **0** | NaN | NaN | NaN | -32567726.0 | NaN | NaN | NaN |
| **1** | NaN | NaN | NaN | -24543278.0 | NaN | NaN | NaN |
| **2** | NaN | NaN | NaN | -25321383.0 | NaN | NaN | NaN |
| **3** | NaN | NaN | NaN | 8786756.0 | NaN | NaN | NaN |
| **4** | NaN | NaN | NaN | -9973633.0 | 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 [9]:
```
data.tail()
```

Out[9]:

| | 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 [11]:
```
data.groupby(['customer_id']).count()
```

Out[11]:

| 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 [12]:
```
data1.groupby(['customer_id']).count()
```

Out[12]:

| 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

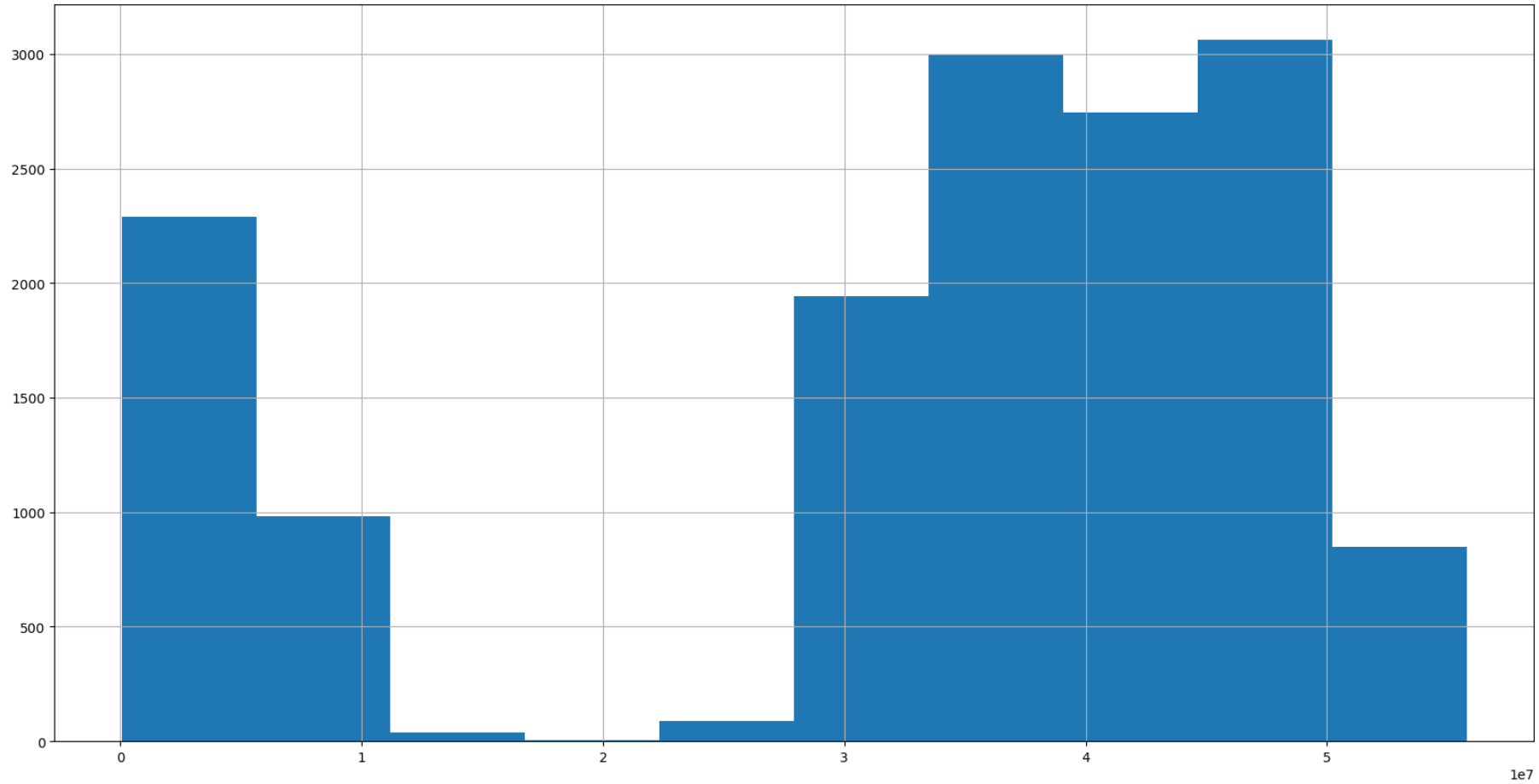
In [13]:
```
data1.tail()
```

Out[13]:

| | 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 [14]:
```
data1['product_id'].hist(figsize=(20,10))
```

Out[14]: <Axes: >



In [16]:
```
test=pd.merge(data, data1, on = "customer_id")
test
```

Out[16]:

| | 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 [17]:
```
test.describe()
```

Out[17]:

| | 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 [18]:
```
test.customer_id.unique()
```

Out[18]:
```
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 [19]:
```
data1.groupby(['product_id'])['basket_count'].sum().sort_values(ascending=False)
```

Out[19]:
```
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 [20]:
```
test.groupby(['customer_age']).count()
```

Out[20]:

| 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 [21]:
```
!pip3 install seaborn
```

```
Requirement already satisfied: seaborn in ./anaconda3/lib/python3.10/site-packages (0.12.2)
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: pandas>=0.25 in ./anaconda3/lib/python3.10/site-packages (from seaborn) (1.
5.3)
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: 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: cycler>=0.10 in ./anaconda3/lib/python3.10/site-packages (from matplotlib!=
3.6.1,>=3.1->seaborn) (0.11.0)
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: 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: packaging>=20.0 in ./anaconda3/lib/python3.10/site-packages (from matplotli
b!=3.6.1,>=3.1->seaborn) (22.0)
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: 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: 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)
```

In [26]:
```
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='dwr')
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[26], line 2
      1 import seaborn as sns
----> 2 sns.heatmap(corll,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='dwr')

NameError: name 'corll' is not defined
```

In [27]:
```
cor=data1.corr()
cor
```

```
/tmp/ipykernel_7247/870474124.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is d
eprecated. In a future version, it will default to False. Select only valid columns or specify the value of
numeric_only to silence this warning.
  cor=data1.corr()
```

Out[27]:

| | customer_id | product_id | basket_count |
|---|---|---|---|
| **customer_id** | 1.000000 | 0.001937 | 0.058235 |
| **product_id** | 0.001937 | 1.000000 | -0.006407 |
| **basket_count** | 0.058235 | -0.006407 | 1.000000 |

In [28]:
```
import seaborn as sns
sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='dwr')
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
Cell In[28], line 2
      1 import seaborn as sns
----> 2 sns.heatmap(cor,vmax=1,vmin=-1,annot=True,linewidths=5,cmap='dwr')

File ~/anaconda3/lib/python3.10/site-packages/seaborn/matrix.py:446, in heatmap(data, vmin, vmax, cmap, c
enter, robust, annot, fmt, annot_kws, linewidths, linecolor, cbar, cbar_kws, cbar_ax, square, xticklabel
s, yticklabels, mask, ax, **kwargs)
    365 """Plot rectangular data as a color-encoded matrix.
    366
    367 This is an Axes-level function and will draw the heatmap into the
   (...)
    443
    444 """
    445 # Initialize the plotter object
--> 446 plotter = _HeatMapper(data, vmin, vmax, cmap, center, robust, annot, fmt,
    447                       annot_kws, cbar, cbar_kws, xticklabels,
    448                       yticklabels, mask)
    450 # Add the pcolormesh kwargs here
    451 kwargs["linewidths"] = linewidths

File ~/anaconda3/lib/python3.10/site-packages/seaborn/matrix.py:163, in _HeatMapper.__init__(self, data,
 vmin, vmax, cmap, center, robust, annot, fmt, annot_kws, cbar, cbar_kws, xticklabels, yticklabels, mask)
    160 self.ylabel = ylabel if ylabel is not None else ""
    162 # Determine good default values for the colormapping
--> 163 self._determine_cmap_params(plot_data, vmin, vmax,
    164                             cmap, center, robust)
    166 # Sort out the annotations
    167 if annot is None or annot is False:

File ~/anaconda3/lib/python3.10/site-packages/seaborn/matrix.py:217, in _HeatMapper._determine_cmap_param
s(self, plot_data, vmin, vmax, cmap, center, robust)
    215         self.cmap = cm.icefire
    216 elif isinstance(cmap, str):
--> 217     self.cmap = get_colormap(cmap)
    218 elif isinstance(cmap, list):
    219     self.cmap = mpl.colors.ListedColormap(cmap)
```

```
File ~/anaconda3/lib/python3.10/site-packages/seaborn/_compat.py:133, in get_colormap(name)
    131 """Handle changes to matplotlib colormap interface in 3.6."""
    132 try:
--> 133     return mpl.colormaps[name]
    134 except AttributeError:
    135     return mpl.cm.get_cmap(name)

File ~/anaconda3/lib/python3.10/site-packages/matplotlib/cm.py:82, in ColormapRegistry.__getitem__(self,
 item)
     80     return self._cmaps[item].copy()
     81 except KeyError:
---> 82     raise KeyError(f"{item!r} is not a known colormap name") from None

KeyError: "'dwr' is not a known colormap name"
```

In [ ]: