Data in Motion Challenge / week 10.12.2022

Pandas challenge

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

In [2]:
```
url = 'https://raw.githubusercontent.com/datasets/investor-flow-of-funds-us/master/data/

df = pd.read_csv(url)
```

In [3]:
```
df.head()
```

Out[3]:

	Date	Total Equity	Domestic Equity	World Equity	Hybrid	Total Bond	Taxable Bond	Municipal Bond	Total
0	2012-12-05	-7426	-6060	-1367	-74	5317	4210	1107	-2183
1	2012-12-12	-8783	-7520	-1263	123	1818	1598	219	-6842
2	2012-12-19	-5496	-5470	-26	-73	103	3472	-3369	-5466
3	2012-12-26	-4451	-4076	-375	550	2610	3333	-722	-1291
4	2013-01-02	-11156	-9622	-1533	-158	2383	2103	280	-8931

In [4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44 entries, 0 to 43
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   Date             44 non-null     object
 1   Total Equity     44 non-null     int64 
 2   Domestic Equity  44 non-null     int64 
 3   World Equity     44 non-null     int64 
 4   Hybrid           44 non-null     int64 
 5   Total Bond       44 non-null     int64 
 6   Taxable Bond     44 non-null     int64 
 7   Municipal Bond   44 non-null     int64 
 8   Total            44 non-null     int64 
dtypes: int64(8), object(1)
memory usage: 3.2+ KB
```

What is the frequency of the dataset? (The time period between each row)

Date type is object. We need to change it datetime </br> create first of all a copy of the dataframe

In [5]:
```
df2=df.copy()
```

In [6]:
```
df2.Date=pd.to_datetime(df2.Date)
```

In [7]:
```
print(df2.Date.dtypes)
```

```
datetime64[ns]
```

In [8]:
```
freq=df2.Date.diff()
```

In [9]:
```
print(freq)
```

```
0         NaT
1      7 days
2      7 days
3      7 days
4      7 days
5      7 days
6    448 days
7      7 days
8      7 days
9      7 days
10     7 days
11     7 days
12     7 days
13     7 days
14     7 days
15     7 days
16     7 days
17     7 days
18     7 days
19     7 days
20     7 days
21    21 days
22     7 days
23     7 days
24     7 days
25     7 days
26     7 days
27     7 days
28    56 days
29     7 days
30     7 days
31     6 days
32    43 days
33     7 days
34     7 days
35     7 days
36     7 days
37     7 days
38    21 days
39     7 days
40     7 days
41     7 days
42     7 days
43     7 days
Name: Date, dtype: timedelta64[ns]
```

What is the data type of the index?

In [10]:
```
df2.index
```

Out[10]:
```
RangeIndex(start=0, stop=44, step=1)
```

In [11]:
```
df2.index.dtype
```

Out[11]:
```
dtype('int64')
```

Set the index to a Datetime.

In [12]:
```
df2.index = pd.to_datetime(df2.index)
```

In [13]:
```
print(df2.index.dtype)
```

```
datetime64[ns]
```

In [14]:
```
df2.head()
```

Out[14]:

	Date	Total Equity	Domestic Equity	World Equity	Hybrid	Total Bond	Taxable Bond	Municipal Bond	Total
1970-01-01 00:00:00.000000000	2012-12-05	-7426	-6060	-1367	-74	5317	4210	1107	-2183
1970-01-01 00:00:00.000000001	2012-12-12	-8783	-7520	-1263	123	1818	1598	219	-6842
1970-01-01 00:00:00.000000002	2012-12-19	-5496	-5470	-26	-73	103	3472	-3369	-5466
1970-01-01 00:00:00.000000003	2012-12-26	-4451	-4076	-375	550	2610	3333	-722	-1291
1970-01-01 00:00:00.000000004	2013-01-02	-11156	-9622	-1533	-158	2383	2103	280	-8931

Change the frequency to monthly, sum the values and assign it to new variable called monthly.

In [15]:
```
monthly=freq/np.timedelta64(1, 'M')
monthly
```

Out[15]:

```
0           NaN
1      0.229984
2      0.229984
3      0.229984
4      0.229984
5      0.229984
6     14.718988
7      0.229984
8      0.229984
9      0.229984
10     0.229984
11     0.229984
12     0.229984
13     0.229984
14     0.229984
15     0.229984
16     0.229984
17     0.229984
18     0.229984
19     0.229984
20     0.229984
21     0.689953
22     0.229984
23     0.229984
24     0.229984
25     0.229984
26     0.229984
27     0.229984
28     1.839874
29     0.229984
30     0.229984
```

```
31     0.197129
32     1.412760
33     0.229984
34     0.229984
35     0.229984
36     0.229984
37     0.229984
38     0.689953
39     0.229984
40     0.229984
41     0.229984
42     0.229984
43     0.229984
Name: Date, dtype: float64
```

In [ ]: