profiler

July 31, 2023

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_palette("viridis")
```

[ ]:
```
df = pd.read_csv('../data/merged_interpolated.csv')
df.datetime = df.datetime.astype('datetime64')
df.head()
```

[ ]:

	datetime	tempc	cloud8	windk	wdir	humid	rainmm	radkjm2	\
0	2018-03-06 09:30:00	20.75	2.5	14.5	135.0	44.5	0.0	1915.0	
1	2018-03-06 10:00:00	21.50	1.0	16.0	140.0	40.0	0.0	2340.0	
2	2018-03-06 10:30:00	22.25	1.5	15.5	145.0	37.0	0.0	2570.0	
3	2018-03-06 11:00:00	23.00	2.0	15.0	150.0	34.0	0.0	2800.0	
4	2018-03-06 11:30:00	23.55	2.0	13.0	145.0	32.0	0.0	2945.0	

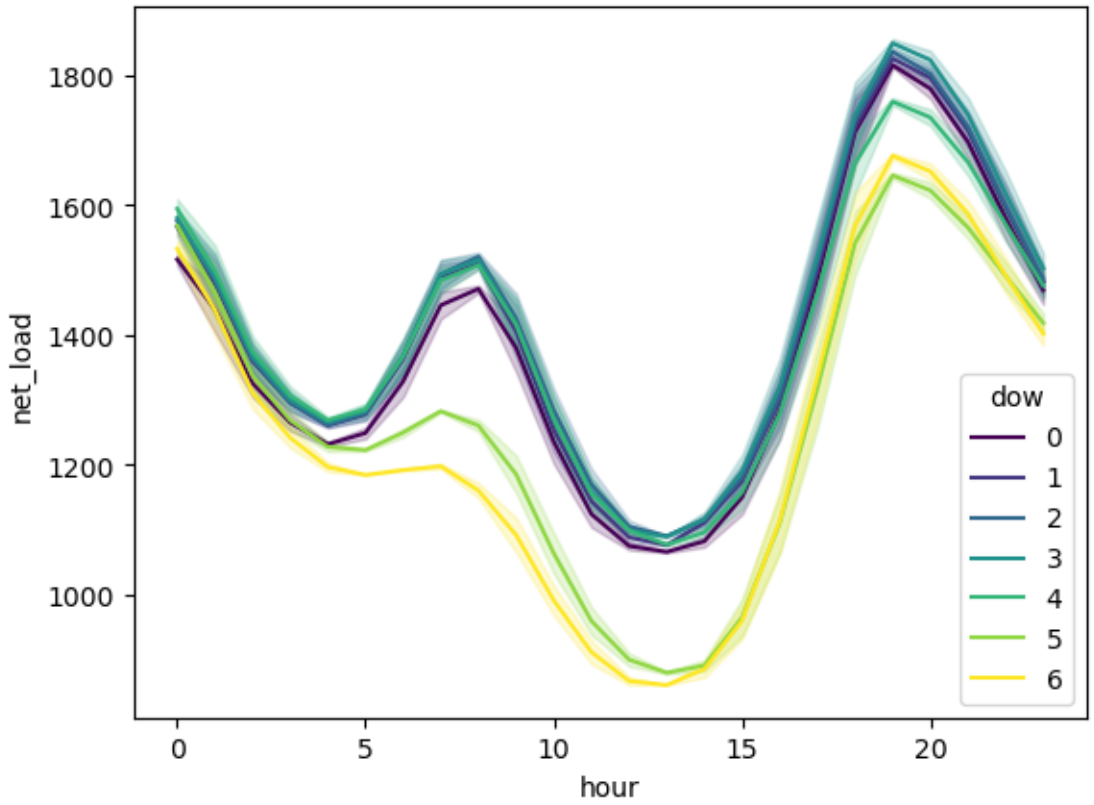
	pv_est	net_load	total_load
0	318.991	1288	1136.79
1	375.231	1237	1054.87
2	430.909	1189	1002.35
3	485.129	1150	971.54
4	523.989	1122	943.68

[ ]:
```
def profiler(data, datetime_col, profile_col):
    df = data.copy().sort_values(datetime_col)
    dt = data[datetime_col].dt
    df['hour'] = dt.hour
    df['minute'] = dt.minute
    grouped = df.groupby(['hour', 'minute'])[[profile_col]].mean()
    sns.lineplot(x='hour', y=profile_col, data=grouped)
    return grouped

def dow_profiler(data, datetime_col, profile_col, palette='viridis'):
    df = data.copy().sort_values(datetime_col)
    dt = data[datetime_col].dt
    df['dow'] = dt.day_of_week
    df['hour'] = dt.hour
    df['minute'] = dt.minute
```

```
    grouped = df.groupby(['dow', 'hour', 'minute'])[[profile_col]].mean()
    sns.lineplot(x='hour', y=profile_col, data=grouped, hue='dow',
                 palette=sns.color_palette(palette, as_cmap=True))

dow_profiler(df, 'datetime', 'net_load')
```



[ ]:
```
df.sort_values('datetime', inplace=True)
dt = df['datetime'].dt
df['year'] = dt.year
df['dow'] = dt.day_of_week
df['hour'] = dt.hour
df['minute'] = dt.minute
grouped = df.groupby(['year', 'dow', 'hour', 'minute'])[['net_load']].mean()
grouped
```

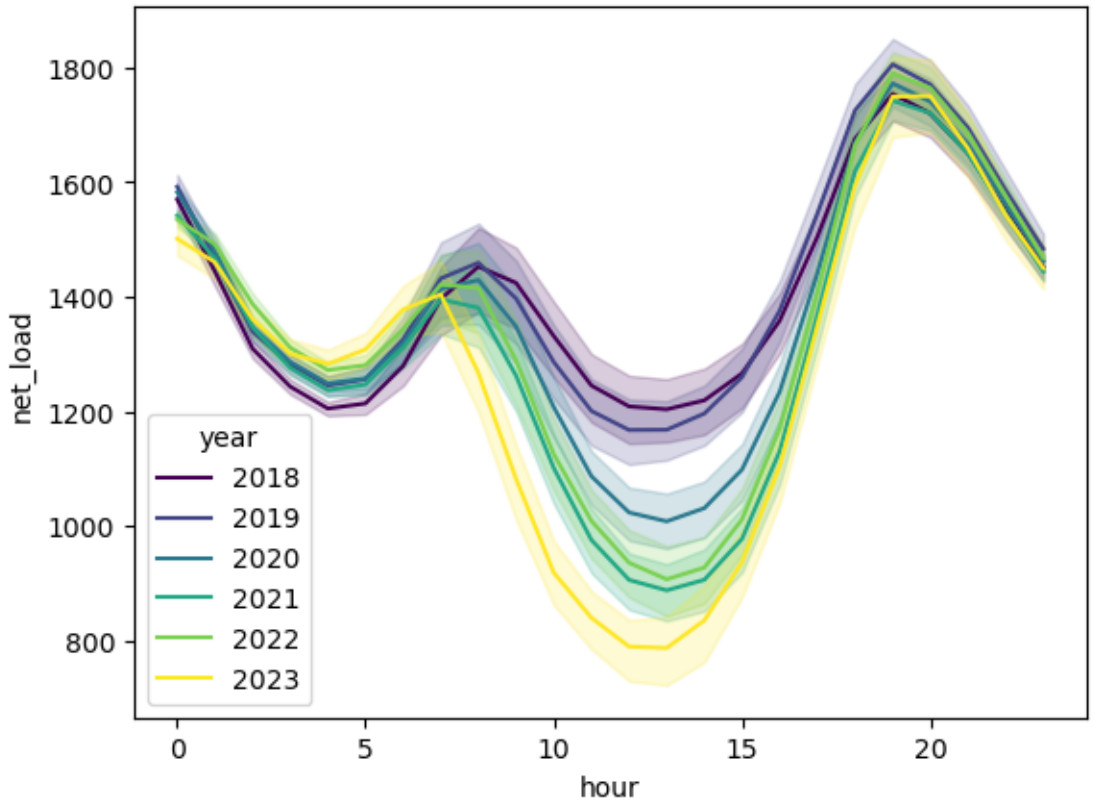
[ ]:
```
                          net_load
year dow hour minute
2018 0   0    0       1534.465116
              30      1497.279070
         1    0       1444.488372
              30      1365.000000
```

```
          2     0        1303.395349
…                                …
2023 6    21    30       1514.800000
          22    0        1466.200000
                30       1414.500000
          23    0        1374.200000
                30       1344.100000

[2016 rows x 1 columns]
```

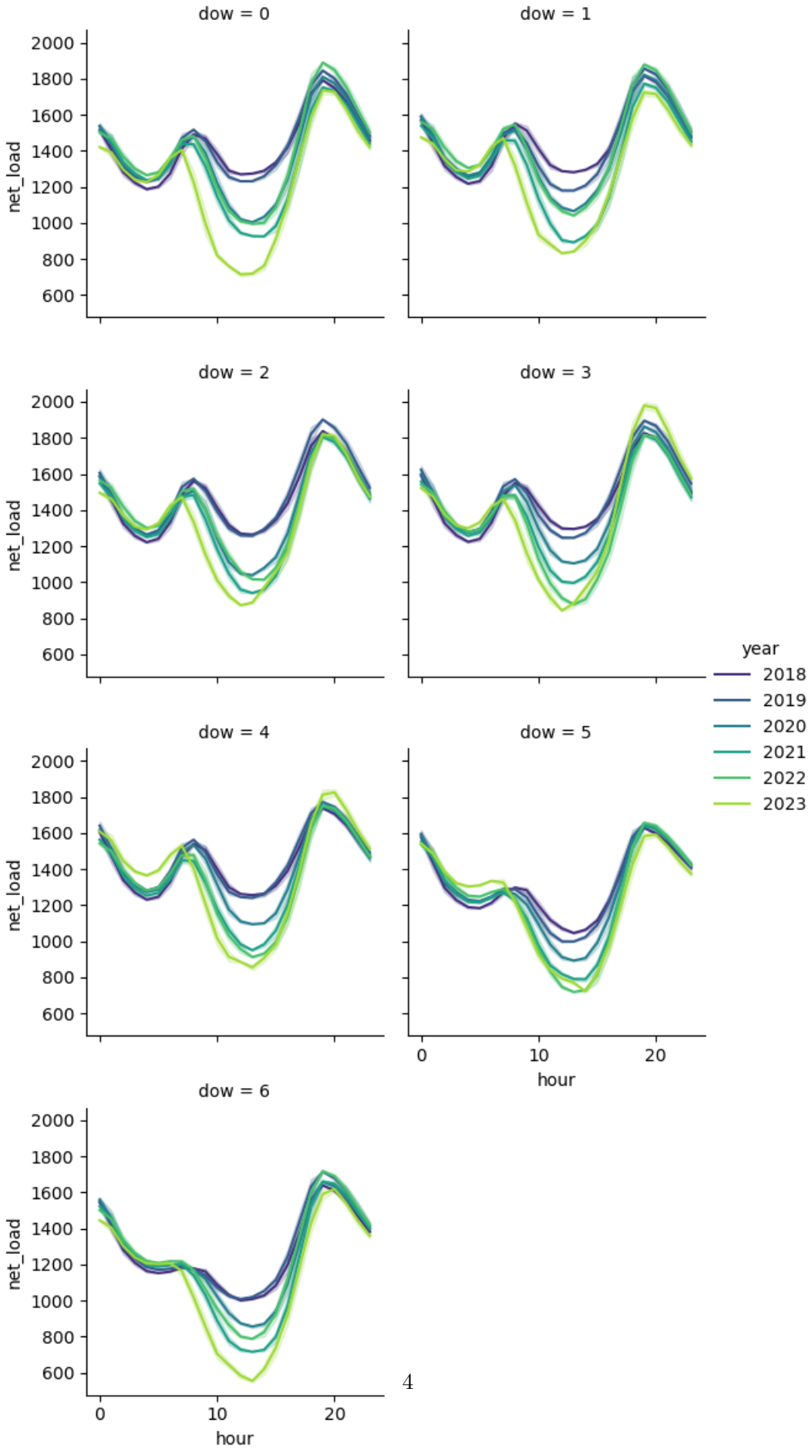
[ ]:
```
sns.lineplot(x='hour', y='net_load', data=grouped, hue='year', palette=sns.
↪color_palette('viridis', as_cmap=True))
```

[ ]:
```
<Axes: xlabel='hour', ylabel='net_load'>
```



[ ]:
```
g = sns.FacetGrid(grouped.reset_index(), col='dow', col_wrap=2,
                  hue='year', palette='viridis')
g.map(sns.lineplot, 'hour', 'net_load')
g.add_legend()
```

[ ]:
```
<seaborn.axisgrid.FacetGrid at 0x7f88eb504f40>
```


dow = 0
dow = 1
dow = 2
dow = 3
dow = 4
dow = 5
dow = 6
net_load
2000
1800
1600
1400
1200
1000
800
600
0
10
20
hour
year
2018
2019
2020
2021
2022
2023

[ ]:
```
df['month'] = dt.month
grouped = df.groupby(['year', 'month', 'dow', 'hour', 'minute'])[['net_load']].
  ↪mean()
grouped
```

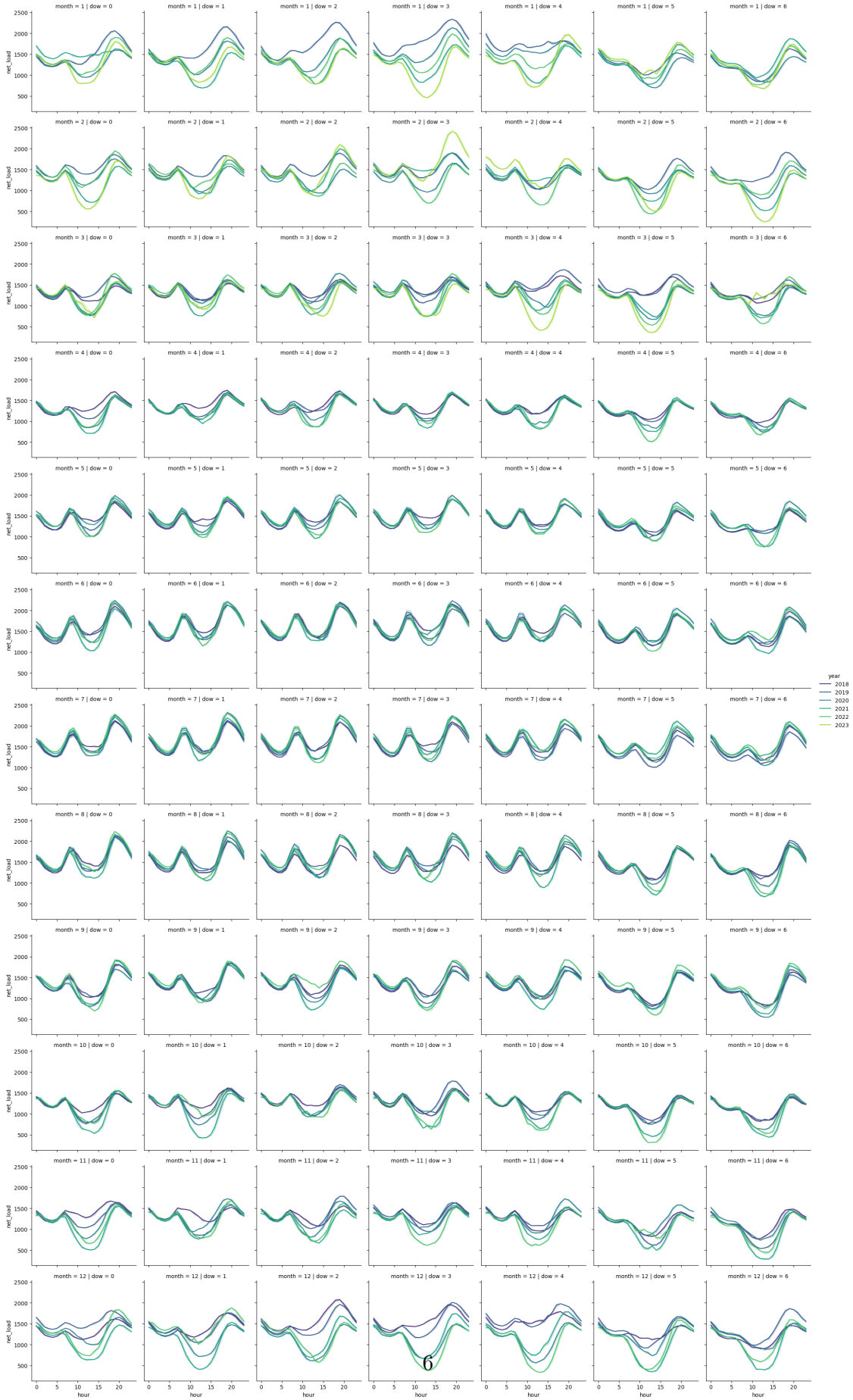
[ ]:

					net_load
year	month	dow	hour	minute	
2018	3	0	0	0	1461.333333
				30	1415.666667
			1	0	1361.333333
				30	1286.666667
			2	0	1245.333333
…					…
2023	3	6	21	30	1433.000000
			22	0	1405.000000
				30	1360.000000
			23	0	1355.000000
				30	1334.000000

[20496 rows x 1 columns]

[ ]:
```
g = sns.FacetGrid(grouped.reset_index(), row='month', col='dow',
                  hue='year', palette='viridis')
g.map(sns.lineplot, 'hour', 'net_load')
g.add_legend()
```

[ ]:
```
<seaborn.axisgrid.FacetGrid at 0x7f88ebc6c220>
```


month = 1 | dow = 0
month = 1 | dow = 1
month = 1 | dow = 2
month = 1 | dow = 3
month = 1 | dow = 4
month = 1 | dow = 5
month = 1 | dow = 6
month = 2 | dow = 0
month = 2 | dow = 1
month = 2 | dow = 2
month = 2 | dow = 3
month = 2 | dow = 4
month = 2 | dow = 5
month = 2 | dow = 6
month = 3 | dow = 0
month = 3 | dow = 1
month = 3 | dow = 2
month = 3 | dow = 3
month = 3 | dow = 4
month = 3 | dow = 5
month = 3 | dow = 6
month = 4 | dow = 0
month = 4 | dow = 1
month = 4 | dow = 2
month = 4 | dow = 3
month = 4 | dow = 4
month = 4 | dow = 5
month = 4 | dow = 6
month = 5 | dow = 0
month = 5 | dow = 1
month = 5 | dow = 2
month = 5 | dow = 3
month = 5 | dow = 4
month = 5 | dow = 5
month = 5 | dow = 6
month = 6 | dow = 0
month = 6 | dow = 1
month = 6 | dow = 2
month = 6 | dow = 3
month = 6 | dow = 4
month = 6 | dow = 5
month = 6 | dow = 6
month = 7 | dow = 0
month = 7 | dow = 1
month = 7 | dow = 2
month = 7 | dow = 3
month = 7 | dow = 4
month = 7 | dow = 5
month = 7 | dow = 6
month = 8 | dow = 0
month = 8 | dow = 1
month = 8 | dow = 2
month = 8 | dow = 3
month = 8 | dow = 4
month = 8 | dow = 5
month = 8 | dow = 6
month = 9 | dow = 0
month = 9 | dow = 1
month = 9 | dow = 2
month = 9 | dow = 3
month = 9 | dow = 4
month = 9 | dow = 5
month = 9 | dow = 6
month = 10 | dow = 0
month = 10 | dow = 1
month = 10 | dow = 2
month = 10 | dow = 3
month = 10 | dow = 4
month = 10 | dow = 5
month = 10 | dow = 6
month = 11 | dow = 0
month = 11 | dow = 1
month = 11 | dow = 2
month = 11 | dow = 3
month = 11 | dow = 4
month = 11 | dow = 5
month = 11 | dow = 6
month = 12 | dow = 0
month = 12 | dow = 1
month = 12 | dow = 2
month = 12 | dow = 3
month = 12 | dow = 4
month = 12 | dow = 5
month = 12 | dow = 6
net_load
2500
2000
1500
1000
500
hour
0
5
10
15
20
year
2018
2019
2020
2021
2022
2023