Sammi Beard

DSC 640 | Weeks 3-4

Setup

```
In [1]:   import warnings
          warnings.simplefilter("ignore")
```

```
In [2]:   import pandas as pd
          import numpy as np
          import matplotlib as plt
          import seaborn as sns
          import functools as ft
```

```
In [3]:   url_2021 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2021_WAVES-ACCESS-RECORDS%20White%20House.c
          url_1_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.01_WAVES-ACCESS-RECORDS.csv'
          url_2_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.02_WAVES-ACCESS-RECORDS.csv'
          url_3_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.03_WAVES-ACCESS-RECORDS-.csv'
          url_4_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.04_WAVES-ACCESS-RECORDS.csv'
          url_5_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.05_WAVES-ACCESS-RECORDS.csv'
          url_6_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.06_WAVES-ACCESS-RECORDS.csv'
          url_7_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.07_WAVES-ACCESS-RECORDS.csv'
          url_8_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.08_WAVES-ACCESS-RECORDS.csv'
          url_9_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.09_WAVES-ACCESS-RECORDS.csv'
          url_10_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.10_WAVES-ACCESS-RECORDS.csv'
          url_11_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.11_WAVES-ACCESS-RECORDS.csv'
          url_12_2022 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2022.12_WAVES-ACCESS-RECORDS.csv'
          url_1_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.01_WAVES-ACCESS-RECORDS.csv'
          url_2_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.02_WAVES-ACCESS-RECORDS.csv'
          url_3_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.03_WAVES-ACCESS-RECORDS.csv'
          url_4_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.04_WAVES-ACCESS-RECORDS.csv'
          url_5_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.05_WAVES-ACCESS-RECORDS.csv'
          url_6_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.06_WAVES-ACCESS-RECORDS.csv'
          url_7_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.07_WAVES-ACCESS-RECORDS.csv'
          url_8_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.08_WAVES-ACCESS-RECORDS.csv'
          url_9_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.09_WAVES-ACCESS-RECORDS.csv'
          url_10_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.10_WAVES-ACCESS-RECORDS.csv'
          url_11_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.11_WAVES-ACCESS-RECORDS.csv'
          url_12_2023 = 'https://raw.githubusercontent.com/Slbeard/DSC640-Data-Pres-Vis/refs/heads/main/data/wk3-4/2023.12_WAVES-ACCESS-RECORDS.csv'
```

```
In [4]:   WH_2021 = pd.read_csv(url_2021)
          WH_1_2022 = pd.read_csv(url_1_2022)
          WH_2_2022 = pd.read_csv(url_2_2022)
          WH_3_2022 = pd.read_csv(url_3_2022)
          WH_4_2022 = pd.read_csv(url_4_2022)
          WH_5_2022 = pd.read_csv(url_5_2022)
          WH_6_2022 = pd.read_csv(url_6_2022)
          WH_7_2022 = pd.read_csv(url_7_2022)
          WH_8_2022 = pd.read_csv(url_8_2022)
          WH_9_2022 = pd.read_csv(url_9_2022)
          WH_10_2022 = pd.read_csv(url_10_2022)
          WH_11_2022 = pd.read_csv(url_11_2022)
          WH_12_2022 = pd.read_csv(url_12_2022)
          WH_1_2023 = pd.read_csv(url_1_2023)
          WH_2_2023 = pd.read_csv(url_2_2023)
          WH_3_2023 = pd.read_csv(url_3_2023)
          WH_4_2023 = pd.read_csv(url_4_2023)
          WH_5_2023 = pd.read_csv(url_5_2023)
          WH_6_2023 = pd.read_csv(url_6_2023)
          WH_7_2023 = pd.read_csv(url_7_2023)
          WH_8_2023 = pd.read_csv(url_8_2023)
          WH_9_2023 = pd.read_csv(url_9_2023)
          WH_10_2023 = pd.read_csv(url_10_2023)
          WH_11_2023 = pd.read_csv(url_11_2023)
          WH_12_2023 = pd.read_csv(url_12_2023)
```

```
In [5]:   list(WH_1_2022.columns)
```

```
Out[5]: ['NAMELAST',
         'NAMEFIRST',
         'NAMEMID',
         'UIN',
         'BDGNBR',
         'ACCESS_TYPE',
         'TOA',
         'POA',
         'TOD',
         'POD',
         'APPT_MADE_DATE',
         'APPT_START_DATE',
         'APPT_END_DATE',
         'APPT_CANCEL_DATE',
         'TOTAL_PEOPLE',
         'LAST_UPDATEDBY',
         'POST',
         'LASTENTRYDATE',
         'TERMINAL_SUFFIX',
         'VISITEE_NAMELAST',
         'VISITEE_NAMEFIRST',
         'MEETING_LOC',
         'MEETING_ROOM',
         'CALLER_NAME_LAST',
         'CALLER_NAME_FIRST',
         'CALLER_ROOM',
         'RELEASEDATE']
```

In [6]:
```python
list(WH_1_2023.columns)
```

```
Out[6]: ['Last Name',
         'First Name',
         'Middle Initial',
         'UIN',
         'BDGNBR',
         'Access Type',
         'TOA',
         'POA',
         'TOD',
         'POD',
         'Appointment Made Date',
         'Appointment Start Date',
         'Appointment End Date',
         'Appointment Cancel Date',
         'Total People',
         'Last Updated By',
         'POST',
         'Last Entry Date',
         'Terminal Suffix',
         'Visitee Last Name',
         'Visitee First Name',
         'Meeting Location',
         'Meeting Room',
         'Caller Last Name',
         'Caller First Name',
         'CALLER_ROOM',
         'RELEASEDATE']
```

In [7]:
```python
count122 = len(WH_1_2022)
count222 = len(WH_2_2022)
count322 = len(WH_3_2022)
count422 = len(WH_4_2022)
count522 = len(WH_5_2022)
count622 = len(WH_6_2022)
count722 = len(WH_7_2022)
count822 = len(WH_8_2022)
count922 = len(WH_9_2022)
count1022 = len(WH_10_2022)
count1122 = len(WH_11_2022)
count1222 = len(WH_12_2022)
count123 = len(WH_1_2023)
count223 = len(WH_2_2023)
count323 = len(WH_3_2023)
count423 = len(WH_4_2023)
count523 = len(WH_5_2023)
count623 = len(WH_6_2023)
count723 = len(WH_7_2023)
count823 = len(WH_8_2023)
count923 = len(WH_9_2023)
count1023 = len(WH_10_2023)
count1123 = len(WH_11_2023)
count1223 = len(WH_12_2023)
```

In [8]:
```python
count_list = [count122, count222, count322, count422, count522, count622, count722, count822, count922, count1022,
              count1122, count1222, count123, count223, count323, count423, count523, count623, count723, count823,
              count923, count1023, count1123, count1223]
```

In [9]:
```python
df_mo_list = ['1_2022', '2_2022', '3_2022', '4_2022', '5_2022', '6_2022', '7_2022', '8_2022',
              '9_2022', '10_2022', '11_2022', '12_2022', '1_2023', '2_2023', '3_2023', '4_2023',
              '5_2023', '6_2023', '7_2023', '8_2023', '9_2023', '10_2023', '11_2023', '12_2023', ]
```

```
In [10]:  df_counts = pd.DataFrame(
              {'DF_Month': df_mo_list,
               'Counts': count_list
              })
```

```
In [11]:  df_counts["Counts"] = pd.to_numeric(df_counts["Counts"])
```

```
In [12]:  ax = sns.barplot(x="Counts", y="DF_Month", data=df_counts, palette='Blues_d').set_title("Number of Visits Per Month")

          plt.pyplot.show()
```
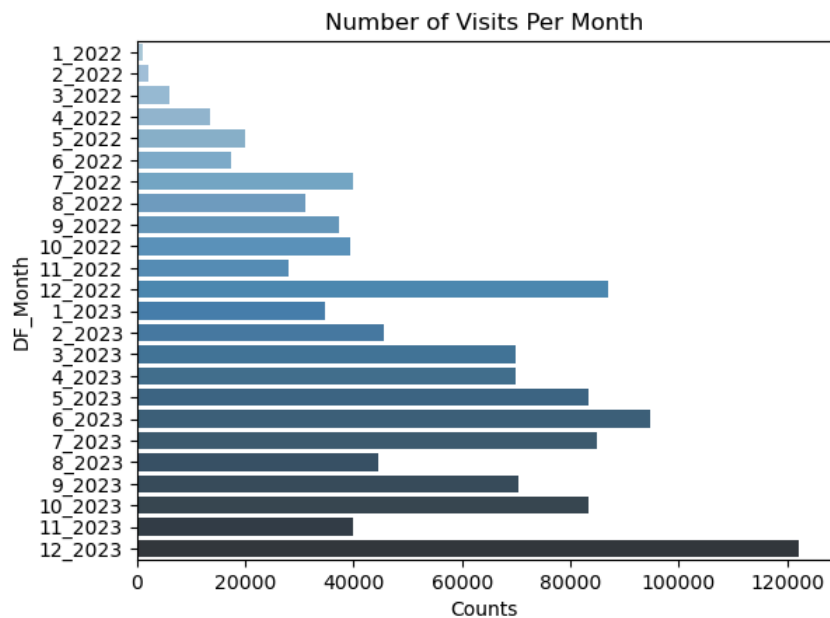


```
In [13]:  df_list_old_cNames = [WH_1_2022, WH_2_2022, WH_3_2022, WH_4_2022, WH_5_2022]
```

```
In [14]:  df_list_new_cNames = [WH_6_2022, WH_7_2022, WH_8_2022, WH_9_2022, WH_10_2022, WH_11_2022, WH_12_2022, WH_1_2023,
                                WH_2_2023, WH_3_2023, WH_4_2023, WH_5_2023, WH_6_2023, WH_7_2023, WH_8_2023,WH_9_2023, WH_10_2023,
                                WH_11_2023, WH_12_2023]
```

```
In [15]:  df_old_cNames = pd.concat(df_list_old_cNames, ignore_index=True)
```

```
In [16]:  df_new_cNames = pd.concat(df_list_new_cNames, ignore_index=True)
```

```
In [17]:  list(df_old_cNames.columns)
```

```
Out[17]:  ['NAMELAST',
           'NAMEFIRST',
           'NAMEMID',
           'UIN',
           'BDGNBR',
           'ACCESS_TYPE',
           'TOA',
           'POA',
           'TOD',
           'POD',
           'APPT_MADE_DATE',
           'APPT_START_DATE',
           'APPT_END_DATE',
           'APPT_CANCEL_DATE',
           'TOTAL_PEOPLE',
           'LAST_UPDATEDBY',
           'POST',
           'LASTENTRYDATE',
           'TERMINAL_SUFFIX',
           'VISITEE_NAMELAST',
           'VISITEE_NAMEFIRST',
           'MEETING_LOC',
           'MEETING_ROOM',
           'CALLER_NAME_LAST',
           'CALLER_NAME_FIRST',
           'CALLER_ROOM',
           'RELEASEDATE',
           'Unnamed: 27',
           'Unnamed: 28']
```

```
In [18]:  list(df_new_cNames.columns)
```

```
Out[18]:   ['Last Name',
           'First Name',
           'Middle Initial',
           'UIN',
           'BDGNBR',
           'Access Type',
           'TOA',
           'POA',
           'TOD',
           'POD',
           'Appointment Made Date',
           'Appointment Start Date',
           'Appointment End Date',
           'Appointment Cancel Date',
           'Total People',
           'Last Updated By',
           'POST',
           'Last Entry Date',
           'Terminal Suffix',
           'Visitee Last Name',
           'Visitee First Name',
           'Meeting Location',
           'Meeting Room',
           'Caller Last Name',
           'Caller First Name',
           'CALLER_ROOM',
           'RELEASEDATE',
           'Caller Room',
           'Release Date']
```

In [19]: `df_old_cNames['Unnamed: 27'].value_counts()`

```
Out[19]:   RL     347
           OHS     23
           FPG     11
           Name: Unnamed: 27, dtype: int64
```

In [20]: `df_old_cNames['Unnamed: 28'].value_counts()`

```
Out[20]:   david.w.nelson@ovp.eop.gov        319
           ashley.n.grove@whmo.mil            30
           charles.powell@whmo.mil            20
           kevin.goodhue@usss.dhs.gov          5
           jessica.p.doiron@whmo.mil           3
           storm.c.horncastle@ovp.eop.gov      1
           Name: Unnamed: 28, dtype: int64
```

In [21]: `df_old_cNames = df_old_cNames.drop(['Unnamed: 27', 'Unnamed: 28', 'RELEASEDATE', 'CALLER_ROOM'], axis=1)`

In [22]:
```
df_old_cNames_changed = df_old_cNames.rename(
    columns={"NAMELAST": "Last Name", "NAMEMID": "First Name", "NAMEFIRST": "Middle Initial", "ACCESS_TYPE": "Access Type",
             "APPT_MADE_DATE": "Appointment Made Date", "APPT_START_DATE": "Appointment Start Date",
             "APPT_END_DATE": "Appointment End Date", "APPT_CANCEL_DATE": "Appointment Cancel Date",
             "TOTAL_PEOPLE": "Total People", "LAST_UPDATEDBY": "Last Updated By","LASTENTRYDATE": "Last Entry Date",
             "TERMINAL_SUFFIX": "Terminal Suffix", "VISITEE_NAMELAST": "Visitee Last Name",
             "VISITEE_NAMEFIRST": "Visitee First Name", "MEETING_LOC": "Meeting Location", "MEETING_ROOM": "Meeting Room",
             "CALLER_NAME_LAST": "Caller Last Name", "CALLER_NAME_FIRST": "Caller First Name",
             "CALLER_ROOM": "Caller Room"})
```

In [23]: `df_new_cNames = df_new_cNames.drop(['RELEASEDATE', 'Release Date', 'Caller Room', 'CALLER_ROOM'], axis=1)`

In [24]:
```
frames = [df_old_cNames_changed, df_new_cNames]

df_final = pd.concat(frames)
```

In [25]: `# df_final = df_final.dropna()`

In [26]: `df_final`

Out[26]:

	Last Name	Middle Initial	First Name	UIN	BDGNBR	Access Type	TOA	POA	TOD	POD	...	Last Updated By	POST	Last Entry Date	Terminal Suffix	Visitee Last Name
0	MALLEY	ROBERT	N	U38641	183020.0	VA	1/1/2022 12:22	NaN	NaN	B04	...	PR	WIN	12/31/2021 1:54	PR	Rajgopal
1	FARLEY	CHRISTOPHER	B	U38643	NaN	VA	NaN	NaN	NaN	NaN	...	RF	WIN	1/2/2022 9:03	RF	Figueroa
2	KALAMBUR	GUHAN	R	U38636	176380.0	VA	1/2/2022 10:59	NaN	NaN	B04	...	SF	WIN	12/30/2021 4:27	SF	Everette
3	KUKLISH	MATILDA	N	U38642	178294.0	VA	1/2/2022 14:01	NaN	NaN	B04	...	AT	WIN	12/31/2021 3:03	AT	Condon
4	SIERRA	RONNEY	F	U38598	NaN	VA	NaN	NaN	NaN	NaN	...	CK	WIN	12/27/2021 4:04	CK	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1123007	PARISH	J	CHARMAINE	U02031	NaN	VA	NaN	NaN	NaN	NaN	...	AG	WIN	12/24/2023 12:36	AG	VPROs
1123008	CARVER	N	DANIELLE	U02031	208431.0	VA	12/25/2023 10:57	VGILL	12/25/2023 22:58	NaN	...	AG	WIN	12/24/2023 12:36	AG	VPROs
1123009	YOUNG	A	MICHAEL	U02031	208409.0	VA	12/25/2023 10:58	VGILL	12/25/2023 15:33	VMAIN	...	AG	WIN	12/24/2023 12:36	AG	VPROs
1123010	KALOGERAS	T	MATTHEW	U03542	NaN	VA	NaN	NaN	NaN	NaN	...	SM	WIN	12/28/2023 12:28	SM	Kalogeras
1123011	JUNKINS	A	CHRISTOPHER	U02588	NaN	VA	NaN	NaN	NaN	NaN	...	CP	WIN	12/26/2023 19:24	CP	Powell

1166046 rows × 25 columns

```
In [27]: df_final.isnull().any()
```

Out[27]:
```
Last Name                 True
Middle Initial            True
First Name                True
UIN                       True
BDGNBR                    True
Access Type               False
TOA                       True
POA                       True
TOD                       True
POD                       True
Appointment Made Date     True
Appointment Start Date    False
Appointment End Date      False
Appointment Cancel Date   True
Total People              True
Last Updated By           False
POST                      False
Last Entry Date           False
Terminal Suffix           False
Visitee Last Name         True
Visitee First Name        True
Meeting Location          False
Meeting Room              True
Caller Last Name          False
Caller First Name         False
dtype: bool
```

```
In [28]: df_final['Total People'] = df_final['Total People'].fillna(1)
```

```
In [29]: df_final['Visitee Last Name'] = df_final['Visitee Last Name'].fillna('None')
         df_final['Visitee First Name'] = df_final['Visitee First Name'].fillna('None')
         df_final['Middle Initial'] = df_final['Middle Initial'].fillna('None')
         df_final['UIN'] = df_final['UIN'].fillna('None')
         df_final['BDGNBR'] = df_final['BDGNBR'].fillna(0)
         df_final['Meeting Room'] = df_final['Meeting Room'].fillna('None')
```

```
In [30]: df_final['POST'].value_counts()
```

Out[30]:
```
WIN    1166046
Name: POST, dtype: int64
```

```
In [31]: df_final = df_final.drop(['Appointment Made Date', 'Appointment Cancel Date', 'TOA', 'POA',
                           'TOD', 'POD', 'Terminal Suffix', 'Last Updated By', 'POST', 'Access Type'], axis=1)
```

```
In [32]: df_final = df_final.dropna()
```

```
In [33]: df_final['Appointment Start Time'] = pd.to_datetime(df_final['Appointment Start Date']).dt.time
         df_final['Appointment End Time'] = pd.to_datetime(df_final['Appointment End Date']).dt.time
         df_final['Appointment Start Date'] = pd.to_datetime(df_final['Appointment Start Date']).dt.date
         df_final['Appointment End Date'] = pd.to_datetime(df_final['Appointment End Date']).dt.date
```

```
In [34]:  df_final['Last Entry Time'] = pd.to_datetime(df_final['Last Entry Date']).dt.time
          df_final['Last Entry Date'] = pd.to_datetime(df_final['Last Entry Date']).dt.date
```

```
In [35]:  df_final['Year'] = pd.to_datetime(df_final['Appointment Start Date']).dt.year
          df_final['Month'] = pd.to_datetime(df_final['Appointment Start Date']).dt.month
          df_final['Day'] = pd.to_datetime(df_final['Appointment Start Date']).dt.day
```

```
In [36]:  df_final['day_of_week'] = pd.to_datetime(df_final['Appointment Start Date']).dt.day_name()
```

```
In [37]:  visits_per_month = df_final.groupby(['Year', 'Month']).size().to_frame('Size')
```

```
In [38]:  visits_per_month = visits_per_month.reset_index()
```

```
In [68]:  sns.lineplot(data=visits_per_month, x="Month", y="Size", palette='Blues_d')
          plt.pyplot.xlabel("Month")
          plt.pyplot.ylabel("Number of Visitor Logs")
```

```
Out[68]:  Text(0, 0.5, 'Number of Visitor Logs')
```
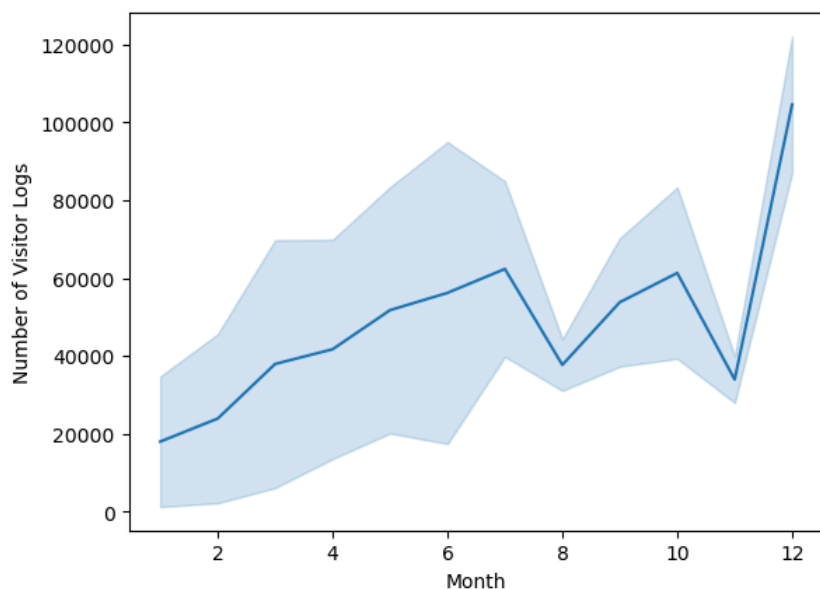


```
In [40]:  visits_per_month_22 = visits_per_month.loc[visits_per_month['Year'] == 2022]
          visits_per_month_23 = visits_per_month.loc[visits_per_month['Year'] == 2023]
```

```
In [41]:  plt.pyplot.bar(visits_per_month_22['Month'], visits_per_month_22['Size'], color='cornflowerblue')
          plt.pyplot.bar(visits_per_month_23['Month'], visits_per_month_23['Size'], bottom=visits_per_month_22['Size'],
                         color='darkblue')

          plt.pyplot.title("Visits Per Month 2022 V 2023")
          plt.pyplot.xlabel("Month")
          plt.pyplot.ylabel("Number of Visitor Logs")
          plt.pyplot.legend(["2022", "2023"])

          plt.pyplot.show()
```
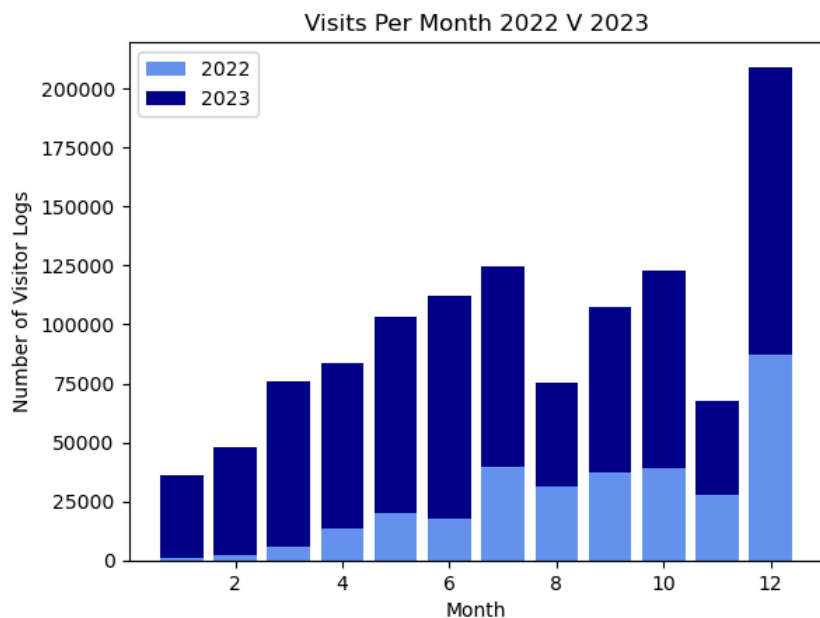
```
In [42]:  group_size_by_month = df_final.groupby(['Year', 'Month', 'Total People']).size().to_frame('Size')
```

```
In [43]:  group_size_by_month = group_size_by_month.reset_index()
```

```
In [44]:  group_size_by_month = group_size_by_month.sort_values(by=['Total People'], ascending=False)
          group_size_by_month
```

Out[44]:

	Year	Month	Total People	Size
817	2022	12	5619.0	4483
483	2022	9	4478.0	4340
1362	2023	6	2775.0	2352
356	2022	7	2424.0	1825
355	2022	7	2162.0	1660
...	...	...	...	...
1363	2023	7	1.0	3510
216	2022	6	1.0	1328
144	2022	5	1.0	1466
1463	2023	8	1.0	2603
0	2022	1	1.0	324

1963 rows × 4 columns

```
In [ ]:
```

```
In [67]:  sns.lineplot(data=group_size_by_month, x="Total People", y="Size", hue="Year", palette='Blues_d', drawstyle='steps-pre')
          plt.pyplot.ylim(0, 1500)
          plt.pyplot.xlim(0, 1500)
          plt.pyplot.title("People Per Visit Log")
          plt.pyplot.xlabel("People Per Reservation")
          plt.pyplot.ylabel("Number of Visitor Logs")
```

Out[67]: Text(0, 0.5, 'Number of Visitor Logs')
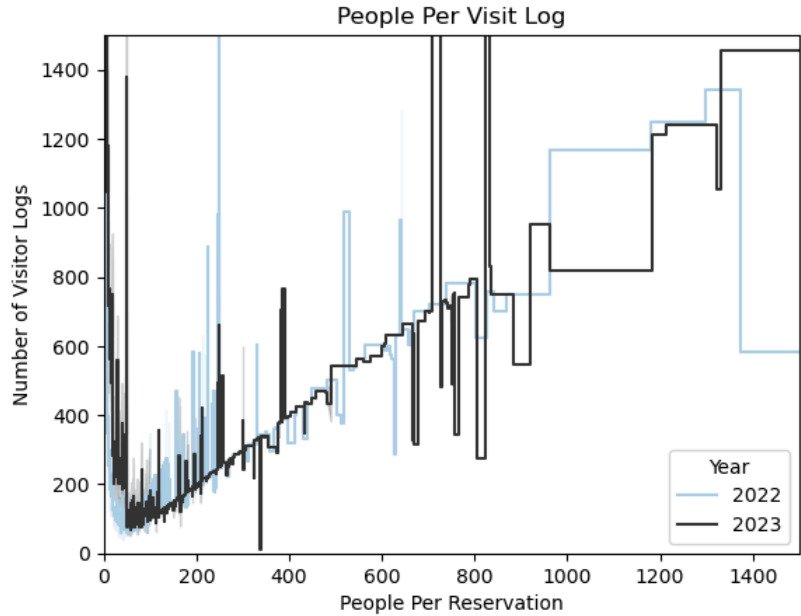

```
In [46]:  visits_by_DOW = df_final.groupby(['Year', 'Month','day_of_week']).size().to_frame('Size')
```

```
In [47]:  visits_by_DOW = visits_by_DOW.reset_index()
```

```
In [48]:  visits_by_DOW
```

	Year	Month	day_of_week	Size
0	2022	1	Friday	277
1	2022	1	Monday	99
2	2022	1	Saturday	47
3	2022	1	Sunday	16
4	2022	1	Thursday	348
...	...	...	...	...
163	2023	12	Saturday	24511
164	2023	12	Sunday	17563
165	2023	12	Thursday	14720
166	2023	12	Tuesday	14100
167	2023	12	Wednesday	15677

168 rows × 4 columns

In [49]:
```python
sns.boxplot(data=visits_by_DOW, x="day_of_week", y="Size", palette='Blues_d')
```

Out[49]: `<AxesSubplot:xlabel='day_of_week', ylabel='Size'>`
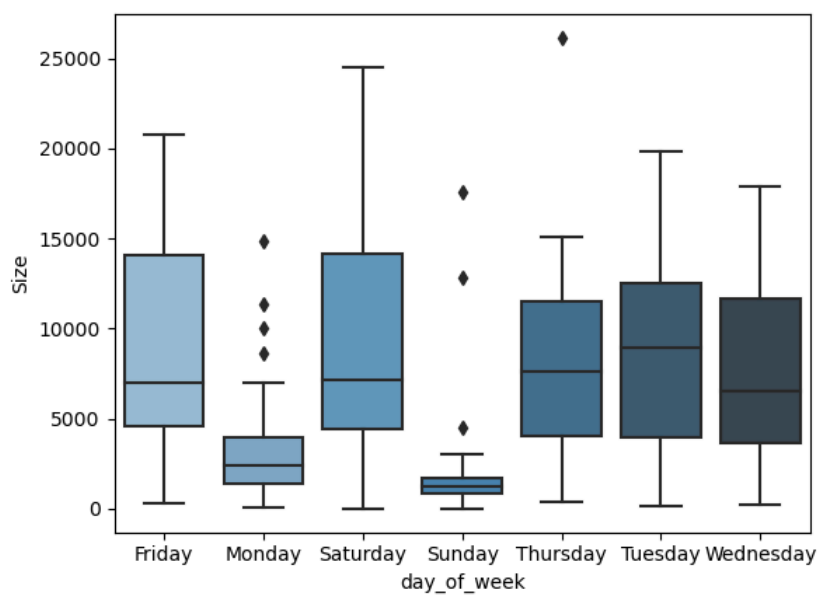
In [50]:
```python
df_final
```

`Out[50]:`

	Last Name	Middle Initial	First Name	UIN	BDGNBR	Appointment Start Date	Appointment End Date	Total People	Last Entry Date	Visitee Last Name	...	Meeting Room	Caller Last Name	Caller First Name	Ap
0	MALLEY	ROBERT	N	U38641	183020.0	2022-01-01	2022-01-01	1.0	2021-12-31	Rajgopal	...	WW-132	RAJGOPAL	PAVAN	
1	FARLEY	CHRISTOPHER	B	U38643	0.0	2022-01-02	2022-01-02	1.0	2022-01-02	Figueroa	...	WW G47	FIGUEROA	RICHARD	
2	KALAMBUR	GUHAN	R	U38636	176380.0	2022-01-02	2022-01-02	1.0	2021-12-30	Everette	...	97	FOWLER	SCOTT	
3	KUKLISH	MATILDA	N	U38642	178294.0	2022-01-02	2022-01-02	1.0	2021-12-31	Condon	...	WHSR	TIBAYAN	ARTURO	
4	SIERRA	RONNEY	F	U38598	0.0	2022-01-02	2022-01-02	1.0	2021-12-27	None	...	97	KLETZIEN	CHRISTOPHER	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1123007	PARISH	J	CHARMAINE	U02031	0.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	
1123008	CARVER	N	DANIELLE	U02031	208431.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	
1123009	YOUNG	A	MICHAEL	U02031	208409.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	
1123010	KALOGERAS	T	MATTHEW	U03542	0.0	2023-12-29	2023-12-29	1.0	2023-12-28	Kalogeras	...	D Section	MCCALL	SHARON	
1123011	JUNKINS	A	CHRISTOPHER	U02588	0.0	2023-12-29	2023-12-29	1.0	2023-12-26	Powell	...	B25	POWELL	CHARLES	

1165966 rows × 22 columns

```
In [51]: visits_by_loc = df_final.groupby(['Meeting Location']).size().to_frame('Size')
```
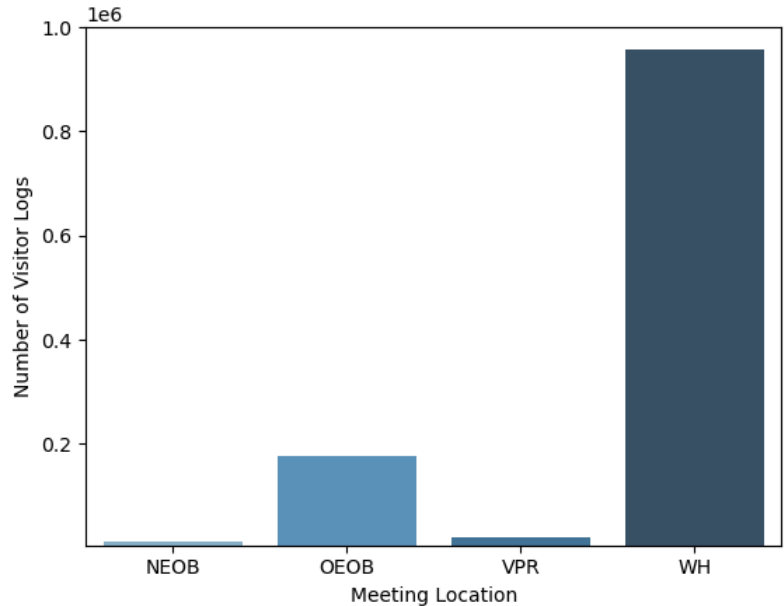
```
In [52]: visits_by_loc = visits_by_loc.reset_index()
```

```
In [53]: visits_by_loc
```

`Out[53]:`

|	Meeting Location	Size
0	NEOB	12214
1	OEOB	175321
2	VPR	20605
3	WH	957826

```
In [69]: sns.barplot(data=visits_by_loc, x='Meeting Location', y='Size', palette='Blues_d')
         plt.pyplot.ylim(5000, 1000000)
         plt.pyplot.xlabel("Meeting Location")
         plt.pyplot.ylabel("Number of Visitor Logs")
```

`Out[69]:` Text(0, 0.5, 'Number of Visitor Logs')


```
In [55]: sns.barplot(data=visits_by_loc, x='Meeting Location', y='Size', palette='Blues_d')
         plt.pyplot.ylim(10000, 200000)
```
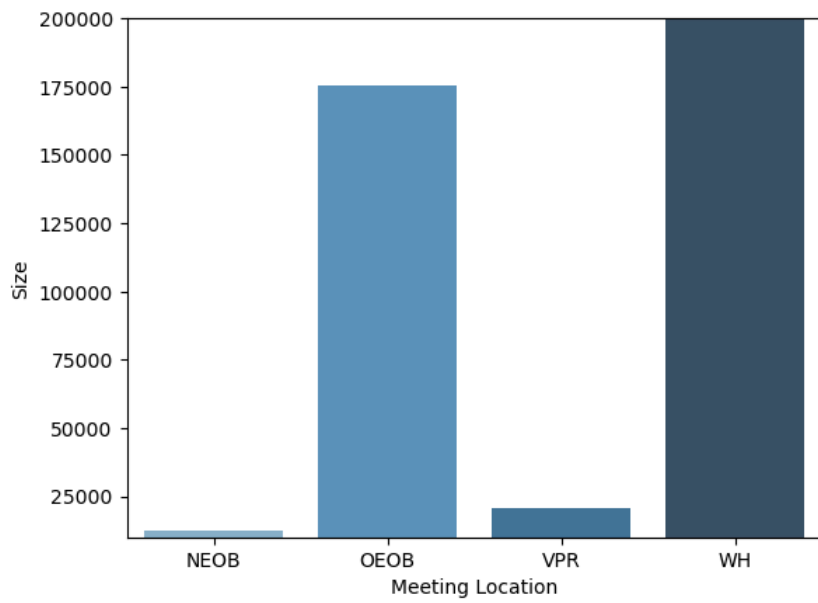
```
In [56]: xmas_day = df_final.loc[(df_final['Month'] == 12) & (df_final['Day'] == 25)]
```

```
In [57]: xmas_day
```

	Last Name	Middle Initial	First Name	UIN	BDGNBR	Appointment Start Date	Appointment End Date	Total People	Last Entry Date	Visitee Last Name	...	Meeting Room	Caller Last Name	Caller First Name	Appointmen Start Tim
207289	CONDON	M	EAMON	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
225218	HILTZ	A	KENNEDY	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
225220	HILTZ	R	MATTHEW	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
225221	HILTZ	C	MAURA	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
225465	HOCHMAN	N	CAROLINE	U09259	0.0	2022-12-25	2022-12-25	5.0	2022-12-19	Hochman	...	Bowling Alley - 037	HOCHMAN	MICHAEL	15:00:0
225468	HOCHMAN	N	JACQUELINE	U09259	0.0	2022-12-25	2022-12-25	5.0	2022-12-19	Hochman	...	Bowling Alley - 037	HOCHMAN	MICHAEL	15:00:0
225470	HOCHMAN	N	MATTHEW	U09259	0.0	2022-12-25	2022-12-25	5.0	2022-12-19	Hochman	...	Bowling Alley - 037	HOCHMAN	MICHAEL	15:00:0
225471	HOCHMAN	N	SARAH	U09259	0.0	2022-12-25	2022-12-25	5.0	2022-12-19	Hochman	...	Bowling Alley - 037	HOCHMAN	MICHAEL	15:00:0
233605	LAABS	A	ALEXANDRA	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
233607	LAABS	A	CHARLES	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
233609	LAABS	K	JULIE	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
233611	LAABS	A	LAURA	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
233614	LAABS	A	OLIVIA	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
233616	LAABS	K	STEPHEN	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
238372	MALONE	M	ERIN	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
238374	MALONE	C	JAMES	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
238384	MALONE	A	MEGHAN	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
238388	MALONE	L	MONICA	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
238390	MALONE	E	PHILIP	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
267287	TORTORA	L	ANN	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
267289	TORTORA	C	MADDALENA	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
267291	TORTORA	J	PAUL	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
267294	TORTORA	E	SARAH	U09260	0.0	2022-12-25	2022-12-25	19.0	2022-12-19	Tortora	...	Bowling Alley - 037	HOCHMAN	MICHAEL	13:00:0
276701	HARMSEN	R	NICHOLAS	U10703	0.0	2022-12-25	2022-12-25	5.0	2022-12-25	Grove	...	1st Floor of Residen	GROVE	ASHLEY	11:30:0
276702	HARMSEN	R	NICHOLAS	U10601	0.0	2022-12-25	2022-12-25	5.0	2022-12-23	Grove	...	1st Floor of Residen	GROVE	ASHLEY	09:30:0
277284	LANE	M	AARON	U10703	0.0	2022-12-25	2022-12-25	5.0	2022-12-25	Grove	...	1st Floor of Residen	GROVE	ASHLEY	11:30:0
277285	LANE	M	AARON	U10601	0.0	2022-12-25	2022-12-25	5.0	2022-12-23	Grove	...	1st Floor of Residen	GROVE	ASHLEY	09:30:0
277766	MCNATT	L	ABRAHAM	U10703	0.0	2022-12-25	2022-12-25	5.0	2022-12-25	Grove	...	1st Floor of Residen	GROVE	ASHLEY	11:30:0
277767	MCNATT	L	ABRAHAM	U10601	0.0	2022-12-25	2022-12-25	5.0	2022-12-23	Grove	...	1st Floor of Residen	GROVE	ASHLEY	09:30:0
278006	MOTT	R	MIKALA	U10703	0.0	2022-12-25	2022-12-25	5.0	2022-12-25	Grove	...	1st Floor of Residen	GROVE	ASHLEY	11:30:0
278007	MOTT	R	MIKALA	U10601	0.0	2022-12-25	2022-12-25	5.0	2022-12-23	Grove	...	1st Floor of Residen	GROVE	ASHLEY	09:30:0
279283	THEEL	I	KALEB	U10703	0.0	2022-12-25	2022-12-25	5.0	2022-12-25	Grove	...	1st Floor of Residen	GROVE	ASHLEY	11:30:0

	Last Name	Middle Initial	First Name	UIN	BDGNBR	Appointment Start Date	Appointment End Date	Total People	Last Entry Date	Visitee Last Name	...	Meeting Room	Caller Last Name	Caller First Name	Appointmen Start Tim
279284	THEEL	I	KALEB	U10601	0.0	2022-12-25	2022-12-25	5.0	2022-12-23	Grove	...	1st Floor of Residen	GROVE	ASHLEY	09:30:0
1123005	MCNATT	L	ABRAHAM	U02031	208487.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	11:30:0
1123006	BROADWATER	K	BRIAN	U02031	208534.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	11:30:0
1123007	PARISH	J	CHARMAINE	U02031	0.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	11:30:0
1123008	CARVER	N	DANIELLE	U02031	208431.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	11:30:0
1123009	YOUNG	A	MICHAEL	U02031	208409.0	2023-12-25	2023-12-25	5.0	2023-12-24	VPROs	...	Basement/1st floor	GROVE	ASHLEY	11:30:0

```
In [ ]: sns.histplot(data=penguins, y="flipper_length_mm")
```

```
In [58]: bowling = df_final.loc[(df_final['Meeting Room'].isin(['Bowling Alley - 037']))]
```

```
In [59]: bowling
```

Out[59]:

	Last Name	Middle Initial	First Name	UIN	BDGNBR	Appointment Start Date	Appointment End Date	Total People	Last Entry Date	Visitee Last Name	...	Meeting Room	Caller Last Name	Caller First Name	Ap
125505	AARON	K	BRYANT	U92621	197377.0	2022-10-27	2022-10-27	14.0	2022-10-25	Ambrose	...	Bowling Alley - 037	LARANCE	MARC	
125551	ABERNATHY	N	NELL	U91032	197558.0	2022-10-20	2022-10-20	11.0	2022-10-20	McBroom	...	Bowling Alley - 037	DEMING	SASHA	
125681	ADAMES	M	JOHANNY	U91873	201325.0	2022-10-27	2022-10-27	19.0	2022-10-21	Perez Fernandez	...	Bowling Alley - 037	MOLINA	JENNIFER	
125780	ADEROJU	O	OMOLARALILIAN	U94474	0.0	2022-10-30	2022-10-30	20.0	2022-10-30	Okunubi	...	Bowling Alley - 037	GONZALEZ	ALEJANDRA	
125781	ADEROJU	O	OMOLOLAVIVIAN	U94474	0.0	2022-10-30	2022-10-30	20.0	2022-10-30	Okunubi	...	Bowling Alley - 037	GONZALEZ	ALEJANDRA	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
497207	WILSONGUARDADO	S	GABRIELLA	U48902	0.0	2023-04-01	2023-04-01	10.0	2023-03-31	Bergemann	...	Bowling Alley - 037	MCLAURIN	JUSCHELLE	
497208	WILSONGUARDADO	S	NOAH	U48902	0.0	2023-04-01	2023-04-01	10.0	2023-03-31	Bergemann	...	Bowling Alley - 037	MCLAURIN	JUSCHELLE	
497391	WITT	N	GABRIELLE	U49100	203840.0	2023-04-01	2023-04-01	17.0	2023-03-31	Smith	...	Bowling Alley - 037	CLARKE	KEZIAH	
497861	WRAYGREENING	N	VICTORIA	U49100	201098.0	2023-04-01	2023-04-01	17.0	2023-03-31	Smith	...	Bowling Alley - 037	CLARKE	KEZIAH	
498561	YOUNGS	B	JON	U48648	201007.0	2023-04-01	2023-04-01	19.0	2023-03-30	Zhong	...	Bowling Alley - 037	CARMICHAEL	TAIWO	

6749 rows × 22 columns

```
In [72]: sns.histplot(data=bowling, y="day_of_week")
plt.pyplot.xlabel("Number of Bowling Reservations")
plt.pyplot.ylabel("Day of Week")
```

Out[72]: Text(0, 0.5, 'Day of Week')
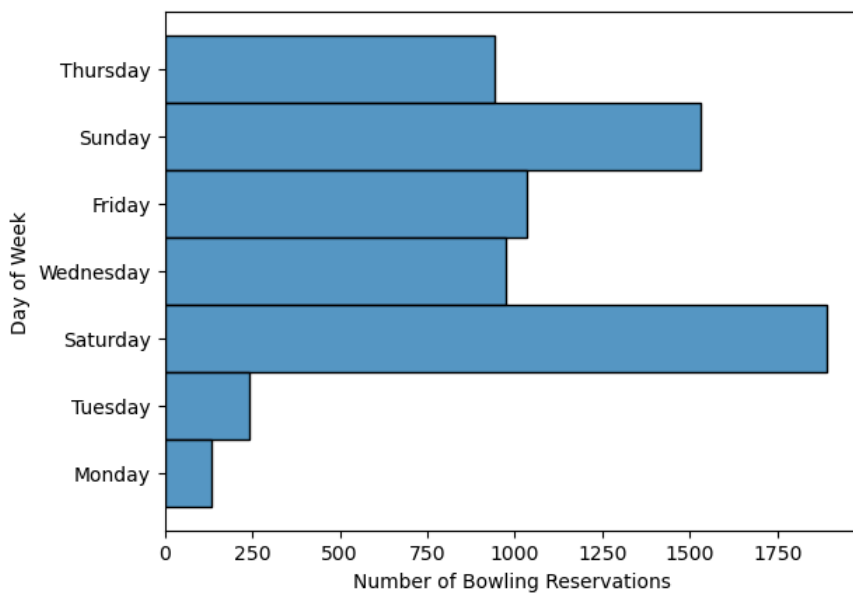
In [61]:
```python
df_final['Appointment Start Time'] = df_final['Appointment Start Time'].astype(str)
df_final['Appointment End Time'] = df_final['Appointment End Time'].astype(str)
df_final['Appointment Start Time'] = df_final['Appointment Start Time'].replace(':','.')
df_final['Appointment End Time'] = df_final['Appointment End Time'].replace(':','.')
```

In [ ]:
```python
fig = plt.pyplot.figure()
ax1 = fig.add_subplot(111)

ax1.step(df_final['Appointment Start time'], df_final['Last Entry Time'], color='cornflowerblue')
plt.pyplot.show()
```

In [ ]:
```python
df_final['Appointment Start Time'] = df_final['Appointment Start Time'].str.replace(':','.')
df_final['Appointment End Time'] = df_final['Appointment End Time'].str.replace(':','.')
```