

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
data = pd.read_csv(r'D:\Career\Udemy\DA 2\Uber Data\uber-pickups-in-new-york-city/uber-raw-
```

In [3]:

```
data.head(2)
```

Out[3]:

	Dispatching_base_num	Pickup_date	Affiliated_base_num	locationID
0	B02617	2015-05-17 09:47:00	B02617	141
1	B02617	2015-05-17 09:47:00	B02617	65

In [4]:

```
data.drop_duplicates(inplace = True)
```

In [5]:

```
data.shape
```

Out[5]:

```
(13372254, 4)
```

which month have max uber pickups

In [6]:

```
data['Pickup_date'] = pd.to_datetime(data['Pickup_date'], format = '%Y-%m-%d %H:%M:%S')
```

In [7]:

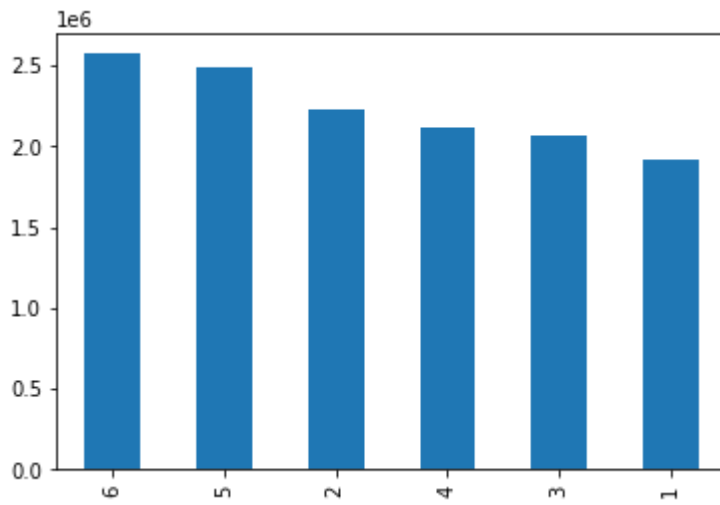
```
data['month'] = data['Pickup_date'].dt.month
```

In [8]:

```
data['month'].value_counts().plot(kind = 'bar')
```

Out[8]:

<AxesSubplot:>



Indepth Analysis of Uber Trips

In [9]:

```
data['week_day'] = data['Pickup_date'].dt.day_name()  
data['day'] = data['Pickup_date'].dt.day  
data['hour'] = data['Pickup_date'].dt.hour  
data['month'] = data['Pickup_date'].dt.month  
data['minute'] = data['Pickup_date'].dt.minute
```

In [10]:

```
data.head()
```

Out[10]:

	Dispatching_base_num	Pickup_date	Affiliated_base_num	locationID	month	week_day	day
0	B02617	2015-05-17 09:47:00	B02617	141	5	Sunday	17
1	B02617	2015-05-17 09:47:00	B02617	65	5	Sunday	17
2	B02617	2015-05-17 09:47:00	B02617	100	5	Sunday	17
3	B02617	2015-05-17 09:47:00	B02774	80	5	Sunday	17
4	B02617	2015-05-17 09:47:00	B02617	90	5	Sunday	17

In [11]:

```
temp = data.groupby(['month', 'week_day'], as_index = False).size()
```

In [12]:

```
temp['month'].unique()
```

Out[12]:

```
array([1, 2, 3, 4, 5, 6], dtype=int64)
```

In [13]:

```
dict_month = {1:'January', 2:'February', 3:'March', 4:'April', 5:'May', 6:'June'}
```

In [14]:

```
temp['month'] = temp['month'].map(dict_month)
```

In [15]:

temp

Out[15]:

	month	week_day	size
0	January	Friday	339285
1	January	Monday	190606
2	January	Saturday	386049
3	January	Sunday	230487
4	January	Thursday	330319
5	January	Tuesday	196574
6	January	Wednesday	245650
7	February	Friday	373550
8	February	Monday	274948
9	February	Saturday	368311
10	February	Sunday	296130
11	February	Thursday	335603
12	February	Tuesday	287260
13	February	Wednesday	286387
14	March	Friday	309631
15	March	Monday	269931
16	March	Saturday	314785
17	March	Sunday	313865
18	March	Thursday	277026
19	March	Tuesday	320634
20	March	Wednesday	256767
21	April	Friday	315002
22	April	Monday	238429
23	April	Saturday	324545
24	April	Sunday	273560
25	April	Thursday	372522
26	April	Tuesday	250632
27	April	Wednesday	338015
28	May	Friday	430134
29	May	Monday	255501
30	May	Saturday	464298
31	May	Sunday	390391
32	May	Thursday	337607
33	May	Tuesday	290004

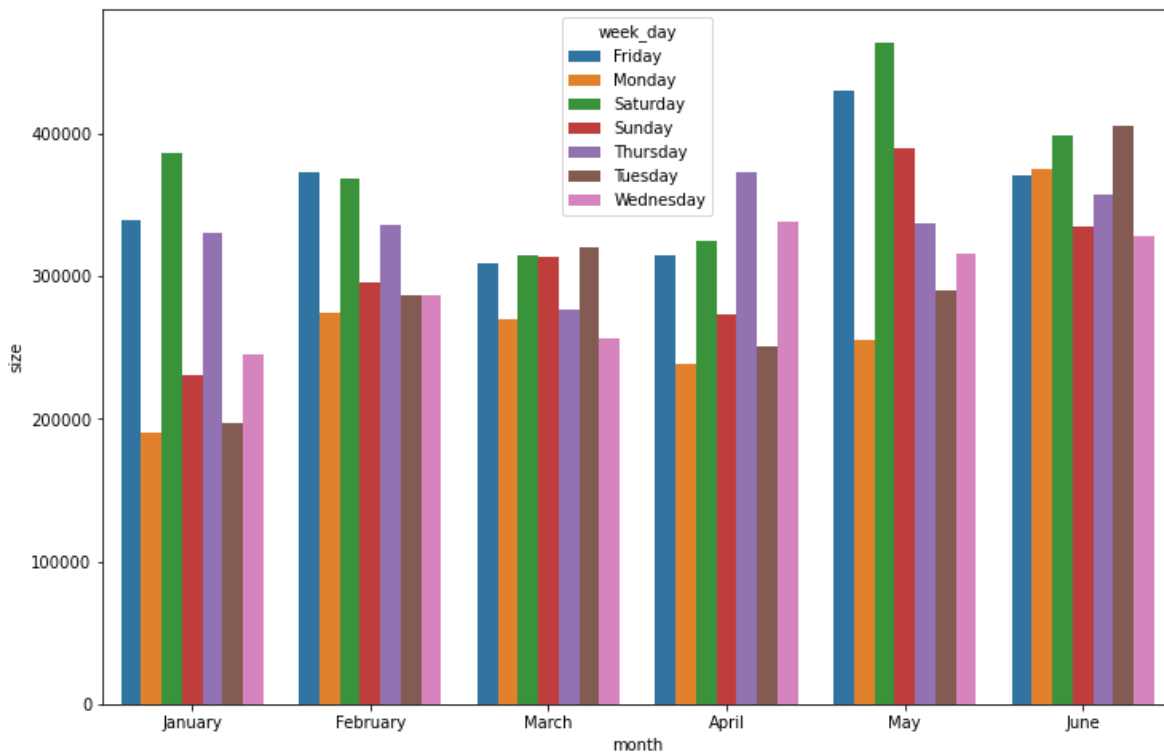
	month	week_day	size
34	May	Wednesday	316045
35	June	Friday	371225
36	June	Monday	375312
37	June	Saturday	399377
38	June	Sunday	334434
39	June	Thursday	357782
40	June	Tuesday	405500
41	June	Wednesday	328141

In [16]:

```
plt.figure(figsize = (12,8))
sns.barplot(x = 'month', y = 'size', hue = 'week_day', data = temp)
```

Out[16]:

<AxesSubplot:xlabel='month', ylabel='size'>



Hourly demand in new york

In [17]:

```
summary = data.groupby(['week_day', 'hour'], as_index = False).size()
```

In [18]:

summary

Out[18]:

	week_day	hour	size
0	Friday	0	79879
1	Friday	1	44563
2	Friday	2	27252
3	Friday	3	19076
4	Friday	4	23049
...
163	Wednesday	19	131317
164	Wednesday	20	123490
165	Wednesday	21	120941
166	Wednesday	22	115208
167	Wednesday	23	91631

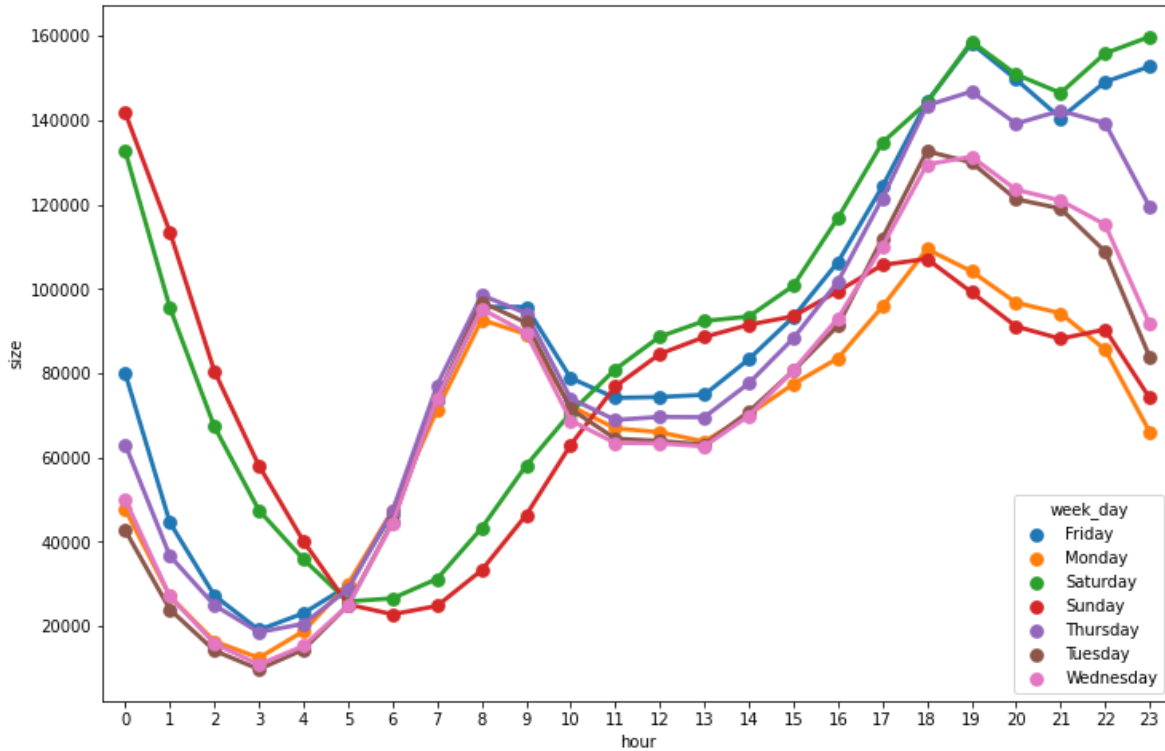
168 rows × 3 columns

In [19]:

```
plt.figure(figsize = (12,8))
sns.pointplot(x = 'hour', y = 'size', hue = 'week_day', data = summary)
```

Out[19]:

<AxesSubplot:xlabel='hour', ylabel='size'>



Most number of active vehicles

In [20]:

```
foil = pd.read_csv(r'D:\Career\Udemy\DA 2\Uber Data\uber-pickups-in-new-york-city\Uber-Jan-foil.head()')
```

Out[20]:

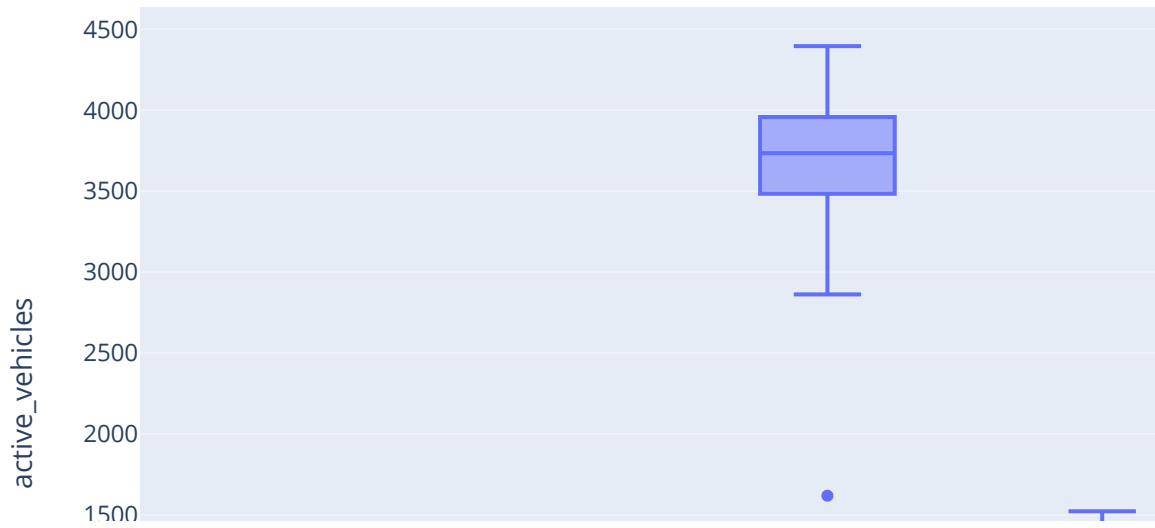
	dispatching_base_number	date	active_vehicles	trips
0	B02512	1/1/2015	190	1132
1	B02765	1/1/2015	225	1765
2	B02764	1/1/2015	3427	29421
3	B02682	1/1/2015	945	7679
4	B02617	1/1/2015	1228	9537

In [21]:

```
import chart_studio.plotly as py
import plotly.graph_objs as go
import plotly.express as px
from plotly.offline import download_plotlyjs, plot, iplot, init_notebook_mode
init_notebook_mode(connected = True)
```

In [22]:

```
px.box(x = 'dispatching_base_number', y = 'active_vehicles', data_frame = foil)
```



In [23]:

```
px.violin(x = 'dispatching_base_number', y = 'active_vehicles', data_frame = foil)
```



Collect Entire Data

In [24]:

```
import os
files = os.listdir(r'D:\Career\Udemy\DA 2\Uber Data\uber-pickups-in-new-york-city')[-7:]
```

In [25]:

```
files
```

Out[25]:

```
['uber-raw-data-apr14.csv',
 'uber-raw-data-aug14.csv',
 'uber-raw-data-janjune-15.csv',
 'uber-raw-data-jul14.csv',
 'uber-raw-data-jun14.csv',
 'uber-raw-data-may14.csv',
 'uber-raw-data-sep14.csv']
```

In [26]:

```
files.remove('uber-raw-data-janjune-15.csv')
```

In [27]:

files

Out[27]:

```
['uber-raw-data-apr14.csv',
 'uber-raw-data-aug14.csv',
 'uber-raw-data-jul14.csv',
 'uber-raw-data-jun14.csv',
 'uber-raw-data-may14.csv',
 'uber-raw-data-sep14.csv']
```

In [28]:

```
path = r'D:\Career\Udemy\DA 2\Uber Data\uber-pickups-in-new-york-city'

final_df = pd.DataFrame()
for file in files:
    current_df = pd.read_csv(path+'/'+file, encoding = 'utf-8')
    final_df = pd.concat([current_df,final_df])
```

In [29]:

final_df.head()

Out[29]:

	Date/Time	Lat	Lon	Base
0	9/1/2014 0:01:00	40.2201	-74.0021	B02512
1	9/1/2014 0:01:00	40.7500	-74.0027	B02512
2	9/1/2014 0:03:00	40.7559	-73.9864	B02512
3	9/1/2014 0:06:00	40.7450	-73.9889	B02512
4	9/1/2014 0:11:00	40.8145	-73.9444	B02512

In [30]:

final_df.drop_duplicates(inplace = True)

In [31]:

final_df.shape

Out[31]:

(4451746, 4)

Location where uber getting rush

In [32]:

rush_uber = final_df.groupby(['Lat', 'Lon'], as_index = False).size()

In [33]:

```
rush_uber.head()
```

Out[33]:

	Lat	Lon	size
0	39.6569	-74.2258	1
1	39.6686	-74.1607	1
2	39.7214	-74.2446	1
3	39.8416	-74.1512	1
4	39.9055	-74.0791	1

In [34]:

```
import folium
```

In [35]:

```
baseMap = folium.Map()
```

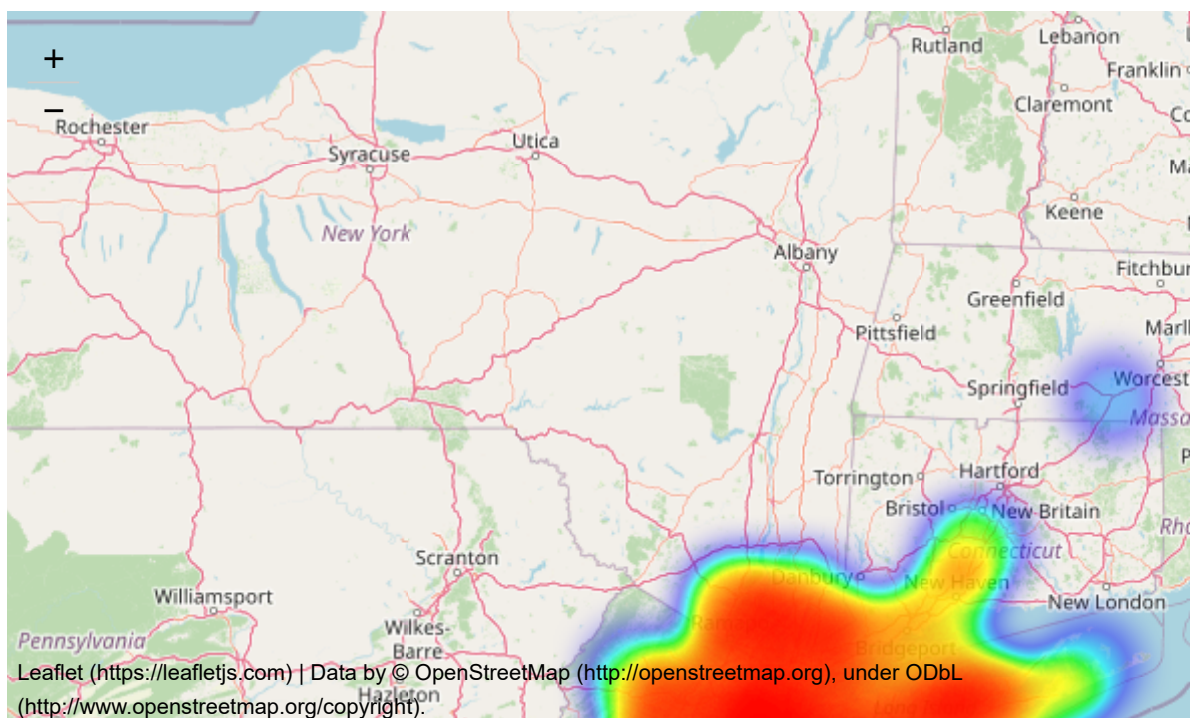
In [36]:

```
from folium.plugins import HeatMap
```

In [37]:

```
HeatMap(rush_uber).add_to(baseMap)  
baseMap
```

Out[37]:



Rush on hour and weekday

In [38]:

```
final_df.head()
```

Out[38]:

	Date/Time	Lat	Lon	Base
0	9/1/2014 0:01:00	40.2201	-74.0021	B02512
1	9/1/2014 0:01:00	40.7500	-74.0027	B02512
2	9/1/2014 0:03:00	40.7559	-73.9864	B02512
3	9/1/2014 0:06:00	40.7450	-73.9889	B02512
4	9/1/2014 0:11:00	40.8145	-73.9444	B02512

In [39]:

```
final_df['Date/Time'] = pd.to_datetime(final_df['Date/Time'], format = '%m/%d/%Y %H:%M:%S')
```

In [40]:

```
final_df.head()
```

Out[40]:

	Date/Time	Lat	Lon	Base
0	2014-09-01 00:01:00	40.2201	-74.0021	B02512
1	2014-09-01 00:01:00	40.7500	-74.0027	B02512
2	2014-09-01 00:03:00	40.7559	-73.9864	B02512
3	2014-09-01 00:06:00	40.7450	-73.9889	B02512
4	2014-09-01 00:11:00	40.8145	-73.9444	B02512

In [41]:

```
final_df['weekday'] = final_df['Date/Time'].dt.day
final_df['hour'] = final_df['Date/Time'].dt.hour
```

In [42]:

```
final_df.head()
```

Out[42]:

	Date/Time	Lat	Lon	Base	weekday	hour
0	2014-09-01 00:01:00	40.2201	-74.0021	B02512	1	0
1	2014-09-01 00:01:00	40.7500	-74.0027	B02512	1	0
2	2014-09-01 00:03:00	40.7559	-73.9864	B02512	1	0
3	2014-09-01 00:06:00	40.7450	-73.9889	B02512	1	0
4	2014-09-01 00:11:00	40.8145	-73.9444	B02512	1	0

In [43]:

```
pivot = final_df.groupby(['weekday', 'hour']).size().unstack()
```

In [44]:

```
pivot
```

Out[44]:

hour	0	1	2	3	4	5	6	7	8	9	...	14	15	
weekday														
1	3178	1944	1256	1308	1429	2126	3664	5380	5292	4617	...	6933	7910	8
2	2435	1569	1087	1414	1876	2812	4920	6544	6310	4712	...	6904	8449	10
3	3354	2142	1407	1467	1550	2387	4241	5663	5386	4657	...	7226	8850	10
4	2897	1688	1199	1424	1696	2581	4592	6029	5704	4744	...	7158	8515	9
5	2733	1541	1030	1253	1617	2900	4814	6261	6469	5530	...	6955	8312	9
6	4537	2864	1864	1555	1551	2162	3642	4766	4942	4401	...	7235	8612	9
7	3645	2296	1507	1597	1763	2422	4102	5575	5376	4639	...	7276	8474	10
8	2830	1646	1123	1483	1889	3224	5431	7361	7357	5703	...	7240	8775	9
9	2657	1724	1222	1480	1871	3168	5802	7592	7519	5895	...	7877	9220	10
10	3296	2126	1464	1434	1591	2594	4664	6046	6158	5072	...	7612	9578	11
11	3036	1665	1095	1424	1842	2520	4954	6876	6871	5396	...	7503	8920	10
12	3227	2147	1393	1362	1757	2710	4576	6250	6231	5177	...	7743	9390	10
13	5408	3509	2262	1832	1705	2327	4196	5685	6060	5631	...	8200	9264	10
14	3748	2349	1605	1656	1756	2629	4257	5781	5520	4824	...	6963	8192	9
15	2497	1515	1087	1381	1862	2980	5050	6837	6729	5201	...	7633	8505	10
16	2547	1585	1119	1395	1818	2966	5558	7517	7495	5958	...	7597	9290	10
17	3155	2048	1500	1488	1897	2741	4562	6315	5882	4934	...	7472	8997	10
18	3390	2135	1332	1626	1892	2959	4688	6618	6451	5377	...	7534	9040	10
19	3217	2188	1604	1675	1810	2639	4733	6159	6014	5006	...	7374	8898	9
20	4475	3190	2100	1858	1618	2143	3584	4900	5083	4765	...	7462	8630	9
21	4294	3194	1972	1727	1926	2615	4185	5727	5529	4707	...	7064	8127	9
22	2787	1637	1175	1468	1934	3151	5204	6872	6850	5198	...	7337	9148	10
23	2546	1580	1136	1429	1957	3132	5204	6890	6436	5177	...	7575	9309	9
24	3200	2055	1438	1493	1798	2754	4484	6013	5913	5146	...	7083	8706	10
25	2405	1499	1072	1439	1943	2973	5356	7627	7078	5994	...	7298	8732	9
26	3810	3065	2046	1806	1730	2337	3776	5172	5071	4808	...	7269	8815	9
27	5196	3635	2352	2055	1723	2336	3539	4937	5053	4771	...	7519	8803	9
28	4123	2646	1843	1802	1883	2793	4290	5715	5671	5206	...	7341	8584	9
29	2678	1827	1409	1678	1948	3056	5213	6852	6695	5481	...	7630	9249	10
30	2401	1510	1112	1403	1841	3216	5757	7596	7611	6064	...	8396	10243	11
31	2174	1394	1087	919	773	997	1561	2169	2410	2525	...	4104	5099	5

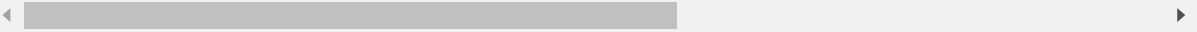
31 rows × 24 columns

In [45]:

```
pivot.style.background_gradient()
```

Out[45]:

hour	0	1	2	3	4	5	6	7	8	9	10	11	12	13
weekday														
1	3178	1944	1256	1308	1429	2126	3664	5380	5292	4617	4607	4729	4930	5790
2	2435	1569	1087	1414	1876	2812	4920	6544	6310	4712	4797	4975	5188	5690
3	3354	2142	1407	1467	1550	2387	4241	5663	5386	4657	4788	5065	5384	6090
4	2897	1688	1199	1424	1696	2581	4592	6029	5704	4744	4743	4975	5193	6170
5	2733	1541	1030	1253	1617	2900	4814	6261	6469	5530	5141	5011	5047	5690
6	4537	2864	1864	1555	1551	2162	3642	4766	4942	4401	4801	5174	5426	6250
7	3645	2296	1507	1597	1763	2422	4102	5575	5376	4639	4905	5166	5364	6210
8	2830	1646	1123	1483	1889	3224	5431	7361	7357	5703	5288	5350	5483	6310
9	2657	1724	1222	1480	1871	3168	5802	7592	7519	5895	5406	5443	5496	6410
10	3296	2126	1464	1434	1591	2594	4664	6046	6158	5072	4976	5415	5506	6520
11	3036	1665	1095	1424	1842	2520	4954	6876	6871	5396	5215	5423	5513	6480
12	3227	2147	1393	1362	1757	2710	4576	6250	6231	5177	5157	5319	5570	6440
13	5408	3509	2262	1832	1705	2327	4196	5685	6060	5631	5442	5720	5914	6670
14	3748	2349	1605	1656	1756	2629	4257	5781	5520	4824	4911	5118	5153	5740
15	2497	1515	1087	1381	1862	2980	5050	6837	6729	5201	5347	5517	5503	6990
16	2547	1585	1119	1395	1818	2966	5558	7517	7495	5958	5626	5480	5525	6190
17	3155	2048	1500	1488	1897	2741	4562	6315	5882	4934	5004	5306	5634	6500
18	3390	2135	1332	1626	1892	2959	4688	6618	6451	5377	5150	5487	5490	6380
19	3217	2188	1604	1675	1810	2639	4733	6159	6014	5006	5092	5240	5590	6360
20	4475	3190	2100	1858	1618	2143	3584	4900	5083	4765	5135	5650	5745	6650
21	4294	3194	1972	1727	1926	2615	4185	5727	5529	4707	4911	5212	5465	6080
22	2787	1637	1175	1468	1934	3151	5204	6872	6850	5198	5277	5352	5512	6340
23	2546	1580	1136	1429	1957	3132	5204	6890	6436	5177	5066	5304	5504	6230
24	3200	2055	1438	1493	1798	2754	4484	6013	5913	5146	4947	5311	5229	5970
25	2405	1499	1072	1439	1943	2973	5356	7627	7078	5994	5432	5504	5694	6200
26	3810	3065	2046	1806	1730	2337	3776	5172	5071	4808	5061	5179	5381	6160
27	5196	3635	2352	2055	1723	2336	3539	4937	5053	4771	5198	5732	5839	6820
28	4123	2646	1843	1802	1883	2793	4290	5715	5671	5206	5247	5500	5486	6120
29	2678	1827	1409	1678	1948	3056	5213	6852	6695	5481	5234	5163	5220	6300
30	2401	1510	1112	1403	1841	3216	5757	7596	7611	6064	5987	6090	6423	7240
31	2174	1394	1087	919	773	997	1561	2169	2410	2525	2564	2777	2954	3280



In [46]:

```
def gen_table(df,col1,col2):  
    pivot = df.groupby([col1,col2]).size().unstack()  
    return pivot.style.background_gradient()
```


In [47]:

```
gen_table(final_df, 'weekday', 'hour')
```

Out[47]:

hour	0	1	2	3	4	5	6	7	8	9	10	11	12	13
weekday														
1	3178	1944	1256	1308	1429	2126	3664	5380	5292	4617	4607	4729	4930	5790
2	2435	1569	1087	1414	1876	2812	4920	6544	6310	4712	4797	4975	5188	5690
3	3354	2142	1407	1467	1550	2387	4241	5663	5386	4657	4788	5065	5384	6090
4	2897	1688	1199	1424	1696	2581	4592	6029	5704	4744	4743	4975	5193	6170
5	2733	1541	1030	1253	1617	2900	4814	6261	6469	5530	5141	5011	5047	5690
6	4537	2864	1864	1555	1551	2162	3642	4766	4942	4401	4801	5174	5426	6250
7	3645	2296	1507	1597	1763	2422	4102	5575	5376	4639	4905	5166	5364	6210
8	2830	1646	1123	1483	1889	3224	5431	7361	7357	5703	5288	5350	5483	6310
9	2657	1724	1222	1480	1871	3168	5802	7592	7519	5895	5406	5443	5496	6410
10	3296	2126	1464	1434	1591	2594	4664	6046	6158	5072	4976	5415	5506	6520
11	3036	1665	1095	1424	1842	2520	4954	6876	6871	5396	5215	5423	5513	6480
12	3227	2147	1393	1362	1757	2710	4576	6250	6231	5177	5157	5319	5570	6440
13	5408	3509	2262	1832	1705	2327	4196	5685	6060	5631	5442	5720	5914	6670
14	3748	2349	1605	1656	1756	2629	4257	5781	5520	4824	4911	5118	5153	5740
15	2497	1515	1087	1381	1862	2980	5050	6837	6729	5201	5347	5517	5503	6990
16	2547	1585	1119	1395	1818	2966	5558	7517	7495	5958	5626	5480	5525	6190
17	3155	2048	1500	1488	1897	2741	4562	6315	5882	4934	5004	5306	5634	6500
18	3390	2135	1332	1626	1892	2959	4688	6618	6451	5377	5150	5487	5490	6380
19	3217	2188	1604	1675	1810	2639	4733	6159	6014	5006	5092	5240	5590	6360
20	4475	3190	2100	1858	1618	2143	3584	4900	5083	4765	5135	5650	5745	6650
21	4294	3194	1972	1727	1926	2615	4185	5727	5529	4707	4911	5212	5465	6080
22	2787	1637	1175	1468	1934	3151	5204	6872	6850	5198	5277	5352	5512	6340
23	2546	1580	1136	1429	1957	3132	5204	6890	6436	5177	5066	5304	5504	6230
24	3200	2055	1438	1493	1798	2754	4484	6013	5913	5146	4947	5311	5229	5970
25	2405	1499	1072	1439	1943	2973	5356	7627	7078	5994	5432	5504	5694	6200
26	3810	3065	2046	1806	1730	2337	3776	5172	5071	4808	5061	5179	5381	6160
27	5196	3635	2352	2055	1723	2336	3539	4937	5053	4771	5198	5732	5839	6820
28	4123	2646	1843	1802	1883	2793	4290	5715	5671	5206	5247	5500	5486	6120
29	2678	1827	1409	1678	1948	3056	5213	6852	6695	5481	5234	5163	5220	6300
30	2401	1510	1112	1403	1841	3216	5757	7596	7611	6064	5987	6090	6423	7240
31	2174	1394	1087	919	773	997	1561	2169	2410	2525	2564	2777	2954	3280

