In [39]:
```
import pandas as pd
import datetime as dt
import numpy as np
```

In [3]:
```
df2101 = pd.read_csv('202101-divvy-tripdata.csv', skipinitialspace= True)
df2102 = pd.read_csv('202102-divvy-tripdata.csv', skipinitialspace= True)
df2103 = pd.read_csv('202103-divvy-tripdata.csv', skipinitialspace= True)
df2104 = pd.read_csv('202104-divvy-tripdata.csv', skipinitialspace= True)
df2105 = pd.read_csv('202105-divvy-tripdata.csv', skipinitialspace= True)
df2106 = pd.read_csv('202106-divvy-tripdata.csv', skipinitialspace= True)
df2107 = pd.read_csv('202107-divvy-tripdata.csv', skipinitialspace= True)
df2108 = pd.read_csv('202108-divvy-tripdata.csv', skipinitialspace= True)
df2109 = pd.read_csv('202109-divvy-tripdata.csv', skipinitialspace= True)
df2110 = pd.read_csv('202110-divvy-tripdata.csv', skipinitialspace= True)
df2111 = pd.read_csv('202111-divvy-tripdata.csv', skipinitialspace= True)
df2112 = pd.read_csv('202112-divvy-tripdata.csv', skipinitialspace= True)
```

In [48]:
```
dfs = pd.concat([df2101,df2102,df2103,df2104,df2105,df2106,df2107,df2108,df2109,df2110,df2111,df2112], axis=0, ignore_index= True)
dfs.head()
```

Out[48]:

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
0	E19E6F188D4C42ED	electric_bike	2021-01-23 16:14:19	2021-01-23 16:24:44	California Ave & Cortez St	17660	NaN	NaN	41.900341	-87.696743	41.89	-87.72	member
1	DC88F20C2C55F27F	electric_bike	2021-01-27 18:43:08	2021-01-27 18:47:12	California Ave & Cortez St	17660	NaN	NaN	41.900333	-87.696707	41.90	-87.69	member
2	EC45C94683FE3F27	electric_bike	2021-01-21 22:35:54	2021-01-21 22:37:14	California Ave & Cortez St	17660	NaN	NaN	41.900313	-87.696643	41.90	-87.70	member
3	4FA453A75AE377DB	electric_bike	2021-01-07 13:31:13	2021-01-07 13:42:55	California Ave & Cortez St	17660	NaN	NaN	41.900399	-87.696662	41.92	-87.69	member
4	BE5E8EB4E7263A0B	electric_bike	2021-01-23 02:24:02	2021-01-23 02:24:45	California Ave & Cortez St	17660	NaN	NaN	41.900326	-87.696697	41.90	-87.70	casual

In [49]:
```
dfs.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5595063 entries, 0 to 5595062
Data columns (total 13 columns):
 #   Column              Dtype  
---  ------              -----  
 0   ride_id             object 
 1   rideable_type       object 
 2   started_at          object 
 3   ended_at            object 
 4   start_station_name  object 
 5   start_station_id    object 
 6   end_station_name    object 
 7   end_station_id      object 
 8   start_lat           float64
 9   start_lng           float64
 10  end_lat             float64
 11  end_lng             float64
 12  member_casual       object 
dtypes: float64(4), object(9)
memory usage: 554.9+ MB
```

In [50]:
```
dfs['started_at'] = pd.to_datetime(dfs['started_at'])
dfs['ended_at'] = pd.to_datetime(dfs['ended_at'])
```

In [51]:
```
dfs['rideable_type'].unique()
dfs['member_casual'].unique()
```

Out[51]:
```
array(['member', 'casual'], dtype=object)
```

In [52]:
```
dfs['rideable_type'] = dfs['rideable_type'].astype('category')
dfs['member_casual'] = dfs['member_casual'].astype('category')
```

In [53]:
```
dfs.dropna(inplace= True)
```

In [54]:
```
dfs.head()
```

Out[54]:

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
9	B9F73448DFBE0D45	classic_bike	2021-01-24 19:15:38	2021-01-24 19:22:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	member
10	457C7F4B5D3DA135	electric_bike	2021-01-23 12:57:38	2021-01-23 13:02:10	California Ave & Cortez St	17660	California Ave & North Ave	13258	41.900406	-87.696733	41.910435	-87.696890	member
11	57C750326F9FDABE	electric_bike	2021-01-09 15:28:04	2021-01-09 15:37:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900374	-87.696688	41.899180	-87.672178	casual
12	4D518C65E338D070	electric_bike	2021-01-09 15:28:57	2021-01-09 15:37:54	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900379	-87.696716	41.899149	-87.672177	casual
13	9D08A3AFF410474D	classic_bike	2021-01-24 15:56:59	2021-01-24 16:07:08	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	casual

In [55]:
```
dfs['Ride_dist'] = dfs.ended_at-dfs.started_at
dfs.head()
```

Out[55]:

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual	Ride_dist
9	B9F73448DFBE0D45	classic_bike	2021-01-24 19:15:38	2021-01-24 19:22:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	member	0 days 00:07:13
10	457C7F4B5D3DA135	electric_bike	2021-01-23 12:57:38	2021-01-23 13:02:10	California Ave & Cortez St	17660	California Ave & North Ave	13258	41.900406	-87.696733	41.910435	-87.696890	member	0 days 00:04:32
11	57C750326F9FDABE	electric_bike	2021-01-09 15:28:04	2021-01-09 15:37:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900374	-87.696688	41.899180	-87.672178	casual	0 days 00:09:47
12	4D518C65E338D070	electric_bike	2021-01-09 15:28:57	2021-01-09 15:37:54	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900379	-87.696716	41.899149	-87.672177	casual	0 days 00:08:57
13	9D08A3AFF410474D	classic_bike	2021-01-24 15:56:59	2021-01-24 16:07:08	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	casual	0 days 00:10:09

In [56]:
```
dfs.rename(columns={'Ride_dist':'Ride_Time'},inplace= True)
```

In [57]:
```
dfs.head()
```

Out[57]:

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual	Ride_Time
9	B9F73448DFBE0D45	classic_bike	2021-01-24 19:15:38	2021-01-24 19:22:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	member	0 days 00:07:13
10	457C7F4B5D3DA135	electric_bike	2021-01-23 12:57:38	2021-01-23 13:02:10	California Ave & Cortez St	17660	California Ave & North Ave	13258	41.900406	-87.696733	41.910435	-87.696890	member	0 days 00:04:32
11	57C750326F9FDABE	electric_bike	2021-01-09 15:28:04	2021-01-09 15:37:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900374	-87.696688	41.899180	-87.672178	casual	0 days 00:09:47
12	4D518C65E338D070	electric_bike	2021-01-09 15:28:57	2021-01-09 15:37:54	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900379	-87.696716	41.899149	-87.672177	casual	0 days 00:08:57
13	9D08A3AFF410474D	classic_bike	2021-01-24 15:56:59	2021-01-24 16:07:08	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	casual	0 days 00:10:09

In [58]:
```
print(dfs.duplicated().value_counts())
```

```
False    4588302
dtype: int64
```

In [59]:
```
dfs['day_of_week'] = dfs['started_at'].dt.dayofweek
```

In [60]:
```
dfs['hour_of_day'] = dfs['started_at'].dt.hour
dfs['month_of_year'] = dfs['started_at'].dt.month
dfs.head()
```

Out[60]:

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual	Ride_Time	day_of_week	hour_o
9	B9F73448DFBE0D45	classic_bike	2021-01-24 19:15:38	2021-01-24 19:22:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	member	0 days 00:07:13	6	
10	457C7F4B5D3DA135	electric_bike	2021-01-23 12:57:38	2021-01-23 13:02:10	California Ave & Cortez St	17660	California Ave & North Ave	13258	41.900406	-87.696733	41.910435	-87.696890	member	0 days 00:04:32	5	
11	57C750326F9FDABE	electric_bike	2021-01-09 15:28:04	2021-01-09 15:37:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900374	-87.696688	41.899180	-87.672178	casual	0 days 00:09:47	5	
12	4D518C65E338D070	electric_bike	2021-01-09 15:28:57	2021-01-09 15:37:54	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900379	-87.696716	41.899149	-87.672177	casual	0 days 00:08:57	5	
13	9D08A3AFF410474D	classic_bike	2021-01-24 15:56:59	2021-01-24 16:07:08	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	casual	0 days 00:10:09	6	

In [61]:
```
dfs.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4588302 entries, 9 to 5595062
Data columns (total 17 columns):
 #   Column              Dtype          
---  ------              -----          
 0   ride_id             object         
 1   rideable_type       category       
 2   started_at          datetime64[ns] 
 3   ended_at            datetime64[ns] 
 4   start_station_name  object         
 5   start_station_id    object         
 6   end_station_name    object         
 7   end_station_id      object         
 8   start_lat           float64        
 9   start_lng           float64        
 10  end_lat             float64        
 11  end_lng             float64        
 12  member_casual       category       
 13  Ride_Time           timedelta64[ns]
 14  day_of_week         int64          
 15  hour_of_day         int64          
 16  month_of_year       int64          
dtypes: category(2), datetime64[ns](2), float64(4), int64(3), object(5), timedelta64[ns](1)
memory usage: 568.8+ MB
```

In [62]:
```
casual_user = dfs[dfs['member_casual'] == 'casual']
membership_user = dfs[dfs['member_casual'] == 'member']
```

In [63]:
```
min_casual_user_ride_time = casual_user['Ride_Time'].min()
min_casual_user_ride_time
```

Out[63]:
```
Timedelta('-1 days +23:04:06')
```

In [64]:
```
min_casual_ride_lenght = casual_user['Ride_Time'].min()
min_member_ride_lenght = membership_user['Ride_Time'].min()

max_casual_ride_lenght = casual_user['Ride_Time'].max()
max_member_ride_lenght = membership_user['Ride_Time'].max()
```

In [65]:
```
print('Minimum casual ride lenght: '+ str(min_casual_ride_lenght))
print('Minimum member ride lenght: '+ str(min_member_ride_lenght))

print('Maximum casual ride lenght: '+ str(min_casual_ride_lenght))
print('Minimum member ride lenght: '+ str(min_member_ride_lenght))
```

```
Minimum casual ride lenght: -1 days +23:04:06
Minimum member ride lenght: -1 days +23:05:55
Maximum casual ride lenght: -1 days +23:04:06
Minimum member ride lenght: -1 days +23:05:55
```

In [66]:
```
mean_casual_ride_lenght = casual_user['Ride_Time'].mean()
mean_member_ride_lenght = membership_user['Ride_Time'].mean()

print('Mean casual rider length: '+str(mean_casual_ride_lenght))
print('Mean member rider length: '+str(mean_member_ride_lenght))
```

```
Mean casual rider length: 0 days 00:32:30.511407800
Mean member rider length: 0 days 00:13:10.983923922
```

In [67]:
```
dfs.head()
```

Out[67]:

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual	Ride_Time	day_of_week	hour_o
9	B9F73448DFBE0D45	classic_bike	2021-01-24 19:15:38	2021-01-24 19:22:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	member	0 days 00:07:13	6	
10	457C7F4B5D3DA135	electric_bike	2021-01-23 12:57:38	2021-01-23 13:02:10	California Ave & Cortez St	17660	California Ave & North Ave	13258	41.900406	-87.696733	41.910435	-87.696890	member	0 days 00:04:32	5	
11	57C750326F9FDABE	electric_bike	2021-01-09 15:28:04	2021-01-09 15:37:51	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900374	-87.696688	41.899180	-87.672178	casual	0 days 00:09:47	5	
12	4D518C65E338D070	electric_bike	2021-01-09 15:28:57	2021-01-09 15:37:54	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900379	-87.696716	41.899149	-87.672177	casual	0 days 00:08:57	5	
13	9D08A3AFF410474D	classic_bike	2021-01-24 15:56:59	2021-01-24 16:07:08	California Ave & Cortez St	17660	Wood St & Augusta Blvd	657	41.900363	-87.696704	41.899181	-87.672200	casual	0 days 00:10:09	6	

In [68]:
```
pivote_by_day_df = dfs.pivot_table(index= 'day_of_week',
                    columns= 'member_casual',
                    values = 'ride_id',
                    aggfunc= len,
                    fill_value= '0')

pivote_by_day_df.set_axis(['mon', 'tues', 'wed', 'thurs', 'fri', 'sat', 'sum'], axis= 0, inplace= True)
pivote_by_day_df
```

Out[68]:

member_casual	casual	member
mon	228936	346488
tues	214937	388132
wed	218134	397720
thurs	224207	373474
fri	290045	365787
sat	468331	357082
sum	403789	311240

In [70]:
```
dfs['ride_id'].value_counts().sum()
```

Out[70]:
```
4588302
```