In [1]:
```
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
```

In [2]:
```
df= pd.read_csv('dailyActivity_merged.csv')
```

In [3]:
```
df
```

Out[3]:

	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActi
0	1503960366	4/12/2016	13162	8.500000	8.500000	0.0	1.88	
1	1503960366	4/13/2016	10735	6.970000	6.970000	0.0	1.57	
2	1503960366	4/14/2016	10460	6.740000	6.740000	0.0	2.44	
3	1503960366	4/15/2016	9762	6.280000	6.280000	0.0	2.14	
4	1503960366	4/16/2016	12669	8.160000	8.160000	0.0	2.71	
...	...	...	...	...	...	...	...	
935	8877689391	5/8/2016	10686	8.110000	8.110000	0.0	1.08	
936	8877689391	5/9/2016	20226	18.250000	18.250000	0.0	11.10	
937	8877689391	5/10/2016	10733	8.150000	8.150000	0.0	1.35	
938	8877689391	5/11/2016	21420	19.559999	19.559999	0.0	13.22	
939	8877689391	5/12/2016	8064	6.120000	6.120000	0.0	1.82	

940 rows × 15 columns

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

Out[4]:

	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActi
935	8877689391	5/8/2016	10686	8.110000	8.110000	0.0	1.08	
936	8877689391	5/9/2016	20226	18.250000	18.250000	0.0	11.10	
937	8877689391	5/10/2016	10733	8.150000	8.150000	0.0	1.35	
938	8877689391	5/11/2016	21420	19.559999	19.559999	0.0	13.22	
939	8877689391	5/12/2016	8064	6.120000	6.120000	0.0	1.82	

In [5]:
```
print(df.isnull().sum())
```

```
Id                          0
ActivityDate                0
TotalSteps                  0
TotalDistance               0
TrackerDistance             0
LoggedActivitiesDistance    0
VeryActiveDistance          0
ModeratelyActiveDistance    0
LightActiveDistance         0
SedentaryActiveDistance     0
VeryActiveMinutes           0
FairlyActiveMinutes         0
LightlyActiveMinutes        0
SedentaryMinutes            0
Calories                    0
dtype: int64
```

In [6]:
```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 940 entries, 0 to 939
Data columns (total 15 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   Id                         940 non-null    int64  
 1   ActivityDate               940 non-null    object 
 2   TotalSteps                 940 non-null    int64  
 3   TotalDistance              940 non-null    float64
 4   TrackerDistance            940 non-null    float64
 5   LoggedActivitiesDistance   940 non-null    float64
 6   VeryActiveDistance         940 non-null    float64
 7   ModeratelyActiveDistance   940 non-null    float64
 8   LightActiveDistance        940 non-null    float64
 9   SedentaryActiveDistance    940 non-null    float64
 10  VeryActiveMinutes          940 non-null    int64  
 11  FairlyActiveMinutes        940 non-null    int64  
 12  LightlyActiveMinutes       940 non-null    int64  
 13  SedentaryMinutes           940 non-null    int64  
 14  Calories                   940 non-null    int64  
dtypes: float64(7), int64(7), object(1)
memory usage: 110.3+ KB
None
```

In [7]:
```
# Changing datatype of ActivityDate

df['ActivityDate']= pd.to_datetime(df['ActivityDate'],format="%m/%d/%Y")
```

In [8]:
```
df['ActivityDate']
```

Out[8]:
```
0     2016-04-12
1     2016-04-13
2     2016-04-14
3     2016-04-15
4     2016-04-16
         ...
935   2016-05-08
936   2016-05-09
937   2016-05-10
938   2016-05-11
939   2016-05-12
Name: ActivityDate, Length: 940, dtype: datetime64[ns]
```

In [9]:
```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 940 entries, 0 to 939
Data columns (total 15 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Id                        940 non-null    int64
 1   ActivityDate              940 non-null    datetime64[ns]
 2   TotalSteps                940 non-null    int64
 3   TotalDistance             940 non-null    float64
 4   TrackerDistance           940 non-null    float64
 5   LoggedActivitiesDistance  940 non-null    float64
 6   VeryActiveDistance        940 non-null    float64
 7   ModeratelyActiveDistance  940 non-null    float64
 8   LightActiveDistance       940 non-null    float64
 9   SedentaryActiveDistance   940 non-null    float64
 10  VeryActiveMinutes         940 non-null    int64
 11  FairlyActiveMinutes       940 non-null    int64
 12  LightlyActiveMinutes      940 non-null    int64
 13  SedentaryMinutes          940 non-null    int64
 14  Calories                  940 non-null    int64
dtypes: datetime64[ns](1), float64(7), int64(7)
memory usage: 110.3 KB
None
```

In [10]:
```
df['TotalMinutes']= df['VeryActiveMinutes']+ df['FairlyActiveMinutes']+ df['LightlyActiveMinutes'] +df['Seder
```

In [11]:
```
df['TotalMinutes']
```

Out[11]:
```
0      1094
1      1033
2      1440
3       998
4      1040
       ...
935    1440
936    1440
937    1440
938    1440
939     931
Name: TotalMinutes, Length: 940, dtype: int64
```

In [12]:
```
df
```

Out[12]:

	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActi
0	1503960366	2016-04-12	13162	8.500000	8.500000	0.0	1.88	
1	1503960366	2016-04-13	10735	6.970000	6.970000	0.0	1.57	
2	1503960366	2016-04-14	10460	6.740000	6.740000	0.0	2.44	
3	1503960366	2016-04-15	9762	6.280000	6.280000	0.0	2.14	
4	1503960366	2016-04-16	12669	8.160000	8.160000	0.0	2.71	
...	...	...	...	...	...	...	...	
935	8877689391	2016-05-08	10686	8.110000	8.110000	0.0	1.08	
936	8877689391	2016-05-09	20226	18.250000	18.250000	0.0	11.10	
937	8877689391	2016-05-10	10733	8.150000	8.150000	0.0	1.35	
938	8877689391	2016-05-11	21420	19.559999	19.559999	0.0	13.22	
939	8877689391	2016-05-12	8064	6.120000	6.120000	0.0	1.82	

940 rows × 16 columns

In [13]:
```
print(df.describe())
```

	Id	TotalSteps	TotalDistance	TrackerDistance
count	9.400000e+02	940.000000	940.000000	940.000000
mean	4.855407e+09	7637.910638	5.489702	5.475351
std	2.424805e+09	5087.150742	3.924606	3.907276
min	1.503960e+09	0.000000	0.000000	0.000000
25%	2.320127e+09	3789.750000	2.620000	2.620000
50%	4.445115e+09	7405.500000	5.245000	5.245000
75%	6.962181e+09	10727.000000	7.712500	7.710000
max	8.877689e+09	36019.000000	28.030001	28.030001

	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActiveDistance
count	940.000000	940.000000	940.000000
mean	0.108171	1.502681	0.567543
std	0.619897	2.658941	0.883580
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.210000	0.240000
75%	0.000000	2.052500	0.800000
max	4.942142	21.920000	6.480000

	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes
count	940.000000	940.000000	940.000000
mean	3.340819	0.001606	21.164894
std	2.040655	0.007346	32.844803
min	0.000000	0.000000	0.000000
25%	1.945000	0.000000	0.000000
50%	3.365000	0.000000	4.000000
75%	4.782500	0.000000	32.000000
max	10.710000	0.110000	210.000000

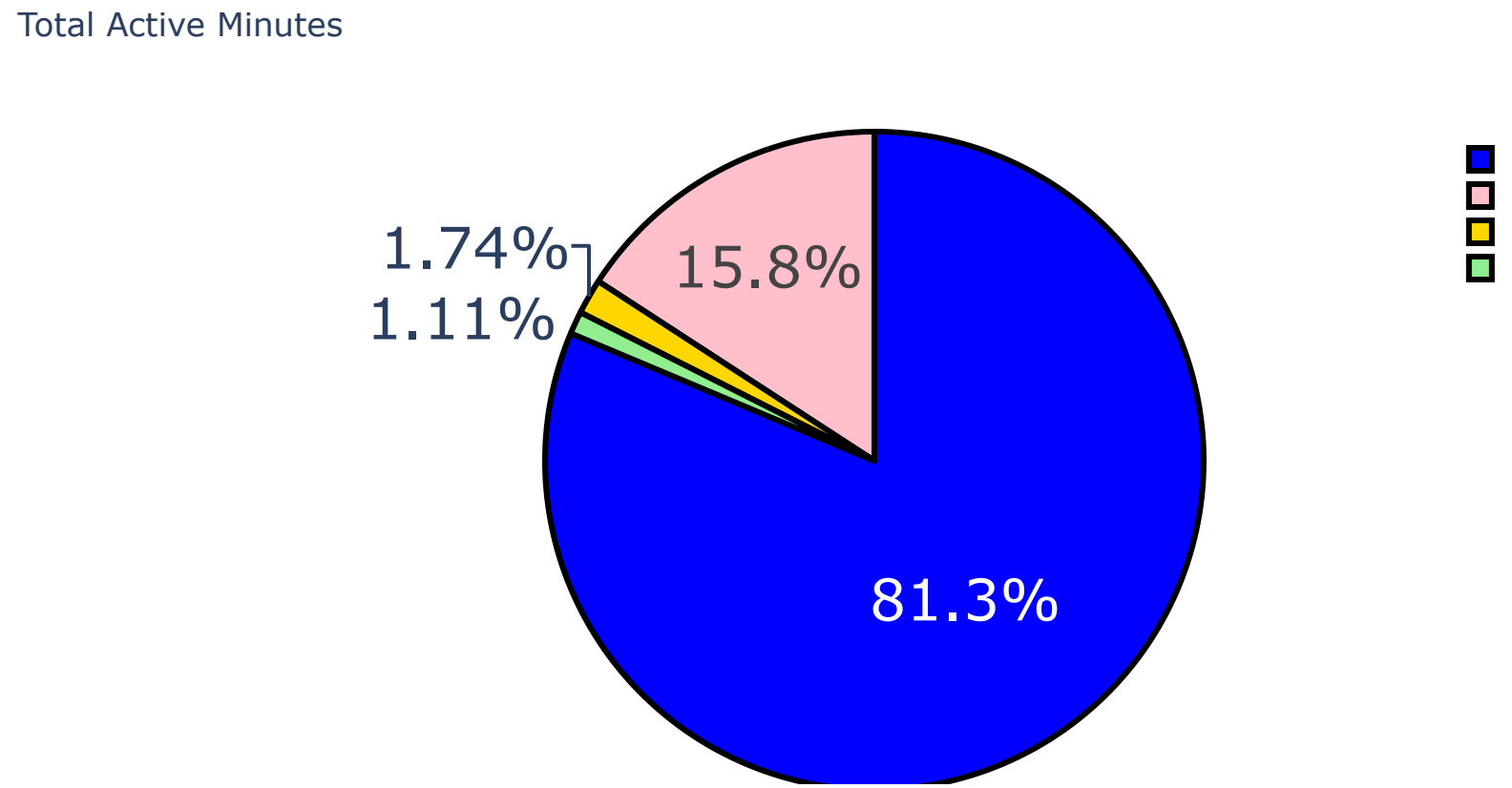
	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes
count	940.000000	940.000000	940.000000
mean	13.564894	192.812766	991.210638
std	19.987404	109.174700	301.267437
min	0.000000	0.000000	0.000000
25%	0.000000	127.000000	729.750000
50%	6.000000	199.000000	1057.500000
75%	19.000000	264.000000	1229.500000
max	143.000000	518.000000	1440.000000

	Calories	TotalMinutes
count	940.000000	940.000000
mean	2303.609574	1218.753191

std	718.166862	265.931767
min	0.000000	2.000000
25%	1828.500000	989.750000
50%	2134.000000	1440.000000
75%	2793.250000	1440.000000
max	4900.000000	1440.000000

In [14]:
```
figure= px.scatter(data_frame=df, x='Calories' , y='TotalSteps', size= 'VeryActiveMinutes',trendline="ols",t
figure.show()
```

In [15]:
```
label = ["Very Active Minutes", "Fairly Active Minutes","Lightly Active Minutes", "Inactive Minutes"]
counts = df[["VeryActiveMinutes", "FairlyActiveMinutes",
               "LightlyActiveMinutes", "SedentaryMinutes"]].mean()
colors = ['gold','lightgreen', "pink", "blue"]
fig = go.Figure(data=[go.Pie(labels=label, values=counts)])
fig.update_layout(title_text='Total Active Minutes')
fig.update_traces(hoverinfo='value', textinfo='percent', textfont_size=30,
                  marker=dict(colors=colors, line=dict(color='black', width=3)))
fig.show()
```

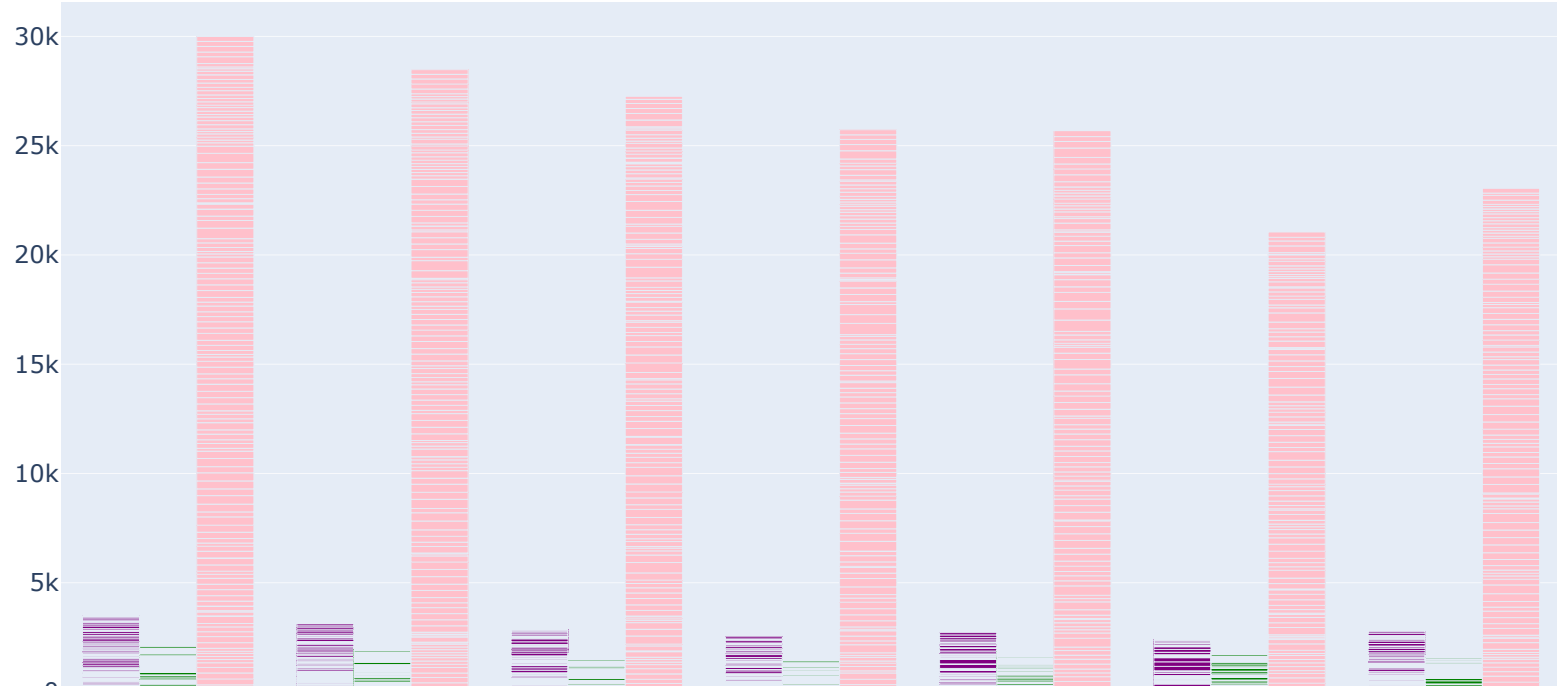


In [16]:
```
df['Date']= df['ActivityDate'].dt.day_name()
```

In [17]:
```
df['Date'].unique()
```

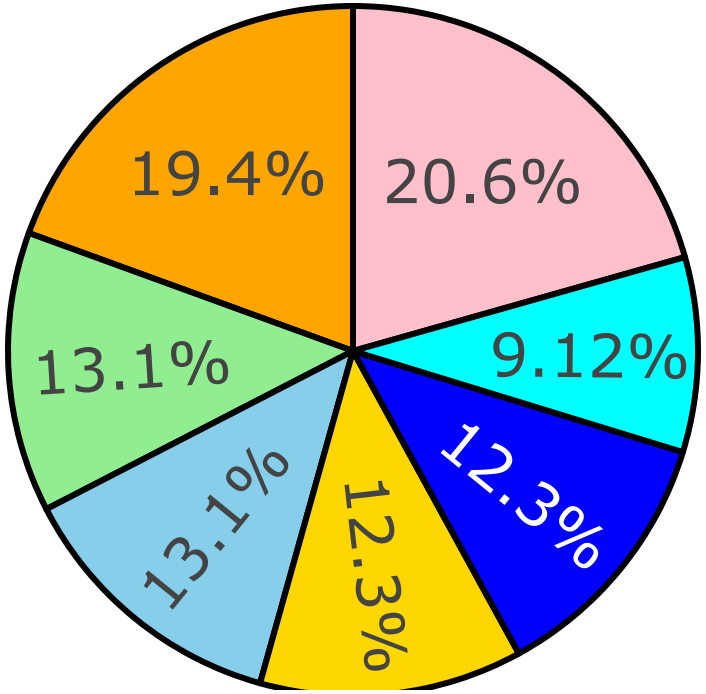
Out[17]:
```
array(['Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday',
       'Monday'], dtype=object)
```

In [18]:
```
fig= go.Figure()
fig.add_trace(go.Bar(x=df['Date'],y=df['VeryActiveMinutes'],name='VeryActive',  marker_color='purple'))
fig.add_trace(go.Bar(x=df['Date'],y=df['FairlyActiveMinutes'],name='Fairly Active',  marker_color='green'))
fig.add_trace(go.Bar(x=df['Date'],y=df['LightlyActiveMinutes'],name='Lightly Active',  marker_color='pink'))
fig.update_layout(barmode='group',xaxis_tickangle=-45)
fig.show()
```

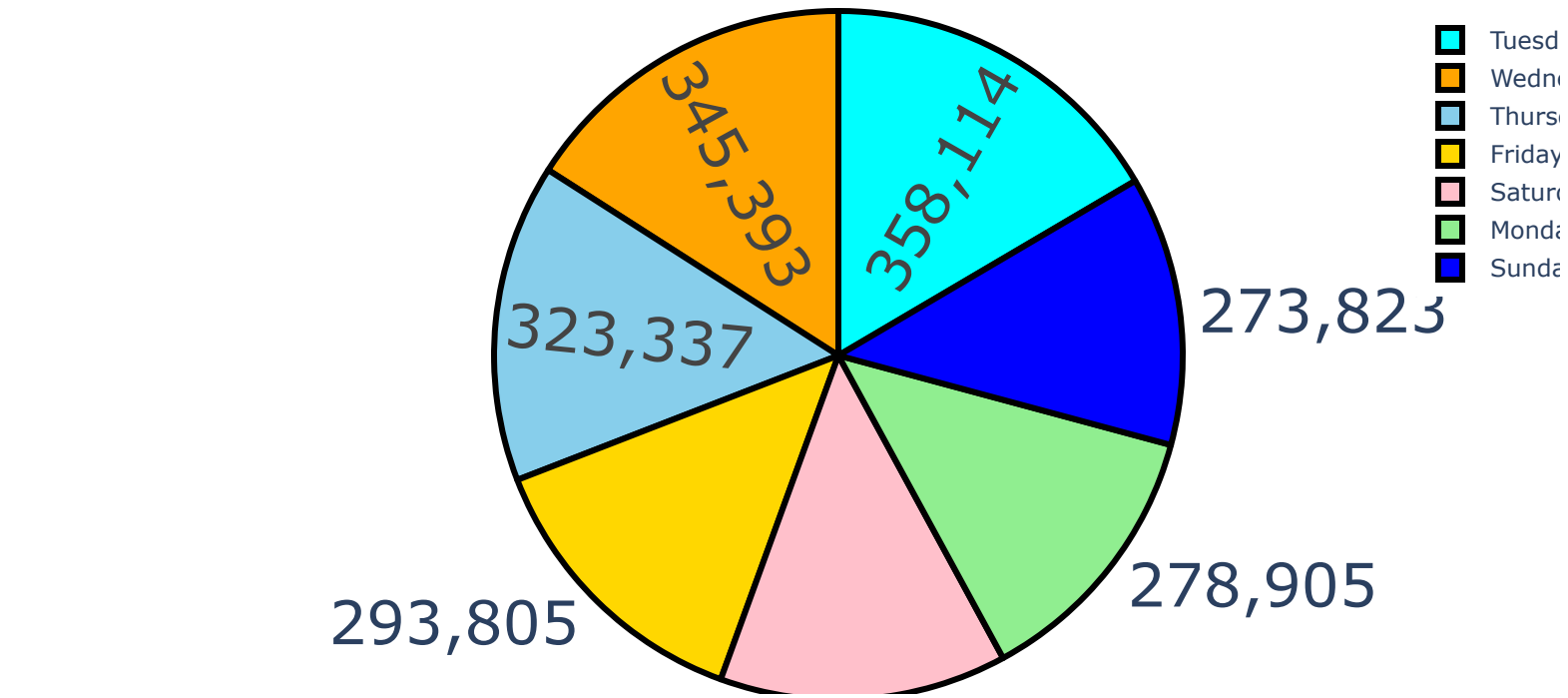


In [20]:
```
day = df['Date'].value_counts()
label = day.index
counts = df['SedentaryMinutes']
colors = ['gold','lightgreen', "pink", "blue", "skyblue", "cyan", "orange"]

fig = go.Figure(data=[go.Pie(labels=label, values=counts)])

fig.update_layout(title_text='Inactive Minutes Daily')
fig.update_traces(hoverinfo='label+percent', textinfo='percent', textfont_size=30,
                  marker=dict(colors=colors, line=dict(color='black', width=3)))

fig.show()
```

Inactive Minutes Daily



In [21]:
```
day = df.groupby('Date')['Calories'].sum()

label = day.index
counts = day.values

colors = ['gold','lightgreen', "pink", "blue", "skyblue", "cyan", "orange"]

# Creating the pie chart
fig = go.Figure(data=[go.Pie(labels=label, values=counts)])

# Updating the layout and traces
fig.update_layout(title_text='Calories Burned Daily (Summed by Date)')
fig.update_traces(hoverinfo='label+percent', textinfo='value', textfont_size=30,
                  marker=dict(colors=colors, line=dict(color='black', width=3)))

# Show the pie chart
fig.show()
```


Calories Burned Daily (Summed by Date)
358,114
273,823
278,905
293,805
323,337
345,393
Tuesd
Wedn
Thurs
Friday
Saturd
Monda
Sunda

In [ ]: