

```
In [20]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from dateutil import parser
```

```
In [4]: crime = pd.read_csv("data.csv")
```

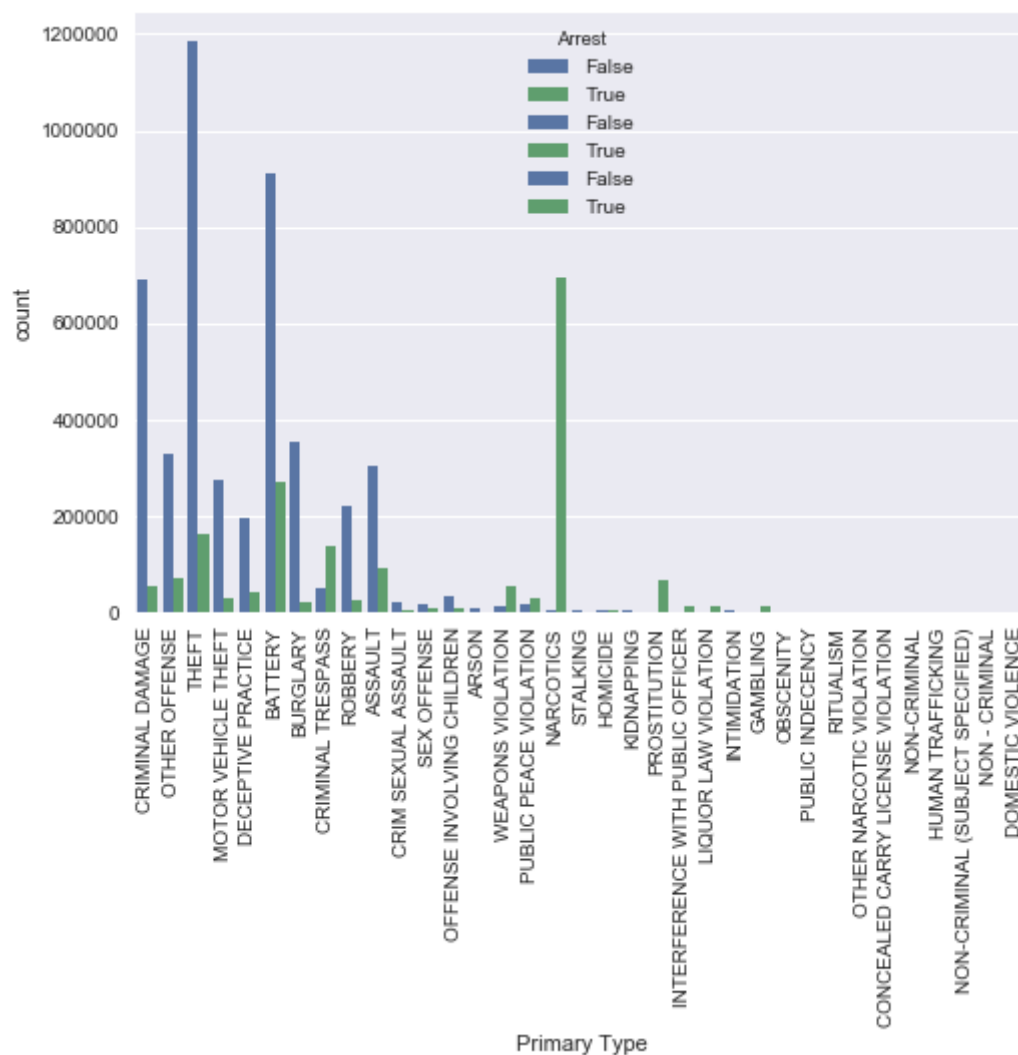
```
In [5]: crime[:3]
```

```
Out[5]:
```

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
0	6530342	HP603653	10/01/2008 03:30:00 PM	076XX S COTTAGE GROVE AVE	0460	BATTERY	SIMPLE	SIDEWALK
1	6530343	HP603773	10/01/2008 05:00:00 PM	031XX S GREEN ST	0620	BURGLARY	UNLAWFUL ENTRY	RESIDENCE
2	6530344	HP600313	09/29/2008 04:15:00 PM	063XX S STONY ISLAND AVE	0890	THEFT	FROM BUILDING	OTHER

3 rows × 22 columns

```
In [6]: plt.xticks(rotation=90)
sns.countplot(x='Primary Type', hue='Arrest', data=crime)
plt.show()
```



```
In [68]: mini_crime = crime[:100]
def convert_to_datetime(row) :
    dt = parser.parse(row["Date"])
    return dt

def label_day_of_week (row) :
    dt = row["Datetime"]
    return dt.weekday()

def label_day_of_month (row) :
    dt = row["Datetime"]
    return dt.day

def label_month(row) :
    dt = row["Datetime"]
    return dt.month

def label_year(row) :
    dt = row["Datetime"]
    return dt.year

def label_hour(row) :
    dt = row["Datetime"]
    return dt.hour
```

```
In [32]: crime["Datetime"] = crime.apply(lambda row : convert_to_datetime(row), a
axis = 1)
crime[:3]
```

Out[32]:

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
0	6530342	HP603653	10/01/2008 03:30:00 PM	076XX S COTTAGE GROVE AVE	0460	BATTERY	SIMPLE	SIDEWALK
1	6530343	HP603773	10/01/2008 05:00:00 PM	031XX S GREEN ST	0620	BURGLARY	UNLAWFUL ENTRY	RESIDENCE
2	6530344	HP600313	09/29/2008 04:15:00 PM	063XX S STONY ISLAND AVE	0890	THEFT	FROM BUILDING	OTHER

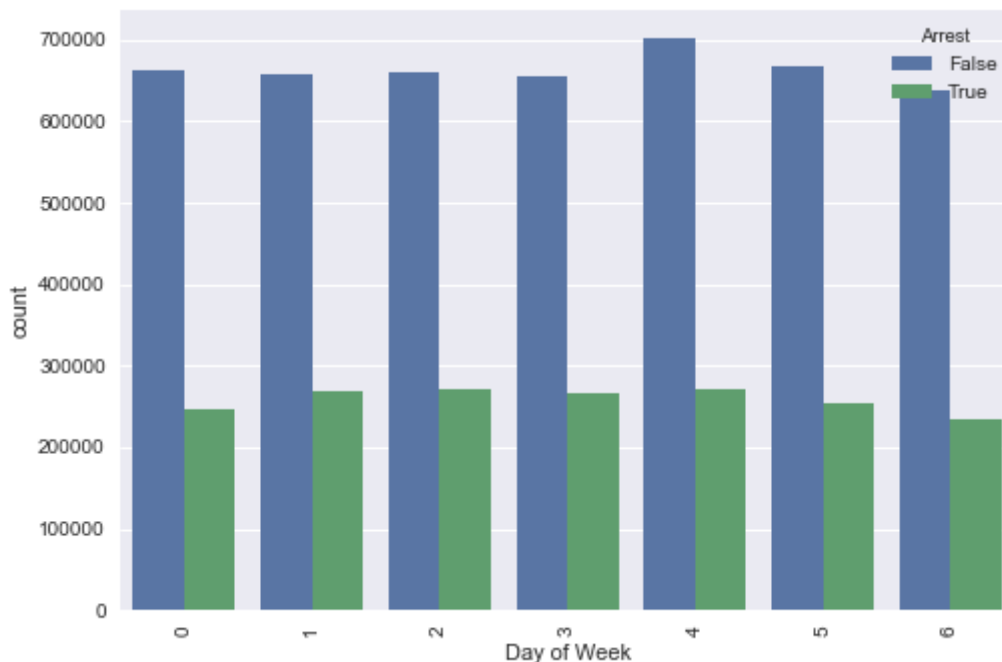
3 rows × 23 columns

```
In [36]: crime["Day of Week"] = crime.apply(lambda row : label_day_of_week(row),
axis = 1)
```

```
In [37]: crime["Day Of Month"] = crime.apply(lambda row :
label_day_of_month(row), axis = 1)
```

```
In [38]: crime["Month"] = crime.apply(lambda row : label_month(row), axis = 1)
```

```
In [39]: plt.xticks(rotation=90)
sns.countplot(x='Day of Week', hue='Arrest', data=crime)
plt.show()
```



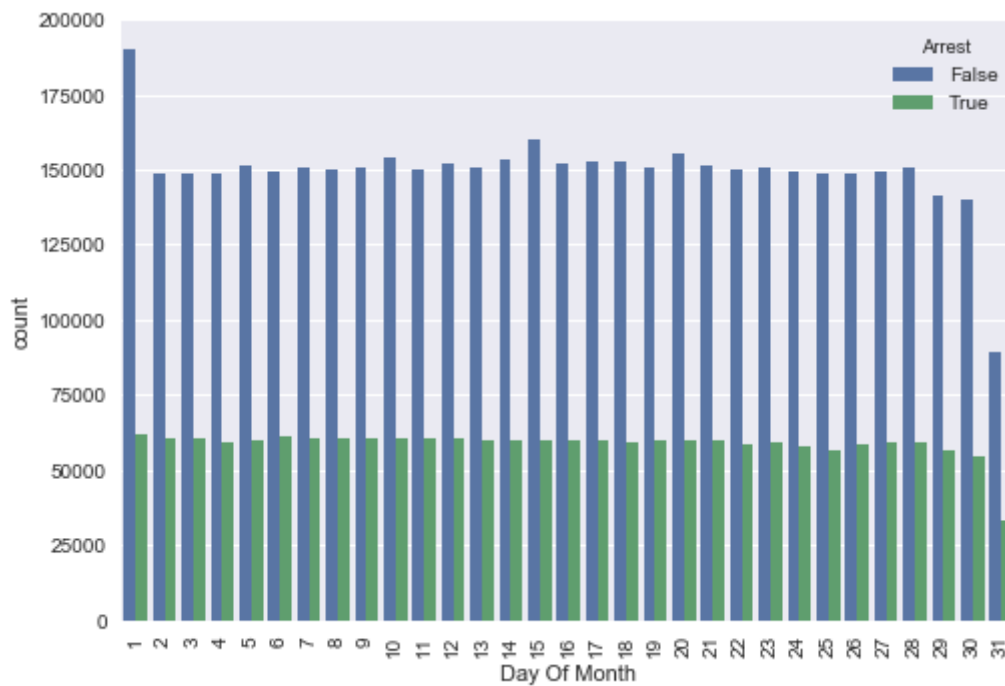
```
In [41]: crime[:3]
```

```
Out[41]:
```

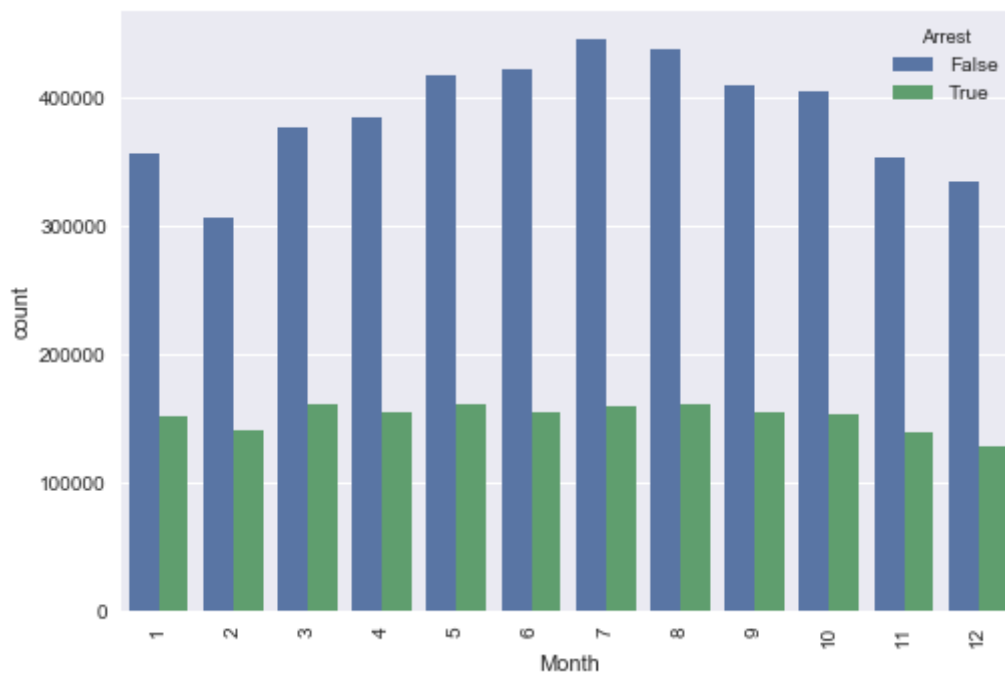
	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
0	6530342	HP603653	10/01/2008 03:30:00 PM	076XX S COTTAGE GROVE AVE	0460	BATTERY	SIMPLE	SIDEWALK
1	6530343	HP603773	10/01/2008 05:00:00 PM	031XX S GREEN ST	0620	BURGLARY	UNLAWFUL ENTRY	RESIDENCE
2	6530344	HP600313	09/29/2008 04:15:00 PM	063XX S STONY ISLAND AVE	0890	THEFT	FROM BUILDING	OTHER

3 rows × 26 columns

```
In [42]: plt.xticks(rotation=90)
sns.countplot(x="Day Of Month", hue='Arrest', data=crime)
plt.show()
```



```
In [43]: plt.xticks(rotation=90)
sns.countplot(x='Month', hue='Arrest', data=crime)
plt.show()
```



```
In [58]: usable_crime = crime.copy(True)
usable_crime["Year"] = usable_crime.apply(lambda row : label_year(row),
axis=1)
usable_crime[0:3]
```

Out[58]:

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
0	6530342	HP603653	10/01/2008 03:30:00 PM	076XX S COTTAGE GROVE AVE	0460	BATTERY	SIMPLE	SIDEWALK
1	6530343	HP603773	10/01/2008 05:00:00 PM	031XX S GREEN ST	0620	BURGLARY	UNLAWFUL ENTRY	RESIDENCE
2	6530344	HP600313	09/29/2008 04:15:00 PM	063XX S STONY ISLAND AVE	0890	THEFT	FROM BUILDING	OTHER

3 rows × 26 columns

```
In [60]: usable_crime = usable_crime[
    (usable_crime["Primary Type"] != "GAMBLING") &
    (usable_crime["Primary Type"] != "LIQUOR LAW VIOLATION") &
    (usable_crime["Primary Type"] != "PROSTITUTION") &
    (usable_crime["Primary Type"] != "NARCOTICS") &
    (usable_crime["Primary Type"] != "PUBLIC INDECENCY")
]
```

```
In [ ]: usable_crime.drop(inplace = True, labels = ["ID", "Case Number", "Date",
    "Block", "Primary Type", "Description", "Beat", "Ward", "FBI Code", "Up
    dated_on", "Location"])
```

```
In [ ]: usable_crime["Day of Week"] = usable_crime.apply(lambda row : label_day_
    of_week(row), axis = 1)
```

```
In [ ]: usable_crime["Day Of Month"] = usable_crime.apply(lambda row : label_day
    _of_month(row), axis = 1)
```

```
In [ ]: usable_crime["Month"] = usable_crime.apply(lambda row :
    label_month(row), axis = 1)
```

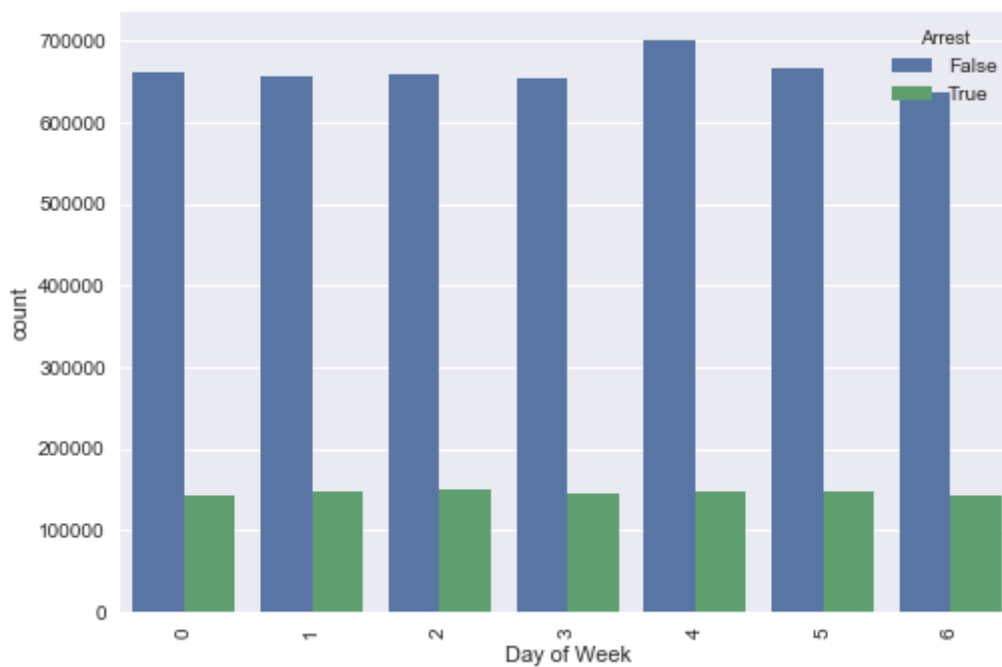
```
In [69]: usable_crime["Hour"] = usable_crime.apply(lambda row : label_hour(row),
axis = 1)
usable_crime[:3]
```

Out[69]:

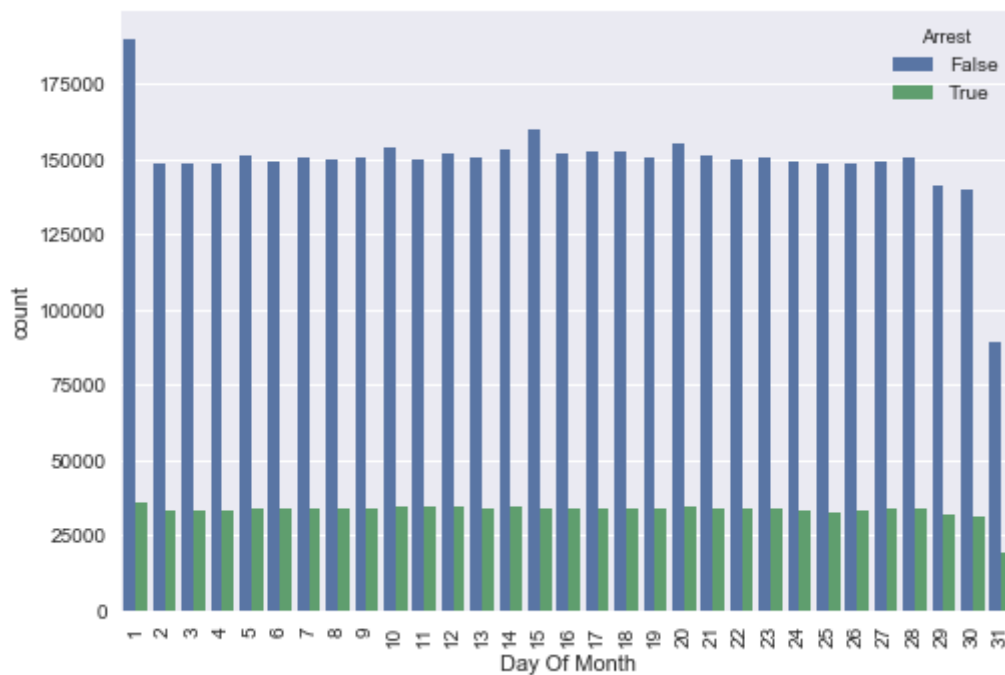
	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
0	6530342	HP603653	10/01/2008 03:30:00 PM	076XX S COTTAGE GROVE AVE	0460	BATTERY	SIMPLE	SIDEWALK
1	6530343	HP603773	10/01/2008 05:00:00 PM	031XX S GREEN ST	0620	BURGLARY	UNLAWFUL ENTRY	RESIDENCE
2	6530344	HP600313	09/29/2008 04:15:00 PM	063XX S STONY ISLAND AVE	0890	THEFT	FROM BUILDING	OTHER

3 rows × 27 columns

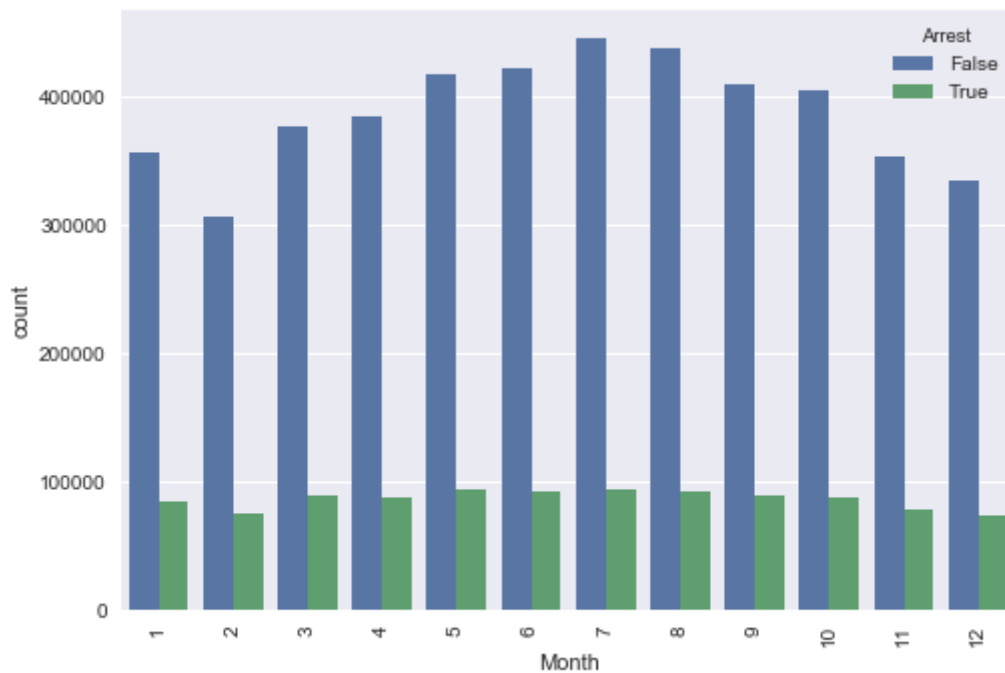
```
In [52]: plt.xticks(rotation=90)
sns.countplot(x='Day of Week', hue='Arrest', data=usable_crime)
plt.show()
```



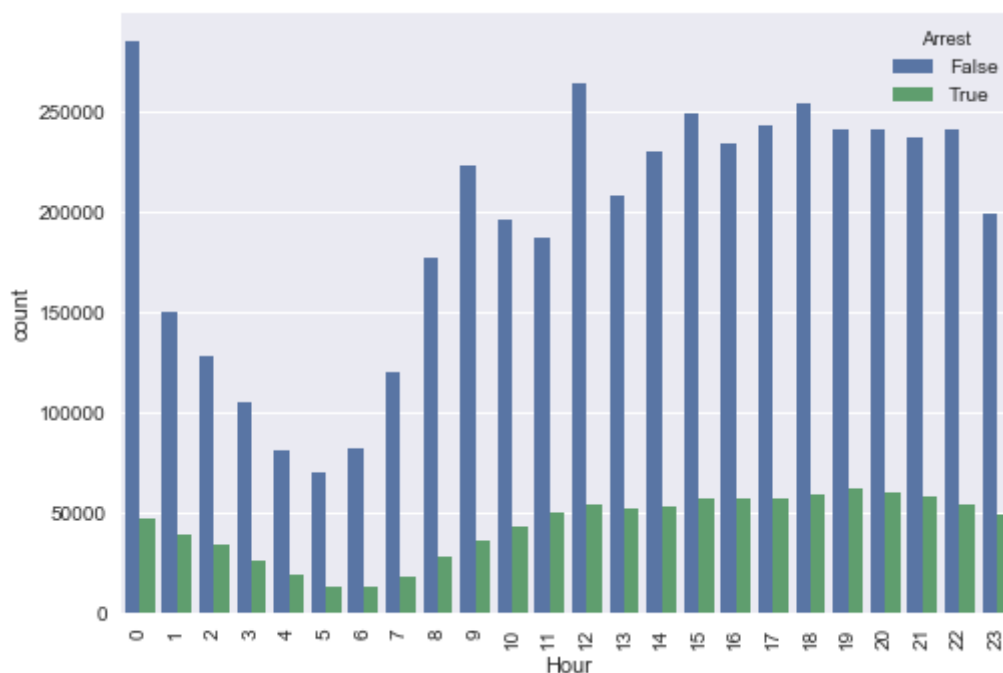
```
In [53]: plt.xticks(rotation=90)
sns.countplot(x="Day Of Month", hue='Arrest', data=usable_crime)
plt.show()
```



```
In [54]: plt.xticks(rotation=90)
sns.countplot(x='Month', hue='Arrest', data=usable_crime)
plt.show()
```




```
In [70]: plt.xticks(rotation=90)
sns.countplot(x='Hour', hue='Arrest', data=usable_crime)
plt.show()
```



```
In [67]: logistic_crime = usable_crime.copy(True)
```

```
Out[67]:
```

	IUCR	Location Description	Arrest	Domestic	District	Community Area	X Coordinate	Y Coordinate	Year
0	0460	SIDEWALK	False	False	6.0	69.0	1182884.0	1854628.0	2001
1	0620	RESIDENCE	False	False	9.0	60.0	1171221.0	1884004.0	2001
2	0890	OTHER	False	False	3.0	42.0	1187961.0	1863261.0	2001

```
In [ ]: logistic_crime = logistic_crime[
    (logistic_crime["Year"] != "GAMBLING") &
    (usable_crime["Primary Type"] != "LIQUOR LAW VIOLATION") &
    (usable_crime["Primary Type"] != "PROSTITUTION") &
    (usable_crime["Primary Type"] != "NARCOTICS") &
    (usable_crime["Primary Type"] != "PUBLIC INDECENCY")
]
```

```
In [ ]: logistic_crime.drop(["ID", "Case Number", "Date", "Block", "Primary Type", "Description", "Beat", "Ward", "FBI Code", "Updated On", "Location"], inplace = True, axis=1)
logistic_crime[:3]

logistic_crime['sin_hour'] =
np.sin(2*np.pi*logistic_crime["Hour"]/24)
logistic_crime['cos_hour'] = np.cos(2*np.pi*logistic_crime["Hour"]/24)

logistic_crime['sin_month'] = np.sin(2*np.pi*(logistic_crime["Month"] - 1)/12)
logistic_crime['cos_month'] = np.cos(2*np.pi*(logistic_crime["Month"] - 1)/12)
```

```
In [ ]:
```