Untitled

November 15, 2020

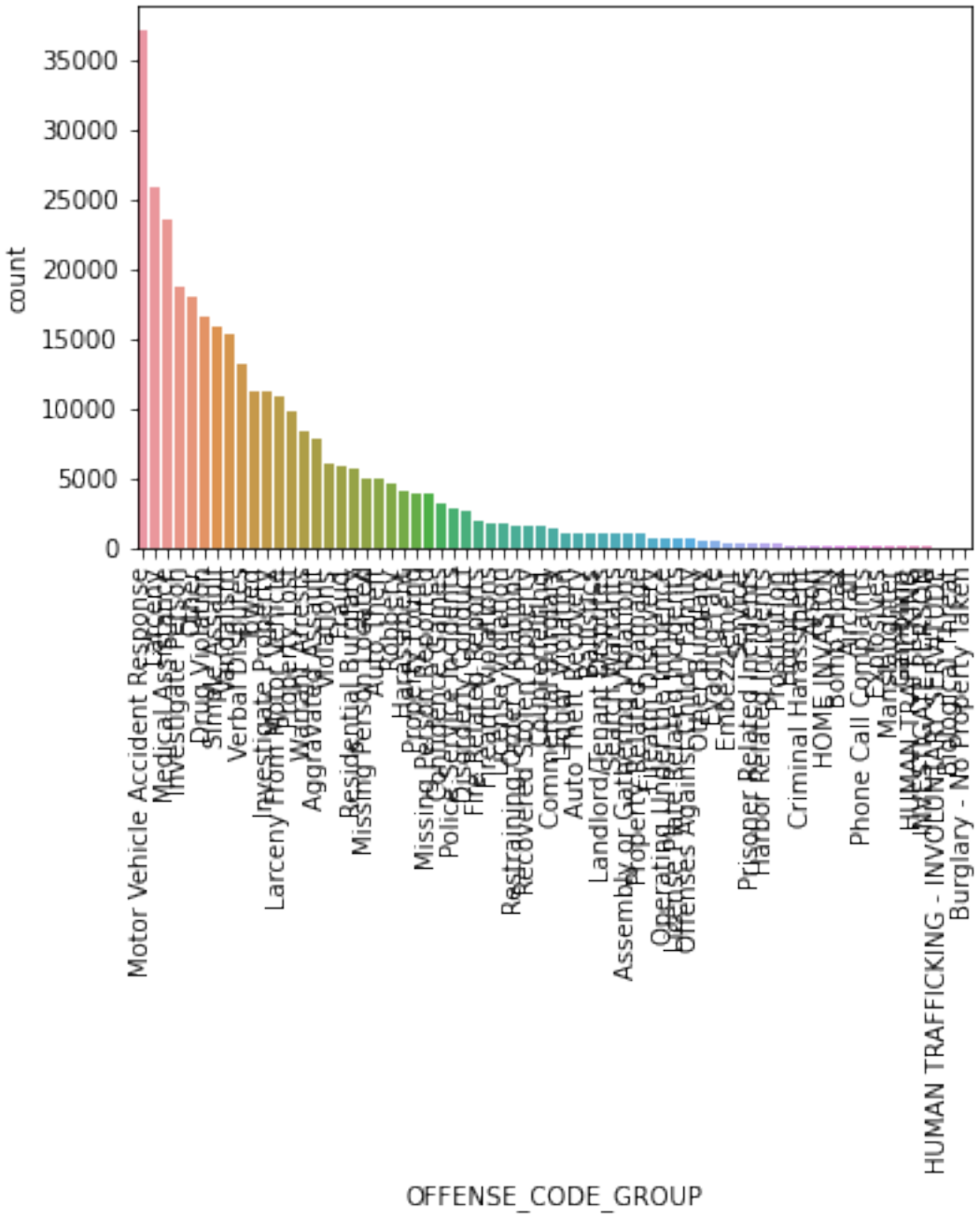
[2]:
```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
```

[3]:
```
# Read data in
df = pd.read_csv("crime.csv", encoding="latin-1")
```

1 Leitfragen 1/3

Welche Straftaten sind am häufigsten?

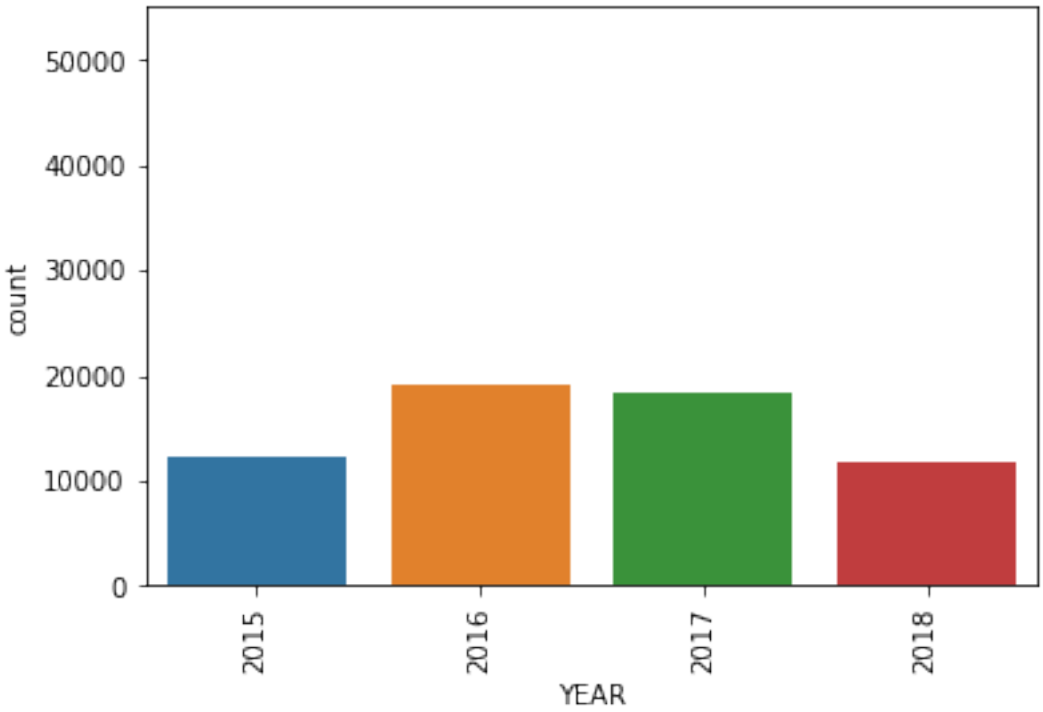
[23]:
```
plot = sns.countplot(x="OFFENSE_CODE_GROUP",
                     data = df,
                     order = df['OFFENSE_CODE_GROUP'].value_counts().index)

_ = plt.setp(plot.get_xticklabels(), rotation=90)
```



1.0.1 Wie hat sich die Zahl der schweren Straftaten im Laufe der Jahre entwickelt?

[5]:
```
table = df.groupby(["OFFENSE_CODE_GROUP"]).INCIDENT_NUMBER.count().to_frame()
table = table.sort_values('INCIDENT_NUMBER', ascending=False)
table = df.groupby(["UCR_PART", "YEAR"]).INCIDENT_NUMBER.count()
table.T.plot(kind='bar', stacked=True)
```

```
table = df[df['UCR_PART'] == "Part One"]
#table = table.groupby(["YEAR"], as_index=False).INCIDENT_NUMBER.count()
plot = sns.countplot(x="YEAR",
                     data = table)
```



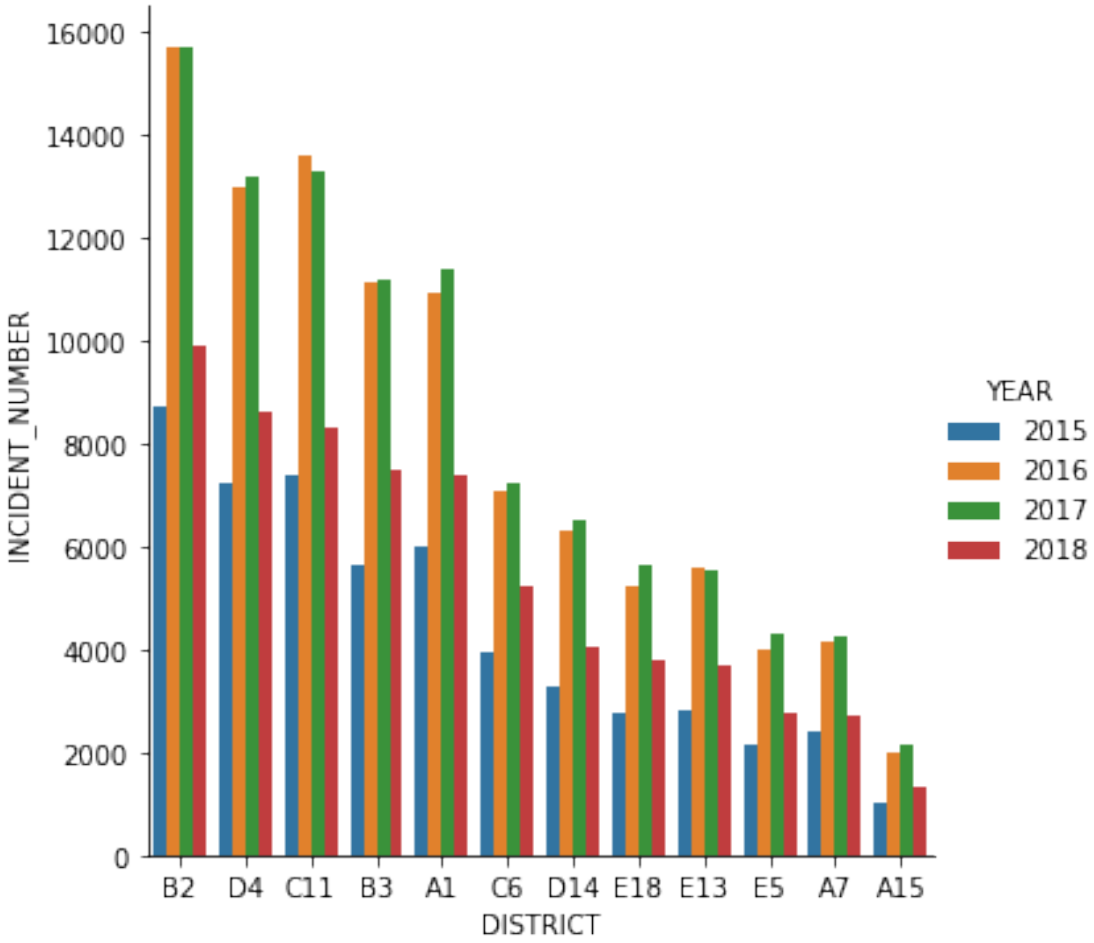
1.0.2 Warum ist die Gesamtzahl 2015 und 2018 (so) niedrig?

[6]:
```
min(df['OCCURRED_ON_DATE'])
max(df['OCCURRED_ON_DATE'])
```

[6]: '2018-09-03 21:25:00'

1.0.3 In welchen Stadtgebieten werden, aufgeschlüsselt nach Jahr, die meisten Straftaten begangen?

[7]:
```
table = df.groupby(["DISTRICT", "YEAR"],  as_index=False).INCIDENT_NUMBER.
 ↪count()
table = table.sort_values(['YEAR', 'INCIDENT_NUMBER'], ascending=False)

plot = sns.catplot(data=table, kind="bar", x="DISTRICT", y="INCIDENT_NUMBER",␣
 ↪hue="YEAR")
```



1.0.4 In welchen Stadtgebieten werden die meisten schweren Straftaten ('Part One') begangen?

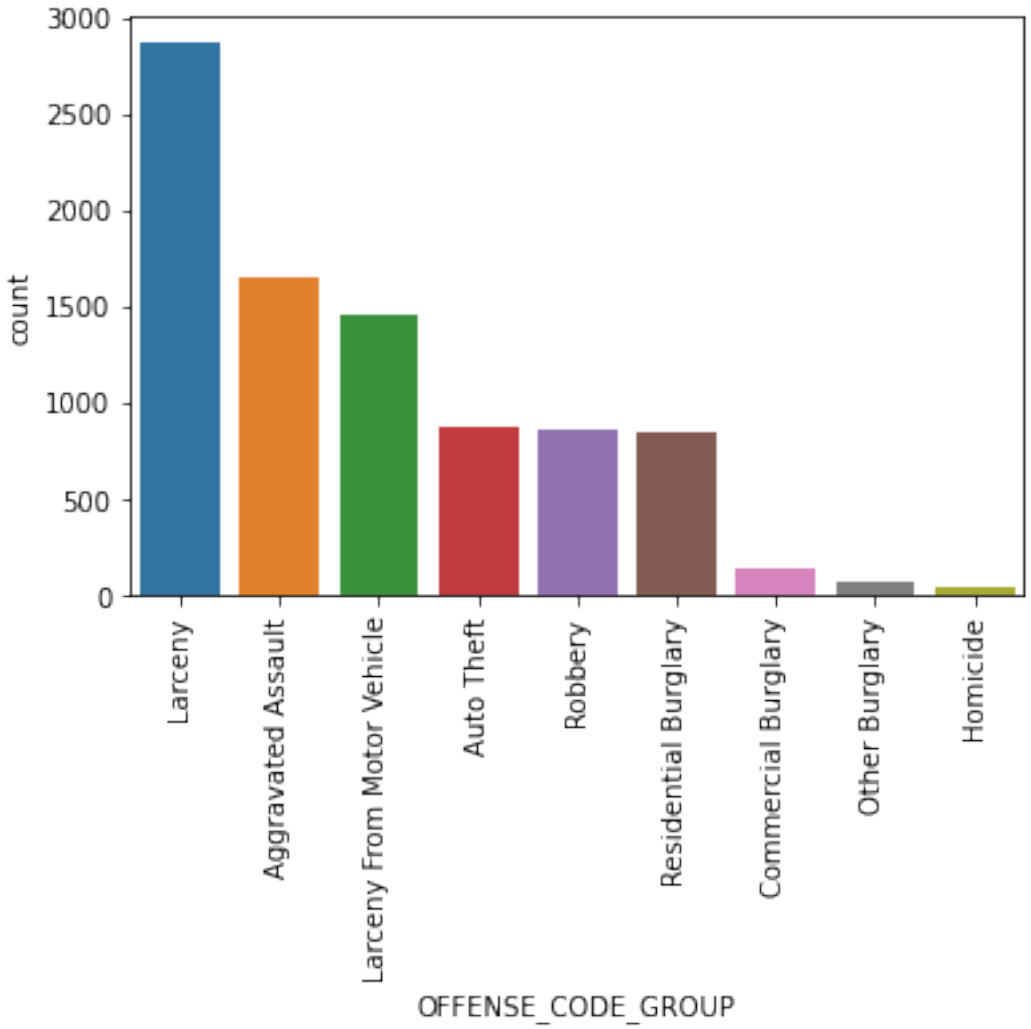
[8]:
```
table = df[df['UCR_PART'] == "Part One"]
plot = sns.countplot(x="DISTRICT", data = table, order =␣
↪table['INCIDENT_NUMBER'].value_counts().index)
```

1.0.5 Welche Arten von schweren Straftaten ('Part One') treten in dem Stadtgebiet 'B2' am häufigsten auf?

[9]:
```
filtered = df[(df['UCR_PART'] == "Part One") & (df['DISTRICT'] == "B2")]

plot = sns.countplot(x="OFFENSE_CODE_GROUP",
              data = filtered,
              order = filtered['OFFENSE_CODE_GROUP'].value_counts().index)
plt.setp(plot.get_xticklabels(), rotation=90)
```

[9]:
```
[None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

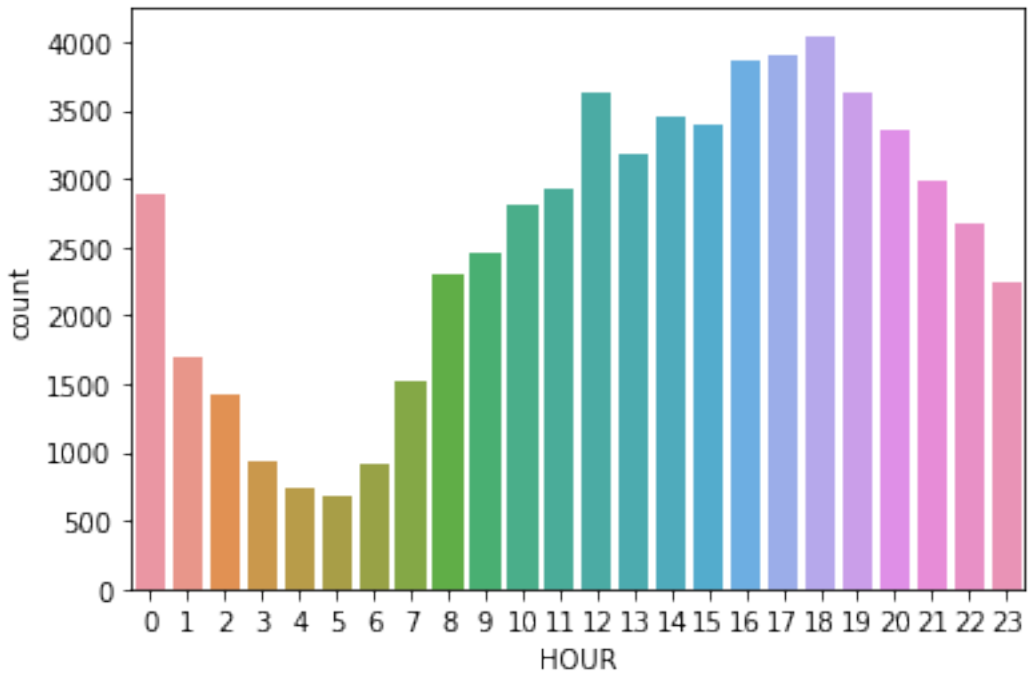
```
None,
None,
None]
```



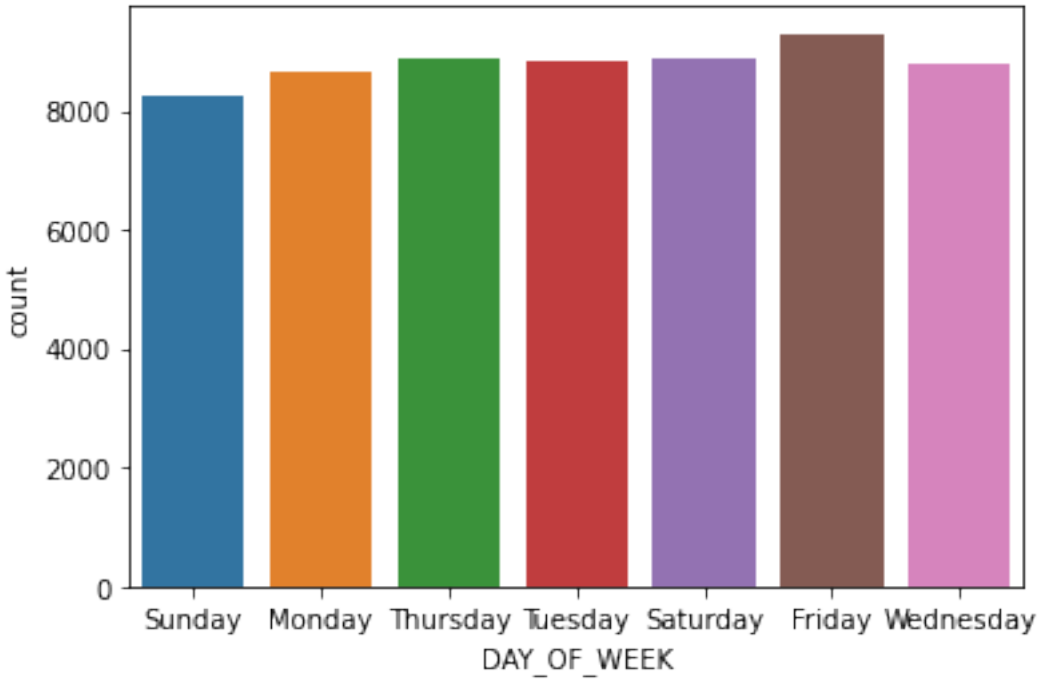
2 Leitfragen 2/3

3. Existieren (a) Uhrzeiten, (b) Tage oder (c) Monate an denen mehr schwere Verbrechen ('Part One') stattfinden? • Finden Straftaten eher nachts oder tagsüber statt? • Wann werden somit die meisten Polizisten benötigt

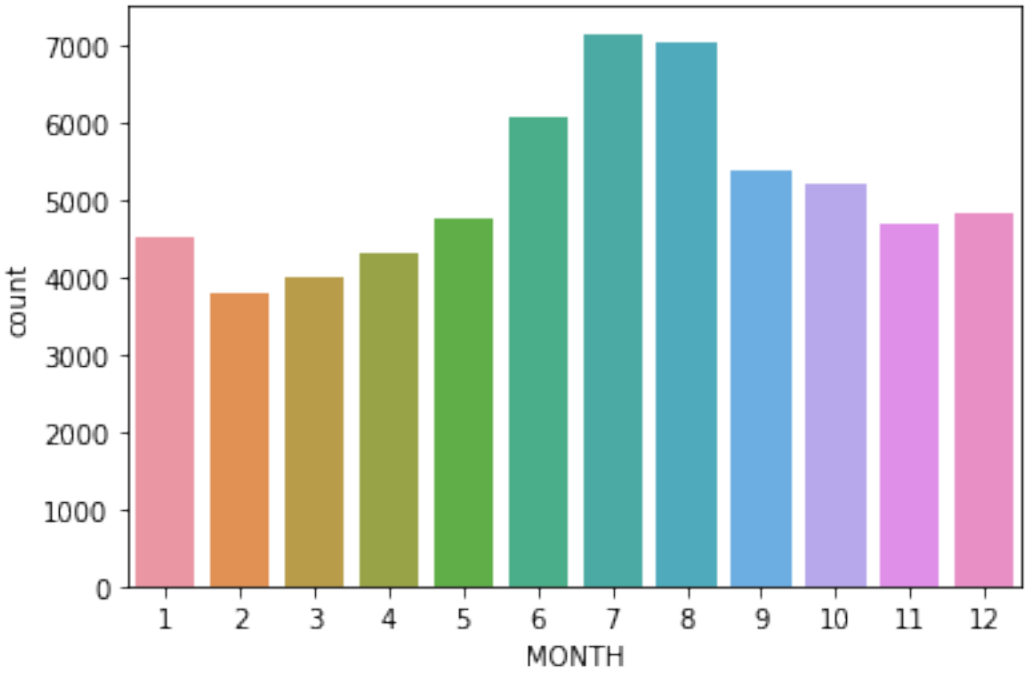
[14]:
```
filtered = df[df['UCR_PART'] == "Part One"]

plot = sns.countplot(x="HOUR", data = filtered)
```



[15]:
```
plot = sns.countplot(x="DAY_OF_WEEK", data = filtered)
```



[16]:
```
plot = sns.countplot(x="MONTH", data = filtered)
```

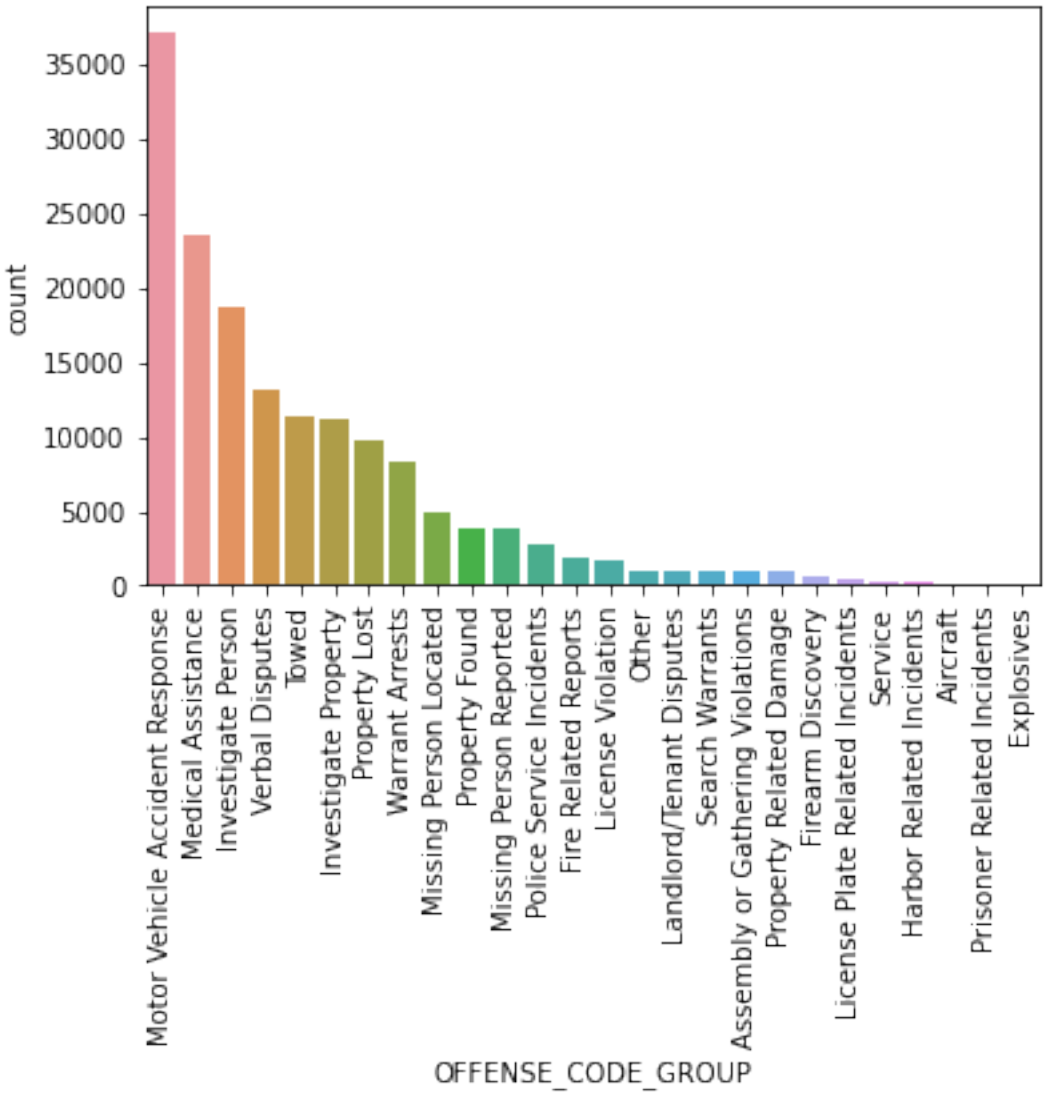


4. Welche leichten Straftaten ('Part Tree') benötigen (vermutlich) Verkehrspolizisten? Welches sind die 5 Straßen, in denen die meisten Verkehrspolizisten benötigt werden?

[11]:
```
filtered = df[df['UCR_PART'] == "Part Three"]

table = filtered.groupby(["OFFENSE_CODE_GROUP"]).INCIDENT_NUMBER.count().
 ↪to_frame()
table = table.sort_values('INCIDENT_NUMBER', ascending=False)

plot = sns.countplot(x="OFFENSE_CODE_GROUP",
              data = filtered,
              order = filtered['OFFENSE_CODE_GROUP'].value_counts().index)
plt.setp(plot.get_xticklabels(), rotation=90)


filtered.OFFENSE_CODE_GROUP.unique()
filtered = filtered[filtered['OFFENSE_CODE_GROUP'].isin(["Motor Vehicle␣
 ↪Accident Response", 'License Plate Related Incidents'])]
table = filtered.groupby(["STREET"]).INCIDENT_NUMBER.count().to_frame()
table = table.sort_values('INCIDENT_NUMBER', ascending=False)
table[0:5]
```

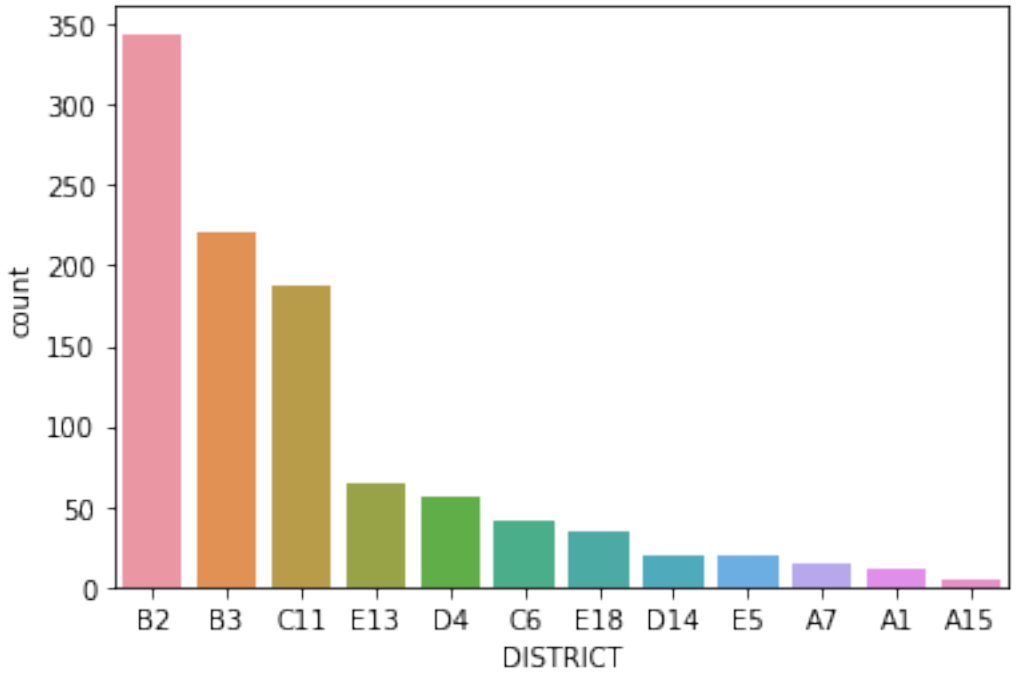
[11]:

STREET	INCIDENT_NUMBER
BLUE HILL AVE	1254
WASHINGTON ST	1142
DORCHESTER AVE	819
CENTRE ST	620
COMMONWEALTH AVE	617

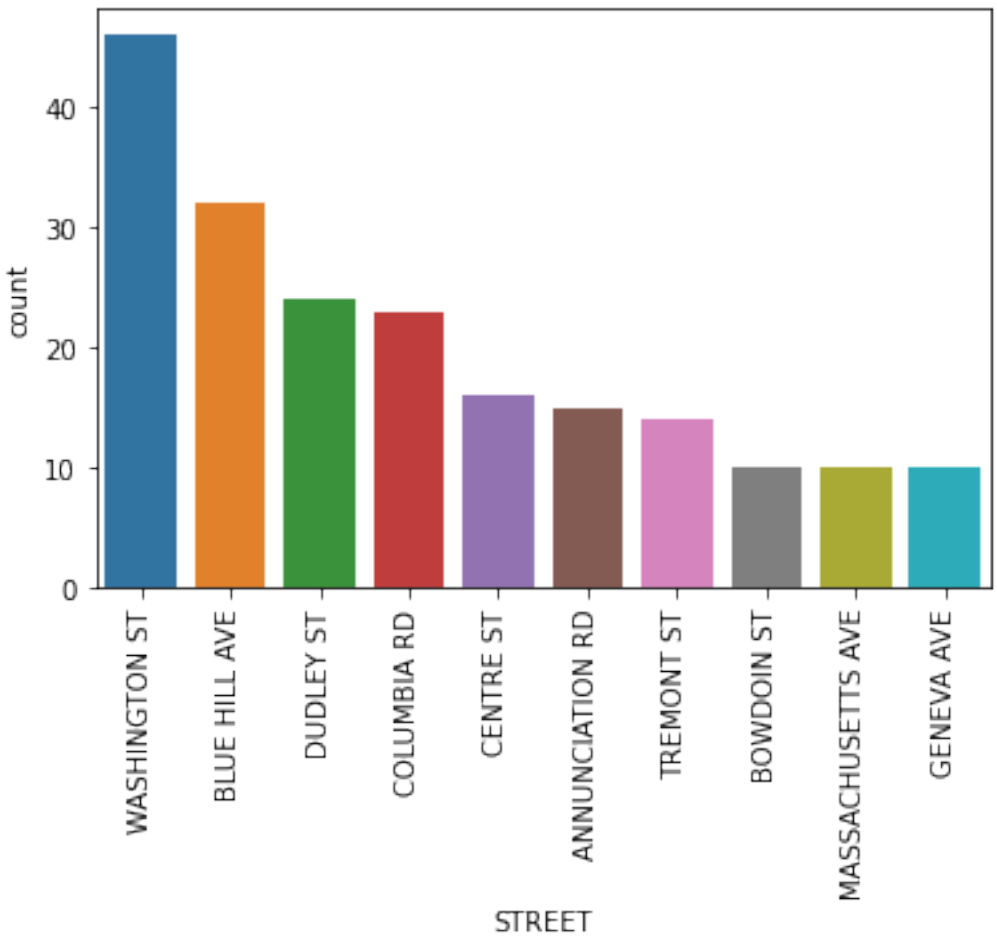


5. Wie hat sich die Anzahl der Schießereien in den letzten Jahren entwickelt? • In welchem Bezirk finden die meisten Schießereien statt? • In welcher Straße finden die meisten Schießereien statt? • Zu welchen Uhrzeiten finden die meisten Schießereien statt?

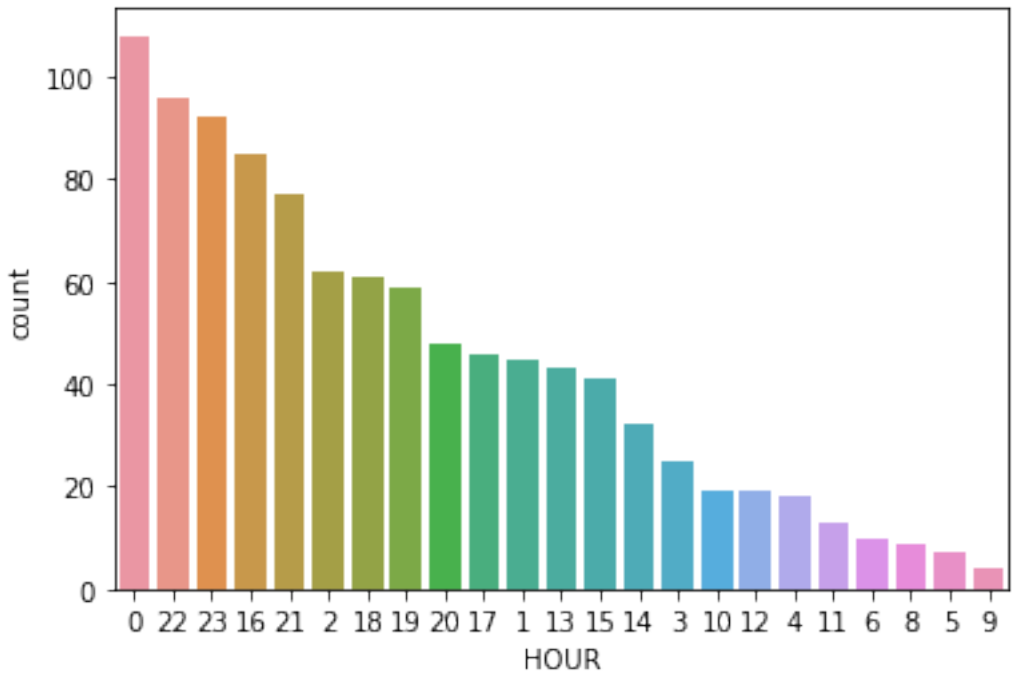
[17]:
```
df.SHOOTING.unique()
filtered = df[df['SHOOTING'] == "Y"]


plot = sns.countplot(x="DISTRICT",
              data = filtered,
              order = filtered['DISTRICT'].value_counts().index)
```



[24]:
```
plot = sns.countplot(x="STREET",
              data = filtered,
              order = filtered['STREET'].value_counts().iloc[:10].index)
_ = plt.setp(plot.get_xticklabels(), rotation=90)
```



[19]:
```
plot = sns.countplot(x="HOUR",
                     data = filtered,
                     order = filtered['HOUR'].value_counts().index)
```


100
80
60
40
20
0
count
0 22 23 16 21 2 18 19 20 17 1 13 15 14 3 10 12 4 11 6 8 5 9
HOUR

[ ]: