IMPORTING THE DATA

```
import pandas as pd
infile = "/content/311_Service_Requests_20240217.csv"
df = pd.read_csv(infile)
```

```
ns (0) have mixed types. Specify dtype option on import or set low_memory=False.
```

```
df.columns
```

```
Index(['Case Reference', 'Open Date', 'Closed Date', 'Status', 'Subject',
       'Reason', 'Type', 'Object Type', 'Address Number', 'Address Line 1',
       'Address Line 2', 'City', 'State', 'Zipcode', 'Property ID', 'Location',
       'Latitude', 'Longitude', 'Council District', 'Council District 2011',
       'Police District', 'Census Tract', 'Census Block Group', 'Census Block',
       'Neighborhood', 'X Coordinate', 'Y Coordinate', '2010 Census Tract',
       '2010 Census Block Group', '2010 Census Block', 'TRACTCE20',
       'GEOID20_tract', 'GEOID20_blockgroup', 'GEOID20_block'],
      dtype='object')
```

```
df.shape
```

```
(118105, 34)
```

```
df.head()
```

	Case Reference	Open Date	Closed Date	Status	Subject	Reason	Type	Object Type
0	1001988123	02/13/2024 10:38:00 AM	02/14/2024 02:49:00 PM	Closed	Dept of Public Works	Engineering - Street Repairs	Pot Hole (Req_Serv)	Property
1	1001988270	02/13/2024 03:19:00 PM	02/14/2024 01:05:00 PM	Closed	Dept of Public Works	Rodent_Pest Control	Rodents (Req_Serv)	Property
2	1001988178	02/13/2024 12:15:00 PM	02/14/2024 01:04:00 PM	Closed	Dept of Public Works	Sanitation	Totes Replace (Req_Serv)	Property
3	1001988102	02/13/2024 10:06:00 AM	02/13/2024 12:54:00 PM	Closed	Dept of Parking	Parking Violations Bureau	Parking Issues (Req_Serv)	Property
4	1001988088	02/13/2024 09:54:00 AM	02/13/2024 02:17:00 PM	Closed	Dept of Public Works	Sanitation	Trash Ordinance Violation (Req_Serv)	Property

5 rows × 34 columns

LOOKING FOR WHICH VARIABLES HAVE LIMITED OUTCOMES

```
variables = ['Status', 'Subject', 'Reason', 'Type', 'Object Type']


for variable in variables:
    print(f"Unique values for '{variable}':")
    print(df[variable].value_counts())
    print()
```

```
Unique values for 'Status':
Closed    116389
Open        1716
Name: Status, dtype: int64

Unique values for 'Subject':
Dept of Public Works                   78048
DPIS                                   18629
Dept of Parking                         7141
Utilities                               6549
Buffalo Police Department               5262
Office of the Mayor                      741
Buffalo Municipal Housing Authority      467
Dept of Law                              425
Office of Strategic Planning             401
```

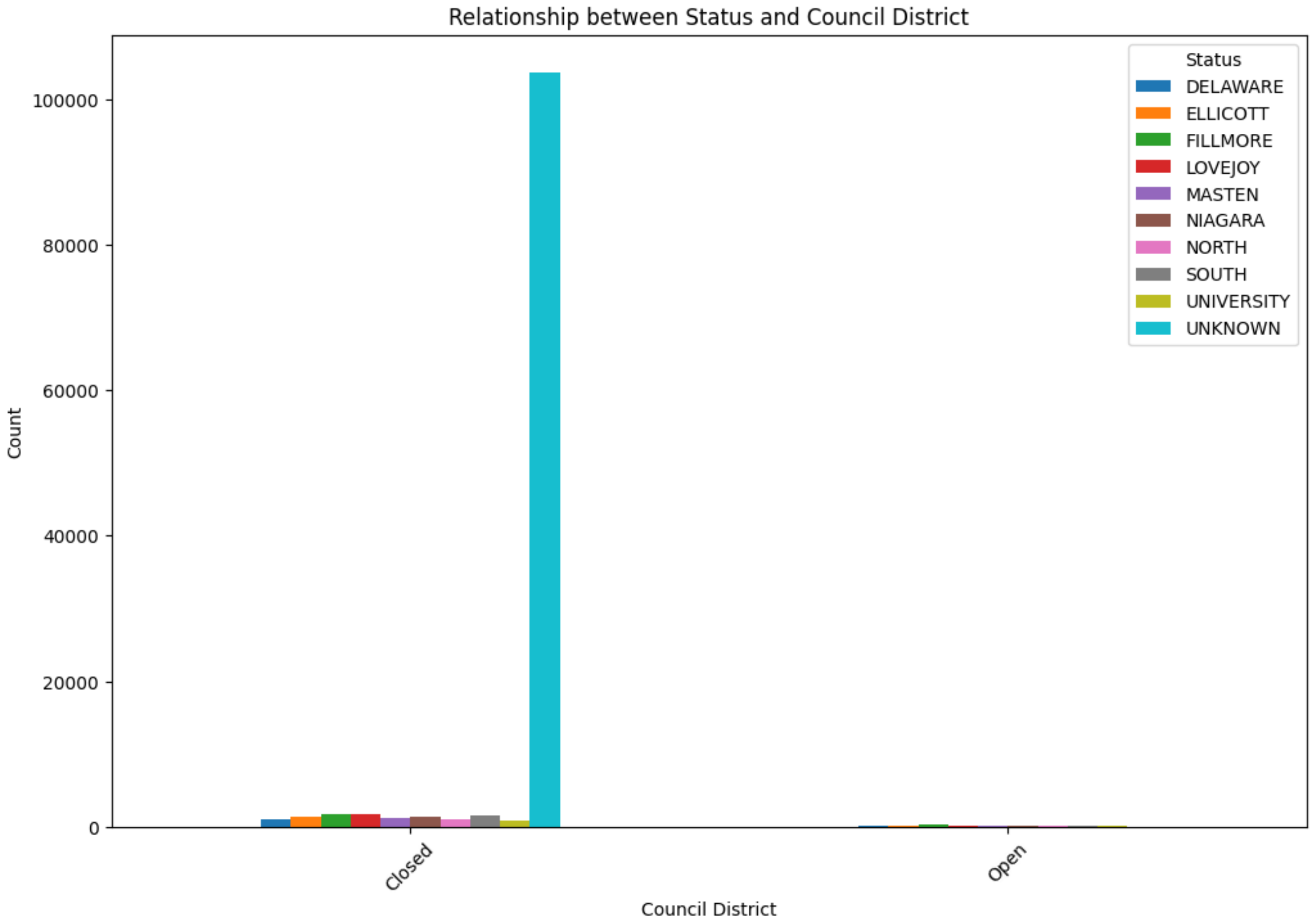
```
Assessment & Taxation                    200
Buffalo Fire Department                  155
City Clerk                                70
Community Services & Rec. Program.        10
Human Resources                            4
New Americans                              2
Management Information Systems             1
Name: Subject, dtype: int64

Unique values for 'Reason':
Sanitation                                 37998
Housing                                    18629
Engineering - Street Repairs               12543
Streets                                     9662
Parking Violations Bureau                   6986
Rodent Control                              5586
Police                                      5262
Forestry                                    4999
Engineering - Traffic                       3655
National Grid                               3492
Animal Shelter                              2323
Buffalo Sewer Authority                     1872
Buffalo Water Authority                     1185
City Parks                                   530
BMHA                                         467
Citizen Services - Quick Response Teams      311
Streets/Sanitation                           291
Rodent_Pest Control                          286
Real Estate                                  260
Freedom of Information                       235
Adjudication - Ordinance Violation           190
Taxation                                     168
Citizen Services - Graffiti                  162
Moving Violations                            155
BFD                                          155
Citizen Services - Save Our Streets          148
Personnel                                    122
Citizens Services - Clean City               119
Administration                               114
City Clerk Issue                              43
Buildings Division                            34
Assessment & Taxation                         30
Licenses                                      27
```

Based on this, it looks like Object Type and Status can be defined by a limited amount of categories. The rest of the variables that were looked at have larger amounts of categories, with Type having the most as it does not even list them all as it has 180.

PRODUCING GRAPHS

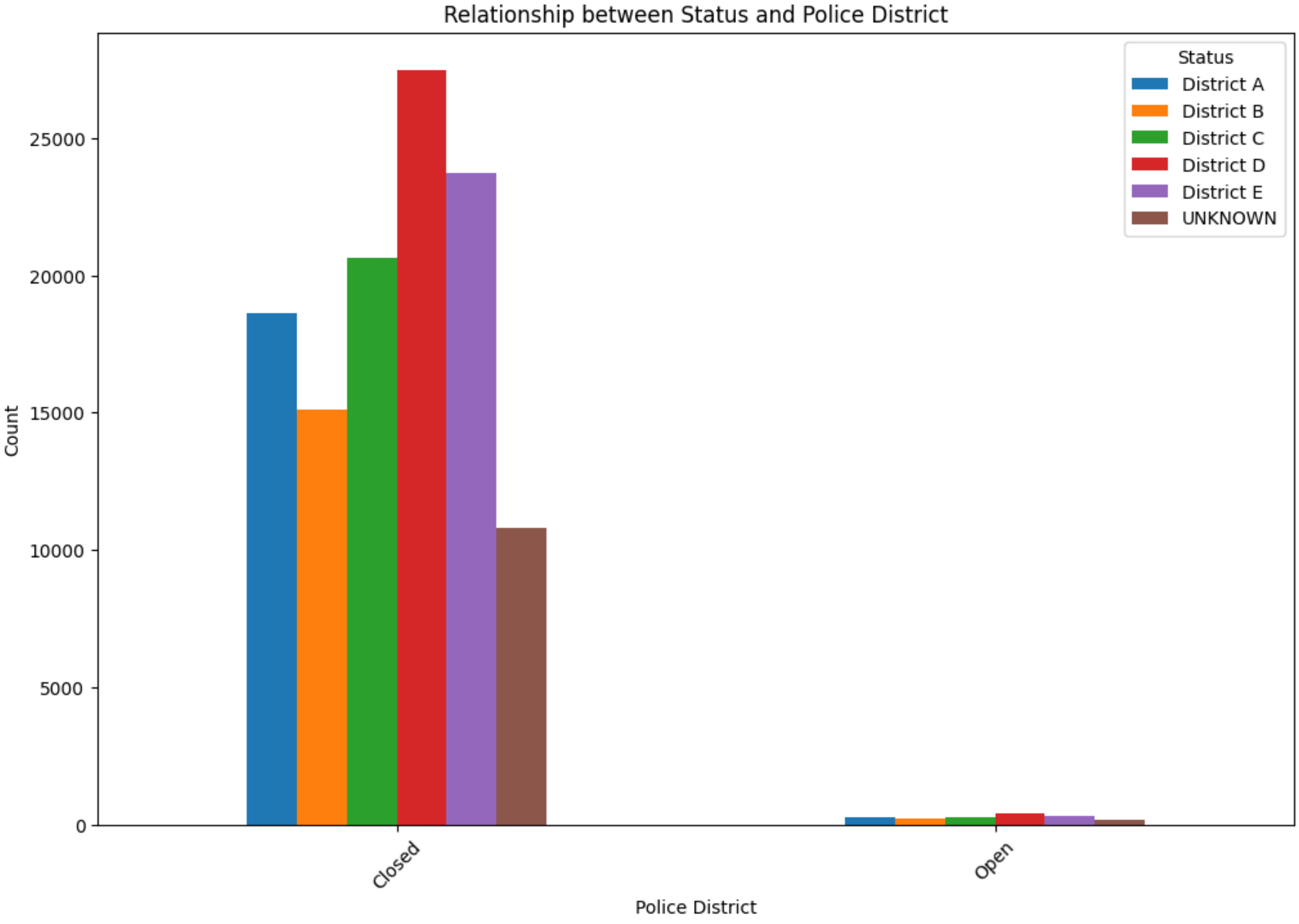
```
import matplotlib.pyplot as plt
```

```
grouped_data = df.groupby(['Status', 'Council District']).size().unstack(fill_value=
grouped_data.plot(kind='bar', figsize=(12, 8))
plt.title('Relationship between Status and Council District')
plt.xlabel('Council District')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Status')
plt.show()
```



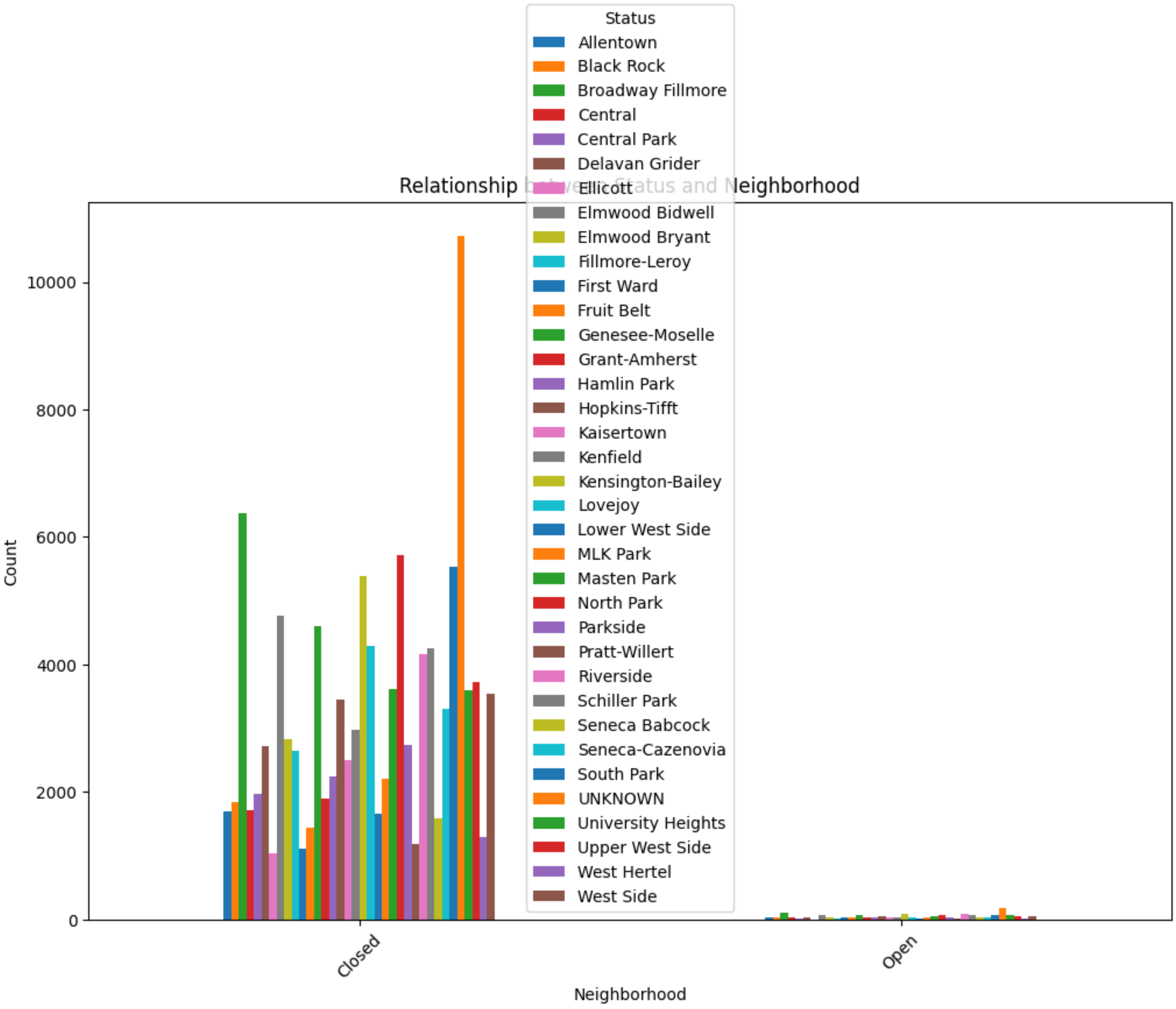
With the severe difference in amount of closed versus open observations it is not a surprise to see much more information on the closed side. The overwhelming majority of closed observations fall under the UNKNOWN council district.

```
grouped_data1 = df.groupby(['Status', 'Police District']).size().unstack(fill_value=
grouped_data1.plot(kind='bar', figsize=(12, 8))
plt.title('Relationship between Status and Police District')
plt.xlabel('Police District')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Status')
plt.show()
```



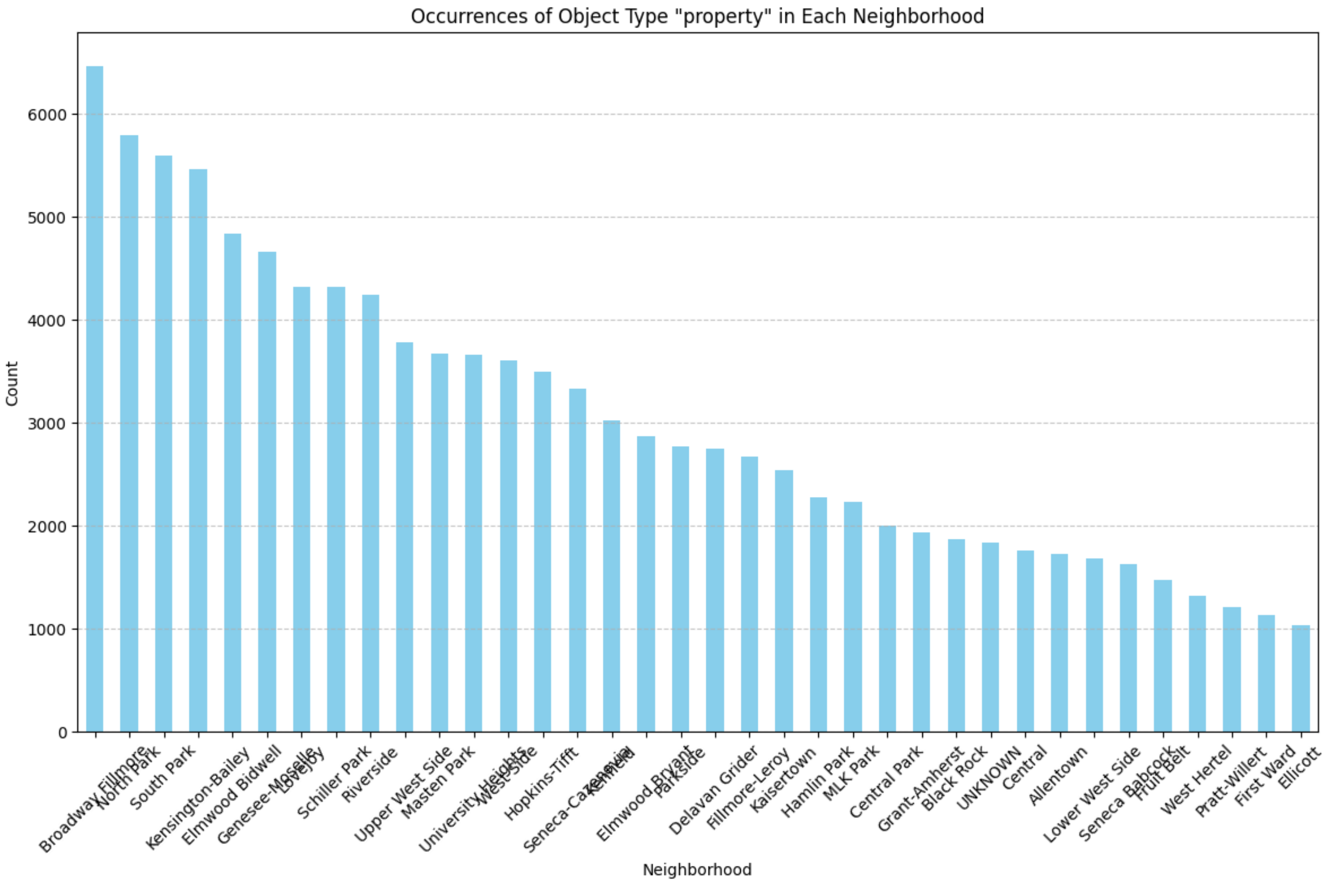
This shows much different results than the council district as there are far less unknowns. District D has the highest number of closed and open status'. District E has the second most for both closed and open.

```
grouped_data2 = df.groupby(['Status', 'Neighborhood']).size().unstack(fill_value=0)
grouped_data2.plot(kind='bar', figsize=(12, 8))
plt.title('Relationship between Status and Neighborhood')
plt.xlabel('Neighborhood')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Status')
plt.show()
```



While this is slightly difficult to read with the amount of neighborhoods that there are, we can still clearly see that UNKNOWN has the highest smount of closed and open status'. It seems to be a clear relationship between the number of closed and open observations as the neighborhoods with high number of closed observations also have high number of open observations.

```
property_data = df[df['Object Type'] == 'Property']
property_counts = property_data['Neighborhood'].value_counts()
plt.figure(figsize=(12, 8))
property_counts.plot(kind='bar', color='skyblue')
plt.title('Occurrences of Object Type "property" in Each Neighborhood')
plt.xlabel('Neighborhood')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



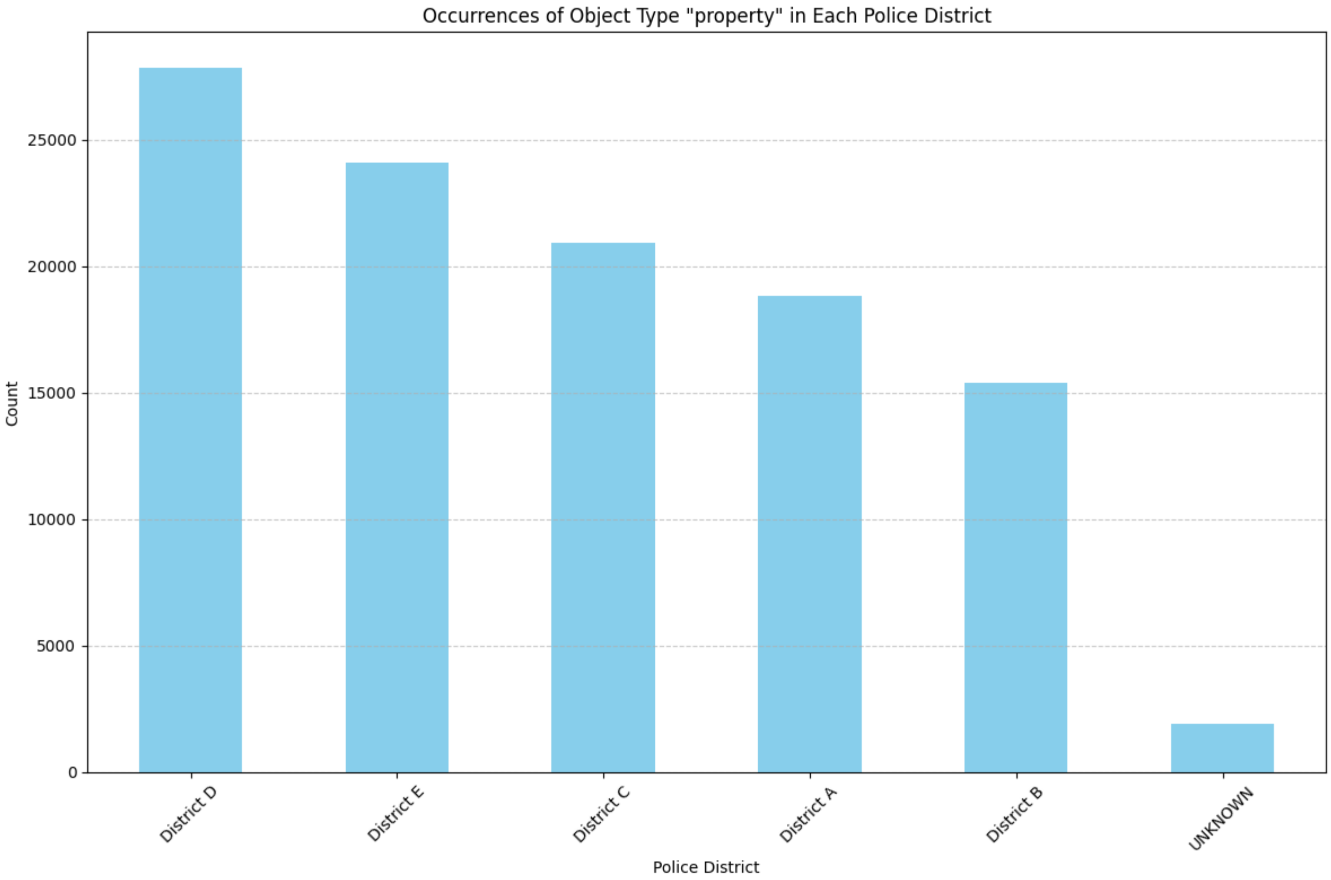
```
property_counts
```

```
    Broadway Fillmore     6469
    North Park            5794
    South Park            5600
    Kensington-Bailey     5469
    Elmwood Bidwell       4842
    Genesee-Moselle       4668
    Lovejoy               4325
    Schiller Park         4323
    Riverside             4249
    Upper West Side       3784
    Masten Park           3669
    University Heights    3667
    West Side             3606
    Hopkins-Tifft         3492
    Seneca-Cazenovia      3336
    Kenfield              3019
    Elmwood Bryant        2869
    Parkside              2776
    Delavan Grider        2747
    Fillmore-Leroy        2675
    Kaisertown            2537
    Hamlin Park           2279
    MLK Park              2228
    Central Park          1999
    Grant-Amherst         1937
    Black Rock            1872
```

```
UNKNOWN               1835
Central               1756
Allentown             1725
Lower West Side       1685
Seneca Babcock        1627
Fruit Belt            1475
West Hertel           1318
Pratt-Willert         1213
First Ward            1134
Ellicott              1037
Name: Neighborhood, dtype: int64
```

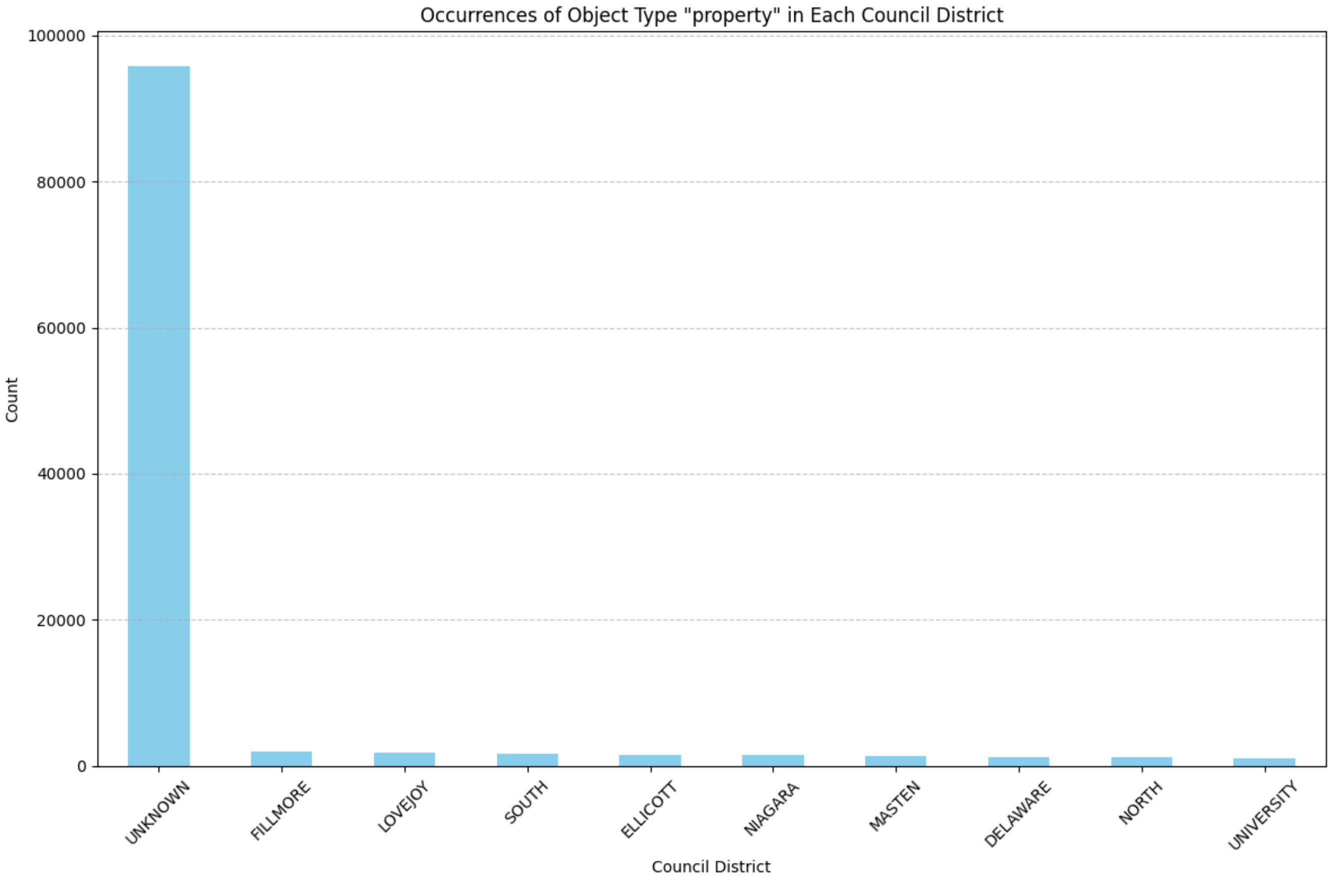
Here we can see that Broadway Fillmore has the most observations where the object type if property. Ellicott has the least.

```
property_counts = property_data['Police District'].value_counts()
plt.figure(figsize=(12, 8))
property_counts.plot(kind='bar', color='skyblue')
plt.title('Occurrences of Object Type "property" in Each Police District')
plt.xlabel('Police District')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



Here we see that District D has the highest number of properties within it. UNKNOWN has the least amount.

```
property_counts = property_data['Council District'].value_counts()
plt.figure(figsize=(12, 8))
property_counts.plot(kind='bar', color='skyblue')
plt.title('Occurrences of Object Type "property" in Each Council District')
plt.xlabel('Council District')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

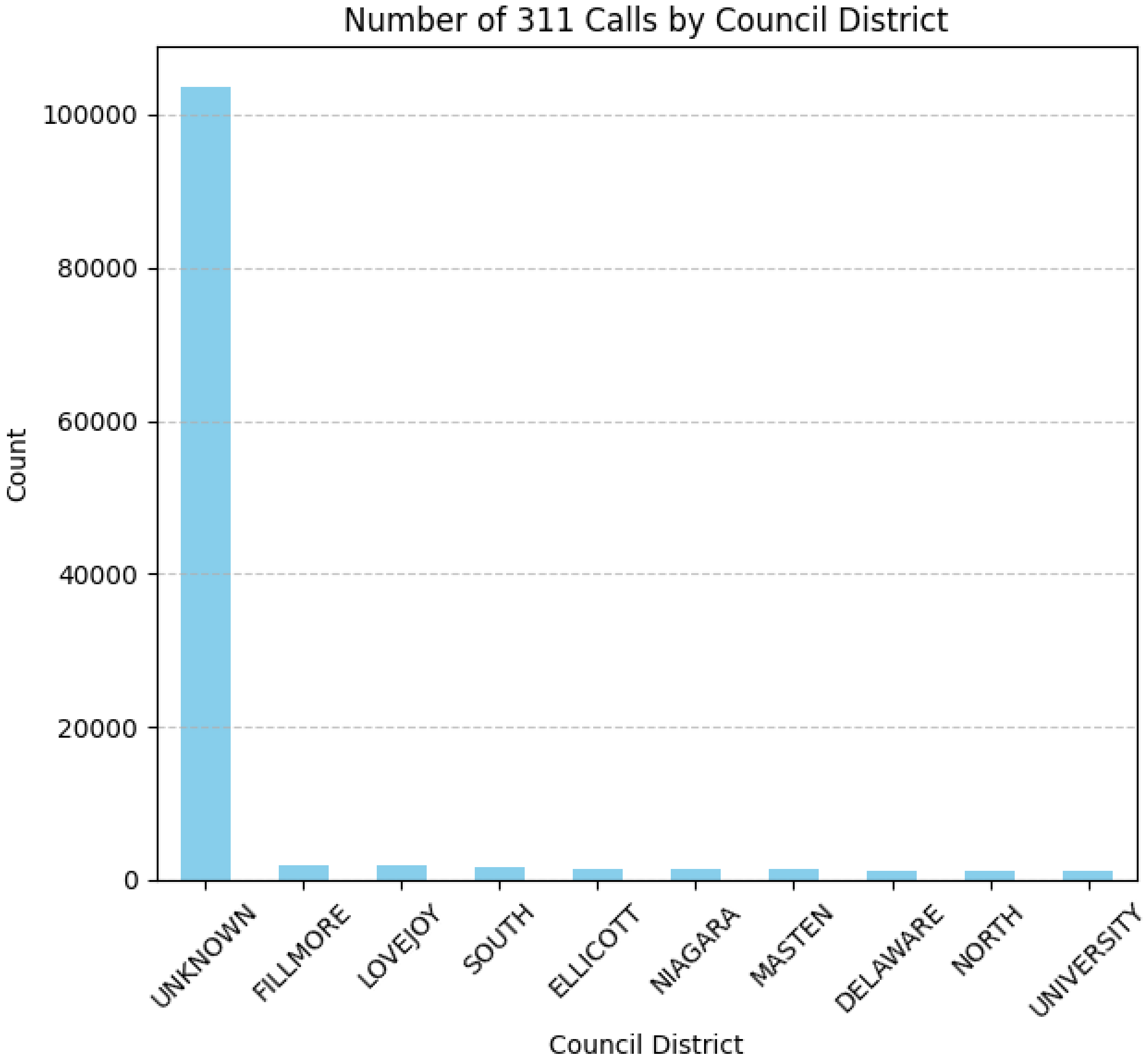


Similarly to how status performed with council district, UNKNOWN has the overwhelming amount of observations while the other council districts have rather similar numbers

NUMBER OF CALLS GRAPHS

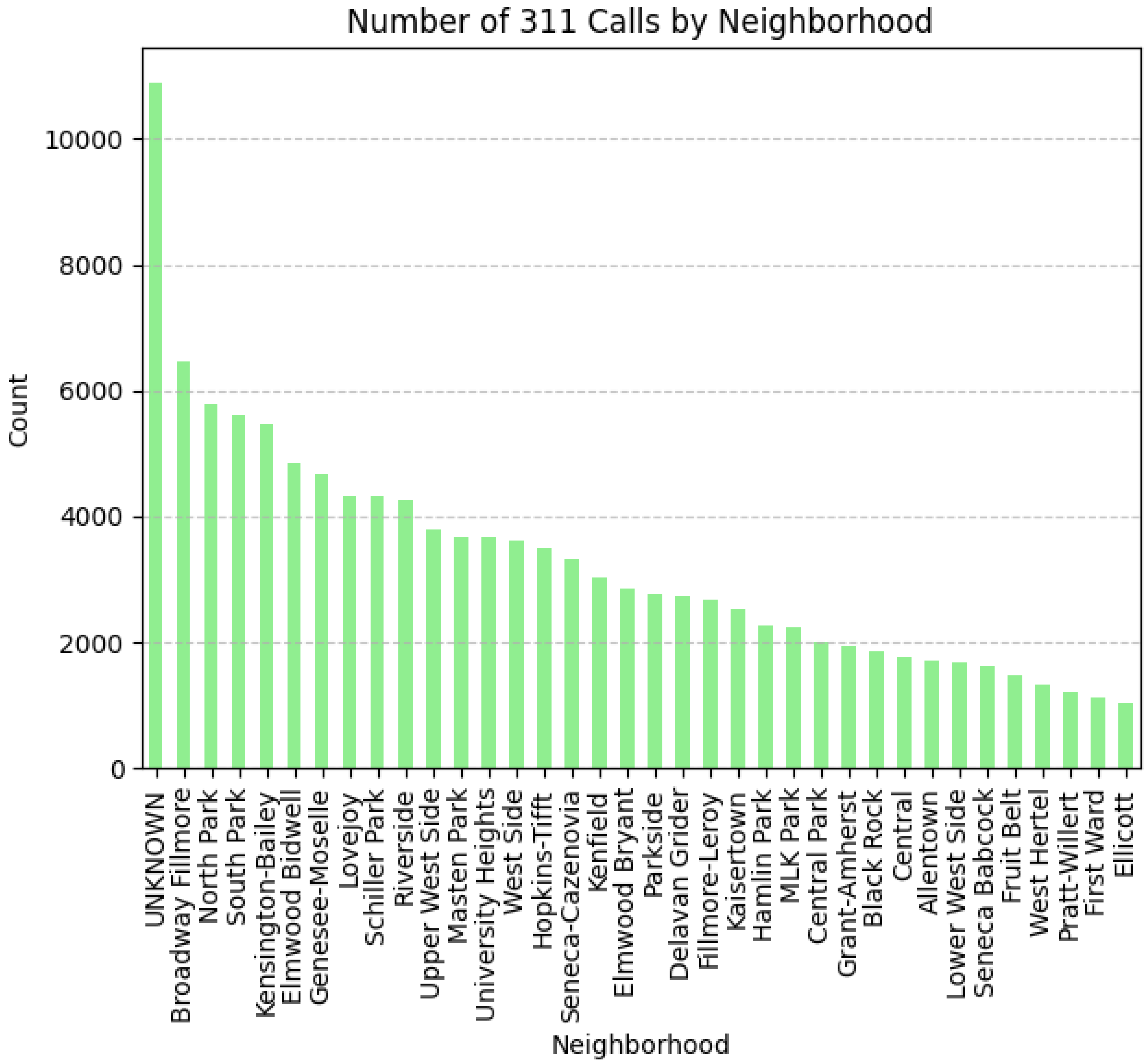
```
district_counts = df['Council District'].value_counts()
neighborhood_counts = df['Neighborhood'].value_counts()
```

```
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
district_counts.plot(kind='bar', color='skyblue')
plt.title('Number of 311 Calls by Council District')
plt.xlabel('Council District')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.tight_layout()
plt.show()
```



It is again evident that UNKNOWN has the highest number of 311 calls when looking at council district. Fillmore has the second highest.

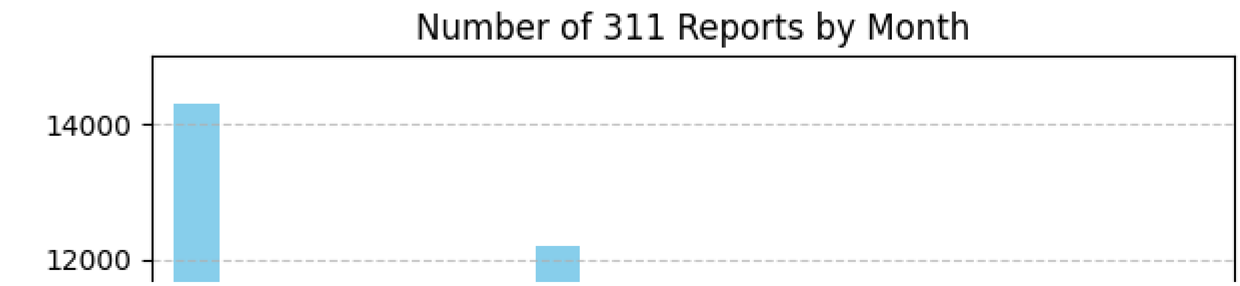
```
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 2)
neighborhood_counts.plot(kind='bar', color='lightgreen')
plt.title('Number of 311 Calls by Neighborhood')
plt.xlabel('Neighborhood')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.tight_layout()
plt.show()
```



Similarly to council district, UNKNOWN has the highest number of 311 calls when looking at it by neighborhood. 2nd is Broadway-Fillmore.

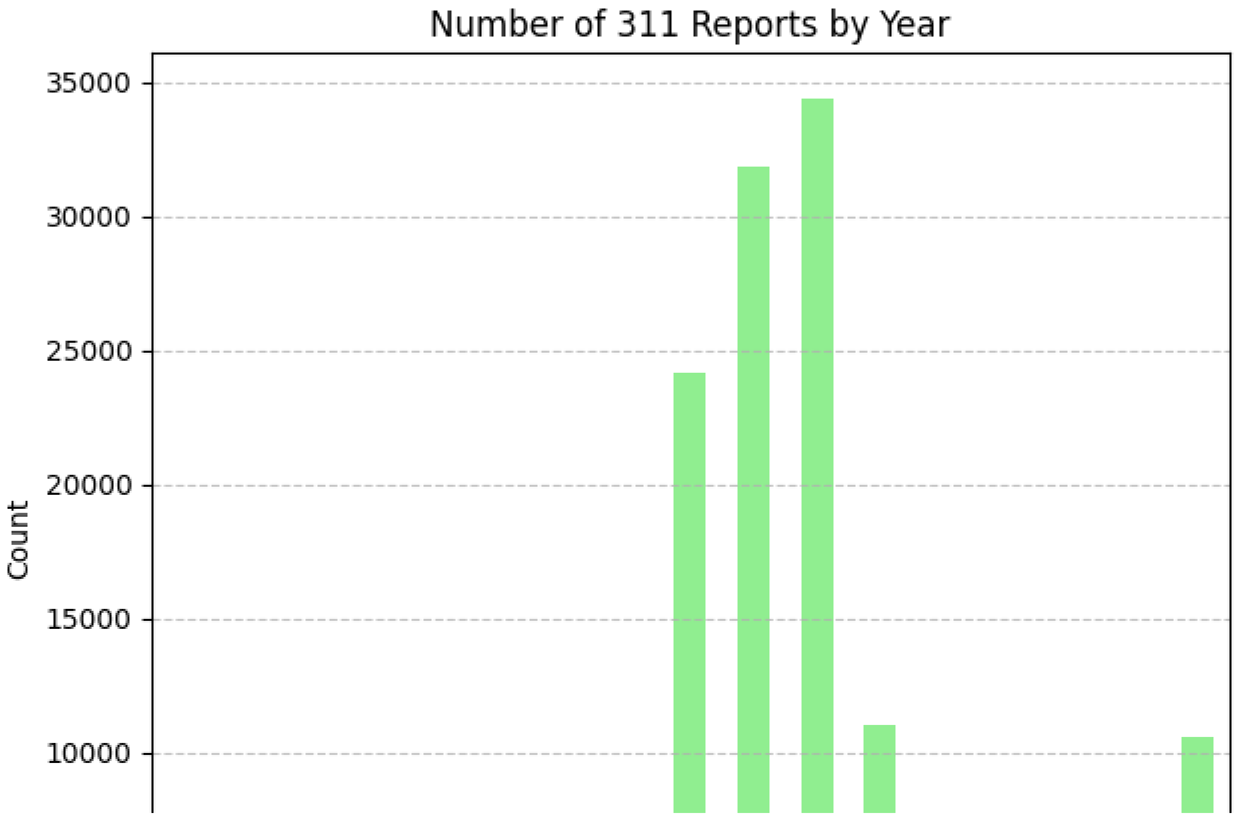
NUMBER OF CALLS BY MONTH AND YEAR

```
df['date'] = pd.to_datetime(df['Open Date'])


df['month'] = df['date'].dt.month
df['year'] = df['date'].dt.year


monthly_counts = df['month'].value_counts().sort_index()
yearly_counts = df['year'].value_counts().sort_index()


plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
monthly_counts.plot(kind='bar', color='skyblue')
plt.title('Number of 311 Reports by Month')
plt.xlabel('Month')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.grid(axis='y', linestyle='--', alpha=0.7)


plt.tight_layout()
plt.show()
```



Here we can see January had the highest number of reported 311 cases. This is followed by the months surrounding and including summer with May, June, July and August all being grouped together much higher than any other month. November has the lowest amount of reported 311 cases.

```
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 2)
yearly_counts.plot(kind='bar', color='lightgreen')
plt.title('Number of 311 Reports by Year')
plt.xlabel('Year')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.tight_layout()
plt.show()
```


Number of 311 Reports by Year
Count