HomeWork 6

December 22, 2021

1 Film Perimit in NYC

1.1 Yuzhen Wang

1.2 Question Statement

During the pandemic, film industry in NYC suffered a lot. Here I want to look deeper into this issue and try to answer the following questions:

Which borough or community is the most popular site to shoot films?

How did film permit change across the city during the pandemic?

What kind of film suffered the most in pandemic?

1.3 Hypothesis

Midtown Manhattan might have the highest number of film permits issued.

After the outbreak of the pandemic, I expect to see a decrease in the volume of film permits issued in NYC. The volume of film permits might have a linear relationship with the daily covid cases.

More specifically, I think among all film permit categories, movies should suffer the most because the relatively high cost, other categories like commercials, might not decrease as much.

```
[1]: ## import data
     import pandas as pd
     import plotly.express as px
     import plotly.io as pio

     pio.renderers.default = "notebook_connected+pdf"

     fp = pd.read_csv('https://data.cityofnewyork.us/api/views/tg4x-b46p/rows.csv?
      ↪accessType=DOWNLOAD')
     fp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72410 entries, 0 to 72409
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   EventID          72410 non-null  int64
```

```
1    EventType            72410 non-null   object
2    StartDateTime        72410 non-null   object
3    EndDateTime          72410 non-null   object
4    EnteredOn            72410 non-null   object
5    EventAgency          72410 non-null   object
6    ParkingHeld          72410 non-null   object
7    Borough              72410 non-null   object
8    CommunityBoard(s)    72392 non-null   object
9    PolicePrecinct(s)    72392 non-null   object
10   Category             72410 non-null   object
11   SubCategoryName      72410 non-null   object
12   Country              72410 non-null   object
13   ZipCode(s)           72392 non-null   object
dtypes: int64(1), object(13)
memory usage: 7.7+ MB
```

2   Data Cleaning

There are missing values in CommunityBoard(s), I decide to delete these observations.

I also decide to drop "PolicePrecinct(s)" & "ZipCode(s)", so that I can use CommunityBoard(s) as the only location variable

```
[2]: fp=fp.sort_values('CommunityBoard(s)').head(72392).reset_index().drop(columns =␣
     ↪['PolicePrecinct(s)','ZipCode(s)'])
```

```
[3]: fp['CommunityBoard(s)'].head(5)
```

```
[3]: 0    0, 1
     1    0, 1
     2    0, 1
     3    0, 1
     4    0, 1
     Name: CommunityBoard(s), dtype: object
```

Note that in CommunityBoards, some cells have mutiple values, so I split the mutiple values into different rows.

```
[4]: for i in range(len(fp)):
         if ',' in fp.loc[i,'CommunityBoard(s)']:
             cb = fp.loc[i,'CommunityBoard(s)'].split(',')
             for k in range(len(cb)):
                 new = fp.loc[i]
                 new['CommunityBoard(s)']=cb[k]
                 fp = fp.append(new,ignore_index=True)
```

```
/opt/conda/envs/python-public-policy/lib/python3.8/site-
packages/pandas/core/series.py:1056: SettingWithCopyWarning:
```

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
[5]:  #delete the original rows with multiple "CommunityBoard(s)"
      def findcomma(row):
          if ',' in row['CommunityBoard(s)']:
              return 1
          else:
              return 0
      fp['comma']=fp.apply(findcomma,axis=1)
      fp = fp[fp['comma']==0].reset_index()
```

```
[6]:  #drop irrelevant variables
      fp=fp.drop(columns=['level_0','index','comma'])
      fp.head(5)
```

[6]:
	EventID	EventType	StartDateTime	EndDateTime \
0	439255	Shooting Permit	08/31/2018 08:00:00 AM	08/31/2018 10:00:00 PM
1	395362	Shooting Permit	02/01/2018 09:00:00 AM	02/02/2018 01:00:00 AM
2	438746	Rigging Permit	08/30/2018 06:00:00 AM	08/30/2018 08:00:00 PM
3	439242	Shooting Permit	08/31/2018 07:00:00 AM	08/31/2018 05:00:00 PM
4	470969	Shooting Permit	03/26/2019 07:00:00 AM	03/26/2019 11:30:00 PM

	EnteredOn	EventAgency \
0	08/29/2018 11:17:47 AM	Mayor's Office of Film, Theatre & Broadcasting
1	01/29/2018 08:08:14 PM	Mayor's Office of Film, Theatre & Broadcasting
2	08/24/2018 07:55:35 PM	Mayor's Office of Film, Theatre & Broadcasting
3	08/29/2018 10:28:23 AM	Mayor's Office of Film, Theatre & Broadcasting
4	03/20/2019 11:06:11 AM	Mayor's Office of Film, Theatre & Broadcasting

	ParkingHeld	Borough \
0	HEYWARD STREET between HARRISON AVENUE and MAR…	Brooklyn
1	HUBERT STREET between COLLISTER STREET and HUD…	Manhattan
2	CALYER STREET between JEWEL STREET and HUMBOLT…	Brooklyn
3	NORMAN AVENUE between SUTTON STREET and KINGSL…	Brooklyn
4	Dupont Street between PROVOST STREET and McGui…	Brooklyn

	CommunityBoard(s)	Category	SubCategoryName	Country
0	1	Film	Feature	United States of America
1	1	Television	Cable-episodic	United States of America
2	1	Television	Episodic series	United States of America
3	1	Television	Cable-episodic	United States of America
4	1	Television	Cable-episodic	United States of America

Now each row only has one community board number.

Next I will creat a borough community board variable for the convenience of mapping, and format the time variables to time series so python can recognize them as time.

```
[7]: # Creat Borough Community Board
     def recode_borocd_counts(row):
       if 'Manhattan' in row["Borough"]:
         return 100 + int(row["CommunityBoard(s)"])
       elif 'Bronx' in row["Borough"]:
         return 200 + int(row["CommunityBoard(s)"])
       elif 'Brooklyn' in row["Borough"]:
         return 300 + int(row["CommunityBoard(s)"])
       elif 'Queens' in row["Borough"]:
         return 400 + int(row["CommunityBoard(s)"])
       elif 'Staten Island' in row["Borough"]:
         return 500 + int(row["CommunityBoard(s)"])

     fp['borocd'] = fp.apply(recode_borocd_counts, axis = 1).astype(str)
     fp.head(5)
```

```
[7]:     EventID         EventType         StartDateTime              EndDateTime  \
     0    439255   Shooting Permit   08/31/2018 08:00:00 AM   08/31/2018 10:00:00 PM
     1    395362   Shooting Permit   02/01/2018 09:00:00 AM   02/02/2018 01:00:00 AM
     2    438746    Rigging Permit   08/30/2018 06:00:00 AM   08/30/2018 08:00:00 PM
     3    439242   Shooting Permit   08/31/2018 07:00:00 AM   08/31/2018 05:00:00 PM
     4    470969   Shooting Permit   03/26/2019 07:00:00 AM   03/26/2019 11:30:00 PM


                     EnteredOn                                    EventAgency  \
     0   08/29/2018 11:17:47 AM   Mayor's Office of Film, Theatre & Broadcasting
     1   01/29/2018 08:08:14 PM   Mayor's Office of Film, Theatre & Broadcasting
     2   08/24/2018 07:55:35 PM   Mayor's Office of Film, Theatre & Broadcasting
     3   08/29/2018 10:28:23 AM   Mayor's Office of Film, Theatre & Broadcasting
     4   03/20/2019 11:06:11 AM   Mayor's Office of Film, Theatre & Broadcasting


                                       ParkingHeld     Borough  \
     0   HEYWARD STREET between HARRISON AVENUE and MAR…    Brooklyn
     1   HUBERT STREET between COLLISTER STREET and HUD…   Manhattan
     2   CALYER STREET between JEWEL STREET and HUMBOLT…    Brooklyn
     3   NORMAN AVENUE between SUTTON STREET and KINGSL…    Brooklyn
     4   Dupont Street between PROVOST STREET and McGui…    Brooklyn


        CommunityBoard(s)     Category   SubCategoryName                   Country  \
     0                  1         Film            Feature   United States of America
     1                  1   Television     Cable-episodic   United States of America
     2                  1   Television    Episodic series   United States of America
     3                  1   Television     Cable-episodic   United States of America
     4                  1   Television     Cable-episodic   United States of America
```

```
    borocd
0      301
1      101
2      301
3      301
4      301
```

[8]:
```python
# creat time series
fp['StartDateTime'] = pd.to_datetime(fp['StartDateTime'], format='%m/%d/%Y %I:
 ↪%M:%S %p')
fp['EndDateTime'] = pd.to_datetime(fp['EndDateTime'], format='%m/%d/%Y %I:%M:%S␣
 ↪%p')
fp.head(5)
```

[8]:
```
    EventID       EventType        StartDateTime         EndDateTime  \
0    439255  Shooting Permit 2018-08-31 08:00:00 2018-08-31 22:00:00
1    395362  Shooting Permit 2018-02-01 09:00:00 2018-02-02 01:00:00
2    438746   Rigging Permit 2018-08-30 06:00:00 2018-08-30 20:00:00
3    439242  Shooting Permit 2018-08-31 07:00:00 2018-08-31 17:00:00
4    470969  Shooting Permit 2019-03-26 07:00:00 2019-03-26 23:30:00

               EnteredOn                                  EventAgency  \
0  08/29/2018 11:17:47 AM  Mayor's Office of Film, Theatre & Broadcasting
1  01/29/2018 08:08:14 PM  Mayor's Office of Film, Theatre & Broadcasting
2  08/24/2018 07:55:35 PM  Mayor's Office of Film, Theatre & Broadcasting
3  08/29/2018 10:28:23 AM  Mayor's Office of Film, Theatre & Broadcasting
4  03/20/2019 11:06:11 AM  Mayor's Office of Film, Theatre & Broadcasting

                                        ParkingHeld    Borough  \
0  HEYWARD STREET between HARRISON AVENUE and MAR…   Brooklyn
1  HUBERT STREET between COLLISTER STREET and HUD…  Manhattan
2  CALYER STREET between JEWEL STREET and HUMBOLT…   Brooklyn
3  NORMAN AVENUE between SUTTON STREET and KINGSL…   Brooklyn
4  Dupont Street between PROVOST STREET and McGui…   Brooklyn

   CommunityBoard(s)    Category  SubCategoryName                   Country  \
0                  1        Film           Feature  United States of America
1                  1  Television    Cable-episodic  United States of America
2                  1  Television   Episodic series  United States of America
3                  1  Television    Cable-episodic  United States of America
4                  1  Television    Cable-episodic  United States of America

    borocd
0      301
1      101
2      301
```

```
3     301
4     301
```
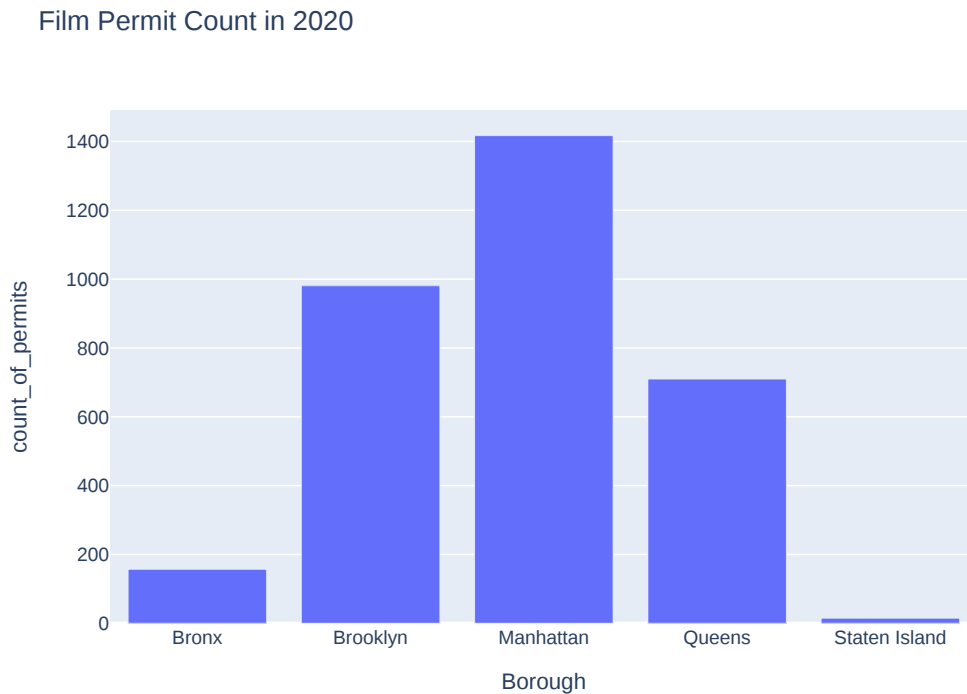
Get the 2020 and 2021 data, becuse they are the dates of interest.

```
[9]: fp20 = fp[(fp['StartDateTime']>='2020-1-1')&(fp['StartDateTime']<='2020-12-31')]
     fp21 = fp[(fp['StartDateTime']>='2021-1-1')]
```
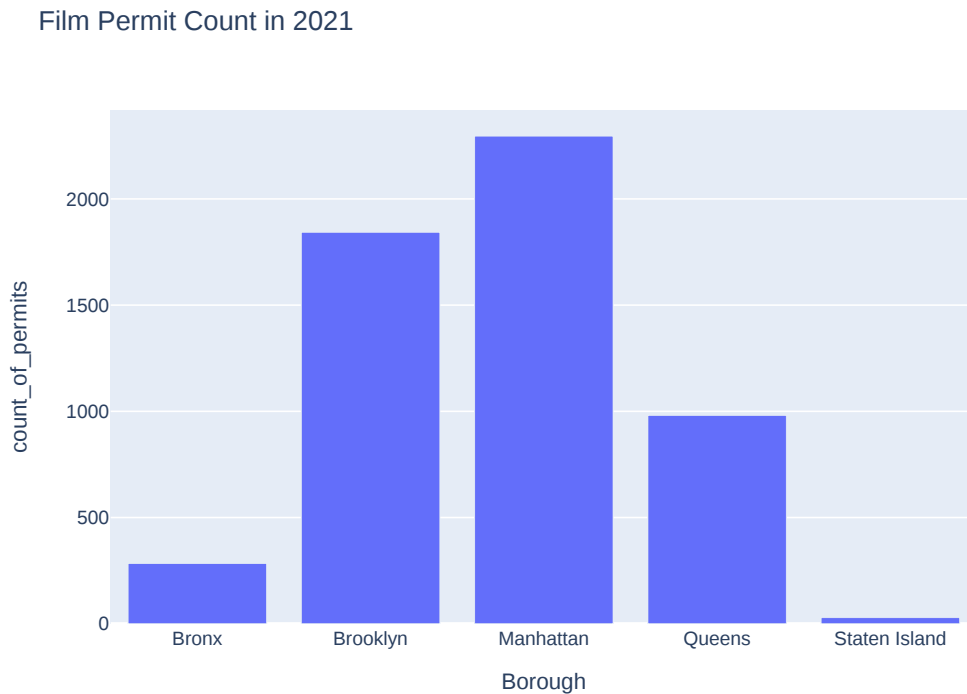
3 Data Analysis

3.1 Film Permits by Borough

```
[10]: borough_counts = fp20.groupby('Borough').size().
      ↪reset_index(name='count_of_permits')
      fig = px.bar(borough_counts, x="Borough", y="count_of_permits", title = 'Film␣
      ↪Permit Count in 2020')
      fig.show()
```

Film Permit Count in 2020


```
[11]: borough_counts = fp21.groupby('Borough').size().
      ↪reset_index(name='count_of_permits')
```

6

```
fig = px.bar(borough_counts, x="Borough", y="count_of_permits", title = 'Film␣
 ↪Permit Count in 2021')
fig.show()
```

Film Permit Count in 2021

3.2 Permit Counts by Community Boards

```
[12]: import requests

      geojson_url = 'https://data.cityofnewyork.us/api/geospatial/yfnk-k7r4?
       ↪method=export&format=GeoJSON'

      response = requests.get(geojson_url)
      geojson_data = response.json()

      fp20_by_cd = fp20.groupby('borocd').size().reset_index(name='count of permits')
      fp21_by_cd = fp21.groupby('borocd').size().reset_index(name='count of permits')
```

```
[13]: fig = px.choropleth_mapbox(fp20_by_cd,
                                 geojson=geojson_url,
                                 locations='borocd',
                                 featureidkey='properties.boro_cd',
```

```
                                color='count of permits',
                                center = {'lat': 40.73, 'lon': -73.98},
                                zoom=9,
                                mapbox_style='carto-positron',
                                title = 'Film Permit Count in 2020')

fig.update_layout(height=700)
fig.show()
```
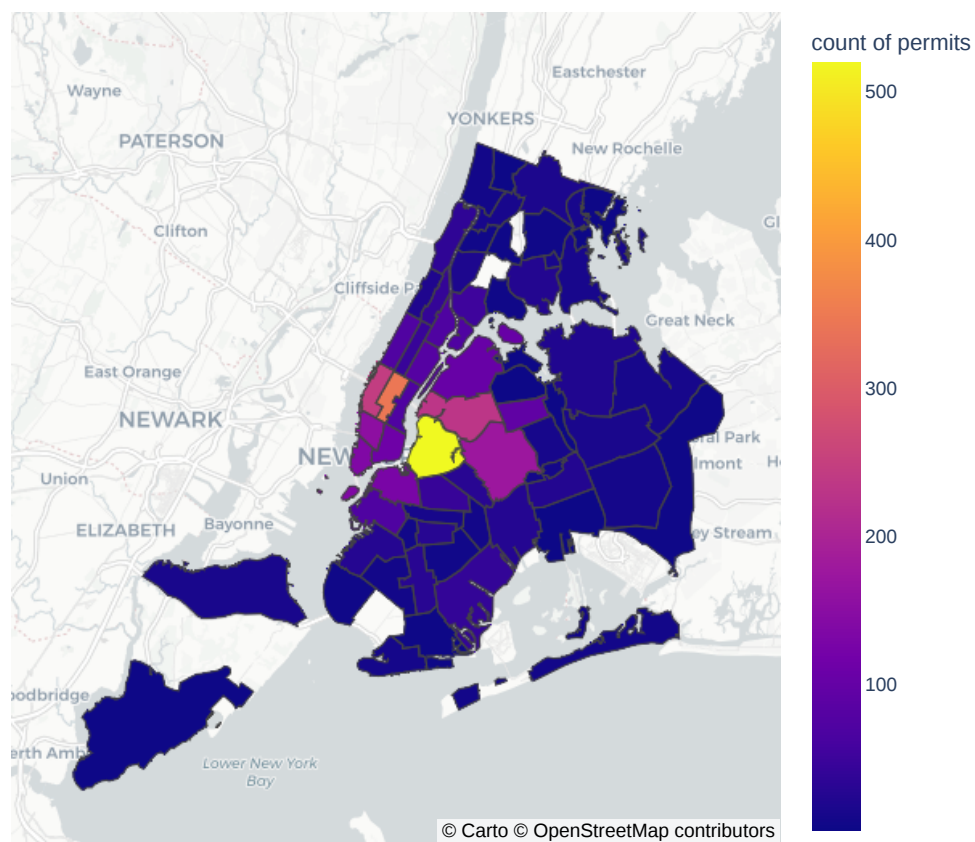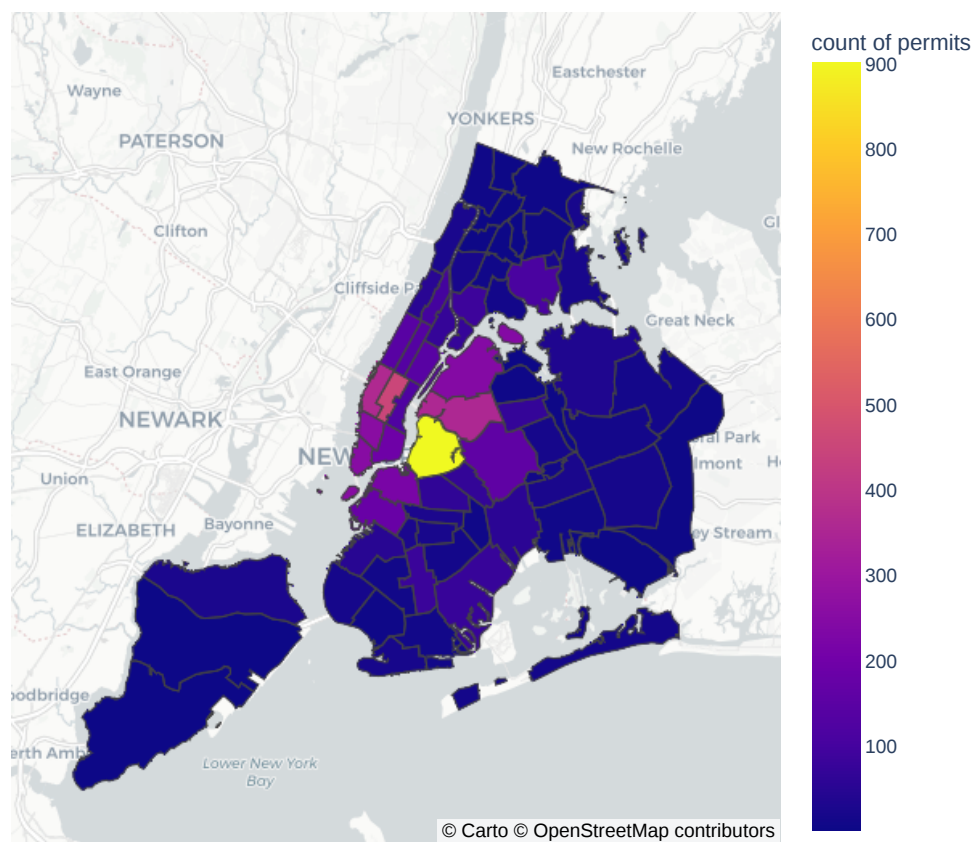
Film Permit Count in 2020


```
[14]: fig = px.choropleth_mapbox(fp21_by_cd,
                             geojson=geojson_url,
                             locations='borocd',
                             featureidkey='properties.boro_cd',
```

```
                    color='count of permits',
                    center = {'lat': 40.73, 'lon': -73.98},
                    zoom=9,
                    mapbox_style='carto-positron',
                    title = 'Film Permit Count in 2021')

fig.update_layout(height=700)
fig.show()
```

Film Permit Count in 2021

3.3 Insight 1

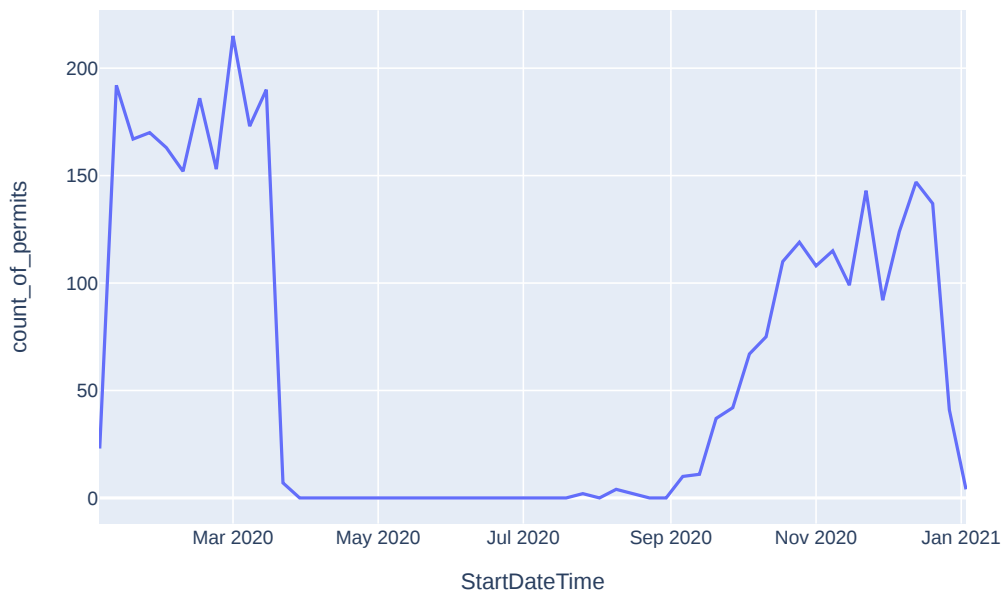In 2020 and 2021, Manhattan was the borough that issued the most permits, followed by Brooklyn and Queens.

At the level of community board, Williamsburg in Brooklyn issued the most permits in 2020 and 2021.

The distribution of permit counts does not vary much between 2020 and 2021.

3.4 Permit Counts by Time

```
[15]: fp20_time = fp20.groupby([pd.Grouper(key='StartDateTime', freq='W')]).size().
      ↪reset_index(name='count_of_permits')
      fig = px.line(fp20_time, x="StartDateTime", y="count_of_permits",title =␣
      ↪'Permit Counts in 2020')
      fig.show()
```

Permit Counts in 2020

3.5 Relationship between permit counts and covid cases.

```
[16]: import plotly.graph_objects as go
      from plotly.subplots import make_subplots

      covidurl = 'https://raw.githubusercontent.com/nychealth/coronavirus-data/master/
      ↪trends/data-by-day.csv'
      covid_by_day = pd.read_csv(covidurl)
```

10

```python
covid_by_day['date_of_interest'] = pd.
 →to_datetime(covid_by_day['date_of_interest'], format='%m/%d/%Y')
covid_by_day =␣
 →covid_by_day[(covid_by_day['date_of_interest']<='2020-12-31')&(covid_by_day['date_of_intere
covid_by_day = covid_by_day.resample('W',on = 'date_of_interest').sum().
 →reset_index()

fig2 = make_subplots(specs=[[{"secondary_y": True}]])
fig2.add_trace(
    go.Scatter(x=covid_by_day['date_of_interest'],␣
 →y=covid_by_day['CASE_COUNT'], name="Case Count"),
    secondary_y=False,
)

fig2.add_trace(
    go.Scatter(x=fp20_time['StartDateTime'], y=fp20_time['count_of_permits'],␣
 →name="Permit Count"),
    secondary_y=True,
)

fig2.update_layout(
    title_text="Film permit count to 311 vs. Covid-19 case counts in NYC"
)

# Set x-axis title
fig2.update_xaxes(title_text="Date")

# Set y-axes titles
fig2.update_yaxes(title_text="Case Count", secondary_y=False)
fig2.update_yaxes(title_text="Permit Count", secondary_y=True)

fig2.show()
```
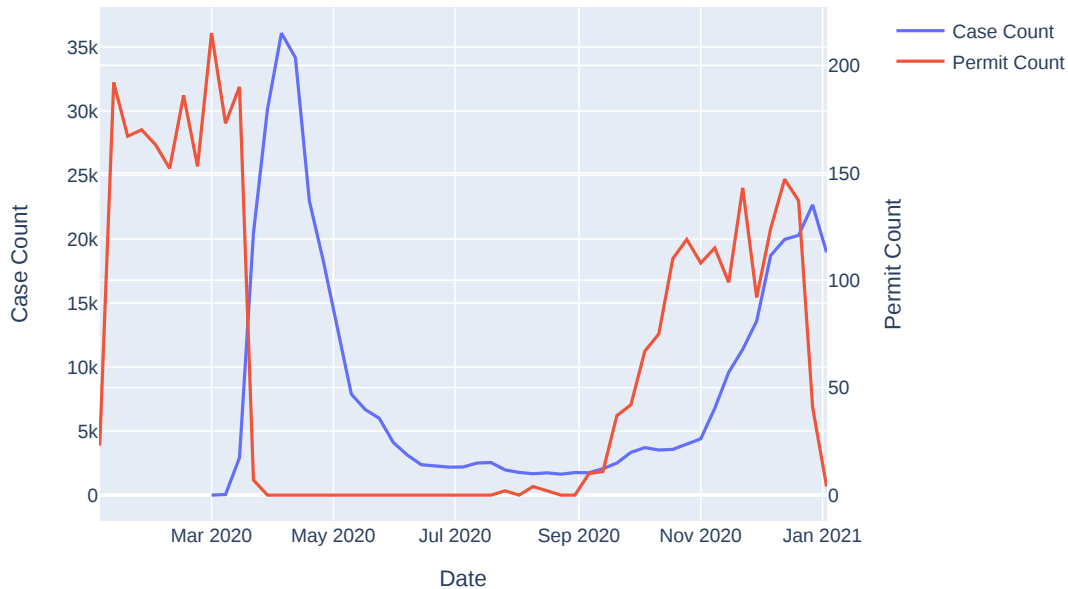
Film permit count to 311 vs. Covid-19 case counts in NYC

3.6 Insight 2

The Permit Counts dropped significantly in March 2020, when the pandamic started.
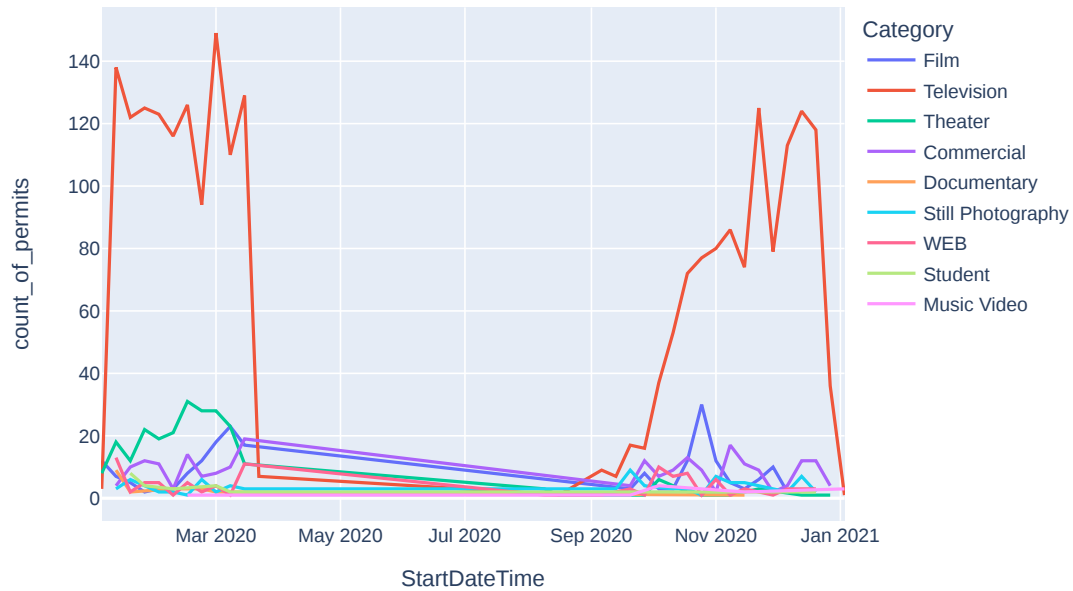
It resumed growth in September 2020.

Film permit counts shared a similar trend with covid case count. In March, increasing covid cases seemed to decrease permit counts, but in September, it seemed to suggest that increasing social activities like shooting films increased the covid cases.
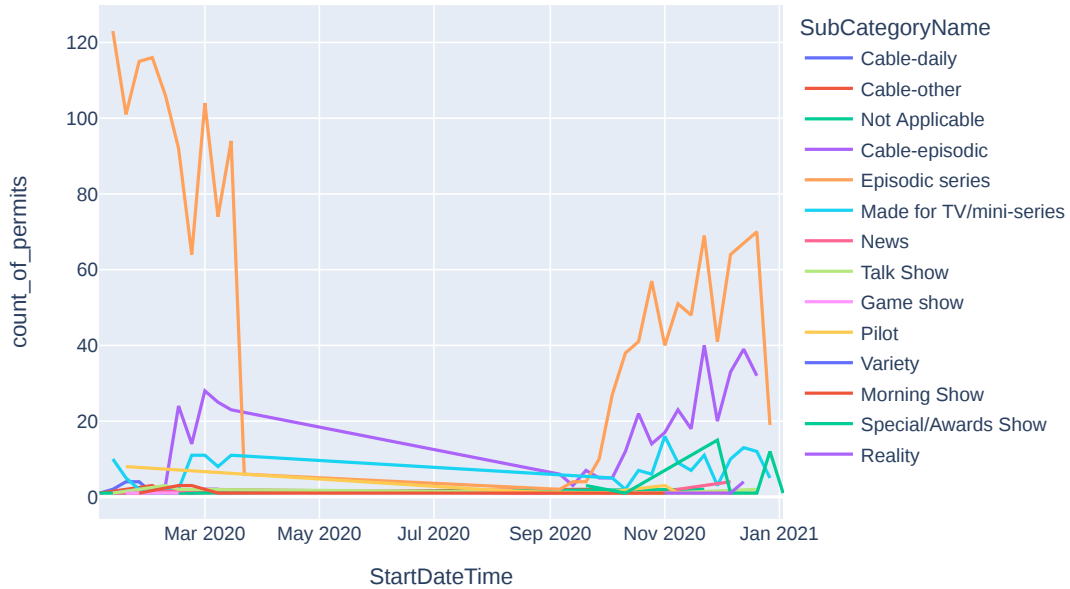
3.7 Categories of Film Permit

```
[17]: fp20_cate = fp20.groupby([pd.Grouper(key='StartDateTime', freq='W'),␣
      ↪'Category']).size().reset_index(name='count_of_permits')
      fig = px.line(fp20_cate, x="StartDateTime",␣
      ↪y="count_of_permits",color='Category', title = 'Permit Counts in 2020')
      fig.show()
```

Permit Counts in 2020


```
[18]: tele = fp20[fp20['Category']=='Television']
      tele_time = tele.groupby([pd.Grouper(key='StartDateTime', freq='W'),␣
       ↪'SubCategoryName']).size().reset_index(name='count_of_permits')
      fig = px.line(tele_time, x="StartDateTime",␣
       ↪y="count_of_permits",color='SubCategoryName', title = 'Permit Counts in␣
       ↪2020')
      fig.show()
```

Permit Counts in 2020

3.8 Insight 3

Television suffered the most, but mostly because it was the most dominant category before the pandemic.

Among all televisions, episodic series experienced the most dramatic fall during the pandemic.

4 Conclusion

4.1 Manhattan is the most popular borough to shoot films, while Williamsburg in Brooklyn is the most popular community board.

4.2 Permit Counts dropped significantly during the pandemic in March 2020, but resumed to growth in September, however, the restart of social activities like fliming seemed to cause another increase of covid cases.

4.3 Television suffered the most during the pandemic, among which the number of episodic series dropped the most. An explaination could be that television was the most dominant category of film shooting in NYC.