# Top 10 Zip Codes according to Permit Data

## Following steps:

1. Import module
2. Import csv file
3. Clean data
4. Filter only the permits we want: ['A2', 'DM', 'NB', 'A1']
5. Find top 10 zip codes by value_counts()

### Import data

In [1]:
```
import pandas as pd
```

In [2]:
```
data=pd.read_csv(filepath_or_buffer =  '../capstone 1/DOB_Permit_Issuance.csv',
```

## Sort Data

***Observation:***

1. The earliest date is December 2020, possibly due to nothing being allowed during the peak of the pandemic.
2. I will do a value counts to see which zip codes have the most activity.

In [3]:
```
data.dropna(inplace=True)
data.sort_values(by=['Issuance Date'])
data
```

Out[3]:

| | BOROUGH | Job Type | Zip Code | Issuance Date |
|---|---|---|---|---|
| **0** | MANHATTAN | A2 | 10020.0 | 12/11/2020 |
| **1** | STATEN ISLAND | A2 | 10301.0 | 12/11/2020 |
| **2** | BROOKLYN | DM | 11209.0 | 06/17/2020 |
| **3** | BROOKLYN | DM | 11226.0 | 06/17/2020 |
| **4** | BROOKLYN | DM | 11210.0 | 06/17/2020 |
| **...** | ... | ... | ... | ... |
| **3747446** | BROOKLYN | A2 | 11231.0 | 05/31/2021 |
| **3747448** | BROOKLYN | A2 | 11205.0 | 05/31/2021 |
| **3747449** | BROOKLYN | A1 | 11230.0 | 05/31/2021 |
| **3747450** | QUEENS | A1 | 11378.0 | 05/31/2021 |
| **3747451** | BROOKLYN | NB | 11231.0 | 05/31/2021 |

3724122 rows × 4 columns

In [47]:
```
# We want to focus only on the recent data andonly want A2 DM NB A1,

# WE DONT WANT A3 and SG

BOROUGH = 'QUEENS'
df = data[data['BOROUGH'] == BOROUGH]
df = df[df['Job Type'].isin(['A2', 'DM', 'NB', 'A1'])]
df['Zip Code'].value_counts().head(10)
list = df['Zip Code'].value_counts().index.to_list()[:10]
list
```

Out[47]:
```
[11101.0,
 11368.0,
 11354.0,
 11355.0,
 11385.0,
 11373.0,
 11357.0,
 11432.0,
 11377.0,
 11691.0]
```

In [48]:
```
BOROUGH = 'MANHATTAN'
df = data[data['BOROUGH'] == BOROUGH]
df = df[df['Job Type'].isin(['A2', 'DM', 'NB', 'A1'])]
df['Zip Code'].value_counts().head(10)
list = df['Zip Code'].value_counts().index.to_list()[:10]
list
```

Out[48]:
```
[10022.0,
 10019.0,
 10013.0,
 10011.0,
 10003.0,
 10017.0,
 10036.0,
 10016.0,
 10001.0,
 10023.0]
```

In [49]:
```
BOROUGH = 'BRONX'
df = data[data['BOROUGH'] == BOROUGH]
df = df[df['Job Type'].isin(['A2', 'DM', 'NB', 'A1'])]
df['Zip Code'].value_counts().head(10)
list = df['Zip Code'].value_counts().index.to_list()[:10]
list
```

Out[49]:
```
[10467.0,
 10456.0,
 10457.0,
 10461.0,
 10469.0,
 10458.0,
 10451.0,
 10459.0,
 10473.0,
 10460.0]
```

In [50]:
```
BOROUGH = 'BROOKLYN'
df = data[data['BOROUGH'] == BOROUGH]
df = df[df['Job Type'].isin(['A2', 'DM', 'NB', 'A1'])]
df['Zip Code'].value_counts().head(10)
list = df['Zip Code'].value_counts().index.to_list()[:10]
list
```

Out[50]:
```
[11201.0,
 11215.0,
 11221.0,
 11211.0,
 11206.0,
 11220.0,
 11207.0,
 11238.0,
 11235.0,
 11219.0]
```

In [51]:
```
BOROUGH = 'STATEN ISLAND'
df = data[data['BOROUGH'] == BOROUGH]
df = df[df['Job Type'].isin(['A2', 'DM', 'NB', 'A1'])]
df['Zip Code'].value_counts().head(10)
list = df['Zip Code'].value_counts().index.to_list()[:10]
list
```

Out[51]:
```
[10314.0,
 10306.0,
 10312.0,
 10309.0,
 10305.0,
 10304.0,
 10301.0,
 10307.0,
 10303.0,
 10308.0]
```