Battle_of_Neighbourhood4

August 1, 2020

Capstone Project - Battle of Neighbourhoods

Introduction:-

1.1 Business Problem

The objective of this Capstone Project is to analyse and choose the safest borough in the New York City based on the total crimes. This will help the people to buy/rent a home who are newly arriving to NYC. Exploring the neighbourhood and select the best among the five boroughs Brooklyn, Queens, Manhattan, The Bronx and Staten Island it has.

1.2 Who will use it?

The target audience for this problem will be all the individual or the families moving to this new place to make a decision of which location is safe and will be suitable for there preferences.

[1]:
```python
import requests
import pandas as pd
import numpy as np
```

[2]:
```python
CLIENT_ID = 'X5DC02PSOJVYTXTIHFY2PGAGVOQZRAUZK3LLRJLWR3IBLLCP'
CLIENT_SECRET = 'MOQER4RYNVWEFMV3CC3NOVAV4KSAPU5E5FE33QIBGJLGCANR'

VERSION = '20180604'
LIMIT = 30

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: X5DC02PSOJVYTXTIHFY2PGAGVOQZRAUZK3LLRJLWR3IBLLCP
CLIENT_SECRET:MOQER4RYNVWEFMV3CC3NOVAV4KSAPU5E5FE33QIBGJLGCANR
```

[3]:
```python
df = pd.read_csv("NYPD_Crime_Data.csv")
```

[4]:
```python
df.head()
```

[4]:
```
   CMPLNT_NUM CMPLNT_FR_DT  RPT_DT  KY_CD                 OFNS_DESC  \
0   574970069       1/1/19  1/1/19    341             PETIT LARCENY
1   695390287       1/1/19  1/1/19    109             GRAND LARCENY
```

```
2   553237569     11/25/18  1/1/19    114                          ARSON
3   320312402       1/1/19  1/1/19    344  ASSAULT 3 & RELATED OFFENSES
4   936158061       1/1/19  1/1/19    578                HARRASSMENT 2

   PD_CD   LAW_CAT_CD    BORO_NM   Latitude  Longitude  \
0  338.0  MISDEMEANOR      BRONX  40.890285 -73.859106
1  411.0       FELONY  MANHATTAN  40.851404 -73.932216
2  264.0       FELONY     QUEENS  40.680003 -73.764022
3  101.0  MISDEMEANOR   BROOKLYN  40.596940 -73.973665
4  637.0    VIOLATION  MANHATTAN  40.856200 -73.934015

                                    Lat_Lon
0   (40.89028471600005, -73.85910627199996)
1  (40.851403574000074, -73.93221569599996)
2   (40.68000300400007, -73.76402239699996)
3   (40.59694042900003, -73.97366455699995)
4   (40.85619961300006, -73.93401465599999)
```

[5]:

```
df['value']=1
```

[6]:

```
df.shape
```

[6]:

```
(482337, 12)
```

[7]:

```
df.columns = ['Crime_No',␣
↪'Crime_DT','Crime_Reported_DT','Classification_Code','Offence_Desc','Internal_Code','Level'
```

[8]:

```
df.head()
```

[8]:

```
     Crime_No  Crime_DT Crime_Reported_DT  Classification_Code  \
0  574970069    1/1/19            1/1/19                  341
1  695390287    1/1/19            1/1/19                  109
2  553237569  11/25/18            1/1/19                  114
3  320312402    1/1/19            1/1/19                  344
4  936158061    1/1/19            1/1/19                  578

                   Offence_Desc  Internal_Code        Level    Borough  \
0                PETIT LARCENY          338.0  MISDEMEANOR      BRONX
1                GRAND LARCENY          411.0       FELONY  MANHATTAN
2                        ARSON          264.0       FELONY     QUEENS
3  ASSAULT 3 & RELATED OFFENSES          101.0  MISDEMEANOR   BROOKLYN
4                HARRASSMENT 2          637.0    VIOLATION  MANHATTAN

    Latitude  Longitude                                   Lat_Lon  \
0  40.890285 -73.859106   (40.89028471600005, -73.85910627199996)
1  40.851404 -73.932216  (40.851403574000074, -73.93221569599996)
2  40.680003 -73.764022   (40.68000300400007, -73.76402239699996)
```

```
3  40.596940 -73.973665   (40.59694042900003, -73.97366455699995)
4  40.856200 -73.934015   (40.85619961300006, -73.93401465599999)

   No_of_crimes
0             1
1             1
2             1
3             1
4             1
```

[9]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 482337 entries, 0 to 482336
Data columns (total 12 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   Crime_No             482337 non-null  int64
 1   Crime_DT             482337 non-null  object
 2   Crime_Reported_DT    482337 non-null  object
 3   Classification_Code  482337 non-null  int64
 4   Offence_Desc         482317 non-null  object
 5   Internal_Code        481968 non-null  float64
 6   Level                482337 non-null  object
 7   Borough              481961 non-null  object
 8   Latitude             475612 non-null  float64
 9   Longitude            475612 non-null  float64
 10  Lat_Lon              475612 non-null  object
 11  No_of_crimes         482337 non-null  int64
dtypes: float64(3), int64(3), object(6)
memory usage: 44.2+ MB
```

[10]:

```
df['Borough'].value_counts()
```

[10]:

```
BROOKLYN         138382
MANHATTAN        121550
BRONX            104825
QUEENS            97201
STATEN ISLAND     20003
Name: Borough, dtype: int64
```

[11]:

```
df['Level'].value_counts()
```

[11]:

```
MISDEMEANOR    254977
FELONY         152691
VIOLATION       74669
Name: Level, dtype: int64
```

```
[12]: NYPD_crime = pd.pivot_table(df,values=['No_of_crimes'],
                               index=['Borough'],
                               columns=['Level'],
                               aggfunc=np.sum,fill_value=0)
      NYPD_crime.head()
```

```
[12]:               No_of_crimes
      Level               FELONY MISDEMEANOR VIOLATION
      Borough
      BRONX                30356       57102     17367
      BROOKLYN             46631       70504     21247
      MANHATTAN            38903       66785     15862
      QUEENS               31369       49857     15975
      STATEN ISLAND         5059       10727      4217
```

```
[13]: NYPD_crime.reset_index(inplace = True)
```

```
[14]: NYPD_crime['Total'] = NYPD_crime.sum(axis=1)
      NYPD_crime.head(33)
```

```
[14]:              Borough No_of_crimes                          Total
      Level                      FELONY MISDEMEANOR VIOLATION
      0              BRONX        30356       57102     17367  104825
      1           BROOKLYN        46631       70504     21247  138382
      2          MANHATTAN        38903       66785     15862  121550
      3             QUEENS        31369       49857     15975   97201
      4      STATEN ISLAND         5059       10727      4217   20003
```

```
[15]: NYPD_crime.columns = NYPD_crime.columns.map(''.join)
      NYPD_crime.head()
```

```
[15]:          Borough  No_of_crimesFELONY  No_of_crimesMISDEMEANOR  \
      0          BRONX               30356                    57102
      1       BROOKLYN               46631                    70504
      2      MANHATTAN               38903                    66785
      3         QUEENS               31369                    49857
      4  STATEN ISLAND                5059                    10727

         No_of_crimesVIOLATION   Total
      0                  17367  104825
      1                  21247  138382
      2                  15862  121550
      3                  15975   97201
      4                   4217   20003
```

```
[16]: NYPD_crime.columns = ['Borough','Felony', 'Misdemeanor','Violation','Total']
      NYPD_crime.head()
```

```
[16]:          Borough  Felony  Misdemeanor  Violation   Total
      0          BRONX   30356        57102      17367  104825
      1       BROOKLYN   46631        70504      21247  138382
      2      MANHATTAN   38903        66785      15862  121550
      3         QUEENS   31369        49857      15975   97201
      4  STATEN ISLAND    5059        10727       4217   20003
```

```
[17]: !conda install -c anaconda lxml --yes
      !conda install -c anaconda beautifulsoup4 --yes
      #from bs4 import BeautifulSoup
      import requests
      from bs4 import BeautifulSoup
      import xml
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/jupyterlab/conda/envs/python

  added / updated specs:
    - lxml


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    ca-certificates-2020.6.24  |                0         133 KB  anaconda
    certifi-2020.6.20          |           py36_0         160 KB  anaconda
    libxml2-2.9.10             |       he19cac6_1         1.3 MB  anaconda
    libxslt-1.1.34             |       hc22bd24_0         573 KB  anaconda
    lxml-4.5.2                 |   py36hefd8a0e_0         1.4 MB  anaconda
    openssl-1.1.1g             |       h7b6447c_0         3.8 MB  anaconda
    ------------------------------------------------------------
                                           Total:         7.3 MB

The following NEW packages will be INSTALLED:

  libxslt            anaconda/linux-64::libxslt-1.1.34-hc22bd24_0
  lxml               anaconda/linux-64::lxml-4.5.2-py36hefd8a0e_0

The following packages will be UPDATED:

  ca-certificates    conda-forge::ca-certificates-2020.6.2~ --> anaconda::ca-
certificates-2020.6.24-0
```

```
  libxml2             conda-forge::libxml2-2.9.9-h13577e0_2 -->
anaconda::libxml2-2.9.10-he19cac6_1

The following packages will be SUPERSEDED by a higher-priority channel:

  certifi            conda-forge::certifi-2020.6.20-py36h9~ -->
anaconda::certifi-2020.6.20-py36_0
  openssl            conda-forge::openssl-1.1.1g-h516909a_0 -->
anaconda::openssl-1.1.1g-h7b6447c_0



Downloading and Extracting Packages
openssl-1.1.1g       | 3.8 MB    | ##################################### | 100%
libxml2-2.9.10       | 1.3 MB    | ##################################### | 100%
ca-certificates-2020 | 133 KB    | ##################################### | 100%
certifi-2020.6.20    | 160 KB    | ##################################### | 100%
lxml-4.5.2           | 1.4 MB    | ##################################### | 100%
libxslt-1.1.34       | 573 KB    | ##################################### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/jupyterlab/conda/envs/python

  added / updated specs:
    - beautifulsoup4


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    beautifulsoup4-4.9.1       |           py36_0         168 KB  anaconda
    soupsieve-2.0.1            |             py_0          33 KB  anaconda
    ------------------------------------------------------------
                                           Total:         201 KB

The following NEW packages will be INSTALLED:

  beautifulsoup4     anaconda/linux-64::beautifulsoup4-4.9.1-py36_0
  soupsieve          anaconda/noarch::soupsieve-2.0.1-py_0
```

```
Downloading and Extracting Packages
soupsieve-2.0.1      | 33 KB     | ##################################### | 100%
beautifulsoup4-4.9.1 | 168 KB    | ##################################### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

[19]:

```python
wikipedia_link='https://www.citypopulation.de/en/usa/newyorkcity/'
raw_wikipedia_page= requests.get(wikipedia_link).text

soup = BeautifulSoup(raw_wikipedia_page,'xml')
table=soup.find('table')
print(soup.prettify())
```

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE html>
<html lang="en">
 <head>
  <meta charset="utf-8">
   <meta content="New York City Boroughs (USA): Boroughs with population
statistics, charts and maps." name="description">
    <title>
     New York City Boroughs (USA): Boroughs - Population Statistics, Charts and
Map
    </title>
    <link href="/favicon.ico" rel="shortcut icon">
     <script>
      var pagemode = 'adminpage'; var pagecat = 'admin_city'; var isAdmin =
false; var pageid = 'usa-newyorkcity'; var pagelang = 'en'; var pagelabel = "New
York City Boroughs"; var popDate = 'E 2019-07-01'; var popcolnum  = 4; var
start_x = -73.975; var start_y = 40.705; var start_level = 10; var swap_width =
1132; var hor_percent = 40; var vert_percent = 42; var swap = 'true'; var
mapcopyright = 'U.S. Census Bureau.'; var objid = ''; var objtype = ''; var
startmap = 'street'; var lev_num = 1; var edit_mode = ''; var placeLocale =
'en'; var nativeName = false; var wikiFromWD = false
     </script>
     <script src="/js/countries/usa.js"/>
     <script src="/jquery/jquery-3.1.1.min.js"/>
     <script src="/js/cp_data_m.js"/>
     <script src="/js/cp_phpbase_v3.js"/>
     <script>
      load_resources()
     </script>
     <script src="/js/cp_menu.js"/>
     <style>
      article#admtable { top: 42%; }
div#admmap { height: 42%; }
```

```
@media all and (min-width: 1132px) {
        article#admtable { top: 26px; left: calc(180px + 40%); }
        div#admmap { height: auto; bottom: -1px; width: 40% }
        header.admpage { left: calc(180px + 40%) }
        div#headline { left: calc(162px + 40%) }
}
     </style>
    </link>
    <body itemscope="" itemtype="http://schema.org/City" onload="init_data();
start_maps()">
     <script>
      writeMenu('en')
     </script>
     <div class="mobiadv">
      <script>
       show_mobiadv();
      </script>
     </div>
     <div class="hor" id="headline">
      <div id="orient">
       <a href="/">
        Home
       </a>
       →
       <span itemprop="containedIn" itemscope=""
itemtype="http://schema.org/Continent">
        <a href="/America.html" itemprop="url">
         <span itemprop="name">
          America
         </span>
        </a>
       </span>
       →
       <span itemprop="containedIn" itemscope=""
itemtype="http://schema.org/Country">
        <a href="/en/usa/" itemprop="url">
         <span itemprop="name">
          USA
         </span>
        </a>
       </span>
      </div>
      <div id="social">
       <div class="changelang">
        <a href="javascript:cp.changePageLang('en','de')">
         <img alt="" src="/images/icons/de.svg" title="Deutsch"/>
        </a>
       </div>
      </div>
      <div class="info" id="inforowdiv" style="display:none"/>
      <article class="cpage swapped" id="admtable">
       <header class="admpage">
        <a href="javascript:openMap()">
         <img alt="Show Map" id="smap" src="/images/smaps/usa-cities.png"
title="Show Map"/>
         <h1>
          <a href="/en/usa/">
           USA
          </a>
          :
          <span class="smalltext" itemprop="name">
           New York City Boroughs
          </span>
         </h1>
        </a>
        <script>
         handleArticleResize()
        </script>
        <h2>
         <span class="noviz">
          Contents:
         </span>
         Boroughs
        </h2>
        <p>
         The population of the boroughs of New York City according to census
results and latest official estimates.
        </p>
        <p class="small noprint">
         The
         <img alt="Details" class="infoicon"
src="/images/icons/separate_wb.svg">
          icon links to further information about a selected division including
its population structure (gender, age groups, age distribution, »race«,
ethnicity).
         </img>
         <table class="data" id="ts">
          <thead>
           <tr>
            <th class="rname" data-coltype="name"
onclick="javascript:sort('ts',0,false)">
             <a href="javascript:sort('ts',0,false)">
              Name
             </a>
            </th>
            <th class="rstatus" data-coltype="status"
```

```
onclick="javascript:sort('ts',1,false)">
             <a href="javascript:sort('ts',1,false)">
              Status
             </a>
            </th>
            <th class="rpop prio4" data-coldate="1990-04-01" data-colhead="C
1990-04-01" data-coltype="pop" onclick="javascript:sort('ts',2,true)">
             <a href="javascript:sort('ts',2,true)">
              Population
             </a>
             <br>
              <span class="unit">
               Census
               <br>
                1990-04-01
               </br>
              </span>
              <th class="rpop prio3" data-coldate="2000-04-01" data-colhead="C
2000-04-01" data-coltype="pop" onclick="javascript:sort('ts',3,true)">
               <a href="javascript:sort('ts',3,true)">
                Population
               </a>
               <br>
                <span class="unit">
                 Census
                 <br>
                  2000-04-01
                 </br>
                </span>
                <th class="rpop prio2" data-coldate="2010-04-01" data-colhead="C
2010-04-01" data-coltype="pop" onclick="javascript:sort('ts',4,true)">
                 <a href="javascript:sort('ts',4,true)">
                  Population
                 </a>
                 <br>
                  <span class="unit">
                   Census
                   <br>
                    2010-04-01
                   </br>
                  </span>
                  <th class="rpop prio1" data-coldate="2019-07-01" data-
colhead="E 2019-07-01" data-coltype="pop"
onclick="javascript:sort('ts',5,true)">
                   <a href="javascript:sort('ts',5,true)">
                    Population
                   </a>
                   <br>
```

```
                    <span class="unit">
                     Estimate
                     <br>
                      2019-07-01
                     </br>
                    </span>
                    <th class="sc" data-coltype="other"/>
                   </br>
                  </th>
                  <tbody>
                   <tr class="rname" itemscope=""
itemtype="http://schema.org/Place" onclick="javascript:sym('36005')">
                    <td class="rname" data-area="108.91" data-density="13021.75"
data-wd="Q18426" data-wiki="The Bronx" id="i36005">
                     <a href="javascript:sym('36005')">
                      <span itemprop="name">
                       Bronx
                      </span>
                     </a>
                    </td>
                    <td class="rstatus">
                     Borough
                    </td>
                    <td class="rpop prio4">
                     1,203,789
                    </td>
                    <td class="rpop prio3">
                     1,332,244
                    </td>
                    <td class="rpop prio2">
                     1,384,580
                    </td>
                    <td class="rpop prio1">
                     1,418,207
                    </td>
                    <td class="sc">
                     <a href="/en/usa/newyorkcity/36005__bronx/" itemprop="url">
                      →
                     </a>
                    </td>
                   </tr>
                   <tr class="rname" itemscope=""
itemtype="http://schema.org/Place" onclick="javascript:sym('36047')">
                    <td class="rname" data-area="180.81" data-density="14157.68"
data-wd="Q18419" data-wiki="Brooklyn" id="i36047">
                     <a href="javascript:sym('36047')">
                      <span itemprop="name">
                       Brooklyn
                      </span>
                     </a>
                     (
                     <span itemprop="name">
                      Kings County
                     </span>
                     )
                    </td>
                    <td class="rstatus">
                     Borough
                    </td>
                    <td class="rpop prio4">
                     2,300,664
                    </td>
                    <td class="rpop prio3">
                     2,465,689
                    </td>
                    <td class="rpop prio2">
                     2,504,721
                    </td>
                    <td class="rpop prio1">
                     2,559,903
                    </td>
                    <td class="sc">
                     <a href="/en/usa/newyorkcity/36047__brooklyn/"
itemprop="url">
                      →
                     </a>
                    </td>
                   </tr>
                   <tr class="rname" itemscope=""
itemtype="http://schema.org/Place" onclick="javascript:sym('36061')">
                    <td class="rname" data-area="58.68" data-density="27756.53"
data-wd="Q11299" data-wiki="Manhattan" id="i36061">
                     <a href="javascript:sym('36061')">
                      <span itemprop="name">
                       Manhattan
                      </span>
                     </a>
                     (
                     <span itemprop="name">
                      New York County
                     </span>
                     )
                    </td>
                    <td class="rstatus">
                     Borough
                    </td>
                    <td class="rpop prio4">
                     1,487,536
                    </td>
                    <td class="rpop prio3">
                     1,538,096
                    </td>
                    <td class="rpop prio2">
                     1,586,381
                    </td>
                    <td class="rpop prio1">
                     1,628,706
                    </td>
                    <td class="sc">
                     <a href="/en/usa/newyorkcity/36061__manhattan/"
itemprop="url">
                      →
                     </a>
                    </td>
                   </tr>
                   <tr class="rname" itemscope=""
itemtype="http://schema.org/Place" onclick="javascript:sym('36081')">
                    <td class="rname" data-area="281.71" data-density="8000.69"
data-wd="Q18424" data-wiki="Queens" id="i36081">
                     <a href="javascript:sym('36081')">
                      <span itemprop="name">
                       Queens
                      </span>
                     </a>
                    </td>
                    <td class="rstatus">
                     Borough
                    </td>
                    <td class="rpop prio4">
                     1,951,598
                    </td>
                    <td class="rpop prio3">
                     2,229,394
                    </td>
                    <td class="rpop prio2">
                     2,230,619
                    </td>
                    <td class="rpop prio1">
                     2,253,858
                    </td>
                    <td class="sc">
                     <a href="/en/usa/newyorkcity/36081__queens/"
itemprop="url">
                      →
                     </a>
                    </td>
                   </tr>
                   <tr class="rname" itemscope=""
itemtype="http://schema.org/Place" onclick="javascript:sym('36085')">
                    <td class="rname" data-area="150.68" data-density="3159.96"
data-wd="Q18432" data-wiki="Staten Island" id="i36085">
                     <a href="javascript:sym('36085')">
                      <span itemprop="name">
                       Staten Island
                      </span>
                     </a>
                     (
                     <span itemprop="name">
                      Richmond County
                     </span>
                     )
                    </td>
                    <td class="rstatus">
                     Borough
                    </td>
                    <td class="rpop prio4">
                     378,977
                    </td>
                    <td class="rpop prio3">
                     443,762
                    </td>
                    <td class="rpop prio2">
                     468,730
                    </td>
                    <td class="rpop prio1">
                     476,143
                    </td>
                    <td class="sc">
                     <a href="/en/usa/newyorkcity/36085__staten_island/"
itemprop="url">
                      →
                     </a>
                    </td>
                   </tr>
                  </tbody>
                  <tbody class="admin0">
                   <tr>
                    <td class="rname">
                     New York City
                    </td>
                    <td class="rstatus">
                     City
```

```
   </td>
   <td class="rpop prio4">
    7,322,564
   </td>
   <td class="rpop prio3">
    8,009,185
   </td>
   <td class="rpop prio2">
    8,175,031
   </td>
   <td class="rpop prio1">
    8,336,817
   </td>
   <td class="sc"/>
  </tr>
 </tbody>
</br>
<section id="sourcesection">
 <p class="source">
  <strong>
   Source:
  </strong>
  U.S. Census Bureau (web).
 </p>
</section>
<hr id="hraddinfo">
 <h3>
  Further information about the population structure:
 </h3>
 <div id="chartgrid">
  <section class="addinfo">
   <div class="addchart" id="addchart0"/>
   <table class="data">
    <thead>
     <tr>
      <th colspan="2">
       Gender (E 2019)
      </th>
     </tr>
    </thead>
    <tbody>
     <tr>
      <td>
       Males
      </td>
      <td class="rpop">
       3,978,439
      </td>
     </tr>
     <tr>
      <td>
       Females
      </td>
      <td class="rpop">
       4,358,378
      </td>
     </tr>
    </tbody>
   </table>
  </section>
  <section class="addinfo">
   <div class="addchart" id="addchart1"/>
   <table class="data">
    <thead>
     <tr>
      <th colspan="2">
       Age Groups (E 2019)
      </th>
     </tr>
    </thead>
    <tbody>
     <tr>
      <td>
       0-14 years
      </td>
      <td class="rpop">
       1,451,817
      </td>
     </tr>
     <tr>
      <td>
       15-64 years
      </td>
      <td class="rpop">
       5,604,595
      </td>
     </tr>
     <tr>
      <td>
       65+ years
      </td>
      <td class="rpop">
       1,280,405
      </td>
     </tr>
    </tbody>
   </table>
  </section>
  <section class="addinfo">
   <div class="addchart" id="addchart2"/>
   <table class="data">
    <thead>
     <tr>
      <th colspan="2">
       Age Distribution (E 2019)
      </th>
     </tr>
    </thead>
    <tbody>
     <tr>
      <td>
       0-9 years
      </td>
      <td class="rpop">
       1,008,031
      </td>
     </tr>
     <tr>
      <td>
       10-19 years
      </td>
      <td class="rpop">
       883,550
      </td>
     </tr>
     <tr>
      <td>
       20-29 years
      </td>
      <td class="rpop">
       1,273,671
      </td>
     </tr>
     <tr>
      <td>
       30-39 years
      </td>
      <td class="rpop">
       1,335,563
      </td>
     </tr>
     <tr>
      <td>
       40-49 years
      </td>
      <td class="rpop">
       1,043,319
      </td>
     </tr>
     <tr>
      <td>
       50-59 years
      </td>
      <td class="rpop">
       1,033,138
      </td>
     </tr>
     <tr>
      <td>
       60-69 years
      </td>
      <td class="rpop">
       878,204
      </td>
     </tr>
     <tr>
      <td>
       70-79 years
      </td>
      <td class="rpop">
       543,337
      </td>
     </tr>
     <tr>
      <td>
       80+ years
      </td>
      <td class="rpop">
       338,004
      </td>
     </tr>
    </tbody>
   </table>
  </section>
  <section class="addinfo">
   <div class="addchart" id="addchart3"/>
   <table class="data">
    <thead>
     <tr>
      <th colspan="2">
       »Race« (E 2019)
      </th>
     </tr>
    </thead>
    <tbody>
     <tr>
      <td>
       White
      </td>
      <td class="rpop">
       4,393,042
      </td>
     </tr>
     <tr>
      <td>
       Black/African American
      </td>
      <td class="rpop">
       2,093,874
      </td>
     </tr>
     <tr>
      <td>
       Indigenous
      </td>
      <td class="rpop">
       116,497
      </td>
     </tr>
     <tr>
      <td>
       Asian
      </td>
      <td class="rpop">
       1,256,584
      </td>
     </tr>
     <tr>
      <td>
       Pacific Islander
      </td>
      <td class="rpop">
       17,682
      </td>
     </tr>
     <tr>
      <td>
       2 or more
      </td>
      <td class="rpop">
```

```
                        258,314
                       </td>
                      </tr>
                     </tbody>
                    </table>
                   </section>
                   <section class="addinfo">
                    <div class="addchart" id="addchart4"/>
                    <table class="data">
                     <thead>
                      <tr>
                       <th colspan="2">
                        Ethnicity (E 2019)
                       </th>
                      </tr>
                     </thead>
                     <tbody>
                      <tr>
                       <td>
                        Hispanic or Latino
                       </td>
                       <td class="rpop">
                        2,423,590
                       </td>
                      </tr>
                      <tr>
                       <td>
                        Other
                       </td>
                       <td class="rpop">
                        5,913,227
                       </td>
                      </tr>
                     </tbody>
                    </table>
                   </section>
                  </div>
                  <script>
                   var addChartData = [{"name":"Gender","type":"pie","data":[["G
ender","Persons"],["Males",3978439],["Females",4358378]]},
{"name":"Age Groups","type":"pie","data":[["Age Groups","Persons"],["0-14
years",1451817],["15-64 years",5604595],["65+ years",1280405]]},
{"name":"Age Distribution","type":"column","data":[["Age
Distribution","Persons"],["0-9 years",1008031],["10-19 years",883550],["20-29
years",1273671],["30-39 years",1335563],["40-49 years",1043319],["50-59
years",1033138],["60-69 years",878204],["70-79 years",543337],["80+
years",338004]]},
{"name":"»Race«","type":"pie","data":[["»Race«","Persons"],["White",4393042],["B
lack/African
American",2093874],["Indigenous",116497],["Asian",1256584],["Pacific
Islander",17682],["2 or more",258314]]},
{"name":"Ethnicity","type":"pie","data":[["Ethnicity","Persons"],["Hispanic or
Latino",2423590],["Other",5913227]]}]
                  </script>
                  <script>
                   var addMapData = ["genderM","genderF","ageX","ageX","ageO"];
var addMapMetadata = [{ "maptype":"genderM", "date":"E 2019-07-01" },{
"maptype":"ageX", "date":"E 2019-07-01" }]
                  </script>
                  <div class="mobiadv">
                   <script>
                    show_mobiadv()
                   </script>
                  </div>
                  <div id="advhor">
                   <script>
                    show_adv('h');
                   </script>
                  </div>
                  <hr>
                   <section class="ytvideosec">
                    <h2>
                     Greater New York: COVID-19 cases, incidence rates and
growth by counties
                    </h2>
                    <iframe allow="accelerometer; autoplay; encrypted-media;
gyroscope; picture-in-picture" class="ytvideo"
src="https://www.youtube.com/embed/mKAGHkMhlXU"/>
                   </section>
                   <script>
                    writeFooter('2020-07-11')
                   </script>
                  </hr>
                  <div id="admmap" itemprop="geo" itemscope=""
itemtype="http://schema.org/GeoCoordinates">
                   <meta content="40.705" itemprop="latitude">
                    <meta content="-73.975" itemprop="longitude">
                     <div id="mapcontainer">
                      <div id="mapdiv">
                       <div id="maplconrl"/>
                       <div id="maprconrl"/>
                      </div>
                     </div>
                    </meta>
                    <div id="alert"/>
                    <div id="helpdiv"/>
                    <div id="adv">
                     <script>
                      show_adv();
                     </script>
                    </div>
                    <script>
                     var startChartID = "NYC"; var startChartType = "adm1";
        var minlat = 40.49; var minlng = -74.26; var maxlat = 40.92; var maxlng
= -73.69;
        var admCount = { "adm1": 5, "adm2": 0 };
        if (cp.getVizMode() == cp.VIZMODE_DESKTOP) cp.social.addSocial(false);
                    </script>
                   </meta>
                  </div>
                  <!-- create time: 0.0037448406219482 countries -->
                  <!-- cache time: 0.00015115737915039 -->
                 </hr>
                </th>
               </br>
              </th>
             </br>
            </th>
           </tr>
          </thead>
         </table>
        </p>
       </header>
      </article>
     </div>
    </body>
   </meta>
  </meta>
 </head>
</html>
```

[20]:

```
NYC_table = pd.read_html('https://www.citypopulation.de/en/usa/newyorkcity/')
NYC=NYC_table[0]
NYC.head()
```

[20]:

```
                              Name  Status  PopulationCensus1990-04-01  \
0                            Bronx  Borough                     1203789
1          Brooklyn (Kings County)  Borough                     2300664
2      Manhattan (New York County)  Borough                     1487536
3                           Queens  Borough                     1951598
4  Staten Island (Richmond County)  Borough                      378977

   PopulationCensus2000-04-01  PopulationCensus2010-04-01  \
0                     1332244                     1384580
1                     2465689                     2504721
2                     1538096                     1586381
3                     2229394                     2230619
4                      443762                      468730

   PopulationEstimate2019-07-01 Unnamed: 6
0                       1418207          →
1                       2559903          →
2                       1628706          →
3                       2253858          →
4                        476143          →
```

[21]:

```
NYC.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 7 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Name                          6 non-null      object
 1   Status                        6 non-null      object
 2   PopulationCensus1990-04-01    6 non-null      int64
 3   PopulationCensus2000-04-01    6 non-null      int64
 4   PopulationCensus2010-04-01    6 non-null      int64
 5   PopulationEstimate2019-07-01  6 non-null      int64
 6   Unnamed: 6                    5 non-null      object
dtypes: int64(4), object(3)
memory usage: 464.0+ bytes
```

[22]:

```
NYC.columns = ['Borough',␣
 ↪'Status','Population-1990','Population-2000','Population-2010','Population-2019','Unnamed']
NYC.head()
```

[22]:

```
                           Borough   Status  Population-1990  Population-2000  \
0                            Bronx  Borough          1203789          1332244
1          Brooklyn (Kings County)  Borough          2300664          2465689
2      Manhattan (New York County)  Borough          1487536          1538096
3                           Queens  Borough          1951598          2229394
4  Staten Island (Richmond County)  Borough           378977           443762

   Population-2010  Population-2019 Unnamed
0          1384580          1418207       →
1          2504721          2559903       →
2          1586381          1628706       →
3          2230619          2253858       →
4           468730           476143       →
```

[23]:

```
NYC_Census=NYC.
 ↪drop(['Unnamed','Population-1990','Population-2000','Population-2010'],␣
 ↪axis=1)
NYC_Census.head()
```

[23]:

```
                             Borough   Status  Population-2019
0                              Bronx  Borough          1418207
1          Brooklyn (Kings County)  Borough          2559903
2      Manhattan (New York County)  Borough          1628706
3                             Queens  Borough          2253858
4  Staten Island (Richmond County)  Borough           476143
```

[24]:

```
NYC_Census=NYC_Census.rename(columns={'B':'Borough'})
```

[25]:

```
NYC_Census.head()
```

[25]:

```
                             Borough   Status  Population-2019
0                              Bronx  Borough          1418207
1          Brooklyn (Kings County)  Borough          2559903
2      Manhattan (New York County)  Borough          1628706
3                             Queens  Borough          2253858
4  Staten Island (Richmond County)  Borough           476143
```

[26]:

```
NYC_Census["Borough"].replace({"Bronx":"BRONX",
                               "Brooklyn (Kings County)":"BROOKLYN",
                               "Manhattan (New York County)":"MANHATTAN",
                               "Queens":"QUEENS",
                               "Staten Island (Richmond County)":"STATEN␣
 ↪ISLAND",
                               "New York City":"NYC"},inplace=True)
```

[27]:

```
print(NYC_Census)
```

```
         Borough   Status  Population-2019
0          BRONX  Borough          1418207
1       BROOKLYN  Borough          2559903
2      MANHATTAN  Borough          1628706
3         QUEENS  Borough          2253858
4  STATEN ISLAND  Borough           476143
5            NYC     City          8336817
```

[28]:

```
NYC_Crime_Table = pd.merge(NYPD_crime, NYC_Census, on='Borough')
NYC_Crime_Table.head()
```

[28]:

```
          Borough  Felony  Misdemeanor  Violation   Total   Status  \
0           BRONX   30356        57102      17367  104825  Borough
1        BROOKLYN   46631        70504      21247  138382  Borough
```

```
2      MANHATTAN   38903        66785      15862  121550  Borough
3         QUEENS   31369        49857      15975   97201  Borough
4  STATEN ISLAND    5059        10727       4217   20003  Borough

   Population-2019
0          1418207
1          2559903
2          1628706
3          2253858
4           476143
```

[29]:
```
NYC_Crime_Table = NYC_Crime_Table[['Borough','Felony','Misdemeanor','Violation',
                 'Status','Population-2019','Total']]
NYC_Crime_Table
```

[29]:
```
         Borough  Felony  Misdemeanor  Violation   Status  Population-2019  \
0          BRONX   30356        57102      17367  Borough          1418207
1       BROOKLYN   46631        70504      21247  Borough          2559903
2      MANHATTAN   38903        66785      15862  Borough          1628706
3         QUEENS   31369        49857      15975  Borough          2253858
4  STATEN ISLAND    5059        10727       4217  Borough           476143

    Total
0  104825
1  138382
2  121550
3   97201
4   20003
```

[30]:
```
NYC_Crime_Table.describe()
```

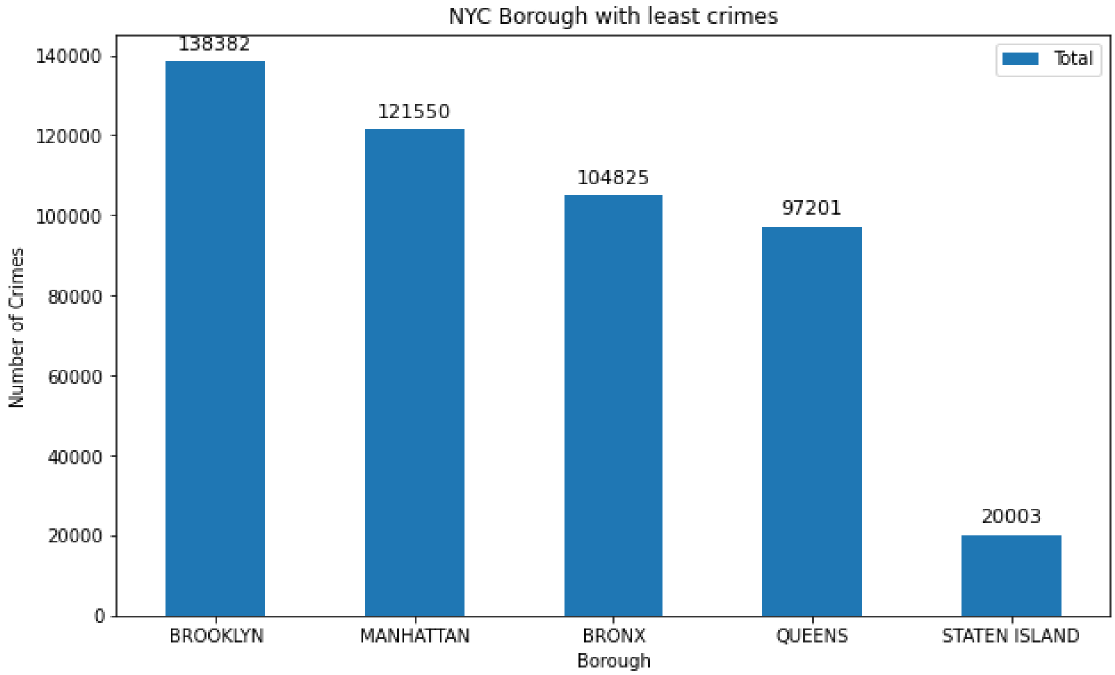
[30]:
```
             Felony   Misdemeanor     Violation  Population-2019  \
count      5.000000      5.000000      5.000000     5.000000e+00
mean   30463.600000  50995.000000  14933.600000     1.667363e+06
std    15643.156037  23927.115883   6375.194334     8.098120e+05
min     5059.000000  10727.000000   4217.000000     4.761430e+05
25%    30356.000000  49857.000000  15862.000000     1.418207e+06
50%    31369.000000  57102.000000  15975.000000     1.628706e+06
75%    38903.000000  66785.000000  17367.000000     2.253858e+06
max    46631.000000  70504.000000  21247.000000     2.559903e+06

              Total
count      5.000000
mean   96392.200000
std    45560.768087
min    20003.000000
25%    97201.000000
```

```
50%     104825.000000
75%     121550.000000
max     138382.000000
```

```
[31]: NYC_Crime_Table.sort_values(['Total'], ascending = False, axis = 0, inplace =␣
       ↪True )
      NYC_Crime_Table
```

```
[31]:          Borough  Felony  Misdemeanor  Violation   Status  Population-2019  \
      1       BROOKLYN   46631        70504      21247  Borough          2559903
      2      MANHATTAN   38903        66785      15862  Borough          1628706
      0          BRONX   30356        57102      17367  Borough          1418207
      3         QUEENS   31369        49857      15975  Borough          2253858
      4  STATEN ISLAND    5059        10727       4217  Borough           476143

          Total
      1  138382
      2  121550
      0  104825
      3   97201
      4   20003
```

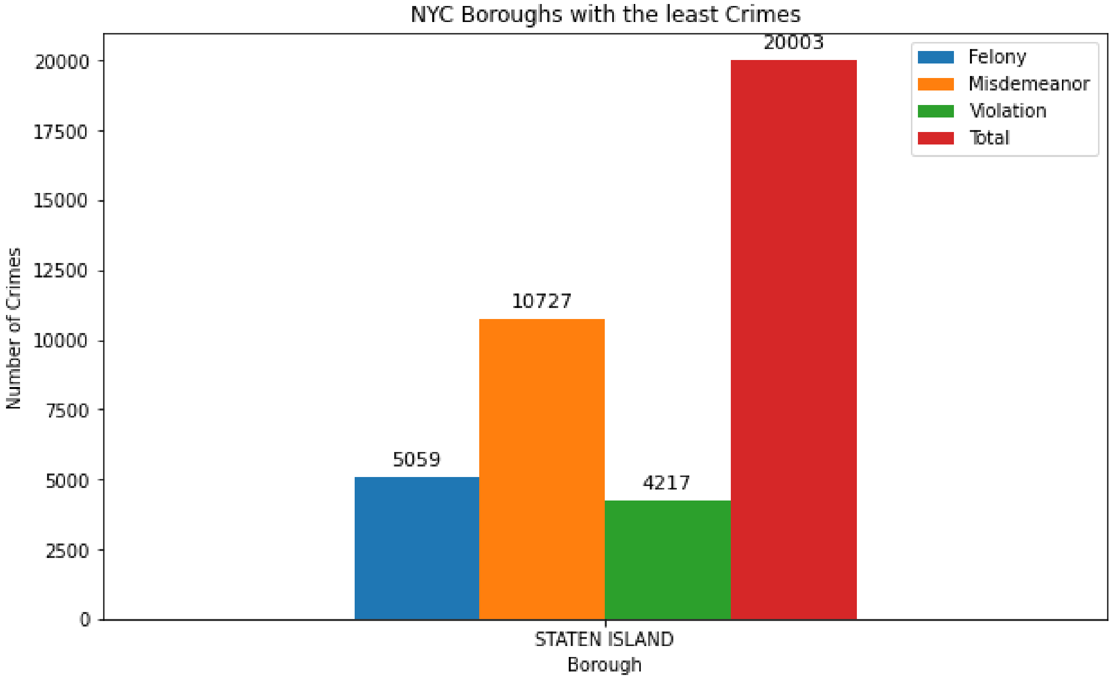
```
[32]: import matplotlib.pyplot as plt
      NYC_V = NYC_Crime_Table[['Borough','Total']]

      NYC_V.set_index('Borough',inplace = True)

      a = NYC_V.plot(kind='bar', figsize=(10, 6), rot=0)

      a.set_ylabel('Number of Crimes')
      a.set_xlabel('Borough')
      a.set_title('NYC Borough with least crimes')

      for p in a.patches:
          a.annotate(np.round(p.get_height(),decimals=2),
                      (p.get_x()+p.get_width()/2., p.get_height()),
                      ha='center',
                      va='center',
                      xytext=(0, 10),
                      textcoords='offset points',
                      fontsize = 11
                     )

      plt.show()
```



[33]:

```
NYC_V1 =  NYC_Crime_Table[NYC_Crime_Table['Borough'] == 'STATEN ISLAND']

NYC = NYC_V1[['Borough','Felony','Misdemeanor','Violation',
                 'Status','Total']]


NYC.set_index('Borough',inplace = True)

a = NYC.plot(kind='bar', figsize=(10, 6), rot=0)

a.set_ylabel('Number of Crimes')
a.set_xlabel('Borough')
a.set_title('NYC Boroughs with the least Crimes')



for p in a.patches:
    a.annotate(np.round(p.get_height(),decimals=2),
                (p.get_x()+p.get_width()/2., p.get_height()),
                ha='center',
                va='center',
                xytext=(0, 10),
                textcoords='offset points',
                fontsize = 11
               )
```

```
plt.show()
```


NYC Boroughs with the least Crimes
Number of Crimes
0
2500
5000
7500
10000
12500
15000
17500
20000
5059
10727
4217
20003
🟦 Felony
🟧 Misdemeanor
🟩 Violation
🟥 Total
STATEN ISLAND
Borough