COVID19Cases_DataDescription

April 5, 2021

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


COVIDdf = pd.read_csv('COVID19 cases.csv')
#Creating COVIDdf dataframe
#Reading COVID19 cases csv file and adding into dataframe
```

[2]:
```
COVIDdf.dtypes
#Identifying data types of each value in dataset
```

[2]:
```
_id                        int64
Assigned_ID                int64
Outbreak Associated       object
Age Group                 object
Neighbourhood Name        object
FSA                       object
Source of Infection       object
Classification            object
Episode Date              object
Reported Date             object
Client Gender             object
Outcome                   object
Currently Hospitalized    object
Currently in ICU          object
Currently Intubated       object
Ever Hospitalized         object
Ever in ICU               object
Ever Intubated            object
dtype: object
```

[3]:
```
COVIDdf.info()
#providing quantitatvie information of dataset and data type
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 77872 entries, 0 to 77871
Data columns (total 18 columns):
```

```
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   _id                     77872 non-null  int64
 1   Assigned_ID             77872 non-null  int64
 2   Outbreak Associated     77872 non-null  object
 3   Age Group               77837 non-null  object
 4   Neighbourhood Name      76433 non-null  object
 5   FSA                     76813 non-null  object
 6   Source of Infection     77872 non-null  object
 7   Classification          77872 non-null  object
 8   Episode Date            77872 non-null  object
 9   Reported Date           77872 non-null  object
 10  Client Gender           77872 non-null  object
 11  Outcome                 77872 non-null  object
 12  Currently Hospitalized  77872 non-null  object
 13  Currently in ICU        77872 non-null  object
 14  Currently Intubated     77872 non-null  object
 15  Ever Hospitalized       77872 non-null  object
 16  Ever in ICU             77872 non-null  object
 17  Ever Intubated          77872 non-null  object
dtypes: int64(2), object(16)
memory usage: 10.7+ MB
```

[4]:
```
COVIDdf = COVIDdf.drop(["Outbreak Associated", "FSA", "Source of Infection",␣
↪"Classification", "Outcome",
                        "Currently Hospitalized","Currently in ICU","Currently␣
↪Intubated", "Ever Hospitalized","Ever in ICU",
                        "Ever Intubated"], axis =1)

#Dropping unwanted columns
```

[5]:
```
print(COVIDdf.isnull().sum())

#Checking dataset for any null values
```

```
_id                     0
Assigned_ID             0
Age Group              35
Neighbourhood Name    1439
Episode Date             0
Reported Date            0
Client Gender            0
dtype: int64
```

[6]:
```
COVIDdf = COVIDdf.dropna()
COVIDdf.count()
```

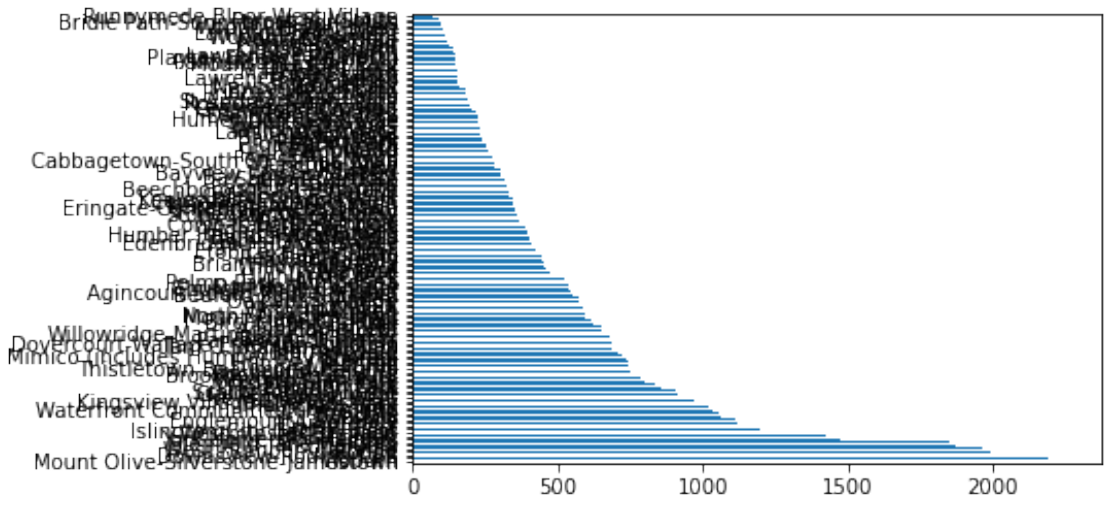
```
#Dropping null values to clean dataset
```

[6]:
```
_id                  76410
Assigned_ID          76410
Age Group            76410
Neighbourhood Name   76410
Episode Date         76410
Reported Date        76410
Client Gender        76410
dtype: int64
```

[7]:
```
print(COVIDdf["Neighbourhood Name"].value_counts())
COVIDdf["Neighbourhood Name"].value_counts().plot(kind = 'barh')

#finding value count of each neighbourhood
#plotting as bar graph
```

```
Mount Olive-Silverstone-Jamestown    2264
Woburn                               2189
Downsview-Roding-CFB                 1997
Rouge                                1995
West Humber-Clairville               1963
                                     ...
Blake-Jones                           105
Woodbine-Lumsden                       95
Bridle Path-Sunnybrook-York Mills      94
Forest Hill South                      92
Runnymede-Bloor West Village           66
Name: Neighbourhood Name, Length: 140, dtype: int64
```

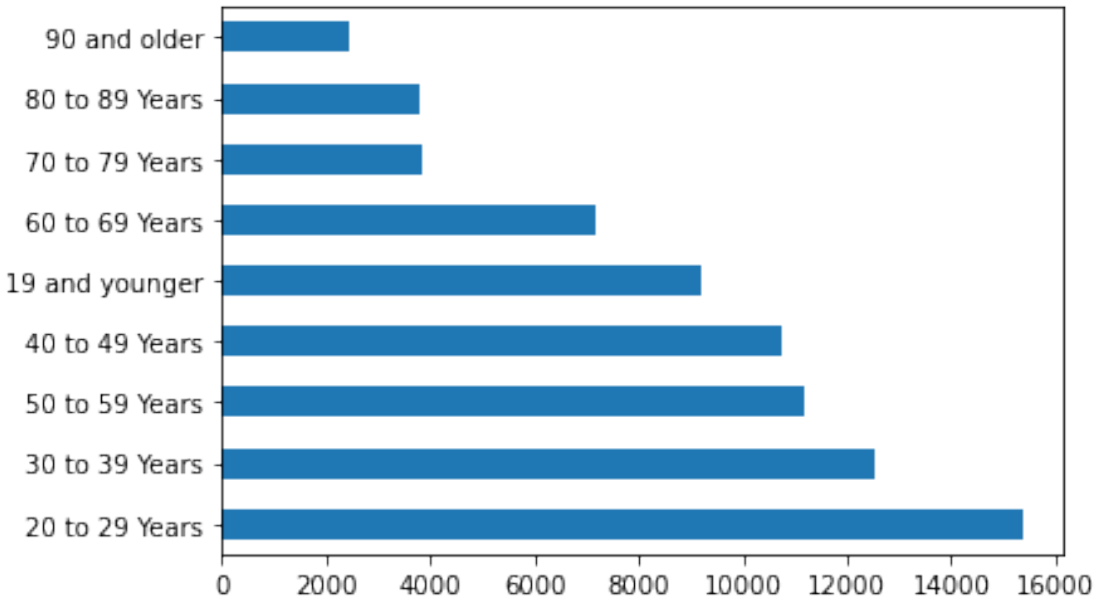
[7]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9229fe5290>
```



[8]:
```
print(COVIDdf["Age Group"].value_counts())
COVIDdf["Age Group"].value_counts().plot(kind = 'barh')

#finding value count of each age group
#plotting as bar graph
```

```
20 to 29 Years    15390
30 to 39 Years    12547
50 to 59 Years    11206
40 to 49 Years    10766
19 and younger     9221
60 to 69 Years     7193
70 to 79 Years     3846
80 to 89 Years     3774
90 and older       2467
Name: Age Group, dtype: int64
```

[8]: `<matplotlib.axes._subplots.AxesSubplot at 0x7f922a387910>`

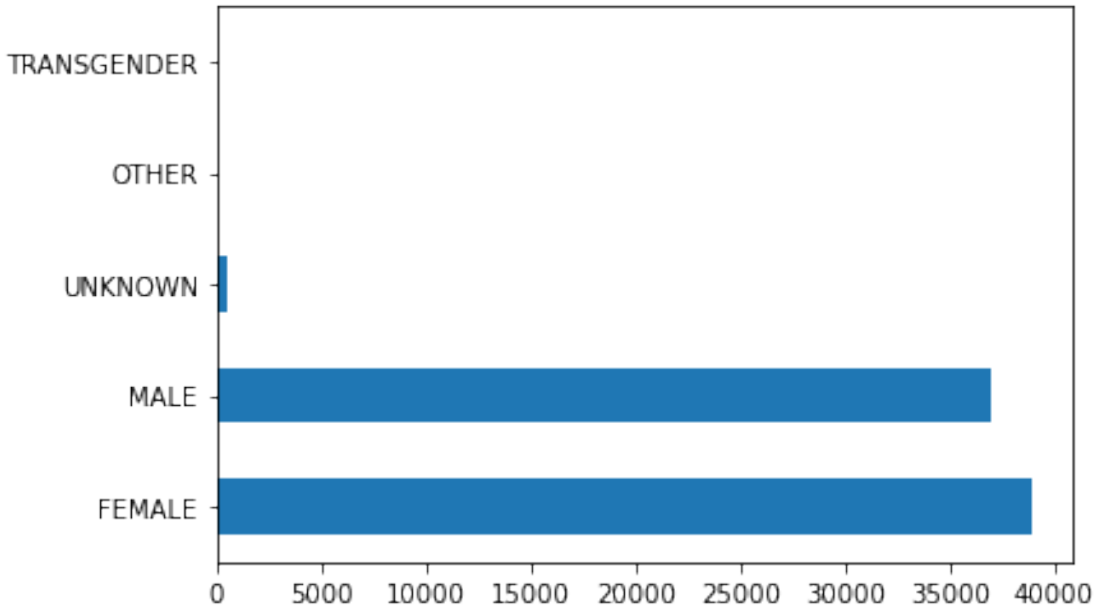


[9]:
```
print(COVIDdf["Client Gender"].value_counts())
COVIDdf["Client Gender"].value_counts().plot(kind = 'barh')

#finding value count of each gender group
#plotting as bar graph
```

```
FEMALE          38882
```

```
MALE           36990
UNKNOWN          513
OTHER             15
TRANSGENDER       10
Name: Client Gender, dtype: int64
```

[9]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9225f71d50>
```



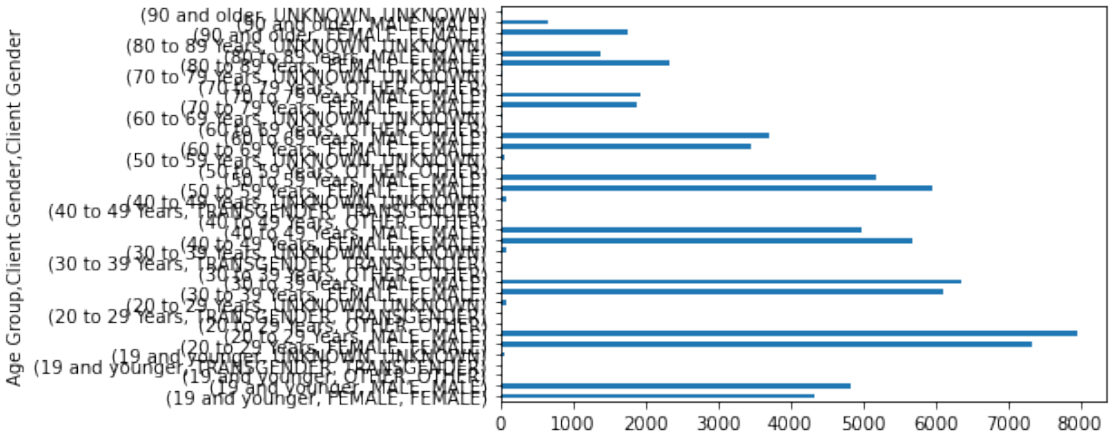
[10]:
```
agesexTable = COVIDdf.groupby(["Age Group", "Client Gender"])["Client Gender"].
↪value_counts()
agesexTable

#finding the gender group breakdown within each age
```

[10]:
```
Age Group       Client Gender  Client Gender
19 and younger  FEMALE         FEMALE          4329
                MALE           MALE            4834
                OTHER          OTHER              2
                TRANSGENDER    TRANSGENDER        1
                UNKNOWN        UNKNOWN           55
20 to 29 Years  FEMALE         FEMALE          7331
                MALE           MALE            7973
                OTHER          OTHER              4
                TRANSGENDER    TRANSGENDER        2
                UNKNOWN        UNKNOWN           80
30 to 39 Years  FEMALE         FEMALE          6105
                MALE           MALE            6350
```

```
                    OTHER          OTHER              3
                    TRANSGENDER    TRANSGENDER        4
                    UNKNOWN        UNKNOWN           85
40 to 49 Years      FEMALE         FEMALE          5682
                    MALE           MALE            4981
                    OTHER          OTHER              3
                    TRANSGENDER    TRANSGENDER        3
                    UNKNOWN        UNKNOWN           97
50 to 59 Years      FEMALE         FEMALE          5970
                    MALE           MALE            5181
                    OTHER          OTHER              1
                    UNKNOWN        UNKNOWN           54
60 to 69 Years      FEMALE         FEMALE          3457
                    MALE           MALE            3704
                    OTHER          OTHER              1
                    UNKNOWN        UNKNOWN           31
70 to 79 Years      FEMALE         FEMALE          1893
                    MALE           MALE            1928
                    OTHER          OTHER              1
                    UNKNOWN        UNKNOWN           24
80 to 89 Years      FEMALE         FEMALE          2347
                    MALE           MALE            1384
                    UNKNOWN        UNKNOWN           43
90 and older        FEMALE         FEMALE          1768
                    MALE           MALE             655
                    UNKNOWN        UNKNOWN           44
Name: Client Gender, dtype: int64
```

[11]:
```
agesexTable.plot(kind='barh')
```

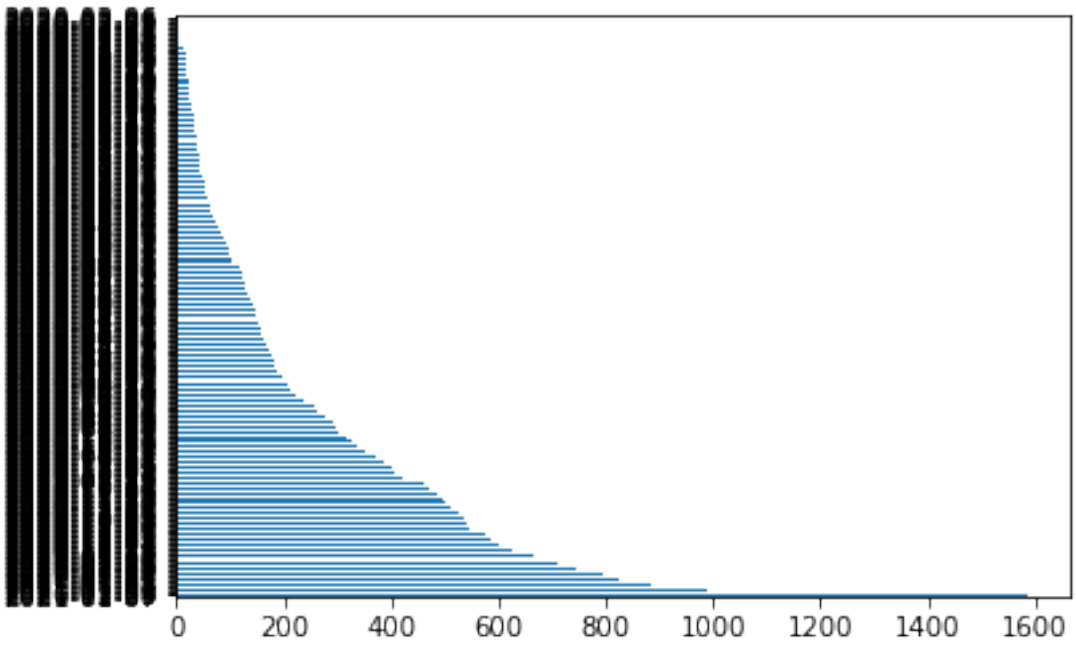
[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9225ed1250>



[12]:
```
print(COVIDdf["Reported Date"].value_counts())
COVIDdf["Reported Date"].value_counts().plot(kind = 'barh')

#finding the number of cases from the time of COVID-19 in Toronto
```

```
2021-01-07    1588
2021-01-08    1115
2020-12-29     994
2021-01-06     990
2021-01-12     978
              ...
2020-03-02       1
2020-03-03       1
2020-02-29       1
2020-02-28       1
2020-03-06       1
Name: Reported Date, Length: 331, dtype: int64
```

[12]: `<matplotlib.axes._subplots.AxesSubplot at 0x7f922a382fd0>`



[13]:
```
COVIDdf
```

[13]:

	_id	Assigned_ID	Age Group	Neighbourhood Name	\
0	873349	1	50 to 59 Years	Willowdale East	
1	873350	2	50 to 59 Years	Willowdale East	
2	873351	3	20 to 29 Years	Parkwoods-Donalda	

```
3      873352            4  60 to 69 Years  Church-Yonge Corridor
4      873353            5  60 to 69 Years  Church-Yonge Corridor
…       …           …              …                    …
77867  951216        79415  19 and younger    Stonegate-Queensway
77868  951217        79416  40 to 49 Years       Newtonbrook East
77869  951218        79417  50 to 59 Years      Banbury-Don Mills
77870  951219        79418  20 to 29 Years     Don Valley Village
77871  951220        79419  80 to 89 Years                Malvern

      Episode Date Reported Date Client Gender
0       2020-01-22    2020-01-23        FEMALE
1       2020-01-21    2020-01-23          MALE
2       2020-02-05    2020-02-21        FEMALE
3       2020-02-16    2020-02-25        FEMALE
4       2020-02-20    2020-02-26          MALE
…            …             …            …
77867   2021-01-15    2021-01-16          MALE
77868   2020-12-27    2021-01-08        FEMALE
77869   2021-01-15    2021-01-17        FEMALE
77870   2021-01-07    2021-01-12        FEMALE
77871   2021-01-07    2021-01-12          MALE

[76410 rows x 7 columns]
```

[16]:
```
len(COVIDdf['Neighbourhood Name'].unique())
```

[16]:
```
140
```

[23]:
```
COVIDdf['NeighbourhoodAvgIncome'] = 0
```

[22]:
```
for i in range(0, len(COVIDdf)):
    neighbourhoodName = COVIDdf.iloc[i]['Neighbourhood Name']

    if neighbourhoodName in FinalIncomeProfile.columns:
        Coviddf.iloc[i]['NeighbourhoodAvgIncome'] = FinalIncomeProfile.
↪loc['Sum'][neighbourhoodName]
```

[ ]:
```

```