miniuri.com/2hw

Carpeta compartida

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

In [104]:
```
import matplotlib.pyplot as plt
```

In [18]:
```
data_e = pd.read_csv("int_etnias.csv")
```

In [19]:
```
data_e
```

Out[19]:

	CLAVE DE ENTIDAD	ENTIDAD	CENTRO DE REINSERCIÓN SOCIAL	AMUZGO	CHATINO	CHINANTECO	CHOOL	CORA	CUCÁPAHS	CUICATECA	...	TAI
0	20	OAXACA	CENTRO DE REINSERCIÓN SOCIAL #4 "CUICATLÁN".	0	0	0	0	0	0	0	...	
1	20	OAXACA	CENTRO DE REINSERCIÓN SOCIAL #8 "MATÍAS ROMERO".	0	0	0	0	0	0	0	...	
2	20	OAXACA	CENTRO DE REINSERCIÓN SOCIAL #1 "PENITENCIARIA...	1	1	1	0	0	0	0	...	
3	20	OAXACA	CENTRO DE REINSERCIÓN SOCIAL #9 "JUCHITÁN".	0	0	0	0	0	0	0	...	
4	20	OAXACA	CENTRO DE REINSERCIÓN SOCIAL #6 "TUXTEPEC".	0	0	1	0	0	0	0	...	

pandas.read_csv

`pandas.`**`read_csv`**(*filepath_or_buffer*, *sep=','*, *delimiter=None*, *header='infer'*, *names=None*, *index_col=None*, *usecols=None*, *squeeze=False*, *prefix=None*, *mangle_dupe_cols=True*, *dtype=None*, *engine=None*, *converters=None*, *true_values=None*, *false_values=None*, *skipinitialspace=False*, *skiprows=None*, *nrows=None*, *na_values=None*, *keep_default_na=True*, *na_filter=True*, *verbose=False*, *skip_blank_lines=True*, *parse_dates=False*, *infer_datetime_format=False*, *keep_date_col=False*, *date_parser=None*, *dayfirst=False*, *iterator=False*, *chunksize=None*, *compression='infer'*, *thousands=None*, *decimal=b'.'*, *lineterminator=None*, *quotechar='"'*, *quoting=0*, *escapechar=None*, *comment=None*, *encoding=None*, *dialect=None*, *tupleize_cols=None*, *error_bad_lines=True*, *warn_bad_lines=True*, *skipfooter=0*, *doublequote=True*, *delim_whitespace=False*, *low_memory=True*, *memory_map=False*, *float_precision=None*) [source]

Read CSV (comma-separated) file into DataFrame

Also supports optionally iterating or breaking of the file into chunks.

Additional help can be found in the online docs for IO Tools.

filepath_or_buffer : *str, pathlib.Path, py._path.local.LocalPath or any* \

object with a read() method (such as a file handle or StringIO)

The string could be a URL. Valid URL schemes include http, ftp, s3, and file. For file URLs, a host is expected. For instance, a local file could be file://localhost/path/to/table.csv

sep : *str, default ','*

Delimiter to use. If sep is None, the C engine cannot automatically detect the separator, but the Python parsing engine can, meaning the latter will be used and automatically detect the separator by Python's builtin sniffer tool, `csv.Sniffer`. In addition, separators longer than 1 character and different from `'\s+'` will be interpreted as regular expressions and will also force the use of the Python parsing engine. Note that regex delimiters are prone to ignoring quoted data. Regex example: `'\r\t'`

delimiter : *str, default* `None`

Alternative argument name for sep.

delim_whitespace : *boolean, default False*

Specifies whether or not whitespace (e.g. `' '` or `'\t'`) will be used as the sep. Equi-

In [128]:
```
data_e.loc[[34,33,13,25]]
```

Out[128]:

	CLAVE DE ENTIDAD	ENTIDAD	CENTRO DE REINSERCIÓN SOCIAL	AMUZGO	CHATINO	CHINANTECO	CHOOL	CORA	CUCÁPAHS	CUICATECA	...	TARASCO	TE
34	9	DISTRITO FEDERAL	CENTRO FEMENIL DE READAPTACIÓN SOCIAL TEPEPAN.	0	0	0	0	0	0	0	...	0	
33	9	DISTRITO FEDERAL	RECLUSORIO FEMENIL SANTA MARTHA ACATITLA.	0	0	0	0	0	0	0	...	0	
13	7	CHIAPAS	CENTRO ESTATAL DE REINSERCIÓN SOCIAL PARA SENT...	0	0	0	9	0	0	0	...	0	
25	26	SONORA	CENTRO DE REINSERCIÓN SOCIAL "HERMOSILLO".	0	0	0	0	0	6	0	...	0	

In [130]:
```
data_e["CENTRO DE REINSERCIÓN SOCIAL"][:10]
```

Out[130]:
```
0         CENTRO DE REINSERCIÓN SOCIAL #4 "CUICATLÁN".
1     CENTRO DE REINSERCIÓN SOCIAL #8 "MATÍAS ROMERO".
2    CENTRO DE REINSERCIÓN SOCIAL #1 "PENITENCIARIA...
3          CENTRO DE REINSERCIÓN SOCIAL #9 "JUCHITÁN".
4          CENTRO DE REINSERCIÓN SOCIAL #6 "TUXTEPEC".
5          CENTRO DE REINSERCIÓN SOCIAL #5 "COSOLAPA".
6              CENTRO DE REINSERCIÓN SOCIAL #2 "ETLA".
7        CENTRO DE REINSERCIÓN SOCIAL #3 "MIAHUATLÁN".
8        CENTRO DE REINSERCIÓN SOCIAL #13 "HUAJUAPAN".
9      CENTRO DE REINSERCIÓN SOCIAL FEMENIL "TANIVET".
Name: CENTRO DE REINSERCIÓN SOCIAL, dtype: object
```

In [33]:
```
data_e[["CENTRO DE REINSERCIÓN SOCIAL"]][:10]
```

Out[33]:

	CENTRO DE REINSERCIÓN SOCIAL
0	CENTRO DE REINSERCIÓN SOCIAL #4 "CUICATLÁN".
1	CENTRO DE REINSERCIÓN SOCIAL #8 "MATÍAS ROMERO".
2	CENTRO DE REINSERCIÓN SOCIAL #1 "PENITENCIARIA...
3	CENTRO DE REINSERCIÓN SOCIAL #9 "JUCHITÁN".
4	CENTRO DE REINSERCIÓN SOCIAL #6 "TUXTEPEC".
5	CENTRO DE REINSERCIÓN SOCIAL #5 "COSOLAPA".
6	CENTRO DE REINSERCIÓN SOCIAL #2 "ETLA".
7	CENTRO DE REINSERCIÓN SOCIAL #3 "MIAHUATLÁN".
8	CENTRO DE REINSERCIÓN SOCIAL #13 "HUAJUAPAN".
9	CENTRO DE REINSERCIÓN SOCIAL FEMENIL "TANIVET".

In [131]:
```
cuentas=data_e["ENTIDAD"].value_counts()
```

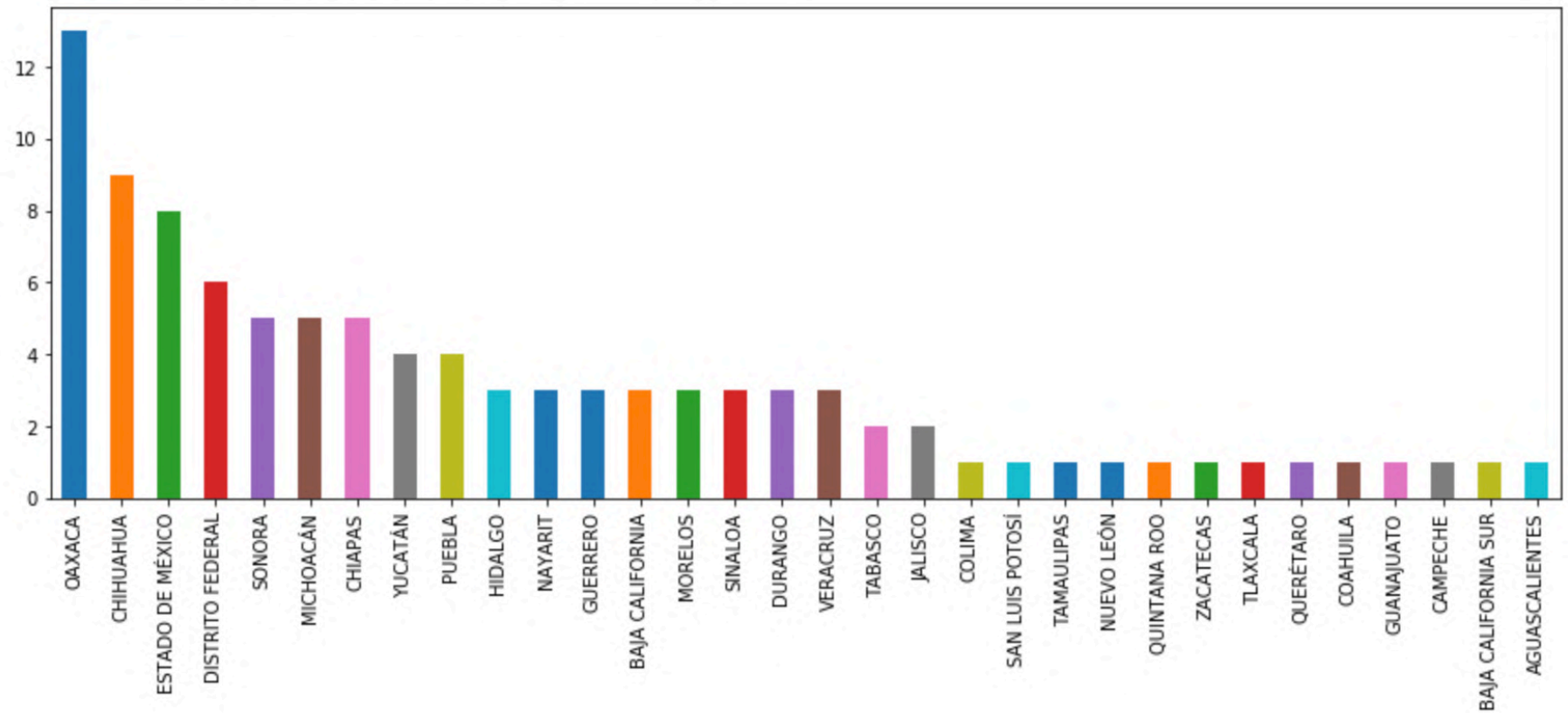
In [132]:
```
cuentas
```

Out[132]:

OAXACA	13
CHIHUAHUA	9
ESTADO DE MÉXICO	8
DISTRITO FEDERAL	6
SONORA	5
MICHOACÁN	5
CHIAPAS	5
YUCATÁN	4
PUEBLA	4
HIDALGO	3
NAYARIT	3
GUERRERO	3
BAJA CALIFORNIA	3
MORELOS	3
SINALOA	3
DURANGO	3
VERACRUZ	3
TABASCO	2
JALISCO	2
COLIMA	1
SAN LUIS POTOSÍ	1
TAMAULIPAS	1
NUEVO LEÓN	1
QUINTANA ROO	1
ZACATECAS	1
TLAXCALA	1
QUERÉTARO	1
COAHUILA	1

In [38]:
```
cuentas.plot(kind="bar")
```

Out[38]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x122466358>
```



In [42]:
```
datac = pd.read_csv("311-service-requests.csv")
```

```
/anaconda3/lib/python3.6/site-packages/IPython/core/interactiveshell.py:2785: DtypeWarning: Columns (8) have mixe
d types. Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

In [43]:
```
datac.info()
```

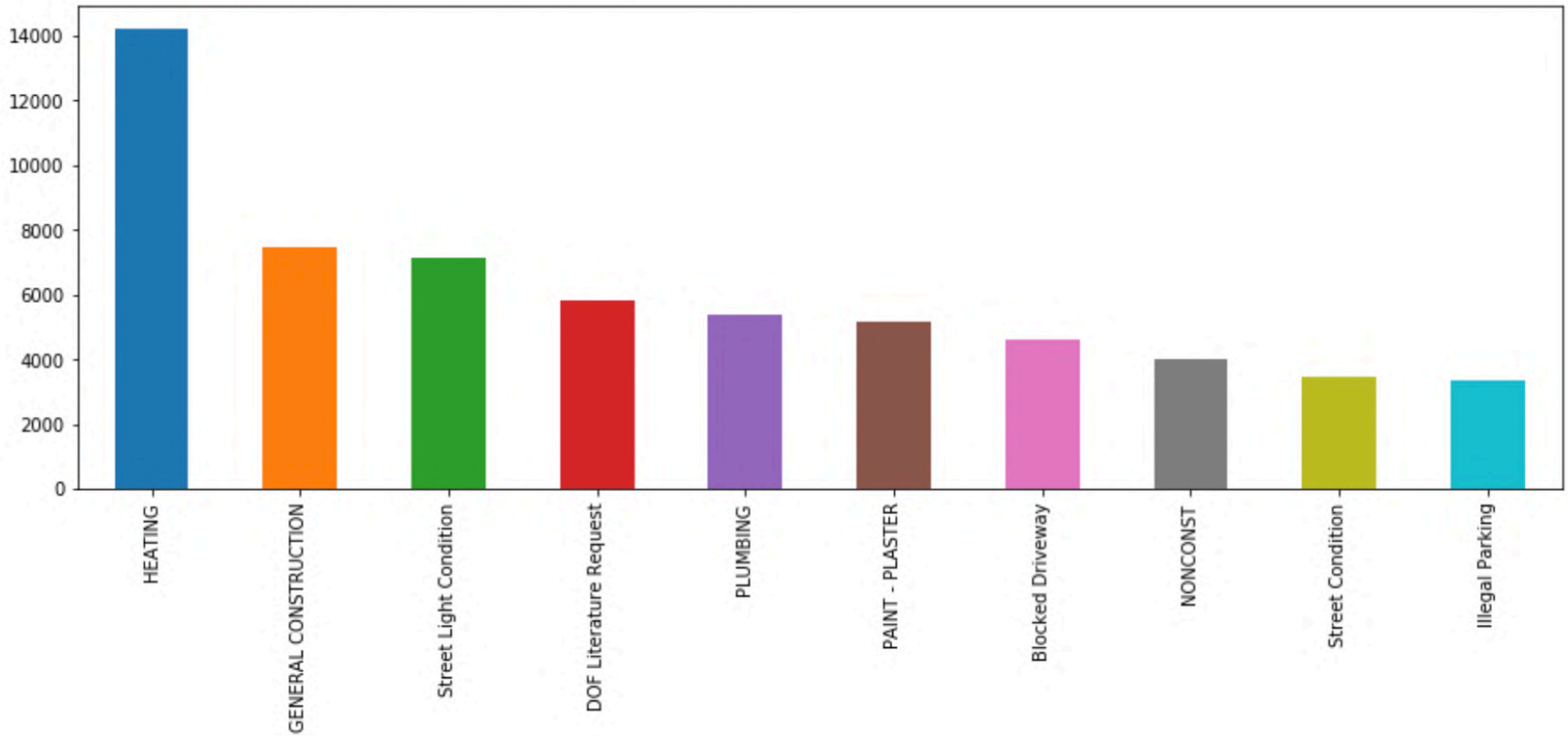
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 111069 entries, 0 to 111068
Data columns (total 52 columns):
Unique Key                        111069 non-null int64
Created Date                      111069 non-null object
Closed Date                       60270 non-null object
Agency                            111069 non-null object
Agency Name                       111069 non-null object
Complaint Type                    111069 non-null object
Descriptor                        110613 non-null object
Location Type                     79022 non-null object
Incident Zip                      98807 non-null object
Incident Address                  84441 non-null object
Street Name                       84432 non-null object
Cross Street 1                    84728 non-null object
Cross Street 2                    84005 non-null object
Intersection Street 1             19364 non-null object
Intersection Street 2             19366 non-null object
Address Type                      102247 non-null object
City                              98854 non-null object
Landmark                          95 non-null object
Facility Type                     19104 non-null object
Status                            111069 non-null object
Due Date                          39239 non-null object
Resolution Action Updated Date    96507 non-null object
Community Board                   111069 non-null object
Borough                           111069 non-null object
X Coordinate (State Plane)        98143 non-null float64
Y Coordinate (State Plane)        98143 non-null float64
```

In [45]:
```
datac["Complaint Type"].value_counts()
```

Out[45]:
```
HEATING                                14200
GENERAL CONSTRUCTION                    7471
Street Light Condition                  7117
DOF Literature Request                  5797
PLUMBING                                5373
PAINT - PLASTER                         5149
Blocked Driveway                        4590
NONCONST                                3998
Street Condition                        3473
Illegal Parking                         3343
Noise                                   3321
Traffic Signal Condition                3145
Dirty Conditions                        2653
Water System                            2636
Noise - Commercial                      2578
ELECTRIC                                2350
Broken Muni Meter                       2070
Noise - Street/Sidewalk                 1928
Sanitation Condition                    1824
Rodent                                  1632
Sewer                                   1627
Taxi Complaint                          1227
Consumer Complaint                      1227
Damaged Tree                            1180
Overgrown Tree/Branches                 1083
Graffiti                                 973
Missed Collection (All Materials)        973
Building/Use                             942
Root/Sewer/Sidewalk Condition            836
Derelict Vehicle                         803
                                        ...
```

In [52]:
```
datac["Complaint Type"].value_counts()[:10].plot(kind="bar")
```

Out[52]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1231bfe48>
```



In [137]:
```
datac["Complaint Type"][55:65]
```

Out[137]:
```
55      Noise - Commercial
56    Sanitation Condition
57                ELECTRIC
58                PLUMBING
59                 HEATING
60                ELECTRIC
61                 HEATING
62                 HEATING
63    GENERAL CONSTRUCTION
64                 HEATING
Name: Complaint Type, dtype: object
```

In [138]:
```
datab = datac["Complaint Type"] == "HEATING"
```

In [139]:
```
datab[55:65]
```

Out[139]:
```
55    False
56    False
57    False
58    False
59     True
60    False
61     True
62     True
63    False
64     True
Name: Complaint Type, dtype: bool
```

In [59]:
```
datab.value_counts()
```

Out[59]:
```
False    96869
True     14200
Name: Complaint Type, dtype: int64
```

In [142]:
```
data_heat = datac[datab][:5]
```

In [143]:
```
data_heat
```

Out[143]:

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Bridge Highway Name	E Hig Dire
59	26591688	10/31/2013 12:00:00 AM	NaN	HPD	Department of Housing Preservation and Develop...	HEATING	HEAT	RESIDENTIAL BUILDING	10453	150 WEST 179 STREET	...	NaN	
61	26593638	10/31/2013 12:00:00 AM	NaN	HPD	Department of Housing Preservation and Develop...	HEATING	HEAT	RESIDENTIAL BUILDING	10456	1175 MORRIS AVENUE	...	NaN	
62	26594900	10/31/2013 12:00:00 AM	NaN	HPD	Department of Housing Preservation and Develop...	HEATING	HEAT	RESIDENTIAL BUILDING	10469	917 MACE AVENUE	...	NaN	
64	26591211	10/31/2013 12:00:00 AM	NaN	HPD	Department of Housing Preservation and Develop...	HEATING	HEAT	RESIDENTIAL BUILDING	10032	505 WEST 161 STREET	...	NaN	
3323	26593331	10/30/2013 12:00:00 AM	NaN	HPD	Department of Housing Preservation and Develop...	HEATING	HEAT	RESIDENTIAL BUILDING	11206	33 MONTROSE AVENUE	...	NaN	

5 rows × 52 columns

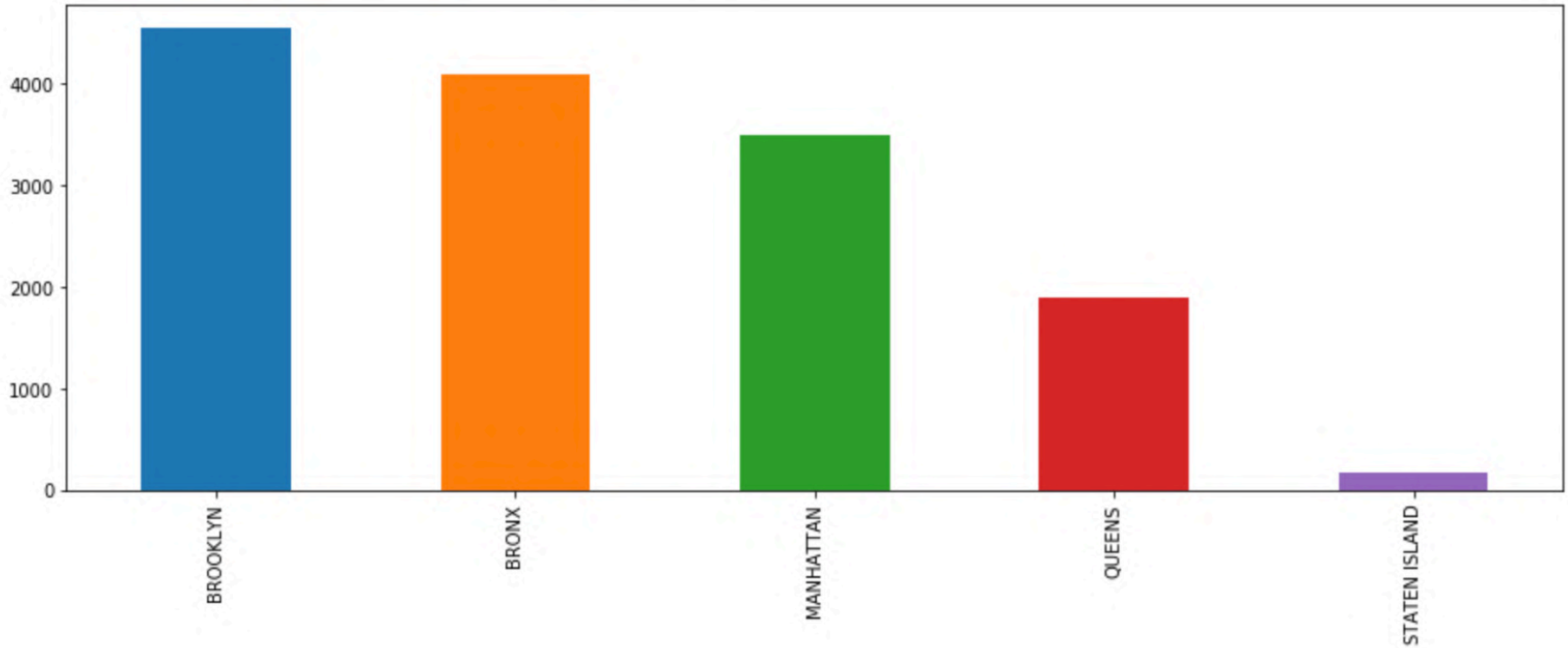
In [145]:
```
ruido=datac[datab]["Borough"].value_counts()
```

In [149]:
```
ruido
```

Out[149]:
```
BROOKLYN         4548
BRONX            4093
MANHATTAN        3494
QUEENS           1896
STATEN ISLAND     169
Name: Borough, dtype: int64
```

In [151]:
```
ruido.plot(kind="bar")
```

Out[151]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x12712ec88>
```



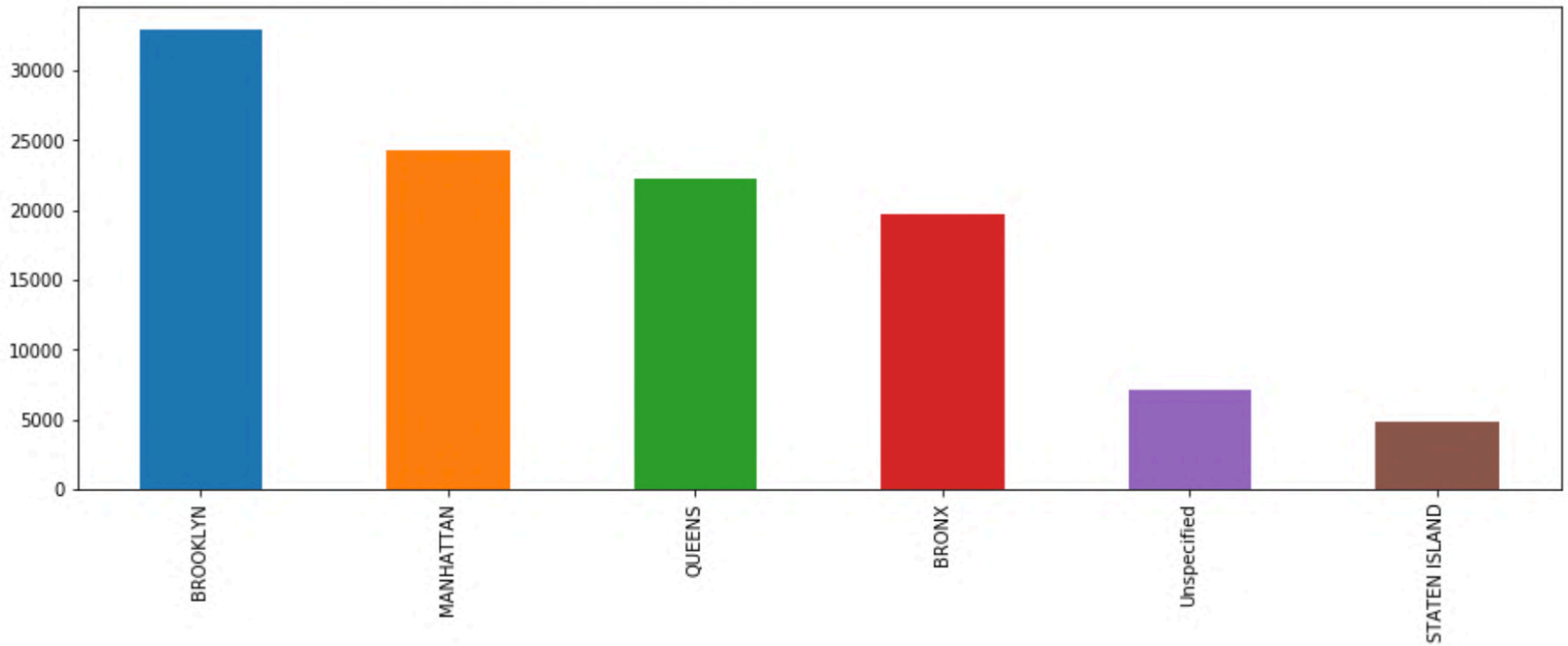
In [67]:
```
quejas=datac["Borough"].value_counts()
```

In [152]:
```
quejas
```

Out[152]:
```
BROOKLYN         32890
MANHATTAN        24288
QUEENS           22281
BRONX            19686
Unspecified       7107
STATEN ISLAND     4817
Name: Borough, dtype: int64
```

In [153]:
```
quejas.plot(kind="bar")
```

Out[153]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1271ef860>
```



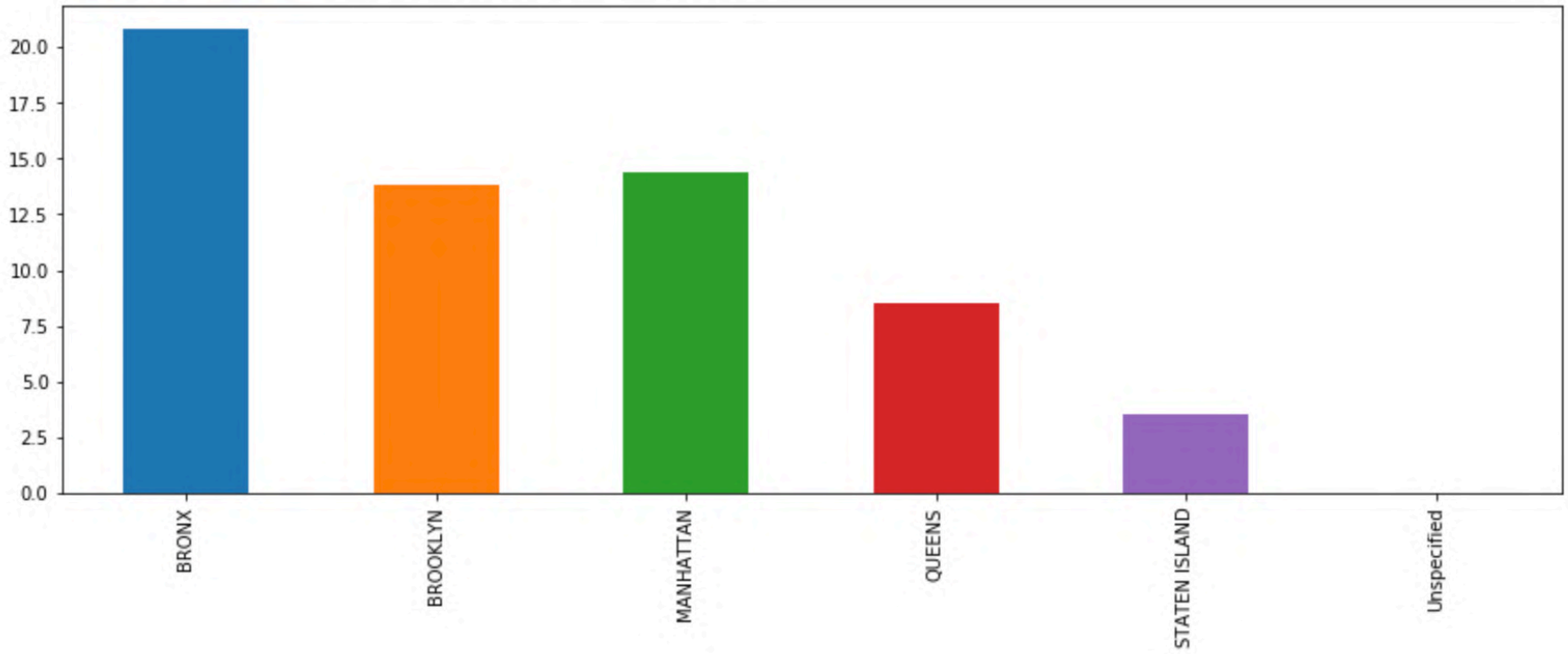
In [75]:
```
norm=(ruido / quejas)*100
```

In [76]:
```
norm
```

Out[76]:
```
BRONX            20.791425
BROOKLYN         13.827911
MANHATTAN        14.385705
QUEENS            8.509492
STATEN ISLAND     3.508408
Unspecified            NaN
Name: Borough, dtype: float64
```

In [147]:
```
norm.plot(kind="bar")
```

Out[147]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x125ebd8d0>
```



In [77]:
```
norm.plot(kind="bar")
```

Out[77]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x12378c0f0>
```