BucovIA Project

Artificial Intelligence Tool to Detect Bullying ans Stress due to Covid-19

Exploratory and Data Analysis

I. Load Libraries

First, install and import the libraries, functions and classess we will use.

In [1]:
```
pip install numpy
```

```
Requirement already satisfied: numpy in c:\users\jonb\anaconda3\envs\bucovia\lib\site-pa
ckages (1.20.3)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\jonb\anaconda3\envs\bucovia\lib\site-p
ackages (1.2.4)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\jonb\anaconda3\envs\bu
covia\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: numpy>=1.16.5 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from pandas) (1.20.3)
Requirement already satisfied: pytz>=2017.3 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from pandas) (2021.1)
Requirement already satisfied: six>=1.5 in c:\users\jonb\anaconda3\envs\bucovia\lib\site
-packages (from python-dateutil>=2.7.3->pandas) (1.15.0)
Note: you may need to restart the kernel to use updated packages.
```

In [3]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\users\jonb\anaconda3\envs\bucovia\lib\si
te-packages (3.4.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\jonb\anaconda3\envs\bucovia
\lib\site-packages (from matplotlib) (1.3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from matplotlib) (0.10.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\jonb\anaconda3\envs\buco
via\lib\site-packages (from matplotlib) (2.8.1)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\jonb\anaconda3\envs\bucovia
\lib\site-packages (from matplotlib) (2.4.7)
Requirement already satisfied: pillow>=6.2.0 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from matplotlib) (8.2.0)
Requirement already satisfied: numpy>=1.16 in c:\users\jonb\anaconda3\envs\bucovia\lib\s
ite-packages (from matplotlib) (1.20.3)
Requirement already satisfied: six in c:\users\jonb\anaconda3\envs\bucovia\lib\site-pack
ages (from cycler>=0.10->matplotlib) (1.15.0)
Note: you may need to restart the kernel to use updated packages.
```

In [4]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\users\jonb\anaconda3\envs\bucovia\lib\site-
packages (0.11.1)
Requirement already satisfied: pandas>=0.23 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from seaborn) (1.2.4)
Requirement already satisfied: matplotlib>=2.2 in c:\users\jonb\anaconda3\envs\bucovia\l
ib\site-packages (from seaborn) (3.4.2)
Requirement already satisfied: scipy>=1.0 in c:\users\jonb\anaconda3\envs\bucovia\lib\si
te-packages (from seaborn) (1.6.3)
Requirement already satisfied: numpy>=1.15 in c:\users\jonb\anaconda3\envs\bucovia\lib\s
ite-packages (from seaborn) (1.20.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\jonb\anaconda3\envs\buco
via\lib\site-packages (from matplotlib>=2.2->seaborn) (2.8.1)
Requirement already satisfied: cycler>=0.10 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from matplotlib>=2.2->seaborn) (0.10.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\jonb\anaconda3\envs\bucovia
\lib\site-packages (from matplotlib>=2.2->seaborn) (2.4.7)
Requirement already satisfied: pillow>=6.2.0 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from matplotlib>=2.2->seaborn) (8.2.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\jonb\anaconda3\envs\bucovia
\lib\site-packages (from matplotlib>=2.2->seaborn) (1.3.1)
Requirement already satisfied: six in c:\users\jonb\anaconda3\envs\bucovia\lib\site-pack
ages (from cycler>=0.10->matplotlib>=2.2->seaborn) (1.15.0)
Requirement already satisfied: pytz>=2017.3 in c:\users\jonb\anaconda3\envs\bucovia\lib
\site-packages (from pandas>=0.23->seaborn) (2021.1)
Note: you may need to restart the kernel to use updated packages.
```

In [5]:

```
pip install openpyxl
```

```
Requirement already satisfied: openpyxl in c:\users\jonb\anaconda3\envs\bucovia\lib\site
-packages (3.0.7)
Requirement already satisfied: et-xmlfile in c:\users\jonb\anaconda3\envs\bucovia\lib\si
te-packages (from openpyxl) (1.1.0)
Note: you may need to restart the kernel to use updated packages.
```

In [6]:

```
# NumPy for numerical computing
import numpy as np

# Pandas for DataFrames
import pandas as pd
pd.set_option('display.max.columns',100)

# Matplotlib for visualization
from matplotlib import pyplot as plt
%matplotlib inline

# Seaborn for easier visualization
import seaborn as sns
sns.set_style('darkgrid')
```

Next, read in the dataset.

- The file name is `'BucovIA_v2.xlsx'`

In [7]:

```
df = pd.read_excel('BucovIA_v2.xlsx')
```

II. EDA (Exploratory and Data Analysis) - Basic Information

First, let's look at the dimensions of the dataset.

In [8]:
```
df.shape
```

Out[8]:
```
(187, 107)
```

Next, let's take a look at the data types of our features.

In [9]:
```
df.dtypes
```

Out[9]:
```
Classroom          object
Age                object
Gender             object
Pain1              object
Pain2              object
                    ...
Self-analaysis      int64
Interactions        int64
Victim             object
Bully              object
Observer           object
Length: 107, dtype: object
```

Convert the object type features (questions) into numeric values

In [10]:
```
df.replace(['SÍ', 'Sí', 'ESTOY DE ACUERDO', 'DE ACUERDO', 'BAI', 'ADOS NAGO'], 1, inpla
df.replace(['NO', 'ESTOY EN DESACUERDO', 'EN DESACUERDO', 'EZ', 'EZ NAGO ADOS', 'EZ ADO
```

Some example observations from the dataset.

In [11]:
```
df.head()
```

Out[11]:

	Classroom	Age	Gender	Pain1	Pain2	Pain3	Pain4	Pain5	Pain6	Intimidation1	Intimidation2
0	2º ESO	14 años	Femenino	1	1	1	1	1	0	0	0
1	2º ESO	13 años	Femenino	1	1	1	1	1	0	1	1
2	2º ESO	14 años	Femenino	0	0	1	0	1	0	0	1
3	2º ESO	14 años	Femenino	0	0	0	0	0	0	0	0
4	2º ESO	14 años	Masculino	1	1	1	1	1	0	0	0

5 rows × 107 columns

Properties of the dataset.

In [12]:
```
df.describe()
```

Out[12]:

	Pain1	Pain2	Pain3	Pain4	Pain5	Pain6	Intimidation1	Intimidati
count	187.000000	187.000000	187.000000	187.000000	187.000000	187.000000	187.000000	187.00(
mean	0.438503	0.545455	0.593583	0.524064	0.513369	0.053476	0.320856	0.36:
std	0.497536	0.499266	0.492483	0.500761	0.501163	0.225585	0.468059	0.48:
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00(
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00(
50%	0.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.00(
75%	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000	1.00(
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.00(

8 rows × 104 columns

Distributions of the categorical features

In [13]:
```
df.describe(include=['object'])
```

Out[13]:

	Classroom	Age	Gender
count	187	187	187
unique	8	11	6
top	3º ESO	14 años	Masculino
freq	71	56	79

Let's plot a histogram of all the numerical features

In [14]:
```
# Plot histogram grid
df.hist(figsize=(20,20))

# Clear the text "residue"
plt.show()
```

III. Data Cleaning

There are a lot of features which can be gathered into a single group.

First, data cleaning is performed to manage the NaN gaps.

In [15]:
```
# First drop the possible duplicates
df = df.drop_duplicates()
print( df.shape )
```

```
(187, 107)
```

In [16]:
```
# Fill missing numerical (NaN) values as 0 as it has no influence at all.
for column in df.select_dtypes(include=['number']):
    df[column].fillna(0, inplace=True)
```

In [17]:
```
# Display number of missing values by feature (numerical)
```

```
df.select_dtypes(include=['number']).isnull().sum()
```

Out[17]:
```
Pain1             0
Pain2             0
Pain3             0
Pain4             0
Pain5             0
                 ..
Self-analaysis    0
Interactions      0
Victim            0
Bully             0
Observer          0
Length: 104, dtype: int64
```

In [18]:
```
# Drop NaN categorical values (drop the entire row)
#for column in df.select_dtypes(include=['object']):
df=df.dropna()
```

In [19]:
```
# Display number of missing values by feature (categorical)
df.select_dtypes(include=['object']).isnull().sum()
```

Out[19]:
```
Classroom    0
Age          0
Gender       0
dtype: int64
```

In [20]:
```
df.head()
```

Out[20]:

	Classroom	Age	Gender	Pain1	Pain2	Pain3	Pain4	Pain5	Pain6	Intimidation1	Intimidation2
0	2º ESO	14 años	Femenino	1	1	1	1	1	0	0	0
1	2º ESO	13 años	Femenino	1	1	1	1	1	0	1	1
2	2º ESO	14 años	Femenino	0	0	1	0	1	0	0	1
3	2º ESO	14 años	Femenino	0	0	0	0	0	0	0	0
4	2º ESO	14 años	Masculino	1	1	1	1	1	0	0	0

5 rows × 107 columns

In [21]:
```
# Check out the names of each feature
list(df.columns.values)
```

Out[21]:
```
['Classroom',
 'Age',
 'Gender',
 'Pain1',
```

```
'Pain2 ',
'Pain3 ',
'Pain4 ',
'Pain5 ',
'Pain6 ',
'Intimidation1',
'Intimidation2',
'Humiliation1',
'Humiliation2',
'Humiliation3',
'Humiliation4',
'Humiliation5',
'Ignore1',
'Ignore2',
'Ignore3',
'Ignore4',
'Ignore5',
'Ignore6',
'Digital1',
'Digital2',
'Digital3',
'Digital4',
'Digital5',
'Digital6',
'Digital7',
'Digital8',
'Digital9',
'Digital10',
'Digital11',
'Digital12',
'Digital13',
'OcurrToMe1',
'OcurrToMe2',
'OcurrToMe3',
'OcurrToMe4',
'OcurrToMe5',
'OcurrToMe6',
'OcurrToMe7',
'OcurrToMe8',
'OcurrToMe9',
'OcurrToMe10',
'Provoke1',
'Provoke2',
'Provoke3',
'Provoke4',
'Provoke5',
'Provoke6',
'Provoke7',
'Provoke8',
'PassiveObserve1',
'PassiveObserve2',
'PassiveObserve3',
'PassiveObserve4',
'PassiveObserve5',
'PassiveObserve6',
'PassiveObserve7',
'PassiveObserve8',
'PassiveObserve9',
'OcurrToMe11',
'OcurrToMe12',
'OcurrToMe13',
'OcurrToMe14',
'OcurrToMe15',
'OcurrToMe16',
'OcurrToMe17',
```

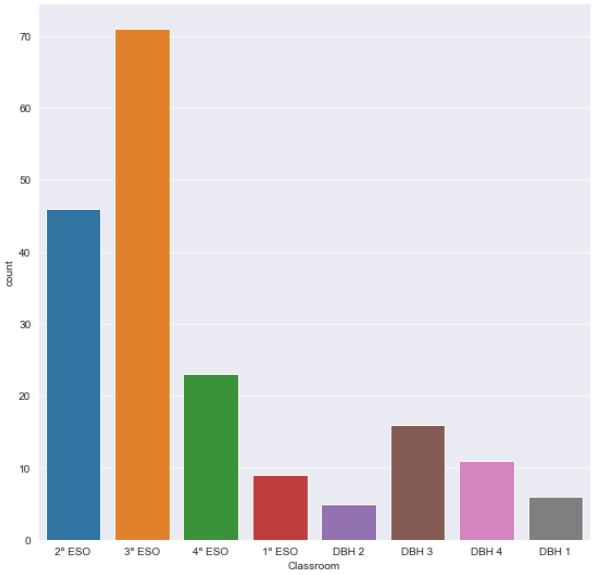
```
 'SituationGeneration1',
 'SituationGeneration2',
 'SituationGeneration3',
 'SituationGeneration4',
 'SituationGeneration5',
 'SituationGeneration6',
 'SituationGeneration7',
 'SituationGeneration8',
 'SituationGeneration9',
 'ActiveObserve1',
 'ActiveObserve2',
 'ActiveObserve3',
 'ActiveObserve4',
 'ActiveObserve5',
 'ActiveObserve6',
 'ActiveObserve7',
 'ActiveObserve8',
 'ActiveObserve9',
 'Behaviour1',
 'Behaviour2',
 'Behaviour3',
 'Behaviour4',
 'Behaviour5',
 'Behaviour6',
 'Feeling1',
 'Feeling2',
 'Feeling3',
 'Feeling4',
 'Thinking1',
 'Thinking2',
 'Thinking3',
 'Thinking4',
 'Thinking5',
 'Self-analaysis',
 'Interactions',
 'Victim',
 'Bully',
 'Observer']
```

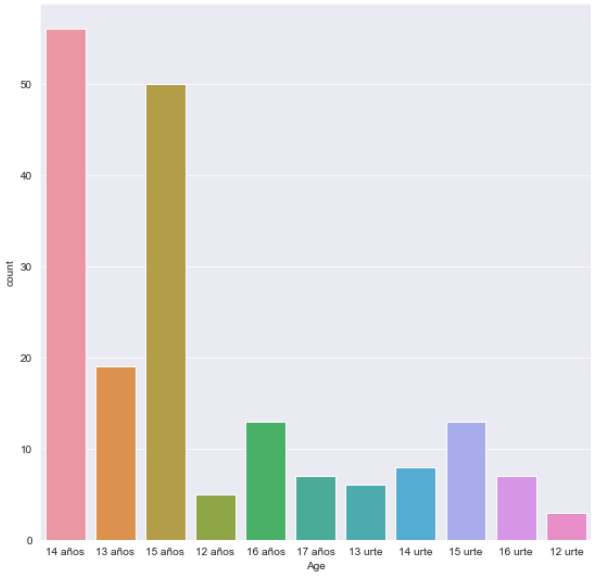
In [22]:

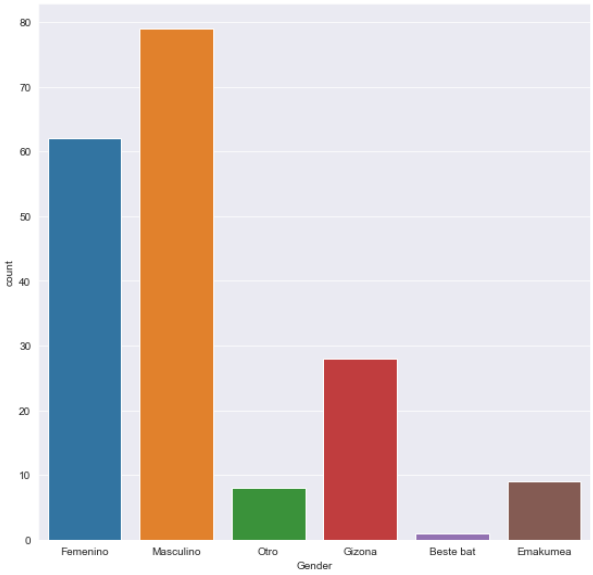
```python
# Plot the bar diagrams of the categorical features to find duplicates or similar categ
fig,ax =plt.subplots(2,2,figsize=(20,20))
sns.countplot(df['Classroom'], ax=ax[0,0])
sns.countplot(df['Age'], ax=ax[0,1])
sns.countplot(df['Gender'], ax=ax[1,0])
fig.show()
```

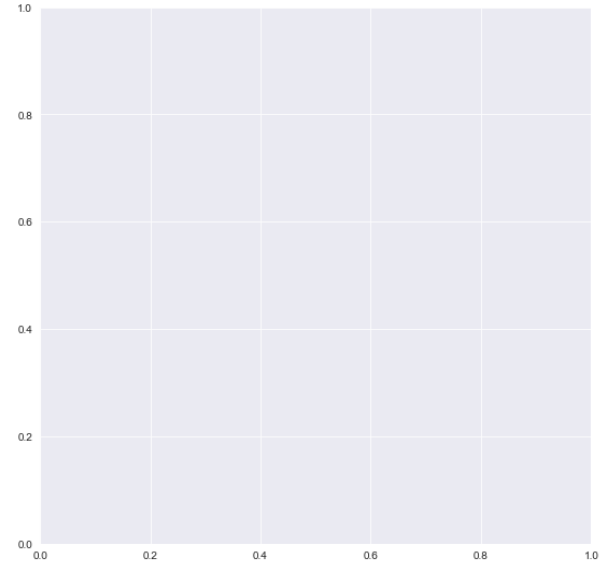
```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
<ipython-input-22-fc94709ab363>:6: UserWarning: Matplotlib is currently using module://i
```

```
pykernel.pylab.backend_inline, which is a non-GUI backend, so cannot show the figure.
  fig.show()
```









In [23]:

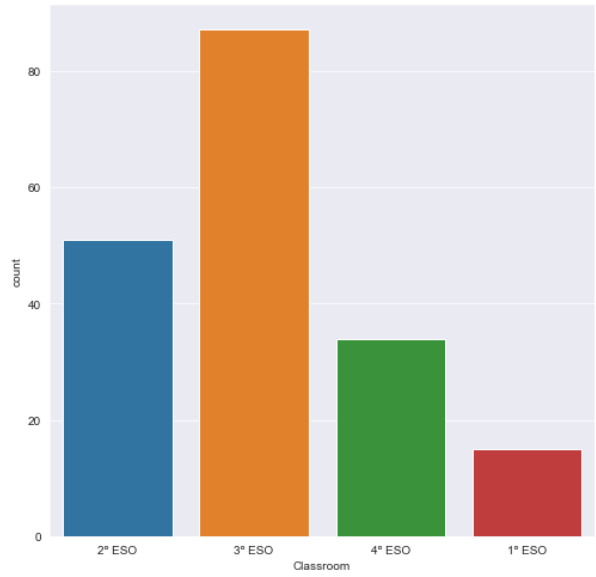
```
# Femeni0 and Emakumea should be 'Femenino'
df.Gender.replace(['Femeni0', 'Emakumea'], 'Femenino', inplace=True)

# Masculi0 and Gizona should be 'Masculino'
df.Gender.replace(['Masculi0', 'Gizona'], 'Masculino', inplace=True)

# Beste bat should be 'Otro'
df.Gender.replace('Beste bat', 'Otro', inplace=True)

# LANBIDE HEZIKETA 1go maila should be '1º FORMACION PROFESIONAL'
df.Classroom.replace(['LANBIDE HEZIKETA 1go maila','1º FORMACIÓN PROFESIONAL'], '1º FOR

# The same with DBH
df.Classroom.replace('DBH 1', '1º ESO',inplace=True)
df.Classroom.replace('DBH 2', '2º ESO',inplace=True)
df.Classroom.replace('DBH 3', '3º ESO',inplace=True)
df.Classroom.replace('DBH 4', '4º ESO',inplace=True)

# The same procedure for the "age" feature
```

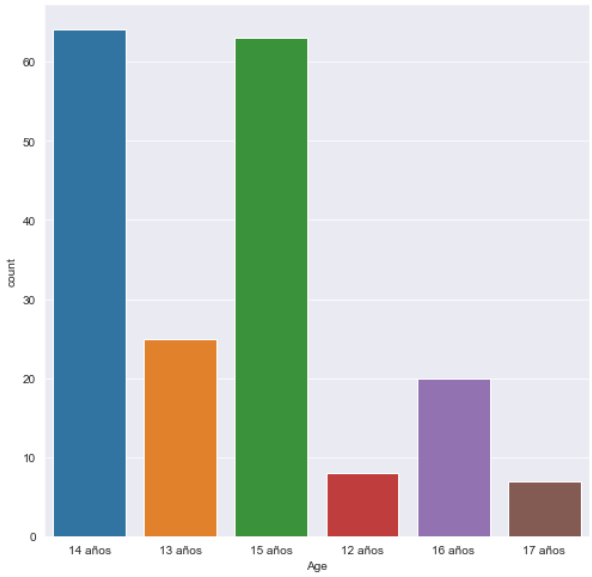
```
df.Age.replace('12 urte', '12 años', inplace=True)
df.Age.replace('13 urte', '13 años', inplace=True)
df.Age.replace('14 urte', '14 años', inplace=True)
df.Age.replace('15 urte', '15 años', inplace=True)
df.Age.replace('16 urte', '16 años', inplace=True)
df.Age.replace('17 urte', '17 años', inplace=True)
df.Age.replace('18 urte', '18 años', inplace=True)
df.Age.replace('19 urte', '19 años', inplace=True)
df.Age.replace('20 urte', '20 años', inplace=True)
df.Age.replace('21 urte', '21 años', inplace=True)
df.Age.replace('21 10 gehiago', 'más de 21', inplace=True)
```

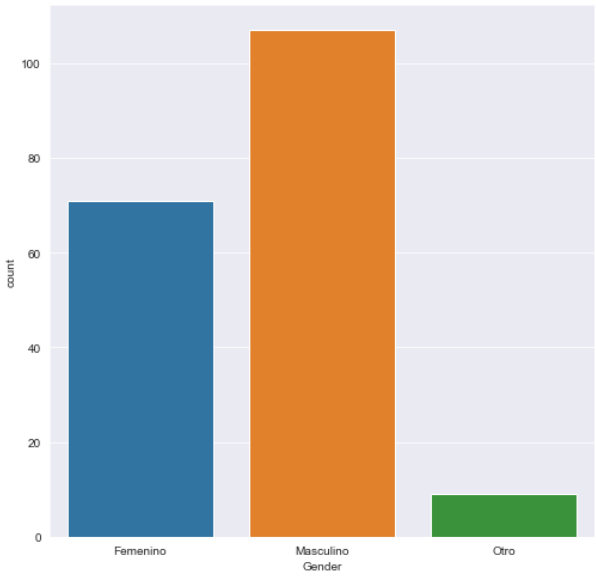
In [24]:

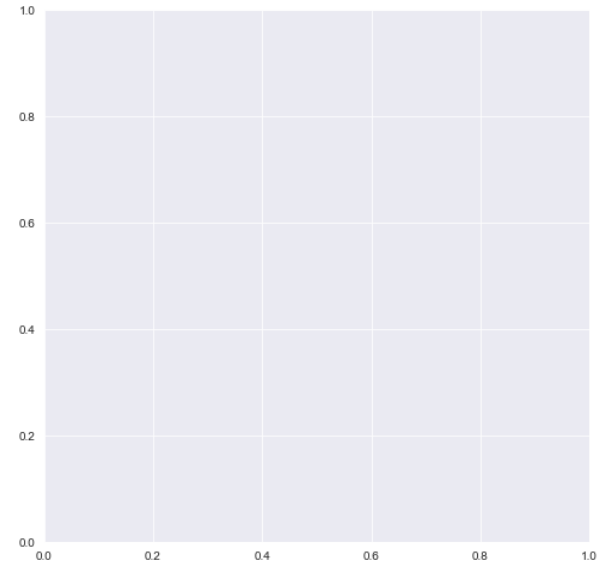
```
# Plot the bar diagrams of the gathered categorical features
fig,ax =plt.subplots(2,2,figsize=(18,18))
sns.countplot(df['Classroom'], ax=ax[0,0])
sns.countplot(df['Age'], ax=ax[0,1])
sns.countplot(df['Gender'], ax=ax[1,0])
fig.show()
```

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
<ipython-input-24-20eedbe973ea>:6: UserWarning: Matplotlib is currently using module://i
pykernel.pylab.backend_inline, which is a non-GUI backend, so cannot show the figure.
  fig.show()
```









IV. Feature Engineering

Next, group all the similar features into groups

In [25]:
```
# Columns with similar content/meaning are gathered up creating a new column (sum of th

df['Pain'] = df['Pain1'] + df['Pain2 '] + df['Pain3 '] + df['Pain4 '] + df['Pain5 '] +

df['Intimidation'] = df['Intimidation1'] + df['Intimidation2']

df['Humiliation'] = df['Humiliation1'] + df['Humiliation2'] + df['Humiliation3'] + df['

df['Ignore'] = df['Ignore1'] + df['Ignore2'] + df['Ignore3'] + df['Ignore4'] + df['Igno

df['Digital'] = df['Digital1'] + df['Digital2'] + df['Digital3'] + df['Digital4'] + df[

df['OcurrToMe'] = df['OcurrToMe1'] + df['OcurrToMe2'] + df['OcurrToMe3'] + df['OcurrToM

df['Provoke'] = df['Provoke1'] + df['Provoke2'] + df['Provoke3'] + df['Provoke4'] + df[
```

```
df['PassiveObserve'] = df['PassiveObserve1'] + df['PassiveObserve2'] + df['PassiveObser

df['SituationGeneration'] = df['SituationGeneration1'] + df['SituationGeneration2'] + d

df['ActiveObserve'] = df['ActiveObserve1'] + df['ActiveObserve2'] + df['ActiveObserve3'

df['Behaviour'] = df['Behaviour1'] + df['Behaviour2'] + df['Behaviour3'] + df['Behaviou

df['Feeling'] = df['Feeling1'] + df['Feeling2'] + df['Feeling3'] + df['Feeling4']

df['Thinking'] = df['Thinking1'] + df['Thinking2'] + df['Thinking3'] + df['Thinking4']

df.head()
```

Out[25]:

	Classroom	Age	Gender	Pain1	Pain2	Pain3	Pain4	Pain5	Pain6	Intimidation1	Intimidation2
0	2º ESO	14 años	Femenino	1	1	1	1	1	0	0	0
1	2º ESO	13 años	Femenino	1	1	1	1	1	0	1	1
2	2º ESO	14 años	Femenino	0	0	1	0	1	0	0	1
3	2º ESO	14 años	Femenino	0	0	0	0	0	0	0	0
4	2º ESO	14 años	Masculino	1	1	1	1	1	0	0	0

5 rows × 120 columns

Remove unnecessary columns

In [26]:

```
df.drop(df.columns.difference(['Classroom','Age','Gender','Self-analaysis','Interaction
                               'Humiliation','Ignore','Digital','OccurrToMe','Provoke',
                               'SituationGeneration','ActiveObserve','Behaviour','Feeli
        1, inplace=True)
df.head()
```

Out[26]:

	Classroom	Age	Gender	Self-analaysis	Interactions	Victim	Bully	Observer	Pain	Intimidation	Hu
0	2º ESO	14 años	Femenino	8	8	0	0	0	5	0	
1	2º ESO	13 años	Femenino	10	10	0	0	1	5	2	
2	2º ESO	14 años	Femenino	2	2	0	0	0	2	1	
3	2º ESO	14 años	Femenino	6	5	0	0	0	0	0	
4	2º ESO	14 años	Masculino	10	10	0	0	0	5	0	

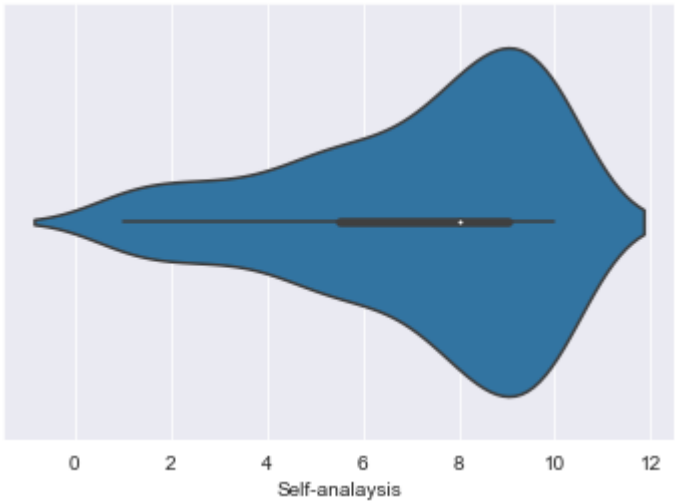
In [27]:
```
# Plot histogram grid with the new features
df.hist(figsize=(20,20))

# Clear the text "residue"
plt.show()
```

Searching for outliers

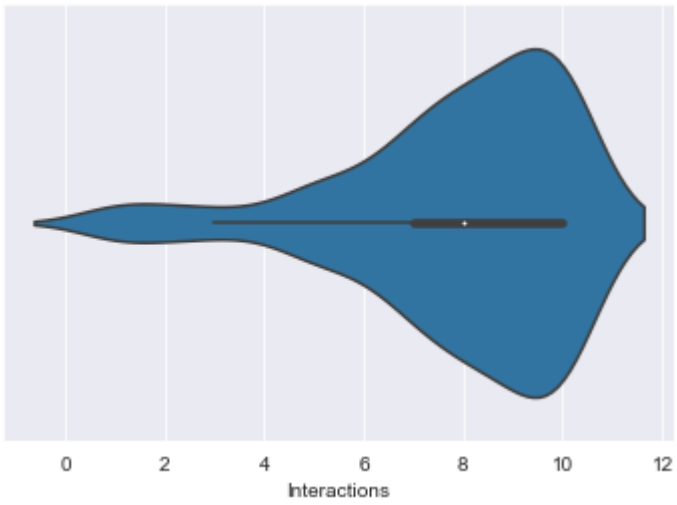
In [28]:
```
# Violin plot of numeric features

for column in df.select_dtypes(include=['number']):
    plt.figure()
    sns.violinplot(df[column])
    plt.show()
```

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
```

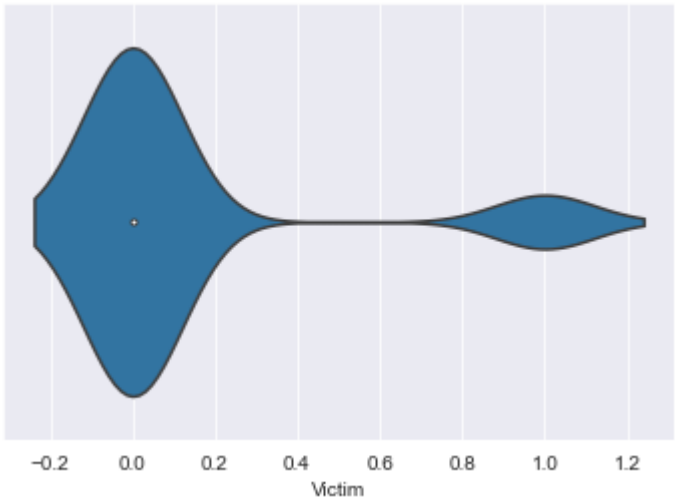
```
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



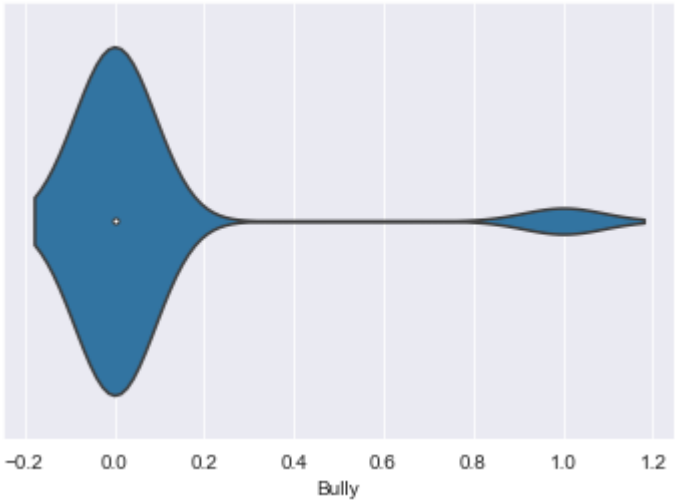

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



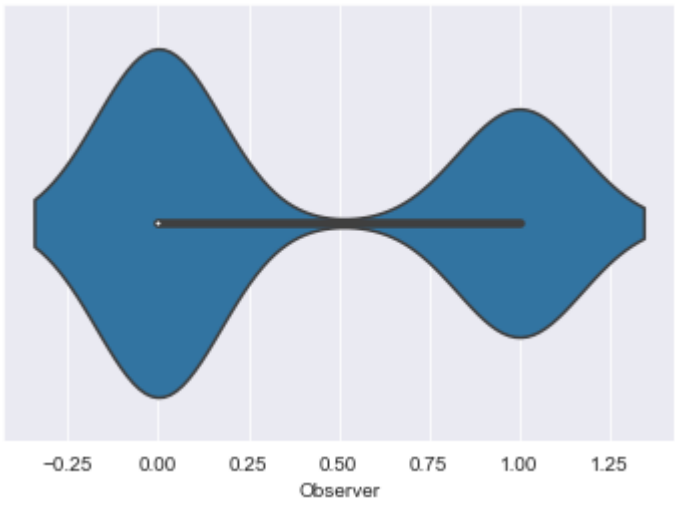

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```




```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

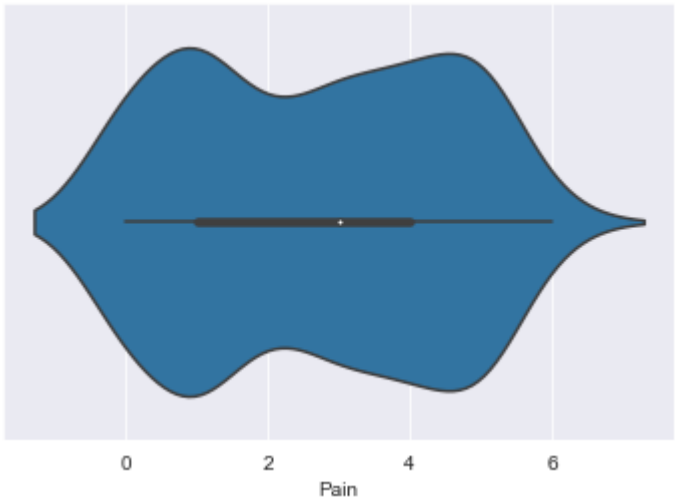



```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

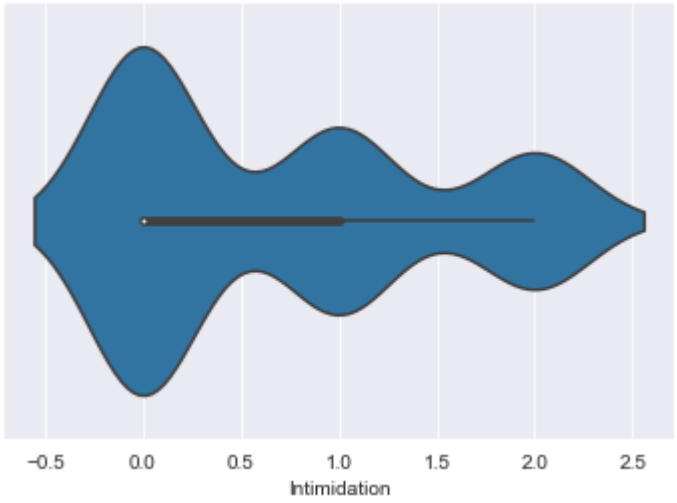



```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
```

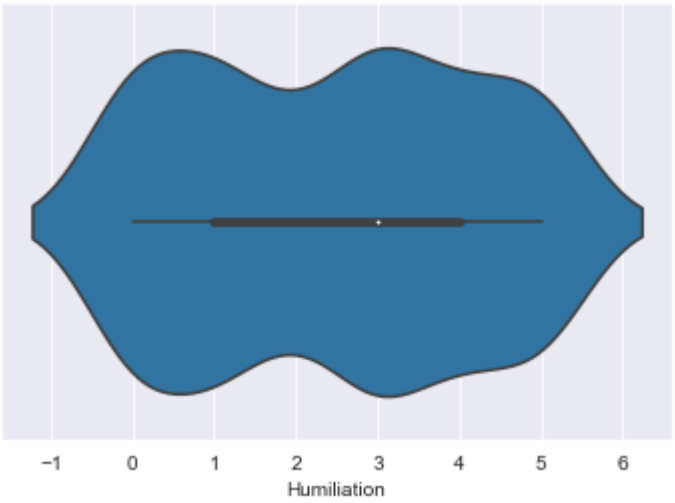
```
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



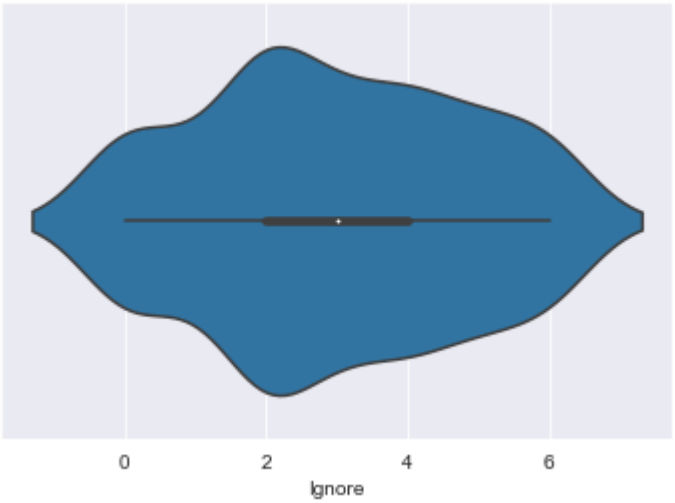

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



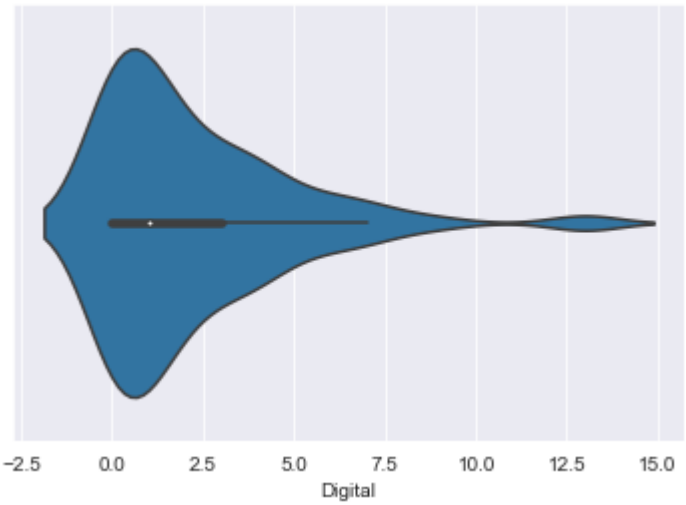

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```




```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

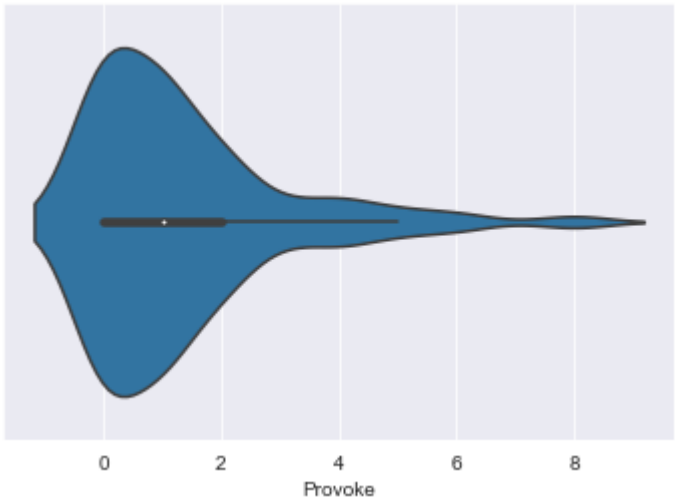



```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

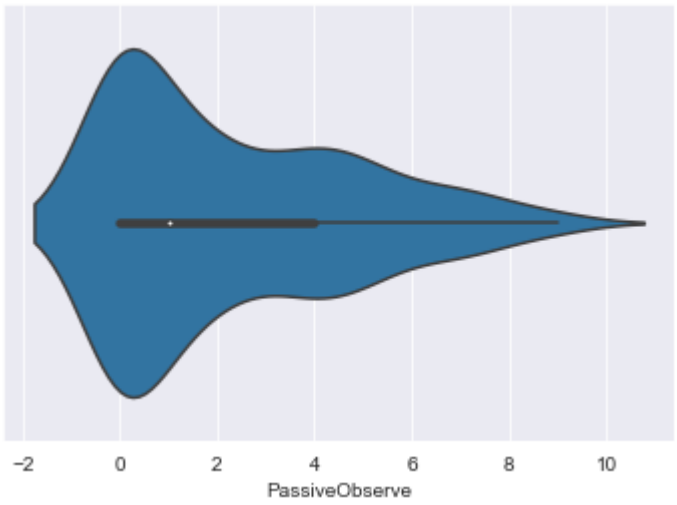



```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
```

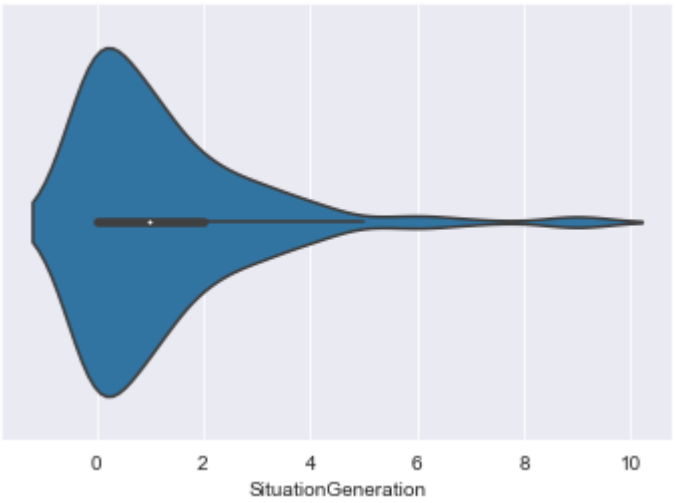
```
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



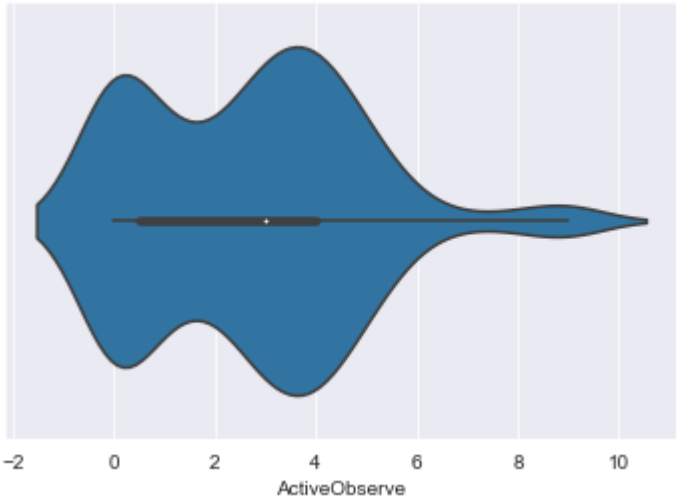

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



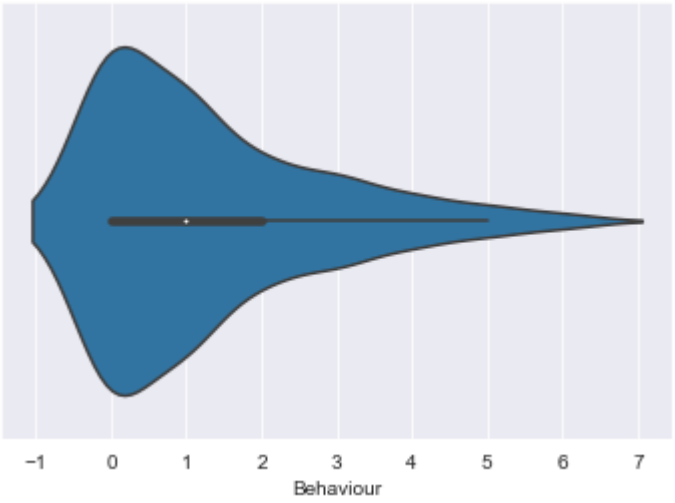

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```




```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

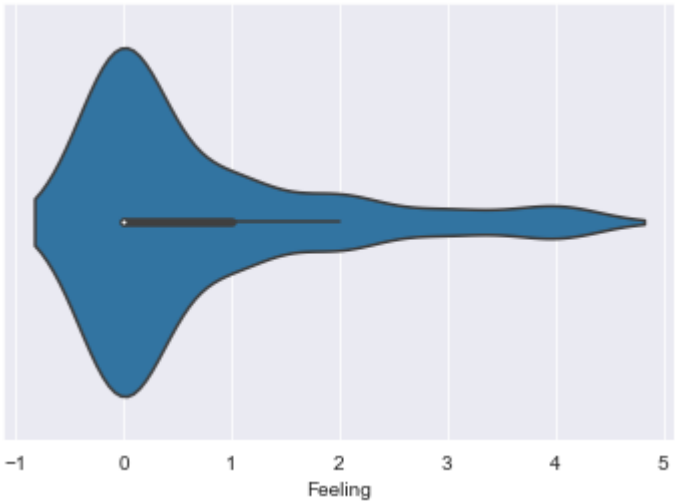



```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

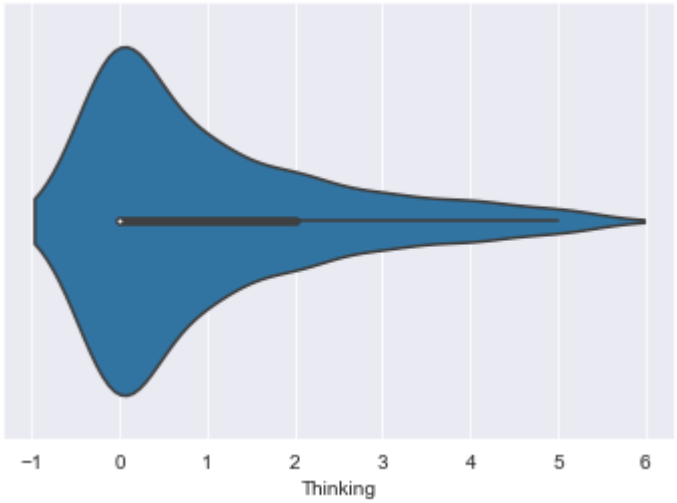



```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
```

```
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



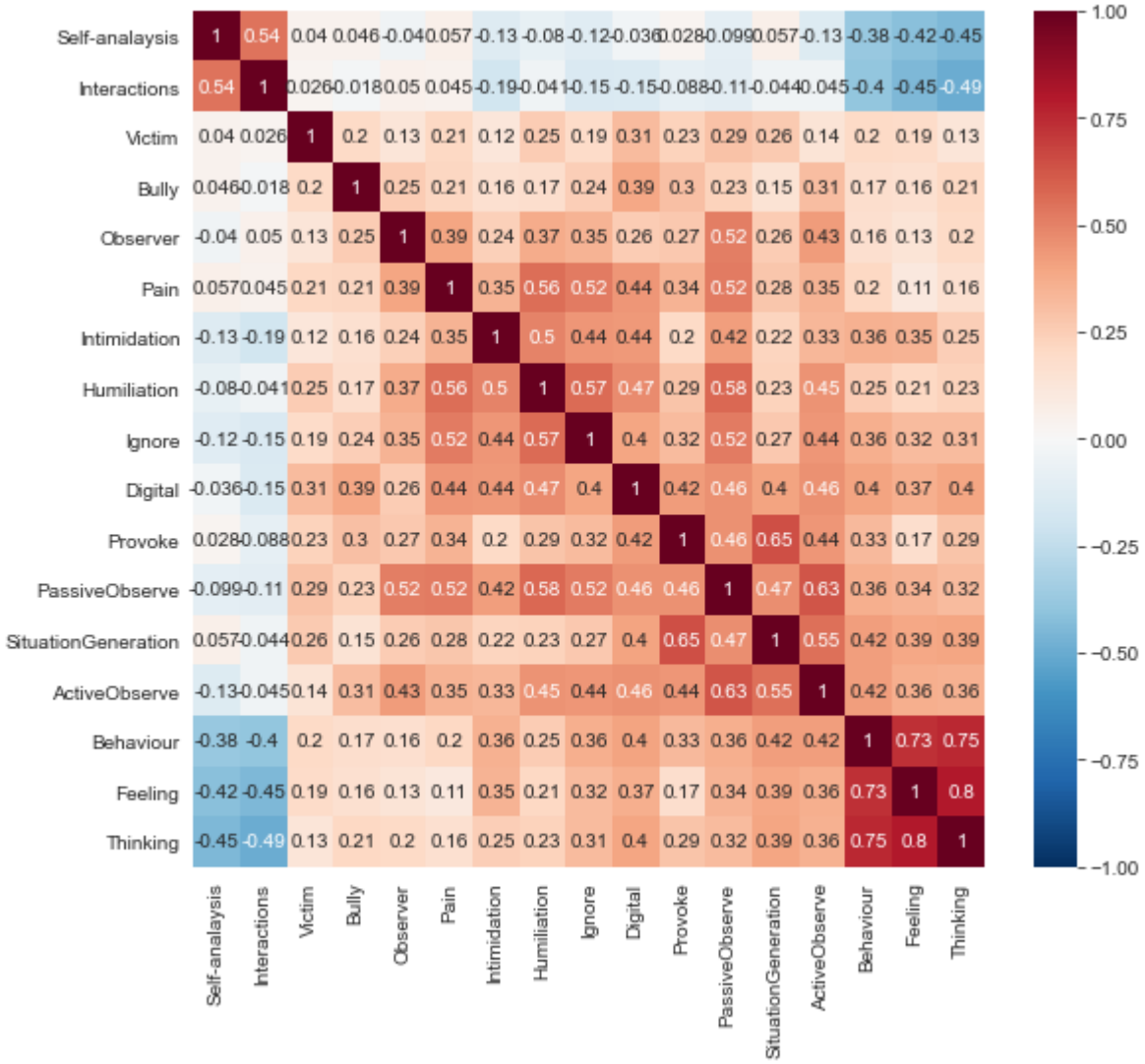

```
C:\Users\JonB\anaconda3\envs\BucovIA\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variable as a keyword arg: x. From version 0.12, the only va
lid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```



Show the correlations among the features

In [29]:

```
corr = df.corr()

plt.figure(figsize=(9,8))
sns.heatmap(corr,
            annot=True,
            cmap='RdBu_r',
            vmin=-1,
            vmax=1)
plt.show()
```



V. Save the Analytical Base Table (ABT)

Save ABT

In [30]:

```
# Save analytical base table
df.to_csv('analytical_base_table.csv', index=None)
```