


```
from google.colab import files
import pandas as pd

uploaded = files.upload()

df = pd.read_csv('EEG.machinelearning_data_BRMH.csv')

df.head()
```

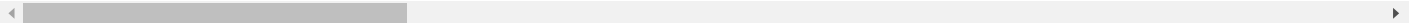
 Choose Files EEG.machi..._BRMH.csv

- **EEG.machinelearning_data_BRMH.csv**(text/csv) - 10537353 bytes, last modified: 8/1/2024 - 100% done

Saving EEG.machinelearning_data_BRMH.csv to EEG.machinelearning_data_BRMH.csv


no.	sex	age	eeg.date	education	IQ	main.disorder	specific.disorder	AB.A.delta.a.FP1	AB.A.delta.b.FP2	...	COH.F.gamma.o.Pz.p.P4	COH.F.gamm
0	1	M	57.0	2012.8.30	NaN	NaN	Addictive disorder	Alcohol use disorder	35.998557	21.717375	...	55.989192
1	2	M	37.0	2012.9.6	6.0	120.0	Addictive disorder	Alcohol use disorder	13.425118	11.002916	...	45.595619
2	3	M	32.0	2012.9.10	16.0	113.0	Addictive disorder	Alcohol use disorder	29.941780	27.544684	...	99.475453
3	4	M	35.0	2012.10.8	18.0	126.0	Addictive disorder	Alcohol use disorder	21.496226	21.846832	...	59.986561
4	5	M	36.0	2012.10.18	16.0	112.0	Addictive disorder	Alcohol use disorder	37.775667	33.607679	...	61.462720

5 rows × 1149 columns



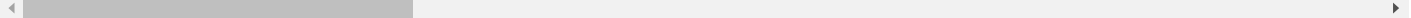
```
df_cleaned = df.dropna()

df_cleaned.head()
```



no.	sex	age	eeg.date	education	IQ	main.disorder	specific.disorder	AB.A.delta.a.FP1	AB.A.delta.b.FP2	...	COH.F.gamma.o.Pz.p.P4	COH.F.gamma.o.P
-----	-----	-----	----------	-----------	----	---------------	-------------------	------------------	------------------	-----	-----------------------	-----------------

0 rows × 1149 columns



```
def reformat_name(name):
    '''
    reformat from XX.X.band.x.channel to band.channel or
    COH.X.band.x.channel1.x.channel2 to COH.band.channel1.channel2
    '''
    splitted = name.split(sep='.')
    if len(splitted) < 5:
        return name
    if splitted[0] != 'COH':
        result = f'{splitted[2]}.{splitted[4]}'
    else:
        result = f'{splitted[0]}.{splitted[2]}.{splitted[4]}.{splitted[6]}'
    return result_type
df.rename(reformat_name, axis=1, inplace=True)
df
```

	no.	sex	age	eeg.date	education	IQ	main.disorder	specific.disorder	delta.FP1	delta.FP2	...	COH.gamma.Pz.P4	COH.gamma.Pz.T6	COH.gamma.P
0	1	M	57.0	2012.8.30	NaN	NaN	Addictive disorder	Alcohol use disorder	35.998557	21.717375	...	55.989192	16.739679	23.45
1	2	M	37.0	2012.9.6	6.0	120.0	Addictive disorder	Alcohol use disorder	13.425118	11.002916	...	45.595619	17.510824	26.77
2	3	M	32.0	2012.9.10	16.0	113.0	Addictive disorder	Alcohol use disorder	29.941780	27.544684	...	99.475453	70.654171	39.13
3	4	M	35.0	2012.10.8	18.0	126.0	Addictive disorder	Alcohol use disorder	21.496226	21.846832	...	59.986561	63.822201	36.47
4	5	M	36.0	2012.10.18	16.0	112.0	Addictive disorder	Alcohol use disorder	37.775667	33.607679	...	61.462720	59.166097	51.46
...
940	941	M	22.0	2014.8.28	13.0	116.0	Healthy control	Healthy control	41.851823	36.771496	...	82.905657	34.850706	63.97
941	942	M	26.0	2014.9.19	13.0	118.0	Healthy control	Healthy control	18.986856	19.401387	...	65.917918	66.700117	44.75
942	943	M	26.0	2014.9.27	16.0	113.0	Healthy control	Healthy control	28.781317	32.369230	...	61.040959	27.632209	45.55
943	944	M	24.0	2014.9.20	13.0	107.0	Healthy control	Healthy control	19.929100	25.196375	...	99.113664	48.328934	41.24
944	945	M	21.0	2015.10.23	13.0	105.0	Healthy control	Healthy control	65.195346	69.241972	...	78.600293	68.255430	70.68

945 rows × 1149 columns

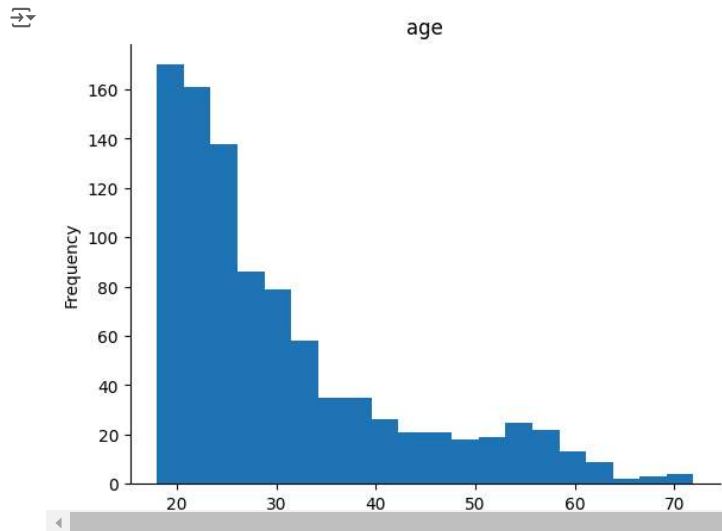
```
patients_info = df.loc[:, 'sex': 'specific.disorder']
patients_info
```

	sex	age	eeg.date	education	IQ	main.disorder	specific.disorder
0	M	57.0	2012.8.30	NaN	NaN	Addictive disorder	Alcohol use disorder
1	M	37.0	2012.9.6	6.0	120.0	Addictive disorder	Alcohol use disorder
2	M	32.0	2012.9.10	16.0	113.0	Addictive disorder	Alcohol use disorder
3	M	35.0	2012.10.8	18.0	126.0	Addictive disorder	Alcohol use disorder
4	M	36.0	2012.10.18	16.0	112.0	Addictive disorder	Alcohol use disorder
...
940	M	22.0	2014.8.28	13.0	116.0	Healthy control	Healthy control
941	M	26.0	2014.9.19	13.0	118.0	Healthy control	Healthy control
942	M	26.0	2014.9.27	16.0	113.0	Healthy control	Healthy control
943	M	24.0	2014.9.20	13.0	107.0	Healthy control	Healthy control
944	M	21.0	2015.10.23	13.0	105.0	Healthy control	Healthy control

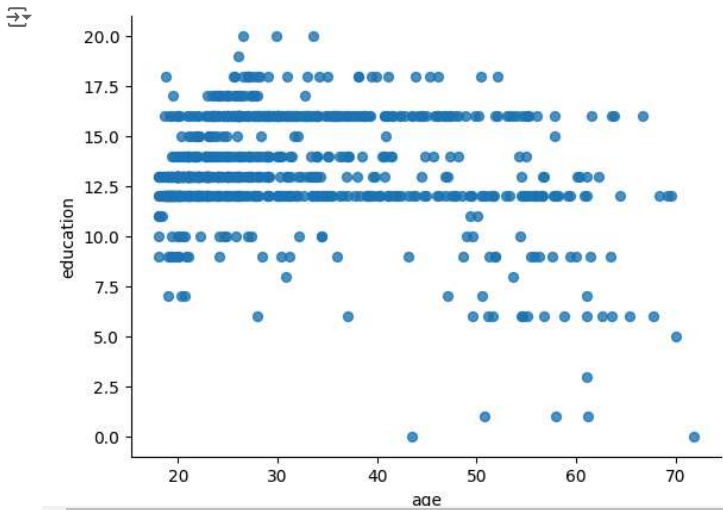
945 rows × 7 columns

Next steps: [Generate code with patients_info](#) [View recommended plots](#) [New interactive sheet](#)

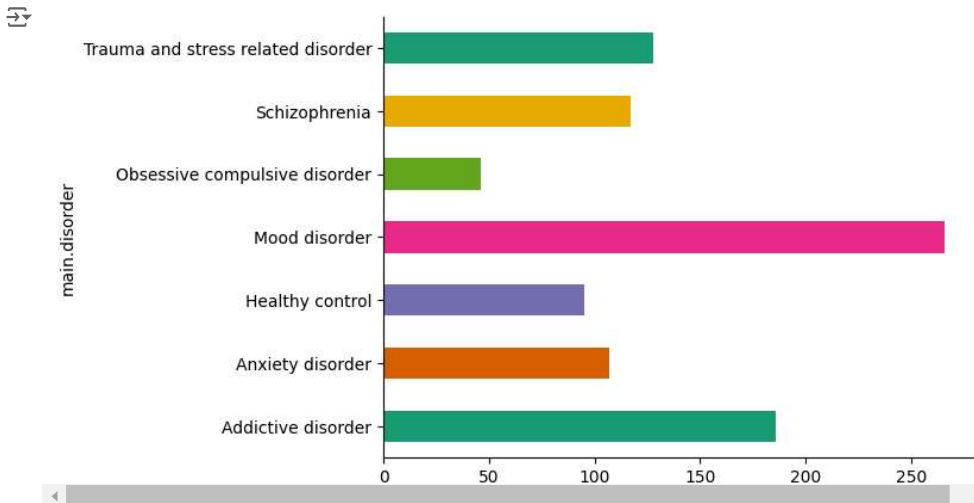
```
from matplotlib import pyplot as plt
patients_info['age'].plot(kind='hist', bins=20, title='age')
plt.gca().spines[['top', 'right']].set_visible(False)
```



```
from matplotlib import pyplot as plt
patients_info.plot(kind='scatter', x='age', y='education', s=32, alpha=.8)
plt.gca().spines[['top', 'right']].set_visible(False)
```



```
from matplotlib import pyplot as plt
import seaborn as sns
patients_info.groupby('main.disorder').size().plot(kind='barh', color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right']].set_visible(False)
```



```
patients_info = df.loc[:, 'sex': 'specific.disorder']
patients_info
```

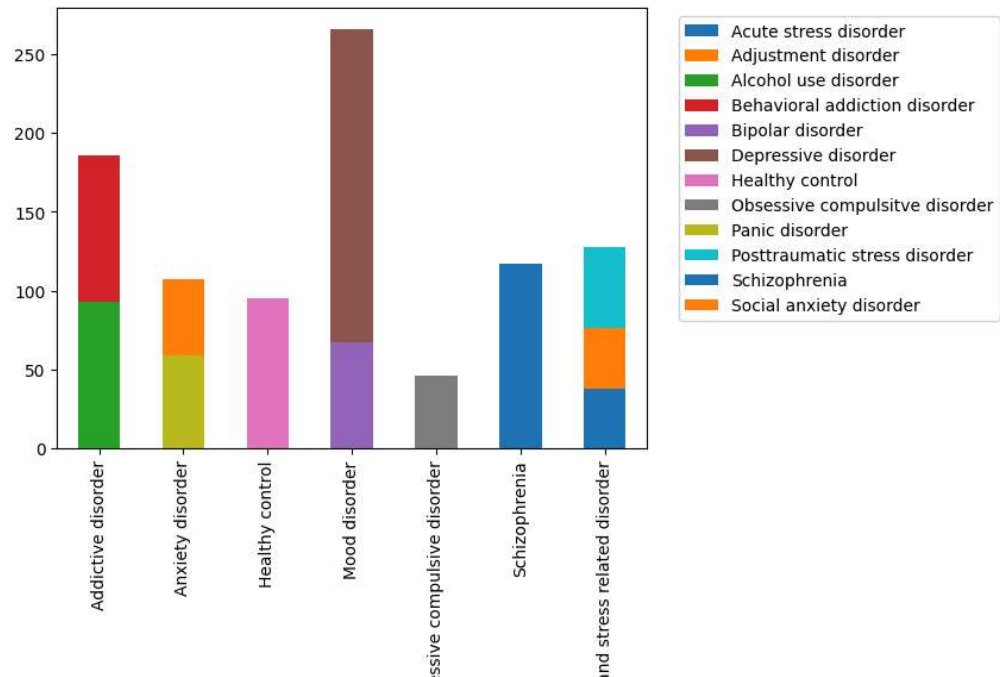
	sex	age	eeg.date	education	IQ	main.disorder	specific.disorder
0	M	57.0	2012.8.30	NaN	NaN	Addictive disorder	Alcohol use disorder
1	M	37.0	2012.9.6	6.0	120.0	Addictive disorder	Alcohol use disorder
2	M	32.0	2012.9.10	16.0	113.0	Addictive disorder	Alcohol use disorder
3	M	35.0	2012.10.8	18.0	126.0	Addictive disorder	Alcohol use disorder
4	M	36.0	2012.10.18	16.0	112.0	Addictive disorder	Alcohol use disorder
...
940	M	22.0	2014.8.28	13.0	116.0	Healthy control	Healthy control
941	M	26.0	2014.9.19	13.0	118.0	Healthy control	Healthy control
942	M	26.0	2014.9.27	16.0	113.0	Healthy control	Healthy control
943	M	24.0	2014.9.20	13.0	107.0	Healthy control	Healthy control
944	M	21.0	2015.10.23	13.0	105.0	Healthy control	Healthy control

945 rows x 7 columns

```
fig,ax = plt.subplots()
disorder_counts = df.groupby(['main.disorder', 'specific.disorder']).size()
display(disorder_counts.sort_values())
disorder_counts.unstack('specific.disorder')\
    .plot(kind='bar', stacked=True, ax=ax)
ax.legend(bbox_to_anchor=(1.04, 1), loc="upper left")
ax.xaxis.grid(False)
plt.show()
```

		0
main.disorder		specific.disorder
Trauma and stress related disorder	Acute stress disorder	38
	Adjustment disorder	38
Obsessive compulsive disorder	Obsessive compulsive disorder	46
Anxiety disorder	Social anxiety disorder	48
Trauma and stress related disorder	Posttraumatic stress disorder	52
Anxiety disorder	Panic disorder	59
Mood disorder	Bipolar disorder	67
Addictive disorder	Alcohol use disorder	93
	Behavioral addiction disorder	93
Healthy control	Healthy control	95
Schizophrenia	Schizophrenia	117
Mood disorder	Depressive disorder	199

dtype: int64



Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.