Importing essential python libraries and datafile.

```
import pandas as pd
import matplotlib.pylab as plt
import seaborn as sns
plt.style.use('ggplot')

df = pd.read_csv('Myers Briggs Table_S1.csv')

pd.set_option("display.max_columns", 50)

df
```

```
    S No  AGE  HEIGHT  WEIGHT     SEX ACTIVITY LEVEL  PAIN 1  PAIN 2
PAIN 3  \
0      1   53      62     125  Female            Low     0.0     0.0
0.0
1      2   52      69     157    Male           High     7.0     8.0
5.0
2      3   30      69     200    Male           High     0.0     0.0
0.0
3      4   51      66     175    Male       Moderate     9.5     9.5
9.5
4      5   45      63     199  Female       Moderate     4.0     5.0
2.0
..   ...  ...     ...     ...     ...            ...     ...     ...
...
92    93   16      58     100    Male       Moderate     0.0     0.0
0.0
93    94   45      62     134  Female       Moderate     0.0     4.0
0.0
94    95   43      69     188    Male       Moderate     2.0     0.0
0.0
95    96   28      67     180  Female            Low     0.0     0.0
0.0
96    97   43      69     188    Male       Moderate     4.0     0.0
0.0

    PAIN 4  MBTI   E   I   S   N   T   F   J   P POSTURE
0      0.0  ESFJ  18   3  17   9   9  13  18   4       A
1      3.0  ISTJ   6  15  14  12  21   3  13   9       B
2      0.0  ESTJ  15   6  16  10  15   9  12  10       A
3      1.5  ISTJ   6  15  21   5  13  11  19   3       D
4      2.0  ENFJ  14   7  20   6   9  15  16   6       A
..     ...   ...  ..  ..  ..  ..  ..  ..  ..  ..     ...
92     3.0  ESTP  19   2  22   4  19   5   2  20       B
93     0.0  ESFJ  11  10  17   9   6  18  13   9       B
94     0.0  ENFP  12   9   9  17   6  18   2  20       A
95     0.0  ESFJ  11  10  22  14   8  16  14   8       B
96     0.0  ENFP  12   9   9  17   6  18   2  20       A
```

```
[97 rows x 20 columns]
```

```
df.dtypes
```

```
S No                    int64
AGE                     int64
HEIGHT                  int64
WEIGHT                  int64
SEX                    object
ACTIVITY LEVEL         object
PAIN 1                float64
PAIN 2                float64
PAIN 3                float64
PAIN 4                float64
MBTI                   object
E                       int64
I                       int64
S                       int64
N                       int64
T                       int64
F                       int64
J                       int64
P                       int64
POSTURE                object
dtype: object
```

```
df.columns
```

```
Index(['S No', 'AGE', 'HEIGHT', 'WEIGHT', 'SEX', 'ACTIVITY LEVEL',
'PAIN 1',
       'PAIN 2', 'PAIN 3', 'PAIN 4', 'MBTI', 'E', 'I', 'S', 'N', 'T',
'F', 'J',
       'P', 'POSTURE'],
      dtype='object')
```

Making copy of the original dataframe in case something goes wrong.

```
df2 = df.copy()
```

Changing units from pounds to kilograms and inches to centimeters.

```
df2['HEIGHT'] = df2['HEIGHT'].astype(float)
df2['HEIGHT-CM'] = df['HEIGHT'] * 2.54
df2['HEIGHT-CM'] = df2['HEIGHT-CM'].round()

df2['WEIGHT-KGs'] = df2['WEIGHT'].astype(float)
df2['WEIGHT-KGs'] = df2['WEIGHT'] * 0.45
df2['WEIGHT-KGs'] = df2['WEIGHT-KGs'].round()

df2
```

```
    AGE     SEX  WEIGHT-KGs  HEIGHT-CM ACTIVITY LEVEL  PAIN 1  PAIN 2
PAIN 3  \
0    53  Female        56.0      157.0            Low     0.0     0.0
0.0
1    52    Male        71.0      175.0           High     7.0     8.0
5.0
2    30    Male        90.0      175.0           High     0.0     0.0
0.0
3    51    Male        79.0      168.0       Moderate     9.5     9.5
9.5
4    45  Female        90.0      160.0       Moderate     4.0     5.0
2.0
..  ...     ...         ...        ...            ...     ...     ...
...
92   16    Male        45.0      147.0       Moderate     0.0     0.0
0.0
93   45  Female        60.0      157.0       Moderate     0.0     4.0
0.0
94   43    Male        85.0      175.0       Moderate     2.0     0.0
0.0
95   28  Female        81.0      170.0            Low     0.0     0.0
0.0
96   43    Male        85.0      175.0       Moderate     4.0     0.0
0.0

    PAIN 4  MBTI   E   I   S   N   T   F   J   P POSTURE
0      0.0  ESFJ  18   3  17   9   9  13  18   4       A
1      3.0  ISTJ   6  15  14  12  21   3  13   9       B
2      0.0  ESTJ  15   6  16  10  15   9  12  10       A
3      1.5  ISTJ   6  15  21   5  13  11  19   3       D
4      2.0  ENFJ  14   7  20   6   9  15  16   6       A
..     ...   ...  ..  ..  ..  ..  ..  ..  ..  ..     ...
92     3.0  ESTP  19   2  22   4  19   5   2  20       B
93     0.0  ESFJ  11  10  17   9   6  18  13   9       B
94     0.0  ENFP  12   9   9  17   6  18   2  20       A
95     0.0  ESFJ  11  10  22  14   8  16  14   8       B
96     0.0  ENFP  12   9   9  17   6  18   2  20       A

[97 rows x 19 columns]
```

```
df2.drop(['WEIGHT', 'HEIGHT'], inplace=True, axis=1)
```

```
df2.columns
```

```
Index(['S No', 'AGE', 'SEX', 'ACTIVITY LEVEL', 'PAIN 1', 'PAIN 2',
'PAIN 3',
       'PAIN 4', 'MBTI', 'E', 'I', 'S', 'N', 'T', 'F', 'J', 'P',
'POSTURE',
       'WEIGHT-KGs', 'HEIGHT-CM'],
      dtype='object')
```

Re-ordering the columns

```
df2 = df2.reindex(columns=['S No', 'AGE', 'SEX', 'WEIGHT-KGs', 
'HEIGHT-CM','ACTIVITY LEVEL', 'PAIN 1', 'PAIN 2', 'PAIN 3', 
       'PAIN 4', 'MBTI', 'E', 'I', 'S', 'N', 'T', 'F', 'J', 'P', 
'POSTURE',
       ])

df2.drop(['S No'], inplace=True, axis=1)
```

Looking for missing values.

```
df2.isna().sum()

Age                 0
Sex                 0
Weight kg           0
Height cm           0
Activity level      0
Pain 1              0
Pain 2              0
Pain 3              0
Pain 4              0
MBTI                0
E                   0
I                   0
S                   0
N                   0
T                   0
F                   0
J                   0
P                   0
Posture             0
dtype: int64
```

Looking for duplicates.

```
df2.loc[df2.duplicated()]

    Age     Sex  Weight kg  Height cm Activity level  Pain 1  Pain 2  
Pain 3  \
95   28  Female       81.0      170.0            Low     0.0     0.0   
0.0   

    Pain 4  MBTI   E   I   S   N  T   F   J  P Posture  
95     0.0  ESFJ  11  10  22  14  8  16  14  8       B  

# index 87 and 95 are duplicates
df2.tail(10)
```

```
    Age     Sex  Weight kg  Height cm Activity level  Pain 1  Pain 2
Pain 3  \
87   28  Female       81.0      170.0            Low     0.0     0.0
0.0
88   22  Female       87.0      165.0            Low     5.0     7.0
7.0
89   56  Female       68.0      170.0            Low     0.0     7.0
0.0
90   29  Female       56.0      165.0       Moderate     2.0     0.0
0.0
91   16  Female       58.0      175.0       Moderate     5.0     0.0
5.0
92   16    Male       45.0      147.0       Moderate     0.0     0.0
0.0
93   45  Female       60.0      157.0       Moderate     0.0     4.0
0.0
94   43    Male       85.0      175.0       Moderate     2.0     0.0
0.0
95   28  Female       81.0      170.0            Low     0.0     0.0
0.0
96   43    Male       85.0      175.0       Moderate     4.0     0.0
0.0

    Pain 4  MBTI   E   I   S   N   T   F   J   P Posture
87     0.0  ESFJ  11  10  22  14   8  16  14   8       B
88     0.0  ESFJ  17   4  14  12   7  17  15   7       B
89     0.0  ISFP   9  12  15  11   4  20   5  17       C
90     4.0  ENFP  19   2  13  13  12  12  10  12       A
91     7.0  ENFJ  19   2   9  17   2  22  12  10       B
92     3.0  ESTP  19   2  22   4  19   5   2  20       B
93     0.0  ESFJ  11  10  17   9   6  18  13   9       B
94     0.0  ENFP  12   9   9  17   6  18   2  20       A
95     0.0  ESFJ  11  10  22  14   8  16  14   8       B
96     0.0  ENFP  12   9   9  17   6  18   2  20       A
```

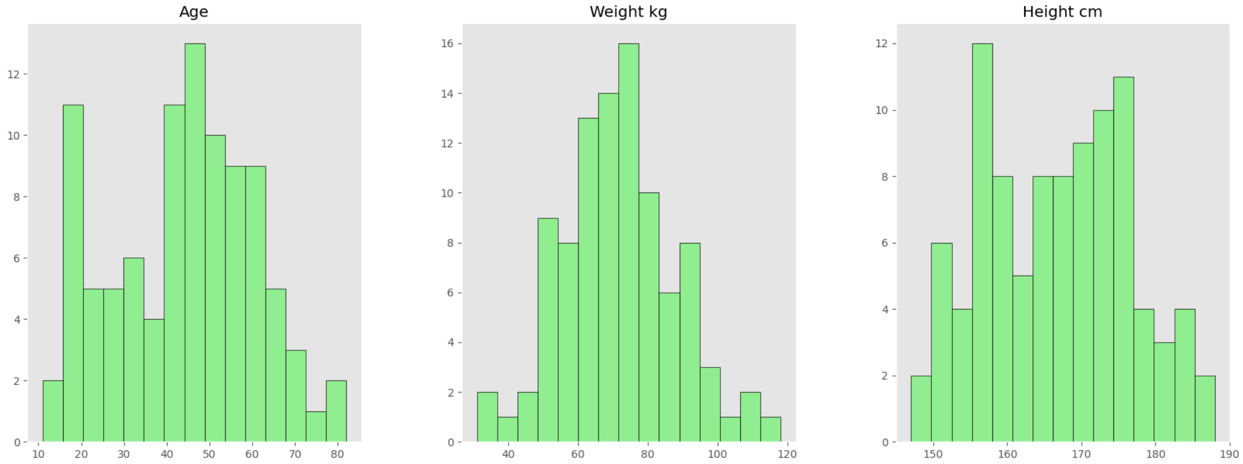
```
df2 = df2.drop(df.index[95])
```

Renaming columns to a more pleasing format.

```
df2 = df2.rename(columns={
    'AGE': 'Age',
    'SEX': 'Sex',
    'WEIGHT-KGs': 'Weight kg',
    'HEIGHT-CM': 'Height cm',
    'ACTIVITY LEVEL' : 'Activity level',
    'PAIN 1' : 'Neck pain',
    'PAIN 2' : 'Thoracic pain',
    'PAIN 3' : 'Lumbar pain',
    'PAIN 4' : 'Sacral pain',
    'POSTURE' : 'Posture'
})

df2 = df2.reset_index(drop=True)
```

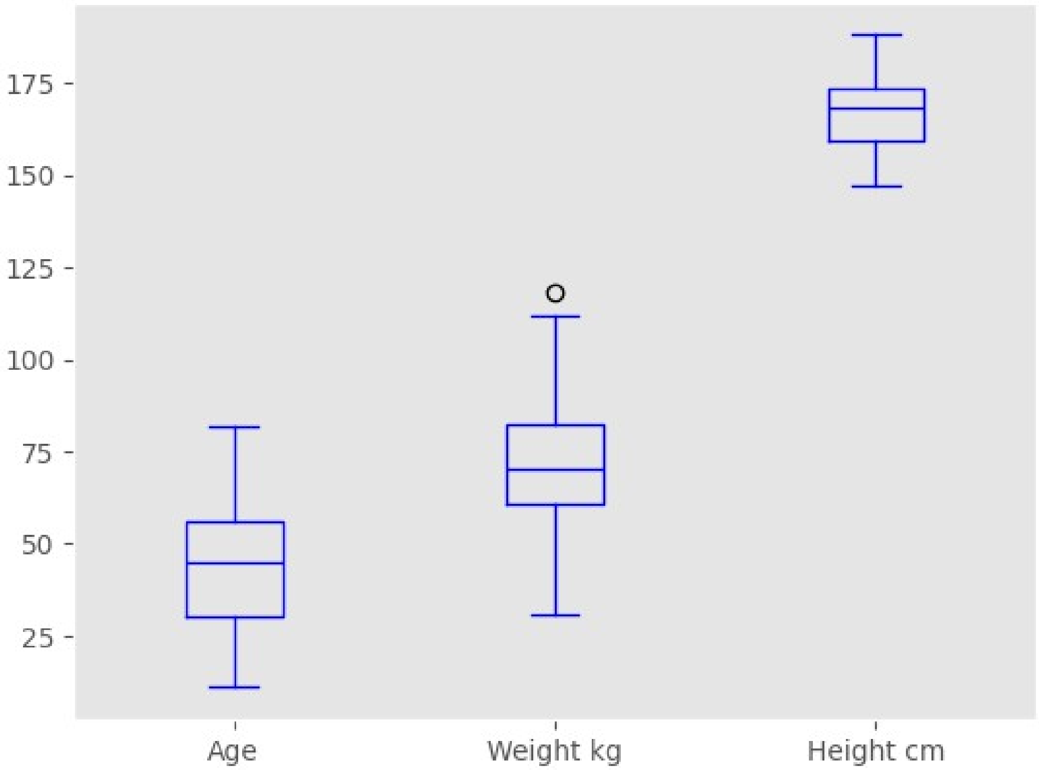
Distribution of age, weight and height of participants.

```
tables = ['Age', 'Weight kg', 'Height cm']
ax = df2[tables].hist(figsize = (20, 7), color = 'lightgreen',
edgecolor = 'black', layout = (1, 3)
                ,grid=False, bins = 15)
```

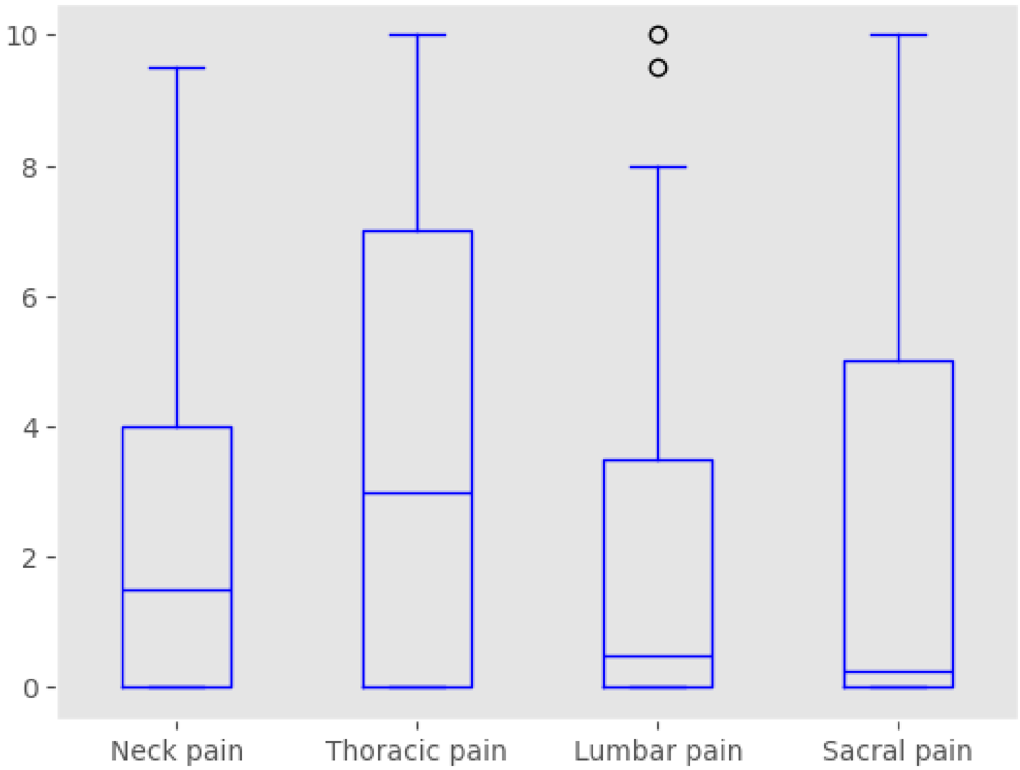


Looking for outliers in terms of age, weight, height and back pain types

```
df2[['Age', 'Weight kg', 'Height cm']].plot(kind='box', grid=False,
color='blue')
plt.show()
```



```
df2[pain_type].plot(kind='box', grid=False, color='blue')
plt.show()
```



```
df2.sort_values(by='Lumbar pain', ascending=False).head(5)

    Age     Sex  Weight kg  Height cm Activity level  Neck pain  \
6    62    Male      118.0      173.0            Low        7.0   
3    51    Male       79.0      168.0       Moderate        9.5   
12   48  Female       57.0      163.0            Low        5.0   
17   33    Male       77.0      183.0       Moderate        4.0   
61   25    Male       72.0      170.0            Low        5.0   

    Thoracic pain  Lumbar pain  Sacral pain  MBTI   E   I   S   N   T
F   J  \
6            10.0         10.0         10.0  ISTP   7  14  20   6  14
10   9
3             9.5          9.5          1.5  ISTJ   6  15  21   5  13
11  19
12            7.0          8.0          7.0  ESFJ  13   8  14  12   9
15  14
17            9.0          8.0          0.0  ENFP  17   4  10  16  11
13   4
61            0.0          8.0          0.0  ESTP  16   5  19   7  19
5   7

     P            Posture  Sum of pain MBTI group
6   13  kyphosis-lordosis         37.0   Explorer
3    3          sway-back         30.0   Sentinel
12   8          sway-back         27.0   Sentinel
```

```
17  18  kyphosis-lordosis        21.0   Diplomat
61  15       ideal posture       13.0   Explorer
```

Decided to drop two outling male participants

```
df3 = df2.drop([df2.index[6], df2.index[3]])
```

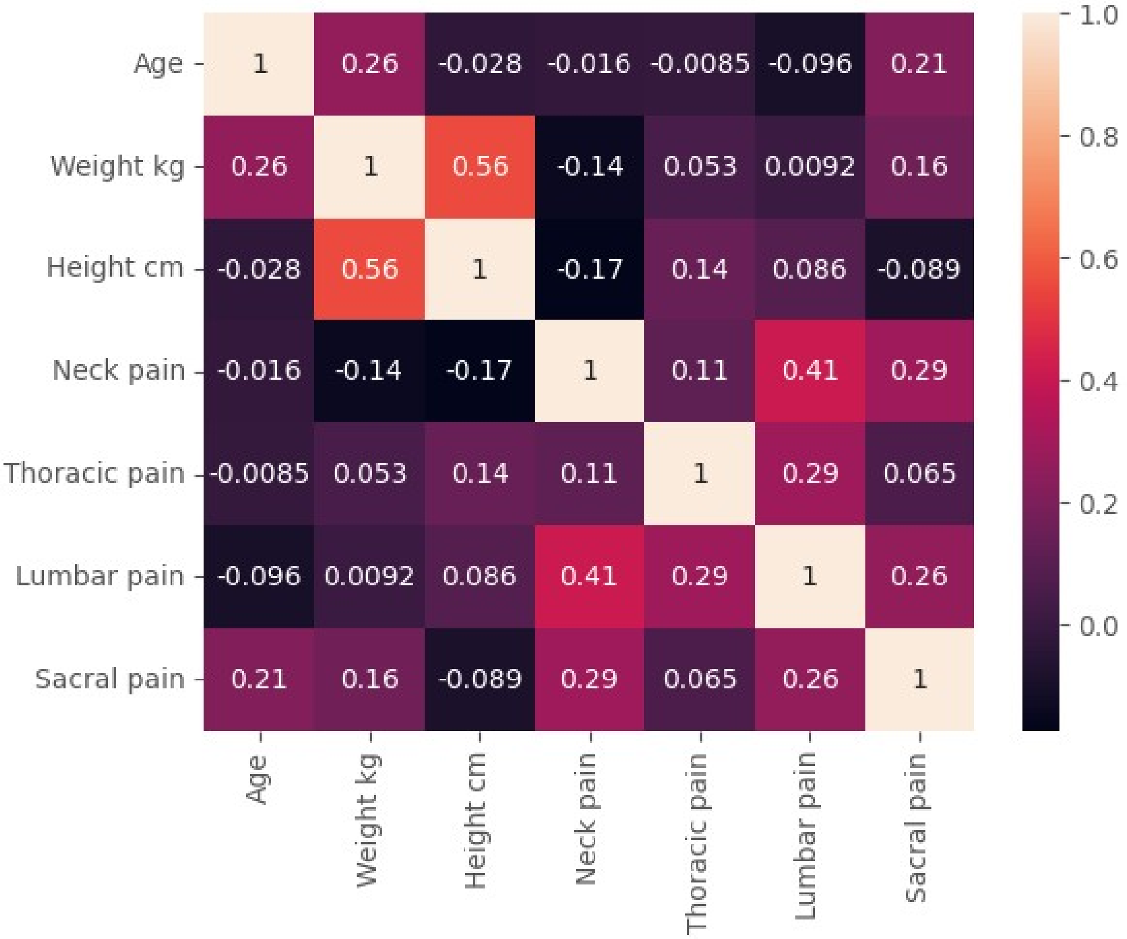
Looking for correlations between variables.

```
corr = df3[['Age', 'Weight kg', 'Height cm', 'Neck pain', 'Thoracic
pain', 'Lumbar pain', 'Sacral pain']].corr()
```

There is a weak correlation between age and sacral pain (0.21), weight and sacral pain (0.16).

There is also a moderate correlation between lumbar pain and neck pain (0.41), and also weak correlation between neck pain and sacral pain (0.29), thoracic pain and lumbar pain (0.29), sacral pain and lumbar pain (0.26).

```
sns.heatmap(corr, annot=True)
plt.show()
```



Adding the total sum of pain column for each participant

```
df3['Sum of pain'] = df3['Neck pain'] + df3['Thoracic pain'] +
df3['Lumbar pain'] + df3['Sacral pain']
```

The mean age of participants is 44 years, mean weight is 71.1 kg, mean height is 166.8 cm.

The youngest participant is 11 years old, the oldest one is 82 years old.

```
df3.describe().round(2)

         Age  Weight kg  Height cm  Neck pain  Thoracic pain  Lumbar
pain  \
count  94.00      94.00      94.00      94.00          94.00
94.00
mean   43.76      71.10     166.78       2.03           3.66
1.80
std    16.84      15.77       9.68       2.44           3.09
2.35
```

```
min    11.00      31.00     147.00       0.00           0.00
0.00
25%    30.00      60.25     157.75       0.00           0.00
0.00
50%    45.00      70.00     168.00       1.00           3.00
0.25
75%    56.00      82.00     174.50       4.00           7.00
3.00
max    82.00     112.00     188.00       8.00          10.00
8.00

       Sacral pain      E      I      S      N      T      F      J
P  \
count        94.00  94.00  94.00  94.00  94.00  94.00  94.00  94.00
94.00
mean          2.49  12.84   8.14  14.95  11.13  10.47  13.48  10.20
11.80
std           3.05   5.73   5.71   4.79   4.73   5.39   5.36   5.73
5.72
min           0.00   2.00   0.00   5.00   1.00   0.00   2.00   0.00
2.00
25%           0.00   9.00   3.00  11.25   8.00   6.00   9.25   5.00
6.25
50%           0.00  13.50   7.50  15.00  11.00  11.00  13.00  10.50
11.50
75%           5.00  18.00  12.00  18.75  14.75  14.75  18.00  15.75
17.00
max          10.00  21.00  19.00  25.00  21.00  22.00  24.00  20.00
22.00

       Sum of pain
count        94.00
mean          9.98
std           7.09
min           0.00
25%           4.00
50%           9.00
75%          15.75
max          31.00
```

Grouping the dataframe by sex.

```
group_sex = df3.groupby('Sex')

df3['Sex'].value_counts()

Sex
Female    48
Male      46
Name: count, dtype: int64
```

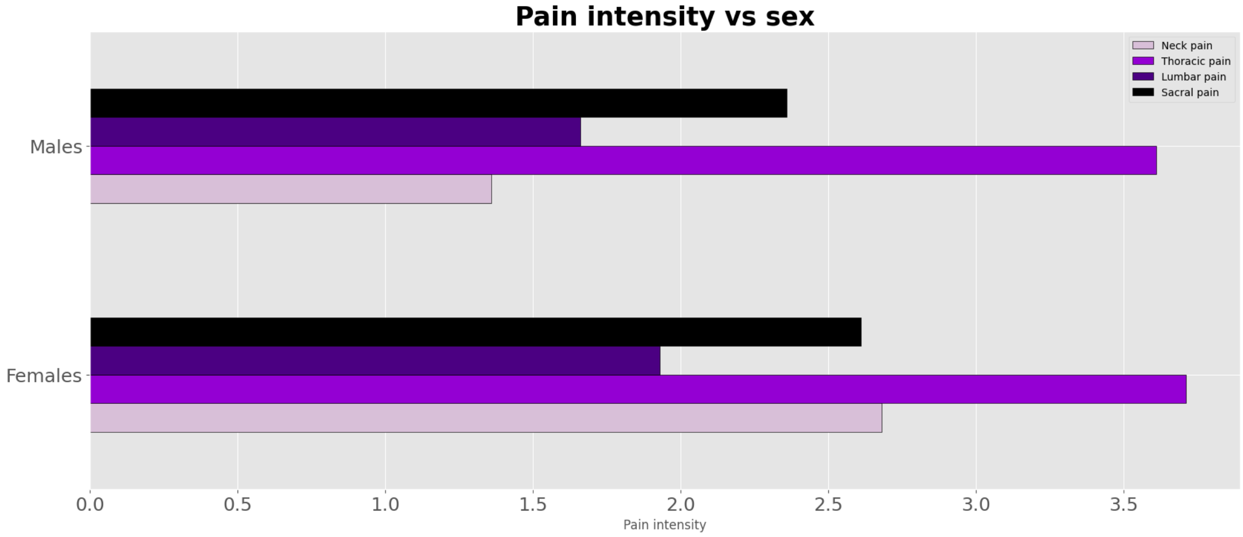
```
group_sex.mean(numeric_only=True).round(2)
```

```
          Age  Weight kg  Height cm  Neck pain  Thoracic pain  Lumbar
pain  \
Sex

Female  43.31      63.27     161.62       2.68           3.71
1.93
Male    44.22      79.26     172.15       1.36           3.61
1.66

        Sacral pain      E     I      S      N      T      F      J
P  \
Sex

Female         2.61  12.02  8.96  16.00  10.15   8.29  15.60  11.25
10.77
Male           2.36  13.70  7.28  13.85  12.15  12.74  11.26   9.11
12.87

        Sum of pain
Sex
Female        10.93
Male           8.99
```

The pain scale consisted of a numbers between 0 (no pain) to 10 (very high pain).

Looking at the mean values of four types of back pain, all of them are higher for females. Also, the one type that stands out is neck pain, which on average is significantly more prominent for females (2.68 on average in pain scale vs 1.36 for males).

```
ax = group_sex[['Neck pain', 'Thoracic pain',
                'Lumbar pain','Sacral
pain']].mean(numeric_only=True).round(2).plot.barh(figsize = (20, 8),
                                                         edgecolor =
'black', grid=True, xlabel='Pain intensity',
                                        rot=0, fontsize=18,
color=['thistle', 'darkviolet', 'indigo', 'black'],

ylabel='')
ax.set_title('Pain intensity vs sex', weight='bold', size=25)

plt.yticks([0,1],['Females', 'Males'])
plt.show()
```



Replacing posture types to a more descriptive ones.

```
# Posture - ideal posture (A), kyphosis-lordosis (B), flat back (C),
sway-back (D)
df3['Posture'].replace({'A': 'ideal posture',
                        'B' : 'kyphosis-lordosis',
                        'C' : 'flat back',
                        'D' : 'sway-back'}, inplace=True)
```

Grouping the participants by their posture type.

```
group_posture = df3.groupby('Posture')
```

Most dominant posture type in this particular sample kyphosis-lordosis. Only 23% of participants have ideal posture

```
df3['Posture'].value_counts()

Posture
kyphosis-lordosis    34
ideal posture        22
sway-back            19
flat back            19
Name: count, dtype: int64
```

People with kyphosis-lordosis are on average the youngest, while people with sway-back are on average the oldest.

```
group_posture.mean(numeric_only=True).round(2)

                     Age  Weight kg  Height cm  Neck pain  Thoracic
pain  \
Posture
```

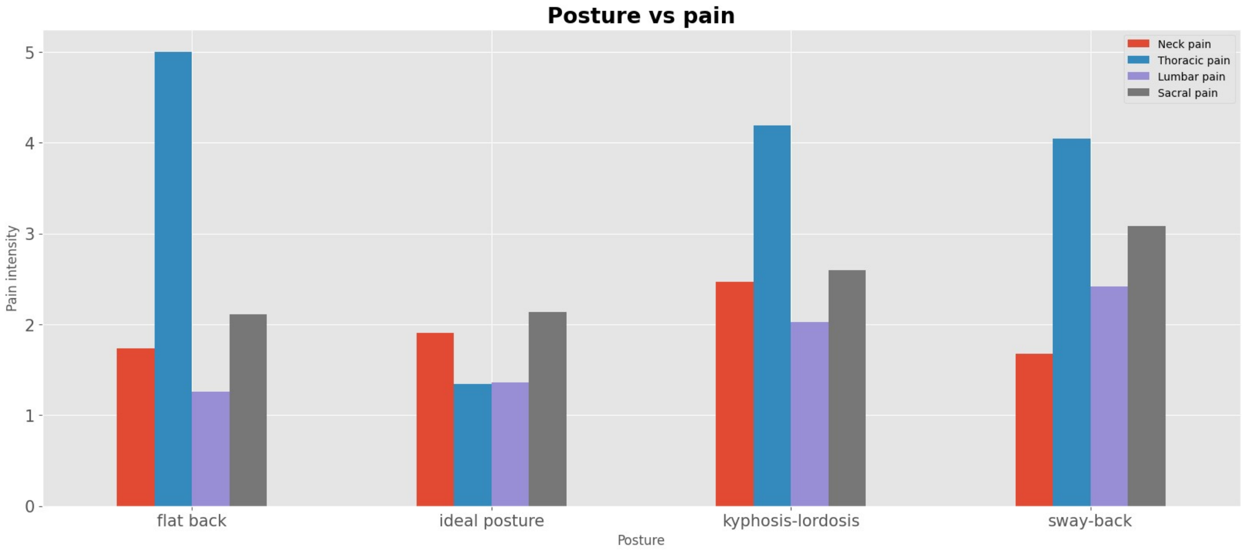
```
flat back            46.37       73.79      169.74        1.74
5.00
ideal posture        44.41       73.95      165.73        1.91
1.34
kyphosis-lordosis    39.15       71.12      166.38        2.47
4.19
sway-back            48.63       65.05      165.74        1.68
4.05

                     Lumbar pain  Sacral pain      E       I       S
N  \
Posture

flat back                   1.26         2.11  10.63   10.42   14.47
11.47
ideal posture               1.36         2.14  16.59    4.41   14.59
11.41
kyphosis-lordosis           2.03         2.60  13.97    6.97   15.06
11.18
sway-back                   2.42         3.08   8.68   12.26   15.63
10.37

                         T       F       J       P  Sum of pain
Posture
flat back             9.89   14.05   12.00   10.05        10.11
ideal posture         9.05   14.86    6.82   15.18         6.75
kyphosis-lordosis    11.26   12.68   10.85   11.12        11.29
sway-back            11.26   12.74   11.16   10.84        11.24
```

The most prominent kind of pain for each posture type: flat back - thoracic pain, ideal posture - sacral pain, kyphosis-lordosis - thoracic pain, sway-back - thoracic pain.

As can be seen on the graphic below, the thoracic pain dominates in three types of unhealthy posture types. The ideal posture seems to protect an individual from this type of pain quite well.

If we sum up the mean values of pain types, the results are as follows: ideal posture - 6.75, flat back - 10.11, sway-back - 11.24, kyphosis-lordsosis - 11.29. So having a good posture indeed protects from suffering from various types of back pain, while people with the kyphosis-lordosis and sway-back posture are the most at risk of suffering from back pain.

```
pain_type = ['Neck pain', 'Thoracic pain', 'Lumbar pain', 'Sacral
pain']
ax = group_posture[pain_type].mean().round(2).plot(kind='bar',
ylabel='Pain intensity', rot=0, figsize=(20,8),
                                                  fontsize=15)
ax.set_title('Posture vs pain', weight='bold', fontsize=20)
plt.show()
```



Grouping the dataframe by participants' day-to-day activity level (low, moderate or high)

```
group_activity = df3.groupby('Activity level')
```

Highly active participants are the youngest on average, while participants with low activity level are the oldest on average

```
group_activity.mean(numeric_only=True).round(2)
```

```
                  Age  Weight kg  Height cm  Neck pain  Thoracic pain
\
Activity level

High            32.50      68.67     172.17       2.33           4.33

Low             46.47      71.11     165.99       2.07           3.72

Moderate        35.75      71.94     168.31       1.75           3.12


                Lumbar pain  Sacral pain      E     I      S      N
T  \
Activity level

High                   1.00         0.50  13.17  7.83  12.67  13.33
12.67
Low                    1.92         2.71  12.44  8.53  15.56  10.54
```

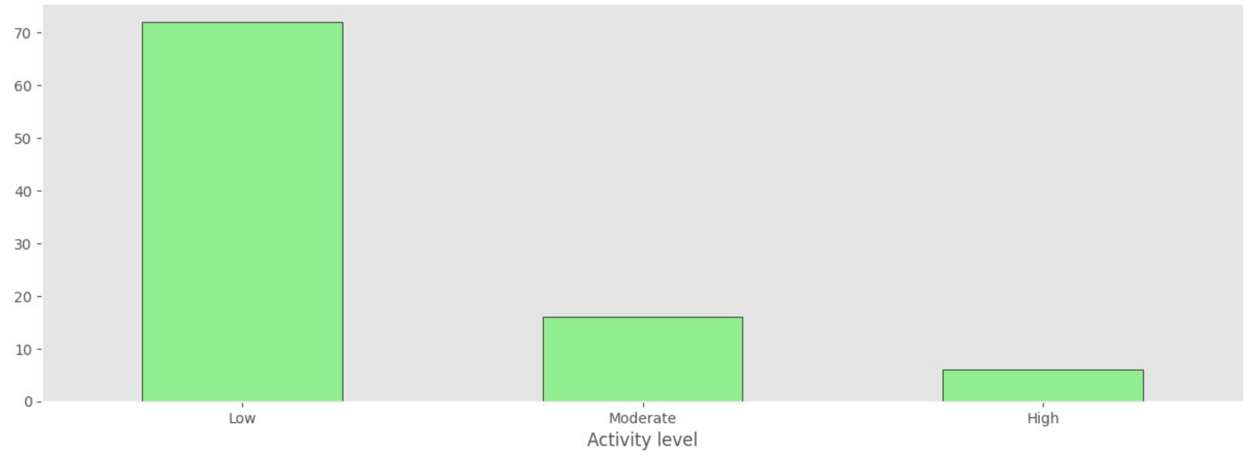
```
10.62
Moderate                   1.56          2.25  14.50  6.50  13.06  12.94
8.94

                    F      J      P  Sum of pain
Activity level
High            11.33  10.67  11.33         8.17
Low             13.31  10.67  11.33        10.42
Moderate        15.06   7.94  14.06         8.69
```

```
df3['Activity level'].value_counts()
```

```
Activity level
Low         72
Moderate    16
High         6
Name: count, dtype: int64
```

Majority of participants declared low activity level, only six of them declared high activity level

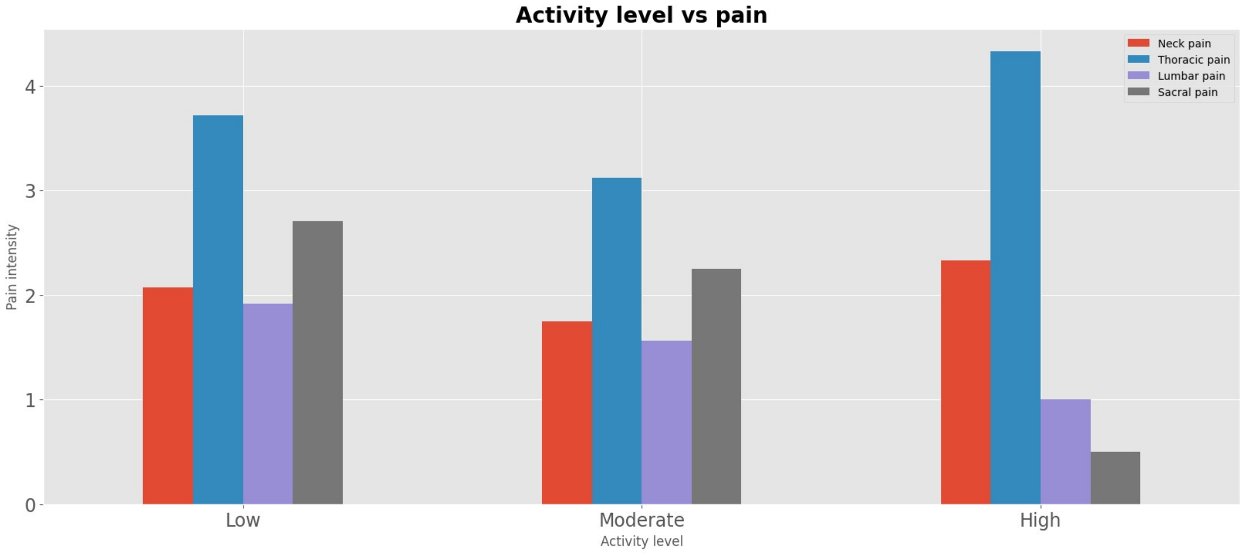
```
df3['Activity level'].value_counts().loc[['Low', 'Moderate',
'High']].plot.bar(rot=0, grid=False,

edgecolor='black',

figsize=(15,5), color='lightgreen')
plt.show()
```



Judging by the graph, the back pain type mean values are slightly lower for low and moderate activity levels. High activity protects from lumbar and sacral pain but active people are still prone to moderate neck pain, and, to a surprise, more prone to a thoracic pain than the other groups.

At the same time we have to remember that the sample for moderate activity was small, and for high one even smaller (only six participants declared high activity). Therefore it is advised to be careful of drawing solid conclusions from the graph below.

```
ax = group_activity[pain_type].mean().round(2).loc[['Low', 'Moderate', 'High']].plot(kind='bar',
                                                                ylabel='Pain intensity', rot=0, figsize=(20,8),
                                                                fontsize=17)
ax.set_title('Activity level vs pain', weight='bold', fontsize=20)
plt.show()
```



Taking a look into MBTI personality type distribution.

```
df3['MBTI'].value_counts()

MBTI
ESFP    12
ESFJ    10
ESTP    10
ENFP    10
ESTJ     7
ISFP     7
ISTJ     5
```

```
ENFJ      5
ENTP      5
ISFJ      5
ENTJ      5
INFP      5
INFJ      4
ISTP      3
INTJ      1
Name: count, dtype: int64
```

The four most prominent MBTI personality types in this particular sample are:

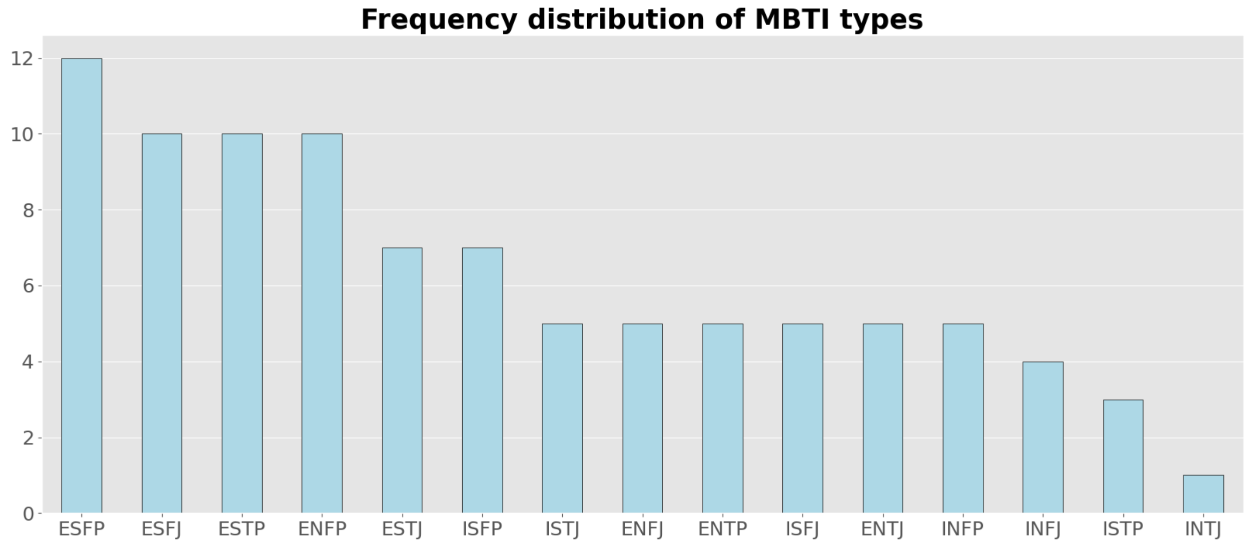
ESFP (Entertainer): these people love to experience new things and find pleasure in discovering the unknown. They are social, and enjoy shared activities.

ESFJ(Consul): they are people-focused, and they enjoy taking part in their social community. They willingly offer guidance to others.

ESTP (Entrepreneur): They are action-oriented, and constantly look for new opportunities, whether it's socializing with others or more personal pursuits.

ENFP (Campaigner): These people tend to have a lot of enery and share that energy with others. They embrace big ideas and actions that reflect their sense of goodwill toward others.

```python
ax = df3['MBTI'].value_counts().plot.bar(figsize = (20, 8), color = 
'lightblue', edgecolor = 'black',
                                        rot=0, fontsize=18)
ax.set_title('Frequency distribution of MBTI types', weight='bold', 
size=25)
ax.xaxis.label.set_visible(False)
ax.grid(axis='x')
plt.show()
```



```
group_mbti = df3.groupby('MBTI')

group_mbti.mean(numeric_only=True).round(2).sort_values(by='Sum of
pain',

ascending=False)
```

```
        Age  Weight kg  Height cm  Neck pain  Thoracic pain  Lumbar
pain  \
MBTI

INFJ  36.25      61.50     167.00       2.50           4.88
5.00
ENFJ  33.00      73.40     168.00       2.80           5.20
2.80
ESTJ  55.29      74.29     166.86       1.71           3.43
2.93
ISFP  55.86      71.14     167.57       1.43           6.21
1.64
ESFJ  47.50      69.40     164.20       2.60           2.85
2.15
INFP  36.20      59.20     164.40       2.00           4.40
2.60
ENTP  47.80      87.40     166.60       1.80           3.70
0.60
ESFP  40.58      69.50     164.83       2.33           3.92
1.12
ESTP  39.80      73.50     170.40       1.20           3.20
1.70
ENTJ  36.20      73.80     172.20       1.10           4.30
1.70
ISTJ  52.80      68.40     168.00       3.40           3.40
```

```
1.00
ENFP  42.60      71.70     165.10       2.20           2.00
1.50
ISFJ  41.40      65.00     163.80       2.20           3.00
0.40
ISTP  47.67      77.67     164.33       0.83           2.17
1.50
INTJ  24.00      70.00     185.00       2.00           3.00
0.00

      Sacral pain      E      I      S      N      T      F      J
P  \
MBTI

INFJ         2.75   5.25  15.25  13.75  12.25   8.00  16.00  15.25
6.75
ENFJ         3.40  16.80   4.20  11.20  14.40   5.80  18.00  15.20
6.80
ESTJ         4.79  14.29   6.71  17.43   8.57  14.43   9.57  16.43
5.57
ISFP         2.43   6.57  14.43  17.71   8.29   8.29  15.71   6.14
15.86
ESFJ         4.00  15.40   5.60  17.50   9.50   8.00  15.80  15.00
7.00
INFP         2.60   6.20  14.80  10.20  15.80   7.40  16.60   6.60
15.40
ENTP         4.40  16.60   4.40   9.20  16.80  17.60   6.40   6.40
15.40
ESFP         1.83  16.08   4.92  17.00   8.92   7.08  16.75   5.92
16.08
ESTP         2.85  16.20   4.80  19.10   6.90  16.10   7.90   6.80
15.20
ENTJ         1.50  17.20   3.80  10.40  15.60  12.60  11.40  15.40
6.60
ISTJ         0.60   5.20  15.80  16.80   9.20  18.80   5.20  16.80
5.20
ENFP         1.20  17.60   3.40   8.80  17.20   5.50  18.50   3.50
18.50
ISFJ         1.00   4.40  16.60  19.80   6.20   8.60  15.40  15.80
6.40
ISTP         0.83   6.00  15.00  16.33   9.67  14.67   9.33   6.00
16.00
INTJ         0.00   5.00  16.00   9.00  17.00  14.00  10.00  17.00
5.00

      Sum of pain
MBTI
INFJ        15.12
ENFJ        14.20
ESTJ        12.86
```

```
ISFP        11.71
ESFJ        11.60
INFP        11.60
ENTP        10.50
ESFP         9.21
ESTP         8.95
ENTJ         8.60
ISTJ         8.40
ENFP         6.90
ISFJ         6.60
ISTP         5.33
INTJ         5.00
```

Two MBTI types most prone to suffer from back pain are:

ENFJ (Protagonist): Those people love helping others, and they tend to have strong ideas and values.

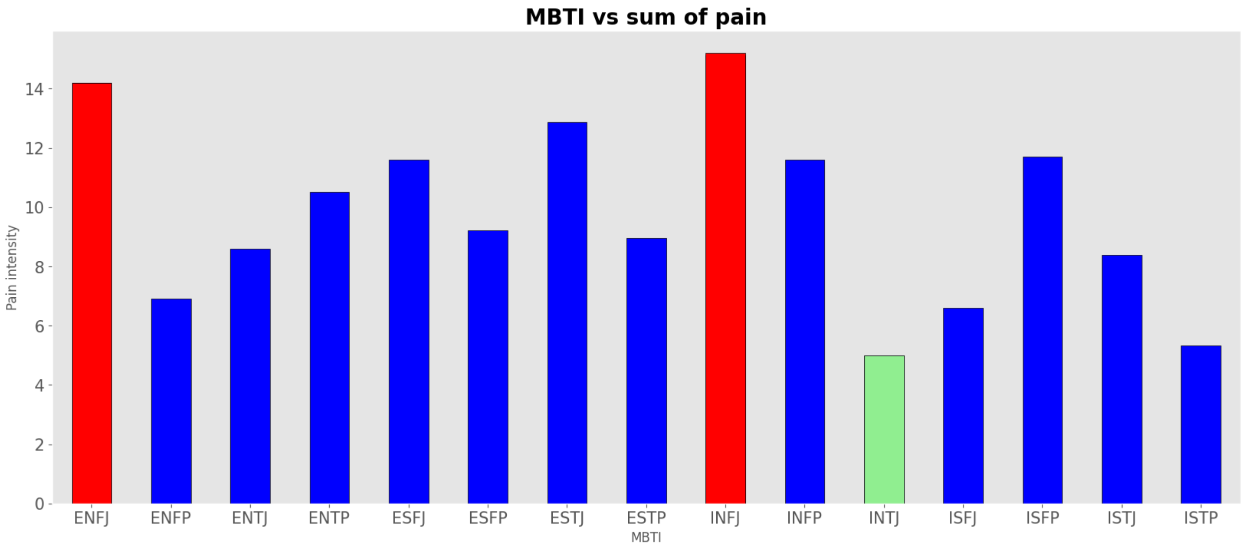
INFJ (Advocate): They approach life with deep thoughtfulness and imagination. What guides them through life is their inner vision and personal values.

The least prone back pain are types:

INTJ (Architect): Those people are perfectionists, they apply creativity and rationality to everything they do. Their inner world is complex and private

ISTP (Virtuoso) Strongly indvidualistic mindset, they are pursuing goals with intrinsic motivation and for their own satisfaction.

```
ax = group_mbti['Sum of pain'].mean().round(2).plot(kind='bar',
ylabel='Pain intensity', rot=0, figsize=(20,8),
                                                  fontsize=15,
grid=False, edgecolor='black', color='blue')
ax.set_title('MBTI vs sum of pain', weight='bold', fontsize=20)
ax.bar([8],[15.2], color='red', edgecolor='black', width=0.5)
ax.bar([0],[14.2], color='red', edgecolor='black', width=0.5)
ax.bar([10],[5], color='lightgreen', edgecolor='black', width=0.5)
plt.show()
```



Grouping MBTI types into 4 main personality categories:

Analysts: rational and intellectual

Diplomats: emphatic and idealistic

Sentinels: stable and self-motivated

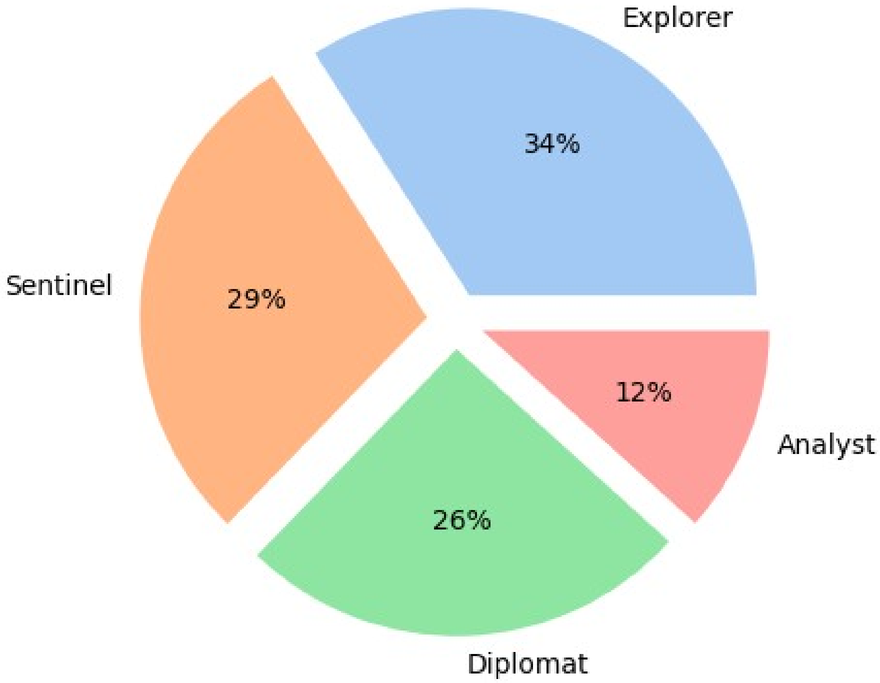
Explorers: spontaneus and flexible

```
def MBTI_group(MBTI):
    if MBTI in ['INTJ', 'INTP', 'ENTJ', 'ENTP']:
        return 'Analyst'
    elif MBTI in ['INFJ', 'INFP', 'ENFJ', 'ENFP']:
        return 'Diplomat'
    elif  MBTI in ['ISTJ', 'ISFJ', 'ESTJ', 'ESFJ']:
        return 'Sentinel'
    else:
        return 'Explorer'

df3['MBTI group'] = df2['MBTI'].map(MBTI_group)
```

In the sample the dominating group are explorers, while the least represented are analysts

```
df3['MBTI group'].value_counts()

MBTI group
Explorer    32
Sentinel    27
Diplomat    24
Analyst     11
Name: count, dtype: int64
```

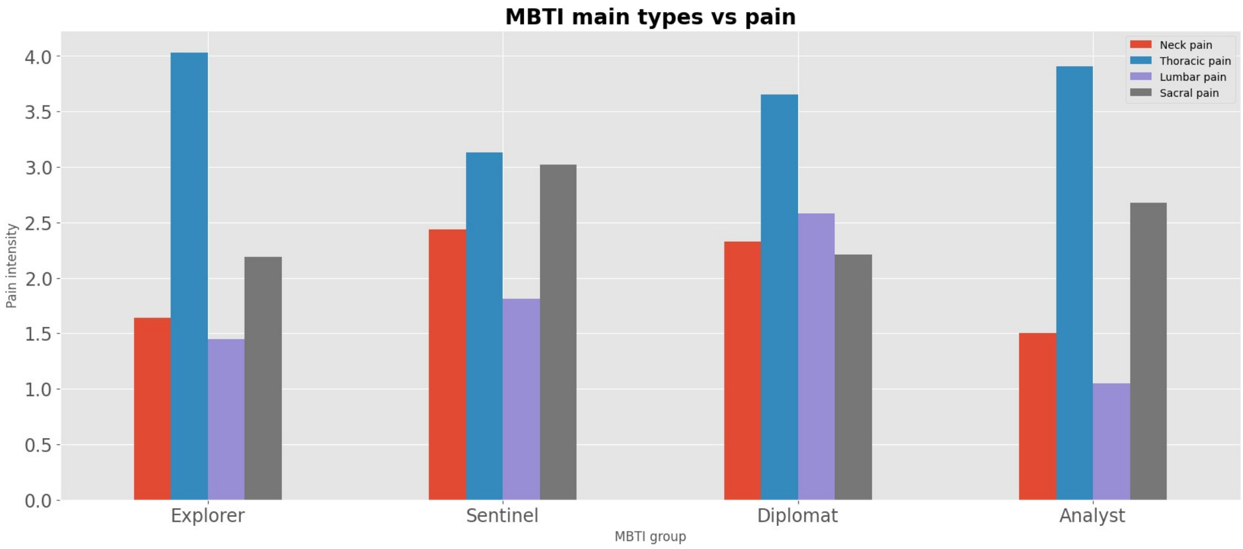
```
data = df3['MBTI group'].value_counts()
labels= ['Explorer', 'Sentinel', 'Diplomat', 'Analyst']
explode = [0.1, 0.1, 0.1, 0.1]
colors = sns.color_palette('pastel')[0:4]
plt.pie(data, labels=labels, colors = colors, autopct='%.0f%%',
explode=explode)
plt.show()
```



```
MBTI_main_types = df3.groupby('MBTI group')
```

Most prominent back pain type for all four MBTI categories is thoracic pain. Least prominent for almost all categories is lumbar pain, with the exception of diplomats, where mean pain values for lumbar pain are higher than sacral and neck pain mean values.

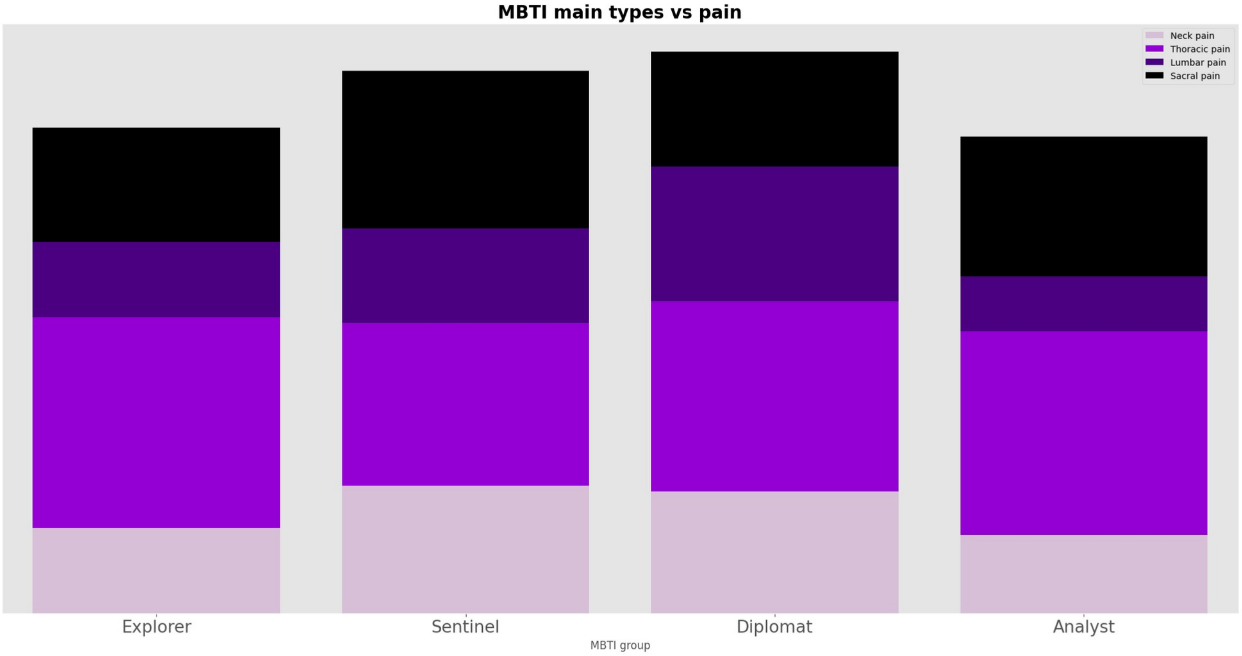
```
ax = MBTI_main_types[pain_type].mean().round(2).loc[['Explorer',
'Sentinel', 'Diplomat', 'Analyst']].plot(kind='bar',
                                                ylabel='Pain
intensity', rot=0, figsize=(20,8),
                                                fontsize=17)
ax.set_title('MBTI main types vs pain', weight='bold', fontsize=20)
plt.show()
```



```
MBTI_main_types[pain_type].mean().round(2)

            Neck pain  Thoracic pain  Lumbar pain  Sacral pain
MBTI group
Analyst          1.50           3.91         1.05         2.68
Diplomat         2.33           3.65         2.58         2.21
Explorer         1.64           4.03         1.45         2.19
Sentinel         2.44           3.13         1.81         3.02

import numpy as np
```

Representing the same data about four MBTI categories and pain with use of stacked bar chart

```
main_types = ['Explorer', 'Sentinel', 'Diplomat', 'Analyst']
y1 = np.array([1.64, 2.44, 2.33, 1.5])
y2 = np.array([4.03, 3.13, 3.65, 3.91])
y3 = np.array([1.45, 1.81, 2.58, 1.05])
y4 = np.array([2.19, 3.02, 2.21, 2.68])

colors = ['thistle', 'darkviolet', 'indigo', 'black']

ax = MBTI_main_types[pain_type].mean().round(2).loc[['Explorer',
'Sentinel', 'Diplomat', 'Analyst']].plot(kind='bar',
                                                            rot=0,
figsize=(25,12),
                                                           fontsize=19,
grid=False, color = colors)
plt.bar(main_types, y1, color=colors[0])
plt.bar(main_types, y2, bottom=y1, color=colors[1])
plt.bar(main_types, y3, bottom=y1+y2, color=colors[2])
plt.bar(main_types, y4, bottom=y1+y2+y3, color=colors[3])
plt.legend(["Neck pain", "Thoracic pain", "Lumbar pain", "Sacral
pain"])
plt.title("MBTI main types vs pain", fontsize=20, weight='bold')
plt.yticks([])
plt.show()
```



To quickly summarize:

- if a person suffers from one type of back pain there is a slightly bigger chance they will also suffer from other types of back pain;

- age and weight could slightly affect chance of suffering from sacral pain;

- women are more prone to back pain than men, especially neck pain;

- having a good posture can protect individual to some degree from back pain;

- being higly active could lead to better protection from lumbar and sacral pain, but as was highlighted earlier in the project, the sample for highly active individuals was very small, so the aspect of activity vs types of back pain would have to be replicated with a more representative research sample;

- the MBTI personality types that most susceptible to back pain are ENFJ (Protagonists) and INFJ (Advocates). The two least likely to be affected by back pain are INTJ (Architects) and ISTP (Virtuosos);

- the thoracic pain dominates in all four main MBTI categories.

Thank You!