

DA 380 - HW #6

Lucas Hasting

February 22, 2026

Contents

1	Descriptive Statistics	2
1.1	General Health	2
1.2	Positive Mental Health Days	2
1.3	Exercise Participation	3
2	Graphics	3
2.1	Distribution of obesity status	3
2.2	Distribution of days with poor mental health	4
2.3	Distribution of BMI by education level	6
2.4	Relationship between days with poor mental health and height	7
2.5	Average BMI for those with and without perfect physical health	8
3	Appendix	9
3.1	Python File	9
3.2	Python File Output	12

1 Descriptive Statistics

1.1 General Health

First, the overall mean and standard deviation for GENHLTH (the variable associated with general health) are 2.62 and 1.03 respectively. Next, the mean and standard deviation for GENHLTH grouped by INCOME3 (the variable associated with income group) are approximately 2.23 and 0.87 respectively when $\text{INCOME3} > 7$ (greater than the median). Moreover, the mean and standard deviation for GENHLTH grouped by INCOME3 are approximately 2.88 and 1.08 respectively when $\text{INCOME3} \leq 7$ (less than or equal to the median). From these results, we see that on average, everyone's general health is between very good (2) and good (3), with general health being lower when income is greater than the median. This result implies that the more money you make does not mean you will generally be healthier.

1.2 Positive Mental Health Days

First, the overall mean and standard deviation for ment_healthy_days_30 (the variable associated with positive mental health days) are 25.59 and 8.53 respectively. Next, the mean and standard deviation for ment_healthy_days_30 grouped by bmi_cat = "Underweight", "Normal", "Overweight", "Obese" (the variable representing each BMI category) are shown (approximately) in Table 1. From these results, we see that on average, people who are underweight are the least mentally healthy, however, we also see that these people have the highest standard deviation, and if some people are on medicine to become skinny, these results make sense as the side effects could be hurting their mental health. Moreover, we see that people who are overweight, but not obese, (on average) have the best mental health, which makes sense, as they are not to the point of hurting themselves over their weight, but may not necessarily be doing anything to decrease their weight which would give more time for hobbies.

	Mean	Standard Deviation
Underweight	22.29	11.23
Normal	25.11	8.94
Overweight	26.44	7.76
Obese	25.01	8.94

Table 1: ment_healthy_days_30 Mean and Standard Deviation grouped by BMI

1.3 Exercise Participation

First, the overall count and percentages for EXERANY2 (the variable associated with exercise participation) are that 1,118 (74.68%) people that responded to the survey said that they exercised outside of their regular job and 379 (25.32%) said that they did not. Next, using the variable sex_cat (the variable associated with sex category) to group the results, of the 1,118 people that said they exercised outside of their regular job, 561 (51%) reported they were male, and 539 (49%) reported they were female. Moreover, of the 379 said that they did not exercise outside of their regular job, 169 (45.68%) reported they were male, and 201 (54.32%) reported they were female. From these results, we see that there were a few more people who are female that reported they did not exercise outside of their job, however, this is not by a large amount. Moreover, this is not enough data to say anything regarding if sex (male/female) has anything to do with exercising outside of their regular job, but we can see that out of the people that responded to the survey, there are many more people that say they do exercise outside of their regular job.

2 Graphics

2.1 Distribution of obesity status

To begin, in Figure 1, we show the distribution of obesity status. Since it is a discrete (0/1) variable, we use a bar chart where the y-axis represents the number of people with the associated status. From the figure, we see that there are more people who are not obese than people who are obese, however, since there are 3 categories within the “Not Obese” category, it is surprising that there are more than half of the “Not Obese” category of people that are considered obese.

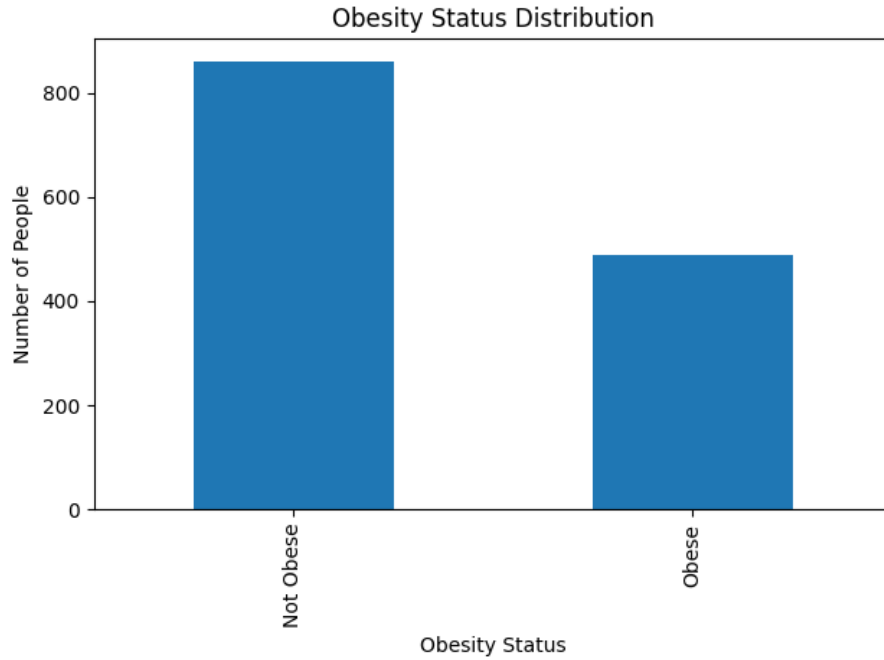


Figure 1

2.2 Distribution of days with poor mental health

To begin, in Figure 2, we show the distribution of days with poor mental health. Since it is a discrete (0-30) variable, we use a bar chart where the y-axis represents the number of people with the associated number of days. Moreover, there are a large number of people who had perfect mental health, which skews the distribution heavily to the right, so Figure 3 shows the results excluding those people to better view the distribution in those areas. From these figures we see that many people selected increments of 5, if not less 5 or equal to 7, instead of giving an exact number. This could imply (and is likely) that these results are just a guess, as increments of 5 tend to feel right to a user answering a survey, and thus should be interpreted as such.

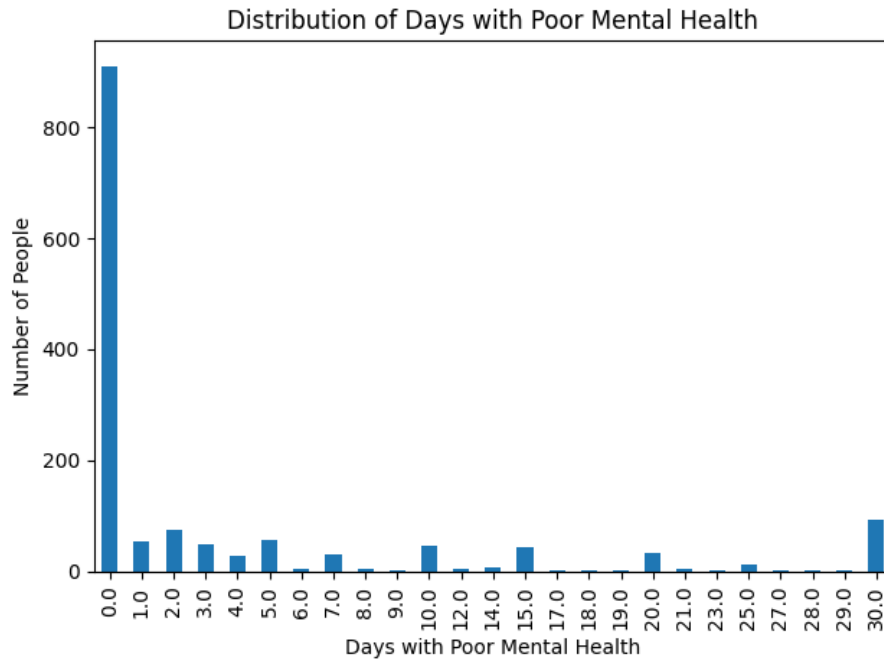


Figure 2

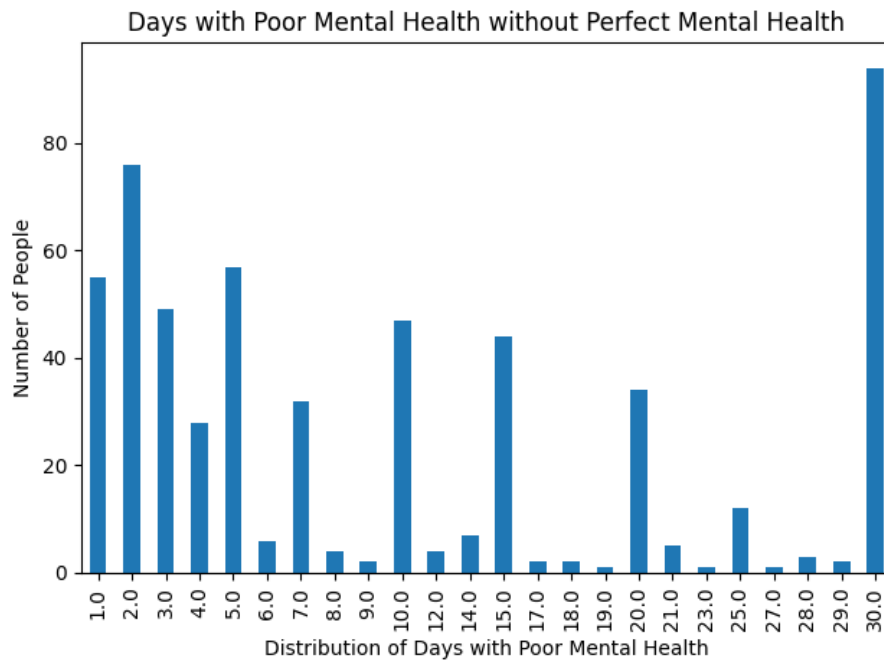


Figure 3

2.3 Distribution of BMI by education level

To begin, in Figure 4, we show the distribution of education level. Since education level is a discrete (1-6) variable, we use a bar chart where the y-axis represents the number of people with that education level. Next, to show the dispersion of BMI in this distribution, we use a box and whisker plot of BMI for each education level, which is shown in Figure 5. From these figures we see that most people that reported BMI had higher education (levels 4 and above), with there being 3 observations for level 1. Besides level 1, we see that the mean/quartiles are consistent with each level, which implies the education level did not have much (if any) effect on the result of their BMI score.

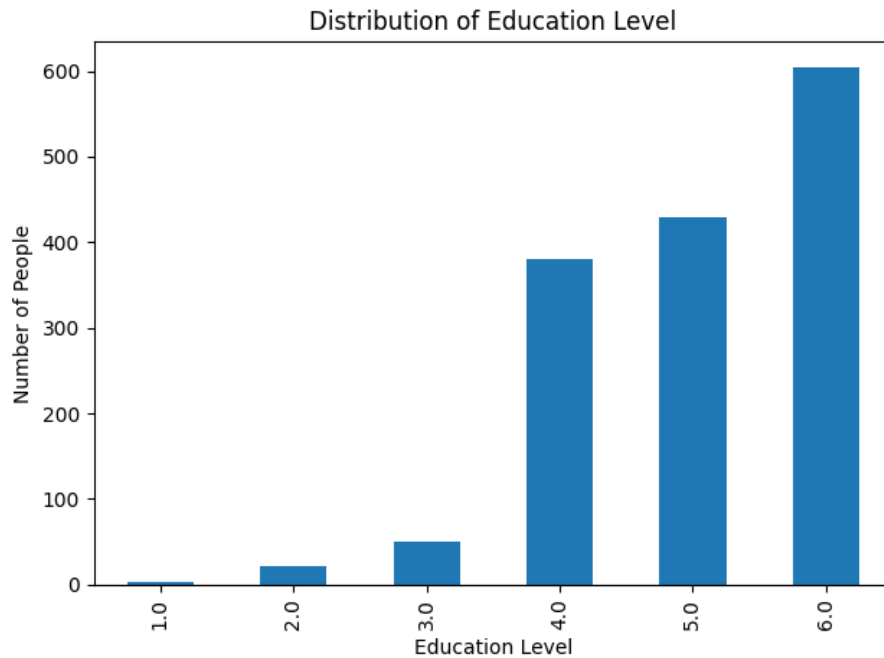


Figure 4

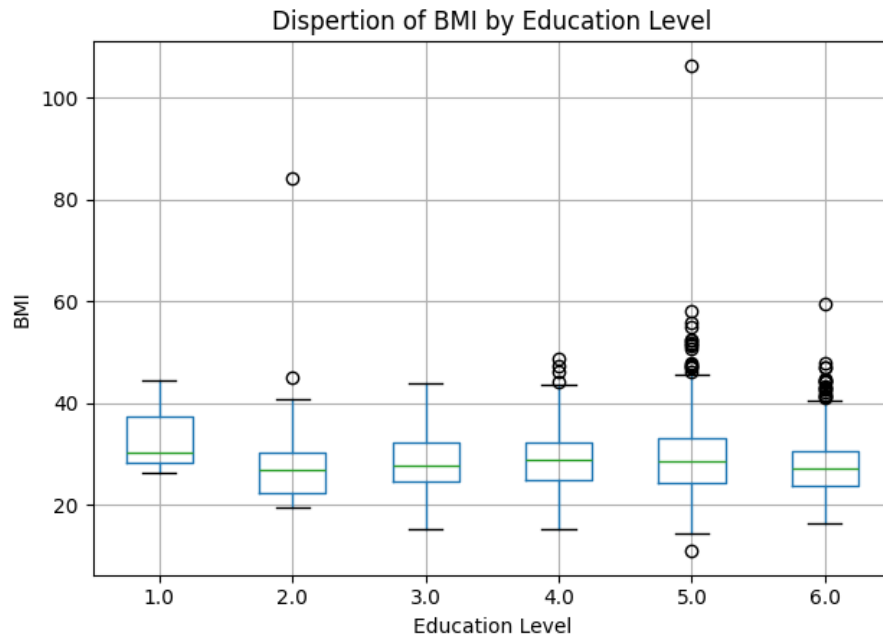


Figure 5

2.4 Relationship between days with poor mental health and height

To begin, in Figure 6, we show the relationship of both poor mental health days and height (in inches). We do this with a scatter plot to determine if there is a correlation between the two. From this figure, we see that there is no real correlation between the two variables, the results remain mostly constant during the change of either variable.

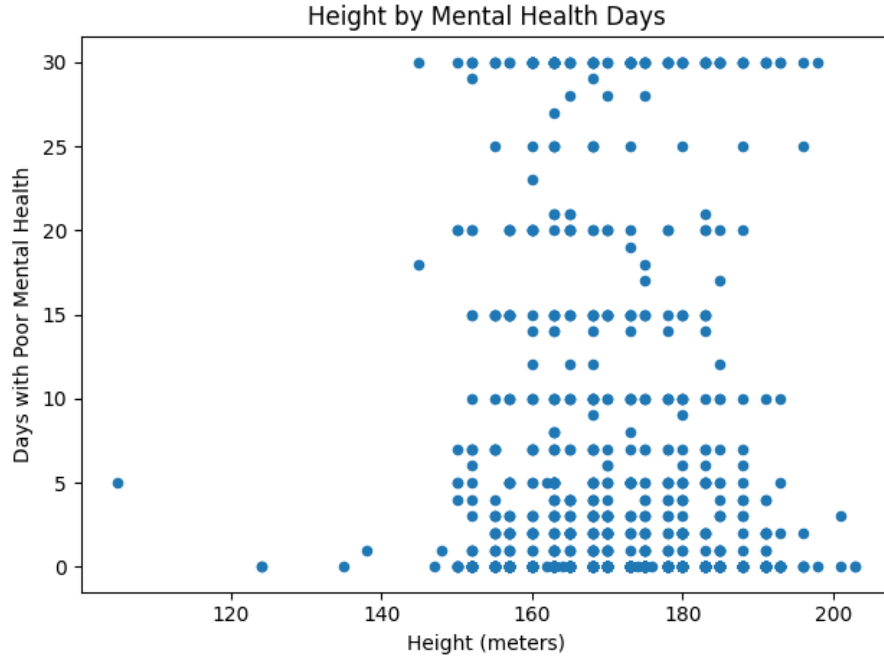


Figure 6

2.5 Average BMI for those with and without perfect physical health

To begin, in Figure 7, we show the dispersion of BMI with respect to people with/without perfect physical health. Since we are concerned with the average, we will use a box and whisker plot. More specifically, we compute the average BMI to be 28.25 for people with perfect physical health and 29.4 for people without perfect physical health. From the figure, we see that both categories of BMI have similar results, with people of perfect physical health having more outliers. These results imply that the reported physical health does not have much impact on BMI, but we expect it would have an impact, since having perfect physical health implies a healthy amount of body fat.

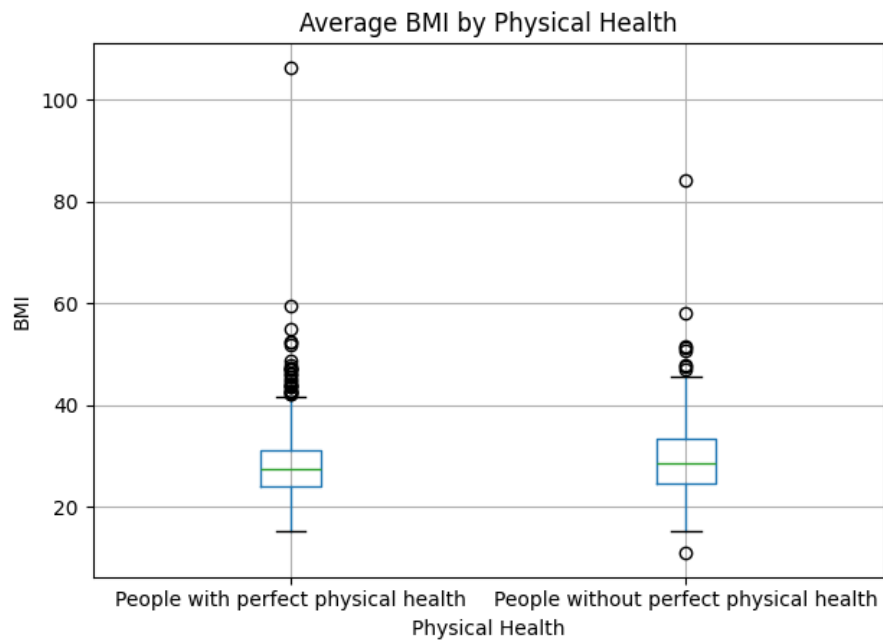


Figure 7

3 Appendix

3.1 Python File

```

1 '''
2 Name: Lucas Hasting
3 Course: DA 380
4 Instructor: Dr. Michael Floren
5 Date: 2/19/2026
6 Description: Get descriptive stas/graphs from brfss1_cleaned.csv
7 '''
8
9 #import needed libraries
10 import pandas as pd
11 import numpy as np
12 import matplotlib.pyplot as plt
13
14 #read in data
15 df = pd.read_csv("brfss1_cleaned.csv")
16
17 #--DESCRIPTIVE STATS
18
19 #General Health
20 print(f"GENHLTH mean: {round(df["GENHLTH"].mean(),2)}")
21 print(f"GENHLTH std: {round(df["GENHLTH"].std(),2)}")
22
23 print(f"\nmean - GENHLTH by {round(df.groupby("income_bin")["GENHLTH"].
    mean(),2)}")

```

```

24 print(f"\nstd - GENHLTH by {round(df.groupby("income_bin")["GENHLTH"].std
    (),2)}",end="\n\n")
25
26 #Positive Mental Health Days
27 print(f"ment_healthy_days_30 mean: {round(df["ment_healthy_days_30"].mean
    (),2)}")
28 print(f"ment_healthy_days_30 std: {round(df["ment_healthy_days_30"].std()
    ,2)}")
29
30 print(f"\nmean - ment_healthy_days_30 by {round(df.groupby("bmi_cat")["
    ment_healthy_days_30"].mean(),2)}")
31 print(f"\nstd - ment_healthy_days_30 by {round(df.groupby("bmi_cat")["
    ment_healthy_days_30"].std(),2)}",end="\n\n")
32
33 #Exercise Participation
34 print(df["EXERANY2"].value_counts(),end="\n\n")
35 print(round(df["EXERANY2"].value_counts(normalize=True)*100,2),end="\n\n")
36 print(df.groupby("EXERANY2")["sex_cat"].value_counts(),end="\n\n")
37 print(round(df.groupby("EXERANY2")["sex_cat"].value_counts(normalize=True)
    *100,2),end="\n\n")
38
39 #BMI by Physical Health
40 print(f"mean - BMI by {round(df.groupby("perf_phys_health")["bmi"].mean()
    ,2)}")
41
42 #--GRAPHICS
43
44 #Obesity Status
45 df["obese"].map({0: 'Not Obese', 1: 'Obese'}).value_counts().plot(kind="
    bar")
46 plt.ylabel("Number of People")
47 plt.xlabel("Obesity Status")
48 plt.title("Obesity Status Distribution")
49
50 plt.tight_layout()
51 plt.savefig("obesity.PNG")
52 plt.clf() #clear the plot
53 plt.close()
54
55 #Days with Poor Mental Health - 1
56 df["MENTHLTH"].value_counts().sort_index().plot(kind="bar")
57 plt.ylabel("Number of People")
58 plt.xlabel("Days with Poor Mental Health")
59 plt.title("Distribution of Days with Poor Mental Health")
60
61 plt.tight_layout()
62 plt.savefig("ment_health.PNG")
63 plt.clf() #clear the plot
64 plt.close()
65
66 #Days with Poor Mental Health - 2
67 df[df["MENTHLTH"] > 0]["MENTHLTH"].value_counts().sort_index().plot(kind="
    bar")
68 plt.ylabel("Number of People")

```

```

69 plt.xlabel("Distribution of Days with Poor Mental Health")
70 plt.title("Days with Poor Mental Health without Perfect Mental Health")
71
72 plt.tight_layout()
73 plt.savefig("ment_health_2.PNG")
74 plt.clf() #clear the plot
75 plt.close()
76
77 #BMI by education level
78 df.boxplot(column="bmi", by="EDUCA")
79 plt.suptitle('')
80 plt.ylabel("BMI")
81 plt.xlabel("Education Level")
82 plt.title("Dispersion of BMI by Education Level")
83
84 plt.tight_layout()
85 plt.savefig("bmi_edu_1.png")
86 plt.clf() #clear the plot
87 plt.close()
88
89 #Education level distribution
90 df["EDUCA"].value_counts().sort_index().plot(kind="bar")
91 plt.ylabel("Number of People")
92 plt.xlabel("Education Level")
93 plt.title("Distribution of Education Level")
94
95 plt.tight_layout()
96 plt.savefig("bmi_edu_2.png")
97 plt.clf() #clear the plot
98 plt.close()
99
100 #Relationship between days with poor mental health and height
101 df.plot(kind="scatter", x="HTM4", y="MENTHLTH")
102 plt.ylabel("Days with Poor Mental Health")
103 plt.xlabel("Height (meters)")
104 plt.title("Height by Mental Health Days")
105
106 plt.tight_layout()
107 plt.savefig("HTIN4_ment_hlth.png")
108 plt.clf() #clear the plot
109 plt.close()
110
111 #BMI by Physical Health
112 df["perf_phys_health_cat"] = df["perf_phys_health"].map({0: "People
    without perfect physical health", 1: "People with perfect physical
    health"})
113 df.boxplot(column="bmi", by="perf_phys_health_cat")
114 plt.suptitle('')
115 plt.ylabel("BMI")
116 plt.xlabel("Physical Health")
117 plt.title("Average BMI by Physical Health")
118
119 plt.tight_layout()
120 plt.savefig("bmi_perf_health.png")

```

```

121 plt.clf() #clear the plot
122 plt.close()

```

3.2 Python File Output

```

1 GENHLTH mean: 2.62
2 GENHLTH std: 1.03
3
4 mean - GENHLTH by income_bin
5 0.0      2.88
6 1.0      2.23
7 Name: GENHLTH, dtype: float64
8
9 std - GENHLTH by income_bin
10 0.0      1.08
11 1.0      0.87
12 Name: GENHLTH, dtype: float64
13
14 ment_healthy_days_30 mean: 25.59
15 ment_healthy_days_30 std: 8.53
16
17 mean - ment_healthy_days_30 by bmi_cat
18 Normal      25.11
19 Obese        25.01
20 Overweight   26.44
21 Underweight  22.29
22 Name: ment_healthy_days_30, dtype: float64
23
24 std - ment_healthy_days_30 by bmi_cat
25 Normal      8.94
26 Obese        8.94
27 Overweight   7.76
28 Underweight  11.23
29 Name: ment_healthy_days_30, dtype: float64
30
31 EXERANY2
32 1.0      1118
33 2.0       379
34 Name: count, dtype: int64
35
36 EXERANY2
37 1.0      74.68
38 2.0      25.32
39 Name: proportion, dtype: float64
40
41 EXERANY2  sex_cat
42 1.0      Male      561
43         Female     539
44 2.0      Female     201
45         Male       169
46 Name: count, dtype: int64
47
48 EXERANY2  sex_cat
49 1.0      Male      51.00

```

```
50           Female      49.00
51 2.0         Female      54.32
52           Male       45.68
53 Name: proportion, dtype: float64
54
55 mean - BMI by perf_phys_health
56 0.0      29.40
57 1.0      28.25
58 Name: bmi, dtype: float64
```