

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: mh = pd.read_csv("D:\Test\Mental Health.csv")
mh
```

Out[2]:

	state	floor	room	empcode	hostel	card_number	m_health	educ_standard	inc
0	Karnataka	0.0	6.0	2.021610e+07	Blue Hostel	6.001322e+07	15.0	7.0	37488.66
1	Karnataka	1.0	9.0	2.037530e+07	Red Hostel	6.022154e+07	NaN	8.0	36193.73
2	Karnataka	1.0	14.0	2.024461e+07	Blue Hostel	6.004956e+07	4.0	10.0	29309.51
3	Karnataka	0.0	10.0	2.023884e+07	Blue Hostel	6.004278e+07	7.0	10.0	37583.10
4	Karnataka	3.0	20.0	2.015258e+07	Red Hostel	5.243690e+05	10.0	10.0	35087.14
...	...	...	...	...	...	...	...	...	...
351	Odisha	4.0	409.0	2.255189e+07	Yellow Hostel	6.278446e+07	NaN	9.0	39899.77
352	Odisha	3.0	20.0	2.279604e+07	Red Hostel	6.248399e+07	10.0	10.0	28738.29
353	Odisha	1.0	13.0	2.247466e+07	Blue Hostel	6.249702e+07	8.0	9.0	38016.14
354	Odisha	3.0	19.0	2.278762e+07	Red Hostel	6.247238e+07	5.0	10.0	32688.77
355	Odisha	1.0	4.0	2.270211e+07	Purple Hostel	6.262206e+07	4.0	5.0	39513.51

356 rows × 10 columns



This Dataset has 356 Rows and 10 Columns.

```
In [4]: mh.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 356 entries, 0 to 355
Data columns (total 10 columns):
#   Column              Non-Null Count  Dtype
---  -
0   state                356 non-null    object
1   floor                297 non-null    float64
2   room                 297 non-null    float64
3   empcode              356 non-null    float64
4   hostel               297 non-null    object
5   card_number          356 non-null    float64
6   m_health              295 non-null    float64
7   educ_standard        333 non-null    float64
8   income               356 non-null    float64
```

```
9 treatment      356 non-null int64
dtypes: float64(7), int64(1), object(2)
memory usage: 27.9+ KB
```

```
In [5]: mh.isnull().sum()
```

```
Out[5]: state      0
floor      59
room       59
empcode    0
hostel     59
card_number 0
m_health   61
educ_standard 23
income     0
treatment  0
dtype: int64
```

In this Dataset some column has null values. Column named floor, room and hostel has 59 Null values each. and column named m\_health has 61 and educ\_standard has 23 Null Values.

```
In [6]: mh.describe()
```

```
Out[6]:
```

	floor	room	empcode	card_number	m_health	educ_standard	income
<b>count</b>	297.000000	297.000000	3.560000e+02	3.560000e+02	295.000000	333.000000	356.000000
<b>mean</b>	2.006734	96.888889	2.121494e+07	5.803929e+07	6.369492	9.573574	32353.089767
<b>std</b>	1.171143	137.566487	9.390945e+05	1.334187e+07	3.772766	1.592005	4689.723814
<b>min</b>	0.000000	3.000000	2.003698e+07	2.301300e+04	0.000000	4.000000	24043.711841
<b>25%</b>	1.000000	9.000000	2.037461e+07	6.022083e+07	3.000000	9.000000	28294.474742
<b>50%</b>	2.000000	17.000000	2.144678e+07	6.144782e+07	6.000000	10.000000	32146.225847
<b>75%</b>	3.000000	211.000000	2.176989e+07	6.174495e+07	9.000000	10.000000	36905.708410
<b>max</b>	4.000000	410.000000	2.279735e+07	6.279740e+07	16.000000	12.000000	39984.185647

From above data we can get brief information about columns in the Dataset which has numerical values. We can get count, mean, std, min, max, 50% etc from above table.

## Q1 - a. How many participants are there in the treatment group? And how many in the control group?

```
In [7]: mh['treatment'].value_counts()
```

```
Out[7]: 0    183
        1    173
        Name: treatment, dtype: int64
```

From above we can say that 183 participants are in treatment group and 173 participants are in Control Group.

b. Answer the below questions only for the “Red hostel” in “Karnataka” state.

In [8]:

```
Karnat = mh[mh['state'] == 'Karnataka']
Karnat
```

Out[8]:

	state	floor	room	empcode	hostel	card_number	m_health	educ_standard	incomm
0	Karnataka	0.0	6.0	20216096.0	Blue Hostel	60013218.0	15.0	7.0	37488.66700
1	Karnataka	1.0	9.0	20375302.0	Red Hostel	60221544.0	NaN	8.0	36193.73276
2	Karnataka	1.0	14.0	20244610.0	Blue Hostel	60049561.0	4.0	10.0	29309.51838
3	Karnataka	0.0	10.0	20238843.0	Blue Hostel	60042775.0	7.0	10.0	37583.10393
4	Karnataka	3.0	20.0	20152584.0	Red Hostel	524369.0	10.0	10.0	35087.14437
...	...	...	...	...	...	...	...	...	...
164	Karnataka	4.0	403.0	20303094.0	Yellow Hostel	60127953.0	2.0	10.0	27048.49939
165	Karnataka	1.0	4.0	20377231.0	Purple Hostel	60225398.0	3.0	8.0	24566.04248
166	Karnataka	2.0	205.0	20303113.0	Yellow Hostel	60128018.0	2.0	9.0	38261.34368
167	Karnataka	4.0	408.0	20215683.0	Yellow Hostel	60012705.0	11.0	4.0	38597.77665
168	Karnataka	NaN	NaN	22222322.0	NaN	60222450.0	NaN	12.0	34949.68937

169 rows × 10 columns

In [9]:

```
Red = Karnat[Karnat['hostel'] == 'Red Hostel']
Red
```

Out[9]:

	state	floor	room	empcode	hostel	card_number	m_health	educ_standard	incomm
1	Karnataka	1.0	9.0	20375302.0	Red Hostel	60221544.0	NaN	8.0	36193.73276
4	Karnataka	3.0	20.0	20152584.0	Red Hostel	524369.0	10.0	10.0	35087.14437
8	Karnataka	0.0	4.0	20036981.0	Red Hostel	10036971.0	8.0	9.0	38192.47222
10	Karnataka	3.0	19.0	20375223.0	Red Hostel	60221485.0	5.0	10.0	36616.47607
13	Karnataka	2.0	14.0	20229159.0	Red Hostel	60030738.0	9.0	NaN	29269.43431

	state	floor	room	empcode	hostel	card_number	m_health	educ_standard	incomm
15	Karnataka	3.0	20.0	20152684.0	Red Hostel	524323.0	5.0	10.0	24813.66391
17	Karnataka	2.0	17.0	20175589.0	Red Hostel	1016360.0	6.0	9.0	33340.41359
20	Karnataka	2.0	13.0	20036983.0	Red Hostel	750979.0	8.0	8.0	39335.03204
22	Karnataka	3.0	20.0	20152761.0	Red Hostel	524350.0	9.0	12.0	27545.16385
31	Karnataka	3.0	20.0	20152682.0	Red Hostel	524304.0	6.0	10.0	37252.13877
32	Karnataka	3.0	22.0	20152657.0	Red Hostel	524377.0	3.0	10.0	35145.68503
33	Karnataka	2.0	16.0	20175325.0	Red Hostel	1016356.0	2.0	10.0	32138.41430
42	Karnataka	1.0	11.0	20375335.0	Red Hostel	60221586.0	5.0	10.0	28007.62673
55	Karnataka	1.0	11.0	20375343.0	Red Hostel	60221559.0	2.0	8.0	38821.67788
72	Karnataka	0.0	6.0	20229148.0	Red Hostel	60030762.0	8.0	11.0	27702.64877
78	Karnataka	1.0	9.0	20375342.0	Red Hostel	60221559.0	0.0	7.0	33156.96097
94	Karnataka	1.0	11.0	20375369.0	Red Hostel	60221593.0	6.0	10.0	29126.32201
130	Karnataka	1.0	9.0	20375300.0	Red Hostel	60221644.0	3.0	10.0	31259.13092
133	Karnataka	3.0	19.0	20375296.0	Red Hostel	60221549.0	6.0	10.0	39524.52965
143	Karnataka	1.0	9.0	20375305.0	Red Hostel	60221662.0	3.0	10.0	25185.84383
144	Karnataka	3.0	19.0	20375304.0	Red Hostel	60221582.0	7.0	9.0	26336.52298
146	Karnataka	3.0	19.0	20375306.0	Red Hostel	60221571.0	3.0	10.0	32019.48639
148	Karnataka	1.0	9.0	20376088.0	Red Hostel	60221633.0	4.0	6.0	30186.93414
150	Karnataka	1.0	9.0	20376091.0	Red Hostel	60221671.0	5.0	10.0	26582.81284
158	Karnataka	2.0	16.0	20175414.0	Red Hostel	1016422.0	8.0	12.0	39596.96663
159	Karnataka	2.0	17.0	20263630.0	Red Hostel	60076510.0	8.0	9.0	37317.29630



## i. How many missing values are there in the m health variable? How can you deal with the missing values?

```
In [10]: Red.isnull().sum()
```

```
Out[10]: state          0
         floor          0
         room           0
         empcode        0
         hostel          0
         card_number     0
         m_health        1
         educ_standard    1
         income          0
         treatment       0
         dtype: int64
```

Only 1 missing value is there in column named m\_health and educ\_standard. As here we loss only 1 row so we can easily remove the row with missing values.

## ii. How many rooms are there in the hostel?

```
In [11]: Red.room.sum()
```

```
Out[11]: 368.0
```

There are 368 rooms in the hostel

iii. What is the minimum number of individuals in a room? Which room?

## iv. Which floor has the lowest number of rooms?

```
In [13]: rooms = Red.groupby('floor')['room'].sum()
         rooms
```

```
Out[13]: floor
         0.0    10.0
         1.0    87.0
         2.0    93.0
         3.0   178.0
         Name: room, dtype: float64
```

From above output it is clear that floor 0 has 10 rooms.

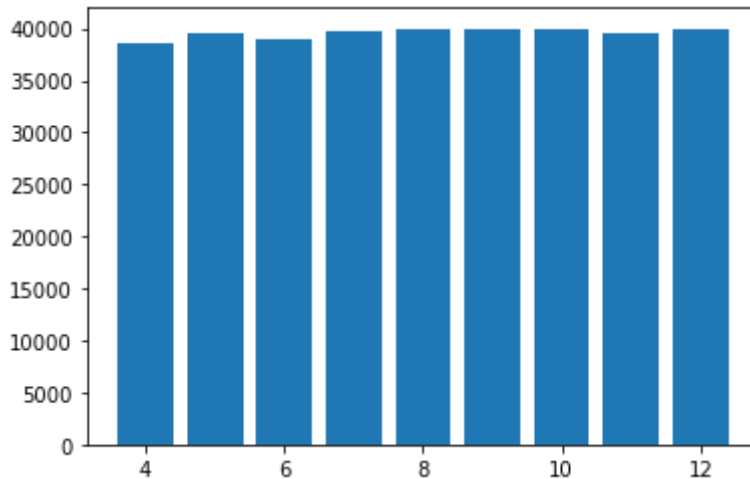
v. What is the average number of people per floor?

## c. One of the variables in this dataset records the participant's education level.

## i. Generate a relationship between income and education.

In [14]: `plt.bar(mh['educ_standard'],mh['income'])`

Out[14]: <BarContainer object of 356 artists>



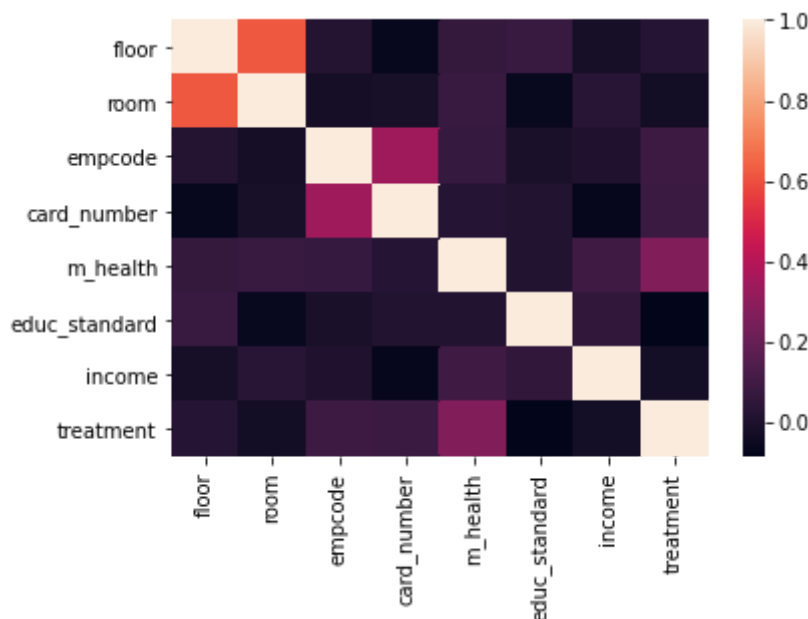
## ii. In the distribution, which level of education has the lowest average income

From above graph we can say that education level 4 has lowest average income.

d. Run any analysis you seem fit between mental health score and treatment indicator to study impact of the intervention. Clearly state the direction of the effect; whether the effect was significant; if yes, the level of significance.

In [16]: `sns.heatmap(mh.corr())`

Out[16]: <AxesSubplot:>



From above heatmap we can say that m\_health is correlated with educ\_standard , income and treatment.

In [ ]:

e. Now, your research manager asks you to control for income in the regression. The income can be classified into 3 categories - low (below 30,000), middle (30,000-35,000) and high (above 35,000)

We can create dummies variable of income and concat it with Dataset for regression as done below

In [17]:

```
income_cat = pd.cut(mh['income'], bins=[0,30000,35000,float('inf')], labels=['low',
```

In [18]:

```
income_dummies = pd.get_dummies(income_cat, prefix='income')
```

In [19]:

```
pd.concat([mh, income_dummies], axis=1)
```

Out[19]:

	state	floor	room	empcode	hostel	card_number	m_health	educ_standard	inc
0	Karnataka	0.0	6.0	2.021610e+07	Blue Hostel	6.001322e+07	15.0	7.0	37488.66
1	Karnataka	1.0	9.0	2.037530e+07	Red Hostel	6.022154e+07	NaN	8.0	36193.73
2	Karnataka	1.0	14.0	2.024461e+07	Blue Hostel	6.004956e+07	4.0	10.0	29309.51
3	Karnataka	0.0	10.0	2.023884e+07	Blue Hostel	6.004278e+07	7.0	10.0	37583.10
4	Karnataka	3.0	20.0	2.015258e+07	Red Hostel	5.243690e+05	10.0	10.0	35087.14
...	...	...	...	...	...	...	...	...	...
351	Odisha	4.0	409.0	2.255189e+07	Yellow Hostel	6.278446e+07	NaN	9.0	39899.77
352	Odisha	3.0	20.0	2.279604e+07	Red Hostel	6.248399e+07	10.0	10.0	28738.29
353	Odisha	1.0	13.0	2.247466e+07	Blue Hostel	6.249702e+07	8.0	9.0	38016.14
354	Odisha	3.0	19.0	2.278762e+07	Red Hostel	6.247238e+07	5.0	10.0	32688.77
355	Odisha	1.0	4.0	2.270211e+07	Purple Hostel	6.262206e+07	4.0	5.0	39513.51

356 rows × 13 columns



Now we have 356 Rows and 13 Columns in Dataset.

In [26]:

```
mh.to_csv('Mental Health Final.csv')
```

```
In [27]: mh.to_excel('Mental Health Final.xlsx')
```

```
In [ ]:
```