

3MTT Cohort 1 Data Analysis

Project Topic:

**Exploring the relationship between Occupational Factors
and health indicators among worker**

By

Hassana Mohammed Shuaibu

(FE/23/46142980)

INTRODUCTION

In today's work environments, understanding the relationship between occupational factors and health indicators is crucial for promoting the well-being of workers.

Problem statement

The prevalence of health issues among workers leads to the need to identify and address occupational factors influencing their well-being.

Aims and objective:

This study aims to investigate the relationship between occupational factors and health indicators.

1. Analyze the relationship between various occupational factors (such as stress level, sleep quality, physical activity level, BMI category, and sleep disorder)
2. Identify trends and patterns in health outcomes across different occupations.
3. Provide insights that can inform the development of targeted interventions and workplace health initiatives aimed at improving health and well-being of workers across diverse professions.

About the dataset

Name: Health and lifestyle dataset

Source: The dataset was collected from online source.

Size: The dataset is a 374*13 and has a total of 4863 data entries

Jupyter 3mttProject Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted | Python 3 (ipykernel)

In [17]: data

Out[17]:

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
...
369	370	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
370	371	Female	59	Nurse	8.0	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
371	372	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
372	373	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea
373	374	Female	59	Nurse	8.1	9	75	3	Overweight	140/95	68	7000	Sleep Apnea

374 rows x 13 columns

This is the view of the dataset

Dataset info:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 374 entries, 0 to 373
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Person ID                            374 non-null    int64
1   Gender                               374 non-null    object
2   Age                                  374 non-null    int64
3   Occupation                           374 non-null    object
4   Sleep Duration                       374 non-null    float64
5   Quality of Sleep                     374 non-null    int64
6   Physical Activity Level               374 non-null    int64
7   Stress Level                         374 non-null    int64
8   BMI Category                         374 non-null    object
9   Blood Pressure                       374 non-null    object
10  Heart Rate                           374 non-null    int64
11  Daily Steps                          374 non-null    int64
12  Sleep Disorder                       374 non-null    object
dtypes: float64(1), int64(7), object(5)
```

The dataset info provides valuable insights into dtypes, column names, non null counts.

Data exploration and cleaning

The next step after accessing your data is the exploration and cleaning. Check for missing data in the data set.

```
In [26]: data.isna().sum()
Out[26]: Person ID      0
         Gender        0
         Age           0
         Occupation    0
         Sleep Duration 0
         Quality of Sleep 0
         Physical Activity Level 0
         Stress Level   0
         BMI Category   0
         Blood Pressure 0
         Heart Rate     0
         Daily Steps    0
         Sleep Disorder 0
         dtype: int64
```

From the table above our dataset does not have any missing values.

Select visualization type(s)

Select data visualization types that will help you understand and explain the data.

Now that we know the data columns to use, it is time to decide which data visualization makes the most sense for EDA of the dataset. What type of data visualization(s) would be most helpful? Consider the distribution of the data.

The following chart would be used in the visualization of the data

- Line graph
- Bar chart
- Histogram
- Heat map
- Scatter plot

THE CONSTRUCTION STAGE

Correlation analysis

Explore the relationships between health indicators using correlation matrices.

Correlation matrix is a table showing correlation coefficients between variables. Each cell in the table represents the correlation between two variables, with values from -1 to +1.

A value of 1 indicates perfect positive correlation while the -1 indicates negative correlation.

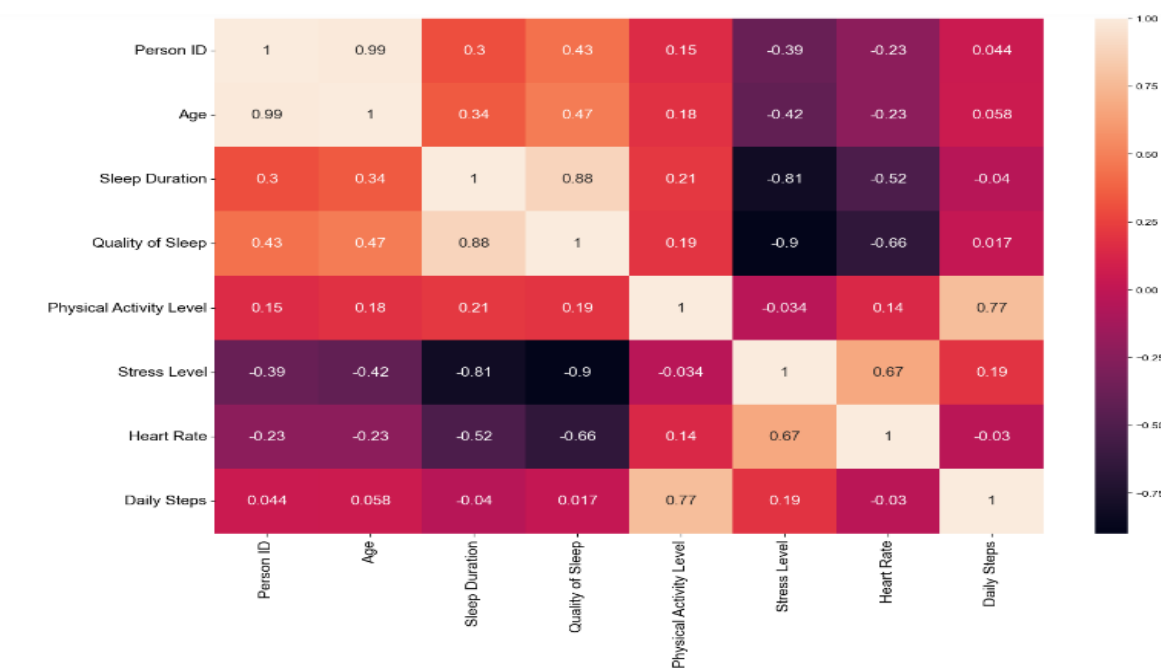
```
In [15]: data.corr()
```

```
Out[15]:
```

	Person ID	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate	Daily Steps
Person ID	1.000000	0.990516	0.296305	0.431612	0.149882	-0.394287	-0.225467	0.043844
Age	0.990516	1.000000	0.344709	0.473734	0.178993	-0.422344	-0.225606	0.057973
Sleep Duration	0.296305	0.344709	1.000000	0.883213	0.212360	-0.811023	-0.516455	-0.039533
Quality of Sleep	0.431612	0.473734	0.883213	1.000000	0.192896	-0.898752	-0.659865	0.016791
Physical Activity Level	0.149882	0.178993	0.212360	0.192896	1.000000	-0.034134	0.136971	0.772723
Stress Level	-0.394287	-0.422344	-0.811023	-0.898752	-0.034134	1.000000	0.670026	0.186829
Heart Rate	-0.225467	-0.225606	-0.516455	-0.659865	0.136971	0.670026	1.000000	-0.030309
Daily Steps	0.043844	0.057973	-0.039533	0.016791	0.772723	0.186829	-0.030309	1.000000

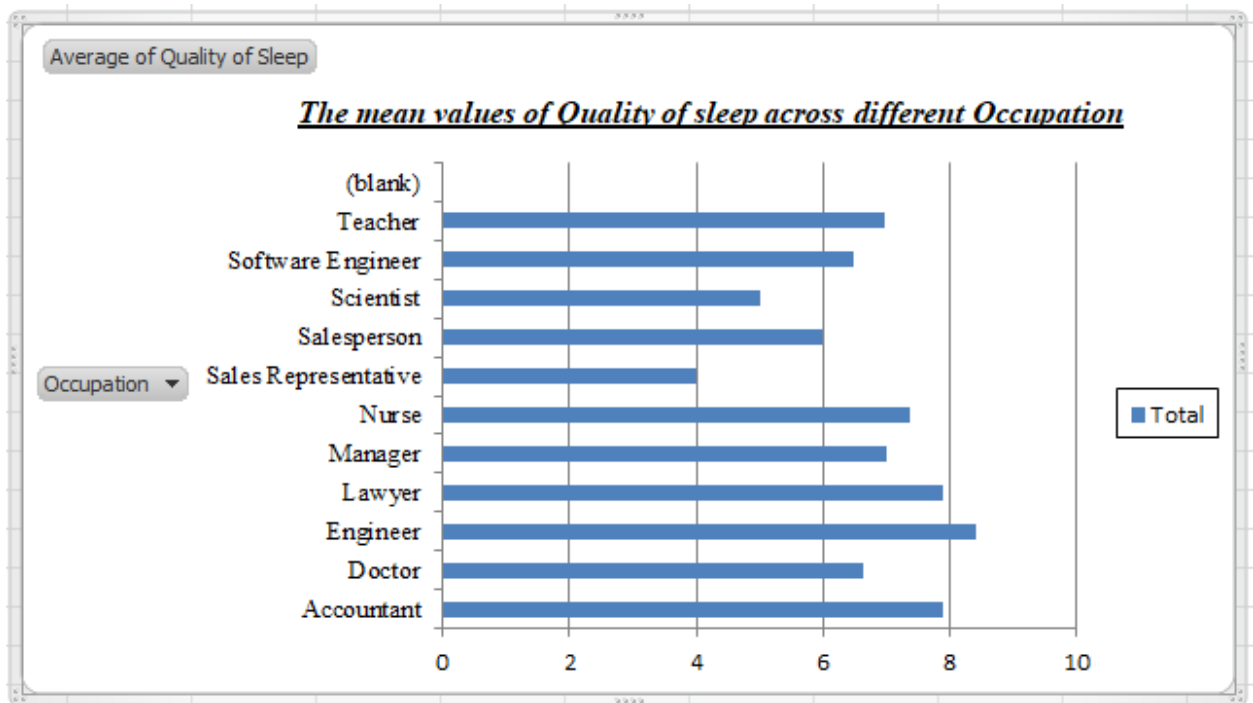
From the table we say that the quality of sleep and sleep duration has 0.88 which means they are positively correlated. Also the same thing with stress level with Heart rate which is 0.67, physical activity level with daily steps.

Below is the visualize correlations using heatmap to identify strong association between different health indicators.



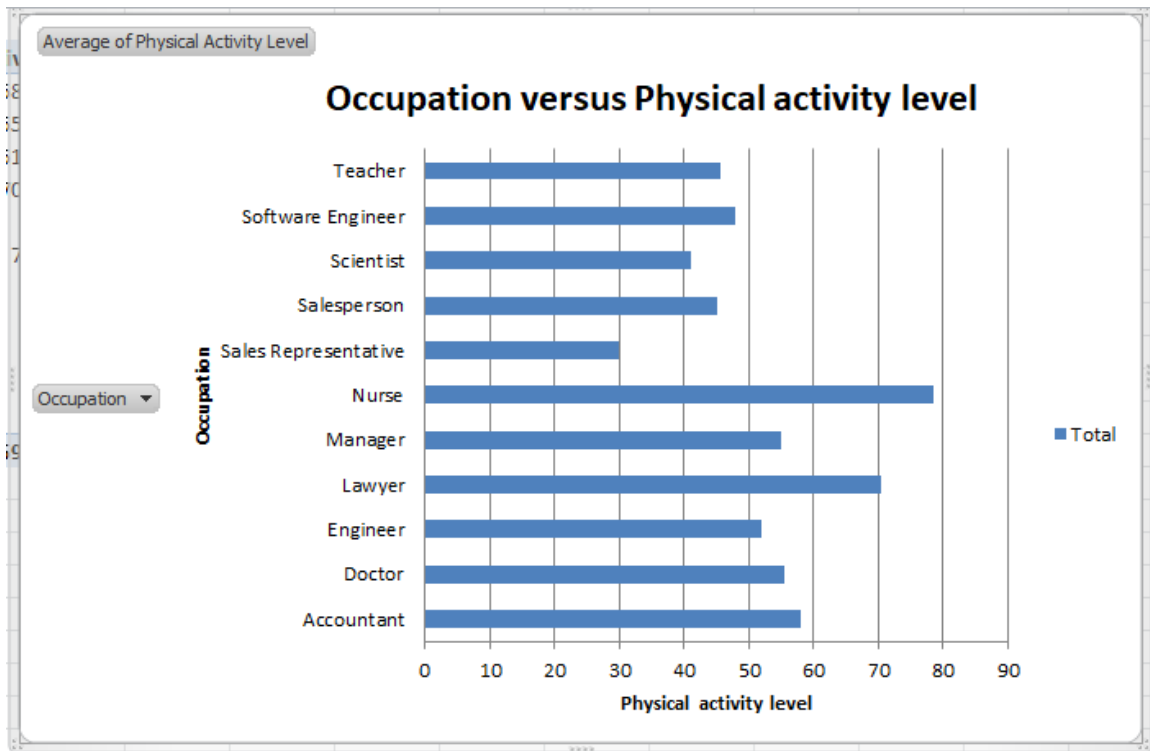
Comparison of Health indicators across occupation

1. Quality of sleep



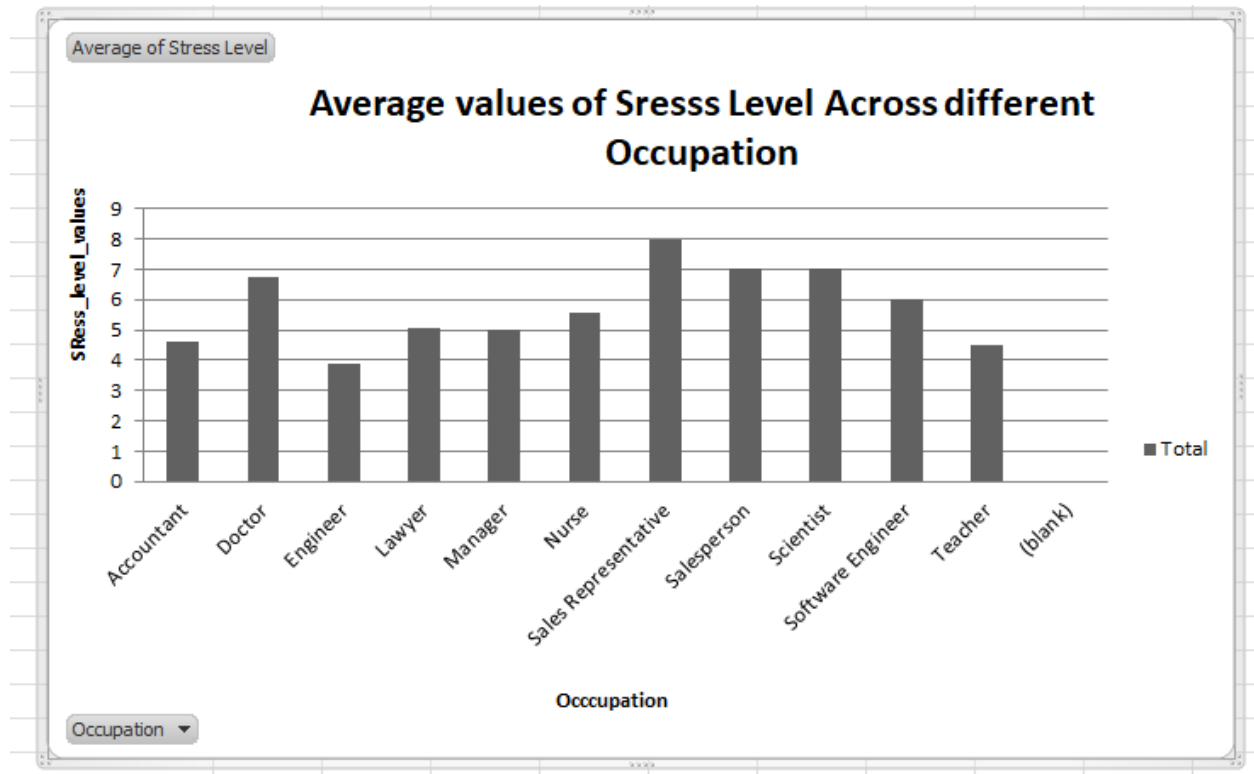
The bar plots show the mean comparison of quality of sleep across the different occupation with Engineering having the highest value of quality of sleep while sales representative has the lowest quality of sleep.

2. Physical Activity level:



The bar plots show the mean comparison of physical activity level across the different occupation with Nursing having the highest value of stress level while sales representative has lowest physical activity level.

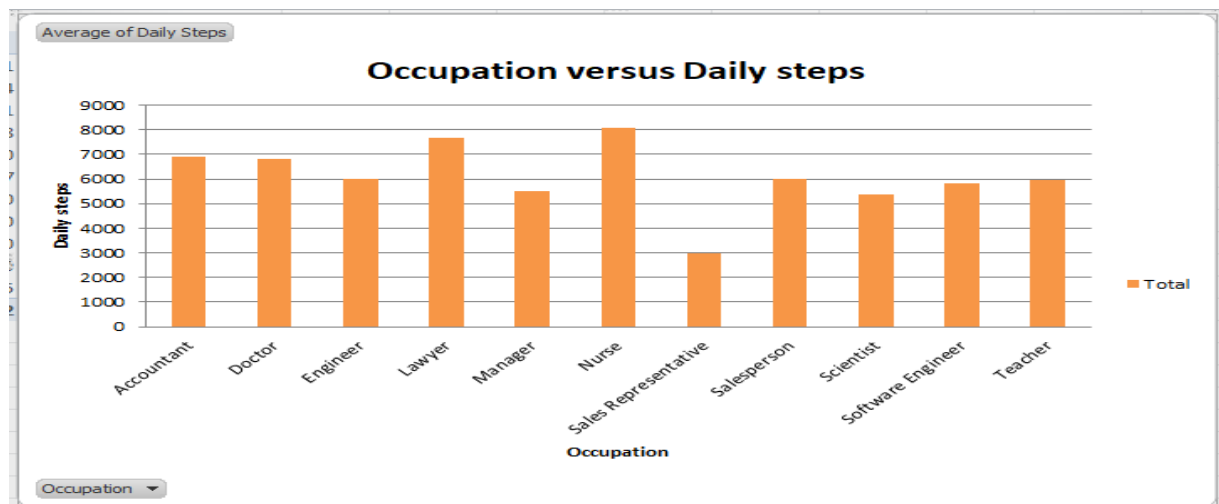
3. Stress Level



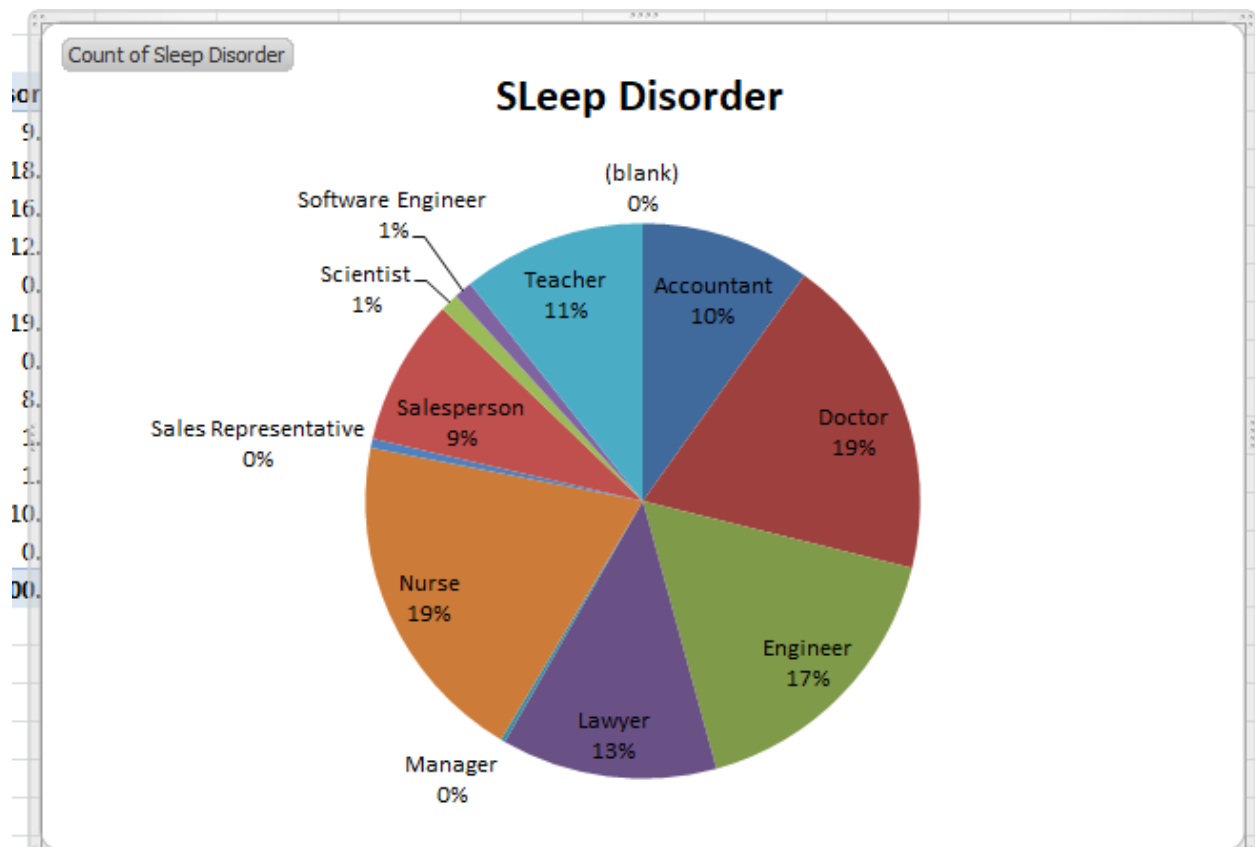
The bar plots show the mean comparison of physical activity level across the different occupation with sales representative having the highest value of stress level while Engineer has lowest stress level.

Comparative analysis of physical activity levels

Daily steps across different occupation



Sleep disorder across different occupation



This is the chart representing different sleep disorder for different occupation.

RESULTS:

Occupation plays role in determining health outcomes, with some occupations associated with higher stress levels and poorer sleep quality.

- Physical activity level was positively correlated with better health outcomes in which the Nurse has high value of activity level
- Stress level emerge as a significant factor influenced by Sleep duration and Quality of sleep
- Stress level were also found prevalent across various occupation

SUMMARY AND CONCLUSION

This analysis aimed to understand the health status of workers across different occupations and identify factors influencing their well-being. We found significant variations in health indicators, including quality of sleep, physical activity level, stress level and BMI category across occupations. Certain occupations exhibited higher level of stress and poorer sleep quality, while others shows healthier status.