

# DSC630\_KristaKnuckey\_Week1

April 5, 2025

```
[3]: #DSC630 Week 1
     #Krista Knuckey
```

```
[24]: #Dataset Summary:
     #The dataset, Sleep Health and Lifestyle Dataset, covers a wide variety of
     ↳ information that pertains to sleep and lifestyle habits. Each
     #individual that participated in this was given an ID and collected many data
     ↳ points, including gender, age, occupation, sleep duration, quality
     #of sleep, physical activity level, stress level, BMI category, blood pressure,
     ↳ heart rate, daily steps, and sleep disorder. After doing a brief
     #overview of what the dataset includes I would like to explore these two
     ↳ questions visually:
     #1) What occupation has the highest stress level?
     #2) Based on the quality of sleep, what is the average heart rate?
```

```
[25]: #Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[8]: #Upload/ preview dataset

sleep_data = pd.read_csv('sleepdata.csv')
sleep_data.head()
```

```
[8]:
```

	Person ID	Gender	Age	Occupation	Sleep Duration	\
0	1	Male	27	Software Engineer	6.1	
1	2	Male	28	Doctor	6.2	
2	3	Male	28	Doctor	6.2	
3	4	Male	28	Sales Representative	5.9	
4	5	Male	28	Sales Representative	5.9	

	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	\
0	6	42	6	Overweight	
1	6	60	8	Normal	
2	6	60	8	Normal	

3	4	30	8	Obese
4	4	30	8	Obese

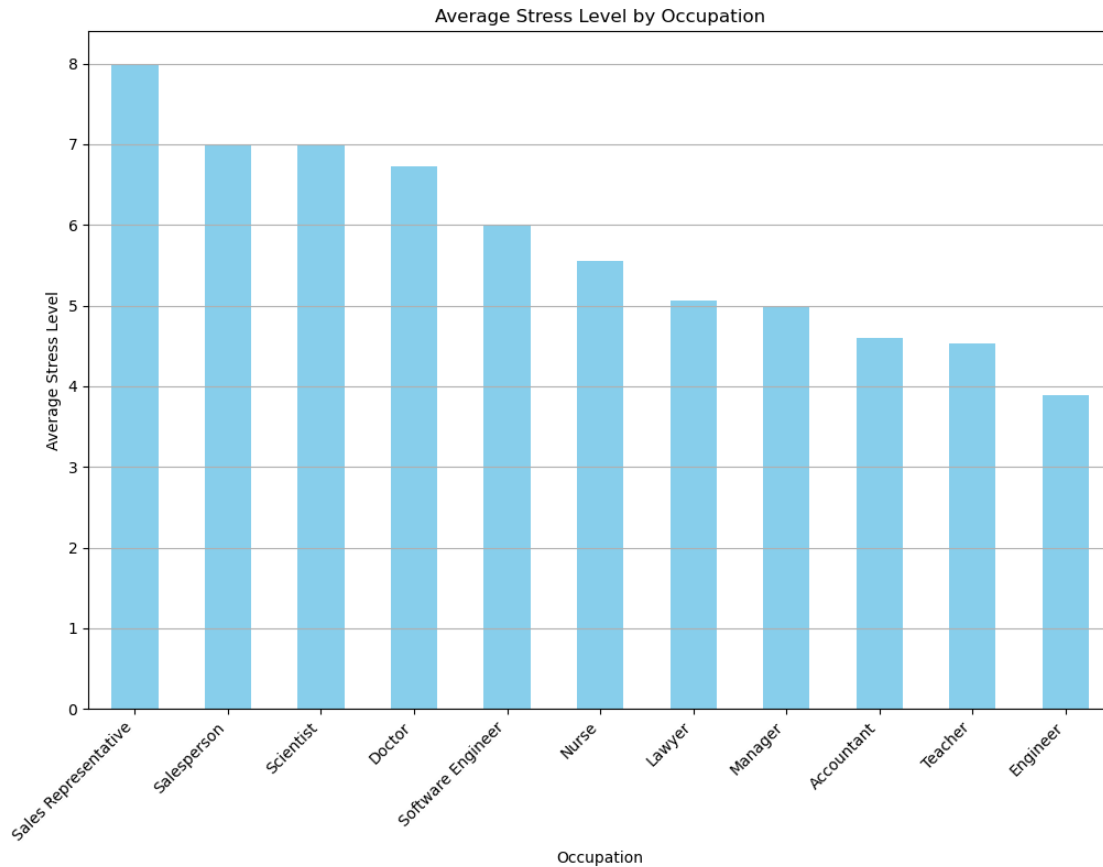
	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	126/83	77	4200	NaN
1	125/80	75	10000	NaN
2	125/80	75	10000	NaN
3	140/90	85	3000	Sleep Apnea
4	140/90	85	3000	Sleep Apnea

```
[10]: #Create a histogram or bar graph of the data- What occupation has the highest
      ↪ stress level?
```

```
occupation_stress = sleep_data.groupby('Occupation')['Stress Level'].mean().
      ↪sort_values(ascending=False)
```

```
plt.figure(figsize=(12, 8))
occupation_stress.plot(kind='bar', color='skyblue')
plt.title('Average Stress Level by Occupation')
plt.xlabel('Occupation')
plt.ylabel('Average Stress Level')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y')

plt.show()
```

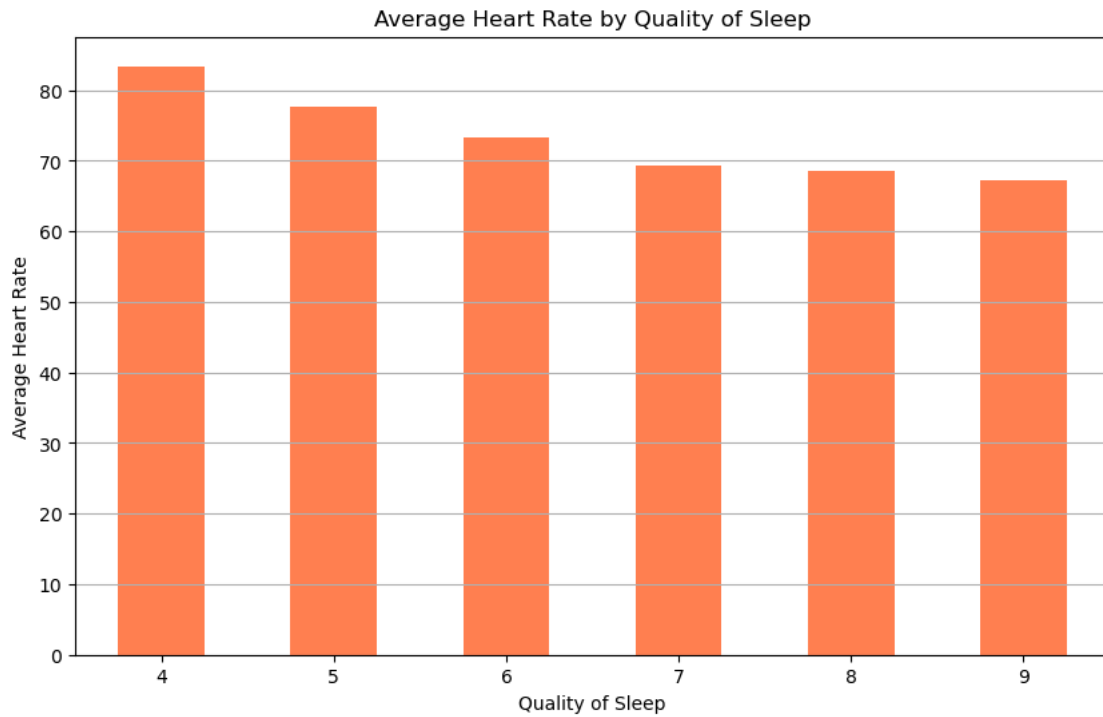


```
[11]: #Create a histogram or bar graph of the data- Based on the quality of sleep,
      ↳ what is the average heart rate?

sleep_quality_heart_rate = sleep_data.groupby('Quality of Sleep')['Heart Rate'].
      ↳ mean()

plt.figure(figsize=(10, 6))
sleep_quality_heart_rate.plot(kind='bar', color='coral')
plt.title('Average Heart Rate by Quality of Sleep')
plt.xlabel('Quality of Sleep')
plt.ylabel('Average Heart Rate')
plt.xticks(rotation=0)
plt.grid(axis='y')

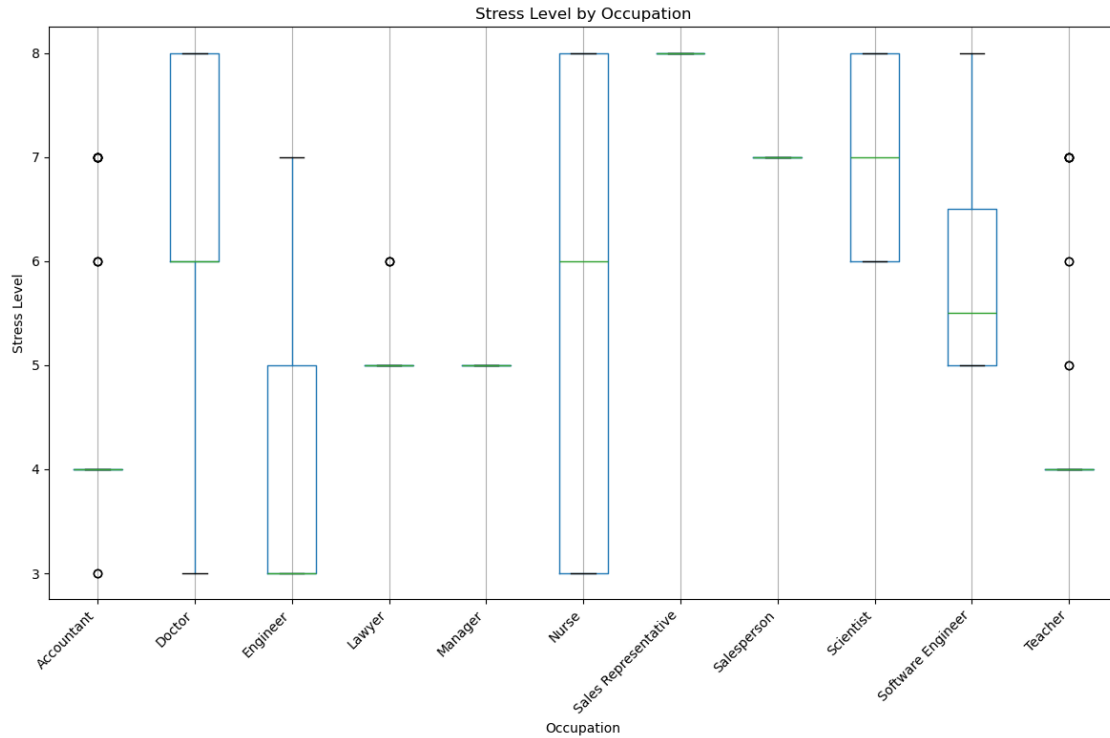
plt.show()
```



[15]: *#Create a boxplot from the data- What occupation has the highest stress level?*

```
plt.figure(figsize=(14, 8))
ax = plt.gca()
sleep_data.boxplot(column='Stress Level', by='Occupation', ax=ax)
plt.title('Stress Level by Occupation')
plt.suptitle('')
plt.xlabel('Occupation')
plt.ylabel('Stress Level')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y')

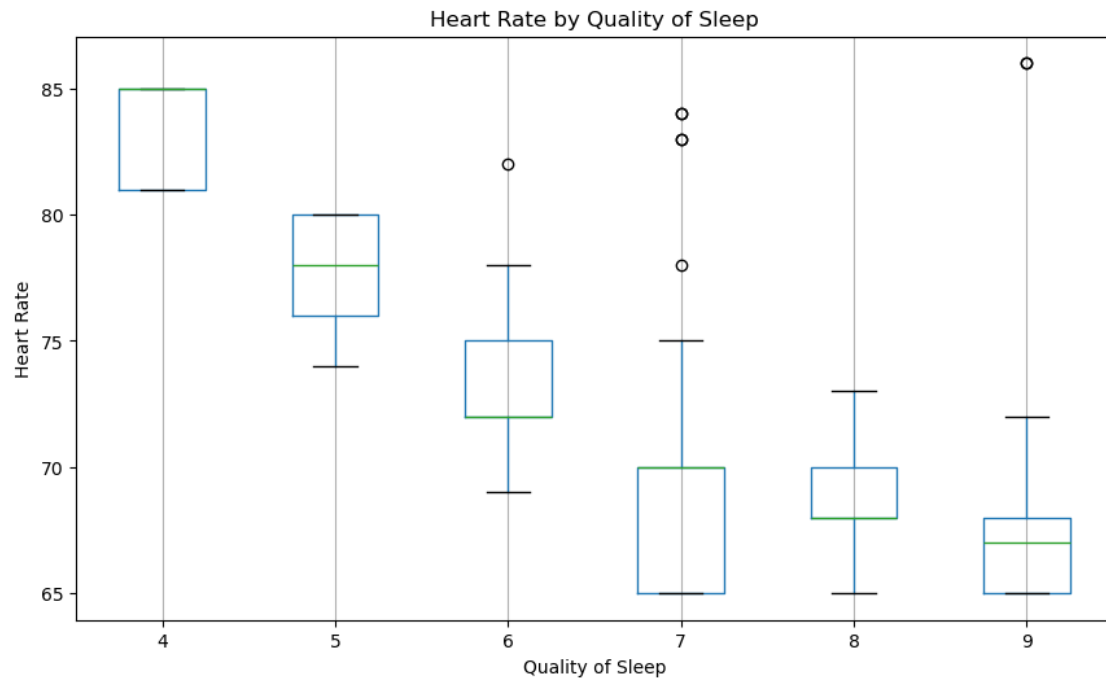
plt.show()
```



[23]: *#Create a boxplot from the data-Based on the quality of sleep, what is the average heart rate?*

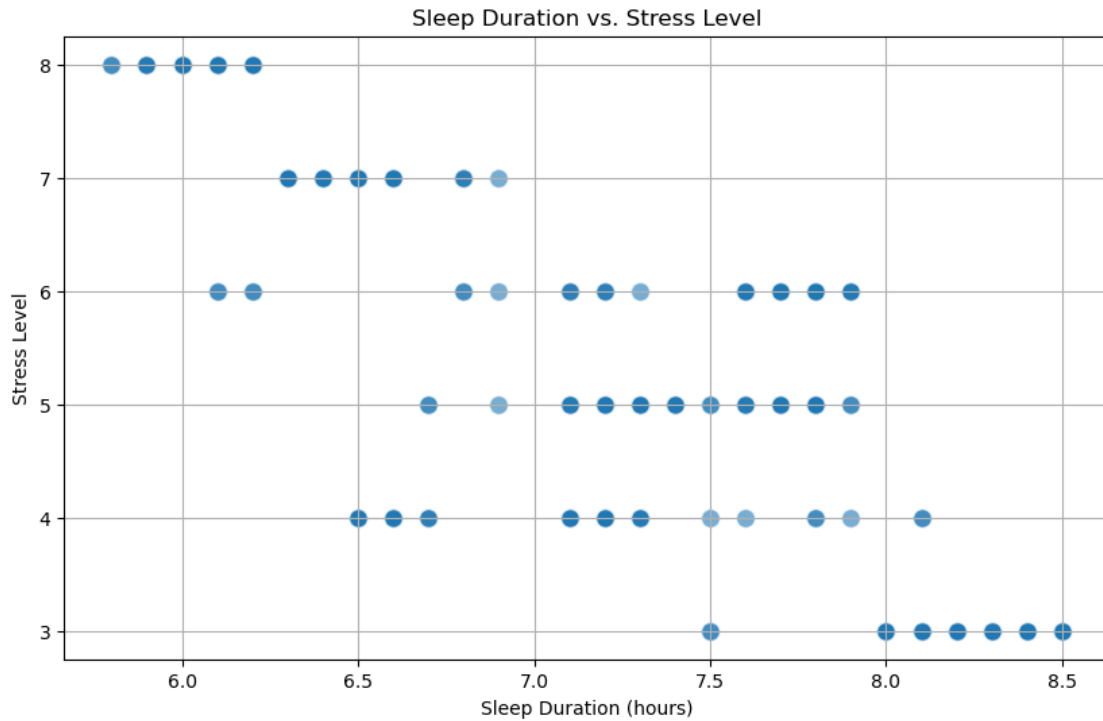
```
plt.figure(figsize=(10, 6))
ax = plt.gca()
sleep_data.boxplot(column='Heart Rate', by='Quality of Sleep', ax=ax)
plt.title('Heart Rate by Quality of Sleep')
plt.suptitle('')
plt.xlabel('Quality of Sleep')
plt.ylabel('Heart Rate')
plt.xticks(rotation=0)
plt.grid(axis='y')

plt.show()
```



```
[17]: #Create a bivariate plot from your data
plt.figure(figsize=(10, 6))
plt.scatter(sleep_data['Sleep Duration'], sleep_data['Stress Level'], alpha=0.
           ↪6, edgecolors='w', s=100)
plt.title('Sleep Duration vs. Stress Level')
plt.xlabel('Sleep Duration (hours)')
plt.ylabel('Stress Level')
plt.grid(True)

plt.show()
```



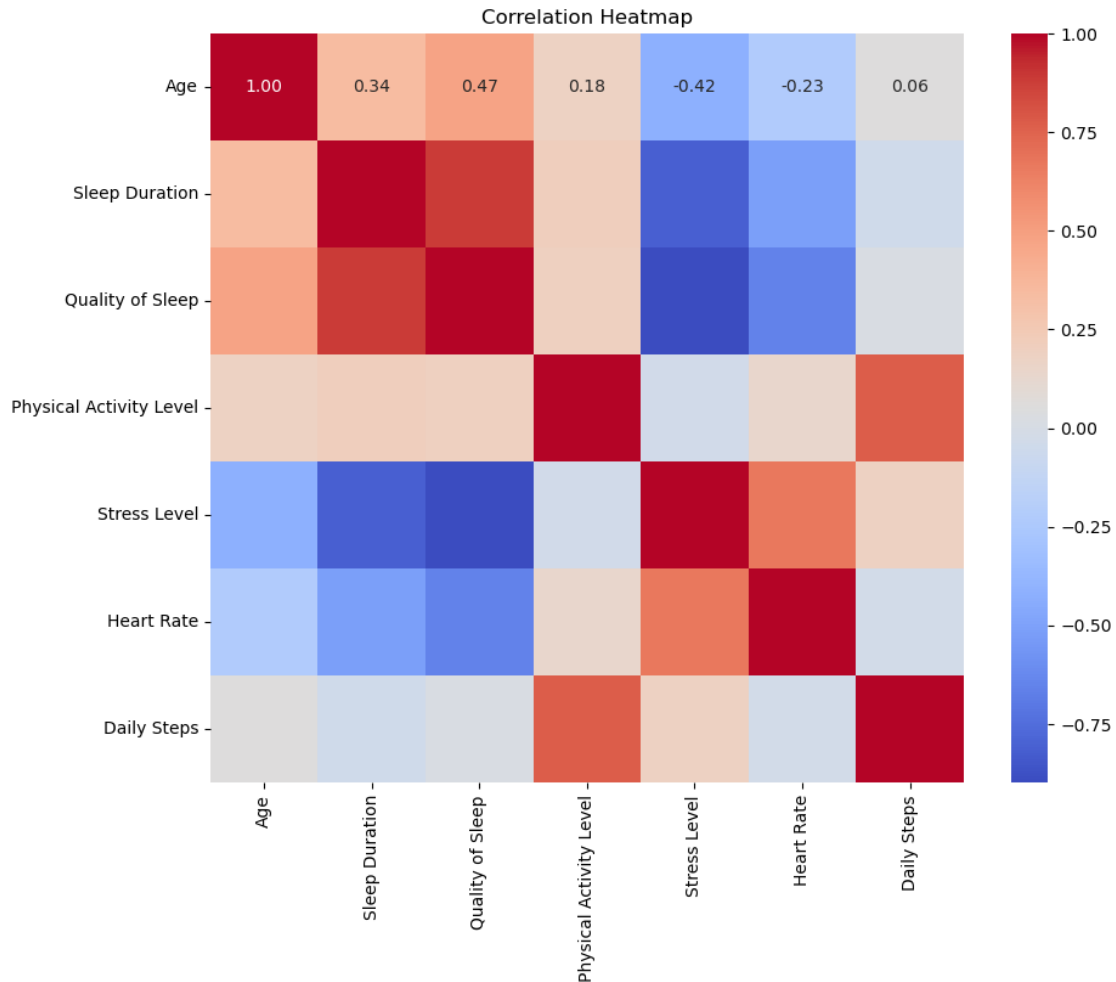
```
[22]: #Create additional visualizations- heatmap

numerical_columns = ['Age', 'Sleep Duration', 'Quality of Sleep', 'Physical_
↳Activity Level', 'Stress Level', 'Heart Rate', 'Daily Steps']
numerical_data = sleep_data[numerical_columns]

correlation_matrix = numerical_data.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')

plt.show()
```



[ ]: *#Conclusion/ Results:*

*#With the dataset provided, there were some surprising results that were given, based on the visualizations. First, we can see that the stress levels of sales workers is much higher compared to those that require more advanced degrees, like doctor, lawyer, and engineer. Also, we can see that the lower amount of hours slept (4 hours) there is a higher heart rate (average of 80). This lets us know that the amount of sleep has direct relation to our health. Also, another interesting point is that those who slept the least also reported higher stress levels. Stress is a huge indication of health, which those who had an average of 8 hours of sleep reported the least amount of stress.*

*#Citation*

*#Sleep health and lifestyle dataset. Kaggle. (2023, September 18). <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>*