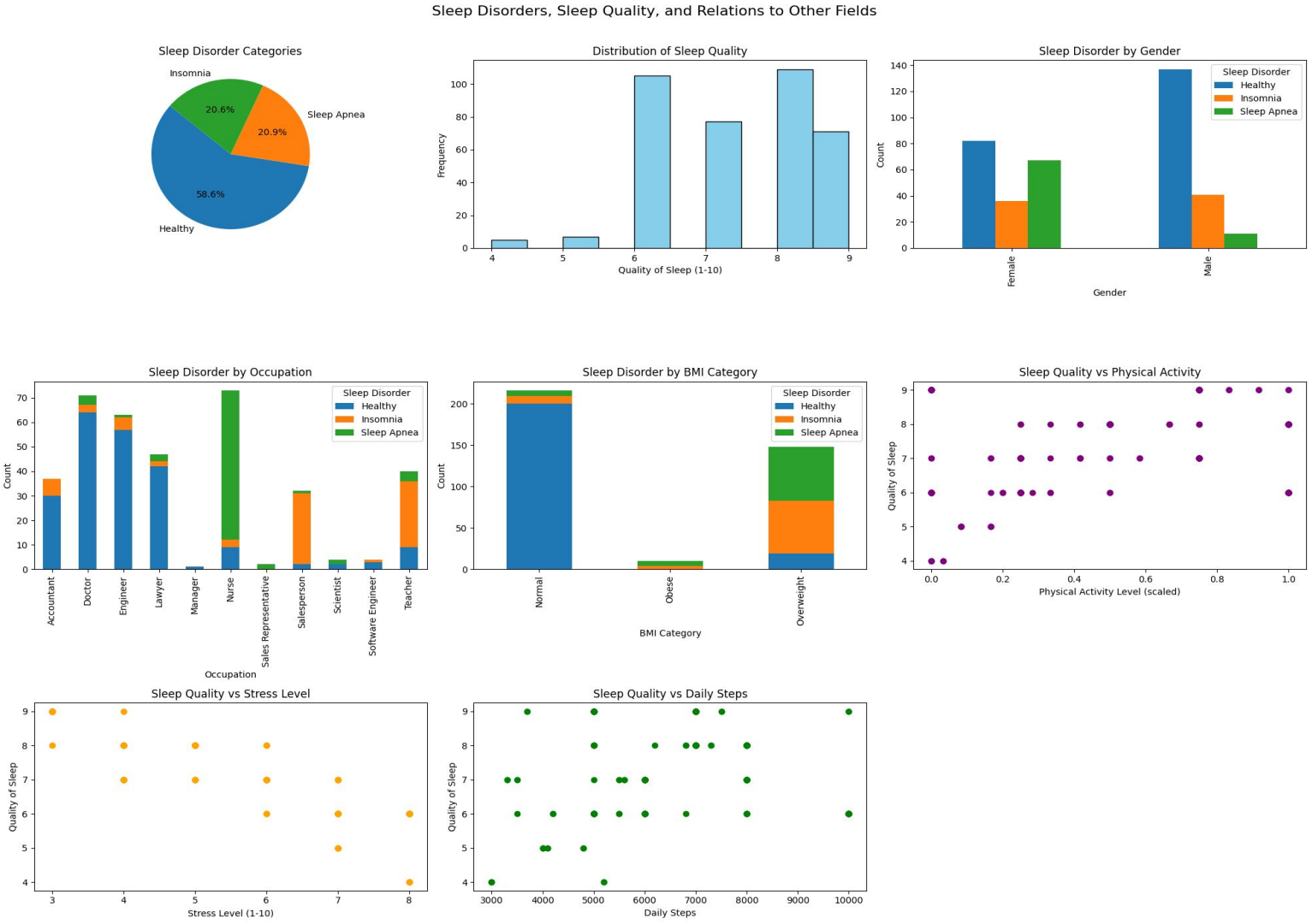
An Analysis of Sleep; The Data Behind Disorders

Daniel Grammer | Emily Hoffman | Isabelle Kramar | John Lin | Rachel Nefesh

The Dataset: An Overview



Quick Note: Data Cleaning

One-Hot Encoded:

- **Gender, Occupation, BMI Category, and Sleep Disorder (both T/F and individual types)**

Min-Max Scaled:

- **Sleep Duration, Physical Activity Level, Age, and Heart Rate**

Custom Encoding:

- ***Blood Pressure:***

Low	Normal	Elevated	High
Systolic < 90 & Diastolic < 60	90 ≥ Systolic < 120 & 60 ≥ Diastolic < 80	120 ≥ Systolic < 140 & 80 ≥ Diastolic < 90	Systolic ≥ 140 & Diastolic ≥ 90

Goals:

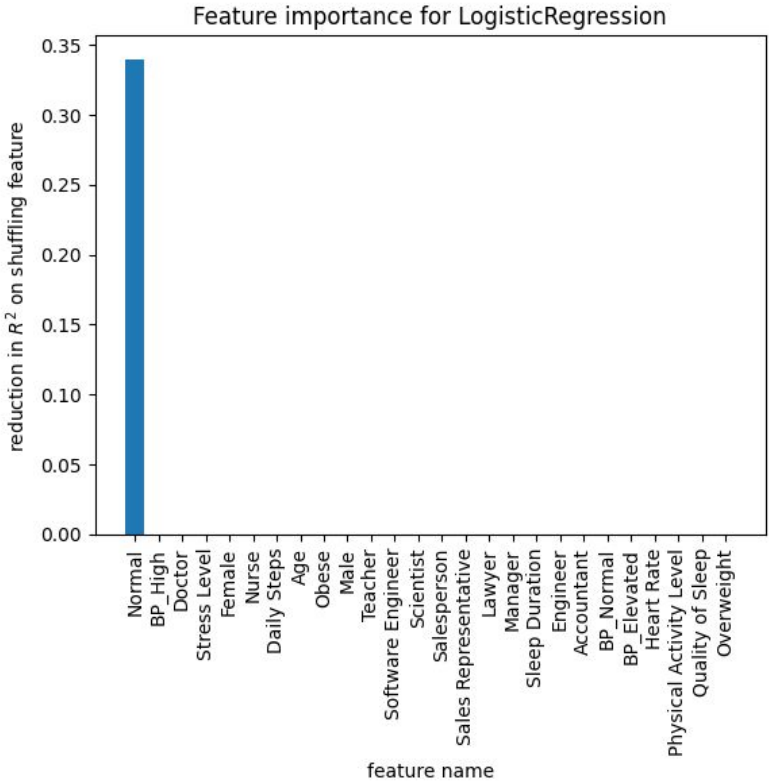
1. **What lifestyle and health factors are the best indicators of a sleep disorder?**
    a. ***Goal:*** **Identify conditions under which a person should be screened for sleep disorders.**

2. **What lifestyle and health factors translate to better quality of sleep?**
    a. ***Goal:*** **Identify lifestyle changes that could lead to a better quality of sleep.**

Question One - Variable Selection

Important Features:

- Age
- Heart Rate
- Daily Steps
- Sleep Duration
- Engineer
- Manager
- Normal
- BP_High
- Stress Level
- Female
- Nurse
- Doctor



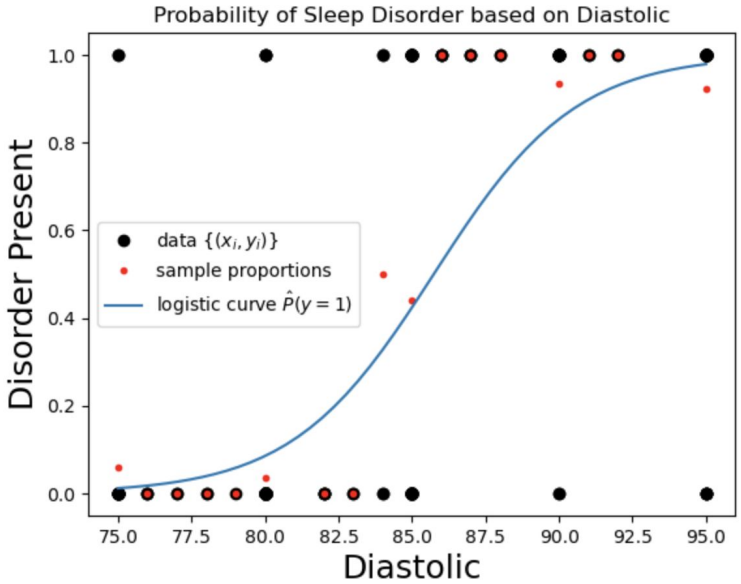
Feature Selection Specs:

- Decision Tree:
  - Max_depth = 5
- LASSO Logistic Regression:
  - C = 0.25

Question 1 Analysis: What lifestyle and health factors are the best indicators of a sleep disorder?

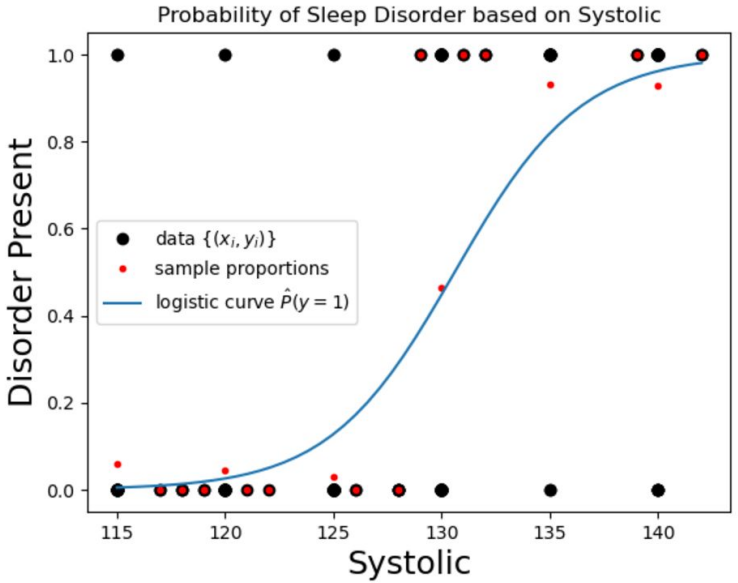
Testing Accuracy: .839

```
intercept=[-35.45430345], slope=[0.41350742], training score=0.839572192513369
<matplotlib.legend.Legend at 0x7fa1b2f08100>
```



Testing Accuracy: .834

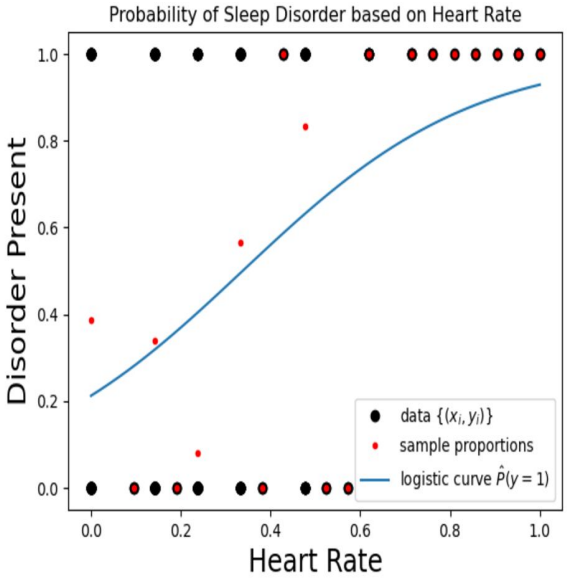
```
intercept=[-44.83087139], slope=[0.34326166], training score=0.8342245989304813
<matplotlib.legend.Legend at 0x14d232900>
```



Question 1 Analysis: What lifestyle and health factors are the best indicators of a sleep disorder?

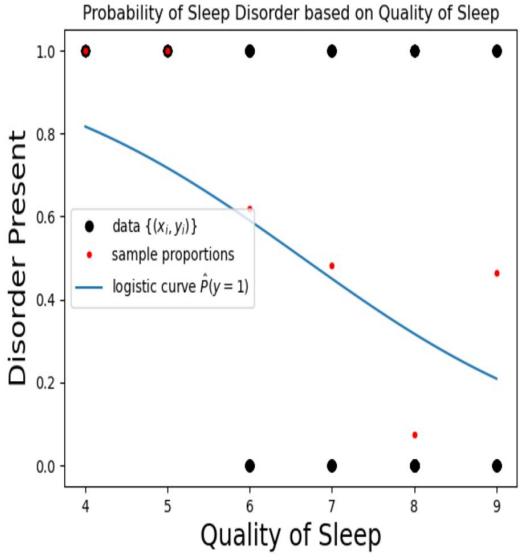
Testing Accuracy: .693

```
intercept=[-1.31306243], slope=[3.88300338], training score=0.6925133689839572
<matplotlib.legend.Legend at 0x14d04b110>
```



Testing Accuracy: .684

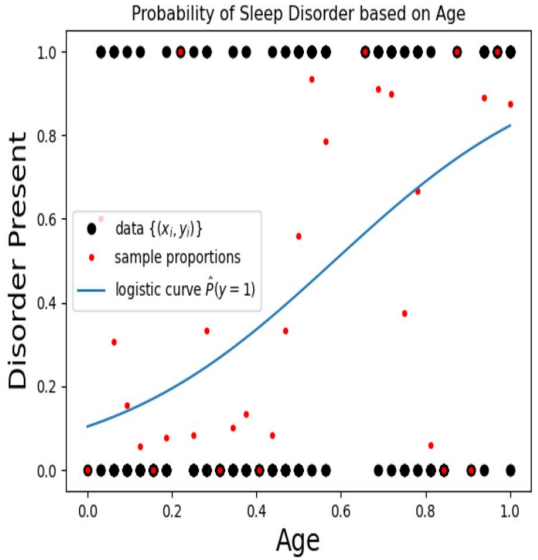
```
intercept=[3.75279173], slope=[-0.56504567], training score=0.6844919786096256
<matplotlib.legend.Legend at 0x14c6b32c0>
```



*** Quality of Sleep is measured on a subjective rating scale**

Testing Accuracy: .671

```
intercept=[-2.1615533], slope=[3.69493605], training score=0.6711229946524064
<matplotlib.legend.Legend at 0x14d7889b0>
```



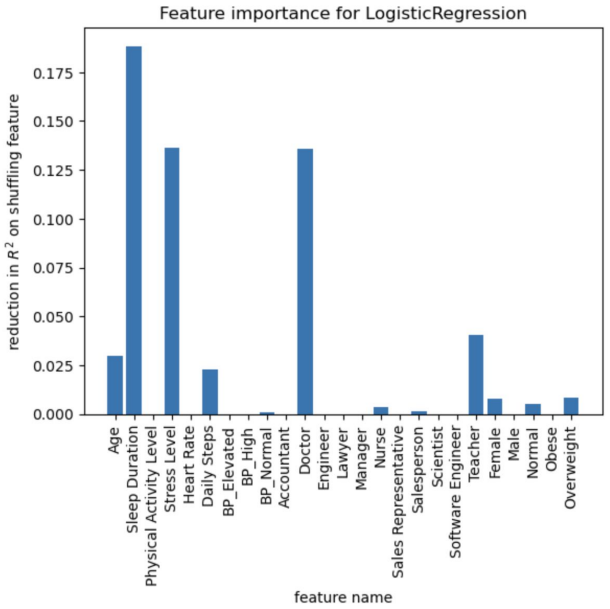
Question 1 - Conclusion

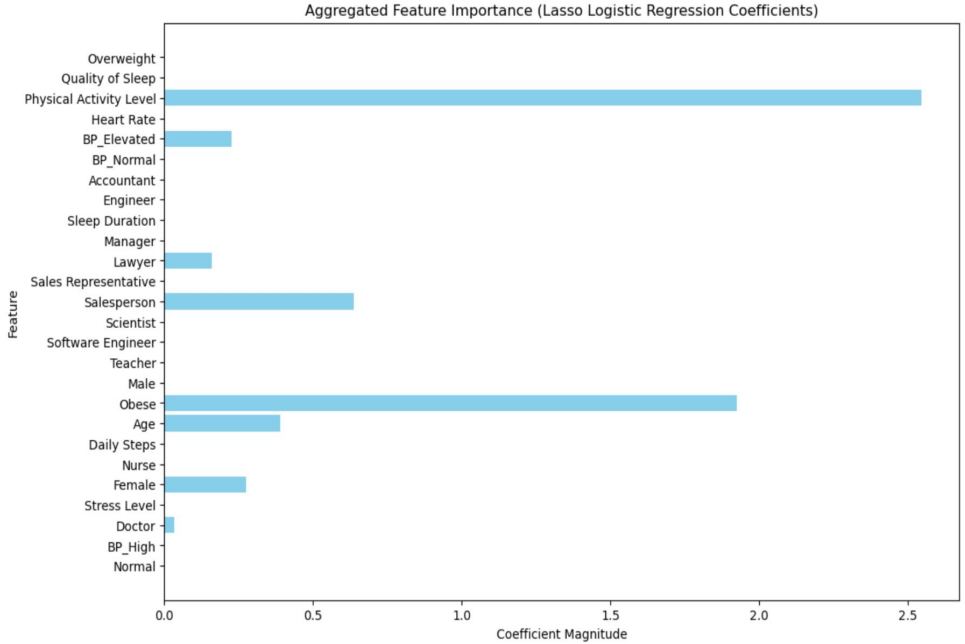
The following factors proved statistically significant:

Diastolic → Score: 0.84
Systolic → Score: 0.83
Heart Rate → Score: 0.69
Quality of Sleep → Score 0.68
Age → Score: 0.67
With all Lifestyle/Health factors → Score: 0.90

If you have high blood pressure, high heart rate, a poor self-report of quality of sleep, and are older, you should consider undergoing further sleep disorder testing.

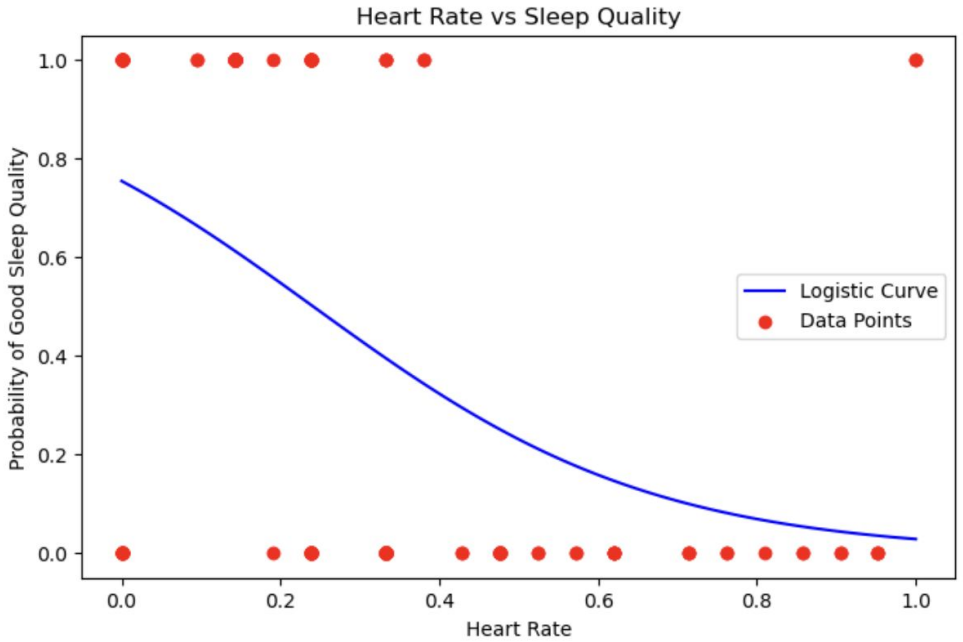
Quality of Sleep Analysis - Variable Selection



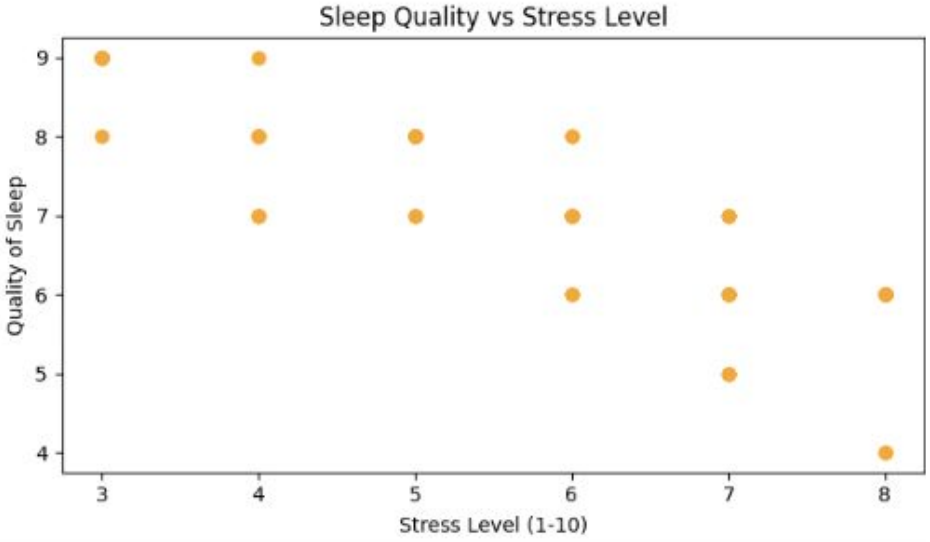


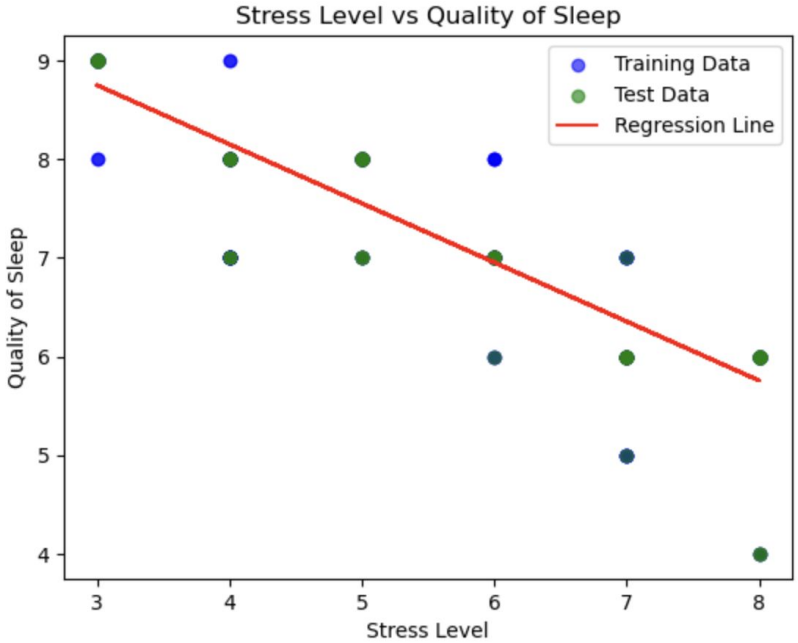
Question 2 Analysis - Sleep Quality vs. Heart Rate

Logistic Regression
Test accuracy: 0.760



Question 2 Analysis - Sleep Quality vs. Stress Level





Linear Regression

- Mean squared error: 0.260
- R^2 score: 0.828

Question 2 Analysis - Sleep Quality vs. Occupation

Logistic Regression

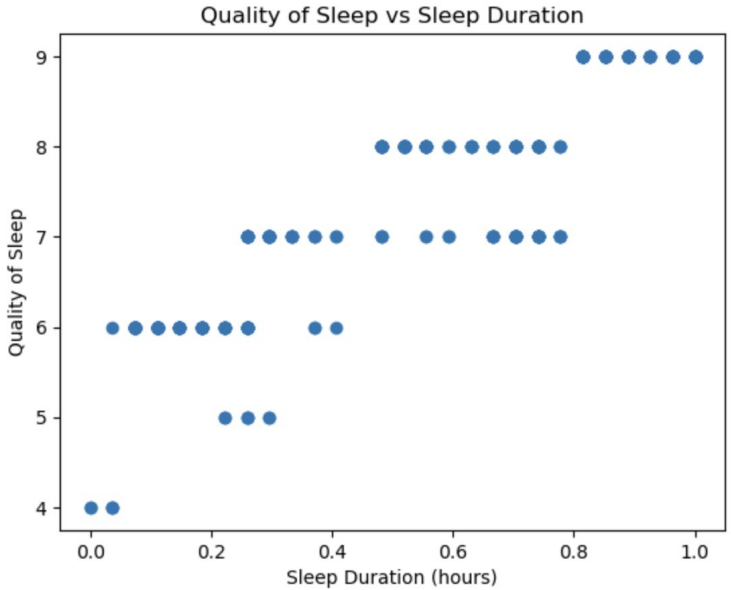
- Sleep quality >= 8 considered good
- Test accuracy: 0.827

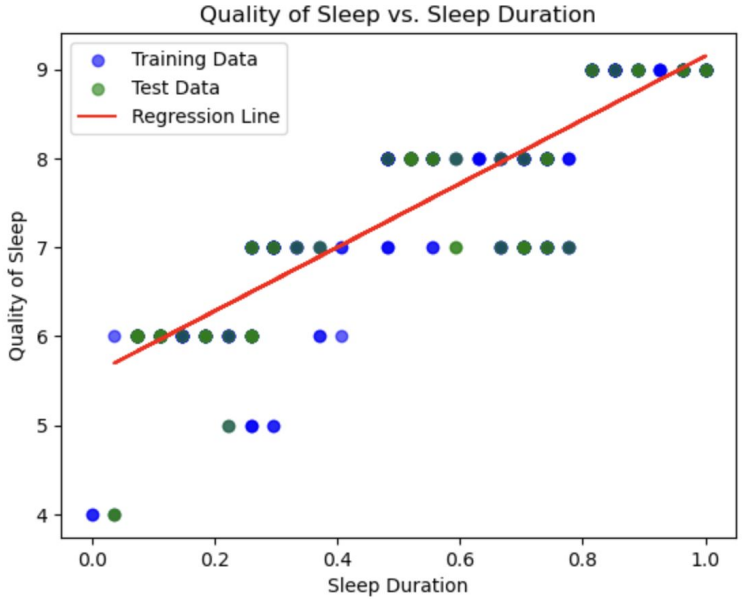
	Occupation	Coefficient
0	Accountant	1.825080
1	Doctor	-2.060721
2	Engineer	2.498459
3	Lawyer	2.194908
4	Manager	-0.337683
5	Nurse	0.214397
6	Sales Representative	-0.581539
7	Salesperson	-2.018636
8	Scientist	-0.767001
9	Software Engineer	0.102777
10	Teacher	-1.119531

Question 2 Analysis - Sleep Quality vs. Sleep Duration

Linear Regression

- R^2 Score = 0.765
- MSE = 0.355

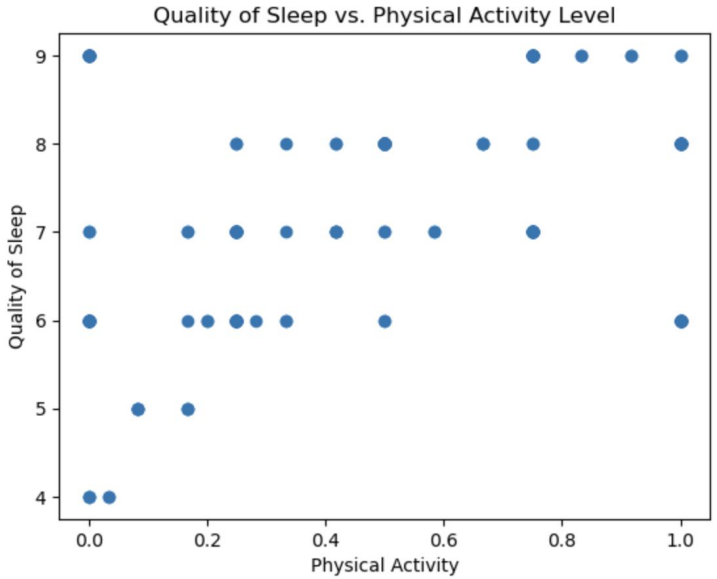


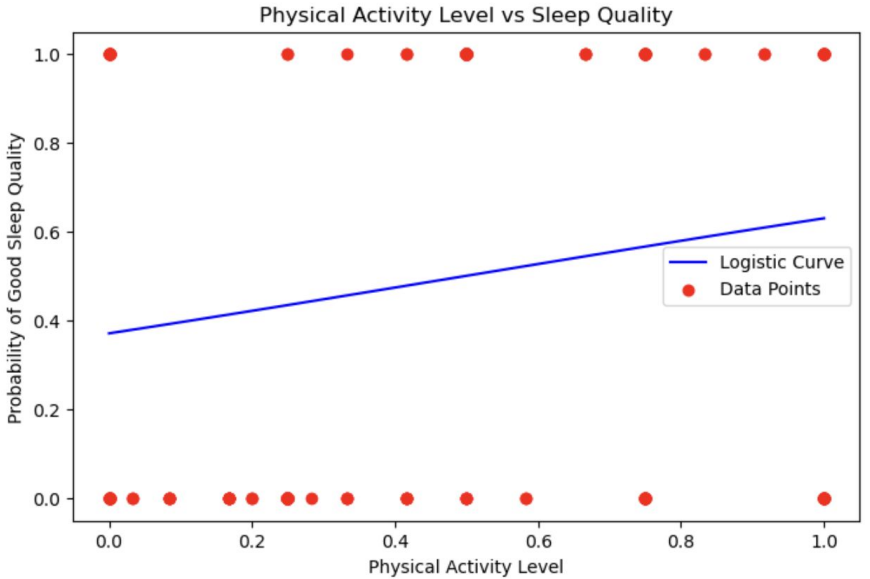


Question 2 Analysis - Sleep Quality vs. Physical Activity

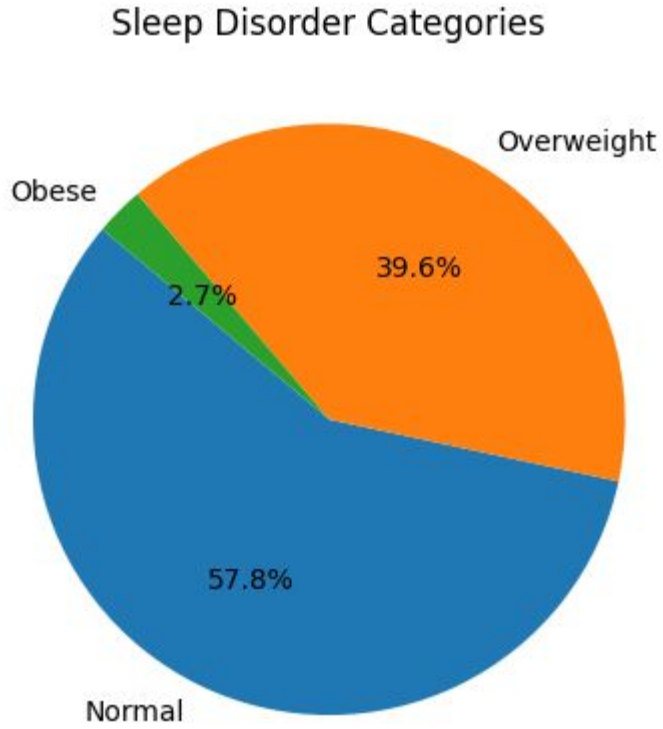
Logistic regression

- Test accuracy: 0.720





Question 2 Analysis - Sleep Quality vs. Weight



Logistic Regression

- Test accuracy = 0.667

	Weight	Coefficient
0	Normal	1.085135
1	Obese	-0.228406
2	Overweight	-0.853617

Question 2 - Conclusion

- Variables most important: heart rate, stress level, occupation, sleep duration, physical activity level, weight

Variable	Direction Associated With Good Sleep Quality
Heart Rate	Lower
Stress Level	Lower
Sleep Duration	Higher
Weight	Lower
Physical Activity	Higher

Weaknesses

- Correlated variables
- Relatively small dataset
- Subjective Variables
  - Quality of sleep
  - Stress Level
- Self-Reported Variables
  - Sleep Duration
  - Physical Activity Level
  - Daily Steps

Questions?