Final Project

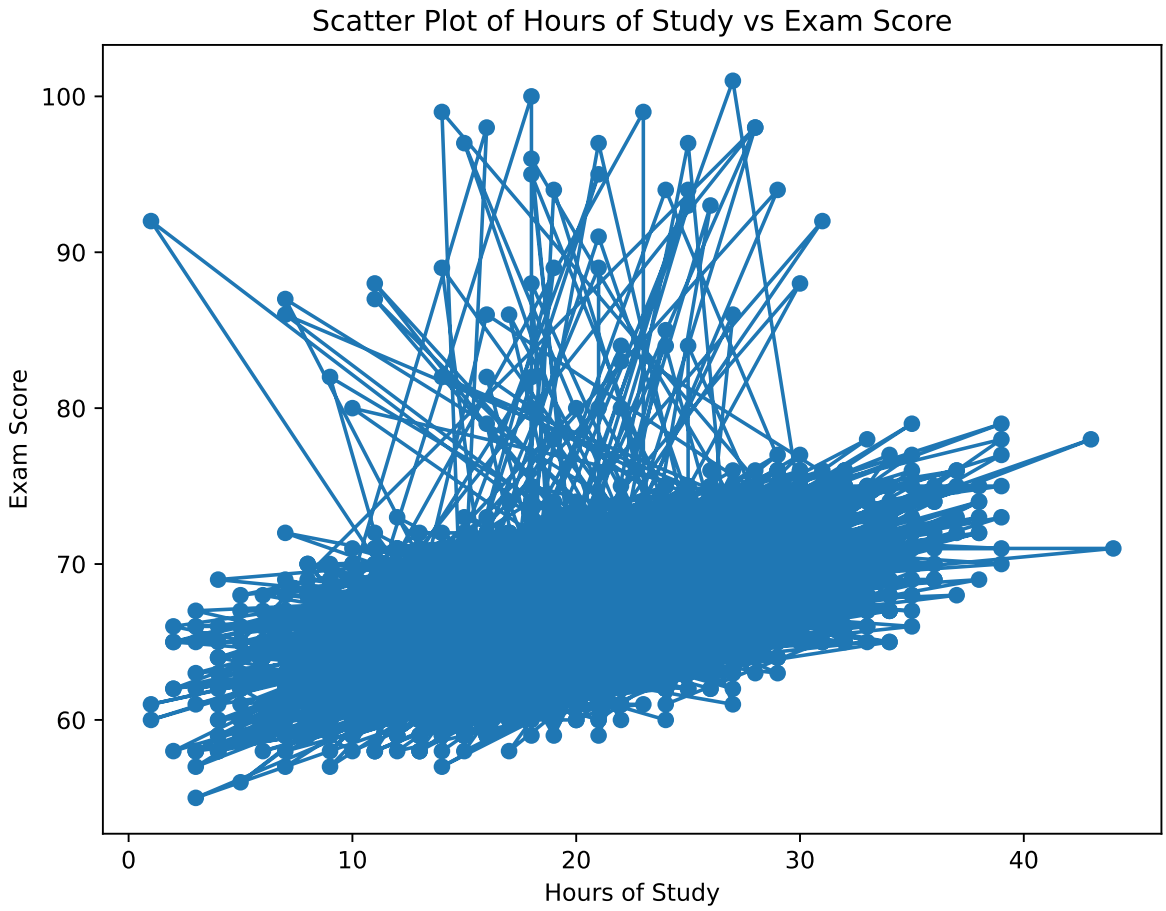
Mba Solomon Akachukwu.

2025-10-03

Data Exploration + Visualization

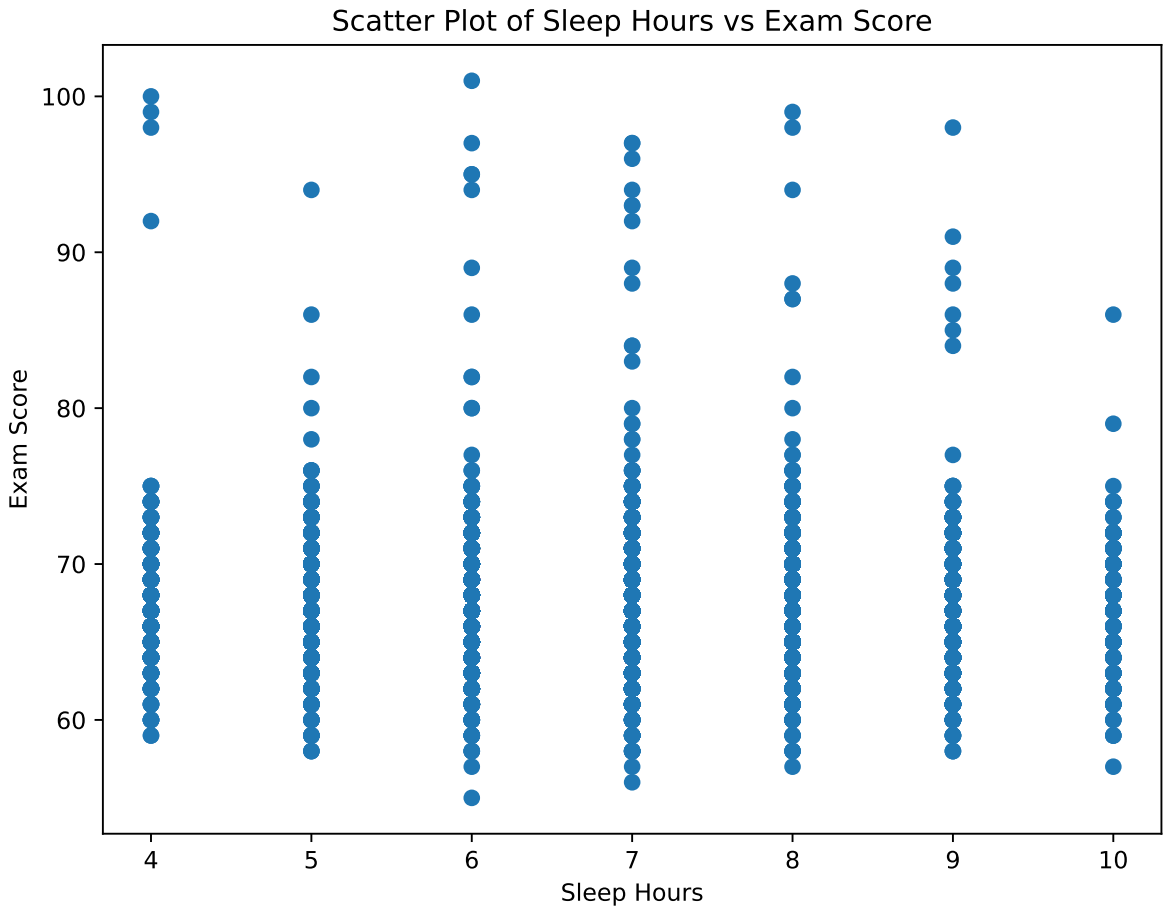
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('StudentPerformanceFactors.csv')
plt.figure(figsize=(8, 6))
plt.scatter(df['Hours_Studied'], df['Exam_Score'])
plt.plot(df['Hours_Studied'], df['Exam_Score'])
plt.title('Scatter Plot of Hours of Study vs Exam Score')
plt.xlabel('Hours of Study')
plt.ylabel('Exam Score')
```

```
Text(0, 0.5, 'Exam Score')
```



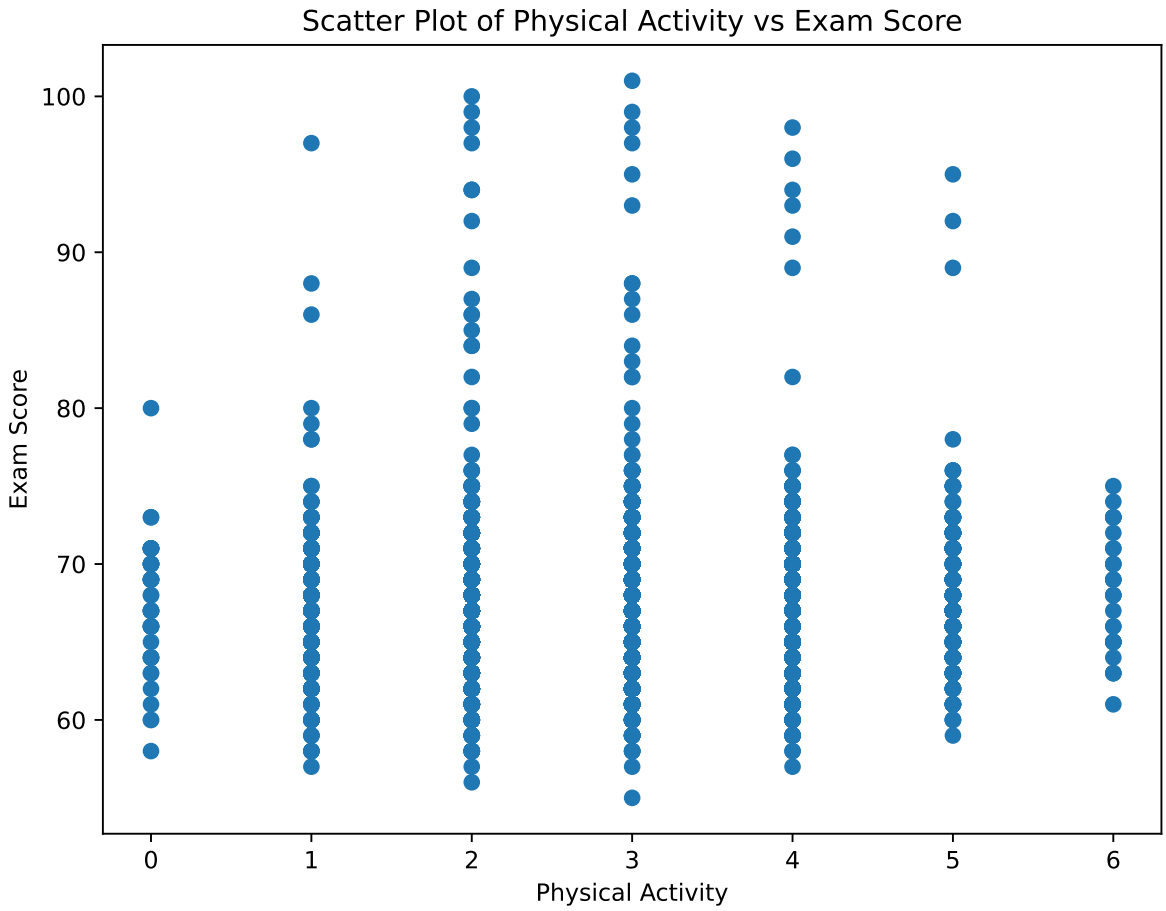

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('StudentPerformanceFactors.csv')
plt.figure(figsize=(8, 6))
plt.scatter(df['Sleep_Hours'], df['Exam_Score'])
plt.title('Scatter Plot of Sleep Hours vs Exam Score')
plt.xlabel('Sleep Hours')
plt.ylabel('Exam Score')
```

```
Text(0, 0.5, 'Exam Score')
```



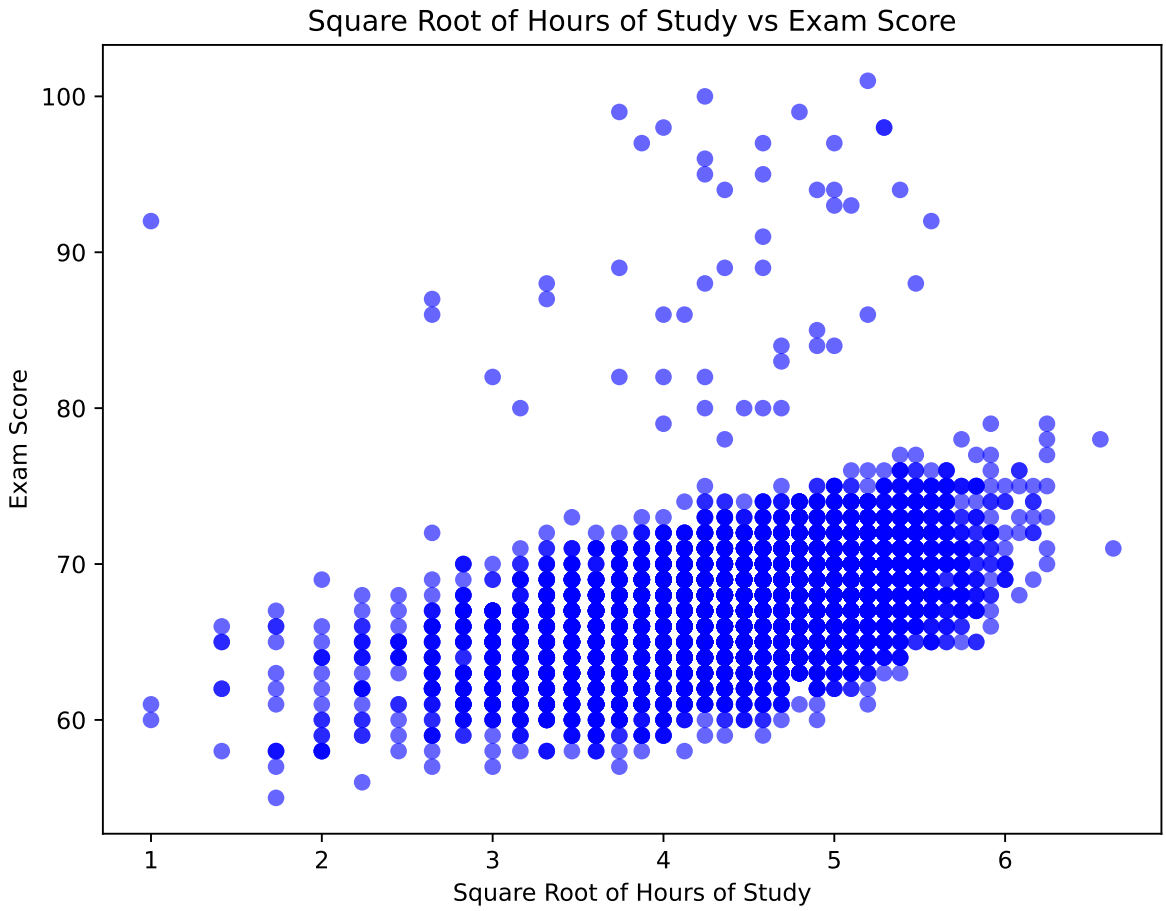

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('StudentPerformanceFactors.csv')
plt.figure(figsize=(8, 6))
plt.scatter(df['Physical_Activity'], df['Exam_Score'])
plt.title('Scatter Plot of Physical Activity vs Exam Score')
plt.xlabel('Physical Activity')
plt.ylabel('Exam Score')
```

```
Text(0, 0.5, 'Exam Score')
```



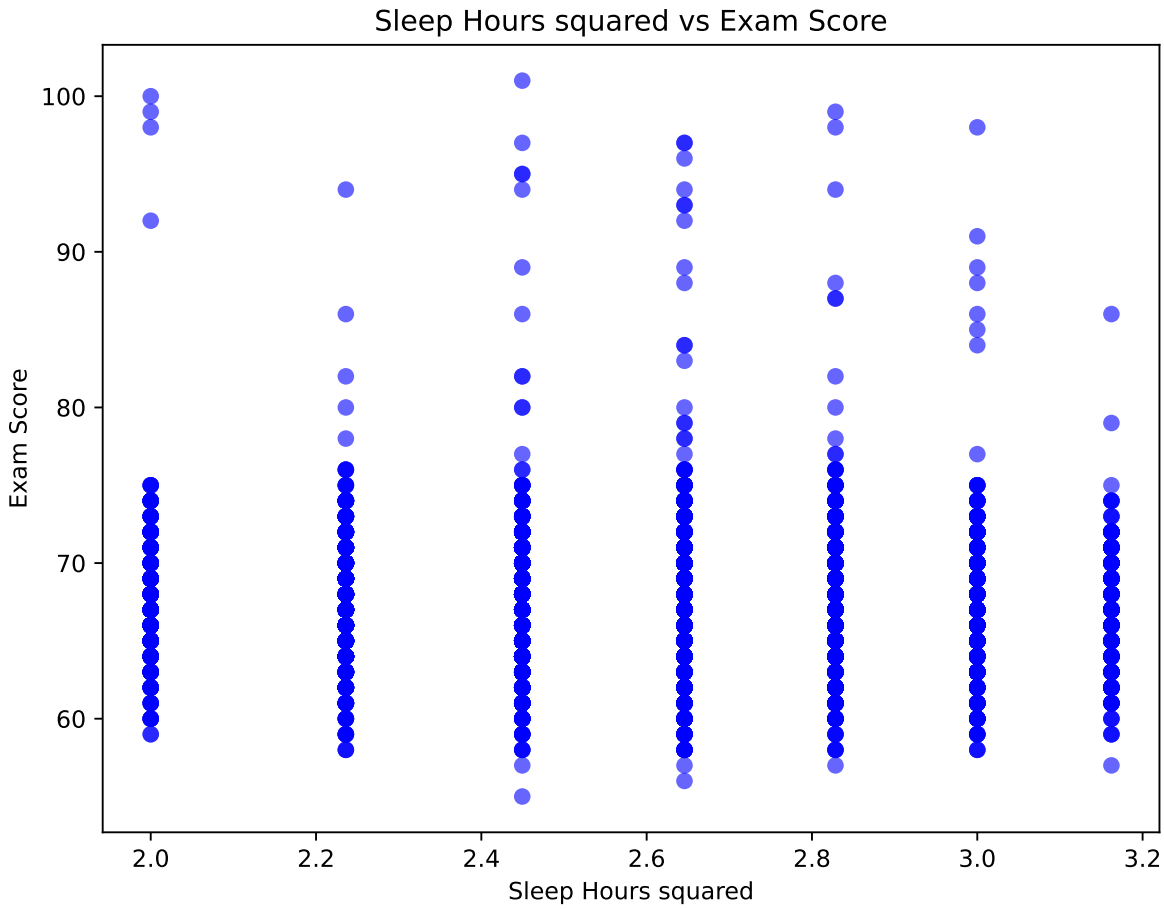
TESTS FOR NON-LINEARITY 1. For Hours Studied vs Exam Score, you can see that the graph ascends at first glance, but it isn't completely consistent since there are some values that appear in places that they would not be predicted to. Therefore I will perform a transformation on hours studied to make it more suitable.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
df['sqrt_Hours_Studied'] = np.sqrt(df['Hours_Studied'])
plt.figure(figsize=(8, 6))
plt.scatter(df['sqrt_Hours_Studied'], df['Exam_Score'],alpha=0.6, color='blue')
plt.title('Square Root of Hours of Study vs Exam Score')
plt.xlabel('Square Root of Hours of Study')
plt.ylabel('Exam Score')
plt.show()
```



2. For sleep vs exam score, there is no linearity since there isnt a clear trend upwards, so the slope is no necessarily positive. In that case, I will try to square the hours of sleep to fix it

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
df['squared_Sleep_Hours'] = np.sqrt(df['Sleep_Hours'])
plt.figure(figsize=(8, 6))
plt.scatter(df['squared_Sleep_Hours'], df['Exam_Score'],alpha=0.6, color='blue')
plt.title('Sleep Hours squared vs Exam Score')
plt.xlabel('Sleep Hours squared')
plt.ylabel('Exam Score')
plt.show()
```


Sleep Hours squared vs Exam Score
Exam Score
100
90
80
70
60
2.0
2.2
2.4
2.6
2.8
3.0
3.2
Sleep Hours squared