ITSC-3162-Intro-to-Data-Mining-Project1

Camron Patel

9/16/24

Dataset: https://www.kaggle.com/code/volkandl/sleep-in-mammals-eda

Introduction

Sleep is one of the most vital mechanisms in life. Understanding sleep in mammals plays a critical role in our knowledge of their health and day-to-day activities. Sleep patterns in mammals can show us lots of information about their behavior. It can reveal how they interact with their environment and use the resources around them. Studying sleep in mammals also helps with conservation efforts. Sleeping patterns can show disruptions in the natural environment in which they inhabit. I am interested in looking into how the different variables in mammals affect the total sleep of the mammal.

- **Does the amount of days in gestation affect the average hours of sleep in a mammal?**

Dataset

The dataset I am using is from Kaggle and is called “Sleep in Mammals” by Volkan Özdemir. This dataset was originally used for the analysis of constitutional and ecological factors and sleeping mammals. It consists of 62 different species of mammals along with 11 variables including the following: species, body_wt, brain_wt, non_dreaming, dreaming, total_sleep, life_span, gestation, predation, exposure, and danger. There are a few downsides to this dataset, since sleep can be affected by a variety of things it is hard to capture all of them. For example, the amount of food the mammal consumes may affect its sleep but it is not listed in this dataset.

This missing data can lead to false conclusions which can be very harmful. It is extremely important to consider these factors when analyzing data.

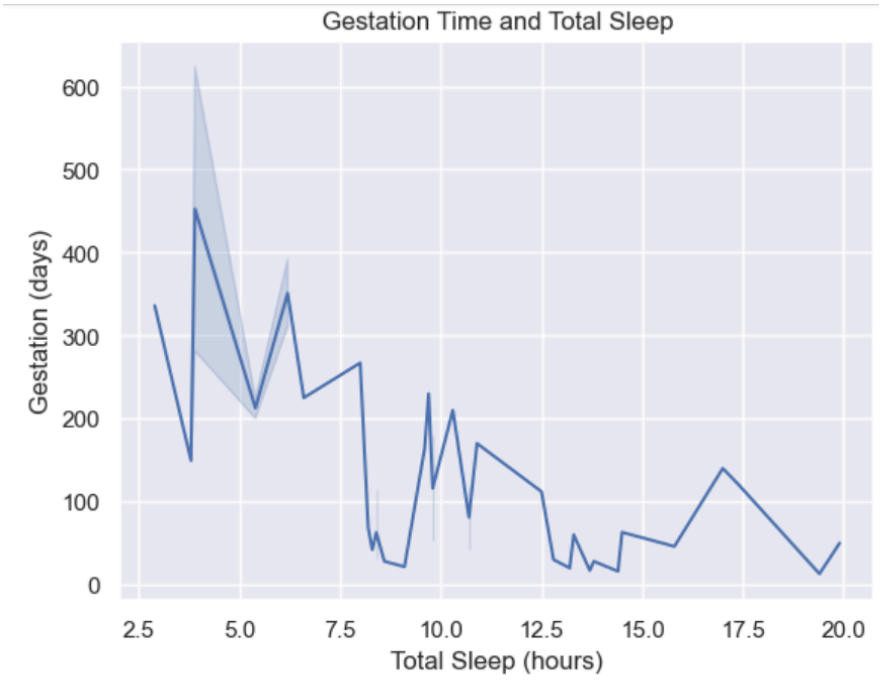
Pre-processing

Before working with this data we must clean it. Data cleaning is important because it guarantees that the data is accurate and standardized. Working with unclean data can lead to misinformation and major consequences.

I removed the species with null values in any column to clean this dataset. I did this because I only wanted to work with species that we had all the information on. Null values in datasets can lead to inconsistent results. Next, I removed the subjective columns (predation, exposure, and danger). This was done because this data was based on ratings by the owner and not any reliable information or sources.

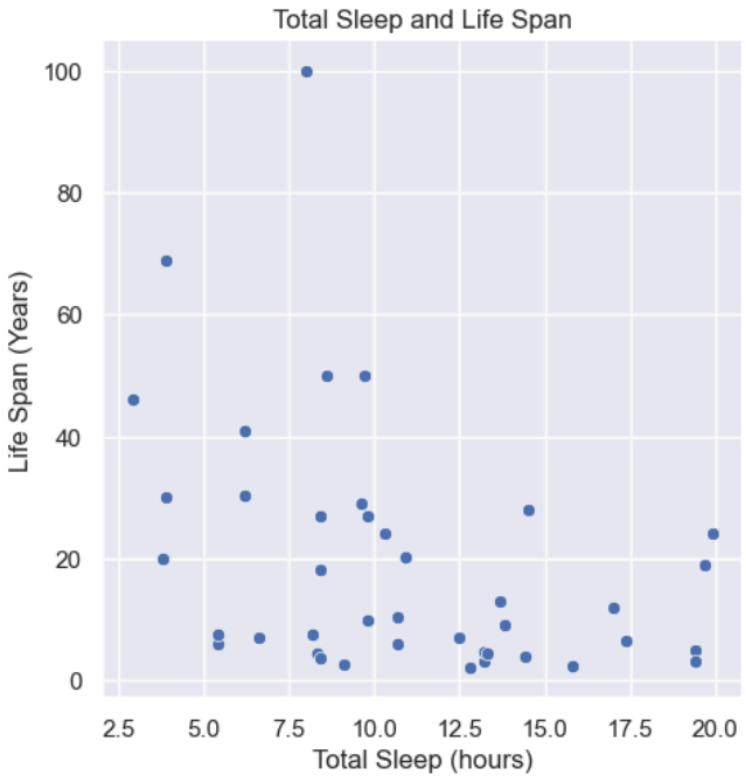
Data Visualization

Fig 1.



The line graph above shows the gestation time in days and the mammal's average total sleep in hours. This graph was created in Python using the Seaborn library. Using this visualization we can see that the lower amount of days in gestation the more total sleep the mammal has and vice versa, the higher amount of days in gestation the less total sleep the mammal has. This graph shows that the time spent in gestation affects a mammal's total sleep.

Fig 2.



The relationship plot above shows the relationship between life span in years and total sleep in hours. This plot was also created in Python using the Seaborn library. As you can see the plot is inconclusive and does not reveal much. This could be because the life span of an animal can be affected by a multitude of things such as predation and environmental factors. This is why we must be careful with our conclusions. I thought this was interesting because I had assumed that mammals with more sleep would live longer on average.

Impacts

The impacts of this data can be both positive and negative. *Fig 1* shows that there is a relationship between gestation days and total sleep this can be further researched. Questions such as, Why do mammals with longer periods of gestation sleep less than mammals with shorter periods of gestation? How else can we predict longer or shorter sleep cycles between mammals? This dataset was also narrowed down to 42 mammals. Since this is such a low number we cannot generalize our conclusions and observations to all mammals. Also working with a variable as complex as sleep we must be careful in our analysis.

References

"Pandas Documentation#." *Pandas Documentation - Pandas 2.2.2 Documentation*, pandas.pydata.org/docs/index.html.

"Sleep in Mammals Eda." *Kaggle*, www.kaggle.com/code/volkandl/sleep-in-mammals-eda.

"Statistical Data Visualization#." *Seaborn*, seaborn.pydata.org/.

Code

Pre-processing

```
#drops rows with null values
mammal_df.dropna(axis = 0, how = 'any', inplace = True)
```

```
#dropping columns
mammal_df.drop('predation', axis= 1, inplace = True)
mammal_df.drop('exposure', axis= 1, inplace = True)
mammal_df.drop('danger', axis= 1, inplace = True)
mammal_df.shape
```

Fig 1

```
sns.set_theme(style = 'darkgrid')
# Line graph for total sleep and gestation
sns.lineplot(x="total_sleep", y="gestation",
             data=mammal_df).set(title = 'Gestation Time and Total Sleep' , xlabel = 'Total Sleep (hours)' , ylabel = 'Gestation (days)')
```

Fig 2

```
#plot for total sleep and lifespan
sns.set_theme(style = 'darkgrid')
sns.relplot(data = mammal_df,x="total_sleep", y="life_span",).set(title = 'Total Sleep and Life Span' , xlabel = 'Total Sleep (hours)',
            ylabel = 'Life Span (Years)')
```