INTRODUCTION TO MACHINE LEARNING COURSE

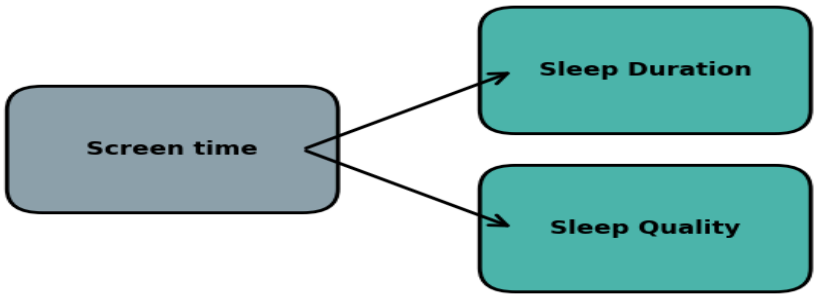
DREAMS & GRADES

PREPARED BY :
ZIED ELLOUZI, ASSIL TEKAYA, YASMINE ALAYA

Graphical models of the main variables of our Project

Here, we examine each dependency separately in order to better understand how every variable influences the sleep-related outcomes. By isolating the effects of Screen Time, Stress, and Caffeine Intake, we can clearly identify their specific impacts on Sleep Duration and Sleep Quality before combining them into the full model. This step-by-step approach allows us to justify each connection in the final graphical model through empirical analysis and scientific evidence.

Screen time → SleepDuration & Sleep Quality :



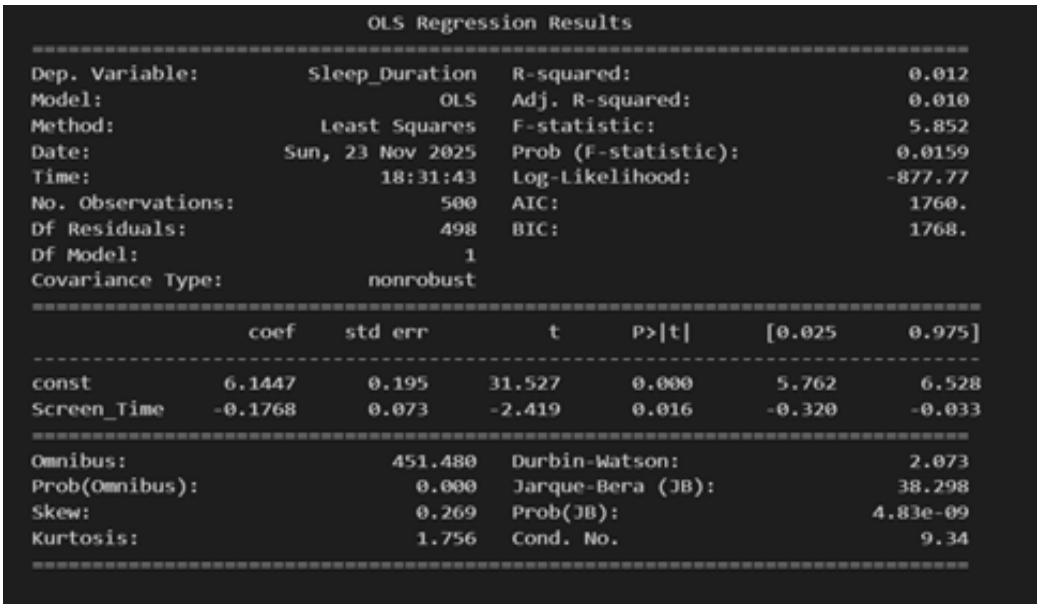
Why ?

- Higher social media use can reduce total sleep duration. Spending excessive time on screens especially before bedtime tends to delay sleep onset and shorten the number of hours slept. Therefore, increased social media use is typically associated with shorter sleep duration.
- Social media use can also negatively impact sleep quality. Screen exposure before sleeping interferes with melatonin production and can disrupt natural sleep cycles, leading to more restlessness and less restorative sleep. As a result, higher Screen time is linked to poorer sleep quality.

  "Screen time and sleep among school-aged children and adolescents: A systematic literature review." ***Sleep Medicine Reviews*****, 21, 50–58.** https://doi.org/10.1016/j.smrv.2014.07.007

➡ This systematic review clearly shows that screen time not only shortens sleep duration but also **significantly impairs sleep quality**, leading to **delayed sleep onset, more nighttime disturbances, and greater daytime tiredness among children and adolescents**.

Below is an example of a linear regression performed between Screen Time and Sleep Duration to illustrate their dependency (using a dataset found on kaggle):



OLS Regression Results

Dep. Variable:	Sleep_Duration	R-squared:	0.012
Model:	OLS	Adj. R-squared:	0.010
Method:	Least Squares	F-statistic:	5.852
Date:	Sun, 23 Nov 2025	Prob (F-statistic):	0.0159
Time:	18:31:43	Log-Likelihood:	-877.77
No. Observations:	500	AIC:	1760.
Df Residuals:	498	BIC:	1768.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.1447	0.195	31.527	0.000	5.762	6.528
Screen_Time	-0.1768	0.073	-2.419	0.016	-0.320	-0.033

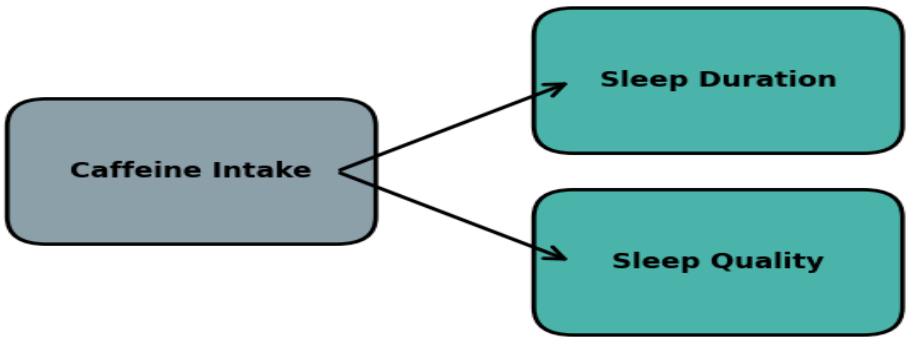
Omnibus:	451.480	Durbin-Watson:	2.073
Prob(Omnibus):	0.000	Jarque-Bera (JB):	38.298
Skew:	0.269	Prob(JB):	4.83e-09
Kurtosis:	1.756	Cond. No.	9.34

The regression indicates a **negative relationship** between Screen Time and Sleep Duration.
The coefficient for Screen_Time is **−0.1768**, meaning that **for each additional hour of screen use, sleep duration decreases by approximately 0.18 hours** (around 11 minutes), on average.

The effect is **statistically significant** (p-value = **0.016**), which supports the idea that Screen Time has a measurable impact on how long students sleep.

Although the **R-squared value is modest (0.012)**—indicating that Screen Time explains a small portion of the variation in sleep duration—this is expected in behavioral data where many external factors influence sleep.

Caffeine intake → SleepDuration & Sleep Quality :



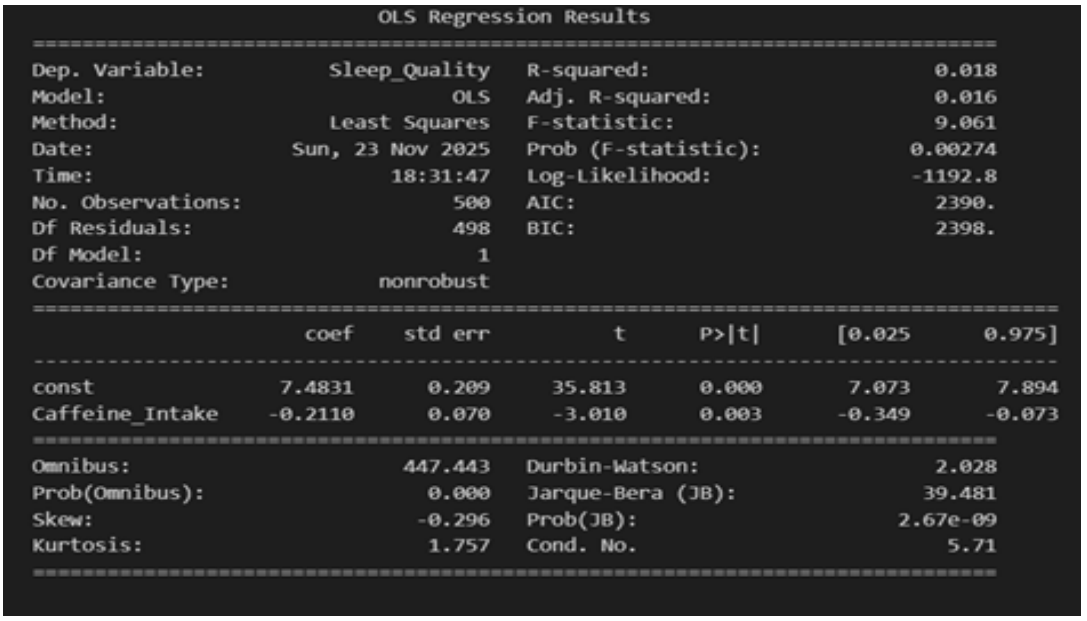
Why ?

- Caffeine blocks adenosine receptors, preventing the feeling of tiredness and delaying the onset of sleep. It also increases alertness and reduces the ability to fall asleep quickly, individuals who consume more caffeine generally sleep fewer hours. This makes caffeine intake a meaningful predictor of sleep duration . **"Caffeine effects on sleep taken 0, 3, or 6 hours before going to bed."** ***Journal of Clinical Sleep Medicine*****, 9(11), 1195–1200.** https://jcsm.aasm.org/doi/10.5664/jcsm.3170

➡ This study shows that caffeine consumed even **6 hours before bedtime significantly disrupts sleep on**

- Caffeine is a stimulant that increases alertness and activates the central nervous system. Consuming caffeine especially later in the day can interfere with the natural process of falling asleep and reduce the depth of sleep. Higher caffeine intake is also associated with more awakenings during the night and reduced time spent in restorative sleep stages. As a result, increased caffeine consumption is typically linked to lower sleep quality.

Below is another example of a linear regression performed between Caffeine intake and Sleep Quality to illustrate their dependency (using the same dataset found on kaggle):



OLS Regression Results

Dep. Variable:	Sleep_Quality	R-squared:	0.018
Model:	OLS	Adj. R-squared:	0.016
Method:	Least Squares	F-statistic:	9.061
Date:	Sun, 23 Nov 2025	Prob (F-statistic):	0.00274
Time:	18:31:47	Log-Likelihood:	-1192.8
No. Observations:	500	AIC:	2390.
Df Residuals:	498	BIC:	2398.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.4831	0.209	35.813	0.000	7.073	7.894
Caffeine_Intake	-0.2110	0.070	-3.010	0.003	-0.349	-0.073

Omnibus:	447.443	Durbin-Watson:	2.028
Prob(Omnibus):	0.000	Jarque-Bera (JB):	39.481
Skew:	-0.296	Prob(JB):	2.67e-09
Kurtosis:	1.757	Cond. No.	5.71

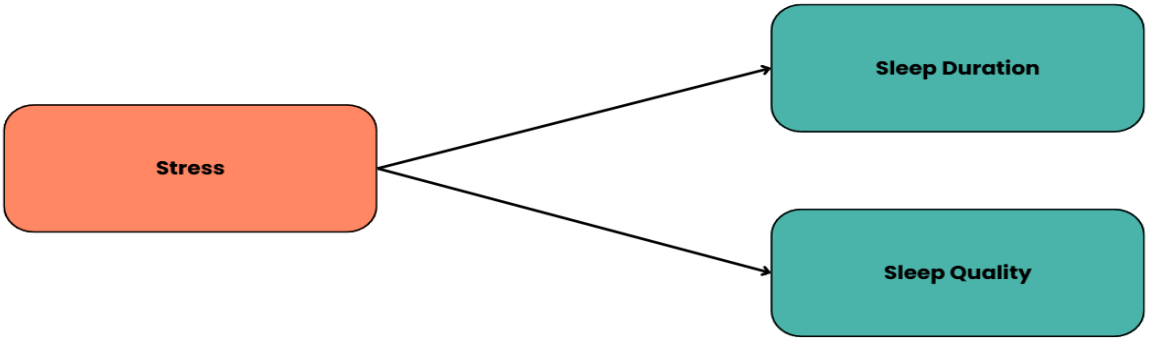
The regression shows a **negative relationship** between Caffeine Intake and Sleep Quality.

The coefficient for Caffeine_Intake is **−0.2110**, meaning that **each additional unit of caffeine intake reduces sleep quality by about 0.21 points on average.**

The effect is **statistically significant** (p-value = **0.003**), indicating that caffeine consumption has a measurable and reliable impact on sleep quality in the dataset.

The **R-squared (0.018)** is small, which is normal in behavioral datasleep quality is influenced by many factors but it still indicates that caffeine plays a non-negligible role in sleep degradation

Stress Level → SleepDuration & Sleep Quality :



. Stress activates the hypothalamic–pituitary–adrenal (HPA) axis, increasing cortisol levels, which makes it harder to fall asleep and reduces total sleep duration. Higher stress levels are therefore typically associated with fewer hours of sleep.

· Stress also affects sleep quality by increasing nighttime awakenings, reducing deep sleep stages, and causing more fragmented sleep. Individuals with higher stress tend to report poorer sleep quality and less restorative sleep

Ref: Current Medicine Research and Practice

Below is an example of a linear regression performed between **Stress Level and Sleep Duration** found on a Kaggle dataset:

```
Regression: Stress Level -> Sleep Duration
                            OLS Regression Results
==============================================================================
Dep. Variable:         Sleep Duration   R-squared:                       0.658
Model:                            OLS   Adj. R-squared:                  0.657
Method:                 Least Squares   F-statistic:                     715.0
Date:                Mon, 24 Nov 2025   Prob (F-statistic):           1.24e-88
Time:                        02:47:21   Log-Likelihood:                -244.18
No. Observations:                 374   AIC:                             492.4
Df Residuals:                     372   BIC:                             500.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          9.0903      0.077    117.903      0.000       8.939       9.242
Stress Level  -0.3636      0.014    -26.739      0.000      -0.390      -0.337
==============================================================================
Omnibus:                        5.513   Durbin-Watson:                   0.667
Prob(Omnibus):                  0.064   Jarque-Bera (JB):                5.291
Skew:                          -0.281   Prob(JB):                       0.0710
Kurtosis:                       3.153   Cond. No.                         18.6
==============================================================================
```

The regression indicates a negative relationship between Stress Level and Sleep Duration.

The coefficient for Stress_Level is **−0.3636**, meaning that for each additional point of stress, sleep duration decreases by approximately **0.36 hours**, on average.

The effect is statistically significant (p-value = 0.000), which supports the idea that stress has a measurable impact on sleep duration.

The R-squared value is **0.658**, indicating that 65.8% of the variation in Sleep Duration is explained by Stress Level which shows a strong dependency between Stress and sleep duration

Below is an example of a linear regression performed between Stress Level and Sleep Quality to illustrate their dependency :

```
Regression: Stress_Level -> Sleep_Quality
                            OLS Regression Results
==============================================================================
Dep. Variable:       Quality of Sleep   R-squared:                       0.808
Model:                            OLS   Adj. R-squared:                  0.807
Method:                 Least Squares   F-statistic:                     1563.
Date:                Mon, 24 Nov 2025   Prob (F-statistic):          2.88e-135
Time:                        03:01:58   Log-Likelihood:                -289.06
No. Observations:                 374   AIC:                             582.1
Df Residuals:                     372   BIC:                             590.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          10.5774      0.087    121.678      0.000      10.406      10.748
Stress Level   -0.6062      0.015    -39.535      0.000      -0.636      -0.576
==============================================================================
Omnibus:                       84.825   Durbin-Watson:                   0.854
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              143.897
Skew:                          -1.329   Prob(JB):                     5.66e-32
Kurtosis:                       4.474   Cond. No.                         18.6
==============================================================================
```

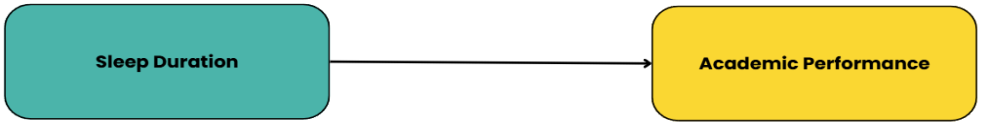
The regression indicates a negative relationship between Stress Level and Sleep Quality.

The coefficient for Stress_Level is **−0.6062**, meaning that for each additional point of stress, sleep quality decreases by approximately **0.61 points**, on average.

The effect is statistically significant (p-value = 0.000), which supports the idea that stress has a measurable impact on how restful and restorative sleep is.

The R-squared value is **0.808**, indicating that stress explains a substantial portion of the variation in sleep quality in this dataset.

Sleep duration → Academic Performance



To investigate the relationship between sleep duration and academic performance, we referred to a large-scale study of 54,102 students that we found in an article, which included both self-reported sleep duration and standardized mathematics scores.

ref: Sleep duration and subject-specific academic performance among adolescents in China - PMC

From this article, we extracted the average math scores corresponding to different sleep duration categories:

Sleep Duration	Original Math Score
<6 h	458
6–7 h	506
7–8 h	506
8–9 h	509
9–10 h	505
>10 h	482

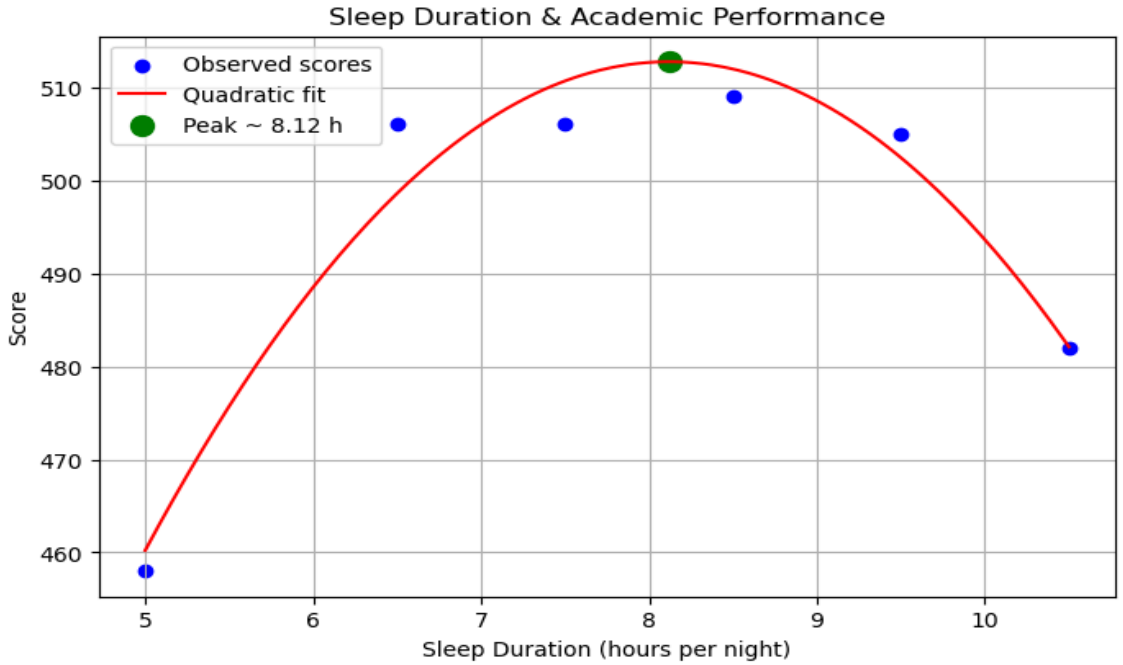
Since this study included more than 54,000 students from different schools and regions, their grading systems and exam difficulties were not identical. To allow fair comparison, the researchers converted all raw scores into standardized scores with a mean of 500 & a standard deviation of 100

Standardization Formula :

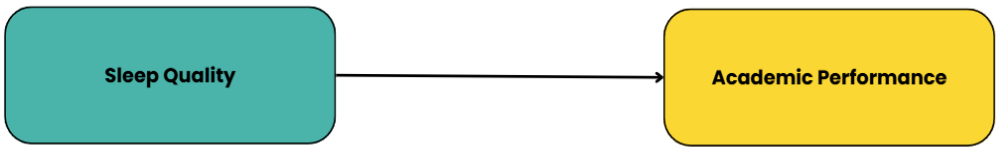
$$\text{standardized score} = (\text{raw score} - \text{group mean}) / \text{standard deviation}$$

The target variable, academic performance, is continuous, so this is a regression problem, and the study shows an inverted-U relationship (performance increases around 8–9 hours of sleep and then decreases).

So to capture this pattern, we use a quadratic function (degree 2)
(LinearRegression with polynomial features to efficiently fit the coefficients)



Sleep Quality → Academic Performance



Based on some researches :

Ref : - Questionnaire-Derived Sleep Habits and Academic Achievement in First Year University Students - PubMed

- Good quality sleep is associated with better academic performance among university students in Ethiopia - PubMed

· Good sleep quality supports memory consolidation, attention, and cognitive processing,

which are essential for learning and academic performance. Students with better sleep quality tend to show higher concentration, faster information processing, and better exam results

· Poor sleep quality, is associated with lower academic achievement. It can increase daytime sleepiness, reduce motivation, and impair cognitive functioning

Summary of Findings:
Below is a summary table based on the trichotomized PSQI scores from the New Zealand study:

Sleep Quality	Description	Sample Size	Average Academic Performance
Good	PSQI ≤ 5	62	Higher GPA
Moderate	PSQI 5–8	63	Medium GPA
Poor	PSQI ≥ 8	68	Lower GPA

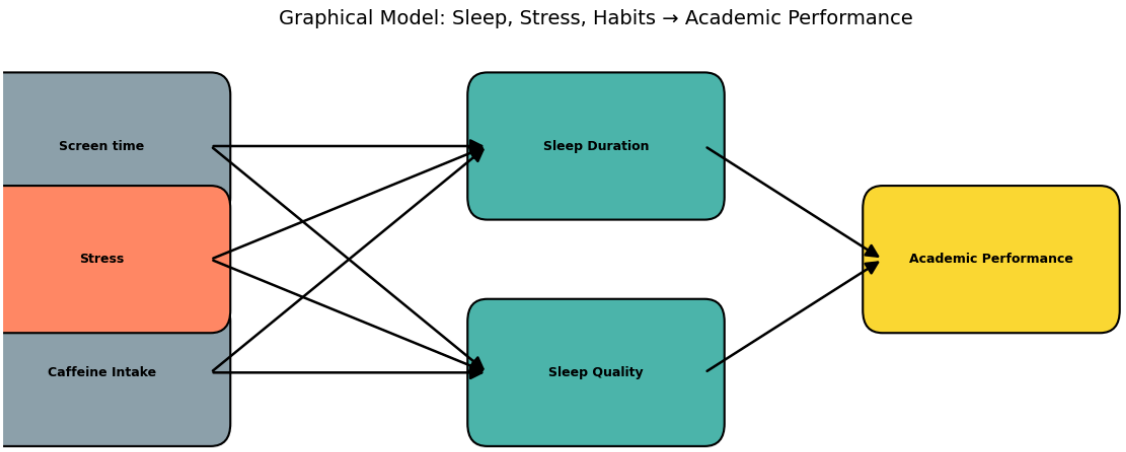
PSQI stands for **Pittsburgh Sleep Quality Index**, It's a widely used **questionnaire to measure sleep quality** over a 1-month period

The evidence indicates a positive relationship between sleep quality and academic performance: students with better sleep quality achieve higher academic outcomes.

Final Graphical Model

Now that we have examined each relationship individually analyzing how Screen Time, Stress, and Caffeine Intake influence Sleep Duration and Sleep Quality, and how these sleep variables in turn affect Academic Performance we are finally able to integrate all these dependencies into one comprehensive view.

The following graphical model summarizes the complete structure of our project: it brings together all the main variables and highlights the direct influences between them exactly as supported by the data analysis and the scientific literature reviewed earlier.



Project plan:

The next steps of our project will focus on strengthening the data preparation process and selecting an appropriate regression model:

1- Generating the Dataset

2- Data preprocessing

- Handle missing values:

  Identify incomplete observations and apply suitable techniques such as implementation or removal depending on the nature of the missing data.

- Encode categorical variables:

  Convert categorical features into numerical formats to ensure compatibility with machine learning algorithms.

- Split the dataset:

  Divide the data into training and test sets to allow for reliable model evaluation and prevent overfitting.

3- Model selection and implementation

- Choose the most suitable regression model:

  compare different regression algorithms (e.g: Bayesian Ridge Regressor, Linear regression…) based on the dataset characteristics.

- Implement and evaluate the model:

  train the model selected and assess its performance using appropriate metrics, and adjust parameters if needed to improve accuracy.