consider whether to remove outliers, based on the type of model you decide to use.

pAce: Analyze Stage

- Perform EDA (analyze relationships between variables)

Reflect on these questions as you complete the analyze stage.

- What did you observe about the relationships between variables?
- What do you observe about the distributions in the data?
- What transformations did you make with your data? Why did you chose to make those decisions?
- What are some purposes of EDA before constructing a predictive model?
- What resources do you find yourself using as you complete this stage? (Make sure to include the links.)
- Do you have any ethical considerations in this stage?

[Double-click to enter your responses here.]

- Relationships between variables:

Observations: You might have observed significant correlations between variables like satisfa
and left, where lower satisfaction levels correlate with higher employee turnover. The relat
between number_project and average_monthly_hours might show that higher workloads correlate
dissatisfaction or departure.

- Distributions in the data:

Observations: The distribution of average_monthly_hours could be bimodal, indicating two grou
the workforce with distinct work patterns. The distribution of satisfaction_level might be sl
suggesting clusters of high and low satisfaction among employees.

- Data transformations:

Reasons: You likely converted categorical variables like department and salary into numerica
(e.g., using one-hot encoding) to prepare them for modeling. Such transformations are crucia
including these variables in machine learning models.

- Purposes of EDA:

Utility: EDA helps in understanding underlying patterns, detecting outliers, and anomalies, a
formulating hypotheses about causative relationships, which is essential before model buildi

- Resources used:

Python documentation, Stack Overflow for coding issues, and possibly sites like Kaggle for i