Scenario

The data consulting firm Automatidata hired you as the newest member of their data analytics team. In a previous email, your new supervisor shared some bullet points from the last Automatidata leadership team meeting. You have just received an additional email from your manager asking you to complete a project proposal for the New York City Taxi and Limousine Commission project. Review the email, then follow the provided instructions to complete the PACE strategy document and the project proposal.

PACE Strategy Document

Plan Stage

- **Audience**: The primary audience for this project is the New York City Taxi and Limousine Commission (TLC), as well as the Automatidata data team members who will be collaborating on the project.
- **Project Goal:** To develop a regression model that accurately predicts taxi cab fares based on factors like distance, time of day, and potentially other variables. This model will help TLC improve fare predictions, aiding both taxi drivers and passengers with better fare transparency.
- **Key Questions:**
    - What are the available data sources, and what data cleaning will be required?
    - What additional variables (e.g., weather, traffic, special events) might impact fare predictions?
    - What tools and resources will we need for model development and deployment?
- **Resources Needed:** Access to historical TLC data, data analysis software (e.g., Python, R), computing resources for model training, and any relevant documentation or guidelines from TLC.
- **Deliverables:**
    - Project proposal document
    - Data cleaning and feature engineering notes
    - Predictive model and evaluation results
    - Final report with insights and recommendations

Analyze Stage

- **Tasks:**
    - Begin exploring and cleaning the data to understand its structure and quality.

- Conduct exploratory data analysis (EDA) to identify trends and correlations between variables.
- Compute descriptive statistics to summarize the data.
- Conduct hypothesis testing to validate the significance of chosen variables in relation to fare predictions.

Construct Stage

- **Tasks:**
  - Develop the regression model by experimenting with different types (e.g., linear, ridge, or lasso regression).
  - Build the machine learning model using training data, and perform model optimization.
  - Compile summary information about the data and model performance to prepare for reporting.

Execute Stage

- **Tasks:**
  - Evaluate the model by testing it with validation data to assess accuracy and reliability.
  - Build visualizations to communicate findings and model insights effectively.
  - Prepare and deliver a final report summarizing the analysis and recommendations for TLC.