# PACE Strategy Document



# Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

# Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

* Complete the PACE Strategy Document to plan your project while considering your audience members, teammates, key milestones, and overall project goal.
* Create a project proposal for the data team.

# Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

* As a new member of a data analytics team, what steps could you take to get 'up to speed' with a current project? What steps would you take? Who would you like to meet with?
* How would you plan an analytics project?
* What steps would you take to translate a business question to an analytical solution?
* Why is actively managing data an important part of a data analytics team's responsibilities?
* What are some considerations you might need to be mindful of when reporting results?

**Reference Guide**

This project has three tasks; the following visual identifies how the stages of PACE are incorporated across those tasks.



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* Who is your audience for this project?

|  |
| --- |
| * New York City TLC Team Members. * Automatidata Team Members |

* What are you trying to solve or accomplish? And what do you anticipate the impact of this work will be on the larger needs of the client?

|  |
| --- |
| * We aim to develop a regression model to estimate taxi fares before a ride using the extensive data collected by TCL. * Anticipated Impact on the client’s larger needs:   + Enhanced Customer Experience   + Operational Improvements   + Data-Driven Decision-Making (fare regulations or zone pricing)   + Strategic Planning (dynamic pricing or revenue forecasting) |

* What questions need to be asked or answered?

|  |
| --- |
| * Client objectives:   + What are TLC’s **primary goals** for this regression model? (e.g., operational efficiency, customer transparency, policy development)   + Are there specific **KPIs or metrics** TLC wants to improve through this project? * Current state:   + How does TLC **currently use** the available data for decision-making?   + Are there any **existing tools** or systems in place for fare estimation? * Data availability and quality:   + What **data fields** are available in the TLC dataset? (e.g., trip distance, pickup/drop-off location, time of day, weather conditions)   + Are there any **known issues** with the data, such as missing values, inconsistencies, or outdated records? * Data relevance:   + Are there **external datasets** that could enhance the model’s accuracy? (e.g., traffic patterns, weather data, public event schedules)   + How is the data **updated or maintained** over time? * Model specifications:   + What **level of accuracy** is required for the fare predictions?   + Are there **constraints or preferences** regarding the type of model? (e.g., simplicity, interpretability, computational efficiency) * Feature engineering:   + Which factors does TLC believe **most influence** taxi fares? (e.g., trip distance, demand surge, time of day)   + Are there **additional features** that could be derived from existing data? (e.g., rush hour indicators, zone-based fees) * Stakeholder input:   + How should the model’s results be presented to ensure they are **actionable** for TLC stakeholders?   + Are there **specific insights or trends** that the TLC team is particularly interested in? * Impact and integration:   + How will this model be **integrated into TLC’s existing workflows** or systems?   + What is the **long-term plan** for maintaining and improving the model after deployment? * Client and customer impact:   + What are the **potential benefits** of fare predictions for passengers and drivers?   + How might **fare transparency** affect customer trust and usage patterns? * Project evaluation:   + How will TLC **evaluate the success** of the project? (e.g., model accuracy, stakeholder satisfaction, operational improvements)   + What **benchmarks or comparisons** will be used to measure the model’s performance? |

* What resources are required to complete this project?

|  |
| --- |
| * Data resources:   + TLC data.   + External datasets (e.g., traffic, weather, public events, etc.) * Technology resources:   + Data Processing Tools   + Data Visualization   + Cloud Computing (for large datasets)   + Database Access * Team expertise:   + Automatidata Team:     - Data analysts and scientists with expertise in regression modeling and data visualization.     - A project manager to oversee timelines, milestones, and communication.   + TLC Team:     - Subject matter experts provide context for the data and validate assumptions.     - Operations and finance representatives to define key metrics and evaluate business impact. * Analytical Framework: Modeling (ML) and validation. * Documentation and communication: Reporting and collaboration tools * Time and budget. |

* What are the deliverables that will need to be created over the course of this project?

|  |
| --- |
| * **Project Proposal Document:** outlines the project scope, goals, milestones, and timeline. Key components:   + Problem statement and project objectives.   + PACE workflow alignment.   + Identified stakeholders and their roles.   + High-level tasks and deliverable roadmap. * **Data Assessment Report:** summarizes the quality, completeness, and relevance of the TLC dataset. Key components:   + Data inventory (fields, types, sample records).   + Issues identified (missing data, inconsistencies, anomalies).   + Recommendations for data cleaning and preprocessing. * **Exploratory Data Analysis (EDA) Report**: shows insights derived from analyzing the dataset to understand trends, patterns, and potential features for the model. Key components:   + Visualizations of key variables (e.g., trip distance vs. fare).   + Correlation analysis between features and fares.   + Identification of potential external factors (e.g., time of day, weather). * **Feature Engineering Document:** Documentation of selected features and their relevance to predicting taxi fares. Key components:   + List of engineered features (e.g., rush hour indicator, zone-based distance metrics).   + Justification for including each feature.   + Feature transformations and scaling methods applied. * **Regression Model:** A fully developed and trained regression model for fare estimation. Key components:   + Model architecture and parameters.   + Training dataset details and validation metrics (e.g., RMSE, MAE).   + Code and documentation for reproducibility. * **Model Evaluation Report:** summarizes the model's performance and any necessary optimizations. Key components:   + Validation results with error metrics.   + Comparison to baseline models (e.g., mean fare prediction).   + Limitations and suggestions for improvement. * **Graphical user interface and deployment plan:** A roadmap for deploying the regression model within TLC’s operational framework. Key components:   + Steps for integrating the model into TLC’s systems.   + Guidelines for model maintenance and updates.   + Training plan for TLC staff to use and interpret the model effectively. * **Final project reports.** * **Presentations for stakeholders.** |

## 

## **THE PACE WORKFLOW**



**[Alt-text: The PACE Workflow with the four stages in a circle: plan, analyze, construct, and execute.]**

You have been asked to demonstrate for the company's data team how you would use the PACE workflow to organize and classify tasks for the upcoming project. Select a PACE stage from the dropdown buttons. A few tasks involve more than one stage of the PACE workflow. Additionally, not every workplace scenario will require every task. Refer back to the Course 1 end-of-course portfolio project overview reading if you need more information about the tasks within the project.

### **Project tasks**

Following is a group of tasks your company’s data team has determined need to be completed within this project. The data analysis manager has asked you to organize these tasks in preparation for the project proposal document. First, identify which stage of the PACE workflow each task would best fit under using the drop down menu. Next, give an explanation of why you selected the stage for each task. Review the following readings to help guide your selections and explanation: [The PACE stages](https://www.coursera.org/learn/foundations-of-data-science/supplement/4OtHr/the-pace-stages) and [Communicate objectives with a project proposal](https://www.coursera.org/learn/foundations-of-data-science/supplement/79Ysh/communicate-objectives-with-a-project-proposal). You will later reorder these tasks within a project proposal.

1. **Evaluating the model:** Construct

Why did you select this stage for this task?

|  |
| --- |
| * Because it focuses on refining and validating the model built in the earlier phase to ensure its reliability and accuracy. |

1. **Conduct hypothesis testing:** Analyze **and** Construct

Why did you select these stages for this task?

|  |
| --- |
| * Hypothesis testing is part of the Analyze stage because it involves investigating relationships within the dataset, validating assumptions, and deriving insights based on statistical tests. * The Construct (Conduct) stage focuses on creating and implementing solutions or outputs based on prior analysis. While hypothesis testing is traditionally an analytical task, it can fit into the Construct stage if it is framed as part of the iterative process of refining and validating models or solutions. |

1. **Begin exploring the data:** Analyze

|  |
| --- |
| * Initial examination of the dataset to identify trends, patterns, and potential issues. * Fits the Analyze stage as it is part of understanding the data. |

1. **Data exploration and cleaning:** Analyze **and** Construct

|  |
| --- |
| Cleaning ensures data quality, while exploration provides initial insights.  Fits the Analyze stage because it involves identifying and resolving data-related challenges. |

1. **Establish structure for project workflow (PACE):** Plan

|  |
| --- |
| This task defines how the project will progress, identifying goals, timelines, and team roles.  Fits the Plan stage because it lays the foundation for all subsequent tasks. |

1. **Communicate final insights with stakeholders:** Execute

Why did you select this stage for this task?

|  |
| --- |
| Sharing project results and actionable recommendations.  Fits the Execute stage as it finalizes the project by delivering results to stakeholders. |

1. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

|  |
| --- |
| Calculating means, medians, and standard deviations to summarize data characteristics.  Fits the Analyze stage because it helps interpret data properties. |

1. **Visualization building:** Analyze **and** Construct

|  |
| --- |
| Creating charts and graphs to reveal patterns and support analysis.  Fits the Analyze stage since it involves interpreting and representing data visually. |

1. **Write a project proposal:** Plan

Why did you select this stage for this task?

|  |
| --- |
| Drafting a proposal ensures clarity on objectives, resources, and deliverables.  Fits the Plan stage as it focuses on organizing the project scope and direction. |

1. **Build a regression model:** Analyze **and** Construct

|  |
| --- |
| Developing a predictive model based on the dataset.  Fits the Construct stage as it involves creating solutions from data analysis. |

1. **Compile summary information about the data:** Analyze

|  |
| --- |
| Consolidating data insights for clarity and communication.  Fits the Analyze stage as it organizes findings for later use. |

1. **Build machine learning model:** Construct

|  |
| --- |
| Designing and training a machine learning model for advanced predictions.  Fits the Construct stage because it requires constructing and optimizing algorithms. |