**Course One**

# Foundations of Data Science



# 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?

The audience for this project is the New York City Taxi and Limousine Commission (TLC). They are the client and the main stakeholders who will benefit from the outcomes of this project.

* 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?

This project aims to develop a regression model that can accurately estimate taxi fares in advance. By accomplishing this, the project aims to provide the TLC with a valuable tool for fare estimation, which can improve customer satisfaction, optimize resource allocation, and enhance overall operational efficiency.

* What questions need to be asked or answered?

What variables and factors have the most significant impact on taxi fares?

How accurately can we estimate taxi fares using the available data?

Are there any specific patterns or trends in the data that can help improve the accuracy of the model?

* What resources are required to complete this project?

The successful completion of this project will require access to the TLC's data on taxi rides, including information such as trip details, fares, and other relevant variables.

Additionally, the project will require the necessary and trustworthy dataset and input from TLC’s stakeholder.

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

Global-level project document outlining the project structure and goals.

Project proposal presenting the project plan, objectives, and anticipated outcomes.

Data files ready for exploratory data analysis (EDA) after proper cleaning and preprocessing.

EDA report providing insights and observations from analyzing the data.

Tableau dashboard/visualizations to present the data in an interactive and informative manner.

Analysis of testing results between important variables to understand their relationship and impact on taxi fares.

Final regression model that accurately estimates taxi fares.

Model evaluation report assessing the performance and effectiveness of the developed model.

Final project report summarizing the findings, recommendations, and overall project outcomes.

## 

## **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 are 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:** Execute

Why did you select this stage for this task?

Because it involves the execution of training the regression model on the test data. We can assess the accuracy of the model based on the performance and by looking at the prediction, and metrics scores.

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

Why did you select these stages for this task?

Because in the Analysis stage, we have to look at the basics EDA to choose variables for hypothesis testing and we assume those variables are significantly related to the fare of the taxi. And during the Construction stage, we conduct tests to confirm our hypothesis if the relationships between the variables we choose are significant.

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

Why did you select this stage for this task?

Explore the data crucial to get familiar with the dataset, understand its structure, and identify any initial patterns or insights.

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

Why did you select these stages for this task?

Data exploration involves an in-depth analysis of the dataset, examining its distribution, identifying outliers or missing values, and gaining a comprehensive understanding of the data's characteristics. Data cleaning involves handling missing values, outliers, and any inconsistencies in the dataset. These stages are essential to ensure the data's quality, integrity, and reliability for further analysis.

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

Why did you select this stage for this task?

This stage allows for the development of a global-level project document that outlines the goals, objectives, milestones, and overall project plan. It provides a framework for effective project management, collaboration among team members, and alignment with the client's needs and expectations.

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

Why did you select this stage for this task?

This is the final steps in PACE. This stage involves summarizing and presenting the findings, results, and recommendations of the project to the relevant stakeholders.

1. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

This stage involves calculating and analyzing various statistical measures, such as mean, median, standard deviation, and correlation, to summarize and describe the key characteristics of the data.

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

Why did you select these stages for this task?

These stages involve selecting appropriate visualization techniques, creating meaningful charts, graphs, or dashboards, and presenting the data in an intuitive and informative manner. This is the final step in analyzing data to wrap-up the key insights of the data before starting of constructing model.

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

Why did you select this stage for this task?

This stage involves outlining the project's objectives, scope, methodology, deliverables, and timeline in a comprehensive document so it has to be done before anything else.

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

Why did you select this stage for this task?

During the analysis stage, the team will assess the dataset, identify relevant features, and perform data preprocessing. The Construct stage involves constructing the regression model using the selected algorithm, training it on the prepared data, and refining the model's parameters.

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

Why did you select this stage for this task?

This stage allows for the organization and documentation of essential information about the dataset, enabling effective communication and reference throughout the project.

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

Why did you select this stage for this task?

The Constructing stage is appropriate for building a machine learning model as it involves the practical implementation of the model-building process, ensuring that the model is operational and ready for further analysis and evaluation.