**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 New York City Taxi and Limousine Commission

* 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 are attempting to create a regression model to calculate the ride fare for customers using TLCs services.

* What questions need to be asked or answered?

Do we have the necessary data to determine ride fare?

Can a regression model deliver reliable results?

What is the best programming language to use?

* What resources are required to complete this project?

TLC’s database

Python IDE

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

Cleaned Dataset

EDA Results

Regression Model

Presentation (Complete with Visuals)

## 

## **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 and 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?

The model is evaluated to prepare it for the stakeholder meeting.

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

Why did you select these stages for this task?

The hypothesis is tested twice: once when the data is analyzed to see if there are relationships between the data points, and once when the model is constructed to see if the results meet expectations.

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

Why did you select this stage for this task?

The data is to be explored after you have planned out what data you need to analyze.

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

Why did you select these stages for this task?

Data is explored during the plan stage as the needed data is identified, and cleaned during the analyze stage as the data is readied for analysis.

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

Why did you select this stage for this task?

A general structure needs to be in place at the beginning of the project.

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

Why did you select this stage for this task?

Final insights are communicated with the stakeholders at the very end of the project.

1. **Compute descriptive statistics: Analyze**

Why did you select this stage for this task?

Descriptive statistics are computed when the data is analyzed.

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

Why did you select these stages for this task?

Visualizations are built during the analyze stage after relationships are identified between the data points.

Visualizations are also built during the construct phase when the results of the model are revealed.

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

Why did you select this stage for this task?

The project proposal belongs in the planning stage because it outlines the major deliverables that are created during the other stages.

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

Why did you select this stage for this task?

A regression model is built during the construct stage using the data from the analyze stage.

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

Why did you select this stage for this task?

The summary information is compiled after the analysis is complete.

1. **Build machine learning model: Select PACE stage**

Why did you select this stage for this task?