**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 are:

• New York City Taxi and Limousine Commission (TLC) 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 are trying to develop a regression model that helps estimate taxi fares before the ride, based on data that TLC has gathered.

The project will provide our client with accurate fare estimate for every ride, this will help increase the credibility and consistency in the clients company and increases their customer’s satisfaction.

* What questions need to be asked or answered?

What is the expected duration of the project?

What is the condition of the provided dataset?

What variables will be the most useful?

Are there trends within the data that can provide insight?

What steps can I take to reduce the impact of bias?

What is the target accuracy of the model?

* What resources are required to complete this project?

The resources required:

• Project proposal

• Relevant dataset

• Python notebook

• Input from stakeholders

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

The deliverables to be created:

• Cleaned data

• Model

• Visualizations

• Executive summary

## 

## **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 “evaluating the model” comes after developing it, developing the model is part of the Construct stage, then evaluation is made during the Execute stage to check if it meets the project goal.

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

Why did you select these stages for this task?

Because in Analyze stage, we get some initial insights/hypothesis from the data which requires testing, while in Construct stage, we build statistical models and test our hypothesis on such models.

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

Why did you select this stage for this task?

Because “exploring the data” will give us deep understanding of the data in the Analyze stage, it provides valuable initial insights about the data that is worth sharing with stakeholders.

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

Why did you select these stages for this task?

Because after collecting the data from primary and secondary sources, we have to make sure that the data is ready for analysis, which happens at the end of the Plan stage and the Analyze stage.

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

Why did you select this stage for this task?

Because this task is a fundamental basis for outlining the workflow of the whole project, so it must be part of the Plan stage.

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

Why did you select this stage for this task?

Because final insights are gained by the end of the Construct stage, this task is part of the Execute stage as it’s the final stage in the PACE workflow.

1. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Because descriptive statistics can play a vital role in the model development in Construct stage, so it must be computed at the Analyze stage to be used afterwards.

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

Why did you select these stages for this task?

Because visualizations begins with data assessment and is created during the construction stage.

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

Why did you select this stage for this task?

Because the project proposal is fundamental for all entities working on the project to understand what the project’s goals and objectives are, it must be done first at the Plan stage.

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

Why did you select this stage for this task?

Because regression models are examined using the clean data during the Analyze stage, we start building them at the Construct stage

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

Why did you select this stage for this task?

Because summary information is gained once we get our data clean and organized which is part of the Analyze stage.

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

Why did you select this stage for this task?

Because machine learning models are built after the hypothesis we have on data are confirmed, this is part of the Construct stage.