**Course One**

# Foundations of Data Science



# Instructions

Use this PACE strategy document to record your decisions and reflections as you work through this end-of-course project. As a reminder, this document is a resource that you can reference in the future and a guide to help consider responses and reflections posed at various points throughout projects.

# Course Project Recap

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

* Understand and assess the proposed scenario
* Demonstrate foundational knowledge of the data science workflow - PACE
* Articulate a data project proposal in the planning stage for cross-functional team members

# 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: Planning Stage**

* Who is your audience for this project?

Juliana Soto and Titus Nelson of the NYC TLC.

* 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?

Our goal is to build a statistical model that predicts taxi and limousine ride durations based on some predictors related to location and time of day. We expect that this model will allow taxis to be routed more efficiently, and that it will also give riders a better estimate of their ride duration.

* What questions need to be asked or answered?

Do we have enough data? What steps need to be taken to clean it? Is it unbiased? Who are the stakeholders for this project? What are the big tasks of this project, and how can we break them into achievable milestones. What visuals do the TLC executives want to see? What is the type of each of the predictor variables (continuous/discrete)? What will be the specifics of the A/B testing?

* What resources are required to complete this project?

The data set itself. Python. A database. The data team. A calendar. Online meeting software.

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

A presentation for the stakeholders. The model itself.

## **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: Automatidata](https://www.coursera.org/learn/foundations-of-data-science/supplement/XxgHa/course-1-end-of-course-portfolio-project-overview-automatidata) 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:** Construct

Why did you select this stage for this task?

We need to be at a point where we have constructed the model in order evaluate it. Moreover, evaluating a model is part of creating a better model.

1. **Conduct a hypothesis test:** Analyze **and** Construct

Why did you select these stages for this task?

Well, this is a statistical analytic technique, so it would take place in either of these two steps.

1. **Understanding the data:** Analyze

Why did you select this stage for this task?

You’ve got to do a preliminary analysis of the data.

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

Why did you select these stages for this task?

These are things you have to do early on in the project in order to make the more intense construction and evaluation steps work.

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

Why did you select this stage for this task?

You plan in the planning phase.

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

You do this in the executive phase.

1. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

This is a statistical step, but not a predictive step.

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

Why did you select these stages for this task?

You make visualizations so that you can understand the data, and so that the final stakeholders can understand the data too.

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

Why did you select this stage for this task?

You have to write a proposal before you can make the project happen.

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

Why did you select this stage for this task?

You build statistical models in these two steps.

1. **Inspect the data set for missing data:** Plan

Why did you select this stage for this task?

This is an early step.

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

You wouldn’t get down into this intense business until you’ve taken care of planning and doing a broad analysis of the data.