**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?

New York City Taxi and Limousine Commission (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?

Develop a regression model that predicts taxi and limousine ride durations. TLC has to regulate all kinds of for hire vehicles in NYC and by predicting ride durations based on various predictors they can monitor different kind of rides and use it for further updating various policies like standard prices and others essential which need the input of rides duration.

* What questions need to be asked or answered?

How the predictive model will help the clint?

What are the relationships between various Predictors?

What kind of model will suit the client needs? Do we even need a model?

What is the Data Pipeline of the Client?

* What resources are required to complete this project?

Python/R

Power BI/Tableau

SQL

TLC Internal Data Set containing Ride Details.

Other BI Tools if needed

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

1. Predictive Model Summary
2. Chosen Predictive Model
3. Visualizations
4. 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: 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?

Evaluating the model is done after constructing the models. We will evaluate different kinds of model and submit the most proper one is Execute phase.

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

Why did you select these stages for this task?

Hypothesis tests can be formed after doing preliminary examination of the data set and using the model we can perform these tests. That’s why chose Analyze and Construct Stage here.

1. **Understanding the data:** Analyze

Why did you select this stage for this task?

Performing Data Cleaning and EDA is essential for forming hypothesis and constructing models. Both of these tasks are done in Analyze phase.

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

Why did you select these stages for this task?

While we deal with the data in Analyze phase Planning phase set the scope of data we will work on.

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

Why did you select this stage for this task?

Plan stage is best for establishing a structure.

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

Final insight can only be shared in Execute phase.

1. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

As said before, EDA is done in Analyze phase.

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

Why did you select these stages for this task?

Visualizations are made during the EDA and in the final stage. First is used in observing the data and Second one is used in communication of the final results to the stakeholders.

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

Why did you select this stage for this task?

Plan stage is best stage to create a project proposal as it gives an overall vision of the project to all kinds of Stakeholders.

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

Why did you select this stage for this task?

Regression model is mainly made in the Construct phase.

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

Why did you select this stage for this task?

Data Cleaning is done in Analyze phase.

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

Models are made in construct phase.