**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 TLC team.

* 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’re trying to build a reliable prediction model for ride duration. The estimated ride duration would help the client identify which ride took more than expected.

* What questions need to be asked or answered?

\*\*We need to think as data professionals\*\*

How is the quality and consistency of the dataset?

Is there any important variable to consider?

Is there any input bias in the data

Budget of the project.

Why predicting ride duration? What are the problems?

What tools should be used?

* What resources are required to complete this project?

Dataset, jupyter notebook and packages

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

Prediction model, EDA - hidden information/relationship,

**Google answer** A data set scrubbed for exploratory data analysis, visualizations, statistical model, regression analysis and/or machine learning model.

## **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: Analyze**

Why did you select this stage for this task?

While evaluating the model, we’re trying to analyze the performance or accuracy of the estimated ride duration with validation data.

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

Why did you select these stages for this task?

During the analyzing stage, it is determined that a statistical test will be needed to check the hypothesis made. During the construction phase, the test is carried out which allows in validating hypotheses.

1. **Understanding the data: Analyze**

Why did you select this stage for this task?

We understand the data by tools like exploratory data analysis and hypothesis testing to understand relationships between features, to know the distribution of features, and more.

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

Why did you select these stages for this task?

Planning takes place first about the methods that need to be performed for data exploration and cleaning. Analyzing takes place when understanding the data using cleaning and exploration.

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

Why did you select this stage for this task?

In the planning stage, the team identifies the problem, solutions, tools requirements, and project workflow.

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

At the end of the analysis and modeling, once we’re confident about the insights, we share the insights with stakeholders with the help of visualization, documentation, and video.

1. **Compute descriptive statistics: Analyze**

Why did you select this stage for this task?

We investigate the data using the descriptive statistics during analysis to understand the data.

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

Why did you select these stages for this task?

We build visualizations to share the insights with the stakeholders and complete the project.

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

Why did you select this stage for this task?

Project proposal is the very first step in any project. It is either used to get an approval from the client and to give an overview of the project (intent and solution) to the higher-level.

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

Why did you select this stage for this task?

We will build regression models in this project to estimate the ride duration. This is help in understanding the data (analyze) through the relationship between a feature and ride duration. Models is the primary aim of this project, and it would be done in construct stage.

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

Why did you select this stage for this task?

We analyze the data identify missing values and clean the dataset

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

Why did you select this stage for this task?

The machine learning model are build at the construct stage for predicting or estimating a dependent feature.