**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 intended audience for Automatidata's newest project that helps estimate taxi fares before the ride is 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?

TLC wants Automatidata to develop an app that helps users to estimate their ride fare before embarking on their journey. This will help the users to know their exact fare and plan their payments in advance. Once this app goes live, our client could pull off in depth insights into the user's experience and continue making improvements to help facilitate their users.

* What questions need to be asked or answered?

What’s the condition of the current dataset?

How do current users of TLC get to know their fares?  
Are the fares fair according to the journey covered?

Does climate change affect the fares?

What is the protocol to be followed when either the rider or captain arrives late?

Does road condition affect fare?

How can we reduce bias if there is any?

* What resources are required to complete this project?

Historical data needs to be gathered from TLC

Python IDE

TLC’s input

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

Proposal, Cleaned data, Tested Machine Learning models, Visualizations/ Reports, 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 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?

Evaluating models against test data comes under the Execute stage. It’s because once the model is created in the construct stage, the next step is to test the model for its accuracy and to make sure it meets the project expectations.

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

Why did you select these stages for this task?

Hypothesis testing comes under Analyze and Construct because the data team needs to test against assumptions as well. Hence, in the analysis stage, they may come up with a statistical scenario which could be tested in the Construct stage to see how our model stands.

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

Why did you select this stage for this task?

In the analysis stage, data analysts need to explore the current state of data. They create a summary of the data showcasing redundancies, missing values, outlets that need to be dealt with in EDA.

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

Why did you select these stages for this task?

EDA happens in the plan and analyze stage where the data team first plans to carry out exploration of the project dataset and then cleans, transforms, explores and analyzes data.

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

Why did you select this stage for this task?

Establishing structure for project workflow is the foremost task of any project, hence this task comes under the plan stage where data professionals lay out the entire project’s schedule.

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

Why did you select this stage for this task?

Communicating insights with stakeholders is the final stage of any project and comes under the Execute stage. Visualization and insights generated need to be shown to the relevant stakeholders to get feedback and give recommendations.

1. **Compute descriptive statistics: Analyze**

Why did you select this stage for this task?

In the analysis stage, data analysts compile results by comparing 2 or more relevant variables that affect the project the most.

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

Why did you select these stages for this task?

Visualization building begins in the analyze stage where data is explored and analysis is made. It’s upon those findings that visualizations are mapped upon in the construct stage.

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

Why did you select this stage for this task?

Project proposal is one of the key initial tasks that needs to be made before the work begins. Hence, the proposal is cultivated in the plan phase.

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

Why did you select this stage for this task?

All the data that has to be fed onto the desired model is finalized in the analyze phase. Once complete, the models are built upon that data in the construct phase.

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

Why did you select this stage for this task?

Data is gathered, explored, cleaned in the analyze phase, hence data overview is compiled in that stage as well.

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

Why did you select this stage for this task?

Construct phase as the name suggests is the stage where models are created.