**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?

 **Automatidata team (technical):** Udo Bankole, Deshawn Washington, Luana Rodriquez, Uli King

 **TLC stakeholders (non-technical):** Juliana Soto, Titus Nelson

* 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?

 Build a regression model to estimate taxi fares before rides.

 Deliver a rider-facing fare estimation tool.

 Anticipated impact: improved transparency, customer trust, and operational insights for TLC.

* What questions need to be asked or answered?

 What variables (distance, time of day, surcharges, locations) impact fares?

 How accurate is the model compared to actual fares?

 What data quality issues exist (missing values, outliers)?

 What success metrics will determine if the project is ready for deployment?

* What resources are required to complete this project?

 Historical TLC trip data.

 Python environment for analysis and modeling.

 Visualization tools (e.g., Tableau/Matplotlib).

 Team expertise (EDA, regression modeling, stakeholder communication).

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

 Global-level project document with milestones.

 Clean dataset ready for EDA.

 EDA report and descriptive statistics.

 Hypothesis test and A/B test results.

 Regression model and machine learning model results.

 Visualizations and dashboards.

 Final report and presentation for stakeholders.

## 

## **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?

Model evaluation requires testing (Construct) and confirming performance for delivery (Execute)

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

Why did you select these stages for this task?

Testing begins with analysis of variables (Analyze) and supports model validation (Construct)

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

Why did you select this stage for this task?

Early stage data exploration helps identify variables and patterns

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

Why did you select these stages for this task?

Data cleaning is part of analysis, while preparing it for modeling is construction

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

Why did you select this stage for this task?

Project planning defines workflow, milestones, and goals

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

Why did you select this stage for this task?

Deliverables and reporting to TLC are part of execution.

1. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Descriptive stats help summarize and understand dataset characteristics

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

Why did you select these stages for this task?

Visualizations support analysis internally and are final deliverables for stakeholders.

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

Why did you select this stage for this task?

Proposal is part of planning before analysis begins

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

Why did you select this stage for this task?

Construct involves building, while Analyze is used to test performance

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

Why did you select this stage for this task?

Summaries are part of understanding the dataset.

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

Why did you select this stage for this task?

**Model building is part of construction phase**