

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:

- ☒ 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

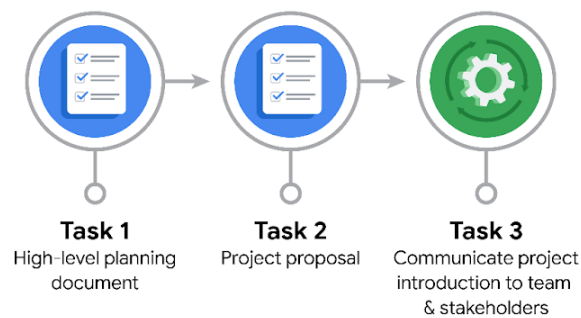
Completing this end-of-course project will empower you to respond to the following interview topics:

- 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: Planning Stage

- Who is your audience for this project?

The audience for this project is New York City TLC Team Members - **Juliana Soto**, Finance and Administration Department Head and **Titus Nelson**, Operations Manager.

- 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?

“Automatidata” and NY City TLC work on a project to predict the duration of rides that will depend on location and time of day. The project is a ML model, and with integration into the company’s system, it would help both clients and taxi/limo drivers to be aware of the situation they are in when ordering/accepting rides.

- What questions need to be asked or answered?

- Where do we get the data from?
- Who is responsible for the data our client provides, in case we have further questions about its quality, validity, etc.
- What are the deadlines? Is the project supposed to be completed within a short amount of time?
- Is there something else we can do to provide some useful information using the data they provided? Maybe they want to know something else?



- What resources are required to complete this project?

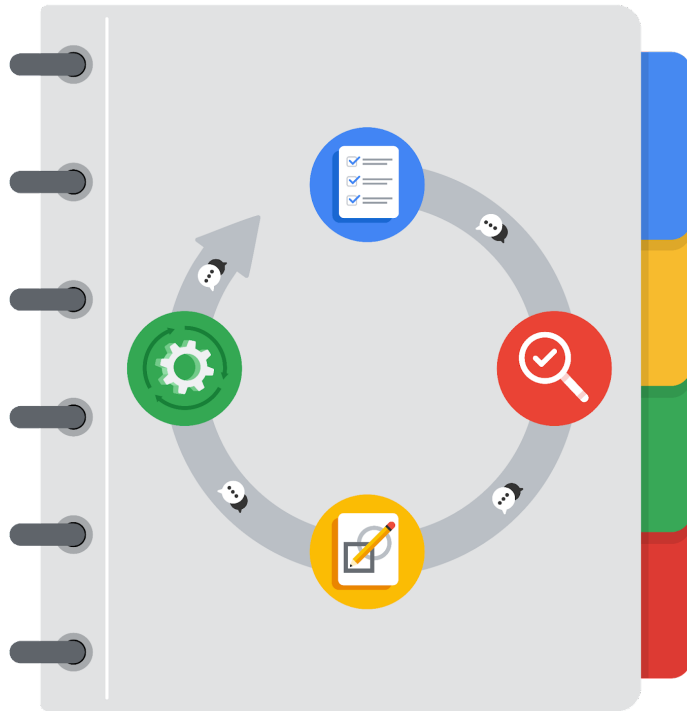
Python - for getting, cleaning, analyzing, visualizing and building the ML model

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

Reports - for both the project team and our stakeholders. Everyone should be in the loop and be updated. Setting up alignment meetings would also help a lot in case there are complications / changes to be made.



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: Automating data](#) 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:** Construct ▾

Why did you select this stage for this task?

Before we present anything to our clients, we need to be sure our model works as well as it should. When we are certain about it being robust - we can move to the next stage and present our work

2. **Conduct a hypothesis test:** Analyze ▾ and Construct ▾

Why did you select these stages for this task?

It is a good idea to go through the data before actually working on a ML model. Maybe it can give us something interesting to know about and we could present it to our clients.

3. **Understanding the data:** Analyze ▾

Why did you select this stage for this task?

After we get the data, we should know what it has or has not. We should define what every observation stands for and what are the features of these observations.

4. **Data exploration and cleaning:** Analyze ▾ and Construct ▾

Why did you select these stages for this task?

Both of these stages require clean and concise data to work with

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

Why did you select this stage for this task?

Before actually doing anything - we should have a plan about when and how we do everything

6. **Communicate final insights with stakeholders:** Execute ▾

After everything was said, done and made, it is time for us to present our work

7. Compute descriptive statistics: **Analyze**

Why did you select this stage for this task?

We can check whether patterns or relationships exist within the data. Also, summarizing the data is always a good idea.

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

Why did you select these stages for this task?

Visualization is a universal language that helps us communicate with each other without giving into technicalities. Visualizations in the Execute stage are meant for our clients. Analyze - for me and our team.

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

Why did you select this stage for this task?

After defining what our client needs we should also define the solution we offer to him before actually getting into work

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

Why did you select this stage for this task?

Is a final product of our project for the stakeholders

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

Why did you select this stage for this task?

Missing data can cause a lot of problems for our model which can be undertrained. We need to address that.

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



Before the final stage and after the data is ready is a perfect spot