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~~

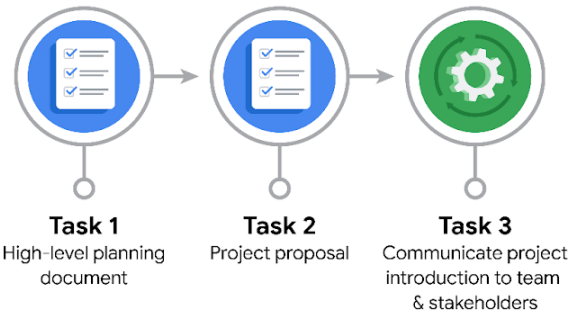
Relevant Interview Questions

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?

  Build a regression model that predicts ride durations based on distance, time of day, season, and additional variables as found necessary.

  It would be reasonable to anticipate that accurate ride duration prediction models would provide valuable insights into demand for taxi rides throughout the city. Variables such as distance, time of day, season and other factors will provide the details necessary for TCL to prepare their fleet of taxis.

- What questions need to be asked or answered?

  It would be responsible for us to confirm any personally identifiable information (PII) has been removed or censored from the data or whether that requires effort on our end. Any additional data stewardship and aggregation procedures need to be considered and documented.

  We also need details on who the users of our deliverables are. Management and operations may want a visual dashboard that outlines the global trends and modeled predictions with interfaces connected to variables and filters that make trend exploration accessible and understandable. Taxi drivers, on the

other hand, may prefer something more straightforward such as push notifications that provide them with a message in anticipation of a change in ride demand based on the model. Sales and marketing teams may prefer a more computational tool that outputs spreadsheets with data on demand, cost, and profit models.

- What resources are required to complete this project?

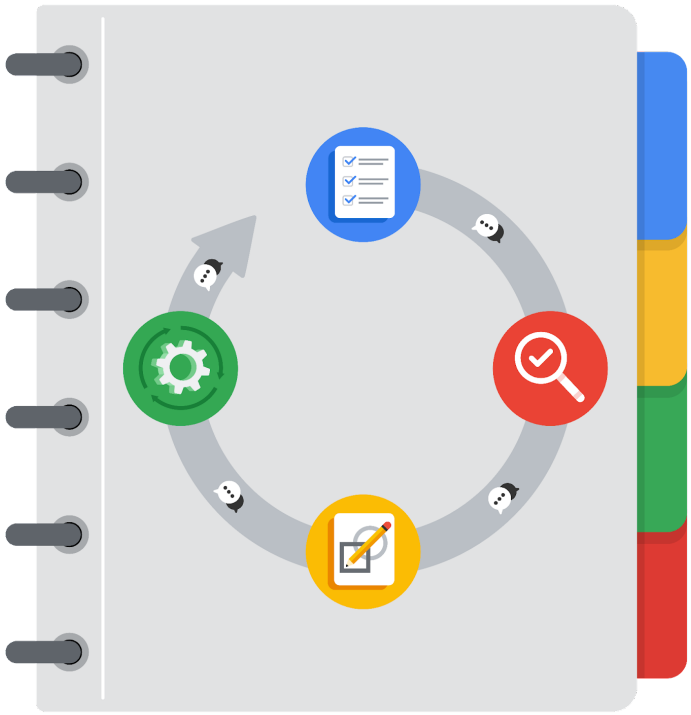
We require access to the aforementioned dataset along with any legal and terms based practices clearly outlined.

We need access to at least one subject matter expert from each area of the data collection team, Taxi drivers, and management for correspondence during the research and development processes.

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

Depending on who the intended audience is for our project [see previous answer in questions that need to be answered section] we will be developing at least one predictive model, as many as three dashboard or insights interfaces, and documentation for the process employed by any predictive models and their limitations, as well as documentation for any dashboards or user facing insights providers that detail how they function, their intended use, and outlining any APIs for integration with any of TLCs larger technical products.

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 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 **and** Execute

Why did you select this stage for this task?

The construction phase will emphasize the building and revising of any and all machine learning models. As we revise each iteration of the models we will be tuning for performance, accuracy, and usability of the results. Model evaluation will be a critical component of the construction phase. Finally, after the model has been constructed, data is run through to evaluate whether it meets the project's expectations and goals.

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

Why did you select these stages for this task?

During the analyzing stage, it is determined that a statistical test will be used. During the construction phase, the test is carried out.

3. **Understanding the data:** Analyze

Why did you select this stage for this task?

When cleaning and formatting, we will gain a deeper understanding of the data.

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

Why did you select these stages for this task?

Planning takes place when we first make choices about the methods needed. The cleaning process then takes place in the analyzing stage.

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

Why did you select this stage for this task?

Creating an initial project PACE document outlines the workflow and helps to plan how to best approach a project.

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

Communication is necessary at various points throughout a project. Final insights are shared with stakeholders in the execute phase of the data project workflow.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

Investigating the statistics within data takes place during the analysis process.

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

Why did you select these stages for this task?

Visualization begins with data sessment and is created during the construction stage

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

Why did you select this stage for this task?

A project proposal is the initial document used to define a project.

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

Why did you select this stage for this task?

The regression model is first constructed. Then, during the analyzing stage the model is examined in detail to be sure it will meet the needs of the task.

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

Why did you select this stage for this task?

Inspecting a data set for missing data would take place while assessing the quality of the data.

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

The building of a data model would take place in the construction stage.