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.~~

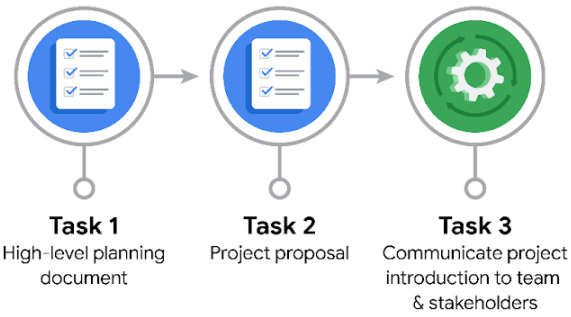
Relevant Interview Questions

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: Planning Stage

- Who is your audience for this project?

  The audience for this project is the New York City Taxi and Limousine Commission.

- 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?

  The problem we are trying to solve is to estimate the duration of rides based on their location and the time of day.

- What questions need to be asked or answered?

  1. What is the condition of the provided dataset?
  2. What variables will be the most useful?
  3. Are there trends within the data that can provide insight?
  4. What steps can we take to reduce the impact of bias?

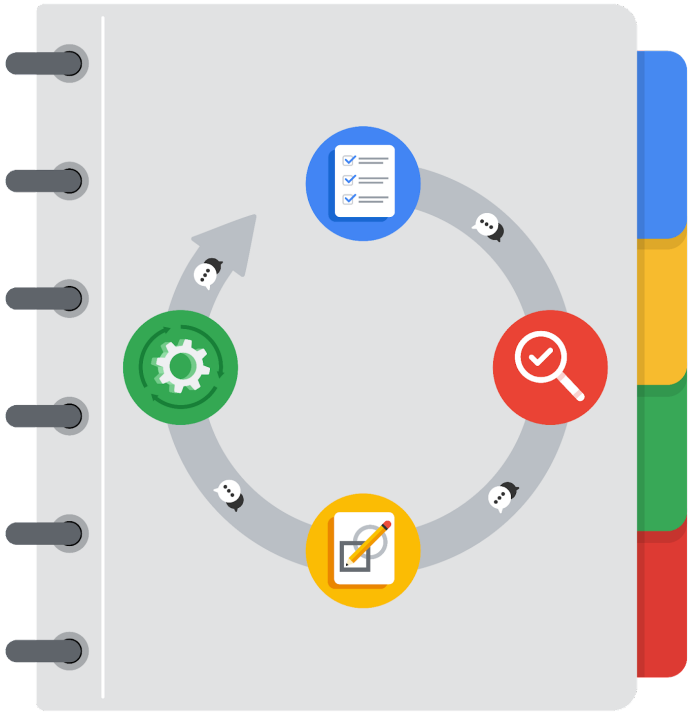
- What resources are required to complete this project?

  1. The project data set,
  2. Python notebook, and
  3. Input from stakeholders.

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

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 to 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 is 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?

Once the model is built, the data is processed to assess whether it aligns with the project's objectives and desired outcomes.

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

   Why did you select these stages for this task?

In the analysis stage, a statistical test is chosen, which is then performed during the construction phase.

3. **Understanding the data: Analyze**

   Why did you select this stage for this task?

During the data cleaning and formatting stage, you can develop a more comprehensive understanding of your data.

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

   Why did you select these stages for this task?

The planning stage involves making decisions about the required methods, while the data cleaning process occurs during the analysis stage.

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

   Why did you select this stage for this task?

During the planning stage, drafting an initial Project PACE document can help define the project's workflow and determine the optimal approach.

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

Communication plays a crucial role at different stages of a project, and in the execution phase of the data project workflow, the final insights are shared with stakeholders.

7. **Compute descriptive statistics: Analyze**

   Why did you select this stage for this task?

The process of investigating the statistics within data occurs during the analysis phase.

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

   Why did you select these stages for this task?

   The creation of visualizations begins with data assessment and continues during the construction stage.

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

   Why did you select this stage for this task?

   During the planning stage, a project proposal serves as the primary document used to define a project.

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

    Why did you select this stage for this task?

    The first step is constructing the regression model, followed by a detailed examination during the analyzing stage to ensure it satisfies the task requirements.

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

    Why did you select this stage for this task?

    While assessing the quality of the data, one of the tasks in the analyzing stage is to inspect the dataset for any missing data.

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

    The construction of a data model is carried out during the construct stage.