

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.

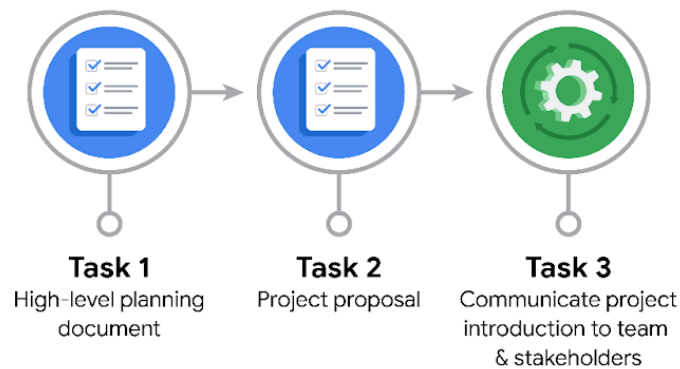
Relevant Interview Questions

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?

Audience will be 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 is to solve the estimation of taxi fares based on the identifying relevant variables given in data

- What questions need to be asked or answered?

1. What is the condition of the data provided by the company?
2. Which variables are most important to get valuable insights from the data?
3. Are these trends within the data helpful for insights?
4. What steps can I take to reduce biases?
5. What are the expectations of the stakeholders?



6. What steps I can take to meet project timeline?
7. Model selection, metrics selection for evaluation, feedback mechanism?

- What resources are required to complete this project?

Required resources are project dataset, human resources, technical resources, financial resources, communication tools, documentation resources etc.

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

The deliverables include dataset/ data collection and preprocessing, exploratory data analysis(EDA), data visualization, statistical models, model development, model evaluation, user interface tool etc.

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?



I selected this task because after the model has constructed, data is run through to evaluate either the model is working fine and meet the project's goals and expectations too.

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

Why did you select these stages for this task?

After Plan stage, in hypothesis testing, data is analyzed to through Statistical strategies and statistical tests are applied to test the relationships between or differences between groups, which is very fundamental process to test process. While in Construct stage, this statistical test is carried out. This stage helps in decision making and for further research process.

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

Why did you select this stage for this task?

During the analyzing phase, we get the deeper understanding of the dataset and we understand the story about the data. We explore the data, identify patterns, apply statistical tests, and visualize the data to get better understanding of the dataset.

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

Why did you select these stages for this task?

Plan, in some extent may be used when we have to select specific methods we needed in the required problem solving. Analyze stage is necessary for both data exploration and analyzing the data. Data cleaning is very crucial step in analyzing phase.

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

Why did you select this stage for this task?

Planning stage is used for creating an initial project PACE document. It outlines the workflow and helps to plan how to best approach a project.



6. Communicate final insights with stakeholders: Execute

Why did you select this stage for this task?

Execute stage is used in this context because communication with the stakeholders is very important throughout the project. So, final insights are shared with stakeholders in execute phase.

7. Compute descriptive statistics: Analyze

Why did you select this stage for this task?

By applying statistical techniques within data to examining the data is the analyzing phase of PACE strategy. Descriptive statistics is the key component in data analyzing. Statistical approaches are very useful for understanding data characteristics.

8. Visualization building: Analyze and Construct

Why did you select these stages for this task?

Visualization is used for data assessment or for data analyzing. It is created during the construction stage.

9. Write a project proposal: Plan

Why did you select this stage for this task?

Plan stage is used in project proposal because it is an initial document to define a project need.

10. Build a regression model: Analyze and Construct

Why did you select this stage for this task?



During the analyze phase in PACE strategy, choose the best model to examine either it meets the requirements of the project. The building of regression model will take place in action during the construct stage where implementation start on the selected regression model.

11. Compile summary information about the data: Analyze

Why did you select this stage for this task?

Analyze stage is used for summary information because key characteristics and initial insights are provided in analyze stage after deeply analyzing the data.

12. Build machine learning model: Construct

Why did you select this stage for this task?

Construct stage of PACE strategy is used for building a machine learning model after analyzing and cleaning the data ready for machine learning model.