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.~~

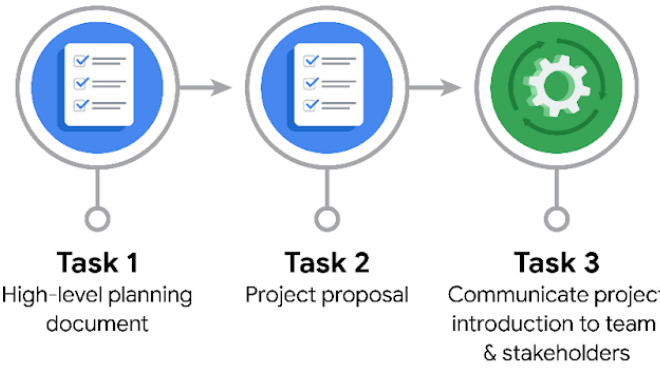
Relevant Interview Questions

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?

The Automatidata data team, including data scientists, project managers, and business stakeholders at NYC 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?

Develop a regression model to predict NYC taxi fare prices using historical TLC data. The goal is to optimize pricing transparency and enhance ride estimation accuracy for both drivers and passengers.

- What questions need to be asked or answered?

Which variables (distance, time, passengers) most influence fare price?

How accurate is our model in predicting fares?

Are there time-based or location-based fare anomalies?

- What resources are required to complete this project?

NYC TLC trip dataset (CSV files)

Python/R for modeling

Jupyter Notebooks or Colab

Tableau/Looker Studio for dashboards

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

Cleaned dataset for EDA

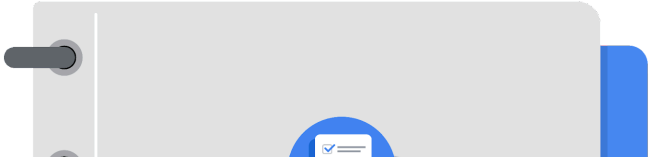
Exploratory data visualizations

Regression model and evaluation report

Tableau dashboard

Stakeholder report

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?

Evaluating the model is part of the final phase where results are assessed against performance metrics like RMSE and R^2. It ensures the model meets business requirements and is ready to be deployed or shared with stakeholders.

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

   Why did you select these stages for this task?

Hypothesis testing is a statistical method used during the data exploration and analysis phase to determine relationships or differences between variables. It helps validate assumptions before modeling.

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

Why did you select this stage for this task?

This is an early analytical task to understand data distributions, patterns, or anomalies. It informs which variables may be useful for modeling and highlights data quality issues.

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

Why did you select these stages for this task?

Exploration is part of analysis, while cleaning (handling missing values, correcting data types) involves transformation, which is part of construction.

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

Why did you select this stage for this task?

Setting up the workflow is a strategic task done at the beginning. It aligns the team on project stages, milestones, and deliverables.

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

Why did you select this stage for this task?

Delivering final insights involves presenting outcomes, visualizations, and recommendations to decision-makers, which happens at the end of the project.

7. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

This step helps summarize and describe dataset features (e.g., mean, median, standard deviation), which informs further modeling decisions.

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

Why did you select these stages for this task?

Creating dashboards and charts involves technical development (construct), and they are used to communicate results to stakeholders (execute).

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

Why did you select this stage for this task?

The project proposal outlines objectives, timeline, resources, and deliverables—core components of the planning phase.

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

Why did you select this stage for this task?

Model building is a hands-on technical task where you apply analytical methods to construct a predictive solution.

11. **Compile summary information about the data:** Plan

Why did you select this stage for this task?

Compiling summary information is typically done at the start to inform the team about data contents, sources, and structure.

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

Why did you select this stage for this task?

This task refers to the actual development of a machine learning algorithm using cleaned and prepared data, which is part of the construction phase.