**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?

This project’s audience consists of the data team (Harriet Hadzic, May Santner, Chidi Ga, Sylvester Esperanza, and myself) as well as Finance and Administration Department Head Emrick Larson and Operations Manager Ursula Sayo.

* 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?

Our goal is to build a machine learning model that predicts user churn, which will help Waze to prevent user churn, improve user retention, and increase growth.

* What questions need to be asked or answered?

We need to inspect the data for cleanliness and completeness.

We need to identify any relationships in the data (hypothesis testing has been suggested), as well as which variables are most useful for prediction.

* What resources are required to complete this project?

We’ll need the dataset itself and Jupyter notebook environment (I’m partial to Jupyter notebook in Visual Studio Code). I think it would be a good idea to set up a git repo using the cookie cutter data science template so that we can all use our notebooks in the same place while also having the option to develop a shared script. A strong line of communication with the non-data team people involved will also be important.

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

Deliverables required to proceed are a cleaned dataset.

The final deliverables are the model itself, and visualizations and talking points to share with executives.

## 

## **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](https://www.coursera.org/learn/foundations-of-data-science/supplement/4OtHr/the-pace-stages) and [Communicate objectives with a project proposal](https://www.coursera.org/learn/foundations-of-data-science/supplement/79Ysh/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?

Evaluation is part of the iterative process of constructing a model. We construct a model and then either accept the model or refine it based on evaluation metrics.

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

Why did you select these stages for this task?

This is a statistical step, so it needs to take place during the analyze and construct stages.

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

Why did you select this stage for this task?

After formulating a plan, we have to explore the data so we can determine which statistical methods are most appropriate for analyzing it.

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

Why did you select these stages for this task?

Our first exploration of the data will likely reveal a multitude of data oddities that will need to be cleaned out. We’ll need to form a cleaning plan, develop it, and then analyze if it was effective and if there was anything we missed.

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

Why did you select this stage for this task?

You’ll want to have some form of plan before diving into a project like this.

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

Why did you select this stage for this task?

Once we do this, we’re essentially telling our executives that we’re done and then giving them a list of insights. They’re then going to act on these insights.

1. **Compute descriptive statistics:** Analyze

Why did you select this stage for this task?

This is part of the output of the data analysis phase.

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

Why did you select these stages for this task?

We’ll need visualizations during the analyze phase to get a grasp on what the data means. Our executives will need visualizations during the execute phase for the same reason. I would also add these to the construct phase, since these can help us evaluate our model.

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

Why did you select this stage for this task?

You can’t start a collaborative project like this without some form of proposal.

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

Why did you select this stage for this task?

In the analyze phase, a regression model can give us a basic view of how the variables in the data set are related.

In the construct phase, a regression model can be one of many candidate models that we use to produce predictions.

1. **Compile summary information about the data:** Analyze

Why did you select this stage for this task?

This is also part of the analysis phase’s output.

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

Why did you select this stage for this task?

Machine learning models usually refer to prediction algorithms at least as complex as a linear regression, likely more so, so we wouldn’t want to attempt this until we’re past the analysis phase.