

# 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 is aimed at providing company leadership with information on business decisions regarding the Waze application.

- 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 team and I are trying to predict user churn on the Waze app, specifically highlighting how many people have uninstalled or stopped using the app. This project will help target users what Waze stands to lose based on predictive models in order to take action to retain them. This effort will support the priorities of Waze with their currently highest priority on retaining current users, growth through taking on more users and increased use of Waze with current user base.

- What questions need to be asked or answered?



- Who are the current users of Waze?
- What software is currently being used to evaluate user numbers?
- What resources are available for this project?
- What is the structure, availability level and storage for the current datasets?
- What factors are most common among Waze users?
- What are common factors among people who do not use Waze?
- Who are the users most in danger of churning?
- Why would a user churn?
- What situations increase the likelihood of a user churning?
- Where are Waze users located or what is its relative popularity geographically?
- Where are the largest number of Waze users?
- What bias is contained in the source data?

- What resources are required to complete this project?

Need access to the dataset.

Need people to support clean data and research model options.

Will need a common IDE, workbook space to work in Python and collaborate with team.

What is the availability of stakeholders for questions?

Will need a common communication channel or strategy for updates.

Will need a data pipeline.

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

Global-level project doc to outline goals and milestones.

Visuals to share with executives.

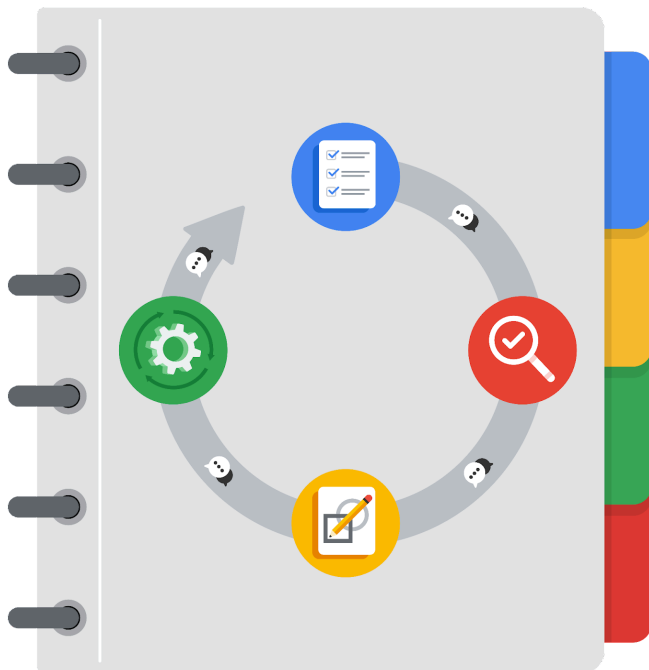
Dataset must be cleaned and approved.

Proposed models must be created and approved.

Presentations to the leadership team.



## 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: Construct ▾



Why did you select this stage for this task?

We will need to have selected a built a model approach before being able to evaluate it and we will want to have approval on the model before presenting it as a final result to leadership.

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

Why did you select these stages for this task?

We will be alternating between making a plan and testing that plan through analysis.

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

Why did you select this stage for this task?

Before performing an formal changes on the data, we will simply need to familiarize ourselves with it in order to anticipate stakeholder's needs.

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

Why did you select these stages for this task?

We will be exploring the data during the overall analysis phase because it is the same stage as when we are formatting the database as a whole and cleaning data or converting it into correct formatting. Exploring and cleaning go hand in hand.

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

Why did you select this stage for this task?

Developing a workflow needs to happen as early on as possible in the project in order to make sure that work is not being repeated and that all teams are aligned.



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

Why did you select this stage for this task?

There would be nothing to communicate to end stakeholders until we have finished the construction stage or building models etc.

7. **Compute descriptive statistics:** **Construct** ▾

Why did you select this stage for this task?

During the construction stage, we will have models available in order to generate these values and statistics and be ready to share in the Execution stage.

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

Why did you select these stages for this task?

We would require visualizations throughout the Construct phase to support the work of the team and clarify understanding. These visuals would then need to be updated, simplified or changed in order to work for stakeholders.

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

Why did you select this stage for this task?

This needs to be the first priority in a project to narrow the focus down and eliminate scope creep.

10. **Build a regression model:** **Construct** ▾ and **Execute** ▾

Why did you select this stage for this task?



**Models will be confirmed in the Construct phase and presented to leadership. Additional models could be requested by leadership within the Execute phase.**

**11. Compile summary information about the data:** **Execute** ▾

Why did you select this stage for this task?

Feedback will be provided at the end of the Execute phase and summaries will be very valuable to outliers who still need to be informed of the project results.

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

Why did you select this stage for this task?

**A machine learning model along with any other models would be built within the Construct phase as we would not have enough information in the Analyze phase and such a model needs to be built and tested before presenting to the stakeholders.**