

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

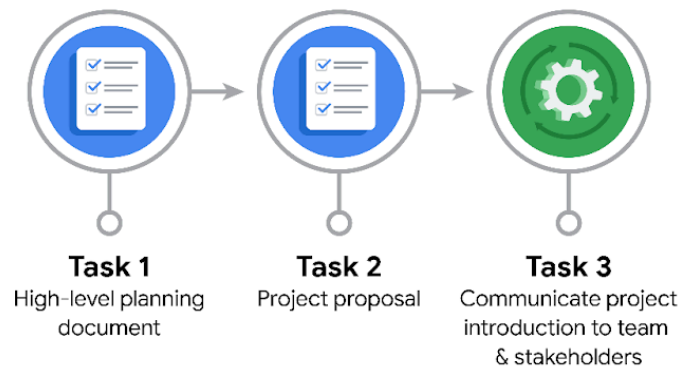
Completing this end-of-course project will empower you to respond to the following interview topics:

- 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?

L'audience principale est composée de l'équipe de direction de Waze et de la responsable de l'analyse des données, **May Santner**.

Ces parties prenantes souhaitent comprendre comment un modèle prédictif peut aider à **réduire le taux de désabonnement des utilisateurs**, à **améliorer la fidélité** et à **soutenir la croissance continue de l'entreprise**

- 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?

L'objectif est de **concevoir un modèle d'apprentissage automatique** capable de **prédire le risque de désabonnement des utilisateurs**.

Ce modèle permettra d'identifier les utilisateurs susceptibles de quitter la plateforme et d'agir à temps grâce à des stratégies de rétention ciblées.

L'impact global sera une **augmentation de la rétention**, une **optimisation des campagnes marketing**, et une **amélioration continue de l'expérience utilisateur**.



- What questions need to be asked or answered?

Quelles variables influencent le désabonnement (fréquence d'utilisation, durée de session, notifications, localisation, type d'appareil) ?

Quel est le pourcentage actuel d'utilisateurs qui se désabonnent ?

Comment définir un "désabonnement" (utilisateur inactif pendant X jours, désinstallation, etc.) ?

Quelle quantité et qualité de données avons-nous ?

Quel type de modèle (régression logistique, arbres de décision, random forest, XGBoost) offre la meilleure précision ?

Comment mesurer les performances du modèle (précision, rappel, F1-score, AUC) ?

- What resources are required to complete this project?

Ressources humaines :

- May Santner (Gestionnaire de l'analyse des données)
- Équipe de data scientists et data analysts
- Ingénieurs machine learning et data engineers

Ressources techniques :

- Python, scikit-learn, TensorFlow
- Google Cloud Platform / BigQuery
- Tableau ou Power BI pour la visualisation

Ressources données :

- Historique d'utilisation des utilisateurs Waze
- Données de session, fréquence, durée, abandon, localisation
- Retours utilisateurs (feedback, désinstallation, notifications)

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

Proposition de projet (document global de planification)

Données nettoyées et prêtes pour l'analyse

Rapport d'exploration et de statistiques descriptives

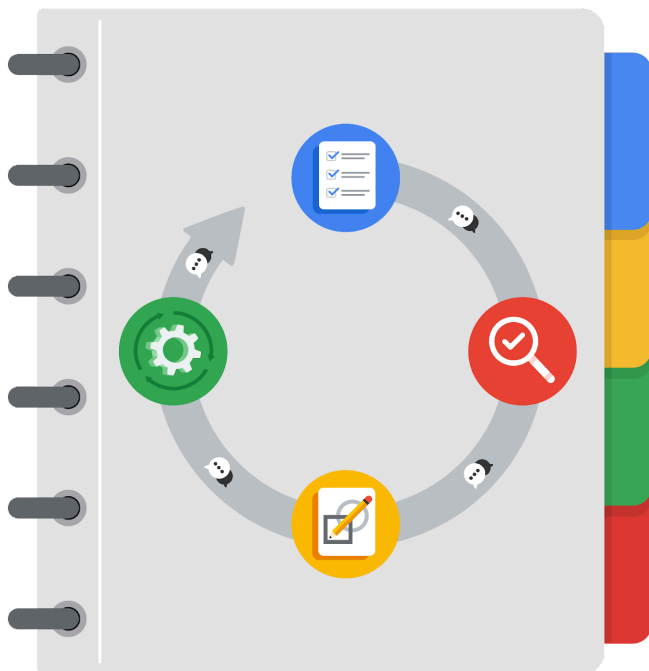
Tests d'hypothèse sur les facteurs de désabonnement

Visualisations et tableau de bord de l'analyse exploratoire

Modèle prédictif de désabonnement

Rapport d'évaluation du modèle et recommandations finales

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: **Analyze** ▾

Why did you select this stage for this task?

On évalue la performance du modèle sur des données de validation (AUC, précision, rappel, F1, courbe PR), on vérifie l'overfitting et on fixe le seuil de décision. La présentation des résultats/actionnables relève d'**Execute**.

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

Why did you select these stages for this task?

On formule et teste des hypothèses (t-test, χ^2 /ANOVA) sur les facteurs de churn (**Analyze**) puis on intègre les enseignements dans la sélection/ingénierie des variables pour la modélisation (**Construct**).

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

Why did you select this stage for this task?

- L'EDA permet de comprendre distributions, valeurs manquantes, types, outliers et premières corrélations avant tout nettoyage ou modélisation.



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

Why did you select these stages for this task?

On nettoie/déduplique, gère les valeurs manquantes, normalise, encode, puis on vérifie l'effet de ces transformations par une EDA de contrôle.

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

Why did you select this stage for this task?

Définition des rôles, jalons, outils (Python/BigQuery), conventions de versioning, risques et planning — c'est la planification du projet.

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

Why did you select this stage for this task?

On restitue résultats et recommandations (impact business, segments à risque, plan d'activation) via présentation/tableau de bord aux décideurs.

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

Why did you select this stage for this task?

Calcul des statistiques de base (moyenne, écart-type, médianes, taux de churn, corrélations) pour cadrer le problème et guider la suite.

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

Why did you select these stages for this task?

On **construit** les visualisations/tableaux de bord (segmentation churn, entonnoirs d'usage) puis on les **partage** pour la prise de décision.



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

Why did you select this stage for this task?

Le document de proposition fixe l'objectif, la portée, les livrables, les ressources et le calendrier selon PACE.

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

Why did you select this stage for this task?

Mise en place d'un modèle de régression (logistique) comme baseline (Construct) et analyse des coefficients/performances pour l'interprétabilité (Analyze).

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

Why did you select this stage for this task?

Synthèse des sources, schéma, dictionnaire de variables, volumétrie, fraîcheur et qualité des données pour cadrer l'étude.

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

Why did you select this stage for this task?

Entraînement et tuning de modèles ML (Random Forest, XGBoost, etc.), validation croisée et préparation de l'artefact de scoring.