

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.

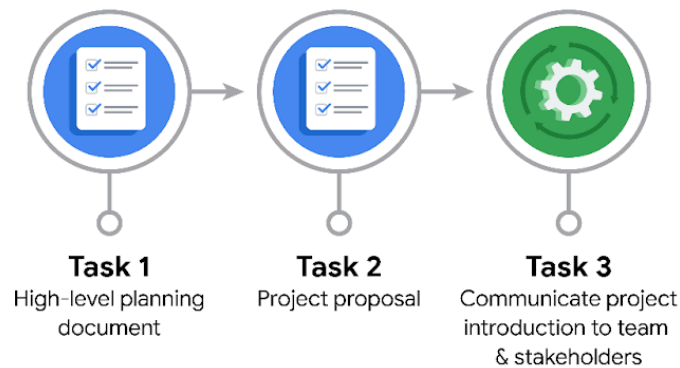
Relevant Interview Questions

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 TikTok, des responsables de la science des données (Willow Jaffey et Rosie Mae Bradshaw), ainsi que des membres du bureau de gestion de projet (Mary Joanna Rodgers). Ces parties prenantes ont besoin de comprendre la stratégie, les livrables et l'impact du modèle d'apprentissage automatique sur le processus de modération du contenu.

- 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 prédictif capable de classer automatiquement les affirmations faites dans les vidéos TikTok afin de réduire le nombre de rapports d'utilisateurs à traiter manuellement. L'impact attendu est une modération plus rapide et plus précise, une amélioration de la sécurité de la plateforme et une meilleure expérience utilisateur grâce à une priorisation efficace du contenu signalé.



- What questions need to be asked or answered?

Quelles sont les caractéristiques (texte, métadonnées, comportements d'utilisateurs) les plus utiles pour détecter les affirmations ?

Quelle quantité de données est disponible et quel est leur niveau de qualité ?

Quel type de modèle d'apprentissage automatique conviendrait le mieux (régression logistique, arbres de décision, réseaux de neurones, etc.) ?

Comment mesurer l'efficacité du modèle (précision, rappel, F1-score) ?

Quels sont les critères éthiques et de confidentialité à respecter pour traiter ces données ?

- What resources are required to complete this project?

Ressources humaines : l'équipe de science des données de TikTok, les ingénieurs machine learning et les chefs de projet.

Ressources techniques : Python, bibliothèques de machine learning (scikit-learn, TensorFlow), environnement cloud sécurisé, outils de visualisation (Tableau).

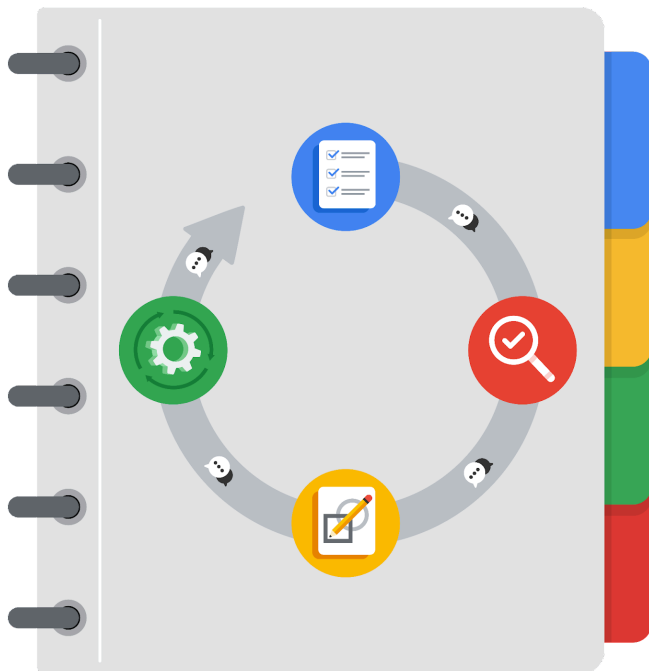
Ressources données : jeux de données TikTok contenant les vidéos, les commentaires, les rapports d'utilisateurs et leurs métadonnées.

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

- Proposition de projet (document global avec jalons).
- Rapport d'analyse exploratoire des données (EDA).
- Résultats de tests d'hypothèse.
- Visualisations des données et tableau de bord.

- Modèle de régression et modèle d'apprentissage automatique final.
- Rapport de performance et recommandations à la direction.

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?

Cette tâche consiste à évaluer la performance du modèle final à l'aide de métriques statistiques (précision, rappel, F1-score). Elle fait partie de l'analyse des résultats et de l'exécution finale du projet pour valider le modèle avant sa mise en production

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

Why did you select these stages for this task?

Cette étape permet de vérifier les relations entre variables et de tester les hypothèses sur les données avant de modéliser. Elle appartient à la phase d'analyse et de construction du modèle pour confirmer les choix statistiques.

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

Why did you select this stage for this task?

L'exploration initiale des données aide à comprendre leur structure, leur cohérence et leurs tendances générales. Cela correspond à la phase d'analyse, car il s'agit d'une première interprétation des jeux de données.

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

Why did you select these stages for this task?

Cette tâche consiste à nettoyer les données, supprimer les valeurs manquantes et préparer les variables pour le modèle. C'est une phase de construction technique et d'analyse de la qualité des données.

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

Why did you select this stage for this task?



Cette tâche correspond à la planification du flux de travail du projet. Elle permet de définir les jalons, les rôles et les responsabilités avant le lancement du travail technique

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

Why did you select this stage for this task?

Cette étape concerne la communication des résultats finaux du modèle aux parties prenantes à travers des présentations et des rapports. Elle appartient à la phase d'exécution car elle clôture le projet.

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

Why did you select this stage for this task?

Le calcul des statistiques descriptives permet de résumer les caractéristiques des données et d'identifier les tendances principales. C'est une étape clé de l'analyse.

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

Why did you select these stages for this task?

La construction de visualisations aide à interpréter et à communiquer les résultats du projet. Cela implique à la fois la phase de construction (création des graphiques) et d'exécution (partage des insights).

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

Why did you select this stage for this task?

Cette tâche est directement liée à la phase de planification car elle définit les objectifs, les ressources et les livrables du projet avant son exécution.

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

Why did you select this stage for this task?



La création d'un modèle de régression nécessite de construire les algorithmes et d'analyser leurs performances. Elle englobe la préparation des variables et la validation statistique.

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

Why did you select this stage for this task?

Cette étape vise à résumer les principales caractéristiques du jeu de données, à fournir un aperçu global et à orienter les futures décisions analytiques.

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

Why did you select this stage for this task?

Cette tâche fait partie de la phase de construction, car elle consiste à développer le modèle d'apprentissage automatique qui permettra de classer automatiquement les affirmations.