

Questions et considérations pour le projet de données



PACE: Plan Stage

- Qui est le public cible de ce projet ?

Le public cible est l'**équipe de données de TikTok et les membres d'équipes transversales**.

- Quel problème essaie-t-on de résoudre ? Quel impact attend-on de ce projet sur les besoins du client

On travaille sur le développement d'un **modèle prédictif** permettant de déterminer si une vidéo contient une **revendication** ou si elle exprime une **opinion**.

- Quelles sont les questions à poser ou à répondre ?

Quel est l'état du jeu de données fourni ?

Quelles sont les variables les plus utiles ?

Y a-t-il des tendances dans les données qui peuvent fournir des insights ?

Quelles étapes peuvent être mises en place pour réduire les biais ?

- Quels sont les ressources nécessaires pour mener à bien ce projet ?

Le **jeu de données du projet**

Un **notebook Python**

Les **contributions des parties prenantes**

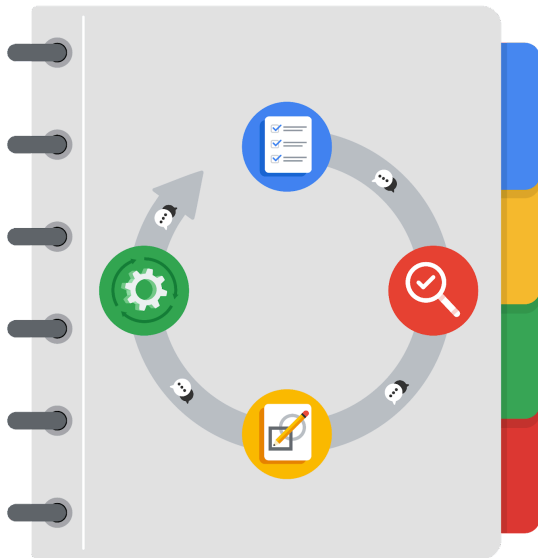
- Quels livrables doivent être produits au cours du projet ?

Un **jeu de données nettoyé** pour l'analyse exploratoire

Des **visualisations**

Un **modèle statistique**

Une **analyse de régression** et/ou un **modèle de machine learning**



Le processus de travail PACE

Tâches du projet

L'équipe de données a défini plusieurs tâches à accomplir dans le cadre de ce projet. Il faut organiser ces tâches en fonction des étapes du **workflow PACE** et expliquer pourquoi elles appartiennent à ces phases.

1. Évaluer le modèle : **Execute** ▾

Après la construction du modèle, les données sont utilisées pour évaluer si le modèle répond aux attentes et objectifs du projet.

2. Effectuer un test d'hypothèse : **Analyze** ▾ et **Construct** ▾

Pendant l'analyse, on détermine si un test statistique est nécessaire. Pendant la construction, on applique le test.

3. Explorer les données : **Analyze** ▾

Pendant cette phase, on cherche à mieux comprendre le jeu de données et les informations qu'il contient.



4. Exploration et nettoyage des données: **Plan** et **Analyze**

La phase de planification consiste à choisir les méthodes à utiliser. Ensuite, le nettoyage se fait pendant l'analyse des données

5. Établir la structure du workflow du projet (PACE): **Plan**

L'établissement du document PACE permet de structurer le projet et d'organiser efficacement les tâches.

6. Communiquer les insights finaux aux parties prenantes: **Execute**

Les insights finaux sont partagés avec les parties prenantes dans la phase d'exécution du projet.

7. Compute descriptive statistics: **Analyze**

Investigating the statistics within data takes place during analysis.

8. Calculer les statistiques descriptives: **Analyze** et **Construct**

L'examen des statistiques du jeu de données fait partie de l'analyse.

9. Créer des visualisations: **Plan**

Les visualisations commencent par l'évaluation des données et sont créées pendant la phase de construction.

10. Construire un modèle de régression: **Analyze** et **Construct**

Pendant l'analyse, on vérifie si le modèle répond aux besoins. Sa construction se fait ensuite dans la phase de développement.



11. Compiler les informations récapitulatives sur les données: **Analyze** ▾

L'inspection des données pour compiler les informations a lieu dans la phase d'analyse.

12. Construire un modèle de machine learning : **Construct** ▾

Le développement d'un modèle de machine learning se fait pendant la phase de construction.