**EXEMPLE DE FICHE : Projet étudiant en Data Engineering sur la Météo**

1. **Description du projet**

L'objectif de ce projet est de développer une application ou un outil qui utilise des techniques de Data Engineering pour traiter et analyser des données météorologiques. Cela pourrait impliquer la collecte, le nettoyage, la transformation, l'analyse et la visualisation de données météorologiques provenant de diverses sources. Le projet devrait démontrer une compréhension des principes fondamentaux du Data Engineering et de leur application dans un domaine concret.

**Analyse des tendances climatiques à long terme:** Ce projet impliquerait la collecte de données météorologiques historiques, l'identification des tendances climatiques et la visualisation des résultats.

**Création d'un outil de visualisation des données météorologiques:** Ce projet impliquerait le développement d'un outil interactif pour visualiser les données météorologiques, telles que les cartes radar, les graphiques de température et les prévisions météorologiques.

1. **Activités**

| **Tâches à réaliser / Étapes / Livrables** | **Consignes** | **Durée** |
| --- | --- | --- |
| Collection de donnée | Se connecter à la plateforme kaggle et télécharger le jeu de données.  Lien du jeu de données : [Dataset meteo](https://www.kaggle.com/datasets/muthuj7/weather-dataset) | 2 jour |
| Traitement de données | Une fois les données collectées, elles doivent être nettoyées et préparées pour l'analyse. Cela implique:   * **Nettoyage des données:** Identification et correction des erreurs, des incohérences et des valeurs manquantes dans les données. * **Transformation des données:** Conversion des formats de données, traitement des valeurs aberrantes et agrégation des données selon les besoins. * **Normalisation des données:** Assurer la cohérence des unités, des mesures et de la représentation des données. | 4 jours |
| Stockage des données | Les données traitées doivent être stockées de manière sécurisée et efficace pour une utilisation future. Cela pourrait impliquer:   * **Bases de données relationnelles:** Stockage de données structurées dans des bases de données SQL traditionnelles comme MySQL ou PostgreSQL. * **Bases de données NoSQL:** Stockage de données non structurées ou semi-structurées dans des bases de données comme MongoDB ou Cassandra. * **Data lakes:** Stockage des données brutes dans leur format natif dans un référentiel centralisé comme Hadoop ou Amazon S3. | 4 jours |
| Analyse des données | L'analyse des données implique l'application de techniques statistiques et analytiques pour extraire des informations des données. Cela pourrait impliquer:   * **Analyse exploratoire des données (EDA):** Résumé des données, identification des modèles de base et visualisation des tendances. * **Analyse statistique:** Utilisation de méthodes statistiques pour tester des hypothèses, identifier des relations et quantifier l'incertitude. * **Apprentissage automatique:** Entraînement d'algorithmes pour faire des prédictions, classifier des données ou identifier des anomalies. | 5 jours |
| Visualisation des données | La visualisation des données est le processus de création de représentations visuelles des données pour communiquer les informations de manière efficace. Cela pourrait impliquer:   * **Création de graphiques et de tableaux:** Utilisation d'outils comme Matplotlib, Seaborn ou Tableau pour visualiser les distributions de données, les tendances et les relations. * **Construction de tableaux de bord:** Création de tableaux de bord interactifs pour surveiller les indicateurs clés de performance (KPI) et suivre les changements au fil du temps. * **Développement de visualisations de données interactives:** Utilisation de bibliothèques JavaScript comme D3.js ou Plotly pour créer des visualisations de données interactives qui permettent aux utilisateurs d'explorer les données en temps réel. | 5 jours |
| Rapport | Faire un rapport qui présente le projet | 2 jours |

1. **Technologies:**

Les technologies spécifiques utilisées dans le projet dépendent des objectifs spécifiques et des compétences des étudiants. Cependant, quelques technologies courantes du domaine du Data Engineering qui pourraient être utilisées incluent:

* **Langages de programmation:** Python
* **Outils de traitement de données:** Spark, Hadoop
* **Bases de données:** NoSQL, SQL
* **Outils de visualisation:** Tableau, Power BI

1. **Compétences:**

Les compétences que les étudiants développeront au cours de ce projet incluent:

* **Collecte et nettoyage de données**
* **Transformation et analyse de données**
* **Visualisation de données**
* **Ingénierie logicielle**
* **Résolution de problèmes**
* **Communication**

**Impact:**

Ce type de projet peut avoir un impact positif sur la communauté en fournissant des informations météorologiques plus accessibles et utiles. Les résultats du projet pourraient être utilisés pour améliorer la préparation aux événements météorologiques, soutenir la prise de décision agricole ou simplement aider les gens à mieux planifier leur journée.

**Évaluation:**

Le projet peut être évalué sur la base des critères suivants:

* **Clarté des objectifs et de la portée**
* **Mise en œuvre efficace des technologies**
* **Qualité de l'analyse et de la visualisation des données**
* **Communication claire et concise des résultats**
* **Impact potentiel sur la communauté**

En plus de ces critères généraux, les évaluateurs peuvent également prendre en compte la créativité, l'originalité et la difficulté du projet.

**Conclusion:**

Un projet d'étudiant en Data Engineering sur la météo peut être une expérience enrichissante et stimulante qui permet aux étudiants d'appliquer leurs compétences à un problème réel et d'avoir un impact positif sur le monde.