

D'accord ! Avec ton objectif d'approfondir le machine learning tout en développant un profil hybride pour une alternance en data analyst, et en tenant compte de ton niveau intermédiaire en Python, machine learning, et statistique, je vais te proposer un **plan d'ensemble détaillé** sur une période de deux semaines. Ce plan intègre des projets des deux documents ("12 Projets pour devenir Data Analyst" et "10 Projets pour un Portfolio Data Science Réussi"), en combinant des compétences variées (Excel, Python, visualisation, déploiement) et en mettant l'accent sur le machine learning. Le plan est structuré pour être réalisable entre aujourd'hui (jeudi 29 mai 2025, 11:15 AM CEST) et dans les deux semaines suivantes, avec des jalons clairs et des ajustements possibles selon ton rythme.

Plan d'ensemble (2 semaines : 29 mai - 12 juin 2025)

Objectifs globaux

- **Machine Learning** : Approfondir les compétences en modélisation supervisée (régression, classification), non supervisée (clustering), et systèmes de recommandation.
- **Profil hybride** : Intégrer Excel (exploration initiale), Python (modélisation, analyse), et outils comme Power BI ou Streamlit (visualisation/déploiement).
- **CV/Portfolio** : Produire des résultats concrets (visualisations, code, tableaux de bord) pour un portfolio et un CV optimisé pour une alternance.

Hypothèses

- Disponibilité : 3-4 heures par jour (ajustable).
 - Matériel : Accès à un ordinateur, internet, et logiciels (Excel, Python, Power BI/Tableau optionnel).
 - Progression : Travail autonome avec possibilité de demander des clarifications.
-

Semaine 1 : Bases du machine learning et exploration hybride (29 mai - 4 juin 2025)

Jour 1 : Préparation et démarrage (29 mai, jeudi)

- **Heure** : 11:15 AM - 3:15 PM (4h).
- **Tâches** :
 - **Installation et configuration** : Vérifie Python (avec pandas, scikit-learn, seaborn, matplotlib), installe Excel si besoin, et télécharge Power BI (version gratuite) ou Tableau Public.
 - **Choix des datasets** :
 - Projet 9 (Doc 1) : Mall Customer Segmentation Dataset.
 - Projet 10 (Doc 1) : Boston Housing Dataset.
 - Projet 7 (Doc 1) : Global Superstore Dataset (pour tableau de bord).
 - **Planification** : Révise les objectifs des projets et note les compétences à travailler (clustering, régression, visualisation).
- **Délivrabile** : Environnement prêt, datasets téléchargés.

Jour 2 : Clustering - Segmentation client (30 mai, vendredi)

- **Heure** : 10:00 AM - 2:00 PM (4h).
- **Tâche** : Projet 9 (Doc 1) - Segmenter les clients avec K-means.
 - **Étape 1 (1h)** : Nettoyage dans Excel (suppression des doublons, vérification des valeurs manquantes).
 - **Étape 2 (2h)** : Modélisation avec Python :
 - Charge les données avec pandas.
 - Normalise avec StandardScaler .
 - Applique K-means (scikit-learn) et utilise la méthode du coude pour choisir le nombre de clusters.
 - **Étape 3 (1h)** : Visualisation avec seaborn (scatter plot des clusters).
- **Délivrabile** : Script Python, graphique des clusters, sauvegarde sur GitHub.
- **Compétence** : Clustering non supervisé.

Jour 3 : Régression - Prédiction des prix (31 mai, samedi)

- **Heure** : 10:00 AM - 2:00 PM (4h).
- **Tâche** : Projet 10 (Doc 1) - Prédire les prix des maisons.
 - **Étape 1 (1h)** : Analyse exploratoire dans Excel (corrélation, histogramme).
 - **Étape 2 (2h)** : Modélisation avec Python :
 - Sépare les données (train/test avec scikit-learn).
 - Entraîne une régression linéaire, évalue avec R^2 et RMSE.
 - Teste une régression multiple avec plusieurs variables.
 - **Étape 3 (1h)** : Visualisation (prédictions vs. réelles avec matplotlib).
- **Délivrabile** : Script Python, graphique des prédictions, sauvegarde sur GitHub.
- **Compétence** : Modélisation supervisée (régression).

Jour 4 : Tableau de bord - Visualisation hybride (1er juin, dimanche)

- **Heure** : 10:00 AM - 2:00 PM (4h).
- **Tâche** : Projet 7 (Doc 1) - Créer un tableau de bord interactif.
 - **Étape 1 (1h)** : Nettoyage initial dans Excel (filtres, tri).
 - **Étape 2 (2h)** : Crée un tableau de bord dans Power BI :
 - Importe le Global Superstore Dataset.
 - Ajoute des KPI (ventes par région, tendances mensuelles) et des slicers.
 - **Étape 3 (1h)** : Exporte une capture d'écran, intègre dans ton portfolio.
- **Délivrabile** : Tableau de bord Power BI, capture d'écran pour LinkedIn/GitHub.
- **Compétence** : Visualisation métier.

Jour 5 : Révision et consolidation (2 juin, lundi)

- **Heure** : 10:00 AM - 1:00 PM (3h).
- **Tâches** :
 - Révise les scripts Python (clustering, régression) pour optimiser (ex. GridSearchCV pour K-means).
 - Ajoute des commentaires dans le code et mets à jour GitHub.
 - Prépare une brève description pour ton CV (ex. "Segmentation client avec K-means, prédition des prix avec régression linéaire").
- **Délivrable** : Code optimisé, CV mis à jour.

Week-end (3-4 juin) : Repos ou approfondissement optionnel

- Optionnel : Explore la régression logistique (Projet 11, Doc 1) si tu as du temps (3-4h).
-

Semaine 2 : Approfondissement du machine learning et projet avancé (5 juin - 12 juin 2025)

Jour 6 : Classification - Détection de défauts (5 juin, jeudi)

- **Heure** : 10:00 AM - 2:00 PM (4h).
- **Tâche** : Projet 11 (Doc 1) - Prédire les défauts de paiement.
 - **Étape 1 (1h)** : Analyse dans Excel (fréquence des défauts).
 - **Étape 2 (2h)** : Modélisation avec Python :
 - Entraîne une régression logistique.
 - Évalue avec matrice de confusion et AUC (scikit-learn).
 - Teste SMOTE si les classes sont déséquilibrées.
 - **Étape 3 (1h)** : Visualisation (courbe ROC avec matplotlib).
- **Délivrable** : Script Python, courbe ROC, sauvegarde sur GitHub.
- **Compétence** : Classification supervisée.

Jour 7 : Systèmes de recommandation - Début (6 juin, vendredi)

- **Heure** : 10:00 AM - 2:00 PM (4h).
- **Tâche** : Projet 10 (Doc 2) - Systèmes de recommandation.
 - **Étape 1 (1h)** : Analyse exploratoire dans Excel (tableau de corrélation).
 - **Étape 2 (2h)** : Modélisation avec Python :
 - Utilise Surprise pour SVD (filtrage collaboratif).
 - Évalue avec RMSE.
 - **Étape 3 (1h)** : Prépare une interface de base avec Streamlit.
- **Délivvable** : Script Python, interface Streamlit partielle.
- **Compétence** : Machine learning avancé (recommandation).

Jour 8 : Systèmes de recommandation - Finalisation (7 juin, samedi)

- **Heure** : 10:00 AM - 2:00 PM (4h).
- **Tâche** : Complète le Projet 10 (Doc 2).
 - **Étape 1 (1h)** : Ajoute un filtrage basé sur le contenu (scikit-learn).
 - **Étape 2 (2h)** : Finalise l'interface Streamlit (affichage des recommandations).
 - **Étape 3 (1h)** : Déploie localement et teste.
- **Délivvable** : Système de recommandation fonctionnel, capture d'écran.
- **Compétence** : Recommandation hybride.

Jour 9 : Révision et déploiement (8 juin, dimanche)

- **Heure** : 10:00 AM - 1:00 PM (3h).
- **Tâches** :
 - Optimise les modèles (ex. hyperparamètres pour Random Forest dans régression).
 - Ajoute une API simple avec Flask pour un projet (ex. régression ou recommandation).
 - Met à jour ton portfolio (GitHub, LinkedIn).
- **Délivvable** : API fonctionnelle (optionnel), portfolio mis à jour.

Jour 10-11 : Projet avancé optionnel (9-10 juin, lundi-mardi)

- **Heure** : 10:00 AM - 2:00 PM (8h sur 2 jours).
- **Tâche** : Projet 12 (Doc 1) ou 6 (Doc 2) - Séries temporelles.
 - **Étape 1 (2h)** : Analyse des tendances dans Excel.
 - **Étape 2 (4h)** : Modélise avec Prophet ou ARIMA (statsmodels).
 - **Étape 3 (2h)** : Visualise les prévisions.
- **Délivrable** : Script Python, graphique des prévisions.
- **Compétence** : Séries temporelles.

Jour 12-13 : Finalisation et présentation (11-12 juin, mercredi-jeudi)

- **Heure** : 10:00 AM - 1:00 PM (6h sur 2 jours).
 - **Tâches** :
 - Compile tes résultats (graphiques, tableaux de bord, code).
 - Rédige une section "Projets" pour ton CV.
 - Prépare une présentation (PDF ou vidéo) pour simuler un entretien.
 - **Délivrable** : CV finalisé, portfolio complet, présentation.
-

Résumé des livrables

- **GitHub** : Scripts Python (clustering, régression, classification, recommandation), README.
- **LinkedIn** : Captures d'écran (clusters, prédictions, tableau de bord, interface recommandation).
- **CV** : "Compétences : Python (scikit-learn, statsmodels), Excel, Power BI, machine learning (clustering, régression, classification, recommandation). Projets : Segmentation client, prédiction des prix, tableau de bord interactif, système de recommandation."
- **Optionnel** : API Flask, série temporelle.

Conseils pour réussir

- **Ressources** : Scikit-learn docs, Kaggle tutorials, YouTube (StatQuest pour les bases).
- **Gestion du temps** : Fais des pauses régulières, ajuste les heures selon ton emploi du temps.
- **Support** : Demande-moi un guide détaillé pour un projet (ex. code K-means) si besoin.

Si tu veux un plan ajusté (ex. plus de temps sur un projet) ou un tutoriel spécifique, dis-le-moi ! On peut démarrer dès aujourd'hui avec la préparation. Que penses-tu ?