



Application data sous forme de data storytelling

Type :	PROJET
Formations :	Ynov Informatique
Promotions :	Bachelor 3
UF :	SPE Data & IA

Cadre du projet

Ce projet permet d'évaluer les compétences acquises dans le cadre de l'UF « Spécialité IA & Data ». Le projet consiste à réaliser une application data exploitant des analyses de données synthétisées sous forme de visualisations au travers d'une histoire.

Effectifs par groupe : **2**

Pour valider ce projet, vous pourrez opter pour un projet personnel ou un projet au choix parmi une liste de sujets proposés dans la partie « Projets au choix ».

Il vous appartient d'effectuer les recherches par vous-même pour trouver les ressources à la réalisation du projet.

Date de début :

Date de rendu :

OBJECTIFS DE FORMATION VISÉS

Vous devrez mettre en application les principales compétences acquises durant les différents modules de la formation, à savoir :

Analyse et exploration de données et Analyse de données avancée

- Savoir **acquérir et structurer** des données pertinentes en utilisant des données ouvertes (open data) et/ou des méthodes de web scraping
- Réaliser une **préparation des données** (formatage des différents types de données, gestion des valeurs manquantes, des doublons, des valeurs aberrantes) afin de les rendre exploitables par des méthodes d'analyses
- Réaliser une **analyse exploratoire de données** afin de mettre en évidence les informations contenues dans les données, savoir les synthétiser sous forme de graphiques et identifier les variables pertinentes à utiliser dans un modèle
- Savoir appliquer **au moins un modèle** de machine learning dans vos analyses en fonction de la catégorie de problème à résoudre (classification, régression, clustering)
- Savoir évaluer les performances d'un modèle en utilisant des métriques adaptées, interpréter ces résultats et si nécessaire, ajuster les paramètres pour comparer plusieurs modèles
- Savoir **interpréter vos résultats** et les **synthétiser** en utilisant des graphiques appropriés et des méthodes de **data storytelling** pour présenter de manière synthétique vos conclusions

Maths pour la DataScience

- Utiliser des métriques pour **quantifier la quantité et la qualité de données** présente dans les échantillons que vous aurez sélectionnés (indicateurs statistiques descriptifs, statistique inférentielle)
- Utiliser vos connaissances en **statistiques** et **probabilités** afin d'établir des **indicateurs** univariés et multivariés, pour la préparation et les analyses de données à réaliser
- Utiliser des méthodes d'analyses de données vues en cours et savoir rechercher de nouvelles méthodes qui pourraient être appropriées à votre sujet (modèles statistiques, modèles de machine learning)

Python avancé

- Savoir implémenter, structurer et documenter du code pour les différentes étapes du projet
- Savoir utiliser des bibliothèques appropriées de l'écosystème python pour la data science (pandas, scikit-learn, seaborn...)

Libraries et Outils

- Utilisation obligatoire de Python, complété par SQL, R, Javascript, etc.
- Notebooks interactifs : Jupyter Notebook
- Datasets : [Hugging Face](#), [Kaggle](#)
- Acquisition, structuration et analyse exploratoire de données : [pandas](#), [numpy](#)
- Modélisation statistiques et machine learning : [scikit-learn](#), [statsmodels](#)
- Exploration et visualisation de données : [matplotlib](#) et/ou [seaborn](#)
- Outils orientés visualisation de données : [Dash \(de la suite logicielle Plotly\)](#), [Panel \(de la suite logicielle Holoviz\)](#), [Bokeh](#)
- Outils orientés "interface application data" : [Streamlit](#), [Anvil](#)
- Versionning et collaboration : Git, GitHub, GitLab
- IDE ou éditeur de code

LIVRABLES

Pour chaque groupe, vous devrez livrer les éléments suivants :

- Dépôt Git avec tout le code et la documentation
- Jupyter Notebook retraçant la démarche et les analyses
- Application de data storytelling déployée localement
- Documentation technique du projet et manuel d'installation et utilisation de l'application

MODALITÉS D'ÉVALUATION DU PROJET

Vous serez évalué sur l'ensemble des productions. L'évaluation prendra la forme d'une présentation orale de synthèse accompagnée d'un support de présentation ainsi que d'une démonstration du travail réalisé et la démarche pour y parvenir.

Le jury sera composé d'une partie des intervenants des cours de l'UF « SPÉ Data ».

- Une note compétences transverses (coef. 1) : évaluée pendant l'oral intermédiaire
- Une note technique (coef. 3) : évaluée pendant l'oral final

Chaque critère sera évalué selon 5 niveaux :

Non acquis : L'étudiant ne démontre pas la compétence attendue ou produit un livrable absent, hors sujet ou inexploitable

En cours d'acquisition : L'étudiant amorce la compétence mais de façon incomplète, imprécise ou avec des erreurs majeures

Partiellement acquis : L'étudiant maîtrise partiellement la compétence, avec un résultat globalement cohérent mais comportant des manques ou imprécisions

Acquis : L'étudiant démontre la compétence de manière satisfaisante, avec une mise en œuvre correcte et fonctionnelle, malgré quelques points perfectibles

Maîtrisé : L'étudiant dépasse les attentes, en mobilisant la compétence de façon approfondie. Je pense qu'il pourra bientôt prendre votre place, monsieur/madame du jury !

Chaque niveau correspond à une note sur 20 :

Non acquis	En cours d'acquisition	Partiellement acquis	Acquis	Maîtrisé
0	8	12	15	20

Pour finir il faudra multiplier la compétence par la pondération pour avoir la note finale.

Oral intermédiaire

Cet oral sert à valider la démarche des étudiants et s'assurer qu'ils vont dans la bonne direction pour mener à bien leur projet. C'est également l'occasion pour jauger s'ils ont les compétences de gérer professionnellement un projet.

Date : ...

Durée : **5 à 10 minutes + 5 minutes de questions**

Grille d'évaluation de l'oral intermédiaire

Compétences transverses			100%
Critères	Description	Module de cours	Pondération
Utiliser Git et GitHub pour le développement collaboratif	L'étudiant est capable d'utiliser efficacement Git et GitHub pour versionner son code, gérer les branches, documenter les commits et collaborer via des pull requests dans un projet partagé	x	3
Gérer un projet avec des outils adaptés	L'étudiant est capable de planifier et suivre l'avancement du projet avec des outils de gestion (Trello, Notion, ClickUp...)	x	4
Justifier sa démarche et ses choix techniques	L'étudiant est capable de formaliser les besoins fonctionnels et techniques du projet et mettre en œuvre une stratégie en choisissant les technologies adaptées	x	5
Présenter un projet à l'oral	L'étudiant est capable de structurer une présentation orale claire et convaincante pour identifier les besoins clients en valorisant les enjeux et ses choix pour le projet	x	5
Adopter une posture professionnelle	L'étudiant fait preuve d'autonomie, de rigueur et de responsabilité dans sa démarche, en respectant les contraintes et les livrables	x	3

Oral final

Date : ...

Durée : 15 minutes + 5 minutes de questions

Grille d'évaluation de l'oral final

GRILLE UF B3 IA & DATA			100%
Critères	Description	Module de cours	Pondération
Acquérir des données	L'étudiant est capable de collecter des données fiables depuis plusieurs sources ouvertes (open data, API, base de données SQL, CSV...)	Analyse et exploration des données, Python avancé	4
Préparer et nettoyer des données	L'étudiant est capable de nettoyer, formater et transformer des données (gestion des doubles, valeurs manquantes, formats incohérents, standardisation...)	Analyse et exploration des données, Python avancé	3
Explorer et analyser des données	L'étudiant est capable de réaliser une analyse pertinente des données à partir des critères sélectionnés	Analyse et exploration des données, Maths pour la Data Science	3
Visualiser des données	L'étudiant est capable de créer des visualisations claires, informatives et adaptées à la nature des données	Analyse et exploration des données, Python avancé	3
Appliquer un modèle de machine learning et l'évaluer	L'étudiant est capable de choisir un modèle adapté (régression, clustering...) et d'évaluer ses performances	Analyse et exploration des données, Maths pour la Data Science	4
Prédire et donner des recommandations à partir des données	L'étudiant est capable de proposer des projections ou des simulation de scénarios futurs à partir des données	Analyse et exploration des données, Maths pour la Data Science	4
Concevoir et déployer une interface interactive	L'étudiant est capable de concevoir une interface fonctionnelle et intuitive permettant à un utilisateur d'explorer les données et de la déployer localement ou en ligne	Python avancé	2
Documenter son projet	L'étudiant est capable de rédiger une documentation claire qui explique la démarche, les choix techniques, les résultats, l'installation et l'utilisation de son travail	Analyse et exploration des données	2

Descriptif du projet

Vous avez la possibilité de choisir entre un projet personnel ou un projet proposé.

Vous trouverez ci-dessous une liste des différents sujets qui pourront être abordés. Suivi de la liste de projets au choix, si vous n'avez pas d'idée

LISTE DES PROJETS AU CHOIX

PROJET PERSONNEL :

Vous avez la possibilité de choisir un sujet libre dans la mesure qu'il vous permette de valider les compétences présentes dans la grille d'évaluation. Travail minimum attendu :

- **Acquisition des données (API, Base de données, CSV...) et préparation (pandas)**
- **Analyse exploratoire des données (Analyse + visualisation)**
- **Utilisation d'un modèle de machine learning (prédiction, recommandations, etc...)**

- Concevoir et déployer une application interactive permettant aux utilisateurs de visualiser les données, prédictions etc...
- Documenter le projet

Sujet 1 : Présidentielle 2027 – Analyse électorale et prédictions

Brief

Dans le cadre d'une mission en freelance, vous êtes sollicité par un cabinet de conseil politique qui souhaite anticiper les dynamiques de l'élection présidentielle française de 2027. Le cabinet vous confie la réalisation d'une application interactive de data storytelling permettant d'analyser les résultats passés, de mettre en lumière les évolutions sociopolitiques majeures et d'explorer différents scénarios de vote possibles pour 2027.

Objectifs

- Collecter et analyser les résultats électoraux des scrutins présidentiels précédents
- Identifier les dynamiques régionales et socio-démographiques des votes
- Développer un modèle prédictif pour explorer des hypothèses sur 2027
- Proposer une application visuelle et interactive pour simuler et présenter vos scénarios

Tâches à réaliser

1. Acquisition et préparation des données

- Récupération des données électorales ([data.gouv](https://data.gouv.fr/))
- Collecte de données démographiques (INSEE)
- Préparation des données (nettoyage, fusion, standardisation...)

2. Analyse exploratoire des scrutins passés

- Visualiser les résultats par région, par candidat, par année, par famille politique etc...
- Analyse de l'évolution des scores, taux de participation, abstention etc...
- Croiser les résultats avec des données sociales (zones favorables à tel ou tel parti en fonction de données démographiques)

3. Scénarios et modélisation

- Identification de variables explicatives (âge, revenus, géographie etc...)

- Simulation de plusieurs scénarios pour 2027 (nouveaux candidats, alliance de partis, abstention de la jeunesse etc...) en utilisant des modèles de machine learning

4. Data storytelling application interactive

- Création d'une timeline interactive des élections passées
- Cartes des votes par région avec filtres
- Intégration d'un simulateur d'élection 2027 (avec paramètres modifiables)
- Utilisation de Streamlit ou Dash

5. Déploiement

- Déploiement local
- Rédaction de la documentation

Sujet 2 : Prévisions Météorologiques avec WeatherForYnov

Brief

WFY France est une entreprise fictive spécialisée dans l'analyse environnementale et climatique. Elle développe des solutions de visualisation et de prévision basées sur des données ouvertes pour aider les citoyens et les collectivités locales à mieux comprendre les impacts du changement climatique dans les différentes régions françaises. L'entreprise souhaite également intégrer des prévisions météorologiques pour que nos régions puissent anticiper des phénomènes extrêmes ou plus généralement l'avenir climatique du territoire.

Objectifs

- Collecter et analyser les données climatiques historiques régionales en France
- Identifier les tendances liées au changement climatique (température, précipitations etc...)
- Développer des modèles de prévision météorologique
- Présenter les résultats sous forme de visualisations interactives dans une application accessible

Tâches à réaliser

1. Acquisition et préparation des données

- Utiliser des données climatiques publiques disponibles en ligne (Météo France, NOAA Climate Data, OpenWeatherMap, Open Meteo API)
- Nettoyer et structurer les données pour les analyses

2. Analyse exploratoire des données

- Réaliser des visualisations des données climatiques (Matplotlib, Seaborn)
- Identifier les tendances et les patterns dans les données

3. Modélisation prédictive

- Appliquer des modèles de machine learning pour la prévision météorologique
- Évaluer la performance des modèles

4. Data storytelling et visualisations

- Créer des graphiques interactifs avec Plotly ou Dash pour présenter les résultats des analyses et des prévisions, carte régionales avec temporalité dynamique, filtres, indicateurs climatiques etc...
- Développer une interface utilisateur basique pour visualiser les prévisions météorologiques

5. Déploiement de l'application

- Déployer l'application localement
- Documenter le processus de déploiement

Sujet 3 : Performances sportives pour les JO 2028 à Los Angeles

Brief

YPerf est une start-up fictive qui anticipe les tendances et performances sportives mondiales en vue des Jeux Olympiques de 2028 à Los Angeles. L'objectif est de construire une application de data storytelling pour explorer les performances passées des athlètes par pays, sport et genre, tout en essayant de prédire les nations ou athlètes à suivre en 2028.

Objectifs

- Analyser les résultats de JO précédents par sport, pays et genre
- Visualiser l'évolution des performances sportives par discipline
- Identifier les athlètes et pays en progression par discipline
- Créer des projectifs sur les JO 2028 basées sur les tendances observées

Tâches à réaliser

1. Acquisition et préparation des données

Données historiques des JO (résultats, médailles, records...)

Infos sur les athlètes (âge, nationalité, palmarès etc...)

Nettoyage, normalisation et structuration des données

2. Analyse exploratoire

- Visualiser les médailles par pays, années, discipline, athlètes
- Analyser les tendances de performance (progression par pays, disciplines dominées etc...)
- Détection des athlètes marquants et des nouvelles générations montantes

3. Modélisation prédictive

- Modèle prédictifs des performance 2028
- Côte des athlètes, pays par disciplines

4. Storytelling visuel

- Création de dashboards dynamiques avec filtres (carte des médailles, évolutions des performances des athlètes, évolution par disciplines...)
- Timeline des records
- Visualisation des côtes des athlètes, ajouts d'avis utilisateurs...

5. Déploiement

- Déploiement local
- Rédaction de la documentation