



# Application data sous forme de datastorytelling

Leçon • 1 backlink



Type : **PROJET**

Formations : Ynov Informatique Promotions : Bachelor 3

UF : SPECIALITE DATA

## CADRE DU PROJET

Ce projet permet l'évaluation des compétences acquises grâce aux modules de l'UF « Spécialité Data ».

Le projet consiste à réaliser une application data exploitant des analyses de données synthétisées sous forme de visualisations organisée pour raconter une histoire, votre histoire commencera par une question à laquelle vous répondrez tout au long de votre data storytelling pour conclure avec un réponse ou une autre question.



**Vous devrez réaliser ce projet par équipe de deux.**

Vous devez soumettre un projet personnel

Dans tous les cas, le contenu et les fonctionnalités devront respecter des conditions décrites dans cette section.

Les projets devront être validés par le responsable de la formation.

Un bonus sera apporté aux projets qui proposeront des méthodes d'analyses ou de visualisation plus poussées. Vous devrez **utiliser obligatoirement du code en python** que vous pouvez compléter par d'autres langages si vous le souhaitez (par exemple : SQL, R, Javascript, ...)

## OBJECTIFS DE FORMATION VISES

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

## PRINCIPES DE L'EXPLORATION & ANALYSE DE DONNEES

- 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 des méthodes d'analyses supplémentaires à faire.
- 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. **MATHEMATIQUES POUR LA DATA SCIENCE**
- 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)
- Savoir quantifier la qualité de vos analyses et de vos modèles. **MACHINE LEARNING**
- Savoir appliquer au moins un modèle de machine learning dans vos analyses en fonction de la catégorie de problème à résoudre (classification, régression, clustering) **PYTHON POUR LA DATA SCIENCE**
- Savoir implémenter, structurer et documenter du code pour les différentes étapes du projet
- Savoir utiliser des librairies appropriées de l'écosystème python pour la data science. Vous pouvez utiliser les librairies de votre choix



### Exemple d'outils

- 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, D3.js
- Outils orientés "interface application data": Streamlit, Anvil

## LIVRABLES

Pour chaque groupe vous devrez livrer les éléments suivants :

- **Un dépôt Git** accessible en ligne (par exemple via Gitlab ou Github) contenant tout le code et la documentation produits pour le projet
- **Un document au format jupyter notebook** (ou équivalent) retraçant votre démarche et les analyses exploratoires que vous aurez mises en place durant le projet.
- **Votre application data finale**, synthétisant sous forme graphique vos analyses et présentant un **data storytelling**. Vous devrez présenter cette application, déployée à minima en local sur votre machine, pendant la soutenance.

## DESCRIPTIF DU PROJET

Le projet consiste à réaliser une application data sur une thématique donnée que vous choisirez vous même (sport, environnement, société, ....)

Votre note sera calculée en fonction des fonctionnalités obligatoires et les fonctionnalités bonus que vous réaliserez.

### Fonctionnalités obligatoires :

- Exploiter des données en décrivant comment vous les avez acquises (open data, scraping, APIs ...)
- Utiliser des méthodes d'analyses de données appropriées aux données recueillies
- Utiliser au moins un modèle de machine learning



#### Algorithmes de machine learning :

- Apprentissage supervisé :
  - Régression linéaire
  - Régression logistique
  - Arbres de décision
  - SVM

- Random Forest
- KNN
- Apprentissage non supervisé :
  - Clustering K-Means
  - Clustering hiérarchique
  - Analyse en composantes principales (ACP)
- Apprentissage par renforcement :
  - Q-Learning
  - Deep Q-Learning
  - Méthodes à gradient de politique

- Organisez vos résultats sous forme de graphiques synthétiques pour raconter une histoire
- En utilisant des principes de data storytelling, présenter les analyses et conclusions que vous avez menées pour répondre au problème posé
- Déployer votre application localement et documenter son utilisation
- Proposer un dépôt Git structuré et documenté

### **Fonctionnalités bonus**

- Exploiter de manière pertinente des données cartographiques
- Exploiter de manière pertinente des données textuelles (blog, réseaux sociaux, ...)
- Ajouter une fonctionnalité à votre application permettant la mise à jour automatique des données et analyses présentées dans l'application
- Déployer votre application sur un serveur distant permettant d'accéder à votre application en ligne via une url

Pour chaque fonctionnalité bonus réalisée un bonus de 0,5 pts sera accordé pour un maximum de 2 pts.

### **Exemples**

[https://www.washingtonpost.com/graphics/2018/investigations/unsolved-homicide-database/?utm\\_term=.a83ca9730741](https://www.washingtonpost.com/graphics/2018/investigations/unsolved-homicide-database/?utm_term=.a83ca9730741)

[https://projects.fivethirtyeight.com/2016-election-forecast/?ex\\_cid=rrpromo&utm\\_source=NL+liste+enti%C3%A8re&utm\\_campaign=a6fb073013-](https://projects.fivethirtyeight.com/2016-election-forecast/?ex_cid=rrpromo&utm_source=NL+liste+enti%C3%A8re&utm_campaign=a6fb073013-)

[EMAIL\\_CAMPAIGN\\_2016\\_11\\_08&utm\\_medium=email&utm\\_term=0\\_880bf51aff-a6fb073013-308942073&ct=t\(Trump\\_ou\\_Clinton\\_Les\\_analyses\\_des\\_inside11\\_8\\_2016\)#plus](mailto:EMAIL_CAMPAIGN_2016_11_08&utm_medium=email&utm_term=0_880bf51aff-a6fb073013-308942073&ct=t(Trump_ou_Clinton_Les_analyses_des_inside11_8_2016)#plus)  
<https://www.kaggle.com/code/artvolgin/exploring-phd-community-with-network-analysis>  
<https://www.kaggle.com/code/erikbruin/storytelling-covid-19>

## **Notation**

Le projet comportera 2 notes :

- Rendu
- Soutenance

Les points bonus seront attribué à la note la plus basse des deux.

Vous trouverez ci-dessous le barème de notation pour ces deux notes :

### **Rendu**

Catégorie	Item	Pts
GIT	Dépôt Git structuré et documenté	2
Notebook	Explication de l'origine des données, et les moyens de récupération	2
Notebook	Notebook retraçant tout le processus de création	2
Application	Application fonctionnelle	2
Application	Modèle de machine learning utilisé	2
Application	Data storytelling	4
Application	Choix des graphiques approprié	4
Structure	Problématique	2
Structure	Conclusion répondant à la problématique	2

## Soutenance

Catégorie	Item	Pts
Oral	Voix cf fiche prise de parole en public	6
Oral	Posture cf fiche prise de parole en public	4
Oral	Dynamique, expression du visage cf fiche prise de parole en public	4
Application	L'application est claire, les graphiques sont lisibles	3
Application	L'application comporte de l'interaction permettant de modifier des graphiques	3