

Projet final du Kit Data Science 2021

Le projet final du Kit Data Science 2021 porte sur les données du **Vendée Globe 2020-2021**.

Le projet se déroule **26 octobre au 7 novembre 2021** date limite pour rendre vos projets respectifs.

Les données du dernier Vendée Globe sont disponibles sous la forme de fichiers Excel avec les classements fournis plusieurs fois par jour par les organisateurs de la course. Il y a également une page web avec une fiche technique par voilier qui contient des informations techniques et qu'il est possible de rapprocher des classements.

Il vous appartient de charger les données en Python, de procéder aux préparations nécessaires et d'effectuer les analyses pertinentes de votre choix.

Le rendu sera un notebook Jupyter fourni aux formats ipynb et HTML.

Barème sur 15 points :

- Acquisition et chargement des données : 3 points
- Préparation des données : 5 points
- Analyses et story telling : 7 points

Exemples de traitements et d'analyses :

- Récupération des fichiers Excel avec les classements
 - i. Mettre en place une copie locale des fichiers Excel afin de ne pas les recharger à chaque run.
 - ii. Vers la fin de la course le format des fichiers Excel change avec les arrivées des voiliers : il est possible de s'arrêter juste avant.
- Extraction des caractéristiques techniques de chacun des voiliers.
- Rapprochement des données des voiliers avec celle des classements.
- Corrélation et régression linéaire entre le classement (rang) et la vitesse utile (VMG) des voiliers.
- Impact de la présence d'un *foil* sur le classement et la vitesse des voiliers.
- Visualisation de la distance parcourue par voilier.
- Cartes avec les routes d'un ou plusieurs voiliers.
- Analyses de séries temporelles.
- Application d'algorithmes statistiques ou de machine learning.
- etc.

Sources des données

- Page web donnant accès aux fichiers Excel des classements du Vendée Globe
: <https://www.vendeeglobe.org/fr/classement>
- Page web avec les fiches techniques des voiliers du Vendée Globe
: <https://www.vendeeglobe.org/fr/glossaire>

- Site web donnant accès à des fichiers avec les formes géométriques des côtes
: <https://www.naturalearthdata.com/> (ou bien utilisez les librairies plotly ou ipyleaflet pour produire des cartes)
- etc.

Questions/Réponses

Les questions et réponses seront publiées ci-après au fil de l'eau :

1. Qu'est-ce qu'un *foil*? <https://www.vendeeglobe.org/fr/actualites/19755/quels-foils-pour-gagner-le-vendee-globe> La présence d'un *foil* est indiqué dans l'attribut "Nombre de dérives" dans les fiches techniques des voiliers.
2. S'agit-il d'un travail individuel ou collectif ? Il s'agit bien d'un travail individuel.
3. Est-il possible de rendre plusieurs notebooks afin de délimiter clairement les différentes étapes du projet ? Tout peut tenir dans un seul notebook mais pourquoi pas.
4. Y a-t-il une norme pour le code ? Vous pouvez regarder la PEP8 (voir Naming Conventions pour les noms de variables et de fonctions) mais le projet n'est particulièrement noté sur ce point.
5. Faut-il créer un unique DataFrame mettant à chaque classement d'afficher les caractéristiques des voiliers ou bien peut-on gérer les 2 types de données indépendamment ? C'est à vous de voir en fonction des analyses que vous voulez produire.
6. Peut-on produire des graphiques supplémentaires à ceux proposés ? Oui, vous avez toute latitude pour produire les analyses de votre choix à partir des données.
7. Comment peut-on télécharger les fichier Excel pour ensuite les lire avec pandas? Voir ci-dessous :

```
import shutil
```

```
url = 'URL du fichier Excel à télécharger'
```

```
filename = 'exemple.xlsx'
```

```
with request.urlopen(url) as response, open(filename, 'wb') as out_file: shutil.copyfileobj(response, out_file)
```

Avertissement

Vous devez publier votre **notebook exécuté aux formats ipynb et HTML** sur votre github **avant le dimanche 7 novembre 2021 à 23h59** et lorsque c'est fait **envoyer une notification par email avec le lien du projet** à l'adresse [contact\(at\)yotta-conseil.fr](mailto:contact(at)yotta-conseil.fr)

Bon projet !