

# Solution de Data visualisation

Ce rapport présente la démarche et la méthodologie employé par notre équipe dans le cadre du concours de Data-visualisation des données de transport aérien organisé par les Assises nationales du transport aérien.

## Notre approche

Tout d'abord, au vu du nombre conséquent de l'équipe et de l'expertise variée de chacun, nous avons divisé l'équipe en deux sous-équipes. Un premier groupe s'est chargé d'explorer les données pour proposer des croisements et corrélations pertinentes à présenter ensuite sous forme d'une visualisation web ou applicative par le second groupe.

L'objectif n'était pas simplement de participer au concours, prendre l'énoncé pour répondre à la problématique demandée avec les sujets proposés. Nous étions tous d'accord pour à la fois tirer profit au maximum des données et ensuite exposer ces liens grâce à une solution user-friendly et la plus complète possible. Nous avons relevé le défi de la meilleure manière possible : des idées, un travail d'équipe, de la technique et de l'innovation.

## Les principes de développement

La solution développée est une interface générique capable de traiter un nombre limité de structures. Elle a permis de produire des vues complexes et intéressantes sans forcément disposer des données au préalable. Cela offre de nombreux avantages, notamment pour permettre des évolutions à moindre coût. Une architecture simple est souvent la meilleure manière de garantir un entretien et un développement progressif non contraignant.

Techniquement, nous avons mis en place un *back-end* utilisant les principes *REST* pour des interfaces claires et simples. Particulièrement utile dans notre cas avec une solution multi-plateformes. Toutes ces plateformes peuvent récupérer les nouvelles données automatiquement en cas de mise à jour, sans installation requise et une disponibilité continue en fonction de la solution d'hébergement (99.999% d'uptime avec une solution cloud).

Dans nos ambitions d'une solution complète, nous avons répondu aux différents défis proposés dans le cadre du concours :

- Défi n°1 (accès à des données numériques avec sélections de période temporelle et/ou de zone géographique et/ou de seuil) : déplacement à travers la carte interactive et échelle de temps
- Défi n°2 (visualisation graphique des évolutions temporelles) : informations contextuelles complémentaires par aéroports et propagation des retards

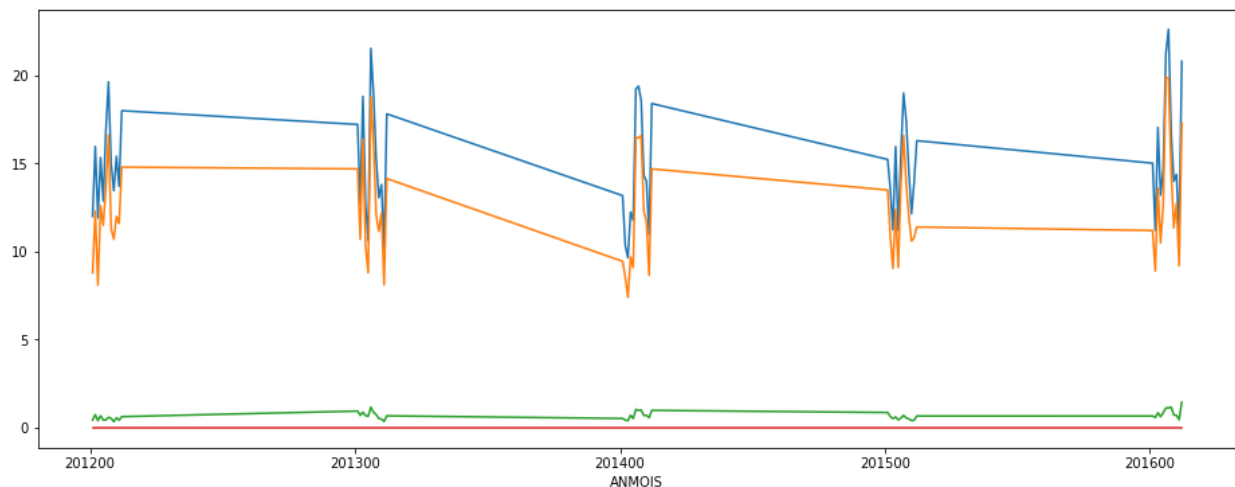
- Défi n° 3 (visualisation de la structure géographique) : affichage d'aéroports et autres objets à travers la carte 3D
- Défi n° 4 (visualisation de corrélations éventuelles) : Propagations des retards et zones d'influence des aéroports
- Défi n°5 (info/avertissement sur des données manquantes) : remontée des erreurs lors de la lecture des fichiers CSV des données et sur l'application par des repères spécifiques

## Les traitements de données

Plusieurs solutions ont été explorées pour valoriser les données fournies.

Nous avons travaillé avec une solution enregistrée dans une base de donnée afin de pouvoir tester différentes requêtes en bénéficiant de tous les avantages propres aux base de données relationnelles (SQL). Cette solution est pertinente pour notre projet car il était important pour nous de séparer le traitement des données de la visualisation. Cette séparation des concepts offre davantage de flexibilité et des mise à jour plus simples.

En parallèle de cette approche, nous avons utilisé des outils de data science (notebook en *Python*) pour étudier les corrélations et identifier des événements intéressants cachés entre les données. Nous avons pu mettre en évidence des mouvements cycliques répétitifs pour les retards. Par exemple, l'évolution des retard moyens pour l'aéroport de Paris "Roissy Charles de Gaulle" ci-dessous :



*Evolution des retards moyens pour l'aéroport Paris - Charles de Gaulle*

Ce diagramme montre clairement des phénomènes cycliques qui se répètent d'années en années avec d'importantes perturbations pendant les périodes de froid, des fêtes de Noël à Février. Cette étude a également permis de créer des relations de dépendances entre aéroports avec des phénomènes de propagation des retards d'un aéroport à l'autre.

# Notre solution

## Description

L'application se présente sous la forme d'une carte 3D de la Terre. Des commandes permettent d'afficher différents types de données :

- Un curseur permettant de choisir l'année des données affichées
- Une liste de représentation de données (ex : émissions de CO2, émissions de particules fines...)
- Une zone d'informations (ex : données contextuelles pour un aéroport en particulier)

## Multi-plateformes

La solution proposée par notre équipe peut être déployée en utilisant *WebGL*, une solution de visualisation en 3D adopté par la plupart des fournisseurs de cartes graphiques pour générer des environnements en 3D sur un navigateur internet.

Avec un environnement en 3 dimensions, il est relativement simple de déployer notre solution sur d'autres types de supports tels que des tablettes et mobiles. Il est également envisageable de proposer des solutions en Réalité Augmenté ou Réalité Virtuelle.

Nous utilisons Unity3D afin de permettre un prototypage rapide au vu des délais de ce concours. Unity n'est pas un logiciel open source mais les solutions que nous déployons sont elles open source, notamment avec *WebGL*.

Voici quelques exemples de notre solution de visualisation sur des supports mobiles :



*Photo de l'application en marche sur un smartphone (Android)*

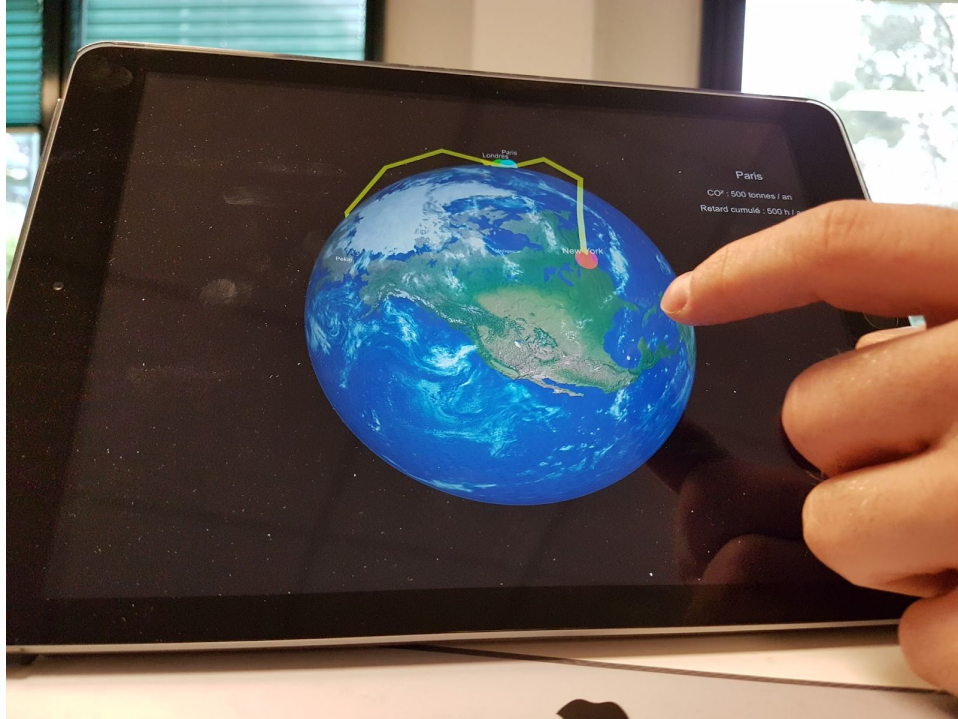


Photo de l'application en marche sur une tablette (iOS)

## Glossaire

Back-end : partie "invisible" d'une application, se chargeant de tout le traitement des données envoyées ensuite à la partie "visible" de l'application.

WebGL : Ensemble de fonctions et méthodes pour les calculs d'objets en 2D et 3D sur un navigateur internet

Notebook Python : environnement de développement interactif pour python

REST : un style d'architecture pour les systèmes hypermédia distribués, créé par Roy Fielding en 2000. Notamment utilisé dans la création de services connectés utilisant Internet.

SQL : langage standardisé servant à enregistrer, manipuler et extraire de l'information dans une base de donnée.