

Printemps 2022

Groupe : ANLO

# [LIFPROJET] Rapport de projet



**ETUDIANT :**

Anne DE JONG

P1905532

Loris MERCIER

P1906860

**ENCADRANT :**

Rémy Cazabet

## Table des matières

1. Introduction .....	1
2. Présentation du projet .....	1
2.1 Qui sommes-nous ? .....	1
2.2 Nos objectifs .....	1
2.3 Nos méthodes de travail .....	2
3. Déroulement du projet .....	3
3.1 De la prise en main à la constitution d'une base de données.....	3
3.2 Exploitation de nos données.....	4
3.3 Visualisation.....	6
4. Conclusion .....	7
5. Annexe : Table des illustrations.....	8

# 1. Introduction

Peu avant la crise covid, nous comptons chaque année plus de 36 millions de vols à travers le monde représentant une moyenne de 1.14 vol par seconde. Avec une telle masse de données, il serait intéressant d'analyser ces chiffres et de visualiser les liaisons majeures entre les différents points du globe.

C'est à partir de ce constat que nous décidons au sein de notre groupe ANLO de travailler sur une exploitation de données du trafic aérien mondial. Avec l'avantage de connaître ce mode de transport, nous pourrions réussir à créer des cartes interactives et itinéraires de vols facilement utilisables permettant ainsi à chaque passager de voir et choisir ses prochaines destinations de vacances !

## 2. Présentation du projet

### 2.1 Qui sommes-nous ?

Le groupe ANLO est constitué de deux étudiants de l'Université Claude Bernard - Lyon 1 en troisième de licence informatique. C'est dans le cadre de notre unité d'enseignement *LIFPROJET* que nous avons choisi de réaliser un projet d'analyse de données baptisé *LifRadarAvion*.

Ayant pour ambition de travailler dans le monde de la donnée dans notre futur, ce projet est pour nous une manière de découvrir une nouvelle application de l'informatique non étudiée au sein de notre licence. C'est également la découverte d'un nouveau langage de programmation qui a porté notre choix sur ce projet de semestre.

### 2.2 Nos objectifs

Très vite après le choix du sujet, nous avons commencé à définir nos objectifs et notre calendrier de travail afin d'avoir une ligne directrice tout au long du semestre.

Notre premier objectif est, comme indiqué auparavant, d'acquérir de nouvelles compétences en prenant un sujet éloigné de nos savoir-faire habituels. Nous risquons de perdre un peu de temps à la prise en main du sujet, mais cela sera très valorisant une fois le projet terminé.

Notre deuxième objectif est d'améliorer notre façon de travailler en équipe en utilisant un gestionnaire de code de façon avancée. (Git dans notre cas). Cela étant

une compétence clé en informatique, ce projet est donc une façon de se préparer au monde professionnel.

Viennent ensuite les objectifs de réalisation. Nous décidons tout d'abord de ne pas partir de données existantes, mais de créer notre propre base de données à partir d'API disponible.

Nous partons ensuite sur trois grands axes d'analyse et de visualisation : les cartes, les graphiques et une exploitation de graphes dans le but de créer des itinéraires.

Une fois tout mis en œuvre, nous aimerions présenter ces résultats sous forme d'application interactive où l'utilisateur pourrait choisir lui-même les paramètres de chaque représentation.

## 2.3 Nos méthodes de travail

Afin d'assurer le bon déroulement d'un projet, il est essentiel de clarifier dès le début les méthodes de travail. Ayant déjà réalisé plusieurs projets en commun, nous avons repris notre façon habituelle de travail.

En premier lieu, notre code est hébergé sur la forge de l'université Lyon 1 ce qui permet de le gérer via le gestionnaire de code *GIT*. Nous avons ensuite chacun un dossier de travail différent permettant d'avancer sur nos parties librement ainsi qu'un code central où nous regroupons au fur et à mesure nos avancées. L'intérêt d'une telle gestion est d'avoir un historique et de pouvoir collaborer facilement à distance.

Concernant la répartition des tâches, nous avons décidé de faire la prise en main du projet ainsi que la récupération des données en commun. C'est une fois notre base de données créée que nous avons divisé le travail pour avoir chacun nos missions. L'avantage de notre projet est qu'il contient 3 axes de développement indépendants ce qui permet une division du travail efficace. Loris a travaillé sur la création de cartes et de graphiques tandis qu'Anne s'est occupée des graphes et de la visualisation finale.

Afin de comparer et/ou mettre en commun nos travaux, nous avons utilisé le créneau du mardi après-midi accordé dans le cadre de l'UE pour faire le point chaque semaine sur l'avancement de notre projet. Cela était également l'occasion d'échanger avec notre encadrant, M. Cazabet. Cependant, nous discutons aussi tout le long de

la semaine au sein du groupe afin de suivre en permanence nos avancées sans attendre notre rendez-vous du mardi.

### 3. Déroulement du projet

#### 3.1 De la prise en main à la constitution d'une base de données

La première difficulté dans un projet comme celui-ci est d'identifier et de se familiariser avec les outils importants. Grâce à l'aide de notre encadrant, nous avons vite compris que notre projet serait réalisé en *python* sur un fichier de type notebook (fichier permettant d'exécuter des portions de codes de façon indépendantes). La première étape a donc été de comprendre ces outils à l'aide des différentes documentations disponibles.

Comme indiqué dans nos objectifs, nous n'avons pas souhaité partir d'une base de données existante. Ces bases auraient certainement été plus complètes mais nous désirions faire un projet complet en données, mêlant toutes les étapes possibles de l'obtention jusqu'à la visualisation des données.

C'est pourquoi à l'aide d'une API reliée au site *FlighRadar24* (site référence pour suivre en temps réel le trafic aérien), nous avons créé un script python permettant de récupérer tous les vols disponibles à l'instant T et de les exporter sous forme d'un fichier csv. Une fois le code confectionné, nous l'avons déposé sur un serveur programmé pour exécuter le script toutes les 30 minutes. Ainsi, après une semaine de requêtes automatiques, nous avons obtenu un fichier csv comprenant tous les vols sans doublons. (La notion de doublon a été défini dans notre projet en tant qu'identifiant du vol. Cela signifie que nous avons obtenu tous les numéros de vols différents des compagnies).

L'inconvénient de notre API est qu'il nous est difficile d'obtenir les dates des vols. Nous avons donc abandonné l'idée d'avoir une chronologie dans nos données.

Finalement, nous nous retrouvons avec un fichier de 118 748 vols aériens à exploiter.

### 3.2 Exploitation de nos données

Notre base de données étant conçue, la partie la plus importante du projet démarre : l'analyse et l'exploitation.

Tout d'abord, nous nous sommes concentrés sur la partie carte. Après quelques difficultés à trouver la bibliothèque idéale, nous avons choisi le module *Folium* pour réaliser toutes nos cartes. Celles-ci ont l'avantage d'être interactives et agréables à visualiser. L'inconvénient est qu'il n'existait pas de fonction pour tracer des "grands cercles" (cercle avec comme centre, le centre de la sphère. Cela correspond au chemin le plus court et réaliste à la surface d'un globe). Nous avons donc créé nous même cette fonction essentielle à nos représentations.

Afin de diversifier nos représentations, nous avons opté pour 4 cartes totalement différentes, utilisant des procédés d'analyse différents comme vous pouvez le voir ci-dessous :



Figure 2: Liaisons nationales françaises



Figure 1: Liaisons aériennes depuis Montréal



Figure 3: Nombre de liaisons aériennes depuis les Pays-Bas



Figure 4: Emplacement des aéroports

Notre deuxième axe de visualisation a été la représentation graphique de nos données. Tout comme les cartes, la première étape fut de trouver la bibliothèque la plus adéquate à notre projet. La plus populaire est sans doute *matplotlib* mais celle-ci n'est pas interactive. Nous avons donc opté pour la bibliothèque *Plotly*.

La difficulté rencontrée ici, qui est commune à toutes nos parties, est qu'il faut trouver pour chaque représentation la bonne manière de manipuler les données pour obtenir ce que l'on veut. Cela peut paraître simple, surtout que python permet d'écrire des choses compliquées en peu de lignes, mais le temps de réfléchir est de trouver la bonne fonction permettant de réaliser parfaitement ce que l'on veut est un procédé assez long lorsque débutons dans ce domaine. Un regroupement de données, suivi d'un filtre avant un renommage final s'écrit en une ligne, mais peut prendre plusieurs heures à concevoir.

Tout comme les cartes, nous avons décidé de présenter les résultats en utilisant des représentations différentes afin de montrer l'étendue des possibilités. Nous avons également essayé d'analyser des caractéristiques pertinentes qui pourraient être interactif :

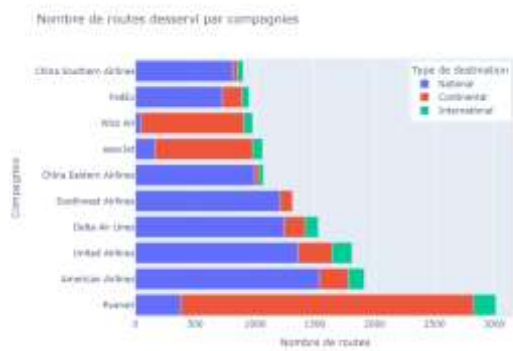


Figure 5: Graphiques à barres du nombre de routes desservi par compagnies

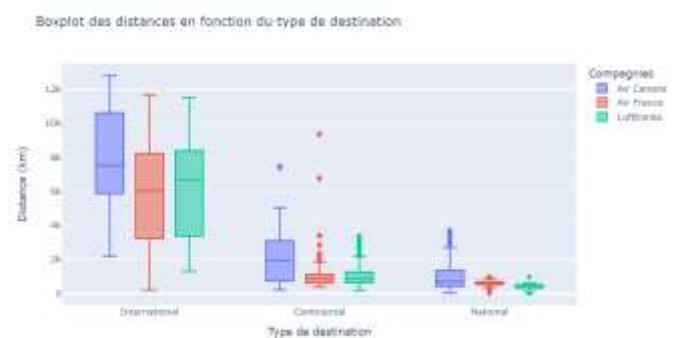


Figure 6: Boxplot des distances en fonction du type de destination

Enfin, nous avons mis nos données sous forme de graphes pour réaliser notre dernière exploitation.

Grâce à des algorithmes de parcours de graphes comme celui de Dijkstra, nous avons pu concevoir des chemins entre deux aéroports partout à travers la planète. En combinant ces résultats aux fonctions d'affichage de cartes écrites précédemment, nous avons pu obtenir des cartes beaucoup plus interactives ne s'appuyant pas sur des données, mais sur un calcul issu de ces données.

Dans le cadre de notre projet, nous avons utilisé la librairie *networkx* pour réaliser des graphes. La partie la plus délicate est ici la transformation de données en graphes. Plusieurs paramètres sont à prendre en compte comme le temps de calcul du graphe car ce dernier peut vite devenir très volumineux au vu du nombre de connexions de chaque aéroport.



Aéroport de départ  
Amsterdam Schiphol Airport

Aéroport d'arrivée  
Los Angeles International Airport

Le chemin le plus court en terme de distance à vol d'oiseau

### 3.3 Visualisation

La grosse difficulté de cette bibliothèque est de l'appréhender afin de réaliser une page web fonctionnelle, agréable à consulter et mettant en valeur nos représentations. De plus, nos objets (cartes, graphiques) n'étant pas issus de bibliothèques standard, il nous a fallu trouver la bonne manière de les afficher afin de garder la notion d'interactivité propre à ces bibliothèques.



Cependant l'avantage est qu'elle permet, une fois prise en main, d'intégrer facilement tous nos résultats depuis les notebooks sans réaliser de multiples requêtes. La comptabilité des widgets déjà présents dans les notebooks standard est également le vrai plus nous ayant fait choisir cette option de visualisation pour nos données.

## 4. Conclusion

Finalement, bien que n'ayant aucune base en analyse de données en début de projet, nous sommes fiers de nos résultats. Nous avons beaucoup appris lors de ces 4 mois et ces connaissances nous seront avantageuses dans la suite de nos études.

Nos objectifs initiaux ont tous été tenus et nous avons pu jouer sur différentes facettes du monde de la donnée : la conception, l'analyse et la visualisation.

Une des difficultés du projet a été le rapport entre le temps consacré à écrire du code et la qualité du rendu. Comme nous le disions au sein de ce rapport, nous passions parfois des journées entières à réaliser une fonctionnalité pour n'obtenir finalement qu'une avancée pas très significative. Mais nous avons vite compris que cela était normal en analyse de données, le plus dur étant de trouver la bonne méthode de formatage des données et non la représentation venant ensuite.

Nous profitons de cette conclusion pour remercier notre professeur encadrant, M. Rémy Cazabet, qui était toujours présent pour nous aider et qui a pu nous orienter vers les bonnes pistes de réflexion lors de nos difficultés.

C'est donc la fin de notre projet LifRadarAvion marquant également la fin de notre semestre. Nous espérons désormais décoller vers de nouveaux projets en Master.

## 5. Annexe : Table des illustrations

Figure 1: Liaisons aériennes depuis Montréal.....	4
Figure 2: Liaisons nationales françaises .....	4
Figure 3: Nombre de liaisons aériennes depuis les Pays-Bas .....	4
Figure 4: Emplacement des aéroports .....	4
Figure 5: Graphiques à barres du nombre de routes desservi par compagnies .....	5
Figure 6: Boxplot des distances en fonction du type de destination .....	5
Figure 7: Itinéraire le plus court entre deux aéroports.....	6