

Contexte :

Les retards de vols coûtent des millions d'euros aux compagnies aériennes (correspondances ratées, trajet suivant aussi en retard, coût d'attente à l'aéroport, ...) et génèrent beaucoup de frustration pour les usagers qui sont sévère envers les compagnies.

IMSD2020 est une compagnie spécialisée dans l'analyse des données issues du ministère public. Elle a pour principaux clients de grandes compagnies aérienne telles que Bla Bla Jet, Creuse AIR LINE ou LeBonAvion. Elle offre un large panel de services autour de la data :visualisation, et prédiction de retards et d'annulations de vols.

Les performances du modèle utilisé pour les prédictions ne sont plus alignées avec la réalité.

Elle vous a donc engagé en tant que data-scientist afin de construire un nouvel algorithme permettant la prédiction des retards et recommander le déploiement fonctionnel de ce modèle pour leurs clients.

Données disponibles

Le directeur de la DSI d' IMSD2020 vous a mis à disposition une base de données contenant les informations de nombreux vols. Cette base de données a été reconstruite à partir d'un backup de notre base principale.

La base de données est un fichier de base [Sqlite](#) (Python inclut une librairie pour lire ce type de fichier).

Documentation des bases de données :

Vous trouverez les deux fichiers contenant des bases de données au format sqlite3.

Fichier batch_1.db :

Le fichier batch_1.db contient 3 tables :

- aeroports
- compagnies
- vols

VOLS

Description

La table contient toutes les informations relatives aux vols.

Nom de la colonne / type

IDENTIFIANT / INTEGER

VOL / INTEGER

CODE AVION / TEXT

AEROPORT DEPART (code IATA) / TEXT

AEROPORT ARRIVEE (code IATA) / TEXT

DEPART PROGRAMME / INTEGER

HEURE DE DEPART / REAL

RETART DE DEPART / REAL

TEMPS DE DEPLACEMENT A TERRE AU DECOLLAGE / REAL

DECOLLAGE / REAL

TEMPS PROGRAMME / REAL

TEMPS PASSE / REAL

TEMPS DE VOL / REAL

DISTANCE / INTEGER

ATTERRISSAGE / REAL
TEMPS DE DEPLACEMENT A TERRE A L'ATTERRISSAGE / REAL
ARRIVEE PROGRAMMEE / INTEGER
HEURE D'ARRIVEE / REAL
RETARD A L'ARRIVEE / REAL
DETOURNEMENT / INTEGER
ANNULATION / INTEGER
RAISON D'ANNULATION / TEXT
RETARD SYSTEM / REAL
RETARD SECURITE / REAL
RETARD COMPAGNIE / REAL
RETARD AVION / REAL
RETARD METEO / REAL
DATE / TEXT
NIVEAU DE SECURITE / INTEGER
COMPAGNIE AERIENNE / TEXT

COMPAGNIES

Description

La table contient toutes les informations relatives aux compagnies aériennes.

Nom de la colonne type

COMPAGNIE / TEXT
CODE / TEXT

AEROPORTS

Description

La table contient toutes les informations relatives aux aéroports.

Nom de la colonne type

CODE IATA (code aéroport) / TEXT
NOM / TEXT
LIEU / TEXT
PAYS / TEXT
LONGITUDE / TEXT
LATITUDE / TEXT
HAUTEUR / REAL
PRIX RETARD PREMIERE 10 MINUTES / INTEGER
PRIS RETARD POUR CHAQUE MINUTE APRES 10 MINUTES / INTEGER

Projet :

Organisation du travail sur le projet

une présentation permettant de
présenter en 20 minutes votre analyse du contexte et la solution technique que vous avez
développé ainsi qu'une présentation des résultats.

En l'absence éventuelle d'informations, vous êtes invité à prendre des hypothèses réalistes
qui seront détaillées dans votre présentation.

Pour ce projet, il n'y a pas de bonne ou mauvaise réponse. Elle a pour objectif de mettre en lumière vos capacités à analyser une situation floue et complexe, à formuler des solutions adaptées au contexte et à restituer vos conclusions à l'écrit. Ce projet a pour but de mettre en application les savoir acquis lors des formations.

But

Vous devez construire le nouveau projet de prévision de retard des vols. Votre objectif est de fournir un maximum d'information que vous estimez utile à notre clientèle.

Le livrable de ce projet sera composé des éléments suivants :

- Une présentation détaillant :
 - la démarche de cadrage expliquant :
 - comment vous avez choisi la colonne à prédire
 - le choix de la métrique d'optimisation
 - une analyse du dataset (un rapide commentaire sur chacune des colonnes)