

Sources de données

Vous trouverez ci-dessous une brève description des données qui seront manipulées pendant le projet. Pour avoir une vision plus précise de la composition de ces fichiers, vous pouvez consulter le dictionnaire des données disponibles sur l'espace Moodle.

- **Source 1. Le fichier de commentaires (ALL DATA ANNOTATE)**

Il s'agit du fichier principal (et le plus volumineux) manipulé par les étudiants. Dans ce fichier, une ligne équivaut à une phrase d'un commentaire. Les commentaires en langue non anglaise sont présents dans le fichier. Ce fichier regroupe toutes les sources de données qui disposent de commentaires.

En réalité, cette source de données se décompose en deux fichiers distincts. Le premier (**source 1a**) contiendra les données non labellisées alors que le second (**source 1b**) aura la même structure que le premier mais contiendra en plus des colonnes associées aux labels à prédire (issue d'une labellisation manuelle).

- **Source 2. Fichier de routes (DATA KAYAK ROUTE)**

Ce fichier énumère les routes opérées par les compagnies aériennes. Ainsi, pour chaque numéro de vol, nous connaissons le nom de la compagnie, l'origine et la destination du vol ainsi que le type d'avion effectuant ce vol.

- **Source 3. Fichier de notes des compagnies (KAYAK AIRLINE RATE GLOBAL)**

Ce fichier représente une vision agrégée (moyenne) de l'opinion des passagers sur quelques caractéristiques d'intérêt (confort, divertissement, nourriture, ...). Nous disposons également du nombre d'évaluations prises en compte pour agréger les notes.

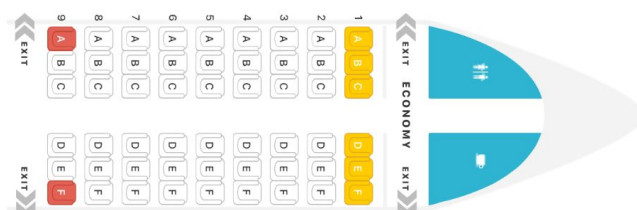
- **Source 4. Fichier équipement (SEATGURU INFO AIRCRAFT)**

Permet de connaître les différentes configurations des avions pour chaque compagnie (plusieurs layout par couple (compagnie, modèle avion) possible).

- **Source 5. Plan des avions (non décrite dans le dictionnaire de données)**

Fichiers image pour connaître la configuration des avions.

Exemple :



- **Source 6. Images sociales (non décrite dans le dictionnaire de données)**

Trois sources ont été scrappées (SeatGuru, Seat Maestro et Instagram) pour récupérer les images associées à des hashtags cible. Pour Instagram, nous disposons également des posts associés.

1. Collecte de données

Enseignant référent : Morgan Séguéla (sequela@irit.fr)

Composition

- R. Sedran (chef de groupe)
- A. Pinson (adjoint)
- A. Diallo (responsable qualité)
- G. Bernard
- B. Anton
- E. Lagaize
- F. Seck
- N. Aider
- N. Marco
- J. Boutibou
- L. Marquez

Description

La mission de ce groupe est de réaliser des robots pour collecter les données des sources suivantes :

- TripAdvisor
- Weibo
- Youtube
- Facebook
- Rome2Rio
- AirlineNet

Données en entrée : N/A

Interaction avec les autres groupes :

- Fournit des entrées pour tous les groupes (sauf les groupes "images")

User Story associée : N/A

Délivrables:

- Données documentées
- Scripts de collecte commentées
- Un rapport descriptif par source (volume, organisation des répertoires, lancement des scripts, mise en place de cron, processus de reprise sur panne, ...)

2. Statistiques

Enseignant référent : Cécile Chouquet (cecile.chouquet@math.univ-toulouse.fr)

Composition

- C. Ollivier (cheffe de groupe)
- L. Krumeich (adjointe)
- E. Martineau (responsable qualité)
- M. Seck
- M. Pace
- D. Cadet
- C. Leroyer
- A. Andrieu
- S. Mirouze
- D. Chalogné
- J. Mora Fernandez

Description

Statistiques descriptives sur toutes les sources de données et explication des modèles.

Données en entrée : Sources 1-4

Interaction avec les autres groupes :

- Prend des entrées du groupe 1 collecte des données + des groupes de prédiction 5
- Fournit des entrées pour le groupe 3 Dashboard (au moins) et 6

User Story associée :

Correlation	see if there is a correlation between comfort and other parameters	(such as price, aircraft, cabin class...)
--------------------	--	---

Délivrables: Notebooks Python

3. Dashboard

Enseignant référent : Shaoyi Yin (Shaoyi.Yin@irit.fr)

Composition

- Q. Marcu (chef de groupe)
- N. Crenner (responsable qualité)
- T. Roy
- R. Yahyaoui
- A. Airault
- V. Bitier
- T. Zeutzius
- C. Pandoupy
- T. Chateau
- T. Saccareau

Description

- Visualisation des données annotées ou prédites sous forme de graphes/ tableaux,...
- Visualisation des commentaires sous forme de nuages de mots, analyse de sentiments
- Tout cela en fonction des compagnies, des jours, des aéroports, ...

Données en entrée : Source 1-4, 6 + sortie des groupes Prédiction et Texte

Interactions avec les autres groupes :

Prend des entrées des entrées du groupe 1 collecte des données + des groupes des prédictions 5 et 6 + groupe 2 des statistiques descriptives + groupe 4 texte

User story associée:

Overall comfort	have a dashboard of perceived economy comfort at aircraft level	(i.e. do people talk positively or negatively about passenger seating for a given aircraft type operating in a given region)
------------------------	---	--

Délivrable :

- Logiciel de Dashboard avec code commenté

4. Texte

Enseignant référent : Karen Pinel-Sauvagnat (sauvagnat@irit.fr)

Composition

- V. Fraysse (chef de groupe)
- C. Gaillard (adjoint)
- F. Rollet (responsable qualité)
- C. Doublet
- N. Messaoudi
- J. Cosio
- M. Bah
- M. Audigier
- C. Bertrand-Orliac
- J. Contarin
- Y. Seba

Description

- Nettoyage du texte (notamment Twitter, Instagram)
- Analyse de sentiments
- Découverte de facettes : clustering qui pourrait aider à la labellisation, au dashboard,

Données en entrée : Sources 1 et 6 (posts Instagram)

Interactions avec les autres groupes :

Prend des entrées des entrées du groupe 1 collecte des données

Fournit des entrées pour les groupes des prédictions 5 et 6 + groupe 3 dashboard

User story associée :

Overall comfort	have a dashboard of perceived economy comfort at aircraft level	(i.e. do people talk positively or negatively about passenger seating for a given aircraft type operating in a given region)
------------------------	---	--

Délivrable : Notebooks Python

5. Prédiction 1

Enseignant référent : Yoann Pitarch (pitarch@irit.fr)

Composition

- L. Bouzidi (cheffe de groupe)
- L. Barthélémy (adjoint)
- K. Niang (responsable qualité)
- F. Horellou
- J. Johann
- K. Belaid
- A. Carlet
- P. Dupuy
- S. Gabriel
- M. Corbeau
- D. Sonnevile

Description

A partir des données annotées, prédire les caractéristiques de confort (voir user stories pour savoir quelles sont ces caractéristiques)

Données en entrée : Sources 1-4

Interactions avec les autres groupes :

Prend des entrées des entrées du groupe 1 collecte des données + groupe 4 texte

Fournit des entrées pour le groupe 3 dashboard

User stories associées:

SEAT	Understand passenger feedback concerning passenger seating for a given airline,aircraft manufacturer and type , for given route (Origin&Destination) and class?	Airline = Air France, EASYJET... Manufacturer = Airbus, BOEING... Type = A321, B777-800... Route = Paris - London Comfort = seat width, legroom...
BED	Understand passenger feedback concerning passenger seating for a given airline,aircraft manufacturer and type , for given route (Origin&Destination) and class?	Airline = Air France, EASYJET... Manufacturer = Airbus, BOEING... Type = A321, B777-800... Route = Paris - London

IFE	Understand passenger feedback concerning the entertainment offered and associated services	Entertainment = Video, audio, Internet, wifi...
FOOD	Measure food services inside the aircraft for a given airline, aircraft manufacturer and type for given route (Origin&Destination) and class	Food services = Meal, beaverage
NOISE	Evaluate the perception of noise within the cabin during flight coming from galleys, lavatories, engines, passenger...	Where is it coming from and how loud it is
TEMPERATURE	Evaluate the perception of temperature within the cabin during flight	Is the temperature comfortable?
HUMIDITY	Evaluate the perception of humidity within the cabin during flight coming from air conditionning, overall aircraft...	Is the humidity comfortable?
CABIN CREW	Evaluate the perception of the cabin crew during flight	Are they kind and resourceful?
LAVATORY SPACE	Evaluate the perception of lavatory space within the cabin during flight	Are the lavatories spacious, clean and comfortable?
ATTRACTIVE AIRCRAFT	Understand passenger feedback concerning aircraft branding for a given airline, aircraft manufacturer and type , for given route (Origin&Destination) and class	spacious, comfortable and storage of hand luggage
Cabin space	Evaluate the perception of cabin space	cabin width, large aisle...

Délivable : Notebooks Python

6. Prédiction 2

Enseignant référent : Max Halford (halford@irit.fr)

Composition

- L. Besnard (cheffe de groupe)
- A. Caminade (adjoint)
- C. Le Brun (responsable qualité)
- T. Dirat
- A. Degaugue
- M. Charry
- F. Azoughag
- V. Blase
- T. Diallo
- A. Godin
- S. Stricot

Description

- A partir des données annotées, prédire les caractéristiques qui ne sont pas liées au confort (liée à la compagnie ou à l'aéroport)

Données en entrée : Sources 1-4

Interactions avec les autres groupes :

Prend des entrées des entrées du groupe 1 collecte des données + groupe 4 texte

Fournit des entrées pour le groupe 3 dashboard

User stories associées :

PRICE	Evaluate the perception of the price of a flight for a given airline,aircraft manufacturer and type , for given route (Origin&Destination) and cabin class	Airline = Air France, EASYJET... Manufacturer = Airbus, BOEING... Type = A321, B777-800... Route = Paris - London Cabin class = Business, Eco..
LOST BAGGAGE	Understand passenger feedback concerning baggage treatment for a given airline,aircraft manufacturer and type , for given route (Origin&Destination)	Baggage/luggage
CHECK IN	Evaluate the perception of the ground staff at airport check in desk	efficient and professional
BOARDING	Evaluate the perception of the ground staff at	efficient and professional

	airport boarding	
PUNCTUALITY	Evaluate the perception of the aircraft punctuality at departure and arrival	efficient and professional
SAV	Evaluate the perception of the airline services offered to passengers	refund policy, communication to passenger

Délivvable : Notebooks Python

7. Images sociales

Enseignant référent : Jérôme Farinas (Jerome.Farinas@irit.fr)

Composition

- C. Latorre (cheffe de groupe)
- P. Sinel-Boucher (adjoint)
- E. Grandgirard (responsable qualité)
- G. Hyvert
- L. Dulinge
- M. Da Ponte
- F. Talmat Ammar
- A. Goulif
- V. Barudio
- C. Diop

Description

À partir des images Instagram, prédire :

- L'avionneur (Airbus vs Boeing)
- Le modèle d'avion
- Le type d'images (intérieur, extérieur, autre)
- Les éléments constituant d'une cabine (ex : siège)

Données en entrée : Source 6 + éventuellement images scrappées d'intérieurs d'avion et d'avions

Interactions avec les autres groupes :

Éventuellement le groupe 1 de collecte des données pour les images

User stories associées :

Instagram	Use instagram posting as indication of aircraft branding	Airspace
-----------	--	----------

Délivrable: Notebooks Python

8. Images cabine

Enseignant référent : Mathieu Serrurier (mathieu.serrurier@irit.fr)

Composition

- V. Dang (chef de groupe)
- W. Azzouza (adjoint)
- C. Marque (responsable qualité)
- C. Marcelo
- S. Bezombes
- J. Daurat
- C. Gaussail
- S. Benhamouche
- H. Hadda
- T. Vedis

Description

À partir des plans d'avion, prédire la position (X,Y) des éléments constituant d'un intérieur d'avions (sièges, WC, bar, ...)

Données en entrée : Source 5

Interactions avec les autres groupes : N/A

User stories associées : N/A

Délivrable : Notebook Python

9. Innovation

Enseignant référent : Gilles Hubert (hubert@irit.fr)

Composition

- D. Jeanneau (chef de groupe)
- A. Gahn (adjoint)
- M. Pouchan (responsable qualité)
- N. Enjalbert
- F. Estermann
- T. Gallou
- M. Lesavourey
- N. Kired
- M. Manson

Description

- Que faire d'original avec ces données ?
- Profilage des utilisateurs ?
- Découverte des routes ?

Données en entrée : Tout

Interactions avec les autres groupes : Aucune

User story associée: N/A

Délivrable: Notebooks Python