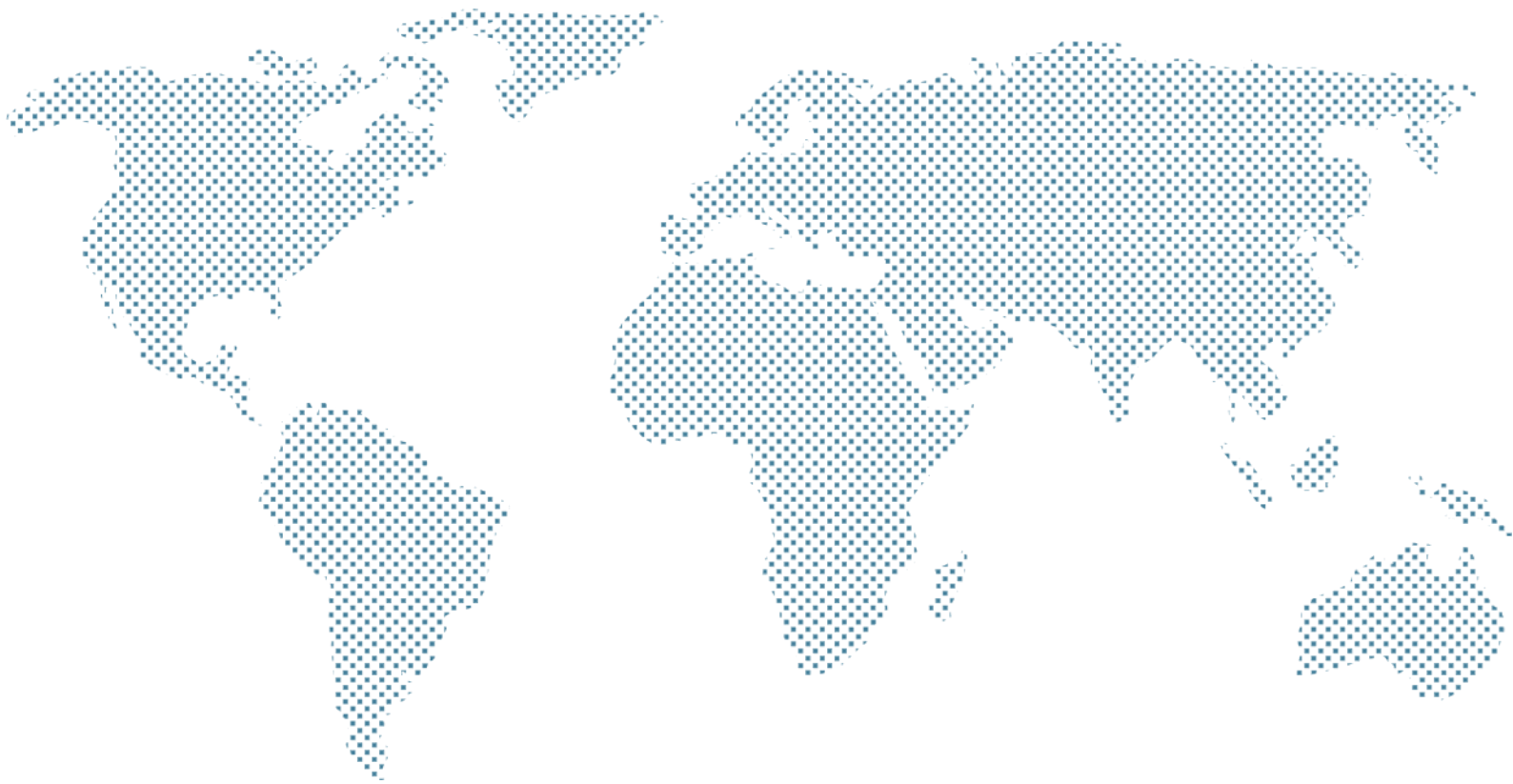


Analyse des retards de vols du 1er janvier 2023

Projet de visualisation de données — Tableau



Hajar Chakir
CHAH29620500

UQAC
UNIVERSITÉ DU QUÉBEC
À CHICOUTIMI

Table des matières

Introduction	3
I-Présentation et justification du jeu de données	4
II-Choix du format : tableau de bord interactif	5
III-Questions de recherche	6
IV-Justification des visualisations et choix de design	6
1. Retards moyens par heure de la journée (graphique en ligne)	7
2. Retards moyens à l'arrivée par compagnie aérienne.....	7
3. Relation entre la distance et le retard à l'arrivée	8
4. Retards moyens à l'arrivée par aéroport (carte)	9
5. Cohérence visuelle et interactivité	10
V-Insights clés.....	10
VI-Limites et pistes d'amélioration.....	11
VII- Conclusion.....	12
Annexe :	13
Lien pour y accéder :	13

Introduction

La ponctualité des vols constitue un enjeu central pour les compagnies aériennes, les aéroports et les passagers. Les retards génèrent des coûts opérationnels importants, perturbent la rotation des appareils et dégradent l'expérience de voyage. Afin de mieux comprendre les facteurs qui influencent ces retards, ce projet propose une analyse visuelle des vols effectués le 1er janvier 2023 aux États-Unis, à partir d'un tableau de bord interactif conçu sous Tableau.

L'objectif est de transformer un jeu de données brut en un outil analytique clair et exploratoire, permettant d'examiner les retards sous différents angles. Le tableau de bord s'appuie pour cela sur plusieurs visualisations complémentaires : tendances temporelles, comparaisons entre compagnies aériennes, analyse géographique et étude de relations entre variables opérationnelles.

L'approche adoptée mobilise les principes essentiels de la visualisation de données, notamment l'usage approprié des encodages visuels, des attributs pré-attentifs et des lois de Gestalt, afin d'assurer une lecture cohérente et efficace des informations. Cette introduction présente ainsi le cadre général du projet et prépare les sections suivantes, consacrées au jeu de données, aux choix de conception, aux analyses réalisées et aux principaux enseignements tirés de l'exploration.

I-Présentation et justification du jeu de données

Le jeu de données utilisé dans ce projet provient de la plateforme Kaggle et correspond à un extrait détaillé des vols effectués le **1er janvier 2023** par plusieurs compagnies aériennes américaines. Il contient **435 352 enregistrements** et **19 variables**, ce qui représente un volume de données largement suffisant pour mener une analyse fiable et statistiquement significative.

Les variables incluent notamment :

- des informations temporelles (*heure de départ, heure d'arrivée, heure prévue, délai de départ, délai d'arrivée, etc.*) ;
- des informations opérationnelles (*numéro de vol, compagnie aérienne, immatriculation de l'appareil*) ;
- des données géographiques (*aéroport d'origine et aéroport de destination*) ;
- des mesures de performance aéronautique (*temps de vol, distance parcourue*).

Ces données permettent d'aborder un enjeu central dans le domaine du transport aérien : **la ponctualité des vols**. Les retards sont un sujet majeur pour les compagnies, les gestionnaires d'aéroports et les usagers, car ils ont un impact direct sur les coûts opérationnels, l'organisation des rotations, la satisfaction des passagers et la fluidité globale du trafic.

Ce dataset est particulièrement pertinent pour une visualisation interactive car :

1. **Il combine plusieurs dimensions d'analyse** (temps, espace, compagnies, distances), ce qui se prête très bien à une exploration via dashboard.
2. **Il contient des mesures numériques précises** permettant de calculer des retards moyens, des distributions ou des relations entre variables.
3. **Il comporte des variables catégorielles faciles à filtrer** (compagnie, aéroport), ce qui enrichit l'interactivité.
4. **Il offre un contexte réel et concret**, facilitant l'interprétation des graphiques pour un public non spécialiste.

Bien que ce tableau de bord puisse être utile à différents publics (gestionnaires aériens, compagnies aériennes ou passagers curieux), la conception s'appuie sur une persona principale issue de la démarche de conception centrée utilisateur (CCU). Cette persona correspond à un analyste opérationnel chargé d'examiner les causes des retards, de comparer la performance des compagnies et d'identifier les aéroports problématiques.

L'objectif n'est pas de restreindre l'usage du tableau de bord à ce seul profil, mais de garantir une structure claire et fonctionnelle : choix des visualisations, niveau de détail, encodages visuels et positionnement des graphiques. Cette approche permet de concevoir un dashboard optimal tout en conservant une utilité pour un public plus large.

Ainsi, ce jeu de données constitue une base solide, riche et cohérente pour analyser les facteurs influençant les retards aériens et pour construire un tableau de bord interactif offrant une exploration multidimensionnelle des retards de vols.

II-Choix du format : tableau de bord interactif

Pour ce projet, j'ai choisi de réaliser un **tableau de bord interactif (dashboard)** plutôt qu'un récit narratif. Ce choix s'explique principalement par la nature du jeu de données, qui contient de nombreuses variables temporelles, catégorielles et numériques, et qui se prête particulièrement bien à une exploration multidimensionnelle.

Le dashboard permet en effet à l'utilisateur :

- **d'explorer librement les données** grâce aux **filtres** (compagnie, aéroport, heure, etc.) ;
- **d'analyser simultanément plusieurs dimensions** (temps, distance, géographie, performance des compagnies) ;
- **de croiser les informations en temps réel**, par exemple en affectant le même filtre à plusieurs visualisations ;
- **de découvrir des insights personnalisés**, en fonction des critères sélectionnés, sans suivre un parcours imposé.

Contrairement au récit narratif, qui impose une lecture linéaire et progressive, le tableau de bord offre une **vision globale, interactive et non séquentielle**, ce qui est particulièrement pertinent pour analyser les retards de vols. Les retards peuvent dépendre de nombreux facteurs (heure de départ, compagnie, aéroport, distance du vol), et un format interactif permet d'observer ces relations de manière plus flexible et plus informative.

Enfin, le dashboard facilite l'intégration de plusieurs types de visualisations complémentaires (carte, ligne temporelle, bar chart comparatif, scatterplot), tout en conservant une **cohérence visuelle et analytique**. Ce format permet ainsi de répondre efficacement aux différentes questions de recherche et de mettre en évidence les principaux facteurs associés aux retards.

III-Questions de recherche

L'analyse menée dans ce projet repose sur cinq questions de recherche principales, définies en fonction de la structure du jeu de données et des enjeux liés à la ponctualité aérienne. Ces questions orientent l'exploration visuelle et permettent d'identifier les facteurs pouvant influencer les retards.

1. **Comment les retards varient-ils selon l'heure de la journée ?**
Cette question permet d'examiner les tendances temporelles et d'identifier les périodes critiques où les retards sont plus fréquents ou plus importants.
2. **Quelles compagnies aériennes présentent les retards moyens les plus élevés ou les plus faibles ?**
Elle vise à comparer la performance opérationnelle des différentes compagnies afin de comprendre les variations de ponctualité selon l'opérateur.
3. **Quels aéroports enregistrent les retards moyens les plus importants ?**
Cette question explore les disparités géographiques et permet de repérer les aéroports les plus affectés par les retards.
4. **La distance ou la durée du vol influence-t-elle les retards observés à l'arrivée ?**
L'objectif est d'étudier la relation entre les caractéristiques du vol et sa ponctualité, afin de voir si les longs trajets sont davantage sujets aux retards.
5. **Les retards au départ entraînent-ils systématiquement des retards à l'arrivée ?**
Cette question examine la propagation des retards et permet d'évaluer si un retard initial tend à se répercuter sur l'arrivée.

Ces questions sont directement liées aux variables du dataset et sont particulièrement adaptées à une exploration visuelle interactive. Elles couvrent des dimensions temporelles, géographiques et opérationnelles, et permettent d'obtenir une compréhension globale des facteurs associés aux retards de vols.

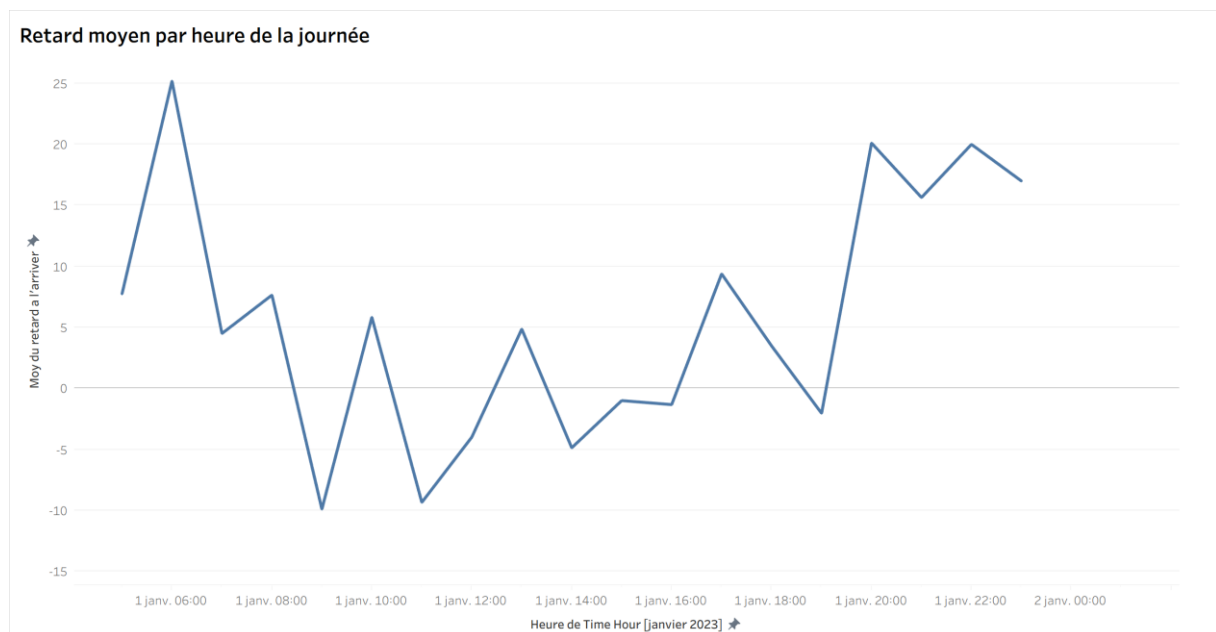
IV-Justification des visualisations et choix de design

Les visualisations présentées dans le tableau de bord ont été sélectionnées en fonction de leur adéquation aux types de données, de leur lisibilité et de leur capacité à répondre efficacement aux questions de recherche. Chaque graphique exploite des encodages visuels adaptés (position, longueur, couleur, orientation), ainsi que des principes perceptifs et cognitifs tels que les attributs pré-attentifs et les lois de Gestalt, afin d'optimiser l'interprétation de l'information.

1. Retards moyens par heure de la journée (graphique en ligne)

Cette première visualisation prend la forme d'un graphique en ligne. Ce choix se justifie par la nature temporelle de la variable analysée : l'heure de la journée constitue une dimension continue, pour laquelle un graphique linéaire permet de représenter efficacement l'évolution d'une mesure au cours du temps. La position horizontale encode l'heure, tandis que la position verticale encode le retard moyen à l'arrivée, ce qui correspond aux encodages visuels les plus précis pour la comparaison quantitative.

La continuité de la courbe facilite la perception immédiate des tendances, notamment l'identification de périodes associées à des retards plus importants. L'orientation et la continuité de la ligne constituent des attributs pré-attentifs qui guident automatiquement l'œil vers les variations les plus marquées. La visualisation s'appuie également sur les lois de Gestalt, en particulier la loi de continuité, qui permet au lecteur de percevoir les points successifs comme faisant partie d'une trajectoire cohérente, ainsi que la loi de proximité, qui associe naturellement les points temporellement proches.

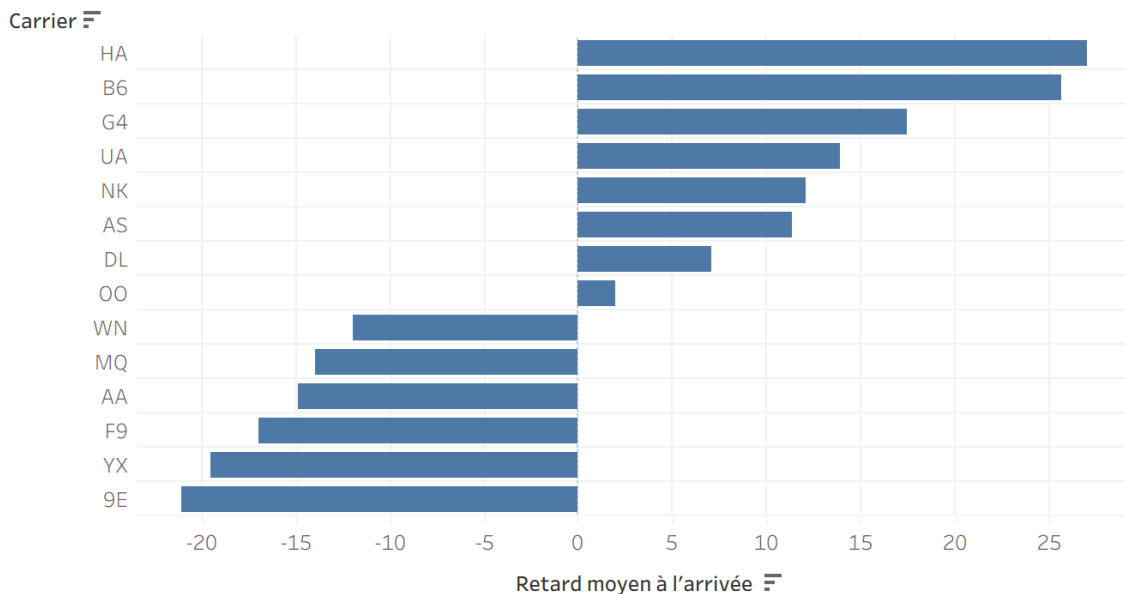


2. Retards moyens à l'arrivée par compagnie aérienne

La comparaison des compagnies aériennes s'effectue à l'aide d'un graphique à barres horizontales. Ce type de visualisation est particulièrement adapté aux données catégorielles nominales, car il permet une comparaison directe entre les longueurs des barres. La longueur, canal visuel très précis, encode ici la valeur du retard moyen. L'alignement horizontal des barres facilite la lecture et réduit la charge perceptive.

Le tri décroissant renforce encore la lisibilité en mettant immédiatement en évidence les compagnies les plus ponctuelles et les moins performantes. Les lois de Gestalt, notamment celles de similitude et d'alignement, permettent au lecteur de percevoir les barres comme un ensemble structuré et comparable. L'utilisation d'une couleur uniforme contribue également à maintenir une cohérence visuelle, évitant toute distraction inutile.

Retard moyen à l'arrivée par compagnie aérienne

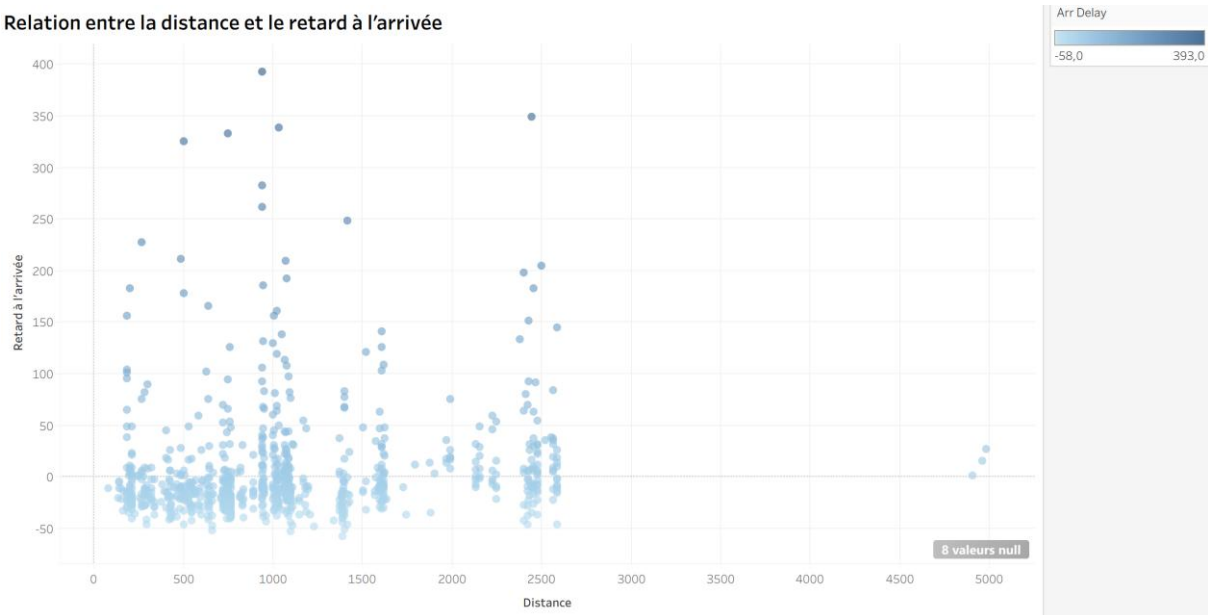


3. Relation entre la distance et le retard à l'arrivée

Pour analyser la relation entre deux variables numériques continues, à savoir la distance du vol et le retard à l'arrivée, un nuage de points a été choisi. Le scatterplot est le type de visualisation le plus pertinent pour examiner des corrélations potentielles, la présence de motifs ou l'existence de groupes distincts. Ici, la position horizontale encode la distance tandis que la position verticale encode le retard.

La position dans l'espace constitue un attribut pré-attentif extrêmement efficace : elle permet de distinguer immédiatement les regroupements, la dispersion ou les valeurs extrêmes. La transparence appliquée aux points améliore la lisibilité dans les zones où les observations se superposent. Les lois de Gestalt, notamment celles de proximité et de regroupement, favorisent l'identification naturelle des clusters, tels que les vols courts présentant généralement des retards plus faibles.

Relation entre la distance et le retard à l'arrivée

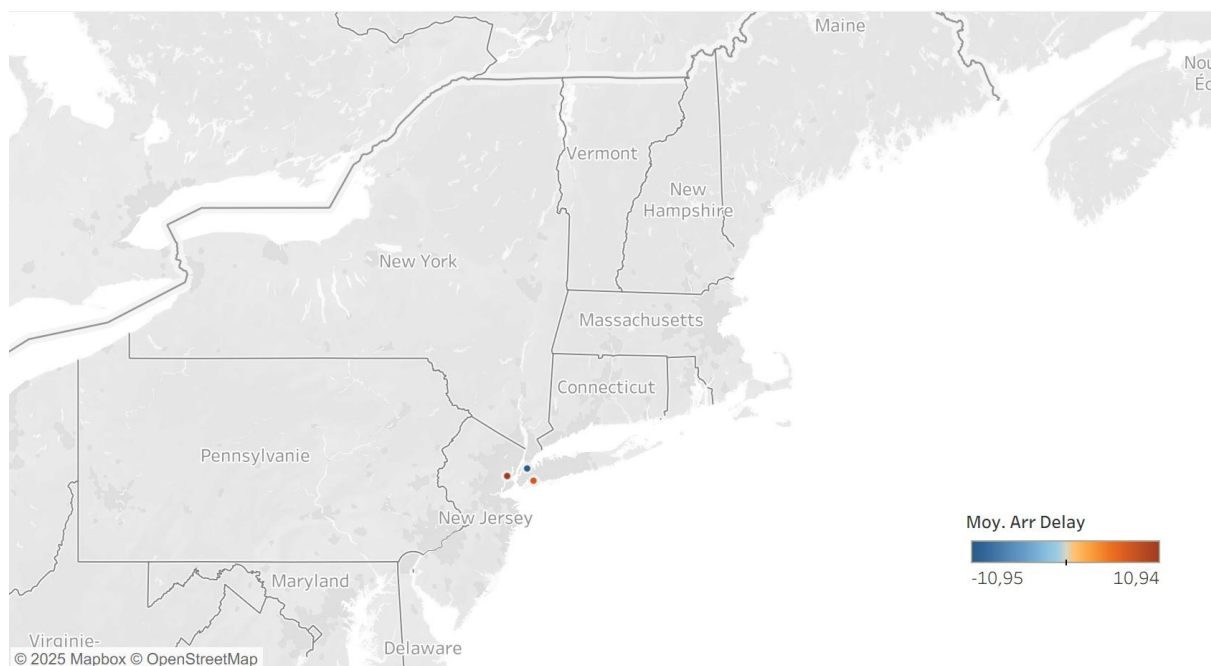


4. Retards moyens à l'arrivée par aéroport (carte)

La représentation cartographique a été retenue pour visualiser les retards moyens selon l'aéroport d'origine. La carte constitue le support le plus intuitif pour toute donnée géolocalisée, car elle associe immédiatement un lieu réel à l'information visuelle. Tableau génère automatiquement les coordonnées géographiques à partir des codes d'aéroport, permettant de positionner précisément chaque point sur la carte.

Dans cette visualisation, la couleur encode le retard moyen à l'arrivée selon une palette divergente, ce qui facilite la détection rapide des aéroports présentant les retards les plus importants. La taille du marqueur a été ajustée de manière à assurer la visibilité tout en évitant la surcharge visuelle. Les attributs pré-attentifs liés à la couleur et à la position permettent au lecteur de repérer instantanément les zones problématiques. La carte mobilise également des lois de Gestalt comme la région commune, car les frontières géographiques structurent naturellement la perception de l'espace.

Retards moyens à l'arrivée par aéroport



5. Cohérence visuelle et interactivité

L'ensemble des visualisations a été harmonisé afin de garantir une lecture fluide et professionnelle du tableau de bord. Une palette de couleurs accessible aux personnes daltoniennes a été utilisée, conformément aux bonnes pratiques d'inclusivité. Les polices et tailles de texte ont été uniformisées afin de renforcer la cohérence visuelle. De plus, l'utilisation de bordures discrètes permet de structurer l'espace sans distraire le lecteur. Les filtres globaux appliqués aux visualisations favorisent une exploration interactive, permettant d'adapter l'analyse à différents besoins et profils d'utilisateurs.

V-Insights clés

L'analyse du tableau de bord met en évidence plusieurs tendances importantes concernant les retards des vols du 1er janvier 2023. Ces observations permettent de mieux comprendre les facteurs associés à la ponctualité et les disparités entre compagnies, horaires et caractéristiques de vol.

Un premier résultat notable est la variation marquée des retards selon l'heure de la journée. Les plages horaires du matin présentent des retards généralement faibles, tandis qu'une augmentation progressive est observée en fin d'après-midi et en soirée. Cette tendance peut s'expliquer par l'effet cumulatif des retards : les perturbations survenues plus tôt dans la journée se répercutent sur les vols suivants, entraînant un accroissement des retards en fin de journée.

La comparaison entre compagnies aériennes révèle également des différences importantes de performance. Certaines compagnies affichent des retards moyens modérés, voire négatifs (arrivées en avance), tandis que d'autres accumulent des retards significatifs. Cette hétérogénéité met en lumière des disparités dans la gestion opérationnelle, notamment en matière de planification, de rotation des avions et de capacité à absorber les perturbations quotidiennes.

L'analyse géographique montre que les trois aéroports étudiés (JFK, EWR, LGA) présentent des niveaux de retard distincts. Certains aéroports sont plus touchés par les retards, ce qui peut s'expliquer par leur niveau de congestion, leurs conditions météorologiques locales ou les particularités de leur gestion du trafic. Bien que le dataset ne couvre que ces trois aéroports, la carte permet néanmoins d'observer clairement leurs différences de performance.

Enfin, l'étude de la relation entre la distance du vol et le retard à l'arrivée suggère qu'il n'existe pas de lien simple entre ces deux variables. Les retards importants touchent indifféremment des vols courts et des vols longs, ce qui implique que la durée ou la distance du trajet ne constitue pas un facteur déterminant en soi. Les retards semblent davantage influencés par des facteurs extérieurs tels que les opérations au sol, la saturation des hubs ou les perturbations rencontrées en amont.

VI-Limites et pistes d'amélioration

Bien que le tableau de bord permette de dégager plusieurs tendances significatives concernant les retards de vols du 1er janvier 2023, l'analyse présente un certain nombre de limites inhérentes au jeu de données et au périmètre du projet. Ces limites doivent être prises en considération pour interpréter les résultats et envisager des améliorations futures.

La première limite concerne la **période d'observation**, restreinte à une seule journée. L'analyse ne permet donc pas d'identifier des tendances saisonnières, hebdomadaires ou mensuelles, ni de distinguer les jours atypiques (fêtes, conditions météorologiques particulières, hausse du trafic). Une extension du dataset à plusieurs semaines ou mois permettrait d'obtenir une vision plus robuste et plus générale de la ponctualité aérienne.

Une deuxième limite tient à la **couverture géographique réduite** du jeu de données. Les vols inclus proviennent essentiellement de trois aéroports majeurs de la région de New York (JFK, EWR et LGA). L'analyse géographique reste donc limitée, et les conclusions ne peuvent pas être généralisées à l'ensemble des aéroports américains ou internationaux. L'intégration de données provenant d'autres hubs offrirait une perspective plus complète.

Les variables disponibles constituent également une limite : le dataset ne contient **aucune information sur les conditions météorologiques**, les niveaux de congestion, les problèmes techniques ou les décisions opérationnelles, qui sont pourtant des facteurs déterminants dans les retards de vols. L'ajout de ces variables permettrait de développer des modèles d'analyse plus explicatifs.

Enfin, certaines analyses visuelles, notamment le scatterplot distance–retard, montrent une grande dispersion, ce qui suggère que **la distance n'est pas un facteur explicatif suffisant**. Cela confirme la nécessité d'intégrer d'autres dimensions ou d'appliquer des méthodes statistiques plus avancées si l'objectif était de prédire les retards.

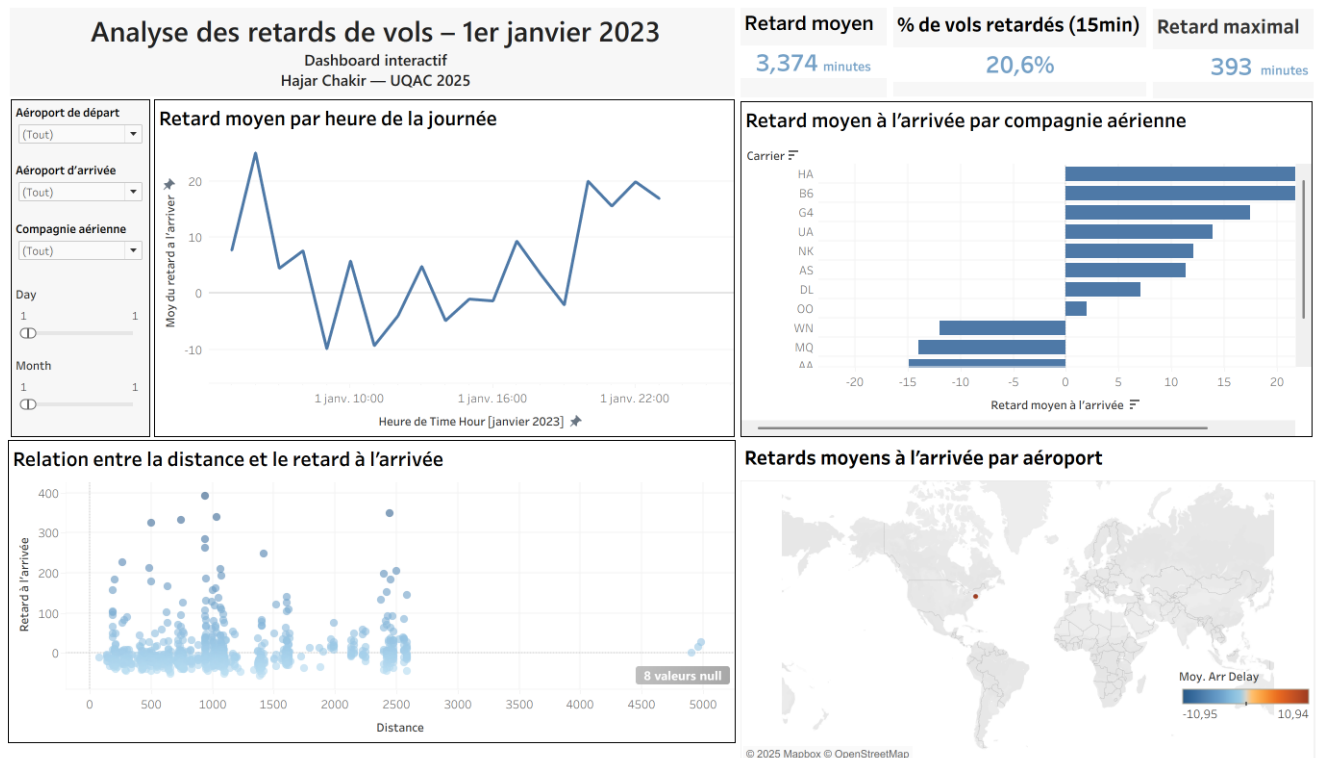
En dépit de ces limites, le tableau de bord constitue une base solide pour l'exploration des retards et peut être enrichi dans de futurs travaux par un dataset plus complet, des variables additionnelles et des analyses plus approfondies.

VII- Conclusion

Ce projet a permis d'analyser les retards de vols du 1er janvier 2023 à l'aide d'un tableau de bord interactif conçu sous Tableau. Les visualisations réalisées — évolution des retards selon l'heure, comparaison entre compagnies, analyse géographique et relation entre distance et retard — ont mis en évidence plusieurs tendances importantes, notamment l'augmentation des retards en fin de journée, des différences significatives entre compagnies aériennes et des variations marquées entre les trois aéroports étudiés.

Bien que le dataset présente certaines limites, notamment une période d'observation restreinte et une couverture géographique limitée à la région de New York, le dashboard a montré son efficacité pour explorer rapidement les données et répondre aux questions de recherche. Ce travail illustre l'intérêt de la visualisation interactive pour comprendre des phénomènes opérationnels complexes comme la ponctualité aérienne.

Annexe :



Lien pour y accéder :

[tableau_bord_hajar_CHAKIR | Tableau Public](https://public.tableau.com/app/profile/hajar.chakir/viz/tableau_bord_hajar_CHAKIR/Tableaud_ebord1?publish=yes)

https://public.tableau.com/app/profile/hajar.chakir/viz/tableau_bord_hajar_CHAKIR/Tableaud_ebord1?publish=yes