2023

Joachim/Pierre/Walid

Simplon

03/11/2023



Determination de la Satisfaction des passagers des compagnies aériennes

Table des matières

[A. Analyse exploratoire des données 3](#_Toc149744201)

[1. Valeurs manquantes 3](#_Toc149744202)

[2. Valeurs aberrantes 4](#_Toc149744203)

[3. Corrélation 6](#_Toc149744204)

## Analyse exploratoire des données

L’analyse des données se trouve dans le rapport html situé à la racine du projet : report.html.

On a choisi d’établir un rapport automatisé à cause du délai très court pour réaliser le projet.

### Valeurs manquantes

On remarque qu’il n’y a qu’une seule colonne où des valeurs sont manquantes, c’est la colonne « Arrival Delay in Minutes »

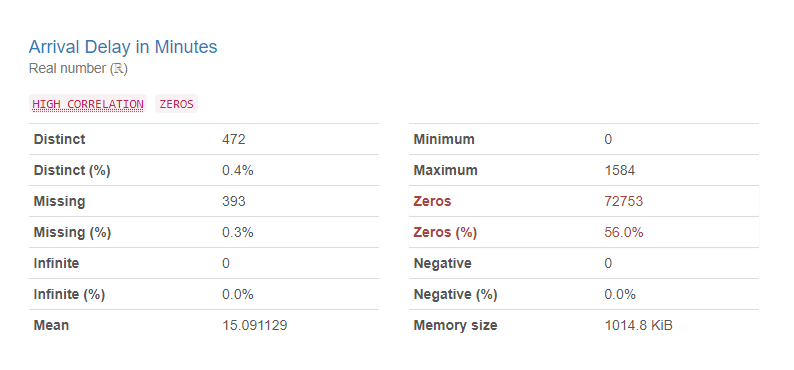


Figure Retard à l'arrivée en minutes.

On voit qu’il y a 0.3% de valeurs manquantes et que la moyenne est d’environ 15 minutes.

D’après la figure 2, plus des trois quarts des valeurs sont comprises entre 0 et 15.

On va donc garder cette colonne et remplacer les valeurs manquantes par la moyenne car cela n’influencera que très faiblement les résultats.

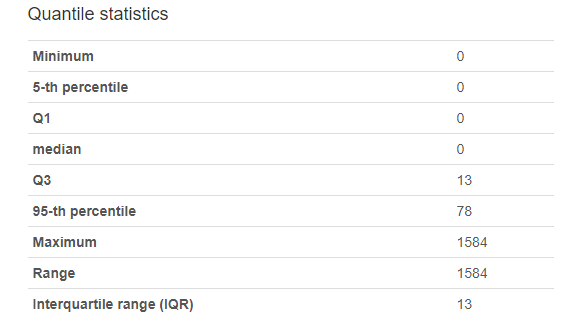
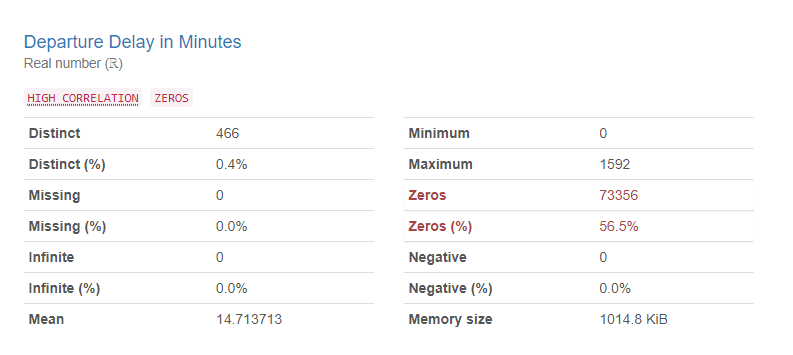


Figure statistiques quantile pour le retard à l’arrivée.

### Valeurs aberrantes

Il existe deux colonnes où il y des valeurs aberrantes (retard de plus de 15 heures) : « Departure Delay in Minutes (fig. 5) », « Arrival Delay in Minutes (fig. 1) ». On remarque que pour la fig. 5 on a moins de 5 % de valeurs aberrantes et pour la fig. 1 moins de 5 % également, on choisit donc là aussi de les remplacer par la moyenne.

Figure Retard au départ en minutes

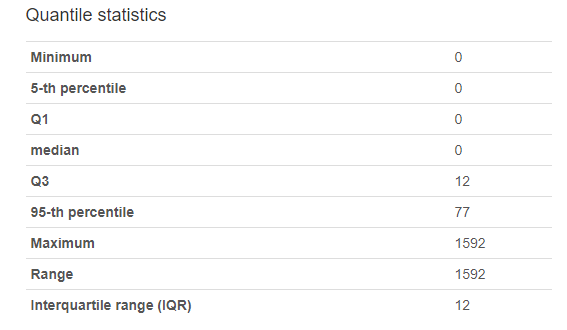


Figure statistiques quantiles pour le retard au départ

### Corrélation

Sur le tableau des corrélations du rapport html, on observe trois fortes corrélation (> 0.5) avec la satisfaction client pour les colonnes : « Inflight wifi service (0.526) », « Online boarding (0.618) » et « class (0.503) ». Elle semble avoir un fort impact sur la satisfaction client, à vérifier.

A l’inverse, on observe de très faibles corrélation (< 0.1) avec la satisfaction client pour les colonnes : « Departure/Arrival time convenient (0.068) », « Departure Delay in Minutes (0.018) », « Arrival Delay in Minutes (0.018) » et « Gender (0.011) ». Ce qui laisse à penser qu’elles n’ont que très peu d’impact voir pas d’impact du tout sur la satisfaction client, à vérifier.

Au vu des exigences du cahier des charges où apparaissent même les colonnes les moins corrélées avec la satisfaction client nous décidons de toutes les garder.