

La base de données `aerien` contient les résultats de sondages pour une compagnie aérienne non identifiée. Le but de cet exercice est de faire une analyse exploratoire afin d'identifier les faits saillants. La base de données contient les variables suivantes, incluant plusieurs niveaux de satisfaction qui sont des échelles de Likert allant de 1 à 5 :

- `id` : identifiant unique du consommateur
- `sexe` : sexe de l'individu, soit homme, soit femme.
- `loyaute_consommateur` : variable binaire indiquant si le consommateur est fidèle à la compagnie, soit oui, soit non.
- `age` : âge (en années)
- `type_deplacement` : motif du déplacement, soit personnel ou pour affaires.
- `classe` : classe du billet, soit affaires, économique (eco) ou économique plus (ecoplus).
- `revenu` : variable catégorielle indiquant le revenu, soit faible (1), moyen (2) ou élevé (3).
- `age` : âge de l'individu (en années).
- `distance_vol` : distance parcourue (en km)
- `service_internet_en_vol` : niveau de satisfaction pour le service internet en vol (0 si pas applicable)
- `temps_arrivee_depart_convenable` : niveau de satisfaction pour l'heure de départ et d'arrivée
- `facilite_reservation_en_ligne` : niveau de satisfaction en rapport à la facilité d'utilisation du service de réservation en ligne
- `localisation_porte` : niveau de satisfaction en lien avec la localisation de la porte d'embarquement
- `nourriture` : niveau de satisfaction en lien avec la nourriture et les breuvages servis en vol.
- `preenregistrement_en_ligne` : niveau de satisfaction en lien avec le pré-enregistrement en ligne
- `confort_siege` : niveau de satisfaction lié au confort du siège
- `divertissement_en_vol` : niveau de satisfaction quant à l'offre de divertissement en vol
- `service_embarquement` : niveau de satisfaction sur l'embarquement du vol
- `service_espace_jambes` : niveau de satisfaction en lien avec l'espace pour les jambes
- `gestion_bagages` : niveau de satisfaction rattaché à la gestion des bagages
- `service_enregistrement` : niveau de satisfaction lié au service d'enregistrement à la porte
- `service_vol` : niveau de satisfaction en lien avec le service durant le vol
- `proprete` : niveau de satisfaction pour la propreté de l'appareil
- `delai_depart_min` : nombre de minutes de retard au décollage
- `delai_arrivee_min` : nombre de minutes de retard à l'arrivée
- `satisfaction` : satisfaction globale, une variable binaire qui vaut 1 si la personne est satisfaite et 0 si la personne est neutre ou insatisfaite
- `test` : variable binaire indiquant si l'observation appartient à l'échantillon test (1) ou à l'échantillon d'entraînement (0).

Effectuez une analyse exploratoire des données `aerien` :

- 1.1 Formulez des questions en lien avec la description de la base de données préliminaires à l'analyse exploratoire.
- 1.2 Examinez la base de données; identifiez le type de variable et leur nature.
- 1.3 Y a-t-il des valeurs manquantes? Si oui, pour quelles variables? Serait-il logique de les imputer par leur moyenne, dans le cas présent?
- 1.4 Calculez les statistiques descriptives pour les variables continues et produisez des tableaux de fréquence ou de contingence pour les variables catégorielles. Que remarquez-vous?
- 1.5 Représentez graphiquement la distribution des variables de la base de données, globalement et selon les différents niveaux de satisfaction.
- 1.6 Résumez les faits saillants de votre analyse exploratoire en cinq à 10 points.