

## Projet de Business Intelligence : Analyse de réservations hôtelières

Le dataset "Hotel Booking Demand" contient des informations détaillées sur les réservations pour deux hôtels : un hôtel urbain (City Hotel) et un hôtel de villégiature (Resort Hotel). Ce jeu de données comprend environ 119 390 observations de réservations effectuées entre juillet 2015 et août 2017, avec 32 variables différentes, notamment :

Colonne	Signification
<b>hotel</b>	Type d'hôtel (Resort Hotel ou City Hotel). Variable catégorique indiquant l'établissement concerné par la réservation.
<b>is_canceled</b>	Indicateur d'annulation. 1 si la réservation a été annulée, 0 si elle a été maintenue. Variable binaire permettant d'identifier les réservations qui ne se sont pas concrétisées.
<b>lead_time</b>	Délai entre la date de réservation et la date d'arrivée exprimé en nombre de jours. Indique l'anticipation avec laquelle les clients effectuent leurs réservations.
<b>arrival_date_year</b>	L'année civile de la date d'arrivée.
<b>arrival_date_month</b>	Le mois de l'année pour la date d'arrivée (January à December)
<b>arrival_date_week_number</b>	Le numéro de la semaine dans l'année pour la date d'arrivée (1 à 53).
<b>arrival_date_day_of_month</b>	Le jour du mois pour la date d'arrivée.
<b>stays_in_weekend_nights</b>	Le nombre de nuits réservées qui tombent un vendredi ou un samedi.
<b>stays_in_week_nights</b>	Le nombre de nuits réservées qui tombent du dimanche au jeudi
<b>adults</b>	Le nombre d'adultes inclus dans la réservation
<b>children</b>	Le nombre d'enfants inclus dans la réservation
<b>babies</b>	Le nombre de bébés inclus dans la réservation
<b>meal</b>	Type de repas réservé : <ul style="list-style-type: none"><li>• BB (Bed &amp; Breakfast)</li><li>• HB (Half Board - petit-déjeuner et un repas)</li><li>• FB (Full Board - tous les repas)</li><li>• SC (Self Catering - sans repas)</li><li>• Undefined (non défini)</li></ul>
<b>country</b>	Pays d'origine du client. Code pays à trois lettres selon la norme ISO 3155-3:2013. Indique la nationalité du client.

<b>market_segment</b>	<p>Segment de marché</p> <ul style="list-style-type: none"> <li>Indique le segment de marché auquel appartient la réservation : <ul style="list-style-type: none"> <li>TA (Travel Agents/Tour Operators)</li> <li>TO (Tour Operators)</li> <li>Direct (réservation directe)</li> <li>Corporate (entreprises)</li> <li>GDS (Global Distribution System)</li> <li>Offline TA/TO (agences de voyage traditionnelles)</li> <li>Complementary (gratuit/complémentaire)</li> <li>Aviation</li> <li>Groups</li> </ul> </li> </ul>
<b>distribution_channel</b>	<p>Canal de distribution. Indique le canal par lequel la réservation a été effectuée :</p> <ul style="list-style-type: none"> <li>Direct</li> <li>Corporate</li> <li>TA/TO (Travel Agents/Tour Operators)</li> <li>GDS (Global Distribution System)</li> <li>Undefined</li> </ul>
<b>is_repeated_guest</b>	<p>Client régulier. 1 si le client a déjà séjourné à l'hôtel auparavant, 0 sinon. Indicateur de fidélité client.</p>
<b>previous_cancellations</b>	<p>Annulations précédentes. Nombre de réservations précédentes que le client a annulées avant cette réservation.</p>
<b>previous_bookings_not_canceled</b>	<p>Réservations précédentes non annulées. Nombre de réservations précédentes que le client n'a pas annulées avant cette réservation.</p>
<b>reserved_room_type</b>	<p>Type de chambre réservée. Code identifiant le type de chambre réservée (A à H).</p>
<b>assigned_room_type</b>	<p>Type de chambre attribuée</p> <ul style="list-style-type: none"> <li>Code identifiant le type de chambre effectivement attribuée (A à H)</li> <li>Peut différer du type réservé en cas de sur-réservation ou upgrade</li> </ul>

<b>booking_changes</b>	Modifications de la réservation. Nombre de modifications apportées à la réservation depuis sa création jusqu'à l'enregistrement.
<b>deposit_type</b>	Type de dépôt <ul style="list-style-type: none"> <li>• No Deposit (aucun dépôt)</li> <li>• Non Refund (dépôt non remboursable)</li> <li>• Refundable (dépôt remboursable)</li> </ul>
<b>agent</b>	Agent <ul style="list-style-type: none"> <li>• ID de l'agence de voyage qui a effectué la réservation</li> <li>• NULL si non applicable</li> </ul>
<b>company</b>	Entreprise <ul style="list-style-type: none"> <li>• ID de l'entreprise/entité qui a effectué la réservation ou est responsable du paiement</li> <li>• NULL si non applicable</li> </ul>
<b>days_in_waiting_list</b>	Nombre de jours où la réservation était en liste d'attente avant d'être confirmée.
<b>customer_type</b>	Type de client <ul style="list-style-type: none"> <li>• Contract (contrat avec allotement ou autre type de contrat)</li> <li>• Group (réservation de groupe)</li> <li>• Transient (client de passage sans contrat)</li> <li>• Transient-party (client de passage avec contrat)</li> </ul>
<b>adr</b>	Average Daily Rate : Tarif journalier moyen calculé en divisant la somme de toutes les transactions d'hébergement par le nombre total de nuits de séjour
<b>required_car_parking_spaces</b>	Nombre de places de stationnement requises par le client
<b>total_of_special_requests</b>	Nombre de demandes spéciales faites par le client (lit supplémentaire, berceau, etc.)
<b>reservation_status</b>	Statut de la réservation

	<ul style="list-style-type: none"> <li>• Canceled (annulée)</li> <li>• Check-Out (client a quitté l'hôtel)</li> <li>• No-Show (client ne s'est pas présenté)</li> </ul>
<b>reservation_status_date</b>	Date du statut de réservation <ul style="list-style-type: none"> <li>• Date à laquelle le dernier statut de réservation a été défini</li> <li>• Cela correspond à la date d'annulation, de check-out ou de constatation de no-show</li> </ul>

Ce dataset est particulièrement intéressant car il contient des données réelles (anonymisées) permettant d'analyser les comportements de réservation et d'annulation dans le secteur hôtelier.

## Objectifs du projet

1. **Conception d'un data warehouse** : Concevoir une architecture efficace pour stocker et organiser les données hôtelières selon un modèle dimensionnel.
2. **Mise en place d'un processus ETL** : Développer un pipeline pour extraire, transformer et charger les données du dataset vers le data warehouse.
3. **Création de tableaux de bord** : Concevoir des visualisations interactives pour analyser les KPIs clés et présenter les insights.
4. **Application de techniques de machine learning** : Utiliser les données du Data warehouse pour développer un modèle prédictif répondant à une problématique business.
5. **Documentation et présentation** : Documenter l'ensemble du processus et présenter les résultats de manière professionnelle.

## KPIs à analyser (proposition à enrichir)

### Performance des réservations

- Taux d'occupation par type d'hôtel, mois et saison
- Revenu moyen par chambre disponible (RevPAR)
- Durée moyenne de séjour par segment de marché
- Répartition des réservations par canal de distribution

### Comportement client

- Taux d'annulation global et par segment
- Délai moyen entre réservation et arrivée
- Taux de fidélisation des clients
- Origine géographique des clients (top 10 pays)

## Efficacité opérationnelle

- Écart entre type de chambre demandée et attribuée
- Nombre moyen de demandes spéciales par réservation
- Variations saisonnières de la demande
- Prévion du taux d'occupation à court terme

## Travail à réaliser

### Phase 1: Analyse et conception (1 semaine)

- Étudier le dataset et comprendre les relations entre les variables
- Concevoir le modèle dimensionnel (schéma en étoile ou flocon ou constellation)
- Définir les dimensions et les faits

### Phase 2: ETL et implémentation du data warehouse (2 semaines)

- Mettre en place l'environnement technique (SQL Server, PostgreSQL, ou autre)
- Développer les scripts d'extraction et de transformation des données
- Implémenter le processus de chargement vers le data warehouse
- Valider l'intégrité des données chargées

### Phase 3: Analyse et visualisation (1 semaine)

- Développer un tableau de bord interactifs (Power BI, Tableau, ou autre)
- Créer des visualisations pour quelques KPIs

### Phase 4: Machine Learning (2 semaines)

- Formuler une problématique business (par exemple : prédiction des annulations)
- Préparer les données pour le machine learning
- Sélectionner les algorithmes à tester
- Évaluer les performances des modèles
- Choisir le meilleur modèle

## Dates importantes

**30/03/2025**

- Un mini rapport contenant **1) la définition des besoins d'affaires** (en suivant la Table 2.1 page 7 du Manifeste BI), **2) l'inspection des données** (section 3.2, page 14 du Manifeste BI), **3) la conception du modèle physique** (en suivant la section 3.3, page 15 du Manifeste BI).

**05/05/2025**

- Un mini rapport détaillant : **1) la définition d'un objectif d'apprentissage** précis (prédiction/description), **2) les différents modèles de machine learning testés** sur les

données du data warehouse, **3) l'approche de validation** utilisée pour évaluer les performances, et **4) la justification du choix du meilleur modèle** selon les résultats obtenus.

**06/05/2025**

- Soutenance technique.