



---

## EDA HOTEL BOOKING

---

Rapport Final

10 DÉCEMBRE 2025

8PRO408 – OUTILS DE PROGRAMMATION POUR LA SCIENCE DES DONNÉES

Remis à :Habba's Ngodjou Doukaga

## Table des matières

<b>Rapport d'Analyse Exploratoire : Demande Hôtelière.....</b>	<b>2</b>
<b>1. Objectif .....</b>	<b>2</b>
<b>2. Méthodologie et Nettoyage.....</b>	<b>2</b>
<b>3. Résultats Clés et Observations.....</b>	<b>2</b>
A. City Hôtel vs Resort Hôtel : Comparaison Globale.....	2
B. Analyse des Prix (ADR) .....	2
C. Saisonnalité et Anticipation .....	2
D. Profils Clients .....	3
<b>4. Conclusion et Recommandations .....</b>	<b>3</b>

# Rapport d'Analyse Exploratoire : Demande Hôtelière

## 1. Objectif

Ce projet vise à analyser les comportements de réservation du dataset hotel\_bookings.csv (119 390 entrées) afin de comparer les performances des City Hotels et Resort Hotels. L'analyse se concentre sur les taux d'annulation, la tarification (ADR), la saisonnalité et les profils clients.

## 2. Méthodologie et Nettoyage

Données : Traitement des variables clés (hotel, adr, is\_canceled, dates, market\_segment).

Nettoyage : Imputation des valeurs manquantes (ex: children=0), conversion des dates, et exclusion des outliers extrêmes (ex: ADR > 5000€).

Outils : Python (Pandas, Seaborn) pour l'analyse statique et Streamlit/Plotly pour l'exploration interactive.

## 3. Résultats Clés et Observations

### A. City Hôtel vs Resort Hôtel : Comparaison Globale

Nous observons une différence marquée entre les deux types d'établissements :

Volume : Le City Hôtel capte la majorité des réservations (~66% du dataset).

Annulations : Le taux d'annulation est significativement plus élevé pour le City Hôtel (~41%) par rapport au Resort Hôtel (~27%). Cela suggère une volatilité plus forte de la clientèle urbaine ou d'affaires.

### B. Analyse des Prix (ADR)

Moyenne : L'ADR moyen global est d'environ 101€.

Tendance : Le Resort Hôtel montre une variabilité de prix beaucoup plus forte, avec des pics très élevés en été (juillet/août), dépassant souvent le City Hôtel, alors qu'il est moins cher le reste de l'année.

Outliers : Des valeurs aberrantes ont été identifiées (ex: transaction à 5400€) et filtrées pour ne pas fausser les moyennes.

### C. Saisonnalité et Anticipation

Périodes de pointe : Le mois d'août enregistre le plus fort volume d'arrivées pour les deux hôtels, suivi de juillet. Les mois d'hiver (janvier/février) sont les plus calmes.

Lead Time : Les clients réservent en moyenne très à l'avance, mais on note une corrélation : plus le délai de réservation (lead\_time) est long, plus la probabilité d'annulation augmente.

## D. Profils Clients

Les couples (2 adultes) représentent le segment dominant.

Les familles (avec enfants/bébés) privilégient le Resort Hôtel, notamment durant la période estivale.

## 4. Conclusion et Recommandations

L'analyse met en évidence que le City Hôtel repose sur un volume élevé mais souffre de nombreuses annulations, tandis que le Resort Hôtel est très saisonnier. Pistes futures :

- Mettre en place des politiques d'annulation plus strictes pour le City Hôtel (dépôts non remboursables) afin de réduire le taux de 41%.
- Dynamiser la basse saison du Resort Hôtel par des offres promotionnelles, l'écart de prix avec l'été étant très marqué.
- Pour une modélisation prédictive (Machine Learning), la variable lead\_time et le deposit\_type seront des "features" déterminantes.

### ***Liens :***

Dépôt Github :

<https://github.com/pollarouch2/DataScienceToolsFinalProject>

Application Streamlit :

<https://datasciencetoolsfinalproject-evmvo8i8n2aqojyumr4b5s.streamlit.app>