**Problem statement**

City Hotel and Resort Hotel aimed to identify the key factors contributing to booking cancellations to enhance their service quality and customer satisfaction.

**Dataset Overview**

The dataset comprises **119,390** bookings recorded between **July 2015 and August 2017**, including both completed stays and cancellations. The data covers two hotel types:

* **City Hotel** (79,330 bookings – 60% of total)
* **Resort Hotel** (40,060 bookings – 40% of total)

**Data Preparation**

**1. Column Selection:**  
From the original 32 columns, **13 relevant fields** were retained for analysis:

* hotel (Hotel type)
* is\_canceled (Cancellation status)
* arrival\_date\_year, arrival\_date\_month (Booking dates)
* adults, children, babies (Guest details)
* country (Customer origin)
* reserved\_room\_type, assigned\_room\_type (Room allocation)
* adr (Average daily rate)
* reservation\_status, reservation\_status\_date (Booking status)

**2. Handling Missing Data:**  
Columns like agent and company had significant null values and were removed.

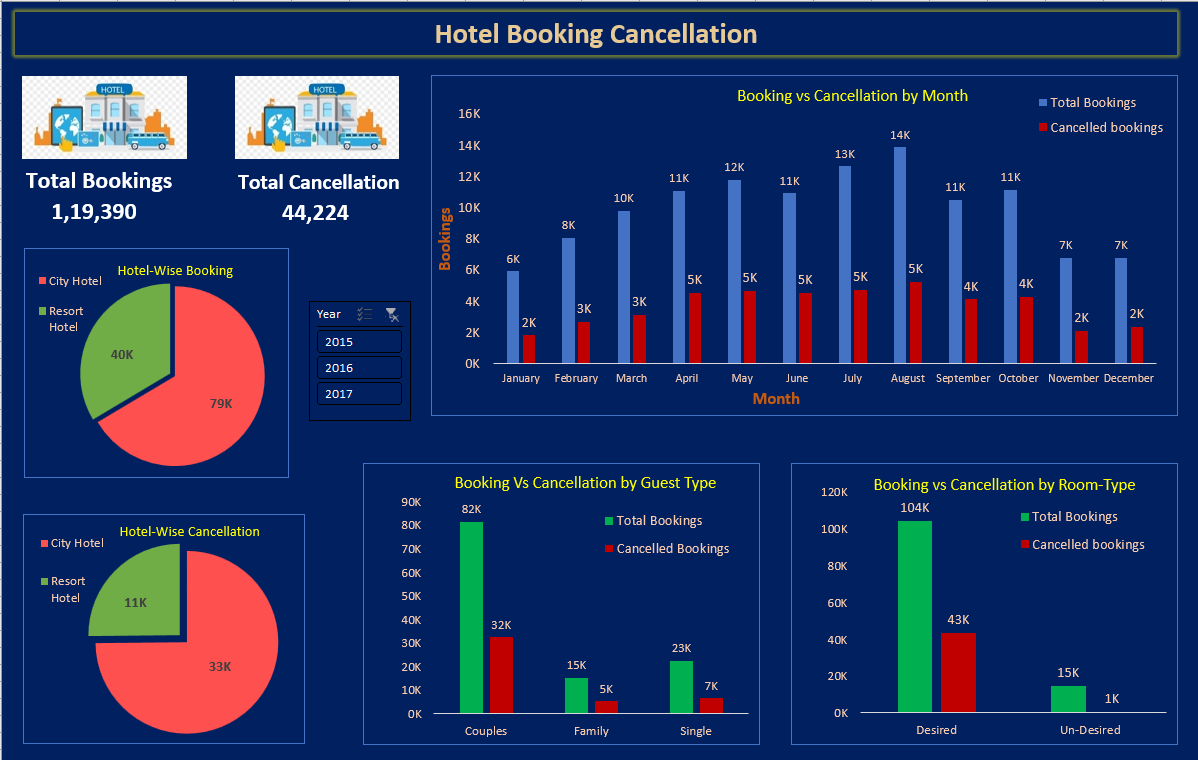
**Feature Engineering**

Three new columns were added for deeper insights:

1. **Room Preference Status:**
   * room\_status = IF(reserved\_room\_type = assigned\_room\_type, "Desired", "Un-Desired")
2. **Guest Category:**
   * guest\_type = IF(adults=2 AND children=0 AND babies=0, "Couples", IF(adults=1 AND children=0 AND babies=0, "Single", "Family"))
3. **Abbreviated Month:**
   * months = LEFT(arrival\_date\_month, 3)

**Key Findings**

1. **Cancellation Rate:**
   * **44,224 bookings (37%)** were cancelled.
   * **City Hotel** accounted for **75% of cancellations (33,101)**, while **Resort Hotel** had **11,123 cancellations (25%)**.
2. **Cancellation by Guest Type:**
   * **Couples** (40% cancellation rate): 32,424 out of 81,560 bookings.
   * **Families** (34% cancellation rate): 5,245 out of 15,253 bookings.
   * **Singles** (29% cancellation rate): 6,555 out of 22,577 bookings.
3. **Room Preference Impact:**
   * **42% of cancellations** occurred even when guests received their **desired room type**, suggesting room allocation has minimal influence on cancellations.
   * Only **5% of cancellations** were due to **unfulfilled room preferences**.
4. **Monthly Trends:**
   * **August** had the **highest bookings and cancellations** in absolute numbers.
   * However, **April, May, and June** showed the **highest cancellation rates (~41%)**, indicating seasonal patterns.



**Conclusion**

* **City Hotel faces significantly higher cancellations** than Resort Hotel.
* **Couples are the most frequent cancellers**, despite being the largest booking segment.
* **Room preference does not strongly impact cancellation decisions**, implying other factors (e.g., pricing, policies, or external events) may play a bigger role.
* **Seasonal trends** (especially mid-year) show higher cancellation risks, suggesting the need for targeted retention strategies during peak months.

This analysis helps hotels refine their booking policies, improve customer engagement, and reduce revenue loss due to cancellations.