# Unternehmensbeschreibung

Das Unternehmen ist eine führende Hotelkette und hat das Ziel eine prädiktive Analyse zur Vorhersage von Stornierungen zu entwickeln. Um seine Effizienz und Rentabilität weiter zu steigern, plant das Unternehmen, auf Basis von Hotelbuchungsdaten ein zuverlässiges Vorhersagemodell für Stornierungen zu implementieren.

# Problem und Fragestellung

Das Hotelgewerbe sieht sich mit der Herausforderung konfrontiert, leere Zimmer aufgrund von Stornierungen zu bewältigen. Insbesondere die Möglichkeit für Kunden, ihre Buchungen kostenlos zu stornieren, trägt zu dieser Problematik bei. Die damit verbundene Unsicherheit in der Kapazitätsplanung führt zu finanziellen Verlusten und kann sich negativ auf die Rentabilität und das Gästeerlebnis auswirken. In Anbetracht dieser Situation lautet die Forschungsfrage: Wie kann das Unternehmen effektiver auf Stornierungen reagieren und sie vorhersehen, um die Betriebsplanung zu optimieren? Durch die Beantwortung dieser Frage können Strategien und Maßnahmen entwickelt werden, um die Auswirkungen von Stornierungen zu minimieren und eine effiziente Ressourcenverwaltung zu gewährleisten.

# Datensatz und ML- Modell

Die Daten, welche zur Entwicklung des Modells benutzt, werden kommen aus dem Dataset [Hotel Reservations Dataset](https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset), auf Kaggle. Das Dataset ist aus 19 Features und 36.000 Datensätzen aufgebaut. Zu jeder stornierten Buchung kommen 2 nicht stornierte Buchungen. Also ist der Datensatz relativ ausbalanciert.  
  
**Feature 1:** Booking\_ID  
dient nur der Zuordnung zu einem Datensatz und ist nicht relevant für das Modell

**Feature 2-18:** Nominal: Booking\_ID, type\_of\_meal\_plan, required\_car\_parking\_space, room\_type\_reserved, market\_segment\_type, repeated\_guest, booking\_status

Kardinal: no\_of\_adults, no\_of\_children, no\_of\_weekend\_nights, no\_of\_week\_nights, lead\_time, no\_of\_previous\_cancellations, no\_of\_previous\_bookings\_not\_canceled, avg\_price\_per\_room, no\_of\_special\_requests

Ordinal: arrival\_year, arrival\_month, arrival\_date

**Feature 19:** Booking\_status – dieses Feature ist die Zielvariable und wird in binär angegeben.

# Methodik

Für die Vorhersage von Stornierungen im Hotelgewerbe bieten sich verschiedene Algorithmen an, darunter Entscheidungsbäume, Random Forest, XGBoost und K-Nearest Neighbors (KNN).

Entscheidungsbäume ermöglichen die Erstellung von Entscheidungsregeln basierend auf den Merkmalen der Buchungsdaten. Durch die Kombination mehrerer Entscheidungsbäume im Random Forest-Algorithmus können robuste und genaue Vorhersagen erzielt werden. Random Forest berücksichtigt die Vorhersagen aller Bäume und aggregiert sie zu einem kohärenten Ergebnis.

XGBoost ist eine Weiterentwicklung des Gradient Boosting-Algorithmus, der speziell auf Effizienz und Genauigkeit ausgelegt ist. Es verwendet Entscheidungsbäume und optimiert die Modellvorhersagen durch das Minimieren einer Kostenfunktion. XGBoost bietet erweiterte Funktionen wie Regularisierung und paralleles Training, um präzise Vorhersagen zu erzielen.

K-Nearest Neighbors (KNN) basiert auf der Idee, dass ähnliche Beispiele tendenziell die gleiche Klassenzugehörigkeit haben. KNN sucht nach den k nächsten Nachbarn eines Datenpunkts und bestimmt die Mehrheitsklasse dieser Nachbarn als Vorhersage. KNN ist besonders nützlich, um Muster und Ähnlichkeiten in den Buchungsdaten zu erkennen und Vorhersagen zu treffen.

# Ausblick

Die erfolgreiche Implementierung eines prädiktiven Modells zur Vorhersage von Stornierungen im Hotelgewerbe bietet vielfältige Auswirkungen und Möglichkeiten für das Unternehmen. Durch die präzise Vorhersage von Stornierungen kann das Unternehmen seine Betriebsplanung optimieren, die Kapazitätsauslastung verbessern und finanzielle Verluste durch leere Zimmer minimieren.