

Name : ANUJAY KALBHOR

ROLL NO : 250240325012

Uncovering Patterns in Hotel Booking Data for Operational Efficiency and Revenue Growth

1. Problem Statement

How can a hotel enhance operational efficiency and maximize revenue using insights from historical booking data?

2. Project Overview

This project conducts a comprehensive exploratory data analysis (EDA) on hotel booking records to uncover trends in customer behavior, booking channels, pricing strategies, and seasonal dynamics. The goal is to extract actionable insights that inform data-driven decision-making across hotel operations and revenue management.

3. Core Objectives

- Analyze customer behavior across channels and over time.
 - Identify key drivers of revenue (ADR), cancellations, and room upgrades.
 - Examine seasonal and regional booking trends.
 - Understand customer segmentation and associated booking characteristics.
 - Validate business assumptions through statistical testing.
-

4. Scope of Work

- **Data Cleaning:** Address missing values, remove duplicates, and manage outliers.

- **Feature Engineering:** Create insightful variables (e.g., `total_guests`, `stay_duration`).
 - **Exploratory Analysis:** Perform univariate, bivariate, and multivariate visualizations.
 - **Statistical Testing:** Use T-tests, ANOVA, and Chi-Square tests to validate patterns.
 - **Business Insight Extraction:** Address 19 key business questions.
-

5. Data Preparation

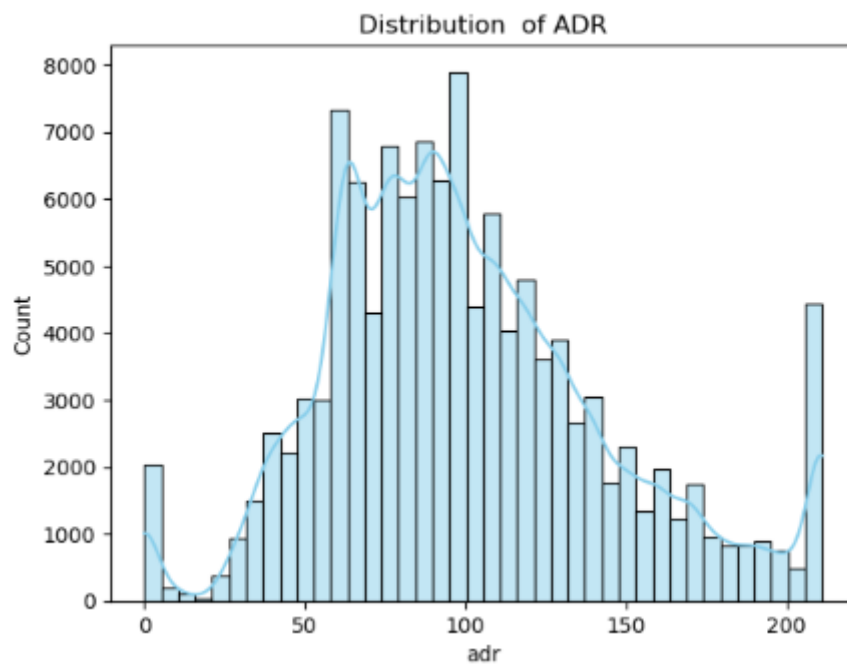
- **Cleaning:** Removed nulls in columns such as `children`, `agent`, `company`, `country`.
 - **De-duplication:** All duplicate entries were eliminated.
 - **Feature Creation:**
 - `total_stay_nights` = `stays_in_weekend_nights` + `stays_in_week_nights`
 - `total_guests` = `adults` + `children` + `babies`
 - `stay_duration` & `booking_date` were derived.
 - **Data Formatting:** Converted object types to categories and standardized date fields.
 - **Outlier Handling:** Applied quantile-based filtering to remove extreme values.
-

6. Exploratory Data Analysis

Univariate Highlights

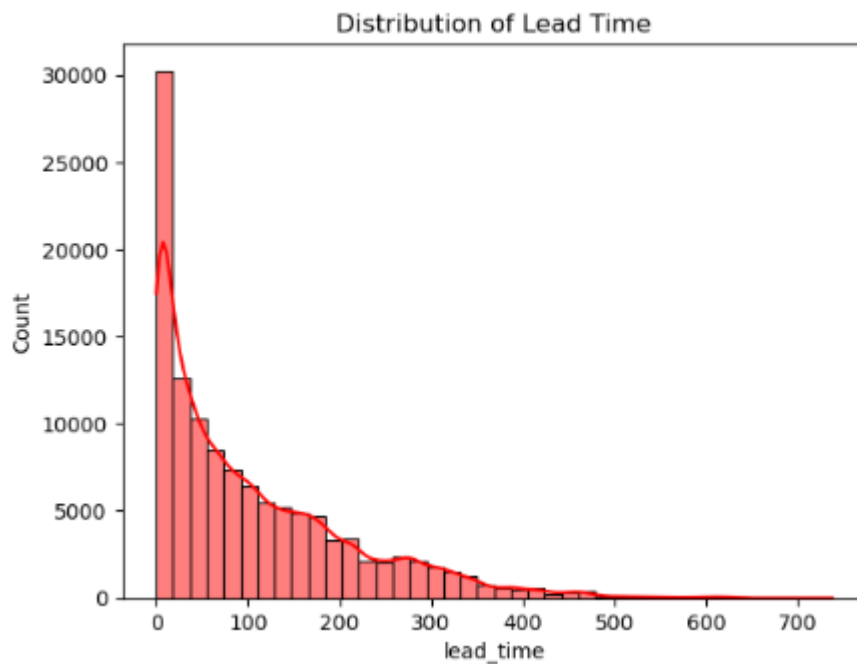
- **ADR (Average Daily Rate):** Majority of values lie between €75–€100.

```
Text(0.5, 1.0, 'Distribution of ADR')
```



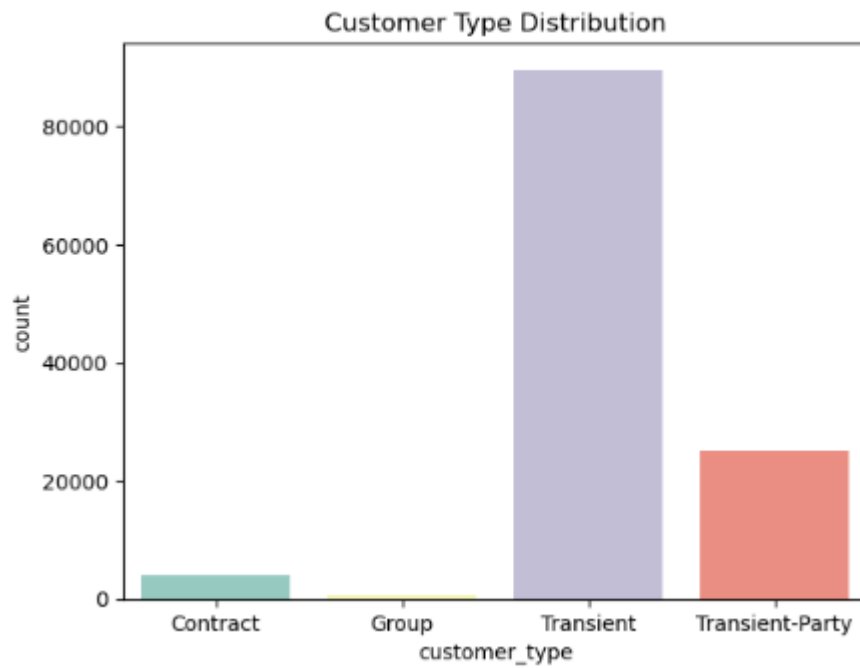
Lead Time: High concentration of bookings within 0–10 days.

```
Text(0.5, 1.0, 'Distribution of Lead Time')
```

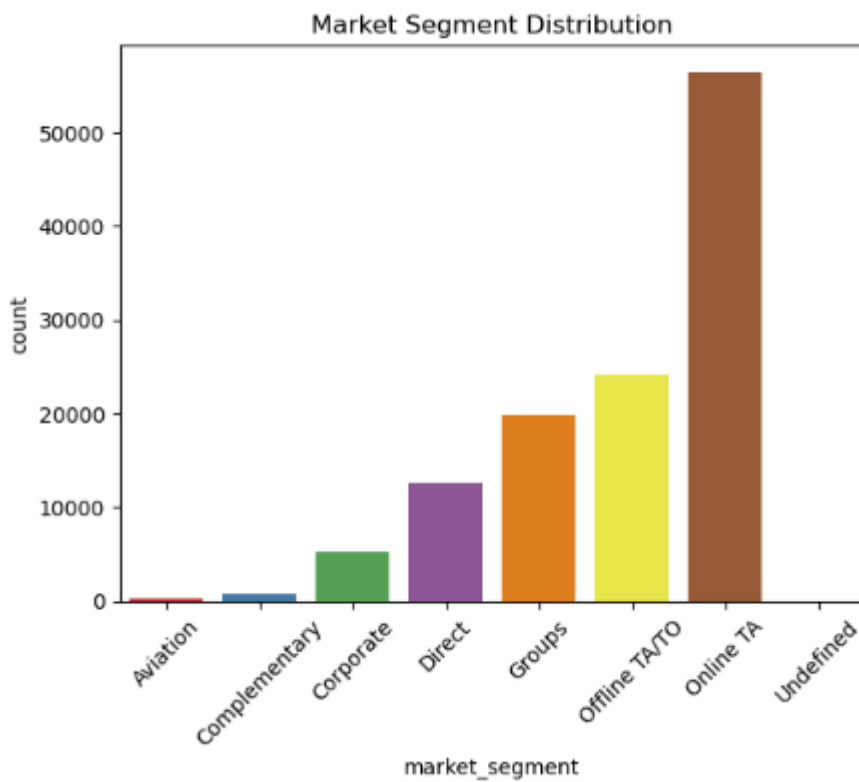


- **Customer Type:** Predominantly transient customers.

```
Text(0.5, 1.0, 'Customer Type Distribution')
```

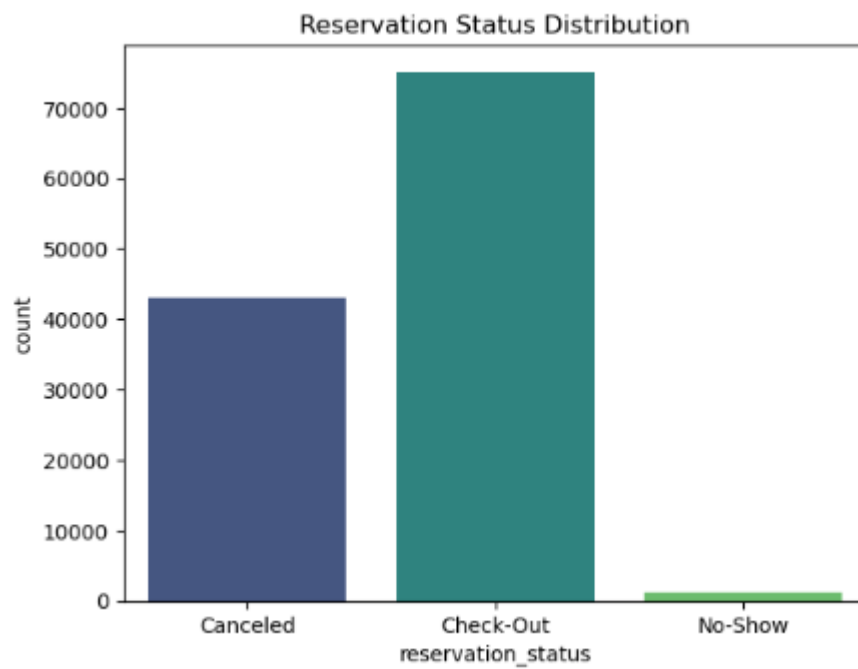


- **Market Segment:** Online travel agencies (OTAs) are the leading channel.



Reservation Status: Most are either check-outs or cancellations.

```
Text(0.5, 1.0, 'Reservation Status Distribution')
```

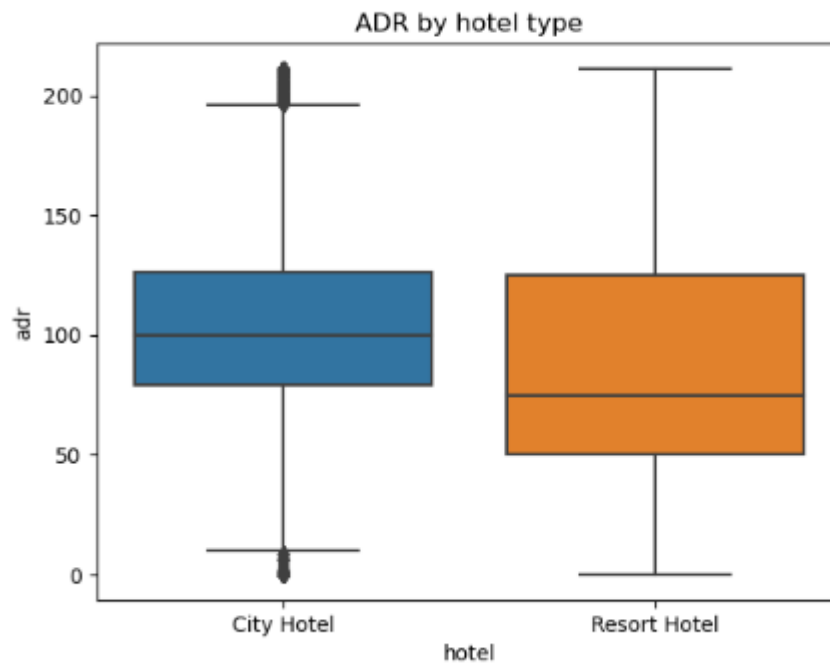


Bivariate Insights

- ADR by Hotel Type :

- **Insight:** ADR is generally higher in city hotels compared to resort hotels.

```
Text(0.5, 1.0, 'ADR by hotel type')
```



- **ADR by Market Segment :**

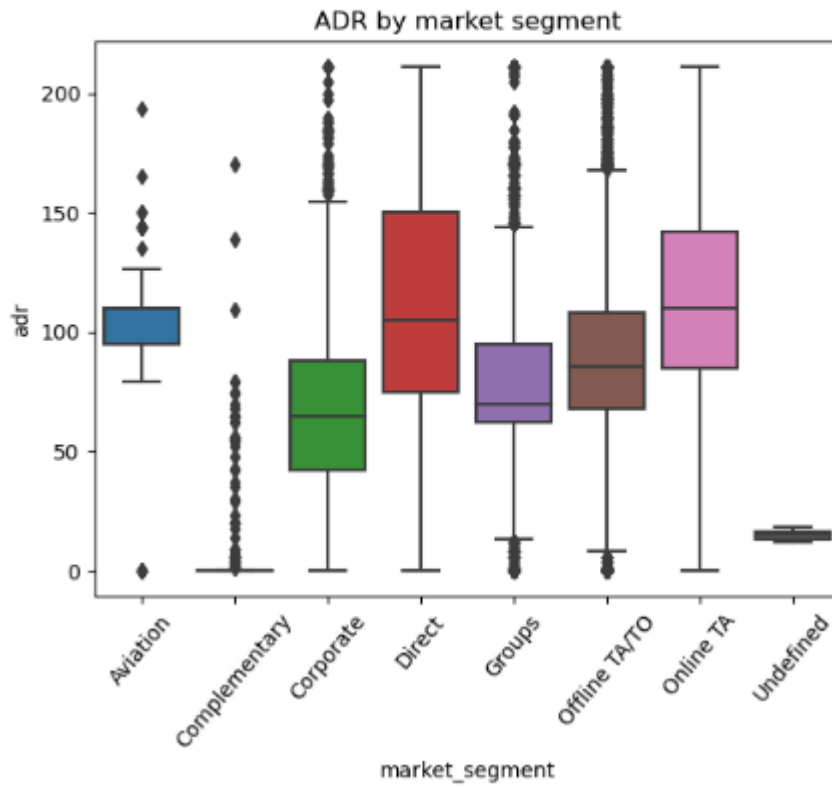
Insight: Corporate and offline TA/TO segments tend to show higher ADR than online TA.

Transient OTA guests yield highest ADR (claimed).

```

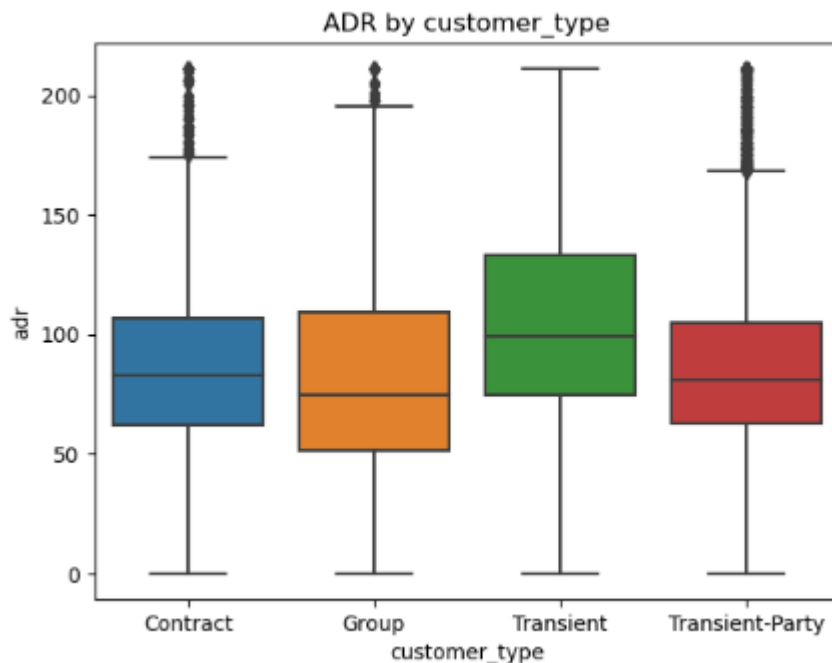
(array([0, 1, 2, 3, 4, 5, 6, 7]),
 [Text(0, 0, 'Aviation'),
  Text(1, 0, 'Complementary'),
  Text(2, 0, 'Corporate'),
  Text(3, 0, 'Direct'),
  Text(4, 0, 'Groups'),
  Text(5, 0, 'Offline TA/TO'),
  Text(6, 0, 'Online TA'),
  Text(7, 0, 'Undefined')])

```



- Customer Type by ADR : Transient OTA guests yield highest ADR (claimed).

```
: Text(0.5, 1.0, 'ADR by customer_type')
```



Multivariate Insights :

1. ADR by Hotel and Customer Type :

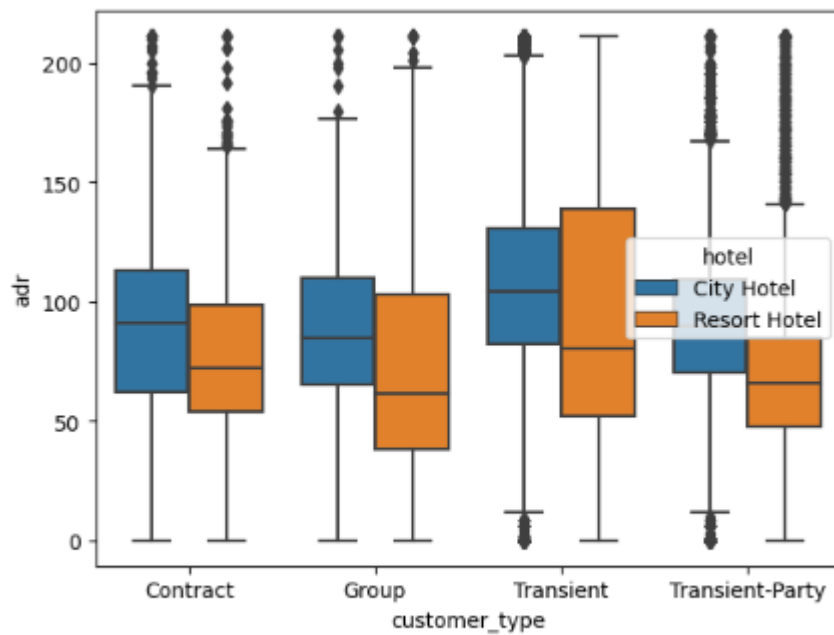
City Hotels tend to command higher ADRs across all customer types compared to Resort Hotels.

Among customer types, Transient customers have the highest ADR in both hotel types, suggesting they are the most valuable segment for pricing.

Contract customers consistently show lower ADRs, especially in Resort Hotels, indicating bulk or corporate pricing strategies.

Group bookings result in moderate ADRs, with some overlap across both hotel types but typically lower than Transients.

<Figure size 1500x2000 with 0 Axes>



<Figure size 1500x2000 with 0 Axes>

2. ADR by Hotel and market_segment :

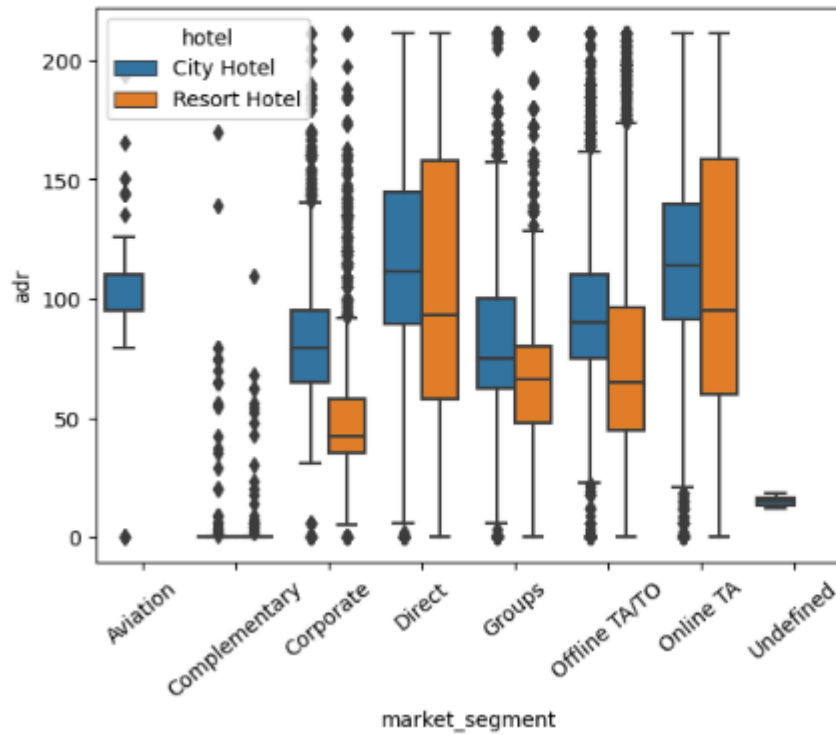
In City Hotels, Corporate and Offline TA/TO segments show significantly higher ADRs than Online TA or Direct bookings.

In Resort Hotels, Offline TA/TO leads ADR, while other segments remain relatively flat or lower.

Online TA is the most frequent channel but often yields lower ADRs, particularly in Resort Hotels.

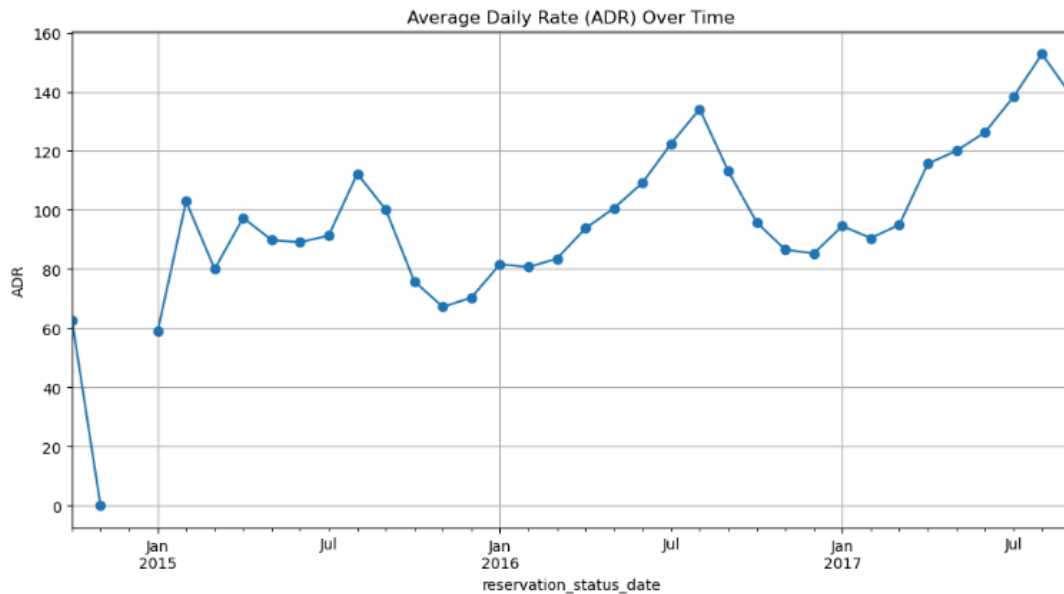
Direct bookings show moderate ADRs, better in City Hotels than Resort ones.

```
(array([0, 1, 2, 3, 4, 5, 6, 7]),
[Text(0, 0, 'Aviation'),
Text(1, 0, 'Complementary'),
Text(2, 0, 'Corporate'),
Text(3, 0, 'Direct'),
Text(4, 0, 'Groups'),
Text(5, 0, 'Offline TA/TO'),
Text(6, 0, 'Online TA'),
Text(7, 0, 'Undefined')])
```



Time-Series Observations

- Peak bookings during summer (June–August).
- Lowest bookings in early year and December.
- Seasonal patterns vary across countries.



7. Correlation Analysis

- **Top Positive Correlations with ADR:**
 - `children` (+0.30)
 - `total_guests` (+0.28)
 - `special_requests` (+0.26)
- **Cancellations:**
 - Positively correlated with `lead_time` (~+0.19)
 - **Negative correlation:** `special_requests` vs. `is_canceled`

8. Hypothesis Testing :

1] H0: There is no difference in ADR between bookings made through Online TA and Direct channels

T-Test 1

- **Test:** Difference in ADR between TA/TO and Direct channels
- **Result:** Statistically significant
- **Insight:** TA/TO bookings have higher ADR on average than Direct bookings

2] H0: Room upgrades are independent of lead time

T-Test 2

- **Test:** Difference in lead time between upgraded and non-upgraded bookings
- **Result:** Statistically significant
- **Insight:** Longer lead times are associated with a higher likelihood of upgrades

3] H0: Average stay duration does not differ between customer types

ANOVA

- **Test:** Difference in stay durations across customer types
- **Result:** Statistically significant
- **Insight:** Customer types have distinct stay length patterns

9. Summary of Graphical Insights

Visualization

ADR Distribution

Key Insight

Majority ADR values fall between €75–€100.

Lead Time	Last-minute bookings are common.
Customer Type	Transient guests form the bulk of bookings.
Market Segment	OTA dominates; Direct and TA/TO follow.
Seasonality	Summer sees peak demand; year-end drops.
ADR by Segment	Higher for OTA-based bookings.
Lead Time by Type	Contract customers plan further ahead.
Guest Count	1–3 guests per booking is typical.
Country Distribution	PT, UK, FR lead in bookings.
ADR by Country	IRL and GBR guests spend more.
Booking Changes	Linked to fewer cancellations, more special requests.

10. Key Business Questions & Data-Driven Answers

- Top ADR Influencers?**
`total_guests`, `children`, `special_requests`, `lead_time`.
- Do early bookers change more?**
Slight positive trend; not strong.
- Country-based ADR differences?**
Yes — IRL and GBR pay more on average.
- Can room upgrades be predicted?**
Yes — based on lead time and special requests.
- Are room types matched?**
Often mismatched, indicating possible upgrades.
- Common guest profiles?**
Groups of 2–3 guests; mostly from PT and UK.
- Customer type behavior?**
Transients book short stays; Contract customers book early.

8. **Lead time variance by type/country?**
Yes — IRL and GBR show longer lead times.
 9. **Impact of lead time on cancellations?**
Longer lead times slightly increase cancellations.
 10. **Average stay duration?**
Typically 3–5 nights.
 11. **Upgrade frequency?**
Rare but present (~2–5%).
 12. **Loyalty and special requests?**
Repeat guests make more requests.
 13. **Highest revenue segments?**
Transient guests via OTAs.
 14. **ADR predictors?**
`children, total_guests, special_requests`.
 15. **Most profitable customer type?**
Transient.
 16. **Does country or lead time affect ADR?**
Yes — both have a measurable impact.
 17. **Do high-ADR guests make more requests?**
Yes.
 18. **Nationality and behavior differences?**
Yes — varies by ADR, stay length, and lead time.
 19. **Do booking changes reduce cancellations?**
Yes — more changes correlate with lower cancellations.
-

11. Final Conclusions

Key Takeaways

- **High Revenue Guests:** Transients from IRL/GBR with children and special requests.
- **Cancellations:** Linked with long lead times and minimal booking modifications.

- **Special Requests:** Good indicator of loyalty and booking value.
- **Operational Efficiency:** Monitor room assignment mismatches for upgrade tracking.
- **Seasonality Awareness:** Plan for summer surges and winter lulls.
- **Channel Strategy:** Increase margins by optimizing Direct bookings.