HOTEL BOOKING CANCELLATION ANALYSIS

Revenue Optimization & Demand Forecasting

Rohit Hiralal Hajare | Python • SQL • Power BI

BUSINESS PROBLEM

- City Hotel and Resort Hotel are experiencing high booking cancellation rates.
- High cancellations are causing:
  - Revenue loss
  - Low room utilization
  - Poor operational efficiency
- Both hotels struggle to accurately forecast demand due to frequent cancellations.
- Empty rooms lead to wasted capacity and reduced profitability.
- Every 1% reduction in cancellations directly improves profitability.

EXECUTIVE SUMMARY

- Cancellation Rate: **37% → Major Revenue Risk**
- High Risk Drivers: **Long lead time, high ADR, City Hotel, Portugal**
- Opportunity: **Dynamic pricing + prepayment can reduce losses**

GOALS & OBJECTIVES

Strategic Goal	Business Impact
Identify cancellation drivers	Reduce revenue loss
Improve demand predictability	Better pricing & staffing
Optimize customer retention	Higher lifetime value

DATASET OVERVIEW

Source: Hotel Booking Cancellation Dataset (CSV)
Total Records: 118,896 bookings
Total Features: 32 variables
Hotel Types: City Hotel, Resort Hotel
Time Period: Multi-year historical booking data

DATA INCLUDES

- **Booking Details:** Lead time, arrival date, length of stay
- **Customer Info:** Adults, children, customer type, repeat guest
- **Room Details:** Reserved vs assigned room type
- **Pricing Data:** ADR (Average Daily Rate)
- **Sales Channels:** Market segment, distribution channel
- **Cancellation Status:** Cancelled / Not Cancelled

TOOLS USED

- **Python**
  - Data cleaning and preprocessing
  - Exploratory Data Analysis (EDA)
  - Feature engineering
  - Statistical analysis
- **Libraries Used:**
- Pandas
- NumPy
- Matplotlib
- Seaborn

- **SQL**
  - Data extraction and transformation
  - Aggregations and filtering
  - KPI calculations
  - Joining and structuring data
- **Power BI**
  - Interactive dashboards
  - KPI cards & visuals
  - Trend and comparison analysis
  - Business reporting & storytelling

SUMMARY STATISTICS

[4]:
```
#Explore Data
```

[5]:
```
df.head(10)
```

[5]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_ni
0	Resort Hotel	0	342	2015	July	27	1	0	
1	Resort Hotel	0	737	2015	July	27	1	0	
2	Resort Hotel	0	7	2015	July	27	1	0	
3	Resort Hotel	0	13	2015	July	27	1	0	
4	Resort Hotel	0	14	2015	July	27	1	0	
5	Resort Hotel	0	14	2015	July	27	1	0	
6	Resort Hotel	0	0	2015	July	27	1	0	
7	Resort Hotel	0	9	2015	July	27	1	0	
8	Resort Hotel	1	85	2015	July	27	1	0	
9	Resort Hotel	1	75	2015	July	27	1	0	

10 rows × 32 columns

KEY METRICS

KPI	Value	Insight
Total Bookings	**119K**	High booking volume across both hotels
Total Cancellations	**44K**	Large portion of bookings are being cancelled
Valid Cancellations	**43K**	Almost all cancellations are genuine
Cancellation Rate	**0.37 (37%)**	Over one-third of all bookings are cancelled
Total Revenue	**12.12M**	Revenue after cancellations
Average Lead Time	**104.31 days**	Customers book far in advance
High Lead Time %	**0.21 (21%)**	One-fifth of bookings have long lead times
Average ADR	**101.95**	Average daily room price
Repeat Guest %	**0.03 (3%)**	Very low customer loyalty
Average Length of Stay	**3.43 days**	Typical stay is 3–4 nights

KEY BUSINESS TAKEAWAYS

- **37%** cancellation rate is a major revenue risk.
- Long lead times strongly contribute to cancellations.
- Low repeat guest **rate (3%)** shows weak customer retention.
- Revenue can be improved by **reducing last-minute cancellations** and **encouraging loyal customers.**

BOOKING CANCELLATION DISTRIBUTION

```
# Data Analysis and Visualizations
```

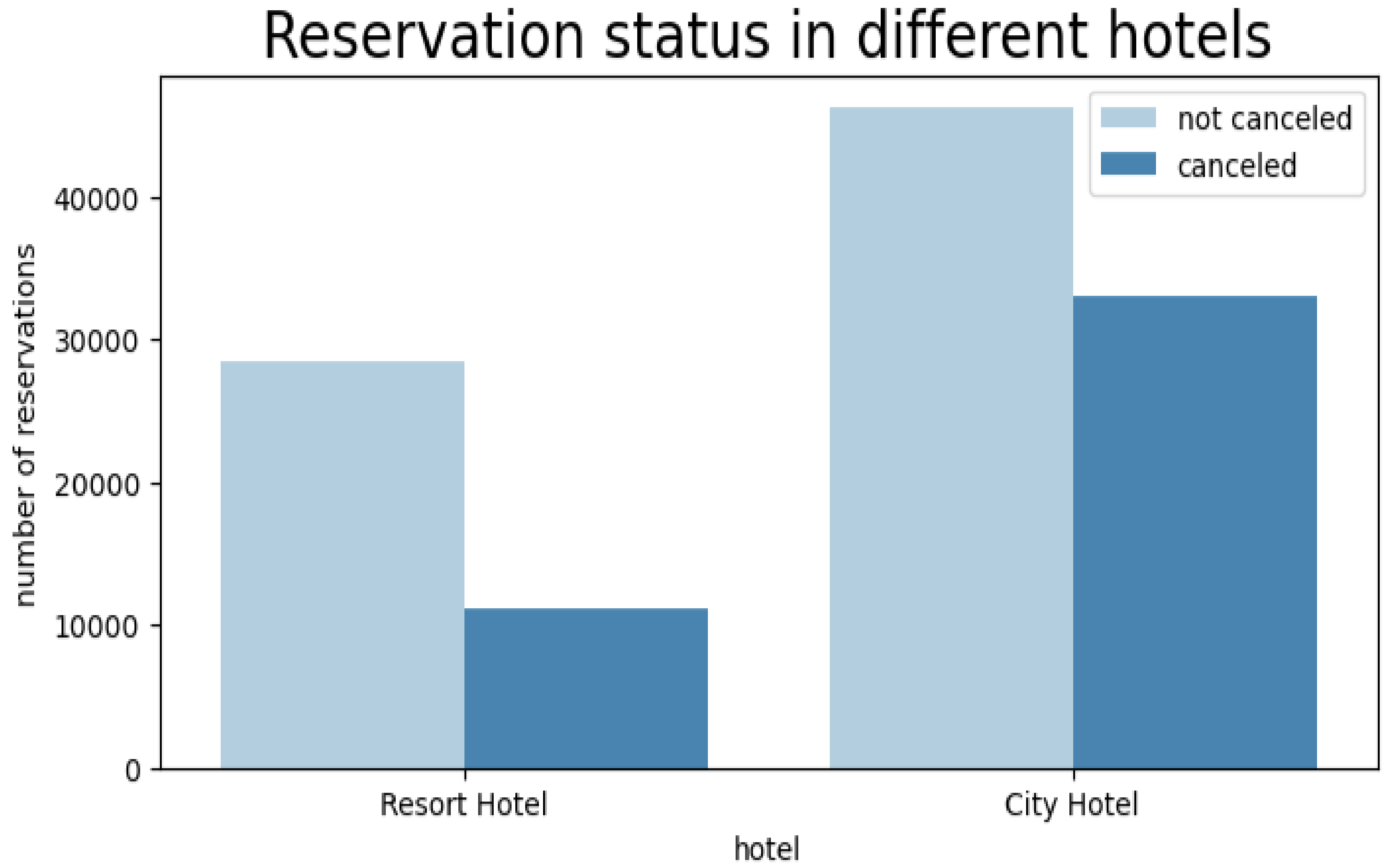
```
cancelled_perc = df['is_canceled'].value_counts(normalize = True)
print(cancelled_perc)

plt.figure(figsize = (5,4))
plt.title('Reservation status count')
plt.bar(['Not canceled','Canceled'],df['is_canceled'].value_counts(), edgecolor = 'k', width = 0.7)
plt.show()
```

```
is_canceled
0    0.628647
1    0.371353
Name: proportion, dtype: float64
```

BUSINESS IMPACT:

- **62.86%** of bookings are not cancelled.
- **37.14%** of bookings are cancelled, which is a very high rate.
- Nearly **1 out of every 3** reservations gets cancelled.
- This high cancellation volume directly impacts:
  - Revenue stability
  - Room utilization
  - Demand forecasting
- Reducing even a small portion of cancellations can significantly **increase revenue and occupancy.**

CANCELLATION COMPARISON: CITY HOTEL VS RESORT HOTEL

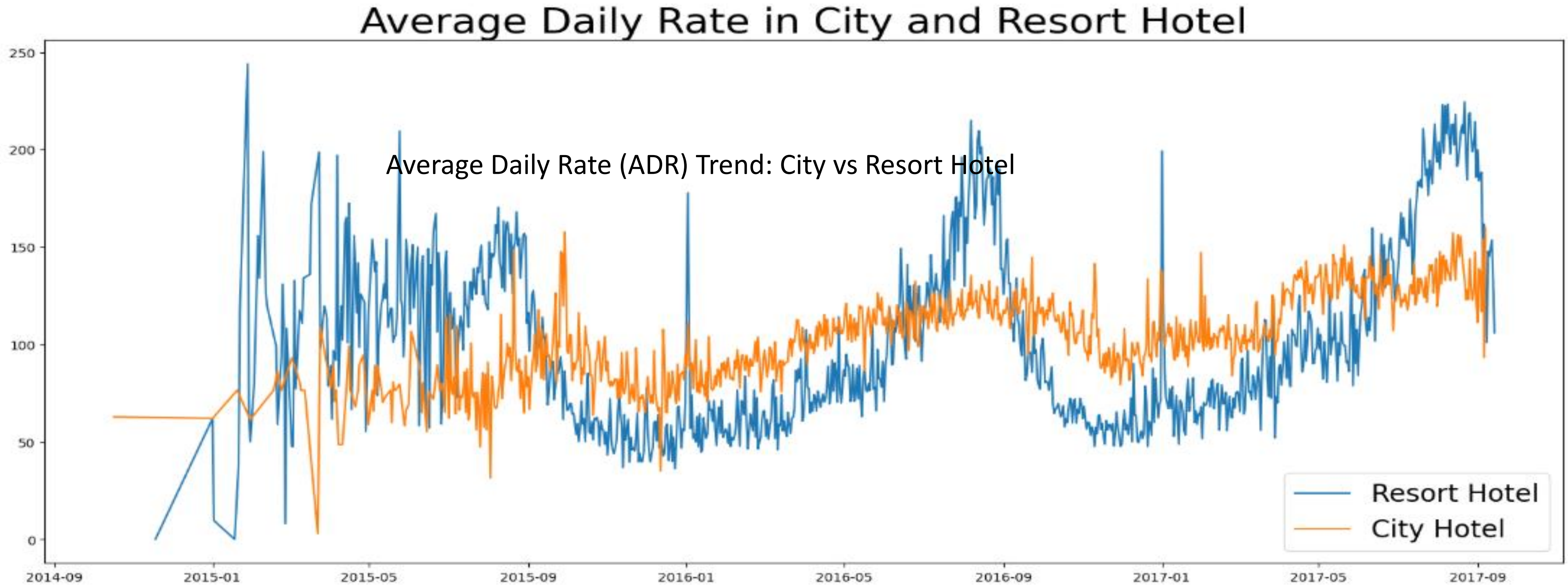


BUSINESS IMPACT:

- City Hotel has the highest number of cancellations compared to Resort Hotel.
- City Hotel also has more total bookings, which increases operational pressure.
- Resort Hotel shows lower cancellations, indicating more stable bookings.
- The cancellation rate is significantly higher in City Hotels, making them the main contributor to revenue loss.
- City Hotels should be the primary focus for cancellation reduction strategies.

AVERAGE DAILY RATE (ADR) TREND: CITY VS RESORT HOTEL

```
plt.figure(figsize=(20, 8))
plt.title('Average Daily Rate in City and Resort Hotel', fontsize=30)
plt.plot(resort_hotel.index, resort_hotel['adr'], label='Resort Hotel')
plt.plot(city_hotel.index, city_hotel['adr'], label='City Hotel')
plt.legend(fontsize=20)
plt.show()
```

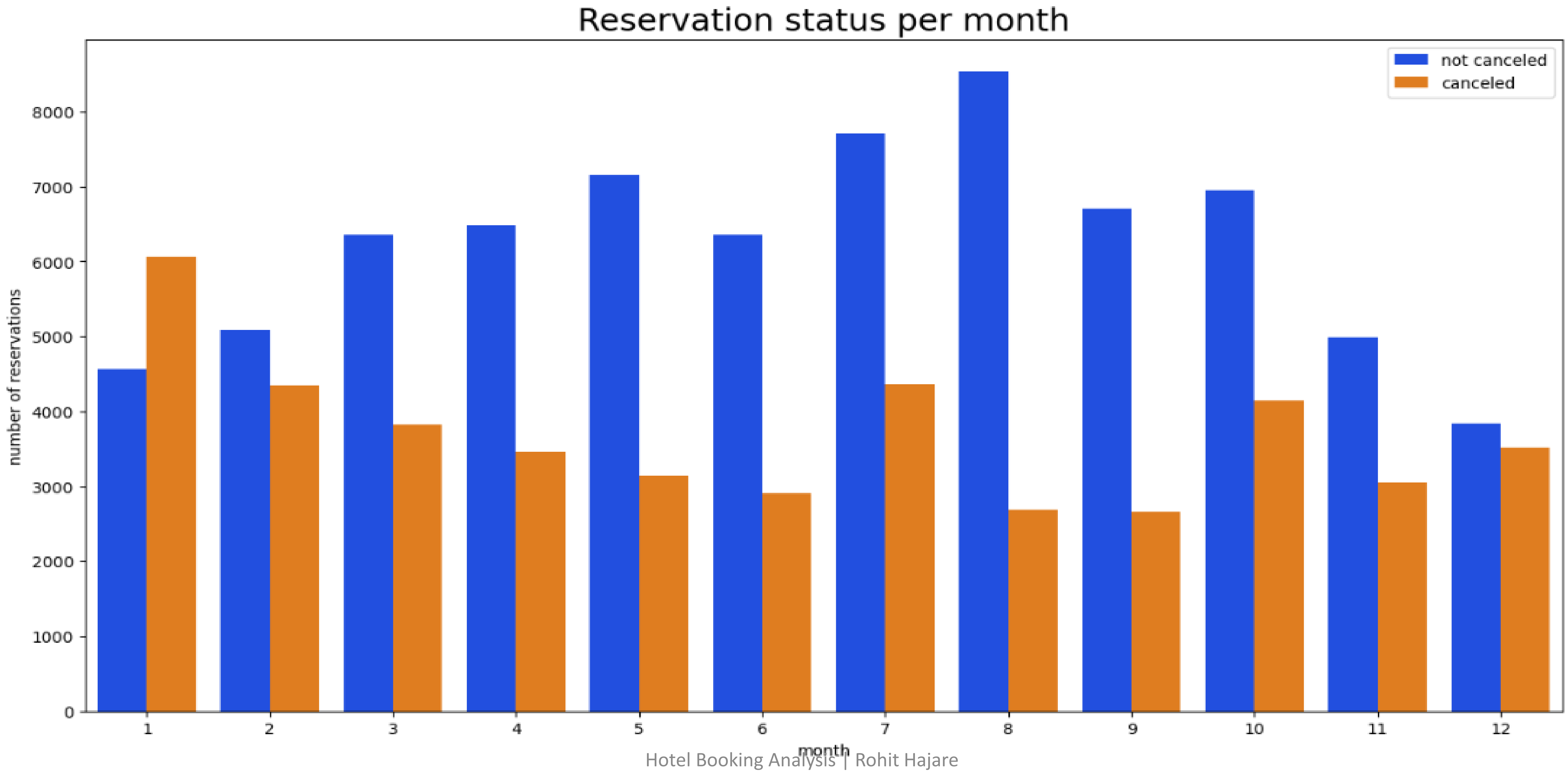


Average Daily Rate (ADR) Trend: City vs Resort Hotel

BUSINESS IMPACT:

- Resort Hotel ADR is more volatile than City Hotel, with sharp price spikes and drops.
- City Hotel ADR is more stable, showing smoother price changes over time.
- Both hotels show seasonal patterns with price peaks during high-demand periods.
- Resort Hotels reach higher peak ADRs, indicating premium seasonal pricing.
- City Hotels maintain moderate but consistent pricing, suitable for business travelers.
- Price fluctuations suggest that demand varies strongly by season, especially for Resort Hotels.

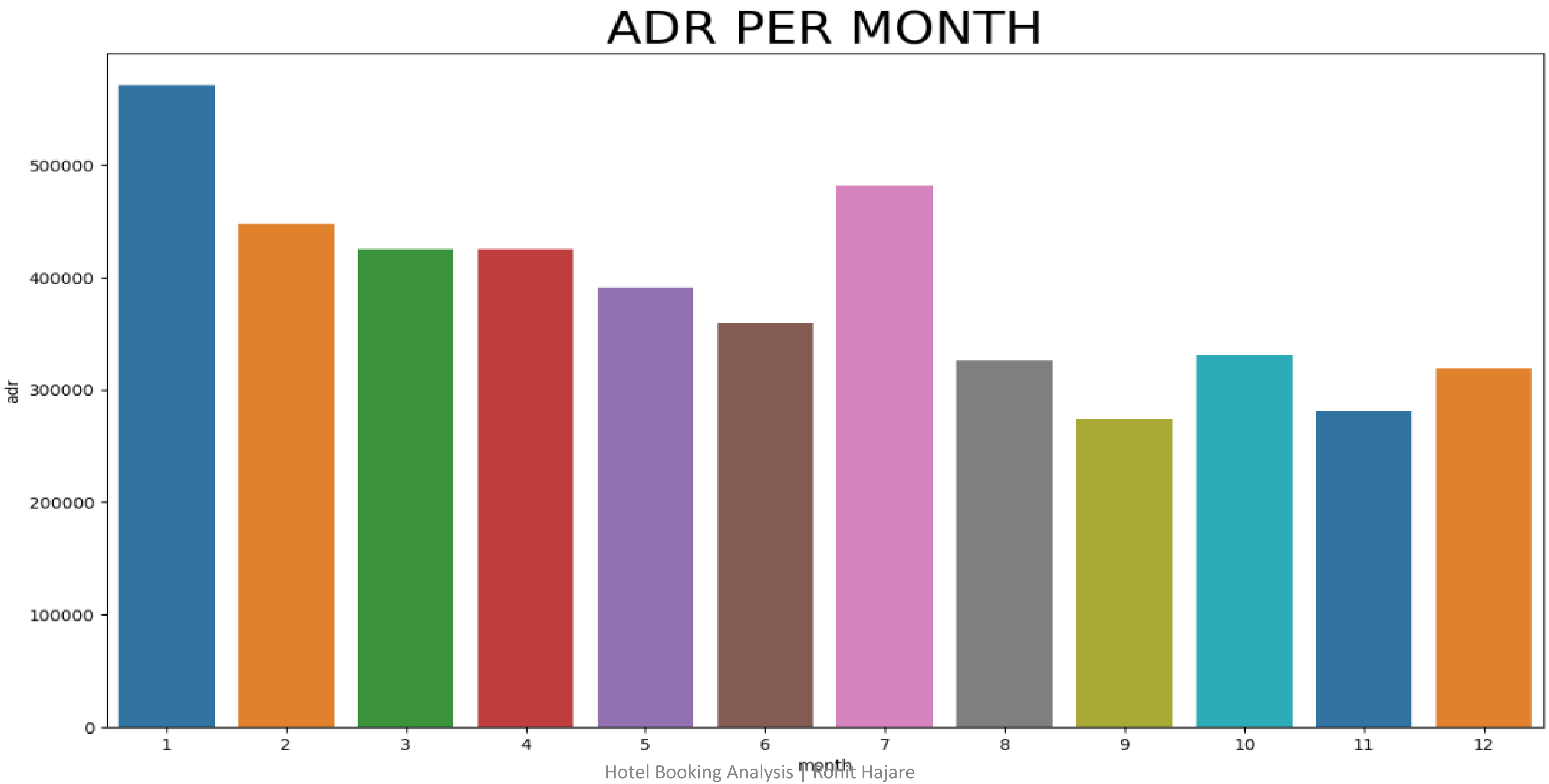
MONTHLY RESERVATION & CANCELLATION TREND



BUSINESS IMPACT

- Peak bookings occur in July and August, showing strong seasonality.
- Cancellations are highest in January and October, indicating planning uncertainty at the start and end of the year.
- Lowest cancellations occur in August and September, when demand is strongest.
- Mid-year months (May–August) show high confirmed bookings with lower cancellation ratios.
- Winter months show lower bookings but relatively higher cancellations, reducing revenue stability.

MONTHLY AVERAGE DAILY RATE (ADR) TREND

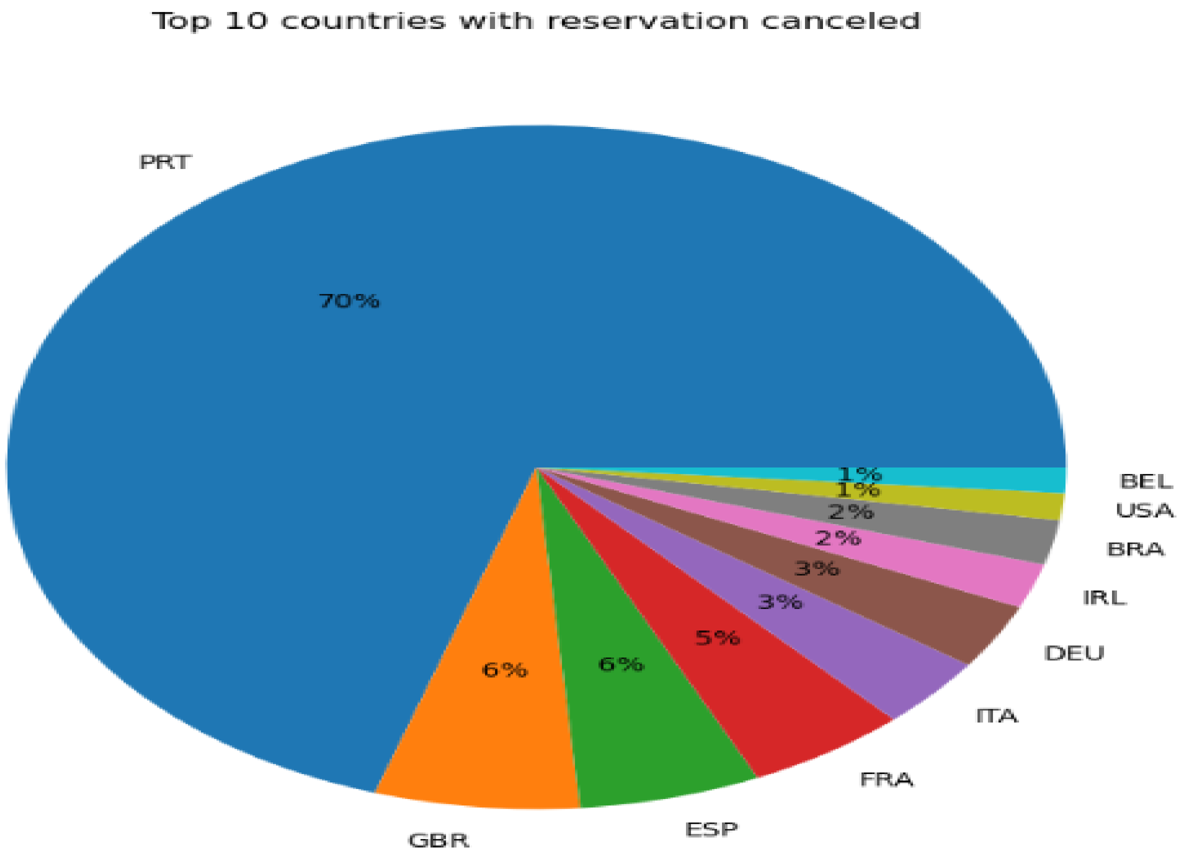


BUSINESS IMPACT

- Highest ADR is in January, indicating premium pricing during peak travel or holiday season.
- ADR peaks again in July, showing strong summer demand.
- Lowest ADR occurs in September, suggesting off-season pricing.
- ADR gradually declines from January to June, then rises again in July.
- After July, ADR drops sharply in August and September, then slightly recovers toward year-end.
- Pricing clearly follows seasonal demand patterns.

TOP 10 COUNTRIES WITH HIGHEST BOOKING CANCELLATIONS

[37]:
```
cancelled_data = df[df['is_canceled'] == 1]
top_10_country = cancelled_data['country'].value_counts()[:10]

plt.figure(figsize=(8,8))
plt.title('Top 10 countries with reservation canceled')

plt.pie(
    top_10_country,
    autopct='%2.f%%',
    labels=top_10_country.index
)

plt.show()
```

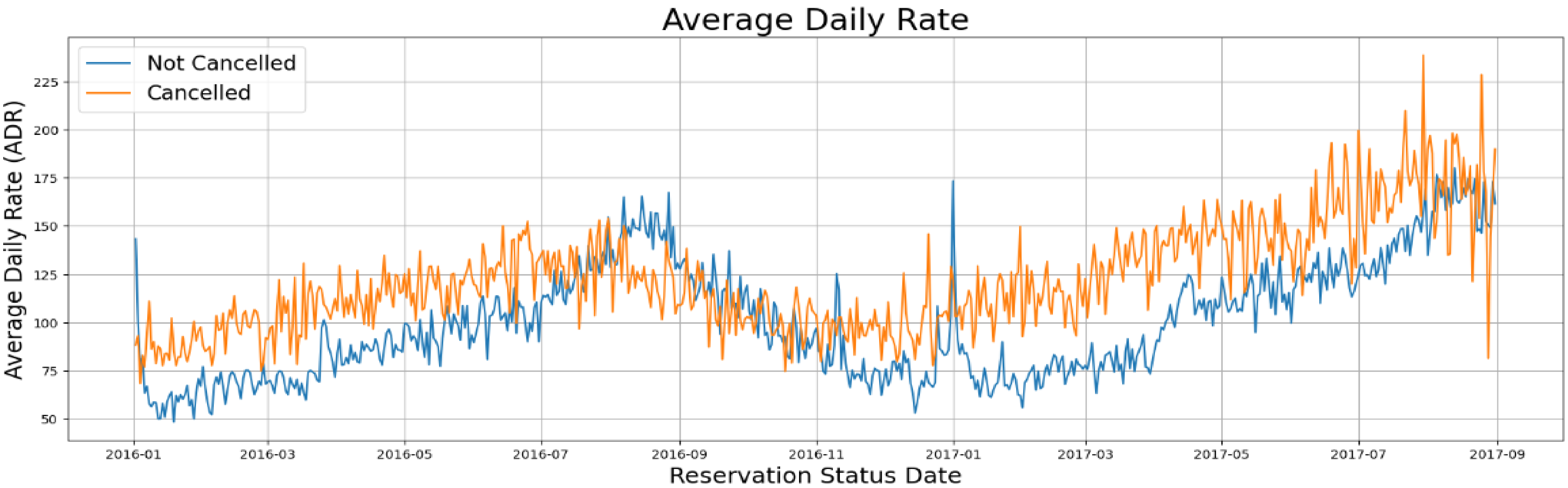


BUSINESS IMPACT

- Portugal (PRT) alone contributes ~70% of all cancellations, making it the dominant source.
- The next highest are:
  - Great Britain (GBR) – ~6%
  - Spain (ESP) – ~6%
  - France (FRA) – ~5%
- Other countries (ITA, DEU, IRL, BRA, USA, BEL) each contribute less than 3%.
- Cancellations are highly concentrated in a few countries, especially domestic travelers (PRT).
- Targeted policies for high-cancellation regions can significantly reduce overall losses.

ADR COMPARISON: CANCELLED VS NOT CANCELLED BOOKINGS

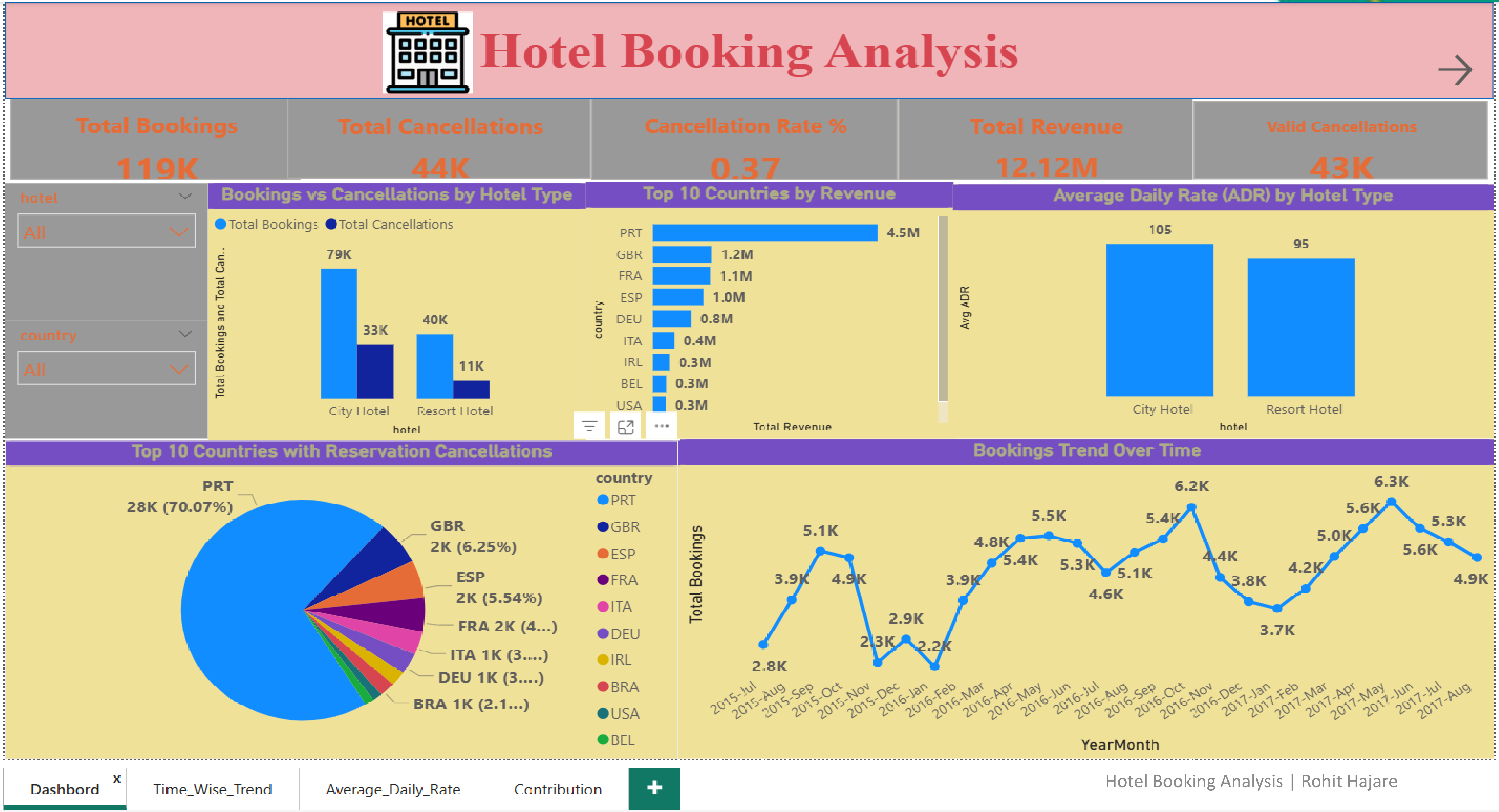
```
import matplotlib.pyplot as plt

plt.figure(figsize=(20, 6))
plt.title('Average Daily Rate', fontsize=24)
plt.plot(not_cancelled_df_adr['reservation_status_date'], not_cancelled_df_adr['adr'], label='Not Cancelled')
plt.plot(cancelled_df_adr['reservation_status_date'], cancelled_df_adr['adr'], label='Cancelled')
plt.legend(fontsize=16)
plt.xlabel('Reservation Status Date', fontsize=18)
plt.ylabel('Average Daily Rate (ADR)', fontsize=18)
plt.grid(True)
plt.show()
```

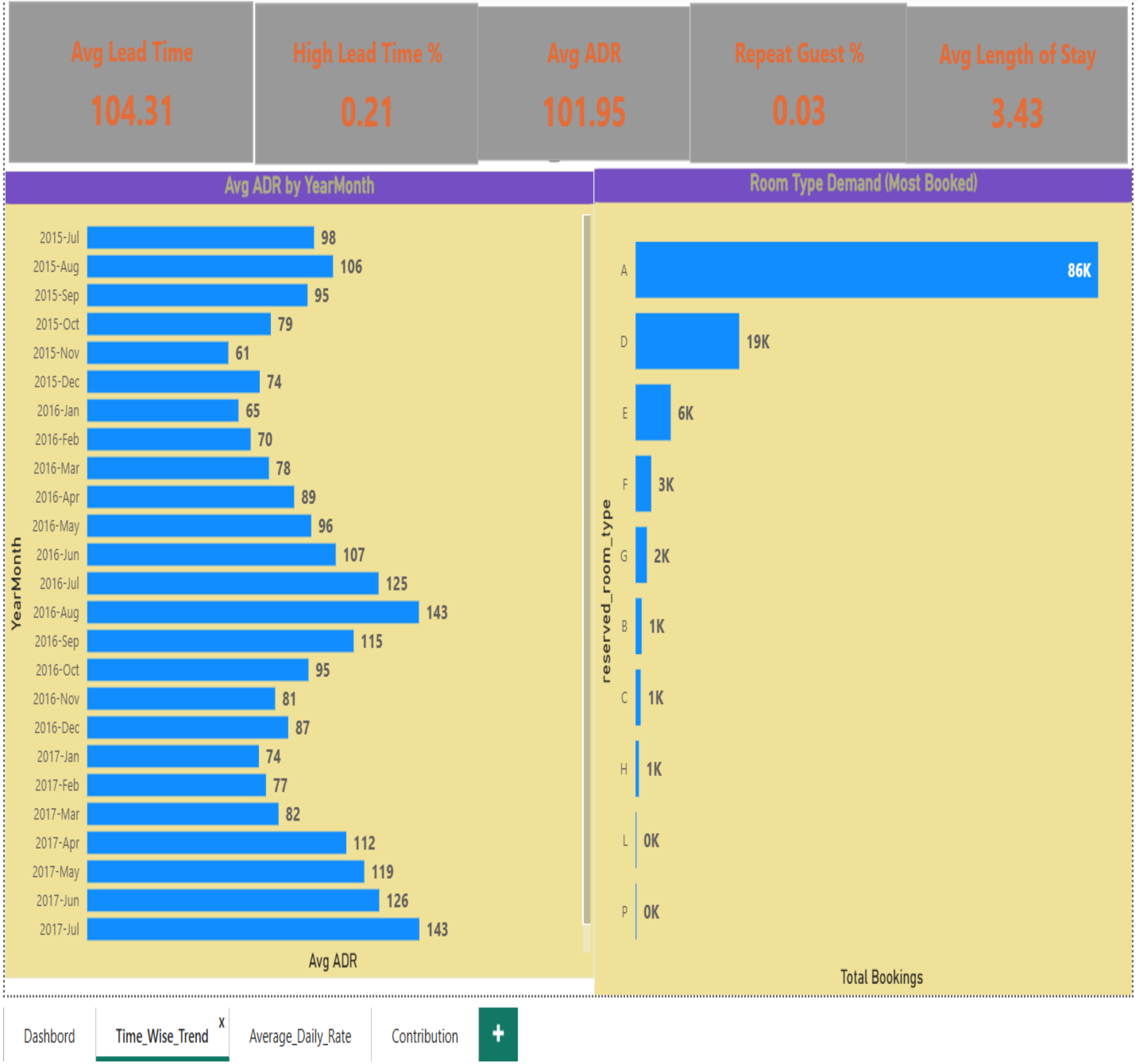


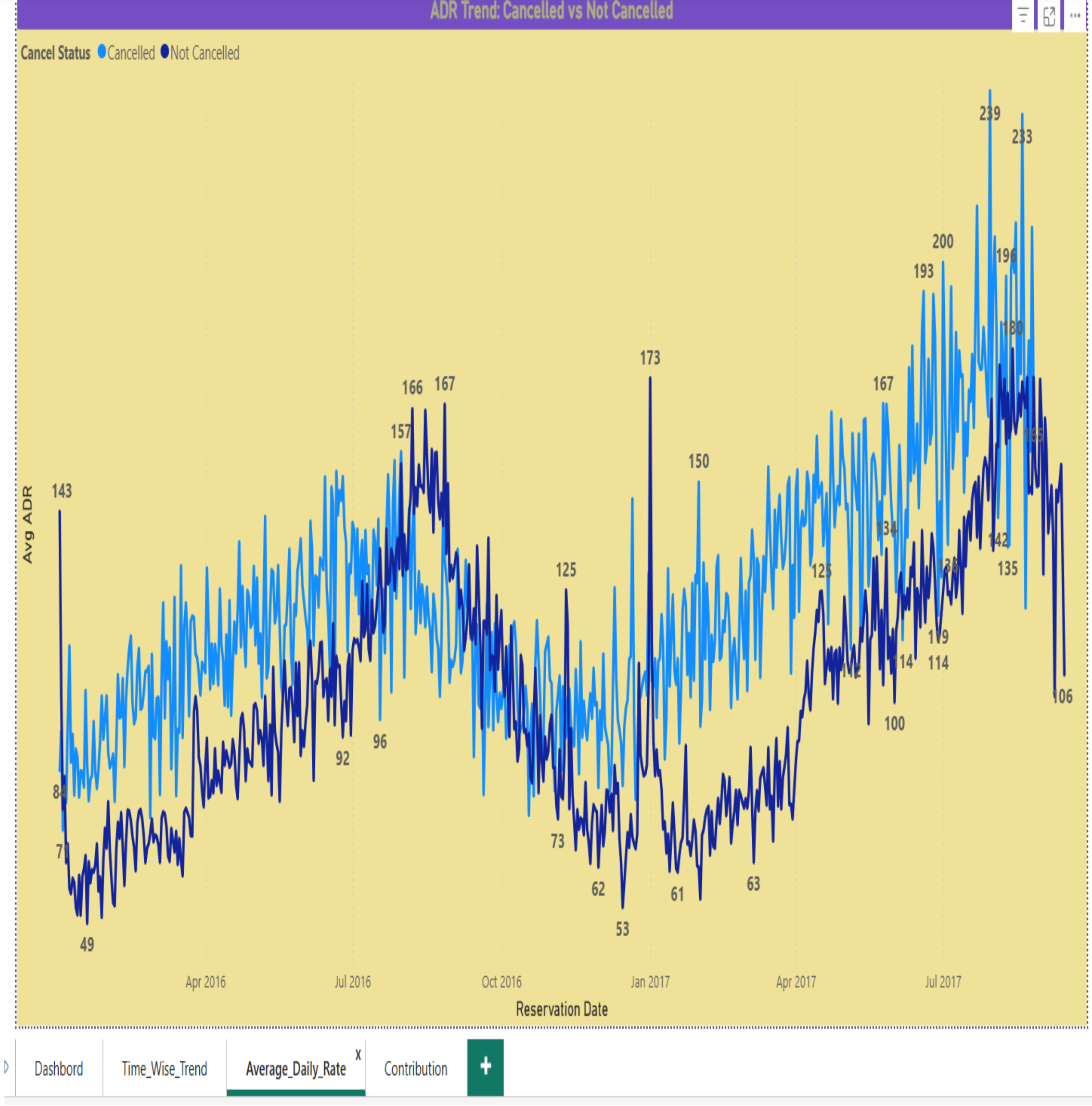
BUSINESS IMPACT

- Cancelled bookings consistently have higher ADR than non-cancelled bookings.
- Guests are more likely to cancel high-priced rooms, showing strong price sensitivity.
- When ADR increases, the gap between cancelled and confirmed bookings widens.
- Lower ADR bookings are more stable and less likely to be cancelled.
- This indicates that dynamic pricing strategies can help reduce cancellations by controlling high-rate volatility.

POWER BI DASHBOARD







***Built an end-to-end Hotel Booking Cancellation Analysis using Python, SQL, and Power BI to identify revenue leakage drivers, reduce cancellation risk, and improve occupancy forecasting*.**

DASHBOARD SECTIONS (SLIDE HEADINGS)

1. Overall Booking Performance
2. Cancellation Analysis Overview
3. Hotel Type Comparison (City vs Resort)
4. Monthly Booking & Cancellation Trends
5. Average Daily Rate (ADR) Analysis
6. Revenue & Loss Impact
7. Customer & Country Insights
8. Lead Time & Cancellation Behavior
9. Seasonality Impact on Bookings
10. Key Business KPIs

STRATEGIC BUSINESS RECOMMENDATIONS

1. Introduce partial prepayment for long lead time bookings.
2. Offer discounts for non-refundable bookings.
3. Apply dynamic pricing during peak seasons.
4. Create loyalty programs to increase repeat guests.
5. Improve cancellation policy transparency.

These actions directly reduce cancellation risk and improve revenue predictability.

PROJECT VALUE DELIVERED

- Identified high-risk cancellation segments
- Improved demand & pricing strategy clarity
- Enabled data-driven revenue optimization
- Built Power BI dashboard for real-time decisions

CONCLUSION

- Cancellations significantly impact revenue.
- Data-driven strategies can reduce risk.
- Power BI dashboard enables real-time monitoring.