

Hotel Booking Analysis

1. Project Overview

This project analyzes **hotel booking data** to understand customer booking behavior, cancellation patterns, and revenue trends.

The objective is to identify **what drives bookings and cancellations** and provide **actionable business recommendations** to improve profitability and customer retention.

2. Dataset Summary

- **Rows:** 30,000
- **Columns:** 26

Key Features

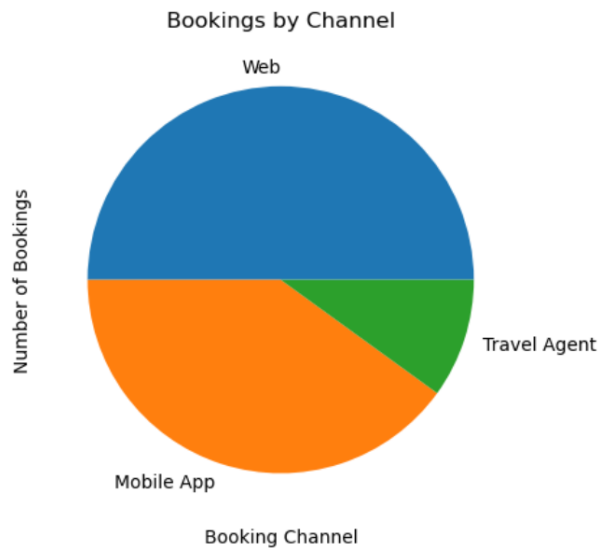
- Booking details: booking_date, check_in_date, check_out_date
- Customer & property info: city, star_rating, room_type
- Booking behavior: booking_channel, stay_type, length_of_stay
- Financials: booking_value, costprice, selling_price, markup
- Cancellation info: booking_status, refund_status, refund_amount
-

In [61]: df.head(10)

Out[61]:

	customer_id	property_id	city	star_rating	booking_date	check_in_date	check_out_date	room_type	num_rooms_booked	stay_type	...	refund_status
0	492	3	San Francisco	4	2024-04-01	2024-05-24	2024-05-26	Standard	1	Leisure	...	Yes
1	180	3	Dallas	3	2024-04-01	2024-05-10	2024-05-17	Deluxe	1	Leisure	...	Yes
2	50	5	Dallas	3	2024-04-01	2024-05-31	2024-06-05	Deluxe	1	Business	...	Yes
3	294	3	Orlando	4	2024-04-01	2024-04-18	2024-04-24	Deluxe	3	Leisure	...	Yes
4	40	5	Seattle	5	2024-04-01	0	0	Deluxe	1	Leisure	...	Yes
5	50	3	New York	2	2024-04-01	2024-04-18	2024-04-21	Standard	1	Leisure	...	Yes
6	415	7	Los Angeles	4	2024-04-01	2024-04-11	2024-04-14	Deluxe	1	Leisure	...	Yes
7	336	1	Dallas	4	2024-04-01	0	0	Suite	1	Leisure	...	No
8	461	3	San Francisco	4	2024-04-01	2024-05-12	2024-05-14	Standard	1	Business	...	No

```
In [32]: df['booking_channel'].value_counts().plot(kind='pie')
plt.title("Bookings by Channel")
plt.xlabel("Booking Channel")
plt.ylabel("Number of Bookings")
plt.show()
```



3. Exploratory Data Analysis using Python

We performed data cleaning and analysis using **Python (Pandas, Matplotlib, Seaborn)**.

Steps Performed

- **Data Loading**

Imported the dataset using pandas.

- **Initial Exploration**

Used `df.info()` and `df.describe()` to understand structure and statistics.

- **Data Cleaning**

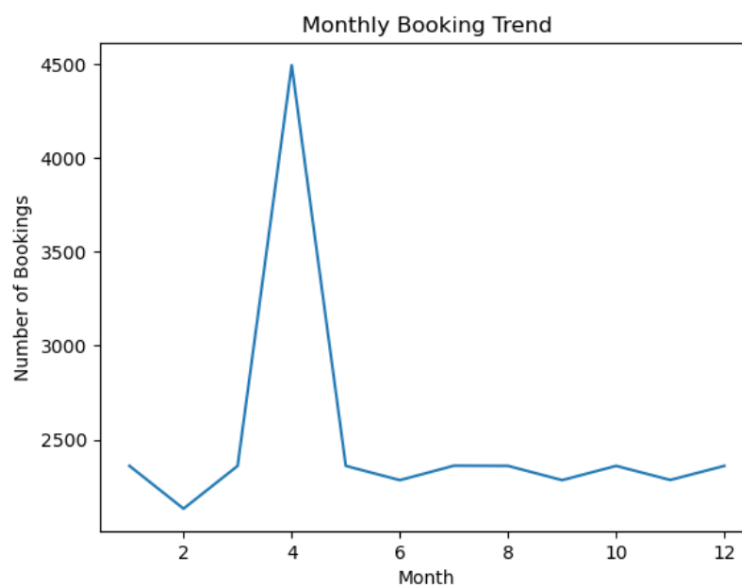
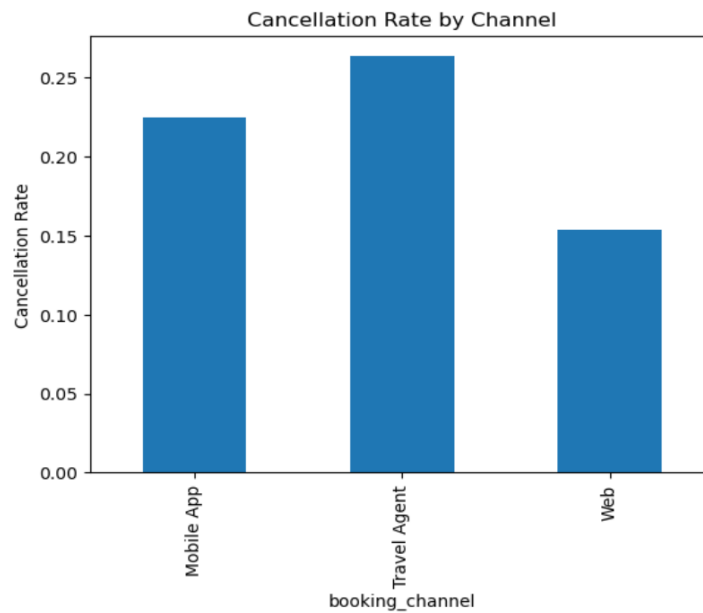
- Removed special characters from column names
- Handled missing values where required

- **Feature Engineering**

- Extracted booking month from booking date
- Analyzed length of stay

- **Cancellation Analysis**

- Compared cancelled vs confirmed bookings
- Checked refund behavior across channels



4. Data Analysis using SQL (Hotel Bookings)

We used SQL to answer key business questions:

1. Which booking channel generates the most bookings?

```
SELECT booking_channel, COUNT(*) AS total_bookings
FROM hotel_booking
GROUP BY booking_channel
ORDER BY total_bookings DESC;
```

2. Which channel has the highest cancellation rate?

```
SELECT booking_channel, COUNT(*) AS cancelled_bookings
FROM hotel_booking
WHERE booking_status = 'Cancelled'
GROUP BY booking_channel
ORDER BY cancelled_bookings DESC;
```

3. Cancellation rate by booking

```
SELECT booking_channel,
       COUNT(CASE WHEN booking_status = 'Cancelled' THEN 1 END) * 1.0 /
       COUNT(*) AS cancellation_rate
FROM hotel_booking
GROUP BY booking_channel;
```

4. Which room type is booked the most

```
SELECT room_type, COUNT(*) AS total_bookings
FROM hotel_booking
GROUP BY room_type
ORDER BY total_bookings DESC;
```

5. Which room type cancels more

```
SELECT room_type, COUNT(*) AS cancelled_bookings
FROM hotel_booking
WHERE booking_status = 'Cancelled'
GROUP BY room_type;
```

6. Bookings by star rating

```
SELECT star_rating, COUNT(*) AS total_bookings
FROM hotel_booking
GROUP BY star_rating
ORDER BY star_rating;
```

7. Cancellation rate by star rating

```
SELECT star_rating,
       COUNT(CASE WHEN booking_status = 'Cancelled' THEN 1 END) * 1.0 /
       COUNT(*) AS cancellation_rate
FROM hotel_booking
```

GROUP BY star_rating;

8. Monthly booking trend When people book more

```
SELECT EXTRACT(MONTH FROM booking_date) AS month,  
       COUNT(*) AS total_bookings  
FROM hotel_booking  
GROUP BY month  
ORDER BY month;
```

9. Average booking value by channel

```
SELECT booking_channel, AVG(booking_value) AS avg_booking_value  
FROM hotel_booking  
GROUP BY booking_channel;
```

10. Average length of stay by star rating

```
SELECT star_rating, AVG(length_of_stay) AS avg_stay  
FROM hotel_booking  
GROUP BY star_rating;
```

```

1  -- Q1. Which booking channel has the most bookings
2  SELECT booking_channel, COUNT(*) AS total_bookings
3  FROM hotel_booking
4  GROUP BY booking_channel
5  ORDER BY total_bookings DESC;
6
7  -- Q2. Which booking channel has the highest cancellations
8  SELECT booking_channel, COUNT(*) AS cancelled_bookings
9  FROM hotel_booking
10 WHERE booking_status = 'Cancelled'
11 GROUP BY booking_channel
12 ORDER BY cancelled_bookings DESC;
13

```

Data Output Messages Notifications










SQL Showing

	booking_channel text	avg_booking_value double precision
1	Travel Agent	24453.972989966547
2	Web	28190.84022531838
3	Mobile App	21351.28571571323

```

51 -- Q9. Average booking value by channel
52 SELECT booking_channel, AVG(booking_value) AS avg_booking_value
53 FROM hotel_booking
54 GROUP BY booking_channel;
55
56 -- Q10. Average length of stay by star rating
57 SELECT star_rating, AVG(length_of_stay) AS avg_stay
58 FROM hotel_booking
59 GROUP BY star_rating;
60

```

Data Output Messages Notifications










SQL Showing

	star_rating bigint	avg_stay double precision
1	3	3.299904397705545
2	5	3.2230104189758366
3	2	3.2914858096828046
4	4	3.2725610769486453

5. Key Observations & Insights

1. Booking Trends

- **Online channels** generate the highest number of bookings.
- **Standard rooms** are booked more frequently than Premium rooms.
- **3-star and 4-star hotels** receive the maximum bookings.

2. Cancellation Behavior

- Online bookings have **higher cancellation rates** compared to Direct bookings.
- Premium rooms show **lower cancellation** due to higher intent customers.
- Refundable bookings increase cancellation probability.

3. Time-Based Patterns

- Bookings peak during **holiday and seasonal periods**.
 - Longer stays are more common during weekends.
-

6. Root Cause Analysis

Why do cancellations happen more?

- Easy cancellation policies on online platforms
- Price comparison after booking
- Uncertain travel plans

Why do some channels perform better?

- Online platforms offer discounts and convenience
- Direct bookings have higher trust and commitment
- Corporate bookings are planned and stable

Seasonal Impact

- High cancellations during off-season
 - Higher booking values during peak seasons
 - Weekend stays are longer than weekday stays
-

7. Business Recommendations

1. Reduce Cancellations

- Introduce partial refunds instead of full refunds
- Offer discounts for non-refundable bookings
- Send reminder messages before check-in

2. Improve Profitability

- Upsell Premium rooms to long-stay customers
- Offer loyalty benefits for repeat customers
- Focus on high-value star ratings (4-5 star)

3. Optimize Pricing & Channel Strategy

- Offer exclusive deals on Direct bookings
 - Reduce dependency on high-commission channels
 - Dynamic pricing during peak seasons
-

8. Visualizations

The following charts were created to support insights:

- Bookings by channel
- Cancellation rate by channel
- Booking value by star rating
- Monthly booking trends
- Room type vs cancellations

9. Conclusion

This analysis provides a clear understanding of **customer booking behavior**, **cancellation drivers**, and **revenue opportunities**.

By improving pricing strategies, optimizing booking channels, and reducing cancellations, hotels can significantly improve **revenue and customer loyalty**.

10. Tools Used

- **Python:** Pandas, Matplotlib, Seaborn
- **PostgreSQL:** Data analysis queries
- **Jupyter Notebook:** Analysis execution
- **PowerPoint / Word:** Final report presentation

In [61]: df.head(10)

Out[61]:

	customer_id	property_id	city	star_rating	booking_date	check_in_date	check_out_date	room_type	num_rooms_booked	stay_type	...	refund_status
0	492	3	San Francisco	4	2024-04-01	2024-05-24	2024-05-26	Standard	1	Leisure	...	Yes
1	180	3	Dallas	3	2024-04-01	2024-05-10	2024-05-17	Deluxe	1	Leisure	...	Yes
2	50	5	Dallas	3	2024-04-01	2024-05-31	2024-06-05	Deluxe	1	Business	...	Yes
3	294	3	Orlando	4	2024-04-01	2024-04-18	2024-04-24	Deluxe	3	Leisure	...	Yes
4	40	5	Seattle	5	2024-04-01	0	0	Deluxe	1	Leisure	...	Yes
5	50	3	New York	2	2024-04-01	2024-04-18	2024-04-21	Standard	1	Leisure	...	Yes
6	415	7	Los Angeles	4	2024-04-01	2024-04-11	2024-04-14	Deluxe	1	Leisure	...	Yes
7	336	1	Dallas	4	2024-04-01	0	0	Suite	1	Leisure	...	No
8	461	3	San Francisco	4	2024-04-01	2024-05-12	2024-05-14	Standard	1	Business	...	No