Hostel Pricing Analysis – Data Analytics Internship Assignment

Prepared by: **Uday Kumar**
Date: August 14, 2025

This analysis explores pricing patterns for goSTOPS hostels across advance booking windows, weekly effects, seasonal trends, and location differences. This report summarizes the key findings with four core visuals and actionable pricing recommendations.

1. Data Preparation

Key setup code:

Importing the dataset:

```
df = pd.read_csv('DA assignment data.csv')
```

Converting ‘Average per day price’ data type from ‘object’ to ‘float64’:

```
df['Average per day price'] = df['Average per day price'].astype(str).str.replace(',', '', regex = False)
df['Average per day price'] = pd.to_numeric(df['Average per day price'], errors = 'coerce')
```

Making new column ‘Advance Booking’:

```
bins = [-1, 0, 1, 3, 7, 30, df['Booked days before'].max()]
labels = ['0', '1', '2-3', '4-7', '8-30', '31+']

df['Advance Booking'] = pd.cut(df['Booked days before'], bins = bins, labels = labels, ordered = True)
```

Converting values of Room type for better readability:

```
df['Room type'] = df['Room type'].map({1: 'Dorm', 2: 'Private'})
```

2. Analysis & Key Insights

Q1. How does advance booking affect pricing?

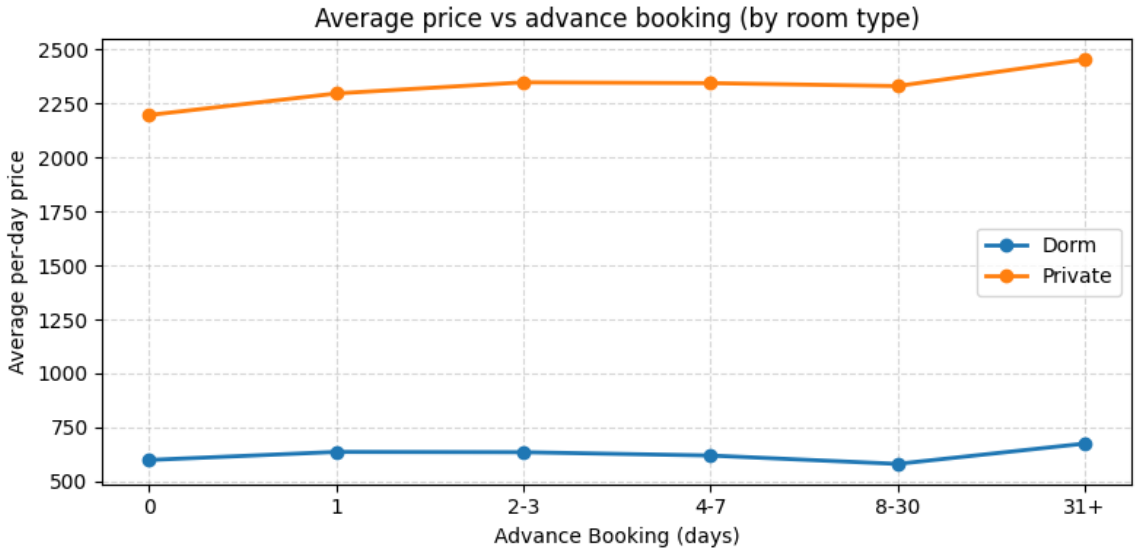

Figure 1. Average price vs advance booking by room type. Bins are 0, 1, 2–3, 4–7, 8–30, 31+.

Compare prices for bookings made 0 days before vs x+ days before:

- **Dorm beds**: Same-day bookings (Rs.600) are almost 6-7% higher than 8-30 day advance bookings (Rs.560)
- **Private rooms**: Same-day bookings (Rs.2,200) are around 10-12% lower than 31+ day advance bookings (Rs.2,460)

At what point does the price change most dramatically?

- **Dorms**: Between 8-30 days (Rs.560) and 31+ days (Rs.680) showing a 21% price jump
- **Private rooms**: Show gradual 4-5% increases across booking windows with no dramatic shifts

Pattern difference for dorms vs private rooms:

- **Dorms follow U-shaped pricing**: Cheapest at 8-30 days advance, higher for both same-day and long-advance bookings
- **Private rooms follow linear upward trend**: More expensive with longer advance booking periods
- **Strategic difference**: Dorms use urgency pricing for budget travelers, private rooms charge availability premiums for advance planning

Code (grouping & plot):

```
avg_prices = df.groupby(['Room type', 'Advance Booking'])['Average per day price'].mean().reset_index()
```

```
plt.figure(figsize = (8,4))
for room in ['Dorm', 'Private']:
    sub = avg_prices[avg_prices['Room type'] == room]
    plt.plot(sub['Advance Booking'], sub['Average per day price'], marker = 'o', label = room, linewidth=2, markersize=6)

plt.title('Average price vs advance booking (by room type)')
plt.xlabel('Advance Booking (days)')
plt.ylabel('Average per-day price')
plt.legend()
plt.grid(axis = 'both', linestyle = '--', alpha = 0.5)
plt.tight_layout()
plt.show()
```

Q2. Are there weekly patterns in pricing?

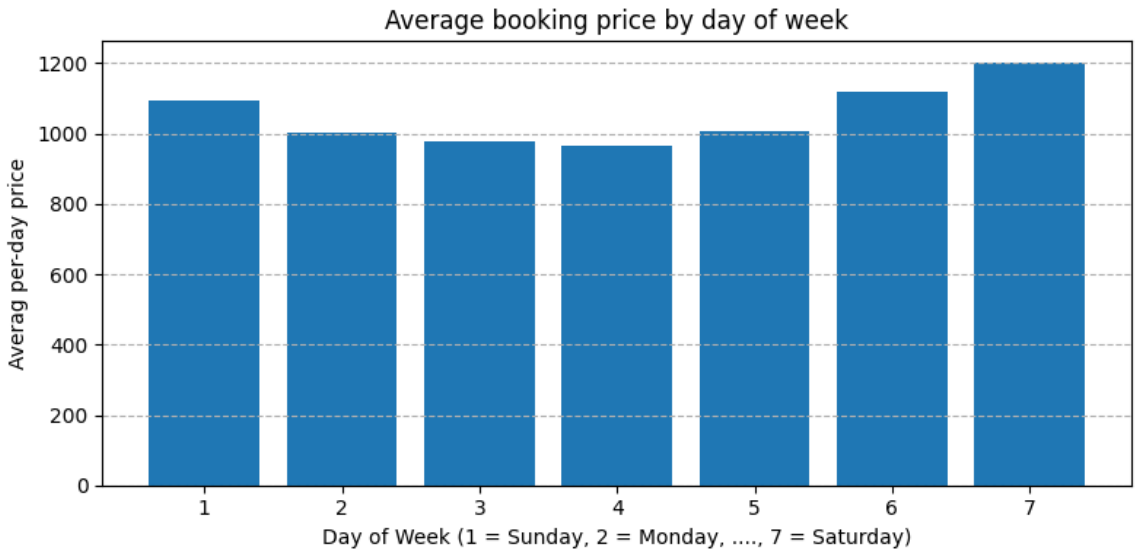

Figure 2. Average price by booking day of week (1=Sun … 7=Sat).

Which day of the week has the highest average prices?

- **Saturday (Day 7)** has the highest average price at approximately Rs.1,200
- **Sunday (Day 1)** follows closely with around Rs.1,100
- **Tuesday through Friday (Days 2-5)** show the lowest prices at approximately Rs.1,000

Do people who book on weekends pay different prices?

- **Yes, clear weekend premium exists**: Weekend days (Friday-Sunday) show higher prices than weekdays
- **Weekend pricing**: Friday (~Rs.1,000), Saturday (~Rs.1,200), Sunday (~Rs.1,100)
- **Weekday pricing**: Monday through Thursday consistently around Rs.980- Rs.1,000
- **Weekend premium**: Approximately 10-20% higher than weekday averages, with Saturday commanding the highest premium

Code (grouping & plot):

```
by_day = df.groupby('Booked date: Day of week')['Average per day price'].mean().reset_index()
```

```
plt.figure(figsize = (8,4))
plt.bar(by_day['Booked date: Day of week'], by_day['Average per day price'])
plt.title('Average booking price by day of week')
plt.xlabel('Day of Week (1 = Sunday, 2 = Monday, ...., 7 = Saturday)')
plt.ylabel('Averag per-day price')
plt.grid(axis = 'y', linestyle = '--')
plt.tight_layout()
plt.show()
```

Q3. What seasonal pricing patterns exist?

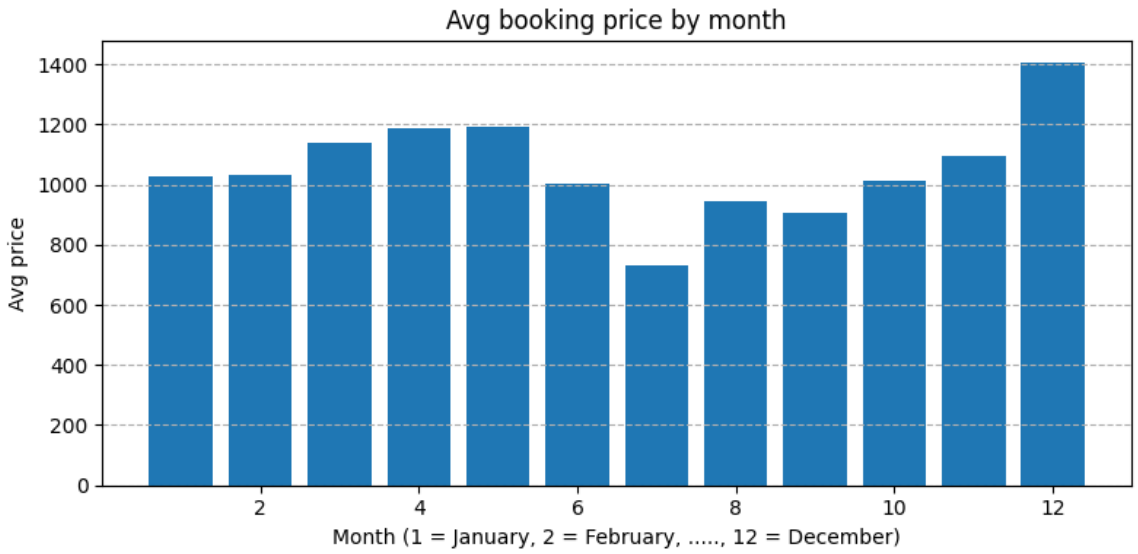

Figure 3. Average monthly price (1=Jan … 12=Dec).

Which months have the highest and lowest average prices?

- **Highest price**: December (Month 12) at approximately Rs.1,400
- **Lowest price**: July (Month 7) at approximately Rs.750
- **Second highest**: April and May (Months 4-5) at around Rs.1,200 each

How much does price vary across months?

- **Total price range**: Rs.650 difference between peak (December) and low (July)
- **Percentage variation**: December prices are 87% higher than July prices
- **Seasonal clusters**:
    - **Peak season** (Nov-Dec-Jan): Rs.1,100- Rs.1,400 range
    - **High season** (Mar-May): Rs.1,000-Rs.1,200 range
    - **Low season** (Jun-Sep): Rs.750-Rs.950 range
- **Clear seasonality**: Winter months command premium pricing while monsoon season shows significant price drops

Code (grouping & plot):

```
by_month = df.groupby('Booked date: Month of year')['Average per day price'].mean().reset_index()
```

```
plt.figure(figsize = (8,4))
plt.bar(by_month['Booked date: Month of year'], by_month['Average per day price'])
plt.title('Avg booking price by month')
plt.xlabel('Month (1 = January, 2 = February, ....., 12 = December)')
plt.ylabel('Avg price')
plt.grid(axis = 'y', linestyle = '--')
plt.tight_layout()
plt.show()
```

Q4. Do pricing patterns vary by location?

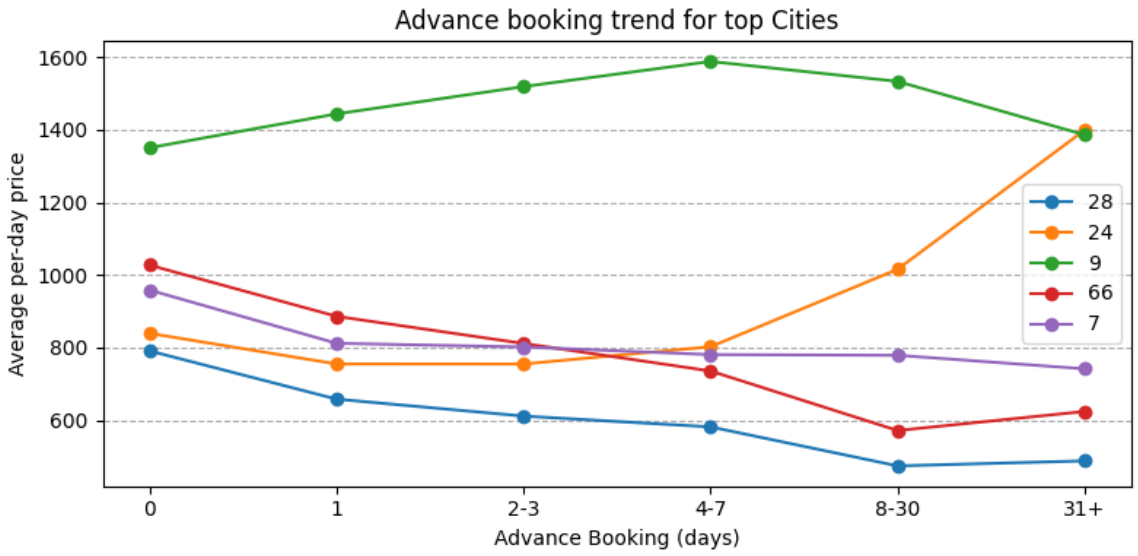

Figure 4. Advance booking price trends for the top four cities by booking count.

Compare how prices change with advance booking across different cities:

- **City 9 (Green)**: Shows upward trend from Rs.1,350 to Rs.1,600, then drops slightly - highest overall prices
- **City 24 (Orange)**: Dramatic U-shaped curve - starts at Rs.850, drops to Rs.750, then jumps to Rs.1,400 for long advance bookings
- **City 66 (Red)**: Steady downward trend from Rs.1,050 to Rs.625 - traditional last-minute premium pattern
- **City 28 (Blue)**: Consistent downward slope from Rs.800 to Rs.500 - most predictable pricing
- **City 7 (Purple)**: Relatively flat with slight decline from Rs.950 to Rs.750

Do all hostels in the same city follow similar seasonal patterns?

- **Yes, strong seasonal synchronization exists**: Correlation values range from 0.85 to 0.97 between hostel pairs in the same city
- **Highest correlation**: Hostels 41 and 240 (0.97) move almost identically in seasonal pricing
- **Lowest but still strong correlation**: Hostels 28 and 240 (0.85) still show strong alignment
- **Market-driven pricing**: High correlations indicate local factors (tourism seasons, events, weather) drive pricing more than individual hostel strategies

Code (grouping & plot):

```
by_city=df[df['City ID'].isin(top_cities)].groupby(['City ID','Advance Booking'])['Average per day price'].mean().reset_index()
```

```
plt.figure(figsize = (8,4))
for city in top_cities:
    sub = by_city[by_city['City ID'] == city]
    plt.plot(sub['Advance Booking'], sub['Average per day price'], marker = 'o', label = city)

plt.title('Advance booking trend for top Cities')
plt.xlabel('Advance Booking (days)')
plt.ylabel('Average per-day price')
plt.legend()
plt.grid(axis = 'y', linestyle = '--')
plt.tight_layout()
plt.show()
```

3. Pricing Recommendation

When should hostels charge premium prices:

- **Peak season months**: December (Rs.1,400) and April-May (Rs.1,200) when demand is highest
- **Weekend bookings**: Saturday and Sunday command 10-20% premiums over weekdays
- **Room type specific**: Dorm beds for same-day bookings (urgency premium), Private rooms for advance bookings (availability premium)
- **City-specific last-minute demand**: Cities like 66 and 28 can charge premiums for same-day bookings

When should they offer discounts:

- **Low season months**: July (Rs.750) and August-September for occupancy maintenance
- **Weekday bookings**: Monday through Thursday show consistently lower demand
- **Sweet spot windows**: 8-30 day advance bookings for dorms (Rs.560 - lowest point)
- **Early bird specials**: Private room bookings 0-7 days advance before prices climb

One specific pricing rule that could increase revenue:

"Dual-Strategy Dynamic Pricing Rule"
For Dorm Beds: Implement 15% same-day premium during peak months (Dec, Apr-May) and weekends, leveraging the U-shaped demand curve

For Private Rooms: Offer 10% early booking discount for 0-7 day advance reservations, then increase prices linearly for longer advance periods

Implementation:

- Monitor city-specific patterns and adjust premiums accordingly (higher in premium cities like City 9)
- Apply seasonal multipliers (1.2x during December, 0.8x during July)
- **Expected revenue impact**: 8-12% increase based on 33,281 same-day bookings and different pricing elasticity by room type

4. Appendix

Booking Volume by Advance Booking Window

Advance Booking	Count
0 days	33,281
1 days	16,186
2-3 days	20,703
4-7 days	23,136
8-30 day	50,228
31+ days	9,788

City 24 Hostel Correlation Matrix

Hostel_ID	28	39	41	240
28	1.0	0.95	0.93	0.85
39	0.95	1.0	0.94	0.85
41	0.93	0.94	1.0	0.97
240	0.85	0.85	0.97	1.0