



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

**Exploratory Data Analysis on Hotel Pricing
Trends (Booking.com Data)**

About me

Anish K

(B.sc in Animation and vfx - 2020)

I have 4 years of experience as a CG Texturing Artist, where I developed strong skills in precision and detail. Seeking a stable, growth-oriented career, I transitioned into data analytics and look forward to applying my analytical mindset to identify opportunities and drive positive outcomes.

LinkedIn: www.linkedin.com/in/anish-gowda-147279312

Business Problem / Use Case

1. Hotels differ widely in pricing depending on location, star rating, amenities, and reviews.
2. Travelers need insights for better decision-making.
3. Businesses (hotels, online travel agencies) can use pricing patterns for competitive advantage.

Objective of the Project

1. To examine pricing variations among hotels in various locations.
 - Consider the impact of:
 - **Destination/Location**
 - The location of a hotel is often the biggest factor in price.
 - **Star Ratings**
 - Hotels are categorized by stars (e.g., 3-star, 5-star). Higher star ratings often mean better services, luxury facilities, and higher pricing.
 - **Reviews & Scores**
 - Customer reviews and ratings also affect hotel pricing.
 - **Room Types**
 - Prices vary by room type (standard, deluxe, suite, family room, etc). Larger or premium rooms naturally cost more.
 - **Booking Features**
 - Some hotels offer extra features like free cancellation, free breakfast, or flexible booking. These benefits can lead to slightly higher prices.

Web Scraping

Website : <https://www.booking.com/>

Process Followed:

The scraping was done using Python libraries:

- **Requests** → To fetch the webpage content.
- **BeautifulSoup** → To parse and extract information from HTML.
- **Pandas** → To organize and structure the scraped data into a usable dataset.
- **Attributes Collected** → Hotel Name, Location, Price, Star Ratings, Review, Review Score, Room Type, Booking Features.

Booking.com site

The screenshot shows the Booking.com homepage with a search query for "India". The search parameters are set for Wednesday, October 1 to Thursday, October 2, for 1 adult, 0 children, and 1 room. The results page displays 64,260 properties found, sorted by "Top Picks for Solo Travelers". Two hotel listings are visible: "Super Townhouse Bidadi Formerly Royal Garuda" in Bangalore and "J Residency Near Nizamuddin Railway Station" in New Delhi. Both listings include details like room types, prices, and booking features.

Data Summary

Category	Data Overview
Rows	62,361
Columns	8
Key Columns	Hotel Name, Location, Price, Star Ratings, Review, Review Score, Room Type, Booking Features
Missing Values	Minimal, handled during cleaning
Price Range	Wide variation across hotels and locations (Min-Max)

Data Snapshot:

	Hotel_Name	Location	Price	Star_Ratings	Review	Review_score	Room_Type	Booking_Features
0	Hotel Akshith Inn	Thanjavūr	₹ 1,755	3	Good	7.8	Superior Double Room	Free cancellation
1	th lagoon	Mayyanād	₹ 3,375	Not a star hotel	Wonderful	9.3	Double or Twin Room with Lake View	Free cancellation
2	Evolve Back Kabini	Begūr	₹ 36,441	5	Wonderful	9.3	Pool Hut-Enjoy ₹5,000 resort credit for 2-nigh...	Cancellation fees
3	Anmol Residency - Near Saket Metro	South Delhi, Jāmb	₹ 878	3	Good	7.2	Deluxe Double Room (2 Adults + 1 Child)	Free cancellation
4	Hotel My Best Stay Patiala !! Parking ! Near B...	Patiāla	₹ 2,993	4	Excellent	8.7	Deluxe Room	Free cancellation

Exploratory Data Analysis:

Data Cleaning Steps

- Removed duplicates in hotel names.
- Cleaned unwanted text in Hotel Name (like “!! Parking”, “- Near...”).
- Converted price to numeric.
- Handled missing star ratings and review scores.

Data Manipulation Steps

- Extracted **Location names** clearly.
- Grouped hotels by **Location** to analyze pricing.
- Categorized **Review Scores** (Excellent, Very Good, Good, etc.).
- Created bins for Price Range.

Raw Data

	Hotel_Name	Location	Price	Star_Ratings	Review	Review_score	Room_Type	Booking_Features
0	Hotel Akshith Inn	Thanjavür	₹ 1,755	3	Good	7.8	Superior Double Room	Free cancellation
1	th lagoon	Mayyanâd	₹ 3,375	Not a star hotel	Wonderful	9.3	Double or Twin Room with Lake View	Free cancellation
2	Evolve Back Kabini	Begür	₹ 36,441	5	Wonderful	9.3	Pool Hut-Enjoy ₹5,000 resort credit for 2-nigh...	Cancellation fees
3	Anmol Residency - Near Saket Metro	South Delhi, Jâmb	₹ 878	3	Good	7.2	Deluxe Double Room (2 Adults + 1 Child)	Free cancellation
4	Hotel My Best Stay Patiala !! Parking ! Near B...	Patialâ	₹ 2,993	4	Excellent	8.7	Deluxe Room	Free cancellation
...
62356	EkanteSrivari StayInn - Tirupati	Narasingâpuram	₹ 1,618	Not a star hotel	Wonderful	9.2	Superior Double Room	Cancellation fees
62357	Super Townhouse Brahma Sarovar Formerly Krishn...	Kurukshetra	₹ 1,215	Not a star hotel	Good	7.6	Standard Double Room	Free cancellation
62358	Ranakpur Safari Resort	Sâdri	₹ 5,130	2	Very Good	8.0	Superior Deluxe Double or Twin Room	Cancellation fees
62359	Hotel City Square	Dindigul	₹ 4,759	5	Wonderful	9.0	Studio with Garden View	Free cancellation
62360	Deccan Suites, Tirupati	Tirupati	₹ 1,980	3	Very Good	8.2	Standard Double Room	Free cancellation

62361 rows x 8 columns

Total Rows : 62,361 , Total Columns : 8

Numerical Columns : 3

1. Price
2. Star Ratings
3. Review Score

Categorical Columns : 5

1. Hotel Name
2. Location
3. Review
4. Room Type
5. Booking Features

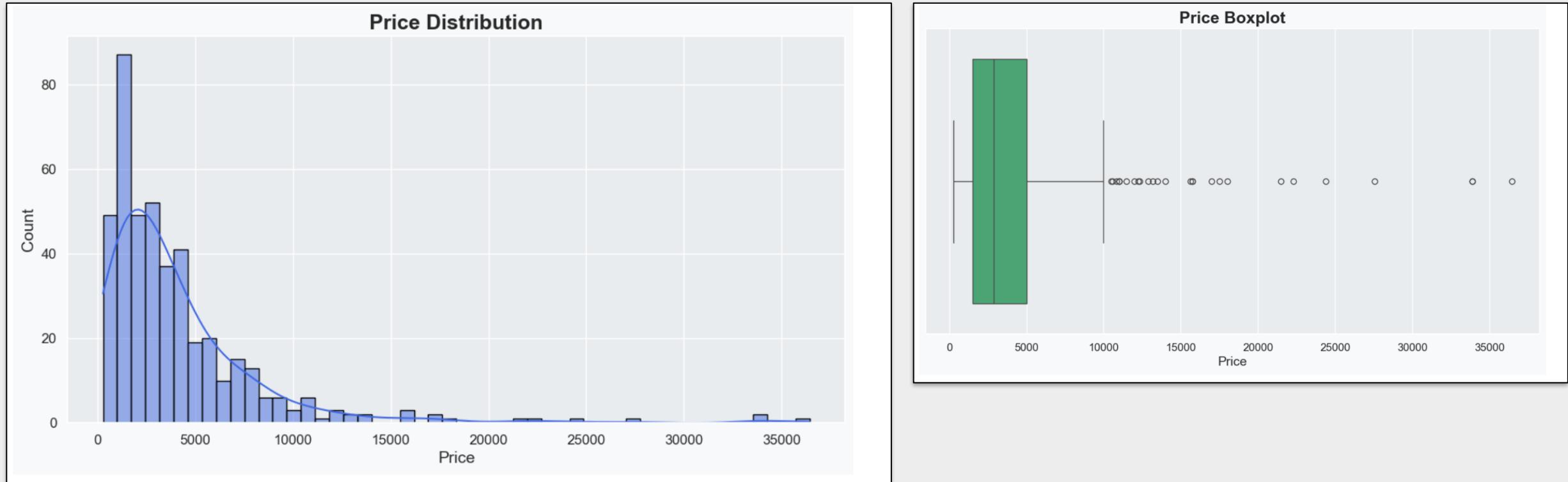
Cleaned Data

	Hotel_Name	Location	Price	Star_Ratings	Review	Review_score	Room_Type	Booking_Features
0	Hotel Akshith Inn	Thanjavür	1755.0	3	Good	7.8	Superior Double Room	Free cancellation
1	th lagoon	Mayyanâd	3375.0	0	Wonderful	9.3	Double or Twin Room with Lake View	Free cancellation
2	Evolve Back Kabini	Begür	36441.0	5	Wonderful	9.3	Pool Hut-Enjoy ₹5,000 resort credit for 2-nigh...	Cancellation fees
3	Anmol Residency - Near Saket Metro	South Delhi, Jâmb	878.0	3	Good	7.2	Deluxe Double Room (2 Adults + 1 Child)	Free cancellation
4	Hotel My Best Stay Patiala !! Parking ! Near B...	Patialâ	2993.0	4	Excellent	8.7	Deluxe Room	Free cancellation
...
429	Shree Vallabh Vilas Lords Plaza Nathdwara - Pu...	Eklingji	4320.0	0	Very Good	8.4	Club Room with Darshan Convenience & Ticket Su...	Cancellation fees
430	Shree Vallabh Vilas Lords Plaza Nathdwara - Pu...	Eklingji	4800.0	0	Very Good	8.4	Club Room with Darshan Convenience & Ticket Su...	Cancellation fees
431	Hotel Skylights Bangalore Airport	Muthugadahalli	2159.0	3	Very Good	8.1	Standard Double Room	Free cancellation
432	APN Residency	Râjapuram	2500.0	0	No review yet	8.3	Double Room	Cancellation fees
433	HOTEL MAA LAXMI PALACE	Deoghar	3699.0	5	Excellent	8.6	Double or Twin Room	Free cancellation

434 rows x 8 columns

Total Rows : 434 , Total Columns : 8

Univariate Analysis

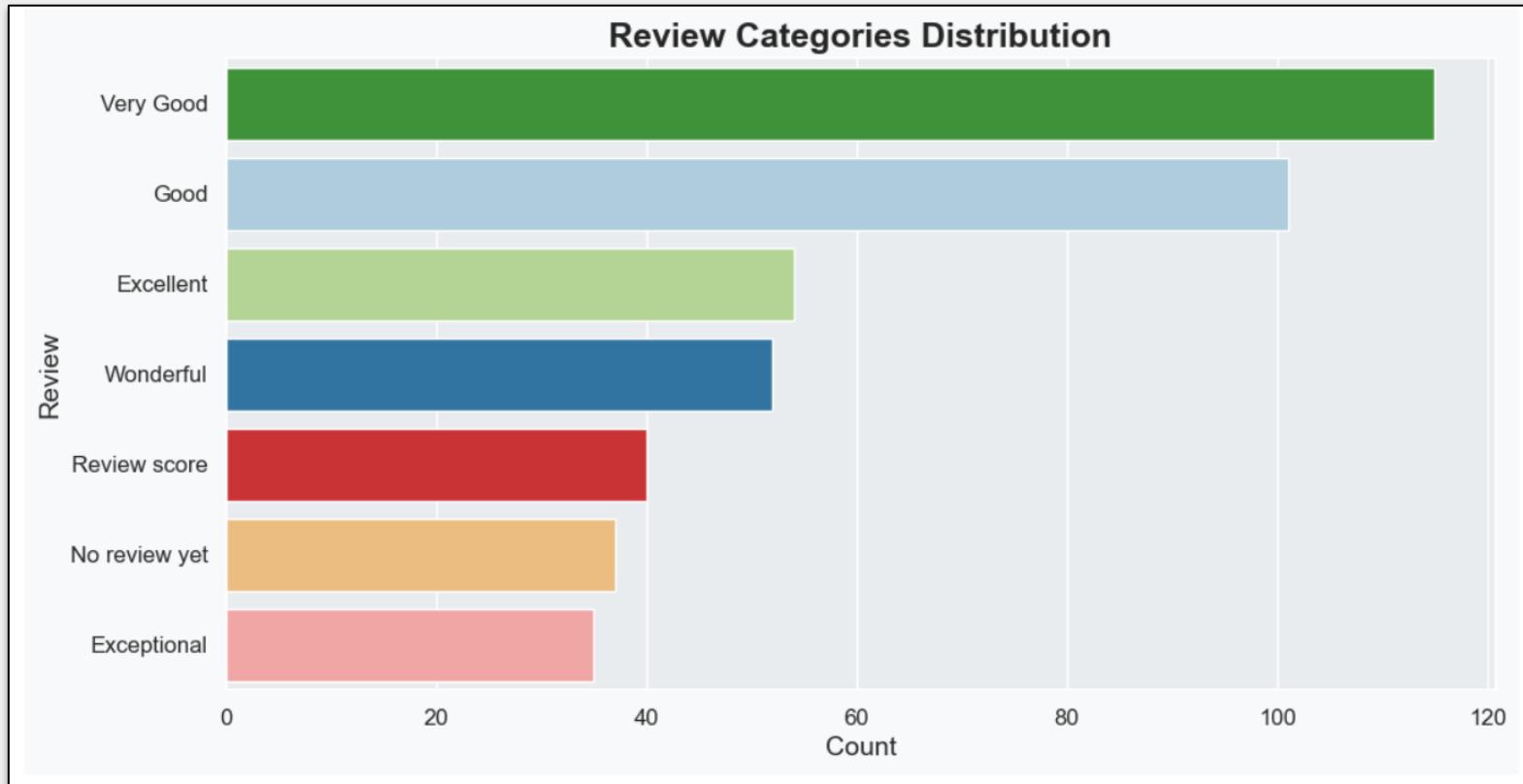


INSIGHTS :

Price Analysis (Boxplot & Distribution):

- Most hotel prices are concentrated below ₹5000, with a long tail of higher prices.
- There are many **outliers above ₹10,000**, showing a few luxury hotels skew the distribution.
- The price distribution is **right-skewed**, meaning most hotels are affordable, but a few very expensive ones pull the average upward.

Univariate Analysis

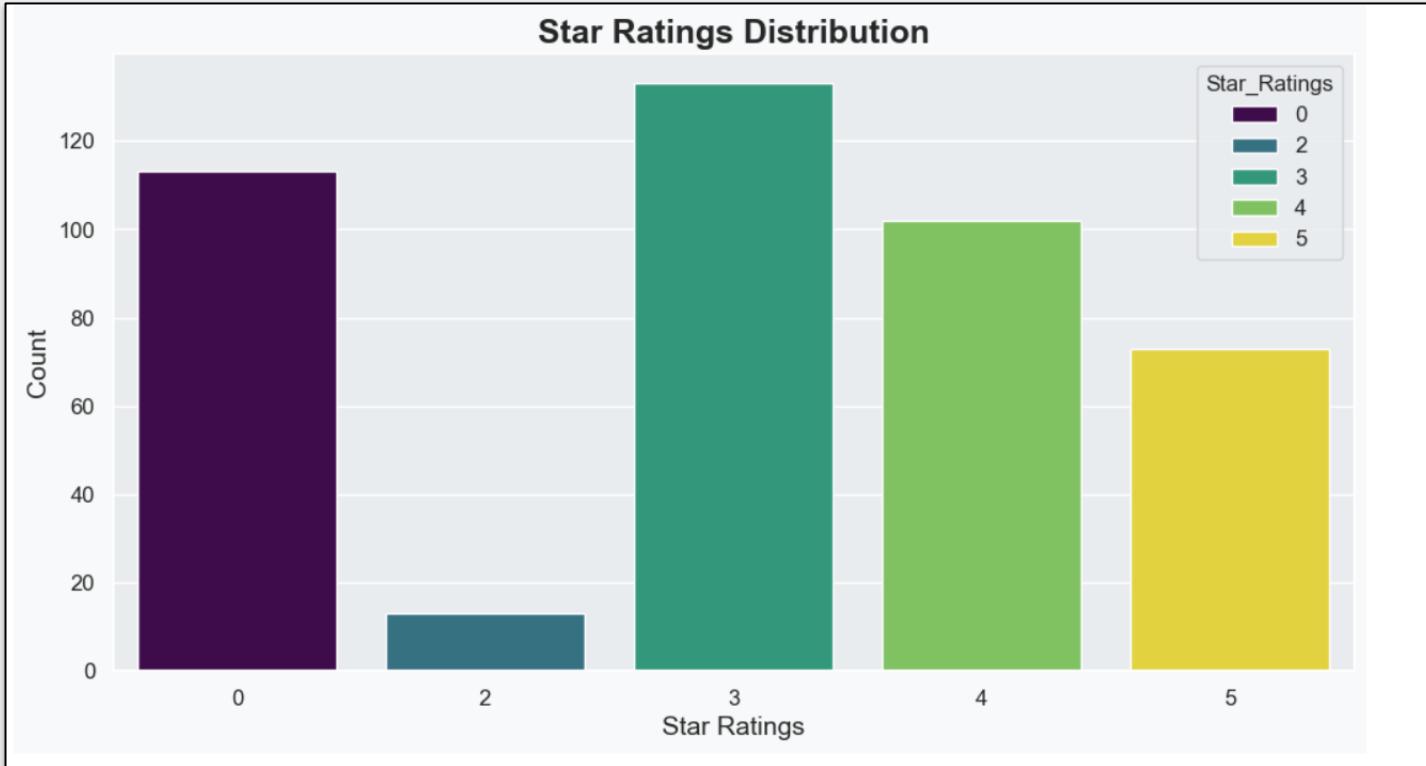


INSIGHTS :

Review Categories:

- The majority of reviews fall into "**Very Good**" and "**Good**", suggesting customer satisfaction is generally positive.
- Fewer hotels are rated "**Exceptional**", meaning only a small portion achieves the highest review standard.

Univariate Analysis



INSIGHTS :

Star Ratings:

- Most hotels are **3-star and 4-star**, with very few 2-star hotels.
- A notable number of hotels have **0 stars (not rated)**, which could be new or unrated listings.
- 5-star hotels are present but fewer compared to 3-star and 4-star.

Bivariate Analysis



INSIGHTS :

Price vs Review Score :

- Most hotels, regardless of price, cluster in the review score range of **7–10**, indicating generally positive reviews.
- Higher-priced hotels (especially above ₹15,000) are mostly **4-star and 5-star hotels**, maintaining strong review scores.
- Review scores alone are not a strong determinant of price - many mid and low-priced hotels also have high reviews.

Bivariate Analysis

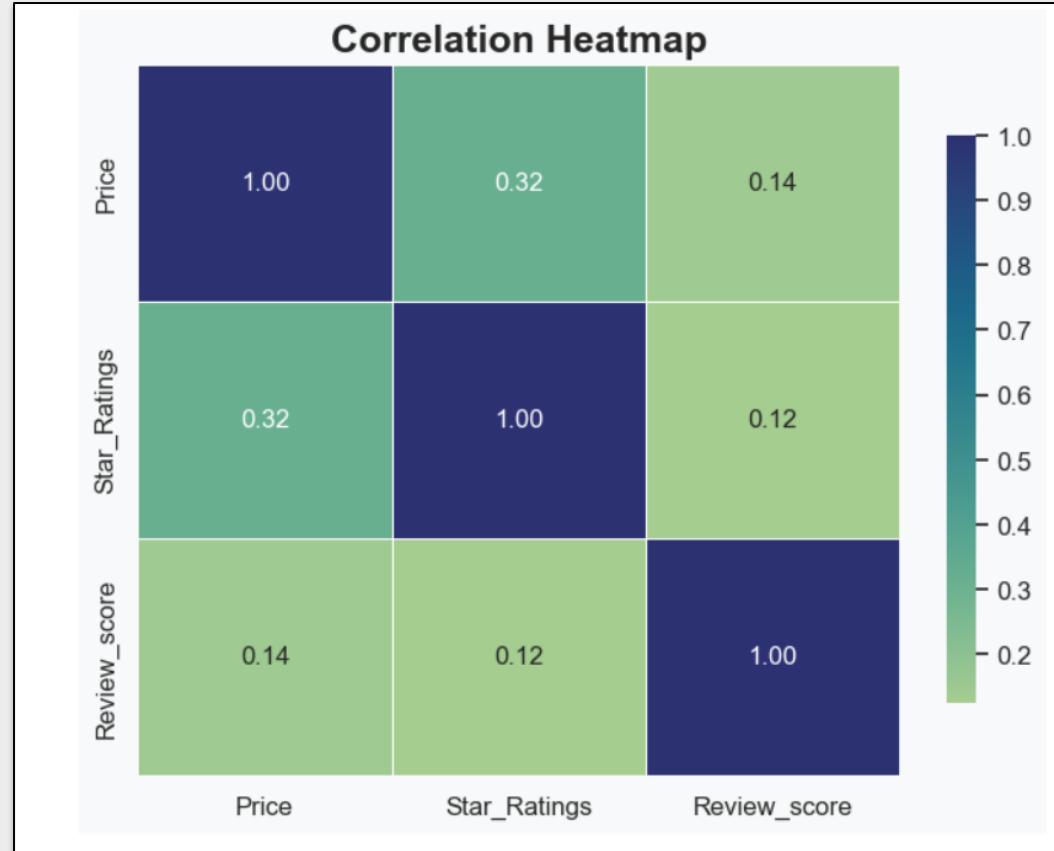


INSIGHTS :

Price vs Star Ratings :

- There is a clear positive relationship between star ratings and price, higher-rated hotels charge more on average.
- 5-star hotels have the widest price range, with significant outliers at very high price levels (above ₹30,000).
- Lower-star hotels (0–2 stars) cluster tightly around lower price ranges, showing less variation.

Multivariate Analysis

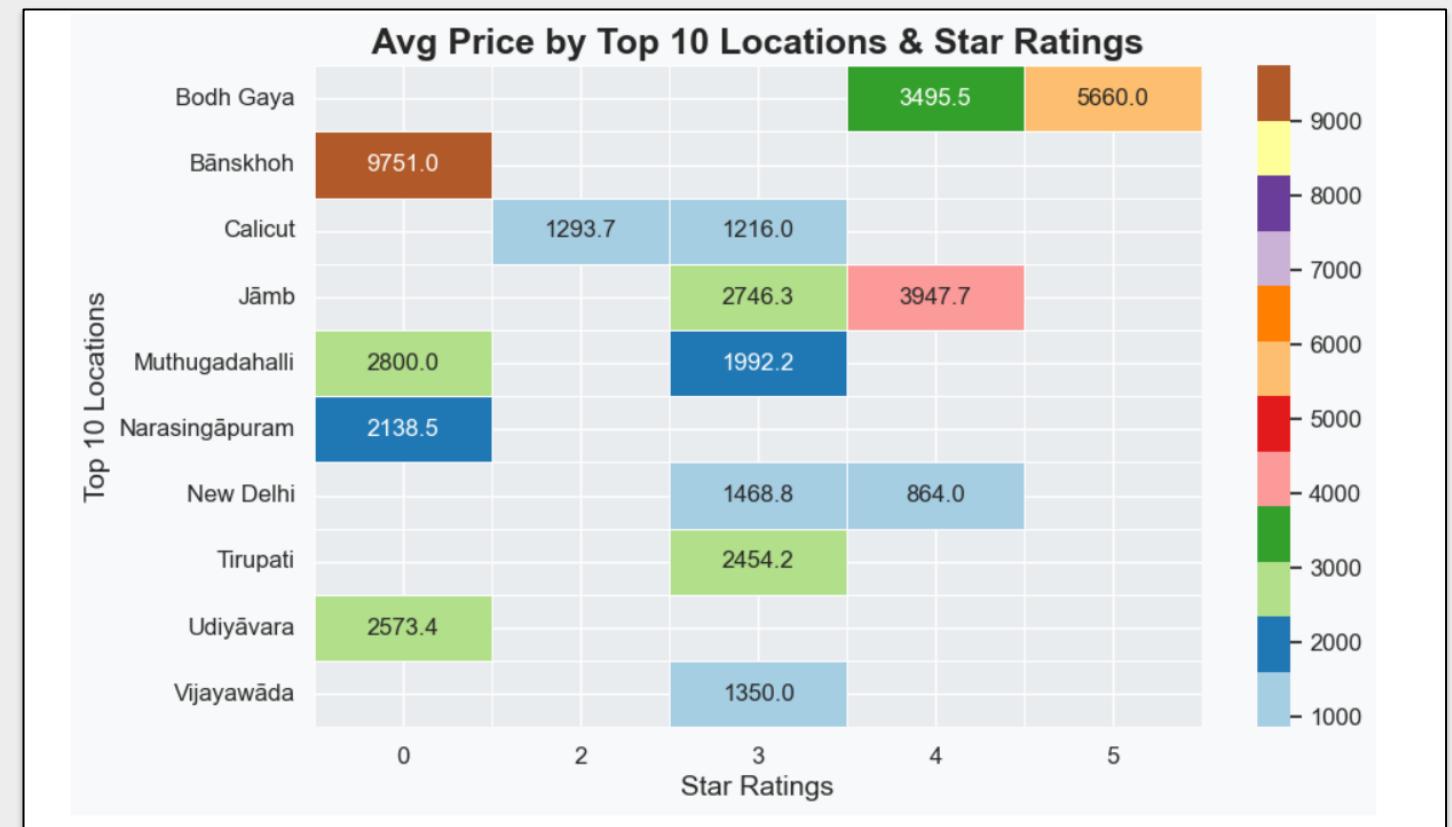


INSIGHTS :

Correlation Heatmap :

- Price & Star Ratings show a moderate positive correlation (0.32) → higher star-rated hotels generally charge higher prices.
- Price & Review Score have only a weak correlation (0.14) → good reviews do not necessarily mean higher prices.
- Star Ratings & Review Score also have a very weak correlation (0.12) → customers give good ratings across all hotel categories, not only premium ones.

Multivariate Analysis



INSIGHTS :

Avg Price by Top 10 Locations & Star Ratings :

- Location strongly impacts pricing :** Cities like Bānskhoh (₹ 9751) and Bodh Gaya (₹ 5660) have much higher average hotel prices compared to places like New Delhi (₹ 864–1469) and Calicut (₹ 1200).
- Star Ratings and Location combined effect :** In cities like Jāmb, prices jump from ₹ 2700 (3-star) to ₹ 3900 (4-star), showing a clear premium for higher category hotels.
- Budget-friendly clusters :** Locations like New Delhi, Vijayawada and Calicut offer lower average prices even for higher star ratings, indicating competitive pricing markets.

Conclusion

- Hotel pricing is influenced more by star ratings and location than by customer review scores.
- Star ratings show a clear positive relationship with price, with 5-star hotels commanding the widest and highest price ranges.
- Review scores are generally high across all hotels, meaning they act more as a baseline expectation than a price driver.
- Location plays a crucial role in price variations in some cities (e.g. Bānskhoh, Bodh Gaya) have significantly higher average prices, while others (e.g. New Delhi, Calicut) remain competitive and budget-friendly.

Key Business Question

1. What factors drive hotel pricing the most → location, star rating, or amenities?
2. How do prices vary across different destinations?
3. Do higher review scores significantly impact hotel prices, or are they just baseline expectations?
4. Which amenities add the greatest value to pricing (e.g., pool, breakfast, parking)?

Experience & Challenges

- **Web Scraping Issues**
 - Encountered **CAPTCHAs** and anti-scraping measures that slowed data collection.
 - Had to adapt to **inconsistent HTML structures** across different hotel pages.
- **Data Cleaning Challenges**
 - Hotel names and room descriptions were often **noisy, duplicated, or unstructured**, requiring extensive preprocessing.
 - Normalizing attributes (like amenities, star ratings, prices) to maintain consistency.
- **Handling Large Dataset**
 - Managed and processed a dataset of **62,361 rows**, which required optimizing code for efficiency.
 - Used **batch processing and panda's optimizations** to ensure smooth analysis.



THANK YOU

