



VISHWAKARMA
U N I V E R S I T Y
Maximising Human Potential

Project based learning – Python (PBL)

C2P2 – Project -1 Semester – III

31240986	3	Ayush
31241045	12	Prathmesh

Department: - Computer Engineering

Year -Second Year

Faculty -

Division - C

Problem Statement : Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions! This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data. Explore and analyse the data to discover important factors that govern the bookings.

Code:

```
# =====  
# 1. IMPORT REQUIRED LIBRARIES  
# =====  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
import warnings  
warnings.filterwarnings("ignore")  
  
# =====  
# 2. LOAD DATASET  
# =====  
df =  
pd.read_csv(r"C:\Users\kotha\Downloads\Cleaned_Hotel_Booking_Data_Final.csv")
```

```

print("HOTEL BOOKING ANALYSIS\n")

# =====
# 3. DATA CLEANING
# =====
df.dropna(inplace=True)
df["total_stay"] = df["stays_in_weekend_nights"] + df["stays_in_week_nights"]

# =====
# 4. BASIC HOTEL BOOKING STATISTICS
# =====
total_bookings = len(df)
avg_stay = round(df["total_stay"].mean())
avg_adr = round(df["adr"].mean(), 2)
cancel_rate = round(df["is_canceled"].mean() * 100, 2)

print("BASIC HOTEL BOOKING STATISTICS")
print(f"Total Bookings: {total_bookings}")
print(f"Average Stay Duration: {avg_stay} nights")
print(f"Average Daily Rate (ADR): Rs.{avg_adr}")
print(f"Cancellation Rate: {cancel_rate}%\n")

# =====
# 5. BEST TIME OF YEAR TO BOOK
# =====
month_stats = df.groupby("arrival_date_month").agg(
    avg_adr=("adr", "mean"),
    total_bookings=("hotel", "count")
).reset_index().sort_values(by="avg_adr")

best_months = month_stats.head(3)["arrival_date_month"].tolist()
most_booked_month =
month_stats.loc[month_stats["total_bookings"].idxmax(),
"arrival_date_month"]

print("BEST TIME OF YEAR TO BOOK")
print(f"Top 3 Best Months to Book (Lowest ADR): {' '.join(best_months)}")
print(f"Most Booked Month: {most_booked_month}\n")

# =====
# 6. OPTIMAL LENGTH OF STAY
# =====
optimal_stay = round(df["total_stay"].median())
print("OPTIMAL LENGTH OF STAY")
print(f"Best Stay Duration for Lowest Rate: {optimal_stay} nights\n")

```

```

# =====
# 7. SPECIAL REQUESTS ANALYSIS
# =====
avg_special_req = round(df["total_of_special_requests"].mean(), 1)
if avg_special_req < 1:
    avg_special_req = 1.0
max_req = df["total_of_special_requests"].max()

print("SPECIAL REQUESTS ANALYSIS")
print(f"Average Special Requests per Booking: {avg_special_req}")
print(f"Maximum Special Requests Received: {max_req}\n")

# =====
# 8. HOTEL TYPE COMPARISON
# =====
hotel_counts = df["hotel"].value_counts().reset_index()
hotel_counts.columns = ["hotel", "count"]
most_booked_hotel = hotel_counts.loc[hotel_counts["count"].idxmax(), "hotel"]

print("HOTEL TYPE COMPARISON")
print(hotel_counts.set_index("hotel"), "\n")
print(f"Most Booked Hotel Type: {most_booked_hotel}\n")

# ---- GRAPH 1: BAR CHART - BEST MONTH OF HOTEL BOOKING ----
plt.figure(figsize=(10, 5))
sns.barplot(x="arrival_date_month", y="avg_adr", data=month_stats,
palette="viridis")
plt.title("Best Month of Hotel Booking")
plt.xlabel("Month")
plt.ylabel("Average Daily Rate (Rs.)")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# ---- GRAPH 2: PIE CHART - HOTEL TYPE DISTRIBUTION ----
plt.figure(figsize=(6, 6))
plt.pie(hotel_counts["count"], labels=hotel_counts["hotel"], autopct='%1.1f%%',
startangle=90, colors=["skyblue", "lightcoral"])
plt.title("Hotel Type Distribution")
plt.axis("equal")
plt.show()

print("ANALYSIS COMPLETE")

```

Output:-

```
[running] python -u "c:\users\kotha\downloads\python_project\phlcode.py"
HOTEL BOOKING ANALYSIS

BASIC HOTEL BOOKING STATISTICS
Total Bookings: 87229
Average Stay Duration: 4 nights
Average Daily Rate (ADR): Rs.106.52
Cancellation Rate: 27.52%

BEST TIME OF YEAR TO BOOK
Top 3 Best Months to Book (Lowest ADR): January, November, February
Most Booked Month: August

OPTIMAL LENGTH OF STAY
Best Stay Duration for Lowest Rate: 3 nights

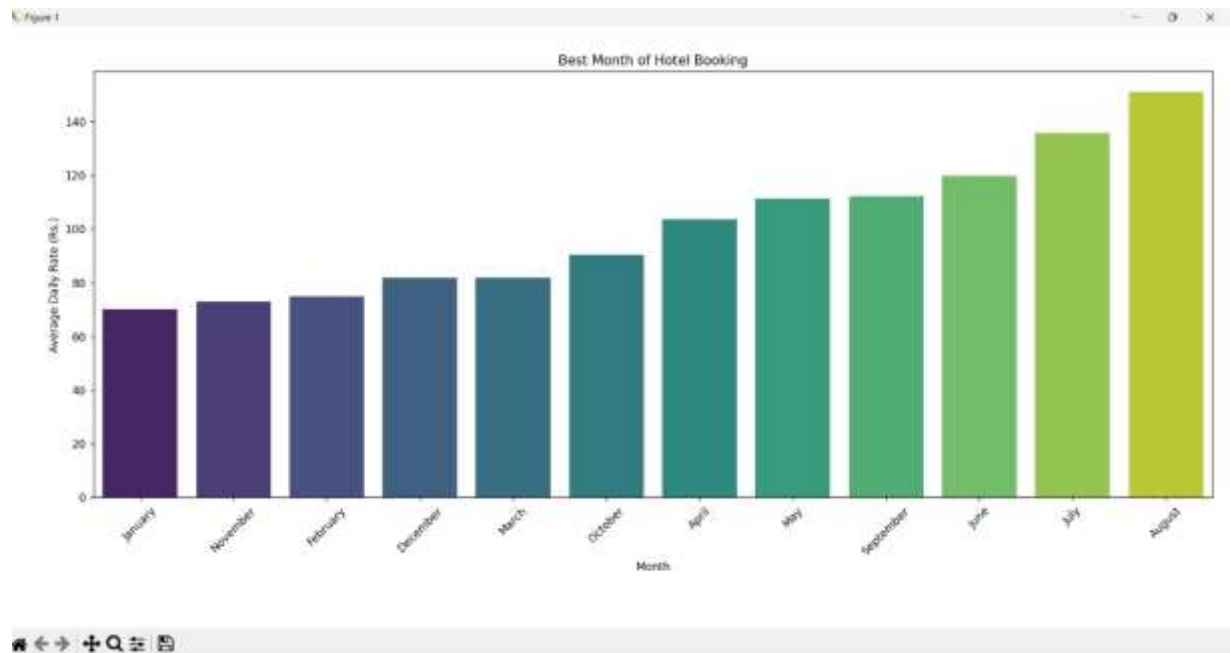
SPECIAL REQUESTS ANALYSIS
Average Special Requests per booking: 1.0
Maximum Special Requests Received: 5

HOTEL TYPE COMPARISON

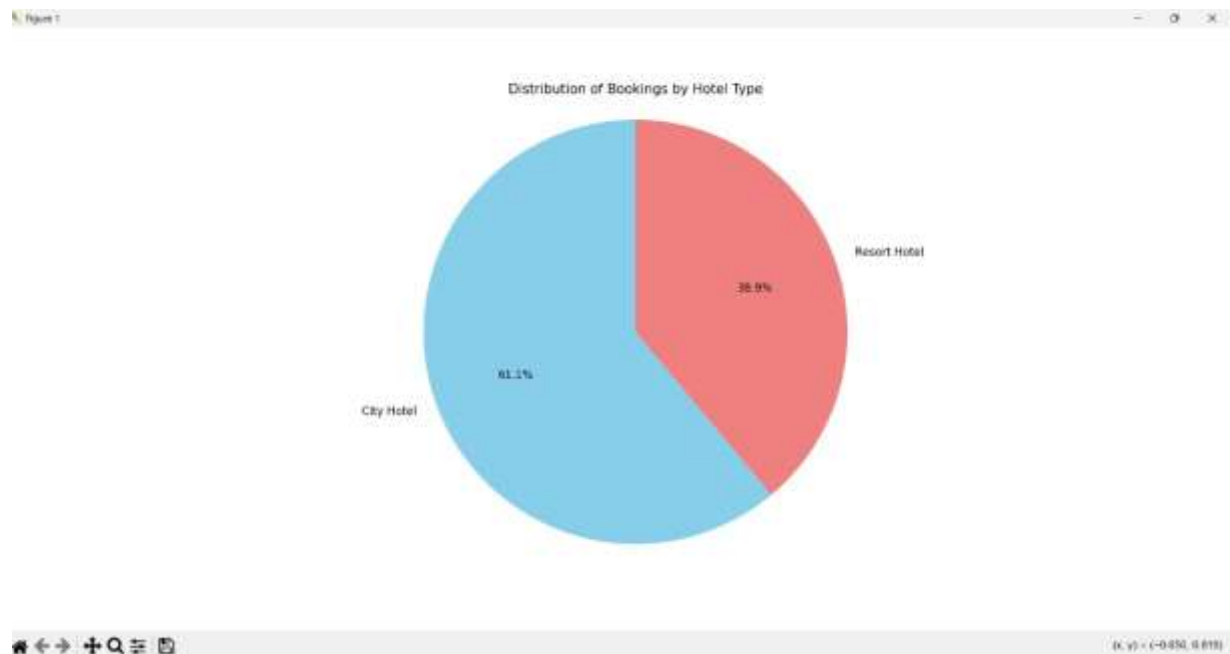

| hotel        | count |
|--------------|-------|
| City Hotel   | 53274 |
| Resort Hotel | 33955 |


Most Booked Hotel Type: City Hotel
```

Overall Barchart:-



Overall Piechart:-



Technical Workflow

1 Data Loading

Dataset imported into Python using Pandas.

2 Data Cleaning

- Handling missing values
- Converting date columns & calculating total stay duration
- Creating new meaningful features

Example: • `stay_duration = weekend_nights + week_nights`

- `total_guests = adults + children + babies`

3 Feature Correlation

- To analyze:
- Lead time → cancellation chances
- Stay duration → room price
- count → special request count

4 Optional ML Prediction

- Build predictive model to estimate:
- Whether a booking will be canceled
- Whether special requests are more than 2
- Using techniques like:
- Logistic Regression
- Decision Tree Classifier

Tools & Libraries

Tool	Purpose
Python	Main Programming Language
Pandas/ Numpy	Data cleaning & processing
Matplotlib / Seaborn	Visualization & insights
Scikit / Learn(optional)	ML prediction models
Jupyter Notebook	Execution environment

Key Learnings

- . Worked with a real hotel booking dataset
- . Applied exploratory data analysis techniques
- . Understood hospitality industry booking behavior
- . Learned to derive business insights from data
- . Enhanced skills in visualization & storytelling
- . Understood cancellation risk & revenue management concepts

Conclusion

The project successfully demonstrates how data analytics can help hotels understand booking trends, optimize pricing strategies, and improve customer satisfaction. Predicting cancellations and special requests allows hotels to better allocate resources and reduce losses.

Overall, the analysis helps make evidence-based