

Association Rule Mining & Customer Segmentation in Hospitality Analytics

Dataset: synthetic_hospitality_dataset.csv

1. Objective

The objective of this assignment is to analyze hospitality booking data using two major techniques:

1. **Association Rule Mining** – To discover hidden patterns, relationships, and frequently co-occurring features in the dataset.
2. **Customer Segmentation** – To cluster customers into meaningful groups based on their booking patterns, expenditure, and preferences.

This exercise will enable students to understand how **data-driven insights** can improve **marketing strategies**, **dynamic pricing**, and **personalized services** in the hospitality industry.

2. Business Context

In today's competitive hospitality market, hotels must optimize pricing, understand customer behavior, and offer tailored services. With the increasing demand for **personalized travel experiences**, hospitality companies rely on **data analytics** to:

- Identify **high-value customers**
- Predict **cancellations**
- Optimize **cross-selling of services**
- Understand **room type and package preferences**

Analyzing booking patterns and customer behaviors allows companies to **improve operational efficiency** and **increase revenue**.

3. Dataset Description

The dataset **synthetic_hospitality_dataset.csv** contains **500 rows** representing hotel booking information.

Columns:

Column Name	Description
Booking_ID	Unique ID for each booking
Customer_ID	Unique customer identifier
Check_In_Date	Check-in date
Check_Out_Date	Check-out date
Room_Type	Type of room booked (e.g., Standard, Deluxe, Suite)
Meal_Plan	Selected meal plan (e.g., Breakfast, Half Board)
Num_Adults	Number of adults in the booking
Num_Children	Number of children in the booking
Booking_Channel	Booking source (e.g., Online, Travel Agent, Direct)
Total_Nights	Number of nights stayed
Total_Cost	Total booking cost
Is_Canceled	Indicates if the booking was canceled (0 = No, 1 = Yes)
Extra_Spa	Whether spa services were purchased
Extra_Airport_Pickup	Whether airport pickup was requested
Extra_City_Tour	Whether city tour was booked
Extra_Sea_View	Whether a sea-view upgrade was requested
Extra_Late_Checkout	Whether late checkout was requested

4. Methodologies

A. Association Rule Mining

Objective:

To find frequent itemsets and identify rules showing the relationships between services, packages, and room preferences.

Example Questions to Answer:

- Do customers booking a **Suite** often request **Sea View** and **Spa** services together?
- Are **Online Booking customers** more likely to add **City Tours**?
- Which **meal plans** are typically booked with **long stays**?

Techniques Used:

- Use the **Apriori Algorithm** or **FP-Growth** to find frequent patterns.
- Use **support, confidence, and lift** to measure rule strength.

Expected Outcome:

- A list of top **association rules** that show relationships between booking behaviors.
- Insights for **cross-selling opportunities** and **service bundling**.

B. Customer Segmentation

Objective:

To cluster customers based on their **spending behavior, booking preferences, and service usage**.

Steps:

1. Select numerical features like **Total_Nights**, **Total_Cost**, **Num_Adults**, and **Num_Children**.
2. Normalize the data for better clustering performance.
3. Apply **K-Means Clustering** to segment customers into groups.
4. Use the **Elbow Method** to choose the optimal number of clusters.

Example Insights:

- **Cluster 1:** Budget travelers → Standard rooms, short stays, fewer extras.
- **Cluster 2:** Family vacationers → Multiple children, long stays, city tours.
- **Cluster 3:** Luxury seekers → Suites, high spending, spa & sea view add-ons.

Expected Outcome:

- Visualized customer clusters using **scatter plots** and **boxplots**.

- Identification of **high-value customer segments** for targeted marketing.

5. Tools & Libraries Required

- **Python 3.x**
- **Pandas** → Data cleaning & manipulation
- **Matplotlib / Seaborn** → Data visualization
- **mlxtend** → Association Rule Mining
- **scikit-learn** → Customer segmentation using K-Means
- **Jupyter Notebook** → For interactive analysis

6. Assignment Tasks

Task 1 — Data Understanding & Preprocessing

- Load the dataset.
- Handle missing values.
- Check data distributions and correlations.
- Visualize room preferences, booking channels, and extra services.

Task 2 — Association Rule Mining

- Convert categorical features into transactional format.
- Apply the **Apriori Algorithm**.
- Generate association rules using support, confidence, and lift.
- Interpret top 5 strongest rules.

Task 3 — Customer Segmentation

- Select relevant numerical and categorical features.
- Standardize the data.
- Use the **Elbow Method** to determine optimal clusters.
- Apply K-Means clustering.

- Visualize the clusters and interpret findings.

Task 4 — Reporting

- Summarize the insights discovered.
- Highlight patterns, customer behaviors, and opportunities for upselling.
- Suggest **marketing strategies** based on findings.

7. Expected Deliverables

- A Jupyter Notebook (hospitality_analysis.ipynb) with:
 - Data preprocessing
 - Visualizations
 - Association rule results
 - Segmentation plots
- A one-page **business insights report** summarizing the findings.
- A CSV or Excel file if new customer segments are created.