HOTEL RECOMMENDATION SYSTEM USING MACHINE LEARNING WITH COMMISION BASED REVENUE MODEL

Sameer

October 28, 2023

Abstract

In this prototype report, a Hotel Recommendation System with a specific focus on its intricate technical model and user-centric features is presented. The system includes user registration, enabling personalized experiences, and showcases a user-friendly interface for viewing comprehensive hotel information. The heart of the prototype lies in the recommendation algorithm, which employs a hybrid approach, combining content-based and collaborative filtering techniques. This unique model processes a rich dataset that incorporates user preferences, hotel attributes, historical bookings, and reviews. It continuously adapts and learns from user interactions, ensuring that recommendations remain relevant.

Furthermore, the report delves into the performance evaluation, emphasizing the use of industry-standard metrics like Precision at K (P@K), Recall at K (R@K), and Mean Average Precision (MAP) to assess the system's recommendation quality. This prototype serves as a vital step in developing a user-centric platform concept that effectively combines data-driven recommendations with commission-based revenue generation.

1.0 Introduction

In today's fast-paced world of travel, one of the most common dilemmas travelers face is the overwhelming abundance of hotel options. It's like standing in front of an all-we-can-eat buffet with so many delicious choices that it's hard to decide where to start. The sheer number of hotels can lead to a problem known as "decision overload," where travelers often end up making subpar hotel choices due to sheer exhaustion from the selection process. This challenge is what we're determined to address with our prototype for a Hotel Recommendation System. Our goal is to simplify the hotel booking process and provide personalized recommendations so we can find the perfect place to stay without all the stress. In this report, we'll introduce our solution and discuss how it can make wer travel planning easier while also explaining our commission-based revenue model to keep this service running smoothly.

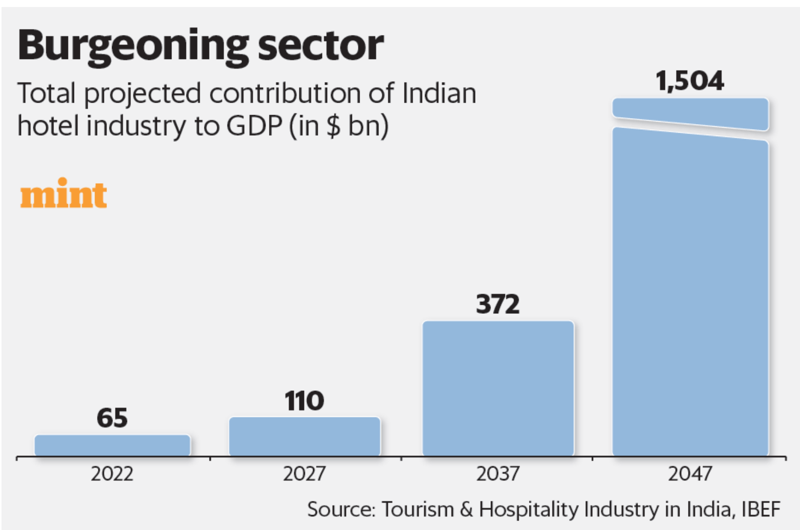
1.1 Problem Statement

This project seeks to build a machine learning-based hotel recommendation system, addressing the challenge of choice overload faced by travelers. The system will provide personalized recommendations, simplifying the hotel selection process, while a commission-based monetization model will ensure its sustainability and financial viability.

2.0 Business Need Assessment for Hotel Recommendation System

With a projected growth rate of 9.7% from 2023 to 2029, the Indian hotel market is set to expand substantially, reaching nearly USD 56.21 million. Furthermore, the hotels market is anticipated to generate USD 7.66 billion in revenue by 2023, with a notable compound annual growth rate (CAGR) of 8.28% from 2023 to 2027, indicative of the industry's robust prospects.

In parallel, the hospitality sector in India is poised for impressive growth, with an expected value of USD 23.5 billion in 2023. It is projected to maintain a healthy CAGR of 4.73%, eventually reaching USD 29.61 billion by 2028. These statistics demonstrate a burgeoning industry with significant economic potential, providing an ideal environment for the development and implementation of innovative solutions like a Hotel Recommendation System, poised to cater to the evolving needs and preferences of both domestic and international travelers.



3.0 External Search (Information Sources)

Selection of a suitable hotel location and reservation of accommodation have become a critical issue for the travelers. The online hotel search has been increased at a very fast pace and became very time-consuming due to the presence of huge amount of online information. Recommender systems (RSs) are getting importance due to their significance in making decisions and providing detailed information about the required product or a service. Hotel recommendation systems leverage vast amounts of data, including user preferences, hotel features, historical bookings, and user reviews.

Machine learning algorithms analyze this data to provide personalized recommendations to users. Machine learning models create user profiles based on their past interactions and preferences. These profiles are used to tailor hotel recommendations to individual users, ensuring a more personalized experience.

Content-based filtering considers the characteristics of both hotels and users. It matches user preferences with hotel attributes like location, price, amenities, and reviews to suggest the most suitable options.

Collaborative filtering identifies users with similar preferences and recommends hotels based on the choices of users with comparable tastes. This method enhances recommendations by considering the wisdom of the crowd.

Many recommendation systems use hybrid models that combine both content-based and collaborative filtering approaches. This combination aims to provide a more comprehensive and accurate recommendation.

3.1 Applicable Regulations

a.) Personal Data Protection Bill (PDPB): India is in the process of implementing the PDPB, which is expected to govern the collection, storage, and processing of personal data. It will require user consent, data protection measures, and data localization, among other provisions.

b.) Information Technology Act, 2000: This act addresses various aspects of electronic commerce and digital signatures. It includes provisions related to data protection, privacy, and cybersecurity.

c.) Payment Card Industry Data Security Standard (PCI DSS): If wer system handles payment information, it must comply with PCI DSS to ensure secure payment processing.

d.) E-commerce Regulations: Various e-commerce regulations may apply, especially if the platform enables hotel bookings and payments. Compliance with pricing transparency, consumer protection, and dispute resolution rules is essential.

e.) User Consent and Data Portability: Ensure that users are informed about data collection and processing practices. Provide options for users to control their data and request data portability.

3.2 Business opportunity

Travelers are increasingly seeking personalized travel services and recommendations. The market demand for such services is on the rise, making this an opportune time to enter the space.

By providing users with tailored recommendations that align with their preferences, the system can boost hotel bookings through the platform. This leads to increased revenue potential.

Offering a sophisticated recommendation system can give wer platform a competitive edge in the crowded travel and booking industry. It can differentiate wer platform from competitors.

Machine learning models can be scaled to handle increasing user demands and data volumes as the business grows.

The model can be extended to other travel-related services, such as restaurant recommendations, activity bookings, and transportation options.

4.0 Concept Generation

The concept of a hotel recommendation model using machine learning emerged from a recognition of the need for personalized, efficient, and data-driven solutions in the hospitality and travel industry. This concept was motivated by the following factual factors:

Market Demand: Extensive research and market analysis revealed a growing demand for personalized travel services. Travelers increasingly seek tailored recommendations for hotels that align with their preferences, making the concept of a recommendation system more relevant and market-driven.

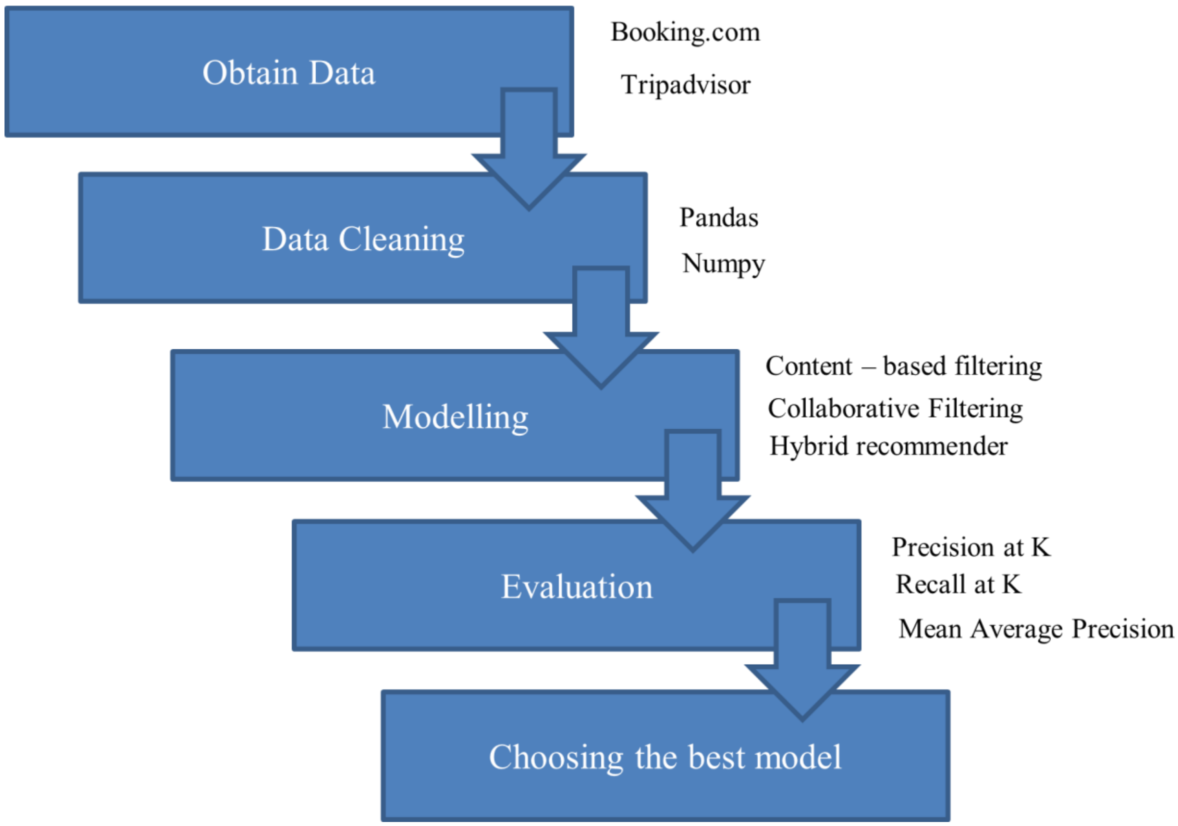
Advancements in Machine Learning: The rapid advancements in machine learning and artificial intelligence offered an opportunity to leverage these technologies for enhancing the travel and hospitality experience. Machine learning's ability to process vast datasets and discern intricate patterns was a key driver.

Data Availability: The availability of diverse and extensive data sources, such as user behavior, hotel attributes, and reviews, presented an opportunity to develop a model that could harness this data to generate valuable recommendations.

Parallel Trends: Observing the success of personalized, predictive models in other industries, such as healthcare, reinforced the notion that a similar approach could benefit the hotel and hospitality sector. This fact-based insight underscored the timeliness and relevance of the concept.

In essence, the concept of a hotel recommendation model using machine learning was driven by market demand, technological advancements, user-centric goals, data availability, and the alignment with trends in personalized and predictive services. These factual factors converged to inspire the development of a solution that enhances the traveler's hotel selection process.

4.1 Concept Development



Data Acquisition: Inititally we start by collecting data from multiple sources, including Booking.com and TripAdvisor. This data likely contains information about hotels, user interactions, reviews, and other relevant details. Combining data from various sources can enrich the recommendation system with a diverse set of information.

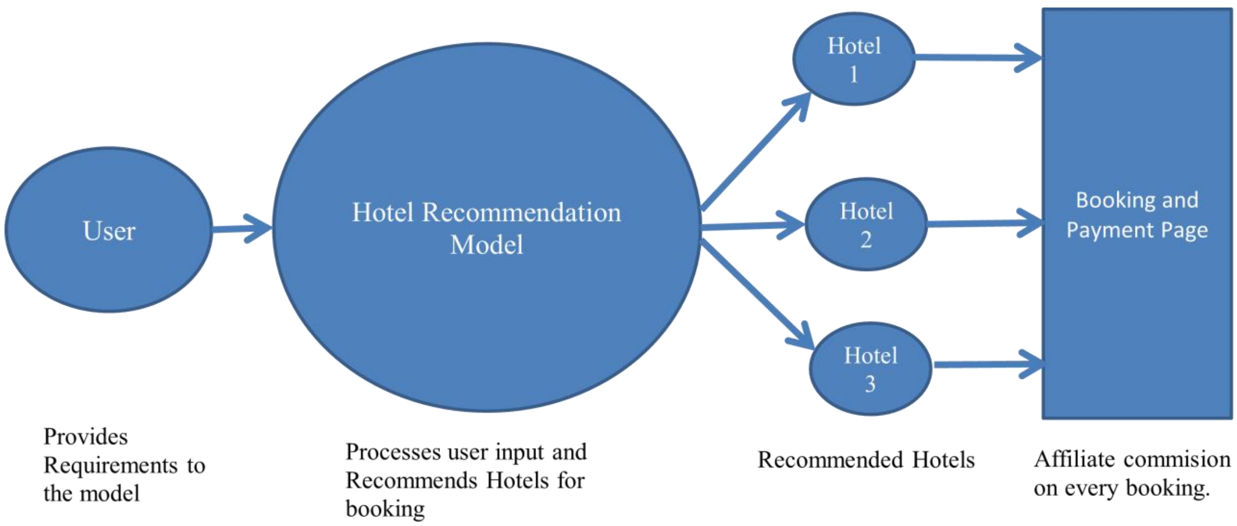
Data Cleaning and Preprocessing: Data preprocessing is crucial to ensure the quality and consistency of the data. In this step, we clean the data by handling missing values, removing duplicates, and addressing any inconsistencies. We also transform and structure the data to make it suitable for modeling. This might include text processing, numerical scaling, and creating user-item interaction matrices.

Modeling: We employ three recommendation techniques: content-based filtering, collaborative filtering, and a hybrid recommender. Content-based filtering leverages hotel characteristics to make recommendations, while collaborative filtering uses user interactions to identify similar users and items. The hybrid model combines the strengths of both approaches to provide more accurate and diverse recommendations. Modeling involves building and training machine learning models based on these techniques.

Evaluation: To assess the performance of wer recommendation models, we employ evaluation metrics such as Precision at K (P@K), Recall at K (R@K), and Mean Average Precision (MAP). These metrics help measure the accuracy, coverage, and quality of wer recommendations. By evaluating the models using these metrics, we gain insights into their strengths and weaknesses.

Model Selection: Based on the evaluation results, we choose the best-performing model. The model that excels in terms of P@K, R@K, and MAP may vary depending on wer specific goals and the characteristics of wer data. The chosen model should strike a balance between precision, recall, and overall recommendation quality.

4.2 Revenue Model:



User Interaction and Hotel Selection: Users visit the website and provide their hotel requirements, preferences, and constraints, such as location, price range, amenities, and travel dates. The recommendation system processes this input and generates a list of the most suitable hotels based on the user's criteria.

Hotel Booking: Users review the recommended hotels, read descriptions, view photos, and check reviews and ratings. Users select a hotel from the recommendations and proceed to the booking and reservation process.

Affiliate Commission: When a user successfully books a hotel through the website's booking platform, an affiliate commission is earned. The commission is a percentage of the total booking value and is typically paid by the hotel or a third-party booking service.

Revenue Generation: Revenue is generated based on the total affiliate commissions earned through successful hotel bookings made by users via the website.

5.0 Final Report Prototype

The product takes the following functions to perfect and provide a good result.

Back-end

Model Development: This must be done before releasing the service. A lot of manual supervised machine learning must be performed to optimize the automated tasks.

1. Performing EDA to realize the dependent and independent features.

2. Algorithm training and optimization must be done to minimize overfitting of the model.

Front End

1. Different user interface: The user must be given many options to choose form in terms of parameters. This can only be optimized after a lot of testing and analysis all the edge cases.

2. Interactive visualization: The data extracted from the trained models will return raw and inscrutable data. This must be present in an aesthetic and an "easy to read" style.

3. Feedback system: A valuable feedback system must be developed to understand the customer's needs that have not been met. This will help us train the models constantly.

6.0 Code Implementation/Validation on Small Scale

6.1 Basic EDA

Data Cleaning and transformations

[13]:
```
del hotel_details['id']
del hotel_rooms['id']
del hotel_details['zipcode']
```

[14]:
```
hotel_details=hotel_details.dropna()
hotel_rooms=hotel_rooms.dropna()
```

[15]:
```
hotel_details.drop_duplicates(subset='hotelid',keep=False,inplace=True)
hotel=pd.merge(hotel_rooms,hotel_details,left_on='hotelcode',right_on='hotelid',how='inner')
```

[17]:
```
hotel.columns
```

[17]:
```
Index(['hotelcode', 'roomamenities', 'roomtype', 'ratedescription', 'hotelid',
       'hotelname', 'address', 'city', 'country', 'propertytype', 'starrating',
       'latitude', 'longitude', 'Source', 'url', 'curr'],
```

```
data_types = hotel.dtypes
data_types
```

[42]:
```
hotelcode          0
roomamenities      0
roomtype           0
ratedescription    0
hotelname          0
address            0
city               0
country            0
propertytype       0
starrating         0
latitude           0
longitude          0
dtype: int64
```

```
null_counts = hotel.isnull().sum()
null_counts
```

[37]:
```
hotelcode          0
roomamenities      0
roomtype           0
ratedescription    0
hotelname          0
address            0
city               0
country            0
propertytype       0
starrating         0
latitude           0
longitude          0
dtype: int64
```

6.2 Data after basic EDA

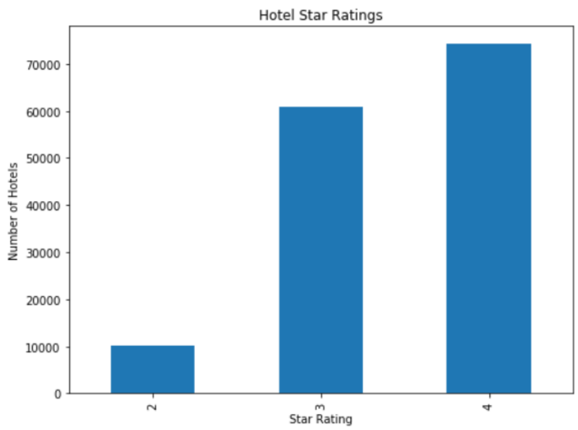
[33]:

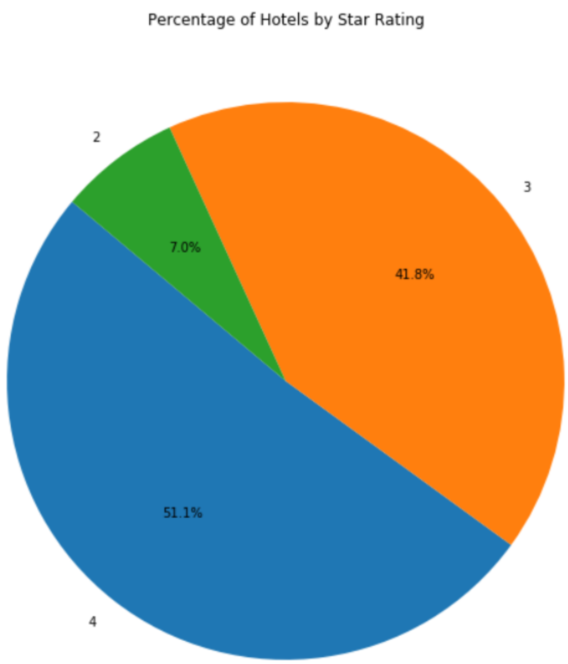
```python
hotel.head()
```

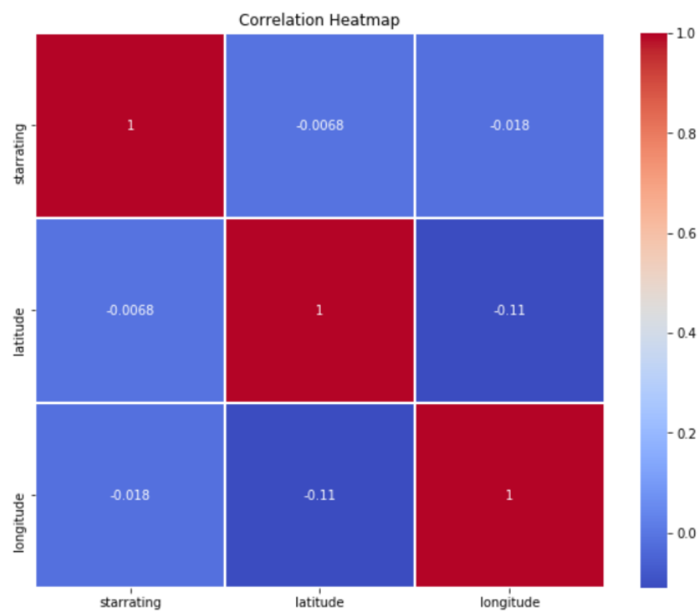
[33]:

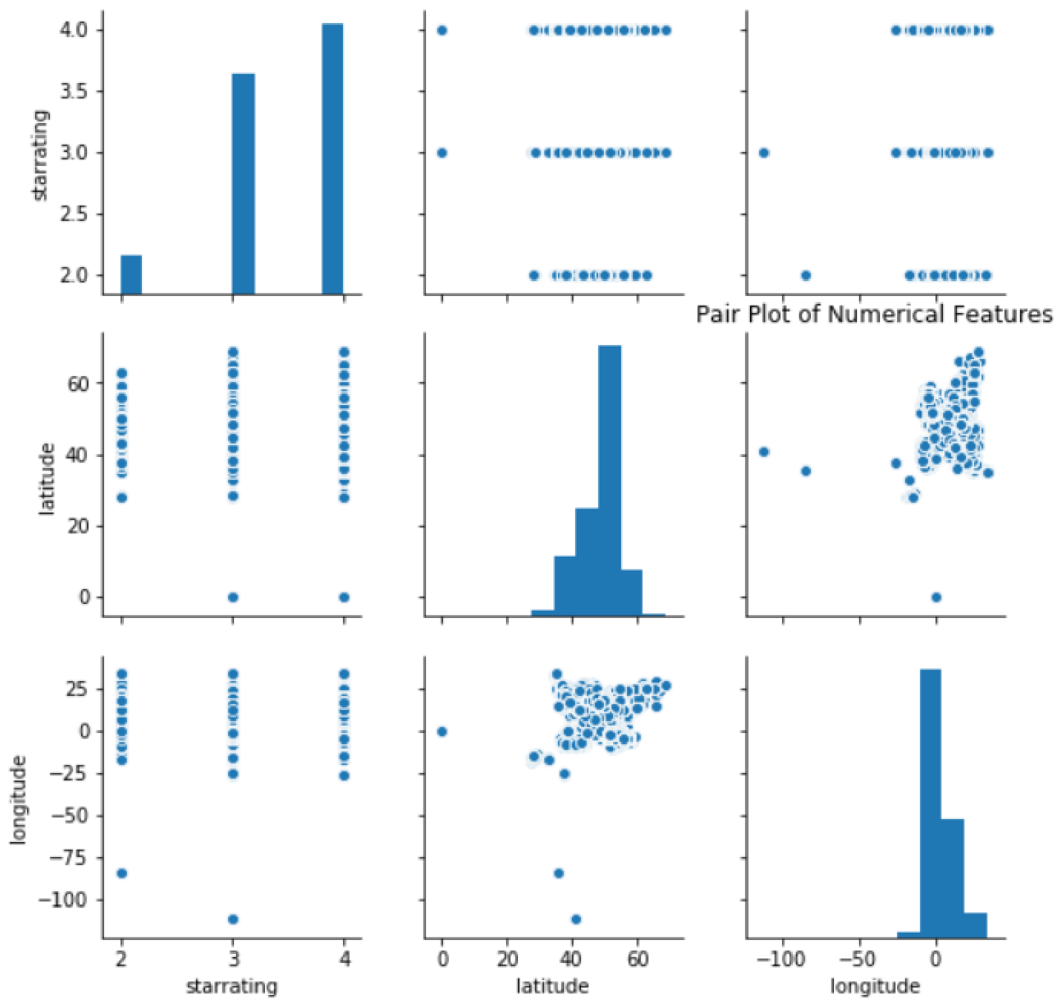
	hotelcode	roomamenities	roomtype	ratedescription	hotelname	address	city	country	propertytype	starrating	latitude	longitude
0	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15 m²/161 ft², Shower, 1 king bed	The Old Cider House	25 Castle Street	Nether Stowey	United Kingdom	Hotels	4	51.150921	-3.15847
1	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15 m²/161 ft², Shower, 1 king bed	The Old Cider House	25 Castle Street	Nether Stowey	United Kingdom	Hotels	4	51.150921	-3.15847
2	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Deluxe Double Room with Shower	Room size: 17 m²/183 ft², Shower, 1 queen bed ...	The Old Cider House	25 Castle Street	Nether Stowey	United Kingdom	Hotels	4	51.150921	-3.15847
3	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Superior Double Room	Room size: 17 m²/183 ft², Shower, 1 double bed	The Old Cider House	25 Castle Street	Nether Stowey	United Kingdom	Hotels	4	51.150921	-3.15847
4	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Standard Double or Twin Room	Room size: 13 m²/140 ft², Shower, 1 queen bed ...	The Old Cider House	25 Castle Street	Nether Stowey	United Kingdom	Hotels	4	51.150921	-3.15847

6.3 Data Visualization






Correlation Heatmap
starrating
latitude
longitude
1
-0.0068
-0.018
-0.0068
1
-0.11
-0.018
-0.11
1
starrating
latitude
longitude
1.0
0.8
0.6
0.4
0.2
0.0


Pair Plot of Numerical Features
starrating
latitude
longitude
4.0
3.5
3.0
2.5
2.0
60
40
20
0
25
0
−25
−50
−75
−100
2
3
4
0
20
40
60
−100
−50
0

6.4 ML Modelling

Recommender system based only on City and ratings about the hotel

[ ]:

```
def citybased(city):
    hotel['city']=hotel['city'].str.lower()
    citybase=hotel[hotel['city']==city.lower()]
    citybase=citybase.sort_values(by='starrating',ascending=False)
    citybase.drop_duplicates(subset='hotelcode',keep='first',inplace=True)
    if(citybase.empty==0):
        hname=citybase[['hotelname','starrating','address','roomamenities','ratedescription']]
        return hname.head()
    else:
        print('No Hotels Available')
```

6.5 Links

Github: [Hotel Recommendation Using machine learning code]

Kaggle Dataset: [Hotel Recommendation Dataset]

7.0 Conclusion

In the world of abundant hotel choices, our Hotel Recommendation System is the key to stress-free bookings, offering tailored suggestions. With our commission-based approach, we ensure both convenience and sustainability in travel planning.

The Indian hotel market is booming, with significant growth expected. This is a great opportunity for our Hotel Recommendation System to serve both local and international travelers with personalized solutions that match their preferences.

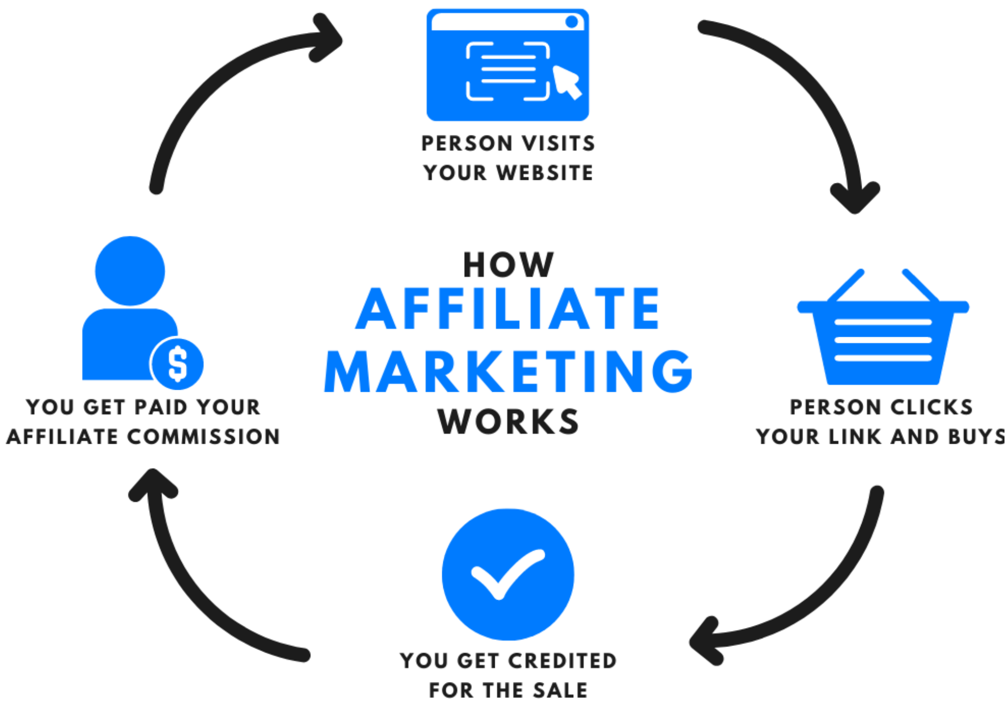
Choosing a hotel can be overwhelming with so many options online. Machine learning helps by giving personalized suggestions based on wer preferences. It makes hotel hunting easier and more enjoyable.

The idea for a hotel recommendation model using machine learning came from recognizing the increasing demand for personalized travel services, advances in technology, and the wealth of available data. These factors converged to create a user-centric solution for easier and more enjoyable hotel choices.

In creating a hotel recommendation system, we gather data from various sources, clean it, build models using content-based, collaborative, and hybrid filtering, and evaluate their performance using metrics like P@K and R@K. The best-performing model is selected based on its ability to provide accurate and diverse recommendations.

Users input their hotel preferences, select a hotel from recommendations, and complete bookings. The website earns revenue through affiliate commissions from successful bookings, ensuring sustainability and user convenience.

BUSINESS MODEL



Affiliate Business Model for Hotel Recommendation system

In the proposed affiliate business model for a hotel recommendation system, user interaction plays a pivotal role in shaping the personalized experience. When users visit the website, they are prompted to input specific hotel requirements, preferences, and constraints, such as preferred location, budgetary considerations, desired amenities, and travel dates. This user-provided information becomes the foundation for the recommendation system's algorithms to process. The recommendation engine then leverages these inputs to generate a curated list of the most suitable hotels, aligning closely with the user's specified criteria.

Upon receiving the list of recommended hotels, users have the opportunity to engage in a comprehensive evaluation process. They can explore detailed hotel descriptions, peruse high-quality photos, and delve into reviews and ratings from fellow travelers. This thorough exploration assists users in making informed decisions when selecting a hotel that best aligns with their preferences. Once a user has made a decision and chosen a hotel from the recommendations, they seamlessly transition to the booking and reservation process.

The revenue generation in this model hinges on the affiliate commission earned through successful hotel bookings facilitated by the website. When a user completes a booking through the platform, an affiliate commission is earned. This commission is typically calculated as a percentage of the total booking value. Importantly, these commissions are paid by the hotel or a third-party booking service, solidifying the affiliate relationship. The more successful bookings that occur through the website, the higher the accumulated affiliate commissions, thereby contributing to the overall revenue generated by the platform.

This revenue generation model is advantageous as it aligns the interests of the hotel recommendation platform with the profitability of the hotels themselves. By facilitating user-friendly interactions, providing comprehensive information, and ensuring a seamless booking process, the platform not only enhances user satisfaction but also maximizes its potential for earning affiliate commissions. Continuous refinement of the recommendation algorithms, strategic partnerships with hotels, and a commitment to user satisfaction are key factors that contribute to the sustained success of this affiliate business model in the competitive landscape of online travel platforms.

Market Segmentation

-Mahesh Tiria

Dataset

This real-world customer dataset with 31 variables describes 83,590 instances (customers) from a hotel in Lisbon, Portugal.The data comprehends three full years of customer personal, behavioural, demographic, and geographical information.

Overview of dataset

Columns in hotel dataset are as follows

```
RangeIndex: 83590 entries, 0 to 83589
Data columns (total 31 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ID                    83590 non-null  int64
 1   Nationality           83590 non-null  object
 2   Age                   79811 non-null  float64
 3   DaysSinceCreation     83590 non-null  int64
 4   NameHash              83590 non-null  object
 5   DocIDHash             83590 non-null  object
 6   AverageLeadTime       83590 non-null  int64
 7   LodgingRevenue        83590 non-null  float64
 8   OtherRevenue          83590 non-null  float64
 9   BookingsCanceled      83590 non-null  int64
 10  BookingsNoShowed      83590 non-null  int64
 11  BookingsCheckedIn     83590 non-null  int64
 12  PersonsNights         83590 non-null  int64
 13  RoomNights            83590 non-null  int64
 14  DaysSinceLastStay     83590 non-null  int64
 15  DaysSinceFirstStay    83590 non-null  int64
 16  DistributionChannel   83590 non-null  object
 17  MarketSegment         83590 non-null  object
 18  SRHighFloor           83590 non-null  int64
 19  SRLowFloor            83590 non-null  int64
 20  SRAccessibleRoom      83590 non-null  int64
 21  SRMediumFloor         83590 non-null  int64
 22  SRBathtub             83590 non-null  int64
 23  SRShower              83590 non-null  int64
 24  SRCrib                83590 non-null  int64
 25  SRKingSizeBed         83590 non-null  int64
 26  SRTwinBed             83590 non-null  int64
 27  SRNearElevator        83590 non-null  int64
 28  SRAwayFromElevator    83590 non-null  int64
 29  SRNoAlcoholInMiniBar  83590 non-null  int64
 30  SRQuietRoom           83590 non-null  int64
```

Data Preprocessing

Removing negative values from ‘age’ and ‘average lead time’.

```
df = df[df['Age'] > 0] # Remove negative & null values
df = df[df['AverageLeadTime'] >= 0]
```

Removing unimportant categorical attributes

```
# Remove Columns
cols = ['DocIDHash', 'NameHash']
df.drop(cols,axis = 1, inplace = True)
```

There are about 188 different nationalities, when we get to the encoding step, this will increase the dimensionality, and the encoded words might not even be correlated to other features (Not very important), therefore, I'll see which nationalities occurred most of the time within the table, and remove the rest of nationalities.

Removing the outliers and ID column

```
# Remove outliars
df = df[df['Age'] < 100]
# Remove ID column
df.drop('ID', axis = 1, inplace = True)
```

Removing outliers in nationalities

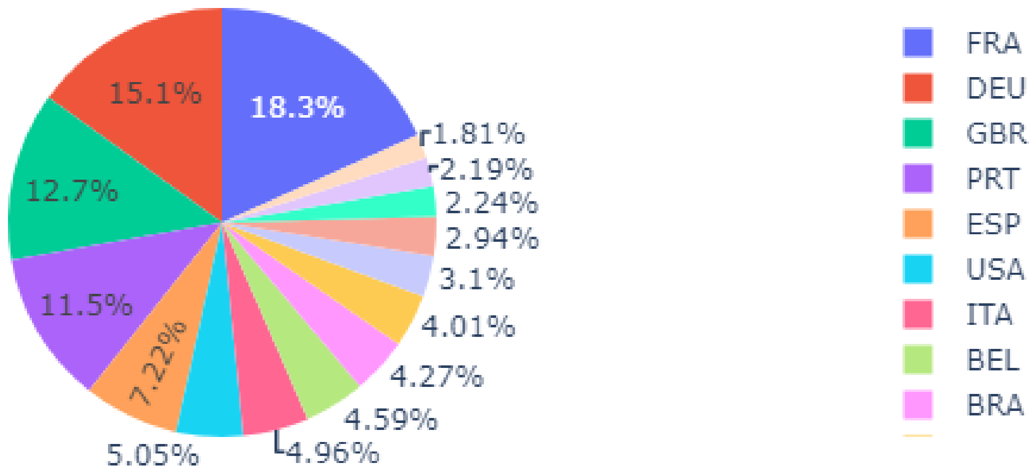
```
# Remove unimportant nationalities
names = df.groupby('Nationality').count().sort_values(by = 'Age', ascending =
top_15_nationalities = list(names[names['Age'] < 1000]['Age'].keys())

for i in top_15_nationalities:
    df['Nationality'] = df['Nationality'].str.replace(i,'')
```

Converting nationalities into a numerical columns via one hot encoding

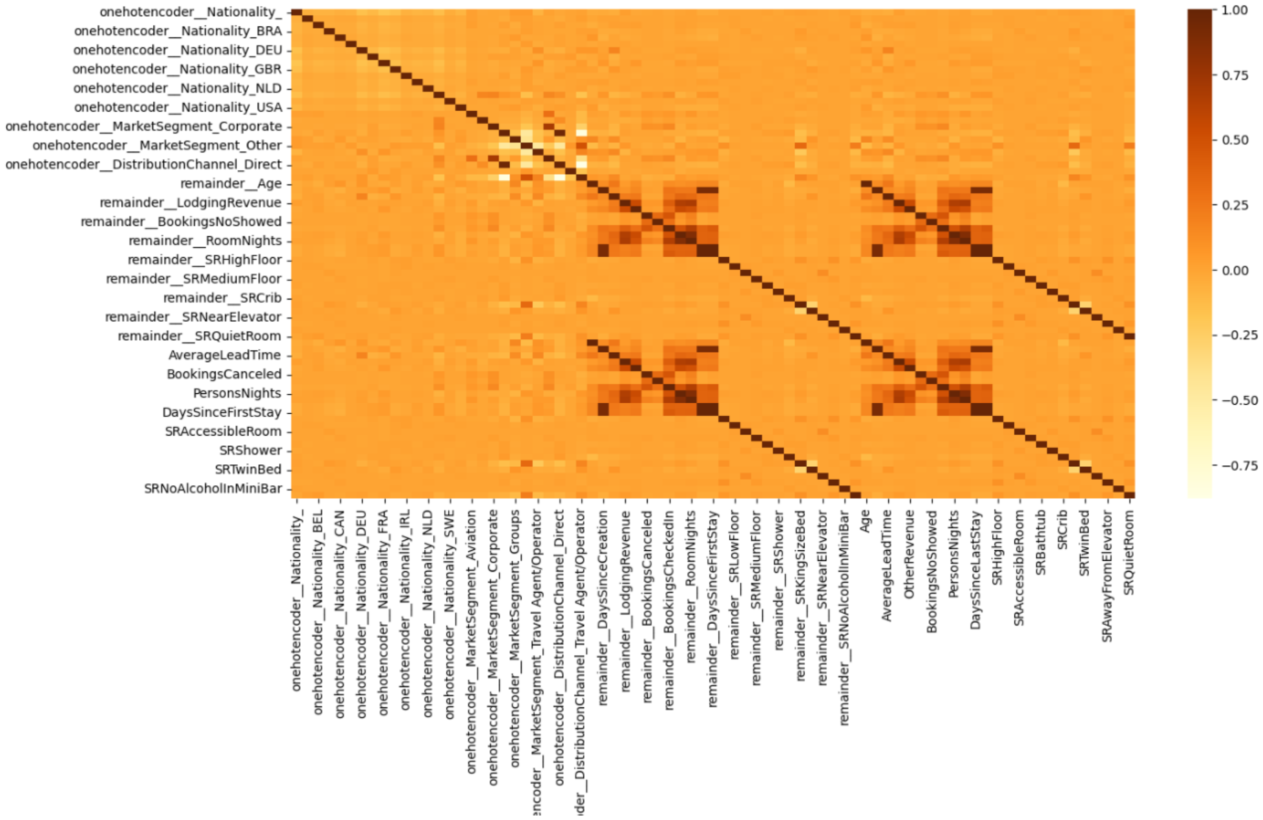
```
# Encoding Variables
transformer = make_column_transformer(
    (OneHotEncoder(sparse=False), ['Nationality', 'MarketSegment', 'DistributionChannel']),
    remainder='passthrough')

transformed = transformer.fit_transform(df)
transformed_df = pd.DataFrame(transformed, columns=transformer.get_feature_names_out())

# Concat the two tables
transformed_df.reset_index(drop=True, inplace=True)
df.reset_index(drop=True, inplace=True)
df2 = pd.concat([transformed_df, df], axis=1)

# Remove old columns
df2.drop(['Nationality', 'MarketSegment', 'DistributionChannel'], axis = 1, inplace = True)
```

Exploring data

Distribution of nationalities

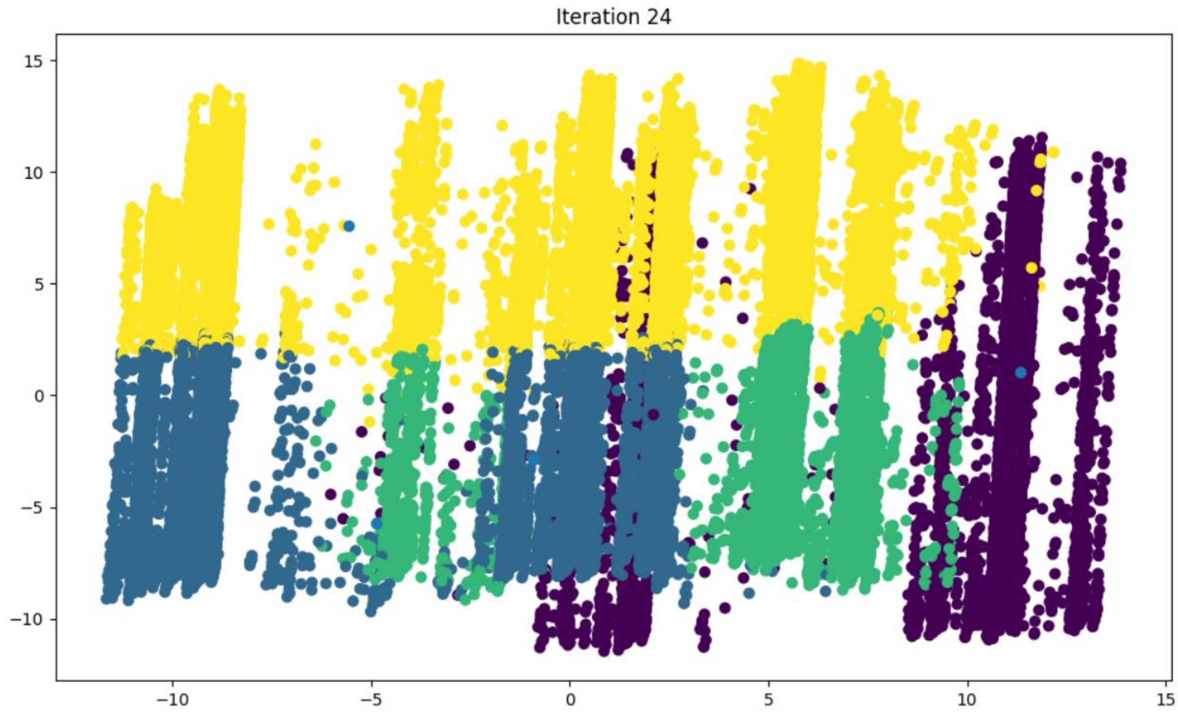


Heatmap showing correlation between various numeric attributes



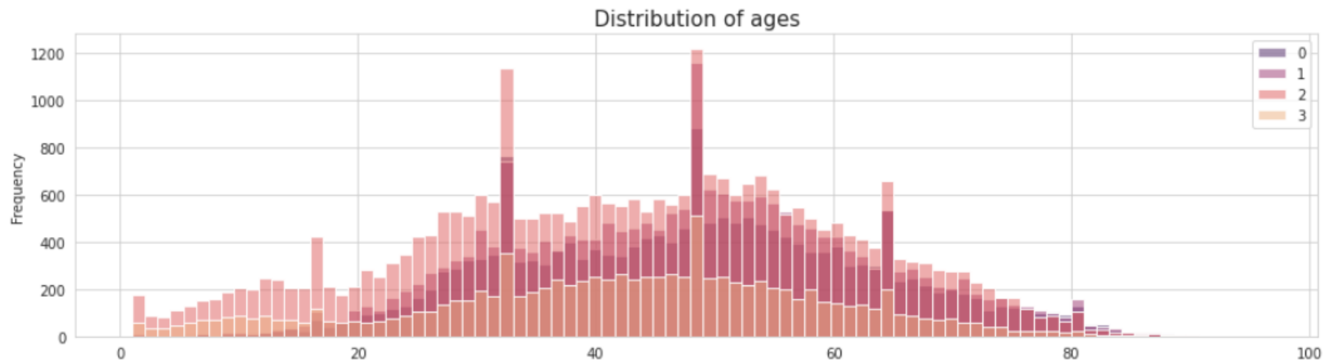
Extracting Segments

Applying Kmeans(K=4) on the processed dataset and using PCA with two components to plot the clusters.

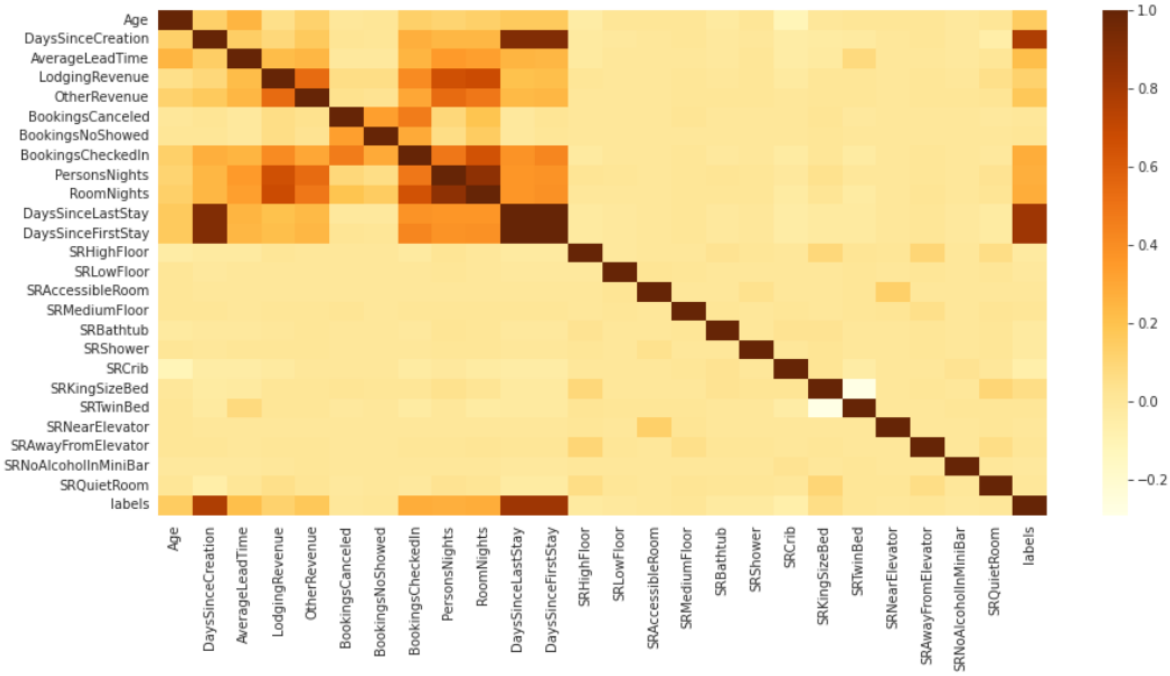


Describing Segments

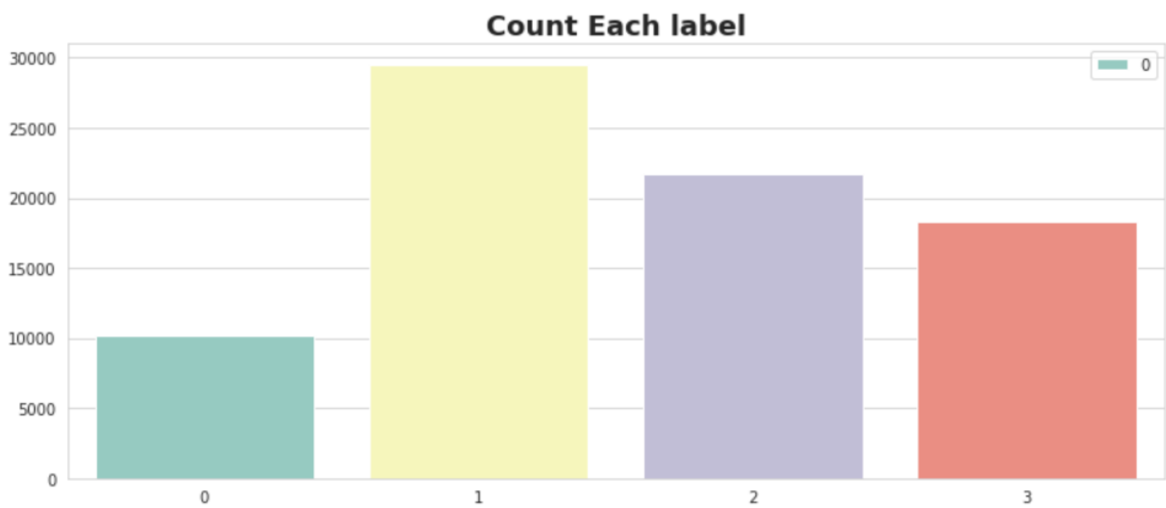
Plotting the distribution of ‘age’ with cluster ‘labels’ and observing relation between age and cluster labels.



Effect of the label column on each feature, observing the correlation value along the ‘label’ column.



Counting the no of each label in the dataset



Conclusion

Here we have successfully segmented the hotel dataset into 4 segments. Segments created are highly correlated to attributes ‘DaysSinceCreation’, ‘DaysSinceFirstDay’ and ‘DaysSinceLastDay’.There is moderate correlation between ‘age’ and ‘label’ generated.

Now we can choose our target segment to select which hotel customer segment to focus for our hotel recommendation services.

HOTEL CUSTOMERS DATASET

SAMARTH R JOSHI

ABSTRACT:

The "Hotel Customers Dataset" contains detailed information about customers who have booked stays at a hotel. The dataset includes various attributes such as nationality, age, booking history, revenue information, room preferences, and other relevant details. It can be used for analyzing customer behavior, preferences, and patterns related to hotel bookings. The dataset's richness in terms of features makes it valuable for exploring trends, identifying patterns, and making informed decisions in the hospitality industry.

Here's a summary of the columns in the dataset:

- ID: Unique identifier for each record.
- Nationality: Nationality of the guest.
- Age: Age of the guest.
- DaysSinceCreation: Number of days since the record was created.
- NameHash, DocIDHash: Hash values for name and document ID.
- AverageLeadTime: Average lead time for booking.
- LodgingRevenue, OtherRevenue: Revenue from lodging and other sources.
- BookingsCanceled, BookingsNoShowed, BookingsCheckedIn: Booking-related information.
- PersonsNights, RoomNights: Number of persons and room nights.
- DaysSinceLastStay, DaysSinceFirstStay: Time-related information about stays.
- DistributionChannel, MarketSegment: Information about booking channels and market segments.
- Specific room preferences and amenities: SRHighFloor, SRLowFloor, SRAccessibleRoom, etc.

NEED FOR CUSTOMER SEGMENTATION

Personalized Guest Experiences:

Customer segmentation allows hotels to understand the diverse needs and preferences of their guests. With this knowledge, hotels can personalize the guest experience by offering tailored services, room preferences, and amenities based on the specific segment to which a guest belongs.

Targeted Marketing and Promotions:

Different customer segments may respond differently to marketing messages and promotions. By segmenting hotel customers, marketing efforts can be targeted more effectively. For example, business travelers may be interested in corporate packages, while leisure travelers may respond better to vacation packages or family-friendly promotions.

Optimized Room Inventory and Pricing:

Customer segmentation helps hotels optimize room inventory and pricing strategies. By understanding the booking behaviors of different segments, hotels can adjust room rates dynamically, offer targeted discounts, and maximize revenue during peak periods.

Enhanced Loyalty Programs:

Loyalty programs are common in the hotel industry, and customer segmentation can help tailor these programs to specific guest preferences. By offering rewards and perks that align with the interests of each segment, hotels can increase loyalty and encourage repeat bookings.

Efficient Operational Planning:

Different customer segments may have distinct patterns of checking in, using amenities, and engaging with hotel services. Understanding these patterns allows hotels to plan their operations more efficiently, ensuring that staff and resources are allocated where they are most needed for each segment.

Customer Retention and Feedback:

Identifying and addressing the specific needs of different customer segments can contribute to higher customer satisfaction and loyalty. Guest feedback from various segments can be used to improve services, resolve issues, and enhance overall guest satisfaction.

Meeting Event and Conference Needs:

For hotels that cater to events and conferences, customer segmentation helps in understanding the requirements of business travelers. This includes meeting facilities, conference rooms, and technology support. Hotels can customize their offerings to attract and accommodate corporate clients effectively.

Adapting to Seasonal Trends:

Seasonal trends may affect different customer segments differently. For example, certain segments may prefer beach resorts during the summer, while others may opt for city hotels.

DATA PREPROCESSING

```
df = df[df['Age'] > 0] # Remove negative & null values
df = df[df['AverageLeadTime'] >= 0]

print('The total number of negative values of age is ', len(df[df['Age'] < 0]))
print('The total number of missing values of age column is', df['Age'].isnull().sum())
print('The total number of negative values of averageLeadTime is ', len(df[df['AverageLeadTime']
< 0]))

#df.shape
```

```
The total number of negative values of age is  0
The total number of missing values of age column is 0
The total number of negative values of averageLeadTime is  0
```

Remove Unimportant Columns & Change Type of ID Column

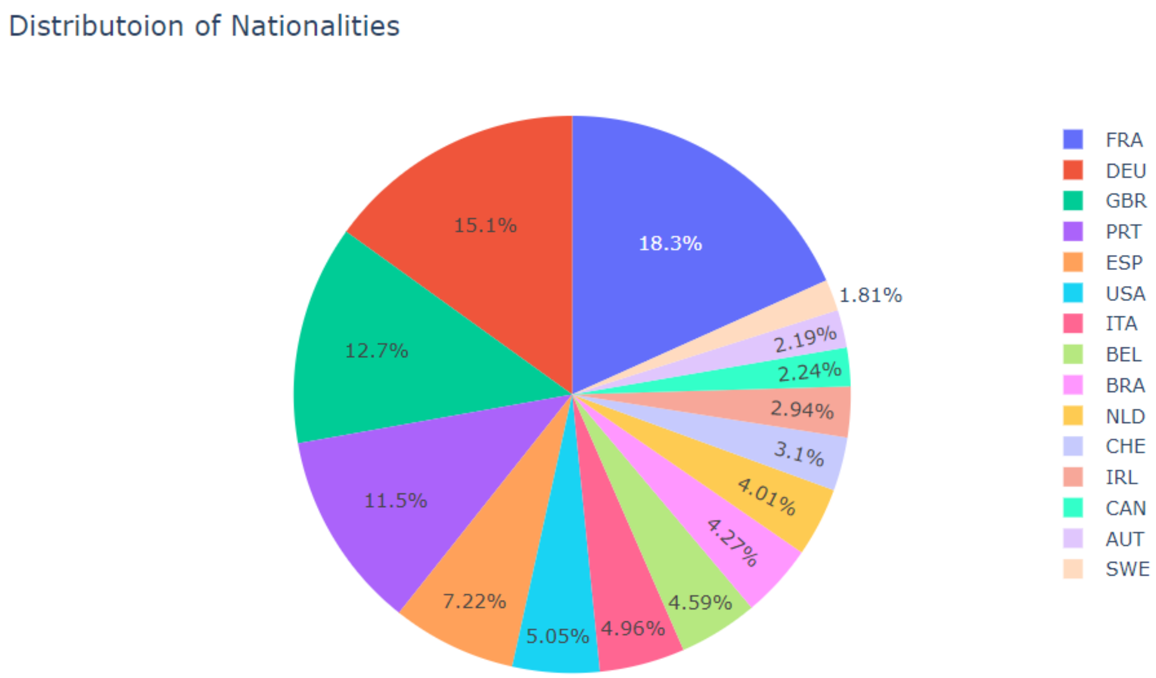
```
# Remove Columns
cols = ['DocIDHash', 'NameHash']
df.drop(cols,axis = 1, inplace = True)

# Change Type
df['ID'] = df['ID'].astype(str)
```

Remove Outliars & ID column

:

```
# Remove outliars
df = df[df['Age'] < 100]
# Remove ID column
df.drop('ID', axis = 1, inplace = True)
```

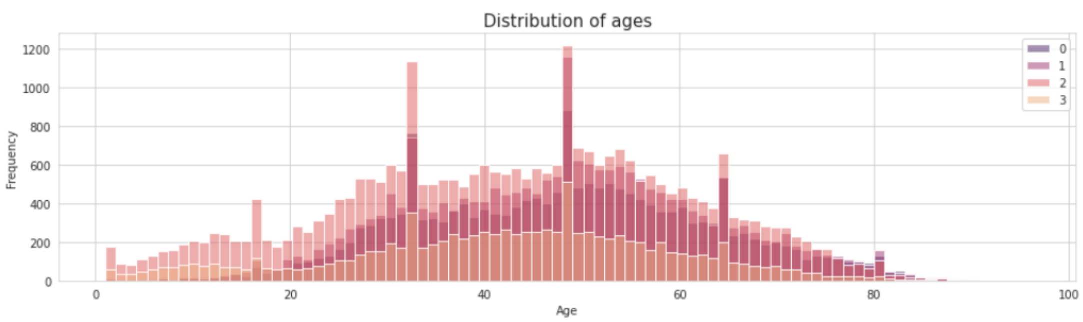
DATA VISUALIZATION

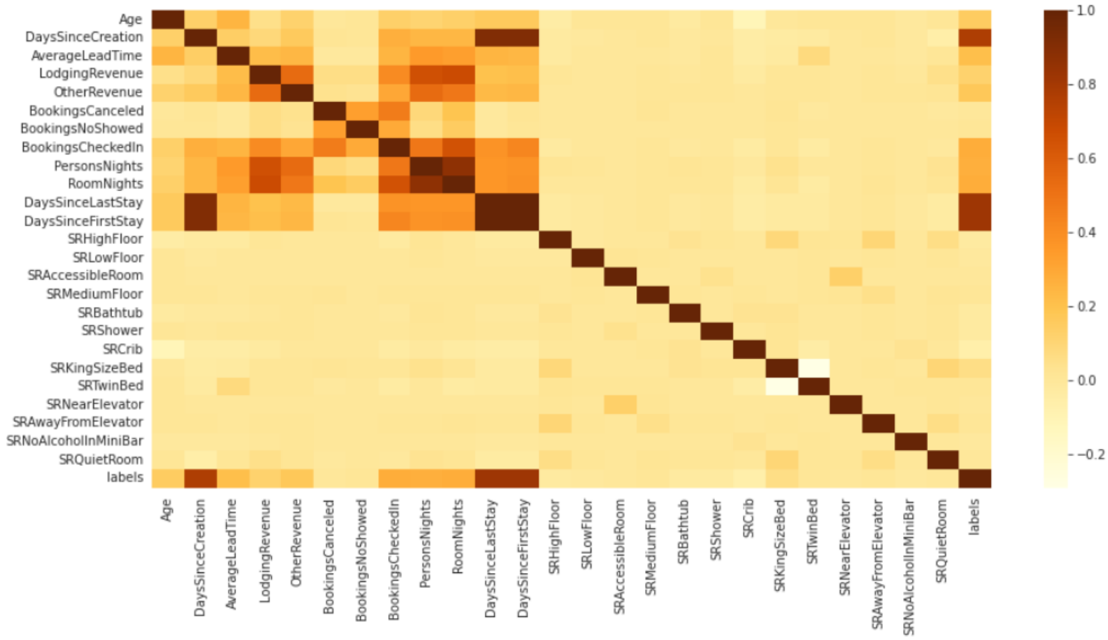


Distrbution of ages for each label

In [27]:

```
plt.figure(figsize = (16,4))
sns.set_style('whitegrid')
sns.histplot(data = df, x="Age", fill = True, color = '#41be78', hue = "labels",palette = 'flar
e' )
plt.title('Distribution of ages', fontsize = 15)
plt.xlabel('Age')
plt.legend(labels = [0,1,2,3])
plt.ylabel('Frequency')
plt.show()
```





LINKS:

GITHUB: [CODE IMPLEMENTATION]

DATASET: [DATASET]

CONCLUSION:

In conclusion, achieving effective customer segmentation in the hotel industry is crucial for tailoring services, optimizing marketing efforts, and enhancing guest satisfaction. The process involves collecting and organizing relevant data, defining segmentation criteria, using data analysis techniques to identify patterns, and developing a segmentation model. By profiling each segment, hotels can tailor marketing strategies, customize services, and communicate with guests in a way that aligns with their unique preferences and behaviors.

The implementation of targeted marketing, personalized offerings, and efficient operational planning based on customer segmentation can lead to improved customer satisfaction, increased loyalty, and optimized revenue streams. The iterative nature of monitoring and adapting ensures that the segmentation strategy remains relevant in the face of evolving customer preferences and market dynamics.

Ultimately, customer segmentation empowers hotels to provide a more personalized and enjoyable experience for their guests, fostering stronger relationships and competitiveness in the hospitality industry.

Financial equation of the hotel management recommendation system

The equation y=x(t)−c to represent the profit in a business context, specifically in the context of a commission-based business where:

- y represents the profit.
- x(t) represents the growth of the customer base at a given time (t).
- c represents the total cost.

In this context, the equation can be expressed as:

Profit=Customer Growth−Total Cost

This equation reflects a simple representation of profit as the difference between the growth of the customer base and the total cost incurred. If the growth of the customer base exceeds the total cost, it contributes positively to the profit.

It's important to note that in a real business scenario, the relationship between profit, customer growth, and total cost is typically more complex. Factors such as revenue per customer, variable costs, and other business expenses may need to be considered for a more accurate representation. Sophisticated financial models and accounting principles are often applied to analyze and optimize profitability in business contexts.

Profit=f(Customer Growth,Revenue per Customer,Cost per Customer,Other Financial Features)−Total Cost

Here:

Customer Growth,Revenue per Customer,Cost per Customer,Other Financial Features)f(Customer Growth,Revenue per Customer,Cost per Customer,Other Financial Features) represents a function that takes into account various financial features related to customer growth, revenue, costs, and other relevant metrics. This function can be developed using machine learning techniques to analyze historical data, predict future trends, and optimize for profit.

Customer Growth :

Customer Growth is still a factor, representing the growth of the customer base, but it's now part of a more sophisticated function.

Revenue per Customer:

Revenue per Customer represents the average revenue generated per customer.

Cost per Customer :

Cost per Customer represents the average cost associated with acquiring and maintaining a customer.

Other Financial Features:

Other Financial Features can include additional factors that impact the financial performance of the business.

Total Cost:

Total Cost still represents the overall costs incurred by the business.

In a machine learning recommendation system, historical data on customer behavior, revenue, and costs can be used to train models that predict the expected financial outcome based on various input features. The recommendation system can then suggest actions or strategies to optimize the financial performance, such as targeting specific customer segments, adjusting pricing strategies, or managing costs more efficiently.

It's important to note that developing a robust machine learning recommendation system for financial optimization requires careful consideration of data quality, feature selection, model training, and validation. Additionally, ongoing monitoring and adaptation are crucial to ensure the system remains effective as business conditions change.

Equation highlights:

The equation $\mathbf{y=x(t)-c}$ to a hotel recommendation system. In this context, let's interpret the equation with suitable terms for a recommendation system:

Let's say:

y represents the overall desirability or suitability of a particular hotel.

x(t) represents various features or factors associated with a hotel at a given time (t). These features could include things like location, price, amenities, user reviews, etc.

c is a constant, which could represent a baseline desirability level or a threshold for a hotel to be considered.

Now, the equation $y=x(t)-c$ can be interpreted as follows:

x(t) represents the positive and negative factors associated with a hotel.

c represents a baseline desirability level or a threshold. If a hotel's overall desirability (y) is greater than this threshold (c), it may be considered a suitable recommendation.

So, in the context of a hotel recommendation system, this equation helps evaluate the desirability of a hotel based on its features relative to a baseline level. This is a simplified representation, and in a real recommendation system, more complex algorithms and machine learning models are often used to analyze and predict user preferences based on a variety of features and historical data.

If you have specific features or criteria you'd like to consider in the recommendation system, feel free to provide more details, and I can offer a more tailored explanation.

Derivation:

The equation y=x(t)−c with respect to time (t).

$$dy/dt= d/dt(x(t)-c)$$

The derivative of x(t) with respect to t is denoted as dx/dt. Using the chain rule, the derivative of x(t)−c is:

Therefore, the derivative of y with respect to (t) is simply the derivative of x(t) with respect to t:

$$dy/dt = dx/dt-0$$

This means the rate of change of y with respect to time is equal to the rate of change of x with respect to time.

$$dy/dt = dx/dt$$

If you have a specific function x(t) or more details about the problem, I can provide a more specific elaboration.

Detailed elaborative equation explained:

- Y=X(t)−c in the context of a customer recommendation system where
- Y represents profit,
- X(t) represents the growth of the customer base at a given time (t), and
- c represents the total cost.

Profit (Y):

Profit is the financial gain obtained after subtracting all costs from revenue. It is a fundamental metric for evaluating the success and sustainability of a business.

Growth of the Customer Base X(t)):

X(t) represents the growth of the customer base, and it involves various factors such as:

New Customer Acquisition: The number of new customers joining the business during a specific time period.

Customer Retention: The ability to retain existing customers over time.

Expansion within Existing Customers: Increasing the revenue generated from existing customers through upselling or cross-selling.

Total Cost (c):

c represents the total cost incurred by the business. This includes:

Operational Costs: Costs associated with day-to-day operations, including employee salaries, rent, utilities, etc.

Marketing Costs: Expenses related to advertising, promotions, and customer acquisition.

Customer Service Costs: Expenses associated with providing customer support and maintaining customer satisfaction.

Technology Costs: Costs related to the use of technology, software, and infrastructure.

Now, let's interpret the equation in the context of a customer recommendation system:

$Y=X(t)-c$ suggests that the profit is determined by the growth of the customer base subtracted by the total cost.

Interpretation:

If the growth of the customer base X(t)) is substantial and contributes positively, it can enhance overall profit (Y).

If the total cost (c) is well-managed and kept in check, it positively impacts profitability. Efficiently managing costs ensures that the revenue generated from the growing customer base contributes more significantly to profit.

Optimizing the Customer Recommendation System:

Strategies for optimizing the customer recommendation system to maximize profit might include:

Personalized Recommendations: Using machine learning algorithms to provide personalized recommendations that enhance customer satisfaction and retention.

Cost-Effective Acquisition: Identifying and utilizing cost-effective customer acquisition channels to ensure that the cost of acquiring new customers is justified by their lifetime value.

Retention Strategies: Implementing strategies to retain existing customers, as retaining customers is often more cost-effective than acquiring new ones.

Machine learning concepts based on the equation:

To incorporate machine learning into the equation to model the relationship between the growth of the customer base X(t)), profit (Y), and total cost (c), you can consider using a predictive model. A common approach is to use a regression model. Here's an example of how you might formulate a machine learning equation:

$Y=f(X(t), c)+ \epsilon$

In this equation:

- Y is the predicted profit.
- X(t) is the growth of the customer base.
- c is the total cost.
- f is a machine learning model that captures the relationship between X(t), c, and Y.
- ϵ is the error term, representing the difference between the predicted and actual profit.

You would train the machine learning model using historical data, where you have information about the growth of the customer base, total cost, and actual profit. The model learns to generalize the underlying patterns in the data, allowing it to make predictions on new, unseen data.

The specific form of the function (f) would depend on the characteristics of your data and the nature of the relationship between the variables. Common machine learning algorithms for regression tasks include linear regression, decision trees, random forests, and neural networks.

Here's a simplified example using linear regression:

$Y=\beta_0+\beta_1 \cdot X(t)+\beta_2 \cdot c+\epsilon$

In this equation:

β_0 is the intercept.

β_1 and β_2 are coefficients representing the weights assigned to $X(t)$ and c, respectively.

Training the model involves finding the values of β_0, β_1, and β_2 that minimize the difference between the predicted and actual profit on the training data.