NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

BC2407 Project Report

Executive Summary

Amidst the rapid growth of the tourism market driven by technological advancements in travel, Expedia, a key player in the industry, finds itself facing both challenges and opportunities. The study delves into the factors that have shaped Expedia's present standing, specifically concerning hotel reservations. It also analyses the necessity for the organisation to devise inventive commercial approaches to restore its competitiveness.

Key findings:

1. **Market Analysis:** We utilised moving average and classical seasonal decomposition to understand Expedia's market position and customer engagement in travel research and hotel booking. These analyses shed light on current trends and customer loyalty.
2. **Demand Forecasting:** To determine the best exponential smoothing method for demand forecasting, we assessed accuracy using Root Mean Square Error (RMSE). The Holt-Winters method emerged as the most accurate choice. Additionally, we employed random forest, a powerful machine learning model, to capture complex seasonality and trends, ensuring the accuracy of demand forecasts.
3. **Competitor Analysis:** Through comprehensive competitor analysis, we examined pricing strategies in the industry to refine Expedia's own pricing strategy. Insights gained from this analysis offer valuable opportunities for Expedia to enhance its competitiveness.

Recommendation:

The findings underscore the importance of accurate forecasting models and the need for Expedia to adapt its strategies accordingly. Specifically, we recommend integrating the Holt-Winters method for demand forecasting and leveraging random forest for trend analysis. Furthermore, insights from competitor analysis should inform revisions to Expedia's pricing strategy. These proactive measures are essential for Expedia to regain its competitive edge in the dynamic tourism market.

In summary, this study highlights actionable steps that Expedia can take to enhance its standing in the industry, given the challenges and opportunities presented by evolving consumer preferences and technological advancements.

1. Project Background

Within the intensely competitive online travel agency market, Expedia faces the challenge of high customer price sensitivity. This is exacerbated by the prevalent use of comparison websites by travellers eager to find the most advantageous deals. Expedia must strike a careful balance in this environment, which means offering competitive pricing tactics to draw in and maintain a varied consumer base. Thus, Expedia’s success is dependent on the ability to regularly assess and modify its pricing strategies according to shifting consumer expectations, rival strategies, and market conditions. This will allow Expedia to be the go-to option for budget-conscious travellers who do not want to forego service excellence or convenience, ensuring its prominence and competitive edge within the worldwide online travel agency space.

Expedia witnessed a remarkable surge in global search volumes throughout 2022 and into early 2024, with a staggering 75% year-over-year increase in the first quarter of 2022 (Expedia Groups, 2022). This trend underscores Expedia's continued relevance in the travel industry, emphasising the importance of strategic planning and market responsiveness.

1.1 Business Problems

While Expedia has been doing seemingly well, maximising revenue has been an obstacle for them in recent times, especially in comparison to rivals like Booking Holdings. In 2019, Expedia Group demonstrated commendable business performance, registering revenues of $12.07 billion, a 7.5% increase from the previous year (Walker, 2021), in contrast to Booking Holdings which reported revenues of $15.07 billion with a 3.7% growth from the previous year (Statista, 2023). Nevertheless, with travel limitations being lifted after the pandemic, Expedia Group's revenue rebound did not match that of Booking Holdings. In 2023, Expedia's revenue was $12.84 billion versus Booking Holdings' impressive $21.37 billion (Statista, 2023). The growth rate of Expedia's revenue decelerated to 6.4%, whereas Booking Holdings witnessed a significant jump to 41.8% in revenue growth since 2019.

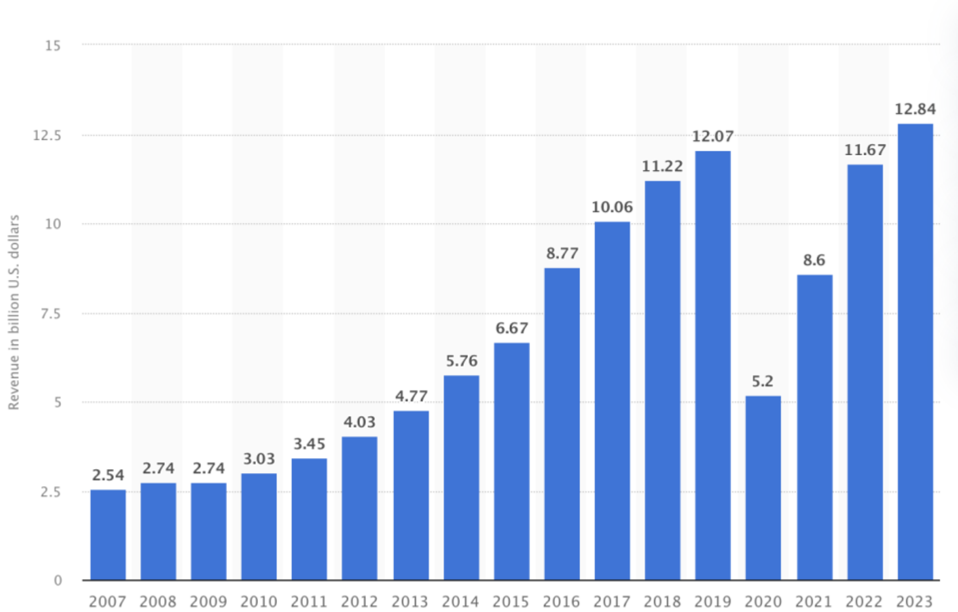


Figure 1: Expedia Group Revenue from 2007 - 2023 (Statista 2023)

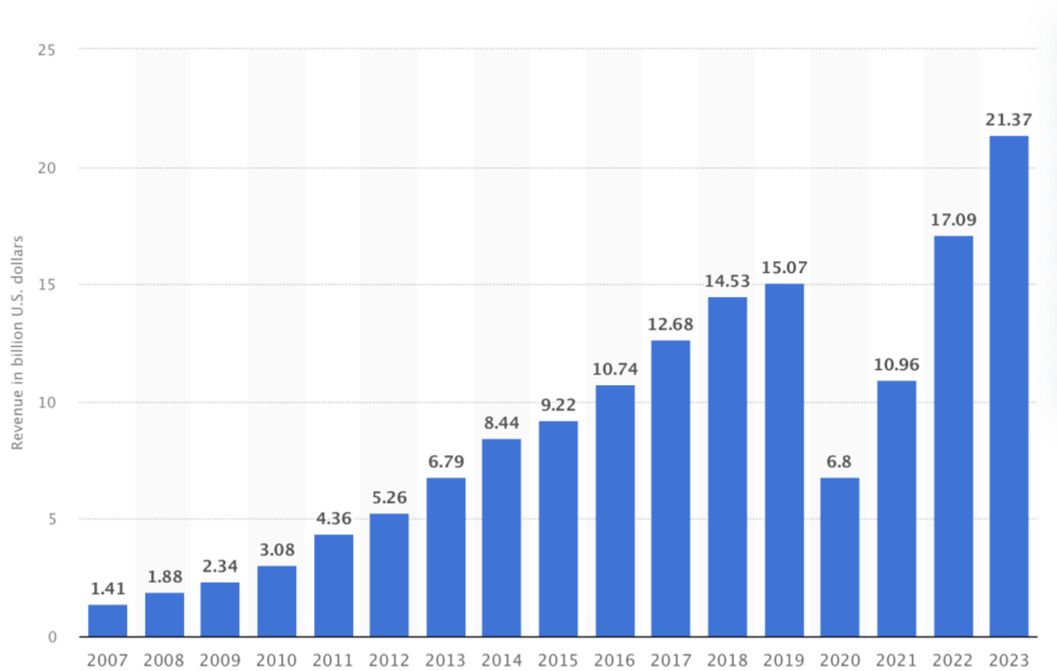


Figure 2: Booking Holdings Revenue from 2007 - 2023 (Statista 2023)

Evidently, there is still much room for development as Expedia's sales performance has not yet reached its fullest potential.

Our team intends to construct effective demand forecasting models for Expedia. Such forecasting is pivotal for making informed business choices, particularly for refining pricing approaches. With insights into future demand, Expedia can take proactive steps to modify its prices, offerings, and marketing strategies to align with the preferences and needs of consumers (Hand, 2024). This approach not only streamlines resource allocation but also sharpens marketing efforts, ultimately enhancing Expedia's market competitiveness and driving revenue growth by pinpointing and capitalising on emerging trends and opportunities.

1.2 Existing Solutions

A variety of demand forecasting methods have been implemented in hotel demand forecasting, mainly falling under two categories: qualitative demand forecasting and quantitative demand forecasting.

Qualitative demand forecasting is subjective as it relies on expert opinion and market research (Chapman, 2021). A frequently used technique is Delphi Technique, questionnaires will be sent out anonymously to experts, for example customers to gather feedback and opinions. After the first round of replies, the result will be analysed and re-circulated (Pennsylvania State University, 2007). This method is useful in analysing long-term trends and developing business strategies. Research conducted by the Rand Corporation suggests that, while the Delphi panel typically converges toward a generally correct consensus view based on current technologies and trends, forecasting accuracy diminishes when predicting new developments (Pennsylvania State University, 2007).

The quantitative forecasting method is statistical as it forecasts future demand based on the analysis of historical data. Techniques such as time series analysis play a crucial role in forecasting demand and future events. Among these methods, exponential smoothing stands out for its effectiveness in demand forecasting. It not only utilises historical data but also captures seasonality and trends, thereby significantly enhancing the accuracy of demand forecasts.

2. Dataset

2.1 Introduction to Dataset

Given the limited availability of existing datasets, we have collected Expedia's booking data from October 2012 to October 2014. Nevertheless, the number of records in the dataset is still sufficient for demand forecasting and understanding the underlying trends and seasonality. Secondly, we will use a general hotel booking dataset for exploratory data analysis to further understand the pricing strategies and customer behaviours.

These datasets are related to the tourism sector and are multivariate, dependent on time series. The main objective of these datasets are to observe the number of searches, number of bookings and competitor price advantage.

2.2 Data Preparation

The first dataset contains 50 columns and 6.6 million rows of records from Expedia. Hence, we will first identify and extract the columns that are most relevant. Our selection of columns is guided by three key areas of analysis for Exponential Smoothing.

2.2.1 Missing Values and Descriptive Statistics

A significant portion of the data, especially those related to competitor rate differences and visitor historical data, is missing, suggesting challenges in data completeness. Descriptive statistics highlight the diversity in hotel ratings and pricing, indicating varied preferences and price sensitivities among customers.

2.2.2 Number of Searches from Expedia

To gauge public interest in utilising Expedia for hotel bookings, we have extracted related data fields, including the search ID and date time columns. This extraction allows us to effectively calculate the frequency of searches conducted on Expedia's platform within each day.

By analysing the search ID and timestamp data, we can discern patterns and trends in user engagement over time. This insight into the volume of daily searches provides a robust indicator of the level of interest and activity surrounding Expedia's hotel booking services. Such analysis empowers us to understand fluctuations in user demand and interest, aiding in strategic decision-making and resource allocation to better cater to customer needs and preferences.

2.2.3 Number of Check-Ins from Expedia

To facilitate demand forecasting, we aim to ascertain the number of check-ins facilitated through Expedia's platform. While there isn't a dedicated column explicitly labelled "Check-In" in our dataset, we can leverage existing data fields such as the search and booking windows.

The search and booking windows represent the number of days from the search date to the commencement of the hotel stay. By adding these durations to the respective search dates, we can derive the check-in dates for each booking. This approach allows us to establish a timeline of when customers intend to check in to their accommodations.

Subsequently, aggregating the number of check-ins per day provides us with valuable insights into the daily demand for Expedia's services. This data can then serve as a foundational element for our demand forecasting endeavours, enabling us to anticipate and plan for fluctuations in user activity and accommodation bookings effectively.

2.2.4 Competitor Analysis

In the dataset, we have 8 sets of competitor price comparisons available for further analysis. However, it is noted that there are missing values within the categorical data columns corresponding to hotel prices comparison. These missing values are distributed across the dataset and are not structured in a time series format.

To ensure the robustness of our analysis, we will proceed by removing the missing values present in the categorical data columns related to competitor hotel prices comparison. This step is crucial for maintaining data consistency and accuracy, allowing us to conduct meaningful comparative analyses among competitors. By systematically eliminating these missing values, we can focus our analysis on complete and reliable categorical data, enabling insightful assessments of pricing dynamics and competitive positioning within the market.

2.3 Exploratory Data Analysis (EDA) - Expedia

2.3.1 Initial Observations

In the ever-evolving landscape of online travel booking, understanding and predicting hotel demand becomes crucial for platforms like Expedia. This report delves into an analysis aimed at predicting hotel demand, leveraging a dataset comprising 6,622,629 records and 50 features. The analysis encompasses several stages, from initial data exploration to model training and evaluation, aiming to uncover patterns and insights that can inform demand prediction strategies.

The dataset is rich, with a mix of numerical and categorical data types, including search details, property information, and competitor comparisons. An initial overview reveals the dataset's breadth, covering a wide range of search parameters and property details.

2.3.2 Outlier Analysis

An analysis of the ‘price_usd’ feature uncovers 1,352 outliers, suggesting extreme price points possibly influenced by specific events or premium offerings.

2.4 Exploratory Data Analysis (EDA) - Hotel Booking

The EDA is aimed at uncovering patterns and insights related to guest behaviour and hotel operations. The analysis covered several key aspects:

2.4.1 Room Prices Over the Year

The analysis revealed price fluctuations throughout the year for both hotel types, with the Resort Hotel prices peaking during the summer months, attributed to the holiday season. Conversely, City Hotel prices were highest during spring and autumn, aligning with business conference seasons and mild weather appealing to tourists. (Figure 3)

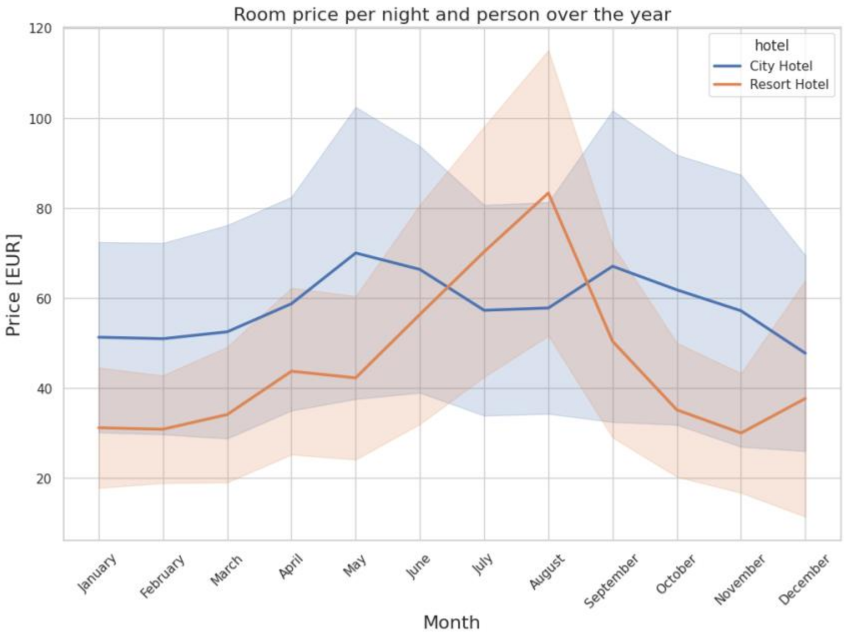


Figure 3: Room price per night per person over the year

2.4.2 Busiest Months

Data indicated varying occupancy rates throughout the year, with City Hotels experiencing peak guest numbers during spring and autumn, while Resort Hotels were busiest during the summer. These patterns closely followed the price trends observed. (Figure 4)

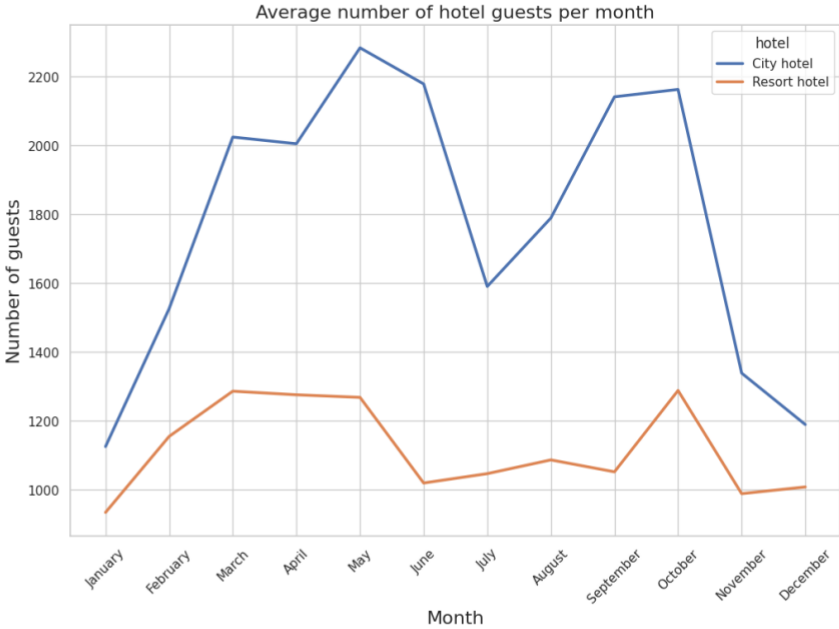


Figure 4: Average number of hotel guests per month

2.4.3 Length of Stay

The duration of stay at hotels revealed distinct preferences, with City Hotel guests favouring shorter stays (1-4 nights) and Resort Hotel guests often opting for longer stays, particularly for a week. (Figure 5)

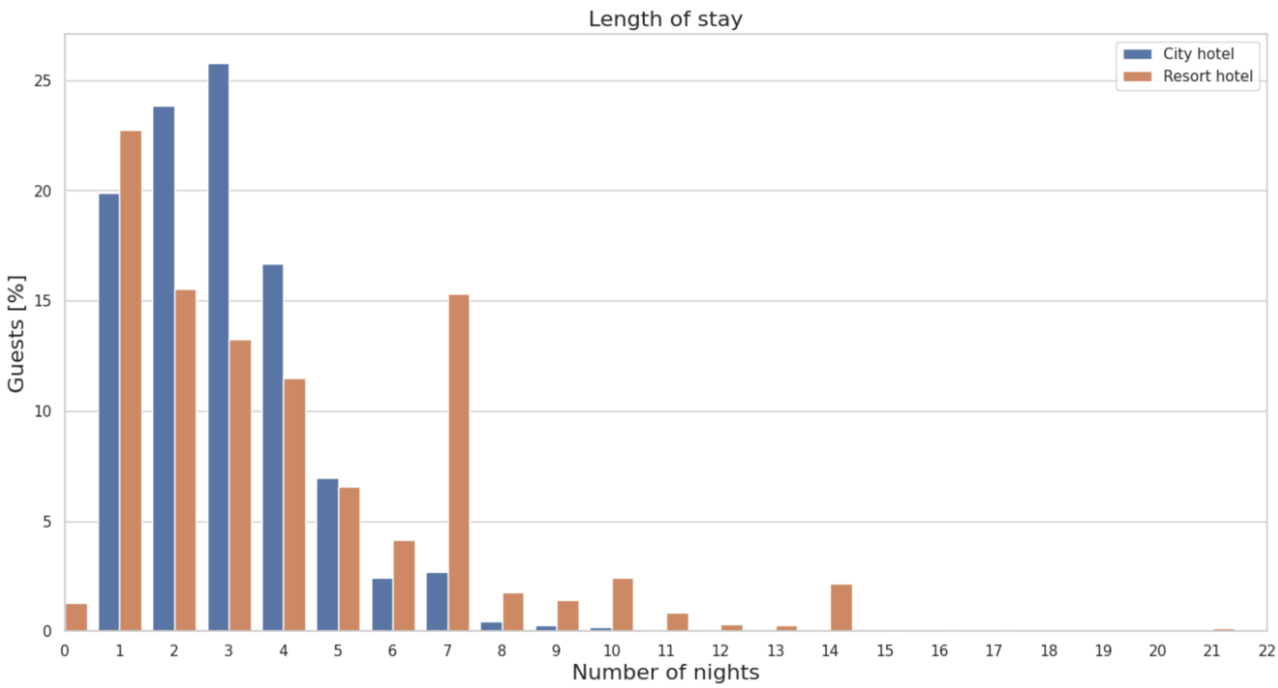


Figure 5: Length of stay distribution

2.4.4 Bookings by Market Segment

An analysis of market segments highlighted the dominance of Online Travel Agencies (OTAs) in booking channels, accounting for a significant portion of the bookings made. (Figure 6)

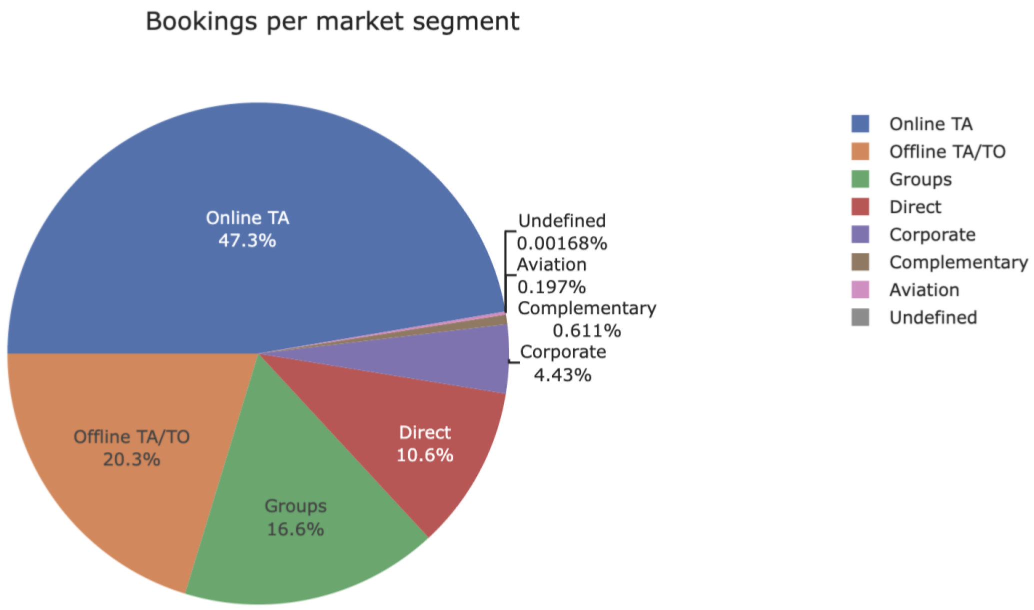


Figure 6: Bookings per market segment

2.4.5 Cancellation Patterns

Cancellations were a notable aspect of hotel operations, with the City Hotel experiencing a higher cancellation rate (42%) compared to the Resort Hotel (28%). The trend in cancellations also varied throughout the year but remained consistently high for the City Hotel. (Figure 7)

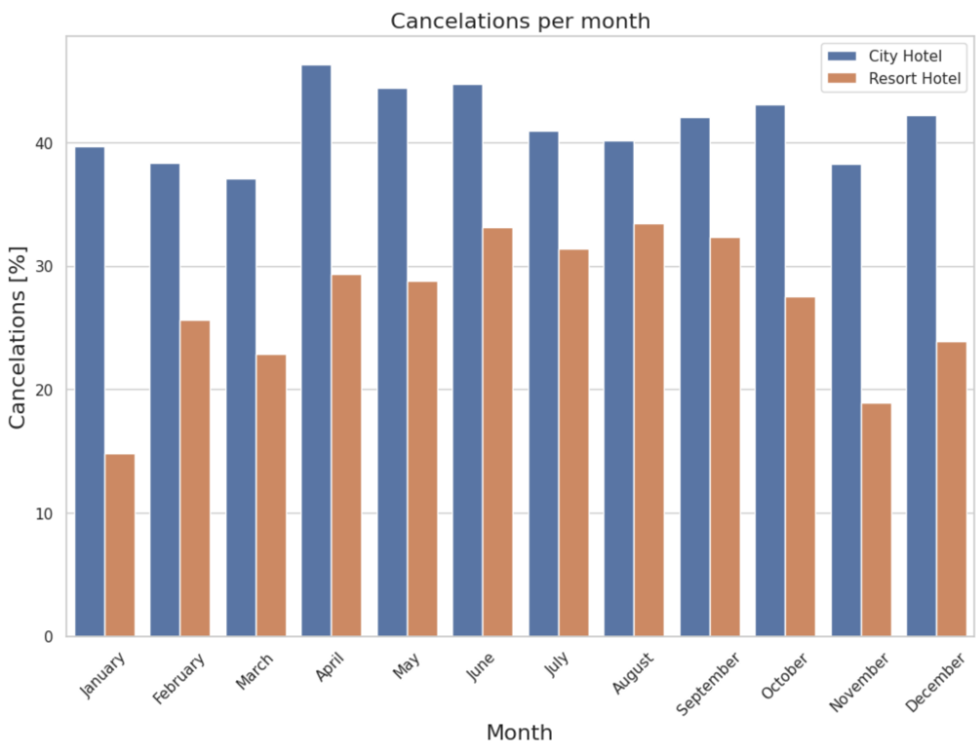


Figure 7: Cancellation rates by month

2.5 Conclusion

The analysis provided valuable insights into guest behaviours, pricing strategies, and cancellation patterns across different hotel types. The findings can inform targeted marketing strategies, pricing adjustments, and improved booking policies to enhance hotel revenue and guest satisfaction. Further, the predictive model for cancellations offers a proactive tool for managing booking uncertainties more effectively.

3. Methodologies Used to Solve the Problem

3.1 Time Series Forecasting

As our team is focusing on demand forecasting based on historical data, we will be using quantitative forecasting methods which include time series analysis. Time series forecasting analyses historical data points collected at regular intervals over time and makes predictions about future values.

Time series forecasting has a few uses that could be beneficial to Expedia. Firstly, as a demand forecasting tool. Expedia can use time series forecasting to predict demand for their services. By analysing historical sales data and accounting for factors such as seasonality, trends, promotions, and economic indicators, Expedia can accurately forecast future sales volumes.

Furthermore, time series forecasting can help Expedia optimise pricing strategies by predicting future price trends and demand elasticity. By analysing historical pricing data, competitor pricing, and market dynamics, Expedia can set optimal prices to maximise profitability and market share.

3.1.1 Moving Average

Moving average is a widely used fundamental technique for forecasting time series data. It involves calculating the average of a specified number of past observations to forecast future values. By adjusting the window size of the moving average, we can control the level of smoothing and responsiveness to changes in the data. We will explore different moving average approaches and select the one that best fits the demand patterns of Expedia.com.

3.1.2 Exponential Smoothing

Exponential smoothing is another widely used technique in time series analysis. It works by assigning exponentially decreasing weights to past observations, with more recent observations weighted more heavily. In our analysis, we will apply exponential smoothing to the historical demand data for Expedia.com. This method allows us to capture trend and seasonality in the data while providing forecasts that adapt to changing patterns over time.

3.2 Random Forest

Random Forest was chosen for predicting hotel demand due to its ability to handle the dataset's complexity, including high dimensionality, missing values, and outliers. This algorithm is robust against overfitting, even with a large number of features, and can process both numerical and categorical data efficiently. Additionally, Random Forest provides valuable insights into which features most significantly impact hotel demand, thanks to its feature importance metrics. It's also adaptable to imbalanced datasets, making it well-suited for accurately identifying patterns in hotel booking behaviours. In essence, Random Forest offers a versatile, powerful, and interpretable modelling solution for the multifaceted challenge of forecasting hotel demand.

4. Insights

4.1 Exponential Smoothings

4.1.1 Number of Searches from Expedia

Before forecasting the number of searches from Expedia, we carried out a time series analysis for the search count from October 2012 to October 2014. It showed a non-constant fluctuation (Figure 8), indicating that it is a multiplicative time series.

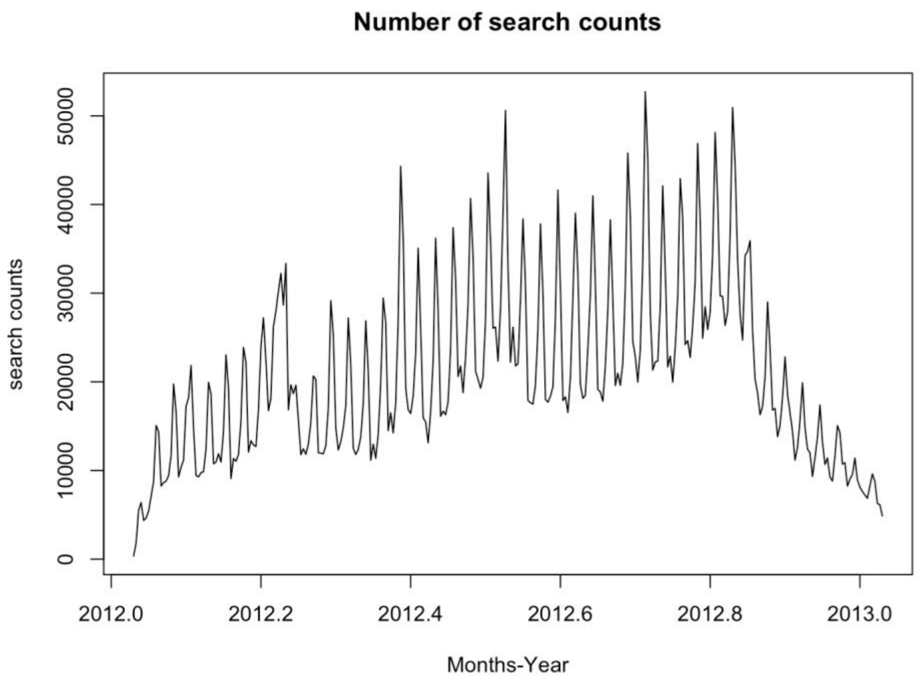


Figure 8: Number of Searches Counts

We conducted moving averages analysis and classical seasonal decomposition. By splitting the time series data with a span of 3 and 7 periods, we were able to capture the short-term and medium-term (Figure 9) fluctuations, and clear downtrends in the data.

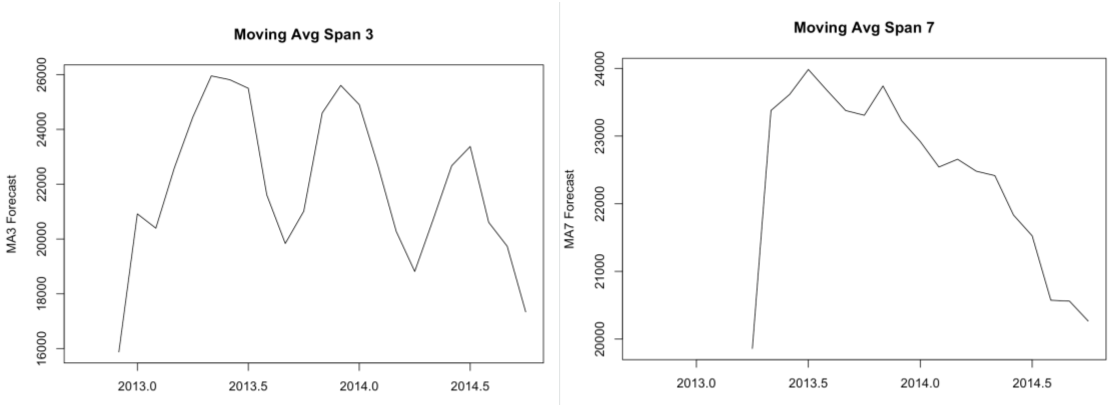


Figure 9: Moving Average Span 3 & 7

From seasonal decomposition (Figure 10), we observed the trend component shows an overall gradual decline over the time period. The seasonal component captures the recurring patterns within the time series data that repeat over a fixed period. Here, we can observe the number of searches getting higher in the months of May, October and November.

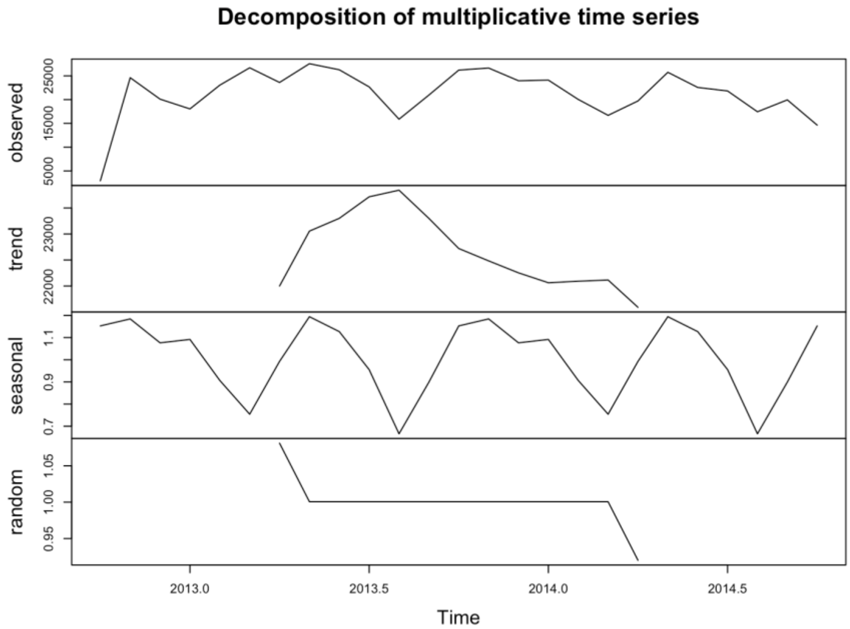


Figure 10: Seasonal Decomposition

To forecast the number of searches, we implemented various exponential smoothing methods such as Simple Exponential Smoothing, Holt's Method, and Holt-Winters method. We compared the Root Squared Mean Error (RMSE) to ensure the accuracy of the methods (Figure 11).

RMSE.holt	7314
RMSE.ses	5759
RMSE.winters	3002

Figure 11: RMSE for Exponential Methods

As the Holt-Winters method has the lowest RMSE, we used this method to forecast the next 4 and 110 periods.

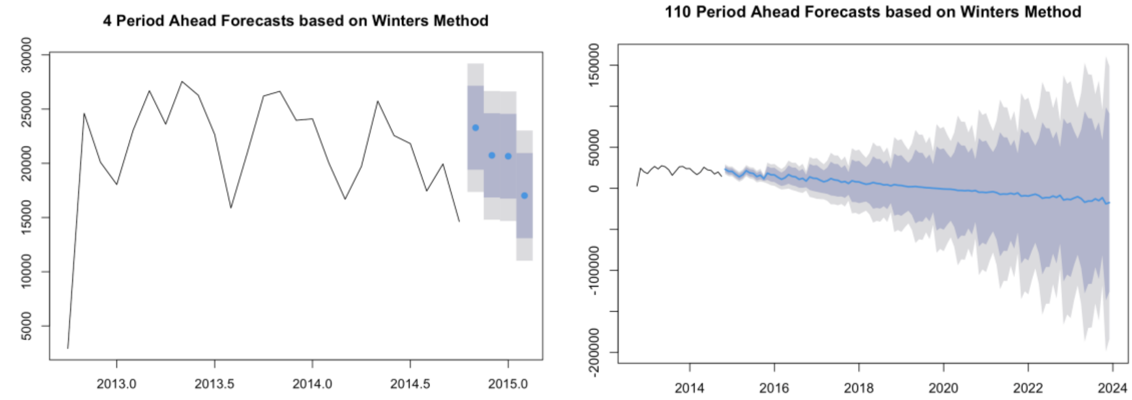


Figure 12: 4 & 110 period ahead forecasts based on Winters Method (Searches)

As shown in Figure 12, short-term forecasted results follow the seasonal patterns closely and align with the historical data's cyclical behaviour. However, as the forecast horizon extends further into the future, the uncertainty increases, and the forecasts become less precise but generally in a gradual decline. The downtrend in Expedia's search volume reflects a decline in customer stickiness. Customers may prefer other Online Travel agencies that provide more options or discounts in travel services.

4.1.2 Number of Check-Ins from Expedia

To examine the number of check-ins from Expedia, we used the same methods as above. We observed a constant fluctuation in the number of Check-In from Exepedia indicating an additive time series (Figure 13).

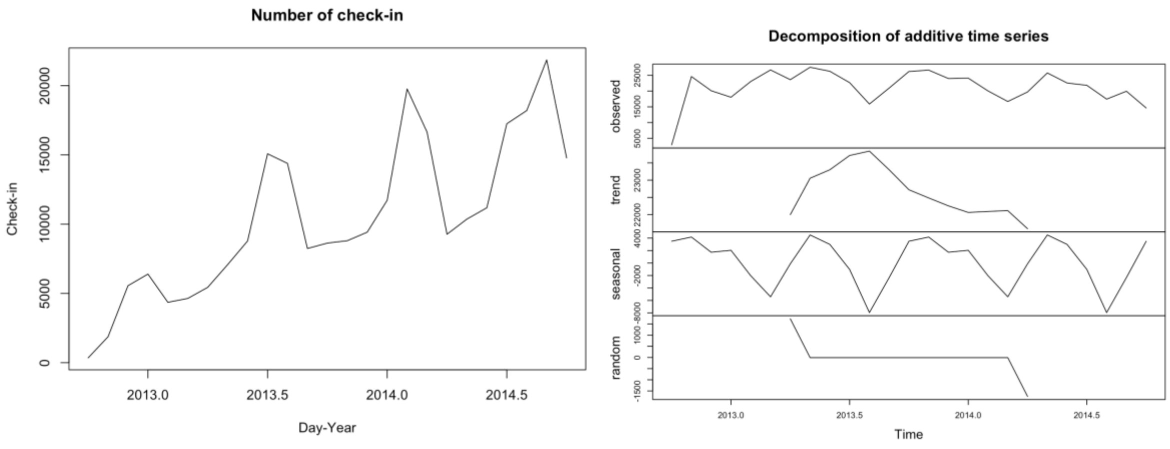


Figure 13: Number of Check-In & Seasonal Decomposition

In the span of 3 and 7 periods, we were able to capture the short-term and medium-term (Figure 14) fluctuations, and clear uptrend in the data.

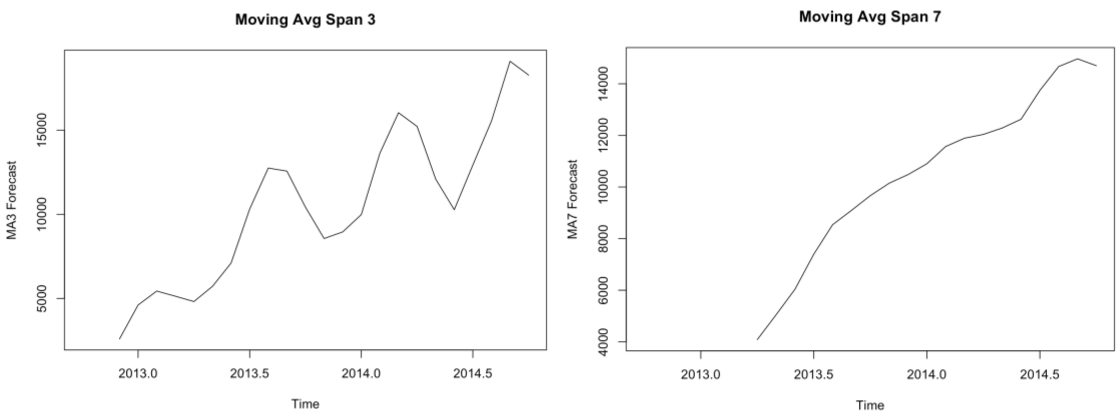


Figure 14: Moving Average Span 3 & 7

As the Holt-Winters method has the lowest RMSE, we will implement this method for forecasting. (Figure 15)

RMSE.holt	4042
RMSE.ses	3777
RMSE.winters	3640

Figure 15: RMSE for Exponential Methods (Check-In)

As shown in Figure 16, short-term forecasts closely mirror the seasonal patterns and cyclical behaviour observed in the historical data. However, as the forecast horizon extends further into the future, a continuous uptrend is observed. This extension leads to increasing uncertainty and the forecasts becoming less precise.

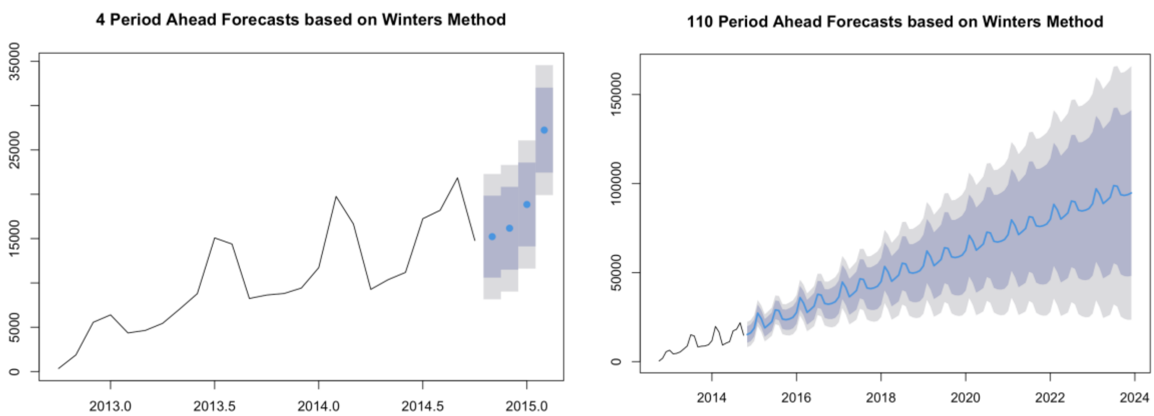


Figure 16: 4 & 110 period ahead forecasting based on Winters Method (Check-In)

4.2 Random Forest

The objective of this analysis is to gain insights into factors influencing hotel demand and develop a predictive model to forecast demand based on search query parameters and hotel property characteristics. The analysis leverages a comprehensive dataset from Expedia, encompassing over 6 million records and a diverse range of features.

Substantial missing data was observed across several features, particularly those related to competitor rates and visitor historical data. To address the issue of missing data, a threshold of 70% was set, and features with missingness exceeding this threshold were removed from the analysis. This decision was driven by the infeasibility of reliable imputation for extensively missing data and the potential introduction of bias or inaccuracies.

4.2.1 Preprocessing Pipeline

A tailored preprocessing pipeline was constructed to prepare the data for modelling with a Random Forest classifier. The pipeline included the following steps:

a) Numerical Data Preprocessing:

Median imputation for handling remaining missing values in numerical features, as it is more robust to outliers compared to mean imputation. Standard scaling to normalise numerical features and ensure equal contribution to the model.

b) Categorical Data Preprocessing:

Imputation of the most frequent category for missing values. One-hot encoding to transform categorical variables into a binary matrix, enabling the model to leverage this information effectively.

The numerical and categorical preprocessing steps were integrated into a ColumnTransformer, ensuring appropriate transformations for each feature type.

4.2.2 Demand Indicator Construction

To facilitate the analysis of demand patterns, a composite ‘search_score’ was constructed based on search parameters such as the number of adults and children, number of rooms, booking window, and whether the search included a Saturday night. This score was normalised between 0 and 1.

A binary ‘high_demand’ indicator was then derived from the normalised search score and price data, with searches classified as high demand if the score exceeded 0.05 and either the price was above $200 USD or a promotion flag was present.

4.2.3 Demand Patterns Analysis

The analysis of the ‘high_demand’ indicator (Figure 17) revealed that approximately 23.79% of searches were classified as high demand, while the remaining 76.21% were low demand.

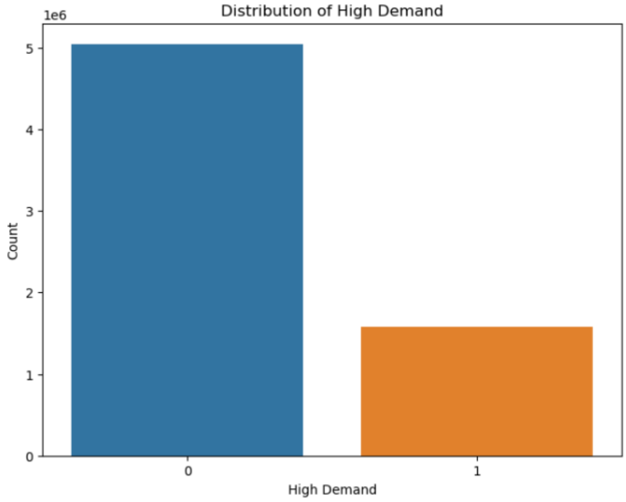


Figure 17: Distribution of High Demand

Comparisons between the two demand categories (Figure 18) highlighted significant differences in average price and average normalised search score:

- Average price for low-demand searches: $158.96 USD
- Average price for high-demand searches: $386.50 USD
- Average normalised search score for low-demand searches: 0.079
- Average normalised search score for high-demand searches: 0.161

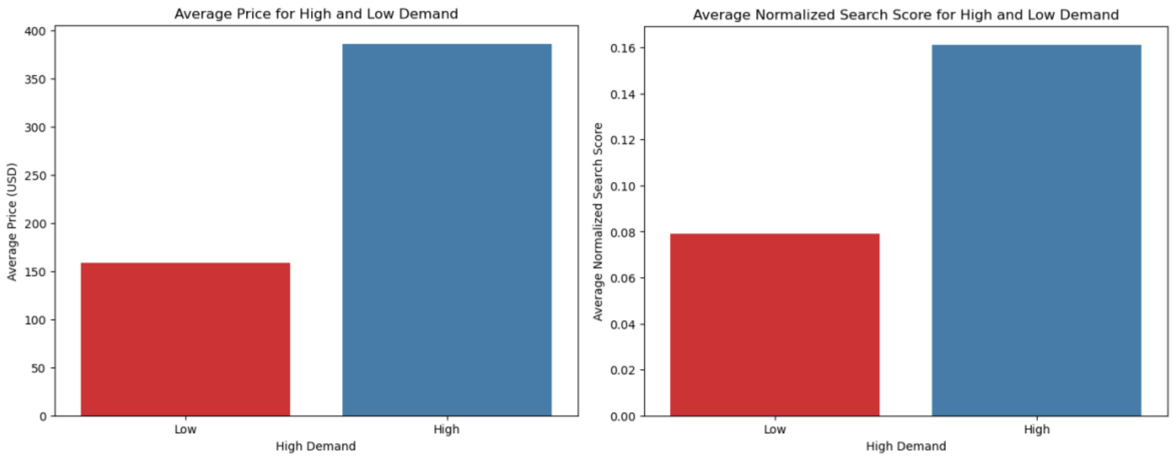


Figure 18: Comparisons between the two demand categories

These findings validate the effectiveness of the constructed search_score and the high_demand indicator in capturing factors influencing hotel demand.

4.2.4 Modelling and Performance Evaluation

Random Forest Classifier

A Random Forest classifier was chosen for its versatility in handling mixed data types and its inherent resilience against overfitting. The model's configuration was fine-tuned to optimise performance, with a focus on the number of estimators (n_estimators) and maximum depth (max_depth) parameters.

Hyperparameter Tuning

The model's performance was evaluated across various configurations of n_estimators and max_depth. Key findings include:

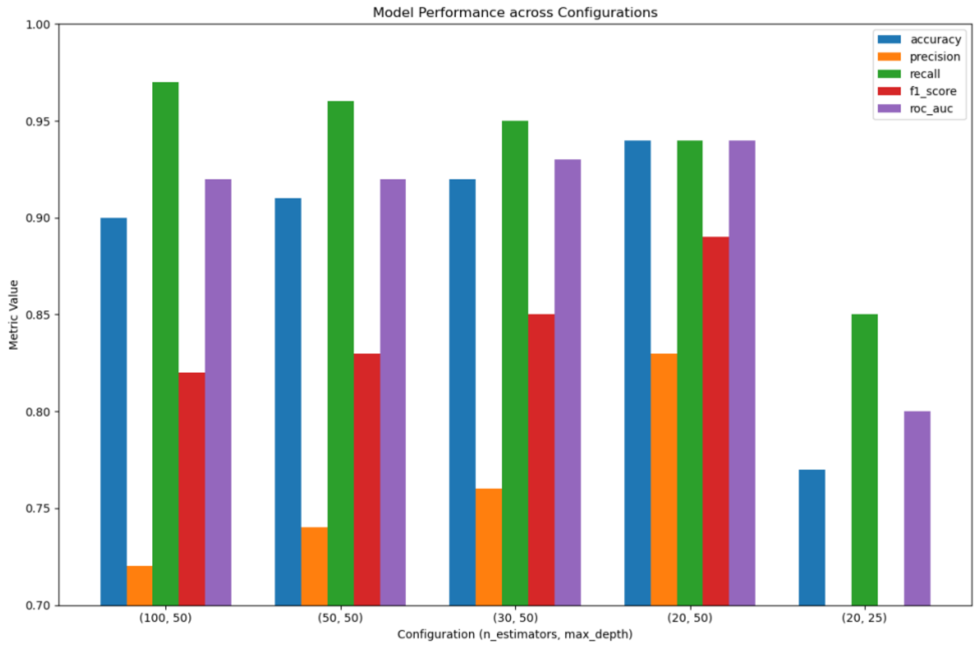
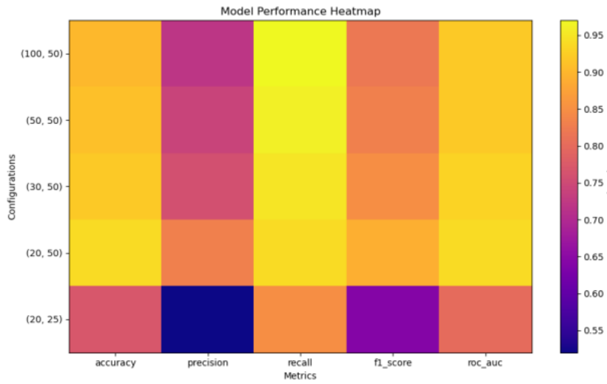


Figure 19: Model Performance Across Configurations

The configuration (Figure 19) with **n_estimators=20 and max_depth=50** emerged as the optimal setup, achieving an **accuracy of 0.94, precision of 0.83, recall of 0.94, F1 score of 0.89, and ROC-AUC score of 0.94.**

Reducing the max_depth to 25 led to a decline in model performance, indicating that a certain level of complexity is necessary for the model to capture underlying patterns effectively.

4.2.5 Conclusion and Recommendations



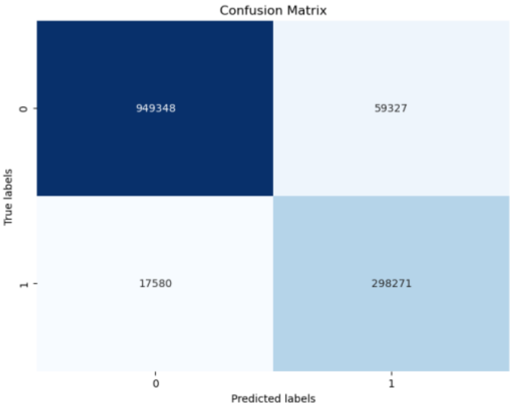


Figure 20: Model Performance Heatmap and Confusion Matrix

The analysis successfully identified critical factors influencing hotel demand, such as search parameters and price sensitivity. The developed Random Forest model, with the optimal configuration of n_estimators=20 and max_depth=50, demonstrated strong predictive performance and the ability to generalise well. (Figure 20)

The findings affirm the significance of search parameters and price sensitivity in determining demand, offering a granular view that can aid in strategic decision-making. Moreover, the high predictive performance of the model indicates its potential utility in forecasting demand scenarios, enabling more informed planning and resource allocation.

4.3 Competitor Analysis

We have conducted a competitor analysis from the dataset, showing the number of hotels from Expedia that have the same price, lower price and higher price as compared to its competitors.

Expedia	**No. Same Price Hotel**	**No. Lower Price Hotel**	**No. Higher Price Hotel**
Company 1	54011	88654	12399
Company 2	2149965	335169	214518
Company 3	1605520	231442	195437
Company 4	278081	66113	73592
Company 5	2165285	439328	371413
Company 6	216188	75719	32726
Company 7	273631	108168	44791
Company 8	1974286	208824	362085

Table 1: Price Table

From Table 1, Expedia has a noticeably higher number of hotels with higher prices compared to Company 4 and Company 8. This suggests that these companies may be offering more competitive pricing options, potentially appealing to cost-conscious travellers. It also indicates that Company 4 and Company 8 may be offering an aggressive pricing strategy to attract customers. This could include discounts, promotions, or negotiated rates with hotel partners to offer more budget-friendly options.

5. Recommendations

According to the data above, it is forecasted that Expedia will continue observing a downward trend concerning the number of searches on the site. However, the forecast for the number of check-ins is expected to increase instead. This could mean that customers of the site might not be new users searching for new destinations to travel to, but rather returning users who are booking places that they have booked and been to before. While having strong customer retention is beneficial, our group has a few recommendations regarding how Expedia can further expand and increase revenue by increasing the number of new users, which would help tackle the problem of the decreasing number of new searches.

5.1 Marketing

By investing in advertisements, social media presence and search engine optimisation, Expedia will be able to increase its visibility on the internet, making its presence known and thus attracting more users to its website when they wish to book a trip. Examples of this would be influencer marketing, Google Ads and YouTube ads. With the increasing usage of social media and other online media platforms among young people, it is imperative that Expedia hops onto the bandwagon to attract a new generation of users.

5.2 Packages and Partnerships using Association Rules

Another way to attract more users to the site would be to offer attractive packages and deals to its users by using Association Rules to find out the frequent items related to travelling packages. As of the present, Expedia does not offer many packages on the site. Expedia may use Association Rules to analyse the frequently purchased items together with hotels, such as flight tickets, travel insurance, parking vouchers or breakfast etc. This could help Expedia to further understand customer preferences and make informed decision-making.

With better understanding of customer preferences, Expedia could partner with attractions in the area to create package deals that users can purchase, which would allow users to not only discover new areas to explore, but also give them ideas on how they can explore the area. An example of this would be to provide vouchers for certain attractions in the area of where users book their accommodations.

Alternatively, Expedia could partner or leverage on events happening in the area. An example of this would be when Klook offered hotel stay packages tagged to tickets for the Taylor Swift Eras Tour in Singapore. Fans could purchase tickets that were tagged to a hotel stay via the Klook website and app, which garnered the attention of many who were vying for tickets to the concert. Packages like these could increase the outreach of Expedia to larger audiences, making it more likely for these people to consider Expedia in their options when planning their trips in the future.

5.3 Corporate Partnerships

Expedia could also reach out to companies and offer corporate packages for company welfare, company trips, business trips etc. Expedia could also offer special packages and privileges to employees of partner organisations. The use of Expedia in the workplace could encourage employees to use the platform for their own personal trips, or even encourage their friends and families to use the platform as well, effectively increasing Expedia’s outreach and likely revenue.

By increasing the number of new users on the platform, it is likely that Expedia would see an increase in the number of searches on the platform, which would lead to platform growth when these searches are converted to bookings.

6. Limitations and Future Considerations

The limited availability and dated nature of the datasets utilised have impacted the accuracy of our demand forecasting models, as evidenced by a larger RMSE than desired. This emphasises the importance of having access to more recent and comprehensive data to improve forecasting accuracy.

To address this, it is recommended for future efforts to focus on enriching the dataset. Leveraging advanced data extraction techniques and tapping into real-time data sources can provide a more current and detailed basis for analysis. This is crucial for ensuring the forecasting models are accurately reflecting current market dynamics and adapting to changing trends.

Additionally, it is imperative to address missing values pertaining to competitors' data within our dataset. Incorporating information about competitors is crucial for gaining a holistic understanding of market conditions and refining the accuracy of our forecasts.

These improvements are necessary to generate precise forecasts, which in turn support better strategic planning and decision-making.

7. Conclusion

The decline in Expedia's search volume reflects a shifting landscape within the tourism market, where competitors have gained traction, thereby splintering Expedia's market share. This trend is underscored by the notable decrease in revenue growth observed since 2021. Despite still holding a significant portion of the market, it's evident that Expedia's current pricing strategies may not be fully optimised to effectively compete in this evolving landscape.

To regain lost ground and maintain its leadership position, Expedia must undertake a comprehensive review of its pricing strategies. This entails not only revisiting pricing models but also considering factors such as value proposition, customer segmentation, and competitor pricing analysis. By leveraging insights gained from these assessments, Expedia can implement targeted pricing adjustments that align with market dynamics and consumer preferences.

References

Andre, J. (2023, August 16). *Q3 2023 Traveler Insights: Traveler enthusiasm holds strong*. Expedia Group Media Solutions. https://advertising.expedia.com/blog/travel-trends/quarterly-traveler-insights-report-traveler-enthusiasm-holds-strong/

Andre, J. (2023, November 14). *Traveler Insights Q4 2023: Traveler confidence steadies*. Expedia Group Media Solutions. https://advertising.expedia.com/blog/travel-trends/traveler-insights-q4-2023-traveler-confidence-steadies/

Andre, J. (2024, February 27). *Traveler insights Q1 2024: Top destinations and event trends*. Expedia Group Media Solutions. https://advertising.expedia.com/blog/travel-trends/traveler-insights-q1-2024-top-destinations-event-travel-trends/

“Booking Holdings: revenue 2023.” *Statista*, Statista, 23 February 2024, https://www.statista.com/statistics/225455/booking-holdings-total-revenue/.

“Booking Holdings: revenue 2023.” *Statista*, Statista, 23 February 2024, https://www.statista.com/statistics/225455/booking-holdings-total-revenue/

Macrotrends. “Expedia Revenue 2010-2023 | EXPE | MacroTrends.” *Macrotrends*, 2023, https://www.macrotrends.net/stocks/charts/EXPE/expedia/revenue.

Walker, Ben. “Online Travel Agencies Market Share Across the World.” *Online Travel Agencies Market Share Across the World - Mize*, 17 February 2021, https://mize.tech/blog/online-travel-agencies-market-share-across-the-world/.

Hand, Rachel. “Demand Forecasting: How to Forecast Demand [+ Examples].” *ShipBob*, 22 February 2024

https://www.shipbob.com/blog/demand-forecasting/

Expedia Groups. (2017, November 9). YouTube: Home. https://partner.expediagroup.com/content/dam/unified/partner/documents/reports/2023-reports/report-path-to-Purchase-2023-final_en-us.pdf

Expedia Groups. (2022). *2022 | Q2 - Traveler Insights Report*. Expedia Group Media Solutions.https://go2.advertising.expedia.com/rs/185-EIA-216/images/Expedia_Group_Media_Solution_Q2_2022_Traveler_Insights_Report.pdf

Expedia Groups. (2022). *2022 | Q4 - Traveler Insights Report*. Expedia Group Media Solutions.https://go2.advertising.expedia.com/rs/185-EIA-216/images/Expedia-Group-Media-Solutions-Q4-2022-Traveler-Insights-Report.pdf

Expedia Groups. (2022, November 9). YouTube: Home. https://go2.advertising.expedia.com/rs/185-EIA-216/images/Q1-2022-MeSo-RecoveryTrendReport-FINAL.pdf

Expedia Groups. (2023). *Introducing the most unexpected travel trends of 2023*. Expedia. https://www.expedia.com.sg/stories/wp-content/uploads/2022/11/Expedia_Travel-Trends-Report_SG.pdf

Expedia Groups. (2023). *Traveler Insights Report*. Expedia Group Media Solutions. https://go2.advertising.expedia.com/rs/185-EIA-216/images/Q1-2023-Traveler-Insights-Report.pdf

Expedia Groups. (2023). *2022 | Q2 - Traveler Insights Report*. Expedia Group Media Solutions. https://go2.advertising.expedia.com/rs/185-EIA-216/images/Expedia_Group_Media_Solution_Q2_2022_Traveler_Insights_Report.pdf

Appendix

Appendix A - Random Forest

1. Insights and Implications

The initial EDA highlights the complexity and richness of the Expedia dataset, underscoring the challenges in managing missing data and outliers. Sparse competitor data and extreme pricing variations suggest potential limitations in capturing comprehensive market dynamics. Addressing these challenges through careful data preprocessing and modelling strategies is essential for deriving meaningful insights and accurate predictions of hotel demand.

2. Creating a Demand Indicator and Analysing Demand Patterns

2.1 Constructing a Demand Indicator

To better understand and predict hotel demand, a composite search_score is introduced based on several search-related parameters, including the number of adults and children, number of rooms, booking window normalised by week, and inclusion of a Saturday night. This score is then normalised to a range between 0 and 1. Subsequently, a binary high_demand indicator is defined based on the normalised search score and price data, categorising searches as high demand if certain conditions are met.

```
df['search_score'] = (
    df['srch_adults_count'] +
    df['srch_children_count'] / 2 +  # assuming children count less towards demand
    df['srch_room_count'] +
    df['srch_booking_window'] / 7 +  # normalized by week
    df['srch_saturday_night_bool']
)

# Normalize the search score to a 0-1 range
max_score = df['search_score'].max()
min_score = df['search_score'].min()
df['normalized_search_score'] = (df['search_score'] - min_score) / (max_score - min_score)

# Step 2: Incorporate price data
# Adjusting the threshold for normalized_search_score and price_usd
df['high_demand'] = ((df['normalized_search_score'] >= 0.05) &
                     ((df['price_usd'] >= 200) | (df['promotion_flag'] == 1))
                    ).astype(int)
```

2.2 Demand Indicator Distribution

Analysis of the distribution of the high_demand indicator reveals that approximately 23.79% of searches are classified as high demand, while the remaining 76.21% are categorised as low demand. This categorization provides a foundation for understanding the characteristics and drivers of high-demand searches.

3. Data Preprocessing and Model Preparation

3.1 Initial Steps and Feature Selection

To address the challenges posed by missing data, a preliminary decision is made to focus on features with substantial completeness. Features with excessive missing data (>70%) are identified and dropped, including detailed visitor history metrics and competitor rate information. This ensures that only informative features with sufficient data are retained for further analysis and modelling.

4. Model Training, Evaluation, and Performance Analysis

4.1 Detailed Process Overview

The model training process involves configuring and training a Random Forest Classifier, followed by evaluating its performance across various hyperparameter settings. The Random Forest Classifier is chosen for its ability to handle complex datasets with mixed data types and its inherent robustness against overfitting.

4.2 Hyperparameter Tuning and Model Performance

To optimise the Random Forest model's performance, various hyperparameter configurations are explored and evaluated. Key hyperparameters include n_estimators, max_depth, class_weight, n_jobs, and random_state.

Configuration	n_estimators	max_depth	Accuracy	Precision	Recall	F1 Score	ROC-AUC Score
n_estimators=100, max_depth=50	100	50	0.90	0.72	0.97	0.82	0.92
n_estimators=50, max_depth=50	50	50	0.91	0.74	0.96	0.83	0.92
n_estimators=30, max_depth=50	30	50	0.92	0.76	0.95	0.85	0.93
n_estimators=20, max_depth=50	20	50	0.94	0.83	0.94	0.89	0.94

n_estimators=20, max_depth=25	20	25	0.77	0.52	0.85	0.64	0.80

4.3 Insights and Optimal Configuration

Analysis suggests a configuration with n_estimators=20 and max_depth=50 as most effective, offering a robust model capable of accurately predicting hotel demand.

4.4 Conclusion and Recommendations

The optimised Random Forest model emerges as a potent tool for predicting hotel demand, enabling informed strategic decisions to optimise hotel offerings and capitalise on market opportunities.

Appendix B - Key Insights from the Expedia Travel Insights Reports for Q1 2022 to Q1 2024:

1. Travel Demand and Search Trends

- Global search volumes saw significant year-over-year increases, up 75% in Q1 2022 and continuing strong growth through 2023 and into early 2024. Quarterly increases were around 25% in Q1 2022 and Q2 2022.
- Searches from most regions showed robust growth, with APAC, EMEA, and NORAM leading the way. APAC in particular rebounded strongly as travel restrictions eased.
- Longer search windows of 61-90 days and 180+ days grew rapidly, indicating increased advance planning for travel. The 180+ day window nearly tripled in Q1 2022 versus the prior year.
- Searches shifted away from the 0-21 day window as travellers planned further ahead. Searches in the 22-60 day range also increased.
- Holiday travel demand spiked, with 60%+ year-over-year growth for Thanksgiving/Christmas 2022 and sustained interest for 2023 holiday periods.

2. Top Destinations

- Major cities like New York, London, Paris remain hugely popular booking destinations across all regions throughout 2022-2024.
- Beach destinations like Cancun, Punta Cana, Honolulu saw high demand, especially for travellers from NORAM and LATAM.
- Asia Pacific destinations like Tokyo, Osaka, Seoul ranked highly in global and APAC regional top bookings lists.

- Fast-growing destinations included several in Italy, Turkey, Australia and Brazil's beach cities based on increasing search demand.

3. Lodging Trends

- Global hotel bookings grew 35% quarter-over-quarter in Q1 2022 and remained strong, hitting record highs in Q2 2022 with APAC leading growth.
- Vacation rental night demand increased substantially, with properties in warm beach locations being top picks across most regions.
- Average daily hotel rates increased year-over-year, while vacation rental length of stay averaged 5.5 days.
- London broke into the top 10 booked vacation rental destinations for APAC and EMEA travellers in 2022.

4. Consumer Behaviour

- Traveller confidence and enthusiasm grew through 2022 into 2023 after pandemic impacts, with most consumers expecting to take multiple leisure trips.
- Demand for inclusive travel options, adventure destinations, and sustainability information increased.
- About 30% showed interest in bleisure (business+leisure) trips and "flexcations" that blend work and travel.
- Influences like friends/family, brands, and social media played a bigger role in inspiring travel.
- Consumers prioritised aspects like refundability, low pricing, flexibility and better overall experiences when making bookings.
- While sustainable travel became more important to consumers, many felt overwhelmed on taking action.

5. Emerging Trends

- Event travel regained popularity, with increases in searches surrounding major sporting events, concerts and tours like Taylor Swift's 2024 Eras Tour.
- Tour tourism saw a rise, with over 70% saying they were more likely to travel for concerts or shows.
- "Flexcations" that combine work and travel became more mainstream, especially for remote workers taking longer trips.
- Mobile usage grew for travel planning, with increases in mobile bookings and app usage share.
- Destinations in China like Shanghai, Guangzhou and Hong Kong saw surging global search interest in early 2024 as travel normalised.

6. Looking Ahead

The travel industry shows incredible resiliency based on the recovery outlined in these reports. Pent-up demand translated to surging searches and bookings through 2022 and 2023. Consumers enthusiastically planned leisure trips, both domestically and internationally, after pandemic limitations were lifted.

As we move through 2024, expect demand to remain robust given the prioritisation of travel and consumers' intent to continue taking multiple trips annually. Emerging trends like event travel, tour tourism, bleisure/flexcations, and an emphasis on inclusivity and sustainability point the way forward.

Marketers should align strategies with these consumer preferences and leverage tools like targeted advertising, mobile marketing, and first-party data to effectively reach an audience that values refundable fares, low pricing, flexibility and enhanced overall experiences when booking travel.