Business Presentation

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Contents

- Problem Definition
- Key findings and Insights
- Solution
- Potential benefits of implementing solution

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Problem Overview and Solution Approach

Perform the data analysis to extract actionable insights from the data on different attributes of customer's booking details in the INN Hotels Group portal.
We will be majorly focusing on the below areas.

- Review the demand for hotel bookings in the INN Hotels portal
- Room preference of the INN Hotels customers
- Understand the impact meal plans might have on hotel bookings or cancellation
- Understand the volume of the bookings over weekdays and weekends
- Estimate the revenue generated by the company
- Review the company policy on booking cancellations
- Review impact of policy of dynamic room pricing
- Proffer solutions on minimizing impact/loss on booking cancellations

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Overview

Variable	Description
Booking ID	Unique identifier for each booking
No of adults No of children	Number of adults Number of children
No of weekend nights No of weeknights	Number of weekend nights Number of weeknights
Type of meal plan Not Selected Meal Plan 1 Meal Plan 2 Meal Plan 3	Meal plan booked by customer No meal plan selected Breakfast Breakfast one other meal Full breakfast, lunch, dinner
Required car-parking space	Parking required
Lead time arrival year Arrival month Arrival date	Days between booking and arrival Arrival year Month of arrival date Date of month
Market segment Repeated guest	Market segmentation designation Customer repeat guest
No of previous cancellations No of previous bookings not cancelled Avg price per room	Previous bookings cancelled Previous bookings not cancelled Average price per day
No of special requests Booking status	Special requests made Flag indicating booking cancellation

Observations	Variables
36275	19

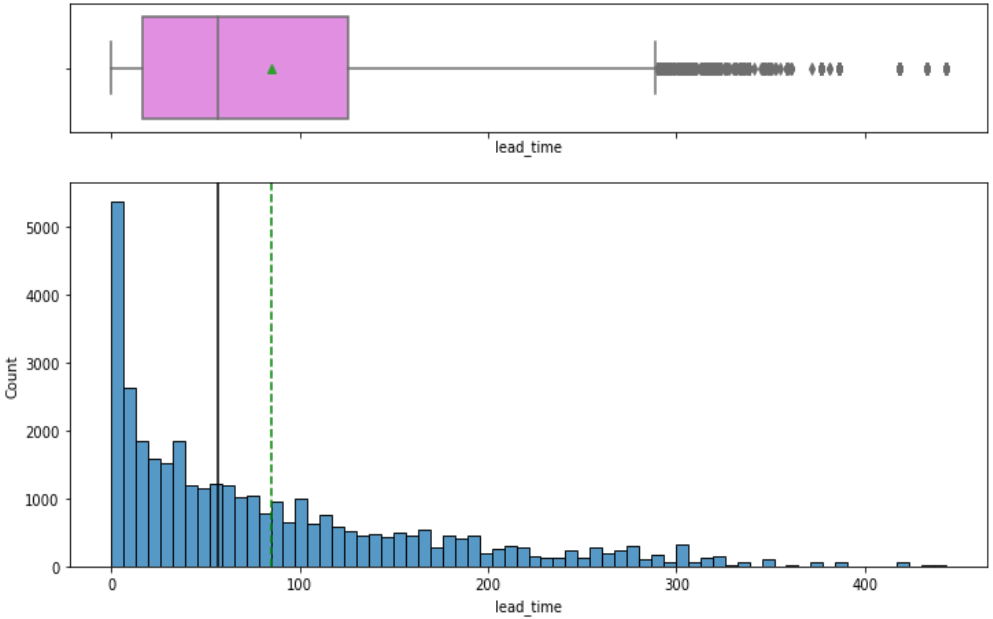
Note:

- There are no missing values in the dataset.
- There are no duplicate entries in the dataset
- The Booking ID which has data type object has been dropped from the database

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univarate Analysis

- Observations on Lead time

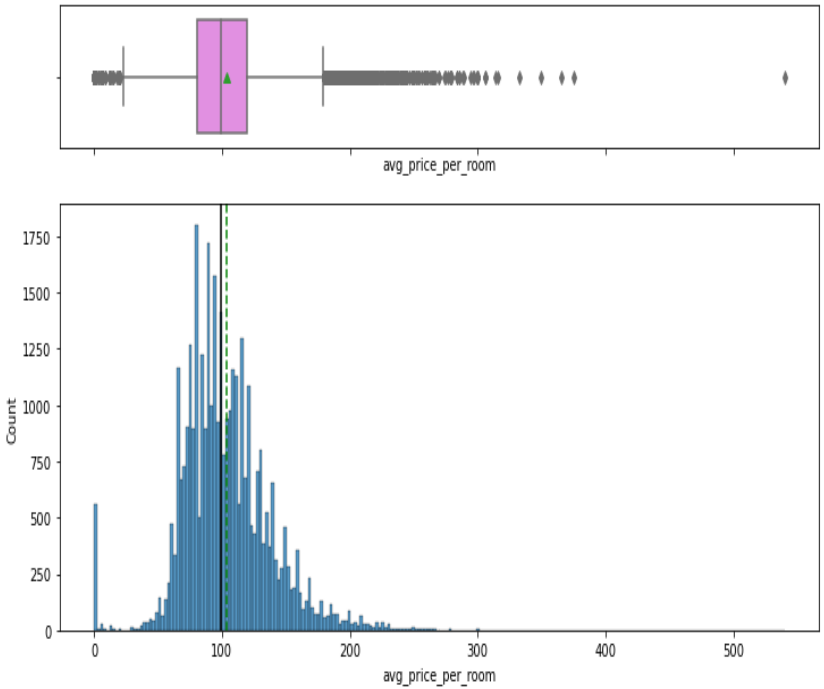


- About 75% if the bookings are below the 3rd quartile while only 25% are below the upper whisker.
- Below 75% of the bookings have a large number of days of approximately 120 days between booking and arrival date
- About 80% of the bookings appear to be below the median lead time
- Suggesting clients make their booking well in advance of their arrival dates

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univarate Analysis

- Observations on Average Price per room

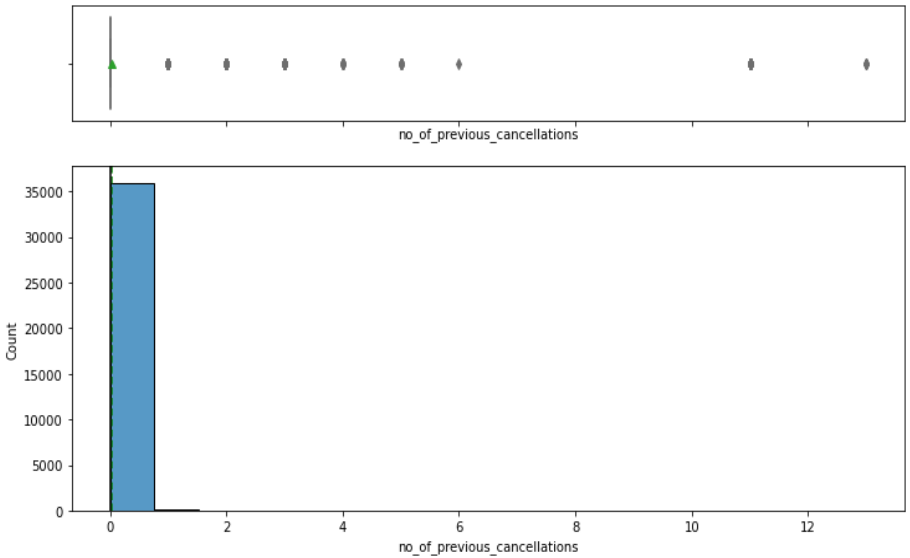


- The distribution of average price per room is slightly right skewed
- There are outliers in this variable
- From the boxplot we can see that the third quartile(Q3) is more than 100 which means that 75% of customers pay more than 100 dollars on the average price per room.
- On average customers pay about 100 dollars per room.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univarate Analysis

- Observations on number of previous booking cancellations

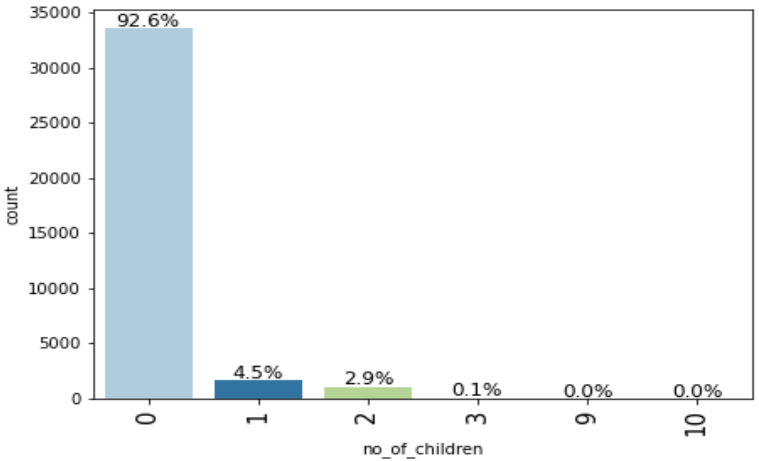


- The distribution on number of previous booking cancellations show most cancellations are not repeat cancellations
- With most cancellations occurring less than 2 times by a customer

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

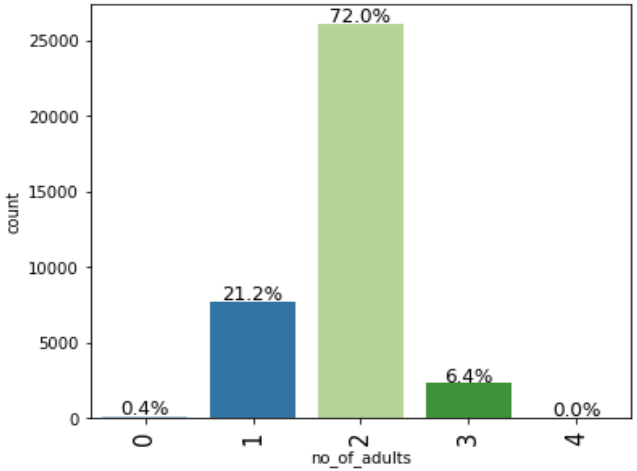
Univarate Analysis

- Number of Children



- 92.6% of the customers who make booking have no children included
- 4.5% of bookings consists of at least one child

- Number of Adults

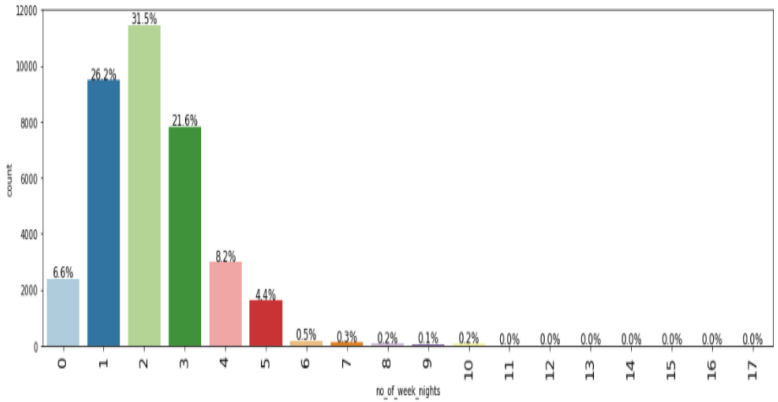


- 72% of the bookings consist of at least 2 adults
- 21.2% of bookings consist of at least 1 adult

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

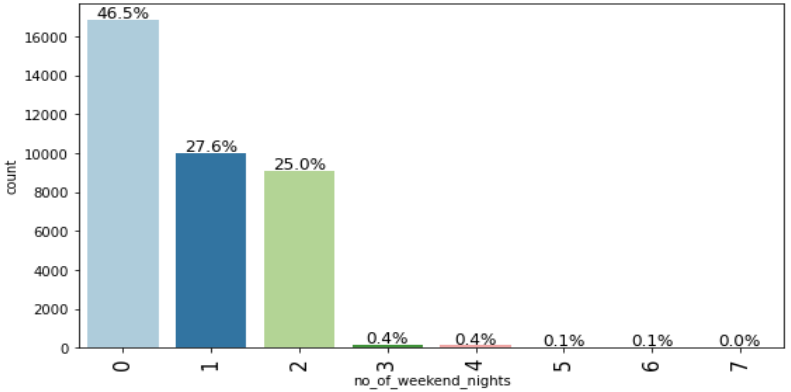
Univarate Analysis

- Observations on number of week nights



- 31.5% of bookings consist of 2 week night bookings
- At least 79.3% of total bookings consist of at least 1 week night booking

- Observations on number of weekend nights

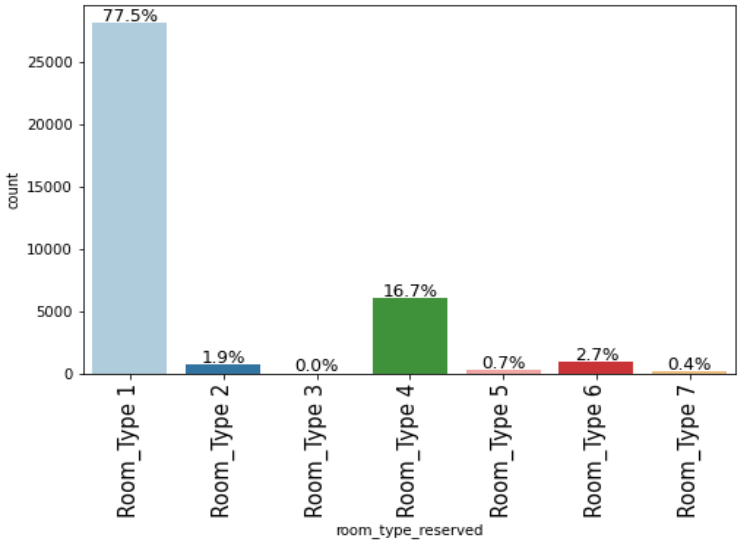


- 46.5% of bookings do not consist of a weekend night booking
- At most 52.6% of bookings consist of one weekend night booking

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univarate Analysis

- Observations on room type reserved



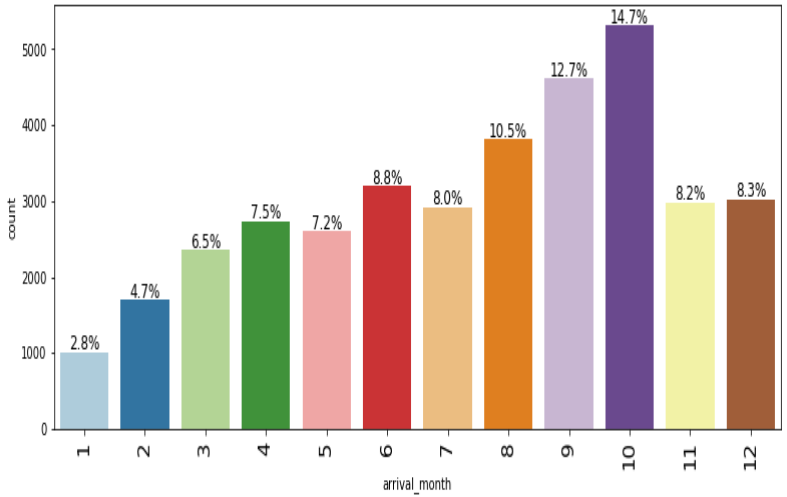
Given the information on room type is encoded by the client only the following can be deciphered from the information provided

- Room type 1 with 77.5% appears to be the most popular type of room reserved by customers
- Room type 4 with 16.7% appears to be the second most popular type of room reserved by customers
- Room type 3 with 0.0% is the least most popular type of room reserved by customers

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

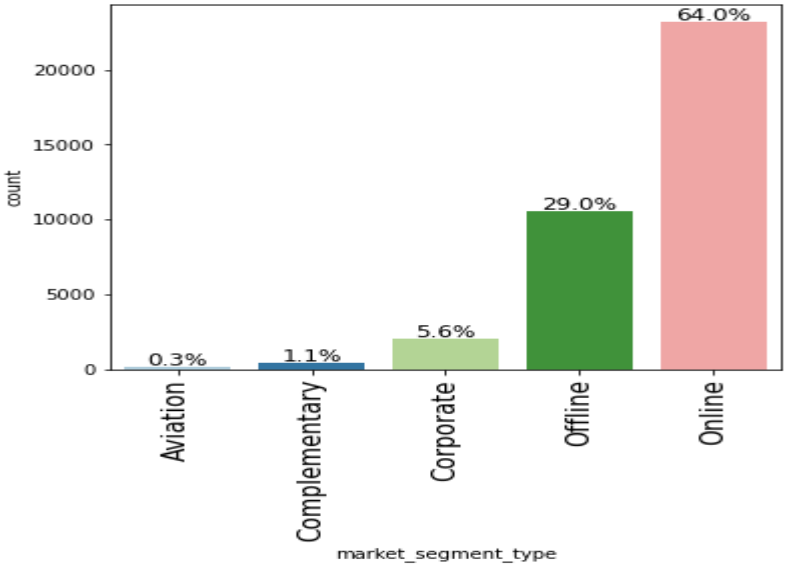
Univarate Analysis

- Observations on arrival month



- The most popular arrival month is month 10 at 14.7% followed by month 9 at 12.7%
- The least most popular month is month 1 at 2.8% followed by month 2 at 4.7%
- Month 7 to Month 12 account for 60% of arrivals

- Observations on market segment type

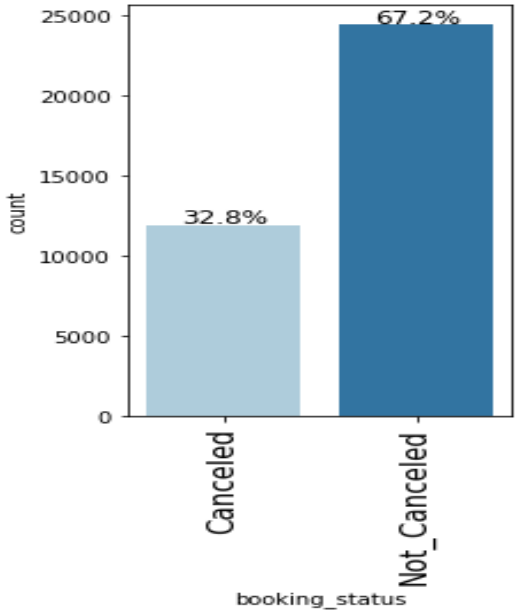


- Online customers account for 64.0% of the market segment followed by offline at 29%
- Aviation customers have the least market segment at 0.3%

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Univarate Analysis

- Observations on Booking Status

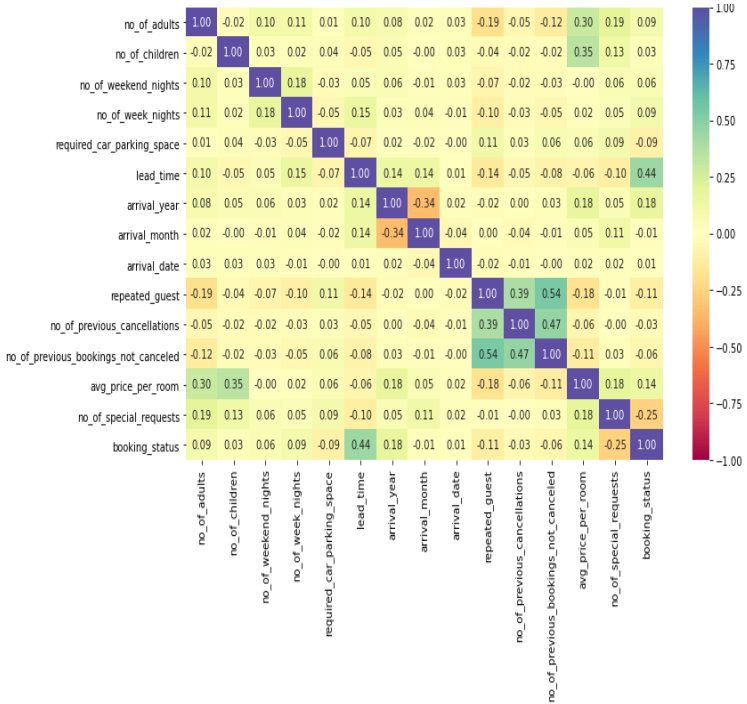


- 67.2% of booking do not cancel
- 32.8% of booking cancel which is a significantly high figure

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Bivarate Analysis

- Heat Map

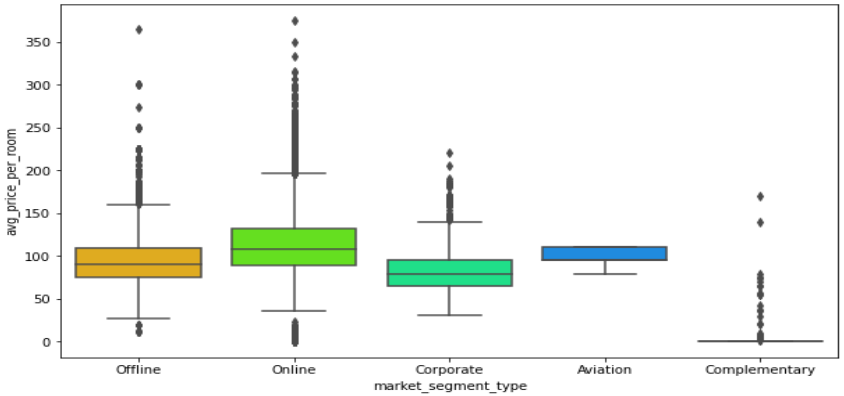


- There appears to be a positive correlation between booking status and lead time at 0.44 thus estimating as lead time increase the booking status would increase
- The correlation between average price per room and no of adults and no of children is at 0.30 and 0.35 thus estimating that as the number of people in the room increases the average price should increase
- There appears to be a positive correlation between bookings not cancelled and repeat guests at 0.54
- There is a negative correlation between repeated guests and avg room pricing, suggesting that an increase in price would affect the repeated guests booking

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

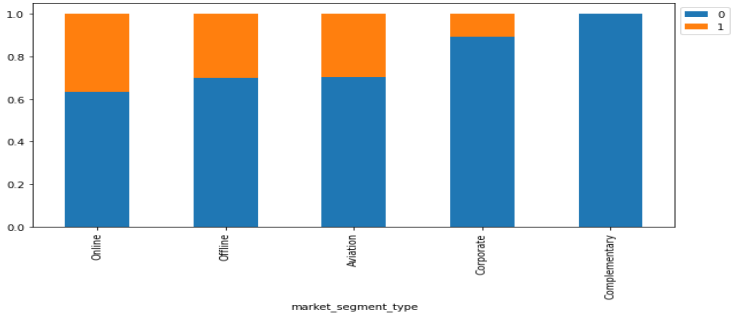
Bivarate Analysis

- Prices vary across different segments



- Online customers have the highest minimum and maximum value average price per room
- Online customers have the highest median price per room with significant outliers
- Corporate customers have the second minimum price followed by offline customers
- Offline customers have significant outliers in their data

- Booking status vary across different segments

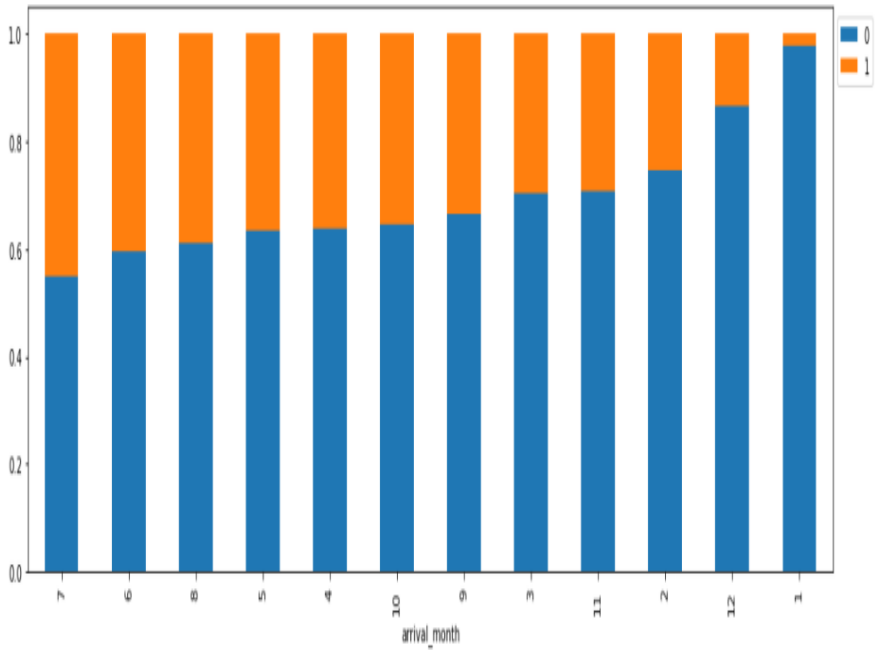


- Online customers appear to have the most booking cancellations
- Corporate customers appear to have the least booking cancellations
- Offline customers and aviation customers are estimated to have relatively same number of cancellations

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Bivarate Analysis

- Percentage of Bookings cancelled each month

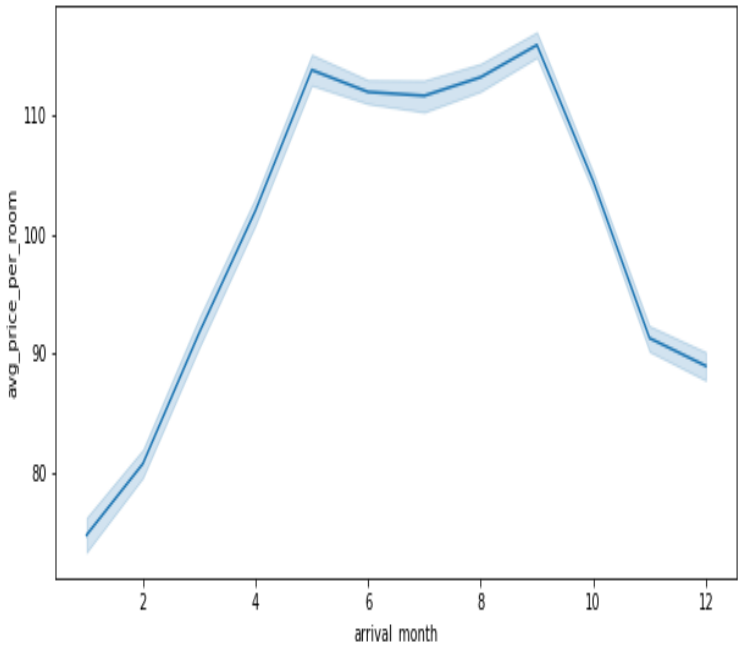


- The highest cancellations appear to be in month 7 and month 6 respectively
- However these month show a drop in level of business
- The lowest cancellation being in month 1 and month 12 respectively
- The busiest month for the hotel are from month 8 to month 11 peaking at month 10
- However these months also have very high cancellation rates

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Bivarate Analysis

- Dynamic Prices across month

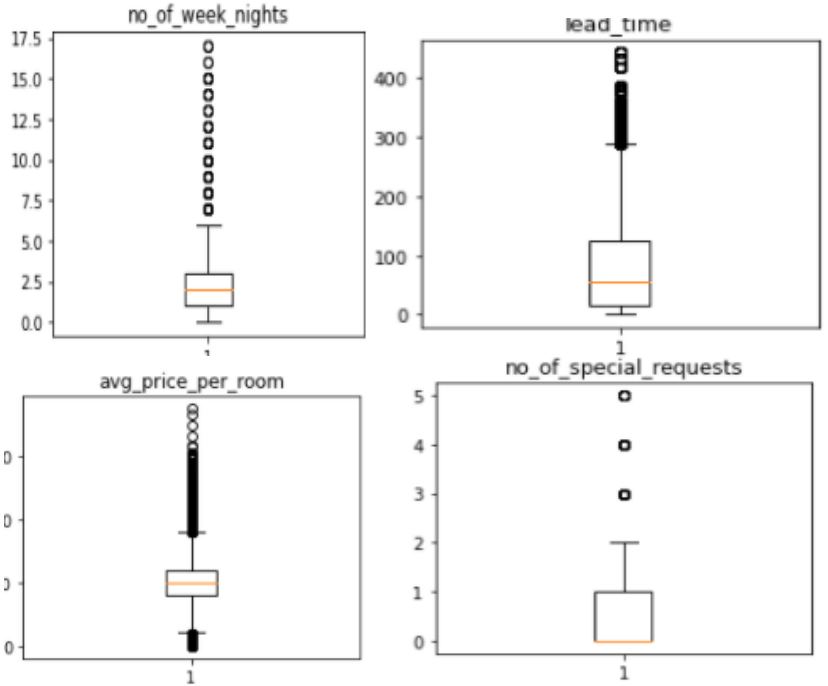


- The highest prices appear to be between month 5 and month 9
- Following which prices begin to fall significantly from month 10 to 12
- However, busiest months are between month 8 and 11 which based on dynamic prices have low prices
- While month 6 and 7 have the highest cancellation rates but also have the higher prices during that period

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Pre-processing

- Missing value treatment: There are no missing values
- Outlier Detection: Number of week nights, lead time, price per room and no of special requests have upper outliers
- Outlier Detection: Average price per room has both upper and lower outliers



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Building-Logistic Regression

Training performance:

	Accuracy	Recall	Precision	F1
0	0.80600	0.63410	0.73971	0.68285

- Accuracy of 0.80 high accuracy of model
- Precision of 0.73 shows correctly predicted positive observation to the total predicted positive observations
- Recall of 0.63 which is the ratio of correctly predicted positive observations to all the observations in that calls. Greater than 0.5 it is satisfactory
- F1 score of 0.68 weighted average of precision and recall is adequate

Logit Regression Results

Dep. Variable:	booking_status	No. Observations:	25392
Model:	Logit	Df Residuals:	25364
Method:	MLE	Df Model:	27
Date:	Fri, 14 Jan 2022	Pseudo R-squ.:	0.3292
Time:	00:01:20	Log-Likelihood:	-10794.
converged:	False	LL-Null:	-16091.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-922.8266	120.832	-7.637	0.000	-1159.653	-686.000
no_of_adults	0.1137	0.038	3.019	0.003	0.040	0.188
no_of_children	0.1580	0.062	2.544	0.011	0.036	0.280
no_of_weekend_nights	0.1067	0.020	5.395	0.000	0.068	0.145
no_of_week_nights	0.0397	0.012	3.235	0.001	0.016	0.064
required_car_parking_space	-1.5943	0.138	-11.565	0.000	-1.865	-1.324
lead_time	0.0157	0.000	58.863	0.000	0.015	0.016
arrival_year	0.4561	0.060	7.617	0.000	0.339	0.573
arrival_month	-0.0417	0.006	-6.441	0.000	-0.054	-0.029
arrival_date	0.0005	0.002	0.259	0.796	-0.003	0.004
repeated_guest	-2.3472	0.617	-3.806	0.000	-3.556	-1.139
no_of_previous_cancellations	0.2664	0.086	3.108	0.002	0.098	0.434
no_of_previous_bookings_not_canceled	-0.1727	0.153	-1.131	0.258	-0.472	0.127
avg_price_per_room	0.0188	0.001	25.396	0.000	0.017	0.020
no_of_special_requests	-1.4689	0.030	-48.782	0.000	-1.528	-1.410
type_of_meal_plan_Meal Plan 2	0.1756	0.067	2.636	0.008	0.045	0.306
type_of_meal_plan_Meal Plan 3	17.3584	3987.814	0.004	0.997	-7798.614	7833.331
type_of_meal_plan_Not Selected	0.2784	0.053	5.247	0.000	0.174	0.382
room_type_reserved_Room_Type 2	-0.3605	0.131	-2.748	0.006	-0.618	-0.103
room_type_reserved_Room_Type 3	-0.0012	1.310	-0.001	0.999	-2.568	2.566
room_type_reserved_Room_Type 4	-0.2823	0.053	-5.304	0.000	-0.387	-0.178
room_type_reserved_Room_Type 5	-0.7189	0.209	-3.438	0.001	-1.129	-0.309
room_type_reserved_Room_Type 6	-0.9501	0.151	-6.274	0.000	-1.247	-0.653
room_type_reserved_Room_Type 7	-1.4003	0.294	-4.770	0.000	-1.976	-0.825
market_segment_type_Complementary	-40.5975	5.65e+05	-7.19e-05	1.000	-1.11e+06	1.11e+06
market_segment_type_Corporate	-1.1924	0.266	-4.483	0.000	-1.714	-0.671
market_segment_type_Offline	-2.1946	0.255	-8.621	0.000	-2.694	-1.696
market_segment_type_Online	-0.3995	0.251	-1.590	0.112	-0.892	0.093

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Performance Evaluation and Improvement- Logistic Regression

- Multicollinearity

Assumptions

- Remove multicollinearity from data to get reliable coefficients and p values
- Variance Inflation factor (VIF) measures inflation in the variances of the regression of coefficient rule:
    - VIF =1 no correlation
    - VIF >5 noderate multicollinearity
    - VIF> = 10 High multicollinearity

Logit Regression Results

Dep. Variable:	booking_status	No. Observations:	25392
Model:	Logit	Df Residuals:	25370
Method:	MLE	Df Model:	21
Date:	Fri, 14 Jan 2022	Pseudo R-squ.:	0.3282
Time:	00:06:01	Log-Likelihood:	-10810.
converged:	True	LL-Null:	-16091.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-915.6391	120.471	-7.600	0.000	-1151.758	-679.520
no_of_adults	0.1088	0.037	2.914	0.004	0.036	0.182
no_of_children	0.1531	0.062	2.470	0.014	0.032	0.275
no_of_weekend_nights	0.1086	0.020	5.498	0.000	0.070	0.147
no_of_week_nights	0.0417	0.012	3.399	0.001	0.018	0.066
required_car_parking_space	-1.5947	0.138	-11.564	0.000	-1.865	-1.324
lead_time	0.0157	0.000	59.213	0.000	0.015	0.016
arrival_year	0.4523	0.060	7.576	0.000	0.335	0.569
arrival_month	-0.0425	0.006	-6.591	0.000	-0.055	-0.030
repeated_guest	-2.7367	0.557	-4.916	0.000	-3.828	-1.646
no_of_previous_cancellations	0.2288	0.077	2.983	0.003	0.078	0.379
avg_price_per_room	0.0192	0.001	26.336	0.000	0.018	0.021
no_of_special_requests	-1.4698	0.030	-48.884	0.000	-1.529	-1.411
type_of_meal_plan_Meal Plan 2	0.1642	0.067	2.469	0.014	0.034	0.295
type_of_meal_plan_Not Selected	0.2860	0.053	5.406	0.000	0.182	0.390
room_type_reserved_Room_Type 2	-0.3552	0.131	-2.709	0.007	-0.612	-0.098
room_type_reserved_Room_Type 4	-0.2828	0.053	-5.330	0.000	-0.387	-0.179
room_type_reserved_Room_Type 5	-0.7364	0.208	-3.535	0.000	-1.145	-0.328
room_type_reserved_Room_Type 6	-0.9682	0.151	-6.403	0.000	-1.265	-0.672
room_type_reserved_Room_Type 7	-1.4343	0.293	-4.892	0.000	-2.009	-0.860
market_segment_type_Corporate	-0.7913	0.103	-7.692	0.000	-0.993	-0.590
market_segment_type_Offline	-1.7854	0.052	-34.363	0.000	-1.887	-1.684

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Performance Evaluation and Improvement- Logistic Regression

Training performance:

	Accuracy	Recall	Precision	F1
0	0.80545	0.63267	0.73907	0.68174

- Accuracy of 0.80 high accuracy of model
- Precision of 0.73 shows correctly predicted positive observation to the total predicted positive observations
- Recall of 0.63 which is the ratio of correctly predicted positive observations to all the observations in that calls. Greater than 0.5 it is satisfactory
- F1 score of 0.68 weighted average of precision and recall is adequate

There are no changes in the data after removing multicollinearity

Logit Regression Results

Dep. Variable:	booking_status	No. Observations:	25392
Model:	Logit	Df Residuals:	25370
Method:	MLE	Df Model:	21
Date:	Fri, 14 Jan 2022	Pseudo R-squ.:	0.3282
Time:	00:06:01	Log-Likelihood:	-10810.
converged:	True	LL-Null:	-16091.
Covariance Type:	nonrobust	LLR p-value:	0.000

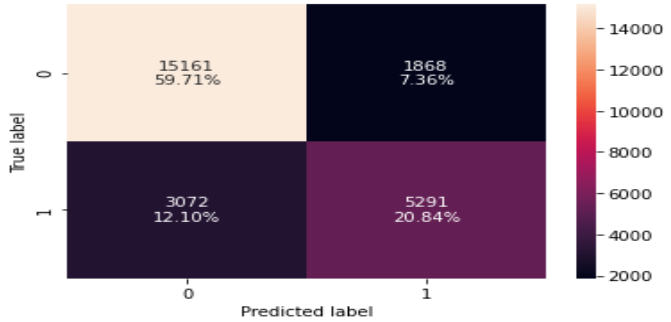
	coef	std err	z	P>|z|	[0.025	0.975]
const	-915.6391	120.471	-7.600	0.000	-1151.758	-679.520
no_of_adults	0.1088	0.037	2.914	0.004	0.036	0.182
no_of_children	0.1531	0.062	2.470	0.014	0.032	0.275
no_of_weekend_nights	0.1086	0.020	5.498	0.000	0.070	0.147
no_of_week_nights	0.0417	0.012	3.399	0.001	0.018	0.066
required_car_parking_space	-1.5947	0.138	-11.564	0.000	-1.865	-1.324
lead_time	0.0157	0.000	59.213	0.000	0.015	0.016
arrival_year	0.4523	0.060	7.576	0.000	0.335	0.569
arrival_month	-0.0425	0.006	-6.591	0.000	-0.055	-0.030
repeated_guest	-2.7367	0.557	-4.916	0.000	-3.828	-1.646
no_of_previous_cancellations	0.2288	0.077	2.983	0.003	0.078	0.379
avg_price_per_room	0.0192	0.001	26.336	0.000	0.018	0.021
no_of_special_requests	-1.4698	0.030	-48.884	0.000	-1.529	-1.411
type_of_meal_plan_Meal Plan 2	0.1642	0.067	2.469	0.014	0.034	0.295
type_of_meal_plan_Not Selected	0.2860	0.053	5.406	0.000	0.182	0.390
room_type_reserved_Room_Type 2	-0.3552	0.131	-2.709	0.007	-0.612	-0.098
room_type_reserved_Room_Type 4	-0.2828	0.053	-5.330	0.000	-0.387	-0.179
room_type_reserved_Room_Type 5	-0.7364	0.208	-3.535	0.000	-1.145	-0.328
room_type_reserved_Room_Type 6	-0.9682	0.151	-6.403	0.000	-1.265	-0.672
room_type_reserved_Room_Type 7	-1.4343	0.293	-4.892	0.000	-2.009	-0.860
market_segment_type_Corporate	-0.7913	0.103	-7.692	0.000	-0.993	-0.590
market_segment_type_Offline	-1.7854	0.052	-34.363	0.000	-1.887	-1.684

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Performance Evaluation and Improvement- Logistic Regression

- Coefficient Interpretation
- Coefficients of average-price per room, no of previous cancellations and type of meal plan selected are positive
- Increase in these figures might lead to a person cancelling the book
- Coefficients of repeat guests, special requests are negative thus increase in these will lead to a decrease in chance of person cancelling reservation

- Model Performance on Training Set



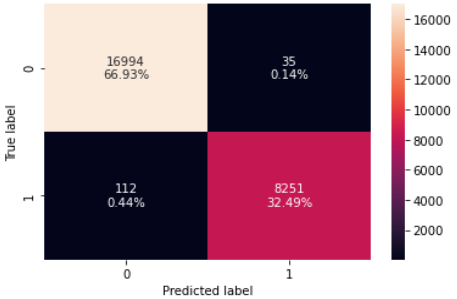
Training performance:

	Accuracy	Recall	Precision	F1
0	0.80545	0.63267	0.73907	0.68174

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Performance Evaluation and Improvement – Decision Tree

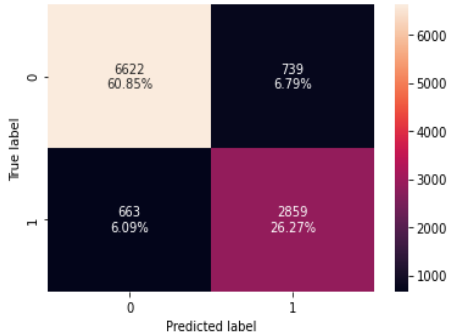
- Checking model performance on training set



	Accuracy	Recall	Precision	F1
0	0.99421	0.98661	0.99578	0.99117

- There is a significant increase in the accuracy, recall, precision and F1 figures

- Checking model performance on test set



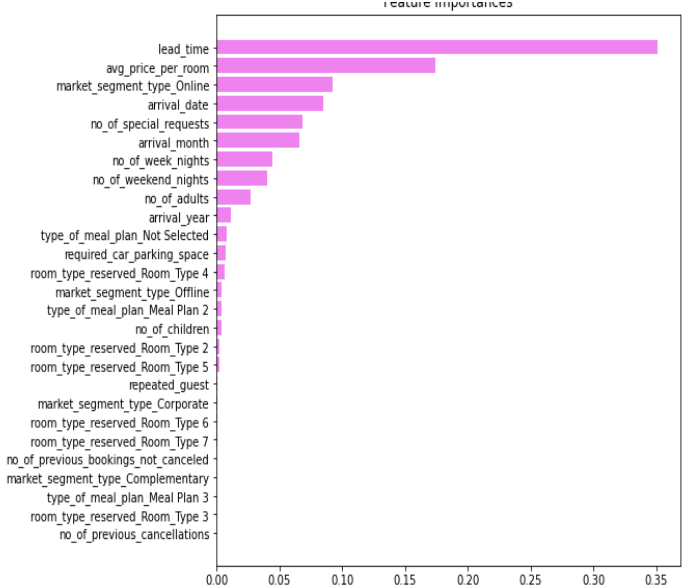
	Accuracy	Recall	Precision	F1
0	0.87118	0.81175	0.79461	0.80309

- There is a significant decrease in the accuracy, recall, precision and F1 figures

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Building- Decision Tree

- Important features before Pruning Tree



- The four most important features are
  - Lead time
  - Average price per room
  - Market segment type Online
  - Arrival date

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Actionable Insights and Recommendations

- Key Takeaways
  - Lead time, Average price per room, Market segment type Online and Arrival date have a significant impact on booking cancellation or non cancellation
  - Dynamic pricing model used to determine average price per room has to be reviewed to take into account the busiest months
  - Repeat customers form an important segment and should be harnessed in terms of average pricing

- Business Recommendation
  - A portion of the average price per room should be non-refundable to minimize fluctuations in revenue earned
  - Review of the average price for corporate segment should be done given the consistency of their bookings
  - Bookings during the busiest month should include a surcharge which is non-refundable
  - The busiest months of 8 to 11 should have a fixed room price which would smooth over fluctuation in the less busy months
  - Return customers should have a loyalty program to ensure rewards for repeat visits

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Potential Benefits of Implementing Business Recommendation

- Repeat customers: Reward programs would ensure continuous business and serve as a form of advertisement which would impact both the profit statement and image of the company
- Corporate customers: Given their low cancellation policy, a slight increase in their fees would offset cancellations by offline customers
- Surcharge on cancellation: Revenue would be based not on total charges but would also include the penalty for cancellations
- Fixed room charge: Fixed room charge during the busiest months would help incorporate the probability of dropped bookings in non busy months which would impact revenue

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

greatlearning
Power Ahead

Happy Learning !