StarHotels Project

Our Solution Approach

Dear CFO, StarHotels, we feel highly honored for the opportunity to present our proposal which we are confident will provide actionable insights to the problem StarHotels faces at the moment.The format of the presentation will be as follows:

Solution strategy

- Stating the problem
- Objective of the project
- Solution approach
  - Visualizing the data:
  - Presenting EDA Highlights
  - Presenting Data preprocessing approach
  - Presenting Logistic regression and Decision tree analysis results
  - Comparison of model results
  - Conclusion
  - Business Insights

Background

Context

- StarHotels faces the problem of rising booking reservation cancellations
- Booking cancellations impact adversely on company revenue
- A redress to the problem is needed

Problem to Solve

- How can we help StarHotel resolve its booking cancellation problem using insights from ML on company data containing history of booking status and other demographics of individuals and entities that made reservations for the hotel

Objective of the project

- Build a Machine Learning based solution to help StarHotels predict booking reservation cancellations

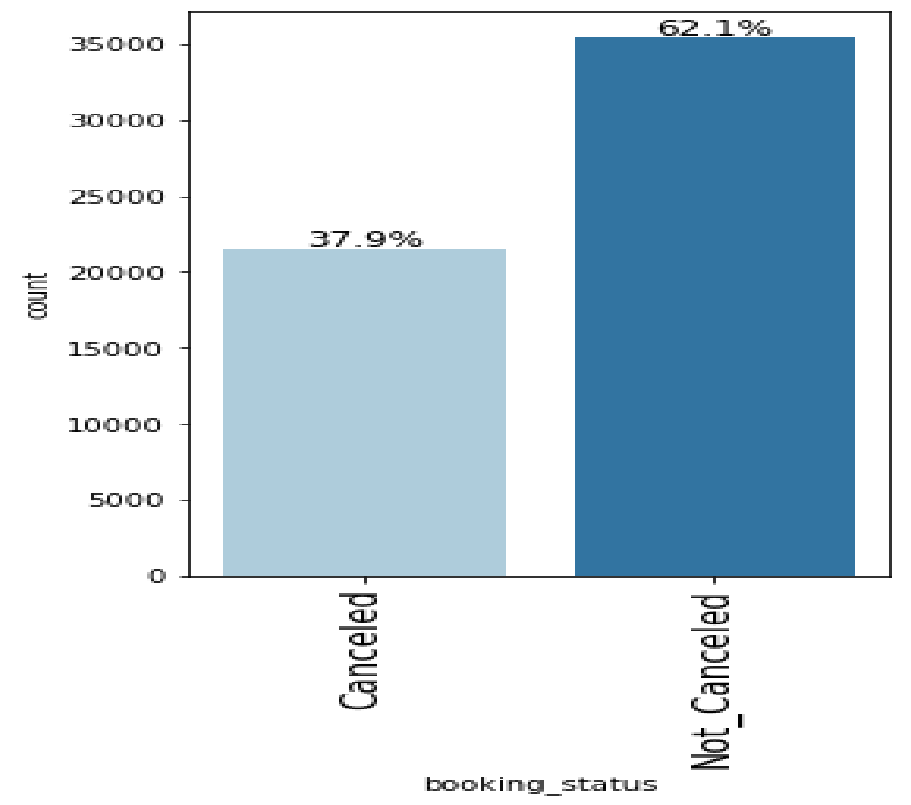
Key questions the solution should answer:

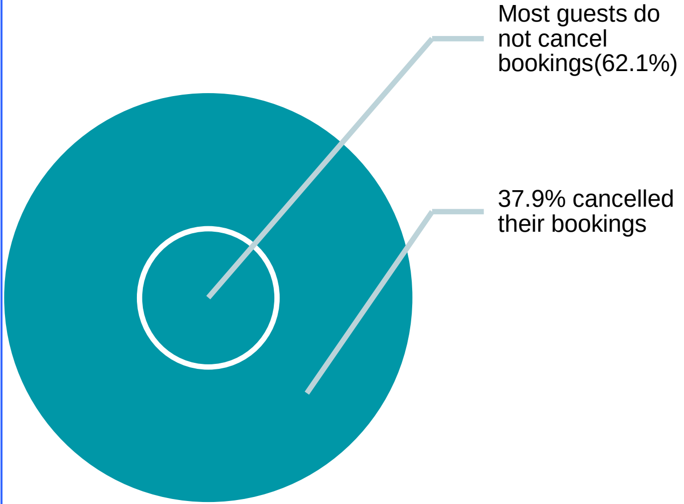
- What factors influence booking cancellation behavior for prospective guests to StarHotels?
- What insights can Starhotel gain from information about determinants of past records on booking cancellation behavior?

Data Overview : StarHotels has 56926 Information on booking reservation status and other 17 details

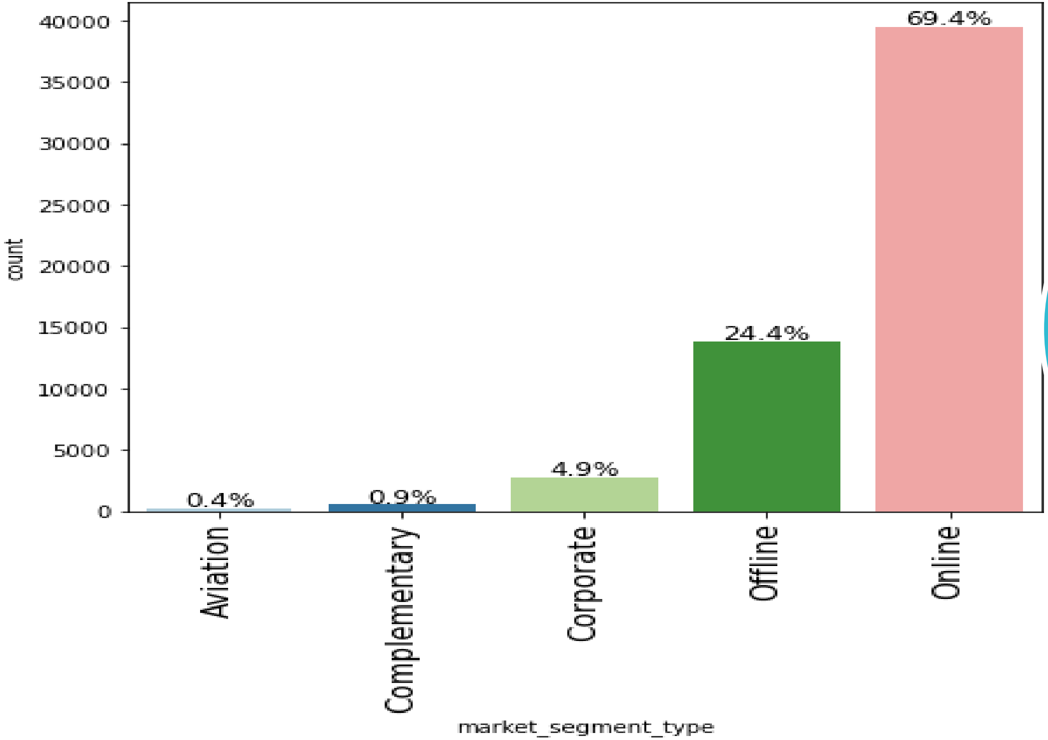
1. no_of_adults: Number of adults
2. no_of_children: Number of Children
3. no_of_weekend_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
4. no_of_week_nights: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
5. type_of_meal_plan: Type of meal plan booked by the customer:
    - Not Selected – No meal plan selected
    - Meal Plan 1 – Breakfast
    - Meal Plan 2 – Half board (breakfast and one other meal)
    - Meal Plan 3 – Full board (breakfast, lunch, and dinner)
6. required_car_parking_space: Does the customer require a car parking space? (0 - No, 1- Yes)
7. room_type_reserved: Type of room reserved by the customer. The values are ciphered (encoded) by Star Hotels.
8. lead_time: Number of days between the date of booking and the arrival date
9. arrival_year: Year of arrival date
10. no_of_previous_cancellations: Number of previous bookings that were canceled by the customer prior to the current booking
11. no_of_previous_bookings_not_canceled: Number of previous bookings not canceled by the customer prior to the current booking
12. avg_price_per_room: Average price per day of the reservation; prices of the rooms are dynamic. (in euros)
13. no_of_special_requests: Total number of special requests made by the customer (e.g. high floor, view from the room, etc)
14. booking_status: Flag indicating if the booking was canceled or not.
15. market_segment_type: Market segment designation.
16. repeated_guest: Is the customer a repeated guest? (0 - No, 1- Yes)
17. arrival_month: Month of arrival date
18. arrival_date: Date of the month

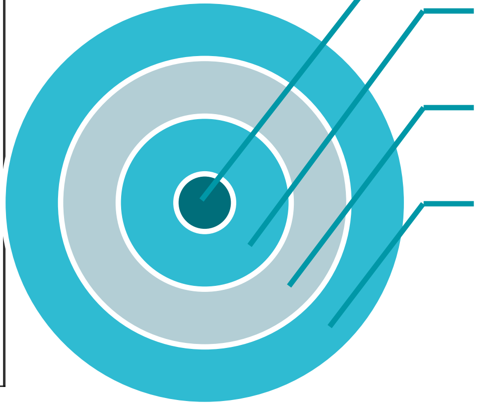
Booking Status





Market Segment





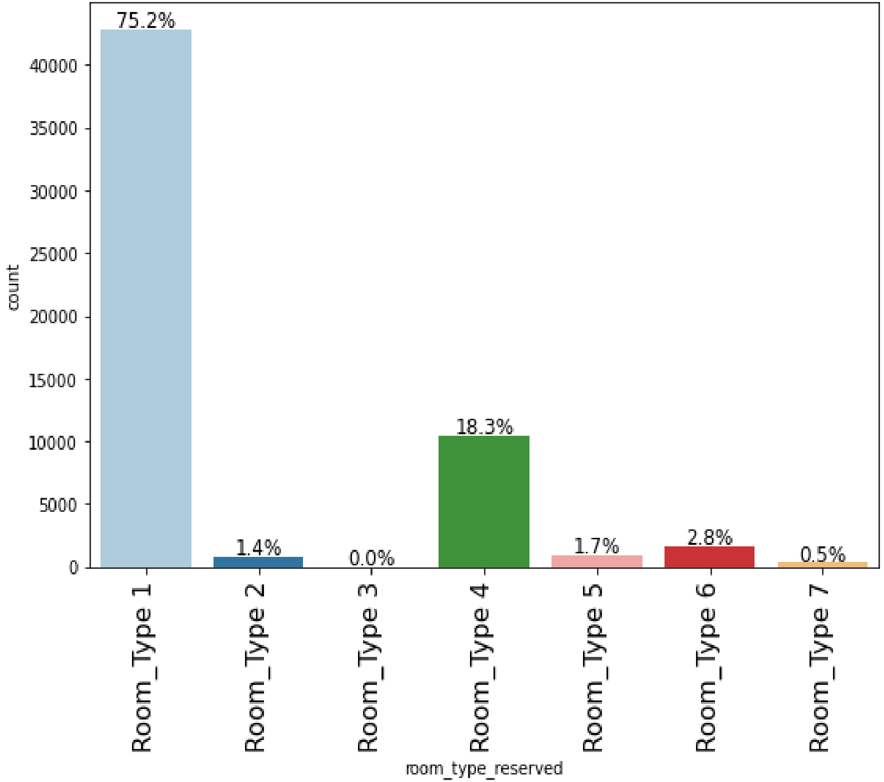
69.4% of guests at StarHotels are from the market segment is online;

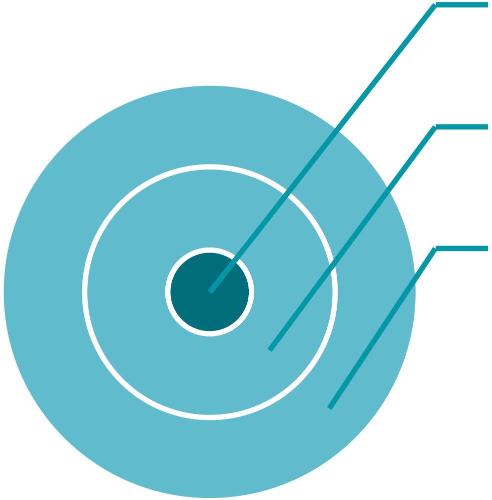
24.4% offline;

4.9% Corporate;

0.9% Complementary

Room_Type Reserved



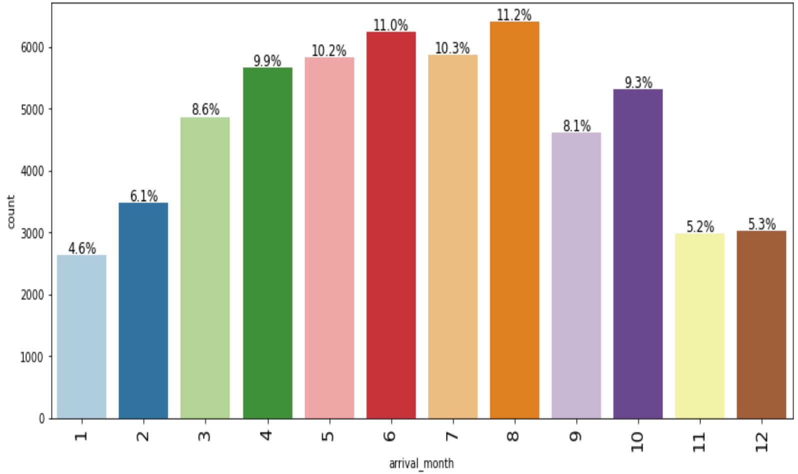


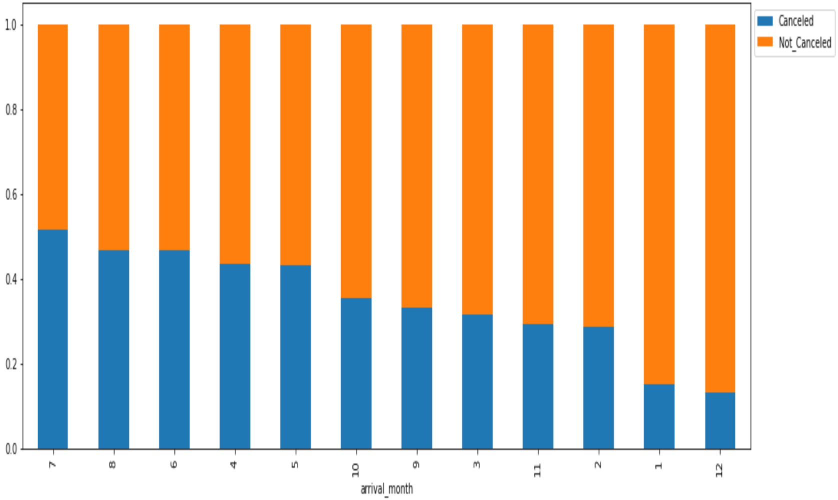
Most (75.2%) guests request for Room_Type1

18.3% request for Room_Type 4

Room_Type 3 and 7 are the least reserved

Arrival_month and booking_status



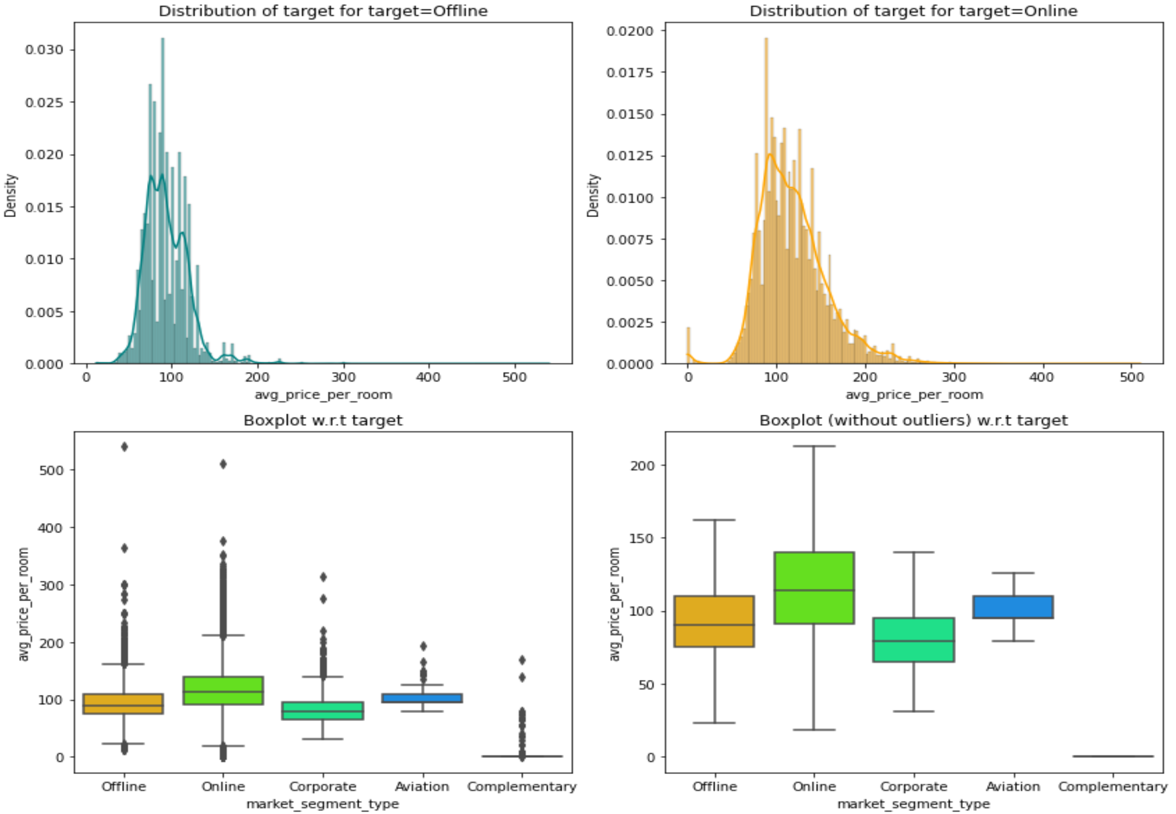


11.2% of arrivals occured in august,followed by June (11.0%) January had the lowest arrivals (4.6%)

arrival_month by booking_status,shows that July and August were the two months with highest cancellations

January and December are the two months with the lowest cancellations

Arg_price_per_room and market_segment_type



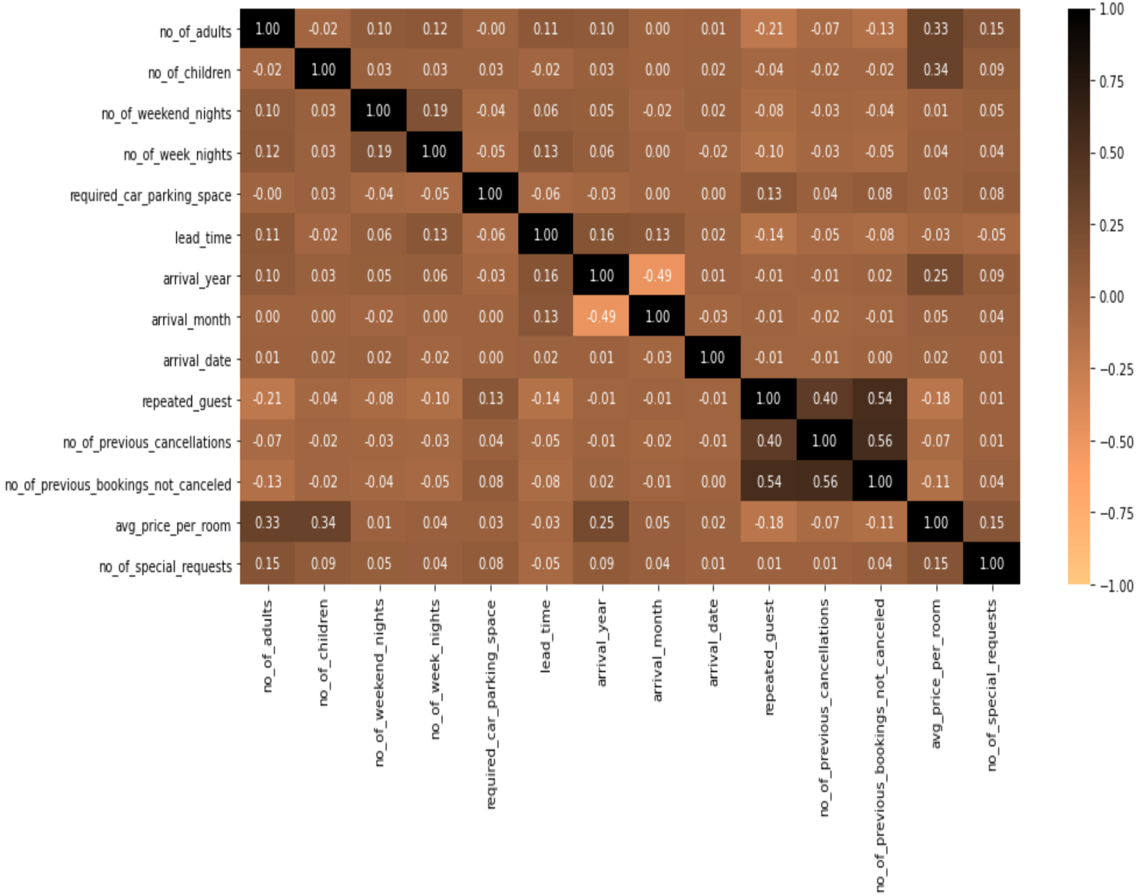
online reservations tend to be higher than offline reservations

average room price for online segment in ~ 120,while that for offline reservation is ~ 80

with the exception of complementary market segment, all reservations made online pay higher room prices than those who offline reservations

complementary reservations pay the lowest average room price

Correlation results



pairwise correlations do not show strong correlations between variables in the dataset

correlations are relatively moderate with the highest being 0.56

There are both positive & negative correlations

Preprocessing approach

-Involved missing value identification and treatment

- Outlier detection and treatment involving some but not all numerical predictor variables

- Converting object dtype variables into category variables

- Preparing data for modeling

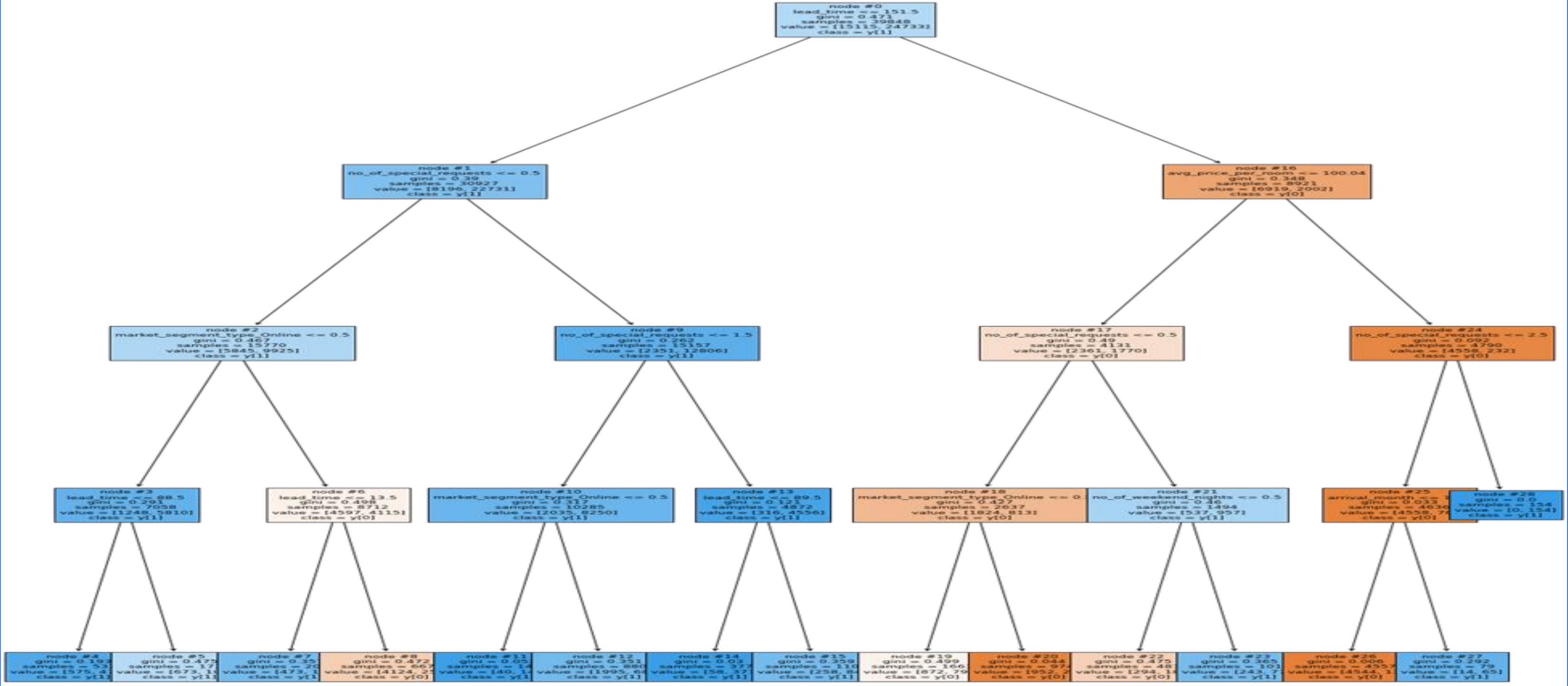
- Removing multicollinearity for sm.OLS models

- created dummies for category variables

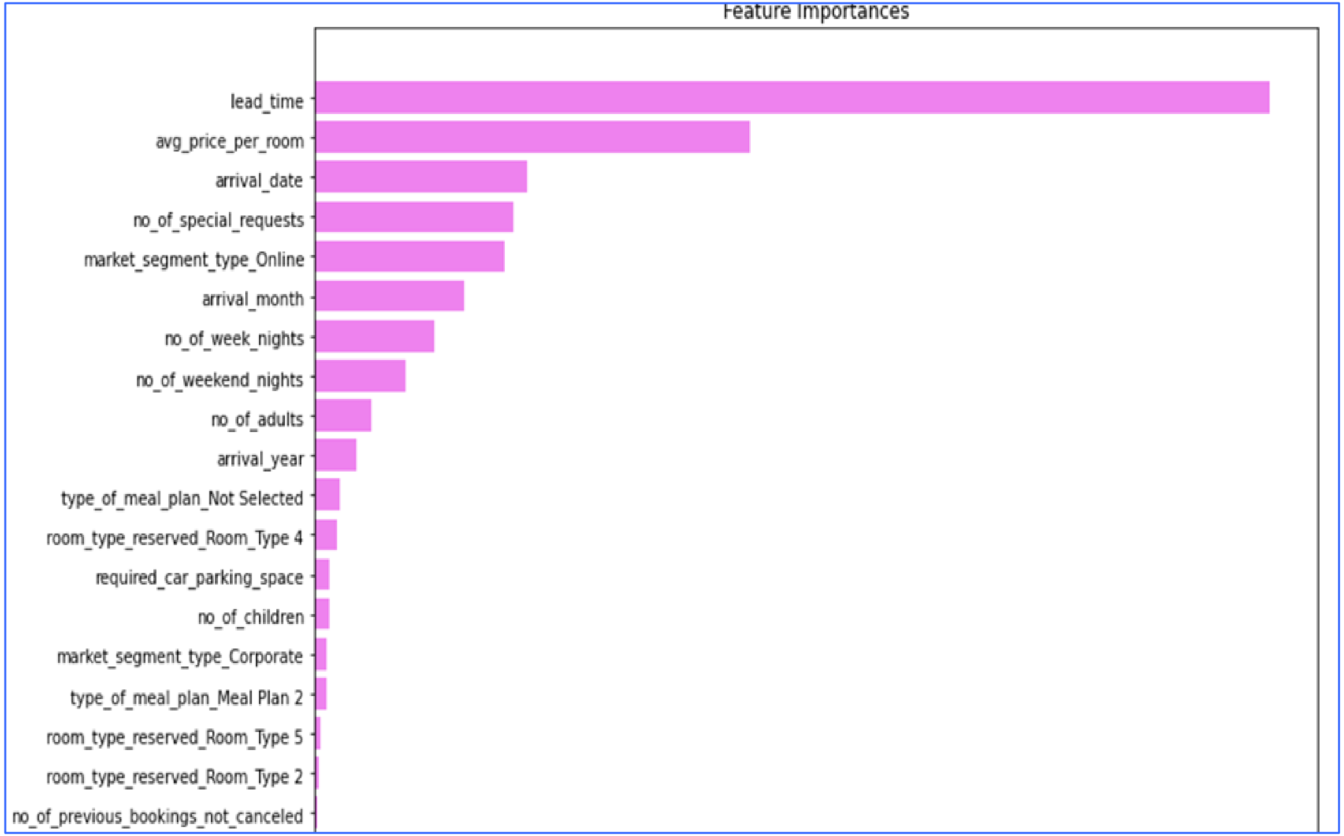
Presentation of results logistic regression results

	coef	std err	z	P>|z|	[0.025	0.975]
const	4.4403	0.074	59.787	0.000	4.295	4.586
no_of_weekend_nights	-0.0951	0.015	-6.234	0.000	-0.125	-0.065
no_of_week_nights	-0.0273	0.010	-2.633	0.008	-0.048	-0.007
required_car_parking_space	1.5013	0.113	13.278	0.000	1.280	1.723
lead_time	-0.0167	0.000	-84.252	0.000	-0.017	-0.016
arrival_month	0.0636	0.005	13.259	0.000	0.054	0.073
repeated_guest	3.1301	0.493	6.350	0.000	2.164	4.096
no_of_previous_cancellations	-0.2918	0.071	-4.117	0.000	-0.431	-0.153
avg_price_per_room	-0.0203	0.001	-38.514	0.000	-0.021	-0.019
no_of_special_requests	1.3336	0.021	62.252	0.000	1.292	1.376
type_of_meal_plan_Meal Plan 2	0.1497	0.053	2.819	0.005	0.046	0.254
type_of_meal_plan_Not Selected	-0.3295	0.038	-8.701	0.000	-0.404	-0.255
room_type_reserved_Room_Type 2	0.3595	0.114	3.164	0.002	0.137	0.582
room_type_reserved_Room_Type 4	0.2794	0.039	7.081	0.000	0.202	0.357
room_type_reserved_Room_Type 5	0.4525	0.101	4.461	0.000	0.254	0.651
room_type_reserved_Room_Type 6	0.4269	0.088	4.873	0.000	0.255	0.599
room_type_reserved_Room_Type 7	0.6102	0.179	3.400	0.001	0.258	0.962
market_segment_type_Corporate	-1.2655	0.081	-15.674	0.000	-1.424	-1.107
market_segment_type_Online	-1.5881	0.041	-38.635	0.000	-1.669	-1.508

Presentation of results Decision tree   results



Presentation of Decision tree results: Classification features



Key classification features:

- Lead_time
- Avg_price_per_room
- Arrival_date
- No_of_special_requests
- Mark_segment_type_online
- Arrival_month
- No_of_week_nights
- No_of_weekend_nights
- No_of_adults
- _arrival_year

Presentation of Logistic Regression (Odds results) Classification features

	Odds	Change_odd%
const	84.8	8380.0
no_of_weekend_nights	0.9	-9.1
no_of_week_nights	1.0	-2.7
required_car_parking_space	4.5	348.7
lead_time	1.0	-1.7
arrival_month	1.1	6.6
repeated_guest	22.9	2187.5
no_of_previous_cancellations	0.7	-25.3
avg_price_per_room	1.0	-2.0
no_of_special_requests	3.8	279.5
type_of_meal_plan_Meal Plan 2	1.2	16.2
type_of_meal_plan_Not Selected	0.7	-28.1
room_type_reserved_Room_Type 2	1.4	43.3
room_type_reserved_Room_Type 4	1.3	32.2
room_type_reserved_Room_Type 5	1.6	57.2
room_type_reserved_Room_Type 6	1.5	53.2
room_type_reserved_Room_Type 7	1.8	84.1
market_segment_type_Corporate	0.3	-71.8
market_segment_type_Online	0.2	-79.6

Key classification features logistic regression:

- no_of_weekend_nights,
- no_of_week_nights
- required_car_parking_space
- lead_time
- arrival_month
- repeated_guest
- no_of_previous_cancellations
- avg_price_per_room
- no_of_special_requests
- type_of_meal_plan_Meal Plan 2
- type_of_meal_plan_Not Selected
- room_type_reserved_Room_Type 2
- room_type_reserved_Room_Type 4
- room_type_reserved_Room_Type 5
- room_type_reserved_Room_Type 6
- room_type_reserved_Room_Type 7
- market_segment_type_Corporate
- market_segment_type_Online

Comparison Logistic Regression and Decision Tree model Performance results

Logistic regression models results

	Logistic Regression statsmod_Train	Logistic Regression-0.67 Threshold	Logistic Regression-0.59 Threshold	logistic Regression statsmod_Test
Accuracy	0.784054	0.766488	0.780692	0.790608
Recall	0.860551	0.738730	0.806817	0.865665
Precision	0.804992	0.865350	0.834379	0.811108
F1	0.831845	0.797042	0.820366	0.837499

Decision tree models performance results

	Decision Tree sklearn	Decision Tree sklearn_8	Decision Tree sklearn_4	Decision Tree (Pre-Pruning)	Decision Tree (Post-Pruning)
Accuracy	0.994454	0.842677	0.801872	0.620684	0.865363
Recall	0.994865	0.901144	0.855820	1.000000	0.936199
Precision	0.996194	0.853554	0.830209	0.620684	0.859439
F1	0.995529	0.876704	0.842820	0.765953	0.896178

Comparison model Performance results

Decision tree uses fewer key features to classify StarHotel data include lead_time, number of special requests made, reserving via online channels, average room price, arrival month, and number of weekend nights reserved.

Since we were not able to adopt the post pruned tree because of its complexity, there is high likelihood that the decision tree we used which has four levels is prone to underfitting.

The implication is that some crucial variables that would have been included in the decision making to classify the dataset were not used.

Thus, logistic regression uses more features to predict booking status, making a more reliable ML approach in this case

Conclusion

Factors increasing booking cancellations

- Booking many requests for car parking space,
- Arrival month especially July and August,
- Making many special requests,
- repeated guests,
- guests who in their booking details request for meal plan 2,
- bookings that make reservation for room type 2, room type 4, room type 5 and room type 6, and room type 7

Factors decreasing booking cancellations

- nights booked on weekends
- nights books on week ,
- previous booking cancellations,
- * lead time of booking prior to arrival especially bookings above 150 days
- Lead_time,
- average price per room,w ith higher room prices associated with low booking cancellations,
- bookings that are made by corporate entities,
- booking via online channels,
- bookings that do not select any meal plans

Business Insights

greatlearning
Power Ahead

- Prospective guests who book many nights on weekends and week nights, have prior records of previous booking cancellations, making their bookings far in advance prior to their arrival date, reserve rooms in a high price range, booking reservations made by corporate entities, booking reservations made via online channels, do not select meal plans are less likely to cancel their bookings.
- As June and August are the most busy months for hotel reservations, while July and August are the months with the highest cancellations StarHotel management can devise arrangements that attract guests to hotel premises during those the peak months and provide incentives for guests who make booking reservations in other months. Spreading out reservations to all months of the year should reduce revenue variability.
- StarHotels should introduce incentives to guests with profiles identified as less likely to cancel bookings for instance by providing surprise gifts, discounts on meals, transport from arrival locations to hotel premises, and link number of guests recommended to stay in hotel premises by such segments to additional services at reduced prices among others
- StarHotels management should pay special attention to guests who make many requests for car parking space, booking many months in advance prior to the arrival month, make many special requests, and guests who in their booking details request for meal plan 2, reserve room type 2, type 4, type 5 and type6, and type 7, as they have higher likelihood of cancelling their bookings than guests who do not request and reserve such 'additionals' in their bookings.
- Taking a deep dive into the social-demographic profiles of hotel guests who book reservations for meal plan 2, room types 4,5,6, and 7 should more granular actionable insights to inform Star Hotel policy on bookings
- To reduce the adverse impact of cancellations on StarHotels finances, the company should disincentive cancellation practice by introducing fines to segments that have been identified as having high likelihood of making cancellations. Refunds should be linked to time to when cancellation of the booking is made prior to arrival date with longer time to arrival date allowing larger refund than those who cancel bookings a few days prior to arrival.
- StarHotels should incentivize online reservations as they are both less susceptible to cancellation and generate higher revenue for StarHotels because of higher average prices for rooms reserved

Great thank you for the opportunity

greatlearning
Power Ahead

Happy Learning !