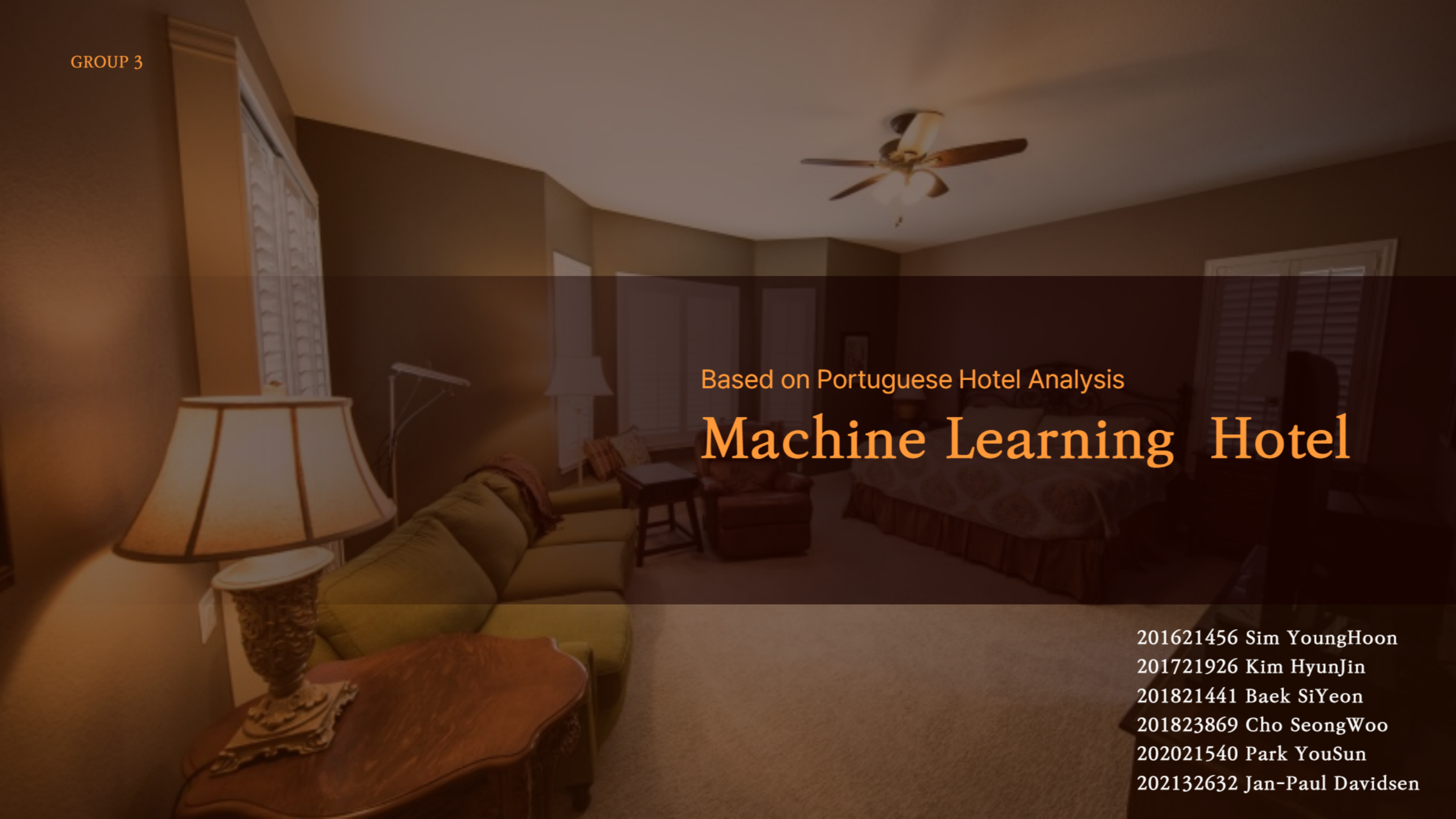
GROUP 3



Based on Portuguese Hotel Analysis

Machine Learning Hotel

201621456 Sim YoungHoon
201721926 Kim HyunJin
201821441 Baek SiYeon
201823869 Cho SeongWoo
202021540 Park YouSun
202132632 Jan-Paul Davidsen

Contents

Introduction

We are the management who runs the hotel business. As the number of travelers is expected to increase due to the easing of COVID-19, we are trying to enter the global market after analyzing hotel data

Goal

Expand business into Europe

Data Selection

Portugal Hotel Data

Target

Travelers traveling in Europe

Business Issues

Cancellations Increase Sales

hotel_bookings

Dataset

We looked for the data with a plan to expand our business to Europe on the premise of City Hotel, taking into account the local situation in Portugal (climate-Mediterranean climate, culture).

This is the actual data on the reservation and cancellation of city and resort hotels from July 2015 to August 2017.

Data for customer identification has been deleted because it is real data, but a total of 119,390 observations and 31 variables are expected to produce significant results. As the number of overseas travelers is expected to increase due to the easing of COVID-19, we will analyze Portugal's hotel industry data to find the key to entering the global market.

Findings 1

- Is_canceled -
Value indicating if the booking was canceled (1) or not (0)
- lead_time -
Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

Findings 2

- ADR -
Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights

Findings 3

- Length of stay -
Number of nights the guest stayed or booked to stay at the hotel

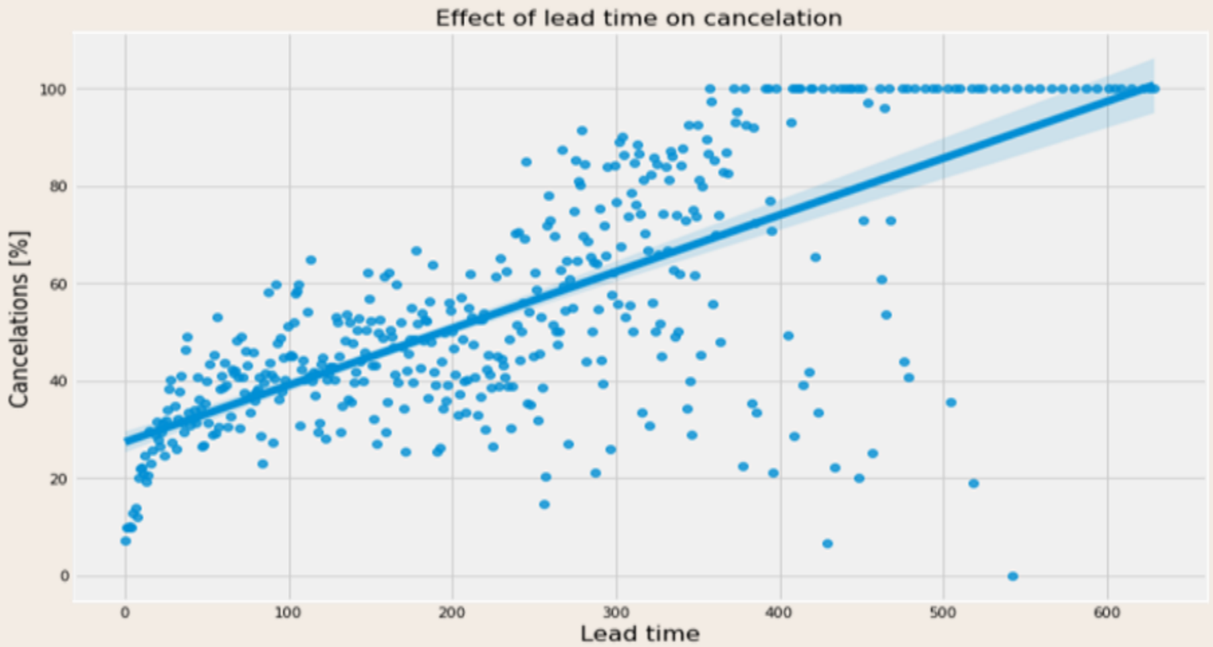
Findings 4

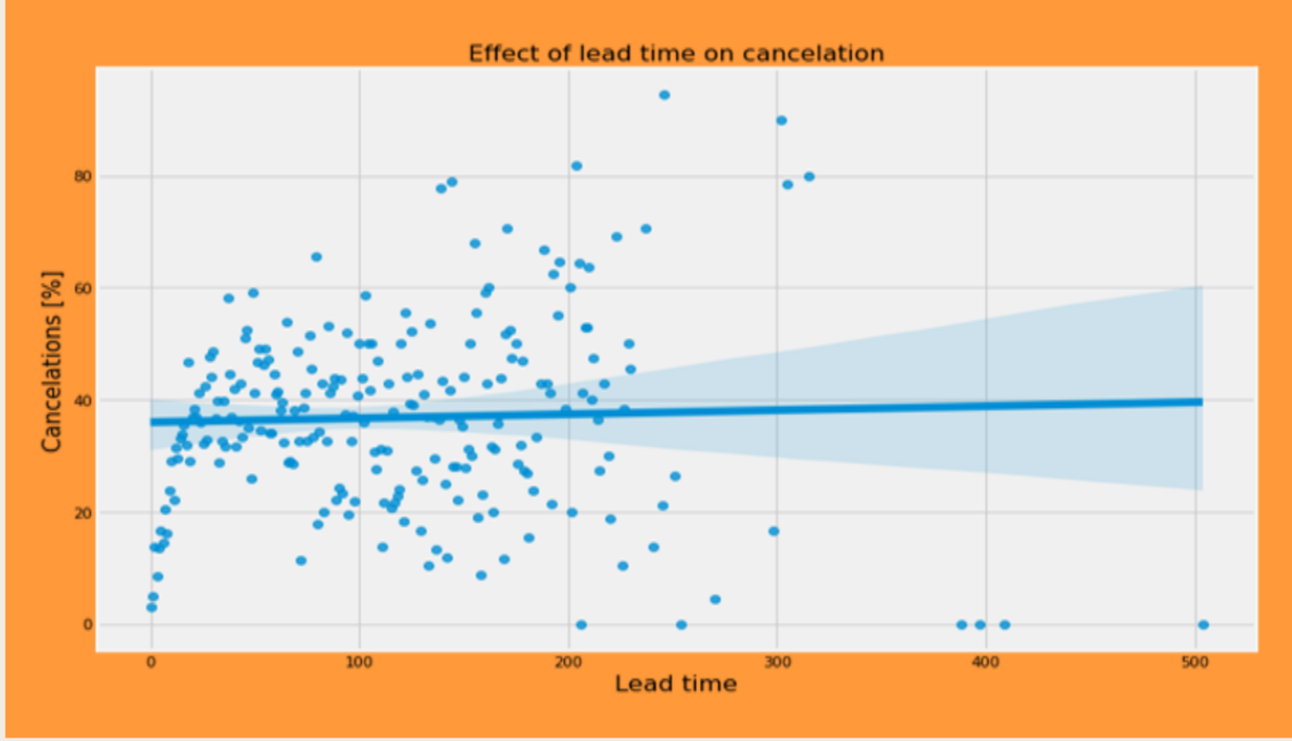
- is_repeated_guest -
Value indicating if the booking n ame was from a repeated guest (1) or not (0)

GROUP 3

Analyses & Findings

Finding 1 is_canceled





Cancellation rate by time of booking

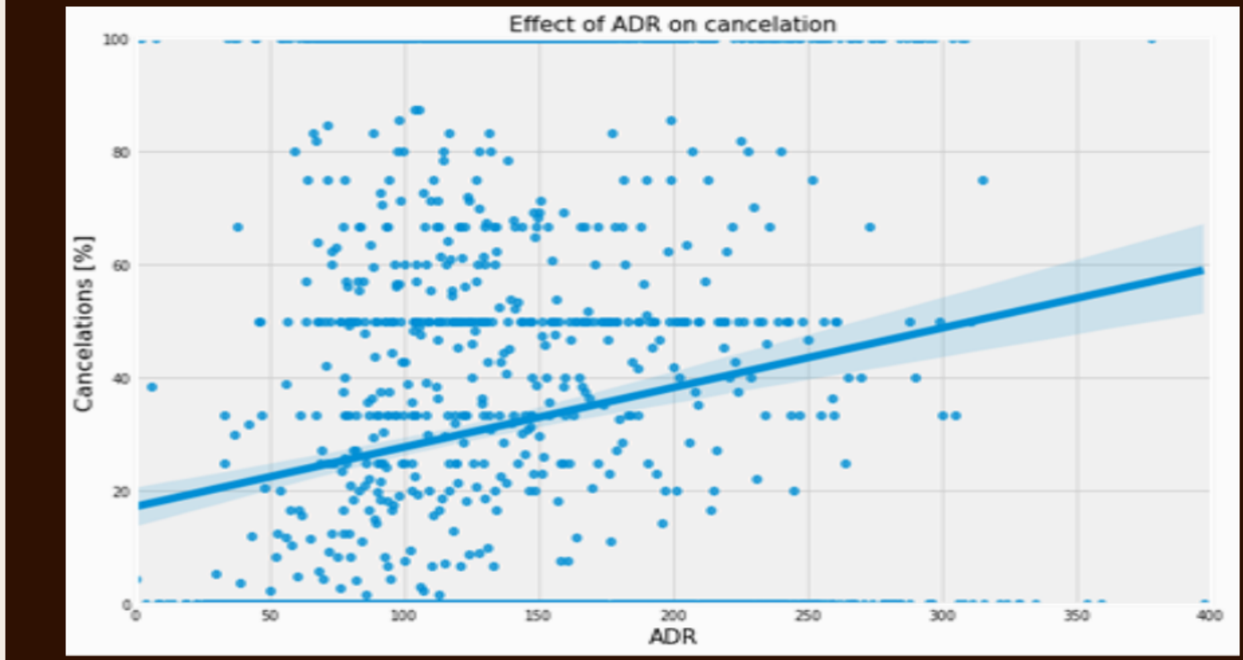
1) In the case of guests from European countries(the graph on the left), there was a big correlation between the lead time and cancelation.
→ The longer you make a reservation, the more likely your plan will change in the middle.

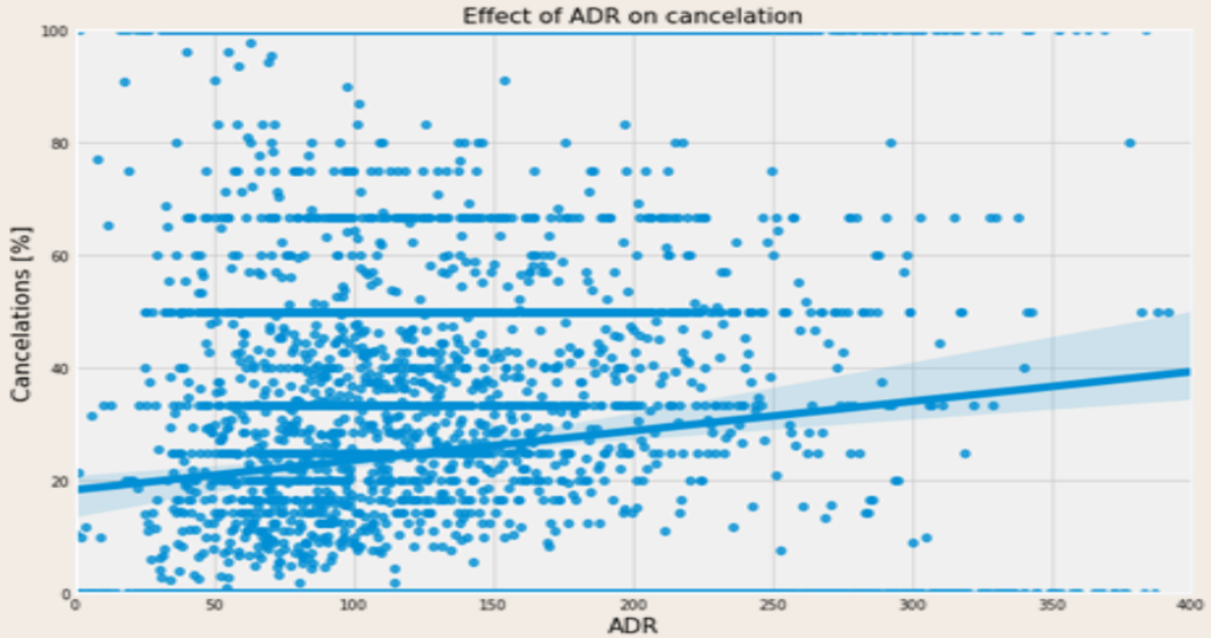
2) However, in the case of non-European countries(the graph on the top), there was not much correlation.
→ The amount of cost and time that is put into travelling Europe is much more than European countries, so they make specific travel plans from a long time ago. Therefore, the cancellation rate according to the time of reservation is relatively low.

GROUP 3

Analyses & Findings

Finding 1 is_canceled





Cancellation rate based on average price

- Contrary to lead time, the price had more effect on non-European countries. (the graph on the top)
- > The overall cost of going to Europe is much higher, so the sensitivity of the price is inevitably high..

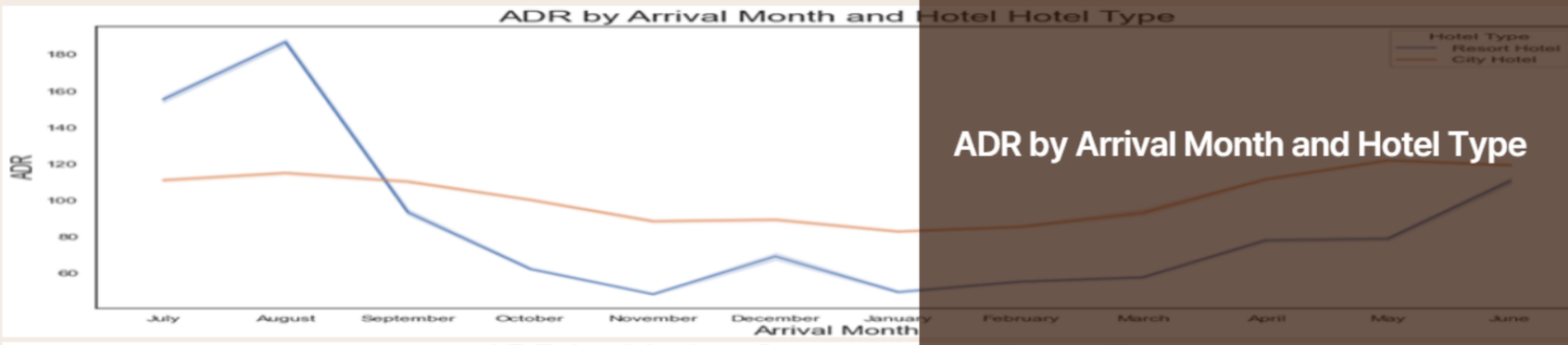
GROUP 3

Analyses & Findings

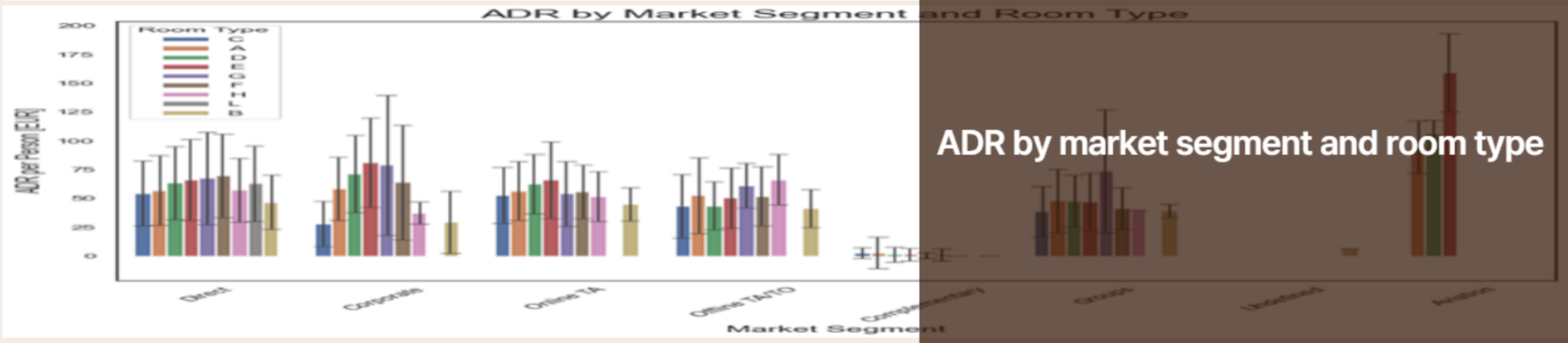
Finding 2 ADR

Based on customer data, the hotel's price promotion is carried out to devise a method to increase the reservation rate and sales. Think of ways to set different prices during the peak and non-peak seasons to increase sales the most, and to collaborate with hotels and travel reservation agencies to progress price promotion to increase customers' reservation rates.

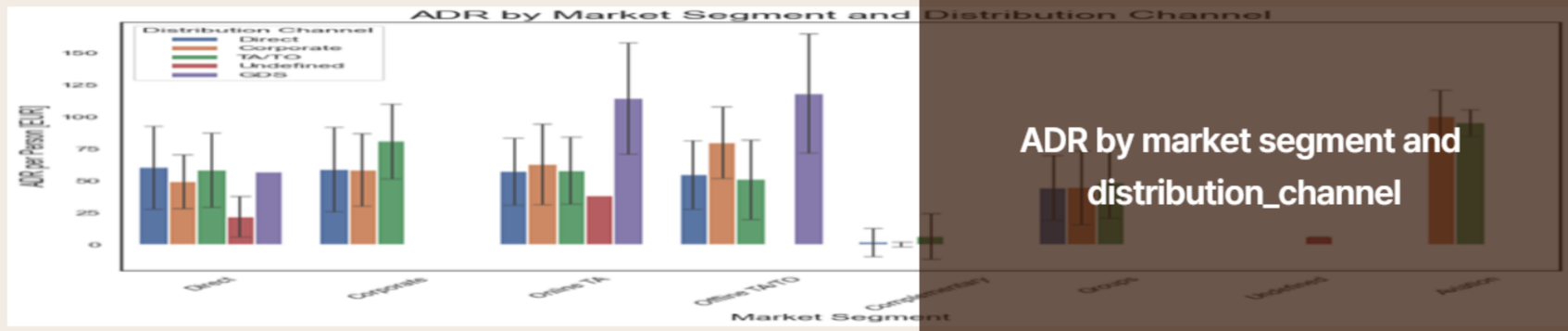
Guest Per Price by Room Type



ADR by Arrival Month and Hotel Type



ADR by market segment and room type

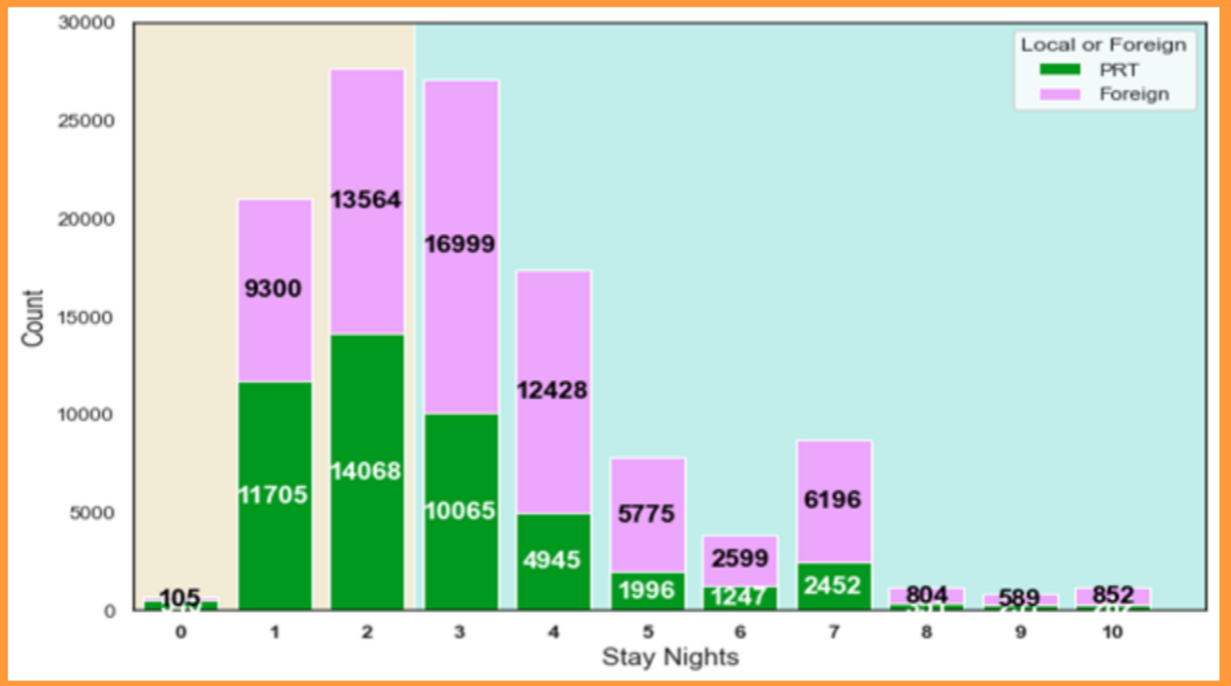


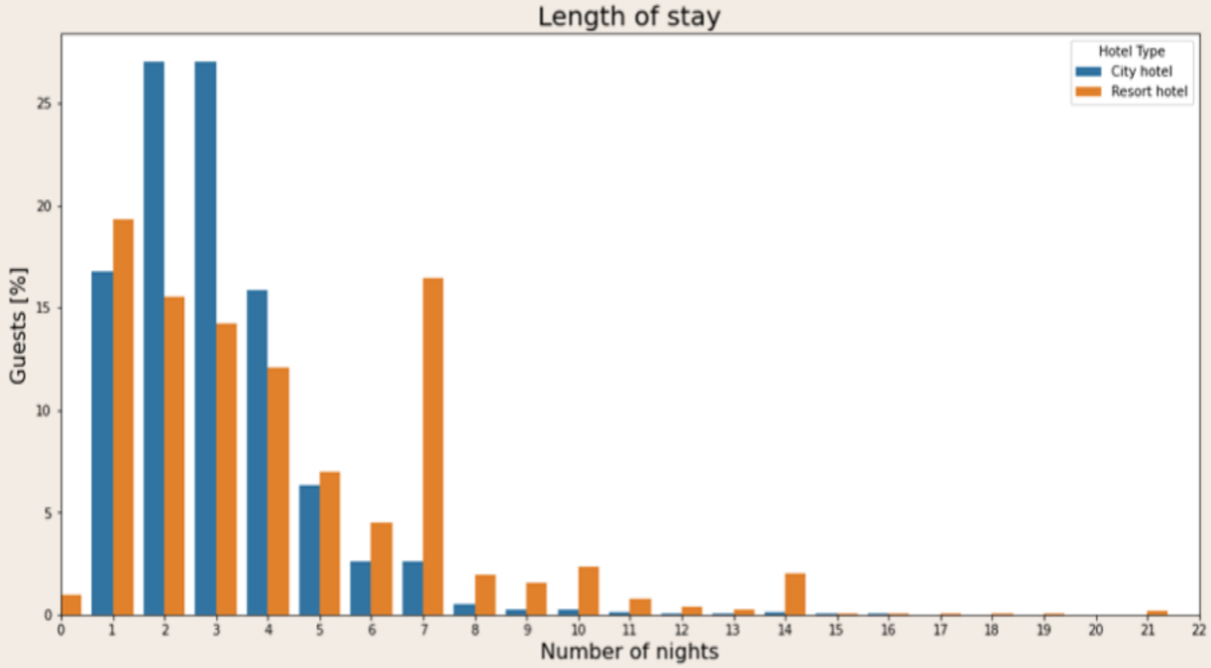
ADR by market segment and distribution_channel

GROUP 3

Analyses & Findings

Finding 3 Length of stay





- Most observations are distributed between the 1st and 4th days of accomm odation in City Hotels, and resort hotels are similar to City Hotels overall.
-> However, the number of observations stayed for seven days was significan tly higher in the Resort Hotel.

- The other graph shows the number of Portuguese customers and foreigners separately
-> There were more long-term accommodations among foreigners than those in Portugal.

GROUP 3

Analyses & Findings

Finding 4 ML Analysis for Business Questions

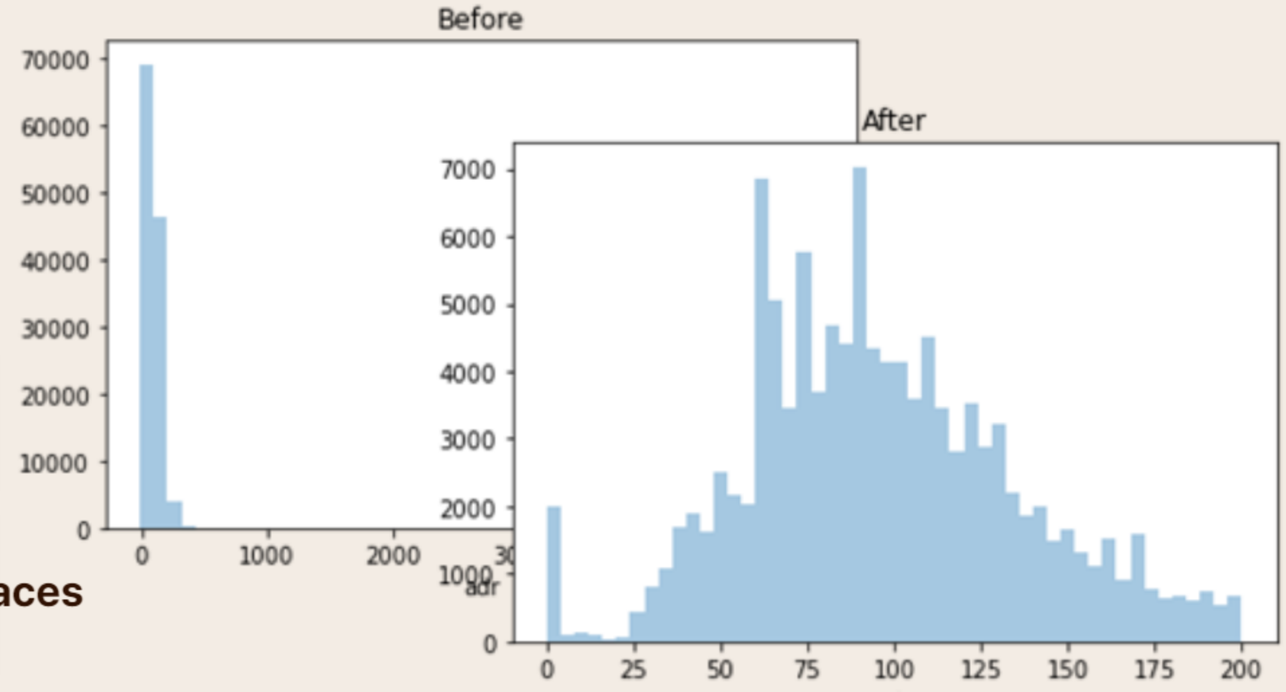
Cancellation(Classification)

- Preprocessing

1)Delete rows with all 0 'adults' & 'children' & 'babies'
2)Only use rows with 3 or less 'adults'
3)'Convert 'Booking_changes', 'special_requests', 'required_car_parking_spaces' into binary variables because the number of reservations and cancellation rates are low er than 1
4)0 <= ADR <= 200
5)Convert Reservation date to datetime variable
6)Convert 'country' to PRT/Foreign
7)Normalize numerical variables
8)delete useless columns

	adults	children	babies
2224	0	0.0	0
2409	0	0.0	0
3181	0	0.0	0
3684	0	0.0	0
3708	0	0.0	0
...	...	...	...
115029	0	0.0	0
115091	0	0.0	0
116251	0	0.0	0
116534	0	0.0	0
117087	0	0.0	0

180 rows × 3 columns



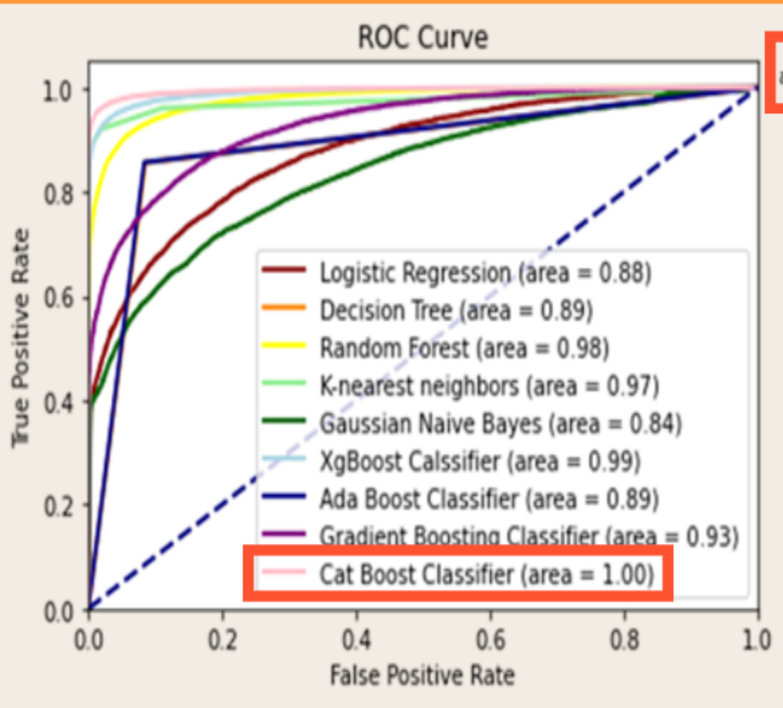
GROUP 3

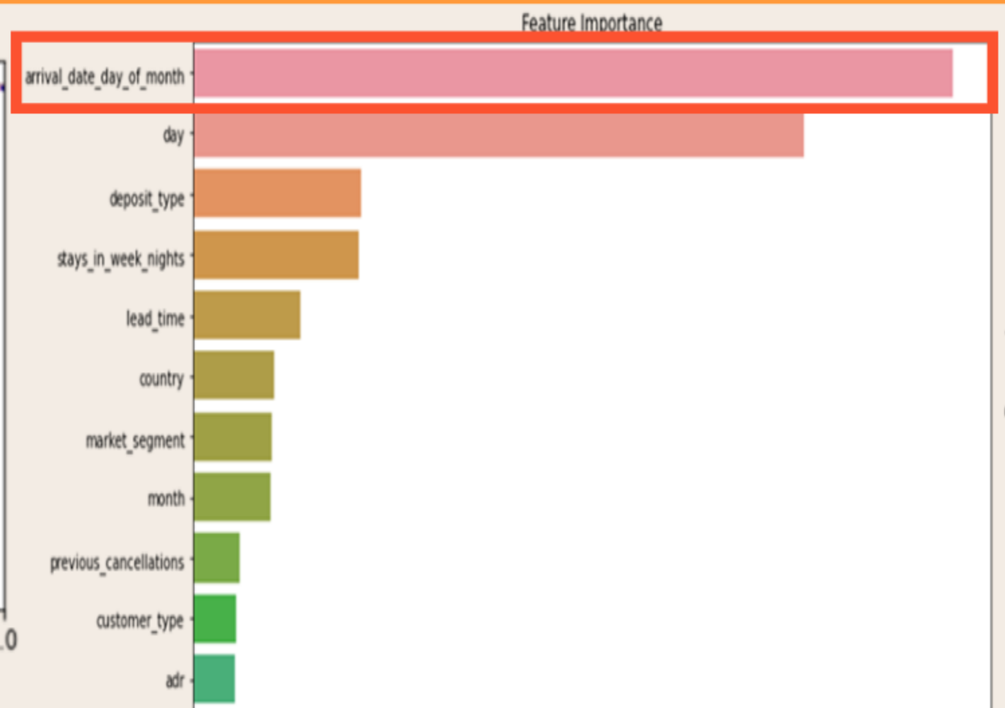
Analyses & Findings

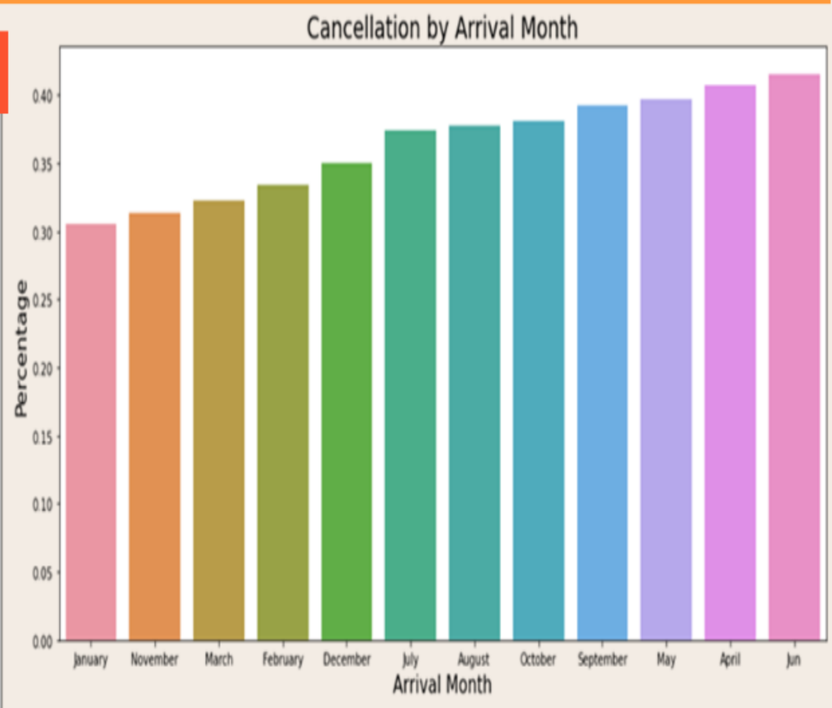
Finding 4 ML Analysis for Business Questions

Cancellation(Classification)

Classification Report	precision	recall	f1-score	support
0	0.97	0.99	0.98	21476
1	0.99	0.94	0.96	12722
accuracy			0.97	34198
macro avg	0.98	0.97	0.97	34198
weighted avg	0.97	0.97	0.97	34198







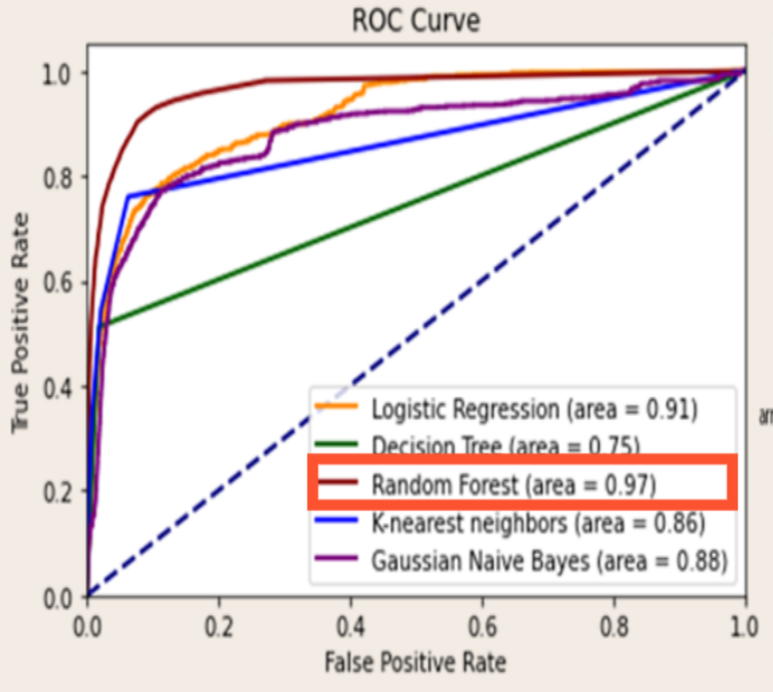
GROUP 3

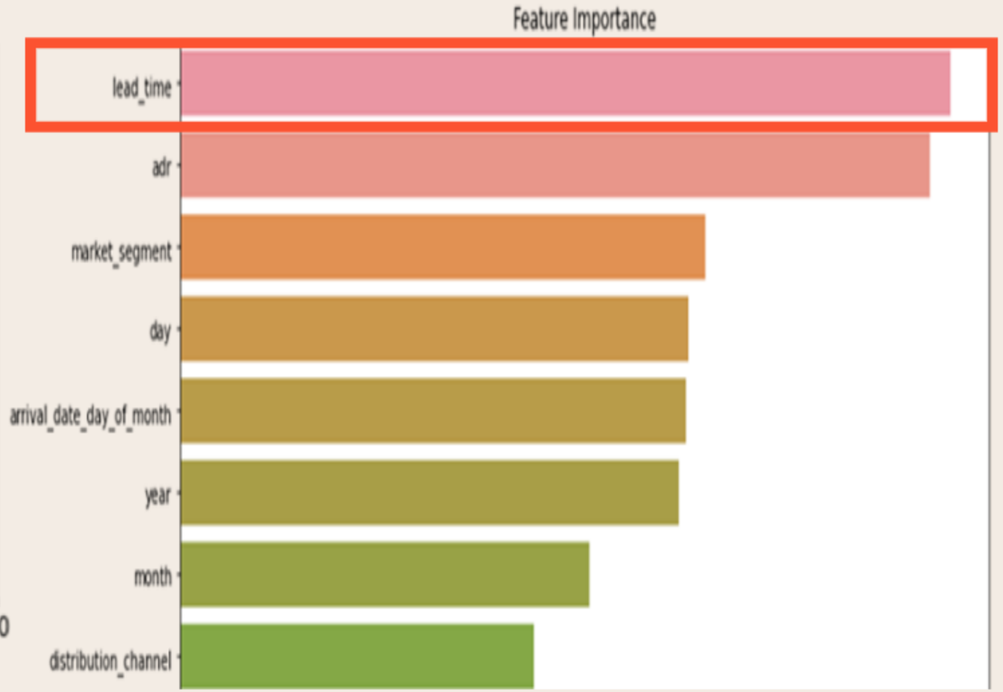
Analyses & Findings

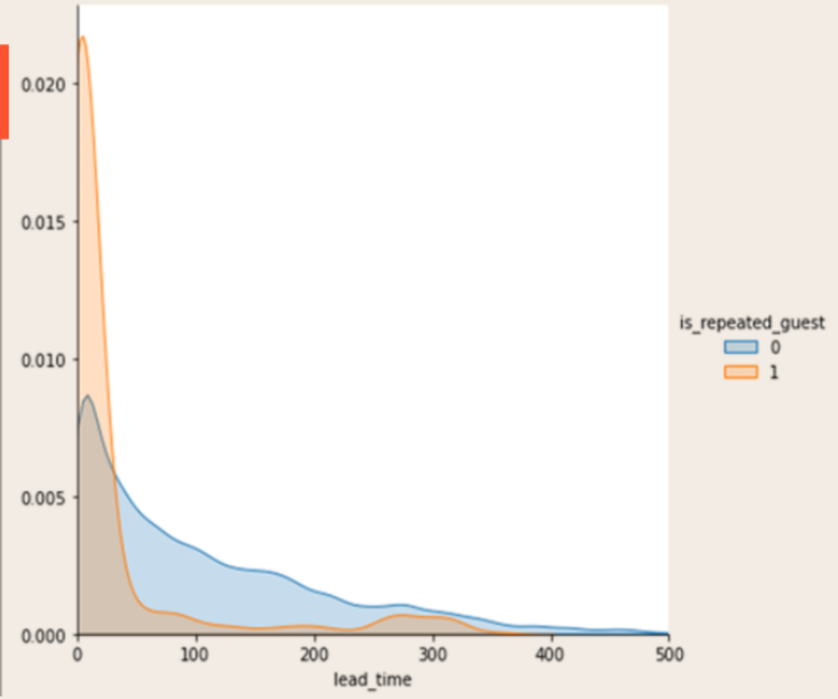
Finding 4 ML Analysis for Business Questions

Revisit(Classification)

Classification Report	precision	recall	f1-score	support
0	0.98	1.00	0.99	34446
1	0.78	0.43	0.55	1133
accuracy			0.98	35579
macro avg	0.88	0.71	0.77	35579
weighted avg	0.97	0.98	0.97	35579







GROUP 3

Analyses & Findings

Finding 4 ML Analysis for Business Questions

ADR_1(Multiple Linear Regression)

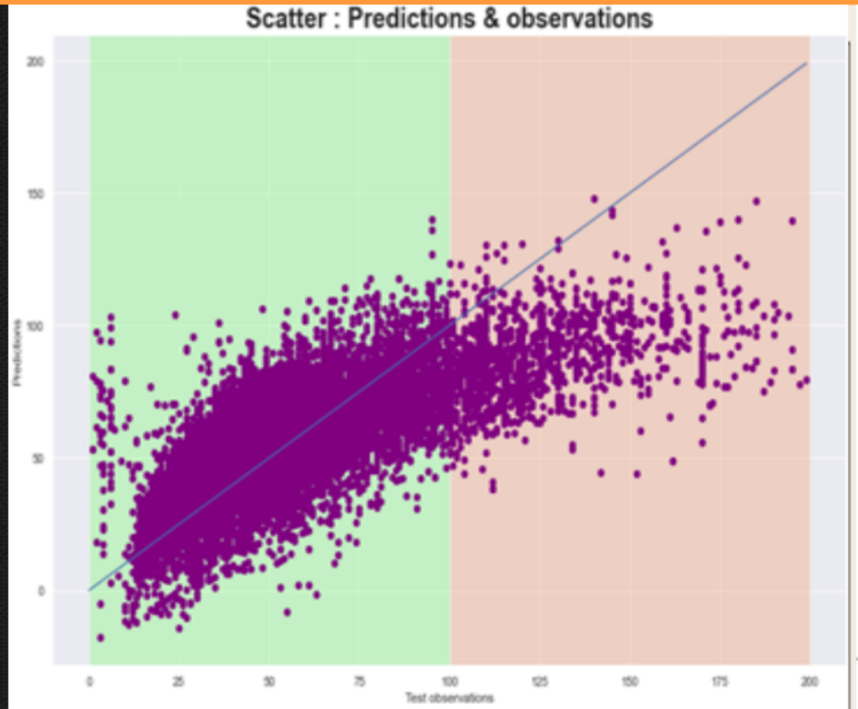
ADR_1: Average daily usage amount per person

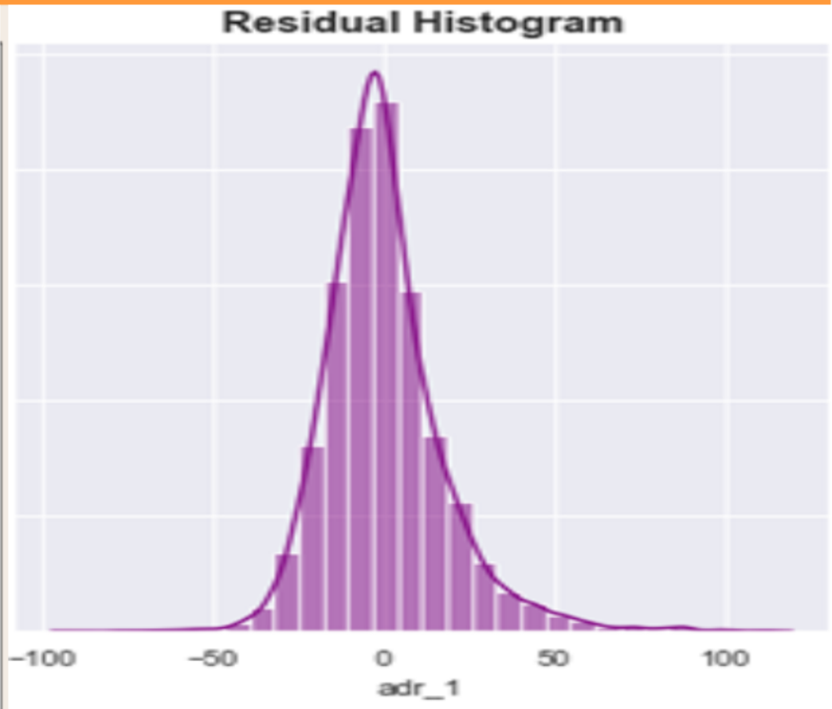
- Preprocessing

1)Is identical to 4.1.
2)Categorize number of guests
3)Create 'adr_1' by dividing adr by total number of guests
4)In the case of the cancelled reservation team, it is considered to be meaningless information that does not help increase sales, so excludes it

Adjusted R-squared	MAE	MSE	RMSE
0.548	12.8	317.4	17.8

	Coefficient
distribution_channel_GDS	35.613625
assigned_room_type_H	31.548673
assigned_room_type_G	26.020550
meal_FB	25.223998
arrival_date_month_August	22.788702
meal_Undefined	22.122882
arrival_date_month_July	18.520021
assigned_room_type_F	16.822670
arrival_date_month_September	16.779157
arrival_date_year_2017	16.722816
	Coefficient
market_segment_Complementary	-49.427673
market_segment_Corporate	-22.117203
required_car_parking_spaces_8	-22.046501
market_segment_Groups	-19.523605
market_segment_Offline TA/TO	-19.123829





GROUP 3

Analyses & Findings

Finding 4 ML Analysis for Business Questions

Additional Association Rules (Apriori)

- Preprocessing

1) Is identical to 4.3.
2) 0 <= ADR_1 <= 400
3) Replace Missing values with appropriate values, such as 0
4) The Undefined variable in 'meal' is the MEAL of SC type and is replaced
5) In the case of a team with one 'is_canceled' (canceled team), it is excluded because it is deemed meaningless to interpret the relevant rules related to actual accommodation
6) Categorize 'addr_1' as low, mid, high

Couple	Transient	Corporate
1) often did not eat meals → need to think of strate gies to increase sales thr ough promotions that incl ude meals	1) Increase the room rotation rate because they stay short 2) have a high ADR →very high potential value	1) single-person customers 2) have lower accommodation co sts than multi-person customers → ADR is small, but ADR_1 is high → indicating that they are high va lue-added customers

GROUP 3

Analyses & Findings

Finding 4 ML Analysis for Business Questions

Additional Association Rules (Apriori)

- Preprocessing

1) Is identical to 4.3.
2) 0 <= ADR_1 <= 400
3) Replace Missing values with appropriate values, such as 0
4) The Undefined variable in 'meal' is the MEAL of SC type and is replaced
5) In the case of a team with one 'is_canceled' (canceled team), it is excluded because it is deemed meaningless to interpret the relevant rules related to actual accommodation
6) Categorize 'addr_1' as low, mid, high

Customers with children

1) ADR was high
2) transient-type customers
→ high value-added customer who paid a high ADR with a fast rotation rate

Customers by country

1) UK → prefer resort hotel
2) major countries from France and Germany → prefer city hotel
3) High ADR - City hotel / Low ADR - Resort hotel
4) Germany → high probability that a special request existed
→ concentrate on the city hotel for France and Germany, it will attract many customers with high added value.
→ Active promotion of resort hotels for UK customers
→ delicate services to French customers with many special requests.

GROUP 3

Implications

Through the above data analysis, we were able to obtain meaningful insights and solutions for business problems and success.

Prevent cancellation

1. special requests
2. longer the lead time
3. accommodation cost

Selling Remaining Rooms

1. lead time was shorter
2. high satisfaction
3. enough information

Increase sales

1. GDS channel
2. during the peak season
3. select the main target
4. promotions by country

Conclusion

Limitation

Since this model is based on data from a specific hotel segment, there is no guarantee that results will be valid to every hotel. In addition, predicting the future with previous data may not be accurate as consumer needs are becoming more diverse and trends are changing rapidly. Therefore, efforts are needed to continuously develop services and promotions by collecting customer data in real time.

Hotel reservation

Derived from behavioral differences in customer characteristics :

operate the room efficiently to create positive results

Hotel Prices

sales can be predicted in advance by estimating the potential value per person

identify expected sales and sales goals in the mid- to long-term

adjust the prices of rooms

Hotel Marketing

Various marketing strategies and significant goals

relatively detailed insights into the various feature values favored by the target

References

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
https://hongl.tistory.com/136
https://romg2.github.io/dss/01_%EB%8D%B0%EC%9D%B4%ED%84%B0-%EC%82%AC%EC%9D%B4%EC%96%B8%EC%8A%A4-%EC%8A%A4%EC%BF%A8-4.9-%EB%AA%A8%ED%98%95%EC%9D%98-%EC%A7%84%EB%8B%A8%EA%B3%BC-%EC%88%98%EC%A0%95/+
https://datascienceschool.net/03%20machine%20learning/04.03%20%EC%8A%A4%EC%BC%80%EC%9D%BC%EB%A7%81.html
https://mingtory.tistory.com/m/140
https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand?sort=votes
lecture note 10 ~13, 16~17

Thank You!