FINAL PROJECT PRESENTATION

HOTEL BOOKING DEMAND

GEDE SATYA DANANJAYA

WHAT IS HOTEL CANCELATION?

Hotel cancellation is the act of canceling a reservation or booking a hotel room. This can be done by the guest who made the reservation or by the hotel due to a policy, change in circumstances, or other reasons.

CANCELLATION AND DEPOSIT

- Cancellation policy: This policy outlines the procedures and conditions under which a guest can cancel their hotel reservation, including the deadline for cancellation, whether a fee will be charged, and any refunds that may be given.
- Deposit policy: This policy outlines the amount of money that a guest must pay as a deposit to secure their reservation when the deposit must be paid, and whether the deposit is refundable or not.
- Prepayment policy: This policy outlines whether a guest must pay for their room in advance, how much they must pay, and whether prepayment is refundable or not.

HOTEL CANCELLATION RATES

In December 2019, Fornova conducted research that included over 200,000 hotels all over the world to gain insights into hotel cancellation rates and policies across the industry. At the time, and well before the pandemic hit, 38% of hotels had free cancellation policies, while 62% had a no-refund policy. In July 2020, they ran the same survey and the results were dramatically different. 58% of the hotels now offer free cancelations, while 42% of those hotels are still refusing to offer a refund. The shift was greatly impacted by the pandemic and now a year later.

Booking cancellations in the hospitality industry not only generate revenue loss and affect pricing and inventory allocation decisions, but they also, in overbooking situations, have the potential to affect the hotel's online social reputation

Therefore we want to prevent room cancelation from happening

In this Data science project, we are given the task to PREDICT whether the guest is canceling their booking or not

If hotel cancellations could be accurately predicted, it would bring several benefits to the hospitality industry:

- Improved Room Utilization
- Better Revenue Management
- Improved Planning and Preparation
- Better Customer Experience

Table Of Contents

1.DATA IMPORTING & CLEANING

2.DATA ANALYSIS & EXPLORATION

3.DEFINING PERFORMANCE METRICS

4.DEALING WITH IMBALANCED DATA

5.MODELLING

6.CONCLUSION

DATA IMPORTING AND CLEANING

HOTEL BOOKING DEMAND

#	Column	Non-Null Count	Dtype
0	hotel	119390 non-null	object
1	is_canceled	119390 non-null	int64
2	lead_time	119390 non-null	int64
3	arrival_date_year	119390 non-null	int64
4	arrival_date_month	119390 non-null	object
5	arrival_date_week_number	119390 non-null	int64
6	arrival_date_day_of_month	119390 non-null	int64
7	stays_in_weekend_nights	119390 non-null	int64
8	stays_in_week_nights	119390 non-null	int64
9	adults	119390 non-null	int64
10	children	119386 non-null	float64
11	babies	119390 non-null	int64
12	meal	119390 non-null	object
13	country	118902 non-null	object
14	market_segment	119390 non-null	object
15	distribution_channel	119390 non-null	object
16	is_repeated_guest	119390 non-null	int64
17	previous_cancellations	119390 non-null	int64
18	previous_bookings_not_canceled	119390 non-null	int64
19	reserved_room_type	119390 non-null	object
20	assigned_room_type	119390 non-null	object
21	booking_changes	119390 non-null	int64
22	deposit_type	119390 non-null	object
23	agent	103050 non-null	float64
24	company	6797 non-null	float64
25	days_in_waiting_list	119390 non-null	int64
26	customer_type	119390 non-null	object
27	adr	119390 non-null	float64
28	required_car_parking_spaces	119390 non-null	int64
29	total_of_special_requests	119390 non-null	int64
30	reservation_status	119390 non-null	object
31	reservation_status_date	119390 non-null	object

DATA IMPORTING AND CLEANING

IMPORTED DATA CONSIST OF 119210 ROWS AND 32 COLUMNS
HOTEL.CSV

REPLACING MISSING VALUES

1. agent: If no agency is given, booking was most likely made without one.
2. company: If none given, it was most likely private.

```
Missing value ratios:
Company: 94.30689337465449
Agent: 13.686238378423655
Country: 0.4087444509590417 7
```

DATA IMPORTING & CLEANING

1."meal" contains values "Undefined", which is equal to SC.
2.Some rows contain entreis with 0 adults, 0 children and 0 babies.
I'm dropping these entries with no guests.

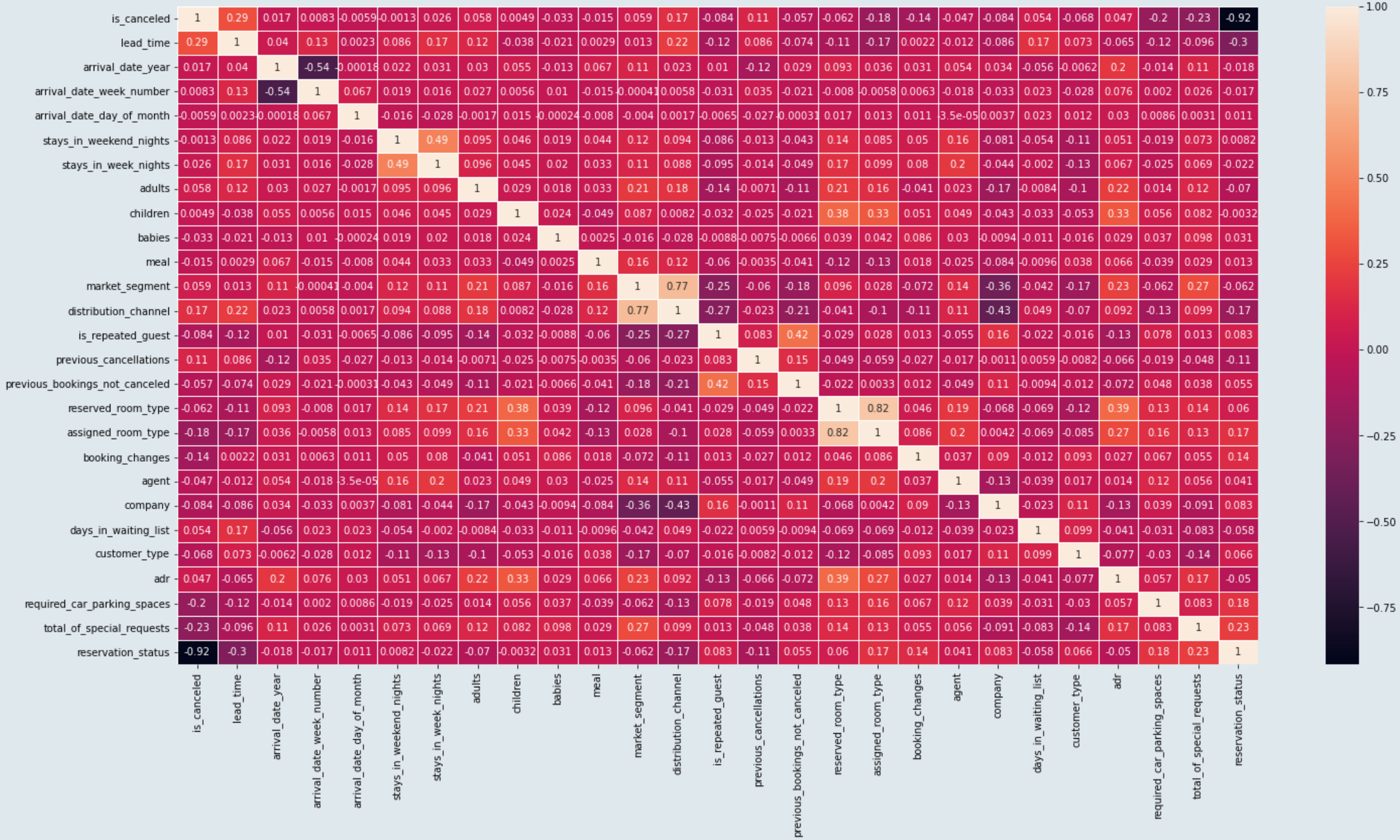
DATA ANALYSIS AND EXPLORATION

HOTEL BOOKING DEMAND

DATA ANALYSIS AND EXPLORATION

We already have function to determine corelation with df.coor. By default, df.corr() uses Pearson correlation coefficient, which measures the linear relationship between numerical variables.

reservation_status	-0.917172
total_of_special_requests	-0.234925
required_car_parking_spaces	-0.195696
assigned_room_type	-0.175841
booking_changes	-0.144821
is_repeated_guest	-0.083740
company	-0.083588
customer_type	-0.068276
reserved_room_type	-0.062218
previous_bookings_not_canceled	-0.057363
agent	-0.046988
babies	-0.032566
meal	-0.015196
arrival_date_day_of_month	-0.005902
stays_in_weekend_nights	-0.001316
children	0.004862
arrival_date_week_number	0.008299
arrival_date_year	0.016694
stays_in_week_nights	0.025549
adr	0.046558
days_in_waiting_list	0.054309
adults	0.058155
market_segment	0.059395
previous_cancellations	0.110147
distribution_channel	0.167650
lead_time	0.292930
is_canceled	1.000000

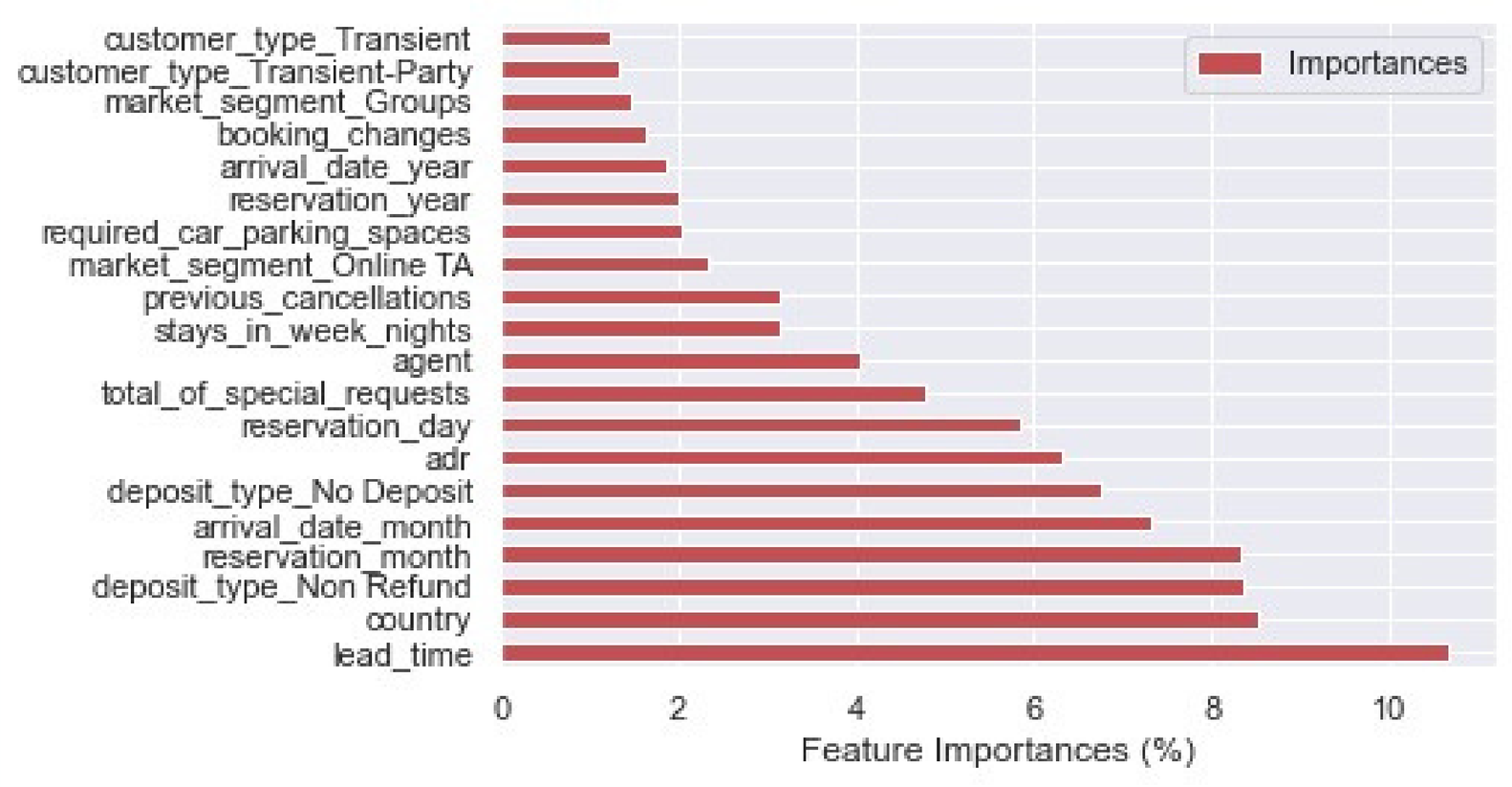

is_canceled 1 0.29 0.017 0.0083 -0.0059 -0.0013 0.026 0.058 0.0049 -0.033 -0.015 0.059 0.17 -0.084 0.11 -0.057 -0.062 -0.18 -0.14 -0.047 -0.084 0.054 -0.068 0.047 -0.2 -0.23 -0.92
lead_time 0.29 1 0.04 0.13 0.0023 0.086 0.17 0.12 -0.038 -0.021 0.0029 0.013 0.22 -0.12 0.086 -0.074 -0.11 -0.17 0.0022 -0.012 -0.086 0.17 0.073 -0.065 -0.12 -0.096 -0.3
arrival_date_year 0.017 0.04 1 -0.54 -0.00018 0.022 0.031 0.03 0.055 -0.013 0.067 0.11 0.023 0.01 -0.12 0.029 0.093 0.036 0.031 0.054 0.034 -0.056 -0.0062 0.2 -0.014 0.11 -0.018
arrival_date_week_number 0.0083 0.13 -0.54 1 0.067 0.019 0.016 0.027 0.0056 0.01 -0.015 -0.00041 0.0058 -0.031 0.035 -0.021 -0.008 -0.0058 0.0063 -0.018 -0.033 0.023 -0.028 0.076 0.002 0.026 -0.017
arrival_date_day_of_month -0.0059 0.0023 -0.00018 0.067 1 -0.016 -0.028 -0.0017 0.015 -0.00024 -0.008 -0.004 0.0017 -0.0065 -0.027 -0.00031 0.017 0.013 0.011 3.5e-05 0.0037 0.023 0.012 0.03 0.0086 0.0031 0.011
stays_in_weekend_nights -0.0013 0.086 0.022 0.019 -0.016 1 0.49 0.095 0.046 0.019 0.044 0.12 0.094 -0.086 -0.013 -0.043 0.14 0.085 0.05 0.16 -0.081 -0.054 -0.11 0.051 -0.019 0.073 0.0082
stays_in_week_nights 0.026 0.17 0.031 0.016 -0.028 0.49 1 0.096 0.045 0.02 0.033 0.11 0.088 -0.095 -0.014 -0.049 0.17 0.099 0.08 0.2 -0.044 -0.002 -0.13 0.067 -0.025 0.069 -0.022
adults 0.058 0.12 0.03 0.027 -0.0017 0.095 0.096 1 0.029 0.018 0.033 0.21 0.18 -0.14 -0.0071 -0.11 0.21 0.16 -0.041 0.023 -0.17 -0.0084 -0.1 0.22 0.014 0.12 -0.07
children 0.0049 -0.038 0.055 0.0056 0.015 0.046 0.045 0.029 1 0.024 -0.049 0.087 0.0082 -0.032 -0.025 -0.021 0.38 0.33 0.051 0.049 -0.043 -0.033 -0.053 0.33 0.056 0.082 -0.0032
babies -0.033 -0.021 -0.013 0.01 -0.00024 0.019 0.02 0.018 0.024 1 0.0025 -0.016 -0.028 -0.0088 -0.0075 -0.0066 0.039 0.042 0.086 0.03 -0.0094 -0.011 -0.016 0.029 0.037 0.098 0.031
meal -0.015 0.0029 0.067 -0.015 -0.008 0.044 0.033 0.033 -0.049 0.0025 1 0.16 0.12 -0.06 -0.0035 -0.041 -0.12 -0.13 0.018 -0.025 -0.084 -0.0096 0.038 0.066 -0.039 0.029 0.013
market_segment 0.059 0.013 0.11 -0.00041 -0.004 0.12 0.11 0.21 0.087 -0.016 0.16 1 0.77 -0.25 -0.06 -0.18 0.096 0.028 -0.072 0.14 -0.36 -0.042 -0.17 0.23 -0.062 0.27 -0.062
distribution_channel 0.17 0.22 0.023 0.0058 0.0017 0.094 0.088 0.18 0.0082 -0.028 0.12 0.77 1 -0.27 -0.023 -0.21 -0.041 -0.1 -0.11 0.11 -0.43 0.049 -0.07 0.092 -0.13 0.099 -0.17
is_repeated_guest -0.084 -0.12 0.01 -0.031 -0.0065 -0.086 -0.095 -0.14 -0.032 -0.0088 -0.06 -0.25 -0.27 1 0.083 0.42 -0.029 0.028 0.013 -0.055 0.16 -0.022 -0.016 -0.13 0.078 0.013 0.083
previous_cancellations 0.11 0.086 -0.12 0.035 -0.027 -0.013 -0.014 -0.0071 -0.025 -0.0075 -0.0035 -0.06 -0.023 0.083 1 0.15 -0.049 -0.059 -0.027 -0.017 -0.0011 0.0059 -0.0082 -0.066 -0.019 -0.048 -0.11
previous_bookings_not_canceled -0.057 -0.074 0.029 -0.021 -0.00031 -0.043 -0.049 -0.11 -0.021 -0.0066 -0.041 -0.18 -0.21 0.42 0.15 1 -0.022 0.0033 0.012 -0.049 0.11 -0.0094 -0.012 -0.072 0.048 0.038 0.055
reserved_room_type -0.062 -0.11 0.093 -0.008 0.017 0.14 0.17 0.21 0.38 0.039 -0.12 0.096 -0.041 -0.029 -0.049 -0.022 1 0.82 0.046 0.19 -0.068 -0.069 -0.12 0.39 0.13 0.14 0.06
assigned_room_type -0.18 -0.17 0.036 -0.0058 0.013 0.085 0.099 0.16 0.33 0.042 -0.13 0.028 -0.1 0.028 -0.059 0.0033 0.82 1 0.086 0.2 0.0042 -0.069 -0.085 0.27 0.16 0.13 0.17
booking_changes -0.14 0.0022 0.031 0.0063 0.011 0.05 0.08 -0.041 0.051 0.086 0.018 -0.072 -0.11 0.013 -0.027 0.012 0.046 0.086 1 0.037 0.09 -0.012 0.093 0.027 0.067 0.055 0.14
agent -0.047 -0.012 0.054 -0.018 3.5e-05 0.16 0.2 0.023 0.049 0.03 -0.025 0.14 0.11 -0.055 -0.017 -0.049 0.19 0.2 0.037 1 -0.13 -0.039 0.017 0.014 0.12 0.056 0.041
company -0.084 -0.086 0.034 -0.033 0.0037 -0.081 -0.044 -0.17 -0.043 -0.0094 -0.084 -0.36 -0.43 0.16 -0.0011 0.11 -0.068 0.0042 0.09 -0.13 1 -0.023 0.11 -0.13 0.039 -0.091 0.083
days_in_waiting_list 0.054 0.17 -0.056 0.023 0.023 -0.054 -0.002 -0.0084 -0.033 -0.011 -0.0096 -0.042 0.049 -0.022 0.0059 -0.0094 -0.069 -0.069 -0.012 -0.039 -0.023 1 0.099 -0.041 -0.031 -0.083 -0.058
customer_type -0.068 0.073 -0.0062 -0.028 0.012 -0.11 -0.13 -0.1 -0.053 -0.016 0.038 -0.17 -0.07 -0.016 -0.0082 -0.012 -0.12 -0.085 0.093 0.017 0.11 0.099 1 -0.077 -0.03 -0.14 0.066
adr 0.047 -0.065 0.2 0.076 0.03 0.051 0.067 0.22 0.33 0.029 0.066 0.23 0.092 -0.13 -0.066 -0.072 0.39 0.27 0.027 0.014 -0.13 -0.041 -0.077 1 0.057 0.17 -0.05
required_car_parking_spaces -0.2 -0.12 -0.014 0.002 0.0086 -0.019 -0.025 0.014 0.056 0.037 -0.039 -0.062 -0.13 0.078 -0.019 0.048 0.13 0.16 0.067 0.12 0.039 -0.031 -0.03 0.057 1 0.083 0.18
total_of_special_requests -0.23 -0.096 0.11 0.026 0.0031 0.073 0.069 0.12 0.082 0.098 0.029 0.27 0.099 0.013 -0.048 0.038 0.14 0.13 0.055 0.056 -0.091 -0.083 -0.14 0.17 0.083 1 0.23
reservation_status -0.92 -0.3 -0.018 -0.017 0.011 0.0082 -0.022 -0.07 -0.0032 0.031 0.013 -0.062 -0.17 0.083 -0.11 0.055 0.06 0.17 0.14 0.041 0.083 -0.058 0.066 -0.05 0.18 0.23 1
is_canceled
lead_time
arrival_date_year
arrival_date_week_number
arrival_date_day_of_month
stays_in_weekend_nights
stays_in_week_nights
adults
children
babies
meal
market_segment
distribution_channel
is_repeated_guest
previous_cancellations
previous_bookings_not_canceled
reserved_room_type
assigned_room_type
booking_changes
agent
company
days_in_waiting_list
customer_type
adr
required_car_parking_spaces
total_of_special_requests
reservation_status
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

DATA ANALYSIS AND EXPLORATION

1. As we see in the sorted list, reservation_status is the most impactful feature. With that, the information accuracy rate should be really high. It can be better to drop the `reservation_status` column to see how other features can predict.
2. Apart from that, I will not use `arrival_date_week_number`, `stays_in_weekend_night` and `arrival_date_day_of_month` since their importances are really low while predicting cancellations.
3. In the Multiple correlation heatmaps we will drop `distribution_channel` and`assigned_room_type` because it's highly correlated with the column `market_segment` and `reserved_room_type` to reduce potential overfitting in modeling

DATA ANALYSIS AND EXPLORATION

Feature Importance



DATA ANALYSIS AND EXPLORATION

1. Feature importance was determined using a Random Forest classifier
2.The top five most important features were lead_time, country, reservation_month, Deposit_type and arrival_date_month with lead_time having the highest importance score of 9.9
3.lead_time being the most important feature suggests that it has the largest impact on the outcome of the model.
4.The number of days that elapsed between the entering date of the booking into the PMS and the arrival date is used in hotels, resorts, vacation rental properties and other types of accommodation businesses as a measure of how early or late a booking is made. It could be used as a feature for predicting the cancellation of a reservation or as a variable in pricing strategies.

DATA ANALYSIS AND EXPLORATION

- Where does the guest come from?

As we can see in the pie chart the top 5 countries that visited our hotels are from: * Portugal

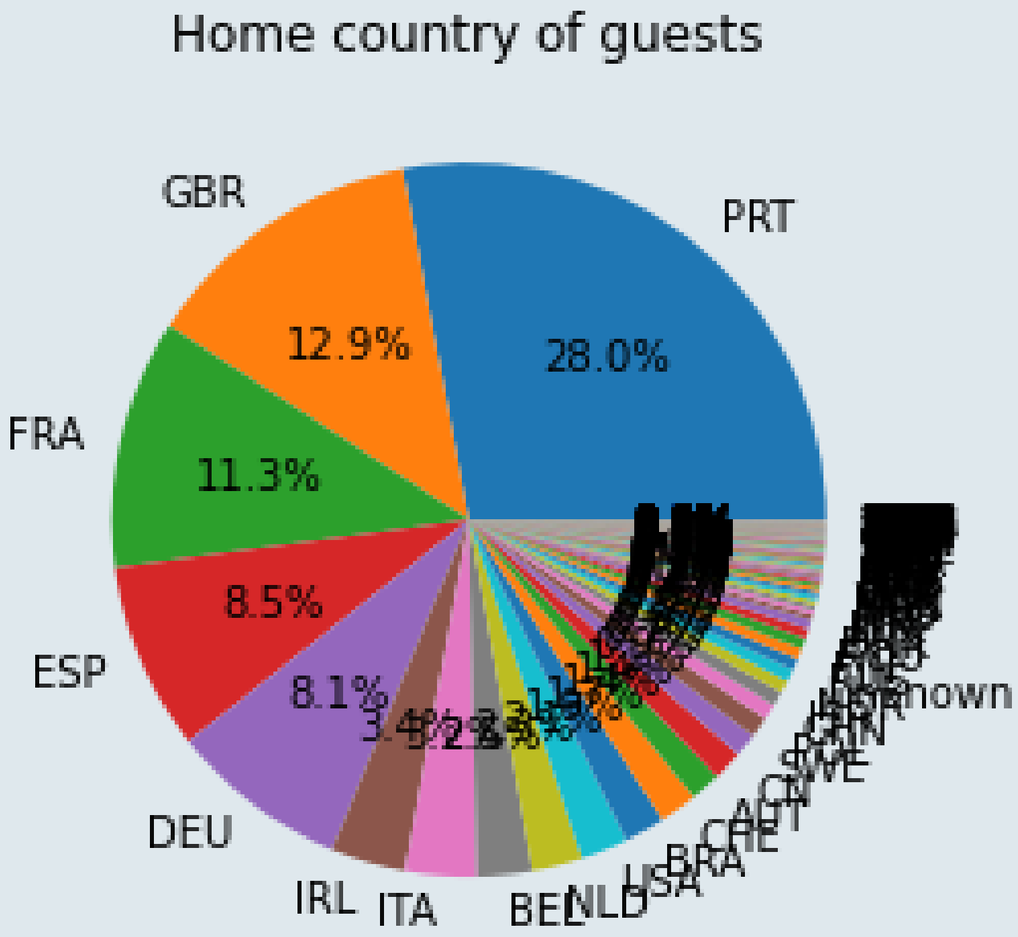
* Uk (United Kingdom)

* France

* Spain

* Germany

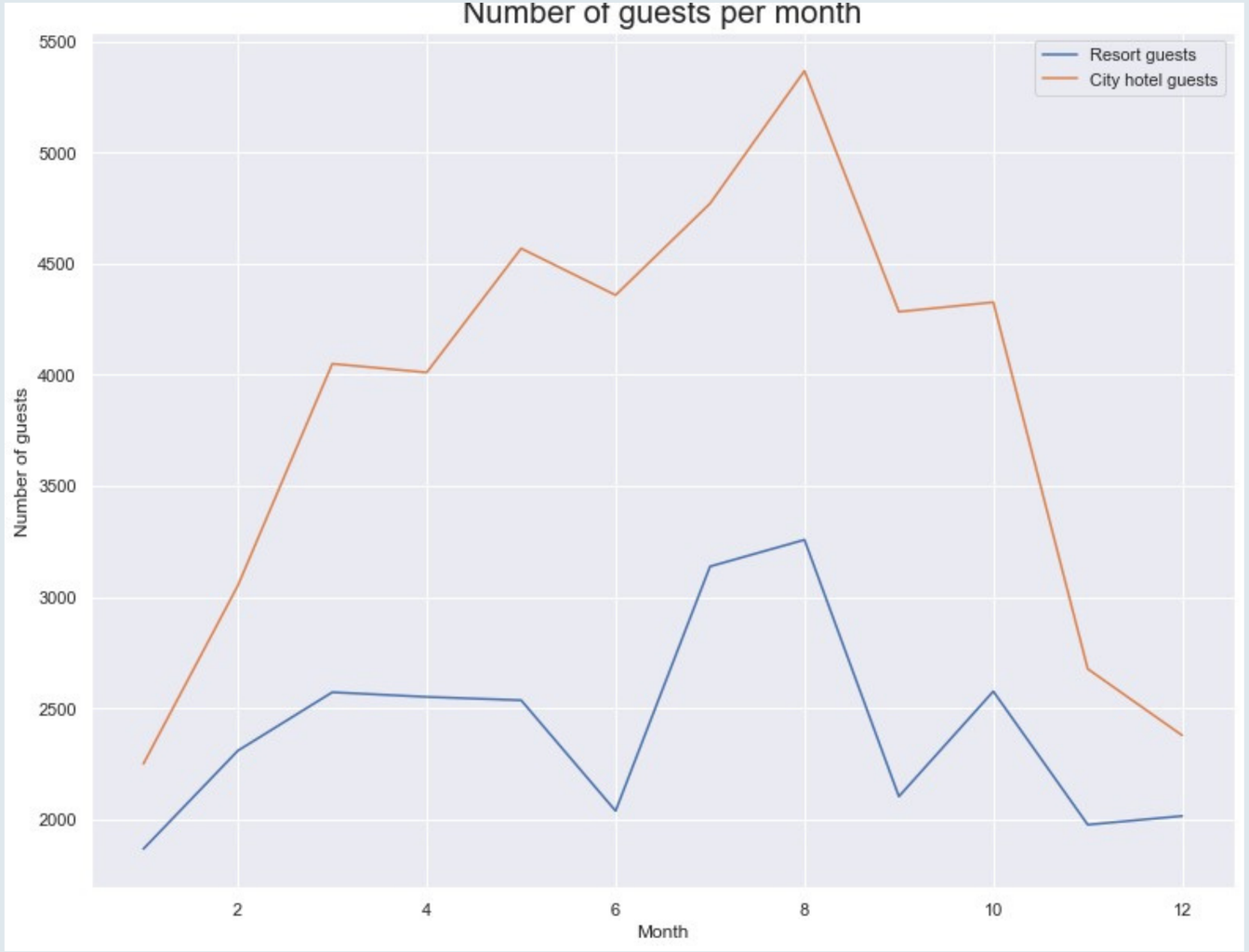
So we can draw the conclusion that most people from all over the world are staying in these two hotels. Most guests are from Portugal and other countries in Europe.



DATA ANALYSIS AND EXPLORATION

- Number of guests per month

From the following table we can see that the middle of the year (August, July, May, October, April, June, September) is a very busy month with customers, this is likely because in Europe it is summer and most people have holidays



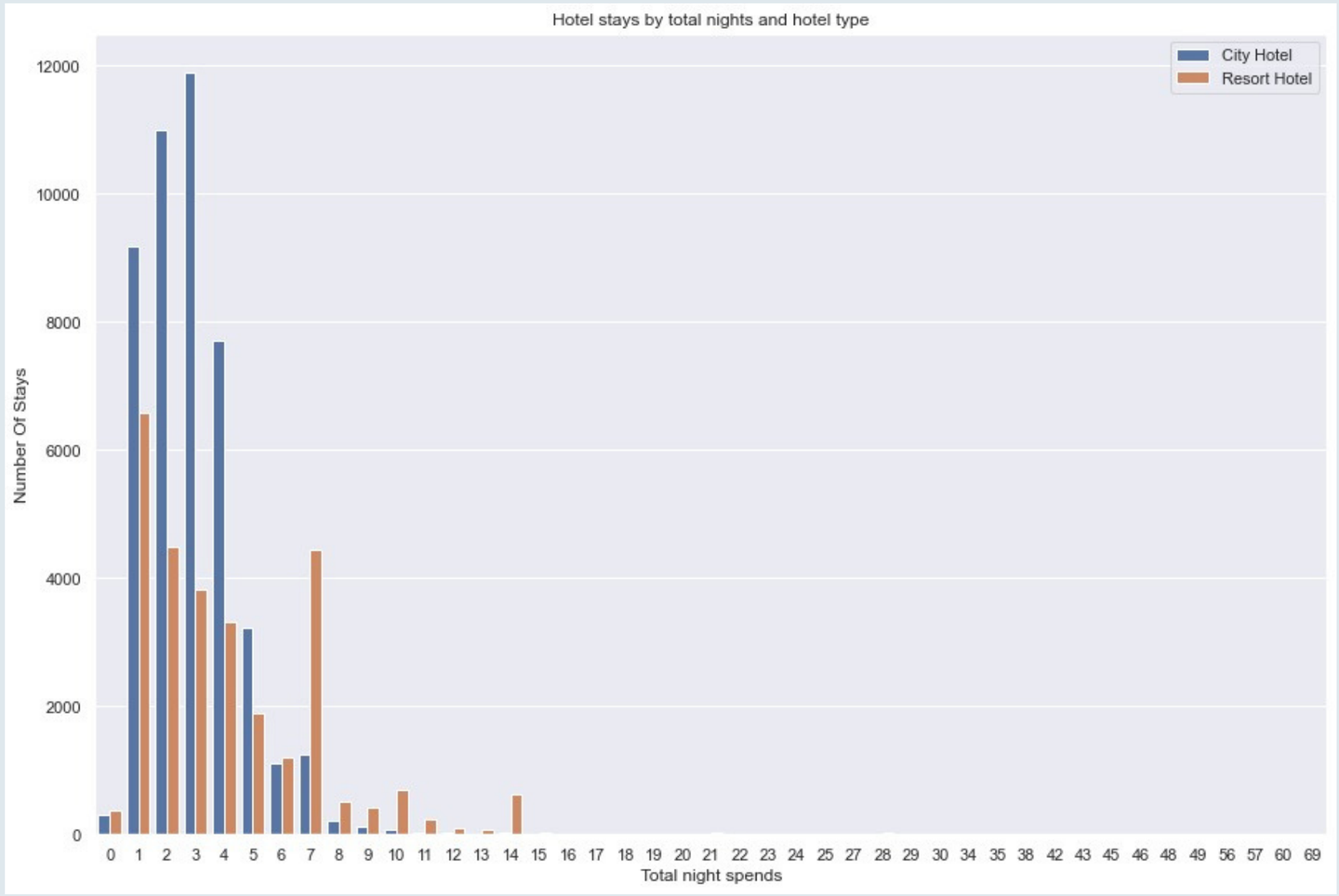
SUGGESTION

1.Offer attractive packages and discounts: Offer special packages and deals to guests during the low season to incentivize them to travel.
2.Promote the destination: Highlight the unique qualities of the destination, such as cultural events, outdoor activities, or local cuisine, to entice guests to visit.

DATA ANALYSIS AND EXPLORATION

- Total night spends by guests

In this bar plot we can see that most of the customers use city hotels compared to resort hotels and the average customer books hotels for a period of 1-14 days with peak bookings for resort hotels for 1 day and city hotels for 3 days.



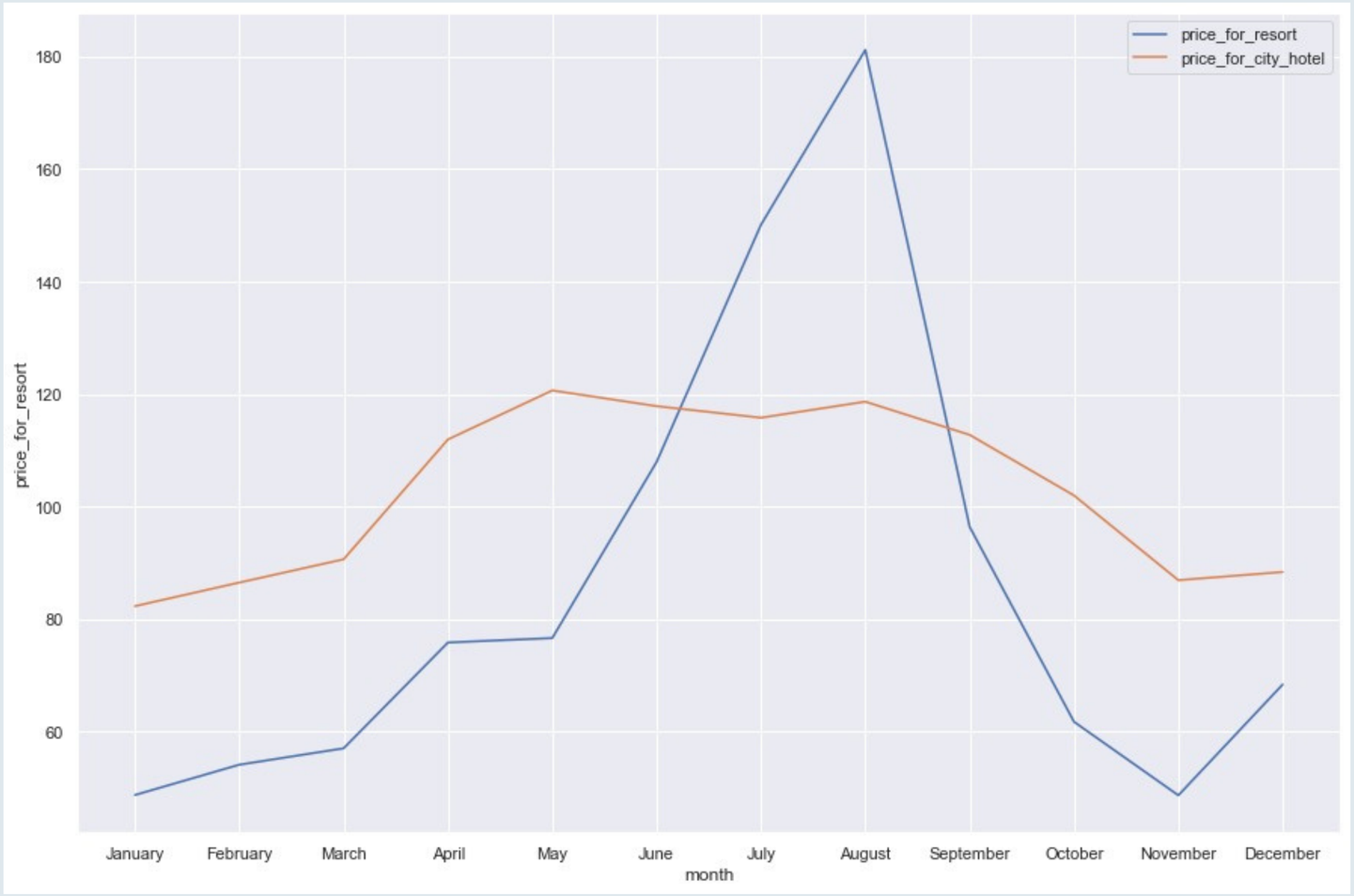
SUGGESTION

1. Offer special packages for extended stays: Offer discounts or special amenities for guests who book extended stays, making it more appealing for them to stay longer.
2. Enhance the guest experience: Invest in improving the overall guest experience, such as updating amenities, adding new services, or providing a more personalized touch. A positive guest experience can lead to increased guest loyalty and repeat stays.

DATA ANALYSIS AND EXPLORATION

- Price per month

From the graph above, we can see that at the beginning of the year, there was a significant increase from January, rising to May and then there was a sharp increase from June to September and starting to decline from October to November. The increase again occurred from November to December.



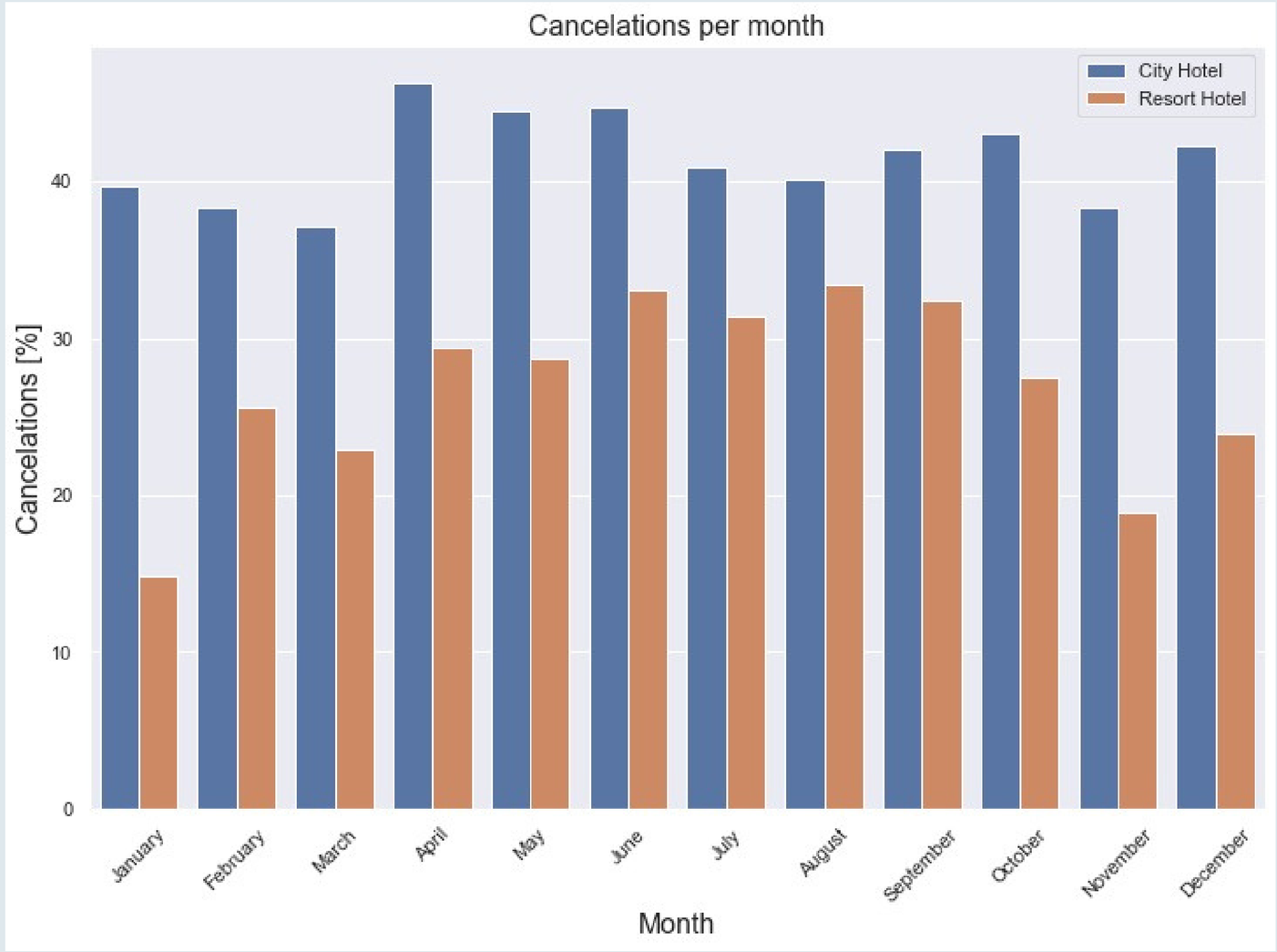
SUGGESTION

To adjust pricing during high season, the following strategies can be considered:
1. Dynamic Pricing: Implement dynamic pricing to adjust prices in real time based on demand, ensuring that prices are optimized to maximize revenue.
2. Use Price Discrimination: Use price discrimination to charge different prices to different segments of the market based on factors such as booking lead time, room type, or length of stay.

DATA ANALYSIS AND EXPLORATION

- Cancelation per month

Most cancellations occur at city hotels compared to resort hotels, city hotels have an average cancellation rate of up to 40% and resort hotels have a fluctuating cancellation rate, increasing in summer and decreasing in winter



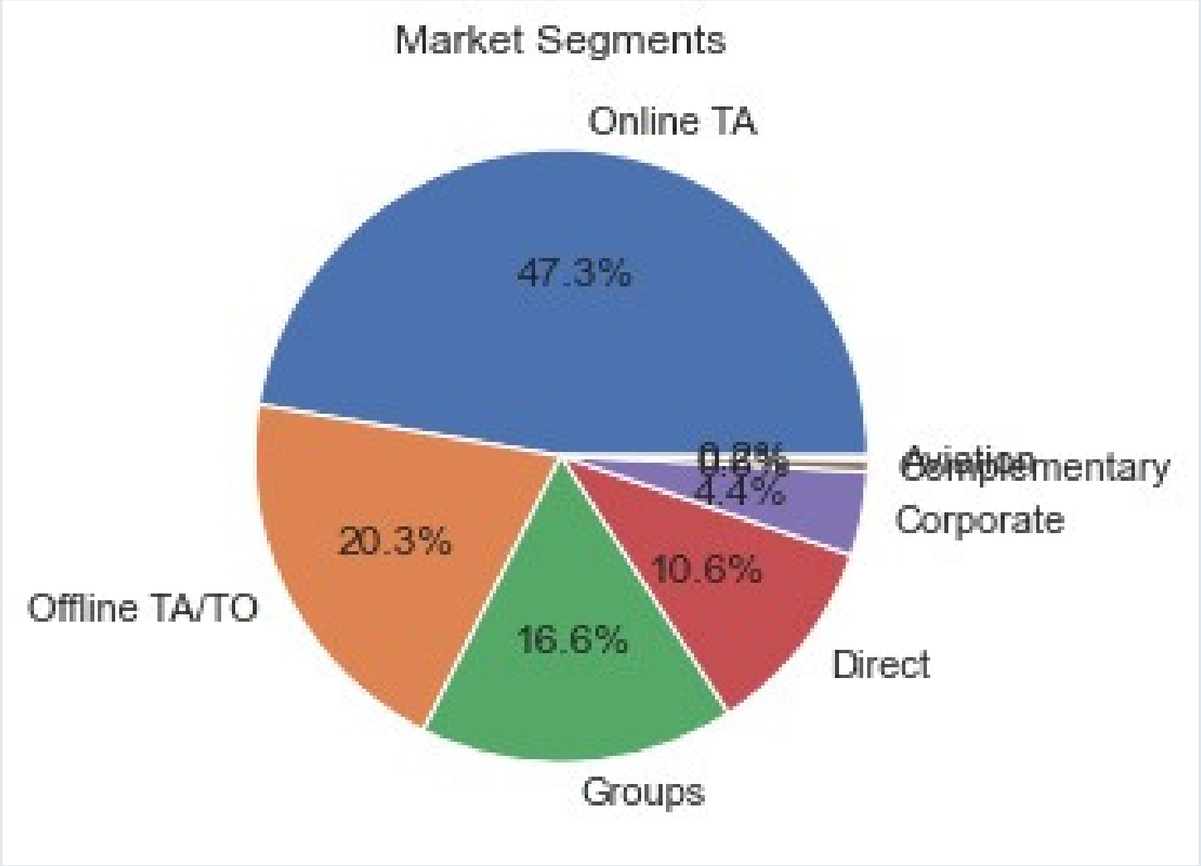
SUGGESTION

1. Offer flexible cancellation policies: Offer flexible cancellation policies, such as allowing cancellations up to a certain number of days before check-in, to reduce the risk of cancellations.
2. Offer alternative solutions: If a cancellation does occur, offer alternative solutions, such as a voucher for a future stay, to encourage the guest to reschedule rather than cancel.

DATA ANALYSIS AND EXPLORATION

- Bookings by market segment

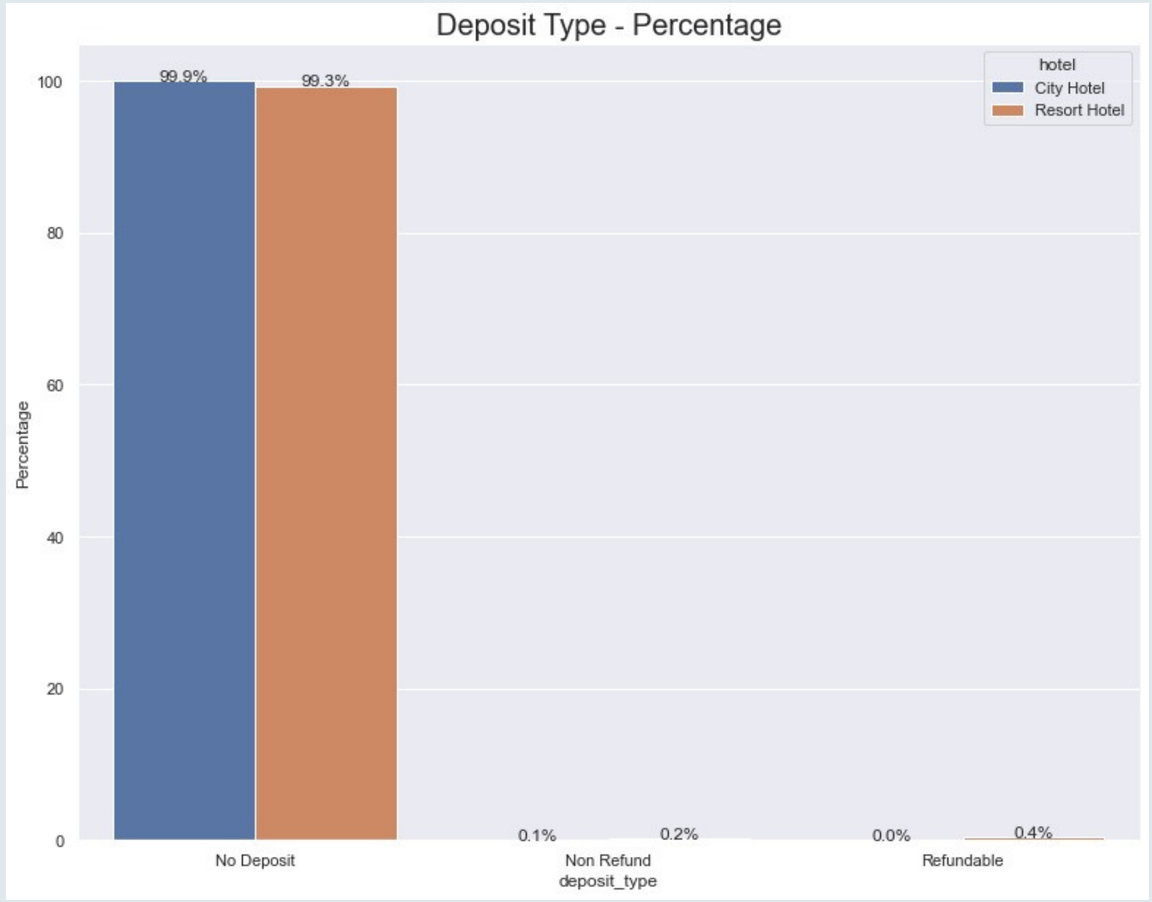
From the following pie chart, online bookings are mostly made by Tour Agency, followed by offline bookings and group bookings.



DATA ANALYSIS AND EXPLORATION

- How do deposit types influence cancelation?

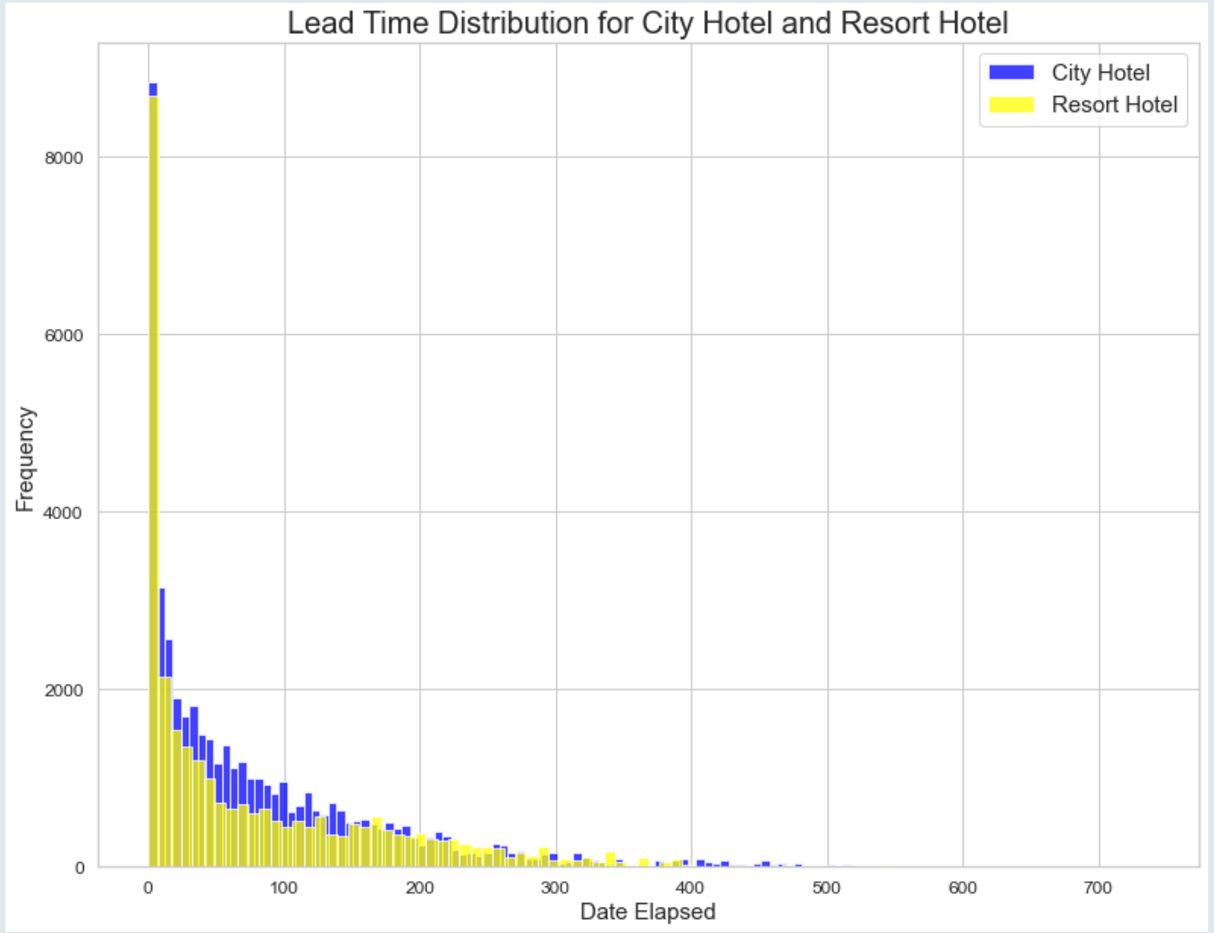
In deposit type no one seems to cancel in the form of refundable type, this is pretty weird considering non-refund has a better choice. this might be happening because no deposit has a lower pricing than refundable option.



DATA ANALYSIS AND EXPLORATION

- Lead Time distribution

on average, bookings made for city hotels have a lead time of 80.7 days, and for resort hotels 78.8 days. This means that most bookings for city hotels are made around 50 days before the arrival date, with some bookings made as early as 12 days in advance and some as late as 121 days in advance. The maximum lead time for a city hotel booking was 518 days.



SUGGESTION

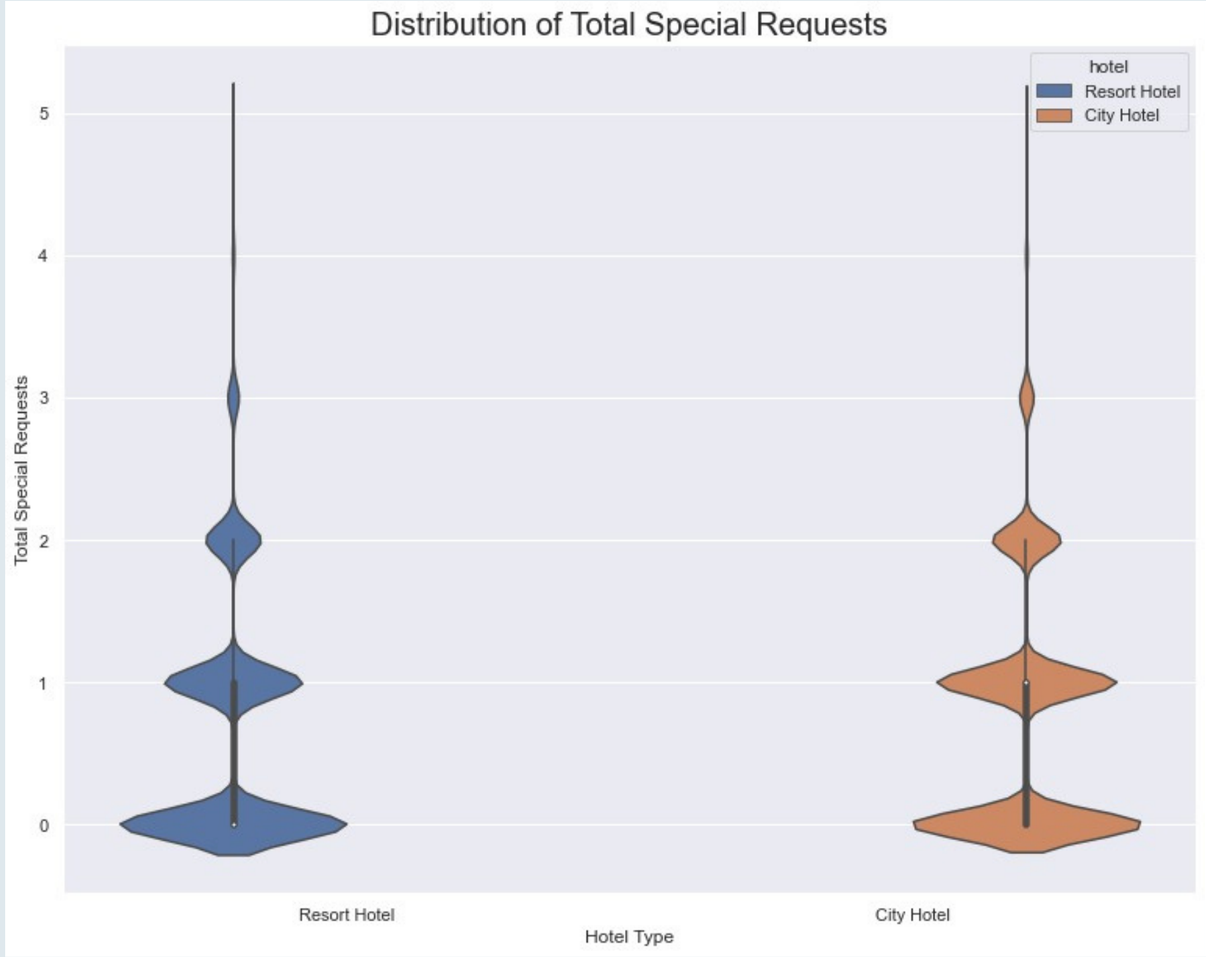
- Based on the average lead time of 80.7 days for city hotels and 78.8 days for resort hotels, the hotel industry can take advantage of these insights by planning their staffing, promotions, and room allocation more effectively.
- For example, they can allocate rooms further in advance for bookings made 50 days before the arrival date to ensure high occupancy levels.
- Additionally, the hotel industry could consider running promotions for last-minute bookings to encourage guests to book even if they're making reservations just a few days in advance.
- To optimize staffing, the hotel industry could also predict the occupancy levels for different lead times and adjust their staffing levels accordingly.

DATA ANALYSIS AND EXPLORATION

- Total special request by guest

Most guest in a resort or city hotel has less than 3 requests. this can be beneficial to hotels because

1. Reduces staff workload
2. Improves financial performance

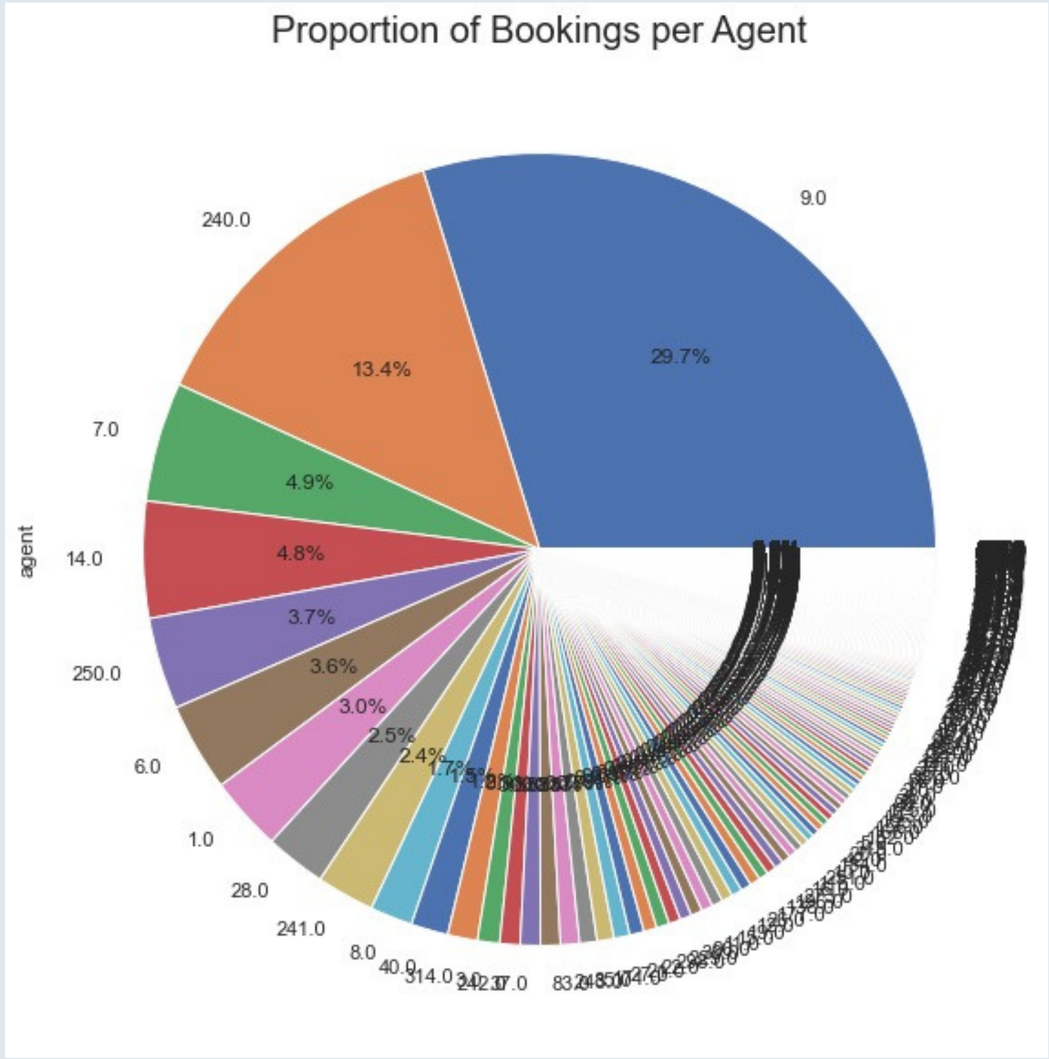


DATA ANALYSIS AND EXPLORATION

- The proportion of Bookings per Agent

Agent id 9 has the most booked agent.

in this case can do providing incentives, such as bonuses or rewards, for top-performing agents can encourage them to continue to perform well and drive more bookings. Recognizing top performers in front of their peers can also help boost morale and motivate others to improve their performance.



DEFINING PERFORMANCE METRICS

HOTEL BOOKING DEMAND

DEFINING PERFORMANCE METRICS

WE NEED TO PAY MORE ATTENTION TO THE CONFUSION MATRIX AND OTHER METRICS SUCH AS:

- PRECISION
- RECALL
- F1-SCORE
- ACCURACY

		Predicted	
		Negative **(N)** -	Positive **(P)** +
Actual	Negative -	True Negative **(TN)**	**False Positive (FP) Type I Error**
	Positive +	**False Negative (FN) Type II Error**	True Positive **(TP)**

DEFINING PERFORMANCE METRICS

- Which label is worse?
- False negative: Predicting guest to not canceled reservation but he/she ACTUALLY HAS CANCELLED, or
- -False Positive: Predicting patient to cancel the reservation but he/she ACTUALLY NOT CANCELING
- -True Positive: Accurately prediction of guest has canceled reservation -
- True Negative: Accurately prediction of the guest not canceled reservation

DEFINING PERFORMANCE METRICS

- Recall is a performance metric that is crucial in evaluating the performance of a model, especially in cases where the consequences of false negatives (cases where a positive event is predicted as negative)
- High recall values indicate that the model has a low rate of false negatives and is able to accurately identify the majority of positive cases.
- The recall, also known as sensitivity or true positive rate, measures: "Out of all guests who actually cancelled their reservation, how many were correctly identified as having cancelled?"

DEALING WITH IMBALANCED DATA

HOTEL BOOKING DEMAND

DEALING WITH IMBALANCED DATA

- There are fewer guests that cancel the booking
- Our data is considered to have a mild imbalance according to google.

	is_canceled	freq	percentage
0	0	75011	62.93
1	1	44195	37.07

DEALING WITH IMBALANCED DATA

The danger of imbalance:

- Poor performance on minority class: Models trained on imbalanced data may have poor performance on the minority class, resulting in low accuracy, precision, and recall for that class.
- In our case, the ML model might face difficulties in predicting guest cancelation, which is bad

To address these issues, it is important to either balancing the dataset or to use appropriate evaluation metrics and techniques, such as stratified cross-validation or class weighting.

DEALING WITH IMBALANCED DATA

Options:

1. Random Under sampling: Random under sampling involves reducing the number of examples in the majority class by randomly selecting a subset of those examples to remove from the dataset.
2. Random Over Sampling: Random oversampling involves increasing the number of examples in the minority class by randomly selecting and duplicating examples from that class.
3. SMOTE (Synthetic Minority Over-sampling Technique): oversampling technique for balancing an imbalanced dataset. It works by generating synthetic examples for the minority class rather than simply duplicating existing examples.

MODELLING

HOTEL BOOKING DEMAND

MODELLING

- First, we need to determine which algorithms we will use.
- We check accuracy using a classifier of 3 algorithms, Random Forest, XGBoost, and KNN.

Model	Accuracy (Imbalanced Data)	Accuracy (Balanced Data)
Random Forest	94.3%	94.7%
XGBoost	94.2%	95.1%
K-Nearest Neighbor	80.2%	82.9%

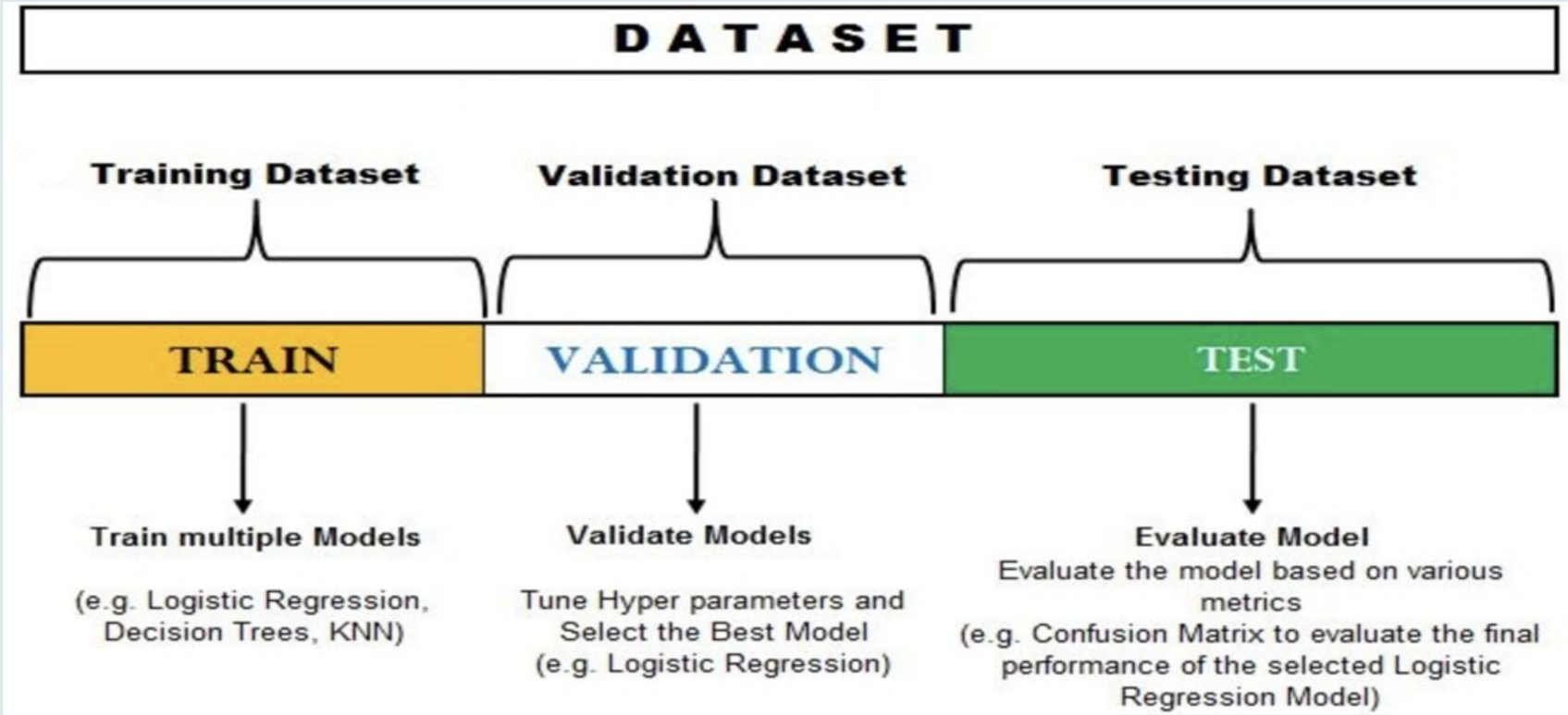
- With imbalanced data handling, we get more than 1% in accuracy in XGB and 2% in KNN

MODELLING

Why did we choose XGBoost?

1. Speed: XGBoost is faster than Random Forest as it uses gradient boosting which is a more efficient method of building decision trees.
2. Scalability: XGBoost can handle large datasets and is scalable in terms of parallel processing, making it suitable for large-scale problems.

MODELLING



We use 70% train data 15% validation and 15% test data

MODELLING

- Before we check the precision, recall, and f1-score we will do hyperparameter tuning first, why do we need to do hyperparameter tuning?
- Hyperparameter tuning is an important step in the machine learning model-building process because it helps to optimize the performance of the model.
- The goal of hyperparameter tuning is to find the best combination of hyperparameters that yield the best results in terms of accuracy, precision, recall, F1 score, or any other performance metric.

MODELLING

Model	Precision	Recall	F1-Score
Random Forest	89%	72%	79%
XGBoost	97%	91%	94%

MODELLING

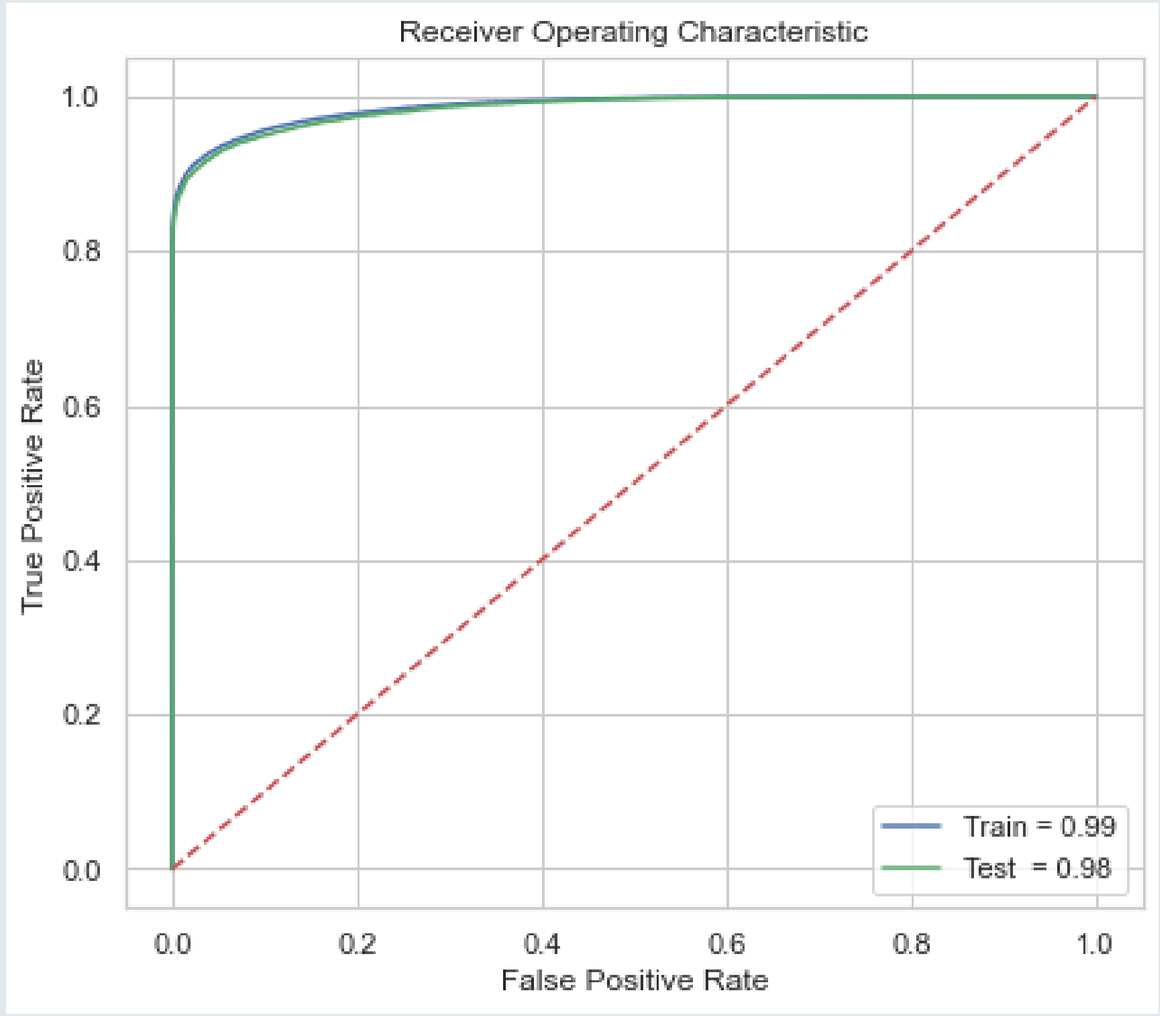
* After tuning XGBoost to get 91% Recall this makes XGBoost the best algorithm for determining the results of the dataset
* However, when using CVgridsearch XGBoost without being randomized, the process takes a very long time
* The model is neither underfit nor overfit because of the high precision and recall scores

MODELLING

- The results of the evaluation of the classification model used in the analysis show that the model has an accuracy of 94% in predicting whether a booking will be cancelled or not. The model has a precision of 0.92 for predicting cancellations and a recall of 0.97 for correctly identifying cancellations.
- The model's F1 score, which is a weighted average of precision and recall, is 0.94, indicating a good balance between precision and recall.
- The results of the evaluation show that the model is effective in accurately predicting cancellations and can be used as a useful tool for hotel management to make informed decisions about resource allocation and risk management.

MODELLING

- Based on the numbers given, the model has a good performance in terms of AUC and accuracy, with AUC scores of 98.68% on the training set and 98.46% on the test set, and accuracy scores of 94.40% on the training set and 94.02% on the test set.
- The recall, specificity, precision, and F1 scores are also high, indicating that the model is able to make correct predictions for both positive and negative classes. The log loss is relatively low, indicating good performance in terms of error.



CONCLUSION

- In conclusion, the analysis of the hotel booking data has shown that the top five most important features affecting bookings are lead_time, country, reservation_month, deposit_type, and arrival_date_month.
- These features provide valuable insights into the booking patterns and trends in the hotel industry. Understanding the lead time of bookings, the countries that the guests are coming from, the months in which the bookings are being made and when the guests are arriving, and the types of deposits being used can help the stakeholders make informed decisions about future bookings, marketing campaigns, and resource allocation.
- By analyzing these features, hotel management can gain a better understanding of their customer base and take steps to improve their offerings and services to meet their needs and preferences.

Gede Satya Dananjaya

Thank you!

linkedin.com/in/satyadananjaya