Business Problem: Hotel Reservation

Richie Tran – Machine Learning Business Analytics Project

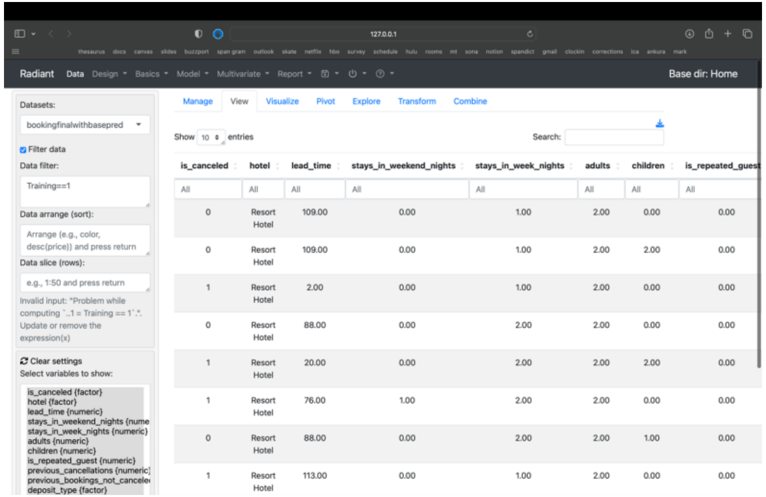
Business Problem

- **Problem Statement:** Hotels are seeing elevated booking cancellations, directly reducing revenue.
  - Broad fixes like marketing, renovations, or blanket customer initiatives are **costly and ineffective**.
- **Goal:** Use existing hotel data to identify the characteristics of guests most likely to cancel.
- **Impact:** Develop predictive ML models that flag high-risk bookings so hotels can intervene early, improve occupancy, and protect revenue.

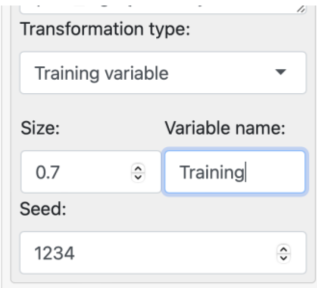
Data Source & Summary Statistics

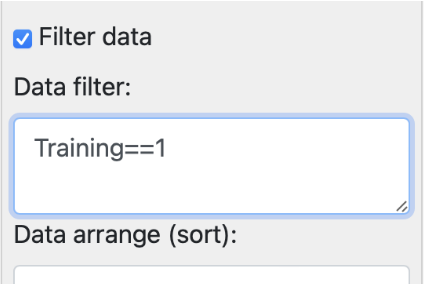
About the Dataset

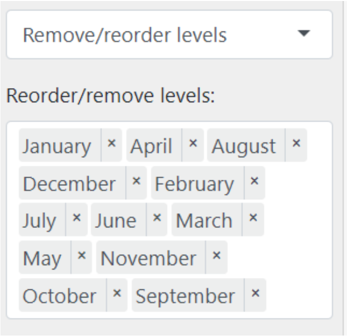
- Combined two datasets: **40,060 resort bookings and 79,330 city hotel bookings**
- Each booking includes **31 descriptive variables**
- Covers the period July 2015 – August 2017
- All personal identifiers were removed for privacy
- Data includes both cancelled and completed bookings.



Data Filtering







- Filtered to 2016 bookings, reducing the dataset to **58,000 observations**
- Created a **training indicator** and split data from the start
- Cleaned and reorganized variables for easier modeling.

	OR	OR%	coefficient	p.value
(Intercept)			-14.566	0.987
country|AGO	5351282.717	535,128,171.7%	15.493	0.986
country|AIA	1.000	0.0%	0.000	1.000
country|ALB	1412144.050	141,214,305.0%	14.161	0.987
country|AND	2118216.075	211,821,507.5%	14.566	0.987
country|ARE	11120634.396	1,112,063,339.6%	16.224	0.985
country|ARG	761233.902	76,123,290.2%	13.543	0.988
country|ARM	1412144.050	141,214,305.0%	14.161	0.987
country|ATA	1.000	0.0%	0.000	1.000
country|AUS	535128.272	53,512,727.2%	13.190	0.988
country|AUT	404239.709	40,423,870.9%	12.910	0.988
country|AZE	2118216.075	211,821,507.5%	14.566	0.987
country|BDI	1.000	0.0%	0.000	1.000
country|BEL	493061.016	49,306,001.6%	13.108	0.988
country|BEN	4486762877271.028	448,676,287,727,002.8%	29.132	0.981
country|BGD	10591080.377	1,059,107,937.7%	16.176	0.985
country|BGR	423643.215	42,364,221.5%	12.957	0.988
country|BHR	1.000	0.0%	0.000	1.000
country|BIH	1059108.038	105,910,703.8%	13.873	0.987
country|BLR	2118216.075	211,821,507.5%	14.566	0.987
country|BOL	1.000	0.0%	0.000	1.000

Variance Inflation Factors

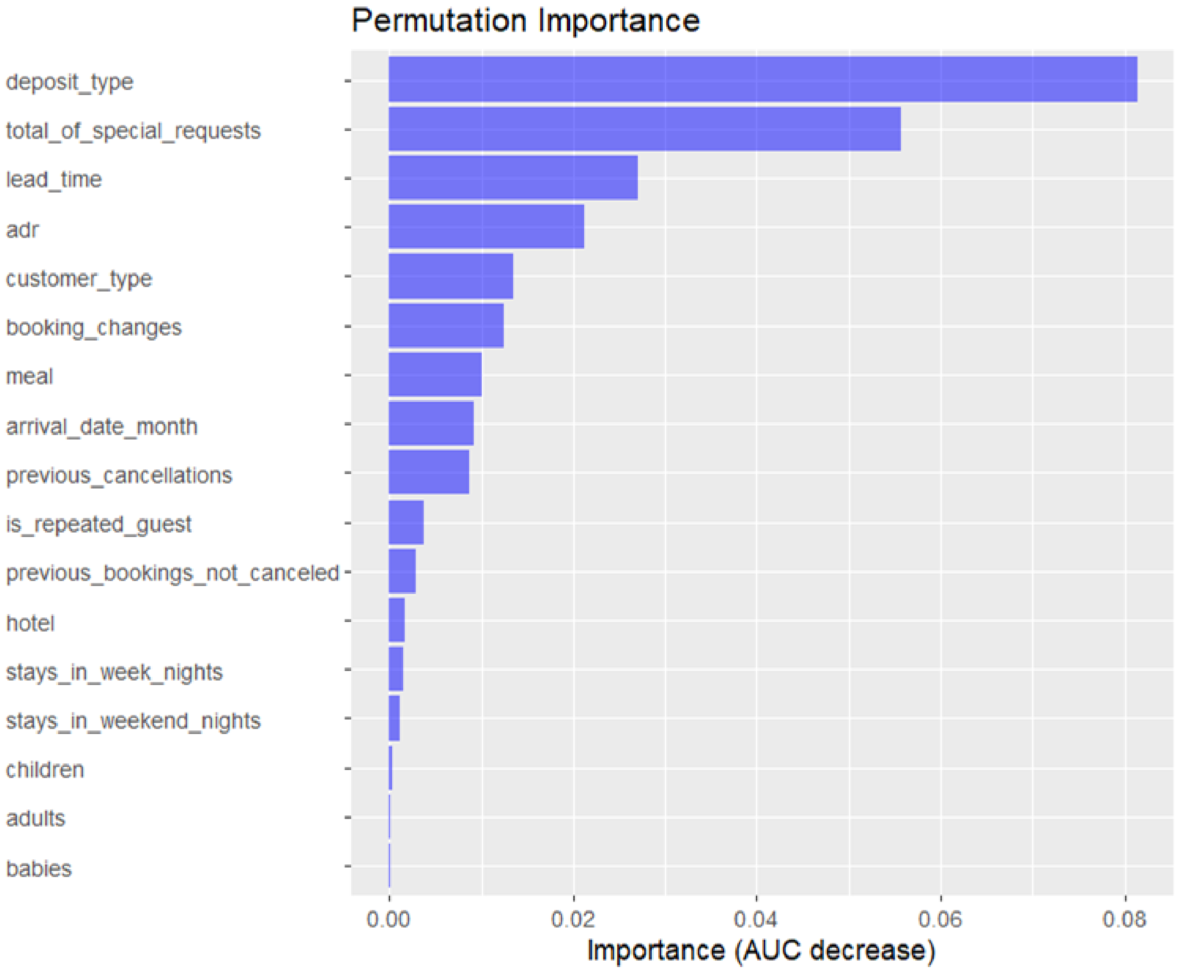
	VIF	Rsq
arrival_date_month	189.263	0.995
arrival_date_week_number	117.048	0.991
adr	1.594	0.373
previous_bookings_not_canceled	1.585	0.369
lead_time	1.541	0.351
previous_cancellations	1.481	0.325
pred_gbtwithcountry	1.451	0.311
is_repeated_guest	1.376	0.273
stays_in_week_nights	1.357	0.263
total_of_special_requests	1.322	0.243
meal	1.285	0.222
customer_type	1.264	0.209
stays_in_weekend_nights	1.246	0.198
hotel	1.191	0.160
days_in_waiting_list	1.100	0.091
deposit_type	1.042	0.040
booking_changes	1.023	0.023

Data Dictionary

Variable Name	Description
Hotel:	Hotel (H1 = Resort Hotel or H2 = City Hotel)
Lead_time:	Number of days between the entering date of the booking and the arrival date
Is_canceled:	Value indicating if the booking was canceled (1) or not (0)
Arrival_date_month:	Month of booking arrival
Stays_in_weekend_nights:	Number of weekend nights (Saturday or Sunday) booked to stay
Stays_in_week_nights:	Number of week nights (Monday to Friday) booked to stay
is_repeated_guest:	Repeated guest (1) or not (0)
previous_cancellations:	Number of previous bookings that were cancelled
Previous_bookings_not_canceled:	Number of previous bookings not cancelled
booking_changes:	Number of amendments made to the booking
deposit_type:	No deposit, non refund, or refundable
days_in_waiting_list:	Number of days the booking was in waiting before confirmation
customer_type:	Four categories: Contract, Group, Transient, Transient-party
adr:	Average Daily Rate
Meals:	SC/Undefined – No Meals; BB – Bed & Breakfast; HB – BB + One Meal; FB - Three Meals
Total_of_special_requests:	Number of special requests

Key Variables That Logically Influence Cancellations

- **Lead_time**
  - Longer time between booking and arrival increases the chance of cancellations due to price changes, plan changes, or unexpected events
- **Deposit_type**
  - Non-refundable deposits reduce cancellations; refundable or no-deposit bookings make cancelling easier
- **Booking_changes**
  - Frequent modifications may signal uncertainty, increasing the likelihood of cancellation.



1. Deposit_type
2. Total_of_special_requests
3. Lead_time
4. ADR
5. Customer_type

Models & Findings

Models Used

- Logistic Regression
- Decile Lift Chart
- Confusion Matrix
- Neural Network
- Gradient Boosted Trees
- Random Forest

Results Analysis: Logistic Regression

```
Confusion matrix
Data        : RT_bookingfinal
Filter      : Train=="1"
Results for: Both
Predictors : pred_logit
Response    : is_canceled
Level       : 1 in is_canceled
Cost:Margin: 1 : 2

     Type  Predictor    TP    FP     TN    FN  total   TPR   TNR precision Fscore
 Training pred_logit 6,437 1,000 24,460 7,798 39,695 0.452 0.961     0.866  0.594
     Test pred_logit 2,781   480 10,430 3,321 17,012 0.456 0.956     0.853  0.594

     Type  Predictor accuracy kappa    profit index  ROME contact   AUC
 Training pred_logit    0.778 0.462 5,437.000 1.000 0.731   0.187 0.809
     Test pred_logit    0.777 0.459 2,301.000 1.000 0.706   0.192 0.812
```

Results Analysis: Logistic Regression

Resort Hotel - (18.29%)

Lead Time - 0.40%

February - 60.32%

March - 25.36%

October - 50.98%

November - 47.55%

December - 92.13%

Stays Weekend - 8.33%

Stays Week - 5.34%

Repeat Guest - (63.76%)

Prev Cancellations - 134.67%

Prev Bookings Not Cancelled - (18.21%)

Booking Changes - (34.62%)

Non Refund - 10141.16%

Refundable - (77.66%)

Transient - 217.40%

Transient Party - 64.21%

ADR - 0.80%

HB Meal - (43.95%)

SC Meal - 45.35%

Special Req - (51.95%)

Filter data

Create pivot table

Categorical variables:

pred_logit_entiredata_025cut {numeric}

is_canceled {factor}

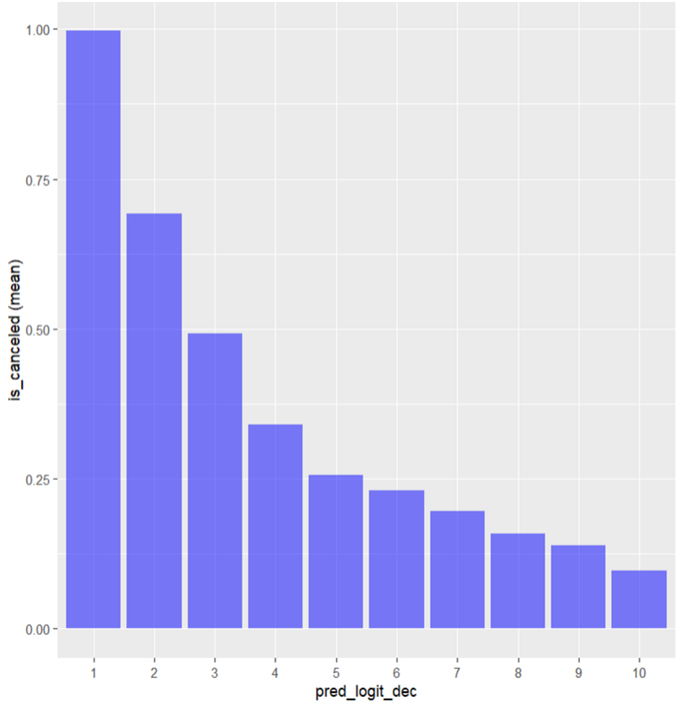
pred_logit_entiredata_025cut

is_canceled	X0	X1	Total
All	All	All	All
0	20,743	15,627	**36,370**
1	3,728	16,609	**20,337**
Total	**24,471**	**32,236**	**56,707**

Sensitivity: (16,609/20,337)*100 = **81.67%**

Specificity: (20,743/36,370)*100 = **57.03%**

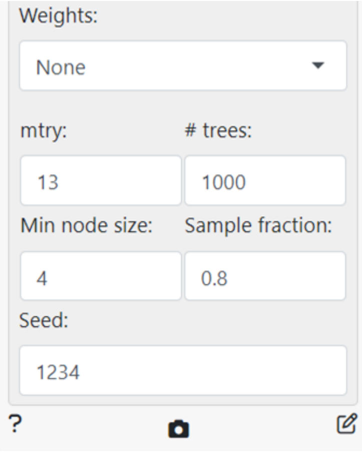
2 deciles have higher than 50% probability of cancelling



Results Analysis: Random Forest

```
Confusion matrix
Data        : RT_bookingfinal
Filter      : Train=="1"
Results for: Both
Predictors : pred_logit, pred_rf
Response    : is_canceled
Level       : 1 in is_canceled
Cost:Margin: 1 : 2
```

Type	Predictor	accuracy	kappa	profit	index	ROME	contact	AUC
Training	pred_logit	0.778	0.462	5,437.000	0.394	0.731	0.187	0.809
Training	pred_rf	0.989	0.976	13,804.000	1.000	0.986	0.353	0.999
Test	pred_logit	0.777	0.459	2,301.000	0.644	0.706	0.192	0.812
Test	pred_rf	0.851	0.668	3,571.000	1.000	0.668	0.314	0.908



Results Analysis: Random Forest

Type	Predictor	accuracy	kappa	profit	index	ROME	contact	AUC
Training	pred_logit	0.778	0.462	5,437.000	0.391	0.731	0.187	0.809
Training	pred_rf	0.989	0.976	13,804.000	0.993	0.986	0.353	0.999
Training	pred_rf2	0.992	0.982	13,906.000	1.000	0.987	0.355	0.999
Test	pred_logit	0.777	0.459	2,301.000	0.644	0.706	0.192	0.812
Test	pred_rf	0.851	0.668	3,571.000	1.000	0.668	0.314	0.908
Test	pred_rf2	0.851	0.666	3,561.000	0.997	0.666	0.314	0.910

```
Confusion matrix
Data        : bookingfinalwithbasepred
Filter      : Train==1
Results for: Both
Predictors  : pred_logit, pred_rf, pred_rf2
Response    : is_canceled
Level       : 1 in is_canceled
Cost:Margin: 1 : 2
```

Type	Predictor	TP	FP	TN	FN	total	TPR	TNR	precision	Fscore
Training	pred_logit	6,437	1,000	24,460	7,798	39,695	0.452	0.961	0.866	0.594
Training	pred_rf	13,904	100	25,360	331	39,695	0.977	0.996	0.993	0.985
Training	pred_rf2	14,001	95	25,365	234	39,695	0.984	0.996	0.993	0.988
Test	pred_logit	2,781	480	10,430	3,321	17,012	0.456	0.956	0.853	0.594
Test	pred_rf	4,457	886	10,024	1,645	17,012	0.730	0.919	0.834	0.779
Test	pred_rf2	4,453	892	10,018	1,649	17,012	0.730	0.918	0.833	0.778

Weights:

None

mtry: 10

trees: 1000

Min node size: 2

Sample fraction: 0.8

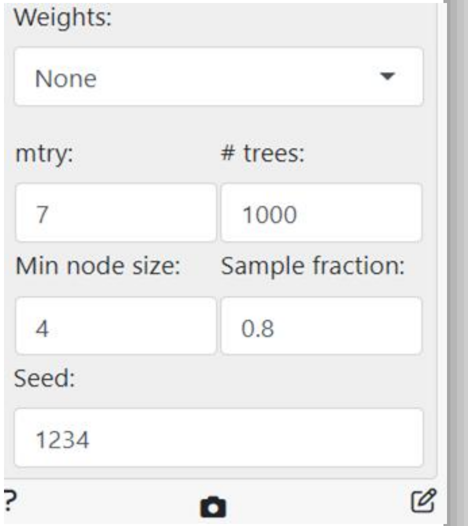
Seed: 1234

?

Results Analysis: Random Forest

	auc (mean)	std	min	max	mtry	min.node.size	num.trees	sample.fraction
1	0.9007128	0.003586407	0.8949845	0.9068864	7	4	500	0.8
2	0.9003412	0.005038050	0.8943738	0.9080599	7	5	500	0.8
3	0.8998140	0.004521882	0.8921382	0.9072942	7	3	500	0.8
4	0.8988100	0.002956099	0.8945491	0.9049659	9	3	500	0.8
5	0.8980661	0.002006282	0.8958051	0.9018199	9	4	500	0.8
6	0.8979671	0.002906688	0.8942659	0.9026324	9	5	500	0.8
7	0.8957692	0.002136930	0.8917583	0.8989563	11	4	500	0.8
8	0.8957631	0.004964185	0.8883107	0.9012223	11	5	500	0.8
9	0.8956985	0.002695795	0.8914092	0.9011373	11	3	500	0.8

Type	Predictor	accuracy	kappa	profit	index	ROME	contact	AUC
Training	pred_logit	0.778	0.462	5,437.000	0.391	0.731	0.187	0.809
Training	pred_rf	0.989	0.976	13,804.000	0.994	0.986	0.353	0.999
Training	pred_rf2	0.991	0.981	13,890.000	1.000	0.987	0.355	0.999
Training	pred_rf_gs	0.986	0.970	13,684.000	0.985	0.982	0.351	0.998
Test	pred_logit	0.777	0.459	2,301.000	0.638	0.706	0.192	0.812
Test	pred_rf	0.851	0.668	3,571.000	0.990	0.668	0.314	0.908
Test	pred_rf2	0.848	0.660	3,509.000	0.973	0.647	0.319	0.907
Test	pred_rf_gs	0.853	0.671	3,606.000	1.000	0.684	0.310	0.910




```
cv.rforest(result, K=5, repeats=2, mtry = seq(7, 11, 2), min.node.size = seq(3, 5, 1), sample.fraction = seq(0.8,0.8,0), num.trees = 500, fun=auc)
```

Results Analysis: Gradient Boosted Trees

```
Confusion matrix
Data        : bookingfinalwithbasepred
Filter      : Train==1
Results for: Both
Predictors : pred_logit, pred_rf, pred_gbt, pred_rf2
Response    : is_canceled
Level       : 1 in is_canceled
Cost:Margin: 1 : 2
```

Type	Predictor	TP	FP	TN	FN	total	TPR	TNR	precision	Fscore
Training	pred_logit	6,437	1,000	24,460	7,798	39,695	0.452	0.961	0.866	0.594
Training	pred_rf	13,904	100	25,360	331	39,695	0.977	0.996	0.993	0.985
Training	pred_gbt	10,076	1,320	24,140	4,159	39,695	0.708	0.948	0.884	0.786
Training	pred_rf2	14,001	95	25,365	234	39,695	0.984	0.996	0.993	0.988
Test	pred_logit	2,781	480	10,430	3,321	17,012	0.456	0.956	0.853	0.594
Test	pred_rf	4,457	886	10,024	1,645	17,012	0.730	0.919	0.834	0.779
Test	pred_gbt	4,095	800	10,110	2,007	17,012	0.671	0.927	0.837	0.745
Test	pred_rf2	4,453	892	10,018	1,649	17,012	0.730	0.918	0.833	0.778

Type	Predictor	accuracy	kappa	profit	index	ROME	contact	AUC
Training	pred_logit	0.778	0.462	5,437.000	0.391	0.731	0.187	0.809
Training	pred_rf	0.989	0.976	13,804.000	0.993	0.986	0.353	0.999
Training	pred_gbt	0.862	0.686	8,756.000	0.630	0.768	0.287	0.931
Training	pred_rf2	0.992	0.982	13,906.000	1.000	0.987	0.355	0.999
Test	pred_logit	0.777	0.459	2,301.000	0.644	0.706	0.192	0.812
Test	pred_rf	0.851	0.668	3,571.000	1.000	0.668	0.314	0.908
Test	pred_gbt	0.835	0.625	3,295.000	0.923	0.673	0.288	0.893
Test	pred_rf2	0.851	0.666	3,561.000	0.997	0.666	0.314	0.910

Weights:
None

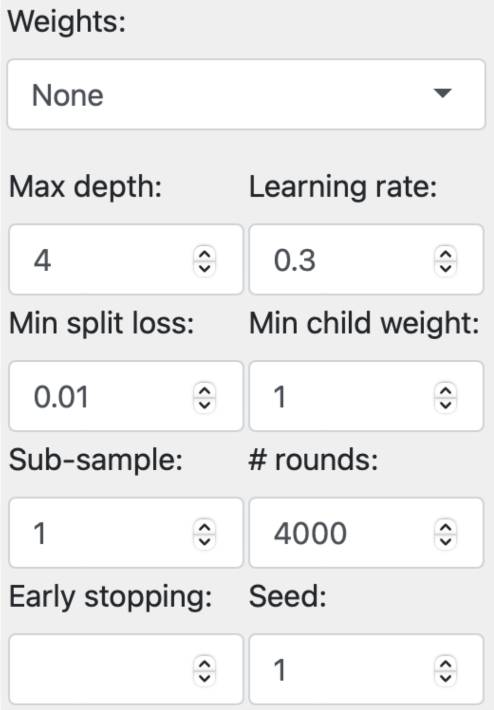
Max depth: 6
Learning rate: 0.3
Min split loss: 0.01
Min child weight: 1
Sub-sample: 1
rounds: 100
Early stopping:
Seed: 1

Results Analysis: Gradient Boosted Trees

```
Confusion matrix
Data        : bookingfinalwithbasepred
Filter      : Train==1
Results for: Both
Predictors : pred_logit, pred_rf, pred_gbt, pred_rf2, pred_gbt2
Response    : is_canceled
Level       : 1 in is_canceled
Cost:Margin: 1 : 2
```

Type	Predictor	TP	FP	TN	FN	total	TPR	TNR	precision	Fscore
Training	pred_logit	6,437	1,000	24,460	7,798	39,695	0.452	0.961	0.866	0.594
Training	pred_rf	13,904	100	25,360	331	39,695	0.977	0.996	0.993	0.985
Training	pred_gbt	10,092	1,290	24,170	4,143	39,695	0.709	0.949	0.887	0.788
Training	pred_rf2	14,001	95	25,365	234	39,695	0.984	0.996	0.993	0.988
Training	pred_gbt2	13,087	437	25,023	1,148	39,695	0.919	0.983	0.968	0.943
Test	pred_logit	2,781	480	10,430	3,321	17,012	0.456	0.956	0.853	0.594
Test	pred_rf	4,457	886	10,024	1,645	17,012	0.730	0.919	0.834	0.779
Test	pred_gbt	4,095	754	10,156	2,007	17,012	0.671	0.931	0.845	0.748
Test	pred_rf2	4,453	892	10,018	1,649	17,012	0.730	0.918	0.833	0.778
Test	pred_gbt2	4,478	1,075	9,835	1,624	17,012	0.734	0.901	0.806	0.768

Type	Predictor	accuracy	kappa	profit	index	ROME	contact	AUC
Training	pred_logit	0.778	0.462	5,437.000	0.391	0.731	0.187	0.809
Training	pred_rf	0.989	0.976	13,804.000	0.993	0.986	0.353	0.999
Training	pred_gbt	0.863	0.689	8,802.000	0.633	0.773	0.287	0.933
Training	pred_rf2	0.992	0.982	13,906.000	1.000	0.987	0.355	0.999
Training	pred_gbt2	0.960	0.912	12,650.000	0.910	0.935	0.341	0.993
Test	pred_logit	0.777	0.459	2,301.000	0.644	0.706	0.192	0.812
Test	pred_rf	0.851	0.668	3,571.000	1.000	0.668	0.314	0.908
Test	pred_gbt	0.838	0.631	3,341.000	0.936	0.689	0.285	0.895
Test	pred_rf2	0.851	0.666	3,561.000	0.997	0.666	0.314	0.910
Test	pred_gbt2	0.841	0.648	3,403.000	0.953	0.613	0.326	0.896



Results Analysis: Gradient Boosted Trees

	nrounds	best_iteration	train_auc_mean	train_auc_std	test_auc_mean	test_auc_std	max_depth	learning_rate	min_split_loss	min_child_
39	1500	1434	0.9581358	0.0004325034	0.8976925	0.004066382	7	0.02	0	
40	1500	1434	0.9581358	0.0004325034	0.8976925	0.004066382	7	0.02	0	
63	2000	1434	0.9581358	0.0004325034	0.8976925	0.004066382	7	0.02	0	
64	2000	1434	0.9581358	0.0004325034	0.8976925	0.004066382	7	0.02	0	
23	1000	423	0.9707211	0.0006120578	0.8967523	0.004466010	7	0.10	0	
24	1000	423	0.9707211	0.0006120578	0.8967523	0.004466010	7	0.10	0	
47	1500	423	0.9707211	0.0006120578	0.8967523	0.004466010	7	0.10	0	
48	1500	423	0.9707211	0.0006120578	0.8967523	0.004466010	7	0.10	0	
71	2000	423	0.9707211	0.0006120578	0.8967523	0.004466010	7	0.10	0	
72	2000	423	0.9707211	0.0006120578	0.8967523	0.004466010	7	0.10	0	
61	2000	1944	0.9524448	0.0007236744	0.8959453	0.003913170	6	0.02	0	
62	2000	1944	0.9524448	0.0007236744	0.8959453	0.003913170	6	0.02	0	
21	1000	504	0.9606094	0.0003227200	0.8955458	0.004226730	6	0.10	0	
22	1000	504	0.9606094	0.0003227200	0.8955458	0.004226730	6	0.10	0	
45	1500	504	0.9606094	0.0003227200	0.8955458	0.004226730	6	0.10	0	
46	1500	504	0.9606094	0.0003227200	0.8955458	0.004226730	6	0.10	0	
69	2000	504	0.9606094	0.0003227200	0.8955458	0.004226730	6	0.10	0	
70	2000	504	0.9606094	0.0003227200	0.8955458	0.004226730	6	0.10	0	
55	2000	2000	0.9445172	0.0005519594	0.8950754	0.003808064	7	0.01	0	
56	2000	2000	0.9445172	0.0005519594	0.8950754	0.003808064	7	0.01	0	
15	1000	1000	0.9443548	0.0005497080	0.8948082	0.003811446	7	0.02	0	
16	1000	1000	0.9443548	0.0005497080	0.8948082	0.003811446	7	0.02	0	
37	1500	1500	0.9428779	0.0006080147	0.8940405	0.003875204	6	0.02	0	

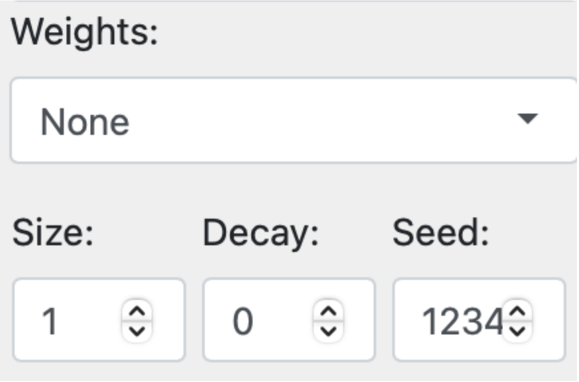
```
cv.gbt(result, K = 5, repeats =2, params = list(max_depth = 4:7, learning_rate = c(0.01, 0.02, 0.1), nrounds=c(1000, 1500, 2000)), fun = auc, maximize =
TRUE)
```

Results Analysis: Neural Network

```
Confusion matrix
Data        : bookingfinalwithbasepred
Filter      : Train==1
Results for: Both
Predictors : pred_logit, pred_rf, pred_rf2, pred_nn1
Response    : is_canceled
Level       : 1 in is_canceled
Cost:Margin: 1 : 2
```

Type	Predictor	TP	FP	TN	FN	total	TPR	TNR	precision	Fscore
Training	pred_logit	6,437	1,000	24,460	7,798	39,695	0.452	0.961	0.866	0.594
Training	pred_rf	13,904	100	25,360	331	39,695	0.977	0.996	0.993	0.985
Training	pred_rf2	14,001	95	25,365	234	39,695	0.984	0.996	0.993	0.988
Training	pred_nn1	7,375	1,472	23,988	6,860	39,695	0.518	0.942	0.834	0.639
Test	pred_logit	2,781	480	10,430	3,321	17,012	0.456	0.956	0.853	0.594
Test	pred_rf	4,457	886	10,024	1,645	17,012	0.730	0.919	0.834	0.779
Test	pred_rf2	4,453	892	10,018	1,649	17,012	0.730	0.918	0.833	0.778
Test	pred_nn1	3,167	677	10,233	2,935	17,012	0.519	0.938	0.824	0.637

Type	Predictor	accuracy	kappa	profit	index	ROME	contact	AUC
Training	pred_logit	0.778	0.462	5,437.000	0.391	0.731	0.187	0.809
Training	pred_rf	0.989	0.976	13,804.000	0.993	0.986	0.353	0.999
Training	pred_rf2	0.992	0.982	13,906.000	1.000	0.987	0.355	0.999
Training	pred_nn1	0.790	0.502	5,903.000	0.424	0.667	0.223	0.810
Test	pred_logit	0.777	0.459	2,301.000	0.644	0.706	0.192	0.812
Test	pred_rf	0.851	0.668	3,571.000	1.000	0.668	0.314	0.908
Test	pred_rf2	0.851	0.666	3,561.000	0.997	0.666	0.314	0.910
Test	pred_nn1	0.788	0.498	2,490.000	0.697	0.648	0.226	0.807



Results Analysis: Neural Network

Type	Predictor	accuracy	kappa	profit	index	ROME	contact	AUC
Training	pred_logit	0.778	0.462	5,437.000	0.391	0.731	0.187	0.809
Training	pred_rf	0.989	0.976	13,804.000	0.993	0.986	0.353	0.999
Training	pred_gbt	0.863	0.689	8,802.000	0.633	0.773	0.287	0.933
Training	pred_rf2	0.992	0.982	13,906.000	1.000	0.987	0.355	0.999
Training	pred_nn1	0.790	0.502	5,903.000	0.424	0.667	0.223	0.810
Training	pred_nn16	0.831	0.613	7,528.000	0.541	0.691	0.275	0.887
Test	pred_logit	0.777	0.459	2,301.000	0.644	0.706	0.192	0.812
Test	pred_rf	0.851	0.668	3,571.000	1.000	0.668	0.314	0.908
Test	pred_gbt	0.838	0.631	3,341.000	0.936	0.689	0.285	0.895
Test	pred_rf2	0.851	0.666	3,561.000	0.997	0.666	0.314	0.910
Test	pred_nn1	0.788	0.498	2,490.000	0.697	0.648	0.226	0.807
Test	pred_nn16	0.823	0.595	3,091.000	0.866	0.654	0.278	0.875

```
Confusion matrix
Data        : bookingfinalwithbasepred
Filter      : Train==1
Results for: Both
Predictors : pred_logit, pred_rf, pred_gbt, pred_rf2, pred_nn1, pred_nn16
Response    : is_canceled
Level       : 1 in is_canceled
Cost:Margin: 1 : 2
```

Type	Predictor	TP	FP	TN	FN	total	TPR	TNR	precision	Fscore
Training	pred_logit	6,437	1,000	24,460	7,798	39,695	0.452	0.961	0.866	0.594
Training	pred_rf	13,904	100	25,360	331	39,695	0.977	0.996	0.993	0.985
Training	pred_gbt	10,092	1,290	24,170	4,143	39,695	0.709	0.949	0.887	0.788
Training	pred_rf2	14,001	95	25,365	234	39,695	0.984	0.996	0.993	0.988
Training	pred_nn1	7,375	1,472	23,988	6,860	39,695	0.518	0.942	0.834	0.639
Training	pred_nn16	9,215	1,687	23,773	5,020	39,695	0.647	0.934	0.845	0.733
Test	pred_logit	2,781	480	10,430	3,321	17,012	0.456	0.956	0.853	0.594
Test	pred_rf	4,457	886	10,024	1,645	17,012	0.730	0.919	0.834	0.779
Test	pred_gbt	4,095	754	10,156	2,007	17,012	0.671	0.931	0.845	0.748
Test	pred_rf2	4,453	892	10,018	1,649	17,012	0.730	0.918	0.833	0.778
Test	pred_nn1	3,167	677	10,233	2,935	17,012	0.519	0.938	0.824	0.637
Test	pred_nn16	3,908	817	10,093	2,194	17,012	0.640	0.925	0.827	0.722

Weights:

None

Size: 16

Decay: 0.1

Seed: 1234

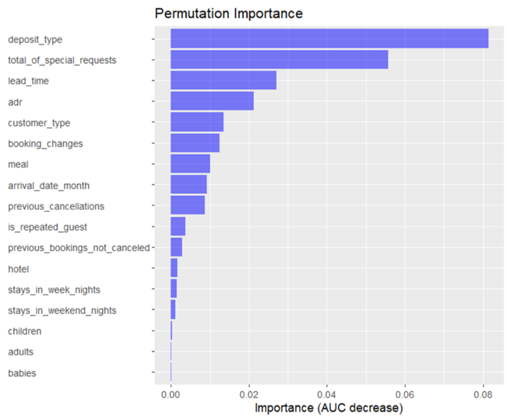
Results Analysis: Neural Network

	auc (mean)	std	min	max	decay	size
1	0.8722634	0.007534552	0.8572860	0.8894517	0.10	16
2	0.8694011	0.007316247	0.8588586	0.8815639	0.01	12
3	0.8687490	0.009087587	0.8367591	0.8763577	0.01	16
4	0.8661605	0.012677093	0.8325696	0.8809241	0.10	12
5	0.8649392	0.007061612	0.8546878	0.8862028	0.10	8
6	0.8627408	0.007330560	0.8491588	0.8716677	0.01	8

Results

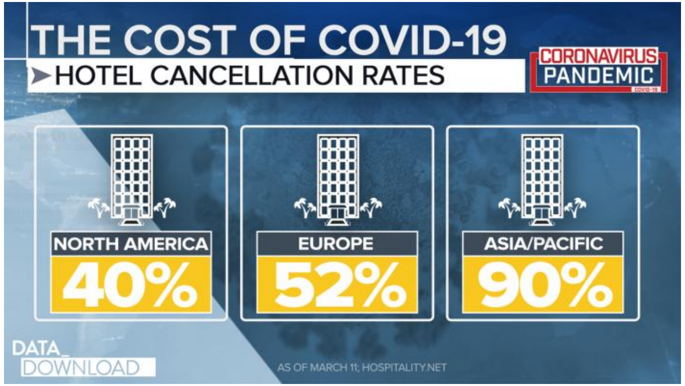
Machine Learning Model Outputs

- Comparative Performance of 3 machine learning models – RF, GBT, and NN
  - RF highest accuracy
  - GBT highest accuracy
  - NN highest accuracy
- Permutation Importance
  - Evaluation of Top Variables
    - Deposit_Type
    - Total_of_special_requests
    - Lead_time



Why are cancellations problematic?

- Rise of cancellations after COVID19 and through Technological Advances.
    - Time Tax, etc.
- Online Travel Agencies (OTAs)
- Revenue per available room is lower when hotel have no shows and cancellations



How can Machine Learning alleviate cancellations?

- **Predict cancellations and timing** using ML models.
- **Adopt dynamic (open) pricing** instead of fixed seasonal pricing.
- **Use trend-based pricing adjustments** to optimize demand.
- **Train staff to identify at-risk guests** based on ML models and offer supportive alternatives.
- **Provide flexible deposit options** to reduce last-minute drop-offs.

Demand Forecasting in Revenue Management

3 Key Executive Actions

- **Deploy a unified ML system** to predict cancellations and guide pricing.
- **Shift to dynamic, data-driven policies** for pricing and deposits.
- **Enable staff with simple playbooks** to act on ML insights and reduce cancellations.

References

References

https://lodgingmagazine.com/forecasting-hotel-reservation-cancellations-with-machine-learning/

https://www.sciencedirect.com/science/article/abs/pii/S0160738316300858?via%3Dihub

https://www.sciencedirect.com/science/article/abs/pii/S0278431920300980

https://www.researchgate.net/publication/333447681_Big_Data_in_Hotel_Revenue_Management_Exploring_Cancellation_Drivers_to_Gain_Insights_Into_Booking_Cancellation_Behavior

https://www.sas.com/en_us/insights/analytics/predictive-analytics.html

https://www.revfine.com/open-pricing/#how-to-implement-hotel-open-pricing

https://www.cvent.com/en/blog/hospitality/open-pricing