A Hotel Bookings Cancellation Prediction Tool

MGT6203 GROUP PROJECT (Team 64)

Alessio Van Keulen, Fei Hui, Rutuja Patil, Yuxin (Sheena) Liu

Introduction

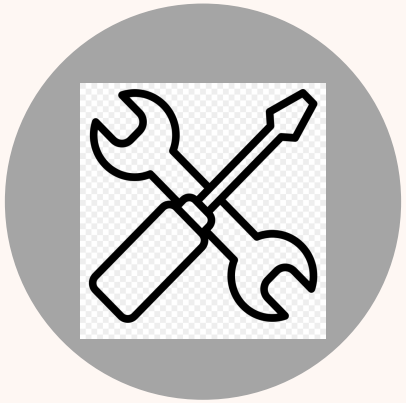
Last minute cancellations of reservations could causes unexpected loss of revenue and difficulty in planning for operations. This study will attempt to understand the underlying forces that drive hotel reservation cancellations.

Dataset

- over 100K hotel reservation records in **Portugal**
- response variable is ”is_canceled” (1 means cancelled, 0 means not)
- 2 types of hotel types: resort hotels vs city hotels
- 31 predictors (both numeric and categorical)

Objectives

REDUCE DIMENSIONALITY IN DATA



CREATE A USEFUL PREDICTION TOOL FOR CANCELLATION

IMPROVE DEMAND FORECAST, BETTER INVENTORY PLANNING AND MANAGEMENT

Dataset

```
'data.frame':   119390 obs. of  35 variables:
 $ hotel                         : Factor w/ 2 levels "City Hotel","Resort Hotel": 2 2 2 2 2 2 2 2 2 2 ...
 $ is_canceled                   : int  0 0 0 0 0 0 0 0 1 1 ...
 $ lead_time                     : int  342 737 7 13 14 14 0 9 85 75 ...
 $ arrival_date_year             : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
 $ arrival_date_month            : Factor w/ 12 levels "April","August",..: 6 6 6 6 6 6 6 6 6 6 ...
 $ arrival_date_week_number      : int  27 27 27 27 27 27 27 27 27 27 ...
 $ arrival_date_day_of_month     : int  1 1 1 1 1 1 1 1 1 1 ...
 $ stays_in_weekend_nights       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ stays_in_week_nights          : int  0 0 1 1 2 2 2 2 3 3 ...
 $ adults                        : int  2 2 1 1 2 2 2 2 2 2 ...
 $ children                      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ babies                        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ meal                          : Factor w/ 5 levels "BB","FB","HB",..: 1 1 1 1 1 1 1 2 1 3 ...
 $ country                       : Factor w/ 8 levels "DEU","ESP","FRA",..: 8 8 4 4 4 4 8 8 8 8 ...
 $ continent                     : Factor w/ 8 levels "Africa","Antarctica",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ market_segment                : Factor w/ 8 levels "Aviation","Complementary",..: 4 4 4 3 7 7 4 4 7 6 ...
 $ distribution_channel          : Factor w/ 5 levels "Corporate","Direct",..: 2 2 2 1 4 4 2 2 4 4 ...
 $ is_repeated_guest             : int  0 0 0 0 0 0 0 0 0 0 ...
 $ previous_cancellations        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ previous_bookings_not_canceled: int  0 0 0 0 0 0 0 0 0 0 ...
 $ reserved_room_type            : Factor w/ 10 levels "A","B","C","D",..: 3 3 1 1 1 1 3 3 1 4 ...
 $ assigned_room_type            : Factor w/ 12 levels "A","B","C","D",..: 3 3 3 1 1 1 3 3 1 4 ...
 $ booking_changes               : int  3 4 0 0 0 0 0 0 0 0 ...
 $ deposit_type                  : Factor w/ 3 levels "No Deposit","Non Refund",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ agent                         : Factor w/ 19 levels "1","14","19",..: 18 18 18 19 6 6 18 19 6 19 ...
 $ company                       : chr  "NULL" "NULL" "NULL" "NULL" ...
 $ days_in_waiting_list          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ customer_type                 : Factor w/ 4 levels "Contract","Group",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ adr                           : num  0 0 75 75 98 ...
 $ required_car_parking_spaces   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ total_of_special_requests     : int  0 0 0 0 1 1 0 1 1 0 ...
 $ reservation_status            : Factor w/ 3 levels "Canceled","Check-Out",..: 2 2 2 2 2 2 2 2 1 1 ...
 $ reservation_status_date       : Factor w/ 926 levels "1/1/2015","1/1/2016",..: 669 669 702 702 735 735 735 735 570 449 ...
 $ ex_rate                       : num  1 1 1 1 1 1 1 1 1 1 ...
 $ temp                          : num  77.8 77.8 77.8 77.8 77.8 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

Approach

I. Data cleaning

- Distribution analysis
- Outlier removal/Null value imputation
- Check for multi-collinearity

II. Fitting logistic model

- Variable selection
- Stepwise regression analysis

III. Conclusion

- Interpret final model
- Validate on test set

Data cleaning

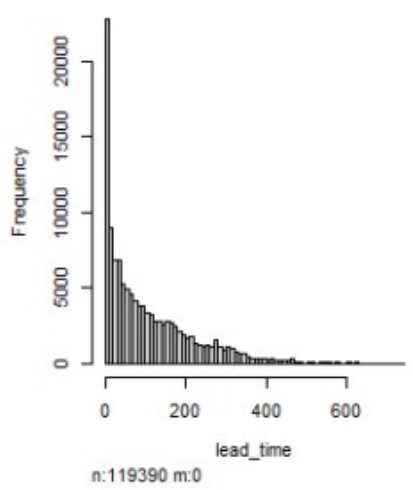
Distribution analysis

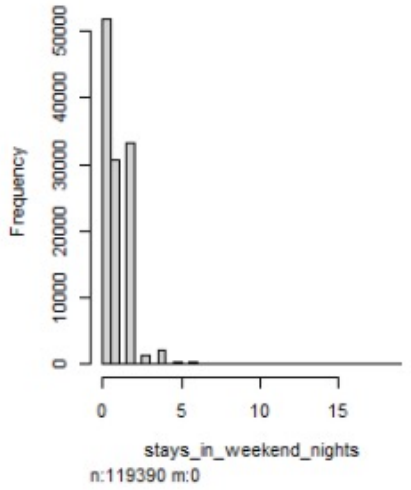
Outlier removal

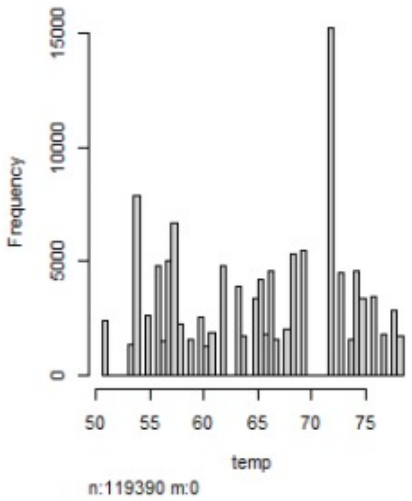
Null value imputation

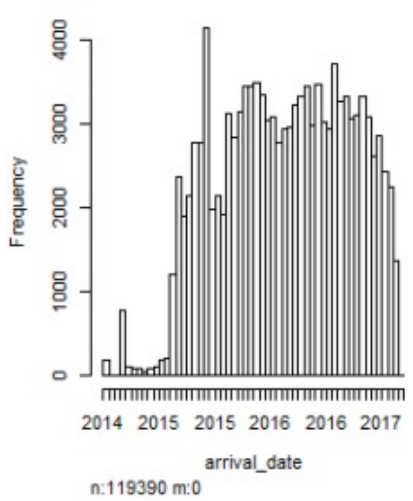
Check for multi-collinearity

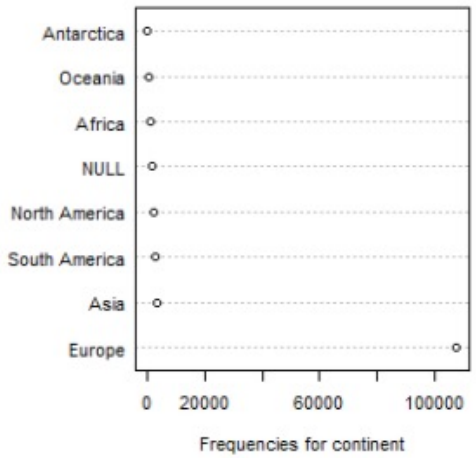
Visualizing data

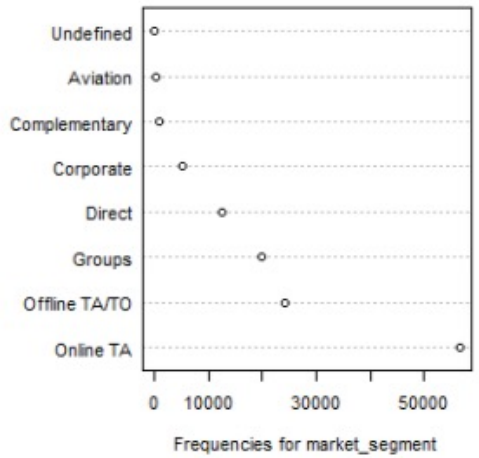


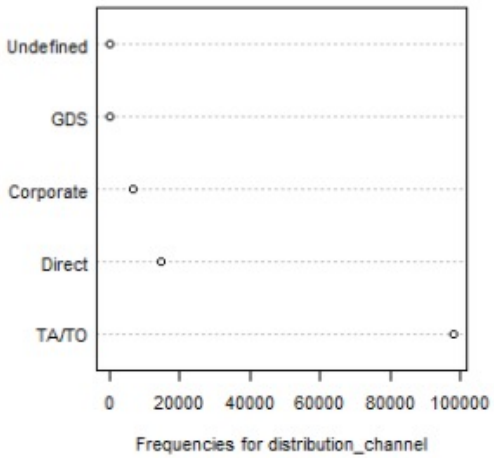




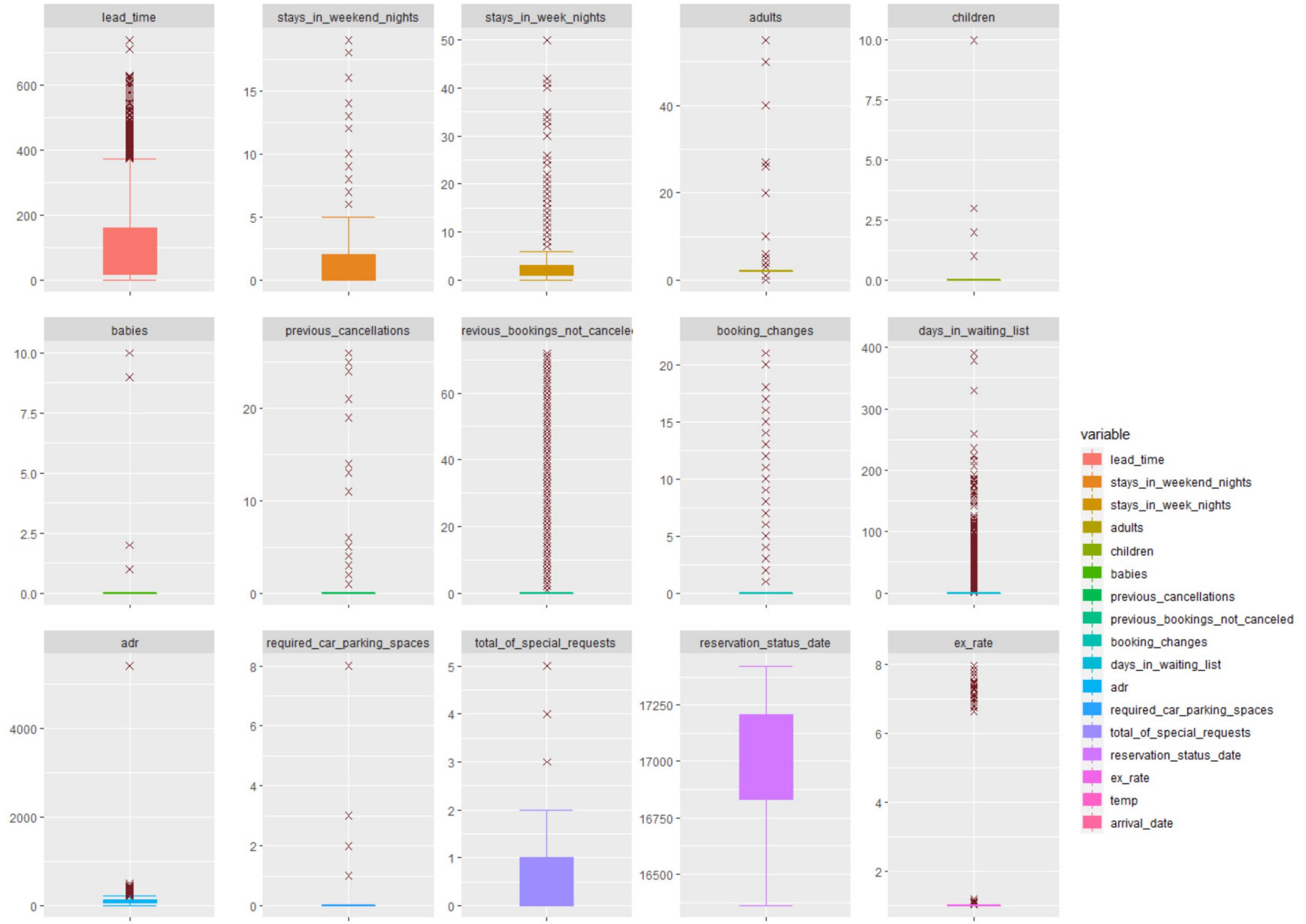






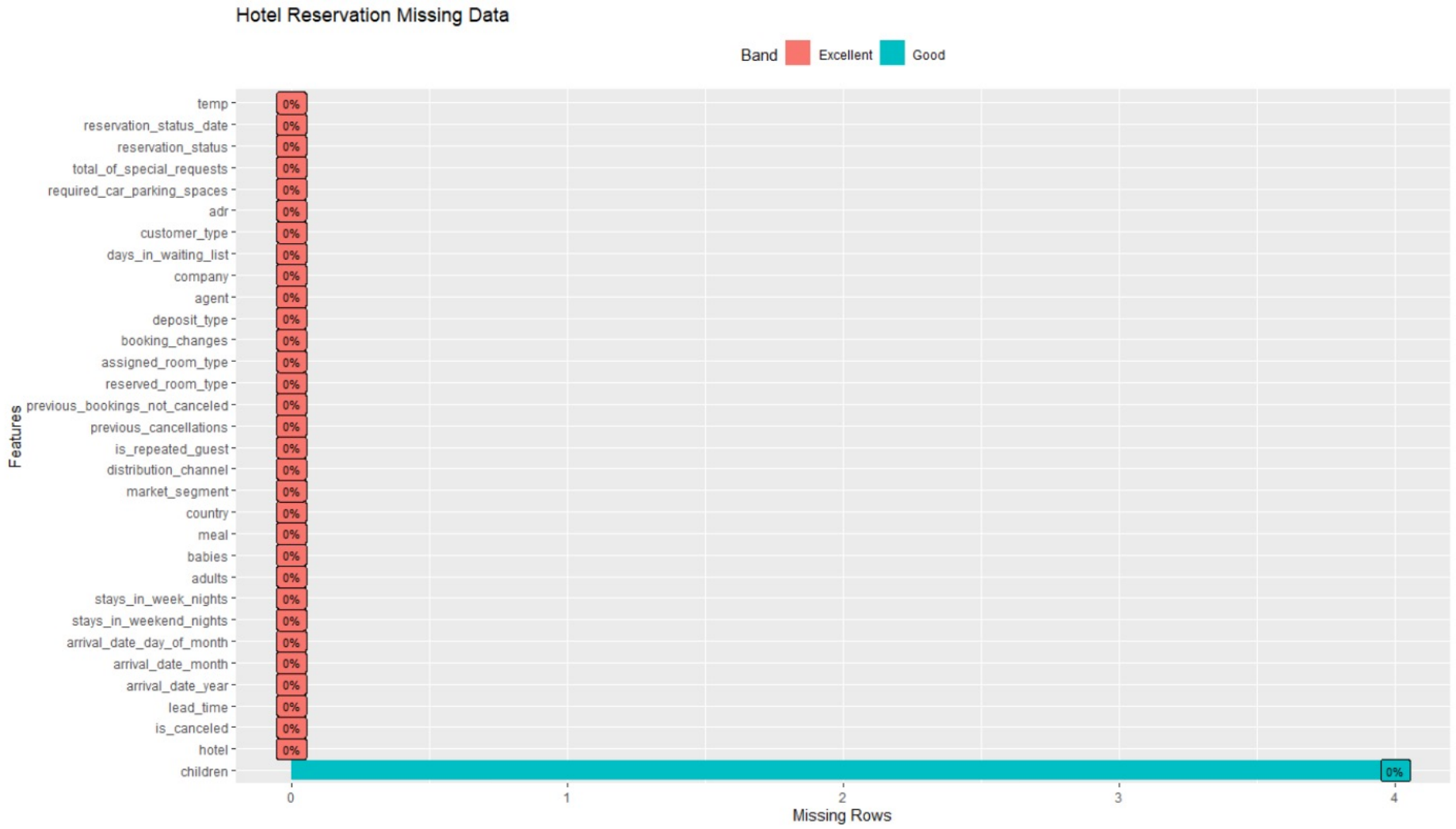


Outlier Detection

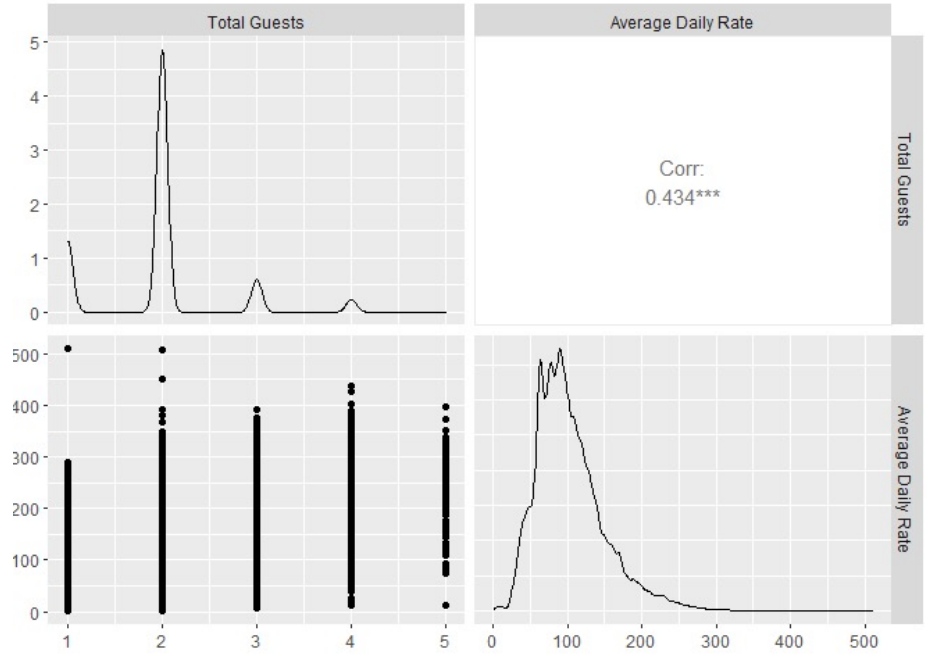


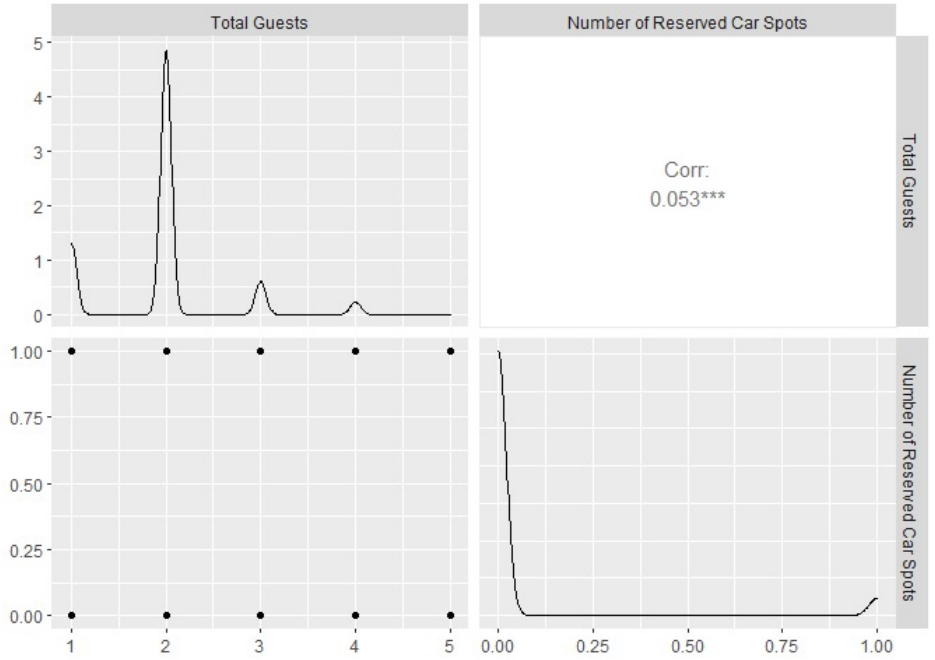
Variable	Range of Values within percentiles 2.5%-97.5%	% of Outliers Detected	Range of Values within percentiles 1%-99%	% of Outliers Detected
lead_time	0 - 374	2.50%	0 - 444	0.99%
stays_in_weekend_nights	0 - 3	1.84%	0 - 4	0.29%
stays_in_week_nights	0 - 7	1.95%	0 - 10	0.34%
adults	1 - 3	0.40%	1 - 3	0.40%
children	0 - 2	0.06%	0 - 2	0.06%
babies	0 - 0	0.77%	0 - 0	0.77%
previous_cancellations	0 - 1	0.36%	0 - 1	0.36%
previous_bookings_not_canceled	0 - 1	1.74%	0 - 3	0.98%
booking_changes	0 - 2	1.32%	0 - 3	0.54%
days_in_waiting_list	0 - 31	2.50%	0 - 75	0.99%
adr	29 - 221.38275	4.94%	0 - 252	0.98%
required_car_parking_spaces	0 - 1	0.03%	0 - 1	0.03%
total_of_special_requests	0 - 2	2.41%	0 - 3	0.32%

Data imputation



Check for Collinearity





Fitting logistic model

Variable selection

Stepwise regression analysis

Further level reduction

Manual variable selection

Remove not useful columns

- Company (94% null, not useful)
- Reservation status date (not running time series models)

Remove columns has repeated information

- Reservation status (repeated from is_cancelled)
- Distribution channel (repeated from market_segment)
- Arrival year/week/day (seasonality can be interpreted by months more)
- Continent (similar as country)
- Previous booking (only keep previous canceled booking)

Combined variables into a more representable variable

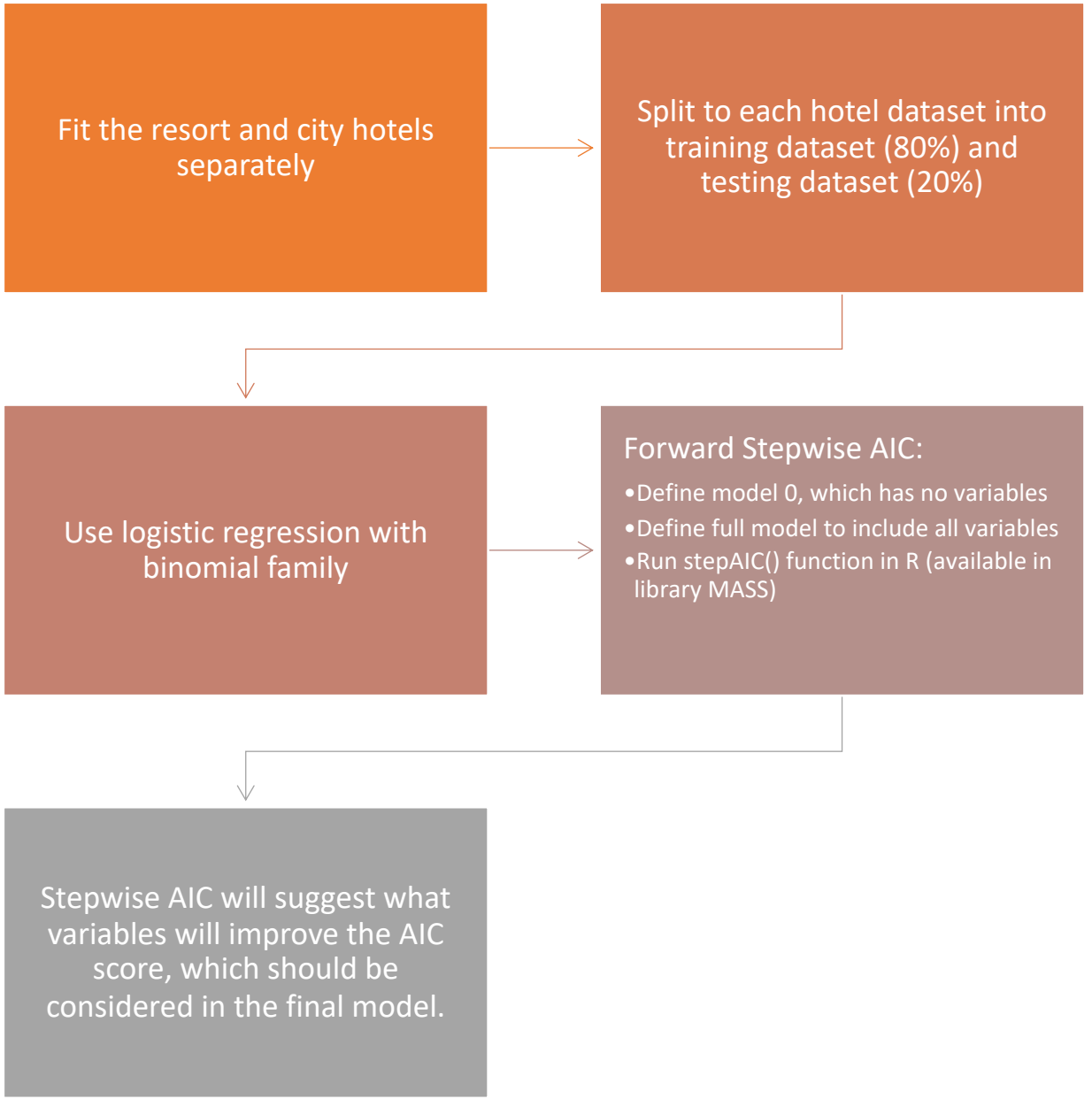
- Total children = # children + # baby
- Total stayed = #weekend stayed + #weekday stayed
- Different room type = if ”assigned room type” is different from “reserved room type”

Reduce levels of categorical variables further more

- Agent group: 9/240/null/others
- meal type: FB/BB/HB (replace undefined as mode)
- booking changes: 0/1+
- Special request: 0/1+
- total children: 0/1+

	manual variable selection
$ hotel	hotel type, done separately
$ is_canceled	response variable
$ lead_time	
$ arrival_date_year	remove
$ arrival_date_month	
$ arrival_date_week_number	remove
$ arrival_date_day_of_month	remove
$ stays_in_weekend_nights	***$ total stayed***
$ stays_in_week_nights	
$ adults	
$ children	***$ total chilren: 0/1+ (2 levels)***
$ babies	
$ meal	***$ meal_grp: FB/BB/HB (3 levels)***
$ country	
$ continent	remove
$ market_segment	
$ distribution_channel	remove
$ is_repeated_guest	
$ previous_cancellations	
$ previous_bookings_not_canceled	remove
$ reserved_room_type	***$ mis.match_room_type***
$ assigned_room_type	
$ booking_changes	***$ booking change grp: 0/1+ (2 levels)***
$ deposit_type	
$ agent	***19 levels >>> 4 levels: $agent grp: 9/240/null/other***
$ company	remove
$ days_in_waiting_list	
$ customer_type	
$ adr	
$ required_car_parking_spaces	
$ total_of_special_requests	***$ special_requests: 0/1+ 2 levels***
$ reservation_status	remove
$ reservation_status_date	remove
$ temp	

Stepwise AIC to select variables



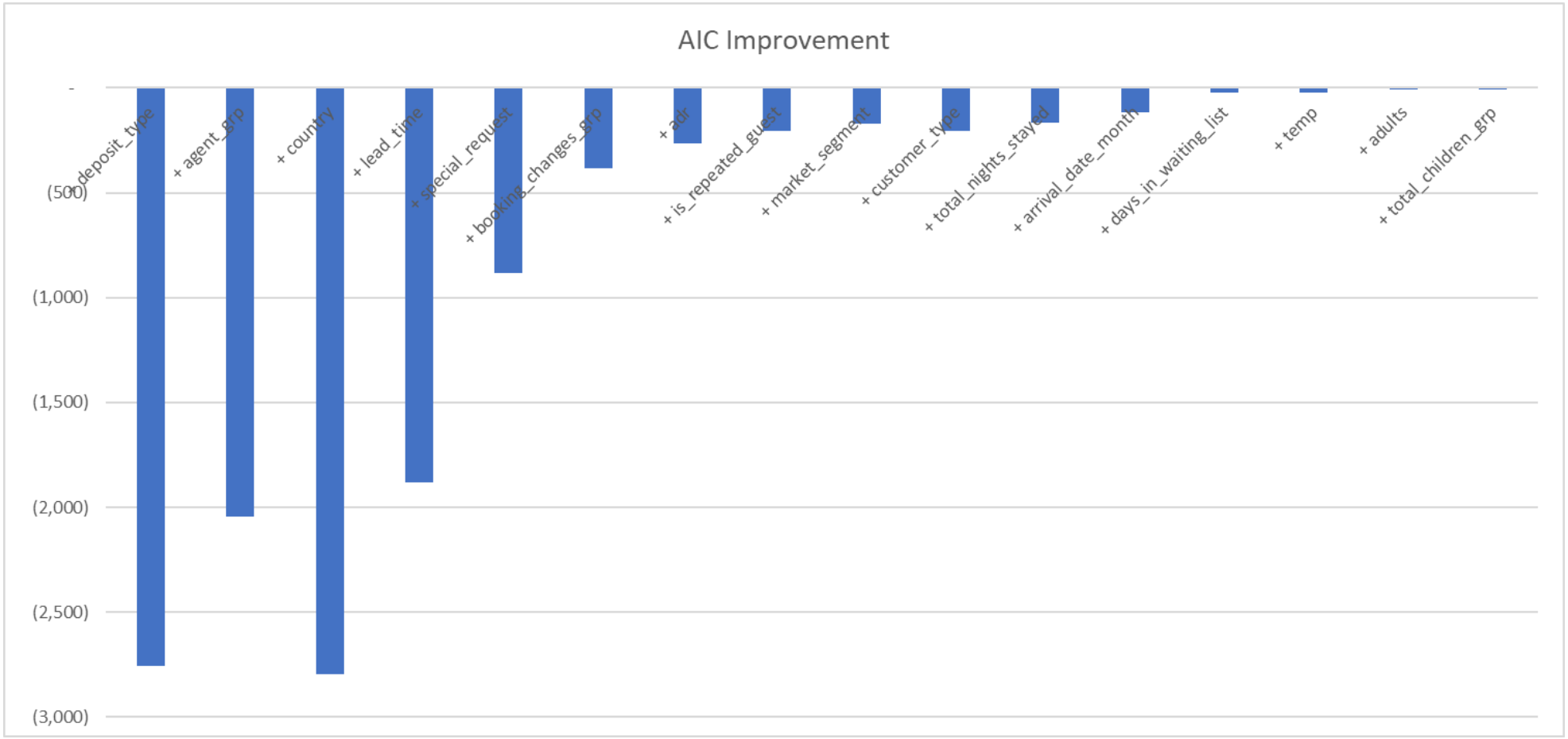
Model Suggested by Stepwise AIC: Resort

ANOVA Tables

Var eliminated

Resort	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1				29831	35455.21	35457.21
2	+ deposit_type	2	2761.159398	29829	32694.05	32700.05
3	+ agent_grp	3	2048.197516	29826	30645.85	30657.85
4	+ country	7	2809.26871	29819	27836.58	27862.58
5	+ lead_time	1	1885.429657	29818	25951.15	25979.15
6	+ special_request	1	883.631338	29817	25067.52	25097.52
7	+ booking_changes_grp	1	384.211538	29816	24683.31	24715.31
8	+ adr	1	266.514166	29815	24416.8	24450.8
9	+ is_repeated_guest	1	207.913339	29814	24208.88	24244.88
10	+ market_segment	5	181.900303	29809	24026.98	24072.98
11	+ customer_type	3	210.63013	29806	23816.35	23868.35
12	+ total_nights_stayed	1	167.922077	29805	23648.43	23702.43
13	+ arrival_date_month	11	140.192227	29794	23508.24	23584.24
14	+ days_in_waiting_list	1	23.184657	29793	23485.05	23563.05
15	+ temp	1	22.371295	29792	23462.68	23542.68
16	+ adults	1	7.831092	29791	23454.85	23536.85
17	+ total_children_grp	1	7.536324	29790	23447.31	23531.31

Resort hotels StepAIC()



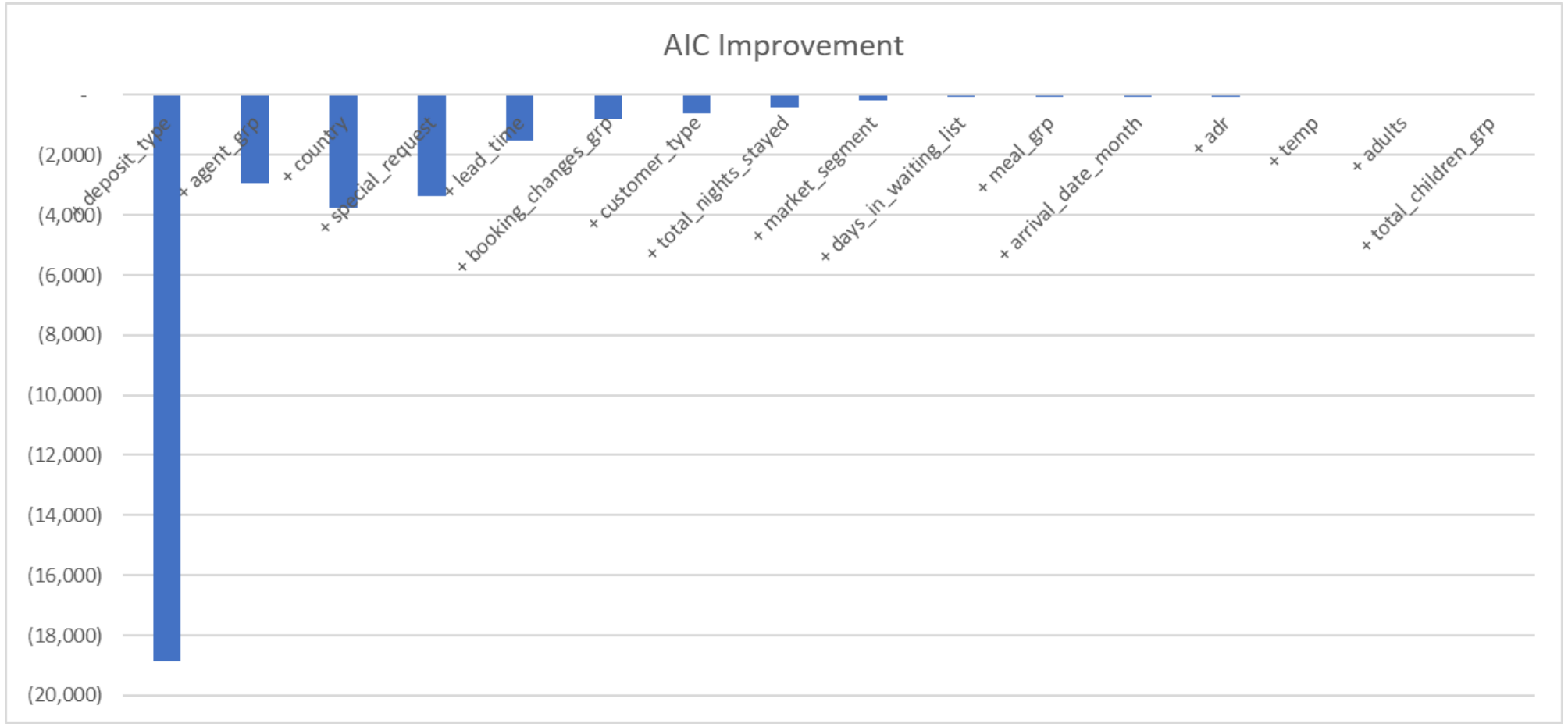
Model Suggested by Stepwise AIC: City Hotels

ANOVA Tables

Var eliminated

City	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1				59017	80414.41	80416.41
2	+ deposit_type	2	18885.45	59015	61528.95	61534.95
3	+ agent_grp	3	2937.809	59012	58591.14	58603.14
4	+ country	7	3760.059	59005	54831.08	54857.08
5	+ special_request	1	3353.252	59004	51477.83	51505.83
6	+ lead_time	1	1531.173	59003	49946.66	49976.66
7	+ booking_changes_grp	1	806.6464	59002	49140.01	49172.01
8	+ customer_type	3	612.5977	58999	48527.42	48565.42
9	+ total_nights_stayed	1	402.258	58998	48125.16	48165.16
10	+ market_segment	7	211.3089	58991	47913.85	47967.85
11	+ days_in_waiting_list	1	64.63683	58990	47849.21	47905.21
12	+ meal_grp	2	49.06928	58988	47800.14	47860.14
13	+ arrival_date_month	11	65.81359	58977	47734.33	47816.33
14	+ adr	1	65.58046	58976	47668.75	47752.75
15	+ temp	1	6.254936	58975	47662.49	47748.49
16	+ adults	1	6.302225	58974	47656.19	47744.19
17	+ total_children_grp	1	3.436558	58973	47652.75	47742.75

City hotels StepAIC()



	manual variable selection	Stepwise on Resort hotels	Stepwise on City hotels
$ hotel	hotel type, done separately		
$ is_canceled	response variable		
$ lead_time		Used	Used
$ arrival_date_year	remove		
$ arrival_date_month		Used	Used
$ arrival_date_week_number	remove		
$ arrival_date_day_of_month	remove		
$ stays_in_weekend_nights	*$ total stayed*	Used	Used
$ stays_in_week_nights			
$ adults		Used	Used
$ children	*$ total chilren: 0/1+ (2 levels)*	Used	Used
$ babies			
$ meal	*$ meal_grp: FB/BB/HB (3 levels)*	eliminated*	Used
$ country		Used	Used
$ continent	remove		
$ market_segment		Used	Used
$ distribution_channel	remove		
$ is_repeated_guest		Used	eliminated*
$ previous_cancellations		eliminated*	eliminated*
$ previous_bookings_not_canceled	remove		
$ reserved_room_type	*$ mis.match_room_type*	eliminated*	eliminated*
$ assigned_room_type			
$ booking_changes	*$ booking change grp: 0/1+ (2 levels)*	Used	Used
$ deposit_type		Used	Used
$ agent	*19 levels >>> 4 levels: $agent grp: 9/240/null/other*	Used	Used
$ company	remove		
$ days_in_waiting_list		Used	Used
$ customer_type		Used	Used
$ adr		Used	Used
$ required_car_parking_spaces		eliminated*	eliminated*
$ total_of_special_requests	*$ special_requests: 0/1+2 levels*	Used	Used
$ reservation_status	remove		
$ reservation_status_date	remove		
$ temp		Used	Used

Further level reduction

```
Coefficients:
                              Estimate Std. Error z value             Pr(>|z|)
(Intercept)                 -4.8011690  1.0385878  -4.623  0.00000378620818313 ***
deposit_typeNon Refund       3.3025989  0.1650396  20.011 < 0.0000000000000002 ***
deposit_typeRefundable      -0.1199955  0.3237944  -0.371              0.710942
agent_grp9                   0.0076538  0.9454002   0.008              0.993541
agent_grpNULL               -1.9798477  0.0876455 -22.589 < 0.0000000000000002 ***
agent_grpOther              -1.3449621  0.0613124 -21.936 < 0.0000000000000002 ***
countryDEU                  -2.6801040  0.1167405 -22.958 < 0.0000000000000002 ***
countryESP                  -1.6014115  0.0593388 -26.988 < 0.0000000000000002 ***
countryFRA                  -2.2244589  0.0993720 -22.385 < 0.0000000000000002 ***
countryGBR                  -2.4415511  0.0597287 -40.877 < 0.0000000000000002 ***
countryIRL                  -2.2294717  0.0796465 -27.992 < 0.0000000000000002 ***
countryITA                  -1.8930103  0.1615868 -11.715 < 0.0000000000000002 ***
countryOTHER                -1.9573611  0.0524918 -37.289 < 0.0000000000000002 ***
lead_time                    0.0073657  0.0002374  31.032 < 0.0000000000000002 ***
special_request1+           -1.0737781  0.0384246 -27.945 < 0.0000000000000002 ***
booking_changes_grp1        -0.9920770  0.0490102 -20.242 < 0.0000000000000002 ***
adr                          0.0045214  0.0005071   8.916 < 0.0000000000000002 ***
is_repeated_guest1          -1.5661675  0.1314186 -11.917 < 0.0000000000000002 ***
market_segmentCorporate      0.6255093  0.7779812   0.804              0.421388
market_segmentDirect        -0.0614948  0.7762906  -0.079              0.936861
market_segmentGroups         0.8287905  0.7791400   1.064              0.287453
market_segmentOffline TA/TO -0.3312661  0.7771569  -0.426              0.669923
market_segmentOnline TA      0.3387315  0.7777855   0.436              0.663194
customer_typeGroup          -0.0372501  0.3706094  -0.101              0.919939
customer_typeTransient       0.9655562  0.1220796   7.909  0.00000000000000259 ***
customer_typeTransient-Party 0.2325898  0.1343043   1.732              0.083307 .
total_nights_stayed          0.0949681  0.0069239  13.716 < 0.0000000000000002 ***
arrival_date_month2          0.3754747  0.1025037   3.663              0.000249 ***
arrival_date_month3          0.0572624  0.1048883   0.546              0.585109
arrival_date_month4         -0.0590401  0.1308486  -0.451              0.651839
arrival_date_month5         -0.1708409  0.1695688  -1.008              0.313694
arrival_date_month6         -0.8517771  0.2488089  -3.423              0.000618 ***
arrival_date_month7         -1.3853840  0.2745245  -5.046  0.00000045001052878 ***
arrival_date_month8         -1.3935849  0.2734408  -5.096  0.00000034603147185 ***
arrival_date_month9         -0.9629611  0.2206762  -4.364  0.00001278908363528 ***
arrival_date_month10        -0.3703699  0.1754059  -2.112              0.034729 *
arrival_date_month11        -0.1068650  0.1262212  -0.847              0.397191
arrival_date_month12         0.0892845  0.1129095   0.791              0.429083
days_in_waiting_list        -0.0430784  0.0140899  -3.057              0.002233 **
temp                         0.0592842  0.0120952   4.901  0.00000095116289836 ***
adults                       0.1231841  0.0448424   2.747              0.006013 **
total_children_grp1          0.1576512  0.0572741   2.753              0.005913 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Can be grouped with baseline.

Too many dimensions, use 95% CI to reduce levels.

Not significant can be removed.

Levels of some variables are grouped further based on the p value significance.

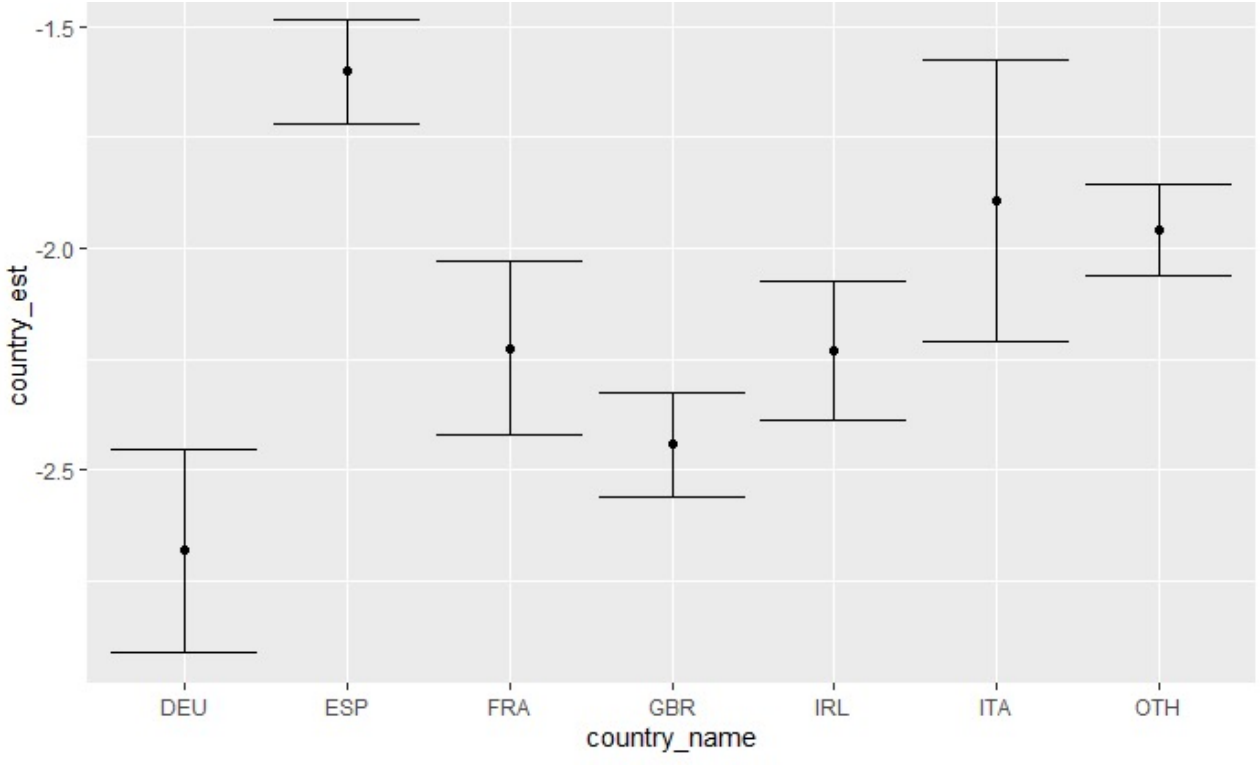
Certain variables are not statically significant (p value>0.5), and these variable will not be included in the final model (e.g., market segment for the resort dataset)

Another way to reduce dimension is to look at the 95% Confidence Interval of all the estimated coefficient. If they overlap with each other, then they can be grouped together.

95% Confidence Interval

```
Coefficients:
                Estimate Std. Error z value             Pr(>|z|)
countryDEU    -2.6801040  0.1167405 -22.958 < 0.0000000000000002 ***
countryESP    -1.6014115  0.0593388 -26.988 < 0.0000000000000002 ***
countryFRA    -2.2244589  0.0993720 -22.385 < 0.0000000000000002 ***
countryGBR    -2.4415511  0.0597287 -40.877 < 0.0000000000000002 ***
countryIRL    -2.2294717  0.0796465 -27.992 < 0.0000000000000002 ***
countryITA    -1.8930103  0.1615868 -11.715 < 0.0000000000000002 ***
countryOTHER  -1.9573611  0.0524918 -37.289 < 0.0000000000000002 ***
```

	country_name <chr>	country_est <dbl>	lower_bound <dbl>	upper_bound <dbl>
countryDEU	DEU	-2.680104	-2.908915	-2.451293
countryESP	ESP	-1.601412	-1.717716	-1.485107
countryFRA	FRA	-2.224459	-2.419228	-2.029690
countryGBR	GBR	-2.441551	-2.558619	-2.324483
countryIRL	IRL	-2.229472	-2.385579	-2.073365
countryITA	ITA	-1.893010	-2.209721	-1.576300
countryOTHER	OTH	-1.957361	-2.060245	-1.854477



The preliminary model suggest all levels of 'country' variable are significant.

We constructed 95% CI graph to see if any of them overlap with each other.

If there are overlaps, that means the estimated coefficients are not statically different from each other. Thus, they can be grouped to reduce dimension.

The resort hotel sample on the left suggested FRA and IRL can be grouped; ITA and OTH can be grouped.

Final model

Resort hotel model

	Estimate	Std. Error	z value	Pr(>|z|)		Coeff
(Intercept)	0.586241	0.18612	3.15	0.001634	**	
deposit_type_grpOth	-3.82656	0.152949	-25.019	2E-16	***	0.021784
agent_grp2NULL	-2.23388	0.057764	-38.673	2E-16	***	0.107112
agent_grp2Other	-1.83115	0.043076	-42.509	2E-16	***	0.16023
country_grpPRT	2.192708	0.038589	56.822	2E-16	***	8.959441
lead_time	0.006565	0.00021	31.321	2E-16	***	1.006587
special_request1+	-1.14614	0.037845	-30.285	2E-16	***	0.317862
booking_changes_grp1	-0.989	0.048956	-20.202	2E-16	***	0.371949
arrival_date_year2016	0.461752	0.048147	9.59	2E-16	***	1.586851
arrival_date_year2017	1.183806	0.05734	20.645	2E-16	***	3.266782
is_repeated_guest1	-1.74638	0.130724	-13.359	2E-16	***	0.174405
arr_momo2_8	0.534245	0.086881	6.149	7.79E-10	***	1.706159
arr_momo9_12	0.985306	0.09364	10.522	2E-16	***	2.678631
total_nights_stayed	0.064826	0.006514	9.952	2E-16	***	1.066974
customer_type_grpTransient	0.485335	0.045371	10.697	2E-16	***	1.624719
adr	0.002651	0.000317	8.361	2E-16	***	1.002655
days_in_waiting_list	-0.03262	0.014018	-2.327	0.019962	*	0.967907
total_children_grp2	0.205908	0.055781	3.691	0.000223	***	1.22864
meal_grpFB	0.609812	0.129035	4.726	2.29046E-06	***	1.840086
meal_grpHB	0.147028	0.043855	3.353	0.000801	***	1.158386

City hotel model

	Estimate	Std. Error	z value	Pr(>|z|)		Coeff
(Intercept)	-2.96264	0.079722	-37.162	2E-16	***	
deposit_typeNon Refund	5.858932	0.279546	20.959	2E-16	***	350.3498407
deposit_typeRefundable	2.448972	0.822651	2.977	0.00291	**	11.57643423
country_grpDEU	-1.9975	0.047131	-42.382	2E-16	***	0.135674045
country_grpESP	-0.99959	0.046946	-21.292	2E-16	***	0.368029309
country_grpFRA	-1.68266	0.040489	-41.558	2E-16	***	0.185879066
country_grpITA	-0.709	0.050141	-14.14	2E-16	***	0.492136235
country_grpOTH	-1.13202	0.029298	-38.638	2E-16	***	0.32238026
special_request1+	-1.04848	0.022868	-45.849	2E-16	***	0.350468377
lead_time	0.005903	0.00014	42.035	2E-16	***	1.005920558
booking_changes_grp1	-1.03798	0.036697	-28.285	2E-16	***	0.354170586
customer_type_grpTransient	1.004393	0.030131	33.334	2E-16	***	2.730249508
total_nights_stayed	0.122782	0.006382	19.238	2E-16	***	1.13063735
mkt_seg_grpOther	1.092412	0.074336	14.696	2E-16	***	2.98145519
days_in_waiting_list	-0.01701	0.001893	-8.984	2E-16	***	0.983134935
meal_grp2HB	0.25296	0.025057	10.095	2E-16	***	1.287831891
arr_mo33	-0.22209	0.02649	-8.384	2E-16	***	0.800841523
arr_mo36	-0.36391	0.032125	-11.328	2E-16	***	0.694951514
arr_mo39	-0.46863	0.04416	-10.612	2E-16	***	0.625861549
adr	0.006804	0.000318	21.411	2E-16	***	1.0068273
adults	0.16849	0.024031	7.011	2.36E-12	***	1.183516273

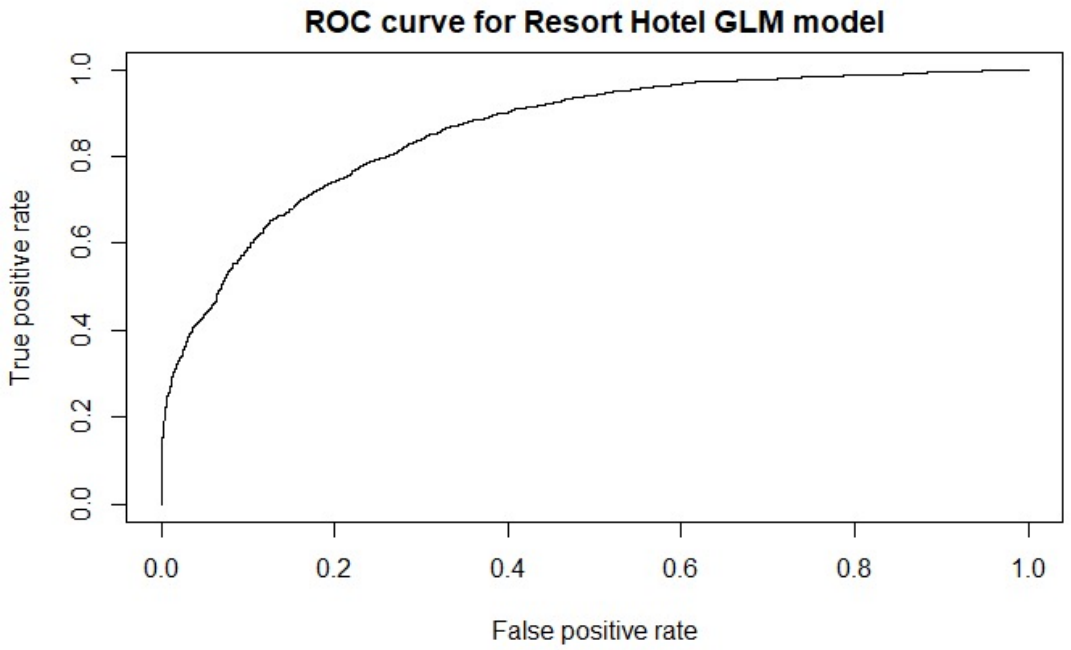
Final Model – Resort Hotel

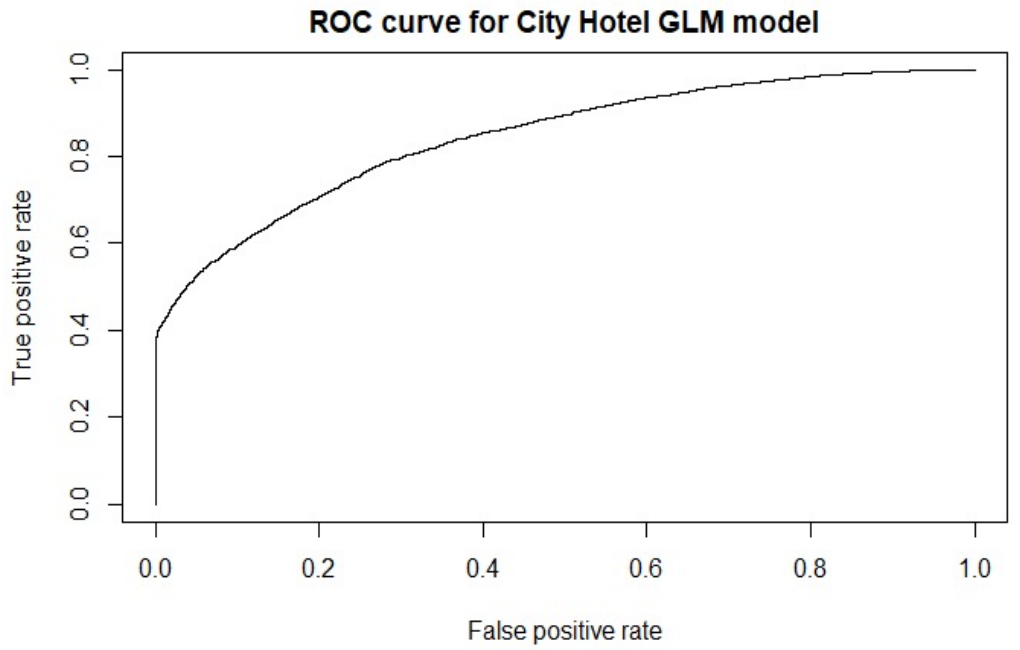
- Deposit Type: “Full deposit” policy actually have higher cancellation probability than “partial deposit” and “No deposit”
- Portuguese visitor has the highest cancellation probability, while German visitor has the lowest.
- One day increase in lead time, cancellation probability increased by 0.8%.
- Customer who has special requests has lower cancellation probability than those who does not.
- Customer who has made booking changes has lower cancellation probability than who didn't.
- Customer who is a repeated guest has lower cancellation probability than first timer.
- Month February to August has higher cancellation rate than January, and Sep to Dec has the highest cancellation probability.
- For one additional night of stay, cancellation prob. increase by 8.2%
- Transient customer has higher cancellation prob.
- $1 increase in adr (average daily rate), cancellation probability increase by 0.5%
- 1 day increase in length on the wait list, the cancellation probability lower by 4%.
- Customer has child/children has higher cancellation probability.
- 1 additional adult cancellation prob. Increased by 10.6%; 1 additional child, cancellation prob.
- Customer has children will have higher cancellation rate.
- 1 Fahrenheit increase in temperature, cancellation rate increase by 4.3%.

Final Model – City Hotel

- Deposit Type: Non Refundable policy actually have higher cancellation probability than “Refundable” and “No Deposit, similar finding as the resort hotel
- Portuguese visitor has the highest cancellation probability, while German visitor has the lowest.
- One day increase in lead time, cancellation probability increased by 0.6%.
- Customer who has special requests has lower cancellation probability than those who does not.
- Customer who has made booking changes has lower cancellation probability than who didn't.
- Winter month (January through February) has higher cancellation rate than the other months. September has the lowest cancellation rate.
- For one additional night of stay, cancellation prob. increase by 12%
- Transient customer has higher cancellation prob, and the magnitude is higher than the resort hotel.
- $1 increase in adr (average daily rate), cancellation probability increase by 0.5%
- 1 day increase in length on the wait list, the cancellation probability lower by 7%.
- 1 additional adult cancellation prob. Increased by 10.6%;
- Variable of whether the customer with children and temperature are not significant in the city hotel model it is excluded.
- Non Contract market segment has higher cancellation rate than Contract market segment.

Validation on test data set



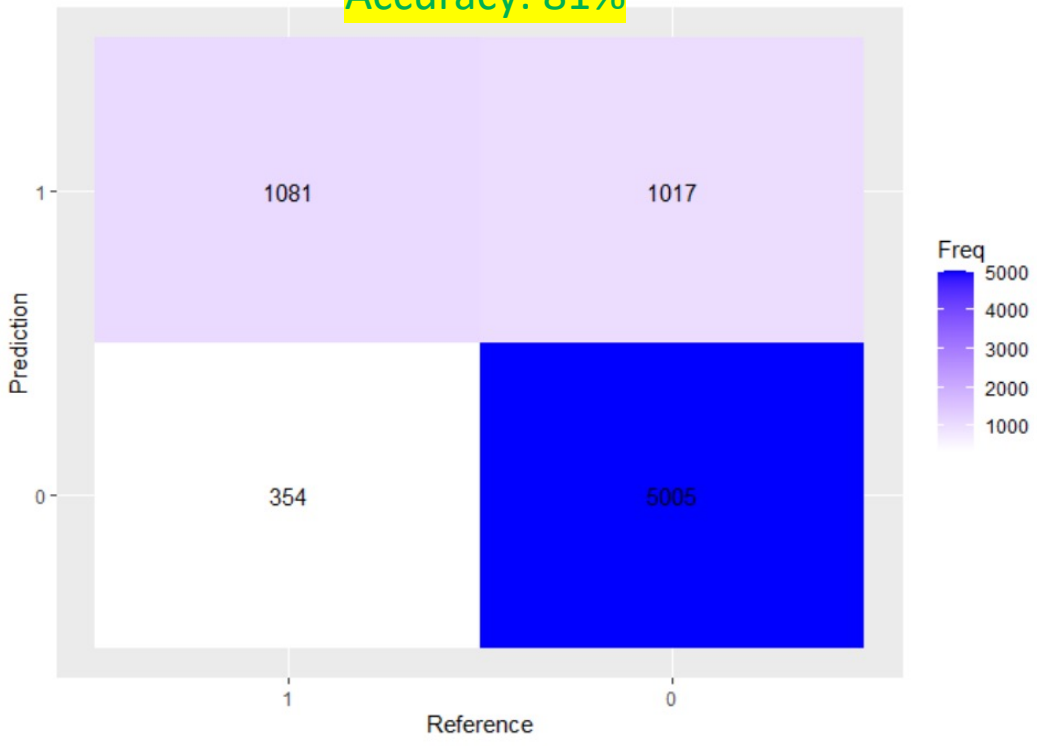


AUC value for Resort hotel GLM is **85.96%**

AUC value for City hotel GLM is **84.52%**

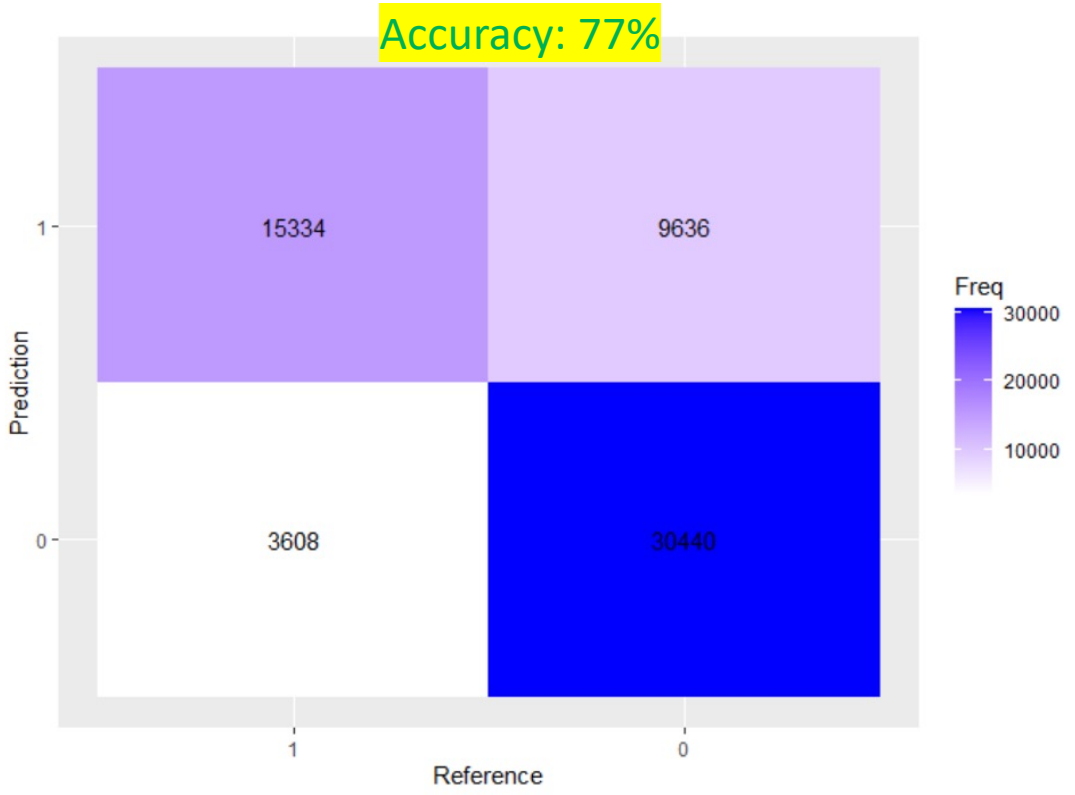
Confusion matrix

Accuracy: 81%



Resort hotels

Accuracy: 77%



City hotels

Future considerations

- Time series analysis
- Cross Validation
- Other modeling
  - Ie. Clustering

Thanks for watching