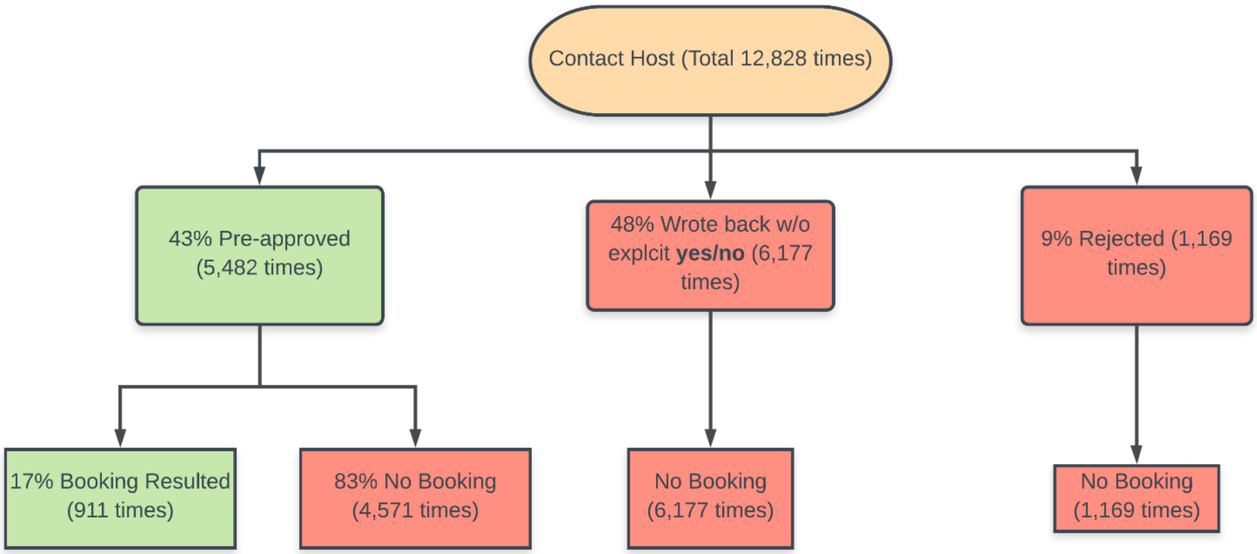
#1 What key metrics would you propose to monitor over time the success of the team's efforts in improving the guest host matching process and why? Clearly define your metric(s) and explain how each is computed.

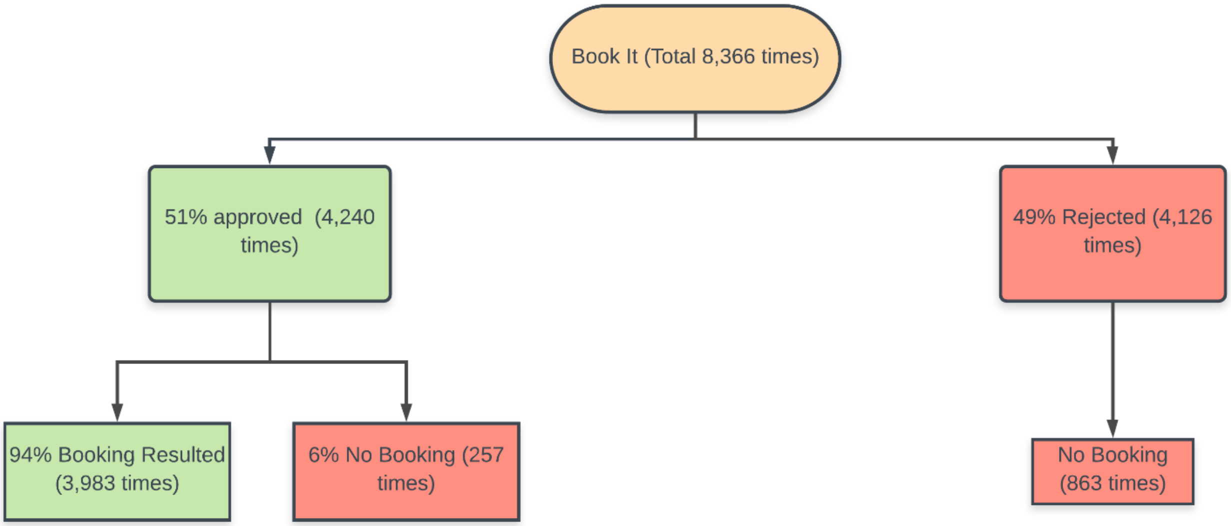
- Booking Funnel metrics split by first contact channel should be tracked at both daily, weekly, and monthly level, as well as DoD, WoW, MoM, YoY % change. Please refer to the 3 funnel diagrams below for all KPIs.



KPIs to track

Monitor them at both daily and weekly level and with WoW & YoY Δ% sliceable by different inquiry's attributes

Total counts of 'Contact Host' = **12828** times
'Contact Host' Overall Booking rate = 911/12828 = **7%**
'Contact Host' Pre-approval rate = 5482/12828 = **43%**
'Contact Host' Wrote-back rate = 5482/12828 = **43%**
'Contact Host' Rejection rate = 5482/12828 = **43%**
'Contact Host' Pre-approved Boooking rate = 911/5482 = **17%**
'Contact Host' Pre-approved No-Boooking rate = 4571/5482 = **83%**



KPIs to track

Monitor them at both daily and weekly level and with WoW & YoY Δ% sliceable by different inquiry's attributes

Total counts of 'Book It' = **8366 times**

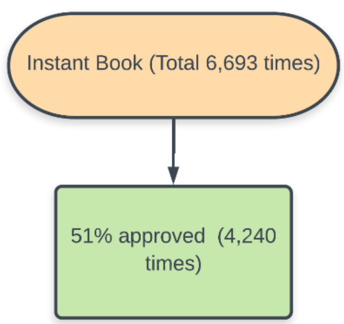
'Book It**'** Overall Booking rate = 3983/8366 = **48%**

'Book It' Approval rate = 4240/8366 = **51%**

'Book It' Rejection rate = 4126/8366 = **49%**

'Book It' Approved Booking rate = 3983/4240 = **94%**

'Book It' Approved No-Boooking rate = 257/4240 = **6%**



KPIs to track

Monitor them at both daily and weekly level and with WoW & YoY Δ% sliceable by different inquiry's attributes

Total counts of 'Instant Book' = **6693 times**

#2 & #3 What areas should we invest in to increase the number of successful bookings in Rio de Janeiro? What segments are doing well and what could be improved? Propose 2-3 specific recommendations (business initiatives and product changes) that could address these opportunities. Demonstrate rationale behind each recommendation AND prioritize your recommendations in order of their estimated impact. There is also interest from executives at Airbnb about the work you are doing, and a desire to understand the broader framing of the challenge of matching supply and demand, thinking beyond the data provided. What other research, experiments, or approaches could help the company get more clarity on the problem?

Recommendations

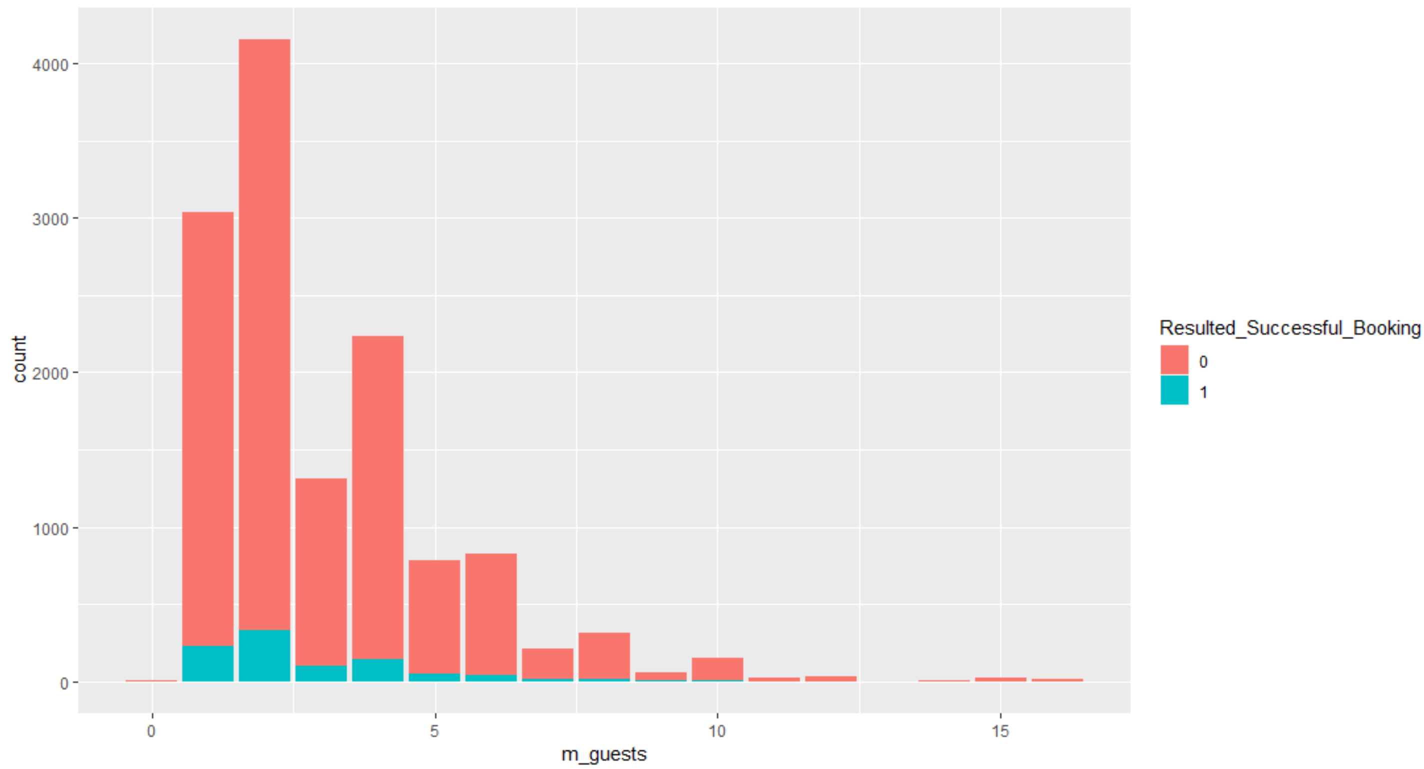
- Even though 'contact me' is a must have feature for guests, this channel has the lowest successful booking rate (7%), which is far lower than that of 'Book It' channel. This unnecessarily imposes messaging time required for both hosts and guests. Rather, think of different programs, features, tutorials, guides to both parties to specify most discussed topics to be explicitly written or easy to understand to minimize the use of this channel if it can be avoided.
    - However, we've noticed that as the number of messages exchanged between hosts and guests, the higher probabilities of the inquiry leading to successful booking, so while maintaining this feature, we might be able to test different things mentioned above, as well as how easily accessible 'contact me' button is located within UI, and try to see if further room for optimization exists.
- Day of checkin, other than Friday and Sunday, other days seem to have much higher successful booking rate. Start some incentive campaigns and try to get guests to book for days other than checkin in on Fridays or Sundays.
    - Similar story with month of checkin, months such as Feb, Aug, and December have the lowest successful booking rate. For the other months, the rates seem vastly different. Since for most cases, un-occupied listing generates $0, run incentive and awareness campaigns to both host and guests to encourage them to leverage other slow months.
        - For example, from the host side, they can specify what entertainment activities or other interesting things to explore on not-so-busy months to attract more guests.
        - From the guest side, Airbnb can do subsidizations for the customers who would be willing to book stays during slow months.
- From the analysis, we've uncovered that how quickly a host replies to the initial contact inquiry is very important to whether an inquiry converts to a successful booking. As the time progresses, much lower chance of it converting.
    - Review the current in-app or email notification services and logics, and try to encourage hosts to reply to guests' inquiries quicker.
- Days from first inquiry to the proposed checkin dates also seem to make quite a big of a difference. The longer ahead the checkin date is, the more likely it will fall through.
    - Embed this into the company's recommendation model
- Past bookers, whom hosts know and think well of, have much higher successful booking rate, from both sides, and it's much quicker conversion turnaround period than from 'contact me' channel. So we can do a few things to take advantage of this:
    - Build features informing to both party about their history with one another, also trying to get social relationships data for the guests, and when we know they have friends on Airbnb as well, then letting the host know that person X is a friend of a guest the host knows and has done business before. This will automatically guarantee a good level of trusts

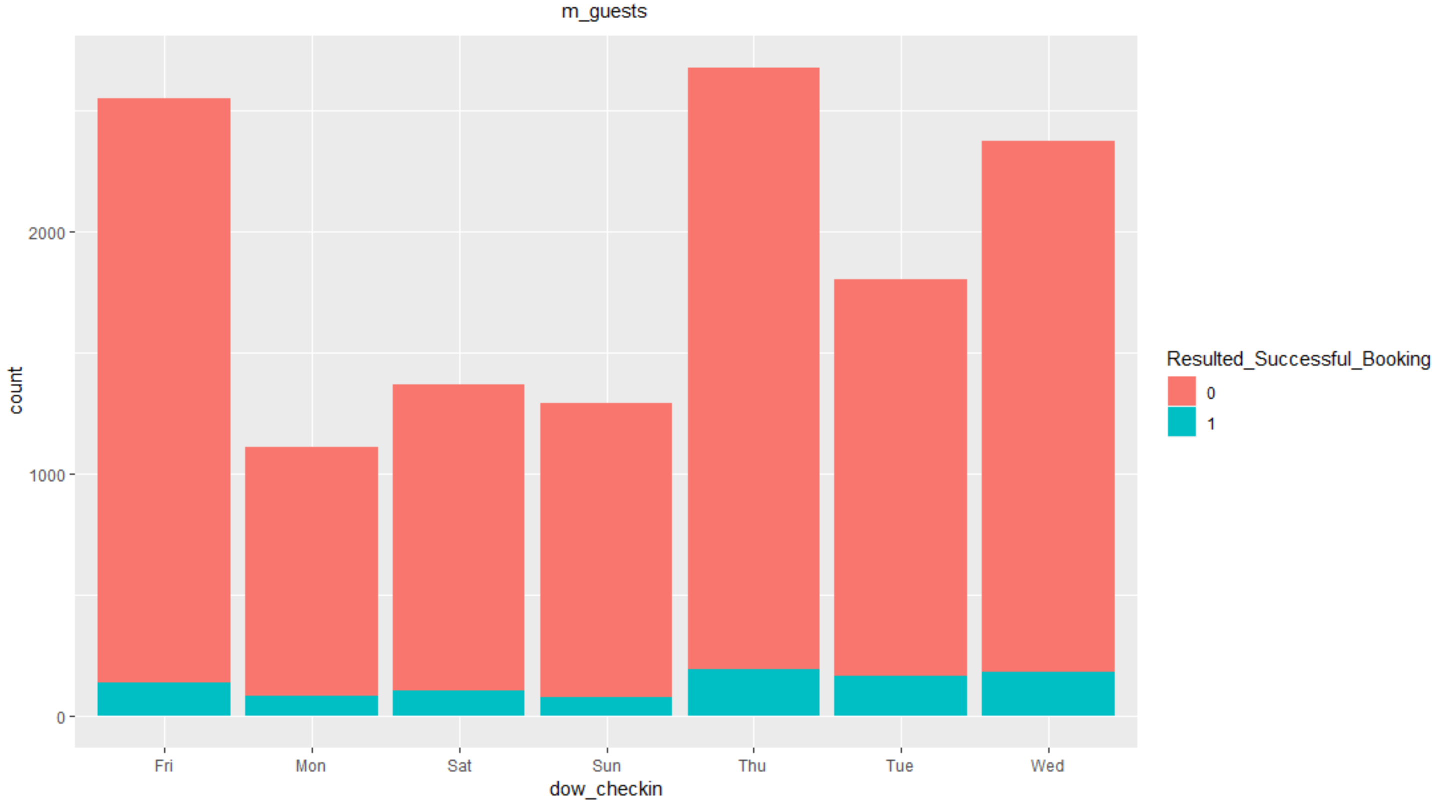
I’ve build a dataset at the individual inquiry level to uncover important metrics related to whether or not a inquiry converts into successful booking.

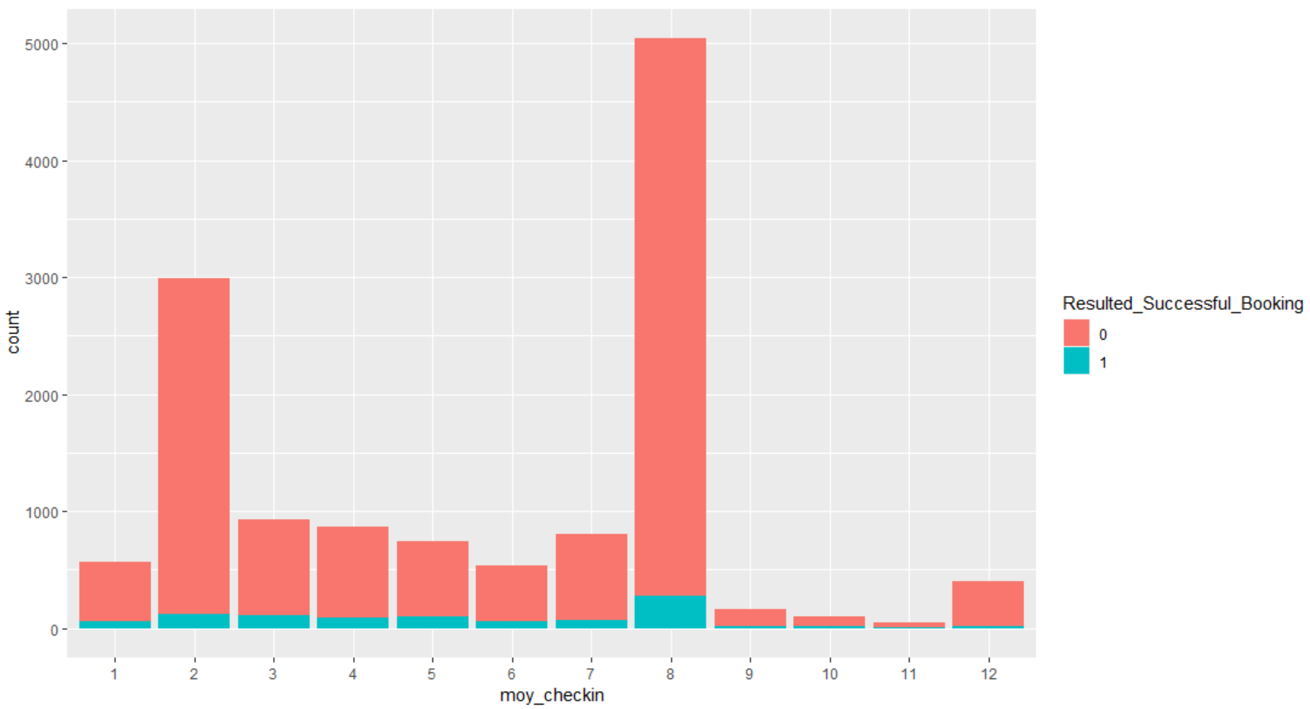
Inquiry (Contact/Booking/Instant) level variables

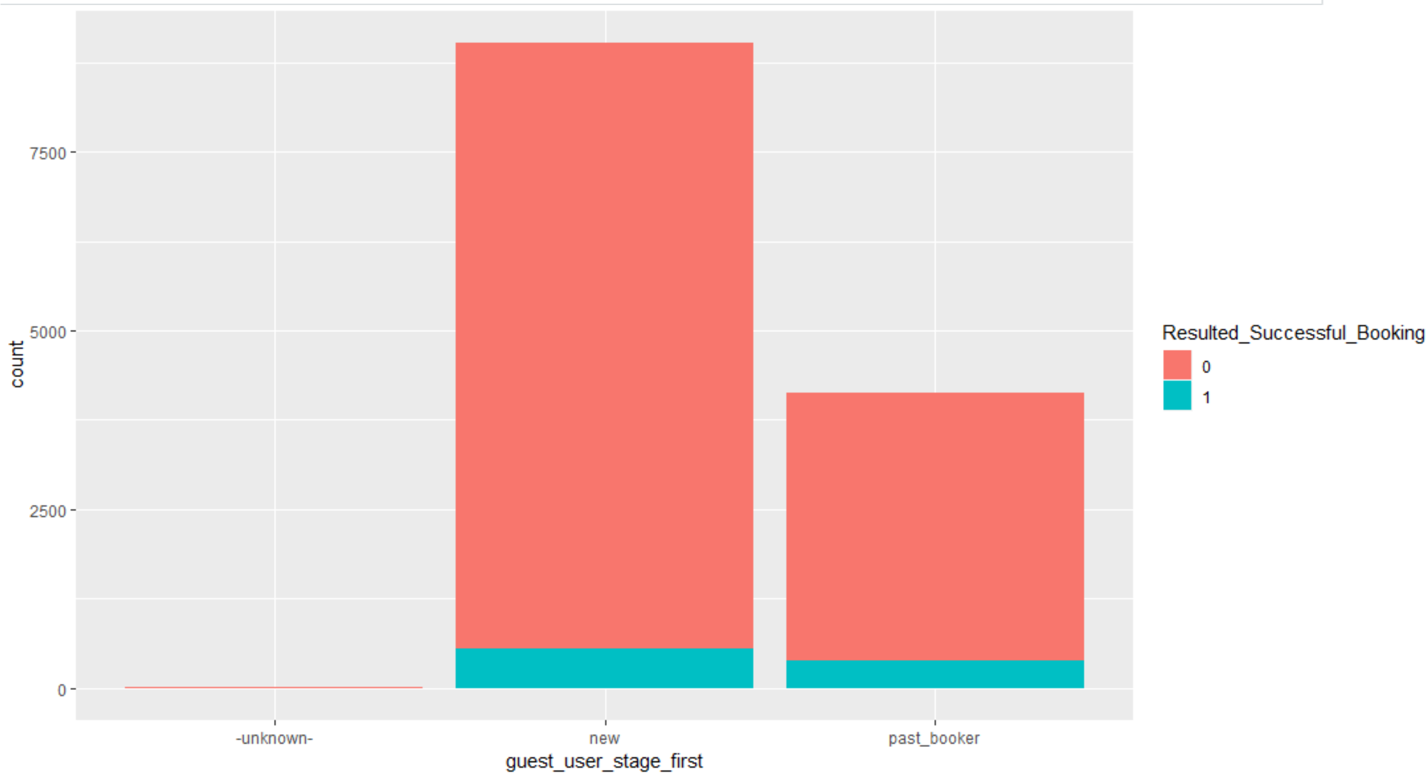
- **Target variable**
  - Whether or not inquiry has turned into legit booking (1,0)
- **Explanatory variable**
  - **Reservation related**
    - Duration of the request stay
    - Number of people requested to stay (M_guest)
    - How far ahead reservation was requested (ds_checkin_first - ts_interaction_first)
    - Whether or not requested checkin date is weekday or weekend (weekly periodicity)
    - Week of the year of requested checkin date (yearly seasonality)
    - *Which holiday overlaps with requested stay (Not here, as there are different countries for the scope of the assessment)*
  - **Guest’s attributes**
    - Whether or not a guest has previously stayed at the listing requested guest_user_stage_first
    - country
    - words_in_user_profile
  - **Host’s attributes**
    - country
    - words_in_user_profile
  - **Listing’s attribute**
    - Total review counts
    - Room type
    - Listing neighborhood
  - **Host and guest Interaction attributes**
    - Time between first interaction (ts_interaction_first) and host’s first reply (ts_reply_at_first) as this tells us how long guest had to wait to get reply from host
    - Time between first reply from host (ts_reply_at_first) and when host approves guests to stay (ts_accepted_at_first) as this tells us how quickly host approved guest after they had replied initially
    - Length of the first message from guest to host (m_first_message_length_in_characters)

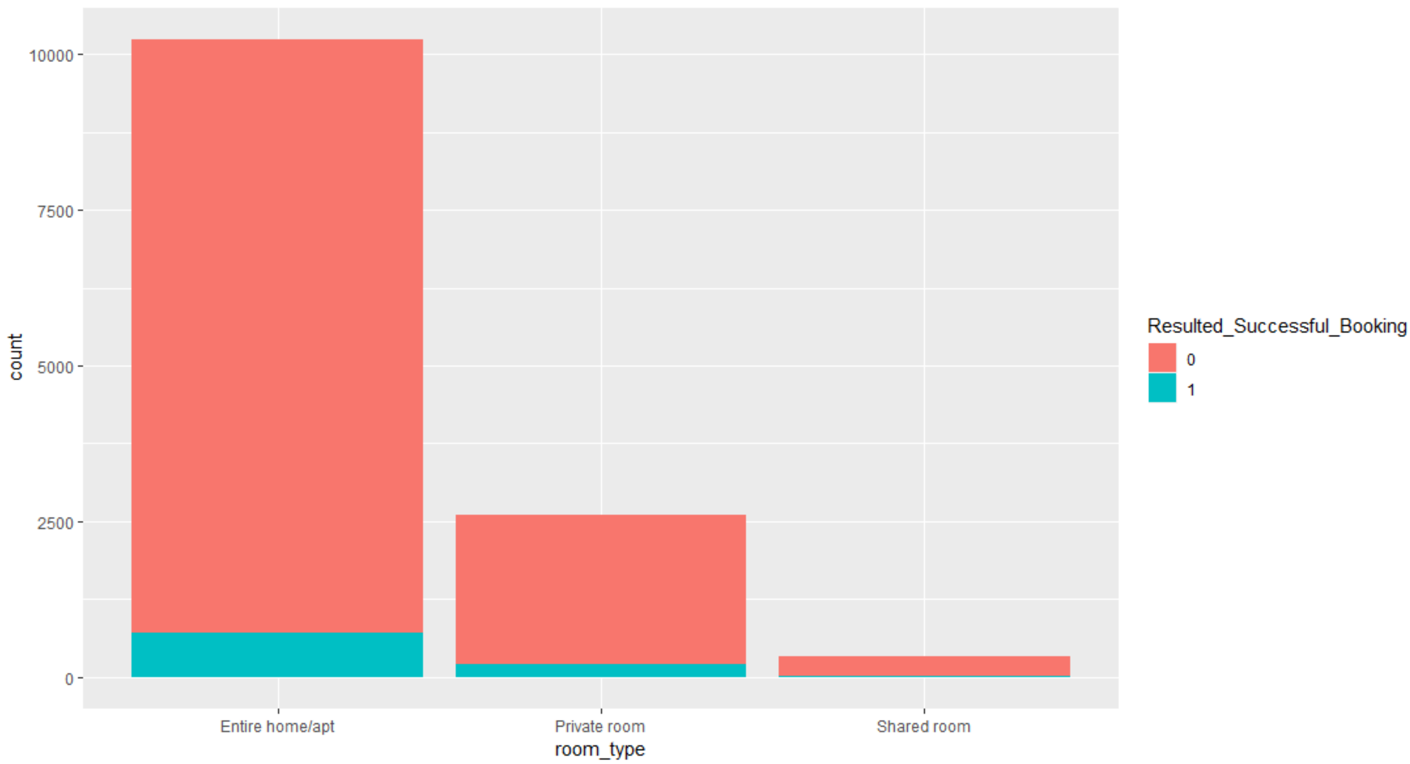
Now let's explore the dataset with some plots for '**contact me'** channel:

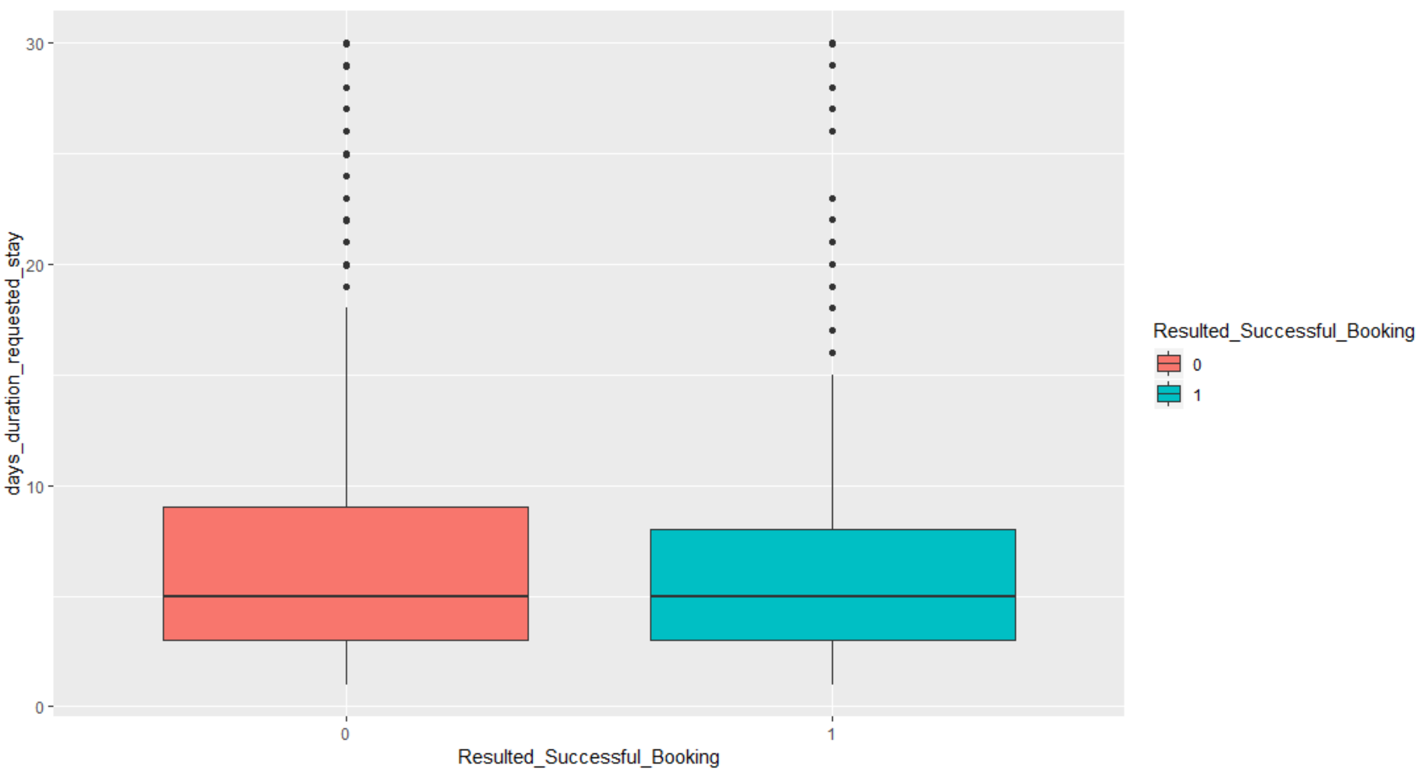


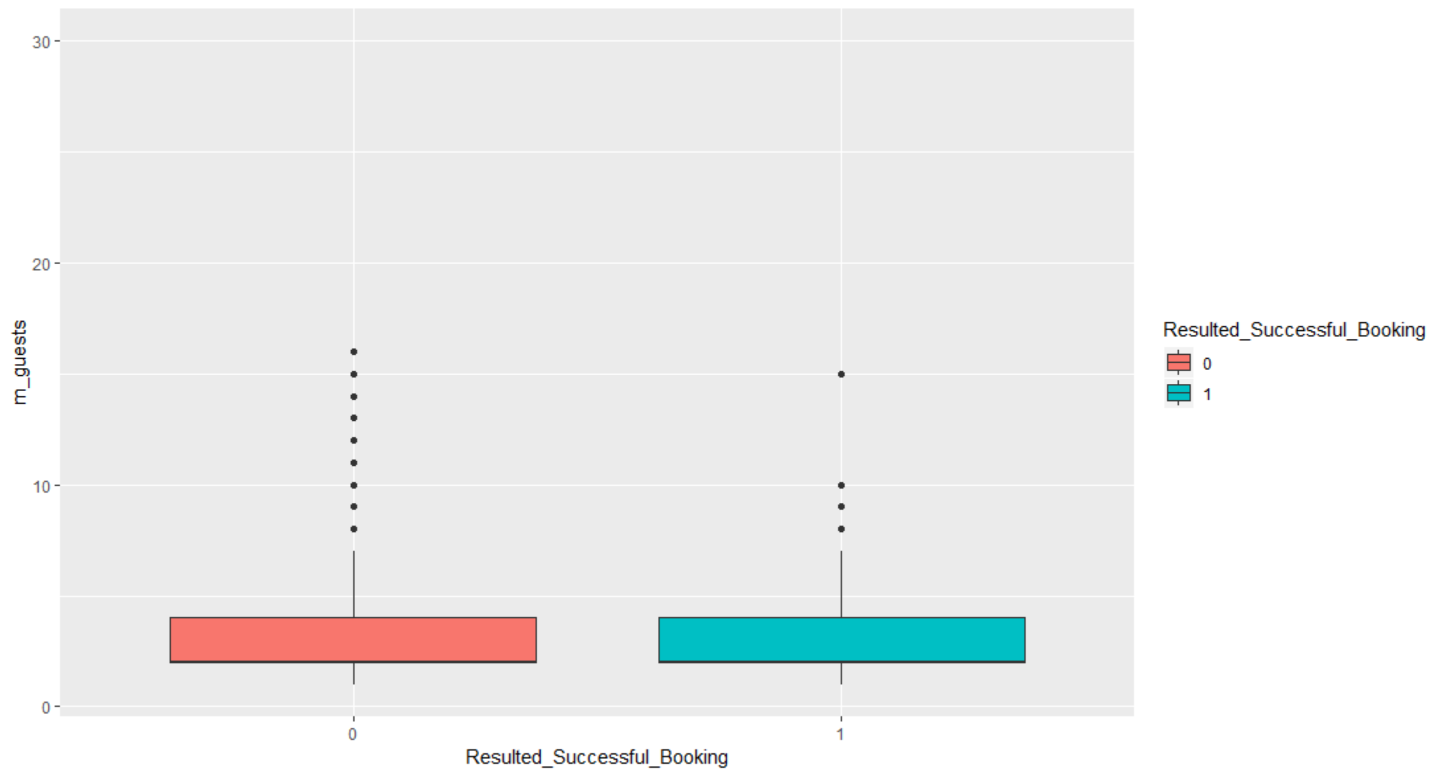


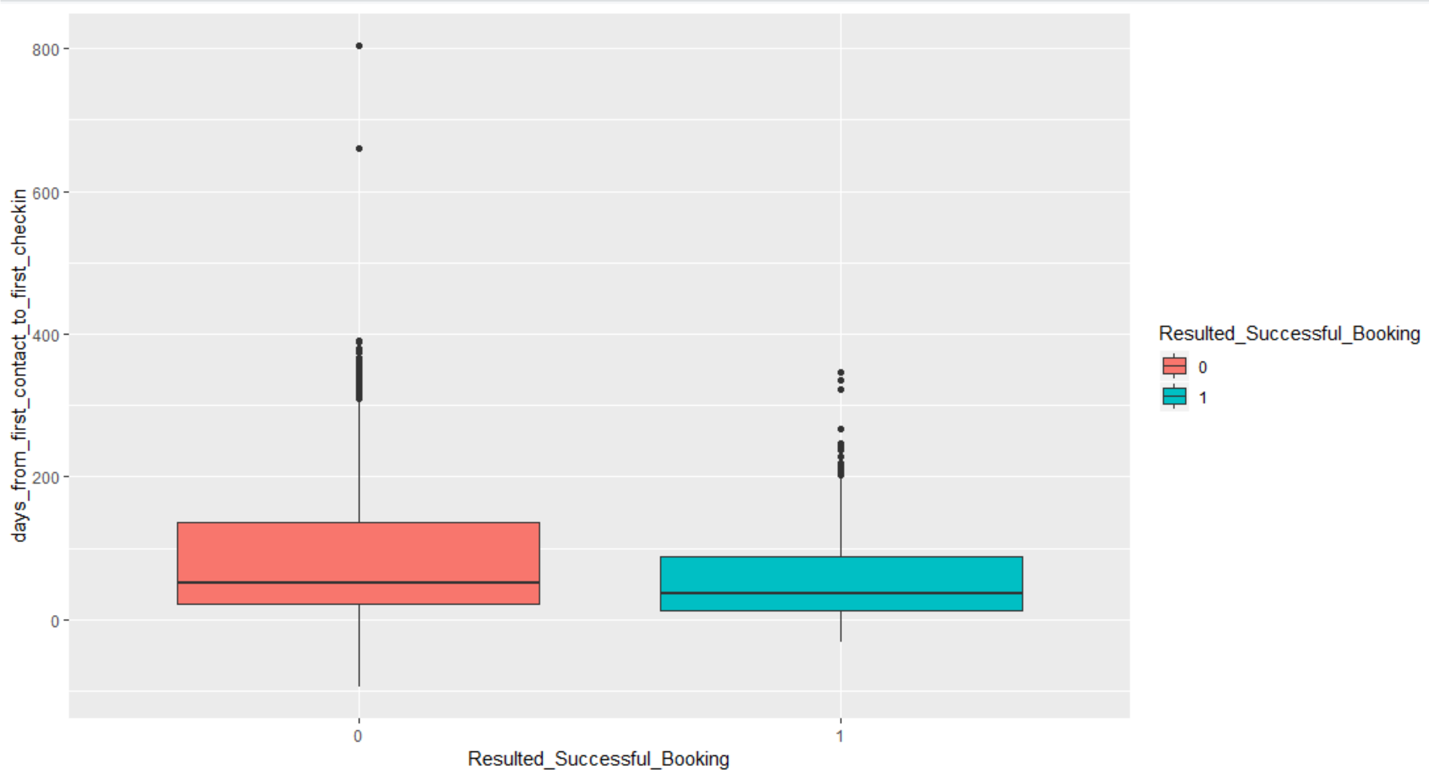

5000
4000
3000
2000
1000
0
count
1
2
3
4
5
6
7
8
9
10
11
12
moy_checkin
Resulted_Successful_Booking
0
1

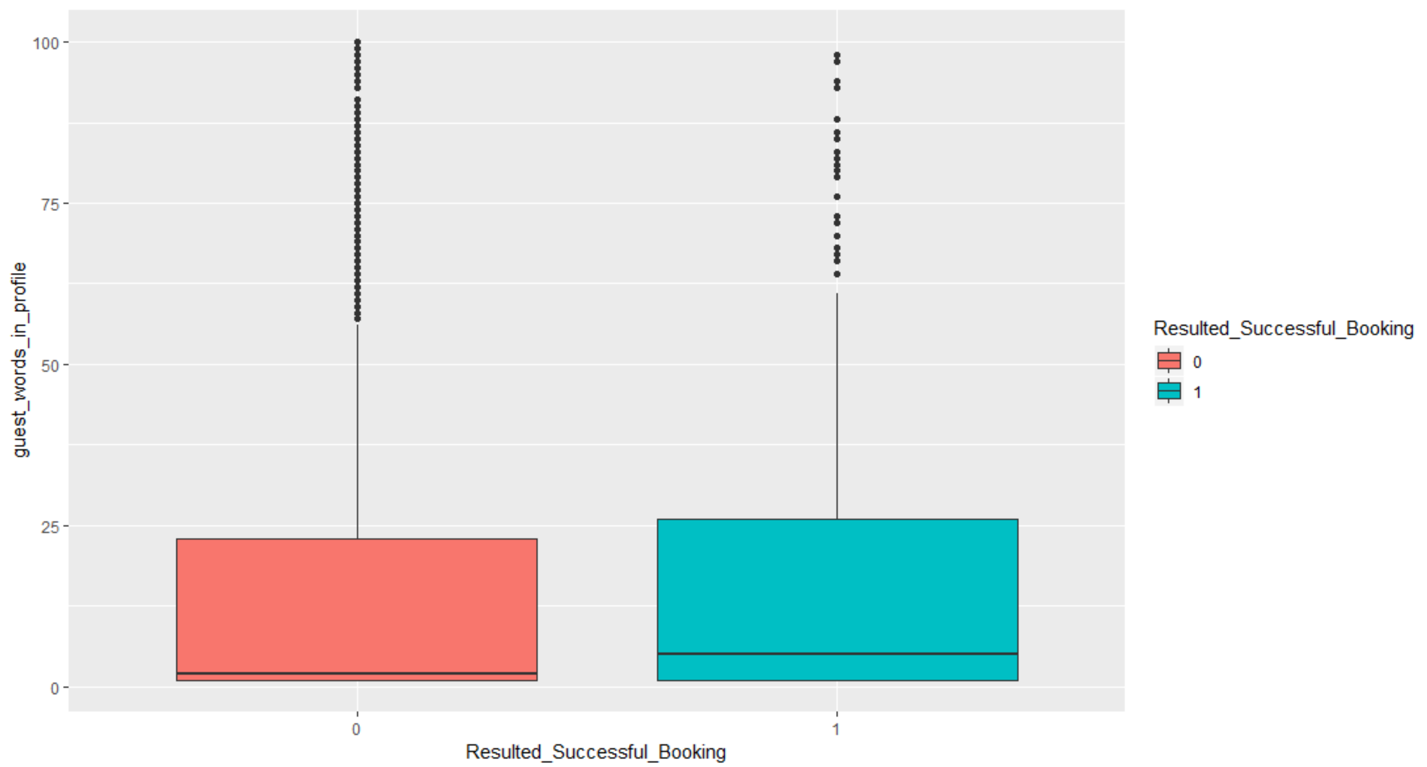

7500
5000
2500
0
count
-unknown-
new
past_booker
guest_user_stage_first
Resulted_Successful_Booking
0
1

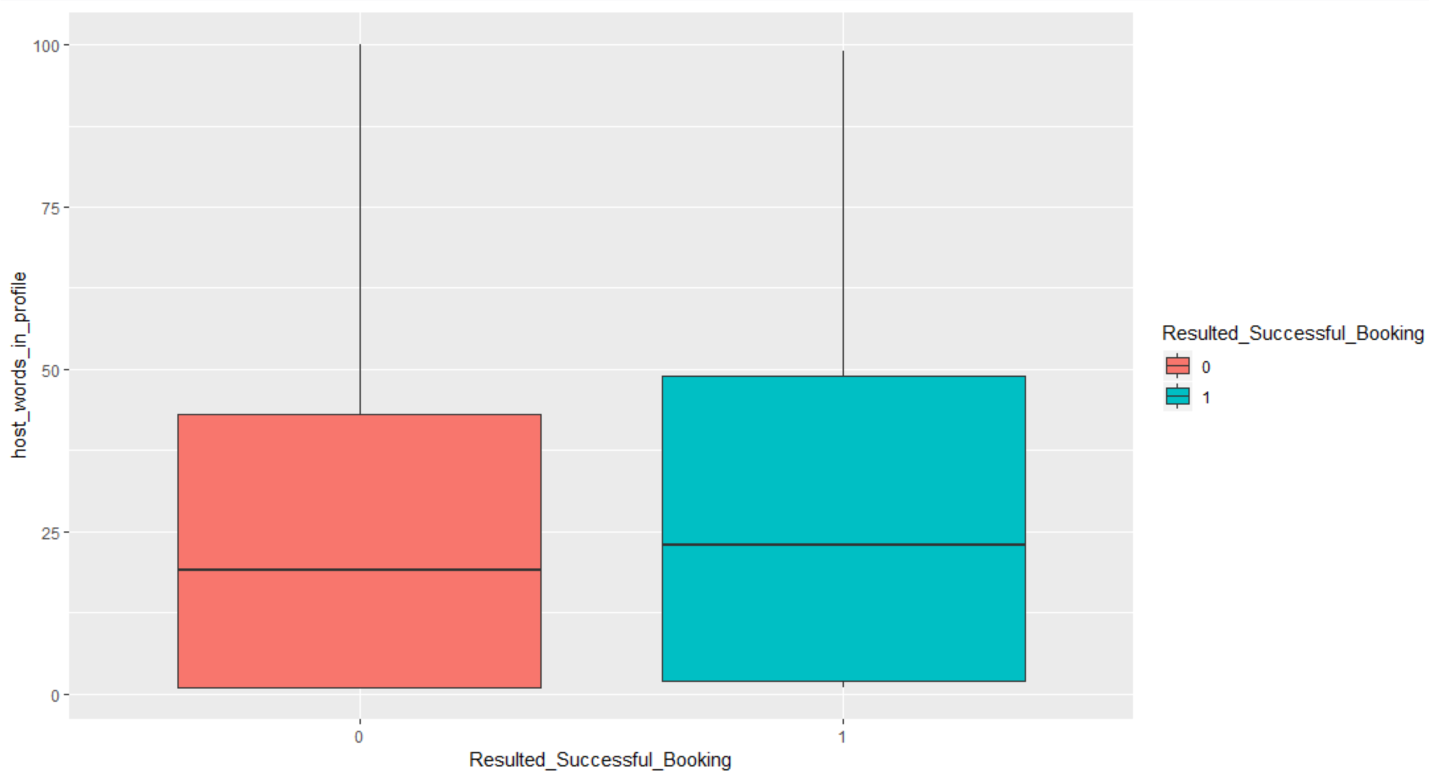

count
10000
7500
5000
2500
0
Entire home/apt
Private room
Shared room
room_type
Resulted_Successful_Booking
0
1

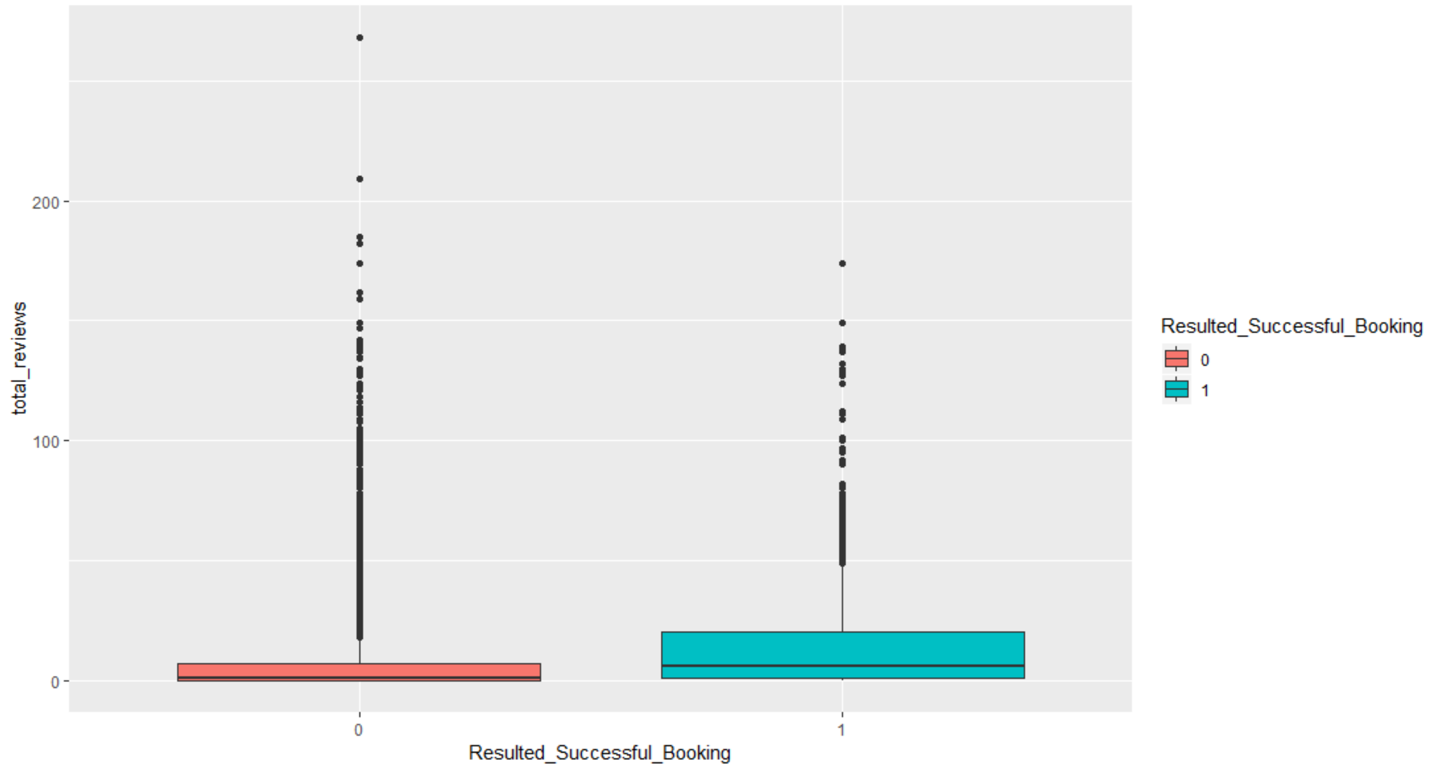

days_duration_requested_stay
30
20
10
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

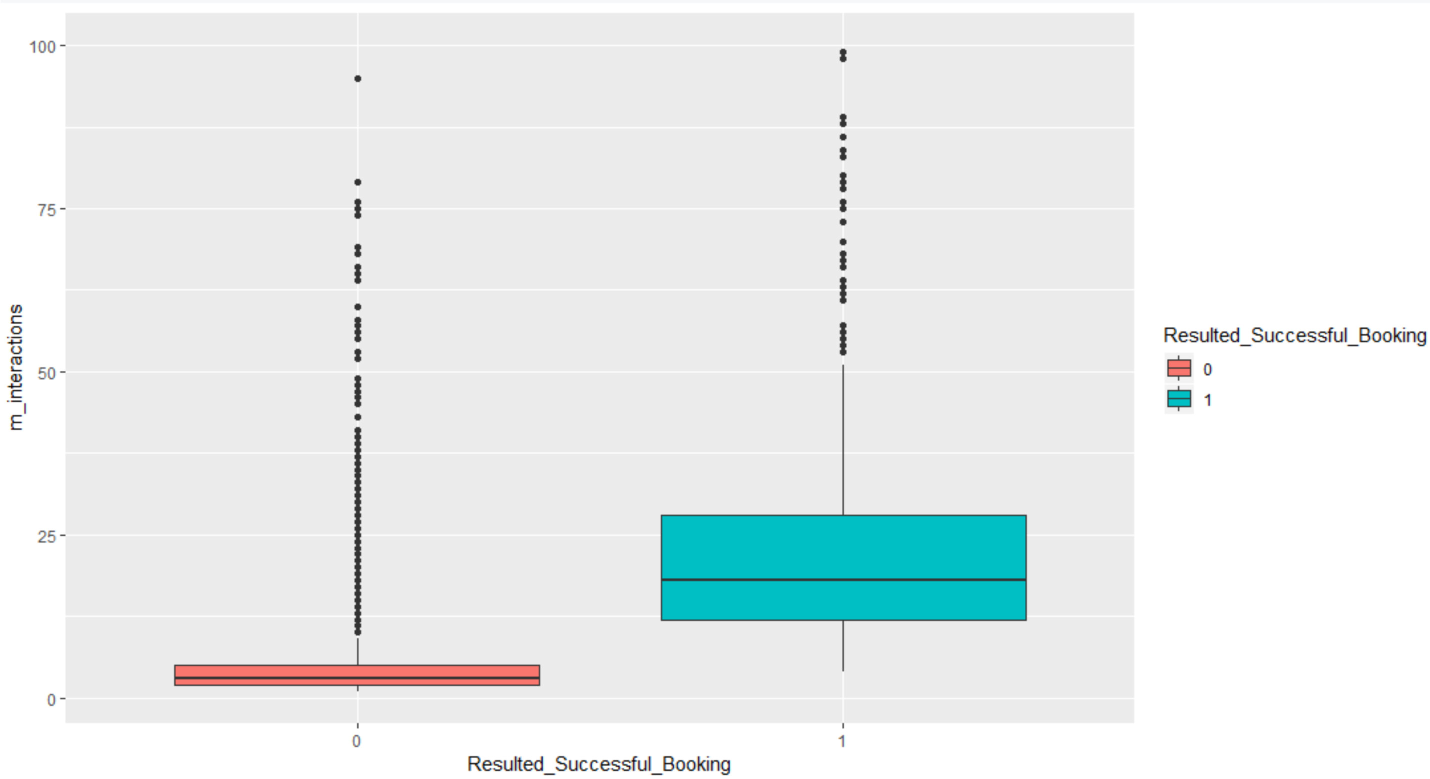

30
20
10
0
m_guests
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

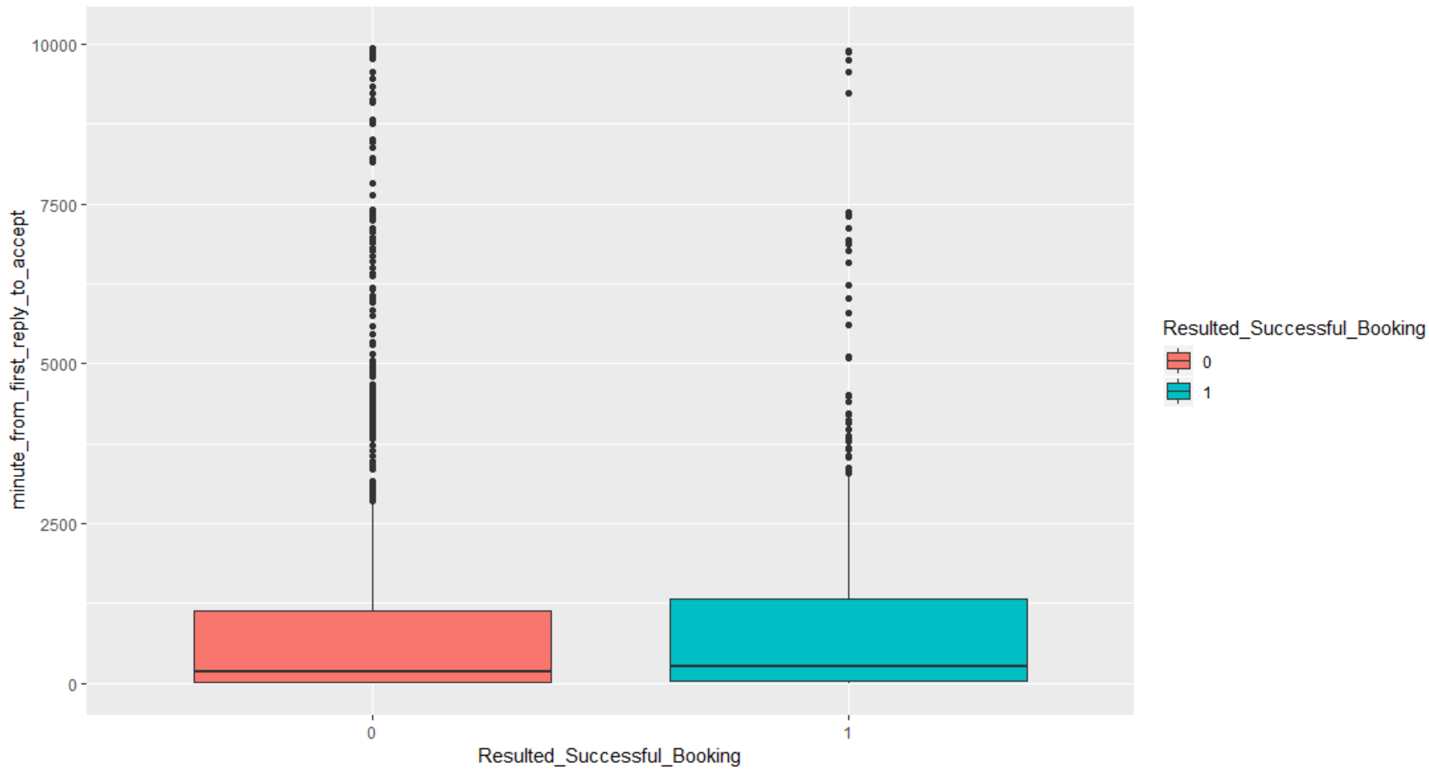

800
600
400
200
0
days_from_first_contact_to_first_checkin
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

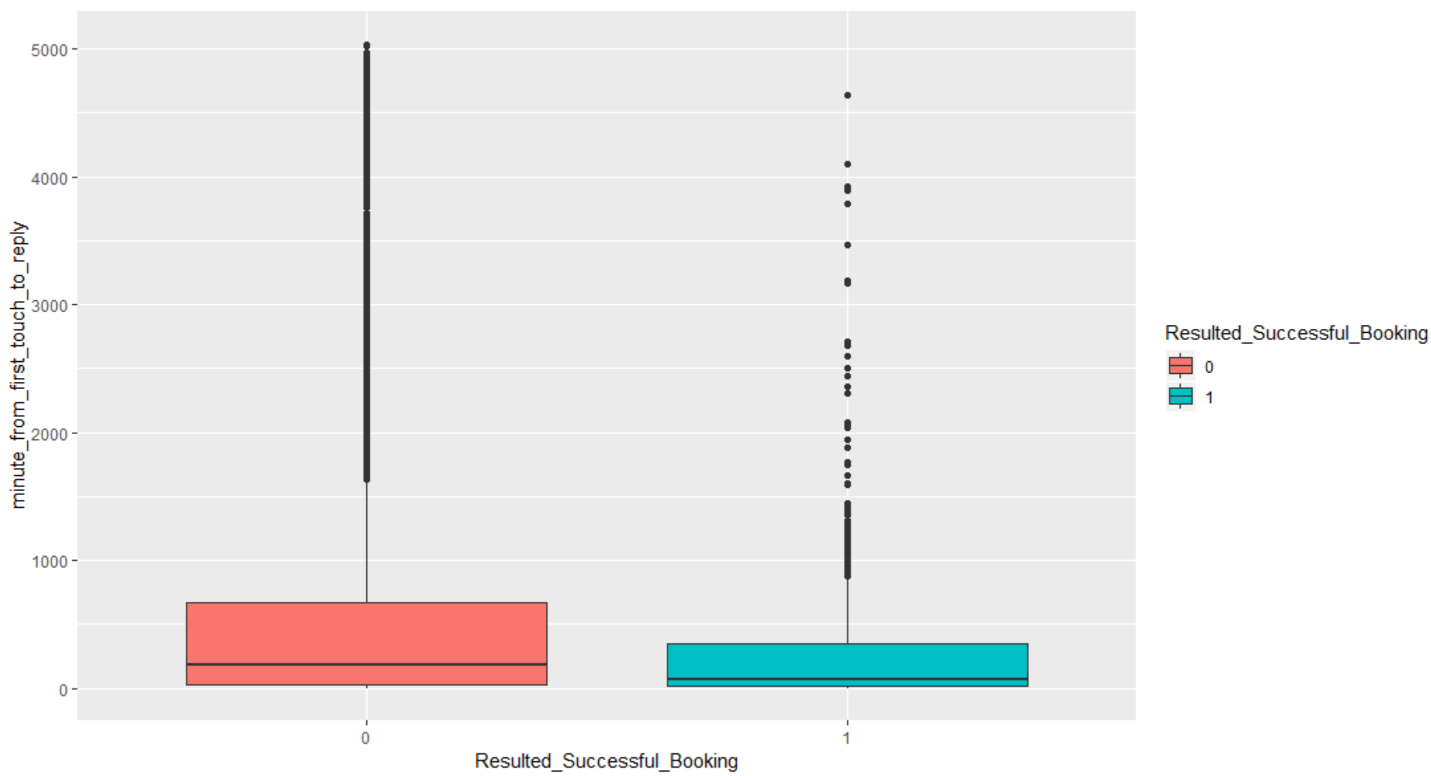

guest_words_in_profile
100
75
50
25
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1


host_words_in_profile
100
75
50
25
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1


total_reviews
200
100
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

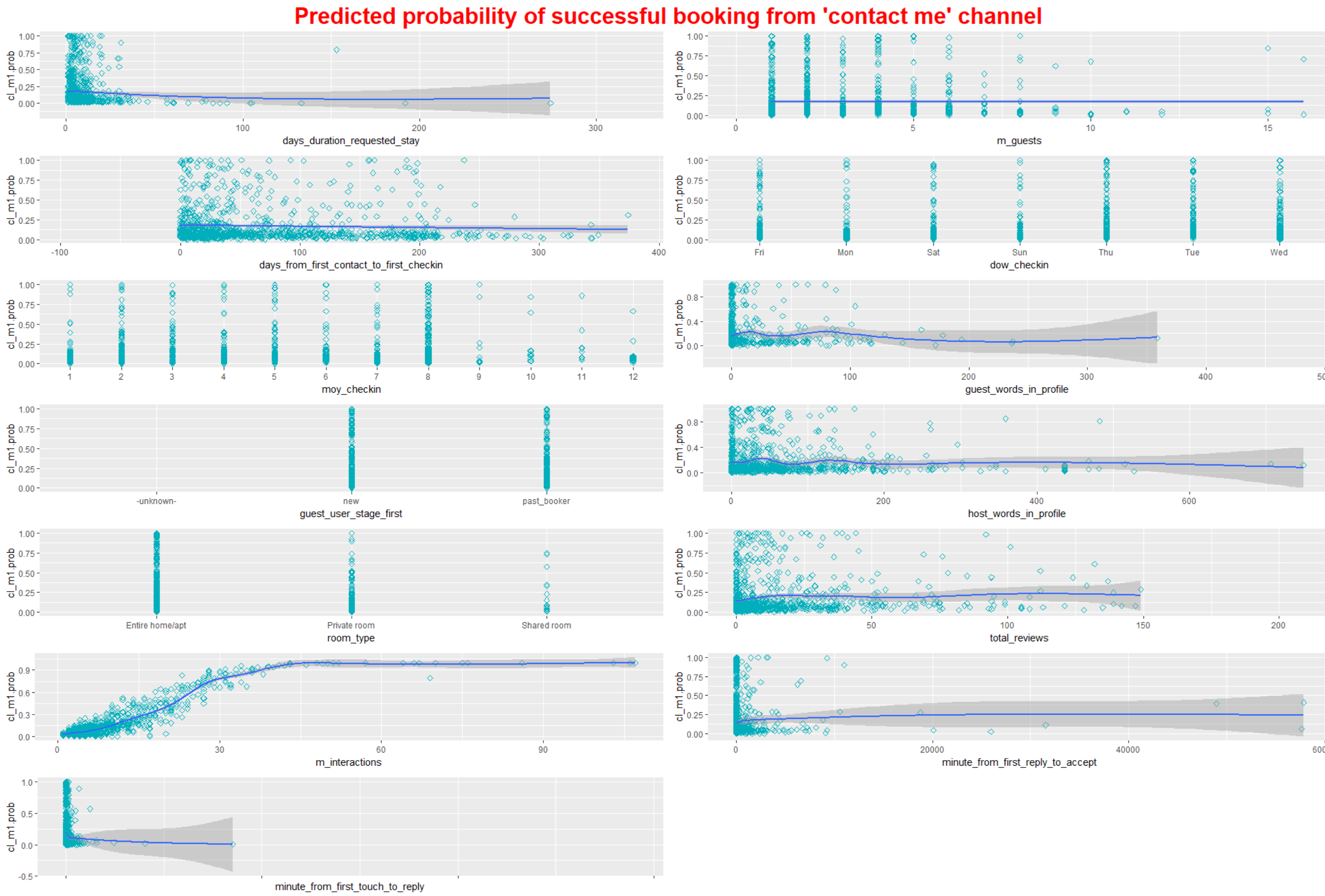

m_interactions
100
75
50
25
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1


10000
7500
5000
2500
0
minute_from_first_reply_to_accept
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

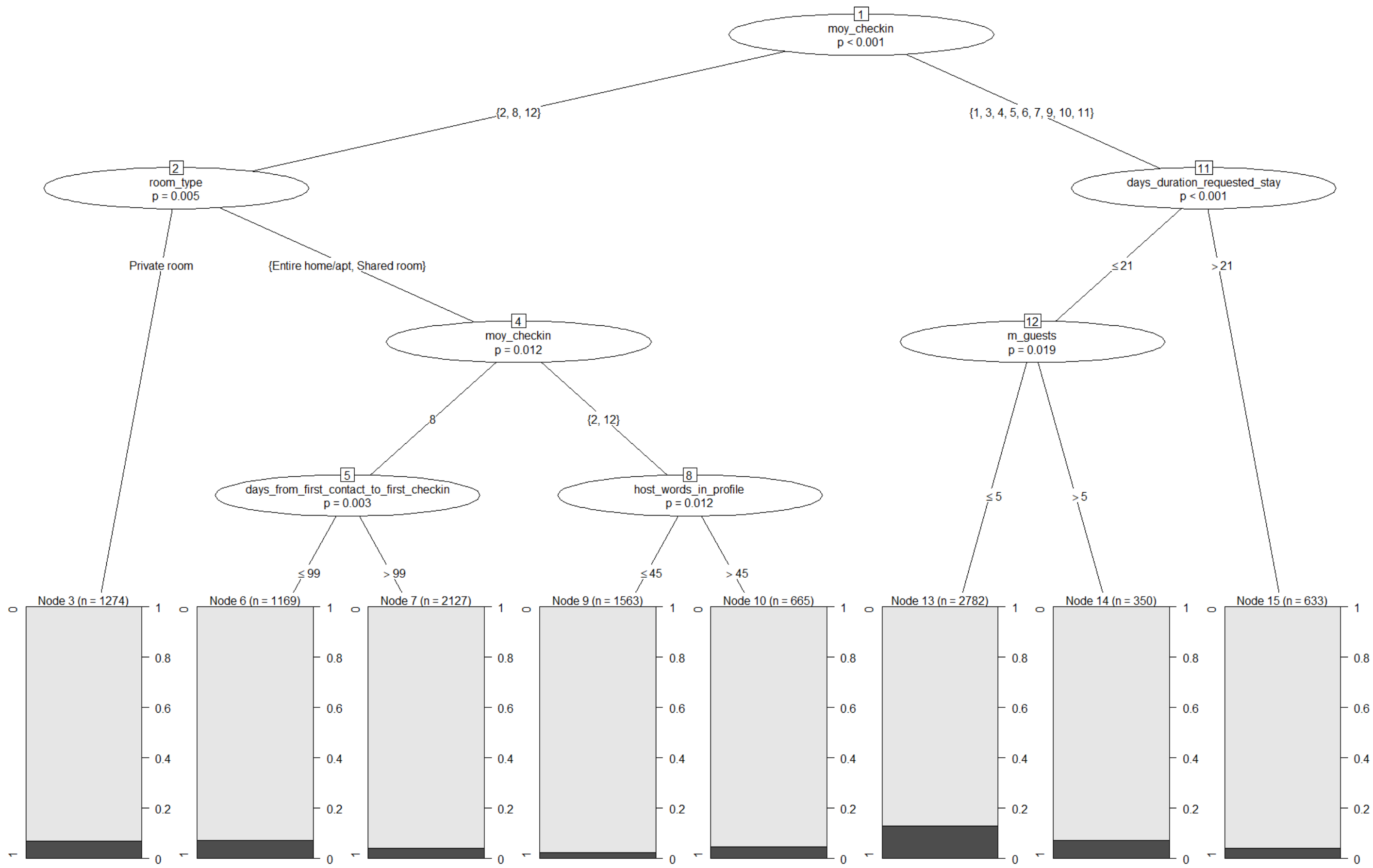

5000
4000
3000
2000
1000
0
minute_from_first_touch_to_reply
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

Next, I have built classification models for inquiries from ‘**contact me’** channel:

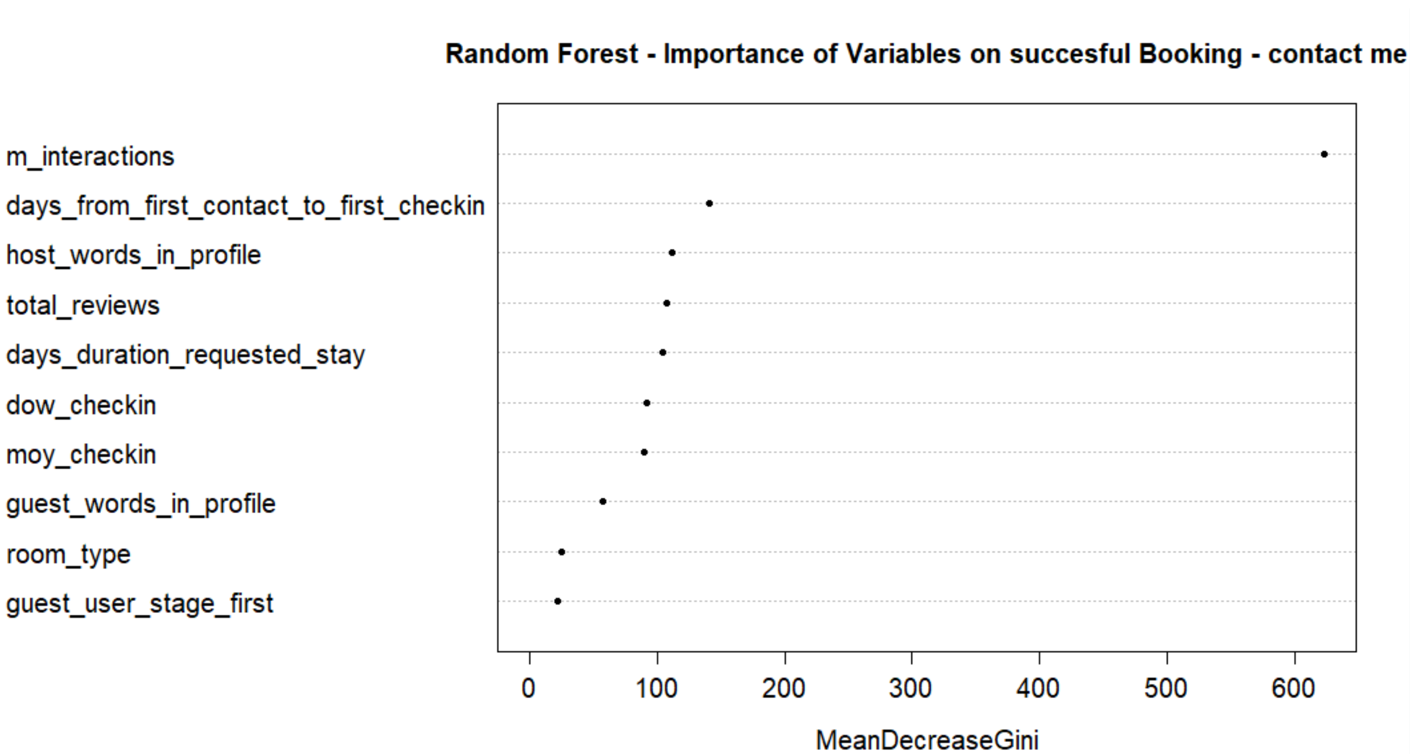
<u>Logistic regression (resulting 88% overall accuracy – more details in code), and below plots are predicted probabilities of successful booking against individual predictors.</u>



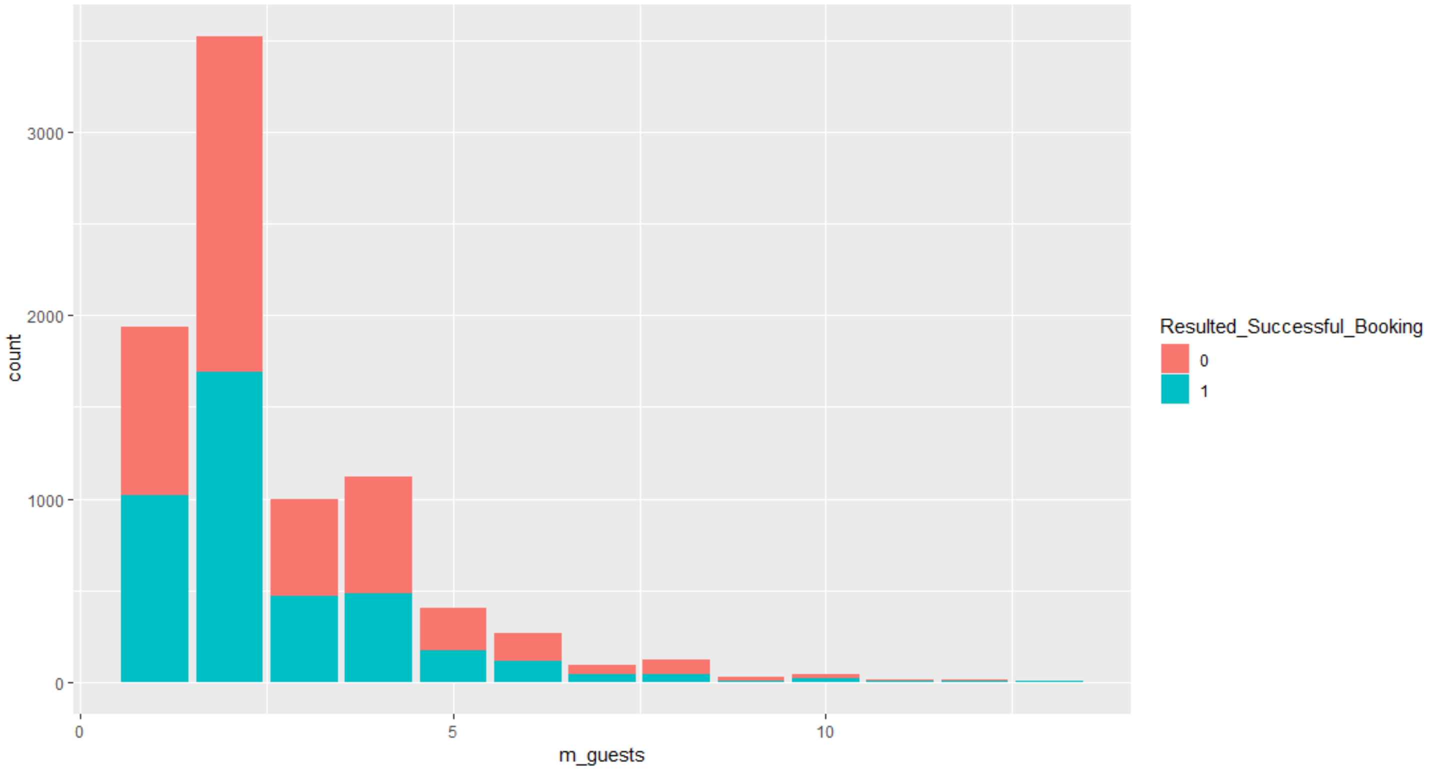
<u>Next Decision Tree:</u>

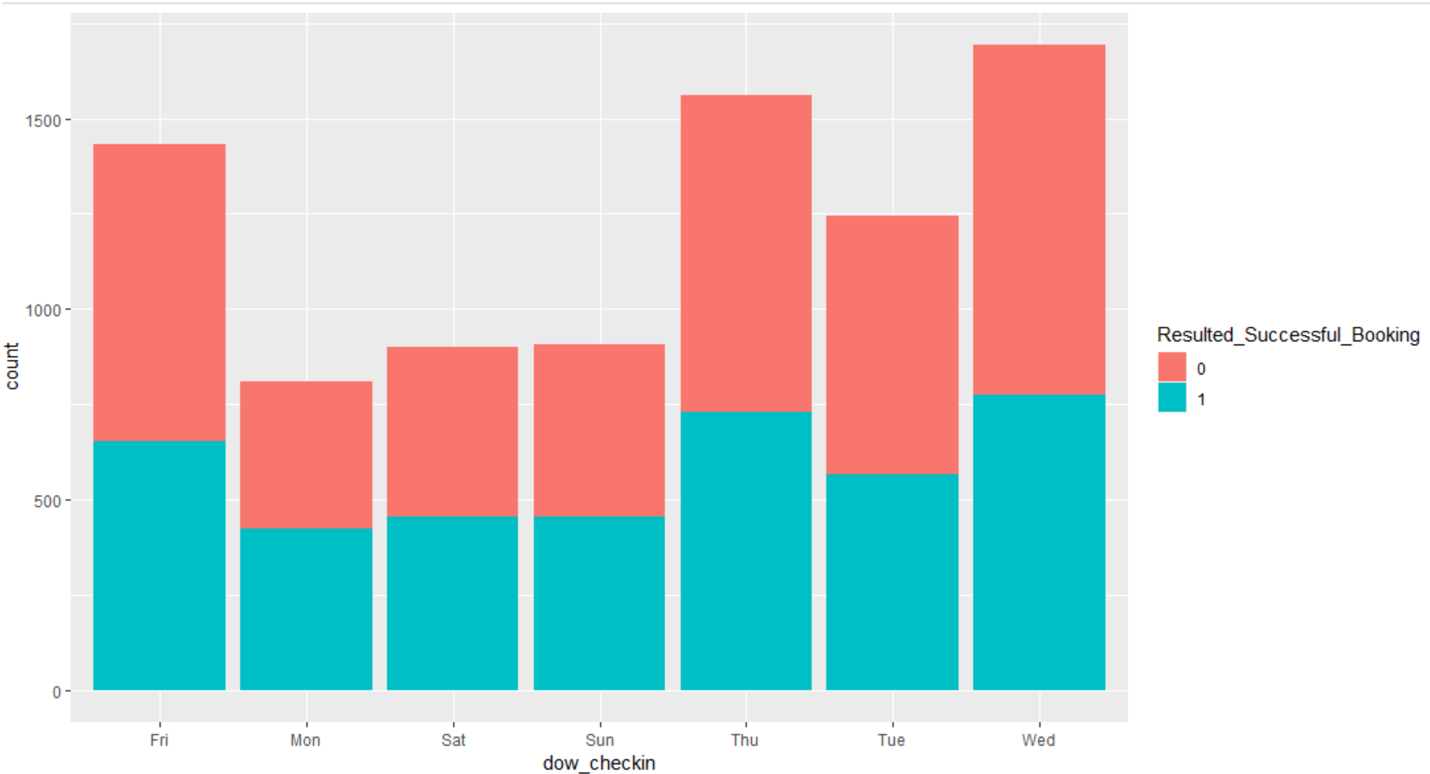


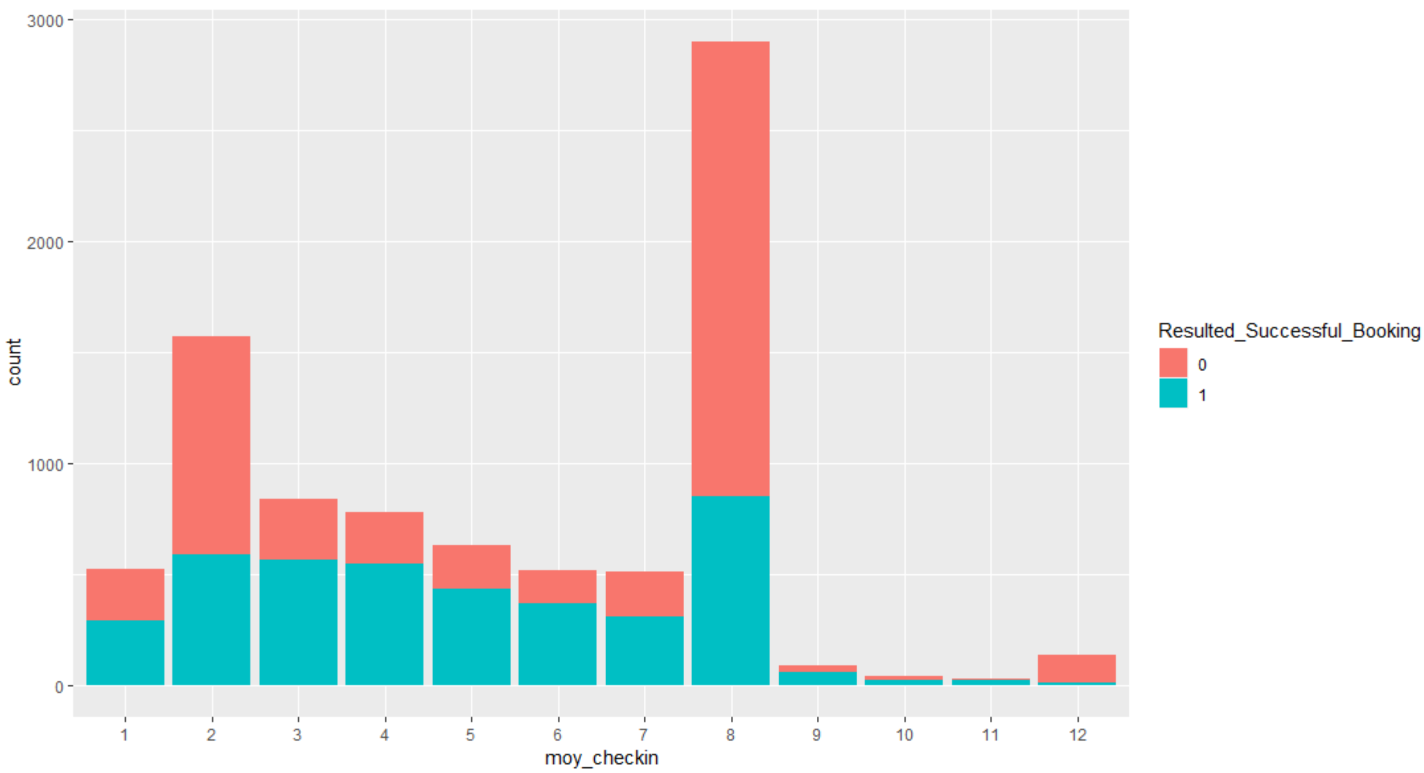
<u>And Finally, a random forest model, which has resulted in (95% accuracy), and here are the important features that influence the odd of an inquiry converting to a successful booking request:</u>

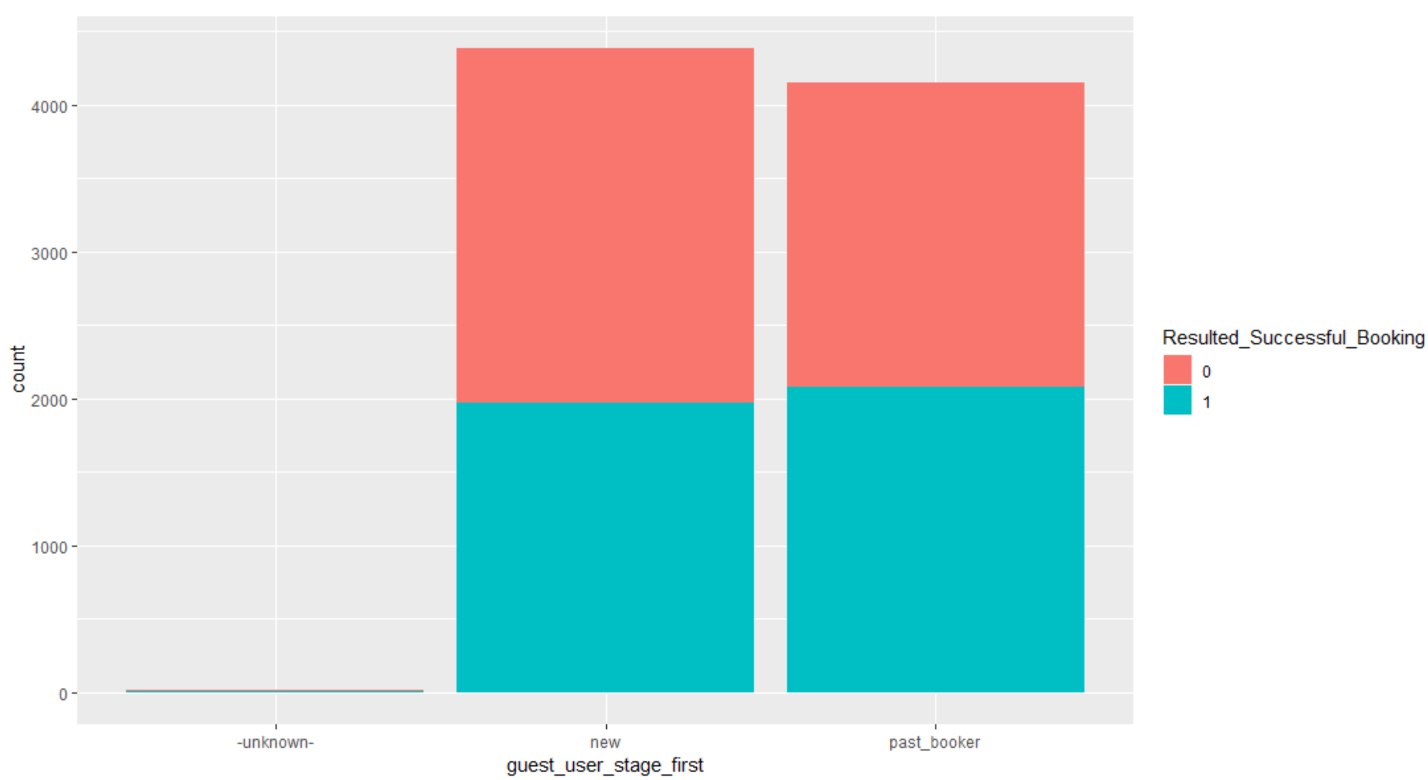


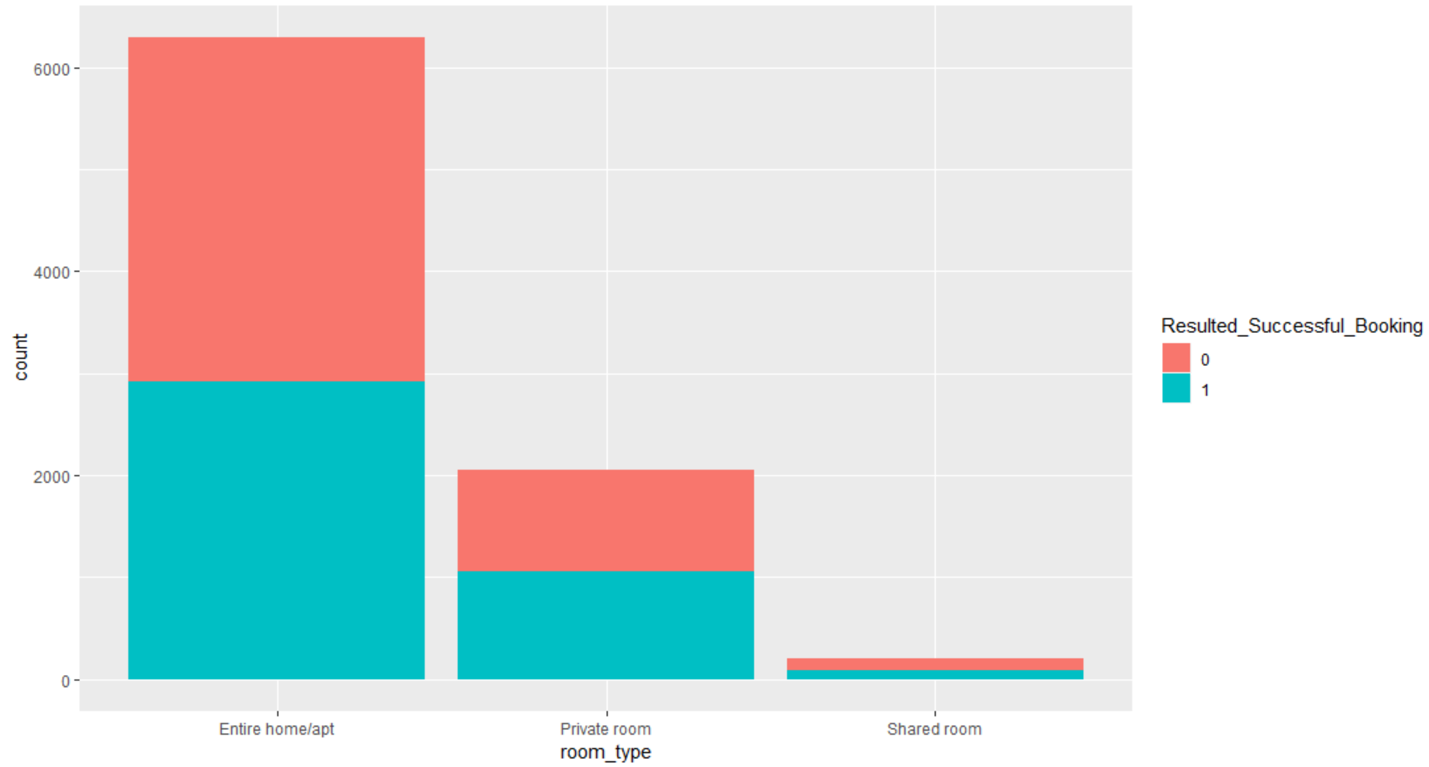
Now let's explore the dataset with some plots for '**Book It'** channel:



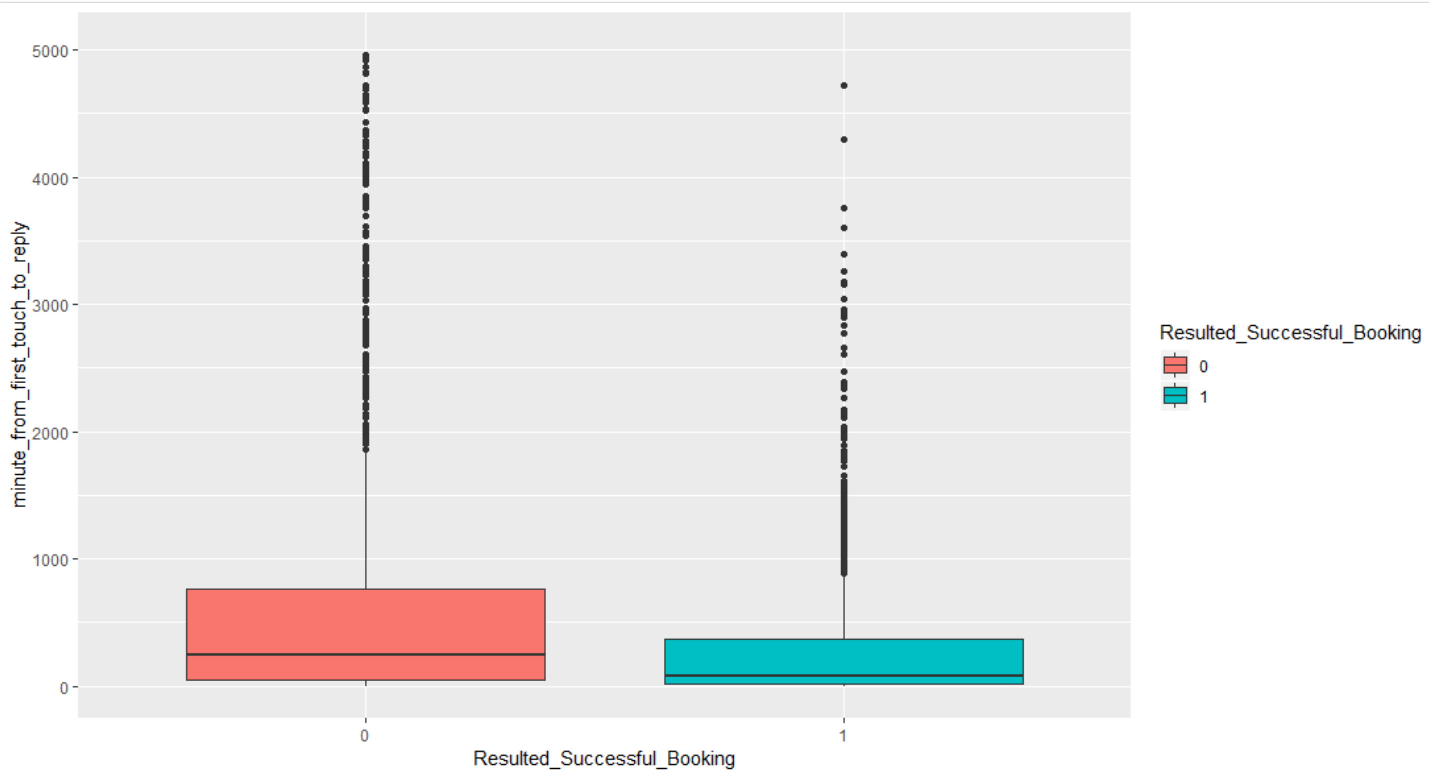



3000
2000
1000
0
count
1
2
3
4
5
6
7
8
9
10
11
12
moy_checkin
Resulted_Successful_Booking
0
1

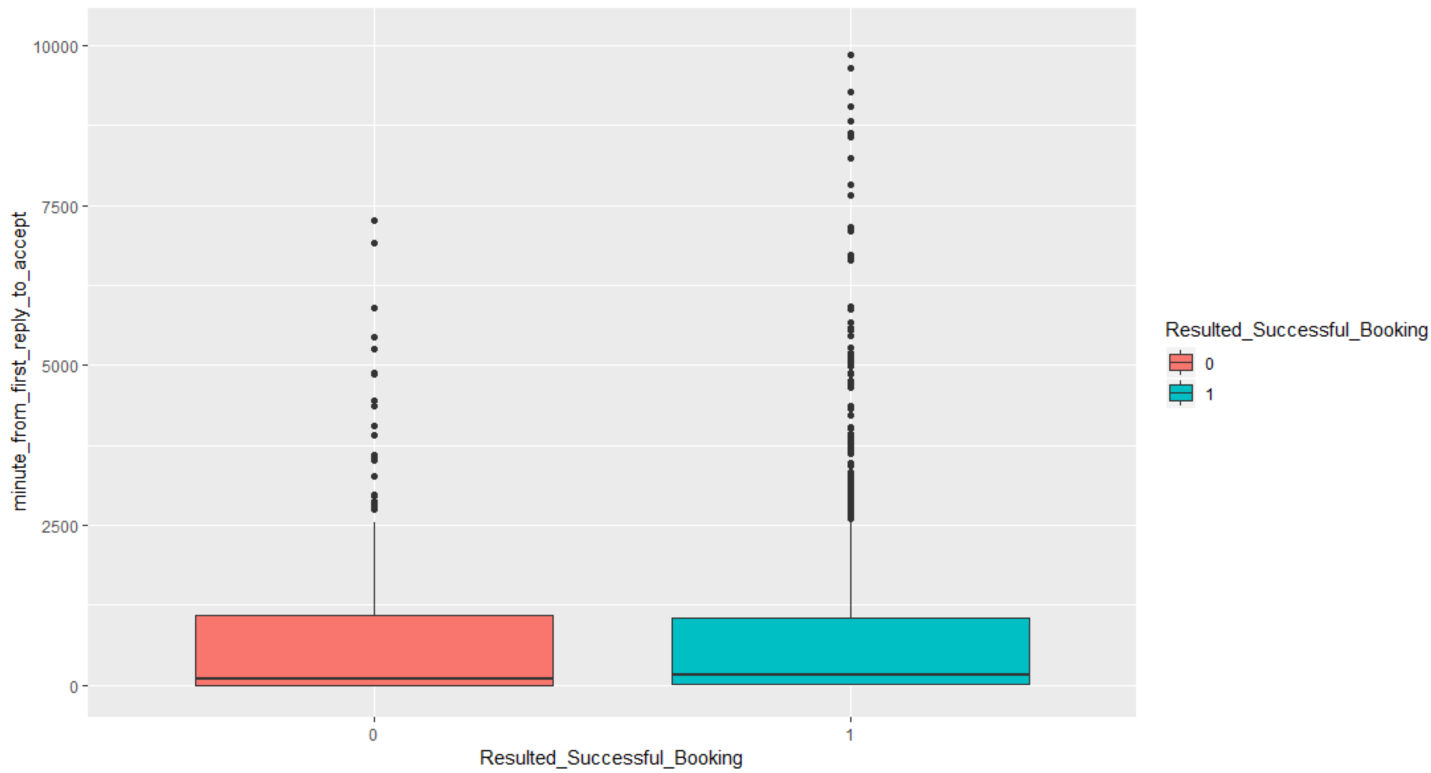

4000
3000
2000
1000
0
count
-unknown-
new
past_booker
guest_user_stage_first
Resulted_Successful_Booking
0
1



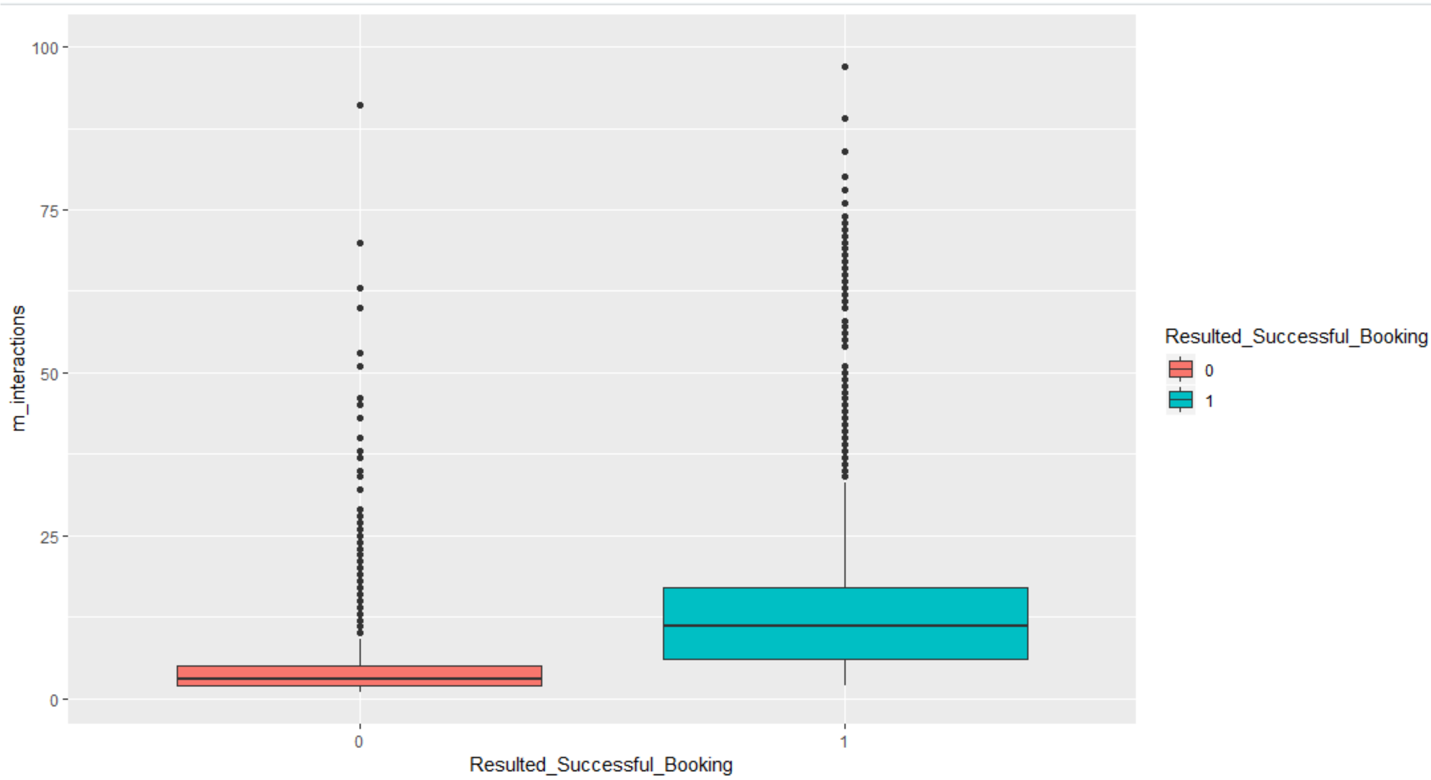
6000
4000
2000
0
count
Entire home/apt
Private room
Shared room
room_type
Resulted_Successful_Booking
0
1



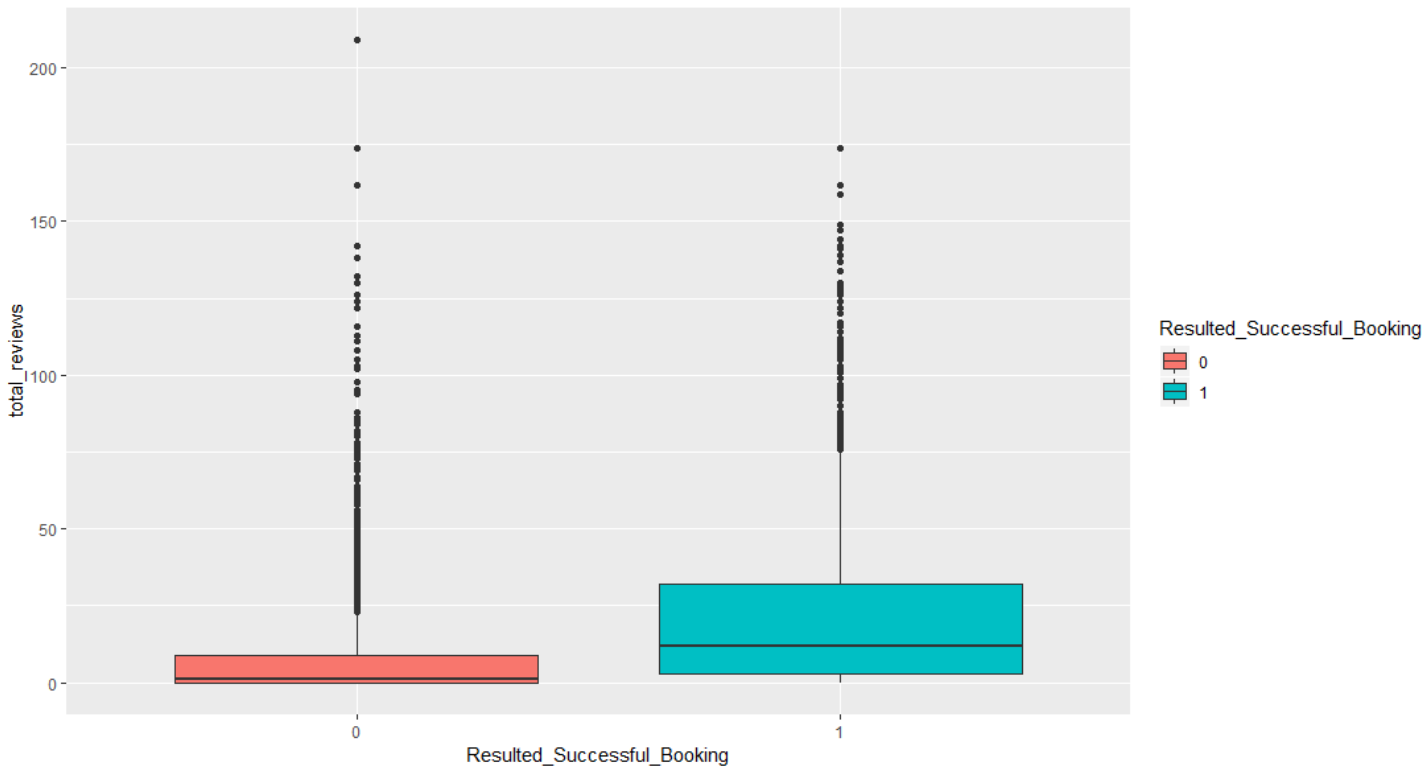
5000
4000
3000
2000
1000
0
minute_from_first_touch_to_reply
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1



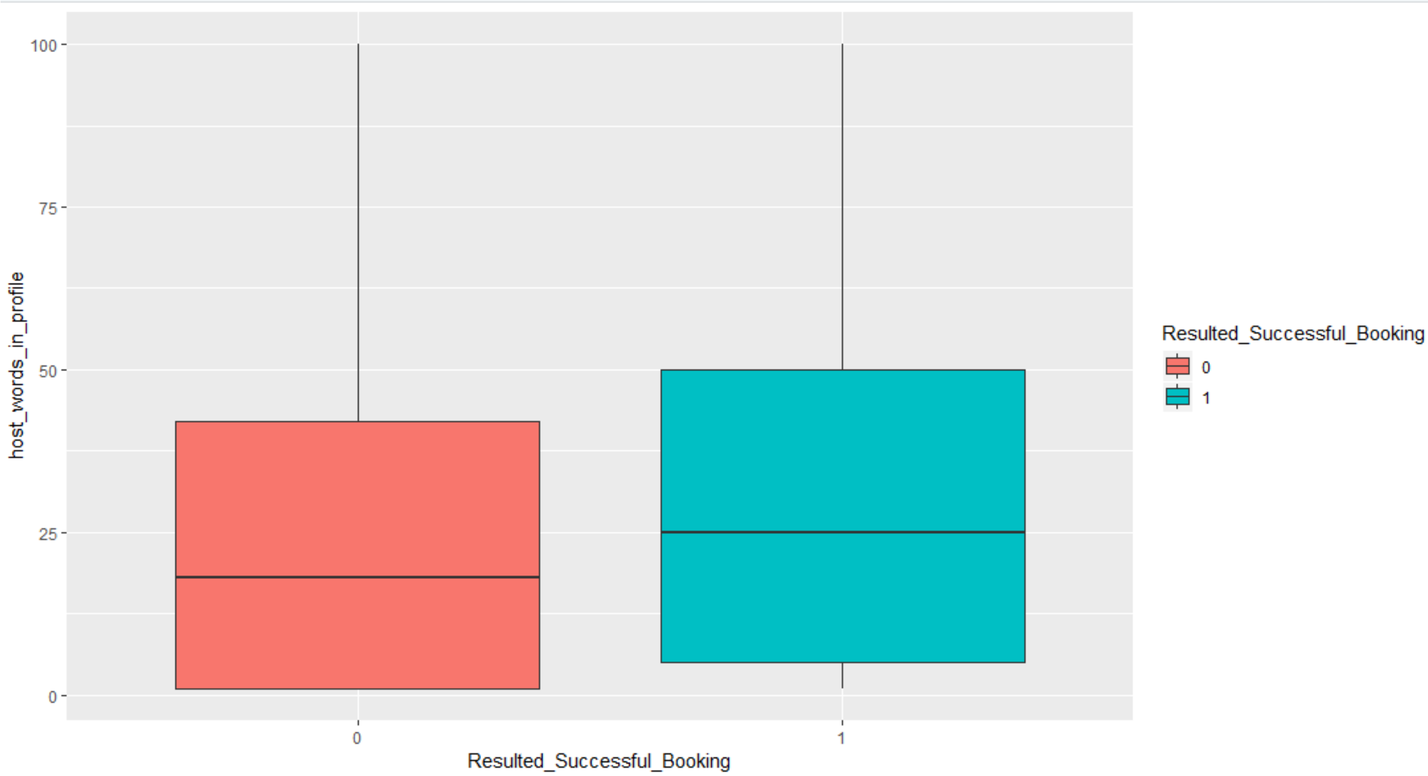
minute_from_first_reply_to_accept
10000
7500
5000
2500
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1



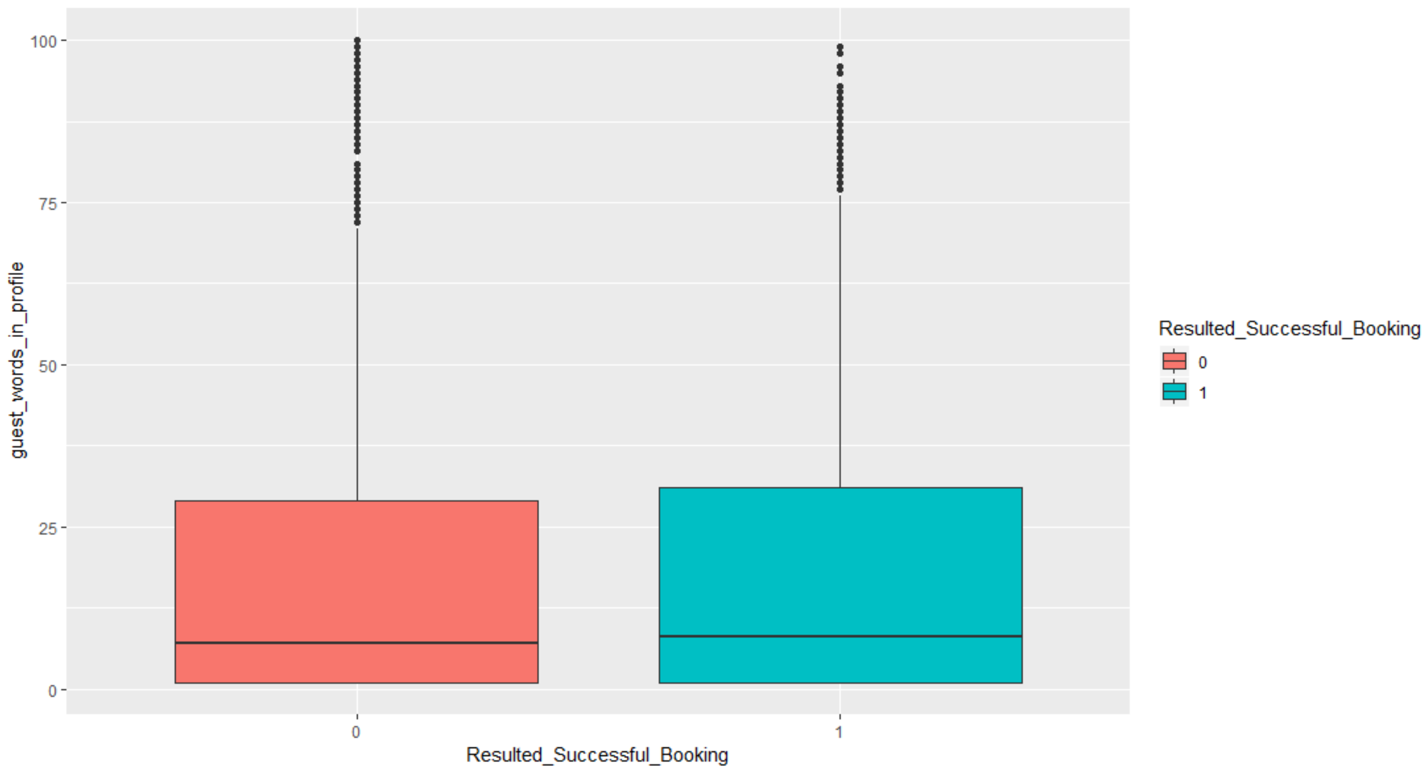
m_interactions
100
75
50
25
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

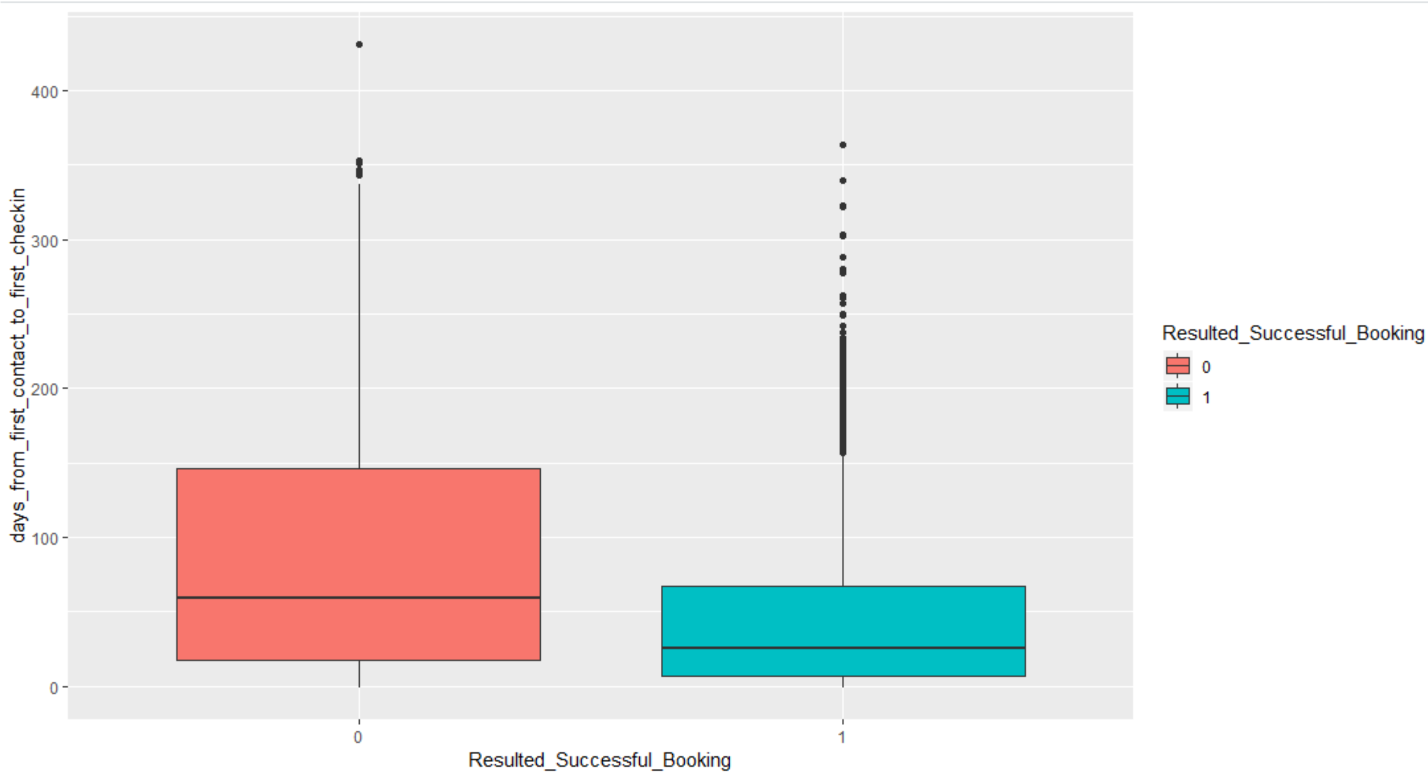


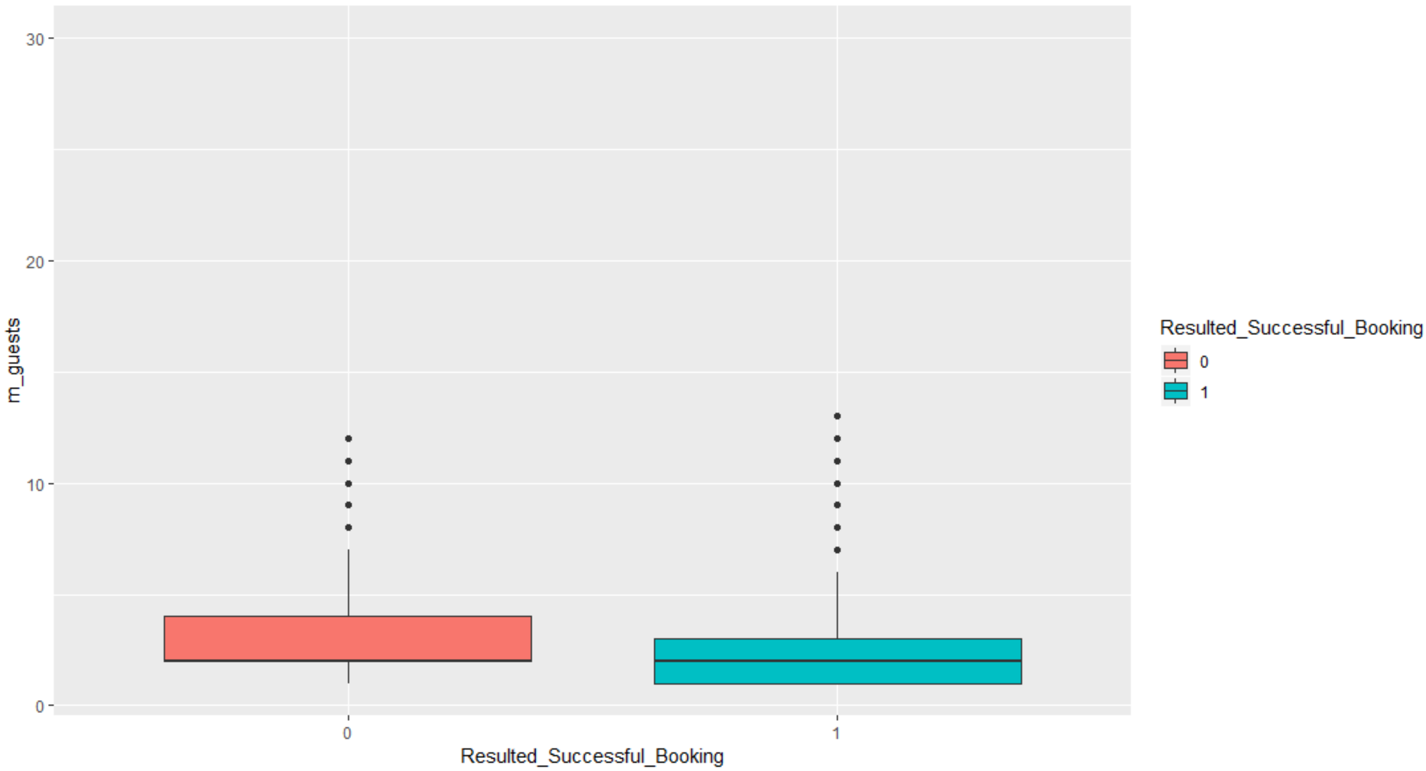
200
150
100
50
0
total_reviews
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1



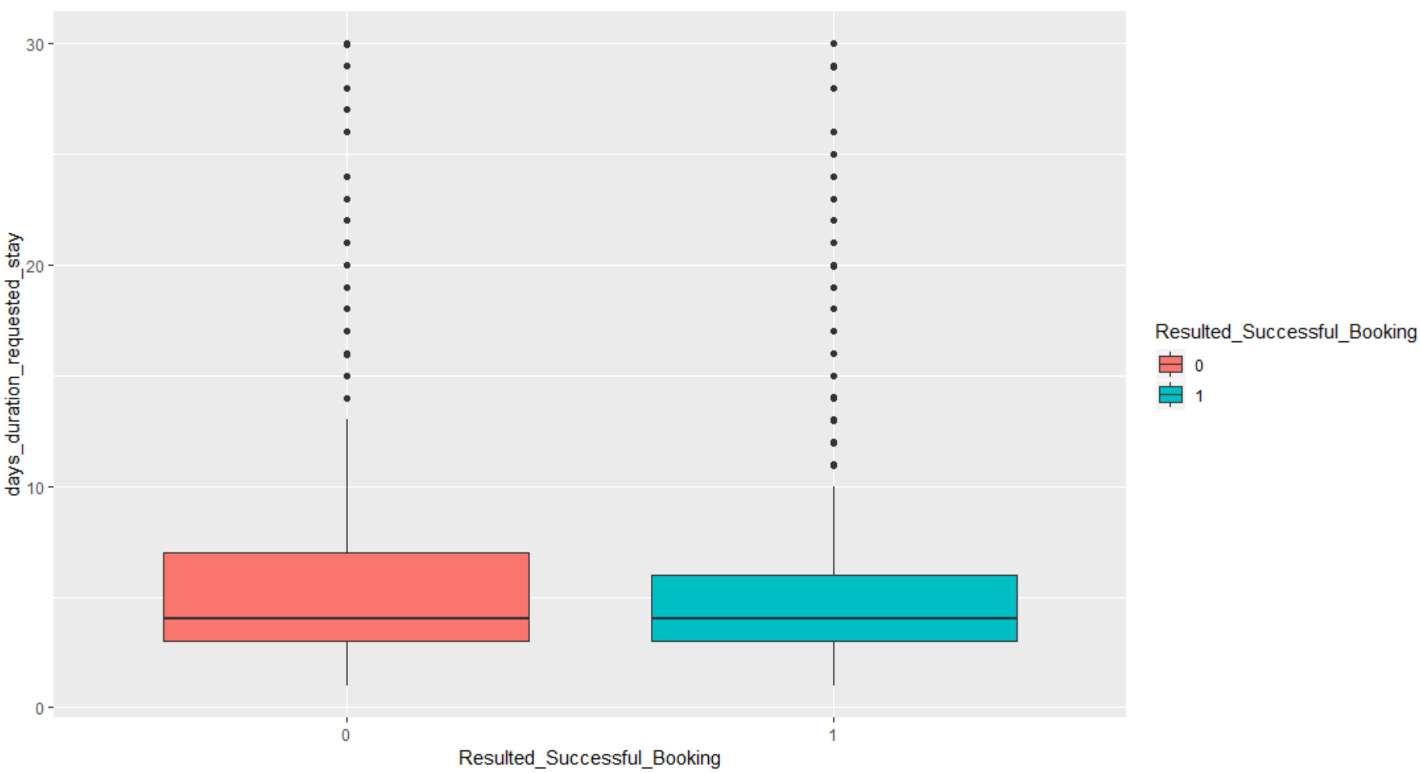
100
75
50
25
0
host_words_in_profile
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1


guest_words_in_profile
100
75
50
25
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1


days_from_first_contact_to_first_checkin
400
300
200
100
0
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1



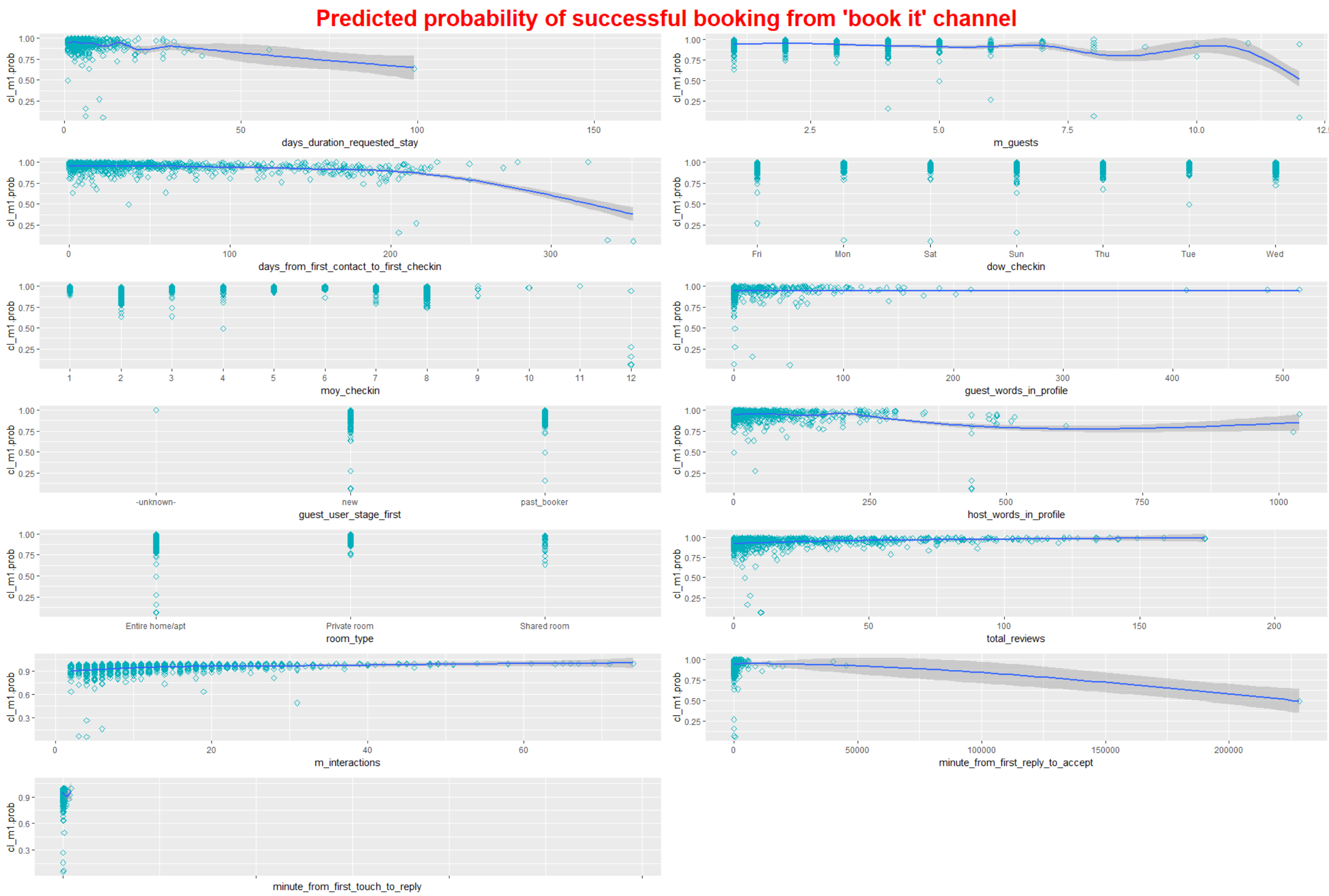
30
20
10
0
m_guests
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1



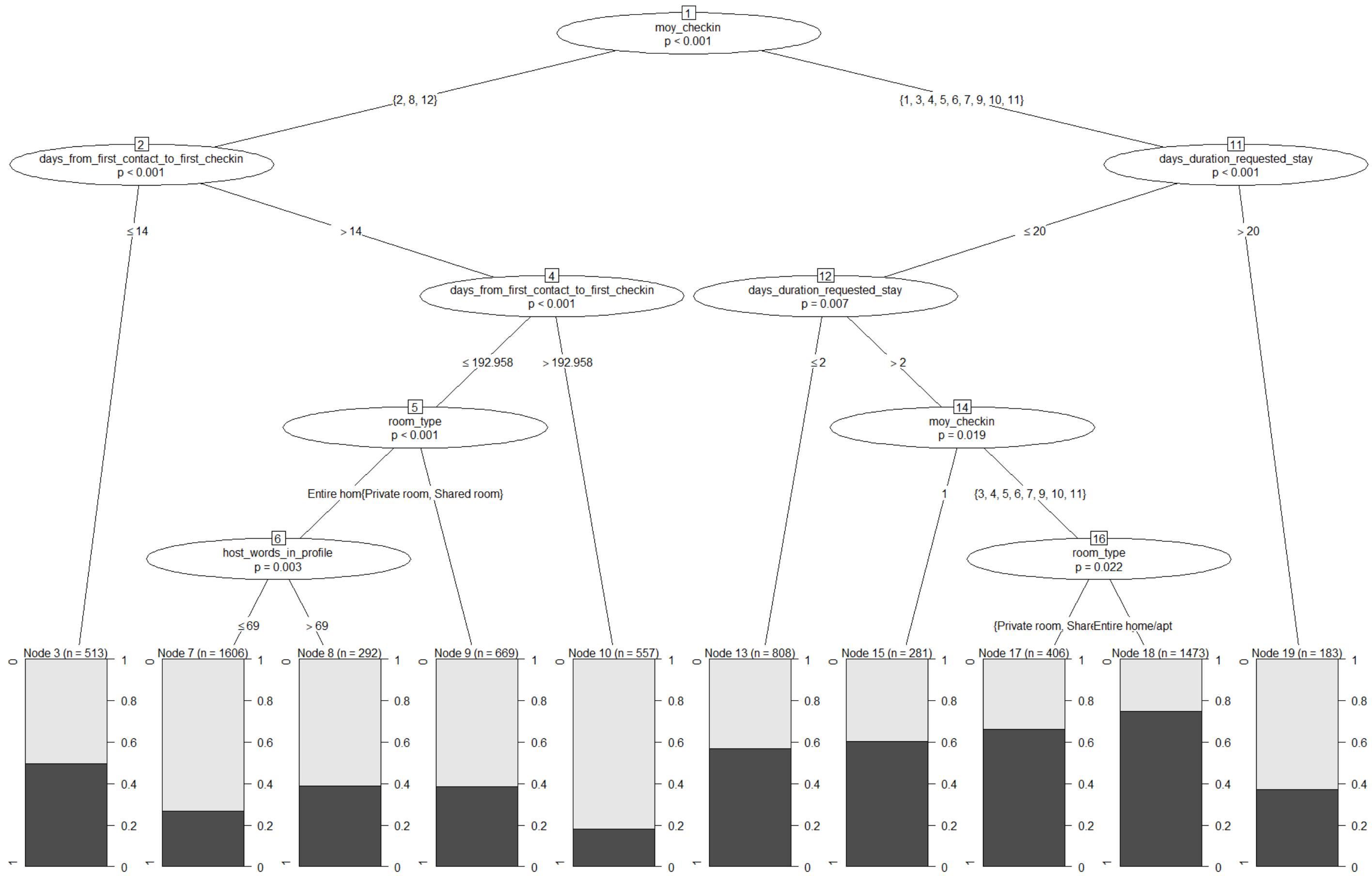
30
20
10
0
days_duration_requested_stay
0
1
Resulted_Successful_Booking
Resulted_Successful_Booking
0
1

Next, I have built classification models for inquiries from ‘**Book It’** channel:

<u>Logistic regression (resulting 95% overall accuracy – more details in code), and below plots are predicted probabilities of successful booking against individual predictors.</u>



<u>Next, decision tree plot:</u>


1
moy_checkin
p < 0.001
{2, 8, 12}
{1, 3, 4, 5, 6, 7, 9, 10, 11}
2
days_from_first_contact_to_first_checkin
p < 0.001
≤ 14
> 14
11
days_duration_requested_stay
p < 0.001
≤ 20
> 20
4
days_from_first_contact_to_first_checkin
p < 0.001
≤ 192.958
> 192.958
12
days_duration_requested_stay
p = 0.007
≤ 2
> 2
5
room_type
p < 0.001
Entire hom{Private room, Shared room}
14
moy_checkin
p = 0.019
1
{3, 4, 5, 6, 7, 9, 10, 11}
6
host_words_in_profile
p = 0.003
≤ 69
> 69
16
room_type
p = 0.022
{Private room, ShareEntire home/apt
Node 3 (n = 513)
Node 7 (n = 1606)
Node 8 (n = 292)
Node 9 (n = 669)
Node 10 (n = 557)
Node 13 (n = 808)
Node 15 (n = 281)
Node 17 (n = 406)
Node 18 (n = 1473)
Node 19 (n = 183)
0
1
1
0.8
0.6
0.4
0.2
0

And Finally, a random forest model, which has resulted in (84% accuracy), and here are the important features that influence the odd of an inquiry converting to a successful booking request: