EDA and Data Preprocessing Overview

Feature Engineering:

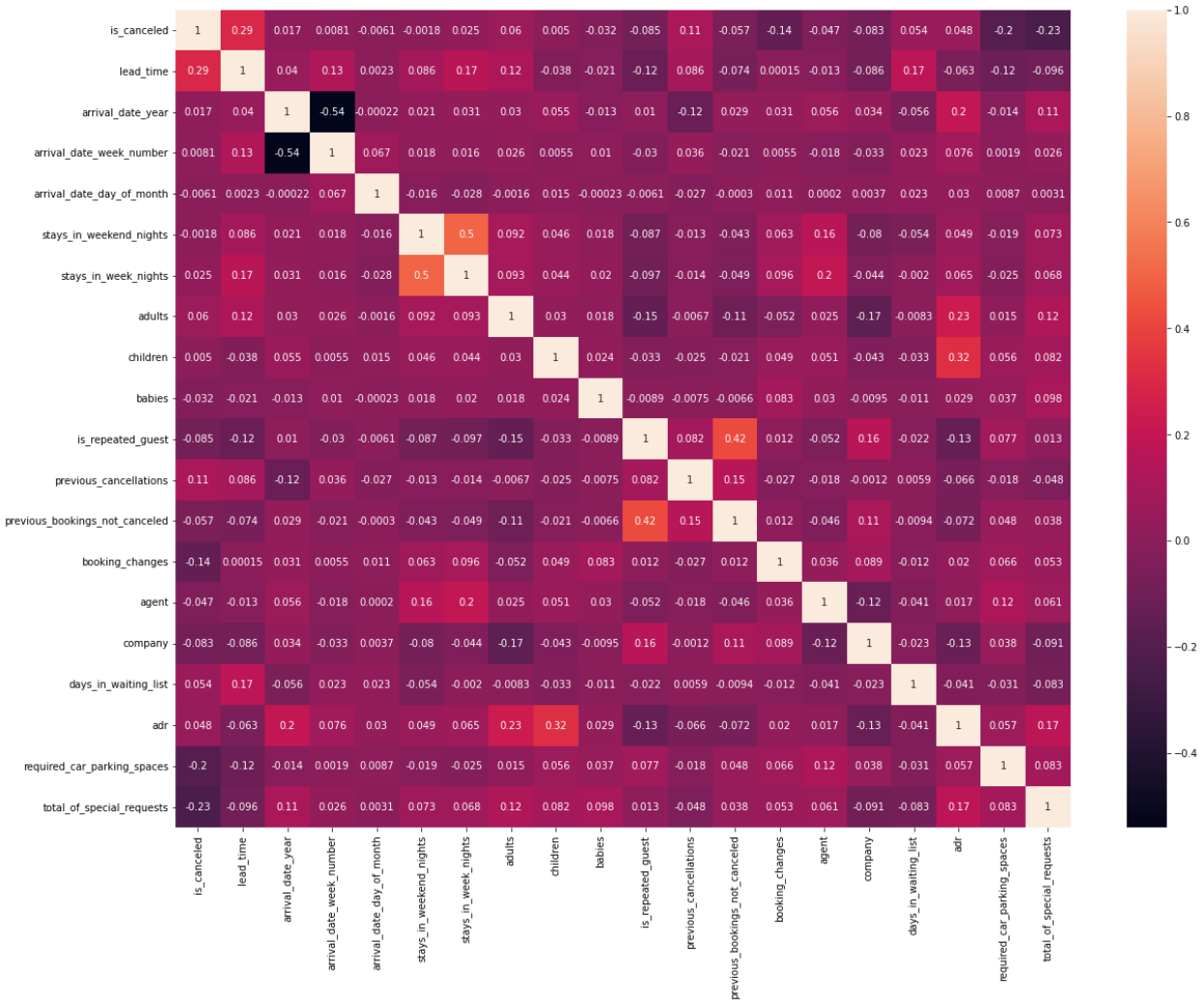
- Drop the duplicate rows .
- Drop the rows with column " adult=0 "
- Converting the datatype of **children** and **agent** from float to integer .
- Remove column **company**.
- Replace null values with 0 in **agent** feature.

Analyses:

In our model we want to predict the probability of the customer to cancel the booking or not.

I want to find out which factors are most important in *canclation* by find out the correlated value in according to "is_canceled" to set up a *LogisticRegression*, so I found out the following features :

```
total_of_special_requests         -23.465777
required_car_parking_spaces       -19.549782
booking_changes                   -14.438099
is_repeated_guest                  -8.479342
company                            -8.299480
previous_bookings_not_canceled     -5.735772
agent                              -4.652945
babies                             -3.249109
arrival_date_day_of_month          -0.613008
stays_in_weekend_nights            -0.179108
children                            0.503625
arrival_date_week_number            0.814807
arrival_date_year                   1.665986
stays_in_week_nights                2.476463
adr                                 4.755660
days_in_waiting_list                5.418582
adults                              6.001721
previous_cancellations             11.013281
lead_time                          29.312336
is_canceled                       100.000000
```


is_canceled 1 0.29 0.017 0.0081 -0.0061 -0.0018 0.025 0.06 0.005 -0.032 -0.085 0.11 -0.057 -0.14 -0.047 -0.083 0.054 0.048 -0.2 -0.23
lead_time 0.29 1 0.04 0.13 0.0023 0.086 0.17 0.12 -0.038 -0.021 -0.12 0.086 -0.074 0.00015 -0.013 -0.086 0.17 -0.063 -0.12 -0.096
arrival_date_year 0.017 0.04 1 -0.54 -0.00022 0.021 0.031 0.03 0.055 -0.013 0.01 -0.12 0.029 0.031 0.056 0.034 -0.056 0.2 -0.014 0.11
arrival_date_week_number 0.0081 0.13 -0.54 1 0.067 0.018 0.016 0.026 0.0055 0.01 -0.03 0.036 -0.021 0.0055 -0.018 -0.033 0.023 0.076 0.0019 0.026
arrival_date_day_of_month -0.0061 0.0023 -0.00022 0.067 1 -0.016 -0.028 -0.0016 0.015 -0.00023 -0.0061 -0.027 -0.0003 0.011 0.0002 0.0037 0.023 0.03 0.0087 0.0031
stays_in_weekend_nights -0.0018 0.086 0.021 0.018 -0.016 1 0.5 0.092 0.046 0.018 -0.087 -0.013 -0.043 0.063 0.16 -0.08 -0.054 0.049 -0.019 0.073
stays_in_week_nights 0.025 0.17 0.031 0.016 -0.028 0.5 1 0.093 0.044 0.02 -0.097 -0.014 -0.049 0.096 0.2 -0.044 -0.002 0.065 -0.025 0.068
adults 0.06 0.12 0.03 0.026 -0.0016 0.092 0.093 1 0.03 0.018 -0.15 -0.0067 -0.11 -0.052 0.025 -0.17 -0.0083 0.23 0.015 0.12
children 0.005 -0.038 0.055 0.0055 0.015 0.046 0.044 0.03 1 0.024 -0.033 -0.025 -0.021 0.049 0.051 -0.043 -0.033 0.32 0.056 0.082
babies -0.032 -0.021 -0.013 0.01 -0.00023 0.018 0.02 0.018 0.024 1 -0.0089 -0.0075 -0.0066 0.083 0.03 -0.0095 -0.011 0.029 0.037 0.098
is_repeated_guest -0.085 -0.12 0.01 -0.03 -0.0061 -0.087 -0.097 -0.15 -0.033 -0.0089 1 0.082 0.42 0.012 -0.052 0.16 -0.022 -0.13 0.077 0.013
previous_cancellations 0.11 0.086 -0.12 0.036 -0.027 -0.013 -0.014 -0.0067 -0.025 -0.0075 0.082 1 0.15 -0.027 -0.018 -0.0012 0.0059 -0.066 -0.018 -0.048
previous_bookings_not_canceled -0.057 -0.074 0.029 -0.021 -0.0003 -0.043 -0.049 -0.11 -0.021 -0.0066 0.42 0.15 1 0.012 -0.046 0.11 -0.0094 -0.072 0.048 0.038
booking_changes -0.14 0.00015 0.031 0.0055 0.011 0.063 0.096 -0.052 0.049 0.083 0.012 -0.027 0.012 1 0.036 0.089 -0.012 0.02 0.066 0.053
agent -0.047 -0.013 0.056 -0.018 0.0002 0.16 0.2 0.025 0.051 0.03 -0.052 -0.018 -0.046 0.036 1 -0.12 -0.041 0.017 0.12 0.061
company -0.083 -0.086 0.034 -0.033 0.0037 -0.08 -0.044 -0.17 -0.043 -0.0095 0.16 -0.0012 0.11 0.089 -0.12 1 -0.023 -0.13 0.038 -0.091
days_in_waiting_list 0.054 0.17 -0.056 0.023 0.023 -0.054 -0.002 -0.0083 -0.033 -0.011 -0.022 0.0059 -0.0094 -0.012 -0.041 -0.023 1 -0.041 -0.031 -0.083
adr 0.048 -0.063 0.2 0.076 0.03 0.049 0.065 0.23 0.32 0.029 -0.13 -0.066 -0.072 0.02 0.017 -0.13 -0.041 1 0.057 0.17
required_car_parking_spaces -0.2 -0.12 -0.014 0.0019 0.0087 -0.019 -0.025 0.015 0.056 0.037 0.077 -0.018 0.048 0.066 0.12 0.038 -0.031 0.057 1 0.083
total_of_special_requests -0.23 -0.096 0.11 0.026 0.0031 0.073 0.068 0.12 0.082 0.098 0.013 -0.048 0.038 0.053 0.061 -0.091 -0.083 0.17 0.083 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
is_repeated_guest
previous_cancellations
previous_bookings_not_canceled
booking_changes
agent
company
days_in_waiting_list
adr
required_car_parking_spaces
total_of_special_requests
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4