

```
In [1]: # read the hotel reservation file
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import plotly.express as px
import numpy as np
import arules
import arules as ar
from arules.utils import five_quantile_based_bins, top_bottom_10, top_5_variant_variable

hotel = pd.read_csv('Hotel Reservations.csv')
pd.options.display.max_colwidth = 100
```

```
In [2]: hotel['lead_time'] = hotel['lead_time'].astype('float64')
```

## Association rules for all attributes (support/confidence experiments)

```
In [3]: rules_3, supp_dict = ar.create_association_rules(hotel,max_cols=3,binning_method=five_qu

True
352
36275
```

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\utils\binning.py:88: F
utureWarning: The series.append method is deprecated and will be removed from pandas in
a future version. Use pandas.concat instead.
    attached_series = binned_series.append(na_series).reindex(series.index)
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\utils\binning.py:88: F
utureWarning: The series.append method is deprecated and will be removed from pandas in
a future version. Use pandas.concat instead.
    attached_series = binned_series.append(na_series).reindex(series.index)
Calculating all relevant supports
100%|██████████| 1159/1159 [01:10<00:00, 16.35it/s]

Calculating all feature level rules per variable level rule
100%|██████████| 6156/6156 [00:57<00:00, 106.20it/s]
Overall # of Rules: 82730
```

```
In [4]: rules_2, supp_dict = ar.create_association_rules(hotel,max_cols=2,binning_method=five_qu

True
352
36275
```

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\utils\binning.py:88: F
utureWarning: The series.append method is deprecated and will be removed from pandas in
a future version. Use pandas.concat instead.
    attached_series = binned_series.append(na_series).reindex(series.index)
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\utils\binning.py:88: F
utureWarning: The series.append method is deprecated and will be removed from pandas in
a future version. Use pandas.concat instead.
    attached_series = binned_series.append(na_series).reindex(series.index)
Calculating all relevant supports
100%|██████████| 190/190 [00:02<00:00, 68.76it/s]

Calculating all feature level rules per variable level rule
100%|██████████| 342/342 [00:01<00:00, 234.63it/s]
Overall # of Rules: 3470
```

# Select only the rules with confidence > 0.7 and support > 0.05

## For rules with 1 antecedent

```
In [5]: rules_2[rules_2['confidence'] > 0.7][rules_2['rule_supp'] > 0.05].sort_values(by='confid
```

```
C:\Users\cezar\AppData\Local\Temp\ipykernel_9916\2089929830.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.  
rules_2[rules_2['confidence'] > 0.7][rules_2['rule_supp'] > 0.05].sort_values(by='confidence', ascending=False)
```

```
Out[5]:
```

	antecedent	consequent	rule print	ant_col
--	------------	------------	------------	---------

438	{'arrival_month': '3'}	{'arrival_year': '2018'}	arrival_month=3 ==> arrival_year=2018	23
439	{'arrival_month': '4'}	{'arrival_year': '2018'}	arrival_month=4 ==> arrival_year=2018	27
1636	{'repeated_guest': '0'}	{'no_of_previous_cancellations': '0'}	repeated_guest=0 ==> no_of_previous_cancellations=0	353
440	{'arrival_month': '5'}	{'arrival_year': '2018'}	arrival_month=5 ==> arrival_year=2018	25
1374	{'repeated_guest': '0'}	{'no_of_previous_bookings_not_canceled': '0'}	repeated_guest=0 ==> no_of_previous_bookings_not_canceled=0	353
...	...	...	...	...
2946	{'lead_time': '(39.0 - 80.0]'}	{'room_type_reserved': 'Room_Type 1'}	lead_time=(39.0 - 80.0] ==> room_type_reserved=Room_Type 1	71
2472	{'arrival_month': '10'}	{'no_of_adults': '2'}	arrival_month=10 ==> no_of_adults=2	53
1170	{'market_segment_type': 'Offline'}	{'booking_status': 'Not_Canceled'}	market_segment_type=Offline ==> booking_status=Not_Canceled	105
2924	{'arrival_month': '7'}	{'type_of_meal_plan': 'Meal Plan 1'}	arrival_month=7 ==> type_of_meal_plan=Meal Plan 1	29
2974	{'no_of_week_nights': '4'}	{'room_type_reserved': 'Room_Type 1'}	no_of_week_nights=4 ==> room_type_reserved=Room_Type 1	29

446 rows × 12 columns

## For rules with 2 antecedents

```
In [6]: rules_3[rules_3['confidence'] > 0.7][rules_3['rule_supp'] > 0.05].sort_values(by='confid
```

```
C:\Users\cezar\AppData\Local\Temp\ipykernel_9916\2973897398.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.  
rules_3[rules_3['confidence'] > 0.7][rules_3['rule_supp'] > 0.05].sort_values(by='confidence', ascending=False)
```

```
Out[6]:
```

	antecedent	consequent	rule print
--	------------	------------	------------

44185	{'lead_time': '(11.0 - 39.0]', 'repeated_guest': '0'}	{'no_of_previous_cancellations': '0'}	lead_time=(11.0 - 39.0], repeated_guest=0 ==> no_of_previous_cancellations=0
44149	{'room_type_reserved': 'Room_Type 1', 'repeated_guest': '0'}	{'no_of_previous_cancellations': '0'}	room_type_reserved=Room_Type 1, repeated_guest=0 ==> no_of_previous_cancellations=0
44079	{'no_of_week_nights': '2', 'lead_time': '(39.0 - 80.0]'}	{'no_of_previous_cancellations': '0'}	no_of_week_nights=2, lead_time=(39.0 - 80.0] ==>

			no_of_previous_cancellations=0
44083	{'no_of_week_nights': '4', 'arrival_year': '2018'}	{'no_of_previous_cancellations': '0'}	no_of_week_nights=4, arrival_year=2018 ==> no_of_previous_cancellations=0
44093	{'no_of_week_nights': '0', 'repeated_guest': '0'}	{'no_of_previous_cancellations': '0'}	no_of_week_nights=0, repeated_guest=0 ==> no_of_previous_cancellations=0
...	...	...	...
51002	{'required_car_parking_space': '0'}	{'no_of_children': '0', 'type_of_meal_plan': 'Meal Plan 1'}	required_car_parking_space=0 ==> no_of_children=0, type_of_meal_plan=Meal Plan 1
70178	{'no_of_week_nights': '4', 'no_of_previous_cancellations': '0'}	{'room_type_reserved': 'Room_Type 1'}	no_of_week_nights=4, no_of_previous_cancellations=0 ==> room_type_reserved=Room_Type 1
70201	{'no_of_week_nights': '4', 'no_of_previous_bookings_not_canceled': '0'}	{'room_type_reserved': 'Room_Type 1'}	no_of_week_nights=4, no_of_previous_bookings_not_canceled=0 ==> room_type_reserved=Room_Type 1
70200	{'no_of_week_nights': '4', 'repeated_guest': '0'}	{'room_type_reserved': 'Room_Type 1'}	no_of_week_nights=4, repeated_guest=0 ==> room_type_reserved=Room_Type 1
32455	{'market_segment_type': 'Offline', 'no_of_previous_bookings_not_canceled': '0'}	{'booking_status': 'Not_Canceled'}	market_segment_type=Offline, no_of_previous_bookings_not_canceled=0 ==> booking_status=Not_Canceled

6473 rows × 12 columns

## Association rules for target attribute (booking\_status)

### For rules with 1 antecedent

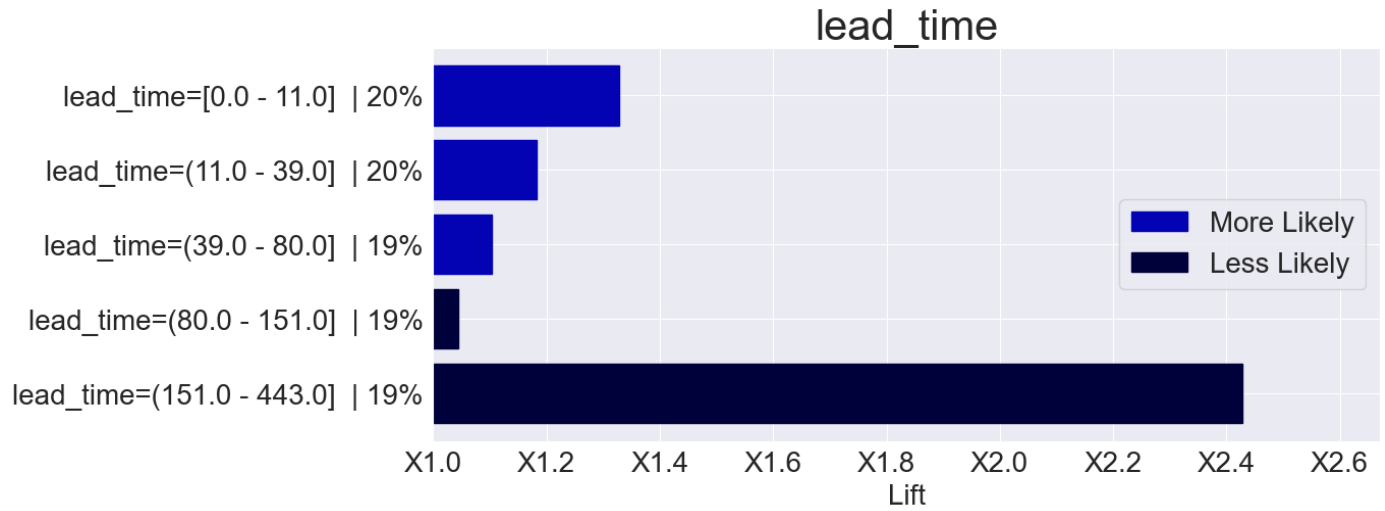
```
In [7]: ar.present_rules_per_consequent(rules_2,consequent={'booking_status':'Canceled', 'bookin
plot=True)
```

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\numpy\core\fromnumeric.py:357
1: FutureWarning: The default value of numeric_only in DataFrame.std is deprecated. In a
future version, it will default to False. In addition, specifying 'numeric_only=None' is
deprecated. Select only valid columns or specify the value of numeric_only to silence th
is warning.
```

```
return std(axis=axis, dtype=dtype, out=out, ddof=ddof, **kwargs)
lead_time
```

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
5	lead_time=[0.0 - 11.0] ==> booking_status=Not_Canceled	0.200055	0.672364	0.178718	0.8933	1.3287
11	lead_time=(11.0 - 39.0] ==> booking_status=Not_Canceled	0.207471	0.672364	0.165100	0.7958	1.1835
16	lead_time=(39.0 - 80.0] ==> booking_status=Not_Canceled	0.196720	0.672364	0.145996	0.7422	1.1038
64	lead_time=(80.0 - 151.0] ==> booking_status=Not_Canceled	0.198704	0.672364	0.127994	0.6441	0.9580
90	lead_time=(151.0 - 443.0] ==> booking_status=Not_Canceled	0.197050	0.672364	0.054555	0.2769	0.4118

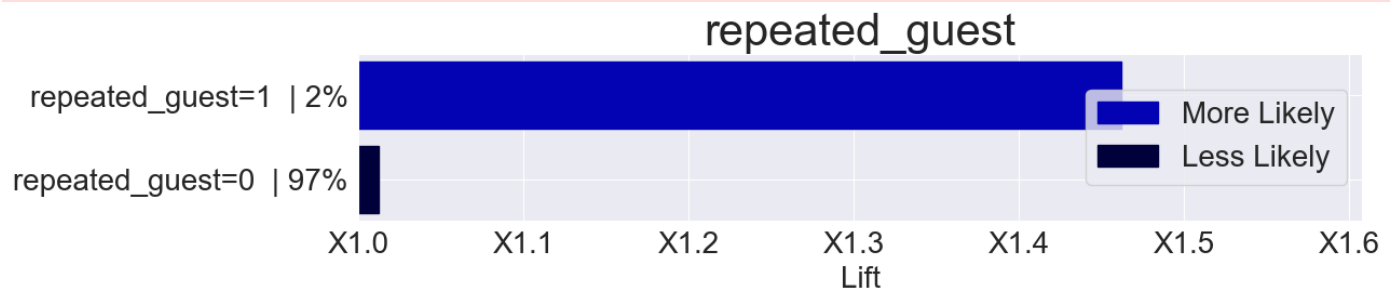
```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:4
09: UserWarning: FixedFormatter should only be used together with FixedLocator
axes.set_xticklabels(labs_x)
```



repeated\_guest

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
2	repeated_guest=1 ==> booking_status=Not_Canceled	0.025637	0.672364	0.025196	0.9828	1.4617
51	repeated_guest=0 ==> booking_status=Not_Canceled	0.974363	0.672364	0.647167	0.6642	0.9879

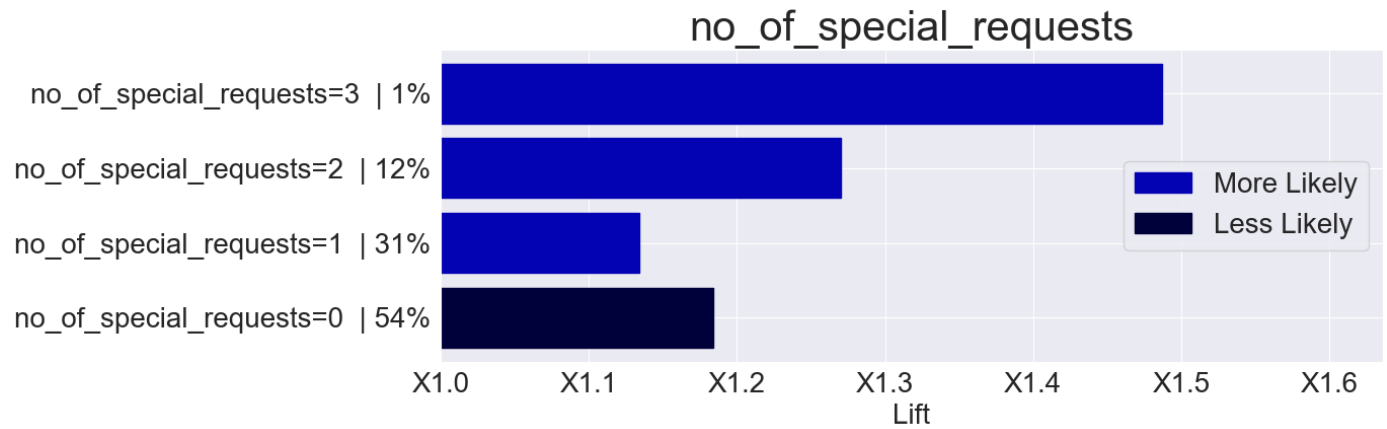
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association\_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator  
 axes.set\_xticklabels(labs\_x)



no\_of\_special\_requests

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
1	no_of_special_requests=3 ==> booking_status=Not_Canceled	0.018608	0.672364	0.018608	1.0000	1.4873
8	no_of_special_requests=2 ==> booking_status=Not_Canceled	0.120303	0.672364	0.102743	0.8540	1.2702
13	no_of_special_requests=1 ==> booking_status=Not_Canceled	0.313522	0.672364	0.239008	0.7623	1.1338
86	no_of_special_requests=0 ==> booking_status=Not_Canceled	0.545196	0.672364	0.309635	0.5679	0.8447

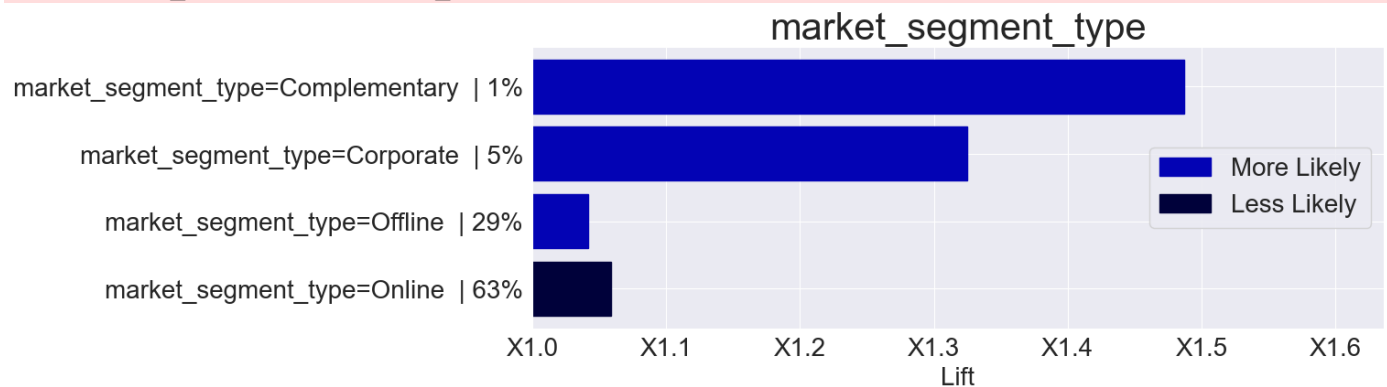
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association\_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator  
 axes.set\_xticklabels(labs\_x)



market\_segment\_type

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
0	market_segment_type=Complementary ==> booking_status=Not_Canceled	0.010779	0.672364	0.010779	1.0000	1.4873
6	market_segment_type=Corporate ==> booking_status=Not_Canceled	0.055603	0.672364	0.049538	0.8909	1.3251
30	market_segment_type=Offline ==> booking_status=Not_Canceled	0.290227	0.672364	0.203308	0.7005	1.0419
70	market_segment_type=Online ==> booking_status=Not_Canceled	0.639945	0.672364	0.406313	0.6349	0.9443

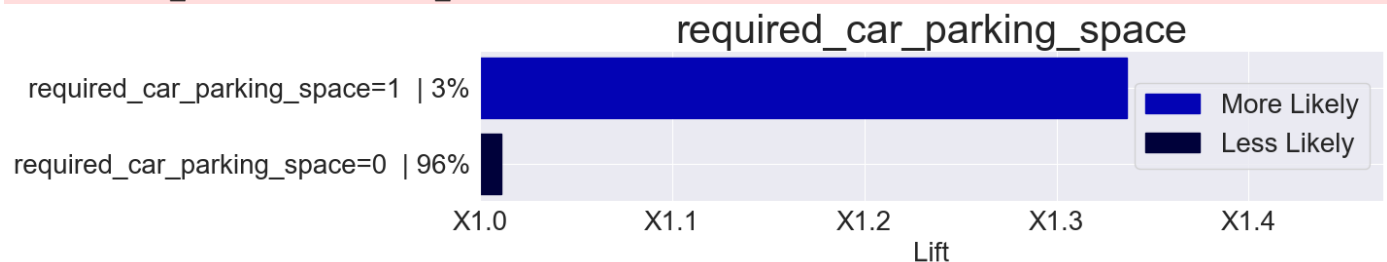
```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator
  axes.set_xticklabels(labs_x)
```



required\_car\_parking\_space

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
4	required_car_parking_space=1 ==> booking_status=Not_Canceled	0.030986	0.672364	0.027843	0.8986	1.3364
49	required_car_parking_space=0 ==> booking_status=Not_Canceled	0.969014	0.672364	0.644521	0.6651	0.9892

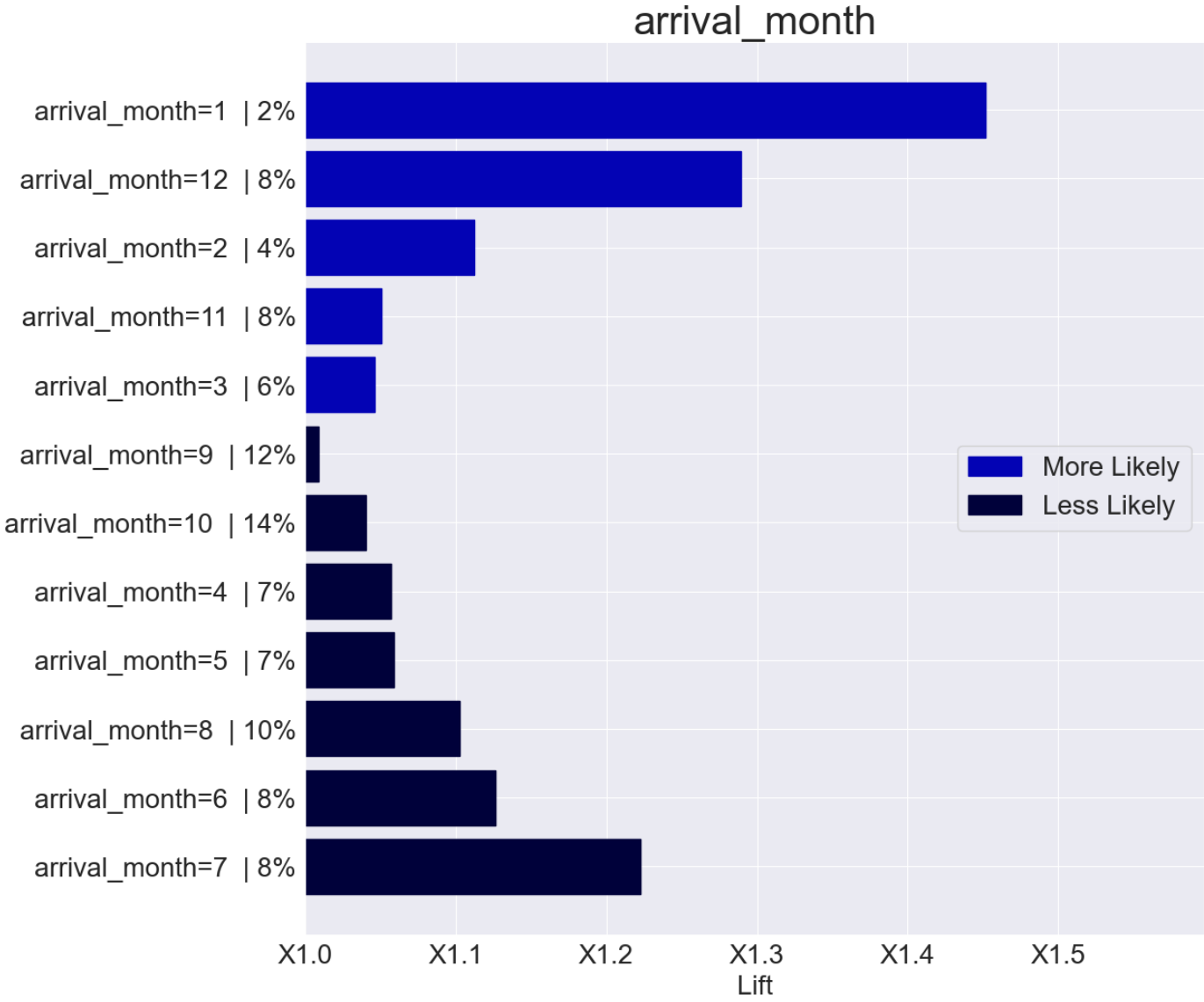
```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator
  axes.set_xticklabels(labs_x)
```



arrival\_month

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
3	arrival_month=1 ==> booking_status=Not_Canceled	0.027953	0.672364	0.027292	0.9763	1.4521
7	arrival_month=12 ==> booking_status=Not_Canceled	0.083280	0.672364	0.072198	0.8669	1.2894
15	arrival_month=2 ==> booking_status=Not_Canceled	0.046975	0.672364	0.035121	0.7477	1.1120
26	arrival_month=11 ==> booking_status=Not_Canceled	0.082150	0.672364	0.058029	0.7064	1.0506
28	arrival_month=3 ==> booking_status=Not_Canceled	0.065003	0.672364	0.045706	0.7031	1.0458
47	arrival_month=9 ==> booking_status=Not_Canceled	0.127112	0.672364	0.084714	0.6664	0.9912
63	arrival_month=10 ==> booking_status=Not_Canceled	0.146575	0.672364	0.094748	0.6464	0.9614
68	arrival_month=4 ==> booking_status=Not_Canceled	0.075424	0.672364	0.047994	0.6363	0.9464
69	arrival_month=5 ==> booking_status=Not_Canceled	0.071620	0.672364	0.045486	0.6351	0.9446
79	arrival_month=8 ==> booking_status=Not_Canceled	0.105114	0.672364	0.064094	0.6098	0.9069
81	arrival_month=6 ==> booking_status=Not_Canceled	0.088298	0.672364	0.052708	0.5969	0.8878
88	arrival_month=7 ==> booking_status=Not_Canceled	0.080496	0.672364	0.044273	0.5500	0.8180

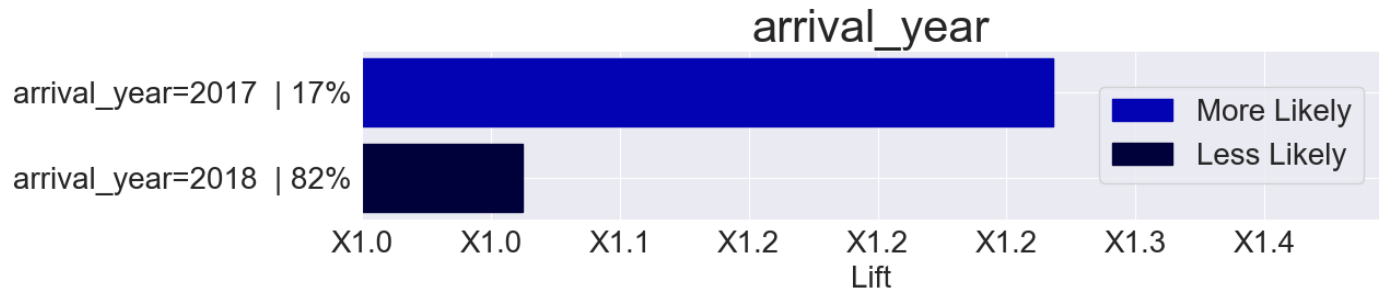
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association\_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator  
axes.set\_xticklabels(labs\_x)



arrival\_year

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
9	arrival_year=2017 ==> booking_status=Not_Canceled	0.179573	0.672364	0.153081	0.8525	1.2679
72	arrival_year=2018 ==> booking_status=Not_Canceled	0.820427	0.672364	0.519283	0.6329	0.9414

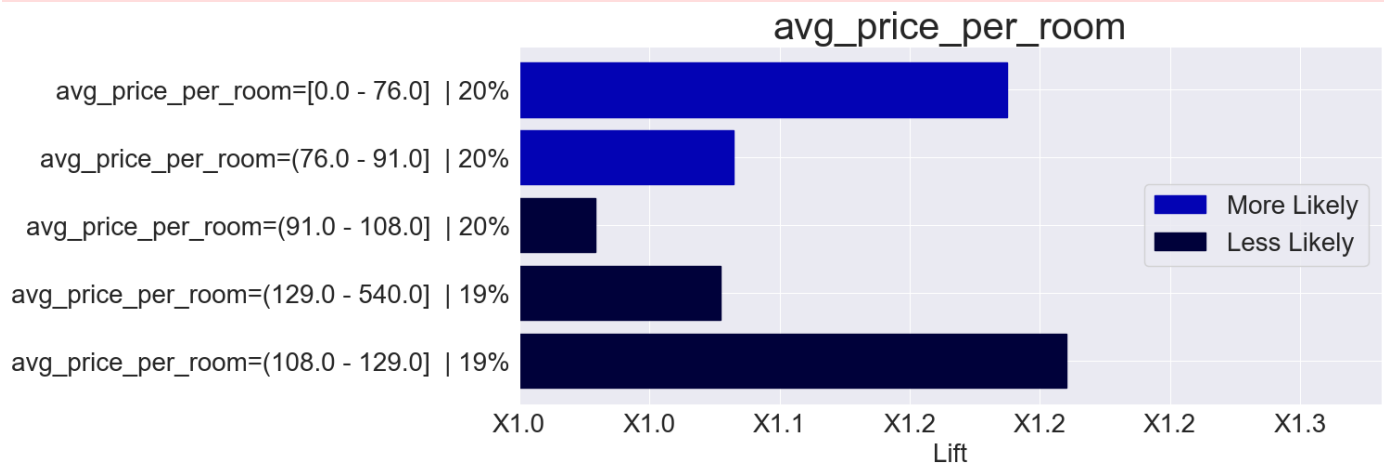
```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator
  axes.set_xticklabels(labs_x)
```



avg\_price\_per\_room

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
10	avg_price_per_room=[0.0 - 76.0] ==> booking_status=Not_Canceled	0.200992	0.672364	0.160469	0.7984	1.1874
20	avg_price_per_room=(76.0 - 91.0] ==> booking_status=Not_Canceled	0.201296	0.672364	0.146492	0.7277	1.0824
58	avg_price_per_room=(91.0 - 108.0] ==> booking_status=Not_Canceled	0.203391	0.672364	0.132874	0.6533	0.9716
75	avg_price_per_room=(129.0 - 540.0] ==> booking_status=Not_Canceled	0.195975	0.672364	0.122316	0.6241	0.9283
87	avg_price_per_room=(108.0 - 129.0] ==> booking_status=Not_Canceled	0.198346	0.672364	0.110214	0.5557	0.8264

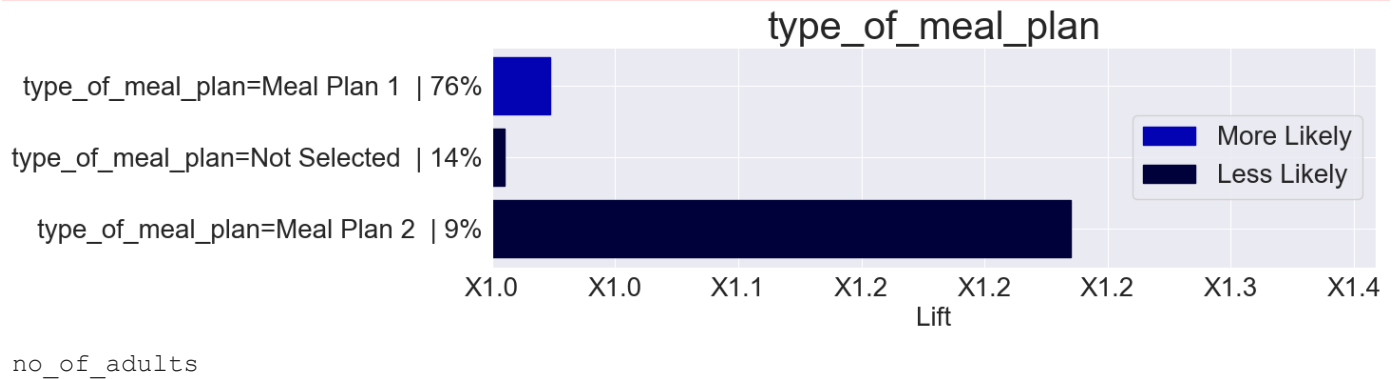
```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator
  axes.set_xticklabels(labs_x)
```



type\_of\_meal\_plan

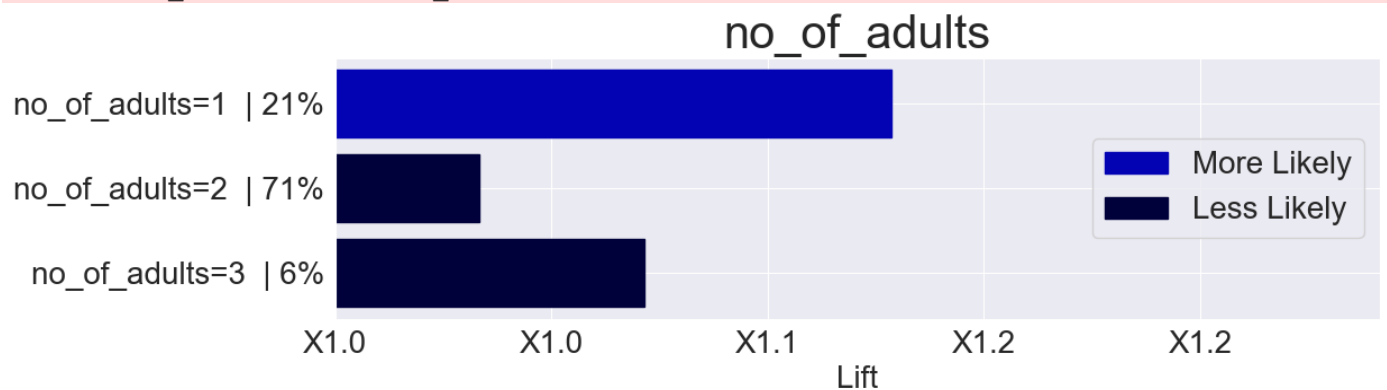
	rule print	ant_supp	con_supp	rule_supp	confidence	lift
36	type_of_meal_plan=Meal Plan 1 ==> booking_status=Not_Canceled	0.767333	0.672364	0.528077	0.6882	1.0236
45	type_of_meal_plan=Not Selected ==> booking_status=Not_Canceled	0.141420	0.672364	0.094583	0.6688	0.9947
89	type_of_meal_plan=Meal Plan 2 ==> booking_status=Not_Canceled	0.091110	0.672364	0.049593	0.5443	0.8096

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:4
09: UserWarning: FixedFormatter should only be used together with FixedLocator
axes.set_xticklabels(labs_x)
```



	rule print	ant_supp	con_supp	rule_supp	confidence	lift
14	no_of_adults=1 ==> booking_status=Not_Canceled	0.212130	0.672364	0.160965	0.7588	1.1286
61	no_of_adults=2 ==> booking_status=Not_Canceled	0.719724	0.672364	0.468339	0.6507	0.9678
74	no_of_adults=3 ==> booking_status=Not_Canceled	0.063873	0.672364	0.040083	0.6275	0.9333

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:4
09: UserWarning: FixedFormatter should only be used together with FixedLocator
axes.set_xticklabels(labs_x)
```



## Interpretation for rules with 1 antecedent

We noticed that the booking status is influenced by the following attributes:

- Lead time : the longer the lead time, the more likely the booking is canceled
- Arrival month : the booking is more likely to be canceled in the months of July, August and June
- Average price per room : we can notice that the booking is more likely to be canceled when the average price per room is between 108 and 540
- Market segment type : the booking is more likely not to be canceled when the market segment type is Complementary and Aviation
- Number of special requests : the booking is more likely not to be canceled when the number of special requests is greater than 0

Regarding the confidence of each rule:

- For lead time column we have 3 rules with confidence > 0.7 that cover around 60% of the possible values
- For repeated guest column we have a rule with confidence = 0.98, but with low coverage, with around 2% of the possible values



- For number of special requests column we have 3 rules with confidence > 0.7 that cover around 46% of the possible values
- For market\_segment\_type column we have 3 rules with confidence > 0.7 that cover around 35% of the possible values
- For required\_car\_parking\_space column we have 1 rule with confidence = 0.89 that covers only around 3% of the possible values
- For arrival\_month column we have 5 rules with confidence > 0.7 that cover around 28% of the possible values
- We consider that arrival year does not supply enough information to be considered as an important attribute
- For average price per room column we have 2 rules with confidence > 0.7 that cover around 40% of the possible values
- For type of meal plan column we have rules with small confidence, so more analysis has to be done to determine if this column is important or not
- For number of adults column we 1 rules with confidence > 0.7 that covers around 21% of the possible values

## For rules with 2 antecedents

```
In [8]: pd.options.display.max_colwidth = 100
ar.present_rules_per_consequent(rules_3,consequent={'booking_status':'Canceled', 'bookin
plot=True)
```

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\numpy\core\fromnumeric.py:357
1: FutureWarning: The default value of numeric_only in DataFrame.std is deprecated. In a
future version, it will default to False. In addition, specifying 'numeric_only=None' is
deprecated. Select only valid columns or specify the value of numeric_only to silence th
is warning.
```

```
return std(axis=axis, dtype=dtype, out=out, ddof=ddof, **kwargs)
lead_time , market_segment_type
```

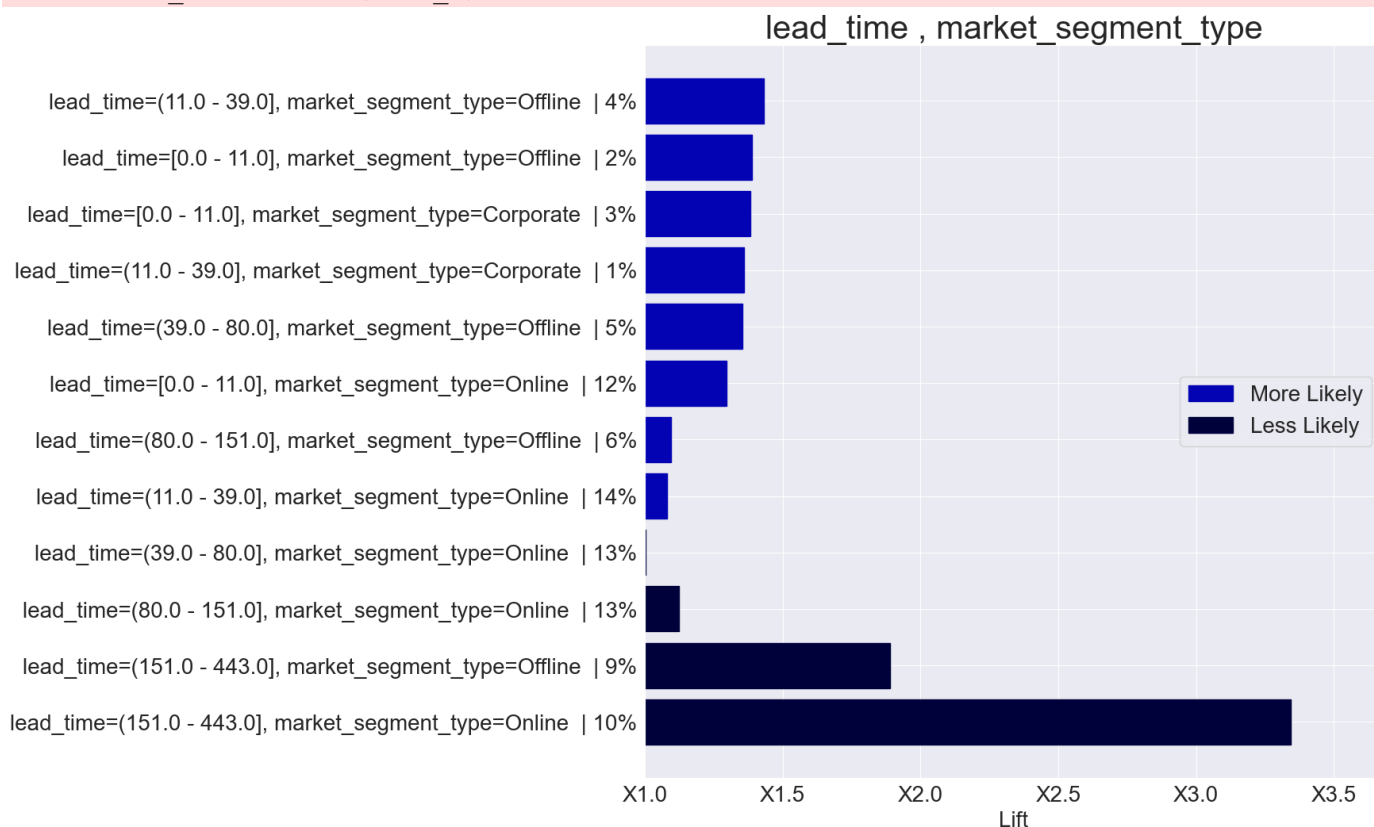
	rule print	ant_supp	con_supp	rule_supp	confidence	lift
<b>53</b>	lead_time=(11.0 - 39.0], market_segment_type=Offline ==> booking_status=Not_Canceled	0.046837	0.672364	0.045100	0.9629	1.4321
<b>67</b>	lead_time=[0.0 - 11.0], market_segment_type=Offline ==> booking_status=Not_Canceled	0.028091	0.672364	0.026216	0.9333	1.3880
<b>68</b>	lead_time=[0.0 - 11.0], market_segment_type=Corporate ==> booking_status=Not_Canceled	0.031757	0.672364	0.029552	0.9306	1.3840
<b>80</b>	lead_time=(11.0 - 39.0], market_segment_type=Corporate ==> booking_status=Not_Canceled	0.015190	0.672364	0.013894	0.9147	1.3604
<b>85</b>	lead_time=(39.0 - 80.0], market_segment_type=Offline ==> booking_status=Not_Canceled	0.057119	0.672364	0.051964	0.9097	1.3531
<b>149</b>	lead_time=[0.0 - 11.0], market_segment_type=Online ==> booking_status=Not_Canceled	0.128298	0.672364	0.111923	0.8724	1.2975
<b>412</b>	lead_time=(80.0 - 151.0], market_segment_type=Offline ==> booking_status=Not_Canceled	0.062302	0.672364	0.045899	0.7367	1.0957
<b>449</b>	lead_time=(11.0 - 39.0], market_segment_type=Online ==> booking_status=Not_Canceled	0.143901	0.672364	0.104700	0.7276	1.0821
<b>744</b>	lead_time=(39.0 - 80.0], market_segment_type=Online ==> booking_status=Not_Canceled	0.135603	0.672364	0.090779	0.6694	0.9957
<b>1232</b>	lead_time=(80.0 - 151.0], market_segment_type=Online	0.131661	0.672364	0.078704	0.5978	0.8891

==> booking\_status=Not\_Canceled

**1420**    lead\_time=(151.0 - 443.0], market\_segment\_type=Offline  
         ==> booking\_status=Not\_Canceled    0.095879    0.672364    0.034128    0.3560    0.5294

**1440**    lead\_time=(151.0 - 443.0], market\_segment\_type=Online  
         ==> booking\_status=Not\_Canceled    0.100482    0.672364    0.020207    0.2011    0.2991

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator
  axes.set_xticklabels(labs_x)
```

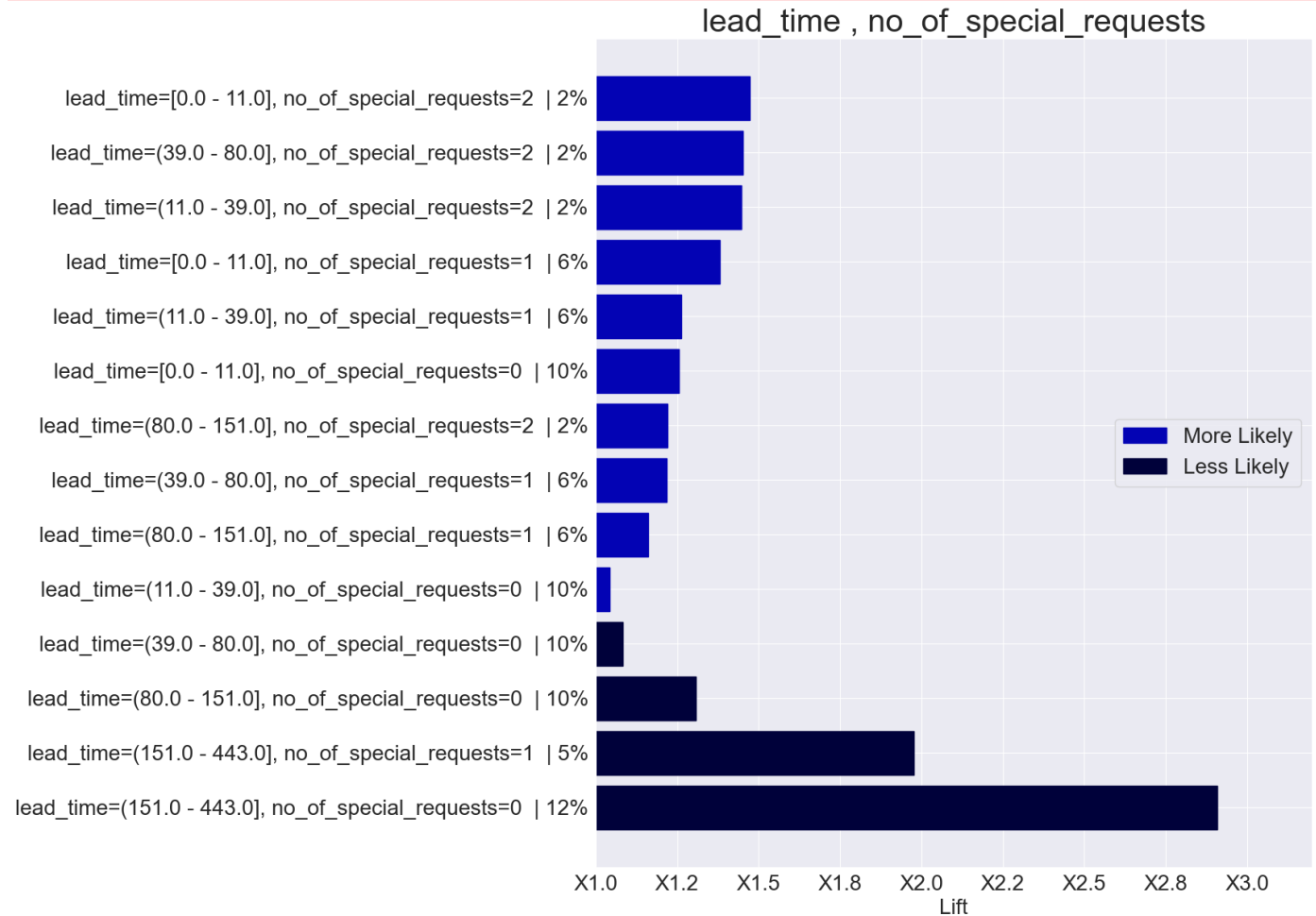


lead\_time , no\_of\_special\_requests

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
22	lead_time=[0.0 - 11.0], no_of_special_requests=2 ==> booking_status=Not_Canceled	0.024700	0.672364	0.024452	0.9900	1.4724
38	lead_time=(39.0 - 80.0], no_of_special_requests=2 ==> booking_status=Not_Canceled	0.025610	0.672364	0.025003	0.9763	1.4521
47	lead_time=(11.0 - 39.0], no_of_special_requests=2 ==> booking_status=Not_Canceled	0.029001	0.672364	0.028229	0.9734	1.4477
72	lead_time=[0.0 - 11.0], no_of_special_requests=1 ==> booking_status=Not_Canceled	0.066051	0.672364	0.061337	0.9286	1.3811
196	lead_time=(11.0 - 39.0], no_of_special_requests=1 ==> booking_status=Not_Canceled	0.069662	0.672364	0.059159	0.8492	1.2630
205	lead_time=[0.0 - 11.0], no_of_special_requests=0 ==> booking_status=Not_Canceled	0.104976	0.672364	0.088601	0.8440	1.2553
244	lead_time=(80.0 - 151.0], no_of_special_requests=2 ==> booking_status=Not_Canceled	0.024259	0.672364	0.019904	0.8205	1.2203
247	lead_time=(39.0 - 80.0], no_of_special_requests=1 ==> booking_status=Not_Canceled	0.064810	0.672364	0.053039	0.8184	1.2172
306	lead_time=(80.0 - 151.0], no_of_special_requests=1 ==> booking_status=Not_Canceled	0.061392	0.672364	0.047912	0.7804	1.1607

<b>553</b>	lead_time=(11.0 - 39.0], no_of_special_requests=0 ==> booking_status=Not_Canceled	0.104176	0.672364	0.073081	0.7015	1.0433
<b>1131</b>	lead_time=(39.0 - 80.0], no_of_special_requests=0 ==> booking_status=Not_Canceled	0.101172	0.672364	0.062826	0.6210	0.9236
<b>1390</b>	lead_time=(80.0 - 151.0], no_of_special_requests=0 ==> booking_status=Not_Canceled	0.108863	0.672364	0.055989	0.5143	0.7649
<b>1422</b>	lead_time=(151.0 - 443.0], no_of_special_requests=1 ==> booking_status=Not_Canceled	0.051606	0.672364	0.017560	0.3403	0.5061
<b>1438</b>	lead_time=(151.0 - 443.0], no_of_special_requests=0 ==> booking_status=Not_Canceled	0.126010	0.672364	0.029139	0.2312	0.3439

D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association\_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator  
 axes.set\_xticklabels(labs\_x)

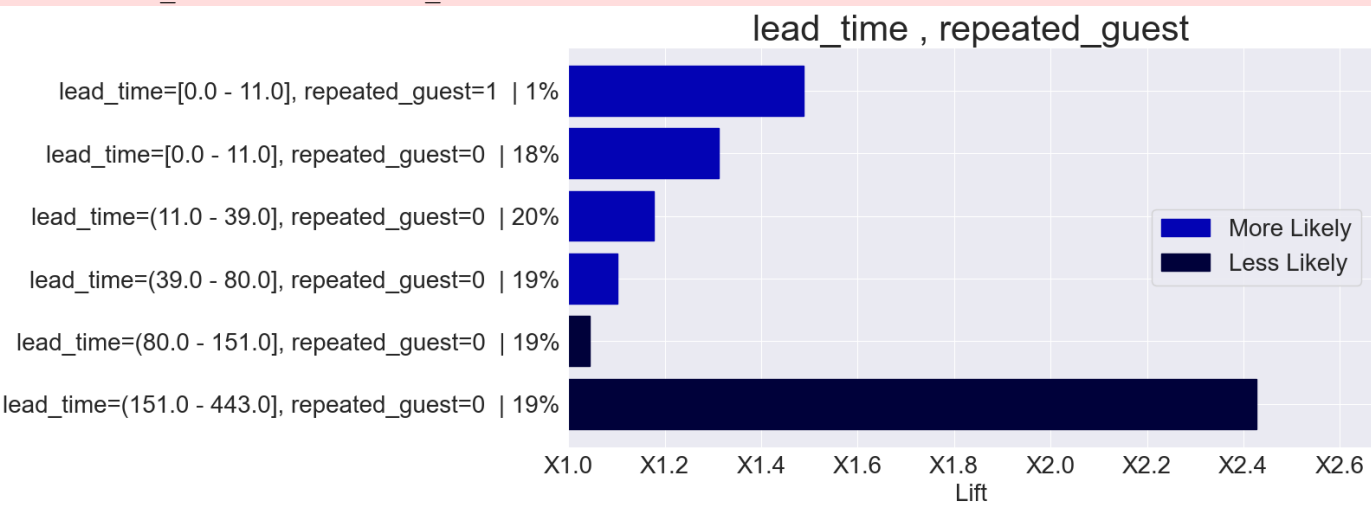


lead\_time , repeated\_guest

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
<b>10</b>	lead_time=[0.0 - 11.0], repeated_guest=1 ==> booking_status=Not_Canceled	0.019325	0.672364	0.019325	1.0000	1.4873
<b>138</b>	lead_time=[0.0 - 11.0], repeated_guest=0 ==> booking_status=Not_Canceled	0.180731	0.672364	0.159394	0.8819	1.3117
<b>288</b>	lead_time=(11.0 - 39.0], repeated_guest=0 ==> booking_status=Not_Canceled	0.202839	0.672364	0.160496	0.7912	1.1768
<b>396</b>	lead_time=(39.0 - 80.0], repeated_guest=0 ==> booking_status=Not_Canceled	0.195865	0.672364	0.145224	0.7414	1.1028
<b>964</b>	lead_time=(80.0 - 151.0], repeated_guest=0 ==> booking_status=Not_Canceled	0.198263	0.672364	0.127581	0.6435	0.9571

1431 lead\_time=(151.0 - 443.0], repeated\_guest=0 ==> 0.196664 0.672364 0.054473 0.2770 0.4120  
booking\_status=Not\_Canceled

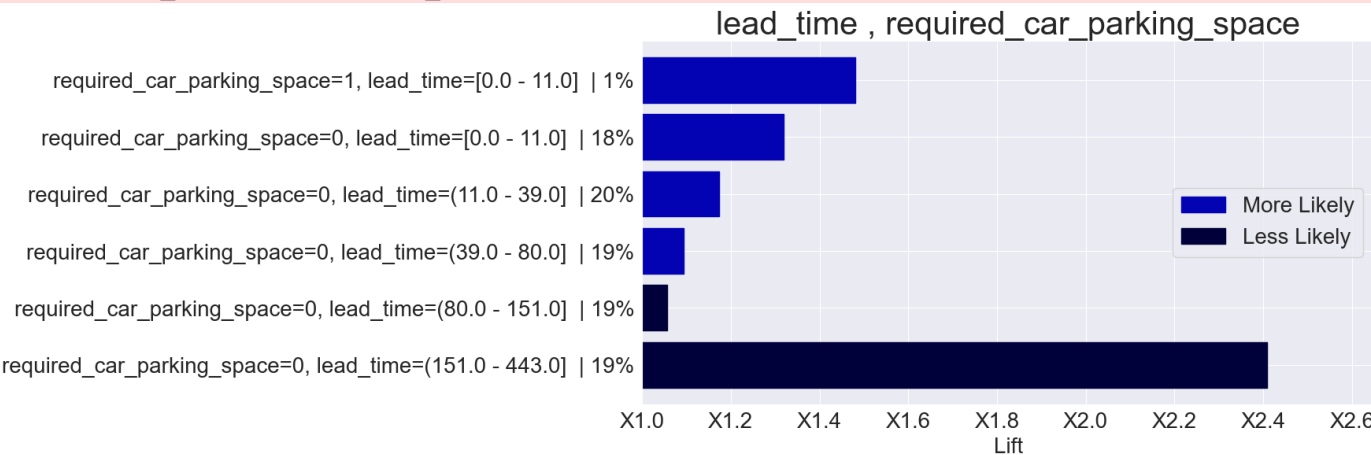
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association\_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator  
axes.set\_xticklabels(labs\_x)



lead\_time , required\_car\_parking\_space

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
19	required_car_parking_space=1, lead_time=[0.0 - 11.0] ==> booking_status=Not_Canceled	0.011992	0.672364	0.011937	0.9954	1.4805
124	required_car_parking_space=0, lead_time=[0.0 - 11.0] ==> booking_status=Not_Canceled	0.188063	0.672364	0.166782	0.8868	1.3190
293	required_car_parking_space=0, lead_time=(11.0 - 39.0] ==> booking_status=Not_Canceled	0.201323	0.672364	0.158980	0.7897	1.1745
420	required_car_parking_space=0, lead_time=(39.0 - 80.0] ==> booking_status=Not_Canceled	0.191868	0.672364	0.141144	0.7356	1.0941
1014	required_car_parking_space=0, lead_time=(80.0 - 151.0] ==> booking_status=Not_Canceled	0.194183	0.672364	0.123584	0.6364	0.9466
1429	required_car_parking_space=0, lead_time=(151.0 - 443.0] ==> booking_status=Not_Canceled	0.193577	0.672364	0.054032	0.2791	0.4151

D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association\_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator  
axes.set\_xticklabels(labs\_x)

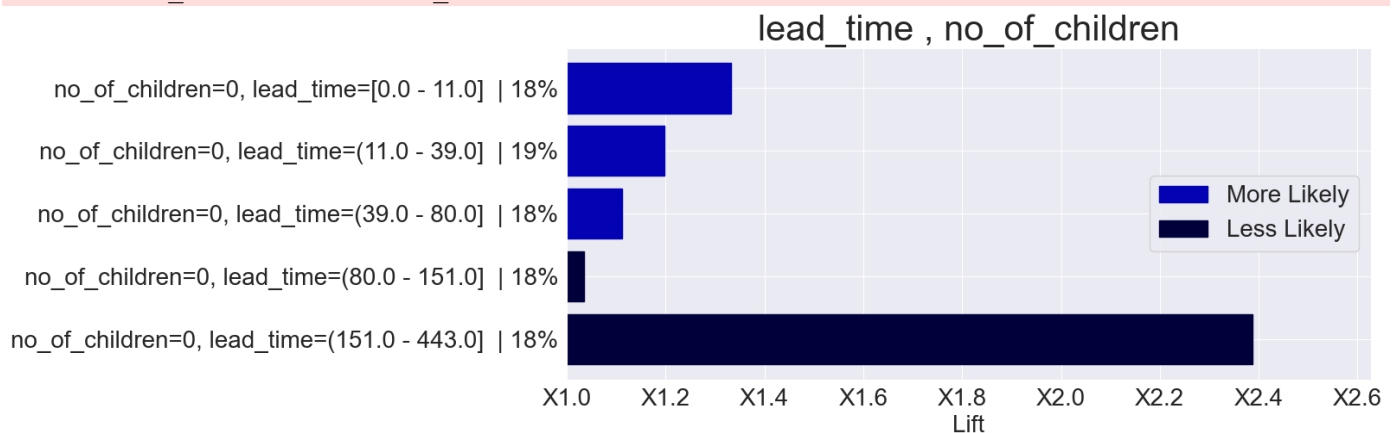


lead\_time , no\_of\_children

	rule print	ant_supp	con_supp	rule_supp	confidence	lift
106	no_of_children=0, lead_time=[0.0 - 11.0] ==> booking_status=Not_Canceled	0.184948	0.672364	0.165734	0.8961	1.3328

<b>262</b>	no_of_children=0, lead_time=(11.0 - 39.0] ==> booking_status=Not_Canceled	0.190958	0.672364	0.153715	0.8050	1.1972
<b>378</b>	no_of_children=0, lead_time=(39.0 - 80.0] ==> booking_status=Not_Canceled	0.180648	0.672364	0.135162	0.7482	1.1128
<b>913</b>	no_of_children=0, lead_time=(80.0 - 151.0] ==> booking_status=Not_Canceled	0.182605	0.672364	0.118539	0.6492	0.9655
<b>1427</b>	no_of_children=0, lead_time=(151.0 - 443.0] ==> booking_status=Not_Canceled	0.186465	0.672364	0.052488	0.2815	0.4187

```
D:\ProgramData\Anaconda3\envs\DataMining\lib\site-packages\arules\association_rules.py:409: UserWarning: FixedFormatter should only be used together with FixedLocator
  axes.set_xticklabels(labs_x)
```



## R code for graphs

```
library(arulesViz)

df <- read.csv("D:\\Facultate\\DM\\Hotel Reservations.csv", header = TRUE, sep
= ",")

rules <- apriori(df, parameter = list(support = 0.005, confidence = 0.5,
maxlen=3, target="rules"),)

plot(rules) #scatter

plot(rules, method = "graph", limit = 20)
```

## Alternative method (without graphs)

```
In [ ]: del hotel[hotel.columns[0]]
hotel = hotel[1:]
```

```
In [7]: hotel.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36275 entries, 0 to 36274
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Booking_ID            36275 non-null  object
1   no_of_adults          36275 non-null  int64
```

2	no_of_children	36275	non-null	int64
3	no_of_weekend_nights	36275	non-null	int64
4	no_of_week_nights	36275	non-null	int64
5	type_of_meal_plan	36275	non-null	object
6	required_car_parking_space	36275	non-null	int64
7	room_type_reserved	36275	non-null	object
8	lead_time	36275	non-null	int64
9	arrival_year	36275	non-null	int64
10	arrival_month	36275	non-null	int64
11	arrival_date	36275	non-null	int64
12	market_segment_type	36275	non-null	object
13	repeated_guest	36275	non-null	int64
14	no_of_previous_cancellations	36275	non-null	int64
15	no_of_previous_bookings_not_canceled	36275	non-null	int64
16	avg_price_per_room	36275	non-null	float64
17	no_of_special_requests	36275	non-null	int64
18	booking_status	36275	non-null	object

dtypes: float64(1), int64(13), object(5)  
memory usage: 5.3+ MB

```
In [ ]: categorical_columns = [
        "type_of_meal_plan",
        "required_car_parking_space",
        "room_type_reserved",
        "market_segment_type",
        "repeated_guest",
        "booking_status"
    ]

hotel = hotel.drop(['Booking_ID'], axis=1)

numerical_columns = hotel.columns.difference(categorical_columns)
```

```
In [111... #implement association rules for every column
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

def discretize(df, column, bins, labels):
    df[column] = pd.cut(df[column], bins=bins, labels=labels)

#discretize average price
bins = [i * 25 for i in range(24)]
labels = ["P" + str(i * 25) + "-" + str((i+1) * 25) for i in range(23)]

hotel[16] = hotel[16].astype('float64')
discretize(hotel, 16, bins=bins, labels=labels)
hotel[16] = hotel[16].astype('object')

bins = [i * 25 for i in range(18)]
labels = ["L" + str(i * 25) + "-" + str((i+1) * 25) for i in range(17)]
hotel[8] = hotel[8].astype('int64')
discretize(hotel, 8, bins=bins, labels=labels)
hotel[8] = hotel[8].astype('object')

col_abvr = ['AD_', 'CH_', 'WEN_', "WKN_", "", "PARK_", "", "", "", "MON_", "DATE_", "",
print(len(col_abvr))
print(hotel.shape[1])
for i in range(len(col_abvr)):
    hotel.iloc[:,i] = col_abvr[i] + hotel.iloc[:,i].astype(str)
```

# Association rules for all attributes (support/confidence experiments)

```
In [112]: from apyori import apriori, load_transactions

print(hotel.shape[0])
print(hotel.shape[1])

# Intializing the list
transacts = []
rows = hotel.shape[0]
columns = hotel.shape[1]
for i in range(rows):
    transacts.append([str(hotel.values[i,j]) for j in range(0, columns)])

print("TRANSCATIONS : ", len(transacts) * len(transacts[0]))
```

```
36275
18
TRANSCATIONS : 652950
```

```
In [113]: rule = apriori(transactions = transacts, min_support = 0.003, min_confidence = 0.1, min_
output = list(rule) # returns a non-tabular output

def inspect(output):
    lhs = [", ".join(tuple(result[2][0][0])) for result in output]
    rhs = [tuple(result[2][0][1])[0] for result in output]
    support = [result[1] for result in output]
    confidence = [result[2][0][2] for result in output]
    lift = [result[2][0][3] for result in output]
    return list(zip(lhs, rhs, support, confidence, lift))

output_DataFrame = pd.DataFrame(inspect(output), columns = ['Left_Hand_Side', 'Right_Han
output_DataFrame
```

```
Out[113]:
```

	Left_Hand_Side	Right_Hand_Side	Support	Confidence	Lift
0		2017	0.179573	0.179573	1.000000
1		2018	0.820427	0.820427	1.000000
2		AD_1	0.212130	0.212130	1.000000
3		AD_2	0.719724	0.719724	1.000000
4		CH_0	0.925624	0.925624	1.000000
...	...	...	...	...	...
33305	WEN_2, SPEC_2	WKN_3	0.007223	0.209265	0.968375
33306	WEN_2, SPEC_2	WKN_5	0.003694	0.107029	2.405494
33307	nan	WKN_1	0.021337	0.443553	3.166054
33308	nan	WEN_0	0.006726	0.139828	0.882901
33309	nan	WEN_1	0.005872	0.122063	2.909223

33310 rows × 5 columns

```
In [99]: output_DataFrame = pd.DataFrame(inspect(output), columns = ['Left_Hand_Side', 'Right_Han
output_DataFrame
```

```
# output_DataFrame.nlargest(n = 50, columns = 'Confidence')
```

Out[99]:

	Left_Hand_Side	Right_Hand_Side	Support	Confidence	Lift
<b>0</b>		2017	0.179573	0.179573	1.000000
<b>1</b>		2018	0.820427	0.820427	1.000000
<b>2</b>		AD_1	0.212130	0.212130	1.000000
<b>3</b>		AD_2	0.719724	0.719724	1.000000
<b>4</b>		CH_0	0.925624	0.925624	1.000000
...	...	...	...	...	...
<b>33305</b>	WEN_2, SPEC_2	WKN_3	0.007223	0.209265	0.968375
<b>33306</b>	WEN_2, SPEC_2	WKN_5	0.003694	0.107029	2.405494
<b>33307</b>	nan	WKN_1	0.021337	0.443553	3.166054
<b>33308</b>	nan	WEN_0	0.006726	0.139828	0.882901
<b>33309</b>	nan	WEN_1	0.005872	0.122063	2.909223

33310 rows × 5 columns

In [104]...

```
#select the rules with canceled and not_canceled as right hand side
output_DataFrame_2 = output_DataFrame[output_DataFrame['Right_Hand_Side'].str.contains("
output_DataFrame_2.nlargest(n = 50, columns = 'Confidence')
```

Out[104]:

	Left_Hand_Side	Right_Hand_Side	Support	Confidence	Lift
<b>735</b>	Complementary	Not_Canceled	0.010779	1.000000	1.487290
<b>2555</b>	PNC_2	Not_Canceled	0.003088	1.000000	1.487290
<b>2567</b>	SPEC_3	Not_Canceled	0.018608	1.000000	1.487290
<b>2557</b>	REP_1	Not_Canceled	0.025196	0.982796	1.461702
<b>2554</b>	PNC_1	Not_Canceled	0.006175	0.982456	1.461197
<b>2156</b>	MON_1	Not_Canceled	0.027292	0.976331	1.452088
<b>2552</b>	PC_1	Not_Canceled	0.005155	0.944444	1.404663
<b>2577</b>	nan	Not_Canceled	0.044907	0.933524	1.388421
<b>686</b>	L375-400	Canceled	0.003529	0.914286	2.790552
<b>2546</b>	P25-50	Not_Canceled	0.007002	0.903915	1.344383
<b>2550</b>	PARK_1	Not_Canceled	0.027843	0.898577	1.336444
<b>757</b>	Corporate	Not_Canceled	0.049538	0.890927	1.325067
<b>2233</b>	MON_12	Not_Canceled	0.072198	0.866931	1.289378
<b>684</b>	L300-325	Canceled	0.015190	0.847692	2.587298
<b>680</b>	L225-250	Canceled	0.013894	0.806400	2.461267
<b>2563</b>	Room_Type 7	Not_Canceled	0.003363	0.772152	1.148414
<b>1132</b>	DATE_2	Not_Canceled	0.028201	0.768595	1.143124
<b>683</b>	L275-300	Canceled	0.011110	0.749071	2.286288



<b>2259</b>	MON_2	Not_Canceled	0.035121	0.747653	1.111976
<b>1720</b>	DATE_9	Not_Canceled	0.023046	0.739823	1.100331
<b>2097</b>	L50-75	Not_Canceled	0.089345	0.737765	1.097270
<b>946</b>	DATE_14	Not_Canceled	0.025224	0.736715	1.095709
<b>682</b>	L250-275	Canceled	0.017919	0.731159	2.231618
<b>2561</b>	Room_Type 5	Not_Canceled	0.005320	0.728302	1.083196
<b>1445</b>	DATE_29	Not_Canceled	0.023598	0.719328	1.069849
<b>1593</b>	DATE_5	Not_Canceled	0.022771	0.715771	1.064559
<b>2571</b>	WKN_0	Not_Canceled	0.047085	0.715543	1.064219
<b>1069</b>	DATE_18	Not_Canceled	0.024645	0.709524	1.055268
<b>818</b>	DATE_10	Not_Canceled	0.021254	0.707989	1.052985
<b>678</b>	L175-200	Canceled	0.029139	0.707023	2.157953
<b>2209</b>	MON_11	Not_Canceled	0.058029	0.706376	1.050586
<b>1380</b>	DATE_27	Not_Canceled	0.020565	0.704438	1.047704
<b>2544</b>	P175-200	Not_Canceled	0.014693	0.704095	1.047194
<b>2280</b>	MON_3	Not_Canceled	0.045706	0.703138	1.045770
<b>1688</b>	DATE_8	Not_Canceled	0.023212	0.702838	1.045324
<b>915</b>	DATE_13	Not_Canceled	0.026189	0.699558	1.040446
<b>852</b>	DATE_11	Not_Canceled	0.021172	0.699454	1.040290
<b>2131</b>	L75-100	Not_Canceled	0.060041	0.695624	1.034595
<b>1537</b>	DATE_31	Not_Canceled	0.011027	0.692042	1.029266
<b>1101</b>	DATE_19	Not_Canceled	0.025196	0.688772	1.024403
<b>1162</b>	DATE_20	Not_Canceled	0.023928	0.677596	1.007781
<b>1195</b>	DATE_21	Not_Canceled	0.021558	0.675302	1.004370
<b>14</b>		Not_Canceled	0.672364	0.672364	1.000000
<b>1656</b>	DATE_7	Not_Canceled	0.020565	0.672072	0.999566
<b>2559</b>	Room_Type 2	Not_Canceled	0.012791	0.670520	0.997258
<b>2514</b>	Not Selected	Not_Canceled	0.094583	0.668811	0.994716
<b>1037</b>	DATE_17	Not_Canceled	0.024728	0.666914	0.991895
<b>2427</b>	MON_9	Not_Canceled	0.084714	0.666450	0.991204
<b>590</b>	CH_1	Not_Canceled	0.029717	0.666255	0.990914
<b>1833</b>	L125-150	Not_Canceled	0.033715	0.664674	0.988563