Hotel Cancellations

Elizabeth Jones, Tia Jones, Chuka Soronnadi, Lagarthucin Legrand

Introduction

- Analyzed dataset was pulled into R from csv file
- 20 Variables 40060 Observations
- Goal : Recognize patterns that predict motives for hotel cancellations
- Goal : Consider and offer solution on how to reduce cancellations

Business Questions

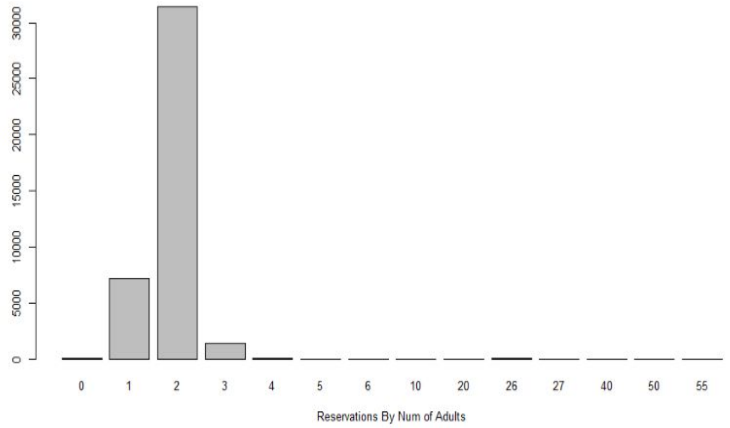
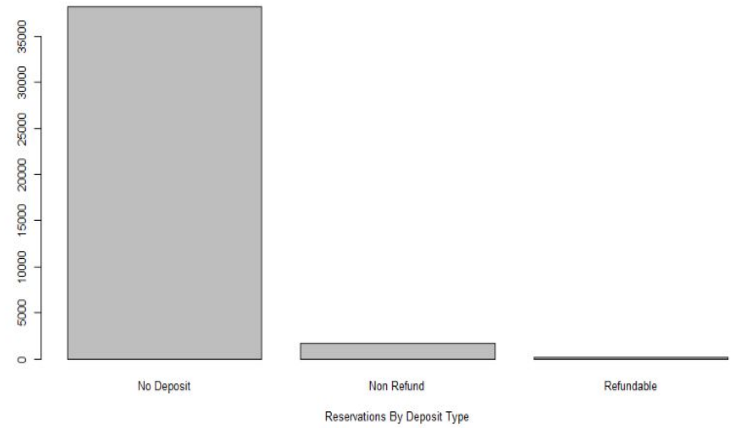
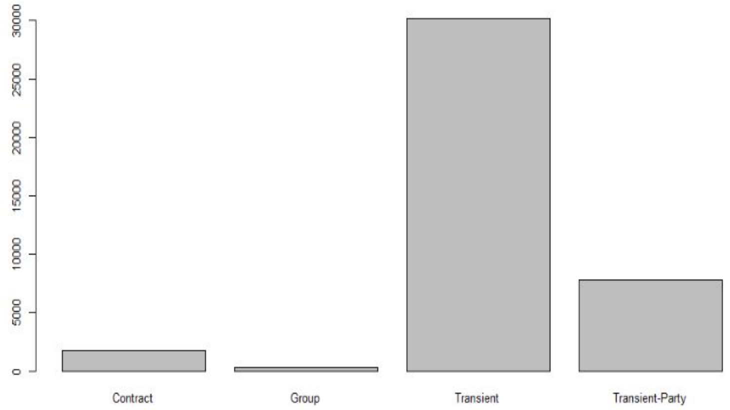
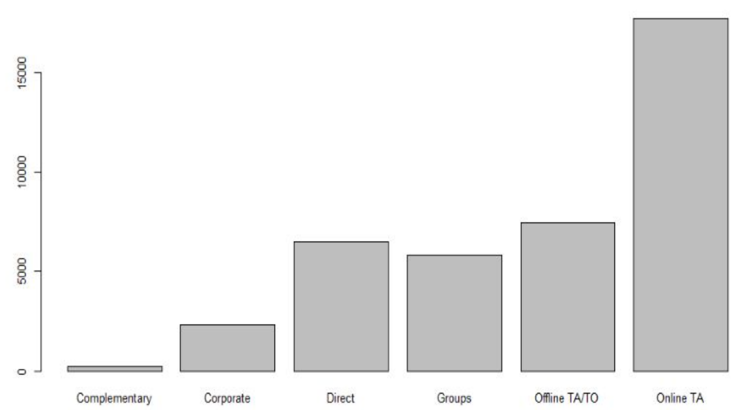
- Where do the most reservations come from?
- What type of person cancels?
- What type of person doesn’t cancel?
- Do attributes exist that may entice customers not to cancel?

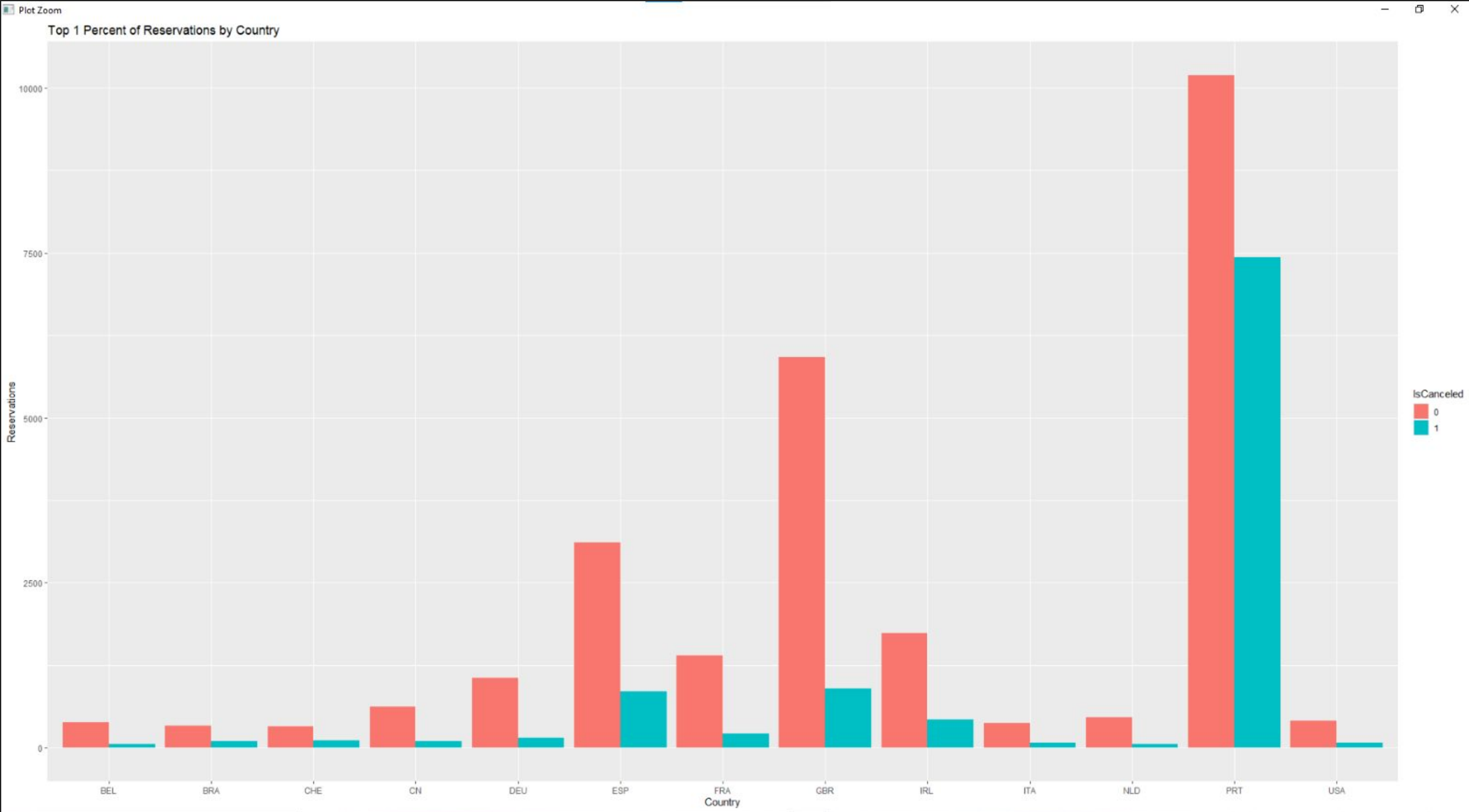
Data Acquisition, Cleansing, Transformation and Munging

- There were no NA’s within any columns that involved numbers
- Dataset missing country names for 432 reservations
- Additional datasets were created for specific analysis
    - Ex. Canceled and Non-canceled reservations
- Datasets that looked at columns as factors not integers/characters
- Added column titled “Room Change”, which compared reserved and assigned room types

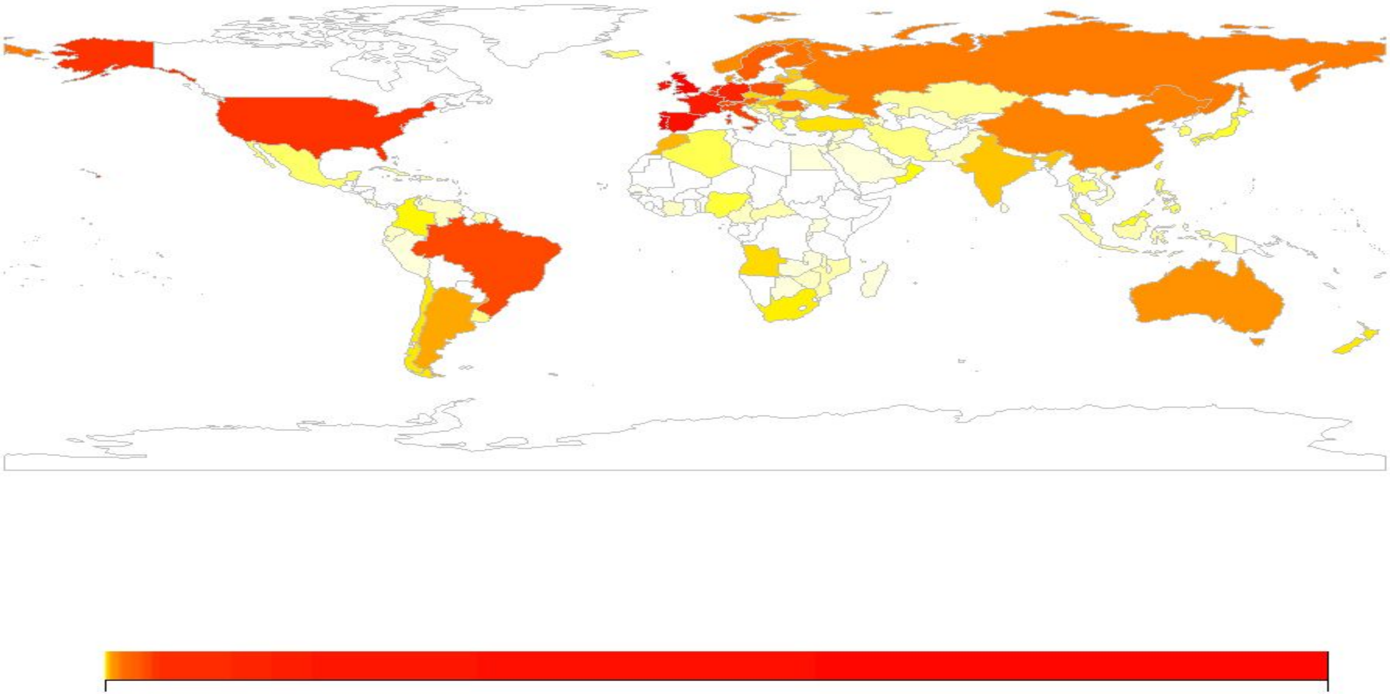
Descriptive Statistics/ Visualizations

- Focused on where we saw the most reservations
- We looked at
  - Adults
  - Deposit Type
  - Customer Type
  - Market Segment
  - Country
- Most of the reservations were from
  - 2 adult
  - No deposit
  - Transient
  - Online Travel Agent
  - Portugal reservations


0
5000
10000
15000
20000
25000
30000
0 1 2 3 4 5 6 10 20 26 27 40 50 55
Reservations By Num of Adults

0
5000
10000
15000
20000
25000
30000
35000
No Deposit
Non Refund
Refundable
Reservations By Deposit Type

0
5000
10000
15000
20000
25000
30000
Contract
Group
Transient
Transient-Party

0
5000
10000
15000
Complementary
Corporate
Direct
Groups
Offline TA/TO
Online TA

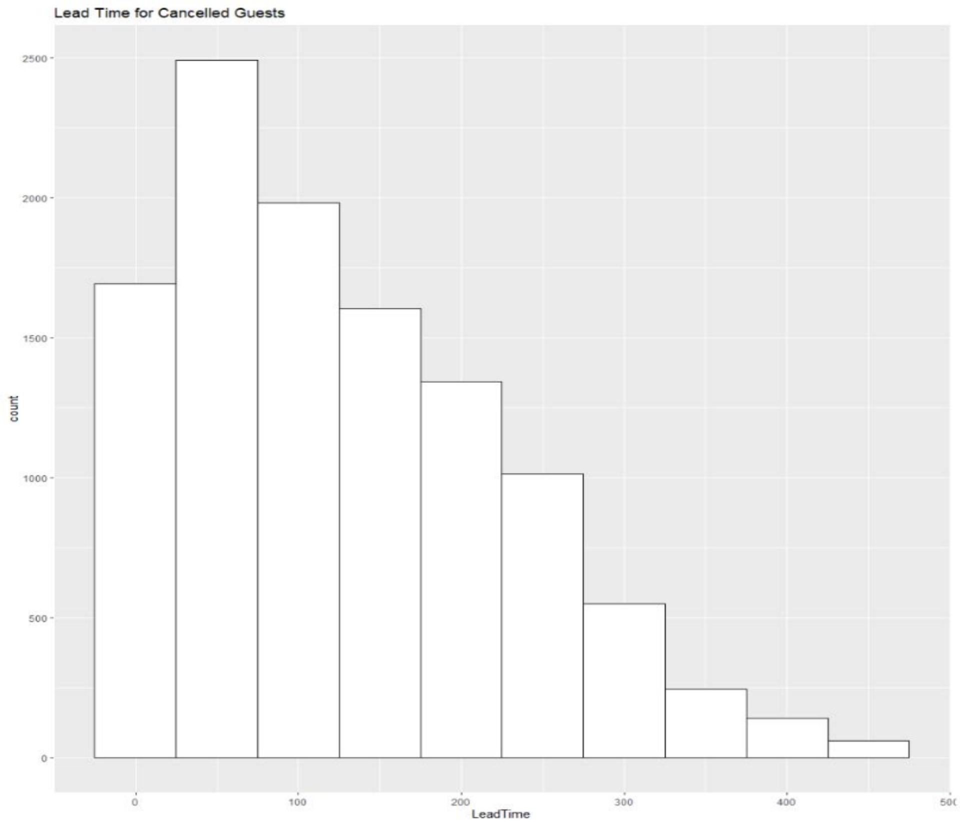

Plot Zoom
Top 1 Percent of Reservations by Country
Reservations
10000
7500
5000
2500
0
BEL
BRA
CHE
CN
DEU
ESP
FRA
GBR
IRL
ITA
NLD
PRT
USA
Country
IsCanceled
0
1

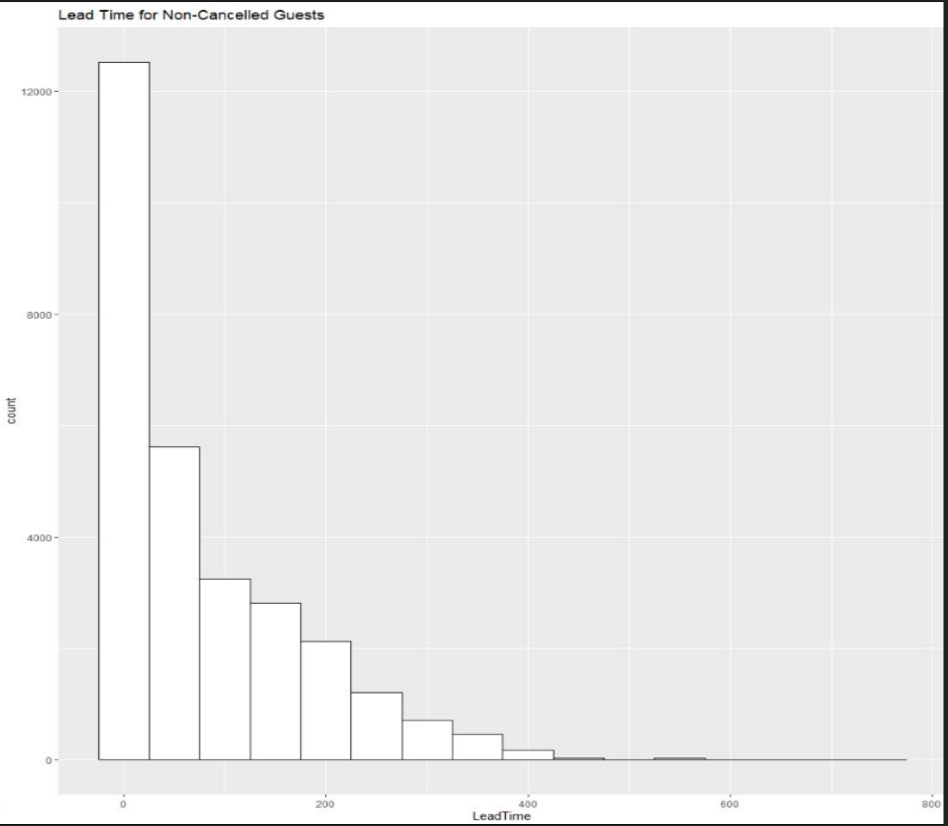
Percentage of Total Reservations



Descriptive Statistics/ Visualizations

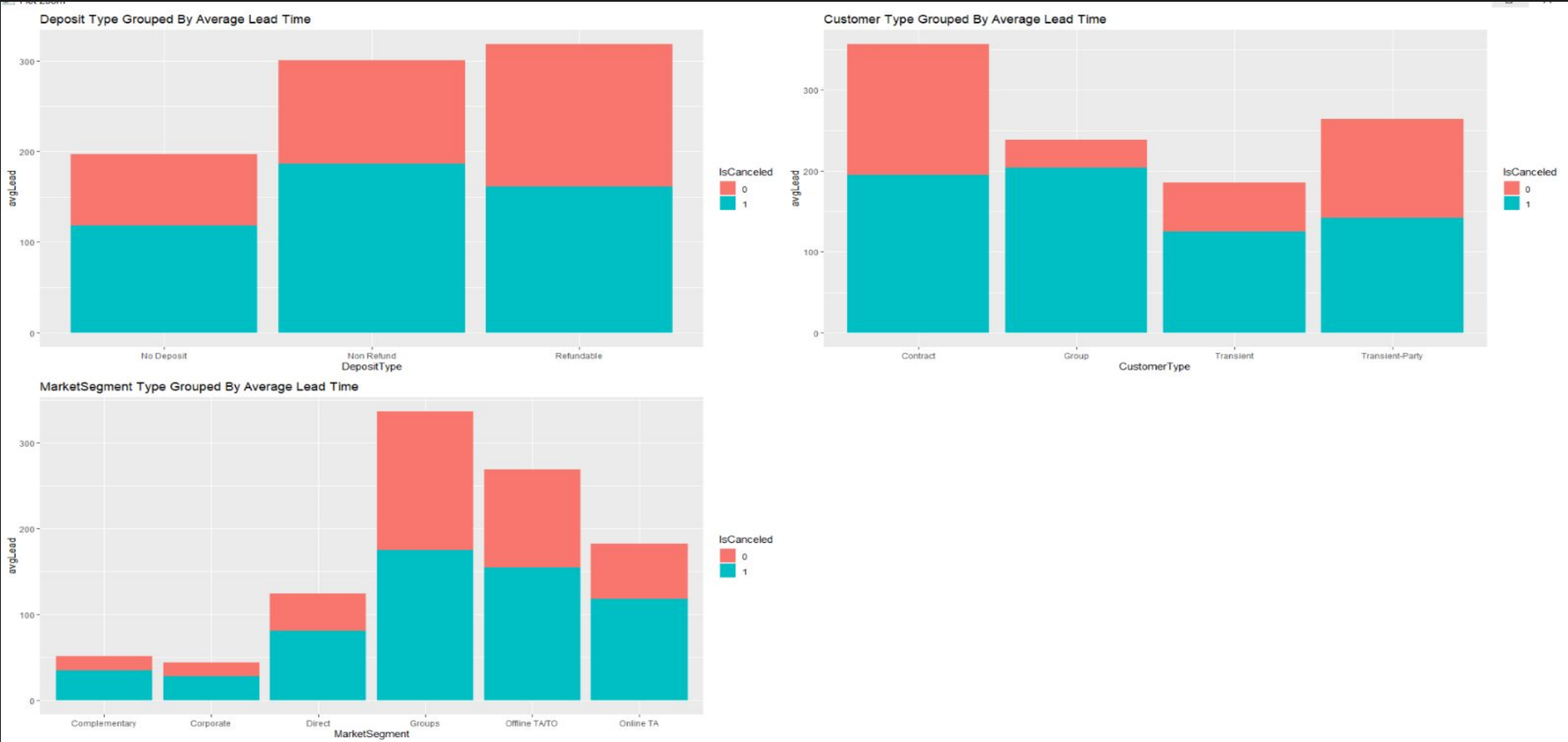
- We analyzed the Lead Time which showed how far in advance people made reservations
- We found a lower Lead Time with non-cancellations
- We found higher Lead Times with cancellations


Lead Time for Cancelled Guests
2500
2000
1500
1000
500
0
count
0
100
200
300
400
500
LeadTime


Lead Time for Non-Cancelled Guests
12000
8000
4000
0
count
0
200
400
600
800
LeadTime

Descriptive Statistics/ Visualizations

- Lead Time is a probable marker for hotel cancellations
- Compared Lead Time against:
  - Deposit Type
  - Market Segment
  - Customer Type
- Less hotel cancellations occurred in these groups when Lead Time was lower



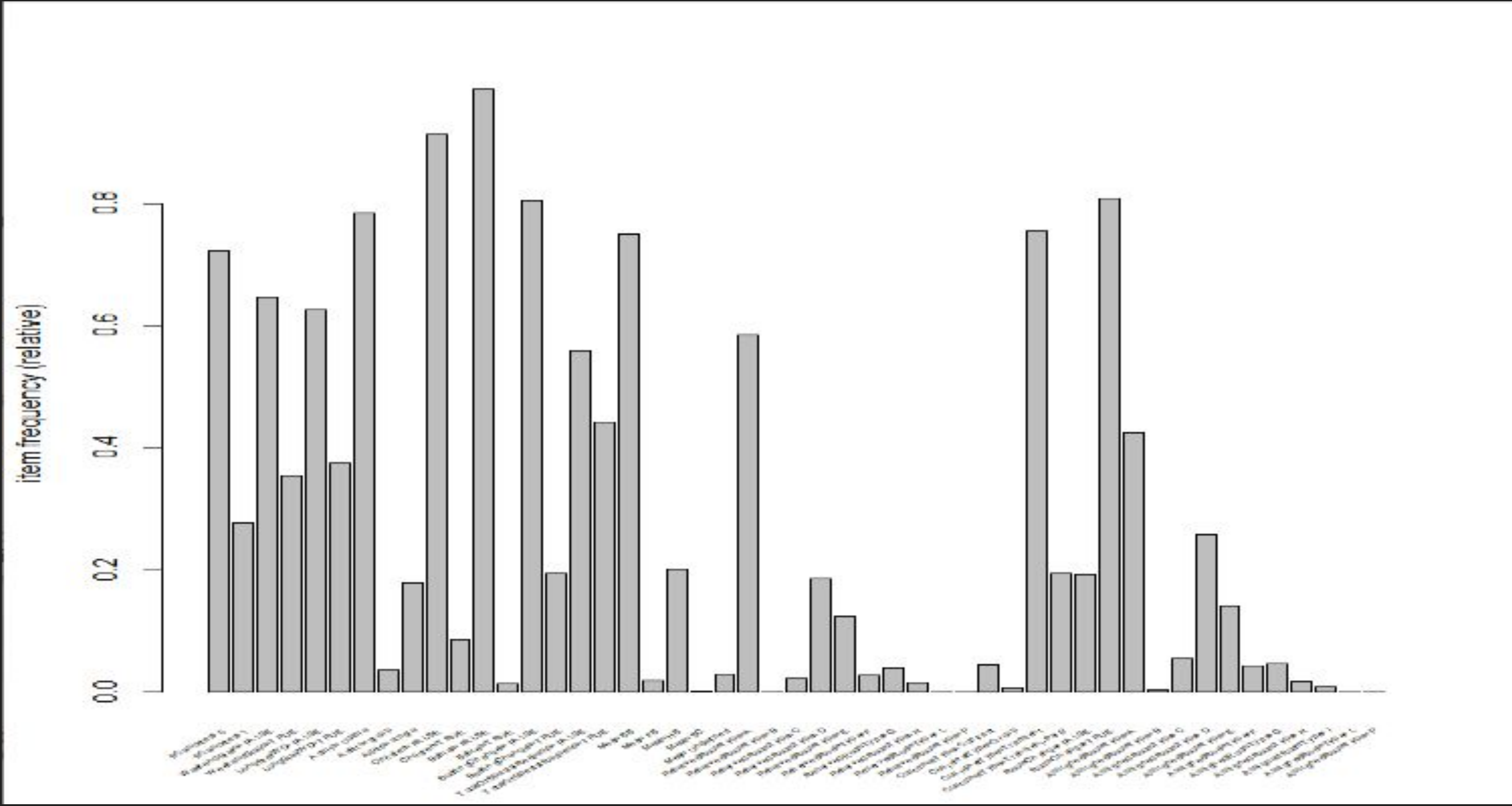
Deposit Type Grouped By Average Lead Time
avgLead
300
200
100
0
No Deposit
Non Refund
Refundable
DepositType
IsCanceled
0
1
Customer Type Grouped By Average Lead Time
avgLead
300
200
100
0
Contract
Group
Transient
Transient-Party
CustomerType
IsCanceled
0
1
MarketSegment Type Grouped By Average Lead Time
avgLead
300
200
100
0
Complementary
Corporate
Direct
Groups
Offline TA/TO
Online TA
MarketSegment
IsCanceled
0
1

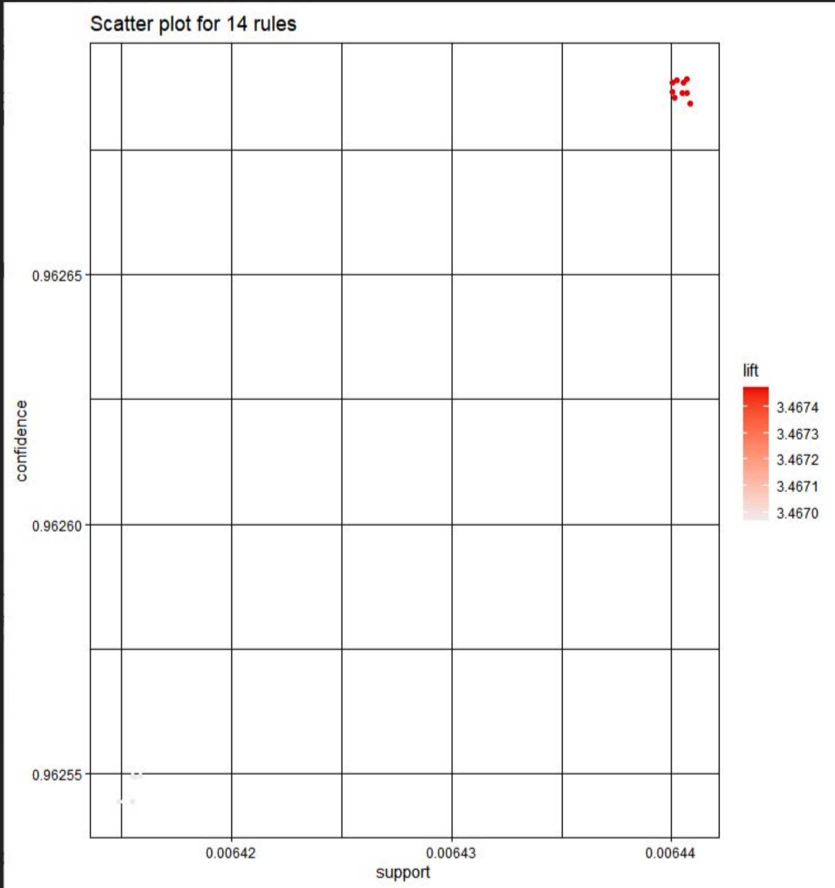
Use of Modeling Techniques and Visualizations

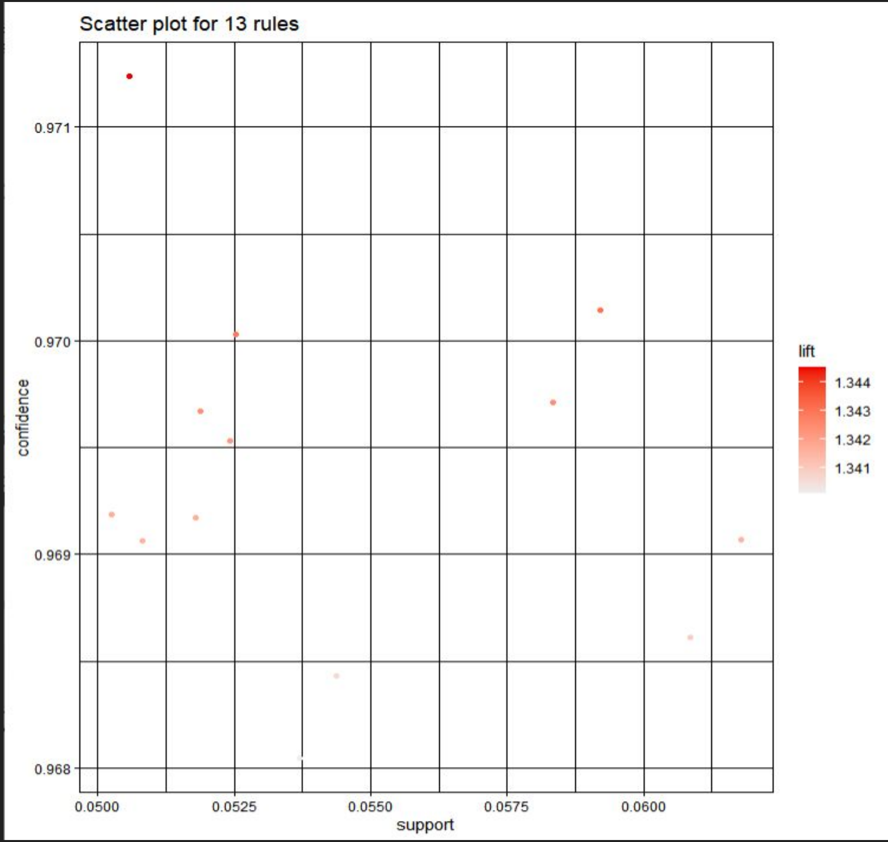
- The first model we created was a Tree model
- We analyzed
  - Country Type
  - Lead Time
  - Room Changes
  - Market Segment
- The Tree model was 81 percent accurate
  - Lead Time was the strongest predictor followed by Market Segment and Room Changes
- Next, we used a SVM Model
- We reviewed the same columns as the Tree Model. Once complete, it showed an accuracy 83% which was slightly more than the Tree model

Use of Modeling Techniques and Visualizations

- Lastly we used an Association Model to recognize any trends among those who did and did not cancel their reservations
- After analyzing the data, we found that customers who cancelled were transient couples with the FB (Full board) meal option who reserved or were assigned to Room A
- Those who did not cancel had the BB (Bed & Breakfast) meal option, no room changes, and had special requests


item frequency (relative)
0.0
0.2
0.4
0.6
0.8


Scatter plot for 14 rules
confidence
0.96265
0.96260
0.96255
support
0.00642
0.00643
0.00644
lift
3.4674
3.4673
3.4672
3.4671
3.4670


Scatter plot for 13 rules
confidence
0.971
0.970
0.969
0.968
support
0.0500
0.0525
0.0550
0.0575
0.0600
lift
1.344
1.343
1.342
1.341

Actionable Insights

- Entice more customers who don’t cancel booking and prevent cancellations from customers who do
- Create New Booking Packages
  - People who didn’t cancel seemed to be couples
  - Had special special requests
  - Booked with Bed and Breakfast
- Advertise Hotel Services
  - The farther someone books in advance the more likely someone will cancel
  - Suggest add-ons to customers at discounted or special rates based to entice them to keep their booking