HOTEL BOOKINGS ANALYSIS

BY JESSICA ATARI-ATA

DATA ACQUISITION

- This Data was gotten from http://www.kaggle.com/jessemostipak/hotel-booking-demand.
- This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Project Highlights:

- To compare the size of the market segments between “City Hotel” & “Resort Hotel”
- To determine the category of customers with the most bookings
- The market segment with the largest number of bookings
- If people with children booked in advance including weekend nights

Project Tool:

- R Programming

Importing Data

```
hotel_bookings <- read.csv("hotel_bookings.csv")
```

Verifying the data

The following code chunks below were used to check the data set and get a list of all the column names:

```
head(hotel_bookings)
```

```
##          hotel is_canceled lead_time arrival_date_year arrival_date_month
## 1 Resort Hotel           0       342               2015               July
## 2 Resort Hotel           0       737               2015               July
## 3 Resort Hotel           0         7               2015               July
## 4 Resort Hotel           0        13               2015               July
## 5 Resort Hotel           0        14               2015               July
## 6 Resort Hotel           0        14               2015               July
```

```
##   arrival_date_week_number arrival_date_day_of_month stays_in_weekend_nights
## 1                       27                         1                       0
## 2                       27                         1                       0
## 3                       27                         1                       0
## 4                       27                         1                       0
## 5                       27                         1                       0
## 6                       27                         1                       0
```

```
##   stays_in_week_nights adults children babies meal country market_segment
## 1                    0      2        0      0   BB     PRT          Direct
## 2                    0      2        0      0   BB     PRT          Direct
## 3                    1      1        0      0   BB     GBR          Direct
## 4                    1      1        0      0   BB     GBR       Corporate
## 5                    2      2        0      0   BB     GBR       Online TA
## 6                    2      2        0      0   BB     GBR       Online TA
##   distribution_channel is_repeated_guest previous_cancellations
## 1               Direct                 0                      0
## 2               Direct                 0                      0
## 3               Direct                 0                      0
## 4            Corporate                 0                      0
## 5                TA/TO                 0                      0
## 6                TA/TO                 0                      0
```

```
##   previous_bookings_not_canceled reserved_room_type assigned_room_type
## 1                              0                  C                  C
## 2                              0                  C                  C
## 3                              0                  A                  C
## 4                              0                  A                  A
## 5                              0                  A                  A
## 6                              0                  A                  A
##   booking_changes deposit_type agent company days_in_waiting_list customer_type
## 1               3   No Deposit  NULL    NULL                    0     Transient
## 2               4   No Deposit  NULL    NULL                    0     Transient
## 3               0   No Deposit  NULL    NULL                    0     Transient
## 4               0   No Deposit   304    NULL                    0     Transient
## 5               0   No Deposit   240    NULL                    0     Transient
```

```
## 6               0   No Deposit   240    NULL                     0     Transient
##   adr required_car_parking_spaces total_of_special_requests reservation_status
## 1   0                           0                         0          Check-Out
## 2   0                           0                         0          Check-Out
## 3  75                           0                         0          Check-Out
## 4  75                           0                         0          Check-Out
## 5  98                           0                         1          Check-Out
## 6  98                           0                         1          Check-Out
##   reservation_status_date
## 1              2015-07-01
## 2              2015-07-01
## 3              2015-07-02
## 4              2015-07-02
## 5              2015-07-03
## 6              2015-07-03
```

```
colnames(hotel_bookings)
```

```
##  [1] "hotel"                         "is_canceled"
##  [3] "lead_time"                     "arrival_date_year"
##  [5] "arrival_date_month"            "arrival_date_week_number"
##  [7] "arrival_date_day_of_month"     "stays_in_weekend_nights"
##  [9] "stays_in_week_nights"          "adults"
## [11] "children"                      "babies"
## [13] "meal"                          "country"
## [15] "market_segment"                "distribution_channel"
## [17] "is_repeated_guest"             "previous_cancellations"
## [19] "previous_bookings_not_canceled" "reserved_room_type"
## [21] "assigned_room_type"            "booking_changes"
## [23] "deposit_type"                  "agent"
## [25] "company"                       "days_in_waiting_list"
## [27] "customer_type"                 "adr"
## [29] "required_car_parking_spaces"   "total_of_special_requests"
## [31] "reservation_status"            "reservation_status_date"
```

Installing Packages

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(ggplot2)
```

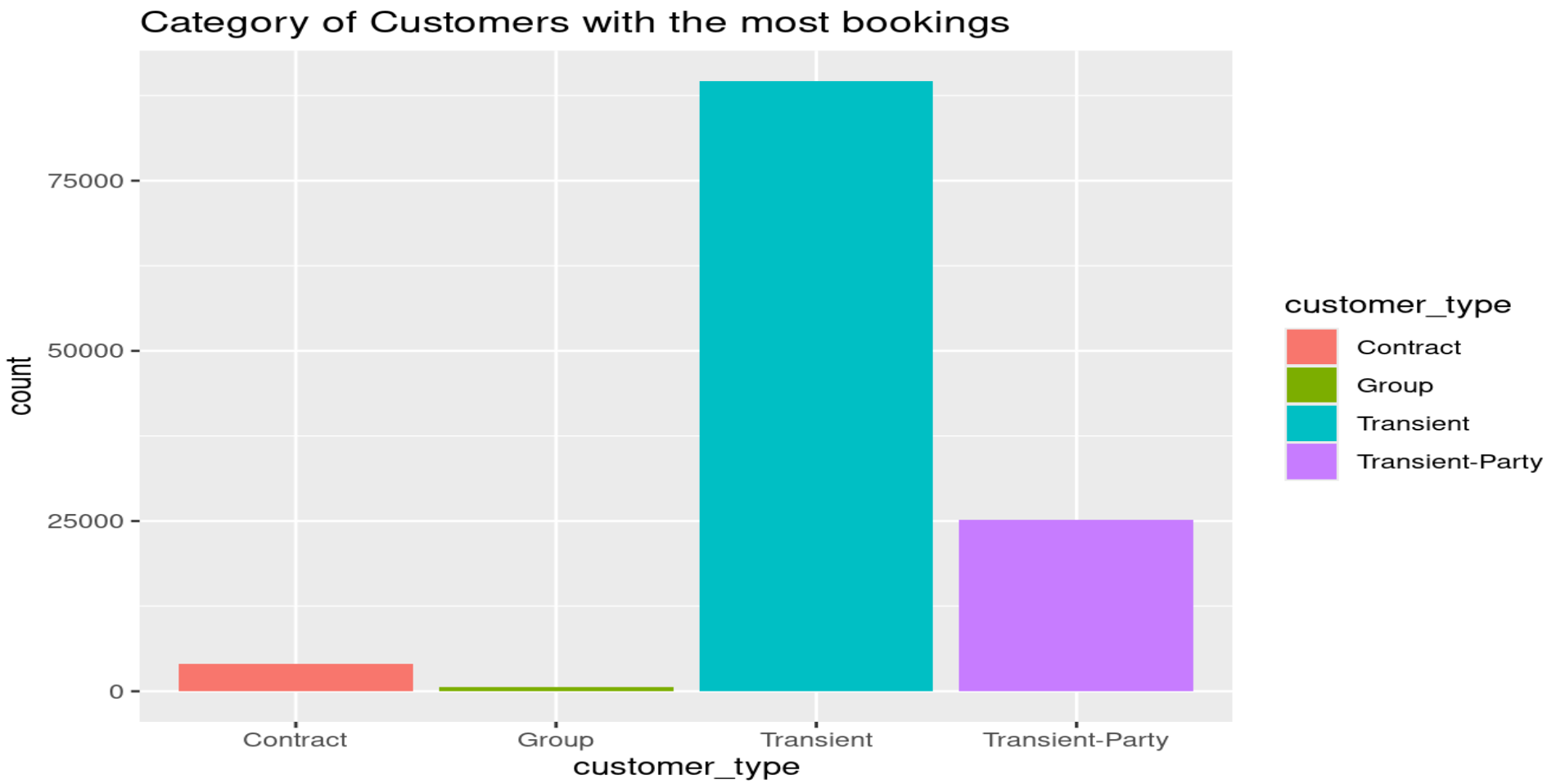
```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ lubridate 1.9.2     ✓ tibble    3.2.1
## ✓ purrr     1.0.1     ✓ tidyr     1.3.0
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
```

1). What category of customers made majority of the bookings?

➔ The bar chart clearly shows that customers who only stay for a short period of time (Transient customers) made most of these bookings.

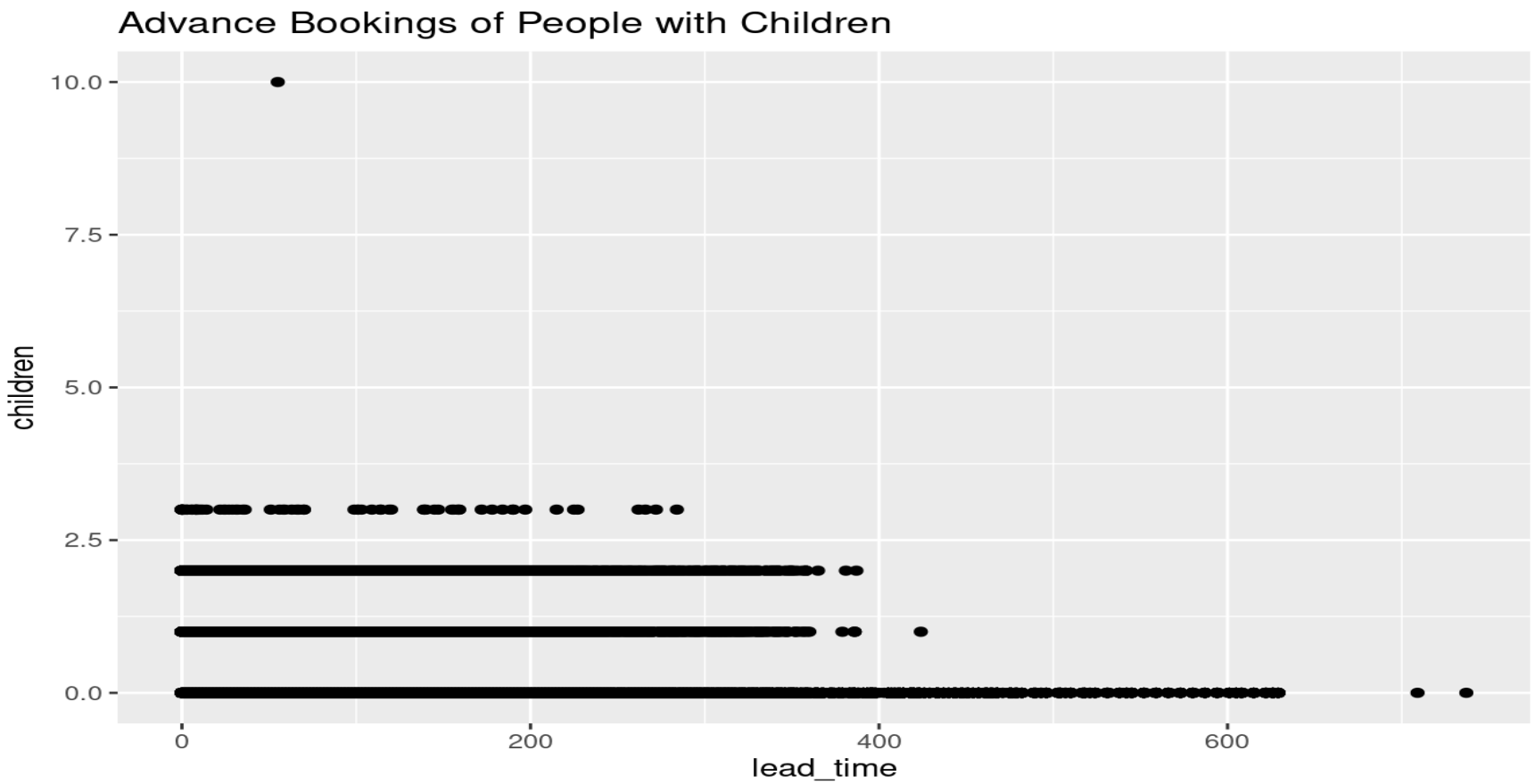
```
ggplot(data=hotel_bookings) + geom_bar(mapping=aes(x=customer_type, fill=customer_type))
+ labs(title = "Category of Customers with the most bookings")
```



2). Determine if people with children booked hotel rooms in advance.

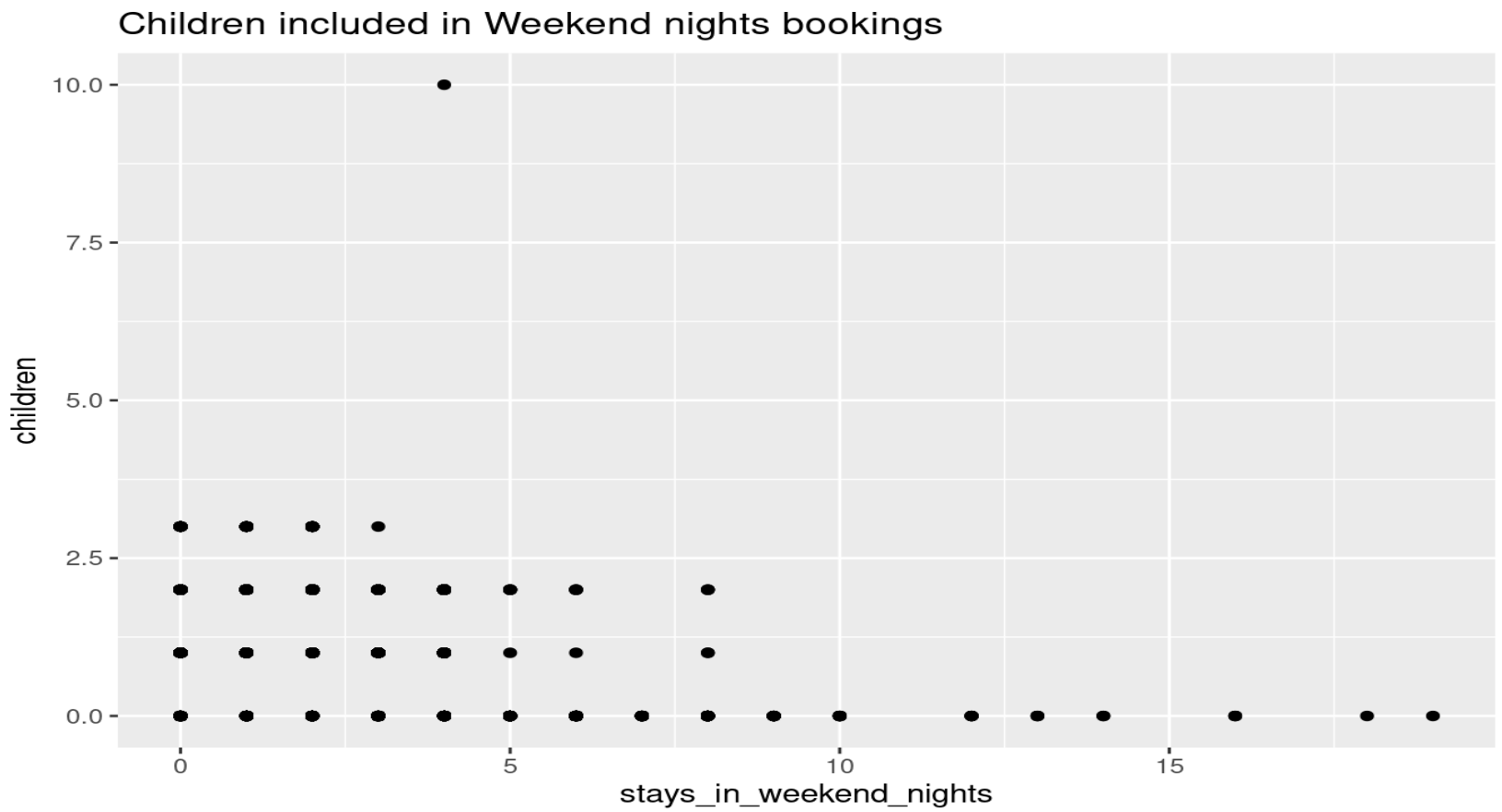
➔ The X-axis shows how far in advance bookings were made and the y-axis shows the number of children included in those bookings.

```
ggplot(data=hotel_bookings)+ geom_point(mapping = aes(x=lead_time,y=children)) + labs
(title="Advance Bookings of People with Children")
```



2b). Determine if children were included in weekend night bookings.

```
ggplot(data=hotel_bookings)+ geom_point(mapping=aes(x=stays_in_weekend_nights,
y=children))+labs(title = "Children included in Weekend nights bookings")
```

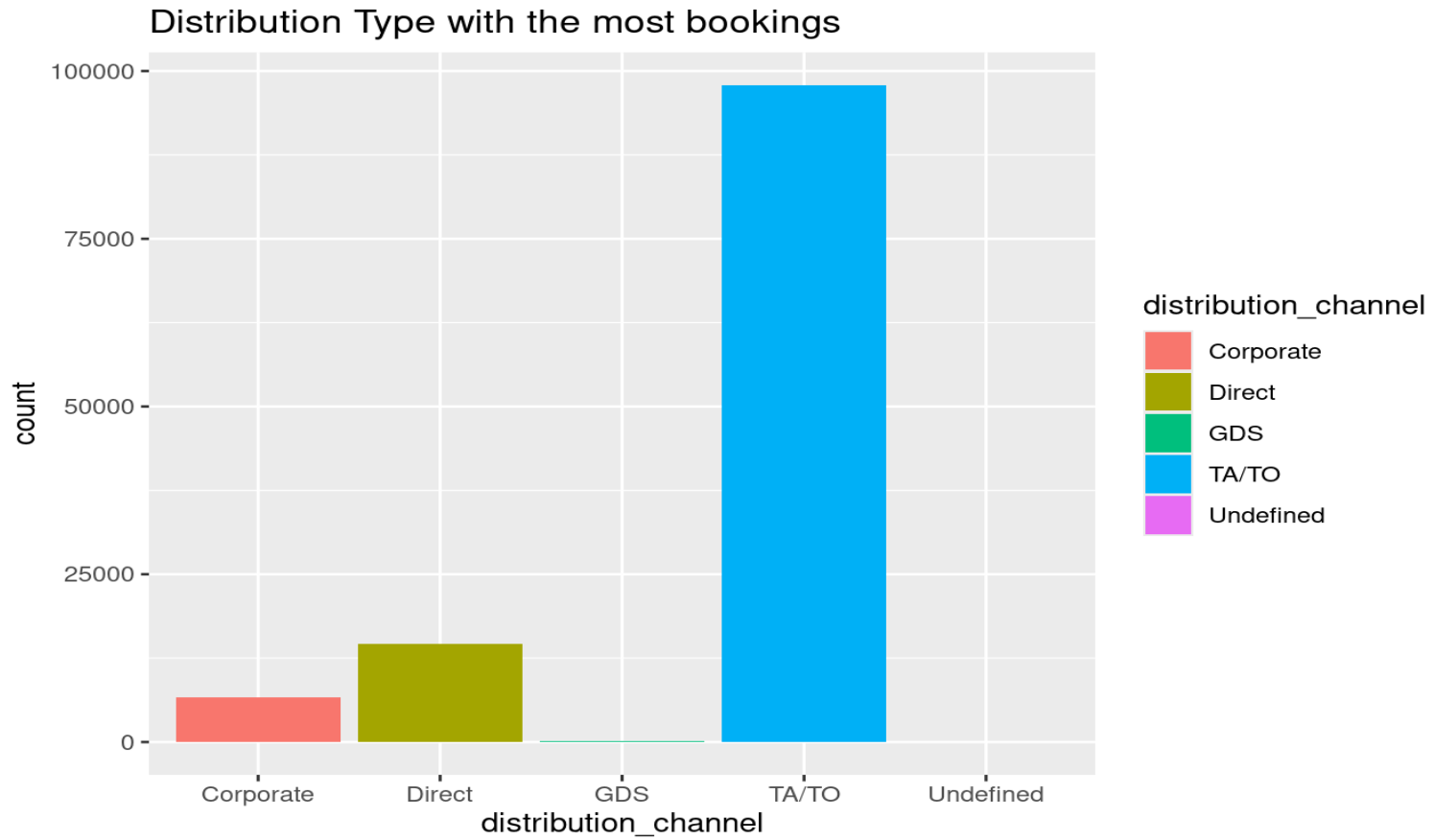


3). Which Distribution type had the most number of bookings?

N.B (TA/TO means Travel Agents/Tour Operators)

The Distribution type with the Most number of bookings are the Travel Agents (TA)/ Tour Operators(TO)

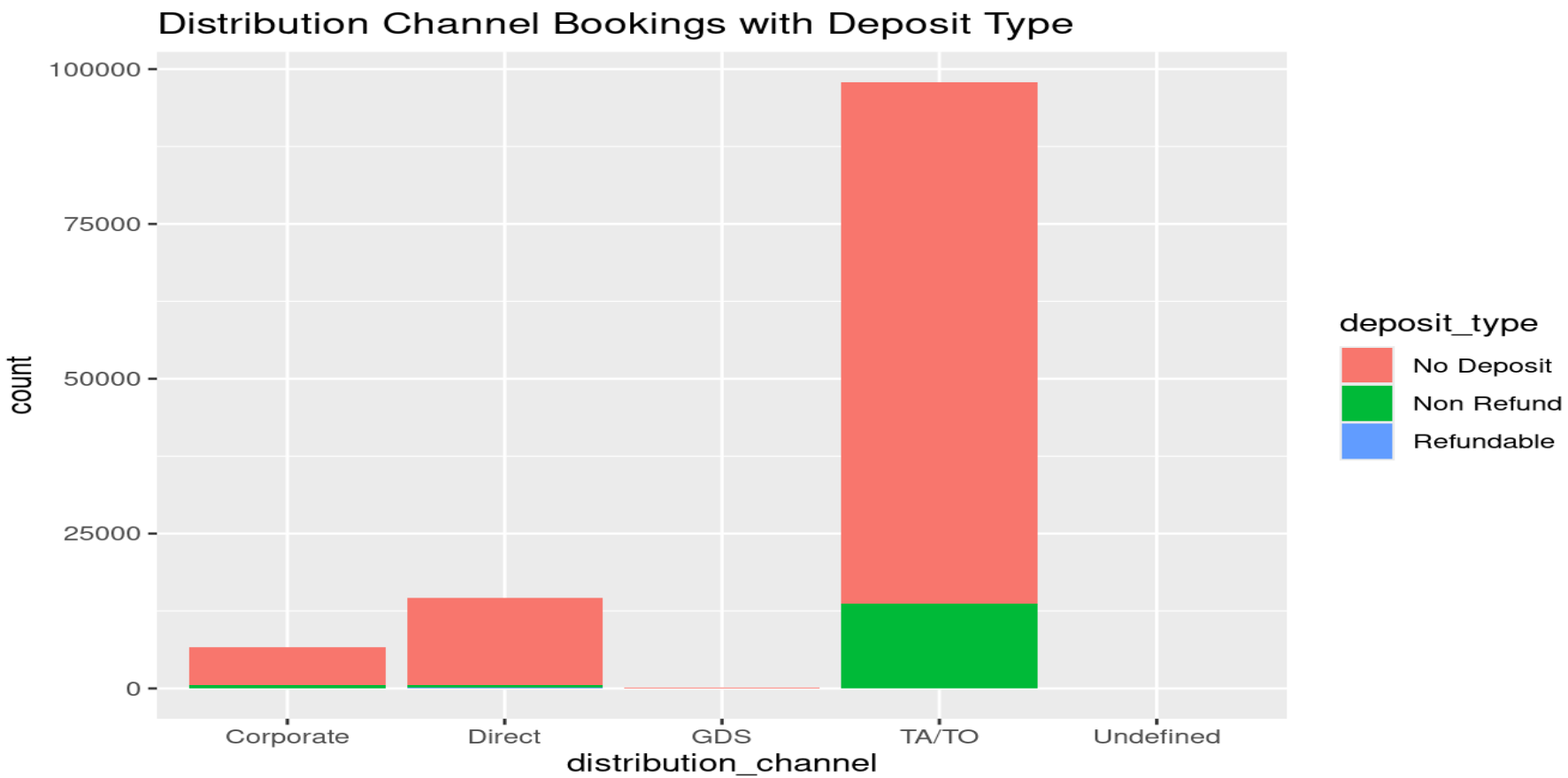
```
ggplot(data=hotel_bookings) + geom_bar(mapping=aes(x=distribution_channel, fill=
distribution_channel))+labs(title="Distribution Type with the most bookings")
```



4a).Find out if the number of bookings for each distribution channel is different depending on whether or not there was a deposit; or what market segment they represent.

N.B (TA/TO means Travel Agents/Tour Operators)

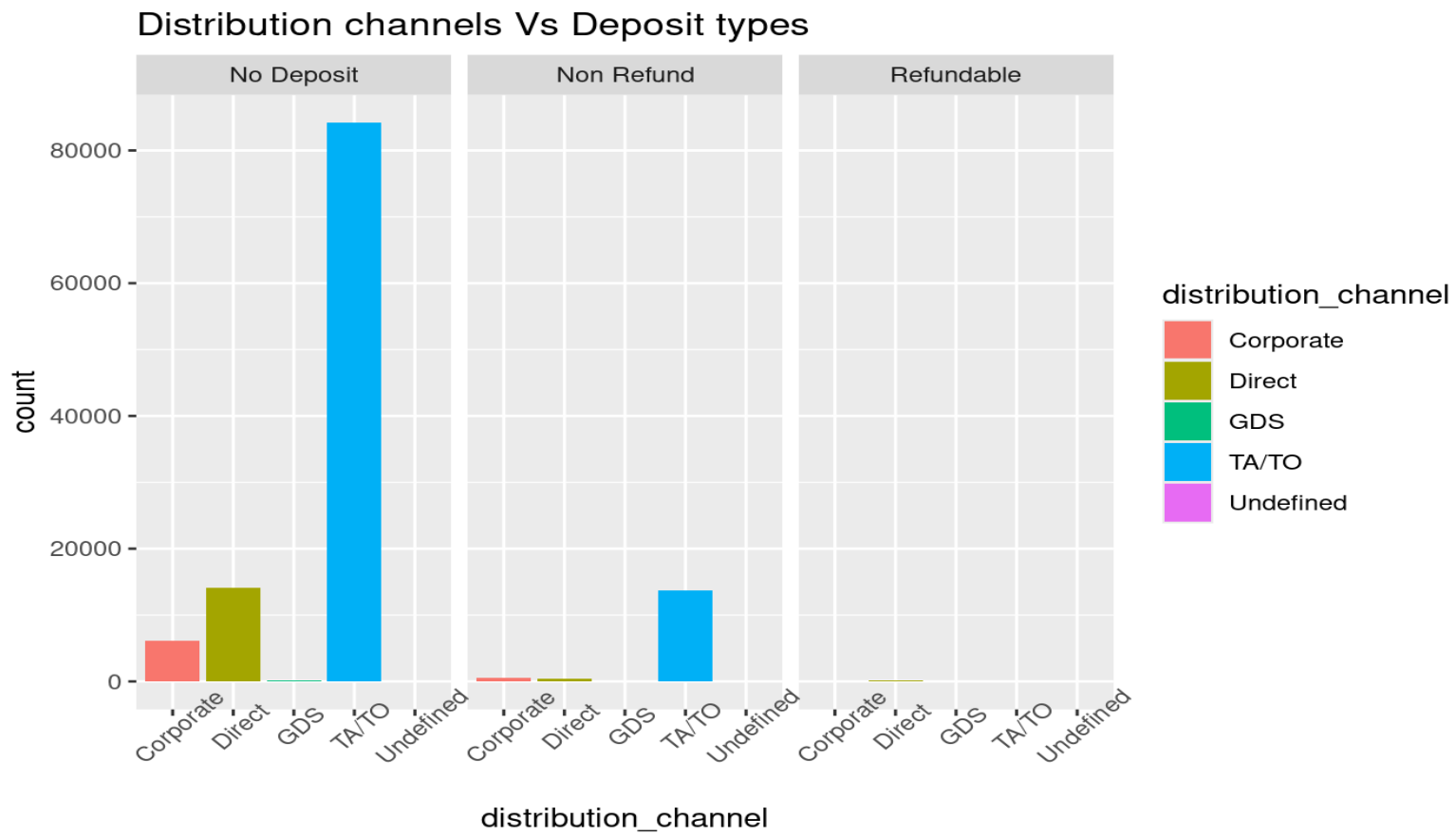
```
ggplot(data=hotel_bookings)+ geom_bar(mapping = aes(x=distribution_channel, fill=
deposit_type)) + labs(title = "Distribution Channel Bookings with Deposit Type")
```



4b.) Compare the deposit type of the distribution channels.

N.B (TA/TO means Travel Agents/Tour Operators)

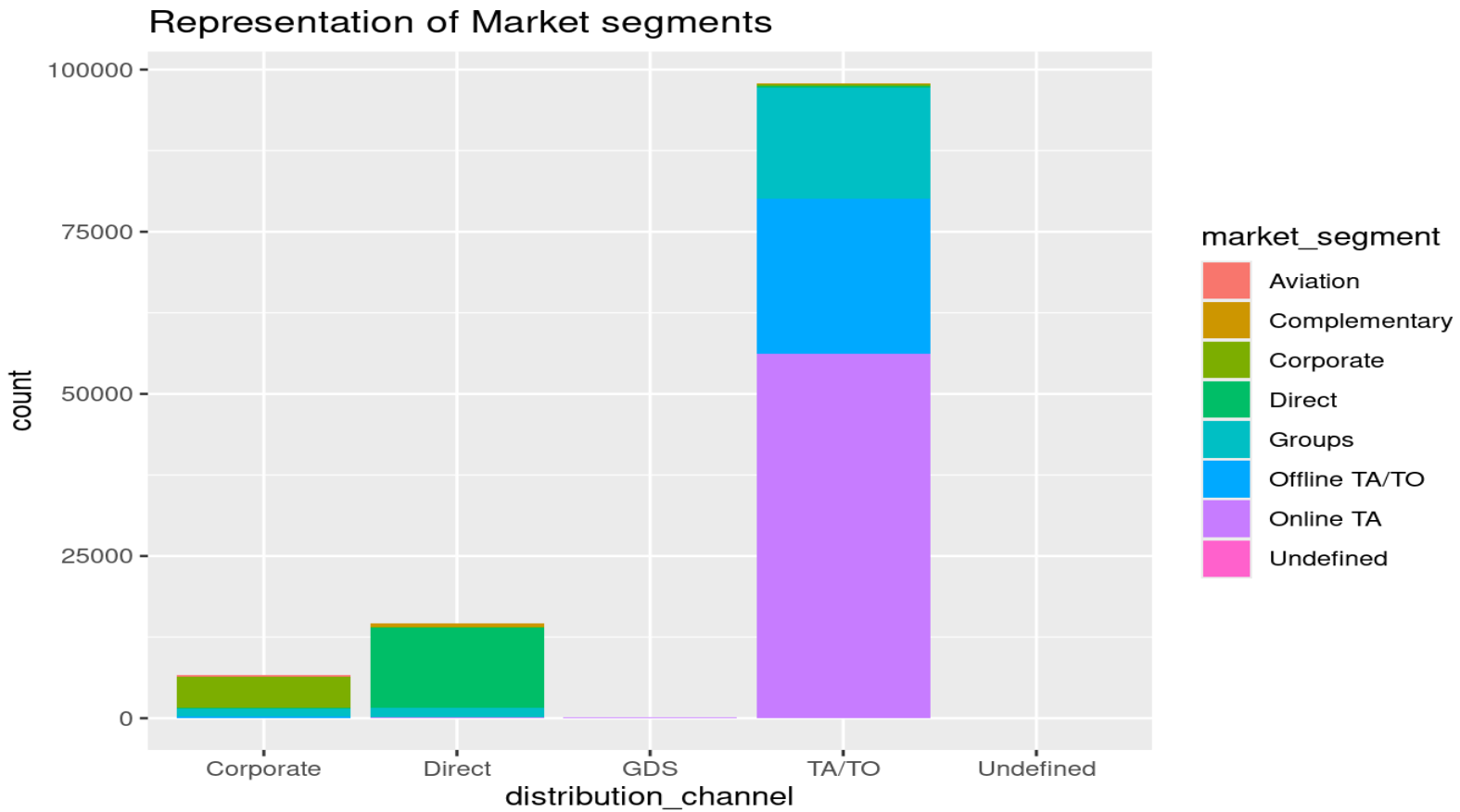
```
ggplot(data=hotel_bookings)+ geom_bar(mapping = aes(x=distribution_channel, fill= distribution_channel)) + facet_grid(~deposit_type) + theme(axis.text.x=element_text(angle=45)) + labs(title= "Distribution channels Vs Deposit types")
```



4c). The market segment they represent

N.B (TA/TO means Travel Agents/Tour Operators)

```
ggplot(data=hotel_bookings) + geom_bar(mapping = aes(x=distribution_channel, fill=
market_segment))+ labs(title = "Representation of Market segments")
```



5). Your Stakeholder wants to run a family- friendly promotion targeting key market segments. They also want to know the market segment that generates the largest number of bookings and where these bookings are made (city hotel or resort hotels).

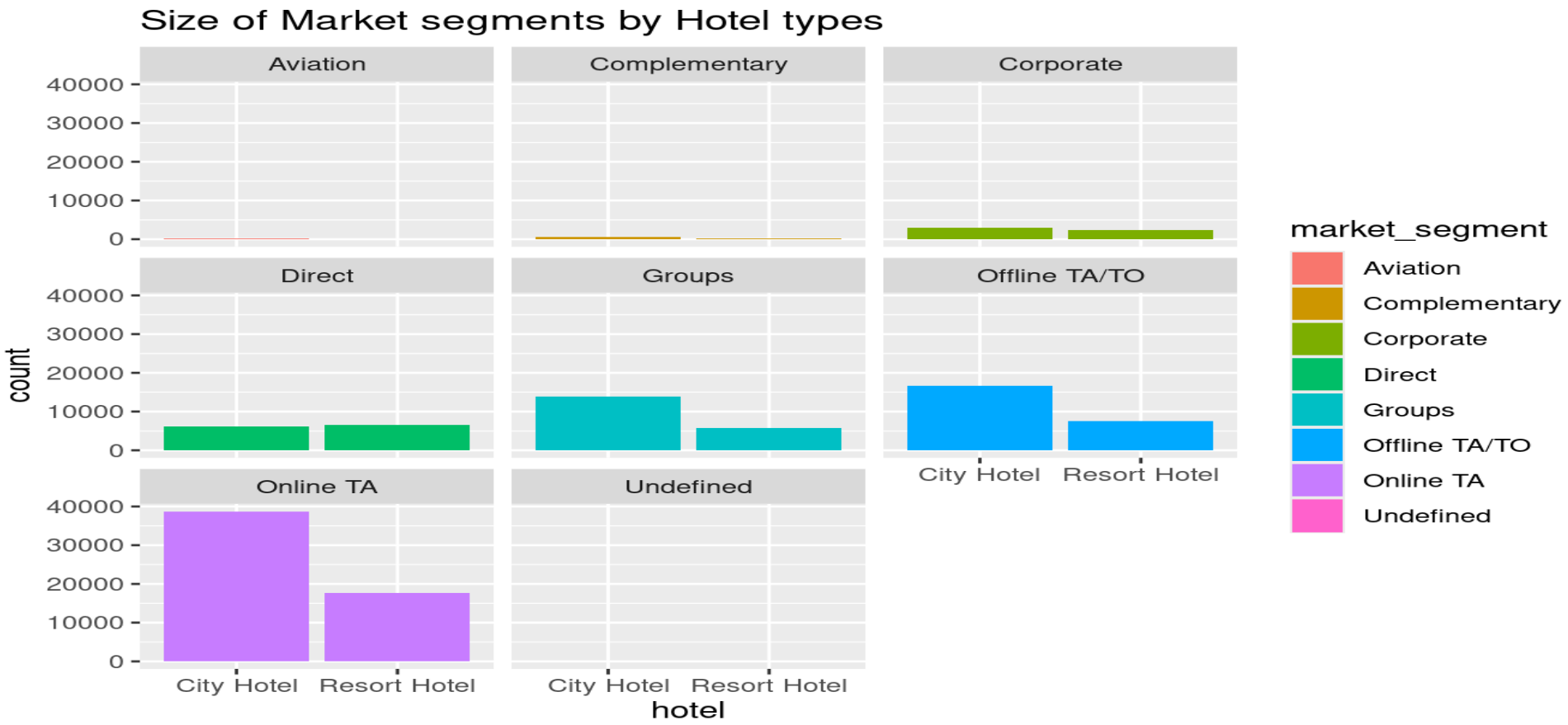
N.B (TA/TO means Travel Agents/Tour Operators)

```
ggplot(data=hotel_bookings) + geom_bar(mapping=aes(x=hotel,fill=market_segment))
+ labs(title="Market segments with the largest number of Bookings")
```

6). Compare the size of the market segment between “City Hotel” and “Resort Hotel”

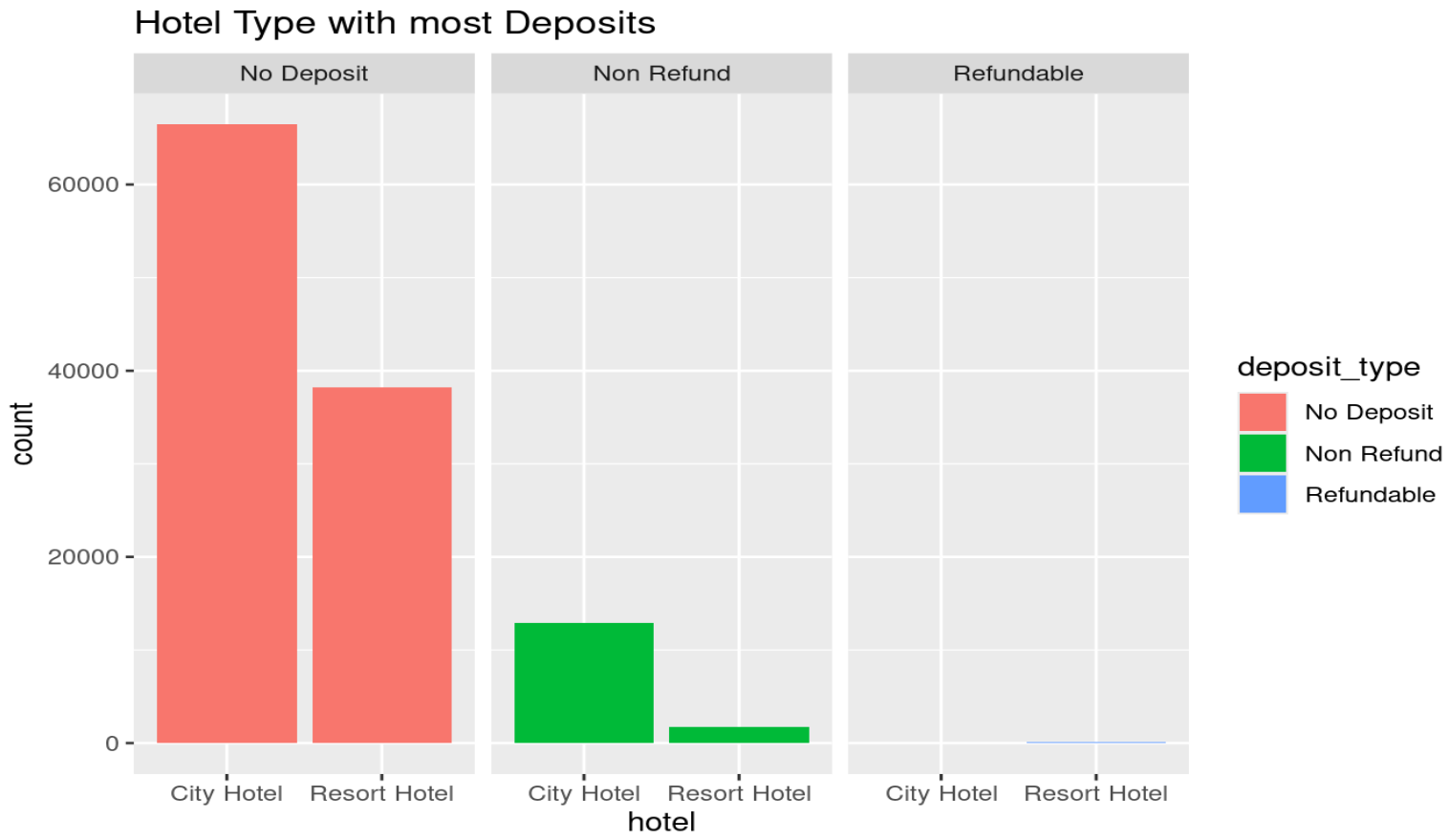
N.B (TA/TO means Travel Agents/Tour Operators)

```
ggplot(data=hotel_bookings) + geom_bar(mapping=aes(x=hotel, fill=market_segment)) +
facet_wrap(~market_segment) + labs(title = "Size of Market segments by Hotel types")
```



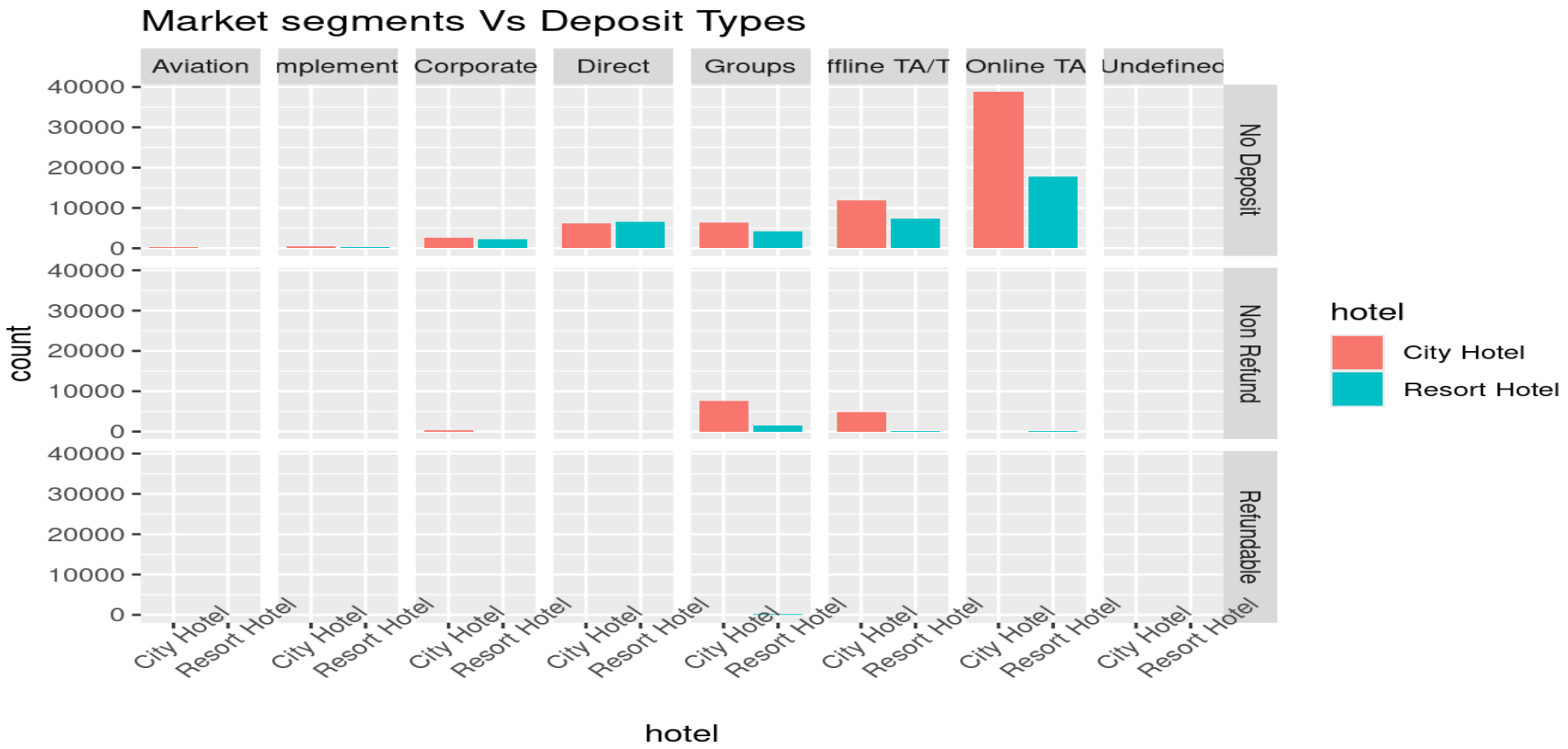
7). Between “City Hotels” and “Resort Hotels” which had more deposits?

```
ggplot(data=hotel_bookings)+ geom_bar(mapping = aes(x= hotel, fill=deposit_type))
+facet_wrap(~deposit_type)+ labs(title = "Hotel Type with most Deposits")
```



8). Compare the different market segments and their deposit types between “City Hotels” and “Resort Hotels”

```
ggplot(data=hotel_bookings)+ geom_bar(mapping = aes(x=hotel, fill= hotel)) +
facet_grid(~deposit_type ~market_segment) + theme(axis.text.x=element_text(an
gle=45)) + labs(title = "Market segments Vs Deposit Types")
```



9). what time period does this data cover?

```
min(hotel_bookings$arrival_date_year)
## [1] 2015
max(hotel_bookings$arrival_date_year)
## [1] 2017
mindate<- min(hotel_bookings$arrival_date_year)
maxdate<- max(hotel_bookings$arrival_date_year)
```

9b). Compare Market Segments by Hotel type

```
ggplot(data=hotel_bookings)+ geom_bar(mapping = aes(x = market_segment,
fill= market_segment)) + facet_wrap(~hotel) + labs(title="comparison of
market segments by hotel type", subtitle=paste0("Data from:" , mindate,
" to ", maxdate),x="Market Segment", y= "Number of Bookings")
```