Hotel Booking System Analytical Model

Sarvesh Kaushik

2/18/2022

Loading Packages:

By effective utilization of R, the Data Cleaning process was started,

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.6     ✓ dplyr   1.0.8
## ✓ tidyr   1.1.4     ✓ stringr 1.4.0
## ✓ readr   2.1.1     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Importing the data:

```
hotel_bookings <- read.csv("/Users/sarveshkaushik/Downloads/hotel.csv", sep = ",")
```

Analyzing the data in the bookings file,

```
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
```

Understanding different column names in the current dataset.

```
colnames(hotel_bookings)
```

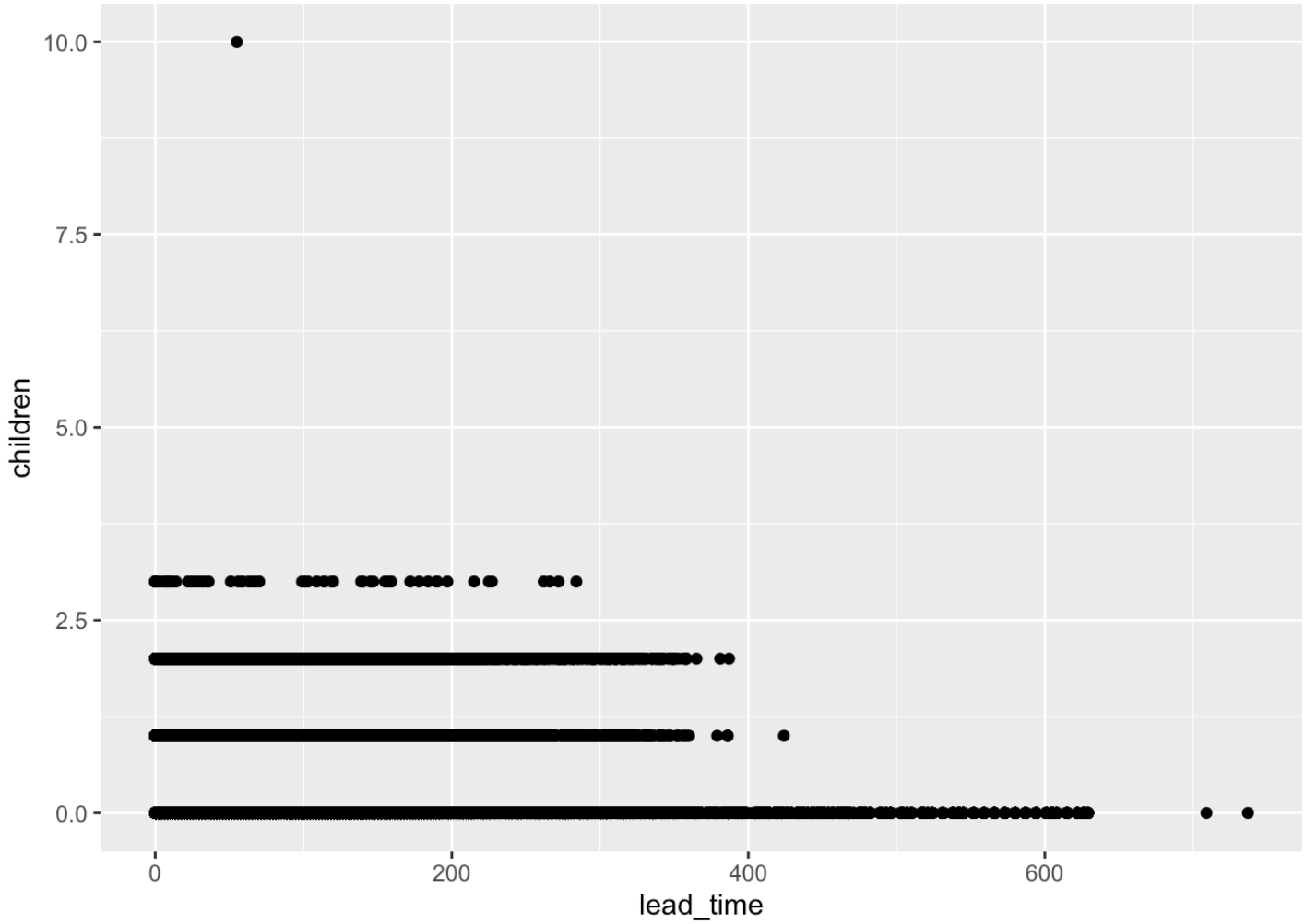
```
##  [1] "hotel"                         "is_canceled"
##  [3] "lead_time"                     "arrival_date_year"
##  [5] "arrival_date_month"            "arrival_date_week_number"
##  [7] "arrival_date_day_of_month"     "stays_in_weekend_nights"
##  [9] "stays_in_week_nights"          "adults"
## [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"
```

StakeHolder's Statement: "I want to target people who book early, and I have a hypothesis that people with children have to book in advance."

```
{r }
```

```
ggplot(data = hotel_bookings) +
  geom_point(mapping = aes(x = lead_time, y = children))
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```



To visualize the Hotel Booking data we will use ggplot package as given below,

install.packages('ggplot2')

library(ggplot2)

Stakeholder Statement: The stakeholder wants to increase weekend bookings, as the most important revenue resource for the hotel.

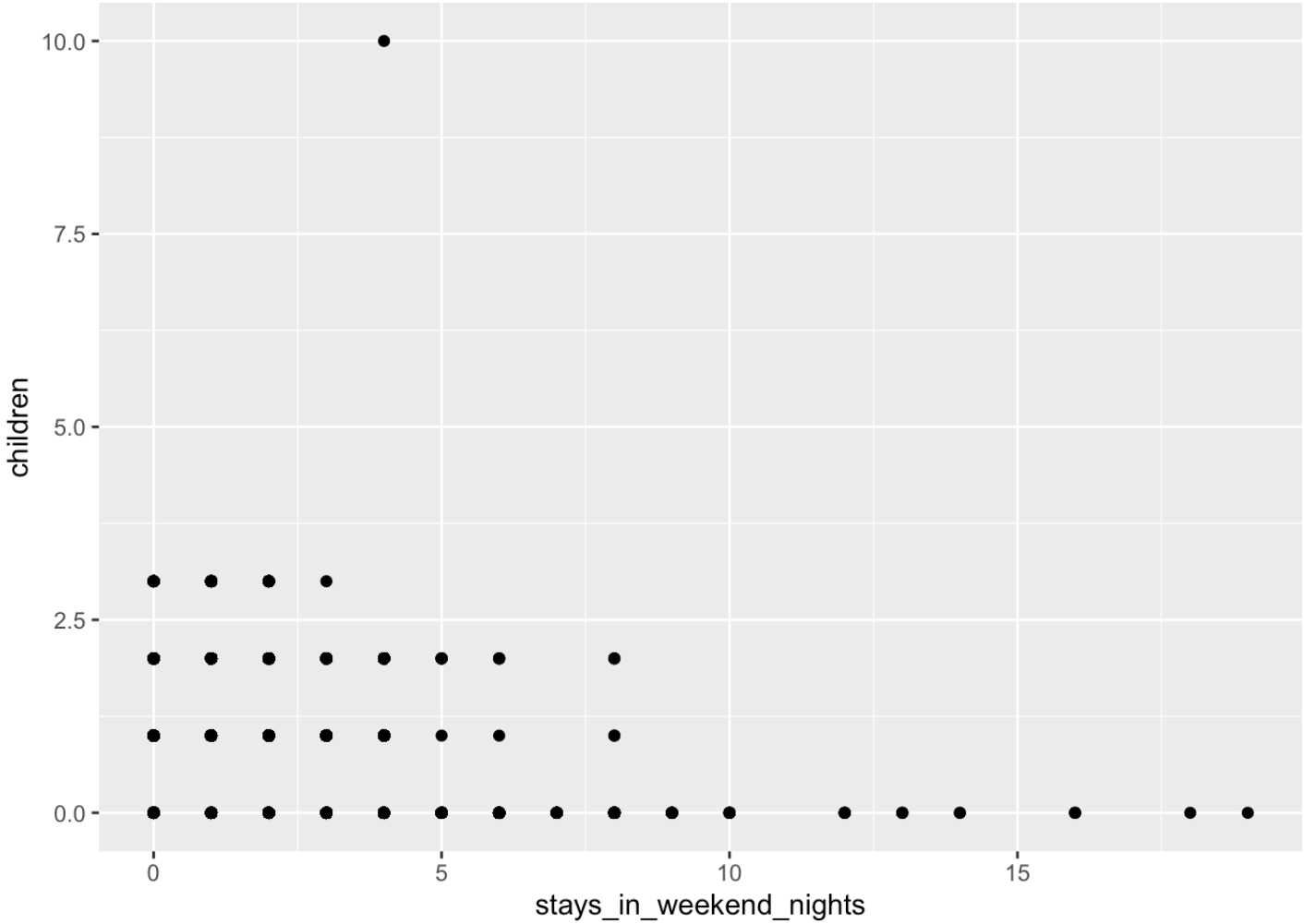
This can be accomplished by knowing, what group of guests book the most weekend nights to target that group in a new marketing campaign

Suggests that guests without children book most weekend nights. We need to verify the hypothesis presented by the Stakeholder.

```
library(ggplot2)

ggplot(data = hotel_bookings) +
 geom_point(mapping = aes(x = stays_in_weekend_nights
 , y = children))
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

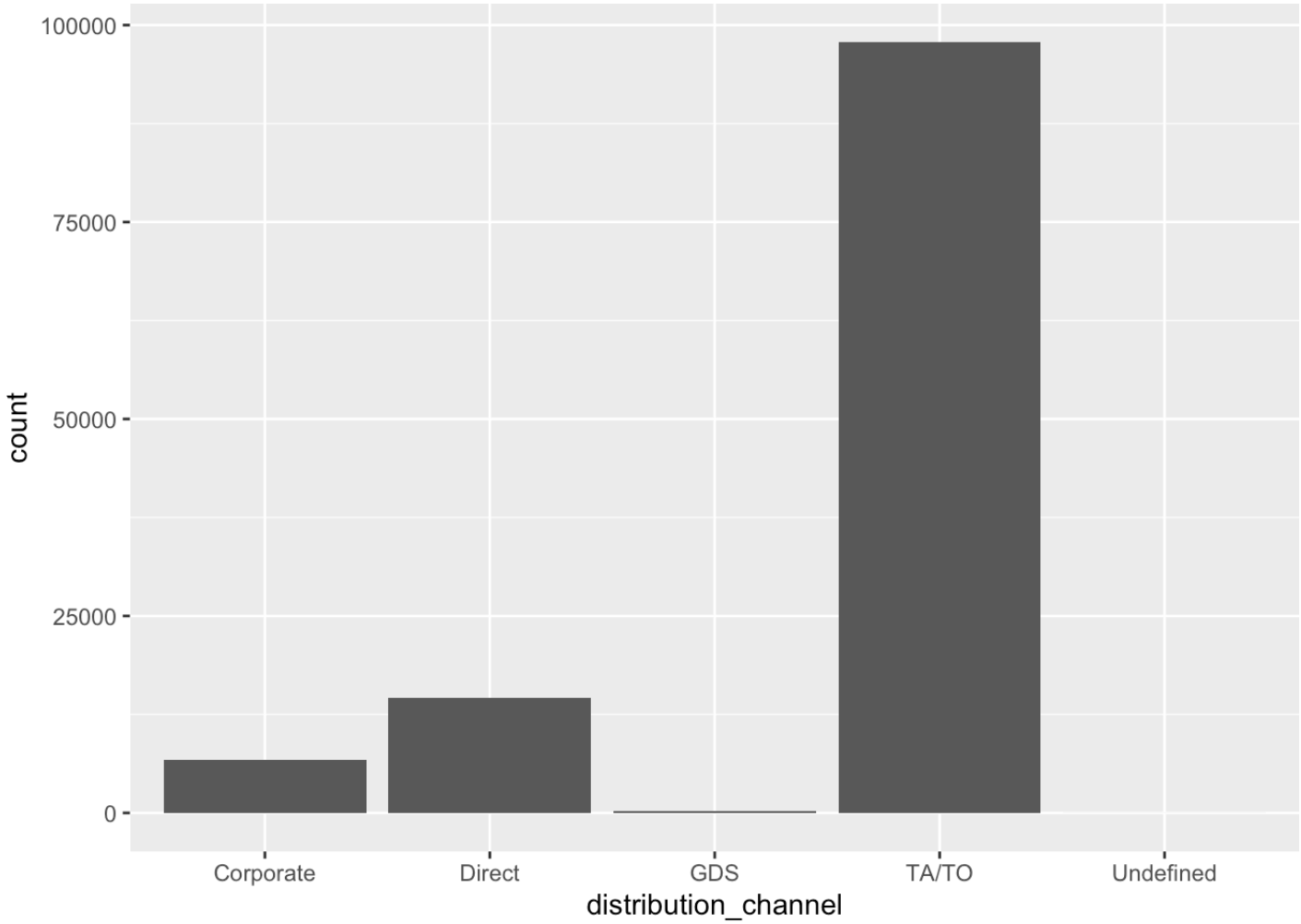


According to the plotted data visualizations, we can see the hypothesis and estimation made by the stakeholder are correct.The guests without children have done the most weekend bookings.

Stakeholder is interested in developing promotions based on different booking distributions, but first they need to know how many of the transactions are occurring for each different distribution type

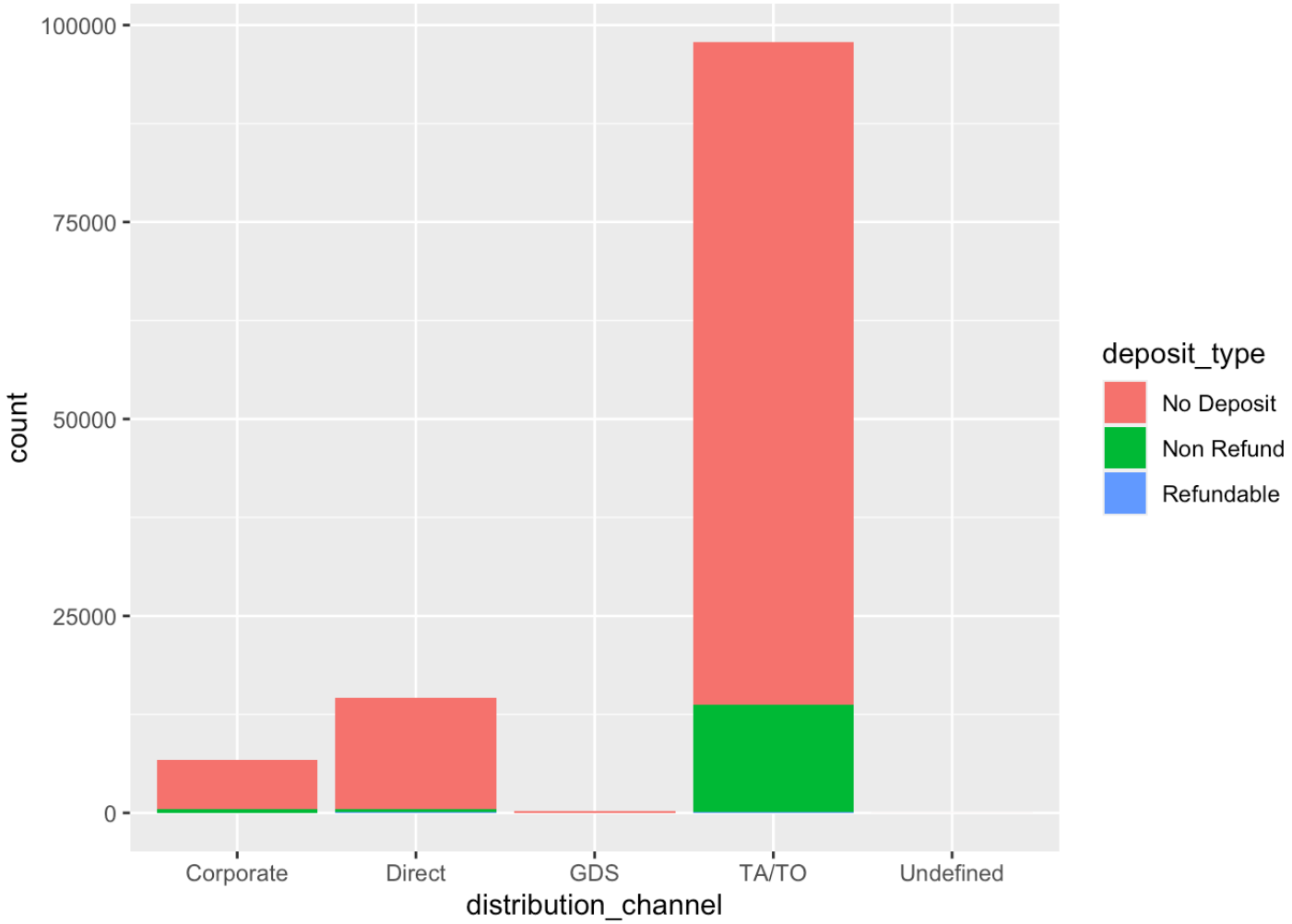
Consider the following bar chart representing transaction count for distribution channels.

```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = distribution_channel))
```



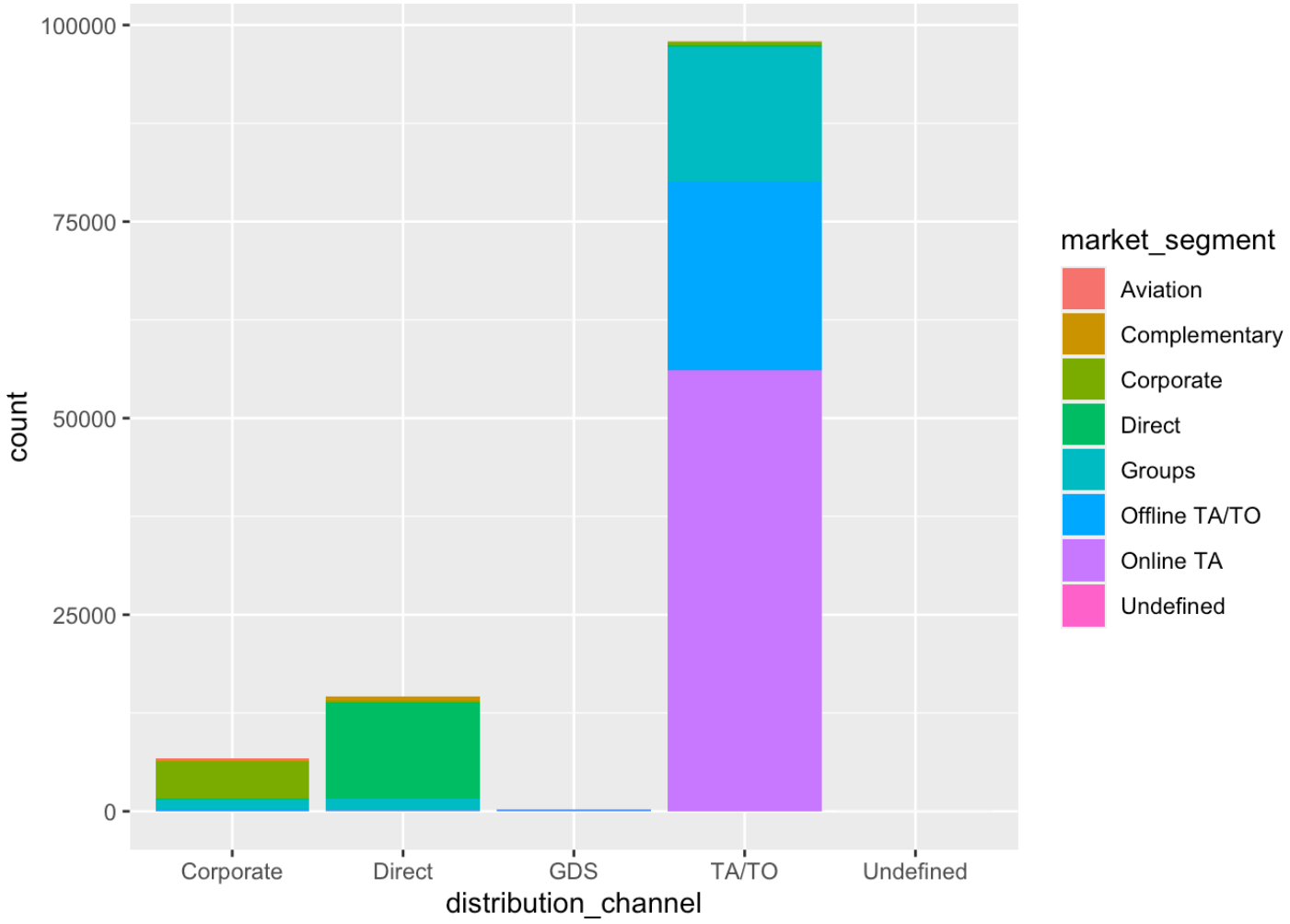
After exaninig the bar chart, The Stakeholder wants to know if the number of bookings for each distribution type is different depending on whether or not there was a deposit or what market segment they represent.

```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = distribution_channel, fill = deposit_type))
```



Consider the report according to the Market Segment as given below

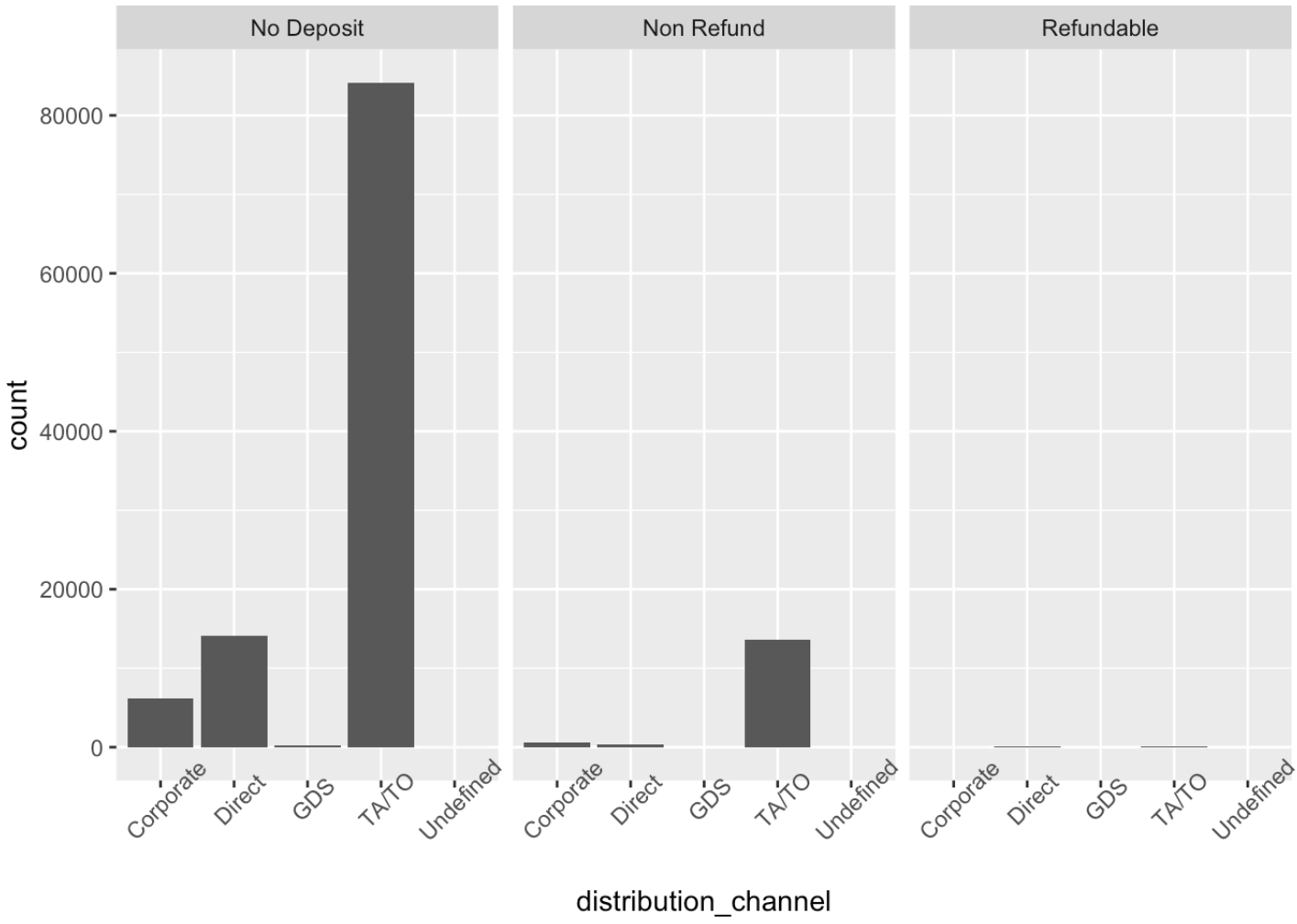
```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = distribution_channel, fill = market_segment))
```



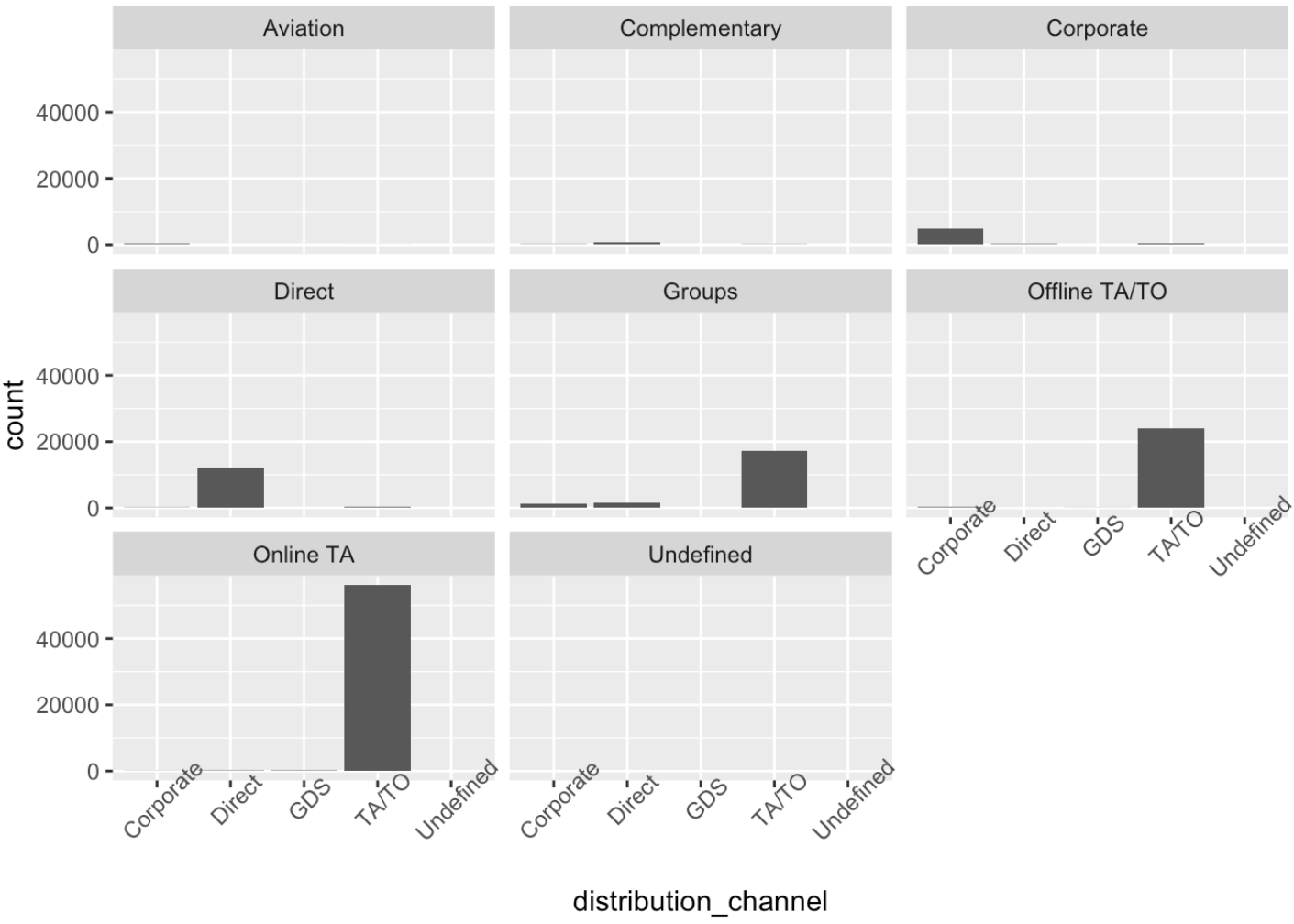
After reviewing the effective data representation stakeholder presents the following statement,

Stakeholder Statement: Stakeholder wants to explore separatley deposit types and market segments to help them understand the differences more clearly.

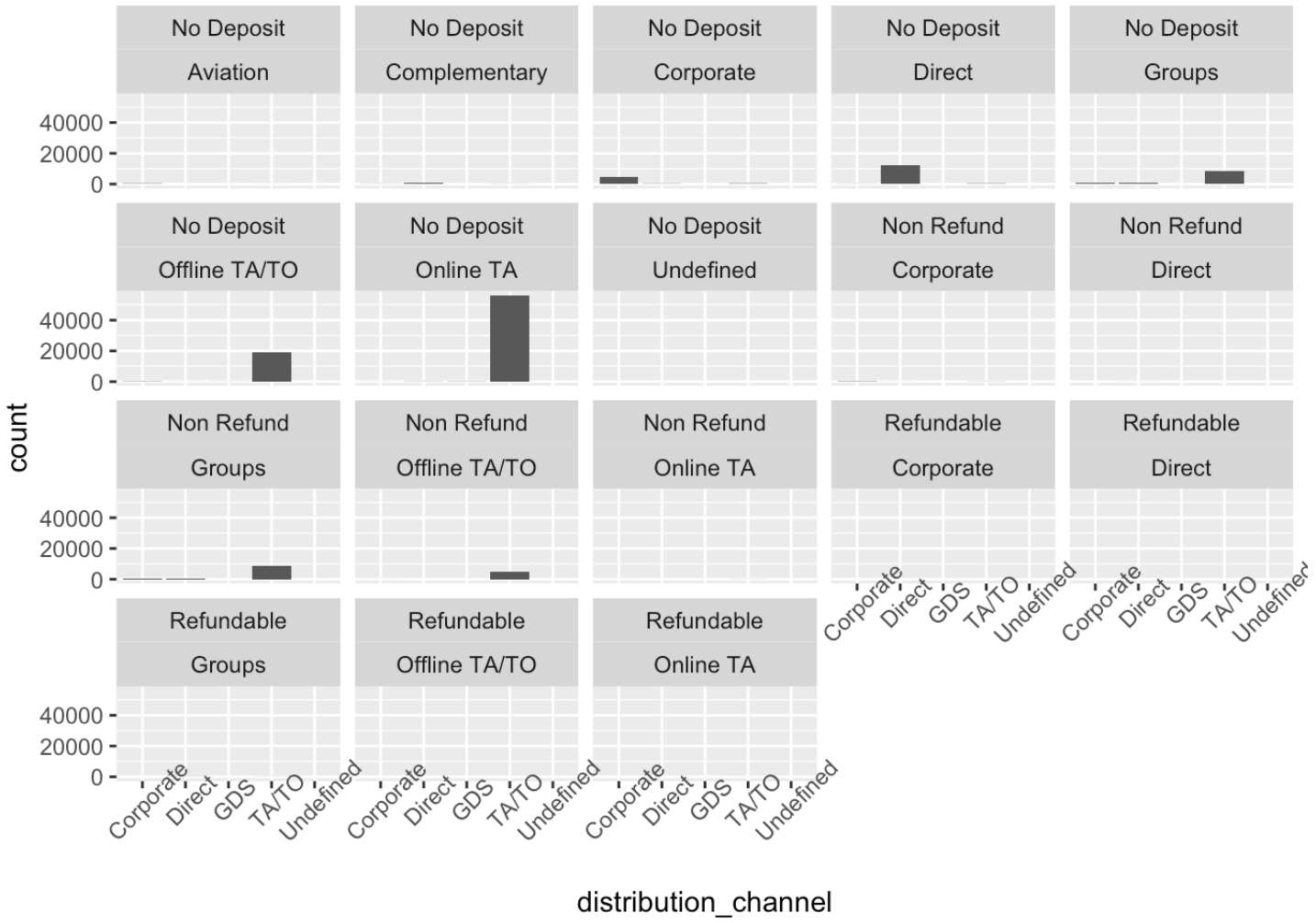
```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = distribution_channel)) +
  facet_wrap(~deposit_type) +
  theme(axis.text.x = element_text(angle = 45))
```




```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = distribution_channel)) +
  facet_wrap(~market_segment) +
  theme(axis.text.x = element_text(angle = 45))
```

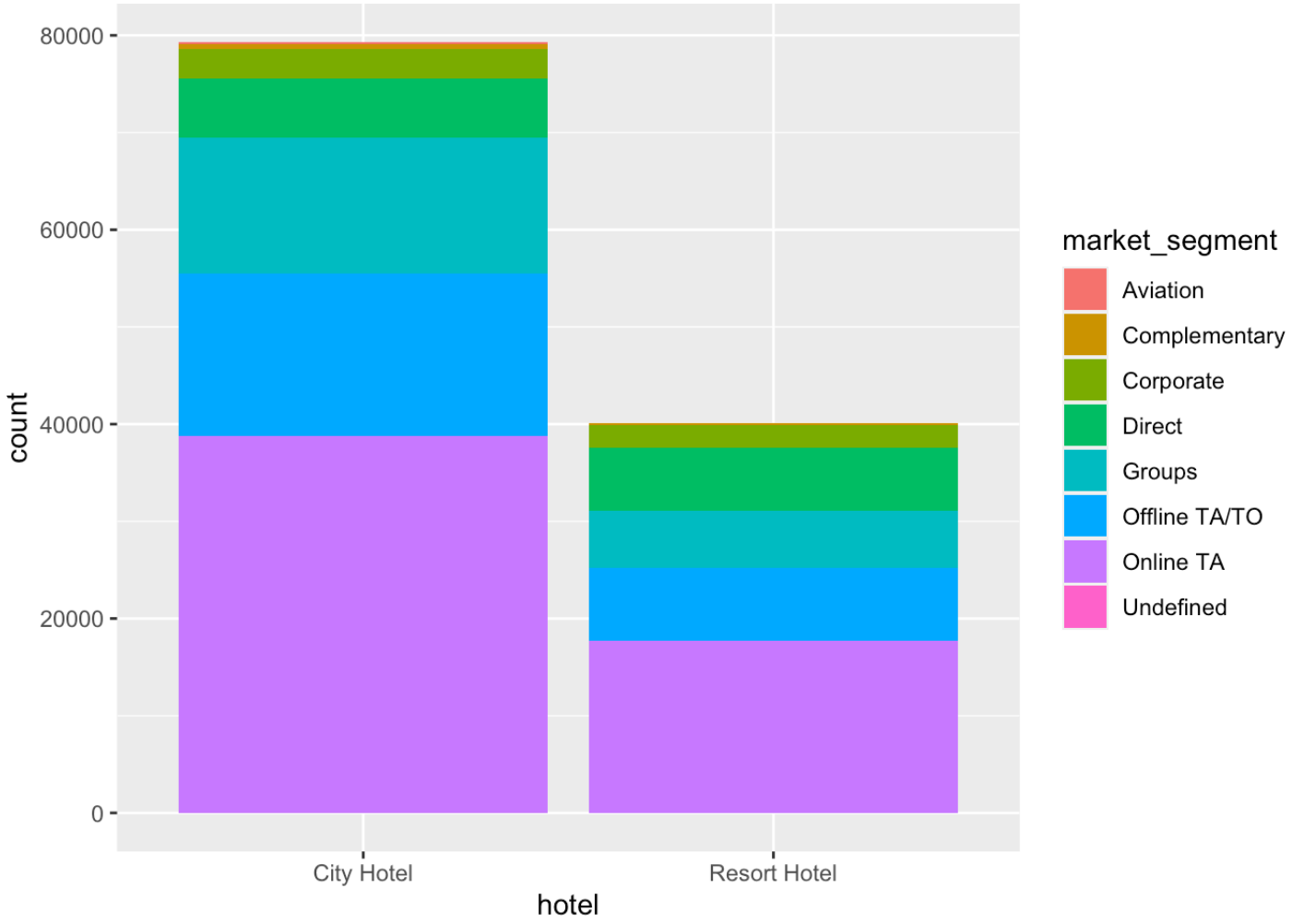



```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = distribution_channel)) +
  facet_wrap(~deposit_type~market_segment) +
  theme(axis.text.x = element_text(angle = 45))
```



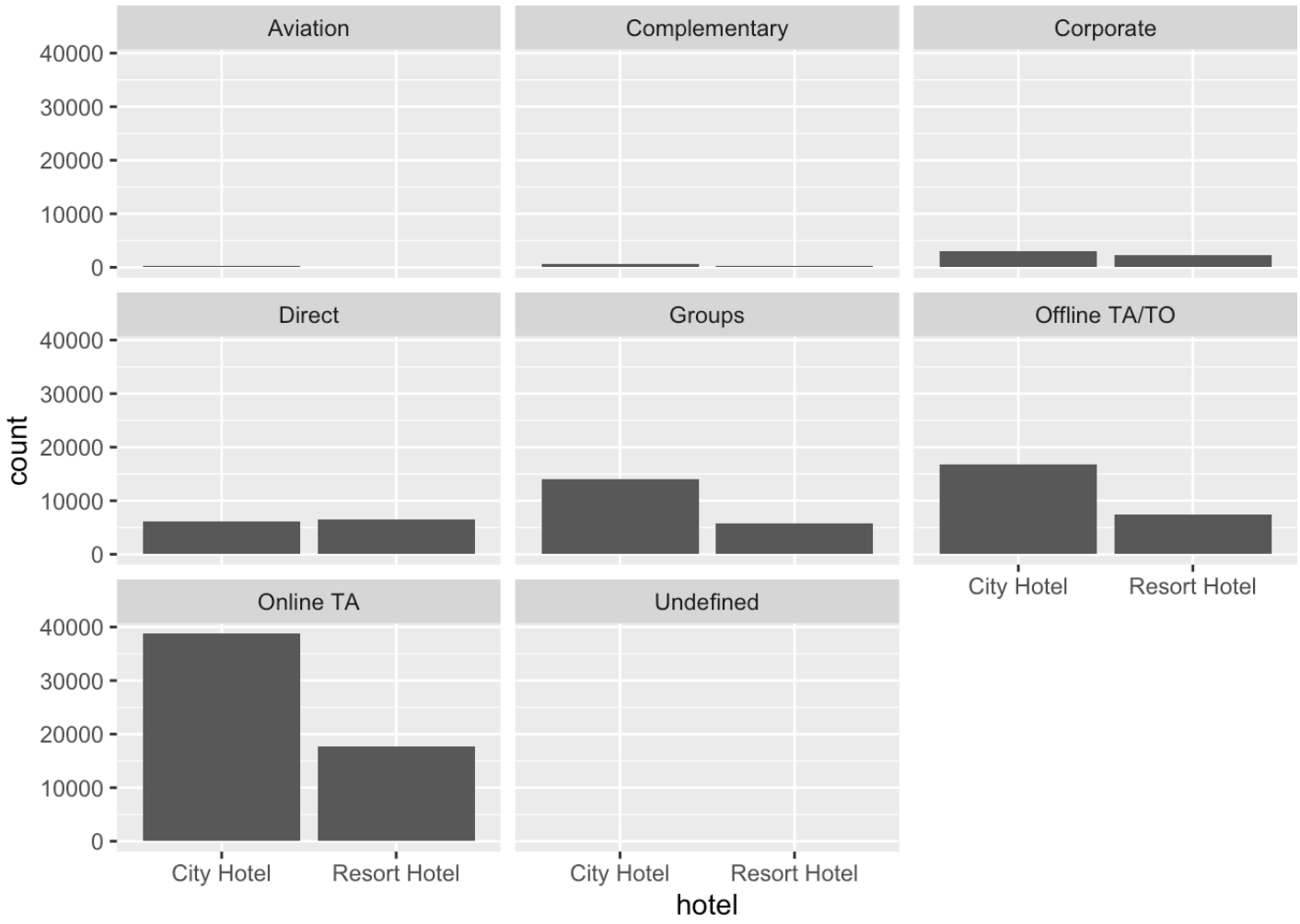
Stakeholder wants to run a family-friendly promotion targeting key market segments. She wants to know which market segments generate the largest number of bookings, and where these bookings are made (city hotels or resort hotels).

```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = hotel, fill = market_segment))
```



From the bar chart it's difficult to compare the size of the market segments at the top of the bars. Consider the follwing chart with modifications.

```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = hotel)) +
  facet_wrap(~market_segment)
```




```
library(tidyverse)
library(dplyr)
```

Stakeholder decides to send the promotion to families that make online bookings for city hotels. The online segment is the fastest growing segment, and families tend to spend more at city hotels than other types of guests.

```
onlineta_city_hotels_v2 <- hotel_bookings %>% filter(hotel=="City Hotel") %>% filter(
market_segment=="Online TA")

head(onlineta_city_hotels_v2)
```

```
##        hotel is_canceled lead_time arrival_date_year arrival_date_month
## 1 City Hotel           1        88              2015               July
## 2 City Hotel           1        65              2015               July
## 3 City Hotel           1        92              2015               July
## 4 City Hotel           1       100              2015               July
## 5 City Hotel           1        79              2015               July
```

```
## 6 City Hotel            1        63               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                      2
## 4                       27                         2                      0
## 5                       27                         2                      0
## 6                       27                         2                      1
##   stays_in_week_nights adults children babies meal country market_segment
## 1                    4      2        0      0   BB     PRT      Online TA
## 2                    4      1        0      0   BB     PRT      Online TA
## 3                    4      2        0      0   BB     PRT      Online TA
## 4                    2      2        0      0   BB     PRT      Online TA
## 5                    3      2        0      0   BB     PRT      Online TA
## 6                    3      1        0      0   BB     PRT      Online TA
##   distribution_channel is_repeated_guest previous_cancellations
## 1                TA/TO                 0                      0
## 2                TA/TO                 0                      0
## 3                TA/TO                 0                      0
## 4                TA/TO                 0                      0
## 5                TA/TO                 0                      0
## 6                TA/TO                 0                      0
##   previous_bookings_not_canceled reserved_room_type assigned_room_type
## 1                              0                  A                  A
## 2                              0                  A                  A
## 3                              0                  A                  A
## 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               0   No Deposit     9    NULL                    0     Transient
## 2               0   No Deposit     9    NULL                    0     Transient
## 3               0   No Deposit     9    NULL                    0     Transient
## 4               0   No Deposit     9    NULL                    0     Transient
## 5               0   No Deposit     9    NULL                    0     Transient
## 6               0   No Deposit     9    NULL                    0     Transient
##    adr required_car_parking_spaces total_of_special_requests reservation_status
## 1 76.5                           0                          1           Canceled
## 2 68.0                           0                          1           Canceled
## 3 76.5                           0                          2           Canceled
## 4 76.5                           0                          1           Canceled
## 5 76.5                           0                          1           Canceled
## 6 68.0                           0                          0           Canceled
##   reservation_status_date
## 1              2015-07-01
## 2              2015-04-30
## 3              2015-06-23
```

```
## 4                2015-04-02
## 5                2015-06-25
## 6                2015-06-25
```