Comprehensive Analysis of Hotel Booking Trends

Introduction

In today's fast-paced travel industry, hotel bookings are influenced by numerous factors such as seasonality, customer preferences, and global events. For instance, imagine a small but popular hotel chain, Sunset Getaways, located in Europe. This hotel saw a sharp increase in bookings in the summer months, primarily driven by vacationers seeking coastal escapes. However, challenges arose as the hotel faced a series of cancellations during the off-peak seasons due to unforeseen events like travel restrictions or changing weather patterns.

To better understand these fluctuations and improve business strategies, Sunset Getaways decided to analyze its hotel booking data. By exploring booking patterns, customer preferences, and seasonal trends, the hotel aimed to optimize its marketing campaigns, improve room allocation, and ensure higher customer satisfaction.

This report delves into the hotel booking dataset, analyzing the trends and insights that can help hoteliers like Sunset Getaways make informed decisions, address operational challenges, and anticipate customer needs.

Dataset Overview

This dataset contains 119390 observations for a City Hotel and a Resort Hotel. Each observation represents a hotel booking between the 1st of July 2015 and 31st of August 2017, including booking that effectively arrived and booking that were canceled.

- **Dataset Name**: Hotel Booking
- **Source**: For more detailed information, check out hotel booking.
- **Time Period**: 2015 to 2017
- **Number of Records**: 119390
- **Number of Variables**: 32

Summery about data

1- **hotel**: Type of hotel (e.g., Resort Hotel, City Hotel).

2- **is_canceled**: A binary field indicating whether a booking was canceled (1 = Canceled, 0 = Not Canceled).

3- **lead_time**: Number of days between the booking date and the arrival date.

4- **arrival_date_year**: Year of arrival.

5- **arrival_date_month**: Month of arrival.

6- **arrival_date_week_number**: The week number of the arrival date.

7- **arrival_date_day_of_month**: The day of the month of the arrival date.

8- **stays_in_weekend_nights**: Number of weekend nights (Friday, Saturday) stayed by guests.

9- **stays_in_week_nights**: Number of weeknights (Monday to Thursday) stayed by guests.

10- **adults**: Number of adults in the booking.

11- **children**: Number of children in the booking.

12- **babies**: Number of babies in the booking.

13- **meal**: Type of meal booked (e.g., BB = Bed & Breakfast, HB = Half Board).

14- **country**: Country of origin of the guest.

15- **market_segment**: Market segment designation (e.g., Direct, Corporate, Online TA).

16- **distribution_channel**: Booking distribution channel (e.g., Direct, TA/TO).

17- **is_repeated_guest:** A binary field indicating whether the guest is a repeat guest (1 = Yes, 0 = No).

18- **previous_cancellations**: Number of previous cancellations by the guest.

19- **previous_bookings_not_canceled**: Number of previous bookings that were not canceled by the guest.

20- **reserved_room_type**: Type of room reserved.

21- **assigned_room_type**: Type of room assigned to the guest upon arrival.

22- **booking_changes**: Number of changes made to the booking.

23**- deposit_type**: Type of deposit made (e.g., No Deposit, Refundable, Non Refundable).

24- **agent**: ID of the booking agent.

25- **company**: ID of the company that made the booking.

26- **days_in_waiting_list**: Number of days the booking was on a waiting list before being confirmed.

27- **customer_type**: Type of customer (e.g., Transient, Contract, Group).

28- **adr**: Average Daily Rate (ADR), a measure of the revenue earned per occupied room.

29- **required_car_parking_spaces**: Number of car parking spaces required.

30- **total_of_special_requests**: Number of special requests made by the guest (e.g., high floor, late check-in).

31- **reservation_status**: The status of the reservation (e.g., Canceled, Check-Out).

32- **reservation_status_date**: Date of the last status change (e.g., when it was canceled or checked out).

Methodology

The analysis was conducted on a hotel booking dataset to explore key trends, patterns, and insights regarding customer behavior, pricing strategies, and operational performance. The following steps outline the methodology used for this analysis:

1. **Data Cleaning and Preparation**
   The dataset was first cleaned to remove any inconsistencies or missing values. Null values in important columns were handled using appropriate strategies such as imputation or removal based on the significance of the missing data. For instance, columns like `reservation_status`, `adr` (Average Daily Rate), and `arrival_date` were critical for analysis, and thus missing values were either replaced or carefully handled.

   The `arrival_date_month` and `arrival_date_year` were converted into appropriate factor levels to ensure proper sorting and analysis for time-based trends.

2. **Exploratory Data Analysis (EDA)**
   Initial exploratory analysis was performed to understand the overall structure of the data and The EDA focused on:

   – **Monthly and Yearly Trends**: The distribution of bookings across different months and years was analyzed using bar charts and line graphs to identify peak booking periods and seasonal variations.

   – **Customer Type and ADR Analysis**: A detailed analysis of the `customer_type` and its relationship with the `adr` (Average Daily Rate) was performed to understand how different customer segments contributed to revenue.

   – **Country-wise ADR Analysis**: Countries were ranked based on their ADR values, and the top 10 countries were visualized using a treemap to highlight the countries generating the highest revenue.

   – **Reservation Status and Cancellations**: Reservation statuses were explored across different months and hotel types to examine patterns in confirmed, canceled, and no-show bookings. Additionally, the cancellation rate was calculated for different market segments to identify trends.

3. **Visualization Techniques**
   Multiple types of visualizations were used to communicate insights effectively:

- **Bar Charts**: Used to visualize monthly booking trends, reservation statuses by hotel type, and cancellation rates by market segment. Labels were added for better interpretability.

- **Line Graphs with Smoothing**: Line graphs with a smoothing function (using `geom_smooth()`) were used to show the ADR trend over time, providing a clear indication of fluctuations and trends in pricing.

- **Treemap**: To visualize the top 10 countries by ADR, a treemap was chosen to illustrate the relative contribution of each country to the total revenue.

- **Pie Charts**: Pie charts were used to represent the proportion of reservation statuses, giving a clear breakdown of confirmed, canceled, and no-show reservations.

4. **Statistical Analysis**
   Key statistical measures such as the mean, median, and standard deviation of ADR were calculated for various customer segments and countries. These measures provided insight into the distribution of revenue across different segments and helped identify outliers or significant trends.

5. **Segmentation and Grouping**

   - **Customer Type**: Grouping customers based on type (e.g., transient, group, contract) to examine differences in behavior and revenue contribution.

   - **Market Segment**: Analyzing how bookings from different market segments (e.g., corporate, direct, online travel agencies) performed in terms of cancellations and ADR.

   - **Reservation Status**: Grouping reservations by status (confirmed, canceled, no-show) to understand operational challenges.

6. **Interpretation of Visualizations**

   - Seasonal peaks in bookings and how these impact revenue.

   - The performance of different market segments in terms of cancellations and revenue.

   - How customer type and country of origin influence the ADR.

   - How cancellation trends vary across market segments and months.

7. **Software and Tools Used** The analysis was performed using R and the following libraries:

   - **ggplot2**: For creating visualizations including bar charts, line graphs, and pie charts.

- **dplyr**: For data manipulation and summarization.
- **treemap**: To create a treemap for country-wise ADR analysis.
- **scales**: For formatting percentages and currency values in the charts.

Key Objectives

- **Analyze Booking Trends** Explore how hotel bookings vary over time, focusing on monthly and yearly trends. Identify peak booking periods to inform marketing and resource allocation strategies.
- **Examine Cancellation Rates** Investigate the cancellation rates among different customer segments (e.g., market segment, lead time). Analyze factors contributing to higher cancellation rates and recommend ways to mitigate them.
- **Understand Customer Demographics and Preferences** Segment customers by demographic factors (e.g., country, customer type) to identify key markets and tailor offerings. Also, analyze whether customers receive their preferred room type and how this impacts satisfaction.
- **Assess Revenue and Pricing Strategies** Analyze the average daily rate (ADR) across customer types, countries, and market segments to identify which segments are most profitable. Suggest strategies for optimizing pricing to maximize revenue.
- **Assess Revenue and Pricing Strategies (ADR)**: Analyze the average daily rate (ADR) across customer types, countries, and market segments to identify which segments are most profitable. Suggest strategies for optimizing pricing to maximize revenue.
- **Evaluate Special Requests and Resource Utilization**: Investigate the distribution of special requests (e.g., parking spaces, additional services) to understand the impact on hotel resources and operational planning.
- **Identify Reservation Status Patterns**: Examine how reservation statuses (confirmed, canceled, no-show) differ by month and hotel type to inform hotel management strategies for improving occupancy rates and reducing no-shows.

Code: Loading Data

```
########################
####  Loading Data  ####
########################
install.packages("openxlsx")
install.packages("readr")
install.packages('treemap')

library(dplyr)
library(tidyr)
library(readr)
```

```
library(ggplot2)
library(openxlsx)
library(treemap)

df <- read_csv("c:\\Users\\Ahmed\\Downloads\\hotel_bookings.csv")
View(df)
```

Code: Data Cleaning

```
#########################
####  Data Cleaning  ####
#########################
df %>%
  mutate(
    hotel = as.character(hotel),
    is_canceled = as.character(is_canceled),
    lead_time = as.numeric(lead_time),
    arrival_date_year = as.numeric(arrival_date_year),
    arrival_date_month = as.character(arrival_date_month),
    arrival_date_week_number = as.numeric(arrival_date_week_number),
    arrival_date_day_of_month = as.numeric(arrival_date_day_of_month),
    stays_in_weekend_nights = as.numeric(stays_in_weekend_nights),
    stays_in_week_nights = as.numeric(stays_in_week_nights),
    adults = as.numeric(adults),
    children = as.numeric(children),
    babies = as.numeric(babies),
    meal = as.character(meal),
    country = as.character(country),
    market_segment = as.character(market_segment),
    distribution_channel = as.character(distribution_channel),
    is_repeated_guest = as.character(is_repeated_guest),
    previous_cancellations = as.numeric(previous_cancellations),
    previous_bookings_not_canceled =
as.numeric(previous_bookings_not_canceled),
    reserved_room_type = as.character(reserved_room_type),
    assigned_room_type = as.character(assigned_room_type),
    booking_changes = as.numeric(booking_changes),
    deposit_type = as.character(deposit_type),
    agent = as.character(agent),
    company = as.character(company),
    days_in_waiting_list = as.numeric(days_in_waiting_list),
    customer_type = as.character(customer_type),
    adr = as.numeric(adr),
    required_car_parking_spaces = as.numeric(required_car_parking_spaces),
    total_of_special_requests = as.numeric(total_of_special_requests),
    reservation_status = as.character(reservation_status),
  )

df %>% mutate(across(where(is.numeric), ~replace_na(., 0)))
```

```r
df <- df %>% mutate(agent = ifelse(agent == "NULL", "Not Found", agent))

df <- df %>% mutate(company = ifelse(company == "NULL", "Not Found",
company))

df %>% distinct(.keep_all=TRUE)

View(df)
```

Code: Data Manipulation

```r
###########################
####  Data Manipulation  ####
###########################

head(df, 10)
############################
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights
	<chr>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>
1	Resort Hotel	0	342	2015	July	27	1	0
2	Resort Hotel	0	737	2015	July	27	1	0
3	Resort Hotel	0	7	2015	July	27	1	0
4	Resort Hotel	0	13	2015	July	27	1	0
5	Resort Hotel	0	14	2015	July	27	1	0
6	Resort Hotel	0	14	2015	July	27	1	0
7	Resort Hotel	0	0	2015	July	27	1	0
8	Resort Hotel	0	9	2015	July	27	1	0
9	Resort Hotel	1	85	2015	July	27	1	0
10	Resort Hotel	1	75	2015	July	27	1	0

```r
str(df)
############################
```

```
> str(df)
tibble [119,390 × 32] (S3: tbl_df/tbl/data.frame)
 $ hotel                         : chr [1:119390] "Resort Hotel" "Resort Hotel" "Resort Hotel" "Resort Hotel" ...
 $ is_canceled                   : num [1:119390] 0 0 0 0 0 0 0 0 1 1 ...
 $ lead_time                     : num [1:119390] 342 737 7 13 14 14 0 9 85 75 ...
 $ arrival_date_year             : num [1:119390] 2015 2015 2015 2015 2015 ...
 $ arrival_date_month            : chr [1:119390] "July" "July" "July" "July" ...
 $ arrival_date_week_number      : num [1:119390] 27 27 27 27 27 27 27 27 27 27 ...
 $ arrival_date_day_of_month     : num [1:119390] 1 1 1 1 1 1 1 1 1 1 ...
 $ stays_in_weekend_nights       : num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ stays_in_week_nights          : num [1:119390] 0 0 1 1 2 2 2 2 3 3 ...
 $ adults                        : num [1:119390] 2 2 1 1 2 2 2 2 2 2 ...
 $ children                      : num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ babies                        : num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ meal                          : chr [1:119390] "BB" "BB" "BB" "BB" ...
 $ country                       : chr [1:119390] "PRT" "PRT" "GBR" "GBR" ...
 $ market_segment                : chr [1:119390] "Direct" "Direct" "Direct" "Corporate" ...
 $ distribution_channel          : chr [1:119390] "Direct" "Direct" "Direct" "Corporate" ...
 $ is_repeated_guest             : num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ previous_cancellations        : num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ previous_bookings_not_canceled: num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ reserved_room_type            : chr [1:119390] "C" "C" "A" "A" ...
 $ assigned_room_type            : chr [1:119390] "C" "C" "C" "A" ...
 $ booking_changes               : num [1:119390] 3 4 0 0 0 0 0 0 0 0 ...
 $ deposit_type                  : chr [1:119390] "No Deposit" "No Deposit" "No Deposit" "No Deposit" ...
 $ agent                         : chr [1:119390] "Not Found" "Not Found" "Not Found" "304" ...
 $ company                       : chr [1:119390] "Not Found" "Not Found" "Not Found" "Not Found" ...
 $ days_in_waiting_list          : num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ customer_type                 : chr [1:119390] "Transient" "Transient" "Transient" "Transient" ...
 $ adr                           : num [1:119390] 0 0 75 75 98 ...
 $ required_car_parking_spaces   : num [1:119390] 0 0 0 0 0 0 0 0 0 0 ...
 $ total_of_special_requests     : num [1:119390] 0 0 0 0 1 1 0 1 1 0 ...
 $ reservation_status            : chr [1:119390] "Check-Out" "Check-Out" "Check-Out" "Check-Out" ...
 $ reservation_status_date       : chr [1:119390] "7/1/2015" "7/1/2015" "7/2/2015" "7/2/2015" ...
```

Alt text

```r
dim(df)
############################
```

```
names(df)
################################

summary(df)
################################
```

```
> dim(df)
[1] 119390     32
> names(df)
 [1] "hotel"                          "is_canceled"                    "lead_time"                      "arrival_date_year"              "arrival_date_month"             "arrival_date_week_number"
 [7] "arrival_date_day_of_month"      "stays_in_weekend_nights"        "stays_in_week_nights"           "adults"                         "children"                       "babies"
[13] "meal"                           "country"                        "market_segment"                 "distribution_channel"           "is_repeated_guest"              "previous_cancellations"
[19] "previous_bookings_not_canceled" "reserved_room_type"             "assigned_room_type"             "booking_changes"                "deposit_type"                   "agent"
[25] "company"                        "days_in_waiting_list"           "customer_type"                  "adr"                            "required_car_parking_spaces"    "total_of_special_requests"
[31] "reservation_status"             "reservation_status_date"
> summary(df)
    hotel            is_canceled       lead_time   arrival_date_year arrival_date_month arrival_date_week_number arrival_date_day_of_month stays_in_weekend_nights stays_in_week_nights     adults
 Length:119390      Min.   :0.0000   Min.   :  0   Min.   :2015      Length:119390      Min.   : 1.00            Min.   : 1.0              Min.   : 0.0000         Min.   : 0.0          Min.   : 0.000
 Class :character   1st Qu.:0.0000   1st Qu.: 18   1st Qu.:2016      Class :character   1st Qu.:16.00            1st Qu.: 8.0              1st Qu.: 0.0000         1st Qu.: 1.0          1st Qu.: 2.000
 Mode  :character   Median :0.0000   Median : 69   Median :2016      Mode  :character   Median :28.00            Median :16.0              Median : 1.0000         Median : 2.0          Median : 2.000
                    Mean   :0.3704   Mean   :104   Mean   :2016                         Mean   :27.17            Mean   :15.8              Mean   : 0.9276         Mean   : 2.5          Mean   : 1.856
                    3rd Qu.:1.0000   3rd Qu.:160   3rd Qu.:2017                         3rd Qu.:38.00            3rd Qu.:23.0              3rd Qu.: 2.0000         3rd Qu.: 3.0          3rd Qu.: 2.000
                    Max.   :1.0000   Max.   :737   Max.   :2017                         Max.   :53.00            Max.   :31.0              Max.   :19.0000         Max.   :50.0          Max.   :55.000

    children           babies               meal             country          market_segment     distribution_channel is_repeated_guest previous_cancellations previous_bookings_not_canceled
 Min.   : 0.0000   Min.   : 0.000000   Length:119390      Length:119390      Length:119390      Length:119390        Min.   :0.00000   Min.   : 0.00000       Min.   : 0.0000
 1st Qu.: 0.0000   1st Qu.: 0.000000   Class :character   Class :character   Class :character   Class :character     1st Qu.:0.00000   1st Qu.: 0.00000       1st Qu.: 0.0000
 Median : 0.0000   Median : 0.000000   Mode  :character   Mode  :character   Mode  :character   Mode  :character     Median :0.00000   Median : 0.00000       Median : 0.0000
 Mean   : 0.1039   Mean   : 0.007949                                                                                 Mean   :0.03191   Mean   : 0.08712       Mean   : 0.1371
 3rd Qu.: 0.0000   3rd Qu.: 0.000000                                                                                 3rd Qu.:0.00000   3rd Qu.: 0.00000       3rd Qu.: 0.0000
 Max.   :10.0000   Max.   :10.000000                                                                                 Max.   :1.00000   Max.   :26.00000       Max.   :72.0000
 NA's   :4
 reserved_room_type assigned_room_type booking_changes   deposit_type          agent             company          days_in_waiting_list customer_type           adr            required_car_parking_spaces
 Length:119390      Length:119390      Min.   : 0.0000   Length:119390      Length:119390      Length:119390      Min.   :  0.000      Length:119390      Min.   :  -6.38   Min.   :0.00000
 Class :character   Class :character   1st Qu.: 0.0000   Class :character   Class :character   Class :character   1st Qu.:  0.000      Class :character   1st Qu.:  69.29   1st Qu.:0.00000
 Mode  :character   Mode  :character   Median : 0.0000   Mode  :character   Mode  :character   Mode  :character   Median :  0.000      Mode  :character   Median :  94.58   Median :0.00000
                                       Mean   : 0.2211                                                            Mean   :  2.321                         Mean   : 101.83   Mean   :0.06252
                                       3rd Qu.: 0.0000                                                            3rd Qu.:  0.000                         3rd Qu.: 126.00   3rd Qu.:0.00000
                                       Max.   :21.0000                                                            Max.   :391.000                         Max.   :5400.00   Max.   :8.00000

 total_of_special_requests reservation_status reservation_status_date
 Min.   :0.0000            Length:119390      Length:119390
 1st Qu.:0.0000            Class :character   Class :character
 Median :0.0000            Mode  :character   Mode  :character
 Mean   :0.5714
 3rd Qu.:1.0000
 Max.   :5.0000
```

```
dataFilteredByCountry <- filter(df, df['country']=='USA')
View(dataFilteredByCountry)
################################

dataFilteredByArrivalDateMonth <- filter(df,
df['arrival_date_month']=='July')
View(dataFilteredByArrivalDateMonth)
################################

dataFilteredByReservationStatus <- filter(df, df['reservation_status'] ==
'Check-Out')
View(dataFilteredByReservationStatus)
################################

write.xlsx(df, file = "c:\\Users\\Ahmed\\Downloads\\my_data.xlsx")
```

Data Visualization

1. Monthly Booking Trends by Years

Insight: While the general seasonal trend persists across years, the overall volume of bookings has varied slightly from year to year. Some years might have experienced higher demand during peak months compared to others. This could be due to external factors such as economic conditions, special events, or promotions affecting customer behavior.

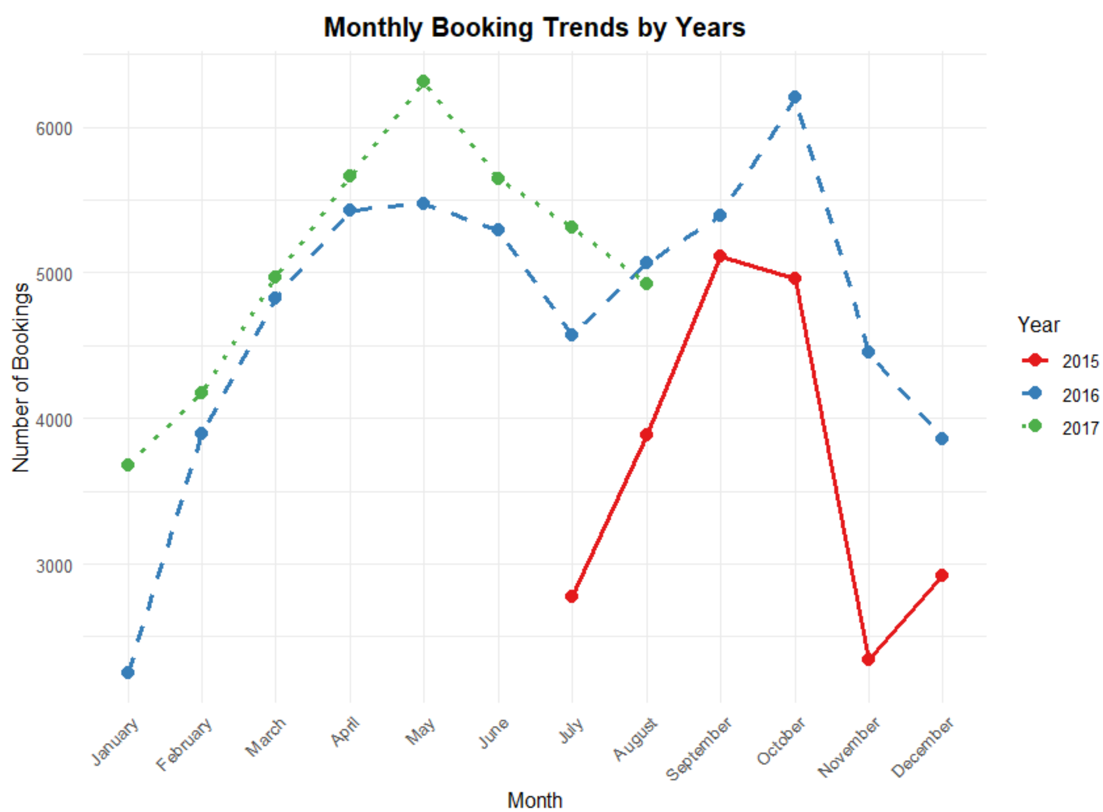

Chart 1 Code

```
## = Monthly Booking Trends by Years = ##
df$arrival_date_year <- as.factor(df$arrival_date_year)

ggplot(df, aes(x = arrival_date_month, group = arrival_date_year, color =
arrival_date_year, linetype = arrival_date_year)) +
  geom_line(stat = 'count', size = 1.2) +
  geom_point(stat = 'count', size = 3) +
  theme_minimal() +
  labs(title = "Monthly Booking Trends by Years",
       x = "Month",
       y = "Number of Bookings",
       color = "Year",
       linetype = "Year") +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
        axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_color_brewer(palette = "Set1") +
  scale_linetype_manual(values = c("solid", "dashed", "dotted"))
```

2. Monthly Booking Trends Over the Years

Insight: Bookings show a seasonal trend, with peaks in certain months like July and August. This suggests that these are the busiest months for hotel bookings, possibly coinciding with holidays or vacation seasons. On the other hand, months like January and November tend to have fewer bookings, indicating low-demand periods.

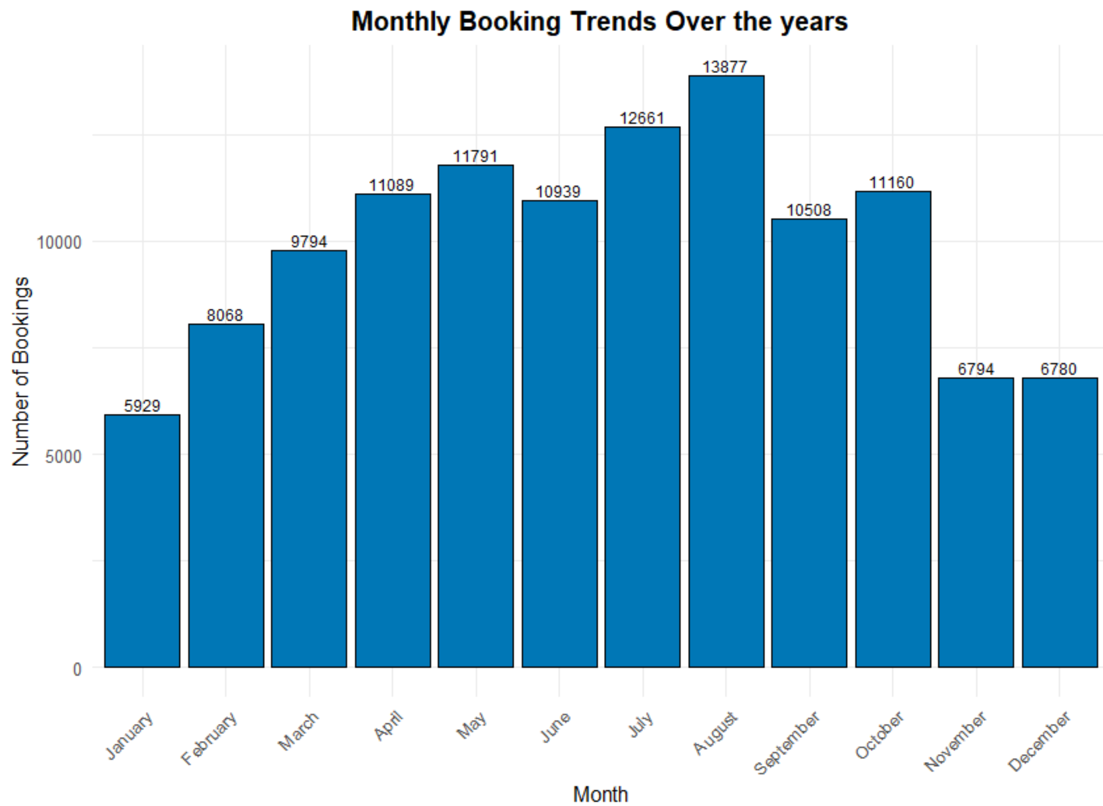

Chart 2 Code

```r
## = Monthly Booking Trends Over the years = ##
df$arrival_date_month <- factor(df$arrival_date_month, levels = c("January",
"February", "March", "April", "May", "June", "July", "August", "September",
"October", "November", "December"))

ggplot(df, aes(x = arrival_date_month)) +
  geom_bar(fill = '#0077b6', color='black') +
  geom_text(stat = 'count', aes(label = ..count..), vjust = -0.3, size = 3) +
  theme_minimal() +
  labs(title = "Monthly Booking Trends Over the years", x = "Month", y =
"Number of Bookings") +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
        axis.text.x = element_text(angle = 45, hjust = 1))
```

3. Average Daily Rate (ADR) by Customer Type

Insight: The average daily rate (ADR) differs across customer types. For example, transient customers may have a higher ADR compared to group bookings. This suggests that individual customers are willing to pay more per day compared to groups, who might benefit from group discounts or promotions.

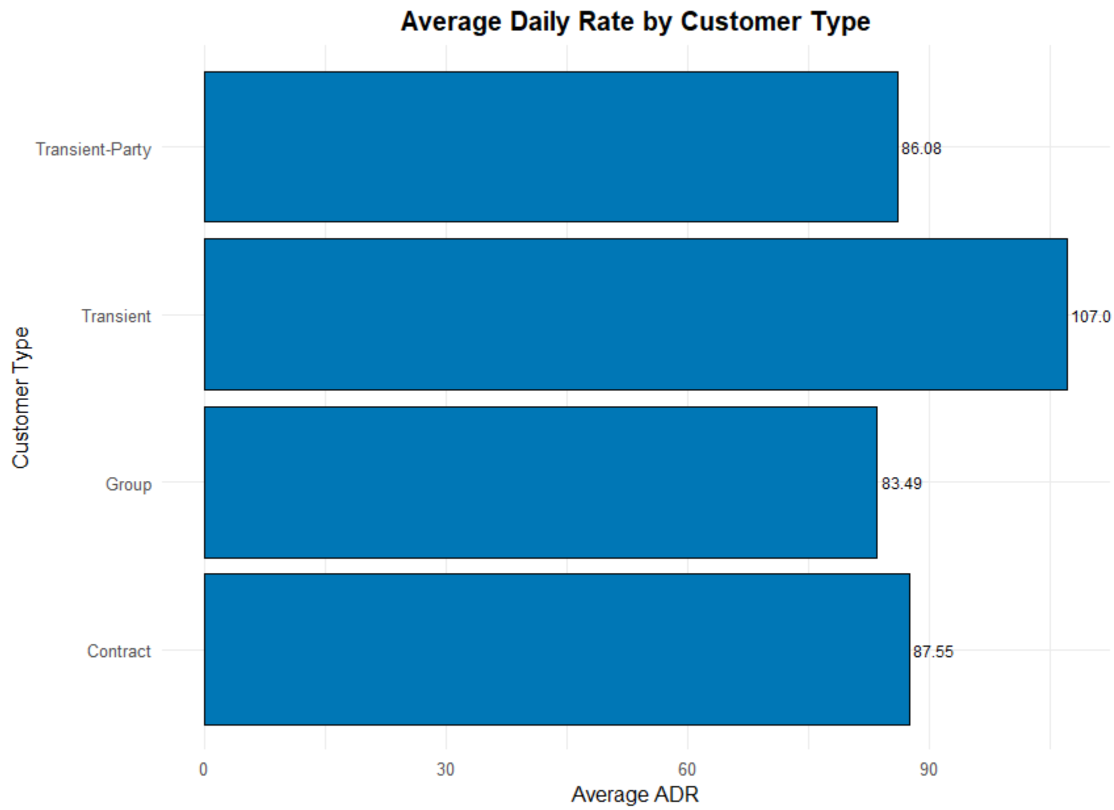

Chart 3 Code

```
## = Average Daily Rate By Customer Type = ##
adr_by_customer <- df %>%
  group_by(customer_type) %>%
  summarize(avg_adr = mean(adr, na.rm = TRUE))

ggplot(adr_by_customer, aes(x = customer_type, y = avg_adr)) +
  geom_bar(stat = "identity", fill = "#0077b6", color = "black") +
  geom_text(aes(label = round(avg_adr, 2)), hjust = -0.1, size = 3) +
  labs(title = "Average Daily Rate by Customer Type", x = "Customer Type", y
= "Average ADR") +
  theme_minimal() +
  coord_flip() +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
```

4. Top Ten Countries by ADR

Insight: The chart shows the countries with the highest ADR. The countries at the top could indicate regions with wealthier travelers or those who typically opt for higher-end accommodation. Understanding the source of high-paying customers can help in targeting marketing efforts.

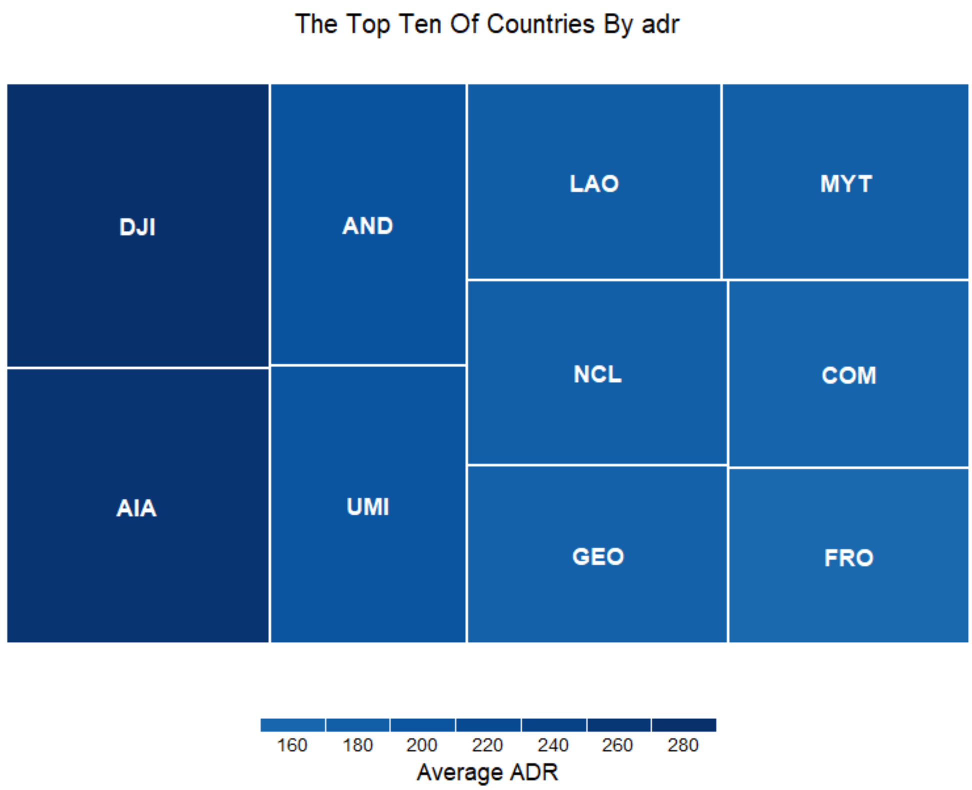

Chart 4 Code

```r
## = The Top Ten Of Countries By ADR = ##
adr_by_country <- df %>%
  group_by(country) %>%
  summarize(avg_adr = mean(adr, na.rm = TRUE)) %>%
  arrange(desc(avg_adr)) %>%
  slice(1:10)

treemap(adr_by_country,
        index = "country",
        vSize = "avg_adr",
        vColor = "avg_adr",
        type = "value",
        palette = "Blues",
        title = "The Top Ten Of Countries By adr",
        fontsize.labels = 12,
        fontsize.title = 14,
        fontcolor.labels = "white",
        border.col = "white",
        title.legend = "Average ADR",
```

```
        draw.labels = TRUE,
#        label.format = list(size = 12, format = function(x) sprintf("%.2f",
x))
)
```

5. ADR Trend Over Months

Insight: The Average Daily Rate (ADR) shows a clear seasonal trend, peaking during the summer months (June to August), while rates tend to dip during the winter months. The smooth line (LOESS curve) highlights that the ADR consistently follows a predictable seasonal cycle, which could help in pricing strategies for future bookings.

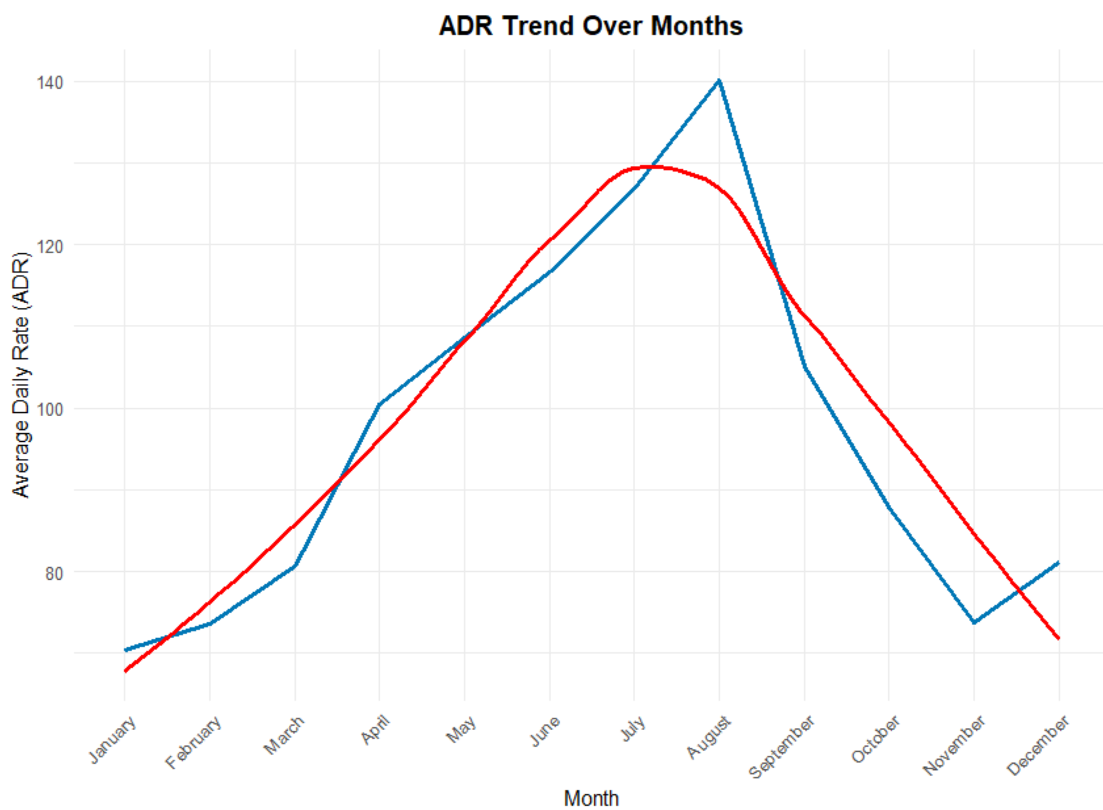

Chart 5 Code

```
## = ADR Trend Over Months = ##
adr_by_month <- df %>%
  group_by(arrival_date_month) %>%
  summarize(avg_adr = mean(adr, na.rm = TRUE))
df$arrival_date_month <- factor(df$arrival_date_month, levels = month.name)

ggplot(adr_by_month, aes(x = arrival_date_month, y = avg_adr, group = 1)) +
  geom_line(color = "#0077b6", size = 1) +
  geom_smooth(method = "loess", se = FALSE, color = "red", size = 1) +
  labs(title = "ADR Trend Over Months", x = "Month", y = "Average Daily Rate
```

```
(ADR)") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(limits = month.name)
```

6. Reservation Status by Month

Insight: The distribution of reservation statuses (e.g., confirmed, canceled) varies by month. There may be a higher number of confirmed reservations during peak months, while cancellation rates could rise during off-season months when customers may reconsider or cancel their travel plans.

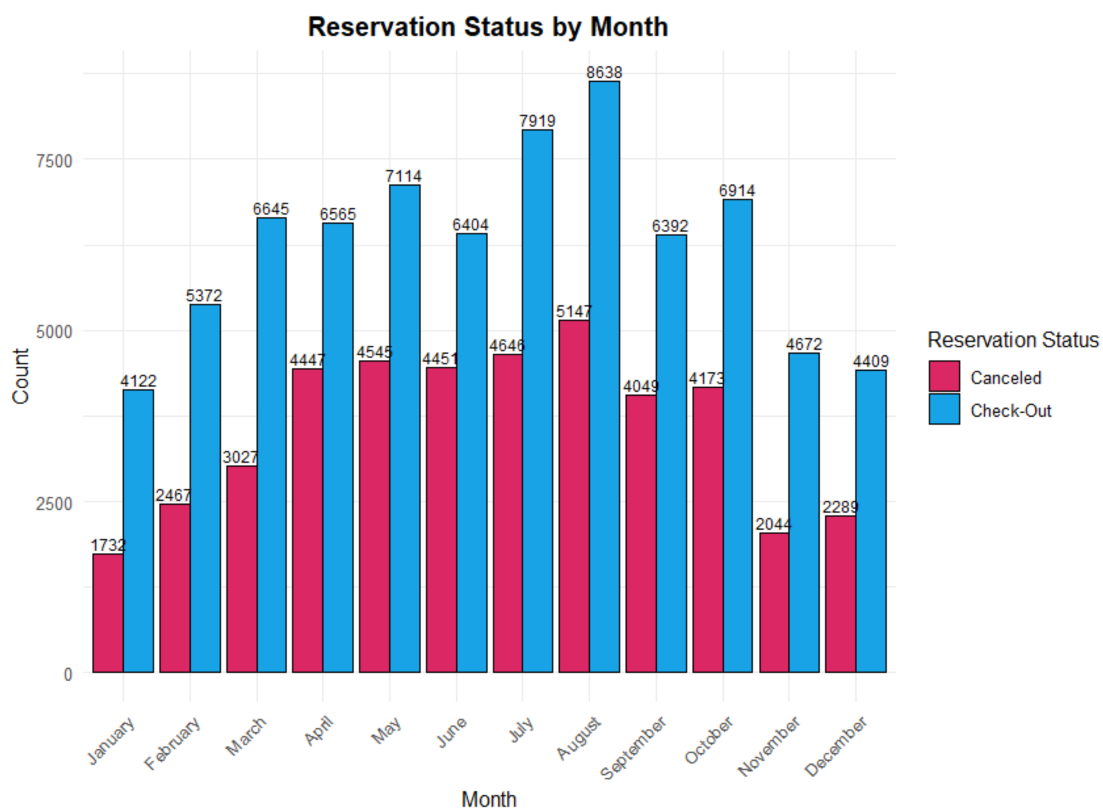

Chart 6 Code

```
## = Reservation status per month = ##
reservation_status_count <- df %>%
  group_by(reservation_status) %>%
  summarize(count = n())

df_filtered <- df %>%
  filter(reservation_status %in% c("Check-Out", "Canceled"))

ggplot(df_filtered, aes(x = arrival_date_month, fill = reservation_status)) +
```

```r
  geom_bar(position = "dodge", color = "black") +
  geom_text(stat = "count", aes(label = ..count..),
            position = position_dodge(width = 0.9), vjust = -0.3, size = 3) +
  labs(title = "Reservation Status by Month",
       x = "Month",
       y = "Count",
       fill = "Reservation Status") +
  theme_minimal() +
 theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
        axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(limits = month.name) +
  scale_fill_manual(values = c("Check-Out" = "#19a3e7", "Canceled" =
"#db2763"))
```

7. Reservation Statuses

Insight: The pie chart reveals the proportion of each reservation status, indicating whether most reservations are confirmed, canceled, or otherwise modified. If a large proportion of bookings are canceled, this could signal the need to review the cancellation policies or customer communication strategies.

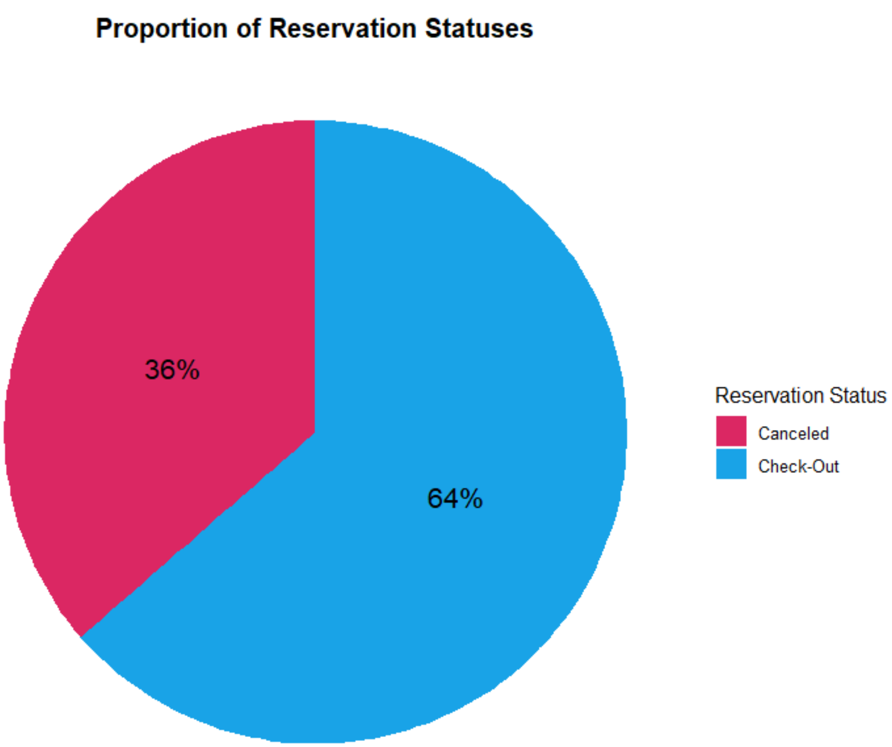

Chart 7 Code

```r
## = Proportion of Reservation Statuses = ##
reservation_status_count <- df %>%
```

```
  group_by(reservation_status) %>%
  filter(reservation_status %in% c("Check-Out", "Canceled")) %>%
  summarize(count = n())

ggplot(reservation_status_count, aes(x = "", y = count, fill =
reservation_status)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(title = "Proportion of Reservation Statuses", fill = "Reservation
Status") +
  theme_void() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 14)) +
  geom_text(aes(label = scales::percent(count/sum(count))), position =
position_stack(vjust = 0.5), size = 5) +
  scale_fill_manual(values = c("Check-Out" = "#19a3e7", "Canceled" =
"#db2763"))
```

8. Reservation Status by Hotel Type

Insight: The reservation status differs between hotel types. For example, city hotels may have higher cancellations compared to resort hotels, which could suggest that city bookings are more short-term or subject to last-minute changes, while resort bookings may be planned well in advance and more stable.

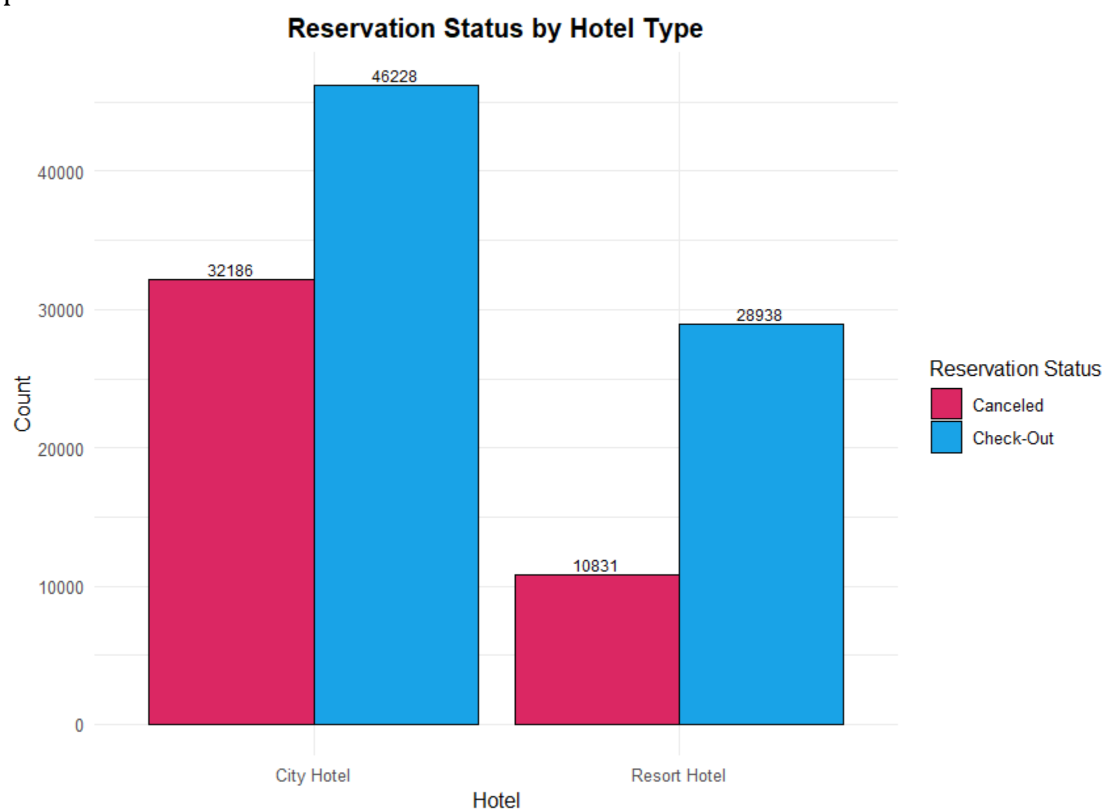

Chart 8 Code

```
## = Reservation Status By Hotel Type = ##
reservation_status_count <- df %>%
  group_by(reservation_status) %>%
  summarize(count = n())

df_filtered <- df %>%
  filter(reservation_status %in% c("Check-Out", "Canceled"))

ggplot(df_filtered, aes(x = hotel, fill = reservation_status)) +
  geom_bar(position = "dodge", color = "black") +
  geom_text(stat = 'count', aes(label = ..count..),
            position = position_dodge(width = 0.9), vjust = -0.3, size = 3) +
  labs(title = "Reservation Status by Hotel Type",
       x = "Hotel",
       y = "Count",
       fill = "Reservation Status") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 14)) +
  scale_fill_manual(values = c("Check-Out" = "#19a3e7", "Canceled" =
"#db2763"))
```

9. Special Requests by Reservation Status

Insight: This chart shows the total number of special requests associated with "Check-Out" and "Canceled" reservations. If "Canceled" reservations have significantly more special requests, it may indicate that guests with more specific needs are more likely to cancel their reservations.

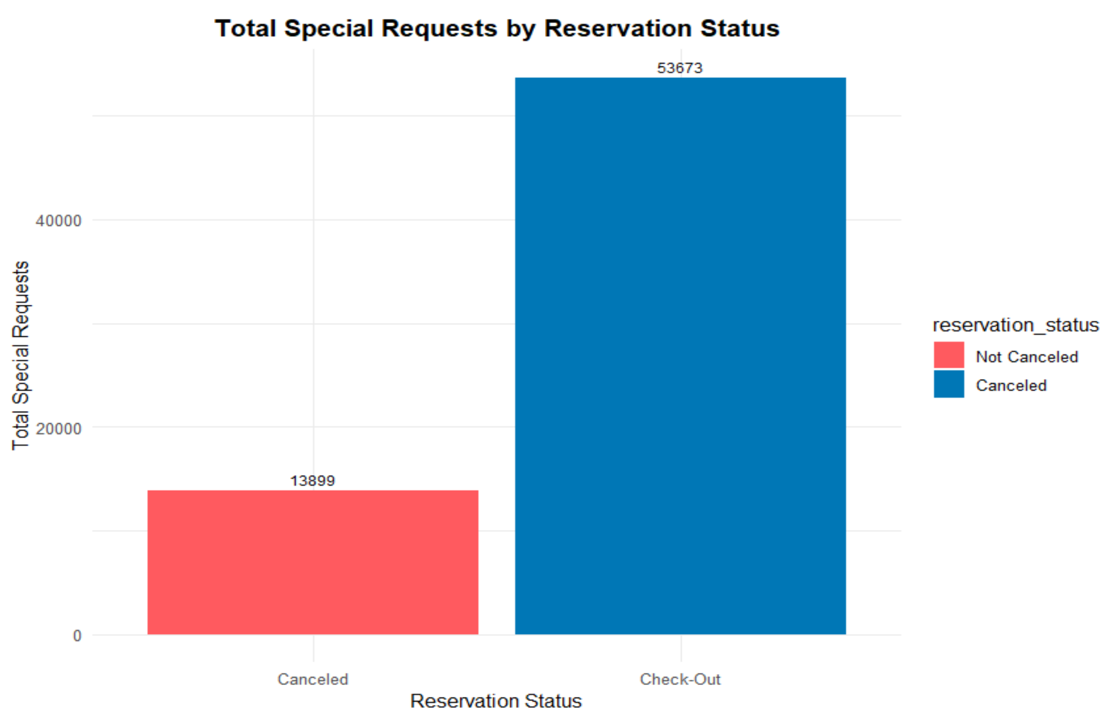

Chart 9 Code

```
## = Special Requests by Reservation Status = ##
df_filtered <- df %>%
  filter(reservation_status %in% c("Check-Out", "Canceled"))

ggplot(df_filtered, aes(x = reservation_status, y =
total_of_special_requests, fill = reservation_status)) +
  geom_bar(stat = "summary", fun = "sum") +
  geom_text(stat = "summary", fun = "sum", aes(label = ..y..), vjust = -0.5,
size = 3) +
  labs(title = "Total Special Requests by Reservation Status", x =
"Reservation Status", y = "Total Special Requests") +
  theme_minimal() +
  scale_fill_manual(values = c("#ff5a5f", "#0077b6"), labels = c("Not
Canceled", "Canceled")) +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
```

10. Reservation Status by Meal Type

Insight: This chart compares the number of reservations with different meal plans ("BB" for Bed and Breakfast, "FB" for Full Board, "HB" for Half Board, and "SC" for Self-Catering) between "Check-Out" and "Canceled" statuses. If a particular meal type is more common among canceled reservations, it could suggest a link between meal preferences and reservation stability.

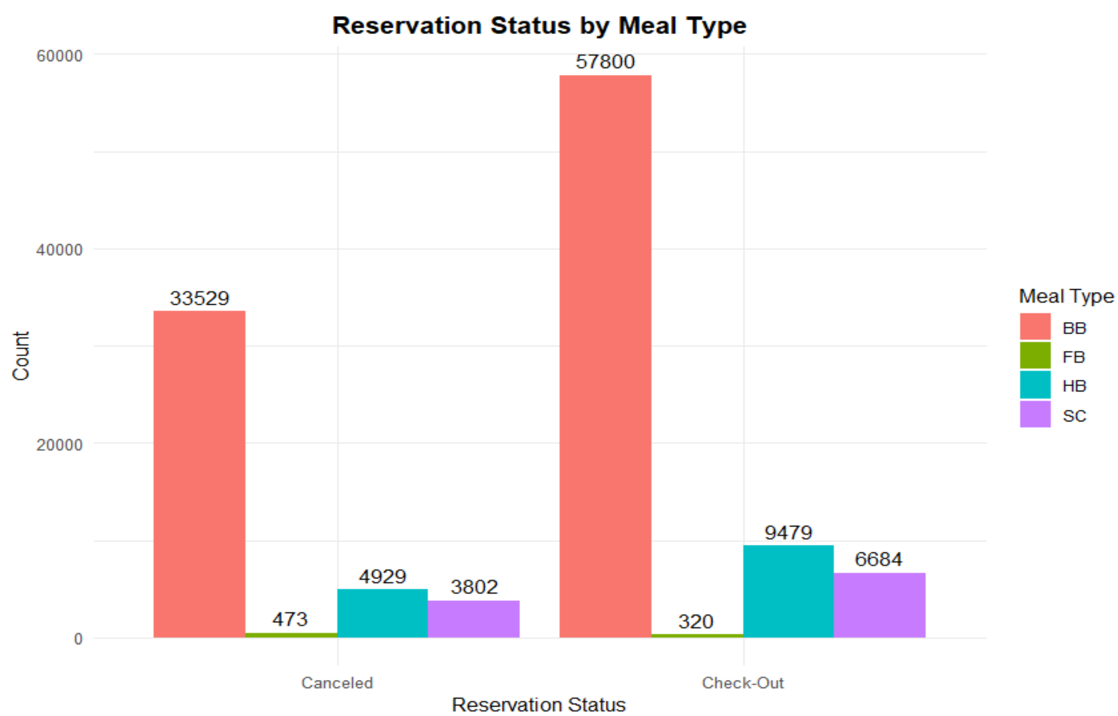

Chart 10 Code

```
## = Reservation Status by Meal Type = ##
df_filtered <- df %>%
  filter(reservation_status %in% c("Check-Out", "Canceled")) %>%
  filter(meal %in% c("BB", "FB", "HB", "SC"))

ggplot(df_filtered, aes(x = reservation_status, fill = meal)) +
  geom_bar(position = "dodge") +
  labs(title = "Reservation Status by Meal Type", x = "Reservation Status", y
= "Count", fill = "Meal Type") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold")) +
  geom_text(stat = "count", aes(label = ..count..), position =
position_dodge(width = 0.9), vjust = -0.5)
```

11. Weekend vs Weekday Stay Preferences by Customer Type

Insight: This chart illustrates the average number of weekend versus weekday nights stayed by different customer types. If certain customer types prefer weekend stays significantly more than weekday stays (or vice versa), it could inform targeted marketing strategies or pricing adjustments.

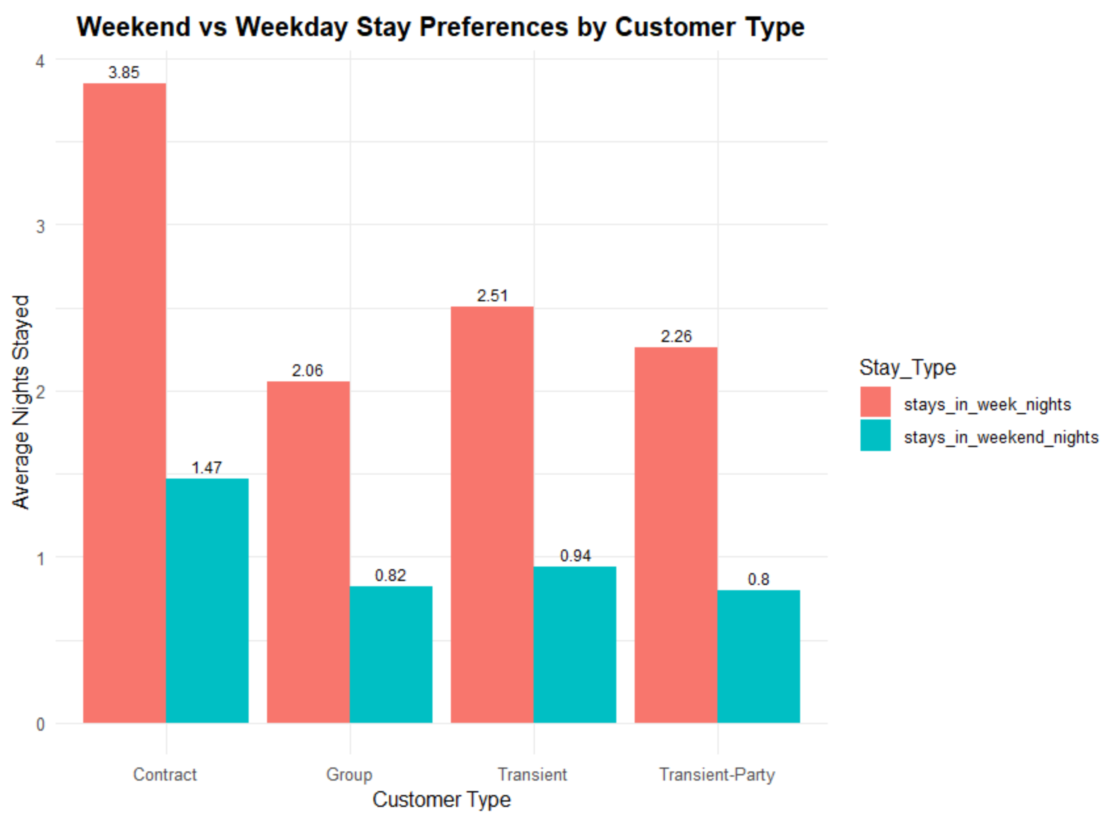

Chart 11 Code

```
## = Weekend vs Weekday Stay Preferences by Customer Type = ##
df_long <- df %>%
  pivot_longer(cols = c(stays_in_weekend_nights, stays_in_week_nights),
names_to = "Stay_Type", values_to = "Nights")

ggplot(df_long, aes(x = customer_type, y = Nights, fill = Stay_Type)) +
  geom_bar(stat = "summary", fun = "mean", position = "dodge") +
  geom_text(stat = "summary", fun = "mean", aes(label = round(..y.., 2)),
position = position_dodge(width = 0.9), vjust = -0.5, size = 3) +
  labs(title = "Weekend vs Weekday Stay Preferences by Customer Type", x =
"Customer Type", y = "Average Nights Stayed") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
```

12. Distribution of Special Requests by Hotel Type

Insight: This chart shows the average number of special requests per hotel type. A higher average in one hotel type suggests that guests at that hotel tend to make more special requests. This can help in understanding the service expectations or preferences at different hotels.

Chart 12 Code

```
## = Distribution of Special Requests by Hotel Type = ##
```

```
ggplot(df, aes(x = hotel, y = total_of_special_requests, fill = hotel)) +
  geom_bar(stat = "summary", fun = "mean", position = "dodge") +
  geom_text(stat = "summary", fun = "mean", aes(label = round(..y.., 2)),
vjust = -0.5, size = 3) +
  labs(title = "Average Special Requests by Hotel Type", x = "Hotel Type", y
= "Average Total Special Requests") +
  theme_minimal() +
  scale_fill_manual(values = c("#0077b6", "#ff5a5f")) +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
```

13. The Count of Parking Spaces Requested by Hotel Type

Insight: This chart displays the total number of parking spaces requested by hotel type. If one hotel type shows a significantly higher count of requested parking spaces, it might indicate a greater reliance on personal vehicles or a larger number of guests with cars at that hotel.

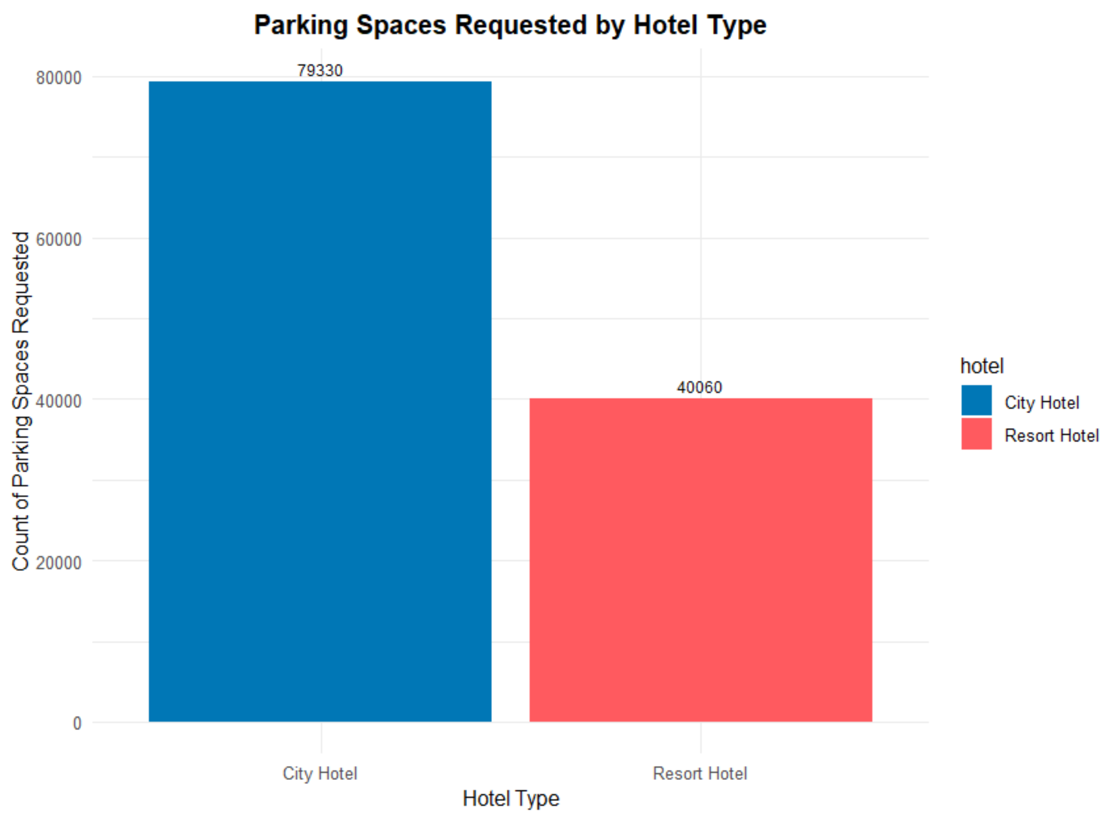

Chart 13 Code

```
## = The Count of Parking Spaces Requested by Hotel Type = ##
ggplot(df, aes(x = hotel, fill = hotel)) +
  geom_bar() +
  geom_text(stat = "count", aes(label = ..count..), vjust = -0.5, size = 3) +
  labs(title = "Parking Spaces Requested by Hotel Type", x = "Hotel Type", y
= "Count of Parking Spaces Requested") +
  theme_minimal() +
```

```r
  scale_fill_manual(values = c("#0077b6", "#ff5a5f")) +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
```

14. Lead Time Distribution by Hotel Type

Insight: This boxplot visualizes the distribution of lead times (the time between booking and check-in) by hotel type. Significant differences in lead time distributions can reveal booking behaviors specific to each hotel type, potentially guiding marketing strategies or operational adjustments.

Chart 14 Code

```r
## = Lead Time Distribution by Hotel Type = ##
ggplot(df, aes(x = hotel, y = lead_time, fill = hotel)) +
  geom_boxplot() +
  labs(title = "Lead Time Distribution by Hotel Type", x = "Hotel Type", y =
"Lead Time (Days)") +
  theme_minimal()+
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
```

15. Cancellation Rate by Deposit Type

Insight: This chart shows the proportion of cancellations relative to deposit types. If a higher cancellation rate is associated with a particular deposit type, it may suggest that guests are more likely to cancel if they have made a specific type of deposit. This insight can guide deposit policies to minimize cancellations.

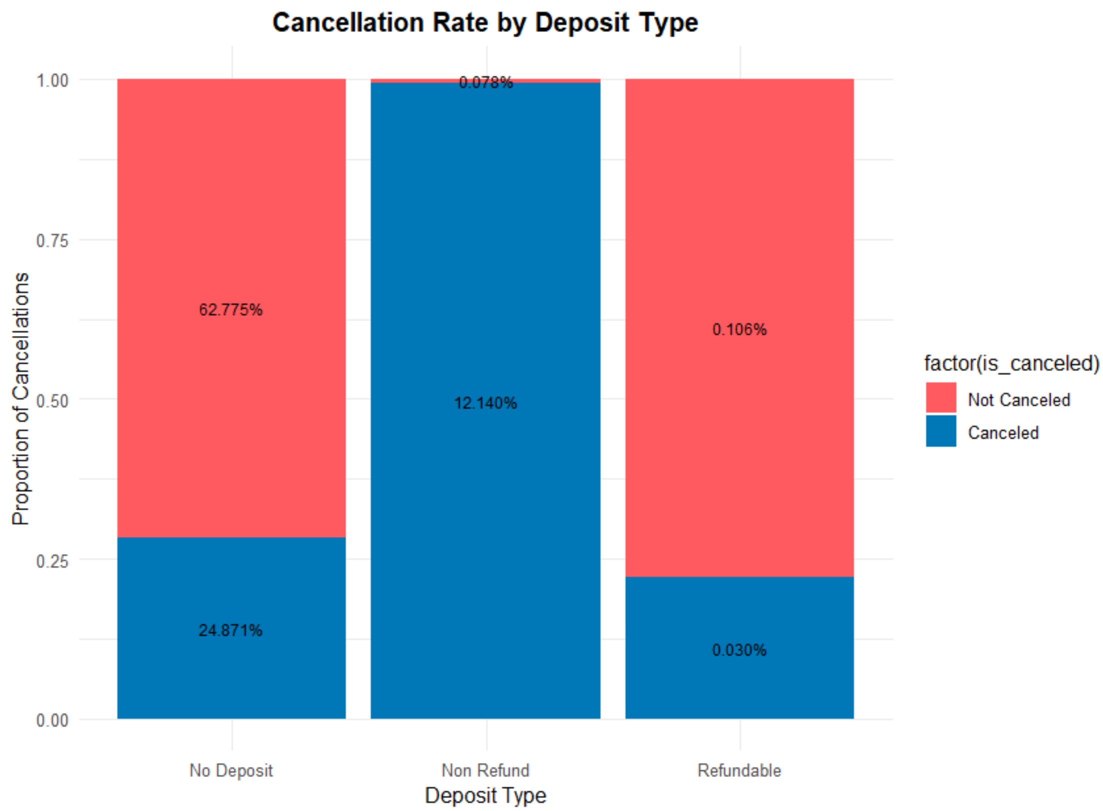

Chart 15 Code

```r
## = Cancellation Rate by Deposit Type = ##
ggplot(df, aes(x = deposit_type, fill = factor(is_canceled))) +
  geom_bar(position = "fill") +
  geom_text(stat = "count", aes(label =
scales::percent(..count../sum(..count..))), position = position_fill(vjust =
0.5), size = 3) +
  labs(title = "Cancellation Rate by Deposit Type", x = "Deposit Type", y =
"Proportion of Cancellations") +
  theme_minimal() +
  scale_fill_manual(values = c("#ff5a5f", "#0077b6"), labels = c("Not
Canceled", "Canceled")) +
  theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
```

16. Cancellation Rate by Market Segment

Insight: The cancellation rate varies significantly by market segment. Segments such as online travel agencies (OTAs) might have a higher cancellation rate compared to other segments like corporate bookings. This indicates that OTA bookings may be more volatile, possibly due to more flexible cancellation policies offered by these platforms.

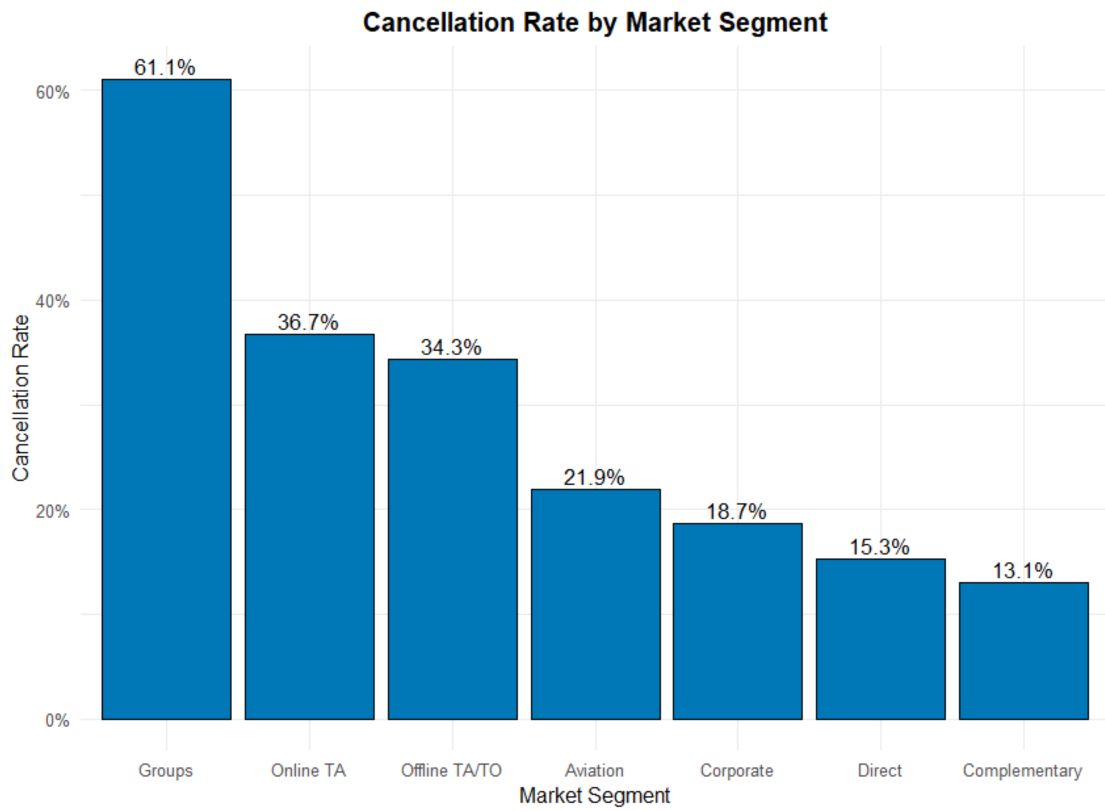

Chart 16 Code

```
## =cancellation rate by market segment = ##
cancellation_rate <- df %>%
  group_by(market_segment) %>%
  summarize(cancellation_rate = mean(is_canceled))

ggplot(cancellation_rate, aes(x = reorder(market_segment, -
cancellation_rate), y = cancellation_rate)) +
  geom_bar(stat = "identity", fill = "#0077b6", color = "black") +
  geom_text(aes(label = scales::percent(cancellation_rate)), vjust = -0.3,
size = 4) +
  labs(title = "Cancellation Rate by Market Segment", x = "Market Segment", y
= "Cancellation Rate") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 14)) +
  scale_y_continuous(labels = scales::percent)
```

Insights and Strategic Suggestions

Booking Patterns

- **Insight**: The monthly booking trends indicate seasonal fluctuations in bookings.
- **Recommendation**: Implement promotional offers or discounts during months with lower booking volumes to increase occupancy rates. Additionally, introduce special packages for off-peak seasons to maintain consistent booking levels.

Cancellation Mitigation

- **Insight**: Certain market segments have higher cancellation rates, particularly as shown in the Cancellation Rate by Market Segment chart.

- **Recommendation**: Reevaluate the cancellation policies for market segments with higher cancellation rates. Consider requiring deposits or offering incentives for confirmed bookings to reduce last-minute cancellations. Enhanced customer engagement could also help lower the cancellation rate.

Customer Targeting

- **Insight**: The Average Daily Rate by Customer Type chart highlights which customer types generate the highest ADR.

- **Recommendation**: Design targeted marketing campaigns to attract more customers from high-ADR segments. Consider offering loyalty programs or tailored experiences for repeat customers from these segments to further boost revenue.

Reservation Process Optimization

- **Insight**: The Reservation Status by Month and Reservation Status by Hotel Type charts reveal variations in booking status, including cancellations and check-ins.

- **Recommendation**: Streamline the reservation process by integrating automated systems that reduce room assignment errors. This could also help address the gap between booked and actual reservations, improving customer satisfaction.

Parking Space Management

- **Insight**: Analysis of special requests, such as for parking spaces, suggests that certain requests are more common in specific periods.

- **Recommendation**: Better allocate parking resources based on high-demand periods. This could involve reserving parking spots in advance or adjusting parking fees during high-occupancy months to optimize space utilization.

Conclusion

In this report, we explored various aspects of hotel booking data using several visualizations. Our analysis revealed critical insights into customer booking behaviors, cancellation trends, and pricing strategies:

- **Booking Trends:** We observed consistent fluctuations in bookings across different months, with certain months showing peak activity. The yearly trends indicated a growth in bookings, particularly in the high season.

- **Customer Preferences:** Analysis of the average daily rate (ADR) by customer type revealed that certain customer segments tend to pay higher prices, possibly indicating a willingness to opt for premium services.

- **Geographical Insights:** The ADR distribution across the top ten countries showed varying revenue contributions, with some countries significantly outperforming others in terms of ADR.
- **Reservation Status:** The data highlighted notable differences in booking cancellations, particularly across different market segments, with some segments having a higher cancellation rate. Additionally, the reservation status per month revealed seasonal patterns affecting customer decisions.
- **Hotel Preferences:** Different hotels exhibited varying performance in terms of reservation status, with certain hotel types attracting more confirmed bookings compared to others.

These findings provide actionable insights that can aid in optimizing pricing strategies, enhancing customer experience, and reducing cancellation rates.

Acknowledgements

I would like to extend my heartfelt thanks to **Dr. Soha Nagy**. for their invaluable support and guidance throughout this project. Your expertise and encouragement have been instrumental in achieving our goals, and I am deeply grateful for your mentorship.

Thank you for your unwavering support and dedication.

Thank You 💙

Made By: **Ahmed Nour**