

# CAPSTONE: Predicting Customer Reservation Cancellation

Nikhil Eatalpacka

2024-04-29

## Analyzing the Hotel Reservations Dataset: Predicting Customer Reservation Cancellation

### Introduction

In this analysis, we aim to predict customer reservation cancellations using the hotel reservations dataset. We will explore various features such as booking lead time, customer demographics, and booking trends to build predictive models.

### R Markdown

```
# Running libraries
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
library(ggplot2)
```

```
# Loading the dataset to data set by using read_csv, with the first 10 rows visible.
```

```
data <- read_csv("hotel_booking.csv")
```

```
head(data, 10)
```

```
##           hotel is_canceled lead_time arrival_date_year arrival_date_month
## 1 Resort Hotel           0        342           2020           July
## 2 Resort Hotel           0        737           2020           July
## 3 Resort Hotel           0          7           2020           July
## 4 Resort Hotel           0         13           2020           July
## 5 Resort Hotel           0         14           2020           July
## 6 Resort Hotel           0         14           2020           July
## 7 Resort Hotel           0          0           2020           July
## 8 Resort Hotel           0          9           2020           July
## 9 Resort Hotel           1         85           2020           July
## 10 Resort Hotel          1         75           2020           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				
## 7	27	1	0				
## 8	27	1	0				
## 9	27	1	0				
## 10	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
## 7	2	2	0	0	BB	PRT	Direct
## 8	2	2	0	0	FB	PRT	Direct
## 9	3	2	0	0	BB	PRT	Online TA
## 10	3	2	0	0	HB	PRT	Offline TA/TO
##	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
## 7	Direct		0				0
## 8	Direct		0				0
## 9	TA/TO		0				0
## 10	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
## 7		0	C				C
## 8		0	C				C
## 9		0	A				A
## 10		0	D				D
##	booking_changes	deposit_type	agent	company	days_in_waiting_list		
## 1	3	No Deposit	NA	NA			0
## 2	4	No Deposit	NA	NA			0
## 3	0	No Deposit	NA	NA			0
## 4	0	No Deposit	304	NA			0
## 5	0	No Deposit	240	NA			0
## 6	0	No Deposit	240	NA			0
## 7	0	No Deposit	NA	NA			0
## 8	0	No Deposit	303	NA			0
## 9	0	No Deposit	240	NA			0
## 10	0	No Deposit	15	NA			0
##	customer_type	adr	required_car_parking_spaces	total_of_special_requests			

```

## 1      Transient    0.0                0                0
## 2      Transient    0.0                0                0
## 3      Transient   75.0                0                0
## 4      Transient   75.0                0                0
## 5      Transient   98.0                0                1
## 6      Transient   98.0                0                1
## 7      Transient  107.0                0                0
## 8      Transient  103.0                0                1
## 9      Transient   82.0                0                1
## 10     Transient  105.5                0                0
##      reservation_status reservation_status_date      name
## 1      Check-Out      7/1/2020      Ernest Barnes
## 2      Check-Out      7/1/2020      Andrea Baker
## 3      Check-Out      7/2/2020      Rebecca Parker
## 4      Check-Out      7/2/2020      Laura Murray
## 5      Check-Out      7/3/2020      Linda Hines
## 6      Check-Out      7/3/2020      Jasmine Fletcher
## 7      Check-Out      7/3/2020      Dylan Rangel
## 8      Check-Out      7/3/2020      William Velez
## 9      Canceled      5/6/2020      Steven Murphy
## 10     Canceled      4/22/2020      Michael Moore
##      email phone.number      credit_card
## 1 Ernest.Barnes31@outlook.com 669-792-1661 *****4322
## 2   Andrea_Baker94@aol.com 858-637-6955 *****9157
## 3 Rebecca_Parker@comcast.net 652-885-2745 *****3734
## 4   Laura_M@gmail.com 364-656-8427 *****5677
## 5   LHines@verizon.com 713-226-5883 *****5498
## 6   JFletcher43@xfinity.com 190-271-6743 *****9263
## 7   Rangel.Dylan@comcast.net 420-332-5209 *****6994
## 8   Velez_William@mail.com 286-669-4333 *****8729
## 9   Steven.Murphy54@aol.com 341-726-5787 *****3639
## 10 MichaelMoore81@outlook.com 316-648-6176 *****9190

# Total no of observations
dim(data)

## [1] 119390      36

# Select specific columns from the dataframe 'data' and store them in 'selected_columns'.
selected_columns <- data[, c(1,4,5,7,9,12,17:23,26)]
head(selected_columns, 5)

##      hotel arrival_date_year arrival_date_month arrival_date_day_of_month
## 1 Resort Hotel      2020      July      1
## 2 Resort Hotel      2020      July      1
## 3 Resort Hotel      2020      July      1
## 4 Resort Hotel      2020      July      1
## 5 Resort Hotel      2020      July      1
##      stays_in_week_nights adults children babies is_repeated_guest
## 1      0      2      0      0      0
## 2      0      2      0      0      0
## 3      1      1      0      0      0
## 4      1      1      0      0      0
## 5      2      2      0      0      0
##      previous_cancellations previous_bookings_not_canceled reserved_room_type
## 1      0      0      C

```

```
## 2          0          0          C
## 3          0          0          A
## 4          0          0          A
## 5          0          0          A
##   assigned_room_type booking_changes deposit_type days_in_waiting_list
## 1          C          3   No Deposit          0
## 2          C          4   No Deposit          0
## 3          C          0   No Deposit          0
## 4          A          0   No Deposit          0
## 5          A          0   No Deposit          0

# Assign the dataframe 'selected_columns' to a new variable 'data1'.
data1 <- selected_columns

# Check if there are any missing values in 'data1' and store the result in 'any_null'.
any_null <- any(is.na(data1))

# Print the result of the check for missing values.
print(any_null)

## [1] TRUE

# Remove rows containing missing values from 'data1' and store the cleaned data in 'clean_data'.
clean_data <- drop_na(data1)

# Assign the cleaned dataframe 'clean_data' to a new variable 'data2'.
data2 <- clean_data

library(lubridate)

##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

# Load the 'lubridate' package for date manipulation.

# Convert month names to month numbers and store them in a new column 'Month_Num' in 'data2'.
data2$Month_Num <- match(tolower(data2$arrival_date_month), tolower(month.name))

# Print the updated dataframe 'data2'.
print(data2)

##           hotel arrival_date_year arrival_date_month
## 1   Resort Hotel             2020              July
## 2   Resort Hotel             2020              July
## 3   Resort Hotel             2020              July
## 4   Resort Hotel             2020              July
## 5   Resort Hotel             2020              July
## 6   Resort Hotel             2020              July
## 7   Resort Hotel             2020              July
## 8   Resort Hotel             2020              July
## 9   Resort Hotel             2020              July
## 10  Resort Hotel             2020              July
## 11  Resort Hotel             2020              July
```

## 12	Resort Hotel	2020	July
## 13	Resort Hotel	2020	July
## 14	Resort Hotel	2020	July
## 15	Resort Hotel	2020	July
## 16	Resort Hotel	2020	July
## 17	Resort Hotel	2020	July
## 18	Resort Hotel	2020	July
## 19	Resort Hotel	2020	July
## 20	Resort Hotel	2020	July
## 21	Resort Hotel	2020	July
## 22	Resort Hotel	2020	July
## 23	Resort Hotel	2020	July
## 24	Resort Hotel	2020	July
## 25	Resort Hotel	2020	July
## 26	Resort Hotel	2020	July
## 27	Resort Hotel	2020	July
## 28	Resort Hotel	2020	July
## 29	Resort Hotel	2020	July
## 30	Resort Hotel	2020	July
## 31	Resort Hotel	2020	July
## 32	Resort Hotel	2020	July
## 33	Resort Hotel	2020	July
## 34	Resort Hotel	2020	July
## 35	Resort Hotel	2020	July
## 36	Resort Hotel	2020	July
## 37	Resort Hotel	2020	July
## 38	Resort Hotel	2020	July
## 39	Resort Hotel	2020	July
## 40	Resort Hotel	2020	July
## 41	Resort Hotel	2020	July
## 42	Resort Hotel	2020	July
## 43	Resort Hotel	2020	July
## 44	Resort Hotel	2020	July
## 45	Resort Hotel	2020	July
## 46	Resort Hotel	2020	July
## 47	Resort Hotel	2020	July
## 48	Resort Hotel	2020	July
## 49	Resort Hotel	2020	July
## 50	Resort Hotel	2020	July
## 51	Resort Hotel	2020	July
## 52	Resort Hotel	2020	July
## 53	Resort Hotel	2020	July
## 54	Resort Hotel	2020	July
## 55	Resort Hotel	2020	July
## 56	Resort Hotel	2020	July
## 57	Resort Hotel	2020	July
## 58	Resort Hotel	2020	July
## 59	Resort Hotel	2020	July
## 60	Resort Hotel	2020	July
## 61	Resort Hotel	2020	July
## 62	Resort Hotel	2020	July
## 63	Resort Hotel	2020	July
## 64	Resort Hotel	2020	July
## 65	Resort Hotel	2020	July

## 66	Resort Hotel	2020	July
## 67	Resort Hotel	2020	July
## 68	Resort Hotel	2020	July
## 69	Resort Hotel	2020	July
## 70	Resort Hotel	2020	July
## 71	Resort Hotel	2020	July
## 72	Resort Hotel	2020	July
## 73	Resort Hotel	2020	July
## 74	Resort Hotel	2020	July
## 75	Resort Hotel	2020	July
## 76	Resort Hotel	2020	July
## 77	Resort Hotel	2020	July
## 78	Resort Hotel	2020	July
## 79	Resort Hotel	2020	July
## 80	Resort Hotel	2020	July
## 81	Resort Hotel	2020	July
## 82	Resort Hotel	2020	July
## 83	Resort Hotel	2020	July
## 84	Resort Hotel	2020	July
## 85	Resort Hotel	2020	July
## 86	Resort Hotel	2020	July
## 87	Resort Hotel	2020	July
## 88	Resort Hotel	2020	July
## 89	Resort Hotel	2020	July
## 90	Resort Hotel	2020	July
## 91	Resort Hotel	2020	July
## 92	Resort Hotel	2020	July
## 93	Resort Hotel	2020	July
## 94	Resort Hotel	2020	July
## 95	Resort Hotel	2020	July
## 96	Resort Hotel	2020	July
## 97	Resort Hotel	2020	July
## 98	Resort Hotel	2020	July
## 99	Resort Hotel	2020	July
## 100	Resort Hotel	2020	July
## 101	Resort Hotel	2020	July
## 102	Resort Hotel	2020	July
## 103	Resort Hotel	2020	July
## 104	Resort Hotel	2020	July
## 105	Resort Hotel	2020	July
## 106	Resort Hotel	2020	July
## 107	Resort Hotel	2020	July
## 108	Resort Hotel	2020	July
## 109	Resort Hotel	2020	July
## 110	Resort Hotel	2020	July
## 111	Resort Hotel	2020	July
## 112	Resort Hotel	2020	July
## 113	Resort Hotel	2020	July
## 114	Resort Hotel	2020	July
## 115	Resort Hotel	2020	July
## 116	Resort Hotel	2020	July
## 117	Resort Hotel	2020	July
## 118	Resort Hotel	2020	July
## 119	Resort Hotel	2020	July

```
## 5259          0          4  2021-4-23 2021-04-23
## 5260          0          4  2021-4-23 2021-04-23
## 5261          0          4  2021-4-23 2021-04-23
## 5262          0          4  2021-4-23 2021-04-23
## 5263          0          4  2021-4-23 2021-04-23
## [ reached 'max' / getOption("max.print") -- omitted 114123 rows ]

# Count the number of bookings per date and store the result in 'daily_booking_counts'.
daily_booking_counts <- table(data2$Date)

# Convert the result to a dataframe and assign column names.
daily_booking_counts_df <- as.data.frame(daily_booking_counts)
names(daily_booking_counts_df) <- c("Date", "Booking_Count")

# Print the dataframe 'daily_booking_counts_df'.
print(daily_booking_counts_df)
```

```
##      Date Booking_Count
## 1  2020-07-01         122
## 2  2020-07-02          93
## 3  2020-07-03          56
## 4  2020-07-04          88
## 5  2020-07-05          53
## 6  2020-07-06          75
## 7  2020-07-07          54
## 8  2020-07-08          69
## 9  2020-07-09          80
## 10 2020-07-10          51
## 11 2020-07-11         103
## 12 2020-07-12          63
## 13 2020-07-13          84
## 14 2020-07-14          35
## 15 2020-07-15         102
## 16 2020-07-16         103
## 17 2020-07-17         151
## 18 2020-07-18         155
## 19 2020-07-19          54
## 20 2020-07-20         125
## 21 2020-07-21          71
## 22 2020-07-22         127
## 23 2020-07-23         125
## 24 2020-07-24          64
## 25 2020-07-25         204
## 26 2020-07-26          47
## 27 2020-07-27         170
## 28 2020-07-28          34
## 29 2020-07-29          50
## 30 2020-07-30          99
## 31 2020-07-31          69
## 32 2020-08-01         110
## 33 2020-08-02          50
## 34 2020-08-03         182
## 35 2020-08-04          51
## 36 2020-08-05         152
## 37 2020-08-06          75
```

## 38	2020-08-07	133
## 39	2020-08-08	206
## 40	2020-08-09	85
## 41	2020-08-10	207
## 42	2020-08-11	117
## 43	2020-08-12	133
## 44	2020-08-13	106
## 45	2020-08-14	329
## 46	2020-08-15	190
## 47	2020-08-16	98
## 48	2020-08-17	188
## 49	2020-08-18	94
## 50	2020-08-19	66
## 51	2020-08-20	110
## 52	2020-08-21	75
## 53	2020-08-22	139
## 54	2020-08-23	79
## 55	2020-08-24	153
## 56	2020-08-25	126
## 57	2020-08-26	122
## 58	2020-08-27	100
## 59	2020-08-28	84
## 60	2020-08-29	109
## 61	2020-08-30	101
## 62	2020-08-31	115
## 63	2020-09-01	100
## 64	2020-09-02	128
## 65	2020-09-03	155
## 66	2020-09-04	192
## 67	2020-09-05	240
## 68	2020-09-06	152
## 69	2020-09-07	133
## 70	2020-09-08	104
## 71	2020-09-09	201
## 72	2020-09-10	139
## 73	2020-09-11	212
## 74	2020-09-12	191
## 75	2020-09-13	65
## 76	2020-09-14	149
## 77	2020-09-15	121
## 78	2020-09-16	170
## 79	2020-09-17	331
## 80	2020-09-18	340
## 81	2020-09-19	195
## 82	2020-09-20	129
## 83	2020-09-21	226
## 84	2020-09-22	101
## 85	2020-09-23	159
## 86	2020-09-24	216
## 87	2020-09-25	275
## 88	2020-09-26	198
## 89	2020-09-27	69
## 90	2020-09-28	107
## 91	2020-09-29	111



## 92	2020-09-30	205
## 93	2020-10-01	121
## 94	2020-10-02	204
## 95	2020-10-03	211
## 96	2020-10-04	185
## 97	2020-10-05	223
## 98	2020-10-06	179
## 99	2020-10-07	190
## 100	2020-10-08	164
## 101	2020-10-09	210
## 102	2020-10-10	250
## 103	2020-10-11	102
## 104	2020-10-12	208
## 105	2020-10-13	98
## 106	2020-10-14	73
## 107	2020-10-15	255
## 108	2020-10-16	356
## 109	2020-10-17	130
## 110	2020-10-18	116
## 111	2020-10-19	229
## 112	2020-10-20	109
## 113	2020-10-21	131
## 114	2020-10-22	88
## 115	2020-10-23	99
## 116	2020-10-24	160
## 117	2020-10-25	148
## 118	2020-10-26	97
## 119	2020-10-27	53
## 120	2020-10-28	141
## 121	2020-10-29	130
## 122	2020-10-30	106
## 123	2020-10-31	191
## 124	2020-11-01	57
## 125	2020-11-02	77
## 126	2020-11-03	70
## 127	2020-11-04	69
## 128	2020-11-05	60
## 129	2020-11-06	101
## 130	2020-11-07	111
## 131	2020-11-08	54
## 132	2020-11-09	82
## 133	2020-11-10	50
## 134	2020-11-11	72
## 135	2020-11-12	65
## 136	2020-11-13	59
## 137	2020-11-14	39
## 138	2020-11-15	28
## 139	2020-11-16	74
## 140	2020-11-17	81
## 141	2020-11-18	173
## 142	2020-11-19	55
## 143	2020-11-20	187
## 144	2020-11-21	86
## 145	2020-11-22	38

## 146	2020-11-23	211
## 147	2020-11-24	69
## 148	2020-11-25	47
## 149	2020-11-26	56
## 150	2020-11-27	132
## 151	2020-11-28	61
## 152	2020-11-29	20
## 153	2020-11-30	56
## 154	2020-12-01	57
## 155	2020-12-02	58
## 156	2020-12-03	83
## 157	2020-12-04	123
## 158	2020-12-05	448
## 159	2020-12-06	63
## 160	2020-12-07	27
## 161	2020-12-08	191
## 162	2020-12-09	152
## 163	2020-12-10	81
## 164	2020-12-11	50
## 165	2020-12-12	43
## 166	2020-12-13	19
## 167	2020-12-14	39
## 168	2020-12-15	29
## 169	2020-12-16	43
## 170	2020-12-17	49
## 171	2020-12-18	47
## 172	2020-12-19	89
## 173	2020-12-20	38
## 174	2020-12-21	47
## 175	2020-12-22	72
## 176	2020-12-23	63
## 177	2020-12-24	90
## 178	2020-12-25	68
## 179	2020-12-26	112
## 180	2020-12-27	123
## 181	2020-12-28	100
## 182	2020-12-29	155
## 183	2020-12-30	238
## 184	2020-12-31	123
## 185	2021-01-01	44
## 186	2021-01-02	183
## 187	2021-01-03	62
## 188	2021-01-04	77
## 189	2021-01-05	47
## 190	2021-01-06	49
## 191	2021-01-07	49
## 192	2021-01-08	59
## 193	2021-01-09	54
## 194	2021-01-10	46
## 195	2021-01-11	40
## 196	2021-01-12	69
## 197	2021-01-13	48
## 198	2021-01-14	96
## 199	2021-01-15	83

##	200	2021-01-16	67
##	201	2021-01-17	37
##	202	2021-01-18	60
##	203	2021-01-19	147
##	204	2021-01-20	64
##	205	2021-01-21	61
##	206	2021-01-22	134
##	207	2021-01-23	70
##	208	2021-01-24	36
##	209	2021-01-25	79
##	210	2021-01-26	69
##	211	2021-01-27	155
##	212	2021-01-28	86
##	213	2021-01-29	86
##	214	2021-01-30	58
##	215	2021-01-31	33
##	216	2021-02-01	62
##	217	2021-02-02	67
##	218	2021-02-03	84
##	219	2021-02-04	105
##	220	2021-02-05	126
##	221	2021-02-06	219
##	222	2021-02-07	129
##	223	2021-02-08	86
##	224	2021-02-09	75
##	225	2021-02-10	79
##	226	2021-02-11	104
##	227	2021-02-12	293
##	228	2021-02-13	183
##	229	2021-02-14	104
##	230	2021-02-15	88
##	231	2021-02-16	78
##	232	2021-02-17	286
##	233	2021-02-18	83
##	234	2021-02-19	247
##	235	2021-02-20	107
##	236	2021-02-21	106
##	237	2021-02-22	108
##	238	2021-02-23	68
##	239	2021-02-24	104
##	240	2021-02-25	154
##	241	2021-02-26	245
##	242	2021-02-27	211
##	243	2021-02-28	194
##	244	2021-03-01	123
##	245	2021-03-02	160
##	246	2021-03-03	109
##	247	2021-03-04	214
##	248	2021-03-05	131
##	249	2021-03-06	67
##	250	2021-03-07	149
##	251	2021-03-08	118
##	252	2021-03-09	143
##	253	2021-03-10	108

##	254	2021-03-11	183
##	255	2021-03-12	131
##	256	2021-03-13	103
##	257	2021-03-14	117
##	258	2021-03-15	112
##	259	2021-03-16	118
##	260	2021-03-17	188
##	261	2021-03-18	199
##	262	2021-03-19	213
##	263	2021-03-20	244
##	264	2021-03-21	183
##	265	2021-03-22	157
##	266	2021-03-23	128
##	267	2021-03-24	295
##	268	2021-03-25	227
##	269	2021-03-26	107
##	270	2021-03-27	105
##	271	2021-03-28	169
##	272	2021-03-29	119
##	273	2021-03-30	166
##	274	2021-03-31	238
##	275	2021-04-01	104
##	276	2021-04-02	102
##	277	2021-04-03	100
##	278	2021-04-04	211
##	279	2021-04-05	115
##	280	2021-04-06	237
##	281	2021-04-07	196
##	282	2021-04-08	182
##	283	2021-04-09	229
##	284	2021-04-10	126
##	285	2021-04-11	188
##	286	2021-04-12	158
##	287	2021-04-13	177
##	288	2021-04-14	203
##	289	2021-04-15	160
##	290	2021-04-16	164
##	291	2021-04-17	157
##	292	2021-04-18	207
##	293	2021-04-19	171
##	294	2021-04-20	224
##	295	2021-04-21	146
##	296	2021-04-22	263
##	297	2021-04-23	171
##	298	2021-04-24	169
##	299	2021-04-25	149
##	300	2021-04-26	174
##	301	2021-04-27	179
##	302	2021-04-28	292
##	303	2021-04-29	266
##	304	2021-04-30	208
##	305	2021-05-01	91
##	306	2021-05-02	263
##	307	2021-05-03	93

## 308	2021-05-04	134
## 309	2021-05-05	271
## 310	2021-05-06	128
## 311	2021-05-07	149
## 312	2021-05-08	148
## 313	2021-05-09	168
## 314	2021-05-10	153
## 315	2021-05-11	132
## 316	2021-05-12	271
## 317	2021-05-13	237
## 318	2021-05-14	154
## 319	2021-05-15	187
## 320	2021-05-16	219
## 321	2021-05-17	228
## 322	2021-05-18	136
## 323	2021-05-19	264
## 324	2021-05-20	227
## 325	2021-05-21	191
## 326	2021-05-22	129
## 327	2021-05-23	134
## 328	2021-05-24	221
## 329	2021-05-25	109
## 330	2021-05-26	276
## 331	2021-05-27	96
## 332	2021-05-28	95
## 333	2021-05-29	185
## 334	2021-05-30	189
## 335	2021-05-31	200
## 336	2021-06-01	133
## 337	2021-06-02	266
## 338	2021-06-03	193
## 339	2021-06-04	151
## 340	2021-06-05	107
## 341	2021-06-06	230
## 342	2021-06-07	146
## 343	2021-06-08	200
## 344	2021-06-09	223
## 345	2021-06-10	196
## 346	2021-06-11	101
## 347	2021-06-12	224
## 348	2021-06-13	130
## 349	2021-06-14	229
## 350	2021-06-15	284
## 351	2021-06-16	149
## 352	2021-06-17	321
## 353	2021-06-18	131
## 354	2021-06-19	122
## 355	2021-06-20	205
## 356	2021-06-21	133
## 357	2021-06-22	116
## 358	2021-06-23	131
## 359	2021-06-24	273
## 360	2021-06-25	79
## 361	2021-06-26	226

## 362	2021-06-27	184
## 363	2021-06-28	133
## 364	2021-06-29	93
## 365	2021-06-30	183
## 366	2021-07-01	173
## 367	2021-07-02	134
## 368	2021-07-03	118
## 369	2021-07-04	148
## 370	2021-07-05	109
## 371	2021-07-06	183
## 372	2021-07-07	170
## 373	2021-07-08	145
## 374	2021-07-09	141
## 375	2021-07-10	106
## 376	2021-07-11	173
## 377	2021-07-12	127
## 378	2021-07-13	118
## 379	2021-07-14	160
## 380	2021-07-15	141
## 381	2021-07-16	176
## 382	2021-07-17	126
## 383	2021-07-18	184
## 384	2021-07-19	106
## 385	2021-07-20	150
## 386	2021-07-21	216
## 387	2021-07-22	127
## 388	2021-07-23	197
## 389	2021-07-24	130
## 390	2021-07-25	173
## 391	2021-07-26	115
## 392	2021-07-27	143
## 393	2021-07-28	153
## 394	2021-07-29	107
## 395	2021-07-30	209
## 396	2021-07-31	114
## 397	2021-08-01	214
## 398	2021-08-02	137
## 399	2021-08-03	144
## 400	2021-08-04	175
## 401	2021-08-05	157
## 402	2021-08-06	158
## 403	2021-08-07	137
## 404	2021-08-08	210
## 405	2021-08-09	122
## 406	2021-08-10	139
## 407	2021-08-11	192
## 408	2021-08-12	172
## 409	2021-08-13	213
## 410	2021-08-14	138
## 411	2021-08-15	177
## 412	2021-08-16	159
## 413	2021-08-17	146
## 414	2021-08-18	232
## 415	2021-08-19	165

##	416	2021-08-20	198
##	417	2021-08-21	154
##	418	2021-08-22	187
##	419	2021-08-23	116
##	420	2021-08-24	150
##	421	2021-08-25	169
##	422	2021-08-26	170
##	423	2021-08-27	141
##	424	2021-08-28	167
##	425	2021-08-29	164
##	426	2021-08-30	132
##	427	2021-08-31	128
##	428	2021-09-01	193
##	429	2021-09-02	174
##	430	2021-09-03	178
##	431	2021-09-04	132
##	432	2021-09-05	185
##	433	2021-09-06	99
##	434	2021-09-07	133
##	435	2021-09-08	212
##	436	2021-09-09	221
##	437	2021-09-10	115
##	438	2021-09-11	172
##	439	2021-09-12	173
##	440	2021-09-13	166
##	441	2021-09-14	128
##	442	2021-09-15	313
##	443	2021-09-16	217
##	444	2021-09-17	121
##	445	2021-09-18	174
##	446	2021-09-19	230
##	447	2021-09-20	209
##	448	2021-09-21	208
##	449	2021-09-22	194
##	450	2021-09-23	140
##	451	2021-09-24	122
##	452	2021-09-25	155
##	453	2021-09-26	304
##	454	2021-09-27	122
##	455	2021-09-28	136
##	456	2021-09-29	182
##	457	2021-09-30	286
##	458	2021-10-01	184
##	459	2021-10-02	150
##	460	2021-10-03	199
##	461	2021-10-04	156
##	462	2021-10-05	189
##	463	2021-10-06	289
##	464	2021-10-07	186
##	465	2021-10-08	129
##	466	2021-10-09	214
##	467	2021-10-10	213
##	468	2021-10-11	182
##	469	2021-10-12	256

## 470	2021-10-13	344
## 471	2021-10-14	195
## 472	2021-10-15	130
## 473	2021-10-16	281
## 474	2021-10-17	209
## 475	2021-10-18	136
## 476	2021-10-19	125
## 477	2021-10-20	247
## 478	2021-10-21	247
## 479	2021-10-22	175
## 480	2021-10-23	192
## 481	2021-10-24	136
## 482	2021-10-25	155
## 483	2021-10-26	156
## 484	2021-10-27	194
## 485	2021-10-28	334
## 486	2021-10-29	239
## 487	2021-10-30	214
## 488	2021-10-31	147
## 489	2021-11-01	212
## 490	2021-11-02	143
## 491	2021-11-03	229
## 492	2021-11-04	237
## 493	2021-11-05	214
## 494	2021-11-06	122
## 495	2021-11-07	366
## 496	2021-11-08	87
## 497	2021-11-09	76
## 498	2021-11-10	147
## 499	2021-11-11	209
## 500	2021-11-12	146
## 501	2021-11-13	126
## 502	2021-11-14	142
## 503	2021-11-15	95
## 504	2021-11-16	131
## 505	2021-11-17	160
## 506	2021-11-18	207
## 507	2021-11-19	140
## 508	2021-11-20	93
## 509	2021-11-21	183
## 510	2021-11-22	77
## 511	2021-11-23	106
## 512	2021-11-24	120
## 513	2021-11-25	191
## 514	2021-11-26	120
## 515	2021-11-27	81
## 516	2021-11-28	118
## 517	2021-11-29	109
## 518	2021-11-30	67
## 519	2021-12-01	174
## 520	2021-12-02	199
## 521	2021-12-03	189
## 522	2021-12-04	87
## 523	2021-12-05	142



## 524	2021-12-06	95
## 525	2021-12-07	119
## 526	2021-12-08	240
## 527	2021-12-09	116
## 528	2021-12-10	99
## 529	2021-12-11	39
## 530	2021-12-12	78
## 531	2021-12-13	77
## 532	2021-12-14	77
## 533	2021-12-15	94
## 534	2021-12-16	142
## 535	2021-12-17	111
## 536	2021-12-18	63
## 537	2021-12-19	99
## 538	2021-12-20	78
## 539	2021-12-21	89
## 540	2021-12-22	125
## 541	2021-12-23	127
## 542	2021-12-24	124
## 543	2021-12-25	72
## 544	2021-12-26	164
## 545	2021-12-27	215
## 546	2021-12-28	149
## 547	2021-12-29	163
## 548	2021-12-30	238
## 549	2021-12-31	76
## 550	2022-01-01	128
## 551	2022-01-02	179
## 552	2022-01-03	124
## 553	2022-01-04	96
## 554	2022-01-05	98
## 555	2022-01-06	112
## 556	2022-01-07	83
## 557	2022-01-08	110
## 558	2022-01-09	77
## 559	2022-01-10	99
## 560	2022-01-11	61
## 561	2022-01-12	138
## 562	2022-01-13	92
## 563	2022-01-14	244
## 564	2022-01-15	67
## 565	2022-01-16	193
## 566	2022-01-17	69
## 567	2022-01-18	219
## 568	2022-01-19	123
## 569	2022-01-20	84
## 570	2022-01-21	115
## 571	2022-01-22	89
## 572	2022-01-23	172
## 573	2022-01-24	87
## 574	2022-01-25	191
## 575	2022-01-26	114
## 576	2022-01-27	150
## 577	2022-01-28	111

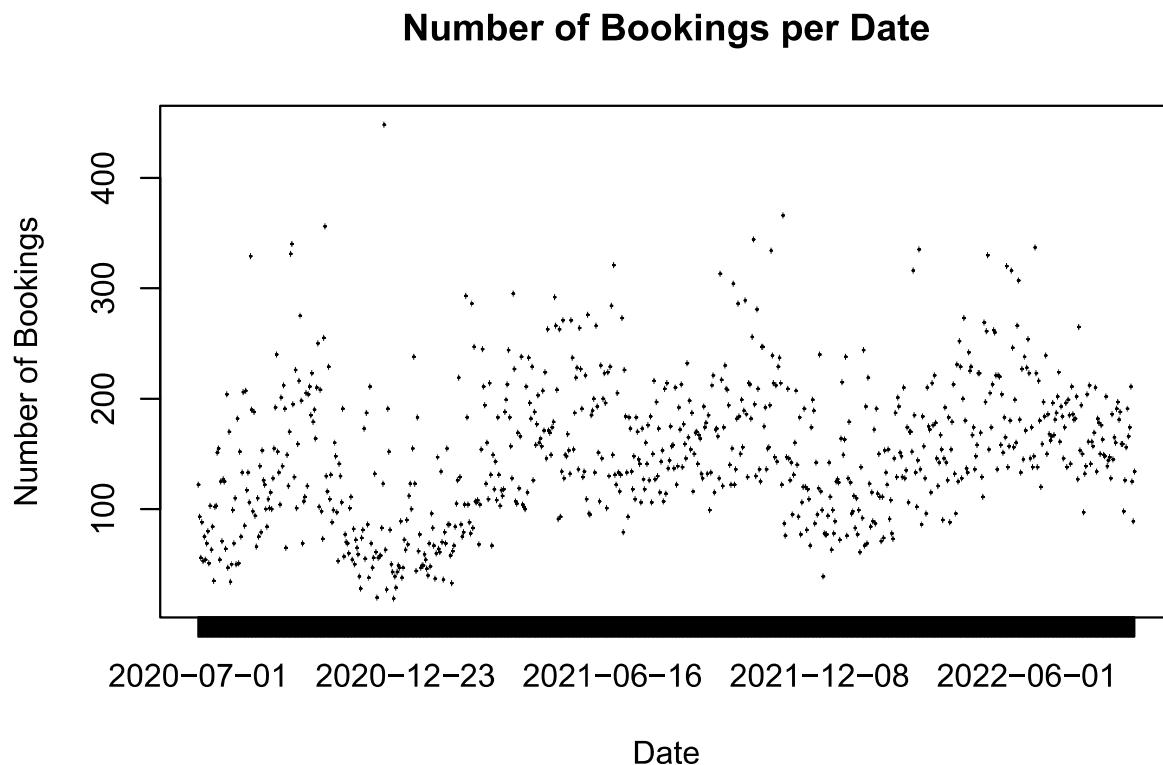
## 578	2022-01-29	71
## 579	2022-01-30	111
## 580	2022-01-31	74
## 581	2022-02-01	153
## 582	2022-02-02	133
## 583	2022-02-03	122
## 584	2022-02-04	154
## 585	2022-02-05	91
## 586	2022-02-06	108
## 587	2022-02-07	78
## 588	2022-02-08	73
## 589	2022-02-09	146
## 590	2022-02-10	187
## 591	2022-02-11	129
## 592	2022-02-12	201
## 593	2022-02-13	193
## 594	2022-02-14	151
## 595	2022-02-15	148
## 596	2022-02-16	129
## 597	2022-02-17	210
## 598	2022-02-18	128
## 599	2022-02-19	135
## 600	2022-02-20	174
## 601	2022-02-21	106
## 602	2022-02-22	170
## 603	2022-02-23	170
## 604	2022-02-24	152
## 605	2022-02-25	316
## 606	2022-02-26	185
## 607	2022-02-27	133
## 608	2022-02-28	102
## 609	2022-03-01	144
## 610	2022-03-02	335
## 611	2022-03-03	136
## 612	2022-03-04	86
## 613	2022-03-05	184
## 614	2022-03-06	157
## 615	2022-03-07	128
## 616	2022-03-08	96
## 617	2022-03-09	210
## 618	2022-03-10	175
## 619	2022-03-11	172
## 620	2022-03-12	115
## 621	2022-03-13	214
## 622	2022-03-14	176
## 623	2022-03-15	221
## 624	2022-03-16	178
## 625	2022-03-17	146
## 626	2022-03-18	121
## 627	2022-03-19	142
## 628	2022-03-20	167
## 629	2022-03-21	154
## 630	2022-03-22	90
## 631	2022-03-23	146

## 632	2022-03-24	192
## 633	2022-03-25	142
## 634	2022-03-26	126
## 635	2022-03-27	183
## 636	2022-03-28	88
## 637	2022-03-29	154
## 638	2022-03-30	179
## 639	2022-03-31	213
## 640	2022-04-01	132
## 641	2022-04-02	96
## 642	2022-04-03	231
## 643	2022-04-04	125
## 644	2022-04-05	252
## 645	2022-04-06	229
## 646	2022-04-07	128
## 647	2022-04-08	200
## 648	2022-04-09	273
## 649	2022-04-10	180
## 650	2022-04-11	137
## 651	2022-04-12	133
## 652	2022-04-13	242
## 653	2022-04-14	225
## 654	2022-04-15	229
## 655	2022-04-16	167
## 656	2022-04-17	174
## 657	2022-04-18	136
## 658	2022-04-19	148
## 659	2022-04-20	158
## 660	2022-04-21	223
## 661	2022-04-22	223
## 662	2022-04-23	169
## 663	2022-04-24	129
## 664	2022-04-25	111
## 665	2022-04-26	269
## 666	2022-04-27	197
## 667	2022-04-28	261
## 668	2022-04-29	330
## 669	2022-04-30	154
## 670	2022-05-01	205
## 671	2022-05-02	174
## 672	2022-05-03	213
## 673	2022-05-04	262
## 674	2022-05-05	260
## 675	2022-05-06	221
## 676	2022-05-07	136
## 677	2022-05-08	221
## 678	2022-05-09	220
## 679	2022-05-10	158
## 680	2022-05-11	204
## 681	2022-05-12	151
## 682	2022-05-13	183
## 683	2022-05-14	165
## 684	2022-05-15	320
## 685	2022-05-16	138

##	686	2022-05-17	157
##	687	2022-05-18	180
##	688	2022-05-19	316
##	689	2022-05-20	246
##	690	2022-05-21	156
##	691	2022-05-22	199
##	692	2022-05-23	160
##	693	2022-05-24	266
##	694	2022-05-25	307
##	695	2022-05-26	181
##	696	2022-05-27	133
##	697	2022-05-28	227
##	698	2022-05-29	145
##	699	2022-05-30	238
##	700	2022-05-31	171
##	701	2022-06-01	228
##	702	2022-06-02	254
##	703	2022-06-03	223
##	704	2022-06-04	146
##	705	2022-06-05	174
##	706	2022-06-06	138
##	707	2022-06-07	148
##	708	2022-06-08	337
##	709	2022-06-09	223
##	710	2022-06-10	138
##	711	2022-06-11	216
##	712	2022-06-12	180
##	713	2022-06-13	120
##	714	2022-06-14	159
##	715	2022-06-15	197
##	716	2022-06-16	184
##	717	2022-06-17	239
##	718	2022-06-18	150
##	719	2022-06-19	186
##	720	2022-06-20	168
##	721	2022-06-21	162
##	722	2022-06-22	166
##	723	2022-06-23	196
##	724	2022-06-24	167
##	725	2022-06-25	181
##	726	2022-06-26	187
##	727	2022-06-27	202
##	728	2022-06-28	224
##	729	2022-06-29	161
##	730	2022-06-30	193
##	731	2022-07-01	184
##	732	2022-07-02	146
##	733	2022-07-03	174
##	734	2022-07-04	196
##	735	2022-07-05	138
##	736	2022-07-06	199
##	737	2022-07-07	185
##	738	2022-07-08	141
##	739	2022-07-09	211

## 740	2022-07-10	186
## 741	2022-07-11	181
## 742	2022-07-12	182
## 743	2022-07-13	202
## 744	2022-07-14	127
## 745	2022-07-15	265
## 746	2022-07-16	153
## 747	2022-07-17	194
## 748	2022-07-18	150
## 749	2022-07-19	97
## 750	2022-07-20	132
## 751	2022-07-21	139
## 752	2022-07-22	204
## 753	2022-07-23	161
## 754	2022-07-24	212
## 755	2022-07-25	141
## 756	2022-07-26	161
## 757	2022-07-27	157
## 758	2022-07-28	169
## 759	2022-07-29	210
## 760	2022-07-30	134
## 761	2022-07-31	182
## 762	2022-08-01	178
## 763	2022-08-02	150
## 764	2022-08-03	176
## 765	2022-08-04	163
## 766	2022-08-05	149
## 767	2022-08-06	148
## 768	2022-08-07	202
## 769	2022-08-08	139
## 770	2022-08-09	144
## 771	2022-08-10	147
## 772	2022-08-11	128
## 773	2022-08-12	186
## 774	2022-08-13	145
## 775	2022-08-14	165
## 776	2022-08-15	160
## 777	2022-08-16	191
## 778	2022-08-17	197
## 779	2022-08-18	181
## 780	2022-08-19	188
## 781	2022-08-20	159
## 782	2022-08-21	159
## 783	2022-08-22	98
## 784	2022-08-23	126
## 785	2022-08-24	156
## 786	2022-08-25	191
## 787	2022-08-26	166
## 788	2022-08-27	174
## 789	2022-08-28	211
## 790	2022-08-29	125
## 791	2022-08-30	89
## 792	2022-08-31	134

```
# Plot the number of bookings per date.
plot(daily_booking_counts_df$Date, daily_booking_counts_df$Booking_Count,
     type = "l", xlab = "Date", ylab = "Number of Bookings",
     main = "Number of Bookings per Date")
```



```
daily_booking_counts_df$Date <- as.Date(daily_booking_counts_df$Date)
# Convert the 'Date' column to Date object.
```

```
# Aggregate data by month to get monthly booking counts.
monthly_booking_counts <- daily_booking_counts_df %>%
  mutate(Month = floor_date(Date, "month")) %>%
  group_by(Month) %>%
  summarise(Booking_Count = sum(Booking_Count))
```

```
# Print the dataframe 'monthly_booking_counts'.
print(monthly_booking_counts)
```

```
## # A tibble: 26 x 2
##   Month      Booking_Count
##   <date>         <int>
## 1 2020-07-01         2776
## 2 2020-08-01         3885
## 3 2020-09-01         5114
## 4 2020-10-01         4957
## 5 2020-11-01         2340
## 6 2020-12-01         2920
## 7 2021-01-01         2248
## 8 2021-02-01         3795
## 9 2021-03-01         4824
## 10 2021-04-01         5428
```

```
## # i 16 more rows
# Fit a linear regression model.
lm_model <- lm(Booking_Count ~ Month, data = monthly_booking_counts)

# Print summary of the linear regression model.
summary(lm_model)

##
## Call:
## lm(formula = Booking_Count ~ Month, data = monthly_booking_counts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1846.38  -796.80   99.02   786.81  1424.06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.262e+04  1.576e+04  -2.704  0.01239 *
## Month        2.508e+00  8.371e-01   2.995  0.00627 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 973.3 on 24 degrees of freedom
## Multiple R-squared:  0.2721, Adjusted R-squared:  0.2418
## F-statistic: 8.972 on 1 and 24 DF,  p-value: 0.006274

# Set smaller margins for plotting.
par(mar = c(5, 4, 4, 2) + 0.1)

# Open a new plotting device with smaller size and save plot as PNG.
png("monthly_bookings_plot.png", width = 800, height = 600)

# Plot the monthly booking counts.
plot(monthly_booking_counts$Month, monthly_booking_counts$Booking_Count,
     type = "l", xlab = "Month", ylab = "Number of Bookings",
     main = "Number of Bookings per Month")

# Add the linear regression line to the plot.
abline(lm_model, col = "red")

# Save and close the plotting device.
dev.off()

## pdf
## 2

# Make predictions using the linear regression model.
predicted <- predict(lm_model, newdata = monthly_booking_counts)
print(predicted)

##      1      2      3      4      5      6      7      8
## 3632.985 3710.720 3788.454 3863.681 3941.415 4016.642 4094.377 4172.111
##      9     10     11     12     13     14     15     16
## 4242.323 4320.057 4395.284 4473.018 4548.245 4625.980 4703.714 4778.941
##     17     18     19     20     21     22     23     24
```

```
## 4856.675 4931.902 5009.637 5087.371 5157.583 5235.317 5310.544 5388.278
##      25      26
## 5463.505 5541.240
```

```
# Calculate residuals.
```

```
residuals <- monthly_booking_counts$Booking_Count - predicted
print(residuals)
```

```
##      1      2      3      4      5      6
## -856.98536 174.28024 1325.54584 1093.31899 -1601.41541 -1096.64225
##      7      8      9     10     11     12
## -1846.37666 -377.11106 581.67722 1107.94282 1082.71597 818.98157
##     13     14     15     16     17     18
##  23.75473 437.02032 690.28592 1424.05907 -402.67533 -1071.90217
##     19     20     21     22     23     24
## -1328.63658 -910.37098 -187.58270 425.68290 1002.45605 258.72165
##     25     26
## -150.50519 -616.23960
```

```
# Square the residuals.
```

```
squared_residuals <- residuals^2
print(squared_residuals)
```

```
##      1      2      3      4      5      6
## 734423.899 30373.602 1757071.765 1195346.423 2564531.316 1202624.231
##      7      8      9     10     11     12
## 3409106.759 142212.752 338348.387 1227537.282 1172273.877 670730.809
##     13     14     15     16     17     18
##  564.287 190986.761 476494.648 2027944.249 162147.421 1148974.266
##     19     20     21     22     23     24
## 1765275.151 828775.321 35187.269 181205.928 1004918.139 66936.892
##     25     26
## 22651.813 379751.242
```

```
# Calculate mean of squared residuals.
```

```
mean_squared_residuals <- mean(squared_residuals)
print(mean_squared_residuals)
```

```
## [1] 874476.7
```

```
# Calculate RMSE.
```

```
rmse <- sqrt(mean_squared_residuals)
print(rmse)
```

```
## [1] 935.1346
```

```
# Print RMSE.
```

```
print(paste("RMSE:", round(rmse, 2)))
```

```
## [1] "RMSE: 935.13"
```

```
# Load necessary libraries
```

```
library(dplyr)
library(ggplot2)
```

```
# Descriptive Analysis
```

```
summary(data2$adults)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```



```
##    0.000    2.000    2.000    1.856    2.000    55.000
```

```
summary(data2$children)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.1039 0.00000 10.0000
```

```
summary(data2$babies)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.000000 0.000000 0.000000 0.007949 0.000000 10.000000
```

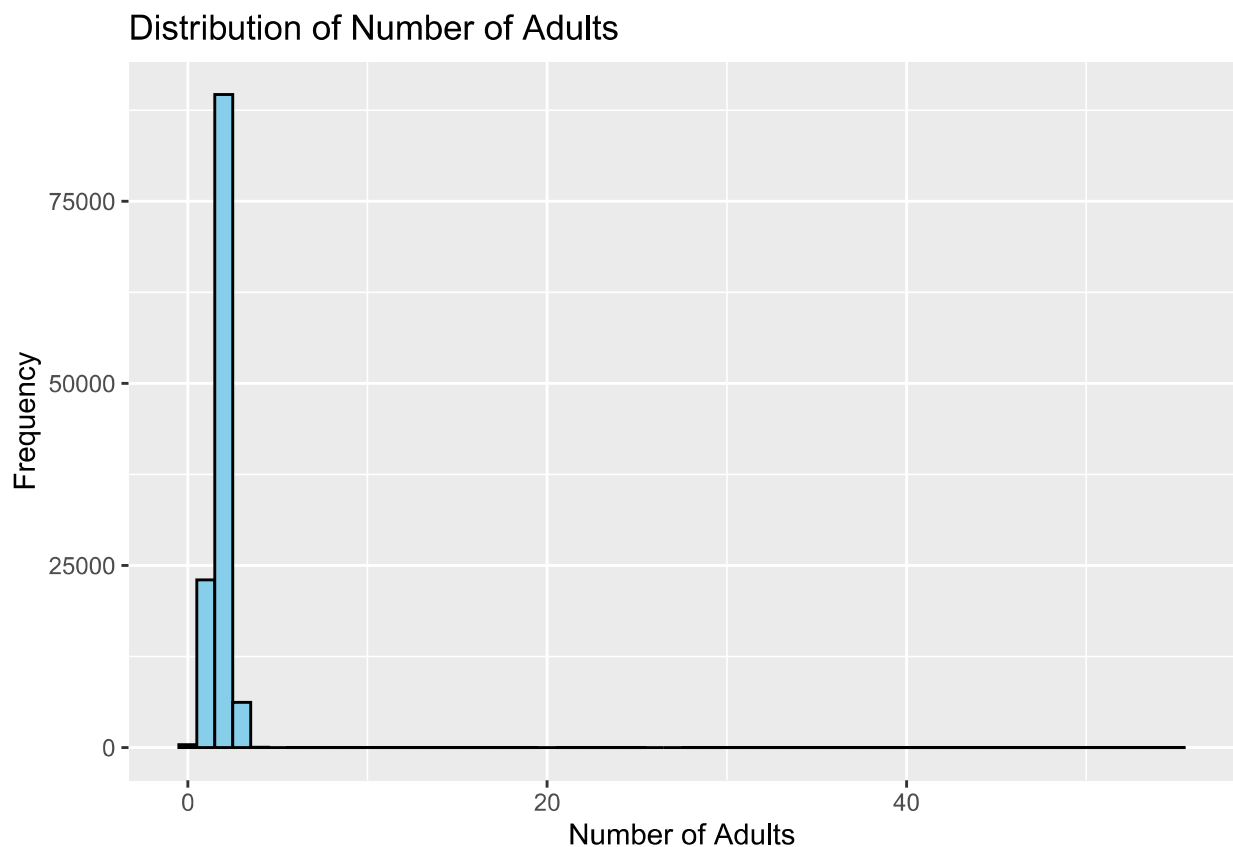
```
summary(data2$stays_in_week_nights)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0      1.0      2.0      2.5      3.0     50.0
```

```
# Visualization
```

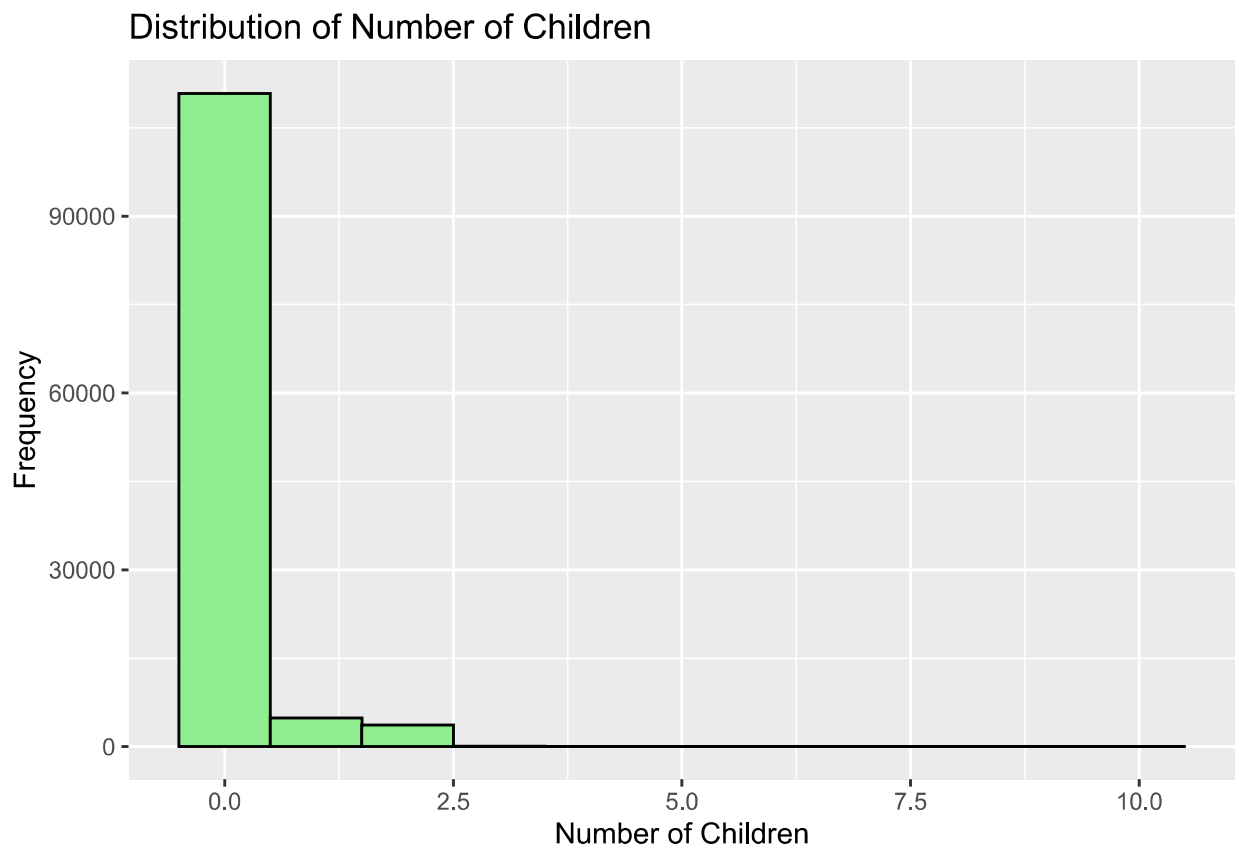
```
# Histogram for number of adults
```

```
ggplot(data2, aes(x = adults)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Number of Adults", x = "Number of Adults", y = "Frequency")
```

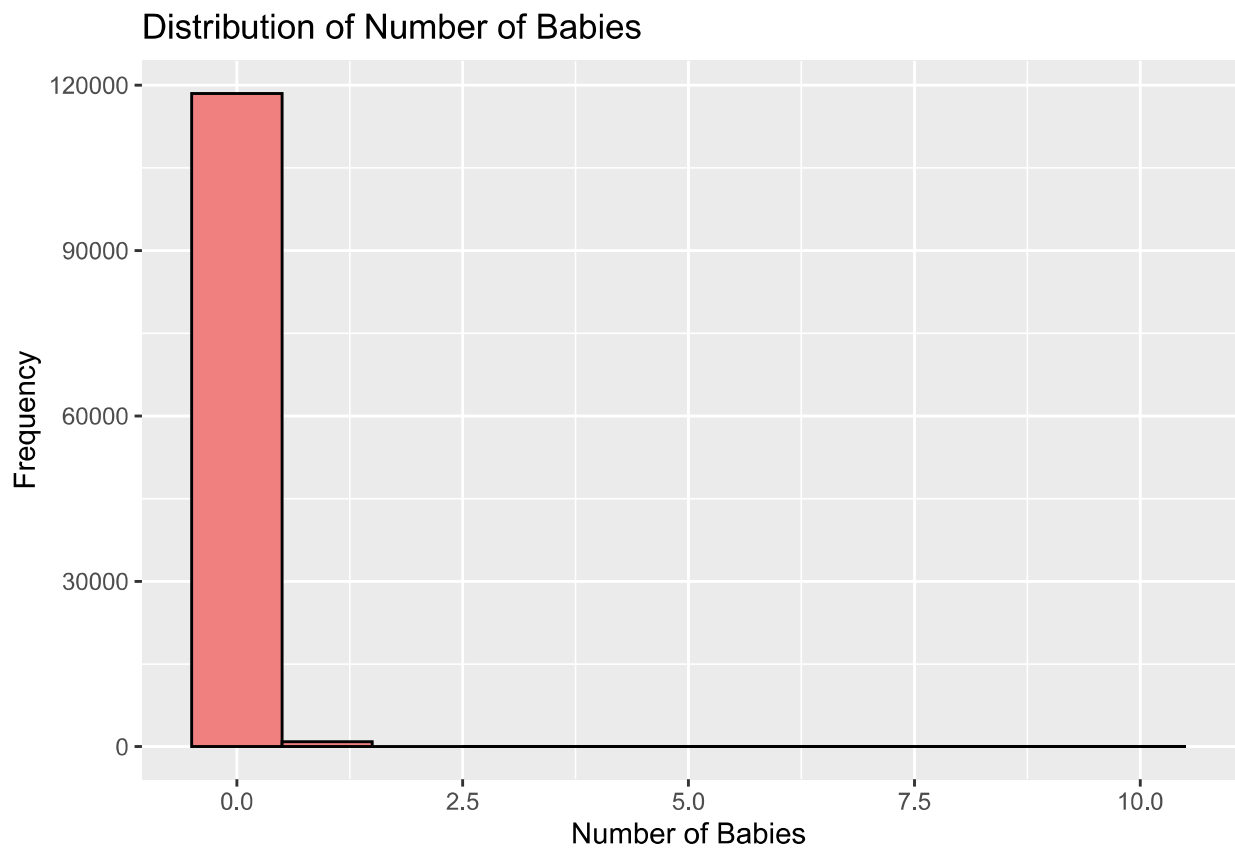


```
# Histogram for number of children
```

```
ggplot(data2, aes(x = children)) +
  geom_histogram(binwidth = 1, fill = "lightgreen", color = "black") +
  labs(title = "Distribution of Number of Children", x = "Number of Children", y = "Frequency")
```

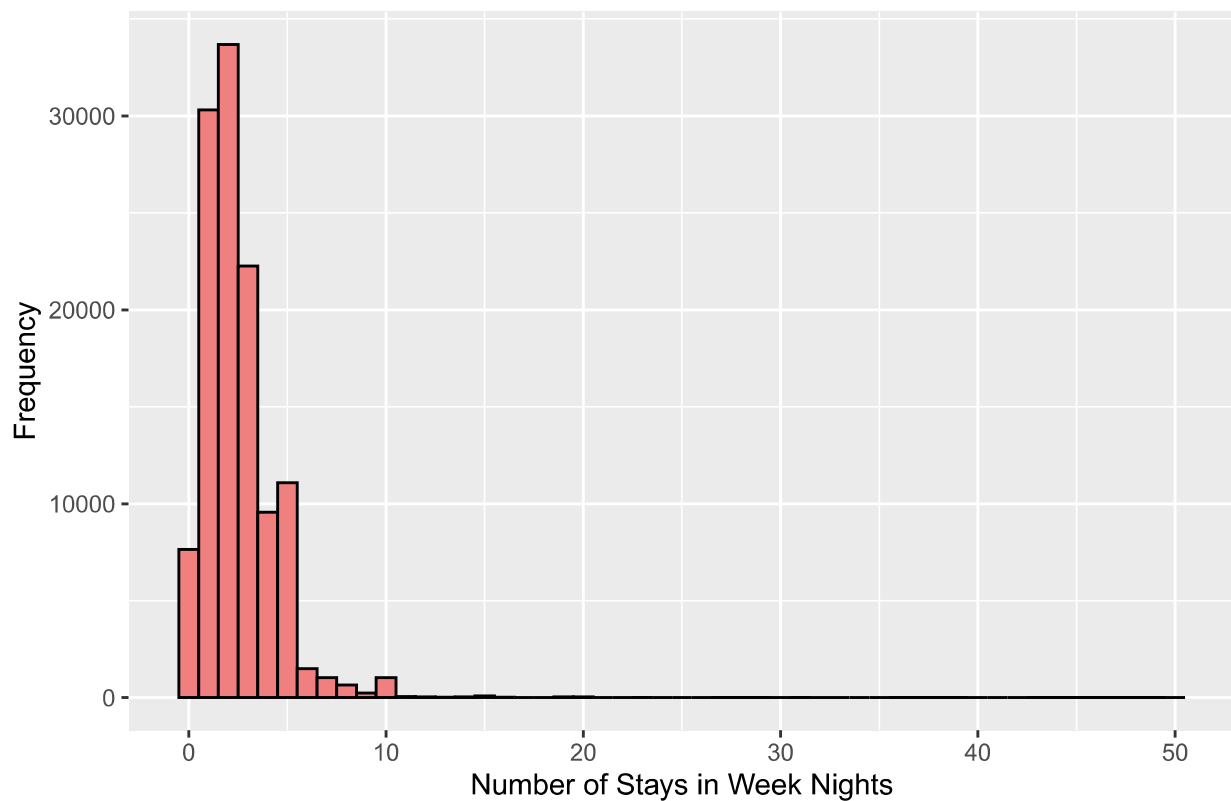


```
# Histogram for number of babies
ggplot(data2, aes(x = babies)) +
  geom_histogram(binwidth = 1, fill = "lightcoral", color = "black") +
  labs(title = "Distribution of Number of Babies", x = "Number of Babies", y = "Frequency")
```



```
# Histogram for number of stays in week nights  
ggplot(data2, aes(x = stays_in_week_nights)) +  
  geom_histogram(binwidth = 1, fill = "lightcoral", color = "black") +  
  labs(title = "Distribution of Number of Stays in Week Nights", x = "Number of Stays in Week Nights")
```

### Distribution of Number of Stays in Week Nights



```
# Correlation Analysis
correlation_matrix <- cor(data2[, c("adults", "children", "babies", "stays_in_week_nights")])
print(correlation_matrix)
```

```
##               adults   children   babies stays_in_week_nights
## adults           1.00000000  0.03044685  0.01814781           0.09298337
## children          0.03044685  1.00000000  0.02402955           0.04420292
## babies            0.01814781  0.02402955  1.00000000           0.02019092
## stays_in_week_nights 0.09298337 0.04420292 0.02019092           1.00000000
```

```
# Regression Analysis
lm_model <- lm(stays_in_week_nights ~ adults + children + babies, data = data2)
summary(lm_model)
```

```
##
## Call:
## lm(formula = stays_in_week_nights ~ adults + children + babies,
##     data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.482  -1.219  -0.520   0.480  47.781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.918117   0.018464 103.882 < 2e-16 ***
## adults        0.301159   0.009489  31.736 < 2e-16 ***
## children      0.196296   0.013793  14.231 < 2e-16 ***
```

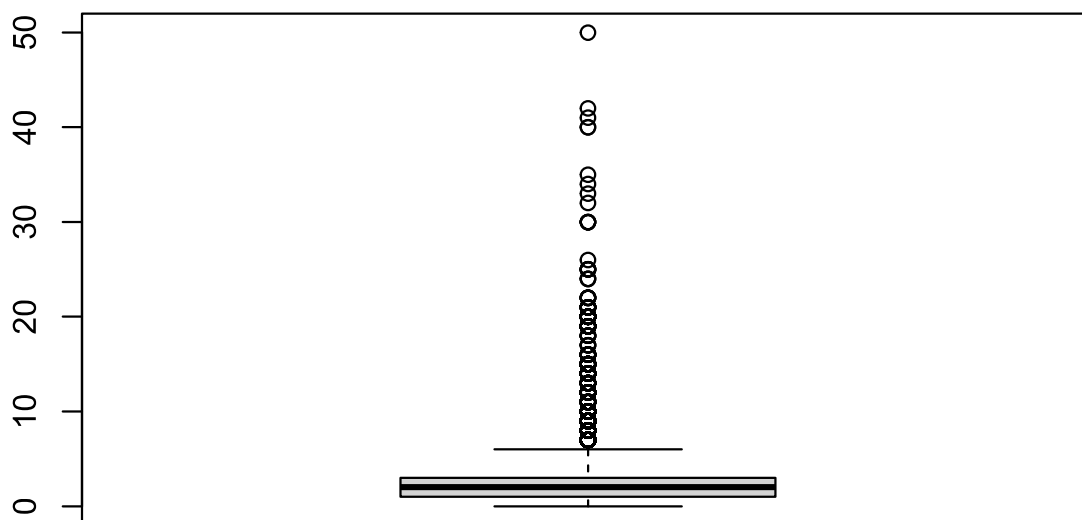
```
## babies      0.343648    0.056404    6.093 1.11e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.898 on 119382 degrees of freedom
## Multiple R-squared:  0.01067,    Adjusted R-squared:  0.01064
## F-statistic:   429 on 3 and 119382 DF,  p-value: < 2.2e-16

# Segmentation Analysis (Example: by number of children)
segment_summary <- data2 %>%
  group_by(children) %>%
  summarise(mean_stays_in_week_nights = mean(stays_in_week_nights),
            median_stays_in_week_nights = median(stays_in_week_nights))
print(segment_summary)

## # A tibble: 5 x 3
##   children mean_stays_in_week_nights median_stays_in_week_nights
##   <int>          <dbl>          <dbl>
## 1     0          2.48            2
## 2     1          2.79            3
## 3     2          2.85            2
## 4     3          2.42            2
## 5    10         10            10

# Outlier Detection (Example: using boxplot)
boxplot(data2$stays_in_week_nights, main = "Boxplot of Number of Stays in Week Nights")
```

## Boxplot of Number of Stays in Week Nights



```
# Validation (Example: using cross-validation)
# (Code for cross-validation would depend on the specific validation technique you want to use)
```