

Analiza rezervacija hotelskog smještaja

Grupa Prosječni

Mario Mrvčić, Rujana Perić, Dorjan Štrbac, Sven Winkler

2023-12-29

OPIS ZADATKA

U današnje vrijeme turistička djelatnost jako je popularna, a tržište puno konkurencije. Zanima nas kako hoteli mogu prilagoditi svoje usluge kako bi poboljšali zadovoljnost gostiju i povećali profitabilnost.

Posjedujući skup podataka koji detaljno dokumentira rezervacije u dva različita hotela želimo istražiti, bolje razumjeti i ispitati veze između boravka u hotelu, godišnjeg doba, socio-ekonomskog statusa gosta, cijena, tipa rezervacija i sličnih parametara danih skupom podataka.

Učitavanje podataka o rezervacijama

Skup podataka sastoji se od 119390 podataka o rezervacijama u dva hotela. Svaki podatak o rezervaciji sastoji se od 31 varijable koje pružaju uvid u trendove dolaska, demografske podatke gostiju, njihove financijske obrasce...

U nastavku je prikazan:

- podatak o broju zapisa rezervacija

```
h1_data = read.csv("Rezervacije/Rezervacije_/H1.csv")
h2_data = read.csv("Rezervacije/Rezervacije_/H2.csv")
data <- rbind(h1_data, h2_data)
dim(data)
```

```
## [1] 119390      31
```

- primjer nekoliko zapisa

```
head(data)
```

```
##   IsCanceled LeadTime ArrivalDateYear ArrivalDateMonth ArrivalDateWeekNumber
## 1          0      342          2015          July              27
## 2          0      737          2015          July              27
## 3          0        7          2015          July              27
## 4          0       13          2015          July              27
## 5          0       14          2015          July              27
## 6          0       14          2015          July              27
##   ArrivalDateDayOfMonth StaysInWeekendNights StaysInWeekNights Adults Children
```

## 1		1		0	0	2	0		
## 2		1		0	0	2	0		
## 3		1		0	1	1	0		
## 4		1		0	1	1	0		
## 5		1		0	2	2	0		
## 6		1		0	2	2	0		
##	Babies	Meal	Country	MarketSegment	DistributionChannel	IsRepeatedGuest			
## 1	0 BB		PRT	Direct	Direct	0			
## 2	0 BB		PRT	Direct	Direct	0			
## 3	0 BB		GBR	Direct	Direct	0			
## 4	0 BB		GBR	Corporate	Corporate	0			
## 5	0 BB		GBR	Online TA	TA/TO	0			
## 6	0 BB		GBR	Online TA	TA/TO	0			
##	PreviousCancellations		PreviousBookingsNotCanceled		ReservedRoomType				
## 1	0				0 C				
## 2	0				0 C				
## 3	0				0 A				
## 4	0				0 A				
## 5	0				0 A				
## 6	0				0 A				
##	AssignedRoomType	BookingChanges		DepositType	Agent	Company			
## 1	C			3 No Deposit	NULL	NULL			
## 2	C			4 No Deposit	NULL	NULL			
## 3	C			0 No Deposit	NULL	NULL			
## 4	A			0 No Deposit	304	NULL			
## 5	A			0 No Deposit	240	NULL			
## 6	A			0 No Deposit	240	NULL			
##	DaysInWaitingList	CustomerType	ADR	RequiredCarParkingSpaces					
## 1	0	Transient	0	0					
## 2	0	Transient	0	0					
## 3	0	Transient	75	0					
## 4	0	Transient	75	0					
## 5	0	Transient	98	0					
## 6	0	Transient	98	0					
##	TotalOfSpecialRequests		ReservationStatus	ReservationStatusDate					
## 1	0		Check-Out	2015-07-01					
## 2	0		Check-Out	2015-07-01					
## 3	0		Check-Out	2015-07-02					
## 4	0		Check-Out	2015-07-02					
## 5	1		Check-Out	2015-07-03					
## 6	1		Check-Out	2015-07-03					

- svi parametri rezervacija

```
names(data)
```

```
## [1] "IsCanceled"           "LeadTime"
## [3] "ArrivalDateYear"      "ArrivalDateMonth"
## [5] "ArrivalDateWeekNumber" "ArrivalDateDayOfMonth"
## [7] "StaysInWeekendNights" "StaysInWeekNights"
## [9] "Adults"               "Children"
## [11] "Babies"               "Meal"
## [13] "Country"              "MarketSegment"
```

## [15]	"DistributionChannel"	"IsRepeatedGuest"
## [17]	"PreviousCancellations"	"PreviousBookingsNotCanceled"
## [19]	"ReservedRoomType"	"AssignedRoomType"
## [21]	"BookingChanges"	"DepositType"
## [23]	"Agent"	"Company"
## [25]	"DaysInWaitingList"	"CustomerType"
## [27]	"ADR"	"RequiredCarParkingSpaces"
## [29]	"TotalOfSpecialRequests"	"ReservationStatus"
## [31]	"ReservationStatusDate"	

Case study: TRAJANJE BORAVKA GOSTA

Slijedi detaljna analiza trajanja boravka gosta s obzirom na različite parametre.

Trajanje prosječnog boravka tijekom različitih godišnjih doba

Dogovorno su godišnja doba definirana na idući način:

ljetno - lipanj, srpanj, kolovoz

jesen - rujan, listopad, studeni

zima - prosinac, siječanj, veljača

proljeće - ožujak, travanj, svibanj

Definiramo četiri pomoćna skupa podataka za svako godišnje doba grupirajući po paramteru “ArrivalDateMonth”.

```
summer = rbind(data[data$ArrivalDateMonth == "June",],
               data[data$ArrivalDateMonth == "July",],
               data[data$ArrivalDateMonth == "August",])

autumn = rbind(data[data$ArrivalDateMonth == "September",],
               data[data$ArrivalDateMonth == "October",],
               data[data$ArrivalDateMonth == "November",])

winter = rbind(data[data$ArrivalDateMonth == "December",],
               data[data$ArrivalDateMonth == "January",],
               data[data$ArrivalDateMonth == "February",])

spring = rbind(data[data$ArrivalDateMonth == "March",],
               data[data$ArrivalDateMonth == "April",],
               data[data$ArrivalDateMonth == "May",])
```

Pomoćnim skupovima dodajemo i parametar “TotalStays” koji predstavlja sumu vrijednosti parametara “StaysInWeekendNights” i “StaysInWeekNights” kako bismo dobili uvid o trajanju boravka gosta neovisno o danima u tjednu.

```
library(magrittr)
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

summer_total <- summer %>%
  mutate(TotalStays = StaysInWeekendNights + StaysInWeekNights)
dim(summer_total)
```

```
## [1] 37477    32
```

```
autumn_total <- autumn %>%  
  mutate(TotalStays = StaysInWeekendNights + StaysInWeekNights)  
dim(autumn_total)
```

```
## [1] 28462    32
```

```
winter_total <- winter %>%  
  mutate(TotalStays = StaysInWeekendNights + StaysInWeekNights)  
dim(winter_total)
```

```
## [1] 20777    32
```

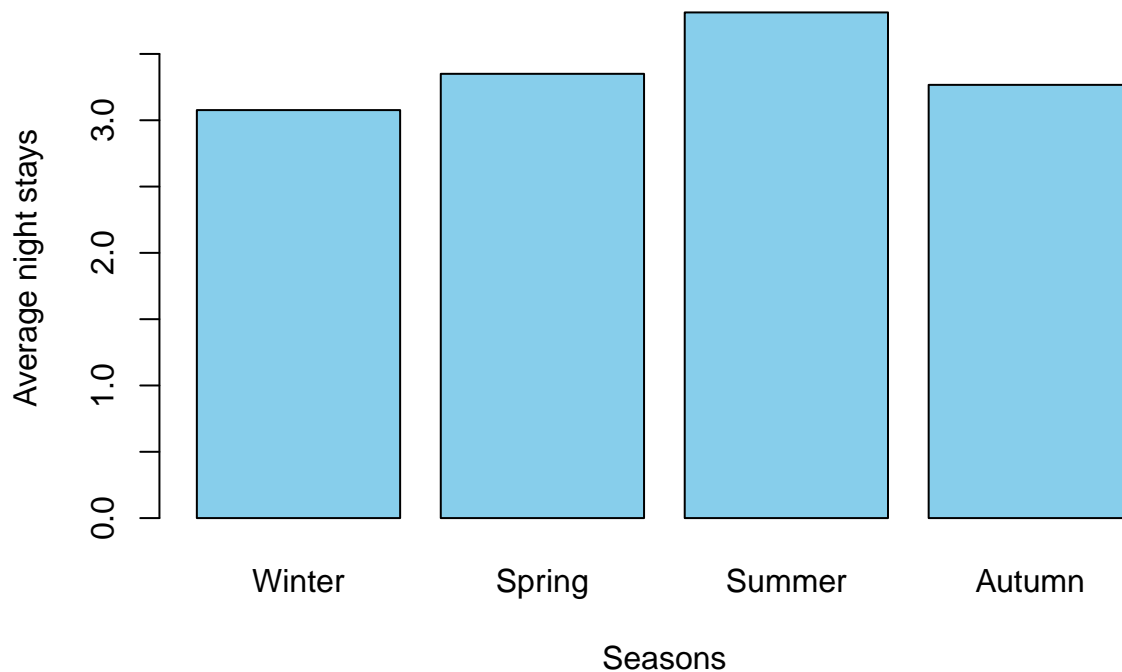
```
spring_total <- spring %>%  
  mutate(TotalStays = StaysInWeekendNights + StaysInWeekNights)  
dim(spring_total)
```

```
## [1] 32674    32
```

Iz novih dimenzija možemo vijeti da je dodan novi parametar svakom zapisu.

Računamo prosječnu duljinu trajanja boravka za svaki pomoćni skup podataka, rezultat prikazujemo stupičastim dijagramom.

```
seasons_means_vector = data.frame(  
  Season = c("Winter", "Spring", "Summer", "Autumn"),  
  Means = c(mean(winter_total$TotalStays),  
            mean(spring_total$TotalStays),  
            mean(summer_total$TotalStays),  
            mean(autumn_total$TotalStays))  
)  
  
barplot(seasons_means_vector$Means, names.arg = seasons_means_vector$Season, col = "skyblue", xlab = "S
```



Iz histograma možemo zaključiti da je prosječni boravak gostiju najduži ljeti, a najkraći zimi.

Trajanje prosječnog boravka i tip iznajmljene sobe

Naše istraživačko pitanje je slijedeće:

Možemo li trajanje boravka povezati s tipom iznajmljene sobe?

Pogledajmo najprije koje sve kategorije soba za iznajmljivanje postoje.

```
#binding adjusted data
data_total_stays = rbind(winter_total, spring_total, summer_total, autumn_total)
#all different values of "ReservedRoomType" variable
```

... objašnjenja što je koja soba...

```
#calculating average "TotalStays" for every room type
```

```
#plotting data...
```

Iz priloženog možemo vidjeti da je prosječna duljina boravka najduža ..., a najkraća ...

```
#box and whisker plot --> mean, rangs, median, variances...
```

... komentar na box and whisker plot...

Imajući detaljan uvid u dane statistike, nas zanima postoji li povezanost između duljine boravka i tipa iznajmljene sobe na razini značajnosti od 5%.

Naša testne hipoteze biti će:

H₀: Prosječna duljina boravka jednaka je za sve tipove soba.

H₁: Za barem jedan tip sobe prosječna duljina boravka različita je od ostalih.

Test koji koristimo za ispirivanje ove hipoteze je ANOVA.

#ANOVA

Zaključak:

.....

Diskusija:

.....

Trajanje prosječnog boravka i kategorija gosta

Naše iduće istraživačko pitanje je slijedeće:

Možemo li trajanje boravka povezati s kategorijom gosta?

Analizirajmo postojeće podatke iz database. Koje sve kategorije gostiju postoje.

#all different values of "CustomerType" variable

... objašnjenja što je koja kategorija gosta...

#calculating average "TotalStays" for every customer type

#plotting data...

Iz priloženog možemo vidjeti da je prosječna duljina boravka najduža ..., a najkraća ...

#box and whisker plot --> mean, rangs, median, variances....

... komentar na box and whisker plot...

Zanima postoji li povezanost između duljine boravka i kategorije gosta na razini značajnosti od 5%.

Naša testne hipoteze biti će:

H₀: Prosječna duljina boravka jednaka je za sve tipove gostiju.

H₁: Za barem jednu kategoriju gosta prosječna duljina boravka različita je od ostalih.

Test koji koristimo za ispirivanje ove hipoteze je ANOVA.

#ANOVA

Zaključak:

.....

Diskusija:

.....

..... ZAKLJUČNA DISKUSIJA ZA CIJELI CASE STUDY

Case study: CIJENE

Case study: PANSION, POLUPANSION

Case study: OTKAZIVANJE REZERVACIJA

Case study: SKUPINE GOSTIJU IZ RAZLIČITIH ZEMALJA