

Mobilnost turista i poslovanje hrvatskog turističkog sektora

tim Paranormalna distribucija

Ivan Kapusta, Antonio Lukić, David Supančić, Leon Stjepan Uroić

2023-04-29

Contents

1	Uvod	2
2	Analiza	3
2.1	Putnici i sredstva	3
2.2	Fundamenti	7
2.3	Veza prometa na graničnim prijelazima s fundamentima poduzeća	11
2.4	Distribucija ostvarenih prodaja turističke tvrtke Arena Hospitality Group d.d.	25
2.5	Analiza efikasnosti poduzeća	35
2.6	Analiza dnevnih kretanja cijene indeksa CROBEXturist	36
3	Zaključak	47

1 Uvod

Hrvatska je omiljena turistička destinacija zbog svoje obale i otoka te privlači mnogo turista svake godine. Turizam predstavlja važan segment u ekonomiji Hrvatske. Zemlja se ističe bogatom poviješću, kulturom i raznolikim prirodnim ljepotama, što ju čini atraktivnom za različite turiste. Kada je riječ o mobilnosti turista, uglavnom se koriste avioni ili automobili, kao što ćemo i vidjeti iz grafova, no postoje i druge opcije, kao što su autobusi, vlakovi ili trajekti. Hrvatska ima dobro razvijenu mrežu cesta i autocesta, što olakšava kretanje unutar zemlje.

U ovome projektnome zadatku opisati ćemo mobilnost turista u Hrvatskoj kroz godine podijeljene na kvartale. Mobilnost je izražena kao broj graničnih prijelaza uz dodatne informacije radi li se o domaćem ili stranom turistu te korištenom prijevoznom sredstvu. Zatim ćemo se baviti prihodima od prodaje i dobitom četiri velika poduzeća u Hrvatskoj: Arena Hospitality Group d.d., Imperial Riviera d.d., Maistra d.d. te Valamar Riviera d.d. Podaci koje analiziramo su iz perioda od 2013. do 2022.

Isražit ćemo razlike među vrstama graničnih prijelaza i promotriti na koji način dan u tjednu te mjesec utječe na broj graničnih prijelaza. Nadamo se uočiti sezonalnosti u fundamentima turističkih tvrtki i objasniti njihovu vezu s brojem graničnih prijelaza. Također želimo objanisliti dnevna kretanja cijene indeksa CROBEXTurist pomoću prijelaza turista preko granica. Isto tako nas zanima postoje li razlike u samim kvartalima te može li se pokazati veza između broja prodaja i godine u kojoj su ostvarene.

2 Analiza

2.1 Putnici i sredstva

Donje prikazane tablice služe kako bi dobili uvid s kakvim podacima radimo. Izvedena je osnovna analiza svakog stupca. Bavimo se cestovnim, željezničkim, riječnim te pomorskim prometom. Ovih par redaka nam je dovoljno da uočimo trend razdiobe korištenih prijevoznih sredstava. Zanimljivo je kako se dosta rijetko koristi riječni promet.

```
kable(head(domaciIzlazPutniciDf))
```

X	Cestovni	Željeznički	Riječni	Pomorski	Zračni	UKUPNO
2013-01-01	23581	334	0	27	590	24532
2013-01-02	46762	687	0	26	1291	48766
2013-01-03	52199	694	0	63	973	53929
2013-01-04	55030	604	0	65	987	56686
2013-01-05	81207	636	0	1	1005	82849
2013-01-06	51806	664	0	119	1464	54053

```
summary(domaciIzlazPutniciDf)
```

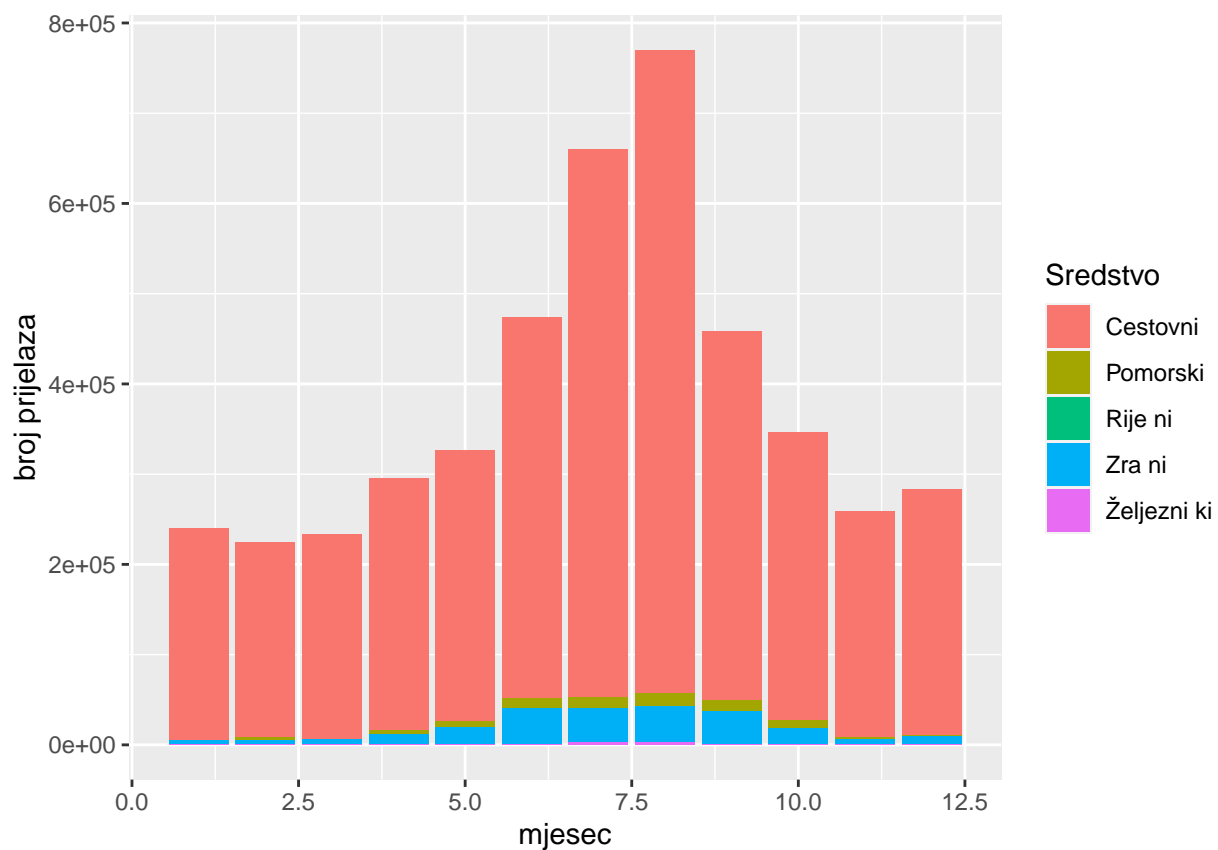
```
##      X              Cestovni      Željeznički      Riječni
## Length:3595      Min.   :    271      Min.   :    0.0      Min.   :    0.000
## Class :character  1st Qu.: 45583      1st Qu.:   65.0      1st Qu.:    0.000
## Mode  :character  Median : 56032      Median :  130.0      Median :    0.000
##                      Mean  : 56097      Mean  :  167.0      Mean   :    1.361
##                      3rd Qu.: 66553      3rd Qu.: 240.5      3rd Qu.:    0.000
##                      Max.   :1394692      Max.   :1024.0      Max.   :3396.000
##      Pomorski      Zračni      UKUPNO
## Min.   :    0.00      Min.   :    0      Min.   :   285
## 1st Qu.:    8.00      1st Qu.:   816      1st Qu.: 47166
## Median :   39.00      Median :  1235      Median : 57695
## Mean   :   51.74      Mean   :   2465      Mean   : 58781
## 3rd Qu.:   74.00      3rd Qu.:  1859      3rd Qu.: 68468
## Max.   :3493.00      Max.   :4000595      Max.   :4072252
```

```
summary(domaciUlazPutniciDf)
```

```
##      X              Cestovni      Željeznički      Riječni
## Length:3595      Min.   :    318      Min.   :    0.0      Min.   :    0.000
## Class :character  1st Qu.: 45560      1st Qu.:   68.0      1st Qu.:    0.000
## Mode  :character  Median : 55804      Median :  130.0      Median :    0.000
##                      Mean  : 56986      Mean  :  169.1      Mean   :    1.137
##                      3rd Qu.: 66528      3rd Qu.: 240.0      3rd Qu.:    0.000
##                      Max.   :4504629      Max.   :1367.0      Max.   :2916.000
##      Pomorski      Zračni      UKUPNO
## Min.   :    0.0      Min.   :    0.0      Min.   :   318
## 1st Qu.:    8.0      1st Qu.: 579.0      1st Qu.: 46654
## Median :   36.0      Median : 919.0      Median : 57074
## Mean   :   326.2      Mean   : 936.1      Mean   : 58419
## 3rd Qu.:   71.0      3rd Qu.:1207.5      3rd Qu.: 68061
## Max.   :1000000.0      Max.   :5934.0      Max.   :4505966
```

Prikažimo ukupan broj graničnih prijelaza i koliki udio čine koje prijevozno sredstvo.

```
zbrojSvihVrstaPoMjesecu <- aggregate(dfZbrojSvihVrsta[,-1],  
                                     list(month(dfZbrojSvihVrsta$X)),  
                                     function(x){round(mean(x),0)})  
  
pivot <- pivot_longer(zbrojSvihVrstaPoMjesecu,  
                      Cestovni:Zračni,  
                      names_to = "Sredstvo",  
                      values_to = "brojPrelazaka")  
  
ggplot(pivot, aes(x = Group.1, y = brojPrelazaka, fill = Sredstvo)) +  
  geom_bar(stat = 'identity') + labs(x = 'mjesec', y = 'broj prijelaza')
```



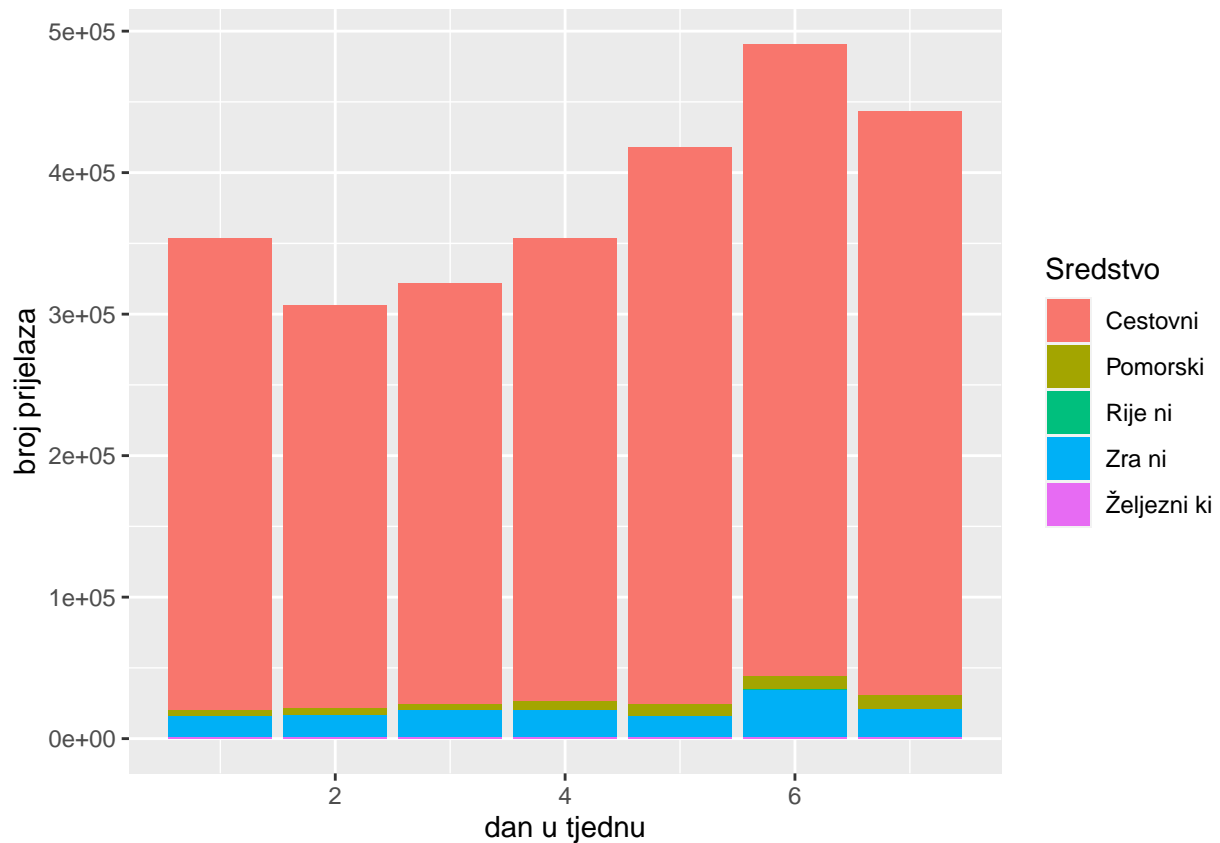
Možemo primjetiti kako cestovni promet čini skoro sav promet preko hrvatskih granica. Zračni promet se nešto češće koristi u 3. kvartalu dok su ostala prijevozna sredstva gotovo zanemariva. Isto tako prijelazi preko granice najviše se događaju u 7. mjesecu odnosno “srcu sezone”.

Slično kao i u gornjem grafu zanima nas broj prijelaza i udio sredstava u odnosu na dan u tjednu.

```
zbrojSvihVrstaPoDanuUTjednu <- aggregate(dfZbrojSvihVrsta[, -1],
                                          list(wday(dfZbrojSvihVrsta$X)),
                                          function(x){round(mean(x), 0)})

pivot <- pivot_longer(zbrojSvihVrstaPoDanuUTjednu,
                      Cestovni:Zračni,
                      names_to = "Sredstvo",
                      values_to = "brojPrelazaka")

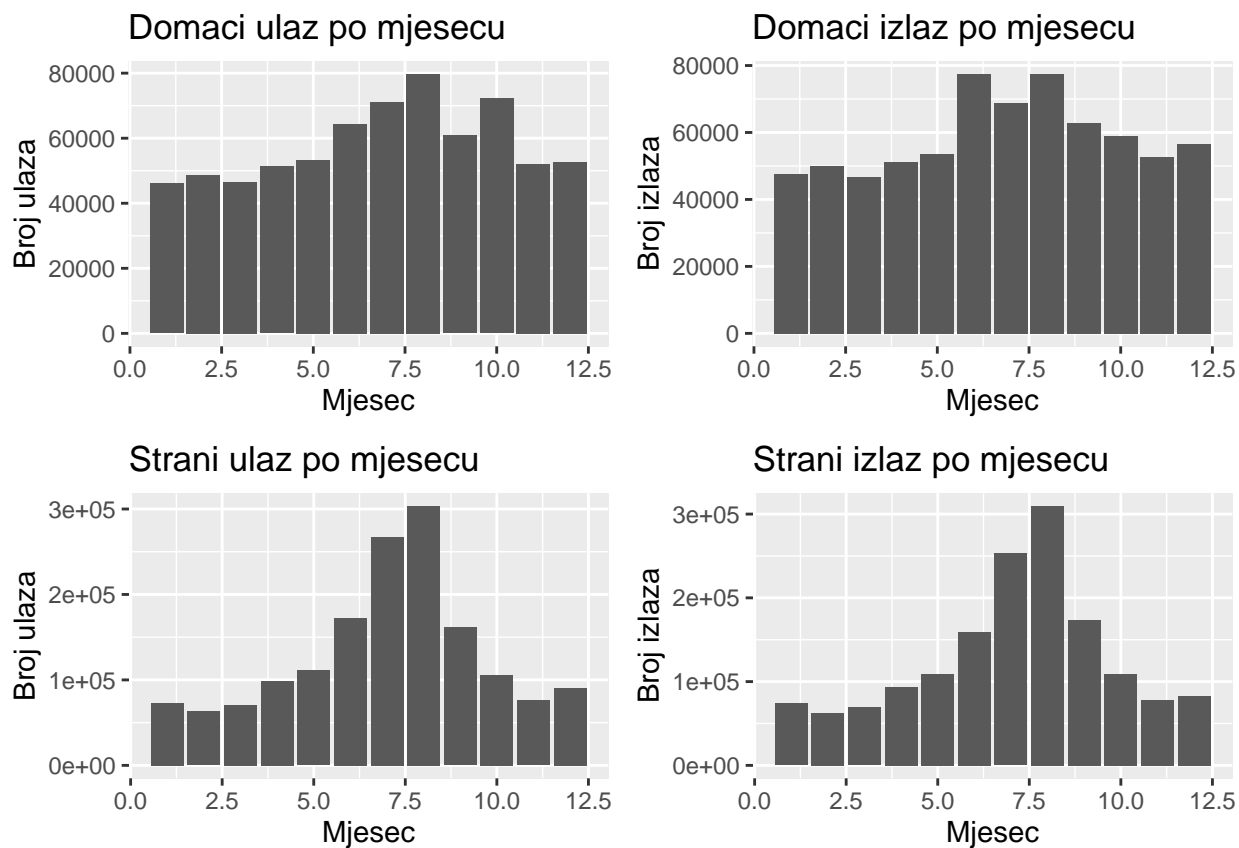
pivot$Group.1[pivot$Group.1 == 1] = 8
pivot$Group.1 = pivot$Group.1 - 1
ggplot(pivot, aes(x = Group.1, y = brojPrelazaka, fill = Sredstvo)) +
  geom_bar(stat = 'identity') + labs(x = 'dan u tjednu', y = 'broj prijelaza')
```



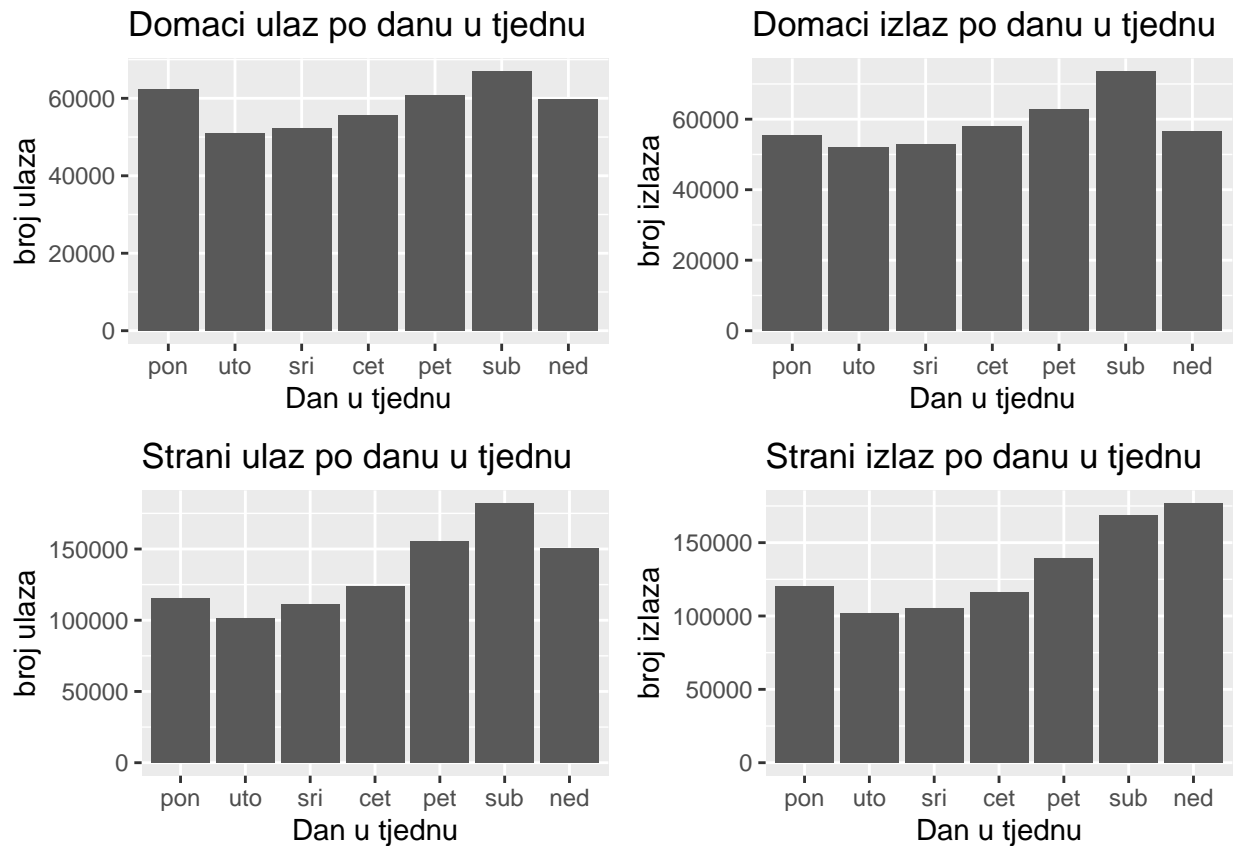
Kao što smo mogli i pretpostaviti najveći broj prijelaza se događa preko vikenda. Isto tako iz ovoga grafa možemo zaključiti da je najbolji dan za izbjeci gužvu utorak.

Mjesečna razdioba stranih ulazaka i izlazaka praktički je jednaka, mali brojevi početkom godine s porastom u 6. mjesecu i potpunim skokom u visine u 7. i 8. mjesecu te smanjivanje u zadnjim mjesecima godine. Sezonalnosti u mjesecima kod domaćih putnika, najveće frekvencije izlazaka mogu se primjetiti na početku ljetne sezone u 6. mjesecu i sredinom sezone u 8. Što se tiče ulazaka njihov broj raste kroz godinu i doseže vrhunac u 8. mjesecu, relativno naglo padne u 9. i ponovno poskoči u 10. mjesecu.

```
grid.arrange(dUlMon,dIzMon,sUlMon,sIzMon, nrow = 2, ncol = 2)
```



```
grid.arrange(dUlDow,dIzDow,sUlDow,sIzDow, nrow = 2, ncol = 2)
```



Broj izlazaka je malen na početku i kraju tjedna, od srijede raste i doseže vrhunac petkom, slična stvar je i kod ulazaka, iako ih vikendom ima više nego početkom tjedna. Izlaz stranih putnika raste kako tjedan napreduje i doseže vrhunac subotom i onda se spušta na niske razine u nedjelju.

2.2 Fundamenti

Slijedeći tip podataka kojima se bavimo su fundamenti, prihodi, broj prodaja i dobit turističkih tvrtki. Analizirati ćemo njihove prihode po kvartalima te pitati se postoje li razlike među kvartalima te kako se mijenjaju kvartali kroz godine.

```
summary(ARNTFundamentiDf)
```

```
##   Published      Year      Quarter      Earnings
## Length:39      Min.   :2013      Min.   :1.000      Min.   : -216378929
## Class :character 1st Qu.:2015      1st Qu.:1.500      1st Qu.: -47695480
## Mode  :character Median :2017      Median :2.000      Median : -34023791
##                  Mean  :2017      Mean  :2.462      Mean  :  4088309
##                  3rd Qu.:2020      3rd Qu.:3.000      3rd Qu.: 22486886
##                  Max.   :2022      Max.   :4.000      Max.   : 220927874
## Sales
## Min.   : 7913339
## 1st Qu.: 27382392
```

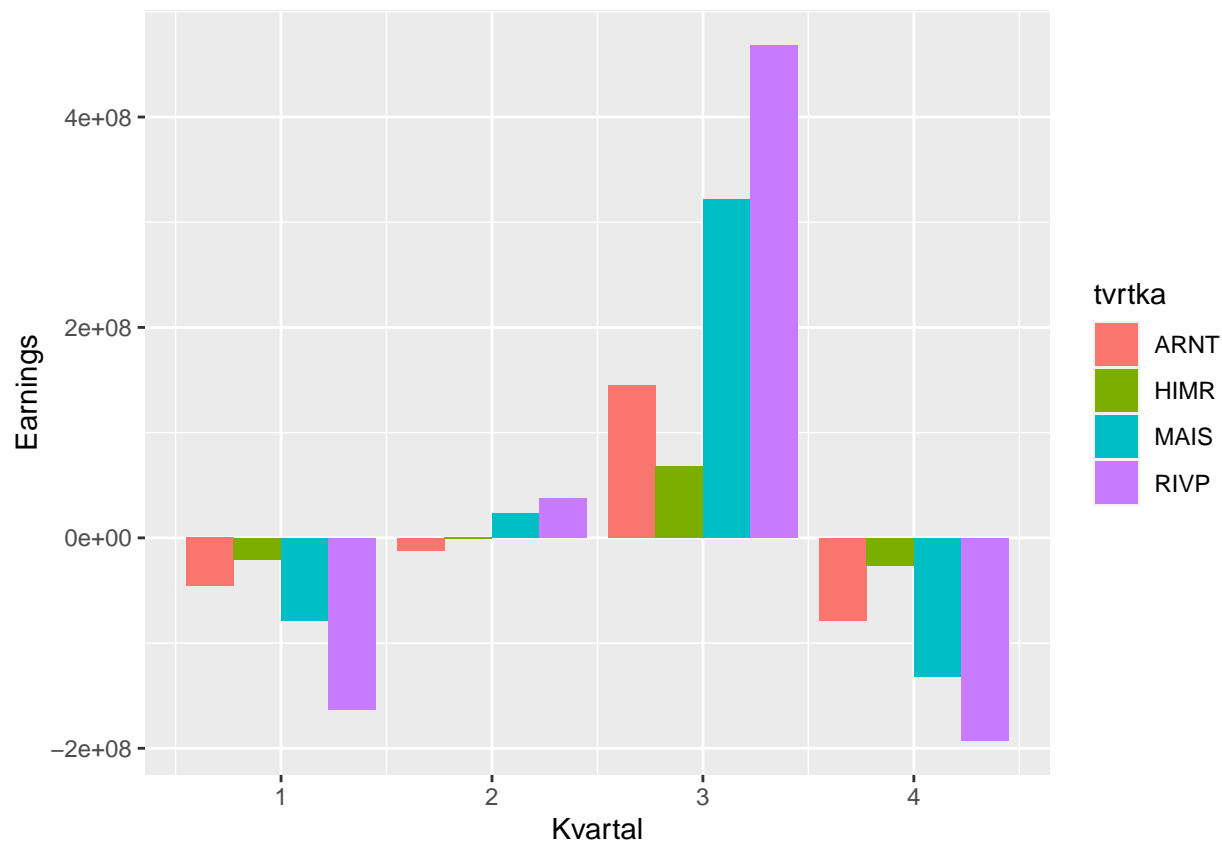
```
## Median : 72383531
## Mean   :129780841
## 3rd Qu.:188087792
## Max.    :508990943
```

```
kable(head(ARNTFundamentiDf))
```

Published	Year	Quarter	Earnings	Sales
2013-05-02	2013	1	-40837883	9310313
2013-07-30	2013	2	-13222299	61366796
2013-10-29	2013	3	107699001	214207984
2014-04-25	2013	4	-52904702	11350774
2014-04-29	2014	1	-38572197	7913339
2014-07-29	2014	2	-12080056	71289588

Pogledajmo prosječne zarade po kvartalu.

```
ARNTFundamentPoKvartalu <- aggregate(ARNTFundamentiDf[,c(-1,-2,-3)],
                                     list(ARNTFundamentiDf$Quarter),
                                     function(x){round(mean(x),0)})
ARNTFundamentPoKvartalu$tvrtka = 'ARNT'
HIMRFundamentPoKvartalu <- aggregate(HIMRFundamentiDf[,c(-1,-2,-3)],
                                     list(HIMRFundamentiDf$Quarter),
                                     function(x){round(mean(x),0)})
HIMRFundamentPoKvartalu$tvrtka = 'HIMR'
MAISFundamentPoKvartalu <- aggregate(MAISFundamentiDf[,c(-1,-2,-3)],
                                     list(MAISFundamentiDf$Quarter),
                                     function(x){round(mean(x),0)})
MAISFundamentPoKvartalu$tvrtka = 'MAIS'
RIVPFundamentPoKvartalu <- aggregate(RIVPFundamentiDf[,c(-1,-2,-3)],
                                     list(RIVPFundamentiDf$Quarter),
                                     function(x){round(mean(x),0)})
RIVPFundamentPoKvartalu$tvrtka = 'RIVP'
sviFundamentiPoKvartalu = rbind(ARNTFundamentPoKvartalu,
                                HIMRFundamentPoKvartalu,
                                MAISFundamentPoKvartalu,
                                RIVPFundamentPoKvartalu)
ggplot(data=sviFundamentiPoKvartalu, aes(x=Group.1, y=Earnings, fill=tvrtka)) +
  geom_bar(stat="identity", position=position_dodge()) + labs(x = "Kvartal", y="Earnings")
```

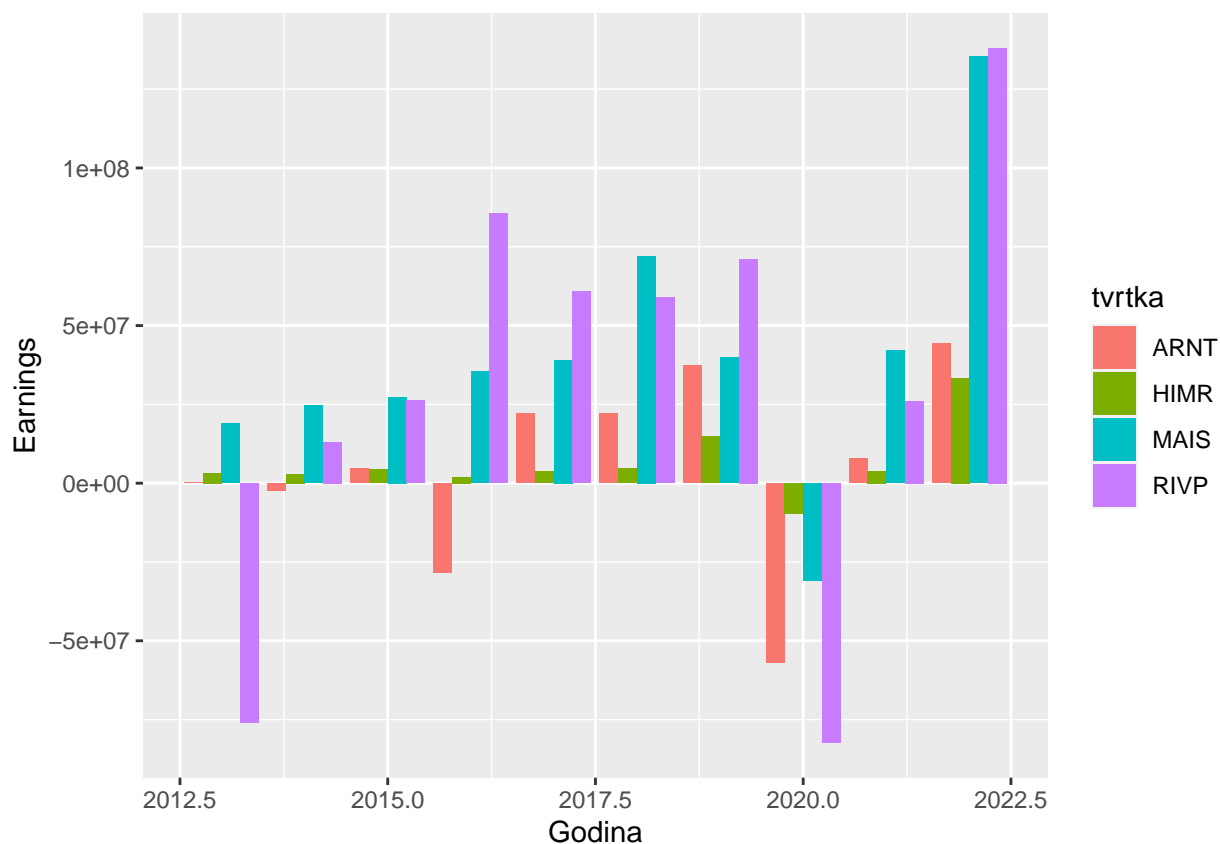



Možemo primjetiti kako profit ostvaren u trećem kvartalu omogućava preživljavanje kompanija kroz ostala tri kvartala. Štoviše, možemo reći da se živi i umire u trećem kvartalu što upućuje na ovisnost o ljetnom turizmu. Tu dominaciju trećeg kvartala ćemo nastaviti uočavati i u ostatku dokumenta.

Sljedeći graf prikazuje zaradu svih tvrtki kroz godine.

```
ARNTFundamentPoGodini <- aggregate(ARNTFundamentiDf[,c(-1,-3)],
                                   list(ARNTFundamentiDf$Year),
                                   function(x){round(mean(x),0)})
ARNTFundamentPoGodini$tvrтка = 'ARNT'
HIMRFundamentPoGodini <- aggregate(HIMRFundamentiDf[,c(-1,-3)],
                                   list(HIMRFundamentiDf$Year),
                                   function(x){round(mean(x),0)})
HIMRFundamentPoGodini$tvrтка = 'HIMR'
MAISFundamentPoGodini <- aggregate(MAISFundamentiDf[,c(-1,-3)],
                                   list(MAISFundamentiDf$Year),
                                   function(x){round(mean(x),0)})
MAISFundamentPoGodini$tvrтка = 'MAIS'
RIVPFundamentPoGodini <- aggregate(RIVPFundamentiDf[,c(-1,-3)],
                                   list(RIVPFundamentiDf$Year),
                                   function(x){round(mean(x),0)})
RIVPFundamentPoGodini$tvrтка = 'RIVP'
sviFundamentiPoGodini = rbind(ARNTFundamentPoGodini,
                              HIMRFundamentPoGodini,
                              MAISFundamentPoGodini,
                              RIVPFundamentPoGodini)
ggplot(data=sviFundamentiPoGodini, aes(x=Group.1, y=Earnings, fill=tvrтка)) +
```

```
geom_bar(stat="identity", position=position_dodge()) + labs(x = "Godina", y="Earnings")
```



Kroz godine se primjećuje lagani rast profita od 2013. do 2019. kada nastupa COVID-19 i profiti su u debelom minusu. 2021. je godina oporavka u kojoj se još uvijek osjete posljedice dok se 2022. ostvaruju rekordni profiti. Pretpostavljamo da je ljudima nedostajalo putovanje.

S obzirom na to da se kod sve četiri tvrtke primjećuju zapravo isti trendovi odlučili smo se uzeti samo jednu tvrtku i nad njom provesti analizu statističkim metodama. Odabrani podaci su prikazani na donjem grafu.

```
ARNTExtendedDf$Quarter <- as.factor(ARNTExtendedDf$Quarter)
ggplot(ARNTExtendedDf, aes(x = Year, y = Sales, fill = Quarter)) + geom_bar(stat = "identity", position
```

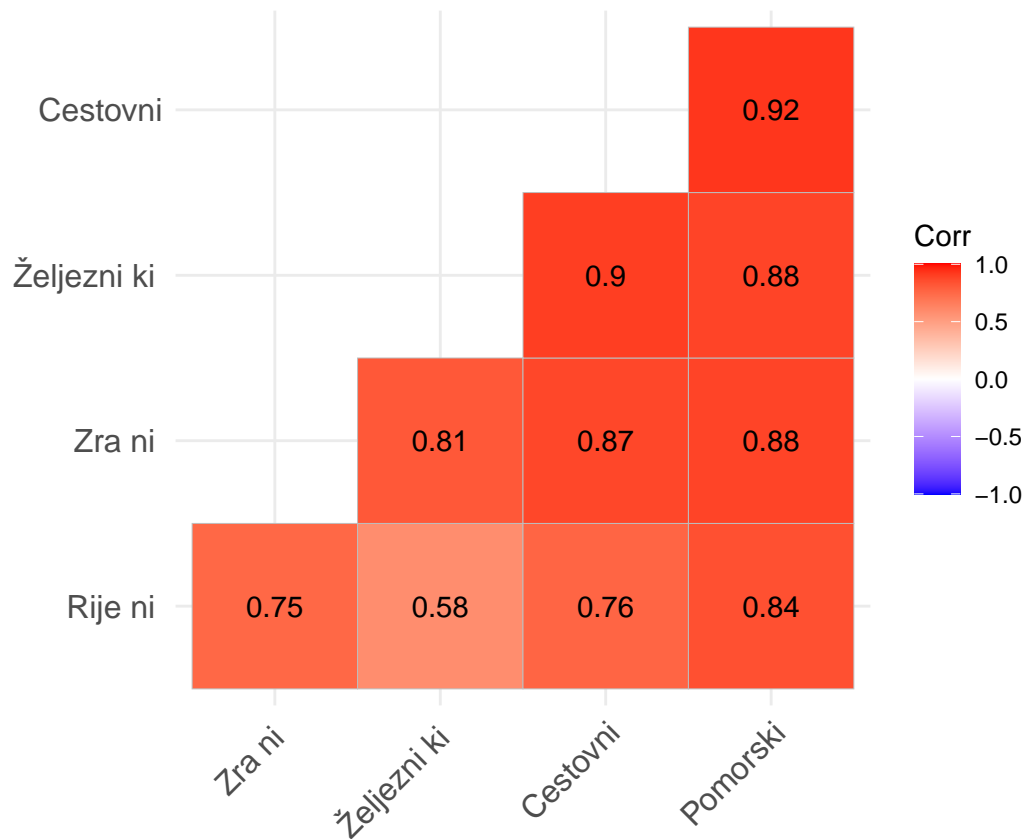


2.3 Veza prometa na graničnim prijelazima s fundamentima poduzeća

Za daljnju analizu ograničavamo se na podatke o ulasku stranih i prihode od prodaje poduzeće Arena Hospitality Group d.d.

Na sljedećem grafu prikazujemo korelaciju broja stranih ulazaka preko različitih vrsta graničnih prijelaza.

```
reduced_data <- subset(putnici.fundamenti.data.d, select = c(3:7))
corr_matrix = round(cor(reduced_data), 2)
ggcorrplot(corr_matrix, hc.order = TRUE, type = "lower", lab = TRUE)
```



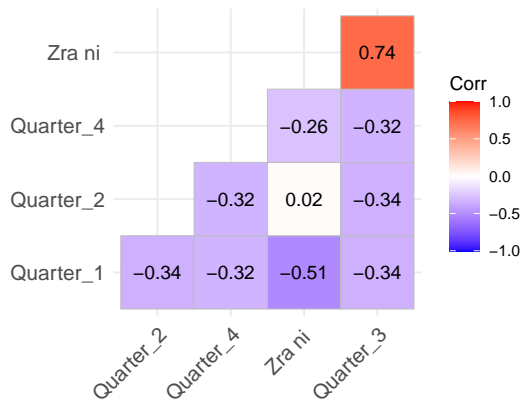
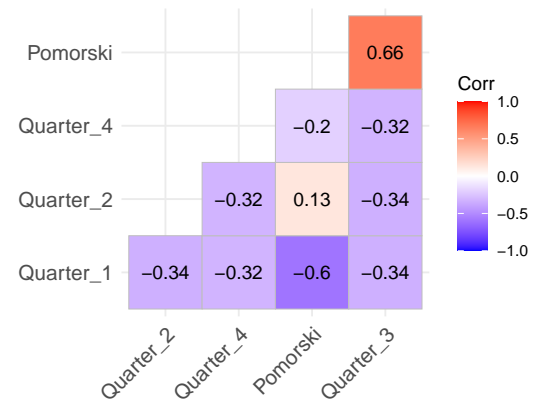
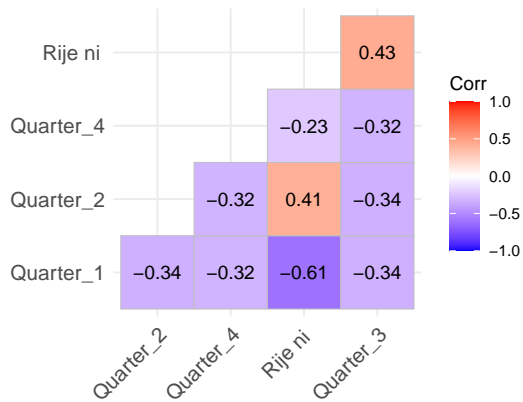
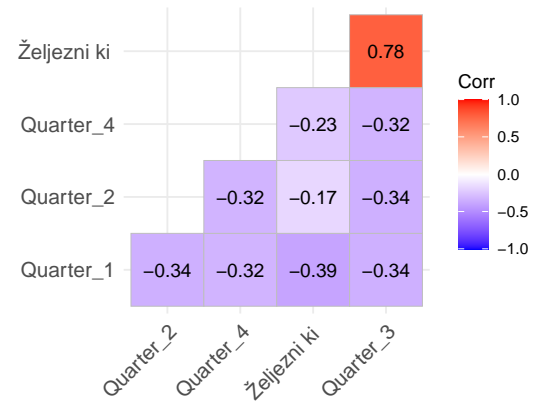
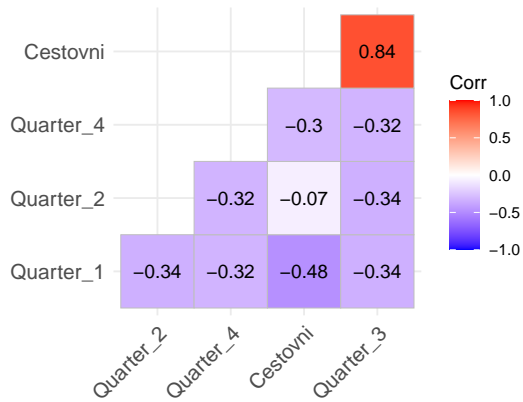
Očekivano postoji visoka korelacija. Ova činjenica nas prisiljava na korištenje više jednostavnih linearnih regresija za svaku vrstu graničnog prijelaza umjesto jedne višestruke.

```
plots <- list()

for (x in 3:7) {
  reduced_data <- subset(putnici.fundamenti.data.d, select = c(x, 11:14))
  corr_matrix = round(cor(reduced_data), 2)

  plots[[x-2]] = ggcorrplot(corr_matrix, hc.order = TRUE, type = "lower", lab = TRUE)
}

grid.arrange(grobs = plots, ncol=2)
```



Iz grafa korelacije broja prelazaka i indikatorskih varijabli kvartala naslućujemo da bi sezonalnost mogla biti prekrivena varijabla koja zapravo objašnjava i broj prelazaka i prihode od prodaje. Kako bi ispitali značajnost ovih regresora u objašnjanju prihoda od prodaje, provodimo višestruku regresiju.

```
fit = lm(Sales ~ Cestovni + Quarter_1 + Quarter_2 + Quarter_3 + Quarter_4, putnici.fundamenti.data.d)
print(summary(fit))
```

```
##
## Call:
## lm(formula = Sales ~ Cestovni + Quarter_1 + Quarter_2 + Quarter_3 +
##     Quarter_4, data = putnici.fundamenti.data.d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

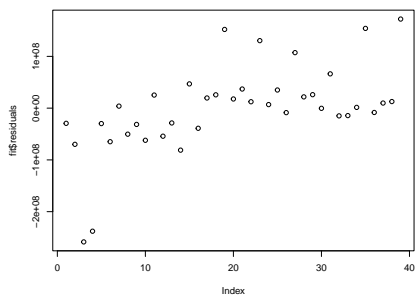
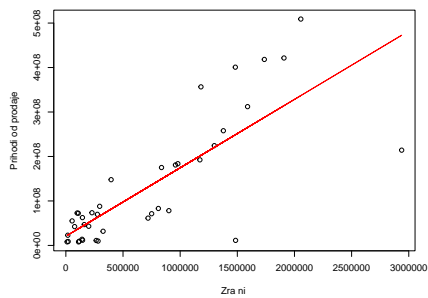
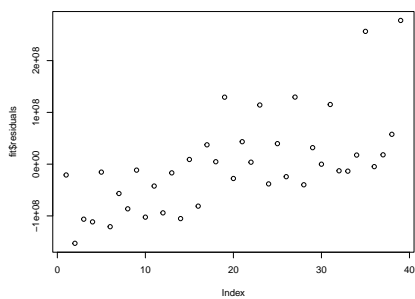
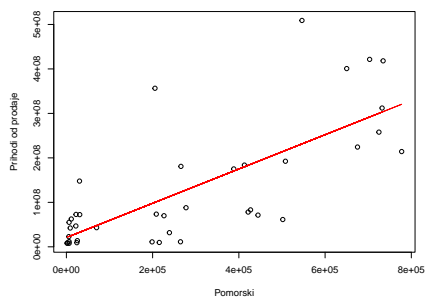
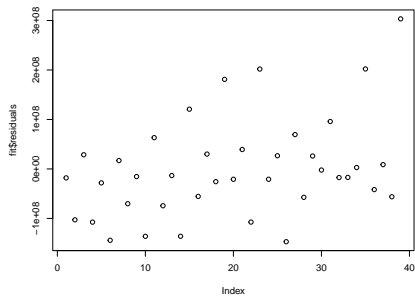
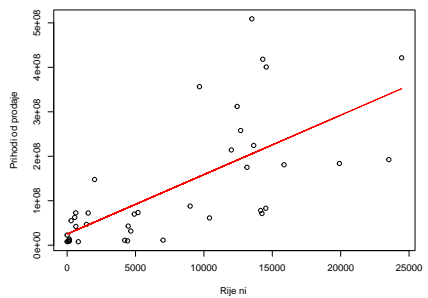
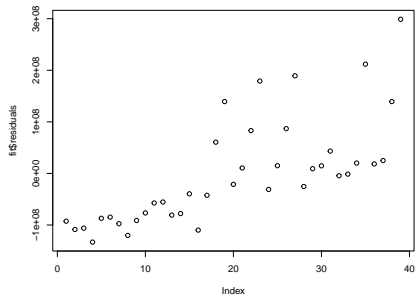
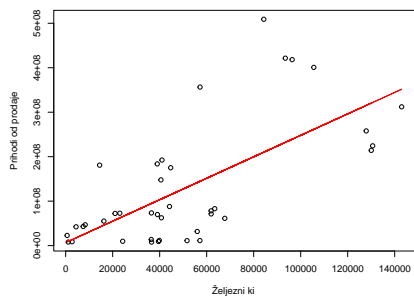
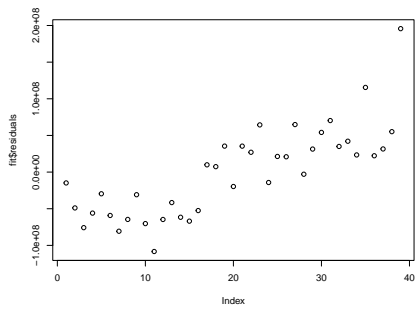
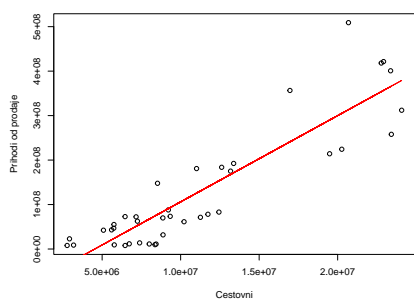
```
## -107888659 -36075695 12462360 31468923 177283016
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.834e+07  3.094e+07  -1.885  0.06795 .
## Cestovni     1.248e+01  3.100e+00   4.026  0.00030 ***
## Quarter_1    1.801e+07  2.730e+07   0.660  0.51391
## Quarter_2    3.749e+07  2.806e+07   1.336  0.19042
## Quarter_3    1.318e+08  4.707e+07   2.800  0.00837 **
## Quarter_4             NA          NA      NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58320000 on 34 degrees of freedom
## Multiple R-squared:  0.8403, Adjusted R-squared:  0.8216
## F-statistic: 44.74 on 4 and 34 DF,  p-value: 4.347e-13
```

Nažalost zbog visoke korelacije četvrtog kvartala i broja prelazaka preko granice, regresija rezultira singularnošću. Odlučujemo u nastavku zato provesti jednostavne regresije pa usporediti R-kvadrat vrijednosti.

Prvo promatramo ovisnost prihoda od prodaje i broja ulazaka stranih turista.

```
par(mfrow = c(5,2))
for (regresor in names(putnici.fundamenti.data.d)[c(3:7)]) {
  model = as.formula(paste("Sales ~", regresor))
  fit = lm(model, putnici.fundamenti.data.d)
  plot(putnici.fundamenti.data.d[[regresor]], putnici.fundamenti.data.d$Sales, xlab = regresor, ylab = "Sales",
  lines(putnici.fundamenti.data.d[[regresor]], fit$fitted.values, col='red')

  plot(fit$residuals)
}
```



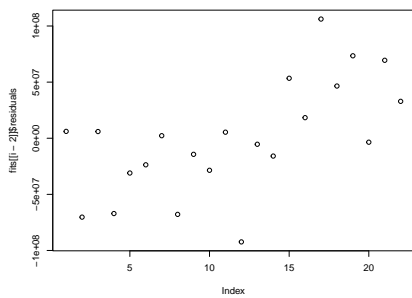
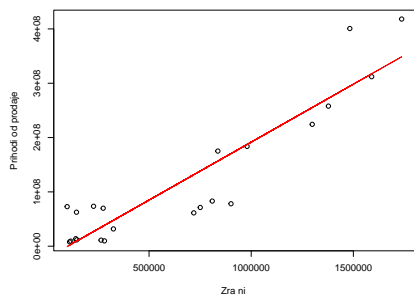
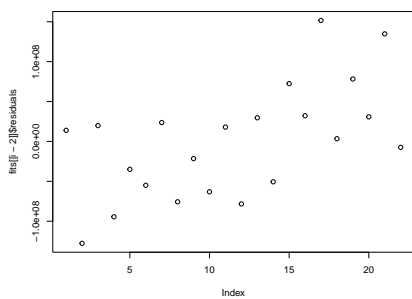
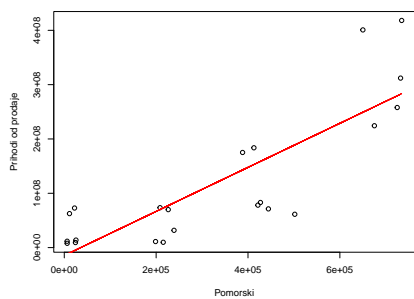
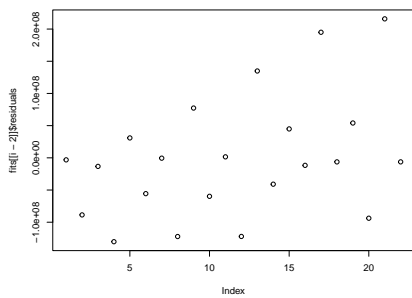
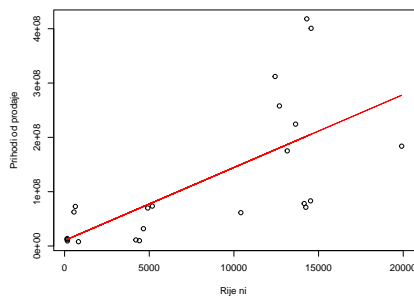
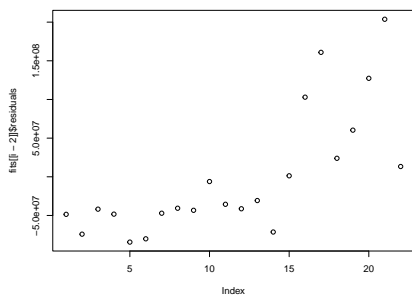
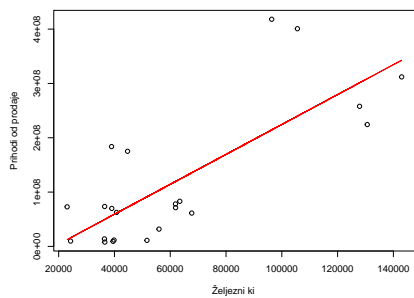
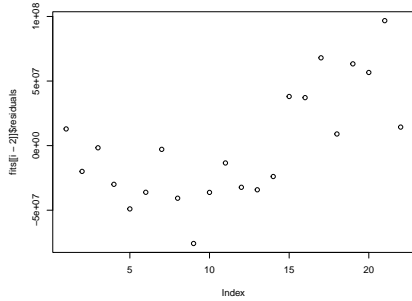
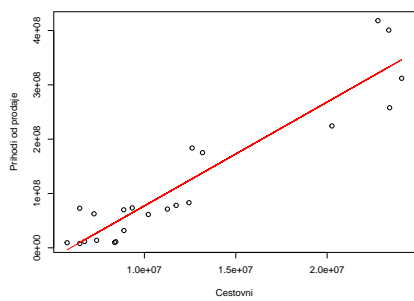
Gornji grafovi reziduala su svakako zabrinjavajući. Uočavamo heterogenost kod nekih i ne normalnost reziduala kod većine.

```
putnici.fundamenti.data.recent = putnici.fundamenti.data.d[putnici.fundamenti.data.d$Year <= 2018,]
putnici.fundamenti.data.recent = putnici.fundamenti.data.recent[-c(3,4), ]

par(mfrow = c(5,2))
fits <- list()

for (i in c(3:7)) {
  regresor = names(putnici.fundamenti.data.recent)[i]
  model = as.formula(paste("Sales ~", regresor))
  fits[[i-2]] <- lm(model, putnici.fundamenti.data.recent)
  plot(putnici.fundamenti.data.recent[[regresor]],
       putnici.fundamenti.data.recent$Sales,
       xlab = regresor, ylab = "Prihodi od prodaje")
  lines(putnici.fundamenti.data.recent[[regresor]], fits[[i-2]]$fitted.values, col='red')

  plot(fits[[i-2]]$residuals)
}
```

```
for (i in c(1:5)) {
  print(names(putnici.fundamenti.data.recent)[i+2])
  print(lillie.test(rstandard(fits[[i]])))
}
```

```
## [1] "Cestovni"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.12185, p-value = 0.5379
##
## [1] "Željeznički"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.25525, p-value = 0.0006408
##
## [1] "Riječni"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.17457, p-value = 0.07965
##
## [1] "Pomorski"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.15166, p-value = 0.2086
##
## [1] "Zračni"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.13927, p-value = 0.3242
```

Izbacivanjem pandemijskih godina iz seta podataka, dobivamo normalno distribuirane rezidualne. Ovo je naravno očekivani rezultat s obzirom da znam koliki je utjecaj pandemija imala na turistička poduzeća. Iznimka navedenom su riječnog prometa gdje svakako nultu hipotezu o normalnosti moramo staviti pod upitnik i željezniki promet gdje hipotezu svakako odbacujemo. Ove dvije vrste prometa nisu od prevelikog značaja za hrvatski turizam pa nas ova činjenica ne zabrinjava.

```
for (i in c(1:5)) {
  print(names(putnici.fundamenti.data.recent)[i+2])
  print(summary(fits[[i]]))
}
```

```
## [1] "Cestovni"
```

```
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -75789406 -33716914 -8163130  31457310  96763682
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.140e+08  2.169e+07  -5.255 3.84e-05 ***
## Cestovni      1.911e+01  1.586e+00   12.055 1.25e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45370000 on 20 degrees of freedom
## Multiple R-squared:  0.879, Adjusted R-squared:  0.873
## F-statistic: 145.3 on 1 and 20 DF,  p-value: 1.255e-10
##
## [1] "Željeznički"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -84454733 -47997703 -38107298  21261327  203722271
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.089e+07  3.637e+07  -1.399   0.177
## Željeznički  2.753e+03  5.113e+02   5.385 2.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 83340000 on 20 degrees of freedom
## Multiple R-squared:  0.5918, Adjusted R-squared:  0.5714
## F-statistic: 29 on 1 and 20 DF,  p-value: 2.855e-05
##
## [1] "Riječni"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -130262073 -58733021 -6275245  41409645  215850834
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10221928  34046138   0.300 0.767094
## Riječni      13430      3314   4.052 0.000623 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 96670000 on 20 degrees of freedom
## Multiple R-squared:  0.4508, Adjusted R-squared:  0.4233
## F-statistic: 16.42 on 1 and 20 DF,  p-value: 0.0006233
##
## [1] "Pomorski"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -127660977 -53869630  8676433  30523570 151619246
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.448e+07  2.538e+07  -0.571    0.575
## Pomorski     4.055e+02  6.067e+01   6.684 1.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 72540000 on 20 degrees of freedom
## Multiple R-squared:  0.6908, Adjusted R-squared:  0.6753
## F-statistic: 44.67 on 1 and 20 DF,  p-value: 1.663e-06
##
## [1] "Zračni"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -92366000 -27336263  -636545  29238186 106187210
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.176e+07  1.757e+07  -1.238    0.23
## Zračni       2.134e+02  2.059e+01  10.361 1.74e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51700000 on 20 degrees of freedom
## Multiple R-squared:  0.843, Adjusted R-squared:  0.8351
## F-statistic: 107.4 on 1 and 20 DF,  p-value: 1.737e-09
```

Sve vrste prometa su signifikantne za objašnjavanje prihoda od prodaje. Cestovni promet objašnjava najveći postotak varijabilnosti, a slijede ga zračni i pomorski.

Sljedeće promatramo ovisnost prodaje o kvartalu.

```
par(mfrow = c(4,2))
fits <- list()

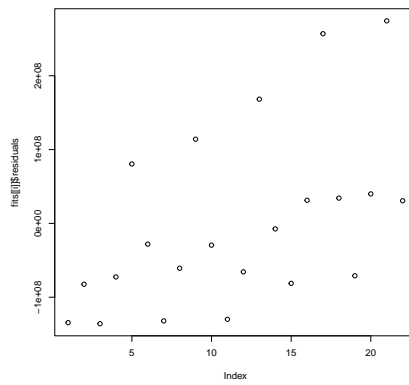
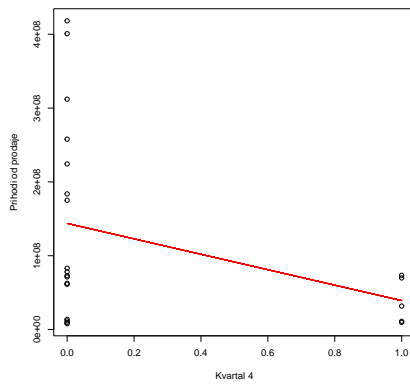
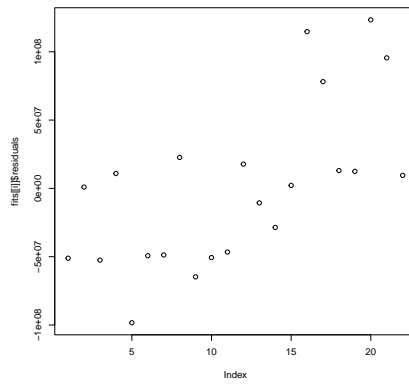
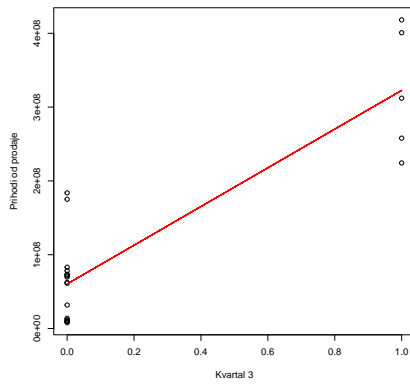
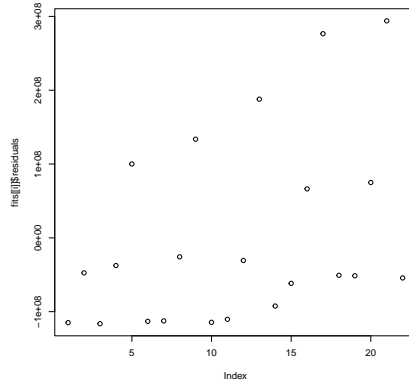
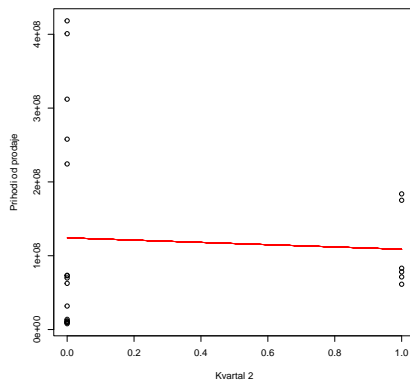
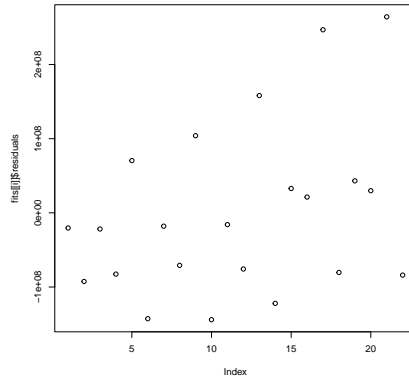
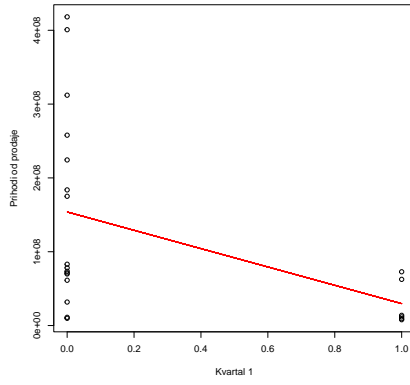
for (i in c(1:4)) {
```

```

regresor = paste("Quarter_", i, sep="")
model = as.formula(paste("Sales ~ Quarter_", i, sep=""))
fits[[i]] <- lm(model, putnici.fundamenti.data.recent)
plot(putnici.fundamenti.data.recent[[regresor]],
     putnici.fundamenti.data.recent$Sales,
     xlab = paste("Kvartal", i),
     ylab = "Prihodi od prodaje")
lines(putnici.fundamenti.data.recent[[regresor]], fits[[i]]$fitted.values, col='red')

plot(fits[[i]]$residuals)
}

```



```
for (i in c(1:4)) {
  print(paste("Kvartal", i))
  print(lillie.test(rstandard(fits[[i]])))
}
```

```
## [1] "Kvartal 1"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.14955, p-value = 0.2258
##
## [1] "Kvartal 2"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.26656, p-value = 0.000272
##
## [1] "Kvartal 3"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.17482, p-value = 0.07874
##
## [1] "Kvartal 4"
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstandard(fits[[i]])
## D = 0.14914, p-value = 0.2293
```

Reziduali su normalno distribuirani za sve kvartale osim drugog pa nastavljamo dalje.

```
for (i in c(1:4)) {
  print(paste("Kvartal", i))
  print(summary(fits[[i]]))
}
```

```
## [1] "Kvartal 1"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -144090872  -82047786  -19186520   40582974  264292108
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  153889297   29210354   5.268 3.72e-05 ***
```

```

## Quarter_1  -124191547  55933570  -2.220  0.0381 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 116800000 on 20 degrees of freedom
## Multiple R-squared:  0.1978, Adjusted R-squared:  0.1576
## F-statistic:  4.93 on 1 and 20 DF,  p-value: 0.03811
##
## [1] "Kvartal 2"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -116313939 -105901852  -49079835   72807369  293954127
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 124227278   32562537   3.815  0.00108 **
## Quarter_2   -15430810   62352512  -0.247  0.80706
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 130300000 on 20 degrees of freedom
## Multiple R-squared:  0.003053, Adjusted R-squared:  -0.04679
## F-statistic: 0.06124 on 1 and 20 DF,  p-value: 0.8071
##
## [1] "Kvartal 3"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -98302816 -49132972   1552877  16568524  123339457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  60420829   14818203   4.077  0.000587 ***
## Quarter_3    262231404   31082925   8.437  5.08e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61100000 on 20 degrees of freedom
## Multiple R-squared:  0.7806, Adjusted R-squared:  0.7697
## F-statistic: 71.17 on 1 and 20 DF,  p-value: 5.081e-08
##
## [1] "Kvartal 4"
##
## Call:
## lm(formula = model, data = putnici.fundamenti.data.recent)
##
## Residuals:

```



```
##           Min           1Q           Median           3Q           Max
## -135864689  -79020168  -28765330   38551575  274403377
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  143778028    29611040   4.856 9.59e-05 ***
## Quarter_4   -104540272    62112642  -1.683   0.108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 122100000 on 20 degrees of freedom
## Multiple R-squared:  0.1241, Adjusted R-squared:  0.08027
## F-statistic: 2.833 on 1 and 20 DF,  p-value: 0.1079
```

Jedino je treći kvartal značajan za objašnjavanje prihoda od prodaje, ali njegova R-kvadrat vrijednost je manja od cestovnog i zračnog prometa.

Na primjer loše ekonomsko stanje u Europi će se odraziti u broju prelazaka, a naravno da neće u indikatorskoj varijabli kvartala.

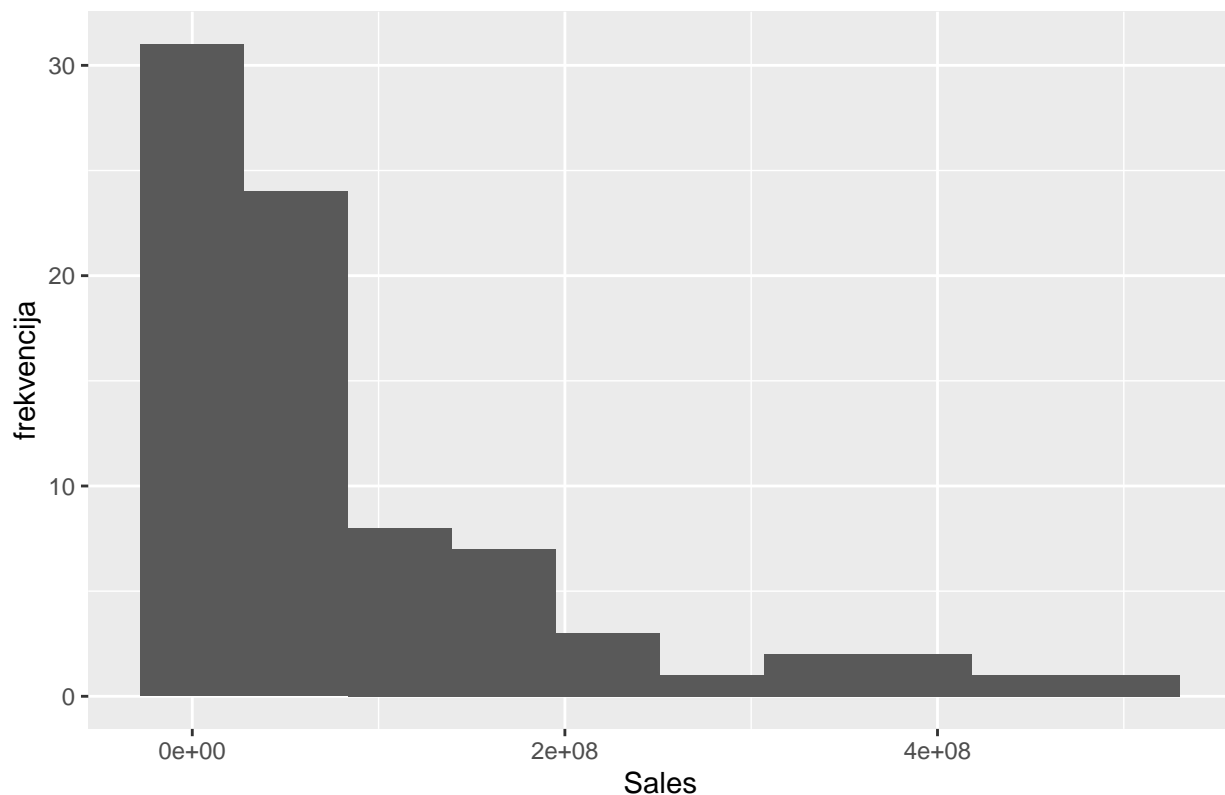
2.4 Distribucija ostvarenih prodaja turističke tvrtke Arena Hospitality Group d.d.

2.4.1 Parametarski pristup

Odgovoriti ćemo na pitanje koja je distribucija ostvarenih prodaja jedne turističke tvrtke od 2002. do 2022. Krenimo s prikazom distribucije prodaja kroz godine.

```
ggplot(ARNTEExtendedDf, aes(x = Sales)) + geom_histogram(bins = 10) +
  labs(title = "Distribucija ostvarenih prodaja", y = "frekvencija")
```

Distribucija ostvarenih prodaja



Vidimo da prodaja ima logaritamsku distribuciju. Ispis prosječnih prodaja:

```
paste0("Prosjecna zarada od ostvarenih prodaja: ", mean(ARNTEextendedDf$Sales), " kn")
```

```
## [1] "Prosjecna zarada od ostvarenih prodaja: 87334981.75 kn"
```

```
filter(ARNTEextendedDf, Quarter == 1) -> q1  
paste0("Prosjecna zarada od ostvarenih prodaja u prvom kvartalu: ", mean(q1$Sales), " kn")
```

```
## [1] "Prosjecna zarada od ostvarenih prodaja u prvom kvartalu: 22363812.25 kn"
```

```
filter(ARNTEextendedDf, Quarter == 2) -> q2  
paste0("Prosjecna zarada od ostvarenih prodaja u drugom kvartalu: ", mean(q2$Sales), " kn")
```

```
## [1] "Prosjecna zarada od ostvarenih prodaja u drugom kvartalu: 74762652 kn"
```

```
filter(ARNTEextendedDf, Quarter == 3) -> q3  
paste0("Prosjecna zarada od ostvarenih prodaja u trecem kvartalu: ", mean(q3$Sales), " kn")
```

```
## [1] "Prosjecna zarada od ostvarenih prodaja u trecem kvartalu: 229697400.6 kn"
```

```
filter(ARNTEextendedDf, Quarter == 4) -> q4  
paste0("Prosjecna zarada od ostvarenih prodaja u cetvrtom kvartalu: ", mean(q4$Sales), " kn")
```

```
## [1] "Prosjecna zarada od ostvarenih prodaja u cetvrtom kvartalu: 22516062.15 kn"
```

Provedimo test normalnosti nad našim podacima:

```
lillie.test(q1$Sales)
```

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: q1$Sales  
## D = 0.39058, p-value = 8.317e-09
```

```
lillie.test(q2$Sales)
```

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: q2$Sales  
## D = 0.24292, p-value = 0.00311
```

```
lillie.test(q3$Sales)
```

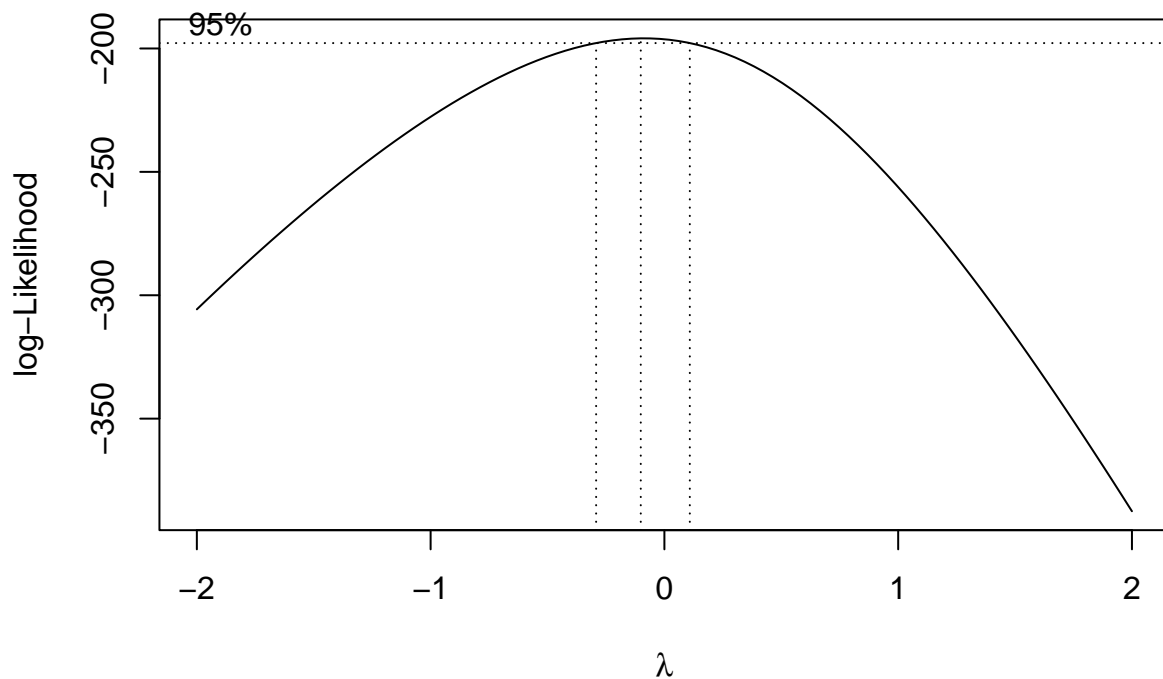
```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: q3$Sales  
## D = 0.18863, p-value = 0.06037
```

```
lillie.test(q4$Sales)
```

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: q4$Sales  
## D = 0.41065, p-value = 8.492e-10
```

Nismo previše zadovoljni s normalnosti podataka. Idemo probati transformirati podatke. Idemo procijeniti parametar λ za Box Cox transformaciju:

```
b <- boxcox(lm(ARNTExtendedDf$Sales ~ 1))
```



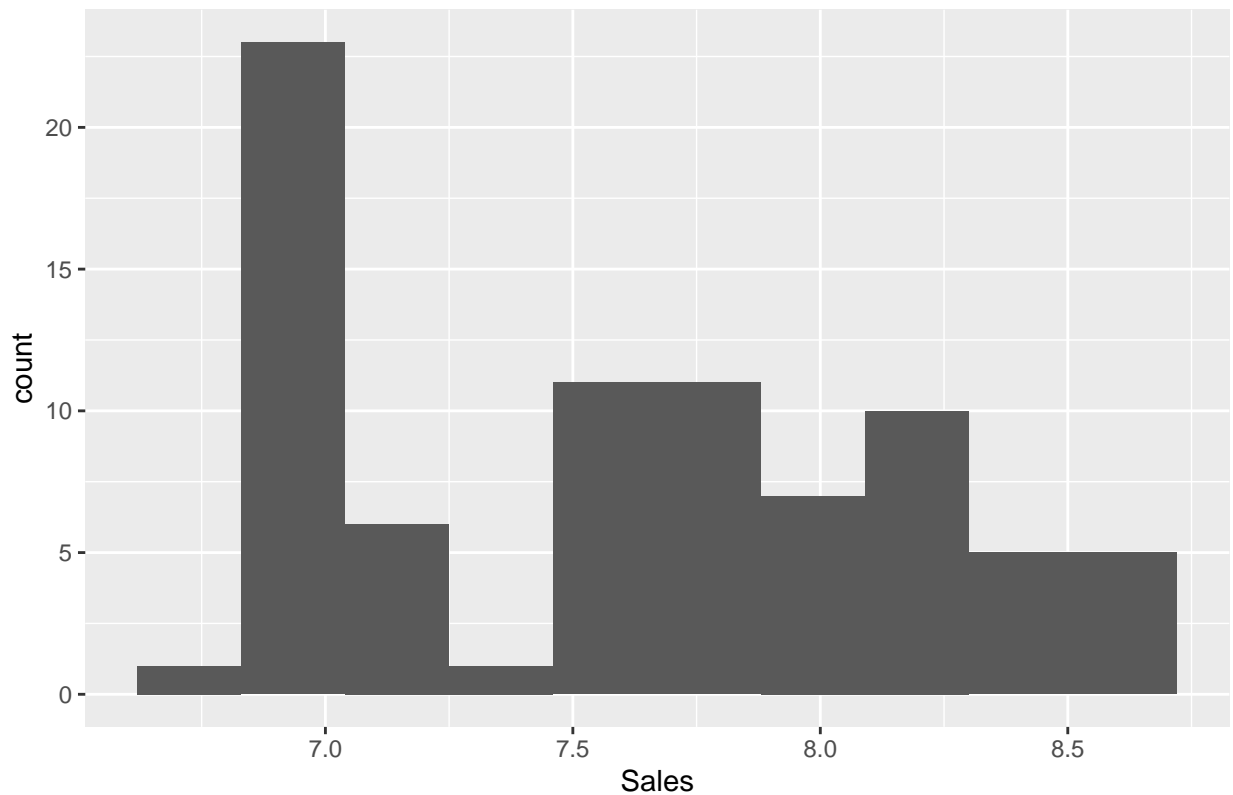
```
# Exact lambda
lambda <- b$x[which.max(b$y)]
lambda
```

```
## [1] -0.1010101
```

Najbolji odabir bi bilo da primjenimo logaritamsku transformaciju nad podacima jer crtež indicira da je 0 unutar 95% intervala pouzdanosti optimalne lambde i zato što je procjena parametra jako bliza 0.

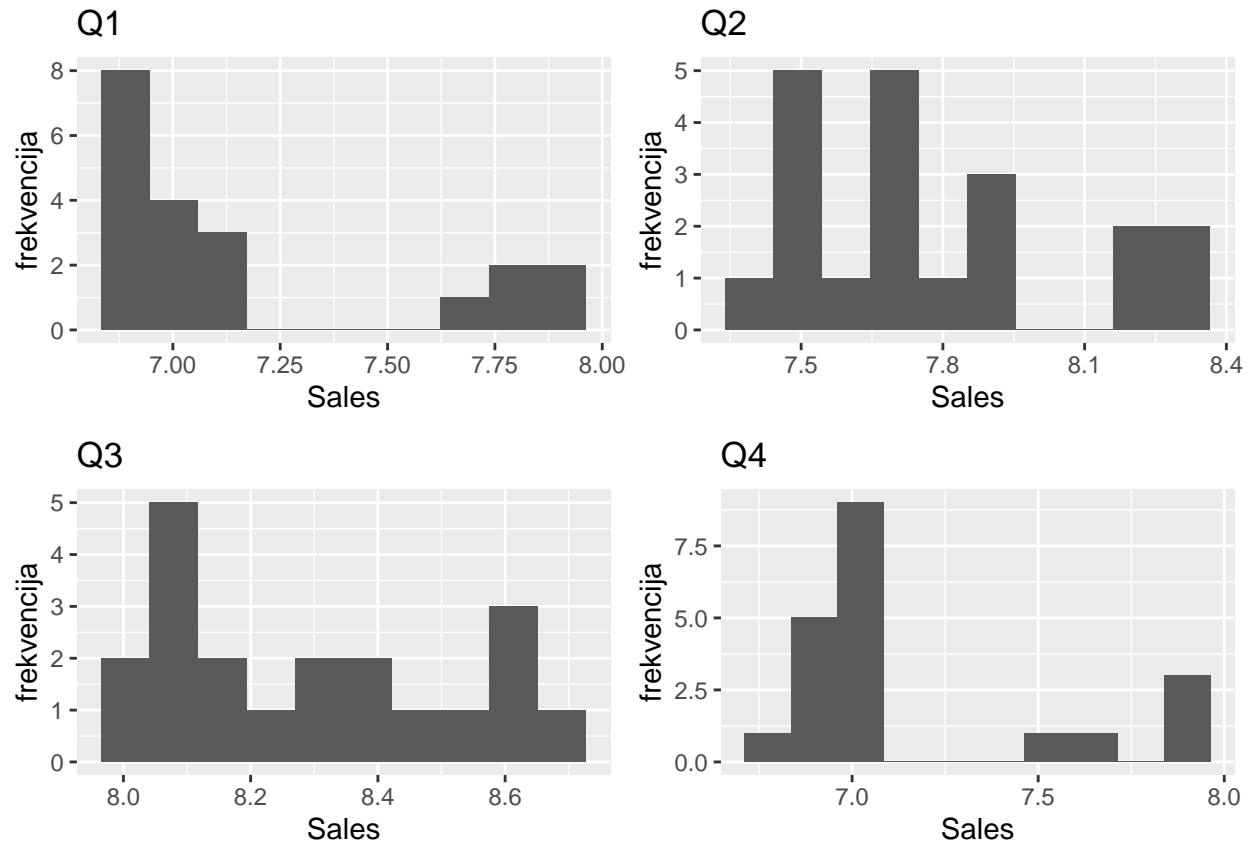
```
transformirano <- ARNTExtendedDf
transformirano$Sales <- log10(ARNTExtendedDf$Sales)
ggplot(transformirano, aes(x=Sales)) + geom_histogram(bins = 10) +
  labs(title = "Transformirani podaci ostvarenih prodaja sa log10")
```

Transformirani podaci ostvarenih prodaja sa log10



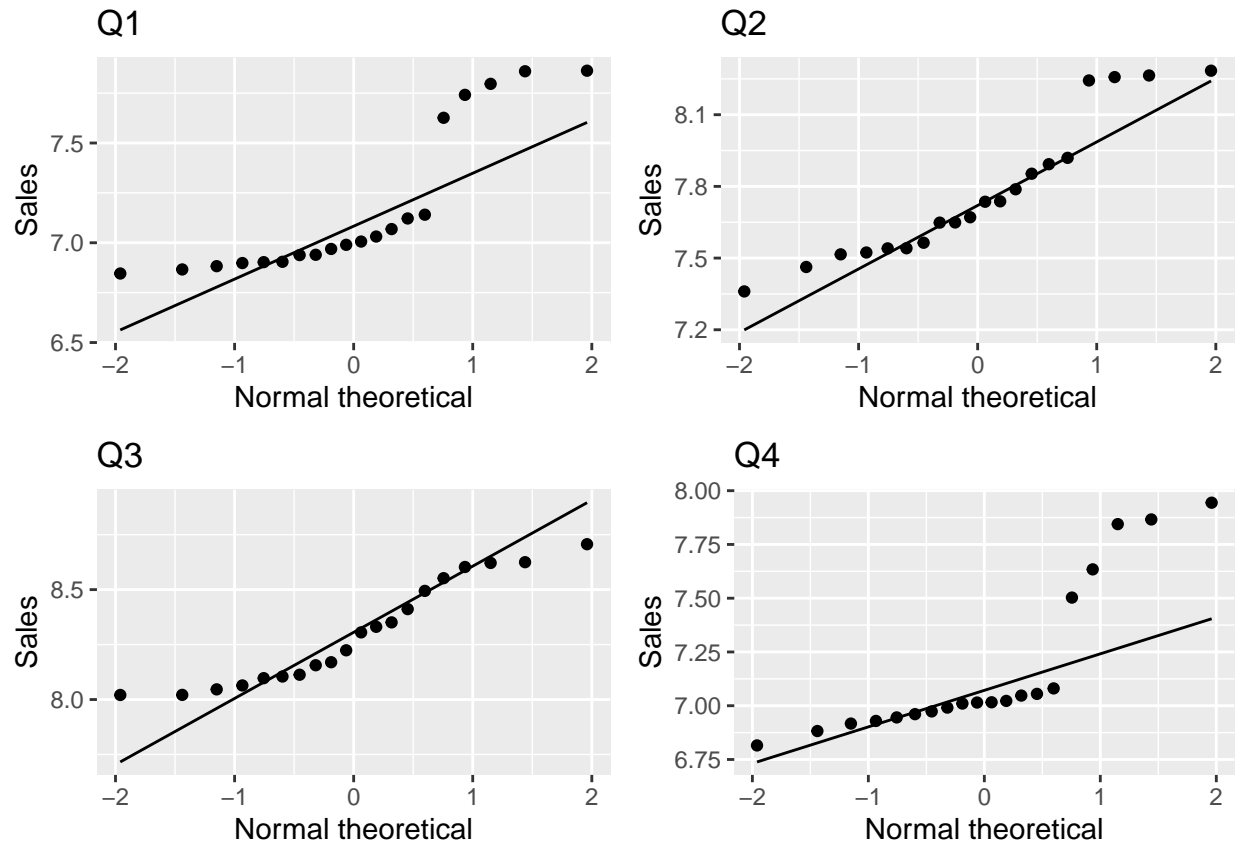
Prikažimo sada distribuciju prodaja po kvartalima:

```
filter(transformirano, Quarter == 1) %>% ggplot(aes(x = Sales)) + geom_histogram(bins = 10) +  
  labs(title = "Q1", y = "frekvencija")-> f1  
filter(transformirano, Quarter == 2) %>% ggplot(aes(x = Sales)) + geom_histogram(bins = 10) +  
  labs(title = "Q2", y = "frekvencija")-> f2  
filter(transformirano, Quarter == 3) %>% ggplot(aes(x = Sales)) + geom_histogram(bins = 10) +  
  labs(title = "Q3", y = "frekvencija")-> f3  
filter(transformirano, Quarter == 4) %>% ggplot(aes(x = Sales)) + geom_histogram(bins = 10) +  
  labs(title = "Q4", y = "frekvencija")-> f4  
grid.arrange(f1,f2,f3,f4, nrow = 2, ncol = 2)
```



Prikaz odstupanja distribucija plaća po kvartalima od normalne razdiobe:

```
qq1 <- filter(transformirano, Quarter == 1)
qq2 <- filter(transformirano, Quarter == 2)
qq3 <- filter(transformirano, Quarter == 3)
qq4 <- filter(transformirano, Quarter == 4)
ggplot(qq1,aes(sample = Sales)) + geom_qq() + geom_qq_line() +
  labs(x= "Normal theoretical", y = "Sales") + labs(title = "Q1") -> qqGraf1
ggplot(qq2,aes(sample = Sales)) + geom_qq() + geom_qq_line() +
  labs(x= "Normal theoretical", y = "Sales") + labs(title = "Q2") -> qqGraf2
ggplot(qq3,aes(sample = Sales)) + geom_qq() + geom_qq_line() +
  labs(x= "Normal theoretical", y = "Sales") + labs(title = "Q3") -> qqGraf3
ggplot(qq4,aes(sample = Sales)) + geom_qq() + geom_qq_line() +
  labs(x= "Normal theoretical", y = "Sales") + labs(title = "Q4") -> qqGraf4
grid.arrange(qqGraf1,qqGraf2,qqGraf3,qqGraf4, nrow = 2, ncol = 2)
```



Provedimo sada Lilliforsov test:

```
lillie.test(qq1$Sales)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  qq1$Sales
## D = 0.28112, p-value = 0.0002192
```

```
lillie.test(qq2$Sales)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  qq2$Sales
## D = 0.14832, p-value = 0.2965
```

```
lillie.test(qq3$Sales)
```

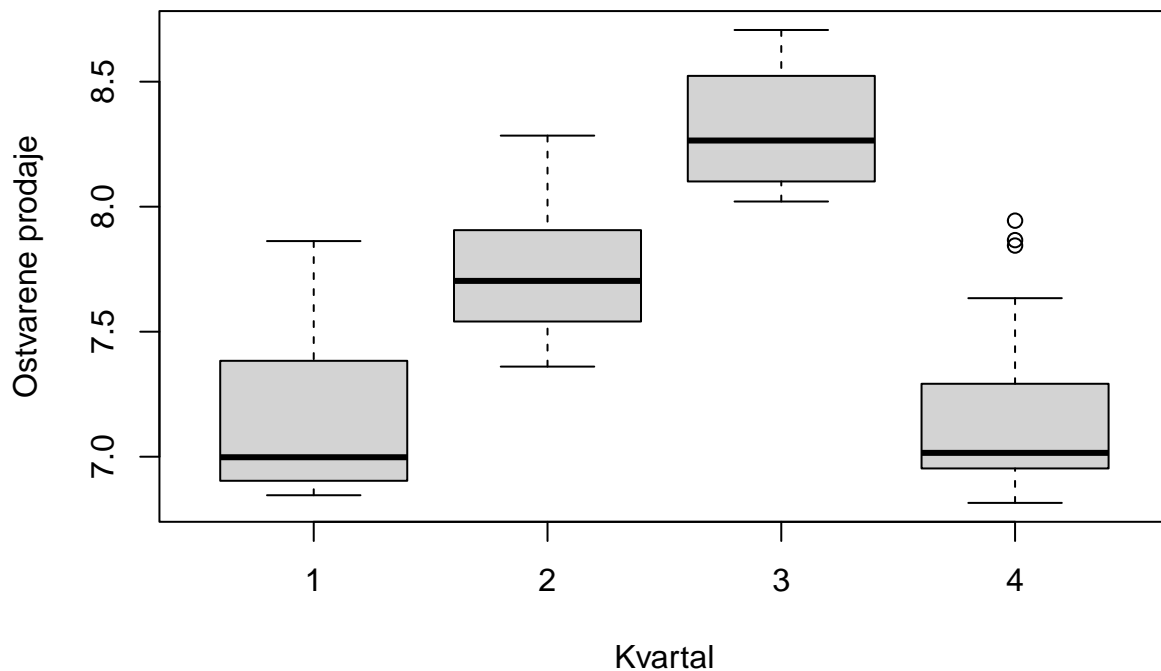
```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  qq3$Sales
## D = 0.16462, p-value = 0.1664
```

```
lillie.test(qq4$Sales)
```

```
##  
##  Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  qq4$Sales  
## D = 0.35043, p-value = 5.427e-07
```

Ovo su sada malo bolji podaci nego prije tranformacije. Prvi i četvrti kvartal nisu prošli test normalnosti, ali ćemo svejedno probati primjeniti ANOVU.

```
boxplot(transformirano$Sales~transformirano$Quarter, xlab = "Kvartal", ylab = "Ostvarene prodaje")
```



Iz grafa možemo naslutiti da varijance iz trećeg i drugog kvartala odskaku u odnosu na druge varijance.

ANOVA pretpostavlja homoskedastičnost odnosno homogenost varijanci svih grupa. Kako bi to provjerili provodimo Bartlettov test:

```
bartlett.test(transformirano$Sales~transformirano$Quarter)
```

```
##  
##  Bartlett test of homogeneity of variances  
##  
## data:  transformirano$Sales by transformirano$Quarter  
## Bartlett's K-squared = 5.0608, df = 3, p-value = 0.1674
```



```
var(qq1$Sales)
```

```
## [1] 0.1377362
```

```
var(qq2$Sales)
```

```
## [1] 0.08377861
```

```
var(qq3$Sales)
```

```
## [1] 0.05351731
```

```
var(qq4$Sales)
```

```
## [1] 0.1311659
```

Na razini značajnosti 0.05 zaključujemo da su vraijance homogene. Provodimo ANOVU za ove dvije hipoteze:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1 : \text{barem dvije sredine se razlikuju}$$

```
anova_results <- aov(Sales~Quarter, data = transformirano)
summary(anova_results)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Quarter      3 17.777    5.926   58.35 <2e-16 ***
## Residuals    76  7.718    0.102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Zaključujemo da se sigurno neke dvije sredine razlikuju na razini značajnosti 0.05. Probajmo izbaciti treći kvartal te ponovo provesti ANOVU:

```
transformirano_no3 <- filter(transformirano, Quarter != 3)
anova_results <- aov(Sales~Quarter, data = transformirano_no3)
summary(anova_results)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Quarter      2  4.824    2.4119   20.52 1.94e-07 ***
## Residuals    57  6.701    0.1176
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I u ovome slučaju zaključujemo da se neke dvije sredine razlikuju na razini značajnosti 0.05. Izbaciti ćemo i treći kvartal te provesti ANOVU, odnosno mogli bismo provesti i f-test:

```
transformirano_no23 <- filter(transformirano_no3, Quarter != 2)
anova_results <- aov(Sales~Quarter, data = transformirano_no23)
summary(anova_results)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Quarter    1  0.000 0.00009   0.001  0.979
## Residuals  38  5.109 0.13445
```

Zaključujemo da prvi i četvrti kvartal imaju jednake sredine na razini značajnosti 0.05.

2.4.2 Neparametraski pristup

Alternativni pristup ANOVI jest neparametraski Kruskal-Wallisov test. Želimo provesti ovaj test zbog toga što imamo mali broj podataka te podaci se ne podudaraju baš s normalnom razdiobom kako bi mi htjeli. Veličina svakog uzorka je veća od 5 pa je ovaj test primjenjiv.

H_0 : medijani prodaja između kvartala se ne razlikuju

H_1 : barem dva medijana prodaja između kvartala se razlikuju

Moramo svakoj vrijednosti prodaja dodijeliti rang. Ako postoje više istih prodaja dodijeljuje im se prosjek. Provjeravamo imamo li iste vrijednosti.

```
table(ARNTExtendedDf$Sales)
```

```
##
## 6536000 7013443 7349587 7625000 7632000 7913339 7994673 8027364
##      1      1      1      1      1      1      1      1
## 8258000 8494764 8674000 8704024 8821113 9135000 9310313 9407000
##      1      1      1      1      1      1      1      1
## 9764000 9798425 10140000 10234689 10346000 10375586 10527000 10744000
##      1      1      1      1      1      1      1      1
## 11146427 11350774 11712148 12035000 13235841 13822659 22931057 29014000
##      1      1      1      1      1      1      1      1
## 31833727 32773000 33364000 34715000 34760159 36650000 42307723 43056801
##      1      1      1      1      1      1      1      1
## 44496637 44562980 46859518 54414365 54629575 55119556 61366796 62580617
##      1      1      1      1      1      1      1      1
## 69912920 71289588 72383531 72847427 73497281 78153377 83074270 87929736
##      1      1      1      1      1      1      1      1
## 104887000 104992000 111224000 115883000 125065000 127275719 129655000 143217859
##      1      1      1      1      1      1      1      1
## 147810089 167465678 175134494 180888641 183760286 192415297 202042470 214207984
##      1      1      1      1      1      1      1      1
## 224349417 257897922 312060986 356527127 400771435 418181405 421442978 508990943
##      1      1      1      1      1      1      1      1
```

moramo izračunati H – statistiku : $H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$

```

salesSorted <- ARNTExtendedDf[order(ARNTExtendedDf$Sales),]
salesSorted$rang <- 1:nrow(salesSorted)
salesSorted %>% group_by(Quarter) %>% summarise(suma = sum(rang)) -> rangSume

Rsume <- 0
for(i in 1:4){
  Rsume <- Rsume + (rangSume[i,2]^2)/nrow(filter(salesSorted, Quarter == i))
}

H <- (12/(nrow(salesSorted)*(nrow(salesSorted)+1))) * Rsume[1,1] - 3*(nrow(salesSorted)+1)
H

```

```
## [1] 52.50981
```

Ova statistika se ravna po hi-kvadrat statistici sa $k - 1$ stupnjeva slobode, gdje je k broj kvartala.

```
qchisq(.95, 3)
```

```
## [1] 7.814728
```

Vidimo da je $H >$ hi-kvadrat odnosno odbacujemo H_0 u prilog H_1 na razini značajnosti 0.05, te zaključujemo da barem dva medijana prodaja između kvadrata se razlikuju.

2.5 Analiza efikasnosti poduzeća

Za dva poduzeća (Arena Hospitality d.d. i Imperial Riviera d.d.) analizirat ćemo vezu između ukupnog ulaska stranih putnika i prihoda od prodaje. Zatim ćemo koristiti bootstrap kako bi testirali jesu li nagibi pravaca nastali regresijom jednaki. Ovako provedenim testom se pitamo jesu li oba poduzeća jednako efikasna u pretvorbi potencijalnih potrošača u prihode od prodaje.

```

putnici.fundamenti = merge(x = straniUlazPutniciAgg,
                           y = ARNTFundamentiDf[-c(1, 4)],
                           by = c("Year", "Quarter"))
colnames(putnici.fundamenti)[colnames(putnici.fundamenti) == 'Sales'] <- 'SalesARNT'
putnici.fundamenti = merge(x = putnici.fundamenti,
                           y = HIMRFundamentiDf[-c(1, 4)],
                           by = c("Year", "Quarter"))
colnames(putnici.fundamenti)[colnames(putnici.fundamenti) == 'Sales'] <- 'SalesHIMR'
set.seed(0)
bldif_function <- function(data, indices) {
  d <- data[indices,]

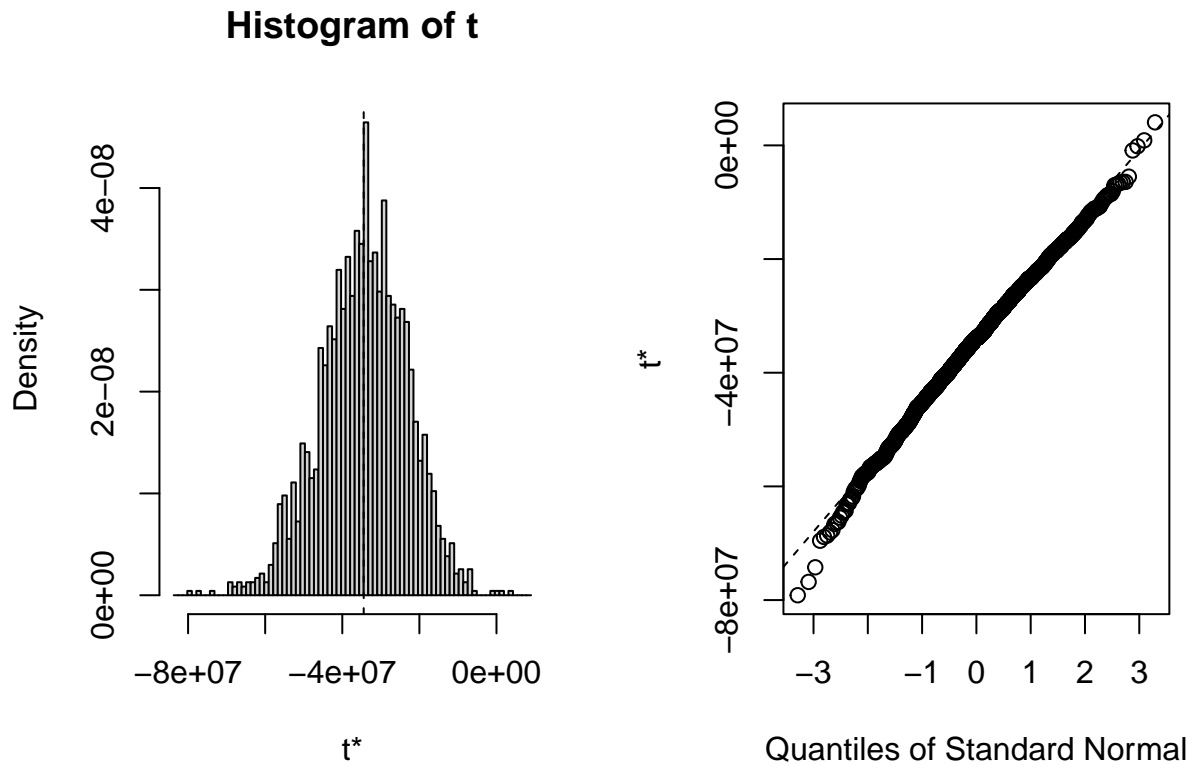
  fit1 <- lm('SalesARNT ~ UKUPNO', data=d)
  fit2 <- lm('SalesHIMR ~ UKUPNO', data=d)

  return(fit1$coefficients[[1]] - fit2$coefficients[[1]])
}

reps <- boot(data=putnici.fundamenti, statistic=bldif_function, R=2000)

plot(reps)

```



Iz histograma i QQ plota vidimo da su uzorci normalno distribuirani pa nastavljamo dalje.

```
boot.ci(reps, type="bca")
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = reps, type = "bca")
##
## Intervals :
## Level      BCa
## 95%      (-59694093, -15745359 )
## Calculations and Intervals on Original Scale
```

Vidimo da nula nije unutar 95% intervala pouzdanosti pa odbacujemo nultu hipotezu o jednakoj efikasnosti poduzeća.

2.6 Analiza dnevnih kretanja cijene indeksa CROBEXturist

```
CROBEXCijeneDf$Date <- as.Date(CROBEXCijeneDf$Date, format = "%Y-%m-%d")
kable(head(CROBEXCijeneDf))
```

Date	Price
2022-11-04	3263.14
2022-11-03	3240.61
2022-11-02	3346.50
2022-10-31	3366.92
2022-10-28	3297.90
2022-10-27	3323.06

```
summary(CROBEXCijeneDf)
```

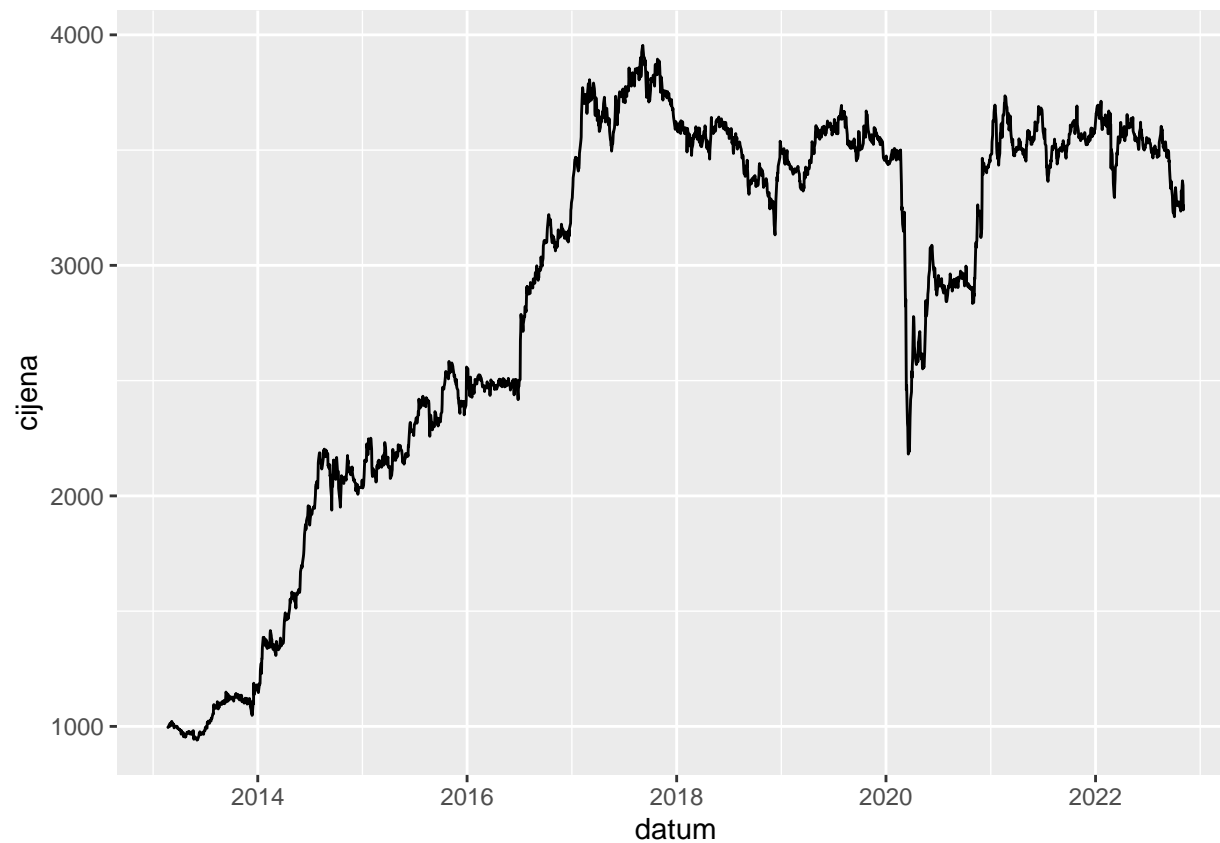
```
##           Date           Price
## Min.      :2013-02-21  Min.    : 940.3
## 1st Qu.:2015-07-29   1st Qu.:2318.1
## Median :2017-12-24   Median :3321.4
## Mean    :2017-12-29   Mean    :2883.3
## 3rd Qu.:2020-06-04   3rd Qu.:3558.0
## Max.    :2022-11-04   Max.    :3954.5
##                                     NA's   :1
```

Možemo vidjeti da je samo jedna cijena NA (not available) pa je možemo maknuti iz skupa podataka. Ne možemo na mjesto NA staviti srednju vrijednost ili medijan jer bi se onda moglo dogoditi da po našim podacima CROBEXturist 2013. godine vrijedi 2883.3 ili 3321.4 što bi bilo tri puta više nego što je tad vrijedio.

```
CROBEXCijeneDf <- na.omit(CROBEXCijeneDf)
```

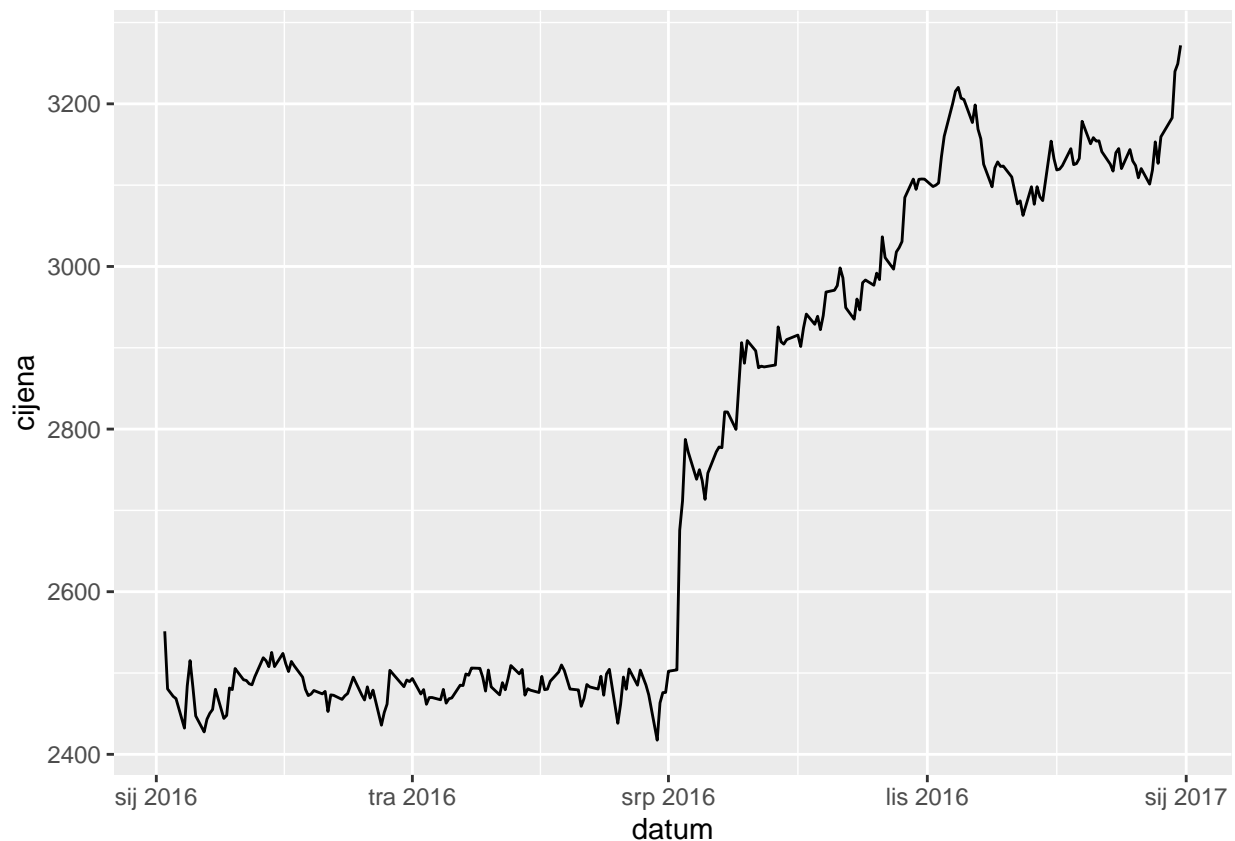
Imamo podatke o cijeni indeksa CROBEXturist iz 2416 dana od 21.2.2013. do 4.1.2022. Idemo vidjeti graf.

```
ggplot(data=CROBEXCijeneDf, aes(x=Date, y=Price) ) + geom_line() + labs(x = 'datum', y = 'cijena')
```



Prikazat ćemo i jednu godinu (npr. 2016.).

```
ggplot(data=subset(CROBEXCijeneDf, year(Date) == 2016), aes(x=Date, y=Price) ) + geom_line() + labs(x =
```



Želimo vidjeti kakav utjecaj na dnevna kretanja indeksa CROBEXtulist ima ulazak i izlazak turista na hrvatskim granicama. Idemo vidjeti primjer tablice prijelaza granice.

```
kable(head(domaciIzlazPutniciDf))
```

X	Cestovni	Željeznički	Riječni	Pomorski	Zračni	UKUPNO
2013-01-01	23581	334	0	27	590	24532
2013-01-02	46762	687	0	26	1291	48766
2013-01-03	52199	694	0	63	973	53929
2013-01-04	55030	604	0	65	987	56686
2013-01-05	81207	636	0	1	1005	82849
2013-01-06	51806	664	0	119	1464	54053

```
summary(domaciIzlazPutniciDf)
```

```
##      X      Cestovni      Željeznički      Riječni
## Length:3595   Min.    :   271   Min.    :   0.0   Min.    :   0.000
## Class :character 1st Qu.: 45583   1st Qu.:  65.0   1st Qu.:   0.000
## Mode  :character Median : 56032   Median : 130.0   Median :   0.000
##              Mean  : 56097   Mean  : 167.0   Mean   :   1.361
##              3rd Qu.: 66553   3rd Qu.: 240.5   3rd Qu.:   0.000
##              Max.   :1394692   Max.   :1024.0   Max.   :3396.000
##      Pomorski      Zračni      UKUPNO
## Min.    :   0.00   Min.    :    0   Min.    :  285
```

```
## 1st Qu.: 8.00 1st Qu.: 816 1st Qu.: 47166
## Median : 39.00 Median : 1235 Median : 57695
## Mean : 51.74 Mean : 2465 Mean : 58781
## 3rd Qu.: 74.00 3rd Qu.: 1859 3rd Qu.: 68468
## Max. :3493.00 Max. :4000595 Max. :4072252
```

```
dim(domaciIzlazPutniciDf)
```

```
## [1] 3595 7
```

```
dim(CROBEXCijeneDf)
```

```
## [1] 2415 2
```

Vidimo da postoji 3595 redova podataka o prijelazu turista, a 2415 o cijeni indeksa. Prvi način koji nam pada na pamet za rješavanje tog problema je izbacivanje redova o prijelazu turista za datume u kojima nemamo cijenu indeksa. Time bi pogriješili jer bi linearna regresija u sljedećem primjeru bila kriva: recimo da 1.1.2015. u Hrvatsku uđe 1 000 000 turista i izađe 1 000, no za taj dan nemamo cijenu indeksa pa ovaj red izbacimo, a 2.1.2015. uđe 1 000 turista, a izađe 1 000 000- u takvom bi slučaju mogli zaključiti da cijene turističkog indeksa rastu izlaskom turista iz države što nema nikakvog smisla. Zbog toga ćemo koristiti linearnu interpolaciju- metodu aproksimacije podataka koje nemamo (npr. ako je 1.1.2016. cijena indeksa bila 1000, a 3.1.2016. cijena 1200, onda ćemo dodati podatak da je 2.1. cijena bila 1100). Interpolacija ne mora nužno dati točne podatke, ali s obzirom na volatilnost ovog indeksa, vjerojatnost je gotovo 0 da CROBEXturist u jednom danu doživi ogroman pad i onda koji dan kasnije ogroman rast nazad na sličnu vrijednost.

```
startDateCROBEX <- min(CROBEXCijeneDf$Date)
endDateCROBEX <- max(CROBEXCijeneDf$Date)

CROBEXallDates <- CROBEXCijeneDf %>%
  complete(Date = seq(from = startDateCROBEX, to = endDateCROBEX, by = "1 day"), fill = list(Price = NA))
```

Sad ćemo nad ovim podacima izvesti linearnu interpolaciju podataka.

```
CROBEXallDates$Price <- na.approx(CROBEXallDates$Price)

dim(CROBEXallDates)
```

```
## [1] 3544 2
```

```
kable(head(CROBEXallDates))
```

Date	Price
2013-02-21	1000.0000
2013-02-22	995.1600
2013-02-23	997.3267
2013-02-24	999.4933
2013-02-25	1001.6600

Date	Price
2013-02-26	1006.3400

Dimenzija CROBEXallDates je manja od dimenzije domaciIzlazPutniciDf.

```
startDateCROBEX
```

```
## [1] "2013-02-21"
```

```
endDateCROBEX
```

```
## [1] "2022-11-04"
```

```
startDateDomaciIzlaz <- min(domaciIzlazPutniciDf$X)
endDateDomaciIzlaz <- max(domaciIzlazPutniciDf$X)
startDateDomaciIzlaz
```

```
## [1] "2013-01-01"
```

```
endDateDomaciIzlaz
```

```
## [1] "2022-11-04"
```

Iz svake tablice podataka prijelaza granica ćemo morati izbaciti podatke od 1.1.2013. do 20.2.2013.

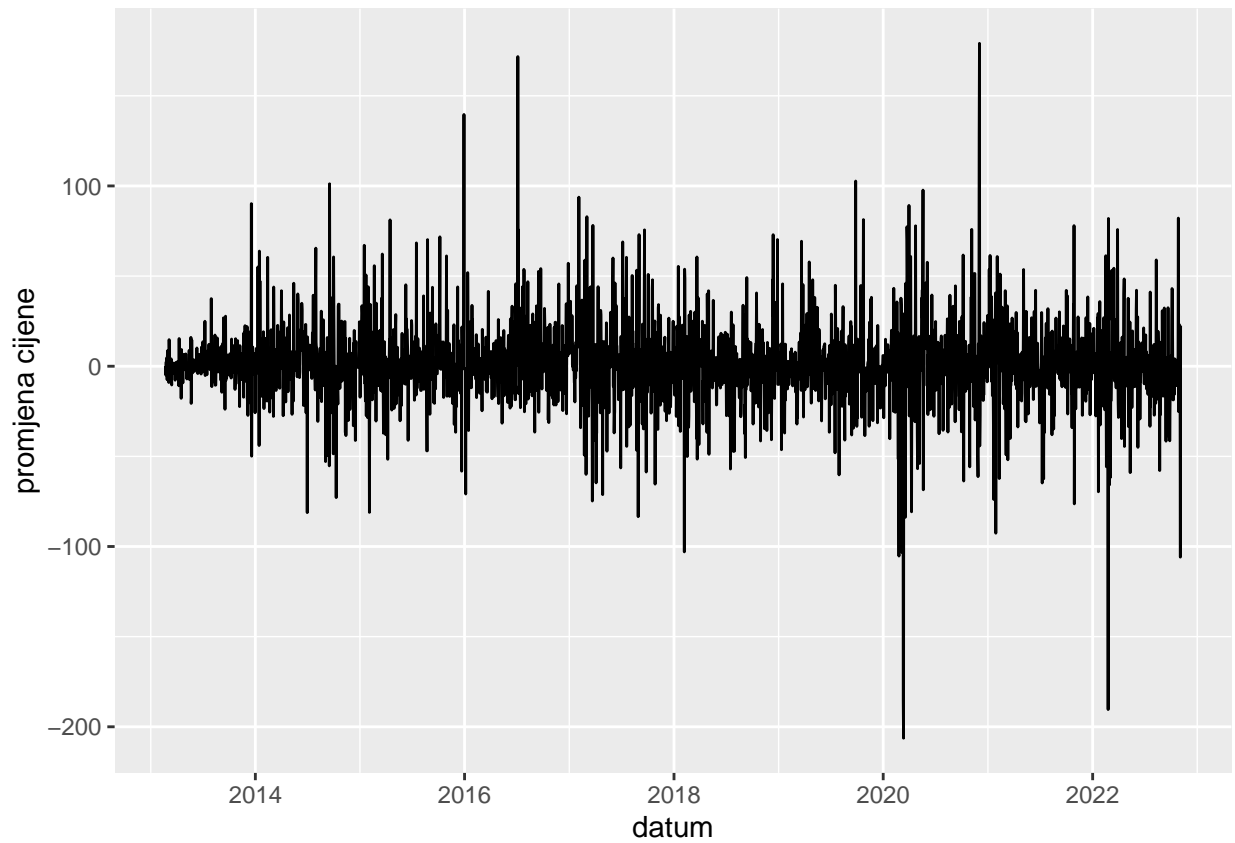
```
domaciIzlazPutniciRemoved<- slice(domaciIzlazPutniciDf, 52:nrow(domaciIzlazPutniciDf))
min(domaciIzlazPutniciRemoved$X)
```

```
## [1] "2013-02-21"
```

```
domaciUlazPutniciRemoved<- slice(domaciUlazPutniciDf, 52:nrow(domaciUlazPutniciDf))
straniIzlazPutniciRemoved<- slice(straniIzlazPutniciDf, 52:nrow(straniIzlazPutniciDf))
straniUlazPutniciRemoved<- slice(straniUlazPutniciDf, 52:nrow(straniUlazPutniciDf))
```

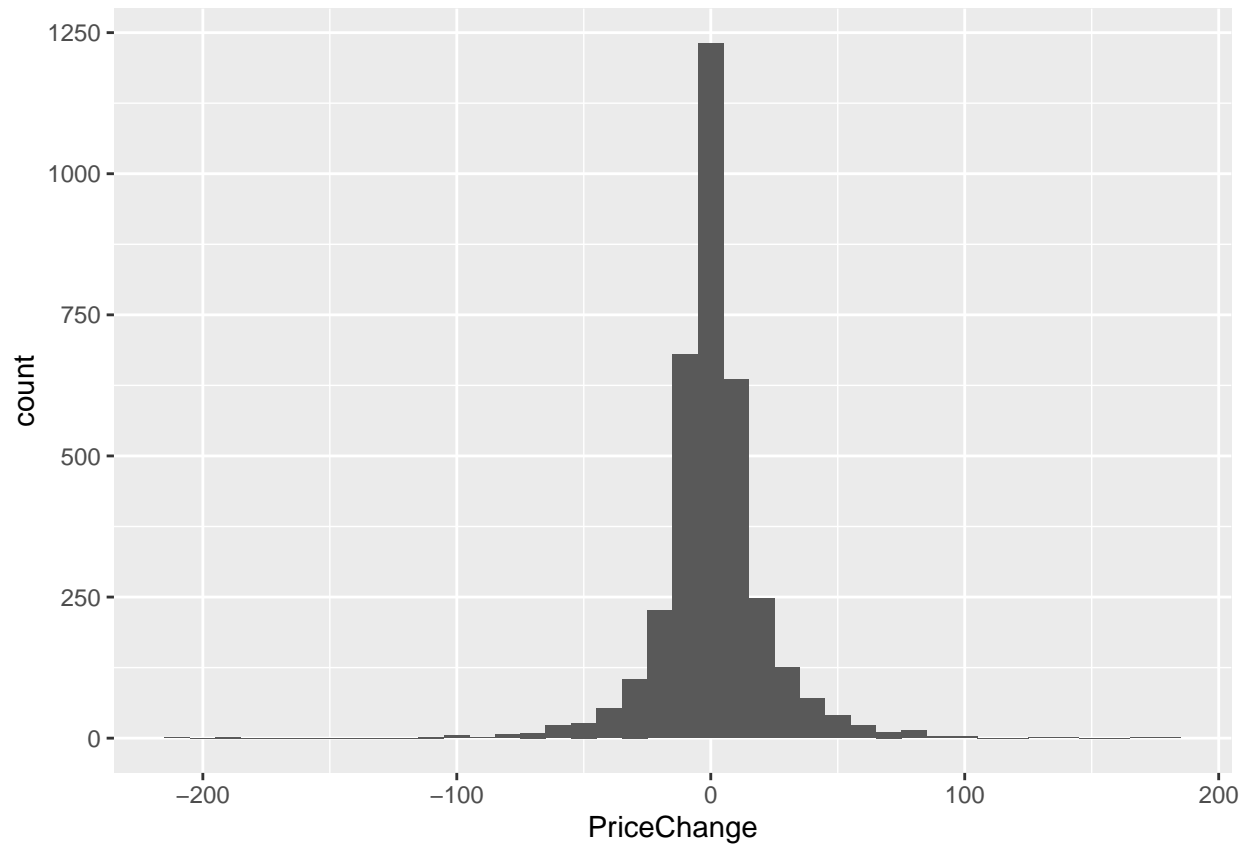
Na cijenu indeksa CROBEXtulist utječu mnogi faktori kao npr. trenutno stanje na tržištu dionica, inflacija... pa bi analiza cijene indeksa dala krive rezultate jer je npr. 2014. mogla biti približno ista turistička godina kao 2017., ali je 2017. cijena 4 puta veća zbog drugih faktora. Kako bi taj problem izbjegli, dodat ćemo za regresora cijenu dionice prijašnjeg dana. Idemo najprije vidjeti kako se kreće promjena cijene indeksa.

```
ggplot(data=CROBEXpriceChange, aes(x=Date, y=PriceChange) ) + geom_line() + labs(x = 'datum', y = 'promjena')
```



Pogledat ćemo histogram promjene cijene indeksa kroz sve godine.

```
ggplot(data = CROBEXpriceChange, aes(x = PriceChange)) + geom_histogram(binwidth = 10)
```

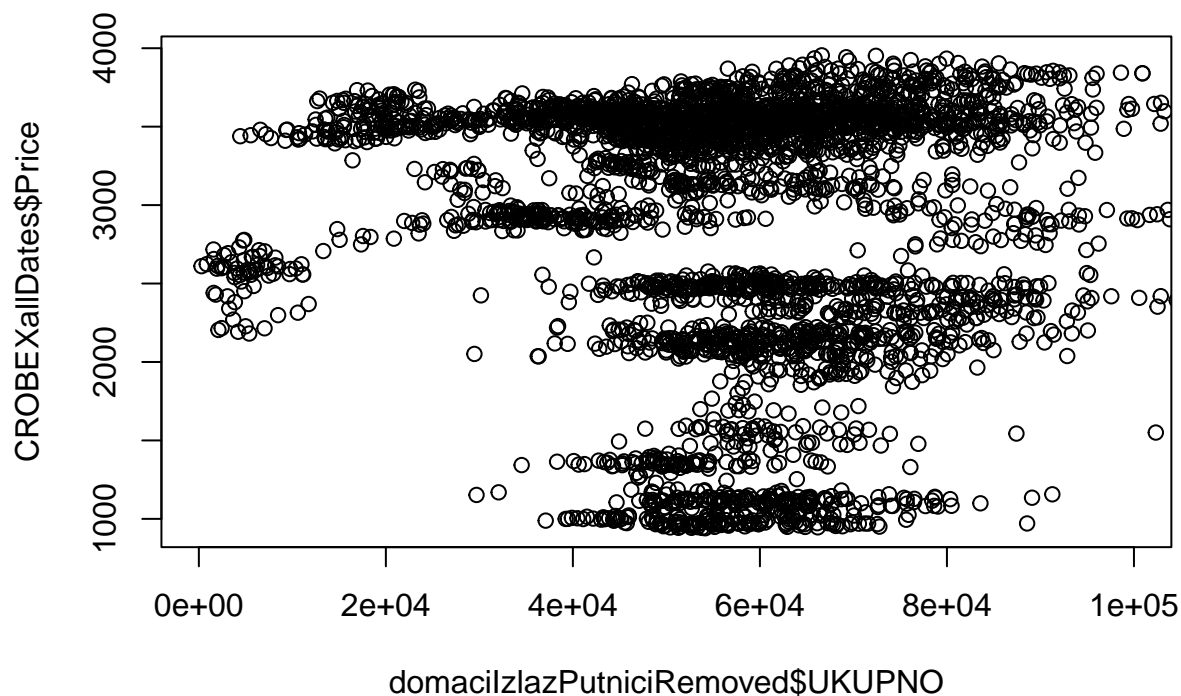


Najprije ćemo napraviti podatke za jedan od regresora- cijenu prijašnjeg dana.

```
CROBEXallDates$priceBefore<- c(NA,CROBEXallDates$Price[1:length(CROBEXallDates$Price)-1])
```

Idemo pogledati dnevnu promjenu cijene s obzirom na kretanja turista.

```
g1 <- plot(domaciIzlazPutniciRemoved$UKUPNO, CROBEXallDates$Price, xlim = c(0, 100000))
```



```
#g2 <- plot(domaciUlazPutniciRemoved$UKUPNO, CROBEXallDates$Price, xlim = c(0, 100000))
#g3 <- plot(straniIzlazPutniciRemoved$UKUPNO, CROBEXallDates$Price, xlim = c(0, 100000))
#g4 <- plot(straniUlazPutniciRemoved$UKUPNO, CROBEXallDates$Price, xlim = c(0, 100000))
```

U grafu ne vidimo linearnost i čini se da ove varijable same po sebi nemaju utjecaj na cijenu CROBEXturist indeksa. Isti zaključak bi izvukli iz ostala tri grafa kretanja ljudi. Spojit ćemo varijable u istu tablicu.

Provest ćemo linearnu regresiju.

```
summary(lm_domaciIzlaz)
```

```
##
## Call:
## lm(formula = Price ~ domaciIzlazPutniciUkupno, data = CROBEXallData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1944.7  -558.1   448.3   672.2  1078.7
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.921e+03  1.849e+01  157.941  < 2e-16 ***
## domaciIzlazPutniciUkupno -6.624e-04  1.990e-04  -3.328  0.000884 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 854.2 on 3542 degrees of freedom
## Multiple R-squared:  0.003117,    Adjusted R-squared:  0.002836
## F-statistic: 11.08 on 1 and 3542 DF,  p-value: 0.0008839
```

```
summary(lm_domaciUlaz)
```

```
##
## Call:
## lm(formula = Price ~ domaciUlazPutniciUkupno, data = CROBEXallData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1943.6   -560.4    447.6    674.4   1073.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.900e+03  1.776e+01  163.317  <2e-16 ***
## domaciUlazPutniciUkupno -3.146e-04  1.789e-04  -1.759   0.0787 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 855.2 on 3542 degrees of freedom
## Multiple R-squared:  0.0008724,    Adjusted R-squared:  0.0005903
## F-statistic: 3.093 on 1 and 3542 DF,  p-value: 0.07874
```

```
summary(lm_straniIzlaz)
```

```
##
## Call:
## lm(formula = Price ~ straniIzlazPutniciUkupno, data = CROBEXallData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1992.1   -581.9    450.1    681.9   1066.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.854e+03  2.425e+01  117.71  <2e-16 ***
## straniIzlazPutniciUkupno 2.063e-04  1.463e-04    1.41   0.159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 855.3 on 3542 degrees of freedom
## Multiple R-squared:  0.0005612,    Adjusted R-squared:  0.000279
## F-statistic: 1.989 on 1 and 3542 DF,  p-value: 0.1586
```

```
summary(lm_straniUlaz)
```

```
##
## Call:
## lm(formula = Price ~ straniUlazPutniciUkupno, data = CROBEXallData)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1941.1  -568.2   448.9   677.9  1071.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.879e+03  2.364e+01 121.816  <2e-16 ***
## straniUlazPutniciUkupno 1.879e-05  1.388e-04   0.135   0.892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 855.5 on 3542 degrees of freedom
## Multiple R-squared:  5.174e-06, Adjusted R-squared:  -0.0002772
## F-statistic: 0.01833 on 1 and 3542 DF,  p-value: 0.8923
```

Iz ovoga je jasno da ove 4 varijable zasebno nemaju (ili imaju neprimjetljivo malen utjecaj) na kretanje CROBEX indeksa. Probat ćemo provesti višestruku regresiju da vidimo hoćemo li dobiti korisne rezultate ukoliko ih spojimo i dodamo varijablu cijene indeksa prijašnjeg dana.

```
lm_visestruka = lm(Price ~ domaciIzlazPutniciUkupno + domaciUlazPutniciUkupno +
                    straniIzlazPutniciUkupno + straniUlazPutniciUkupno , data = CROBEXallData)

summary(lm_visestruka)
```

```
##
## Call:
## lm(formula = Price ~ domaciIzlazPutniciUkupno + domaciUlazPutniciUkupno +
##      straniIzlazPutniciUkupno + straniUlazPutniciUkupno, data = CROBEXallData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2847.1  -564.4   440.6   676.9  1072.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.902e+03  2.641e+01 109.883  < 2e-16 ***
## domaciIzlazPutniciUkupno -7.204e-04  2.034e-04  -3.542 0.000402 ***
## domaciUlazPutniciUkupno  -3.454e-04  1.816e-04  -1.902 0.057194 .
## straniIzlazPutniciUkupno  1.042e-03  3.278e-04   3.179 0.001491 **
## straniUlazPutniciUkupno  -7.167e-04  3.106e-04  -2.307 0.021094 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 852.9 on 3539 degrees of freedom
## Multiple R-squared:  0.006986, Adjusted R-squared:  0.005863
## F-statistic: 6.224 on 4 and 3539 DF,  p-value: 5.478e-05
```

Analizom podataka možemo zaključiti ono što smo uvidom u grafove i samim shvaćanjem kretanja cijena na tržištu kapitala već znali- cijena indeksa ne može se predvidjeti kretanjem turista. Postoje mnogi drugi poznati i nepoznati faktori koji utječu na cijenu CROBEX-a.

3 Zaključak

Analizom veze mobilnosti turista kroz Hrvatsku s uspješnosti hrvatskog turističkog sektora i cijenom dionica turističkih kompanija na Zagrebačkoj burzi došli smo do mnogih zanimljivih zaključaka. Zaključili smo da najviše ljudi prelazi granicu ljeti i većinom petkom, subotom ili nedjeljom. Najviše ljudi dolazi u Hrvatsku cestovnim prometom. Zaključujemo da prihode od prodaje možemo djelomično predvidjeti jednostavno kvartalom u kojem su nastali. Ipak broj stranih ulazaka sadrži više informacija pa je i procjena prihoda bolja. Također smo vidjeli da se cijena indeksa CROBEXturist ne može predvidjeti analizom ulaska i izlaska iz Hrvatske, već ona ovisi i o raznim globalnim varijablama.