

energije

Blaž Povh

Uvod

Motivacija

Od trenutka, ko sem izvedel, da bom moral pri predmetu analiza podatkov s programom R narediti projektno nalogo, sem si zadal cilj to nalogo narediti čim bolj zanimivo in zahtevno ter od sebe oddati dober projekt. Imel sem malo predhodnega znanja s podobnimi računalniškimi programi, zato sem imel željo pridobiti čim več uporabnega znanja tekom te projektne naloge, saj v svetu kakršnem živimo ta znanja postajajo čedalje bolj cenjena ter iskana. Tako sem iskal čimbolj inovativno ter hkrati uporabno temo, ki bi mi lahko še prišla kdaj prav kasneje. Zagrabil sem priložnost na Razvojnem centru Novo mesto, ko so mi predstavili, da iščejo nekoga, ki bi jim pomagal pri projektu. Lotil sem se zanimive teme - analize pretekle porabe in napoved bodoče porabe električne energije za različne porabnike. Tekom dela sem se znašel v veliko napornih trenutkih, ko nisem vedel kaj in kako nadaljevati, vendar me je nekako pripeljala do konca motivacija, da se naučim čim več uporabnih stvari, ki bi mi lahko prišle prav v bodoče in pustiti čim boljši vtis na Razvojnem centru Novo mesto.

Opis problema

V sklopu te naloge sem poskušal analizirati preteklo porabo električne energije in iskati povezave med koledarskimi ter vremenskimi podatki zakaj se poraba razlikuje za različne dneve. Na koncu pa sem želel narediti najpomembnejše/zanimivejše tj. sestaviti preprost napovedni model, ki bi v grobem iz danih podatkov izračunal porabo električne energije za naslednji dan.

Opis podatkov

Kar se tiče podatkov sem imel srečo, saj so mi večino podatkov posredovali iz Razvojnega centra Novo mesto. Glavni vir podatkov predstavlja 50 datotek s podatki o porabi električne energije za različne porabnike (vsaka datoteka za različnega porabnika). Slednje sem prejel v formatu JSON. Vsebujejo pa podatke za prvih 200 dni v letu 2021, kjer beležijo porabo za vsak porabnik na 15 minutni dnevni bazi za vseh 200 dni. Spodaj je mogoče videti kratek izsek surovih podatkov za prvi porabnik.

```
{"date": "2021-01-01", "datetime": "2020-12-31T23:15:00+00:00", "status": "ED0", "values": [1.43, 1.4175, 1.3975
```

Naslednji sklop podatkov predstavlja koledarski podatki za omenjenih 200 dni. Ti podatki vsebujejo informacijo o tem ali je dan dela prost ali ne ter kateri dan v tednu je, prejel pa sem jih v formatu csv. Ti podatki so mi v nadaljevanju zelo uporabni pri analizah. Surovi podatki pa izgledajo nekako takole:

```

["", "date", "dayHours", "dayName", "dayNum", "freeDay", "holidayName", "index", "monthNum", "weekOfMonth", "yearNum"]
["1", "2021-01-01", 24, "Friday", 1, TRUE, "New Year's Day", 0, 1, 1, 2021]
["2", "2021-01-02", 24, "Saturday", 2, TRUE, "New Year's Day", 1, 1, 1, 2021]
["3", "2021-01-03", 24, "Sunday", 3, TRUE, "", 2, 1, 1, 2021]
["4", "2021-01-04", 24, "Monday", 4, FALSE, "", 3, 1, 2, 2021]

```

Zadnji sklop podatkov, ki sem jih prejel iz RCNM, pa so bili še vremenski podatki za sledeče dni. K sreči so vsi porabniki iz Ljubljane, tako da sem opazoval samo vreme v Ljubljani za omenjene dneve in to primerjal s porabo in iskal povezave. Tudi ti podatki sem prejel v formatu JSON. Prilagam kratek izsek surovih podatkov.

```
[{"date": "2021-01-01", "datetime": "2020-12-31T23:00:00+00:00", "humidity": [98, 98, 98, 98, 98, 98, 99, 99, 99, 99]
```

Ko pa sem se lotil vizualizacij, pa sem še potreboval podatke, da sem lahko predstavil podatke grafično. Geografske podatke o Sloveniji in njenih regijah sem pridobil iz **gadm**. V nadaljevanju pa sem potreboval še povprečne letne temperature za različne slovenske kraje. Te podatke pa sem pridobil iz **ARSO**.

Priprava podatkov

Problema sem se lotil sistematično. V prvem sklopu sem se lotil samo podatkov o porabi električne energije za različne porabnike. Podatke za vsah 50 porabnikov sem združil v eno tidy.data tabelo, kjer stolpci predstavljajo meritve od 1 do 96. To namreč predstavljajo izmerjeno porabo električne energije, kjer vsaka naslednja meritev predstavlja porabo čez naslednjih 15 minut. V vrsticah pa se nahajajo porabe električne energije za vseh 50 porabnikov za prvih 200 dni v letu 2021. Zaradi lažje analize sem se odločil odstraniti 28.3.2021, saj se je takrat premaknila ura in nisem imel podatkov za vseh 96 meritev. Ta tabela mi je bila še potem v nadaljevanju projektne naloge uporabljal za temelj vseh analiz.

Naslednja tabela prikazuje vzorec urejenih podatkov

```
data_tidy<-read.table("/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/podatki/zdruzena.csv")
data_tidy[1:5,1:8]
```

##		datum	porabnik	meritve.1	meritve.2	meritve.3	meritve.4	meritve.5
## 1		2021-01-01	1	1.4300	1.4175	1.3975	1.4025	1.3950
## 2		2021-01-02	1	3.2575	1.4075	1.3950	1.3950	1.3950
## 3		2021-01-03	1	1.4450	1.4125	1.4025	1.3975	1.3975
## 4		2021-01-04	1	2.6625	1.4125	1.3950	1.3975	1.4000
## 5		2021-01-05	1	2.0125	2.3675	1.4025	1.3950	1.3950
##		meritve.6						
## 1		1.3950						
## 2		1.3975						
## 3		1.3950						
## 4		1.3975						
## 5		1.3950						

Analiza in vizualizacije

Analiza porabe za različne tipe dni

Ko sem imel to tabelo narejeno sem se lotil prvih analiz. Zanimala me je povprečna poraba za vseh 199 dni. Potem pa sem se lotil podrobnejše analize. S pomočjo uvoza koledarskih podatkov sem pridobil podatek za vsak dan, ali je bil dan dela prost ali ne. Na podlagi tega me je zanimala, za koliko se je razlikovala povprečna poraba za delovne in dela proste dni.

Naslednji prikaz prikazuje kodo s katero sem prišel do povprečne porabe za delovne in dela proste dni.

```
dnevna_poraba <-rowSums(data_tidy[,3:98],is.na(data_tidy) == FALSE)
length(dnevna_poraba)
porabniki <- labels(table(data_tidy$porabnik))[[1]]

povpr_poraba=c()
for (i in 1:length(porabniki)){
```

```

    print(i)
    poraba=dnevna_poraba[data_tidy$porabnik==porabniki[i]]
    povpr_poraba[i]=mean(poraba)
  }
frekvenca <- c(1:51)
data_povpr_poraba <- data.frame(povpr_poraba,frekvenca)

pdf("/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/analiza/prikaz_skupne_povprecne_porabe.pdf")
chol <- data_povpr_poraba
qplot(chol$povpr_poraba,
      geom = "histogram",
      binwidth= 40,
      main = "Povprecna poraba el. energije za razlicne porabnike",
      xlab = "Povprecna poraba",
      ylab = "Frekvenca",
      ylim = c(0,10))
dev.off()

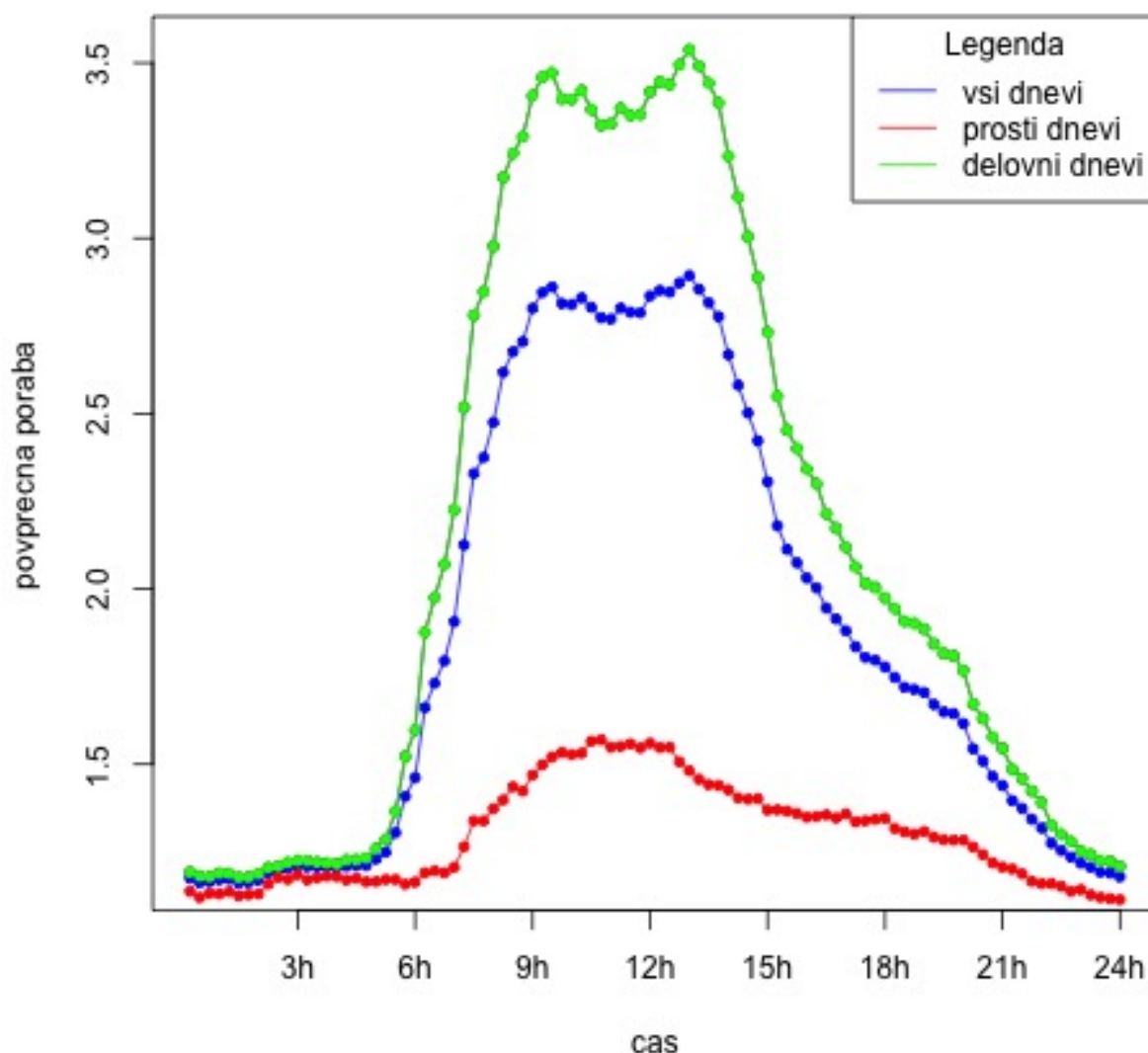
pdf("/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/analiza/prikaz_skupne_povprecne_porabe.pdf")
Tabela_delovnih_dni <- filter(data_tidy, Prosti_dan == "FALSE")
meritev3 <- colMeans(Tabela_delovnih_dni[, 3:98 ],is.na(data_tidy)== FALSE)
p3 <-plot(meritev3, type = "o",pch = 20, xlab="cas",ylab = "povprecna poraba",main = "Graf povprecne porabe")
meritev1 <- colMeans(data_tidy[, 3:98 ],is.na(data_tidy)== FALSE)
p1 <-points(meritev1, type = "o",pch = 20, xlab="cas",ylab = "povprecna poraba",main = "Graf povprecne porabe")
Tabela_prostih_dni <- filter(data_tidy, Prosti_dan == "TRUE")
meritev2 <- colMeans(Tabela_prostih_dni[, 3:98 ],is.na(data_tidy)== FALSE)
p2 <-points(meritev2, type = "o",pch = 20, xlab="cas",ylab = "povprecna poraba",main = "Graf povprecne porabe")
Tabela_delovnih_dni <- filter(data_tidy, Prosti_dan == "FALSE")
meritev3 <- colMeans(Tabela_delovnih_dni[, 3:98 ],is.na(data_tidy)== FALSE)
p3 <-points(meritev3, type = "o",pch = 20, xlab="cas",ylab = "povprecna poraba",main = "Graf povprecne porabe")
xlabels=c("3h","6h","9h","12h","15h","18h","21h","24h")
axis(side=1, at=c(12,24,36,48,60,72,84,96),labels=xlabels)
legend(title = "Legenda", legend = c("vsi dnevi","prosti dnevi","delovni dnevi"),"topright",col = c("black","green","red"))
dev.off()

```

Narisal sem tri grafe, enega za povprečno porabo vseh 199 dni, enega za povprečno porabo delovnih ter enega za dela proste dni. Vse te grafe sem združil v enega, da so razlike vidneše vizualno.

Združen graf pa izgleda takole:

Graf povprečne poraba el. energije za različne tipe dni



Iz grafa lahko razberemo številne (pricakovane) zaključke. Največji porabniki električne energije so pri nas zagotovo tovarne in podjetja, kjer je potrebno veliko elektrike, da vse normalno delujejo. Zato je bilo pričakovati, da bo povprečna poraba mnogo višja za delovne dni kot za dela proste. Ta pričakovanja so se potrdila, saj se vidi občutna razlika na grafu. Ker pa je mnogo več delovnih dni kot dela prostih je tudi povprečna poraba za vseh 199 dosti podobnejša povprečni porabi za delovne dni. Iz grafa pa je moč tudi razbrati za katere ure v dnevu je poraba največja. Iz grafa se lepo vidi da je to nekje med 8 in 18h, ko je večina ljudi v službi, vrh pa se doseže približno ob 13h.

Analiza porabe za različne dni v tednu

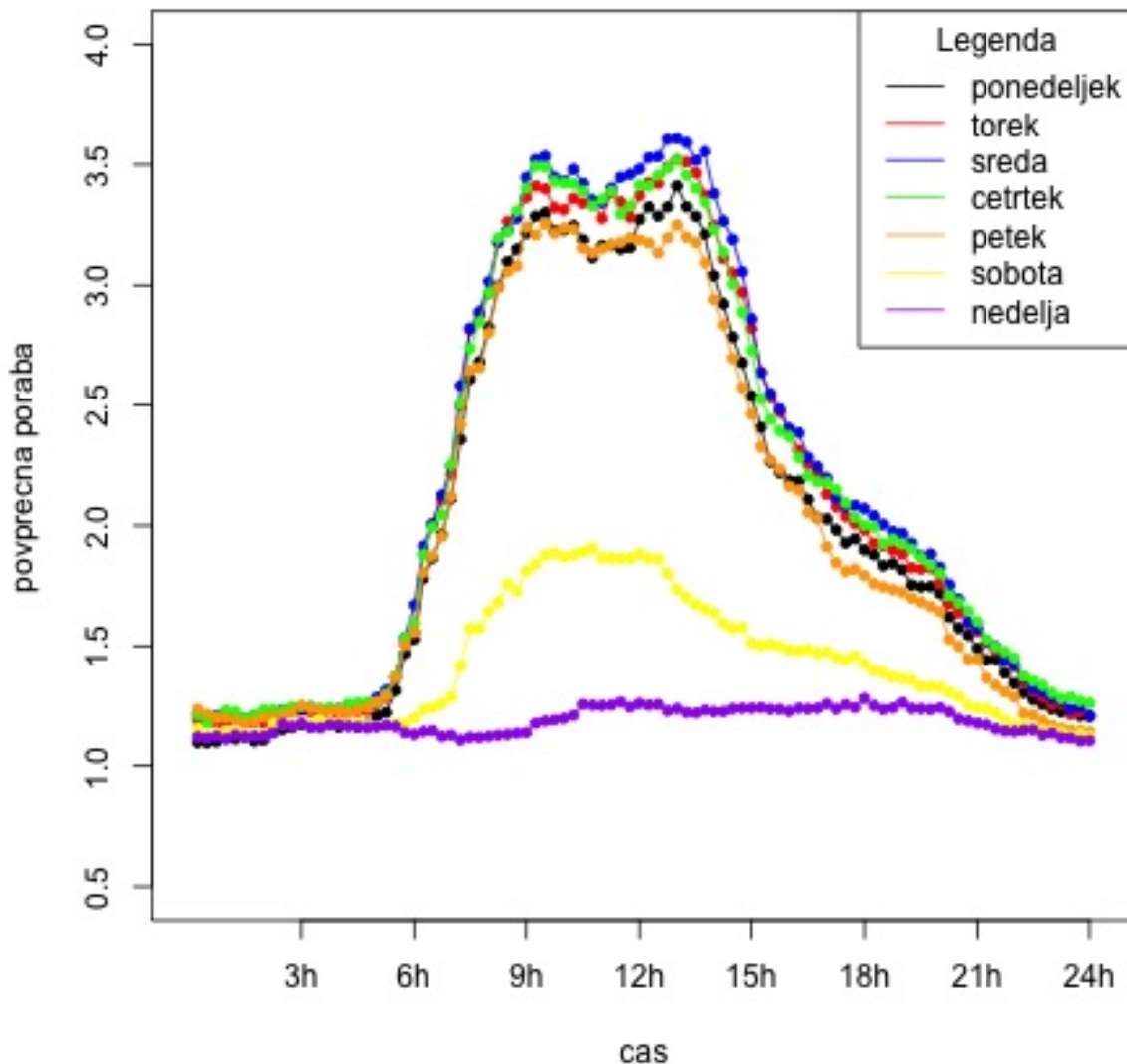
V nadaljevanju sem hotel potrditi prejšnjo analizo, da je ob vikendih poraba občutno manjša kot za prvih 5 dni v tednu. Tako sem izračunal povprečno porabo za vsak dan v tednu in za vsak dan narisal svoj graf in slednje združil v en graf, da so rezultati bolj ilustrativni.

Koda s katero sem si tukaj pomagal je sledeča:

```
pdf("/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/analiza/prikaz_skupne_povprecne")
par(mar=c(5,6,4,1)+.0001)
plot(c(0,97),c(0.5,4),main="Povprecna poraba za dneve v tednu",type = "n",pch=20, xlab = "cas", ylab = "poraba")
povpr_poraba_ponedeljek <- colMeans(filter(data_tidy, Ime_dneva == "Monday")[,3:98])
graf_ponedeljek <- points(povpr_poraba_ponedeljek,type = "o",pch=20,xaxt='n',col="black")
povpr_poraba_torek <- colMeans(filter(data_tidy, Ime_dneva == "Tuesday")[,3:98])
graf_torek <- points(povpr_poraba_torek, type = "o",pch=20,xaxt='n',col="red")
povpr_poraba_sreda <- colMeans(filter(data_tidy, Ime_dneva == "Wednesday")[,3:98])
graf_sreda <- points(povpr_poraba_sreda, type = "o",pch=20,xaxt='n',col="blue")
povpr_poraba_cetrtek <- colMeans(filter(data_tidy, Ime_dneva == "Thursday")[,3:98])
graf_cetrtek <- points(povpr_poraba_cetrtek, type = "o",pch=20,xaxt='n',col="green")
povpr_poraba_petek <- colMeans(filter(data_tidy, Ime_dneva == "Friday")[,3:98])
graf_petek <- points(povpr_poraba_petek, type = "o",pch=20,xaxt='n',col="orange")
povpr_poraba_sobota <- colMeans(filter(data_tidy, Ime_dneva == "Saturday")[,3:98])
graf_sobota <- points(povpr_poraba_sobota, type = "o",pch=20,xaxt='n',col="yellow")
povpr_poraba_nedelja <- colMeans(filter(data_tidy, Ime_dneva == "Sunday")[,3:98])
graf_nedelja <- points(povpr_poraba_nedelja, type = "o",pch=20,xaxt='n',col="purple")
xlabels=c("3h","6h","9h","12h","15h","18h","21h","24h")
axis(side=1, at=c(12,24,36,48,60,72,84,96),labels=xlabels)
legend(title = "Legenda", legend = c("ponedeljek","torek","sreda","cetrtek","petek","sobota","nedelja"),
dev.off())
```

Porabo za vsak dan posebej sem grafično prikazal ter nato združil vseh sedem grafov v enega za lažjo primerjavo.

Povprečna poraba za dneve v tednu



Tudi tukaj nemorem reči, da sem prišel do kakšnih drastičnih novih spoznanj. Kot predvideno je poraba najvišja za dneve od ponedeljka do petka, ko ljudje hodijo v službe ter tovarne obratujejo. V soboto je poraba bistveno manjša kot za prvih 5 dni v tednu a je vseeno nekoliko višja kot v nedeljo. To si razlagam tako, da veliko uspešnih podjetij dela tudi ob sobotah. Hkrati pa so sobote dnevi, ko se še doma nekaj postori. Vsa ta opravila, ki se v nedeljo ponavadi ne delajo porabijo kar nekaj elektrike, kar je tudi po mojem mnenju razlog, zakaj je v soboto vidno višja poraba električne energije kot pa v nedeljo.

Analiza porabe s pomočjo vremenskih podatkov

Za zadnji sklop analiz sem uporabil še vremenske podatke. Ti so mi povzročali največ problemov. Veliko problemov sem imel z uvozom. K sreči so vsi moji porabniki iz Ljubljane, zato sem potreboval vremenske podatke za zadnjih 199 dni le za Ljubljano. Moji vremenski podatki so vsebovali ogromno vremenskih parametrov, jaz pa sem želel stvar malce poenostaviti in sem zato izbral le podatke, ki bi rekli, da imajo največji vpliv na človeka. Tako sem pri vremenskih analizah opazoval temperaturne in padavinske podatke

za prvih 199 dni v Ljubljani.

Koda s katero sem analiziral vpliv vremenskih podatkov pa izgleda takole:

```
data_vreme2_tidy <- read.table("podatki/zdruzeni_podatki/data_vreme2_tidy.txt",sep=" ")
datumi <- (data_vreme2_tidy$datum)
padavine_po_dnevi <- data.frame(rowSums(data_vreme2_tidy[,4:51], is.na(data_vreme2_tidy) == FALSE))
colnames(padavine_po_dnevi) <- c('Kolicina padavin za dan')
rownames(padavine_po_dnevi) <-c(datumi)
write.table(padavine_po_dnevi,"podatki/zdruzeni_podatki/padavine_po_dnevi.txt",sep=" ")

temperature_po_dnevi <- data.frame(rowMeans(data_vreme2_tidy[,52:99], is.na(data_vreme2_tidy) == FALSE))
colnames(temperature_po_dnevi) <- c('Skupna temperatura za dan')
rownames(temperature_po_dnevi) <-c(datumi)
write.table(temperature_po_dnevi,"podatki/zdruzeni_podatki/temperature_po_dnevi.txt",sep=" ")

padavine_po_urah <- data.frame(colSums(data_vreme2_tidy[,4:51]))
colnames(padavine_po_urah) <- c('Povprecna kol padavina za to uro')
write.table(padavine_po_urah,"podatki/zdruzeni_podatki/padavine_po_urah.txt",sep=" ")

temperature_po_urah <- data.frame(colMeans(data_vreme2_tidy[,52:99], is.na(data_vreme2_tidy) == FALSE))
colnames(temperature_po_urah) <- c('Povprecna temperatura za vsako casovno obdobje')
write.table(temperature_po_urah,"podatki/zdruzeni_podatki/temperature_po_urah.txt",sep=" ")
```

Največ časa in potrpljenja pa so mi vzele prostorske predstavitve podatkov. S pomočjo slednjih sem prikazal povprečne letne temperature za vse slovenske statistične regije. Iz **gadm** sem pridobil geografske podatke za statistične regije v Sloveniji. Nato sem iz **ARSO** pridobil podatke za povprečne letne temperature za regije in slednje po nekaj urah spravil na zemljevid in ga obarval po višini letne povprečne temperature.

Temperature smo prikazali z naslednjo R kodo

```
slo.regije.sp <- readRDS("podatki/gadm36_SVN_1_sp.rds")
slo.regije.map = slo.regije.sp %>% spTransform(CRS("+proj=longlat +datum=WGS84")) # pretvorimo v ustrezno projekcijo
print(slo.regije.map@data)
slo.regije.poligoni = fortify(slo.regije.map)
print(slo.regije.poligoni)
slo.regije.poligoni %>%
  left_join(
    rownames_to_column(slo.regije.map@data),
    by = c("id" = "rowname")
  )

slo.regije.poligoni_kratek = slo.regije.poligoni[,c("NAME_1", "long", "lat", "order", "hole", "piece", "area")]

slo.regije.poligoni_kratek %>% write_csv("podatki/OB/OB.shp")

slo.regije.centroidi = slo.regije.map %>% coordinates %>% as.data.frame
colnames(slo.regije.centroidi) = c("long", "lat")

slo.regije.centroidi = slo.regije.centroidi %>% rownames_to_column() %>%
  left_join(
    rownames_to_column(slo.regije.map@data),
    by = "rowname"
  )
slo.regije.centroidi <- slo.regije.centroidi[,c("NAME_1", "long", "lat")]
```



```

jpeg(file="/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/slike/slo_regije1.jpg")
ggplot() +
  geom_polygon(
    data = slo.regije.poligoni_kratek,
    mapping = aes(long, lat, group = group),
    color = 'black',
    fill = 'white'
  ) +
  coord_map() +
  geom_text(
    data = slo.regije.centroidi,
    mapping = aes(x = long, y = lat, label = NAME_1)
  )

dev.off()
jpeg(file="/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/slike/slo_regije2.jpg")

slo.regije.poligoni %>% ggplot() +
  geom_polygon(
    mapping = aes(long, lat, group = group),
    color = "grey",
    fill = "white"
  ) +
  coord_map() +
  geom_text(
    data = slo.regije.centroidi,
    mapping = aes(x = long, y = lat, label = NAME_1),
    size = 3
  ) +
  theme_classic() +
  theme(
    axis.line = element_blank(),
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank()
  )
dev.off()

temperature_po_krajih <- read.csv("podatki/povprecne_temperature.csv", sep = ";")
Gorenjska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Gorenjska regija",2])
Goriška <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Goriska regija",2])
Jugovzhodna_Slovenija <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "JV SLO",2])
Koroška <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Koroska regija",2])
Notranjo_kraška <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Notranjsko-kraska regija",2])
Obalno_kraska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Obalno-kraska regija",2])
Osrednjeslovenska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Osrednjoslovenska regija",2])
Podravska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Podravska regija",2])
Pomurska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Pomurska regija",2])
Savinjska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Savinjska regija",2])
Posavska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Posavska regija",2])
Zasavska <- mean(temperature_po_krajih[temperature_po_krajih[,3] == "Zasavska regija",2])
povprecna_temperatura <- c(Gorenjska,Goriška,Jugovzhodna_Slovenija,Koroška,Notranjo_kraška,Obalno_kraska,Osrednjeslovenska,Podravska,Pomurska,Savinjska,Posavska,Zasavska)

```



```

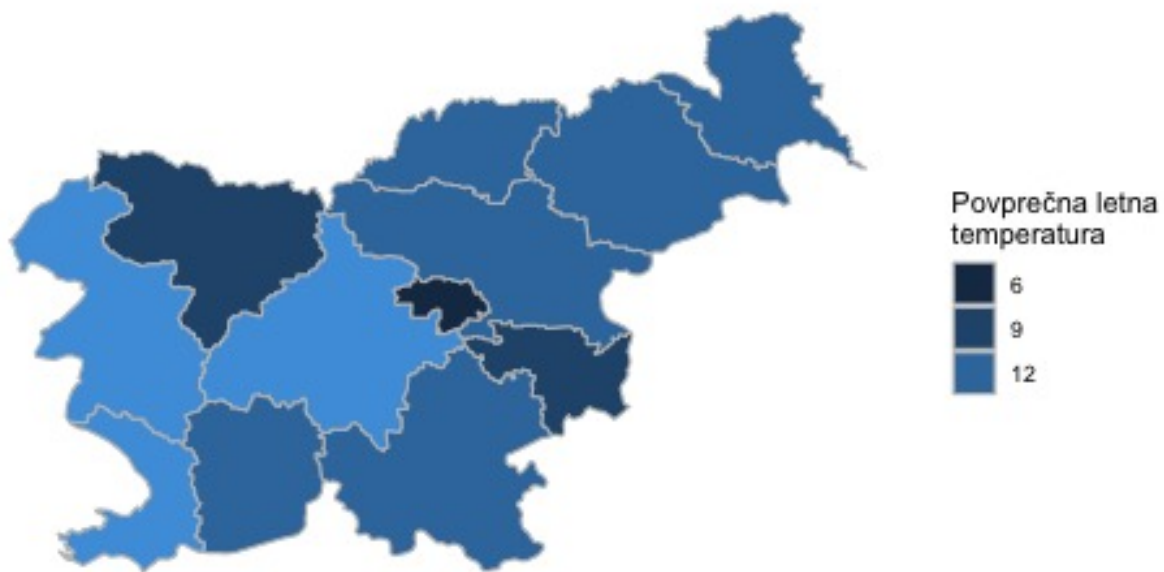
Tabela_regije <- data.frame(NAME_1=Imena_regij,povprecna_temperatura)

jpeg(file="/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/slike/slo_povpr_temp.jpg")

left_join(Tabela_regije,slo.regije.poligoni,
  by = "NAME_1"
) %>%
  ggplot() +
  geom_polygon(
    mapping = aes(long, lat, group = group, fill = povprecna_temperatura),
    color = "grey"
  ) +
  coord_map() +
  scale_fill_binned() +
  theme_classic() +
  theme(
    axis.line = element_blank(),
    axis.ticks = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
  ) +
  guides(fill=guide_legend(title="Povprečna letna \ntemperatura"))

dev.off()

```



V nadaljevanju pa me je zanimalo kakšen vpliv imajo temperature in padavine na povprečno porabo električne energije, kar je bil moj primarni cilj v tem poglavju. Pri tem sem si pomagal z računanjem korelacij med temperaturami in povprečno porabi ter med višino padavine ter povprečno porabo.

Koda, s katero sem izračunal korelaciji pa je taka, kjer za rezultata dobimo $r_{12} = 0.3151918$ ter $r_{13} = -0.0801705$.

```
povpr_poraba_za_vsak_dan <- rowMeans(data_tidy[, 3:98])
povpr_poraba_matrika <- t(matrix(povpr_poraba_za_vsak_dan, 198, 51))
povpr_poraba_za_vsak_dan1 <- as.vector(colSums(povpr_poraba_matrika))
povpr_poraba_za_vsak_dan1 <- povpr_poraba_za_vsak_dan1[-198]
povprecna_T_za_vsak_dan <- temperature_po_dnevih$`Skupna temperatura za dan`
skupne_padavine_za_vsak_dan <- padavine_po_dnevih$`Kolicina padavin za dan`
r12=cor(povpr_poraba_za_vsak_dan1,povprecna_T_za_vsak_dan)
r13=cor(povpr_poraba_za_vsak_dan1,skupne_padavine_za_vsak_dan)
```

S pomočjo teh korelacij sem ugotovil, da imajo temperature mnogo večji vpliv na porabo električne energije kot pa količina padavin. Zame presenetljiv pa je podatek da je pozitivna korelacija med višino temperatur ter povprečno porabo, tega pa si ne znam razložiti. Je pa smiselno, da imajo temperature večji vpliv na porabo kot pa padavine.

Grupiranje podatkov

Ker sem imel 50 različnih porabnikov električne enrgije sem vedel, da v primeru nekaterih skrajnih porabnikov (bodisi zelo visoka bodisi zelo niza poraba) povprečna poraba ni tako uporabna, zato sem se odločil te porabnike razvrstiti v 5 skupin po podobni porabi. Koda, ki sem jo pri tem uporabil pa je naslednja:

```
set.seed(123)
klas <- kmeans(povpr_poraba_matrika[,5])
klasifikacija <- klas$cluster

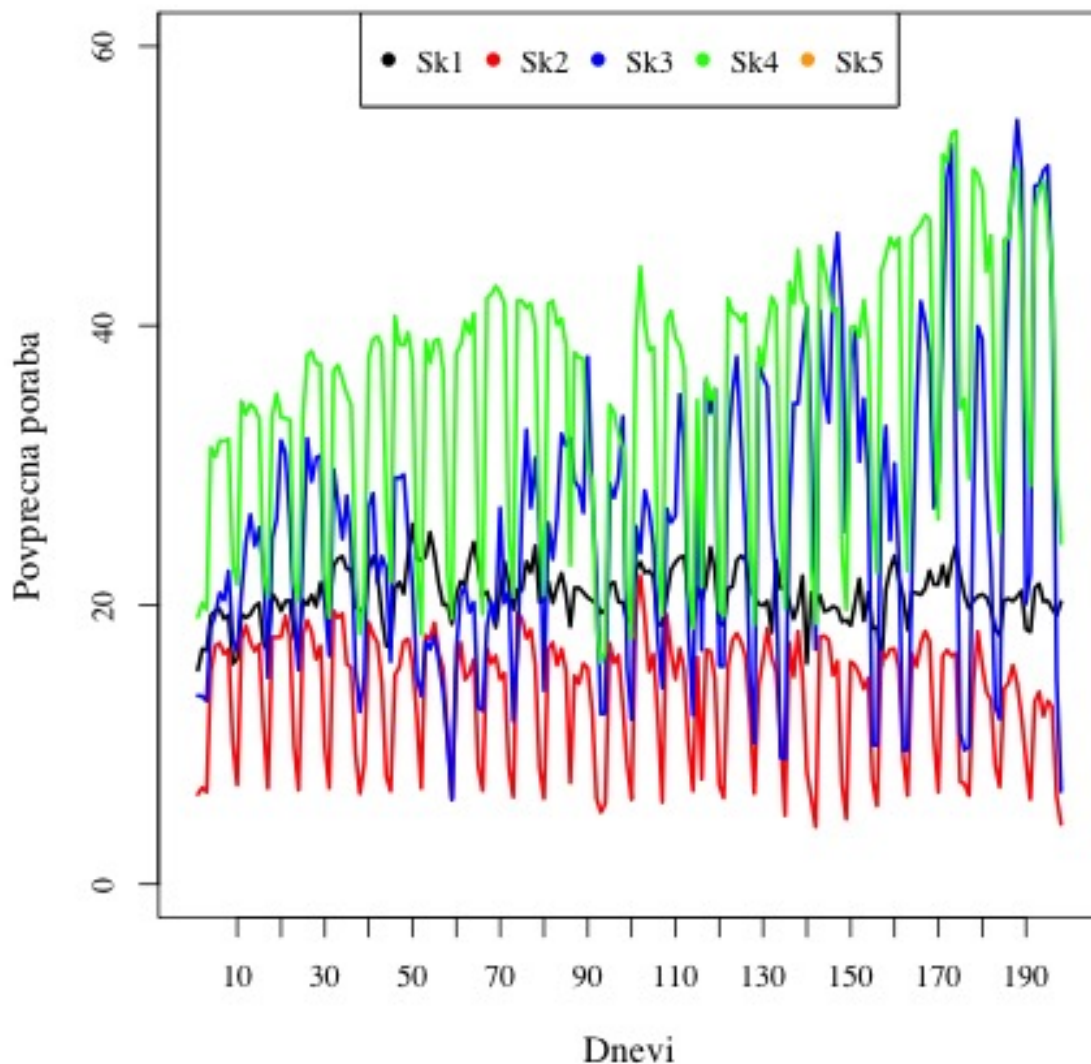
M=max(max(povpr_poraba_matrika[,1:198]))
m=min(min(povpr_poraba_matrika[,1:198]))

Tabela_razred1 <- povpr_poraba_matrika[povpr_poraba_matrika[,199] == "1",1:198]
Tabela_razred2 <- povpr_poraba_matrika[povpr_poraba_matrika[,199] == "2",1:198]
Tabela_razred3 <- povpr_poraba_matrika[povpr_poraba_matrika[,199] == "3",1:198]
Tabela_razred4 <- povpr_poraba_matrika[povpr_poraba_matrika[,199] == "4",1:198]
Tabela_razred5 <- povpr_poraba_matrika[povpr_poraba_matrika[,199] == "5",1:198]

pdf("/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/slike/prikaz_povprecne_porabe_2.pdf")
mar.default <- c(5,2,4,2) + 0.1
par(mar = mar.default + c(0, 4, 0, 0),family="serif") #axis(side, at=, labels=, pos=, lty=, col=, las=,
title="Povprecna poraba za razlicne razrede porabnikov"
plot(c(0,199), c(0,60),yaxt='n',xaxt='n',main=title, type = "n", pch=19,cex.lab=1.2, cex.axis=1.3, cex.l
xlabels=c(1:19)*10
axis(side=1,at=xlabels,labels=xlabels)
axis(side=2,at=c(0:5)*20,labels=as.factor(c(0:5)*20))
lines(colSums(Tabela_razred1), type = "l", col = "black",lwd=2,cex = 1,pch=16)
lines(colSums(Tabela_razred2), type = "l", col = "red",lwd=2,cex = 1,pch=16)
lines(Tabela_razred3, type = "l", col = "blue",lwd=2,cex = 1,pch=16)
lines(colSums(Tabela_razred4), type = "l", col = "green",lwd=2,cex = 1,pch=16)
lines(colSums(Tabela_razred5), type = "l", col = "orange",lwd=2,cex = 1,pch=16)
legend("top", legend=TeX(c("Sk1","Sk2","Sk3","Sk4","Sk5")), pch=c(16,16),cex=c(1,1),col=c("black","red",
dev.off()
```

Nato sem za vsako skupino narisal svoj graf povprečne porabe, da se lažje primerja poraba z drugimi skupinami. Graf pa izgleda takole:

Povprečna poraba za različne razrede porabnikov



Napovedni model

Po tem, ko sem naredil vse potrebne analize, sem se odločil še malo poigrati in sestaviti preprost napovedni model, ki bi znal v grobo napovedati porabo za naslednji dan s pomočjo preteklih porab. Idejo za napovedni model sem dobil pri predmetu numerične metode 1. Tam smo na vajah s pomočjo polinoma sestavili napovedni model, kjer smo iz preteklih vrednosti kriptovalut poizkušali napovedati prihodnjo vrednost kriptovalut. To sem naredil s pomočjo polinoma tretje in četrte stopnje.

Napodeni model sem pa sestavil s sledečo kodo:

```
n <- 3 #stopnja polinoma za aproksimacijo
A <- matrix(c(1,0,0,0,1,1,1,1,1,2,4,8,1,3,9,27), nrow = 4)
m <- length(povpr_poraba_za_vsak_dan1)
t <- matrix(0,length(povpr_poraba_za_vsak_dan1),1)
```

```

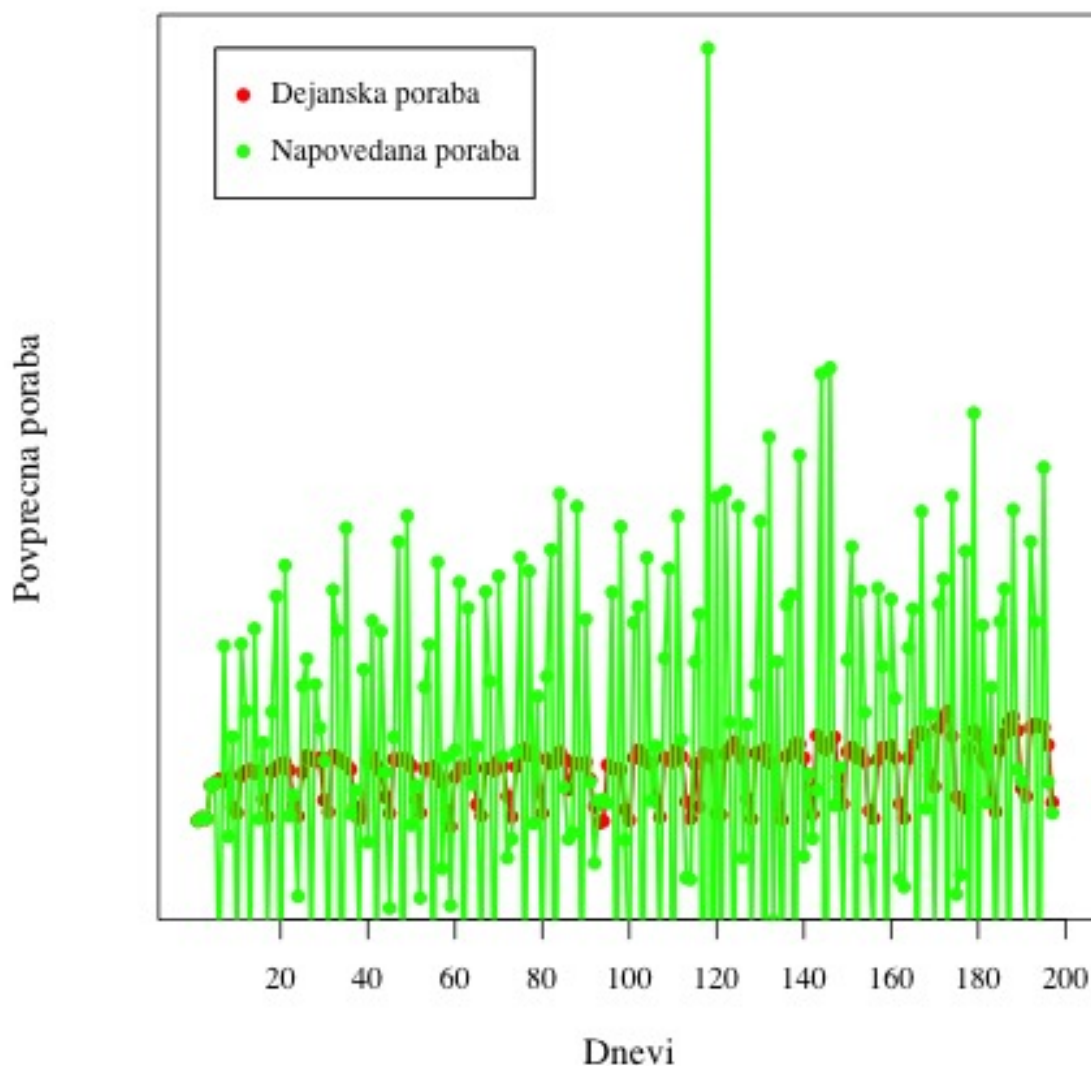
t[1:(n+1)] = povpr_poraba_za_vsak_dan1[1:(n+1)]
for (i in ((n+2):m)) {
  p = solve(A, povpr_poraba_za_vsak_dan1[(i-n-1):(i-1)],)
  t[i,]<- p[1] + p[2]*(n+1)+ p[3]*(n+1)^2 +p[4]*(n+1)^3
}

n1 <- 4 #stopnja polinoma za aproksimacijo
A1 <- matrix(c(1,1,1,1,1,0,1,2,3,4,0,1,4,9,16,0,1,8,27,64,0,1,16,81,256), nrow = 5)
m1 <- length(povpr_poraba_za_vsak_dan1)
t1 <- matrix(0,length(povpr_poraba_za_vsak_dan1),1)
t1[1:(n1+1)] = povpr_poraba_za_vsak_dan1[1:(n1+1)]
for (i in ((n1+2):m1)) {
  p1 = solve(A1, povpr_poraba_za_vsak_dan1[(i-n1-1):(i-1)],)
  t1[i,]<- p1[1] + p1[2]*(n1+1)+ p1[3]*(n1+1)^2 +p1[4]*(n1+1)^3 +p1[5]*(n1+1)^4
}
c <- povpr_poraba_za_vsak_dan1
c_napovedan1 <- t
c_napovedan2 <- t1
slika_aproksimacije <-plot(c, type = "o",pch = 20,col="black")
aproksimacija <- points(c_napovedan1,type = "o",pch = 20, ,col="blue")
aproksimacija1 <- points(c_napovedan2,type = "o",pch = 20, ,col="red")
zgornja_meja <- max(max(c_napovedan),max(c))
jpeg("/Users/blazpovh/Documents/R_projektna_naloga/APPR-2021-22-Blaz-Povh/slike/prikaz_skupne_natancnos
mar.
default <- c(5,2,4,2) + 0.1
par(mar = mar.default + c(0, 4, 0, 0),family="serif") #axis(side, at=, labels=, pos=, lty=, col=, las=,
title="Povprecna poraba in njena napoved"
plot(c(0,200), c(0,zgornja_meja),yaxt='n',xaxt='n',main=title, type = "n", pch=19,cex.lab=1.2, cex.axis=
xlabels=c(1:10)*20
axis(side=1,at=xlabels,labels=xlabels)
#axis(side=2,at=c(0:5)*20,labels=as.factor(c(0:5)*20))
lines(c, type = "o", col = "red",lwd=2,cex = 1,pch=16)
lines(c_napovedan, type = "o", col = "green",lwd=2,cex = 1,pch=16)
legend(5, zgornja_meja/5, legend=TeX(c("Dejanska poraba","Napovedana poraba")), pch=c(16,16),cex=c(1,1)
dev.off()

```

Seveda me je na koncu zanimala, kako dobra je ocena, ki sem jo naredil. Zato sem narisal graf dejanske porabe in napovedane porabe ter oba združil v enega. Končni graf natančnosti ocene pa izgleda takole :

Povprečna poraba in njena napoved



Kot vidimo, je tu še veliko prostora za napredek, saj so ocene preslabe, da bi jih kje lahko uporabili.

Zaključek

V sklopu te projektne naloge sem poleg veliko zanimivih rezultatov prejel še veliko novega znanja iz uvoza, obdelave ter analize in vizualizacije podatkov, kar mi bo prišlo prav tudi pri prihodnjih projektih. Kar se pa tiče napovednih modelov je tu še veliko prostora za napredek in bom še veliko delal na tem, saj je to znanje zelo zašeljeno danes in mnoga podjetja zanimajo napovedi za prihodnost.