



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA DOPRAVNÍ

Bc. Jiřina Lucia Varon Izová

**Analýza dat plovoucích vozidel ve vztahu
k meteorologickým podmínkám**

Příloha 1

R skripty

2023

Obsah

1. Skript – Export z databáze ŘSD.....	2
2. Skript – Kubova huť.....	4
3. Skript – Dálnice D11	16
4. Skript – Predikce datový set 1.....	25
5. Skript – Predikce datový set 2.....	29

1. Skript – Export z databáze ŘSD

```
library(RPostgreSQL)

con <- dbConnect(RPostgres::Postgres(), dbname = "RSD_FCD", host="localhost",
port="5432",
user="izova", password="XXXXX")

tab <- dbListTables(con) #tabulky v databázi pro kontrolu připojení
##KUBOVA HUT
#FCD z roku 2019
vd_tab_19 <- grep("^vd_19", tab, value = TRUE) #tabulky dat z roku 2019
segm_kh <- "TS23363T26307" #filtrování podle TMC segmentu
for (navez_tab in vd_tab_19) {
  query <- paste0("SELECT datum_cas, vd0, vd2, vd5, vd8, vd10 FROM ", navez
_tab, " WHERE vd0 = '", segm_kh, "'") #SQL dotaz
  data <- dbGetQuery(con, query) #použití dotazu v databázi
  write.csv(data, file = paste0(navez_tab, ".csv"), row.names = FALSE)
} #prochází jednotlivé tabulky a ukládá do formátu CSV
#FCD z roku 2020
vd_tab_20 <- grep("^vd_20", tab, value = TRUE) #tabulky dat z roku 2020
for (navez_tab in vd_tab_20) {
  query <- paste0("SELECT datum_cas, vd0, vd2, vd5, vd8, vd10 FROM ", navez
_tab, " WHERE vd0 = '", segm_kh, "'") #SQL dotaz
  data <- dbGetQuery(con, query) #použití dotazu v databázi
  write.csv(data, file = paste0(navez_tab, ".csv"), row.names = FALSE)
} #prochází jednotlivé tabulky a ukládá do formátu CSV
#FCD z roku 2021
vd_tab_21 <- grep("^vd_21", tab, value = TRUE) #tabulky dat z roku 2021
for (navez_tab in vd_tab_21) {
  query <- paste0("SELECT datum_cas, vd0, vd2, vd5, vd8, vd10 FROM ", navez
_tab, " WHERE vd0 = '", segm_kh, "'") #SQL dotaz
  data <- dbGetQuery(con, query) #použití dotazu v databázi
  write.csv(data, file = paste0(navez_tab, ".csv"), row.names = FALSE)
} #prochází jednotlivé tabulky a ukládá do formátu CSV

##DÁLNICE D11 - 17,5 KM
#FCD z roku 2019
vd_tab_19 <- grep("^vd_19", tab, value = TRUE) #tabulky dat z roku 2019
```

```

segm_kh <- "TS25617T01349" #filtrování podle TMC segmentu
for (nazev_tab in vd_tab_19) {
  query <- paste0("SELECT datum_cas, vd0, vd2, vd5, vd8, vd10 FROM ", nazev
_tab, " WHERE vd0 = '", segm_kh, "'") #SQL dotaz
  data <- dbGetQuery(con, query) #použití dotazu v databázi
  write.csv(data, file = paste0(nazev_tab, ".csv"), row.names = FALSE)
} #prochází jednotlivé tabulky a ukládá do formátu CSV
#FCD z roku 2020
vd_tab_20 <- grep("^vd_20", tab, value = TRUE) #tabulky dat z roku 2020
for (nazev_tab in vd_tab_20) {
  query <- paste0("SELECT datum_cas, vd0, vd2, vd5, vd8, vd10 FROM ", nazev
_tab, " WHERE vd0 = '", segm_kh, "'") #SQL dotaz
  data <- dbGetQuery(con, query) #použití dotazu v databázi
  write.csv(data, file = paste0(nazev_tab, ".csv"), row.names = FALSE)
} #prochází jednotlivé tabulky a ukládá do formátu CSV
#FCD z roku 2021
vd_tab_21 <- grep("^vd_21", tab, value = TRUE) #tabulky dat z roku 2021
for (nazev_tab in vd_tab_21) {
  query <- paste0("SELECT datum_cas, vd0, vd2, vd5, vd8, vd10 FROM ", nazev
_tab, " WHERE vd0 = '", segm_kh, "'") #SQL dotaz
  data <- dbGetQuery(con, query) #použití dotazu v databázi
  write.csv(data, file = paste0(nazev_tab, ".csv"), row.names = FALSE)
} #prochází jednotlivé tabulky a ukládá do formátu CSV

#Spojení souborů do jednoho datasetu - Kubova Hut
cesta <- ".../DP/FCD_kubova_hut"
nazvy_s <- list.files(cesta, pattern = "^vd_", full.names = TRUE)
data <- lapply(nazvy_s, read.csv)
spojeni <- do.call(rbind, data)
write.csv(spojeni, "fcd_kh.csv", row.names = FALSE) #novy CSV soubor

#Spojení souborů do jednoho datasetu - Dálnice D11
cesta <- ".../DP/FCD_d11"
nazvy_s <- list.files(cesta, pattern = "^vd_", full.names = TRUE)
data <- lapply(nazvy_s, read.csv)
spojeni <- do.call(rbind, data)
write.csv(spojeni, "fcd_d11.csv", row.names = FALSE) #novy CSV soubor

```

2. Skript – Kubova huť

```
###KUBOVA HUŤ###  
#knihovny  
library(dplyr)  
library(lubridate)  
library(ggplot2)  
library(moments)  
library(stats)  
library(MASS)  
  
##meteorologická data##  
#načtení dat - meteostanice. Kodování bylo upraveno v příkazovém řádku na U  
TF-8  
meteo_kh_19 <- read.csv("../DP/export-202205110909/C010-export-2019_new.cs  
v", sep=";")  
meteo_kh_20 <- read.csv("../DP/export-202205110909/C010-export-2020_new.cs  
v", sep=";")  
meteo_kh_21 <- read.csv("../DP/export-202205110909/C010-export-2021_new.cs  
v", sep=";")  
meteo_kh <- rbind(meteo_kh_19, meteo_kh_20, meteo_kh_21) #sloučení dat do j  
ednoho datasetu  
  
#úprava dat  
names(meteo_kh)[names(meteo_kh) %in% c("Čas.UTC.", "Teplota.vzduchu", "Rosn  
ý.bod", "Tnamrzání", "Srážky", "Rychlost.větru", "Směr.větru")] <-  
  c("cas", "T", "Ros_bod", "T_namrz", "Srazky", "Rychl_v", "Smer_v") #přejm  
enování sloupců  
meteo_kh <- meteo_kh[,-c(1,4,6,7,8,9)] #promazávání sloupců  
meteo_kh <- meteo_kh[,-9:-18]  
meteo_kh$cas <- strptime(meteo_kh$cas, "%d.%m.%Y %H:%M") #převedení na časo  
vý formát  
meteo_kh[,-1] <- lapply(meteo_kh[,-1], gsub, pattern = ",", replacement =  
".", fixed = TRUE) #nahrazení všech čárek na tečku, kromě prvního sloupce  
meteo_kh$T <- as.numeric(meteo_kh$T) #převedení na číselný formát  
meteo_kh$Tvoz <- as.numeric(meteo_kh$Tvoz)  
meteo_kh$Stav <- gsub("mo\u009enost namrzání", "možnost namrzání", meteo_kh  
$Stav)  
meteo_kh$Srazky <- as.numeric(meteo_kh$Srazky)  
meteo_kh$Vlhkost <- as.numeric(meteo_kh$Vlhkost)  
meteo_kh$Tlak <- as.numeric(meteo_kh$Tlak)
```

```

meteo_kh$Rychl_v <- as.numeric(meteo_kh$Rychl_v)
meteo_kh <- meteo_kh %>%
  mutate(mesic = month(cas), rok = year(cas)) #vytvoření nových sloupců

#NA hodnoty
colSums(is.na(meteo_kh)) #počty nulových hodnot v jednotlivých sloupcích
for (i in 2:(nrow(meteo_kh) - 1)) {
  if (is.na(meteo_kh$T[i])) { #pokud je v řádce NA hodnota,nahradí se průměrem z předchozí a následující buňky
    meteo_kh$T[i] <- mean(c(meteo_kh$T[i-1], meteo_kh$T[i+1]), na.rm = TRUE)
  }
}

for (i in 2:(nrow(meteo_kh) - 1)) {
  if (is.na(meteo_kh$Vlhkost[i])) { #pokud je v řádce NA hodnota,nahradí se průměrem z předchozí a následující buňky
    meteo_kh$Vlhkost[i] <- mean(c(meteo_kh$Vlhkost[i-1], meteo_kh$Vlhkost[i+1]), na.rm = TRUE)
  }
}

for (i in 2:(nrow(meteo_kh) - 1)) {
  if (is.na(meteo_kh$Tvoz[i])) { #pokud je v řádce NA hodnota,nahradí se průměrem z předchozí a následující buňky
    meteo_kh$Tvoz[i] <- mean(c(meteo_kh$Tvoz[i-1], meteo_kh$Tvoz[i+1]), na.rm = TRUE)
  }
}

for (i in 2:(nrow(meteo_kh) - 1)) {
  if (is.na(meteo_kh$Tlak[i])) { #pokud je v řádce NA hodnota,nahradí se průměrem z předchozí a následující buňky
    meteo_kh$Tlak[i] <- mean(c(meteo_kh$Tlak[i-1], meteo_kh$Tlak[i+1]), na.rm = TRUE)
  }
}

for (i in 2:(nrow(meteo_kh) - 1)) {
  if (is.na(meteo_kh$Rychl_v[i])) { #pokud je v řádce NA hodnota,nahradí se průměrem z předchozí a následující buňky
    meteo_kh$Rychl_v[i] <- mean(c(meteo_kh$Rychl_v[i-1], meteo_kh$Rychl_v[i+1]), na.rm = TRUE)
  }
}

```

```

for (i in 2:(nrow(meteo_kh) - 1)) {
  if (is.na(meteo_kh$Srazky[i])) { #pokud je v řádku NA hodnota, nahradí se
    průměrem z předchozí a následující buňky
    meteo_kh$Srazky[i] <- mean(c(meteo_kh$Srazky[i-1], meteo_kh$Srazky[i+
1]), na.rm = TRUE)
  }
}
#ošetření NA hodnoty v posledním řádku ve sloupci Srážky
i <- nrow(meteo_kh)
meteo_kh$Srazky[i] <- meteo_kh$Srazky[i-1]
#ošetření chybějících hodnot ve sloupci Stav. Pokud se mohlo jednoznačně ur
čit, že se jedná o výpadek (hodnoty v předchozím a následujícím řádku byly
totožné), byly nahrazeny.
#Zbytek (5 řádků) byl z datasetu odstraněn.
for (i in 2:(nrow(meteo_kh) - 1)) {
  if (meteo_kh$Stav[i]=="" & meteo_kh$Stav[i-1]==meteo_kh$Stav[i+1]) { #pok
ud je v řádku NA hodnota, nahraď ji průměrem z předchozí a následující buňk
y
    meteo_kh$Stav[i] <- meteo_kh$Stav[i-1]
  }
}
meteo_kh <- meteo_kh[meteo_kh$Stav != "", ] #odstranění prázdných řádků

#rozdíl teplot
pom = data.frame(meteo_kh$T, meteo_kh$mesic, meteo_kh$rok, meteo_kh$Vlhkos
t) #pomocný dataframe
pom <- pom %>%
  group_by(meteo_kh.rok, meteo_kh.mesic) %>%
  mutate(delta_t = meteo_kh.T - mean(meteo_kh.T))
meteo_kh <- meteo_kh %>%
  mutate(delta_t = pom$delta_t) #vytvoření nového sloupce
meteo_kh <- meteo_kh[, c(1, 4, 9:10, 2, 11, 3, 5:8)] #přehození sloupců

#vývoj hodnot vlhkosti vzduchu - otestování, zda existuje nějaký vzorec v z
ávislosti na ročním období
ggplot(filter(meteo_kh, meteo_kh$rok==2021), aes(x = factor(mesic), y = Vlh
kost)) +
  geom_boxplot(fill = "royalblue1", color = "black", outlier.color = "red",
outlier.size = 0.3) +
  scale_y_continuous(limits = c(0, 100), expand = c(0, 0)) +

```

```

  labs(x = "Měsíc", y = "Vlhkost vzduchu [%]", title = "Krabicový graf vlhkosti vzduchu v roce 2021 v jednotlivých měsících") +

  theme_bw()

ggplot(filter(meteo_kh, meteo_kh$rok==2020), aes(x = factor(mesic), y = Vlhkost)) +

  geom_boxplot(fill = "royalblue1", color = "black", outlier.color = "red", outlier.size = 0.3) +

  scale_y_continuous(limits = c(0, 100), expand = c(0, 0)) +

  labs(x = "Měsíc", y = "Vlhkost vzduchu [%]", title = "Krabicový graf vlhkosti vzduchu v roce 2020 v jednotlivých měsících") +

  theme_bw()

ggplot(filter(meteo_kh, meteo_kh$rok==2019), aes(x = factor(mesic), y = Vlhkost)) +

  geom_boxplot(fill = "royalblue1", color = "black", outlier.color = "red", outlier.size = 0.3) +

  scale_y_continuous(limits = c(0, 100), expand = c(0, 0)) +

  labs(x = "Měsíc", y = "Vlhkost vzduchu [%]", title = "Krabicový graf vlhkosti vzduchu v roce 2019 v jednotlivých měsících") +

  theme_bw()

#stav

dummymatrix <- model.matrix(~Stav, data = meteo_kh) #převod kategoriální proměnné "Stav" na binární proměnné pomocí funkce model.matrix()

dummymatrix <- dummymatrix[,-1] #odstranění prvního sloupce

meteo_kh <- cbind(meteo_kh, dummymatrix) #sloučení dat do jednoho datasetu

#popisná statistika meteostanice:

#průměrné hodnoty v jednotlivých rocích

meteo_kh_t1 <- meteo_kh %>%

  group_by(Rok = rok) %>%

  summarise(across(.cols = -c(cas, Stav, mesic, rok), .fns = list(Minimum = min, Maximum = max, Průměr = mean, Medián = median))) %>%

  round(2)

write.table(meteo_kh_t1, file = ".../DP/DP_R/meteo_kh_t1.txt", sep = ",", quote = FALSE, row.names = F)

#četnost stavu vozovky

meteo_kh_t2 <- meteo_kh %>%

  group_by(Rok = rok) %>%

  count(Stav) %>%

  mutate("Relativní četnost" = round(n/sum(n) * 100,2)) #relativní četnost v rámci roku

```



```

write.table(meteo_kh_t2, file = ".../DP/DP_R/meteo_kh_t2.txt", sep = ",", q
uote = FALSE, row.names = F)

##FCD data##
fcd_kh <- read.csv("../DP/DP_R/fcd_kh.csv")

#úprava dat
fcd_kh$datum_cas <- strptime(fcd_kh$datum_cas, "%Y-%m-%d %H:%M") #převedení
na časový formát
order_index <- order(as.POSIXct(fcd_kh$datum_cas, format = "%Y-%m-%d %H:%M
")) #seřazení podle času
fcd_kh <- fcd_kh[order_index, ] #seřazení podle času
names(fcd_kh)[names(fcd_kh) %in% c("datum_cas", "vd2", "vd5", "vd8", "vd10
")] <-
  c("cas", "rychl_akt", "rychl_typ", "m_spoleh", "los") #přejmenování sloup
ců
fcd_kh <- fcd_kh[,-2] #promazávání sloupce
fcd_kh[fcd_kh == 0] <- NA #v případě 0 hodnot se nahradí NA hodnotou
fcd_kh <- na.omit(fcd_kh) #promazání NA hodnot
fcd_kh <- fcd_kh %>%
  mutate(mesic = month(cas), rok = year(cas)) %>% #vytvoření nových sloupců
- měsíc a rok
  mutate(delta_rychl = (rychl_akt - rychl_typ)) #vytvoření nového sloupce -
rozdíl rychlosti typické a aktuální
#write.csv(fcd_kh, ".../DP/DP_R/exporty_fcd_kh.csv", row.names=FALSE)

#popisná statistika FCD
#průměrné hodnoty v jednotlivých rocích
fcd_kh_t1 <- fcd_kh %>%
  group_by(Rok = rok) %>%
  summarise("Průměrná vypočtená rychlost" = round(mean(rychl_akt), 3), "Prů
měrná typická rychlost" = round(mean(rychl_typ), 3), "Průměrná míra spolehl
ivosti" = round(mean(m_spoleh), 3), "Průměrný stupeň dopravy" = round(mean
(los), 3))
write.table(fcd_kh_t1, file = ".../DP/DP_R/fcd_kh_t1.txt", sep = ",", quote
= FALSE, row.names = F)

##nový dataframe - spojení meteodat a FCD##
data_kh <- merge(meteo_kh, fcd_kh, by = "cas")

```

```

names(data_kh)[names(data_kh) %in% c("mesic.x", "rok.x", "Stavmokro nasolen o", "Stavmožnost namrzání", "Stavsnih/led", "Stavsucho", "Stavvlhko", "Stav zbytková sůl")] <-
  c("mesic", "rok", "C_moknas", "C_namr", "C_snled", "C_such", "C_vlh", "C_zbsul") #přejmenování sloupců
data_kh <- data_kh[,-c(22,23)] #promazávání sloupců
#write.csv(data_kh, ".../DP/DP_R/expoty/data_kh_exp.csv", row.names=FALSE)

#normalita dat
#histogramy
h1 = hist(data_kh$T, breaks = 30, col = "aquamarine", xlim = c(-20,30), xlab = "Teplota vzduchu", ylab = "Hustota pravděpodobnosti", main = "Histogram teploty vzduchu", freq = FALSE)
nz_t <- fitdistr(data_kh$T, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)
h2 = hist(data_kh$delta_t, breaks = 30, col = "aquamarine", xlim = c(-20,20), xlab = "Rozdil teplot vzduchu", ylab = "Hustota pravděpodobnosti", main = "Histogram rozdílu teplot vzduchu", freq = FALSE)
nz_t <- fitdistr(data_kh$delta_t, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)
h3 = hist(data_kh$Tvoz, breaks = 30, col = "aquamarine", xlim = c(-20,50), xlab = "Teplota vozovky", ylab = "Hustota pravděpodobnosti", main = "Histogram teploty vozovky", freq = FALSE)
nz_t <- fitdistr(data_kh$Tvoz, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)
h4 = hist(data_kh$Srazky, breaks = 50, col = "aquamarine", xlab = "Srážky", ylab = "Hustota pravděpodobnosti", main = "Histogram srážek", freq = FALSE)
nz_t <- fitdistr(data_kh$Srazky, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)
h5 = hist(data_kh$Vlhkost, breaks = 30, col = "aquamarine", xlim = c(10,100), xlab = "Vlhkost", ylab = "Hustota pravděpodobnosti", main = "Histogram vlhkosti vzduchu", freq = FALSE)
nz_t <- fitdistr(data_kh$Vlhkost, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)
h6 = hist(data_kh$Tlak, breaks = 30, col = "aquamarine", xlab = "Tlak", ylab = "Hustota pravděpodobnosti", main = "Histogram tlaku vzduchu", freq = FALSE)
nz_t <- fitdistr(data_kh$Tlak, "normal")

```

```

curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE,
col = "red", lwd = 3)

h7 = hist(data_kh$Rychl_v, breaks = 30, col = "aquamarine", xlab = "Rychlost
v\u00e9tru", ylab = "Hustota pravd\u00e9podobnosti", main = "Histogram rychlosti v\u00e9
tru", freq = FALSE)

nz_t <- fitdistr(data_kh$Rychl_v, "normal")

curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE,
col = "red", lwd = 3)

#histogram stavu vozovky, který ale je jen pro zobrazení četnosti, nelze ur
čit normálnost dat (jen u spojitých veličin)

pom2 <- data_kh %>%
  count(Stav) %>%
  arrange(desc(n))

ggplot(pom2, aes(x = fct_reorder(Stav, n), y = n)) +
  geom_col(fill = "royalblue", width = 0.4) +
  labs(x = "Stav vozovky", y = "Četnost", title = "Histogram stavu vozovky
") +
  theme_bw() +
  scale_y_continuous() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))

#Kvantil-kvantilové grafy (q-q plot)

plot(data_kh$T)

qqnorm(data_kh$T, main = "Kvantil-kvantilový graf teploty vzduchu", xlab = "
Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.
5)

qqline(data_kh$T, col = "red", lwd = 2)

plot(data_kh$delta_t)

qqnorm(data_kh$delta_t, main = "Kvantil-kvantilový graf rozdílu teploty vzdu
chu", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "dark
blue", cex=0.5)

qqline(data_kh$delta_t, col = "red", lwd = 2)

plot(data_kh$Tvoz)

qqnorm(data_kh$Tvoz, main = "Kvantil-kvantilový graf teploty vozovky", xlab
= "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=
0.5)

qqline(data_kh$Tvoz, col = "red", lwd = 2)

plot(data_kh$Srazky)

qqnorm(data_kh$Srazky, main = "Kvantil-kvantilový graf srážek", xlab = "Teor
etický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)

qqline(data_kh$Srazky, col = "red", lwd = 2)

```

```

plot(data_kh$Vlhkost)

qqnorm(data_kh$Vlhkost,main = "Kvantil-kvantilový graf vlhkosti", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)

qqline(data_kh$Vlhkost, col = "red", lwd = 2)

plot(data_kh$Tlak)

qqnorm(data_kh$Tlak,main = "Kvantil-kvantilový graf tlaku", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)

qqline(data_kh$Tlak, col = "red", lwd = 2)

plot(data_kh$Rychl_v)

qqnorm(data_kh$Rychl_v,main = "Kvantil-kvantilový graf rychlosti větru", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)

qqline(data_kh$Rychl_v, col = "red", lwd = 2)

#koeficienty šikmosti a špičatosti

koef_sikmost <- as.data.frame(apply(data_kh[,5:11], 2, skewness)) #koeficient šikmosti

koef_spicatost <- as.data.frame(apply(data_kh[,5:11], 2, kurtosis)) #koeficient špičatosti

##Nový dataframe - spojení meteodat a FCD##

data_kh <- merge(meteo_kh, fcd_kh, by = "cas")

data_kh <- data_kh[,-c(22,23)] #promazávání sloupců

names(data_kh)[names(data_kh) %in% c("mesic.x", "rok.x", "Stavmokro nasolen o", "Stavmožnost namrzání", "Stavsnih/led", "Stavsucho", "Stavvlhko", "Stavzbytková sůl")] <-

  c("mesic", "rok", "C_moknas", "C_namr", "C_snled", "C_such", "C_vlh", "C_zbsul") #přejmenování sloupců

#Korelační analýza

kor_mat <- cor(data_kh[,5:17], method = "spearman") # výpočet korelační matice - provedení Spearmanova testu nezávislosti

kor_mat <- round(kor_mat, 2)

write.table(kor_mat, file = ".../DP/DP_R/heat_kh_1.csv", sep = ",", quote = FALSE, row.names = F)

library(reshape2)

kor_mat_melt <- melt(kor_mat)

kor_mat_melt <- kor_mat_melt %>%

```

```

mutate(Var1 = gsub("C_moknas", "S: mokro nasoleno", Var1),
       Var1 = gsub("C_namr", "S: možnost namrzání", Var1),
       Var1 = gsub("C_snled", "S: sníh/led", Var1),
       Var1 = gsub("C_such", "S: sucho", Var1),
       Var1 = gsub("C_vlh", "S: vlhko", Var1),
       Var1 = gsub("C_zbsul", "S: zbytková sůl", Var1),
       Var1 = gsub("Tvoz", "T vozovky", Var1),
       Var1 = gsub("delta_t", "Rozdíl T", Var1),
       Var1 = gsub("Rychl_v", "Rychlost větru", Var1) )
kor_mat_melt <- kor_mat_melt %>%
  mutate(Var2 = gsub("C_moknas", "S: mokro nasoleno", Var2),
         Var2 = gsub("C_namr", "S: možnost namrzání", Var2),
         Var2 = gsub("C_snled", "S: sníh/led", Var2),
         Var2 = gsub("C_such", "S: sucho", Var2),
         Var2 = gsub("C_vlh", "S: vlhko", Var2),
         Var2 = gsub("C_zbsul", "S: zbytková sůl", Var2),
         Var2 = gsub("Tvoz", "T vozovky", Var2),
         Var2 = gsub("delta_t", "Rozdíl T", Var2),
         Var2 = gsub("Rychl_v", "Rychlost větru", Var2) )

ggplot(kor_mat_melt, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "darkgoldenrod1", mid = "white", high = "red",
midpoint = 0) +
  theme_minimal() +
  labs(x = "", y = "", fill = "Korelace", title = "Heatmapa korelační matic
e", subtitle = "Spearmanův korelační koeficient") +
  theme(axis.text.x=element_text(vjust = 1, hjust = 1, angle = 45),
        axis.text.y=element_text(vjust=0.2))

kor_mat_2 <- cor(data_kh[,5:17]) #pearson
kor_mat_2 <- round(kor_mat_2, 2)

#vykreslení významných závislostí
#vlhkost-delta_t
barvy_mesicu <- rainbow(12) # 12 barev pro 12 měsíců
nazvy_mesicu <- c("Leden", "Únor", "Březen", "Duben", "Květen", "Červen", "
Červenec", "Srpen", "Září", "Říjen", "Listopad", "Prosinec")

```

```

plot(data_kh$delta_t, data_kh$Vlhkost, pch=20, cex=0.5, col=barvy_mesicu[as.integer(data_kh$mesic)], main = "Graf rozdílu teploty a vlhkosti",
      xlab = "Rozdíl teplot [°C]", ylab = "Vlhkost [%]")

legend("topright", legend=nazvy_mesicu, col=barvy_mesicu, pch=20, cex=0.8,
       title="Měsíce")

#koeficienty šikmosti a špičatosti

koef_sikmost <- as.data.frame(apply(data_kh[,5:11], 2, skewness)) #koeficient šikmosti
koef_spicatost <- as.data.frame(apply(data_kh[,5:11], 2, kurtosis)) #koeficient špičatosti

#Regresní analýza
#MODEL 1
modell <- lm(delta_rychl ~ delta_t + Tvoz + Srazky + Vlhkost + Tlak + Rychl_v, data = data_kh)
summary(modell)
plot(modell)

#VYHODNOCENÍ MODELU
reg_koef <- summary(modell)$coefficients[, "Estimate"] #výpis t_hodnot z výsledků modelu
t_hodn <- summary(modell)$coefficients[, "t value"] #výpis t_hodnot z výsledků modelu
st_vol <- 28419 #určení stupňů volnosti z výsledků modelu (počet pozorování - počet odhadovaných parametrů - 1.)
t_0025 <- qt(0.025, st_vol) #výpočet hodnoty kvantilu t_0.025 (hodnota t-hodnoty, která odpovídá 2.5% kvantilu t-rozdělení)
#delta_t
index <- which(names(t_hodn) == "delta_t") #index jednotlivé proměnné v modelu
koef <- summary(modell)$coefficients[index, "Estimate"]
odch <- summary(modell)$coefficients[index, "Std. Error"]
inspol <- c(koef - t_0025 * odch, koef + t_0025 * odch) #výpočet intervalu spolehlivosti
cat("Interval spolehlivosti pro regresní koeficient delta_t:", round(inspol, 5))
qt(0.05, 28419) #určení kritické hodnoty t-testu
qf(0.05, 4, 28419) ##určení kritické hodnoty f-testu
#Tvoz

```

```

index2 <- which(names(t_hodn) == "Tvoz") #index jednotlivé proměnné v model
u
koef2 <- summary(modell1)$coefficients[index2, "Estimate"]
odch2 <- summary(modell1)$coefficients[index2, "Std. Error"]
inspol2 <- c(koef2 - t_0025 * odch2, koef2 + t_0025 * odch2) #výpočet inter
valu spolehlivosti
cat("Interval spolehlivosti pro regresní koeficient Tvoz:", round(inspol2,
5))

#Srazky
index3 <- which(names(t_hodn) == "Srazky") #index jednotlivé proměnné v mod
elu
koef3 <- summary(modell1)$coefficients[index3, "Estimate"]
odch3 <- summary(modell1)$coefficients[index3, "Std. Error"]
inspol3 <- c(koef3 - t_0025 * odch3, koef3 + t_0025 * odch3) #výpočet inter
valu spolehlivosti
cat("Interval spolehlivosti pro regresní koeficient Srazky:", round(inspol
3, 5))

#Vlhkost
index4 <- which(names(t_hodn) == "Vlhkost") #index jednotlivé proměnné v mo
delu
koef4 <- summary(modell1)$coefficients[index4, "Estimate"]
odch4 <- summary(modell1)$coefficients[index4, "Std. Error"]
inspol4 <- c(koef4 - t_0025 * odch4, koef4 + t_0025 * odch4) #výpočet inter
valu spolehlivosti
cat("Interval spolehlivosti pro regresní koeficient Vlhkost:", round(inspol
4, 5))

#Tlak
index5 <- which(names(t_hodn) == "Tlak") #index jednotlivé proměnné v model
u
koef5 <- summary(modell1)$coefficients[index5, "Estimate"]
odch5 <- summary(modell1)$coefficients[index5, "Std. Error"]
inspol5 <- c(koef5 - t_0025 * odch5, koef5 + t_0025 * odch5) #výpočet inter
valu spolehlivosti
cat("Interval spolehlivosti pro regresní koeficient Tlak:", round(inspol5,
5))

#Rychl_v
index6 <- which(names(t_hodn) == "Rychl_v") #index jednotlivé proměnné v mo
delu
koef6 <- summary(modell1)$coefficients[index6, "Estimate"]
odch6 <- summary(modell1)$coefficients[index6, "Std. Error"]
inspol6 <- c(koef6 - t_0025 * odch6, koef6 + t_0025 * odch6) #výpočet inter
valu spolehlivosti

```

```

cat("Interval spolehlivosti pro regresní koeficient Rychl_v:", round(inspol
6, 5))

qt = qt(0.05, 28419) #určení kritické hodnoty t-testu
qf = qf(0.05, 4, 28419) ##určení kritické hodnoty f-testu

#ÚPRAVA MODELU 1 PO VYŘAZENÍ VYBRANÝCH PROMĚNNÝCH
modell1_2 <- lm(delta_rychl ~ delta_t + Srazky, data = data_kh)
summary(modell1_2)

##MODEL 2 - BINOMICKÉ ROZDĚLENÍ
data_kh2 <- data_kh

#nastavení hranice pro rozdíl snížení rychlosti - rozčlenění dat na dvě kat
egorie
data_kh2$ Kateg_r <- ifelse(data_kh2$delta_rychl <= -20, 1, #1: změna rychlo
sti >= -20
                                0) #0: rychlost se snížila o méně než 20 km/h, nezměnil
a se vůbec, nebo se naopak zvýšila

model2 <- glm(kateg_r ~ C_moknas + C_namr + C_snled + C_vlh + C_zbsul, data
= data_kh2, family = binomial()) #binární logistická regrese pomocí GLM
summary(model2)

```


3. Skript – Dálnice D11

```
###DÁLNIČE D11###  
#knihovny  
library(dplyr)  
library(lubridate)  
library(ggplot2)  
library(moments)  
library(stats)  
library(MASS)  
  
##meteorologická data##  
#načtení dat - meteostanice. Kódování bylo upraveno v příkazovém řádku na U  
TF-8  
meteo_d <- read.csv("../DP/export-202205110909/S008-export-2021_new.csv",  
sep=";")  
  
#úprava dat  
l=meteo_d[meteo_d$Srą.ky != "", ]  
meteo = l[,c("Čas.UTC.", "Teplota.vzduchu", "Rosný.bod", "Tvoz", "Stav", "S  
rá.ky", "Vlhkost", "Dohlednost")]  
names(meteo)[names(meteo) %in% c("Čas.UTC.", "Teplota.vzduchu", "Rosný.bod  
, "Srą.ky")] <-  
  c("cas", "T", "Ros_bod", "Srazky") #přejmenování sloupců  
meteo$cas <- strptime(meteo$cas, "%d.%m.%Y %H:%M") #převedení na časový for  
mát  
meteo[,-1] <- lapply(meteo[,-1], gsub, pattern = ",", replacement = ".", fi  
xed = TRUE) #nahrazení všech čárek na tečku, kromě prvního sloupce  
meteo$T <- as.numeric(meteo$T) #převedení na číselný formát  
meteo$Tvoz <- as.numeric(meteo$Tvoz)  
meteo$Stav <- gsub("mo\u009enost namrzání", "možnost namrzání", meteo$Stav)  
meteo$Srazky <- as.numeric(meteo$Srazky)  
meteo$Vlhkost <- as.numeric(meteo$Vlhkost)  
meteo$Dohlednost <- as.numeric(meteo$Dohlednost)  
meteo$Ros_bod <- as.numeric(meteo$Ros_bod)  
  
#NA hodnoty  
colSums(is.na(meteo)) #počty nulových hodnot v jednotlivých sloupcích  
meteo1 = na.omit(meteo) #SADA BEZ NA HODNOT - z toho korelacni matice
```

```

meteo2 = meteo[,c("cas", "T", "Srazky", "Dohlednost")] #SADA s vybranými p
arametry - po VYŘAZENÍ kvůli korelaci

meteo2 = na.omit(meteo2) #promazání NA hodnot

#meteo2 = meteo[,c("cas", "T", "Srazky", "Dohlednost", "mesic", "rok")] #S
ADA s vybranými parametry - po vyrazení kvuli korelaci

#meteo3 = meteo[,c("cas", "T", "Srazky", "Dohlednost", "Stav")] #SADA s vy
branými parametry - PRO PREDIKCI

#meteo3 = na.omit(meteo3) #SADA s vybranými parametry - PRO PREDIKCI

#stav

stav = meteo[,c("cas", "Stav")]

stav <- stav[stav$Stav != "", ] #odstranění prázdných řádků

dummymatrix <- model.matrix(~Stav, data = stav) #převod kategoriální promě
ně "Stav" na binární proměnné pomocí funkce model.matrix()

dummymatrix <- dummymatrix[,-1] #odstranění prvního sloupce (interceptu)

stav <- cbind(stav, dummymatrix) #připojení binárních proměnných k původním
datům

#popisná statistika meteostanice :

#průměrné hodnoty v jednotlivých rocích

meteo_d_t1 <- meteo2 %>% #pred upravou datasetu

  summarise(across(.cols = -c(cas), .fns = list(Minimum = min, Maximum = ma
x, Průměr = mean, Medián = median)) %>%

  round(2)

write.table(meteo_d_t1, file = "meteo_d_t1.txt", sep = ",", quote = FALSE,
row.names = F)

meteo_d_t2 <- stav %>% #četnost stavu vozovky

  count(Stav) %>%

  mutate("Relativní četnost" = n/sum(n) * 100)%>% #relativní četnost v rámc
i roku

  arrange(n)

write.table(meteo_d_t2, file = "meteo_d_t2.txt", sep = ",", quote = FALSE,
row.names = F)

##FCD data##

fcd_d <- read.csv("../DP/DP_R/fcd_d11.csv")

#úprava dat

```

```

fcd_d$datum_cas <- strptime(fcd_d$datum_cas, "%Y-%m-%d %H:%M") #převedení n
a časový formát

order_index <- order(as.POSIXct(fcd_d$datum_cas, format = "%Y-%m-%d %H:%M
"))

fcd_d <- fcd_d[order_index, ]

names(fcd_d)[names(fcd_d) %in% c("datum_cas", "vd2", "vd5", "vd8", "vd10")]
<-
  c("cas", "rychl_akt", "rychl_typ", "m_spoleh", "los") #přejmenování sloup
ců

fcd_d <- fcd_d[,-2] #promazávání sloupce

fcd_d[fcd_d == 0] <- NA #v případě 0 hodnot se nahradí NA hodnotou

fcd_d <- na.omit(fcd_d) #promazání NA hodnot

fcd_d <- fcd_d %>%

  mutate(mesic = month(cas), rok = year(cas)) %>% #vytvoření nových sloupců
- měsíc a rok

  mutate(delta_rychl = (rychl_akt - rychl_typ)) #vytvoření nového sloupce -
rozdíl rychlosti typické a aktuální

#popisná statistika FCD

#průměrné hodnoty v jednotlivých rocích

fcd_d_t1 <- fcd_d %>%

  group_by(Rok = rok) %>%

  summarise("Průměrná vypočtená rychlost" = round(mean(rychl_akt), 3), "Prů
měrná typická rychlost" = round(mean(rychl_typ), 3), "Průměrná míra spolehl
ivosti" = round(mean(m_spoleh), 3), "Průměrný stupeň dopravy" = round(mean
(los), 3))

write.table(fcd_d_t1, file = "fcd_d_t1.txt", sep = ",", quote = FALSE, row.
names = F)

##Nový dataframe - spojení meteodat a FCD##

data_d11 <- merge(meteo2, fcd_d, by = "cas")

####PREDIKCE - export dat

data_pred1 <- merge(meteo1, fcd_d, by = "cas")

write.csv(data_pred1, file = "...DP_2023/Rstudio/data_pred1.csv")

data_pred2 <- merge(meteo3, fcd_d, by = "cas")

data_pred2 <- data_pred2[,-c(1,7,8,9)]

write.csv(data_pred2, file = "...DP_2023/Rstudio/data_pred2.csv")

####

```

```

#normalita dat
#histogramy
h1 = hist(data_d11$T, breaks = 30, ylim = c(0,0.05), col = "aquamarine", xlab = "Teplota vzduchu", ylab = "Hustota pravděpodobnosti", main = "Histogram teploty vzduchu", freq = FALSE)
nz_t <- fitdistr(data_d11$T, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)
h2 = hist(data_d11$Srazky, breaks = 50, col = "aquamarine", xlab = "Srážky", ylab = "Hustota pravděpodobnosti", main = "Histogram srážek", freq = FALSE)
nz_t <- fitdistr(data_d11$Srazky, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)
h3 = hist(data_d11$Dohlednost, breaks = 30, col = "aquamarine", xlab = "Dohlednost", ylab = "Hustota pravděpodobnosti", main = "Histogram dohlednosti", freq = FALSE)
nz_t <- fitdistr(data_d11$Dohlednost, "normal")
curve(dnorm(x, mean = nz_t$estimate[1], sd = nz_t$estimate[2]), add = TRUE, col = "red", lwd = 3)

#histogram stavu vozovky, který ale je jen pro zobrazení četnosti, nelze určit normálnost dat (jen u spojitých veličin)
pom2 <- stav %>%
  count(Stav) %>%
  arrange(n)
ggplot(meteo_d_t2, aes(x = Stav, y = n)) +
  geom_col(fill = "aquamarine", width = 0.5, stat = "identity") +
  labs(x = "Stav vozovky", y = "Četnost", title = "Histogram stavu vozovky") +
  theme_bw() +
  scale_y_continuous() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))
h8 = ggplot(stav, aes(x = Stav)) +
  geom_bar(fill = "aquamarine") +
  labs(x = "Stav vozovky", y = "Četnost", title = "Histogram stavu vozovky") +
  theme_bw()
h8 + scale_y_continuous() +
  theme(axis.text.x = element_text(vjust = 1, hjust = 1, angle = 45)) +

#koeficienty šikmosti a špičatosti

```

```

koef_sikmost <- as.data.frame(apply(data_d11[,5:4], 2, skewness)) #koeficient šikmosti
koef_spicatost <- as.data.frame(apply(data_d11[,2:4], 2, kurtosis)) #koeficient špičatosti

#Kvantil-kvantilové grafy (q-q plot)
plot(data_d11$T)
qqnorm(data_d11$T,main = "Kvantil-kvantilový graf teploty vzduchu", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)
qqline(data_d11$T, col = "red", lwd = 2)
plot(data_d11$delta_t)
qqnorm(data_d11$delta_t,main = "Kvantil-kvantilový graf rozdílu teploty vzduchu", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)
qqline(data_d11$delta_t, col = "red", lwd = 2)
plot(data_d11$Tvoz)
qqnorm(data_d11$Tvoz,main = "Kvantil-kvantilový graf teploty vozovky", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)
qqline(data_d11$Tvoz, col = "red", lwd = 2)
plot(data_d11$Srazky)
qqnorm(data_d11$Srazky,main = "Kvantil-kvantilový graf srážek", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)
qqline(data_d11$Srazky, col = "red", lwd = 2)
plot(data_d11$Vlhkost)
qqnorm(data_d11$Vlhkost,main = "Kvantil-kvantilový graf vlhkosti", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)
qqline(data_d11$Vlhkost, col = "red", lwd = 2)
plot(data_d11$Tlak)
qqnorm(data_d11$Tlak,main = "Kvantil-kvantilový graf tlaku", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)
qqline(data_d11$Tlak, col = "red", lwd = 2)
plot(data_d11$Rychl_v)
qqnorm(data_d11$Rychl_v,main = "Kvantil-kvantilový graf rychlosti větru", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)
qqline(data_d11$Rychl_v, col = "red", lwd = 2)
plot(data_d11$Dohlednost)

```

```

qqnorm(data_d11$Dohlednost, main = "Kvantil-quantilový graf dohlednosti", xlab = "Teoretický kvantil", ylab = "Pozorovaný kvantil", col = "darkblue", cex=0.5)

qqline(data_d11$Dohlednost, col = "red", lwd = 2)

#Korelační analýza
mat_k <- merge(meteo1, stav, by = "cas")
mat_k <- mat_k[,-c(1,5,9)]

kor_mat <- cor(mat_k, method = "spearman") # výpočet korelační matice - provedení Spearmanova testu nezávislosti

kor_mat <- round(kor_mat, 2)

library(reshape2)

kor_mat_melt <- melt(kor_mat)
kor_mat_melt <- kor_mat_melt %>%
  mutate(Var1 = gsub("Stavmokronasoleno", "S: mokro nasoleno", Var1),
         Var1 = gsub("Stavmožnost namrzání", "S: možnost namrzání", Var1),
         Var1 = gsub("Stavsucho", "S: sucho", Var1),
         Var1 = gsub("Stavvlhko", "S: vlhko", Var1),
         Var1 = gsub("Stavzbytková sůl", "S: zbytková sůl", Var1),
         Var1 = gsub("Tvoz", "T vozovky", Var1),
         Var1 = gsub("Ros_bod", "Rosný bod", Var1),
         Var1 = gsub("Srazky", "Srážky", Var1) )

kor_mat_melt <- kor_mat_melt %>%
  mutate(Var2 = gsub("Stavmokronasoleno", "S: mokro nasoleno", Var2),
         Var2 = gsub("Stavmožnost namrzání", "S: možnost namrzání", Var2),
         Var2 = gsub("Stavsucho", "S: sucho", Var2),
         Var2 = gsub("Stavvlhko", "S: vlhko", Var2),
         Var2 = gsub("Stavzbytková sůl", "S: zbytková sůl", Var2),
         Var2 = gsub("Tvoz", "T vozovky", Var2),
         Var2 = gsub("Ros_bod", "Rosný bod", Var2),
         Var2 = gsub("Srazky", "Srážky", Var2) )

ggplot(kor_mat_melt, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "darkgoldenrod1", mid = "white", high = "red", midpoint = 0) +
  theme_minimal() +
  labs(x = "", y = "", fill = "Korelace", title = "Heatmapa korelační matice", subtitle = "Spearmanův korelační koeficient") +

```

```

theme(axis.text.x=element_text(vjust = 1, hjust = 1, angle = 45),
      axis.text.y=element_text(vjust=0.2))

#vykreslení významných závislostí
##ROSNÝ BOD-SRAZKY
barvy_mesicu <- rainbow(12) # 12 barev pro 12 měsíců
nazvy_mesicu <- c("Leden", "Únor", "Březen", "Duben", "Květen", "Červen", "
Červenec", "Srpen", "Září", "Říjen", "Listopad", "Prosinec")
plot(meteo1$Srazky, meteo1$Ros_bod, pch=20, cex=0.5, col=barvy_mesicu[as.in
teger(data_d11$mesic)], main = "Graf rosného bodu a srážek",
      xlab = "Srážky [mm]", ylab = "Rosný bod [°C]")
legend("topright", legend=nazvy_mesicu, col=barvy_mesicu, pch=20, cex=0.8,
title="Měsíce")
##ROSNÝ BOD-TEPLOTA
plot(meteo1$T, meteo1$Ros_bod, pch=20, cex=0.5, col=barvy_mesicu[as.integer
(data_d11$mesic)], main = "Graf rosného bodu a teplot",
      xlab = "Teplota [°C]", ylab = "Rosný bod [°C]")
legend("topright", legend=nazvy_mesicu, col=barvy_mesicu, pch=20, cex=0.8,
title="Měsíce")
##TVOZ-TEPLOTA
plot(meteo1$T, meteo1$Tvoz, pch=20, cex=0.5, col=barvy_mesicu[as.integer(da
ta_d11$mesic)], main = "Graf teplot vozovky a teplot vzduchu",
      xlab = "Teplota vzduchu [°C]", ylab = "Teplota vozovky [°C]")
legend("topright", legend=nazvy_mesicu, col=barvy_mesicu, pch=20, cex=0.8,
title="Měsíce")
##Vlhkost-TEPLOTA
plot(meteo1$T, meteo1$Vlhkost, pch=20, cex=0.5, col=barvy_mesicu[as.integer
(data_d11$mesic)], main = "Graf vlhkosti a teplot vzduchu",
      xlab = "Teplota vzduchu [°C]", ylab = "Vlhkost vzduchu [%]")
legend("topright", legend=nazvy_mesicu, col=barvy_mesicu, pch=20, cex=0.8,
title="Měsíce")

#koeficienty šikmosti a špičatosti
koef_sikmost <- as.data.frame(apply(data_d11[,5:4], 2, skewness)) #koeficie
nt šikmosti
koef_spicatost <- as.data.frame(apply(data_d11[,2:4], 2, kurtosis)) #koefic
ient špičatosti

#Regresní analýza
#MODEL 1
modell <- lm(delta_rychl ~ T + Srazky + Dohlednost, data = data_d11)

```

```

summary(modell)
plot(modell)

qt(0.05, 283680)
qf(0.05, 3, 283680)

#VYHODNOCENÍ MODELU
reg_koef <- summary(modell)$coefficients[, "Estimate"] #výpis t_hodnot z vý
sledků modelu
t_hodn <- summary(modell)$coefficients[, "t value"] #výpis t_hodnot z výsle
dků modelu
st_vol <- 283680 #určení stupňů volnosti z výsledků modelu (počet pozorován
í - počet odhadovaných parametrů - 1.)
t_0025 <- qt(0.025, st_vol) #výpočet hodnoty kvantilu t_0.025 (hodnota t-ho
dnoty, která odpovídá 2.5% kvantilu t-rozdělení)

##T
index <- which(names(t_hodn) == "T") #index jednotlivé proměnné v modelu
koef <- summary(modell)$coefficients[index, "Estimate"]
odch <- summary(modell)$coefficients[index, "Std. Error"]
inspol <- c(koef - t_0025 * odch, koef + t_0025 * odch) #výpočet intervalu
spolehlivosti
cat("Interval spolehlivosti pro regresní koeficient T:", round(inspol, 5))
qt(0.05, 283680) #určení kritické hodnoty t-testu
qf(0.05, 4, 28419) ##určení kritické hodnoty f-testu

##Srazky
index2 <- which(names(t_hodn) == "Srazky") #index jednotlivé proměnné v mod
elu
koef2 <- summary(modell)$coefficients[index2, "Estimate"]
odch2 <- summary(modell)$coefficients[index2, "Std. Error"]
inspol2 <- c(koef2 - t_0025 * odch2, koef2 + t_0025 * odch2) #výpočet inter
valu spolehlivosti
cat("Interval spolehlivosti pro regresní koeficient Srazky:", round(inspol
2, 5))

##Dohlednost
index3 <- which(names(t_hodn) == "Dohlednost") #index jednotlivé proměnné v
modelu
koef3 <- summary(modell)$coefficients[index3, "Estimate"]
odch3 <- summary(modell)$coefficients[index3, "Std. Error"]
inspol3 <- c(koef3 - t_0025 * odch3, koef3 + t_0025 * odch3) #výpočet inter
valu spolehlivosti

```



```

cat("Interval spolehlivosti pro regresní koeficient Dohlednost:", round(inspol3, 5))

##Vlhkost

index4 <- which(names(t_hodn) == "Vlhkost") #index jednotlivé proměnné v modelu
koef4 <- summary(modell1)$coefficients[index4, "Estimate"]
odch4 <- summary(modell1)$coefficients[index4, "Std. Error"]
inspol4 <- c(koef4 - t_0025 * odch4, koef4 + t_0025 * odch4) #výpočet intervalu spolehlivosti

cat("Interval spolehlivosti pro regresní koeficient Vlhkost:", round(inspol4, 5))

#ÚPRAVA MODELU 1 PO VYŘAZENÍ VYBRANÝCH PROMĚNNÝCH
modell1_2 <- lm(delta_rychl ~ T + Srazky, data = data_d11)
summary(modell1_2)
plot(modell1_2)
modell1_3 <- lm(delta_rychl ~ T + Srazky, data = data_d11)
summary(modell1_3)
plot(modell1_3)

##MODEL 2 - BINOMICKE ROZDELENÍ
data_d112 <- data_d11
data_d112 <- merge(data_d112, stav, by = "cas")
names(data_d112)[names(data_d112) %in% c("Stavmokro nasoleno", "Stavmožnost namrzání", "Stavsucho", "Stavvlhko", "Stavzbytková sůl")] <-
  c("C_moknas", "C_namr", "C_such", "C_vlh", "C_zbsul") #přejmenování sloupců

data_d112$kateg_r <- ifelse(data_d112$delta_rychl <= -20, 1, # kategorie 1: snížení rychlosti větší než 20 km/h
  0) # kategorie 2: rychlost se nesnížila

model3 <- glm(kateg_r ~ C_moknas + C_namr + C_such + C_vlh + C_zbsul, data = data_d112, family = binomial()) #binární logistická regrese pomocí GLM
summary(model3)

#ÚPRAVA MODELU 2 PO VYŘAZENÍ VYBRANÝCH PROMĚNNÝCH
model4 <- glm(kateg_r ~ C_namr + C_such + C_vlh, data = data_d112, family = binomial())
summary(model4)

```

4. Skript – Predikce datový set 1

```
library(dplyr)
library(kknn)
library(caret)
library(stats)
library(randomForest)
library(ggplot2)

data1 <- read.csv("../DP_2023/Rstudio/data_pred1.csv")
data1 <- data1[,-c(1,2,11,12,13)] #promazávání sloupců
data1 <- data1[data1$Stav != "", ]
data1 = sample_n(data1, size=20000)

#Rozdělení na trénovací a testovací data
set.seed(123) #generování náhodného čísla

index <- sample(1:nrow(data1), size = 0.8 * nrow(data1)) #koeficient 0.8 ro
zdělí data v poměru 80:20

train_data1 <- data1[index, ] #trénovací data
test_data1 <- data1[-index, ] #testovací data

control <- trainControl(method = "cv", number = 10) #kontrola křížové valid
ace

#převedení kategorických proměnných na faktor
train_data1$Stav <- as.factor(train_data1$Stav)
test_data1$Stav <- as.factor(test_data1$Stav)

#RANDOM FOREST

#model
set.seed(123)

rf_model <- randomForest(rychl_akt ~ ., data = train_data1) #trénování
rf_pred <- predict(rf_model, newdata = test_data1) #predikce
rf_rmse <- sqrt(mean((test_data1$rychl_akt - rf_pred)^2)) #výpočet RMSE

#výstupy
cat("Náhodné lesy RMSE:", rf_rmse, "\n")
print(rf_model)
summary(rf_model)
plot(rf_model)

imp <- round(importance(rf_model),2) #důležitost proměnných
varImpPlot(rf_model, pch = 19, col = "darkgreen", main = "Důležitost proměn
ných")
```

```

mae_rf <- mean(abs(test_data1$rychl_akt - rf_pred)) #chyba MAE
cat("Náhodné lesy MAE:", mae, "\n")

#grafy
n <- 100 #prvních 100 řádků z testovací sady
x = 1:n
plot(x, test_data1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
      main = "Předpověď rychlosti (Random forest)", xlab = "Čas", ylab = "Rychlost")
lines(x, rf_pred[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

##K-NEAREST NEIGHBOR
# Normalizace spojitých proměnných
numeric_vars <- c("T", "Ros_bod", "Tvoz", "Srazky", "Vlhkost", "Dohlednost")
preprocessing_params <- preProcess(train_data1[, numeric_vars], method = c("center", "scale"))
train_data_norm1 <- predict(preprocessing_params, train_data1[, numeric_vars])
test_data_norm1 <- predict(preprocessing_params, test_data1[, numeric_vars])
train_data_norm1$Stav <- train_data1$Stav
train_data_norm1$rychl_akt <- train_data1$rychl_akt
test_data_norm1$Stav <- test_data1$Stav
test_data_norm1$rychl_akt <- test_data1$rychl_akt

#tuneGrid = expand.grid(k = c(5,8,10,15))

#knn_model <- train(rychl_akt ~ ., data = train_data_norm1, method = "knn",
trControl = control, tuneGrid = tuneGrid) #s kolika k má model nejlepší výsledek

#model
knn_model_k1 <- train(rychl_akt ~ ., data = train_data_norm1, method = "knn",
trControl = control, k=1) #trénování - k=1
knn_pred_k1 <- predict(knn_model_k1, newdata = test_data_norm1) #predikce
knn_rmse_k1 <- sqrt(mean((test_data_norm1$rychl_akt - knn_pred_k1)^2)) #výpočet RMSE

knn_model_k5 <- train(rychl_akt ~ ., data = train_data_norm1, method = "knn",
trControl = control, k=5) #trénování - k=5
knn_pred_k5 <- predict(knn_model_k5, newdata = test_data_norm1)

```

```

knn_rmse_k5 <- sqrt(mean((test_data_norm1$rychl_akt - knn_pred_k5)^2))
knn_model_k10 <- train(rychl_akt ~ ., data = train_data_norm1, method = "knn",
trControl = control, k=10) #trénování - k=10
knn_pred_k10 <- predict(knn_model_k10, newdata = test_data_norm1)
knn_rmse_k10 <- sqrt(mean((test_data_norm1$rychl_akt - knn_pred_k10)^2))
mae_knn <- mean(abs(test_data1$rychl_akt - knn_pred_k10)) #chyba MAE

#výstupy
cat("k-NN k=1 RMSE:", knn_rmse_k1, "\n")
cat("k-NN k=5 RMSE:", knn_rmse_k5, "\n")
cat("k-NN k=10 RMSE:", knn_rmse_k25, "\n")
cat("k-NN k=10 MAE:", mae_knn, "\n")
print(knn_model)
summary(knn_model)
plot(knn_model)

#graf pro k = 1
plot(x, test_data_norm1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
      main = "Predikce rychlosti pro k = 1", xlab = "Čas", ylab = "Rychlost")
lines(x, knn_pred_k1[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

#graf pro k = 5
plot(x, test_data_norm1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
      main = "Predikce rychlosti pro k = 5", xlab = "Čas", ylab = "Rychlost")
lines(x, knn_pred_k5[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

#graf pro k = 10
plot(x, test_data_norm1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
      main = "Predikce rychlosti pro k = 10", xlab = "Čas", ylab = "Rychlost")
lines(x, knn_pred_k10[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

```

```

#LINEARNI REGRESE
linear_model <- lm(rychl_akt ~ ., data = train_data1) #Trénování
linear_pred <- predict(linear_model, newdata = test_data1) #predikce
linear_rmse <- sqrt(mean((test_data1$rychl_akt - linear_pred)^2)) #výpočet
RMSE
print(linear_model)
summary(linear_model)
plot(linear_model)
mae_lm <- mean(abs(test_data1$rychl_akt - linear_pred)) #chyba MAE
cat("Lineární regrese RMSE:", linear_rmse, "\n")
cat("Lineární regrese MAE:", mae_lm, "\n")
#graf
plot(x, test_data1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
      main = "Predikce rychlosti (lineární model)", xlab = "Čas", ylab = "Ry
chlost")
lines(x, linear_pred[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost
"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

vysl = data.frame(Typ = c("Random Forest", "KNN", "Lineární"), RMSE = c(rf_
rmse, knn_rmse_k10, linear_rmse), MAE = c(mae_rf, mae_knn, mae_lm))
prumer = mean(data1$rychl_akt)
vysl = vysl %>%
  mutate(rel_ch = RMSE/prumer)

```

5. Skript – Predikce datový set 2

```
library(dplyr)
library(kknn)
library(caret)
library(stats)
library(randomForest)
library(ggplot2)

data1 <- read.csv("../DP_2023/Rstudio/data_pred2.csv")
data1 <- data1[,-c(1)] #promazávání sloupců
data1 <- data1[data1$Stav != "", ]
data1 = sample_n(data1, size=20000)

#Rozdělení na trénovací a testovací data
set.seed(123) #generování náhodného čísla
index <- sample(1:nrow(data1), size = 0.8 * nrow(data1)) #koeficient 0.8 ro
zdělí data v poměru 80:20
train_data1 <- data1[index, ] #trénovací data
test_data1 <- data1[-index, ] #testovací data
control <- trainControl(method = "cv", number = 10) #kontrola křížové valid
ace

#převedení kategorických proměnných na faktor
train_data1$Stav <- as.factor(train_data1$Stav)
test_data1$Stav <- as.factor(test_data1$Stav)

#RANDOM FOREST
#model
set.seed(123)
rf_model <- randomForest(rychl_akt ~ ., data = train_data1) #trénování
rf_pred <- predict(rf_model, newdata = test_data1) #predikce
rf_rmse <- sqrt(mean((test_data1$rychl_akt - rf_pred)^2)) #výpočet RMSE
#výstupy
cat("Náhodné lesy RMSE:", rf_rmse, "\n")
print(rf_model)
summary(rf_model)
plot(rf_model)
round(importance(rf_model),2) #důležitost proměnných
varImpPlot(rf_model, pch = 19, col = "darkgreen", main = "Důležitost proměn
ných")
```

```

mae_rf <- mean(abs(test_data1$rychl_akt - rf_pred)) #chyba MAE
cat("Náhodné lesy MAE:", mae, "\n")
#grafy
n <- 100 #prvních 100 řádků z testovací sady
x = 1:n
plot(x, test_data1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
      main = "Předpověď rychlosti 2 (Random forest)", xlab = "Čas", ylab = "
Rychlost")
lines(x, rf_pred[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost
"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

##K-NEAREST NEIGHBOR
# Normalizace spojitých proměnných
numeric_vars <- c("T", "Srazky", "Dohlednost")
preprocessing_params <- preProcess(train_data1[, numeric_vars], method = c
("center", "scale"))
train_data_norm1 <- predict(preprocessing_params, train_data1[, numeric_var
s])
test_data_norm1 <- predict(preprocessing_params, test_data1[, numeric_var
s])
train_data_norm1$Stav <- train_data1$Stav
train_data_norm1$rychl_akt <- train_data1$rychl_akt
test_data_norm1$Stav <- test_data1$Stav
test_data_norm1$rychl_akt <- test_data1$rychl_akt

#model
knn_model_k1 <- train(rychl_akt ~ ., data = train_data_norm1, method = "kkn
n", trControl = control, k=1) #trénování - k=1
knn_pred_k1 <- predict(knn_model_k1, newdata = test_data_norm1) #predikce
knn_rmse_k1 <- sqrt(mean((test_data_norm1$rychl_akt - knn_pred_k1)^2)) #výp
očet RMSE
knn_model_k5 <- train(rychl_akt ~ ., data = train_data_norm1, method = "kkn
n", trControl = control, k=5) #trénování - k=5
knn_pred_k5 <- predict(knn_model_k5, newdata = test_data_norm1)
knn_rmse_k5 <- sqrt(mean((test_data_norm1$rychl_akt - knn_pred_k5)^2))
knn_model_k10 <- train(rychl_akt ~ ., data = train_data_norm1, method = "kk
nn", trControl = control, k=10) #trénování - k=10
knn_pred_k10 <- predict(knn_model_k10, newdata = test_data_norm1)
knn_rmse_k10 <- sqrt(mean((test_data_norm1$rychl_akt - knn_pred_k10)^2))

```

```

mae_knn <- mean(abs(test_data1$rychl_akt - knn_pred_k10)) #chyba MAE

#výstupy
cat("k-NN k=1 RMSE:", knn_rmse_k1, "\n")
cat("k-NN k=5 RMSE:", knn_rmse_k5, "\n")
cat("k-NN k=10 RMSE:", knn_rmse_k10, "\n")
cat("k-NN k=10 MAE:", mae_knn, "\n")

print(knn_model_k10)
summary(knn_model_k10)
plot(knn_model_k10)

#graf pro k = 1
plot(x, test_data_norm1$rychl_akt[1:n], col = "red", type = "l", lwd=1.5,
     main = "Predikce rychlosti pro k = 1", xlab = "Čas", ylab = "Rychlost")
lines(x, knn_pred_k1[1:n], col = "blue", lwd=1.5)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

#graf pro k = 5
plot(x, test_data_norm1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
     main = "Predikce rychlosti pro k = 5", xlab = "Čas", ylab = "Rychlost")
lines(x, knn_pred_k5[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

#graf pro k = 10
plot(x, test_data_norm1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
     main = "Predikce rychlosti 2 pro k = 10", xlab = "Čas", ylab = "Rychlost")
lines(x, knn_pred_k10[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

#LINEARNÍ REGRESE
linear_model <- lm(rychl_akt ~ ., data = train_data1) #Trénování
linear_pred <- predict(linear_model, newdata = test_data1) #predikce
linear_rmse <- sqrt(mean((test_data1$rychl_akt - linear_pred)^2)) #výpočet RMSE

```



```

summary(linear_model)
plot(linear_model)
mae_lm <- mean(abs(test_data1$rychl_akt - linear_pred)) #chyba MAE
cat("Lineární regrese RMSE:", linear_rmse, "\n")
cat("Lineární regrese MAE:", mae_lm, "\n")

#graf
plot(x, test_data1$rychl_akt[1:n], col = "red", type = "l", lwd=2,
      main = "Predikce rychlosti 2 (lineární model)", xlab = "Čas", ylab = "
Rychlost")
lines(x, linear_pred[1:n], col = "blue", lwd=2)
legend("topright", legend = c("Skutečná rychlost", "Predikovaná rychlost
"),
      fill = c("red", "blue"), col = 2:3, adj = c(0, 0.6))

vysl = data.frame(Typ = c("Random Forest", "KNN", "Lineární"), RMSE = c(rf_
rmse, knn_rmse_k10, linear_rmse), MAE = c(mae_rf, mae_knn, mae_lm))
prumer = mean(data1$rychl_akt)
vysl = vysl %>%
  mutate(rel_ch = RMSE/prumer)
ggplot(vysl, aes(x = Typ, y = RMSE, fill = Typ)) +
  geom_bar(fill = "blue", width = 0.2, stat = "identity") +
  labs(x = "Typ modelu", y = "RMSE", title = "Grafické vyjádření RMSE chyby
") +
  theme_bw()

```