TPC1

Grupo4

2023-09-30

# Nota:

Deve efetuar todos os Save com “Save with encoding UTF-8” de modo a manter palavras acentuadas e caracteres especiais\*\*

## Base de dados:condutores.csv

# Remover tudo!  
  
# Incluir as libraries de que necessita  
library(VIM)  
library(plyr)  
library(ggplot2)  
library(dplyr)  
library(lubridate)  
library(tidyverse)  
library(psych)  
library (corrplot)  
library(ggplot2)  
library(ggfortify)  
library(lsr)# Cramer's V e Eta  
library(factoextra)

# Questão 1 [5 valores]

## Leitura dos dados condutores.csv.

Data <-read.csv("condutores.csv", header=TRUE, stringsAsFactors =T,sep=",",dec=",",check.names=F,na.strings=c("NA","NÃO DEFINIDO"),fileEncoding = "UTF-8")

## Averigue a existência de valores omissos e, caso existam, identifique as respetivas variáveis. Realize a imputação dos valores omissos da variável “Tempo.Condução.Continuada” considerando as variáveis Tipo.Veiculo, Tipo.Serviço e Distrito. Faça upload do sumário de Tempo.Condução.Continuada após a imputação.

summary(is.na(Data))  
Data1<-Data[,c("Tempo Condução Continuada", "Tipo Veiculo", "Tipo Serviço", "Distrito")]  
  
Data1<-kNN(Data1, variable = "Tempo Condução Continuada", k = 200)  
  
summary(is.na(Data1))  
  
Data1<- Data1[, c("Tempo Condução Continuada", "Tipo Veiculo", "Tipo Serviço", "Distrito")]  
  
Data$`Tempo Condução Continuada`<-Data1$`Tempo Condução Continuada`

summary(Data$`Tempo Condução Continuada`)

## Questão 2 [5 valores] Efetue a análise descritiva da variável Ano.matricula, de modo a completar:

#Efetue a análise descritiva da variável Ano.matricula, de modo a completar:  
  
#O veículo mais antigo é do ano 1914 e a percentagem (cumulativa) de veículos com ano de matrícula inferior ou igual a 2007 é 51,5%(percentagem com uma casa decimal); os veículos com ano de matrícula inferior a 1977 (indique o ano) podem ser considerados outliers. Considerando o ano de 2020 para o cálculo da idade dos veículos, observa-se que 25% dos veículos têm menos de 7 anos.

## Questão 3 [5 valores] Crie a variável nominal Estação\_Ano com as classes “Inverno”, “Primavera”, “Verão” e “Outono” e faça upload do correspondente gráfico de barras apresentando as frequências relativas.

Data$Estacao\_ano<- rep(NA,40209)  
Data$Datahora<- ymd\_hms(Data$Datahora)  
Data$Estacao\_ano[which(Data$Datahora > "2020-3-22" & Data$Datahora <= "2020-6-20" )]<- "Primavera"  
Data$Estacao\_ano[which(Data$Datahora > "2020-6-20" & Data$Datahora <= "2020-9-22" )]<- "Verão"  
Data$Estacao\_ano[which(Data$Datahora > "2020-9-22" & Data$Datahora <= "2020-12-21" )]<- "Outono"  
Data$Estacao\_ano[which(Data$Datahora > "2020-12-21" & Data$Datahora <= "2021-1-1" )]<- "Inverno"  
Data$Estacao\_ano[which(Data$Datahora >= "2020-1-1" & Data$Datahora <= "2020-3-22" )]<- "Inverno"  
  
  
freq.rel<-table(Data$Estacao\_ano)/length(Data$Estacao\_ano)  
  
barplot(freq.rel)

## Questão 4 [5 valores] Com base no cálculo de uma medida de associação, estude a relação existente entre a variável Lesoes.a.30.dias e cada uma das seguintes variáveis: Estacao\_ano e Ano.matricula. Faça o upload dos valores das medidas de associação e dos gráficos que ilustram as mesmas associações.

coords <- table(Data$`Lesões a 30 dias`)  
p<-barplot(coords, ylim=c(0,30000),main = "Frequencia",col="lightgrey")  
text(x = p, y = coords, label = coords, cex = 0.8,pos = 3, col = "blue")  
  
x<-table(Data$`Lesões a 30 dias`,Data$Estacao\_ano)  
(x\_1<-round(prop.table(x,2)\*100,1))  
  
coords <- x\_1  
p<-barplot(x\_1,beside=TRUE,legend.text = TRUE, ylim=c  
(0,100), xlab="Estação",ylab="Lesão's %", col= c("dodgerblue3","green", "blue", "red"),args.legend = list(x = "topright",inset =  
c(- 0.05, 0)))  
text(x = p, y = coords, label = coords,cex = 0.8,pos =  
1, col = "white")  
  
cramersV(Data$`Lesões a 30 dias`,Data$Estacao\_ano)  
  
ggplot(Data, aes(x=Data$`Lesões a 30 dias`, y=Data$`Ano matricula`)) +  
geom\_violin(width=1,trim=FALSE)+  
geom\_boxplot(width=0.1, color="grey",  
alpha=0.2)  
  
anova\_ <- aov (Data$`Ano matricula` ~ Data$`Lesões a 30 dias`, Data)  
(Eta\_ <- sqrt(etaSquared(anova\_ )[,1]))

## Tarefa final: Submeta, no Moddle, um ficheiro pdf resultado da compilação do TEMPLATE\_QUIZ1.

## Caso os resultados apresentados não sejam coerentes com as respostas dadas, a classificação será penalizada.