

09

Mão na massa: Modelos Preditivos

Chegou a hora de você executar o que foi visto na aula! Para isso, execute os passos listados abaixo.

1) Baixe [aqui \(https://caelum-online-public.s3.amazonaws.com/799-data-science-preparacao-e-exploracao-dos-dados/05/clima Consolidado_v01_v02.zip\)](https://caelum-online-public.s3.amazonaws.com/799-data-science-preparacao-e-exploracao-dos-dados/05/clima Consolidado_v01_v02.zip) o zip com os CSVs `clima Consolidado_v01.csv` e `clima Consolidado_v02.csv`. E baixe [aqui \(https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/05/Arquivo_Apoio_Aula_05_monta_cenarios.R\)](https://s3.amazonaws.com/caelum-online-public/799-data-science-preparacao-e-exploracao-dos-dados/05/Arquivo_Apoio_Aula_05_monta_cenarios.R) o arquivo `Arquivo_Apoio_Aula_05_monta_cenarios.R` e abra-o no RStudio.

2) Declare a função `geraCenario`, que serve para gerar um arquivo CSV contendo dados candidatos para o modelo preditivo:

```
geraCenario <- function(ent_volume, out_arq_destino)
{
  df_tipo_vs_visualiz = df_clima Consolidado %>%
    select (state:hour) %>%
    filter(!shape %in% c("Other")) %>%
    group_by (shape) %>%
    summarise(quantas = n()) %>%
    filter(quantas > ent_volume) %>%
    arrange (desc(quantas))
  df_OVNI <- df_clima Consolidado[(df_clima Consolidado$shape %in% df_tipo_vs_visualiz$shape),]

  tipos <- unique(df_tipo_vs_visualiz$shape) # tipos[1] é Light
  df_final <- data.frame()
  for (t in tipos){
    x <- df_OVNI[ sample( which( df_OVNI$shape == t ) , ent_volume ) , ]
    df_final <- rbind(df_final,x)
  }
  #rownames(df_final) <- c() # get rid of row names
  df_OVNI <- df_final
  write.csv(rbind(df_OVNI), file = out_arq_destino, row.names=FALSE)
}
```

3) Diga o seu diretório de trabalho, através do `setwd` e coloque a biblioteca `dplyr` em memória:

```
setwd("D:\\R")
library(dplyr)
```

4) Leia o arquivo `clima Consolidado_v01.csv` e faça a primeira chamada à rotina:

```
df_clima Consolidado <- read.csv(file="clima Consolidado_v01.csv", header=TRUE, sep=",")
c1 = geraCenario (500, "Curso_3_Aula_5_Teste_1.csv")
```

5) O arquivo gerado `Curso_3_Aula_5_Teste_1.csv` fica com cerca de 4500 observações. Carregue-o no **Machine Learning | Microsoft Azure** e utilize-o no experimento **Projeto Busca de OVNI**s, feito nos cursos anteriores. Por fim, execute o modelo.

6) Faça uma segunda chamada à rotina e faça o mesmo que no passo anterior, mas com o arquivo

Curso_3_Aula_5_Teste_2.csv:

```
c2 = geraCenario (2000, "Curso_3_Aula_5_Teste_2.csv")
```

7) Por fim, leia o arquivo **clima Consolidado_v02.csv** e faça a mais uma chamada à rotina:

```
df_clima_consolidado <- read.csv(file="clima Consolidado_v02.csv", header=TRUE, sep=",")  
dim(df_clima_consolidado)  
c3 = geraCenario (3500, "Curso_3_Aula_5_Teste_3.csv") # 10.500 observações (este csv será fornecido)
```

Carregue o arquivo **Curso_3_Aula_5_Teste_3.csv** no **Machine Learning | Microsoft Azure** e utilize-o no experimento **Projeto Busca de OVNIs**. Depois disso, execute o modelo.