

ATIVIDADE PARA PROCESSO SELETIVO

Analista de Ciência de Dados

Avaliamos seu currículo e gostamos bastante!! \o/

Agora precisamos que você resolva um desafio pra gente, boraaaa?

CENÁRIO

Suponha que você possui em mãos a seguinte base de dados, denominada “dados”, composta por uma covariável categórica, que assume valores “GrupoA” e “GrupoB”, e uma variável resposta que assume valores “S” ou “N”, ilustrando se um cliente contratou ou não determinado produto do Agibank:

covariavel	contratou
GrupoA	S
GrupoA	S
GrupoA	S
GrupoA	S
GrupoA	N
GrupoA	N
GrupoA	N
GrupoA	N
GrupoB	S
GrupoB	S
GrupoB	S
GrupoB	N
GrupoB	N
GrupoB	N
GrupoB	N
GrupoB	N
GrupoB	N
GrupoB	N

Você pode assumir que essa tabela seja um data.frame/tibble/data.table no R ou um pandas DataFrame no Python.

A sua tarefa é criar uma função denominada “minha_funcao” que retorne o grupo, a quantidade de registros referentes à cada grupo e a taxa de determinado valor da variável resposta para cada grupo. A função deve possuir, no mínimo, dois argumentos sendo um deles a base de dados e um segundo sendo uma string referente ao status de contrato.

De maneira ilustrativa:

- 1) Rodando o código `minha_funcao(dados, status_resposta = 'S')`, a função deve retornar:

covariavel	n	taxa resposta
GrupoA	8	0,5
GrupoB	9	0,3

- 2) Rodando o código `minha_funcao(dados, status_resposta = 'N')`, a função deve retornar:

covariavel	n	taxa resposta
GrupoA	8	0,5
GrupoB	9	0,7

APRESENTAÇÃO DO CASE

Para python: facultativamente, você pode usar pandas, numpy e base Python ≥ 3.0

Para R: facultativamente, você pode usar bibliotecas do tidyverse, R base ou data.table.

Não esqueça que essa é a oportunidade para você mostrar que é capaz para assumir esse desafio! Arrasa!!

Você precisará encaminhar o trabalho desenvolvido para o e-mail gisele.antunes@agibank.com.br no prazo combinado no corpo do e-mail!

Boa sorte!!