

FIA/P GRADUAÇÃO

# 2TBDR

Enterprise Analytics e Big Data

Diógenes Justo  
[profDiogenes.Justo@fiap.com.br](mailto:profDiogenes.Justo@fiap.com.br)

# ***2021 - 1.o Semestre Atividade complementar***

## ***Atividade Complementar 1***

Instruções:

1. Acesse o seguinte repositório:  
[https://github.com/diogenesjusto/FIAP/tree/master/Gradua%C3%A7%C3%A3o/2021/S1/Checkpoint\\_Atividade\\_Complementar](https://github.com/diogenesjusto/FIAP/tree/master/Gradua%C3%A7%C3%A3o/2021/S1/Checkpoint_Atividade_Complementar)
2. Abra o script em R (Script\_base.R)
3. Escolha uma loja e filtre o dataset
4. Você terá, inicialmente, um script para modelar a variável (coluna) GMV que é a venda total
5. Queremos escolher o melhor modelo que consiga prever o GMV com o menor erro possível (variável `erro_med_prev`)
6. Você deve testar algumas variáveis disponíveis para melhorar o resultado, e preencher o quadro anexo.

## ***Atividade Complementar 1***

Parte 1: melhorando o resultado do modelo através de variáveis exógenas disponíveis.

Quadro de tentativas:

Variáveis Exógenas utilizadas	P-valor está abaixo de 0,05?	R <sup>2</sup>	Std Error	Erro Méd Prev

## ***Atividade Complementar 1***

### Parte 2: uso de variáveis externas

1. Vamos agora tentar melhorar o resultado do modelo incorporando alguma variável externa.
2. Você pode escolher uma variável, qualquer (índices econômicos, preço de gasolina, etc)
3. Não é necessário reconstruir a roda - lembre que fizemos um exercício anterior e vocês já utilizaram variáveis externas para modelar o PIB. Podem utilizar estas mesmas.

## ***Atividade Complementar 1***

Parte 2: melhorando o resultado do modelo através de novas variáveis externas.

Quadro de tentativas (pelo menos 1 variável externa)

Variáveis Exógenas utilizadas	P-valor está abaixo de 0,05?	R <sup>2</sup>	Std Error	Erro Méd Prev

## ***Atividade Complementar 1***

Parte 3: uso de um modelo "auto-explicativo"

1. Vamos agora tentar melhorar o resultado do modelo utilizando uma técnica chamada "modelo auto regressivo"
2. Em um modelo autoregressivo utilizamos a própria variável "alvo" como "predictor", porém deslocada no tempo
3. Utilize o script "AR.R" como exemplo. Você terá o GMV deslocado 1 período no tempo (GMV1), 2 períodos (GMV2), e assim por diante. Você poderá utilizar quantos quiser (GMV1+GMV2, GMV1+GMV3+GMV4, etc). Faça alguns testes e registre ao menos 5 tentativas de combinações diferentes.



## ***Atividade Complementar 1***

Parte 3: melhorando o resultado do modelo através de um modelo autoregressivo.

Quadro de tentativas (pelo 5 tentativas)

Variáveis Exógenas utilizadas	P-valor está abaixo de 0,05?	R <sup>2</sup>	Std Error	Erro Méd Prev

## ***Atividade Complementar 1***

Publicação de resultados (individual):

Publique:

- 1) os 3 quadros no formato preferido
- 2) código fonte (pode ser R ou Python)
- 3) Texto explicativo com ao menos 3 parágrafos com conclusões



The image features a solid pink background. In the center, there is a rectangular area with a dark, blurred image of a circuit board. The circuit board has various components, including a chip labeled '11C220T'.

# **TBD**

## **Tecnologia em Banco de Dados**

**Copyright © 2021 Prof. Diógenes Justo**

**Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do professor (autor).**