

**PREDIÇÃO DA ARRECADAÇÃO DE TRIBUTOS:
UM COMPARATIVO ENTRE O *FACEBOOK™ PROPHET* E AS
REDES NEURASIS *LONG SHORT-TERM MEMORY***

Rafael Santos Carvalho



PUC Minas

CONTEXTUALIZAÇÃO

- Predição de receitas como maneira de evitar infrações à Lei de Responsabilidade Fiscal
- Predição de arrecadação como ferramenta de gestão, permitindo o cumprimento de metas fiscais e o equilíbrio entre receitas e despesas

PROBLEMA PROPOSTO

- Ferramentas atuais preocupam-se com a predição de receitas a médio/longo prazo
- A predição a curto ou curtíssimo prazo viabiliza a execução/suspensão/cancelamento de despesas, principalmente as discricionárias

PROBLEMA PROPOSTO

- Uso do estado do Rio Grande do Sul como exemplo
- O uso dos tributos arrecadados por um estado específico como exemplo não limita o estudo, que pode ser aplicado a tributos de outros estados, municípios ou até mesmo da União

COLETA DE DADOS

- Arrecadação: Receita Dados RS (arquivo CSV)
- IGP-M: IpeaData (webscrapping)
- PIB mensal Brasil: IpeaData (arquivo CSV)
- PIB trimestral RS: IpeaData (arquivo XLS)
- Dados Caged: IpeaData (arquivo CSV)

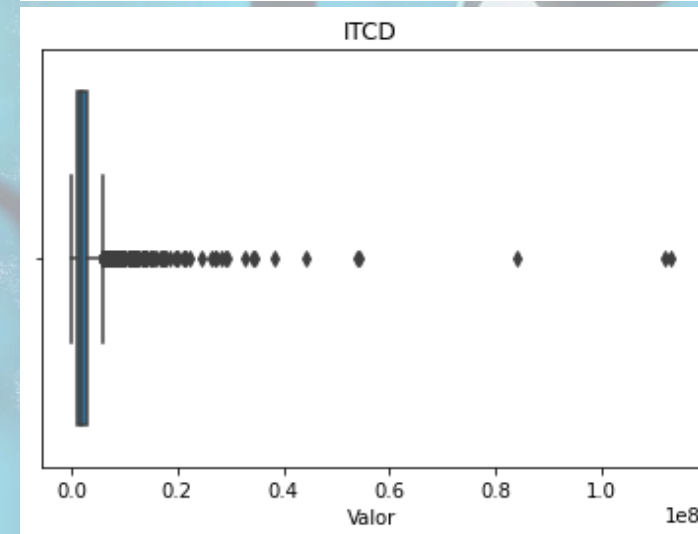
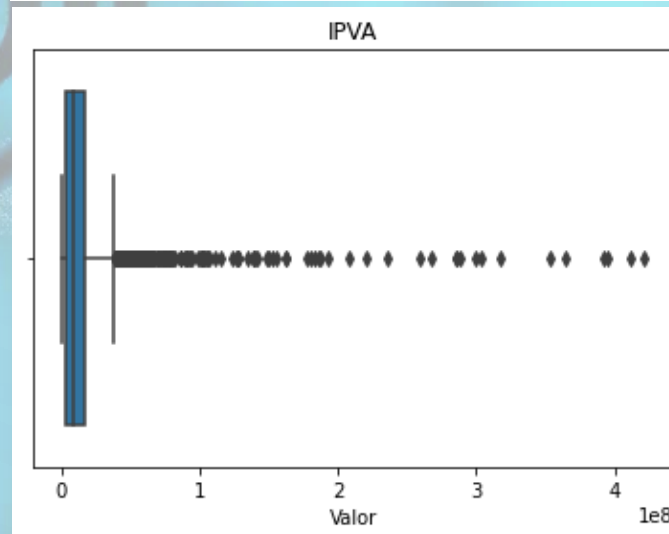
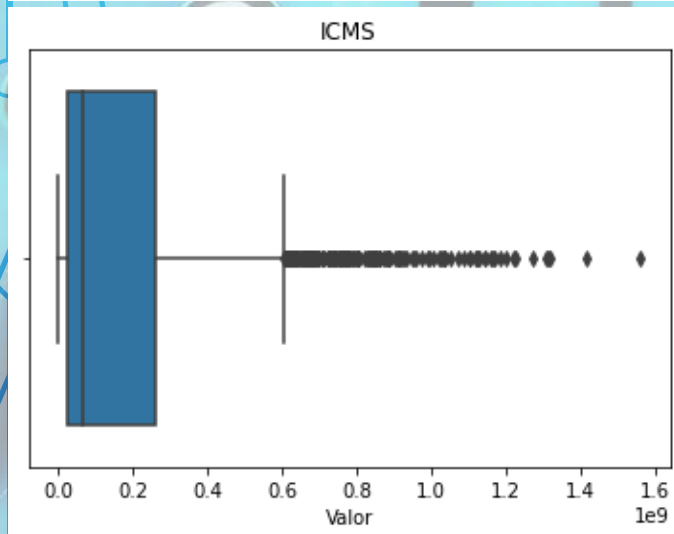
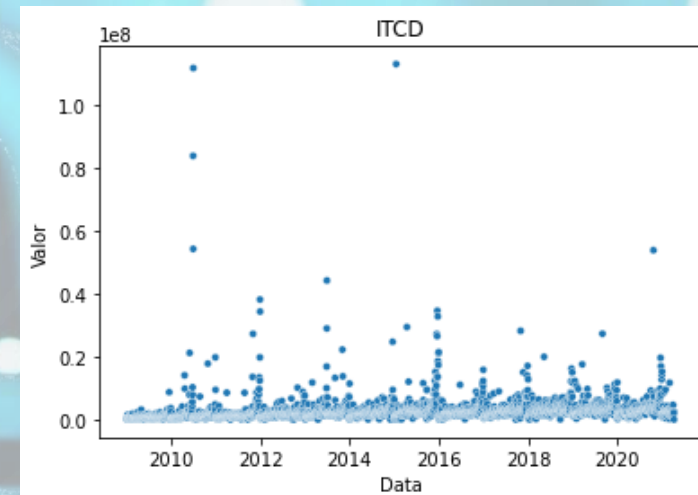
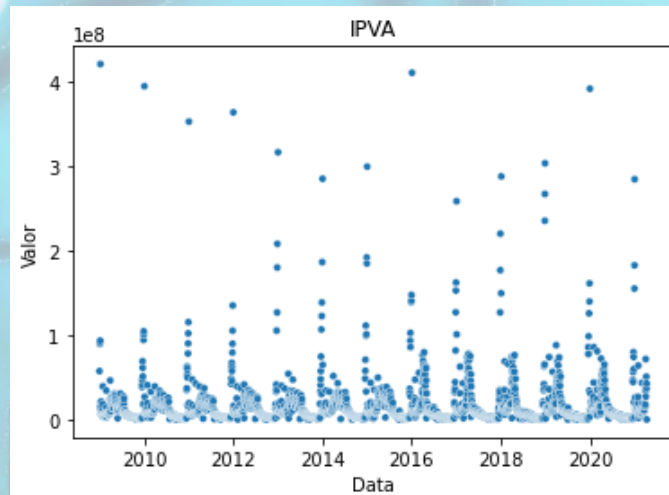
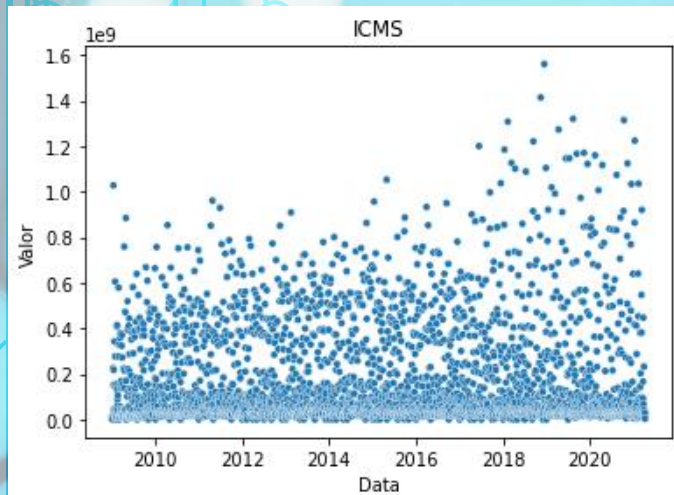
TRATAMENTO DE DADOS

- Ajuste dos valores monetários a valor presente pelo IGP-M (quando os dados coletados não estiverem ajustados)
- Preenchimento de valores NaN nos datasets do Caged

I ANÁLISE E EXPLORAÇÃO DOS DADOS

- ICMS é o tributo com maior variância e com pouca sazonalidade
- IPVA é o tributo com maior sazonalidade, principalmente nos primeiros meses do ano
- ITCD possui menor variância e sazonalidade

ANÁLISE E EXPLORAÇÃO DOS DADOS



MODELOS PREDITIVOS DE ARRECADAÇÃO

- Facebook™ Prophet
- Redes neurais Long Short-Term Memory (LSTM)
- Estudo dividido em duas partes: (1) utilizando apenas a arrecadação passada para predição da arrecadação futura e (2) utilizando outras variáveis preditoras (*features*) para predição da arrecadação futura

MODELOS COM APENAS UMA VARIÁVEL QUANTITATIVA PREDITORA

Prophet

- *Paper Forecasting at Scale* que originou o Prophet recomenda a remoção de *outliers*
- Mesmo seguindo a recomendação do *paper*, modelo sem remoção de *outliers* performou melhor que modelo com remoção de *outliers*
- Modelo de implementação mais fácil que a rede neural LSTM

MODELOS COM APENAS UMA VARIÁVEL QUANTITATIVA PREDITORA

LSTM

- Necessidade de padronização dos dados
- Dados padronizados com o PowerTransformer (método *yeo-Johnson*) resultaram em menor erro que os padronizados com StandardScaler e RobustScaler
- Uso de Embedding em detrimento ao One-Hot Encoder para dados categóricos (datas)

MODELOS COM APENAS UMA VARIÁVEL QUANTITATIVA PREDITORA

LSTM

- Uso do Keras Model subclassing em detrimento do Sequential API
- Keras Model subclassing permite a inserção de *inputs* com quantidades distintas de dimensões (datas e séries temporais de arrecadação)

MODELOS COM APENAS UMA VARIÁVEL QUANTITATIVA PREDITORA

- LSTM foi superior ao Prophet em todos os três tributos
- Superioridade das redes neurais LSTM foi mais significativa nos tributos de maior variância
- No ITCD a superioridade foi praticamente insignificante (pouco mais de 1%)

I MODELOS COM MÚLTIPLAS VARIÁVEIS QUANTITATIVAS PREDITORAS

- Análise restrita ao ICMS pelo fato de haver mais dados público sobre variáveis preditoras
- Novas variáveis preditoras: PIB mensal do Brasil, PIB trimestral do RS, admissões e demissões mensais (Caged)
- Modelo poderia ser melhorado com adição de novas variáveis não-públicas como emissões de notas fiscais eletrônicas e dados sobre inadimplência

I MODELOS COM MÚLTIPLAS VARIÁVEIS QUANTITATIVAS PREDITORAS

- Redes neurais LSTM, assim como no modelo com apenas uma variável quantitativa, teve melhor resultado que o Prophet
- LSTM teve maior redução do erro no modelo multivariado em relação ao modelo multivariado
- A inserção de informações de séries temporais como *inputs* do Prophet reduziu o erro em 7%, mas não foi o suficiente para ter performance superior à LSTM

I CONCLUSÃO

- LSTM tem, em geral, melhor performance que o Prophet
- O fato de ter pior performance não significa que o Prophet deve ser preterido em todos os usos
- A escolha do modelo ideal para cada série de dados cabe ao Cientista de Dados, que analisará os requisitos e frequência de atualização das séries temporais