

# Fontes de Dados e Metodologia de Geração de Dados Sintéticos

A análise de otimização de ofertas em jogos Free-to-Play (F2P) exige o uso de dados transacionais confidenciais (comportamento de gasto individual, LTV). Por razões de **privacidade, segurança e confidencialidade**, o acesso a bases de dados reais é inviável para um projeto acadêmico.

Portanto, a espinha dorsal deste projeto é o uso de **Dados Sintéticos**, seguindo a metodologia de **Prova de Conceito (PoC)**. O objetivo é demonstrar a capacidade de aplicar técnicas robustas de análise e modelagem preditiva, imitando a realidade do mercado.

## 1. Metodologia de Geração de Dados Sintéticos

O conjunto de dados (dados\_jogadores\_sinteticos.csv), composto por \$5000\$ registros, foi gerado via **Script Python** (utilizando as bibliotecas Pandas e NumPy), sendo baseado em premissas de negócio observáveis no setor de jogos.

Variável-Chave	Tipo de Simulação	Premissa de Negócio Refletida
Gasto_Total_BRL	Distribuição Lognormal e Zero-Inflacionada	Simulação da <b>Regra de Pareto (80/20)</b> : apenas $\approx 15\%$ dos jogadores são pagantes, e o gasto é altamente desigual para criar "Baleias" (clientes de alto valor).
COMPROU_OFERTA_X	Distribuição Binomial Ponderada	A <b>Propensão a Gastar</b> foi correlacionada com variáveis de engajamento e histórico de gasto, garantindo que o modelo preditivo encontre relações válidas.
Dias_Desde_Instalacao	Distribuição Exponencial	Simulação do <b>Abandono (Churn)</b> : a maioria dos jogadores abandona o jogo rapidamente, e poucos permanecem por longos períodos.

## 2. Fontes de Dados Secundárias (Enriquecimento Contextual)

Em um cenário real, o dataset sintético seria enriquecido com informações públicas para validar e aprofundar a análise.

Fonte Pública	Método de Coleta	Aplicação Estratégica
Relatórios de Mercado (Newzoo, etc.)	Download Direto	Validação das premissas de LTV e tendências de monetização usadas na simulação.
Análise de Reviews (Steam, App Stores)	Web Scraping / APIs	Identificação de insights sobre o design do jogo ou pontos de atrito que levariam a compras de "solução" (Ex: Oferta X).
APIs de Plataformas de Streaming	APIs Públicas (Twitch, YouTube Gaming)	Correlação entre o consumo de conteúdo externo do jogo e picos de engajamento/compra de itens cosméticos.

## 3. Conclusão da Metodologia

A abordagem de dados sintéticos demonstra a capacidade de construir uma metodologia de otimização robusta, provando que o processo de **Modelagem Preditiva e Segmentação** é viável e eficaz para resolver a problemática da baixa conversão, mesmo sem acesso a dados reais e confidenciais.