



RELATÓRIO TÉCNICO - SIMULADOR FINAL DE ANÁLISE DE DADOS

Nomes: Maryllian de Albuquerque Vieira

Rafaela Santos

Título: Análise do Panorama EJA e Indicadores de Analfabetismo no Brasil

Data de Entrega: 14/11/2025

Base de Dados: Panorama EJA

I. Contexto e Objetivo

O presente relatório tem como objetivo analisar o panorama da Educação de Jovens e Adultos (EJA) no Brasil, com base em dados municipais referentes ao Censo IBGE de 2022 e adesões aos programas de alfabetização (PBA/SBA). A investigação busca compreender os padrões de analfabetismo entre diferentes faixas etárias e regiões, avaliando também a relação entre a oferta de EJA e a incidência do analfabetismo.

II. Pergunta Central

Quais regiões e faixas etárias apresentam os maiores índices de analfabetismo, e como a adesão aos programas EJA e PBA influencia a redução desses índices?

III. Fontes de Dados e LGPD

Os dados utilizados neste estudo foram obtidos a partir da base 'panorama-da-eja-no-brasil.xlsx', composta por informações públicas do IBGE (Censo 2022) e registros de adesão municipal a programas de alfabetização. A base contém variáveis demográficas, educacionais e de adesão institucional.

Os dados são públicos e anonimizados, respeitando integralmente os princípios da LGPD, sem a presença de informações pessoais identificáveis e utilizadas exclusivamente para fins educacionais.

IV. Arquitetura e Ferramentas Utilizadas

O projeto foi desenvolvido em um ambiente Databricks integrado ao GitHub e conectado ao Power BI para visualização final dos resultados. A estrutura de dados foi organizada segundo o modelo de camadas Medallion Architecture, permitindo rastreabilidade e qualidade ao longo do pipeline.

Setup inicial no Databricks:

- Criação de **schemas (databases)** para as três camadas do projeto (*projeto_final_bronze*, *projeto_final_silver*, *projeto_final_gold*) por meio do notebook *create catalog.ipynb*.
- Estruturação de pastas correspondentes às camadas:
 - **Bronze** → ingestão dos dados brutos, com pequenas transformações aplicadas apenas para adequar a base ao formato Delta (*panorama_eja_ingestion.ipynb*);
 - **Silver** → Limpeza, padronização e enriquecimento de variáveis (*transform_ibge.ipynb*);
 - **Gold** → Consolidação e análise final dos dados (*analysis_panorama_eja.ipynb*).

Após o processamento, os dados da **camada Gold** foram importados para o **Power BI**, onde foram criados dashboards interativos para análise visual e apresentação de indicadores.

V. Processo de Análise de Dados

O processo analítico foi conduzido de forma estruturada, integrando boas práticas de ETL, análise exploratória e visualização:

1. Limpeza e Transformação (ETL)

Na camada Silver, os dados passaram por um processo de limpeza, padronização e transformação para garantir qualidade, consistência e integridade. As principais atividades realizadas foram:

- Remoção de registros inconsistentes e linhas totalmente nulas, além de eliminação de duplicatas;
- Padronização de nomes de colunas, tipos de dados e formatação numérica, incluindo remoção de acentos em nomes de estados (UF);
- Criação de campos derivados, como:
- Matrículas totais por nível de ensino (fundamental e médio);
- Matrículas por localização (urbana x rural);
- Número total de docentes por nível;
- Ano de adesão aos programas EJA e PBA;
- Padronização de indicadores de adesão aos programas EJA e PBA (convertendo valores “Sim”/“Não” em binário);
- Consolidação das tabelas e exportação como Delta Table no schema Silver, garantindo integridade e qualidade dos dados para análises posteriores.

2. Análise Exploratória (EDA)

- Avaliação descritiva de variáveis-chave: taxa de analfabetismo, matrículas EJA e adesão a programas federais;
- Cálculo de médias regionais e estaduais;
- Identificação de correlações entre oferta de EJA, população não alfabetizada e faixas etárias;
- Análise comparativa entre municípios com e sem adesão ao PBA.

VII. Insights e Recomendações

Principais Insights:

1. O Nordeste apresenta as maiores taxas médias de analfabetismo.
2. O analfabetismo cresce significativamente nas faixas etárias acima de 55 anos.
3. Municípios com maior número de matrículas EJA tendem a apresentar taxas de analfabetismo ligeiramente menores.
4. A adesão ao Programa Brasil Alfabetizado (PBA) está associada a municípios com esforços estruturados de alfabetização.

VIII. Recomendações

- Intensificar políticas públicas voltadas às faixas etárias adultas e idosas.
- Fortalecer o financiamento de programas municipais de EJA em regiões críticas.
- Utilizar dados educacionais e socioeconômicos integrados para identificar territórios prioritários.
- Monitorar a efetividade das adesões ao PBA por meio de indicadores anuais.
- Promover campanhas de sensibilização sobre o retorno à educação de adultos.

IV. Conclusões

A análise do Panorama EJA reforça a importância da integração entre dados demográficos e educacionais para a formulação de políticas públicas mais direcionadas. Durante o processo, observou-se a relevância da limpeza dos dados e da escolha de indicadores representativos. Para estudos futuros, recomenda-se incluir dados de desempenho e evasão escolar para avaliar o impacto longitudinal dos programas de alfabetização.

X. Referência

- Instituto Brasileiro de Geografia e Estatística (IBGE) – Censo Demográfico 2022. Dados acessados via Base dos Dados.
Link de Acesso: <https://basedosdados.org>
- Ministério da Educação (MEC) – Panorama da EJA no Brasil.
Link de Acesso:
<https://www.gov.br/mec/pt-br/pacto-eja/panorama-da-eja-no-brasil>
- Ministério da Educação (MEC) – Pacto Nacional pela Superação do Analfabetismo e Qualificação da EJA.
Link de Acesso: <https://www.gov.br/mec/pt-br/pacto-eja>