

Uma proposta de ETL e visualização para os dados do Censo Escolar

Aluno: Rafael Júnior Ribeiro

Orientador: Profa. Dra. Renata Galante

Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

Desenvolvimento

Demonstração

Experimentos

Conclusões e Trabalhos futuros

Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

Desenvolvimento

Demonstração

Experimentos

Conclusões e Trabalhos futuros

Exemplos de sucesso para Dados abertos no BR



Portal da Transparência Federal

Disponibiliza dados sobre orçamento, gastos, licitações e contratos. Frequentemente aparece na mídia como material para denúncias de mau uso do dinheiro público



TerraBrasilis

TerraBrasilis

Ferramenta online do INPE que disponibiliza dados abertos sobre desmatamento na Amazônia, permitindo que a sociedade possa monitorar e fiscalizar a ação de desmatadores ilegais

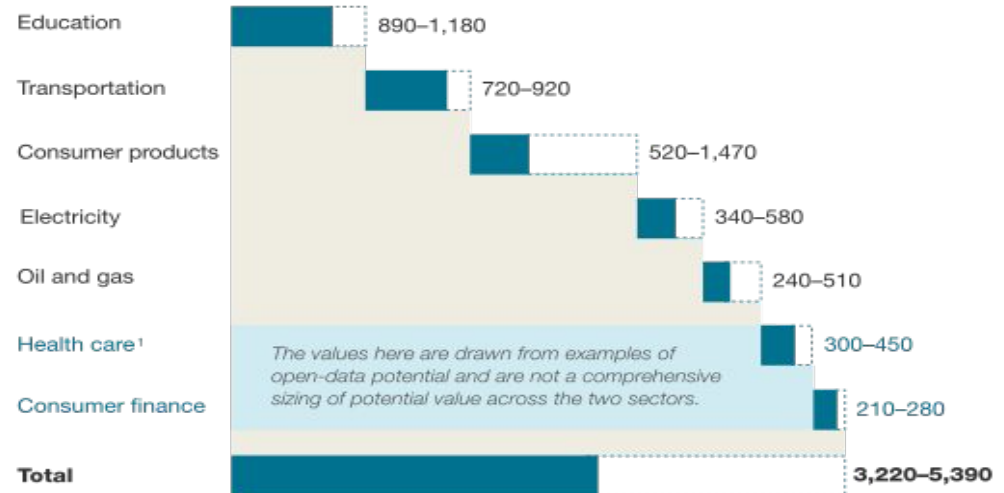
CORONAVÍRUS // BRASIL

Painel Coronavirus

Plataforma online que disponibiliza informações sobre a evolução da doença no Brasil, onde é possível entender de forma interativa seu andamento e prever o seu futuro, orientando políticas públicas que minimizem o impacto da doença

Potencial Econômico (2013)

Potential value in open data, \$ billion



¹Includes US values only.

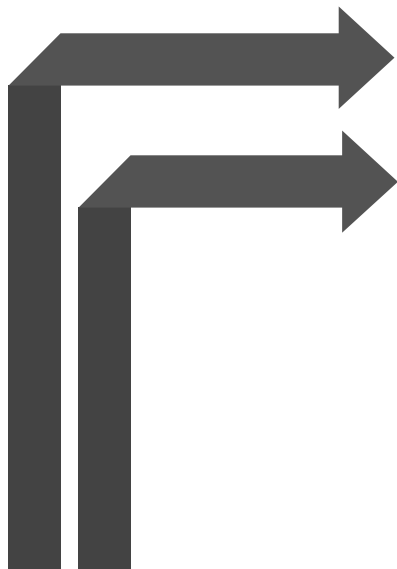
Source: McKinsey Global Institute analysis

Censo Escolar



Principal instrumento de coleta de informações da educação básica e a mais importante pesquisa estatística educacional brasileira

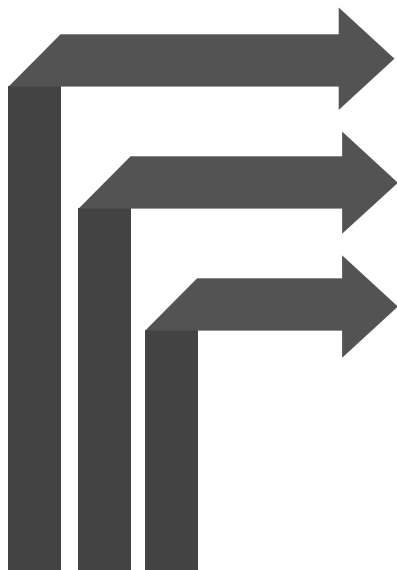
Censo Escolar



Principal instrumento de coleta de informações da educação básica e a mais importante pesquisa estatística educacional brasileira

Coordenado pelo Inep e realizado em regime de colaboração entre as secretarias estaduais e municipais de educação e com a participação de todas as escolas públicas e privadas do país

Censo Escolar



Principal instrumento de coleta de informações da educação básica e a mais importante pesquisa estatística educacional brasileira

Coordenado pelo Inep e realizado em regime de colaboração entre as secretarias estaduais e municipais de educação e com a participação de todas as escolas públicas e privadas do país

Ensino Regular (educação infantil, ensino fundamental e médio), Educação especial, EJA e Educação profissional

Censo Escolar



Principal instrumento de coleta de informações da educação básica e a mais importante pesquisa estatística educacional brasileira

Coordenado pelo Inep e realizado em regime de colaboração entre as secretarias estaduais e municipais de educação e com a participação de todas as escolas públicas e privadas do país

Ensino Regular (educação infantil, ensino fundamental e médio), Educação especial, EJA e Educação profissional

Ferramenta essencial que possibilita aos agentes da educação compreender a realidade educacional do país, das unidades federativas, municípios e Distrito Federal, permitindo um acompanhamento mais efetivo da eficácia das políticas públicas

Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

Desenvolvimento

Demonstração

Experimentos

Conclusões e Trabalhos futuros

Objetivos

Extração dos
microdados e
indicadores do
Censo Escolar a
partir do portal do
INEP



Objetivos

Extração dos
microdados e
indicadores do
Censo Escolar a
partir do portal do
INEP



Transformação dos
dados, adequando
cada coluna ao tipo
mais eficiente, lidando
com dados faltantes e
desnormalizando

Objetivos

Extração dos microdados e indicadores do Censo Escolar a partir do portal do INEP

Carga dos dados transformados em um formato de armazenamento econômico, mas também eficiente para consultas

Transformação dos dados, adequando cada coluna ao tipo mais eficiente, lidando com dados faltantes e desnormalizando



Objetivos

Extração dos microdados e indicadores do Censo Escolar a partir do portal do INEP

Carga dos dados transformados em um formato de armazenamento econômico, mas também eficiente para consultas

Transformação dos dados, adequando cada coluna ao tipo mais eficiente, lidando com dados faltantes e desnormalizando

Visualização dos dados usando gráficos interativos através de uma plataforma Web

Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

Desenvolvimento

Demonstração

Experimentos

Conclusões e Trabalhos futuros

Trabalhos relacionados

	Objetivo	Dados utilizados	Metodologia	Resultados
Determinantes da repetência escolar no Brasil (Oliveira e Soares, 2012)	Determinantes da repetência no Ensino Fundamental	Censos Escolares entre 2007 e 2010	Usaram Logit para relacionar as variáveis do Censo no ano X com a probabilidade de repetir em X+1	Meninos, PNE, usuários de transporte escolar e estudantes de escolas municipais apresentam maior chance de repetência

Trabalhos relacionados

	Objetivo	Dados utilizados	Metodologia	Resultados
O impacto do programa bolsa família sobre a repetência (Oliveira e Soares, 2013)	Mensurar impacto do Bolsa Família na repetição durante o Ensino Fundamental.	CadÚnico, Projeto Frequência e Censo Escolar do ano 2008	Dividiram em 3 modelos de regressão logística, cada usando uma combinação das 3 bases	Alunos do BF tem 11% menos chance de repetir. Meninos, PNE, estudantes de escolas estaduais ou municipais e com família menos escolarizada possuem mais chance de repetir

Trabalhos relacionados

	Objetivo	Dados utilizados	Metodologia	Resultados
Indicadores que influenciam na qualidade da educação da Paraíba (Silva et al., 2022)	Demonstrar as limitações de usar unicamente o Ideb para medir a qualidade da educação básica	Dados de 300 escolas paraibanas no Censo Escolar de 2019	Utilizaram modelos lineares generalizados (GLM) para relacionar o ideb com os demais indicadores educacionais do Censo	Média de horas aula diárias e complexidade de gestão escolar tem um efeito positivo em relação ao Ideb, enquanto que distorção idade-série tem efeito negativo

Trabalhos relacionados

	Objetivo	Dados utilizados	Metodologia	Resultados
Modelagem e Visualização Científica de Dados Educacionais (Barros et al., 2017)	Investigar técnicas de visualização de dados junto de estatísticas e de aprendizagem de máquina, para potencializar a interpretação de dados educacionais	Médias finais de alunos de Licenciatura em Espanhol e Técnico de Segurança do Trabalho do IFRN no período de 2008 a 2016	Plot boxplot, violino e Q-Q para identificar distribuição bimodal, para então usar K-Means para classificar os dados em grupos e então usar novamente Q-Q nos dados agrupados	Violinplot é escolha certa de gráfico para distribuição bimodal. Q-Q demonstrou a normalidade nas distribuições agrupadas após o K-Means, que separou os alunos entre 2 grupos: aprovados e reprovados

Trabalhos relacionados

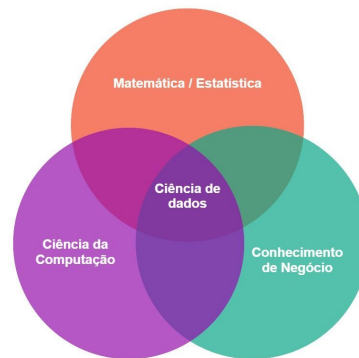
	Objetivo	Dados utilizados	Metodologia	Resultados
Dados Abertos Educacionais Brasileiros: Um Mapeamento Sistemático da Literatura (Ferreira et al., 2021)	Mapeamento sistemático de artigos científicos que abordam dados abertos	Artigos acadêmicos sobre dados abertos entre 2021 e Julho de 2021	Removido manualmente artigos que não são sobre a educação BR ou que não estão em PT ou EN, após isso foram catalogados os dados principais de cada artigo	2019 ano com mais artigos. Enem, Enade, Prova Brasil e Censo são os dados mais estudados. Weka e R ferramentas mais utilizadas, com J48 e Naive Bayes sendo os algoritmos de destaque

Semelhanças

Educação brasileira



Ciência de Dados



Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

Desenvolvimento

Demonstração

Experimentos

Conclusões e Trabalhos futuros

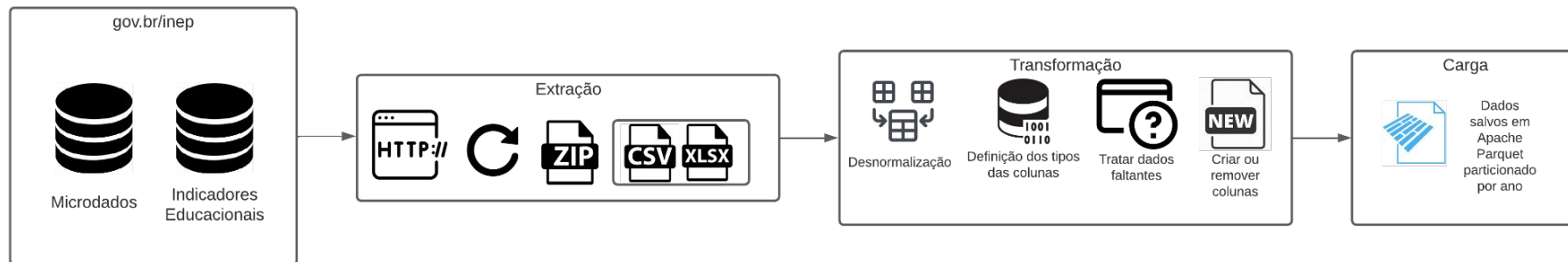
Principal ferramenta

A linguagem de programação Python foi utilizada em todas as etapas de trabalho. DuckDB e Streamlit são bibliotecas de Python.

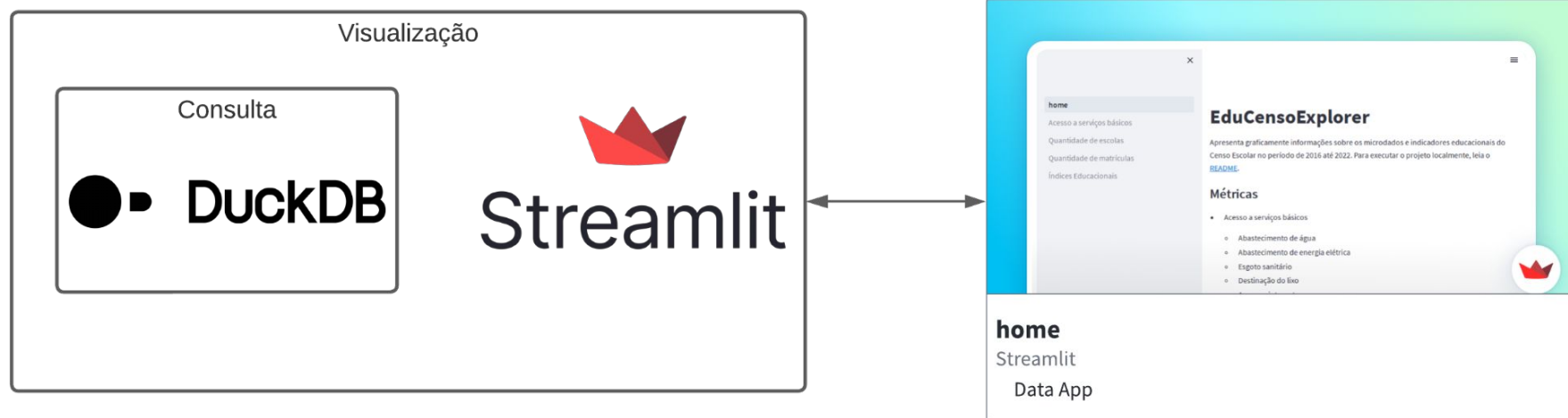
Código fonte disponível em
github.com/rjribeiro/EduCensoExplorer



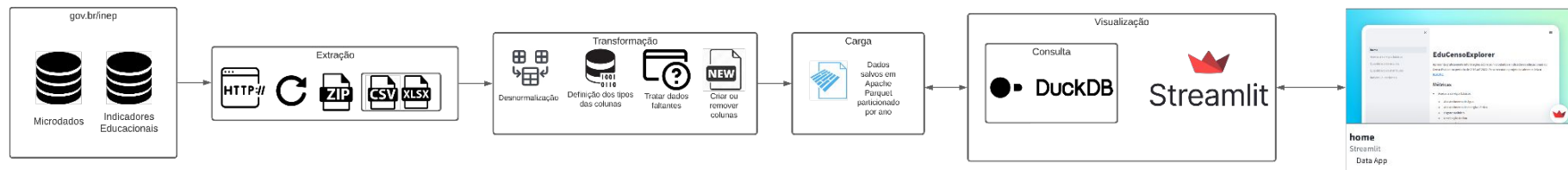
ETL



Visualização



Visão Geral



Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

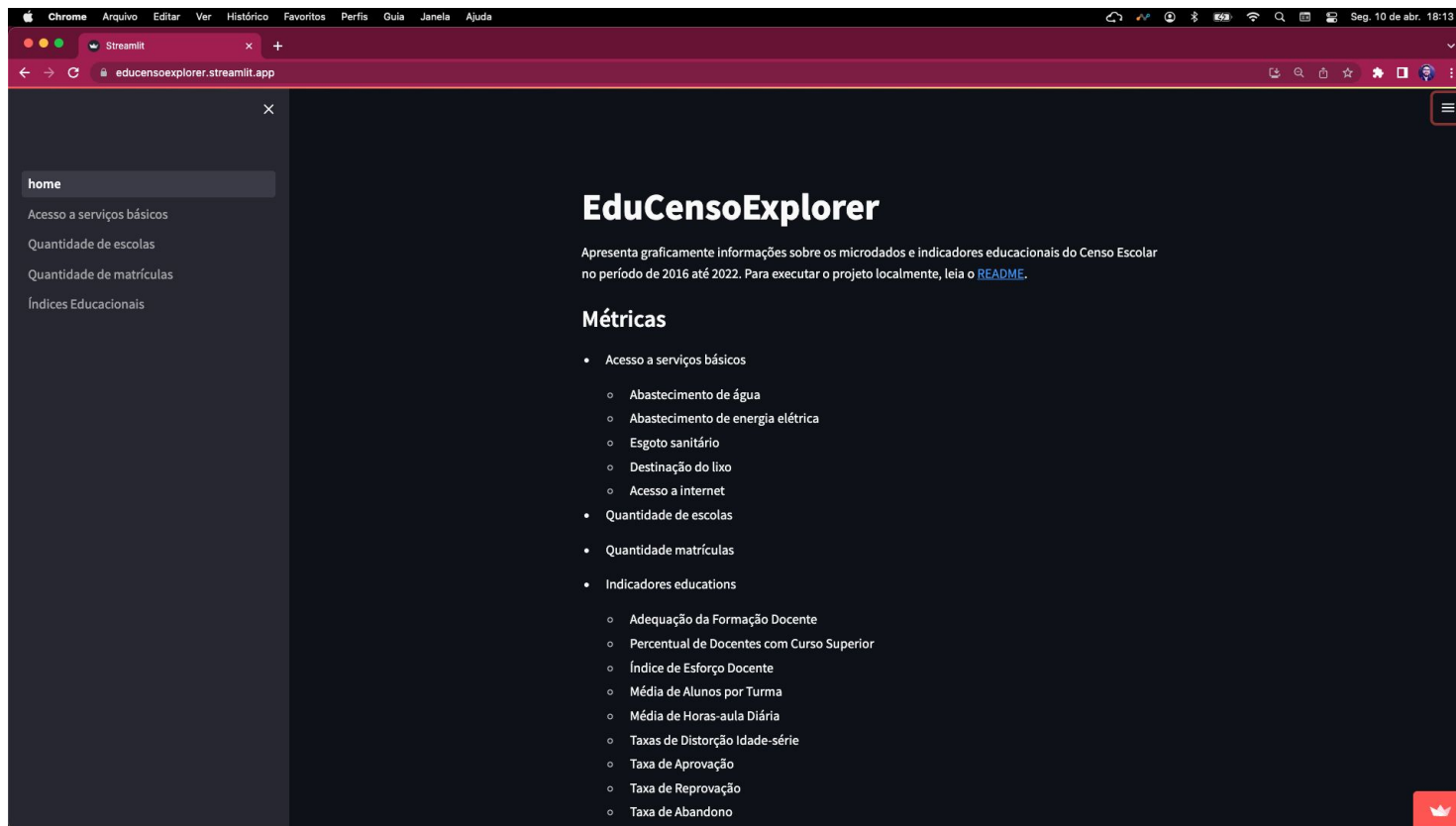
Desenvolvimento

Demonstração

Experimentos

Conclusões e Trabalhos futuros

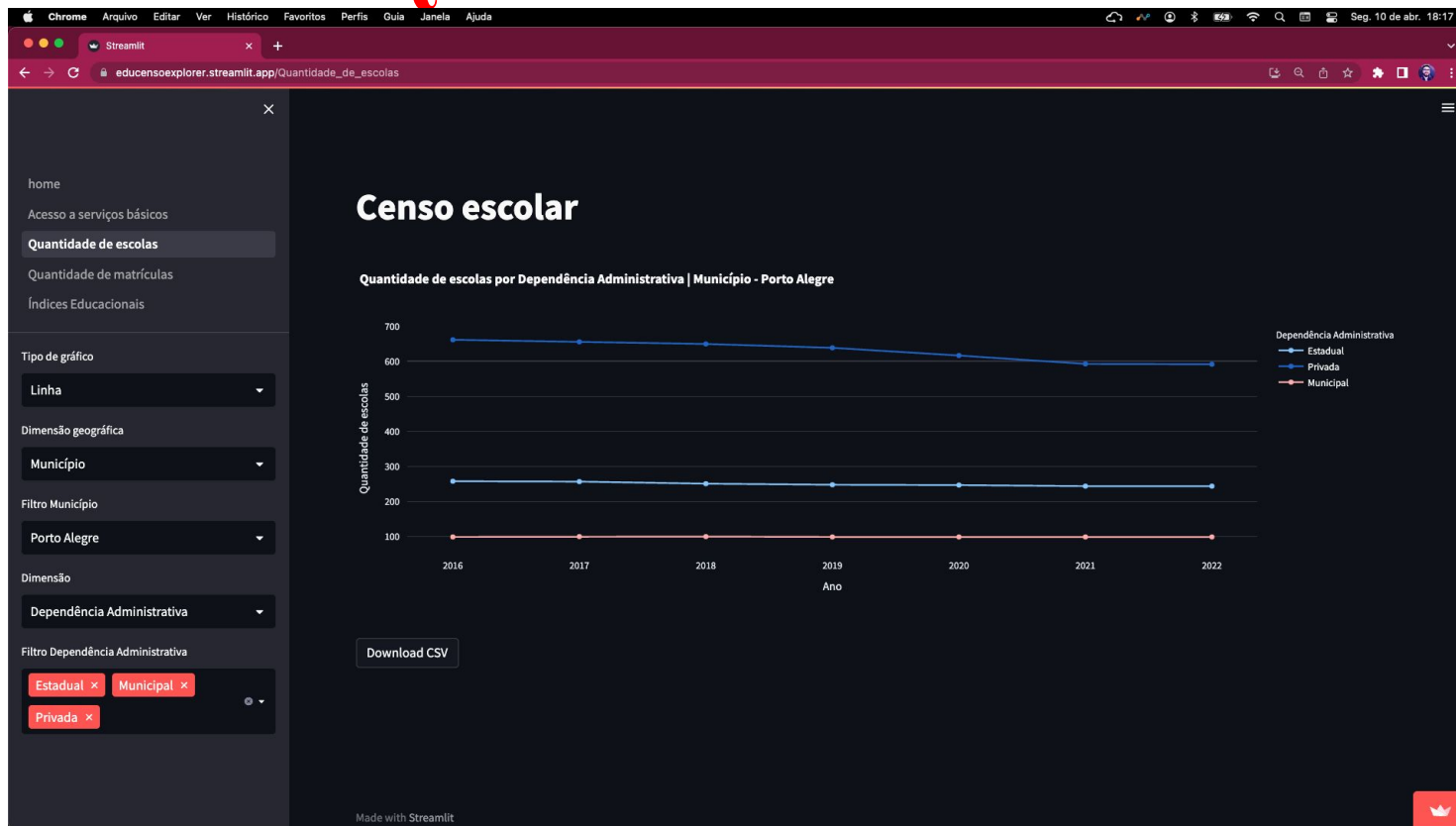
Tela inicial



Acesso a serviços básicos



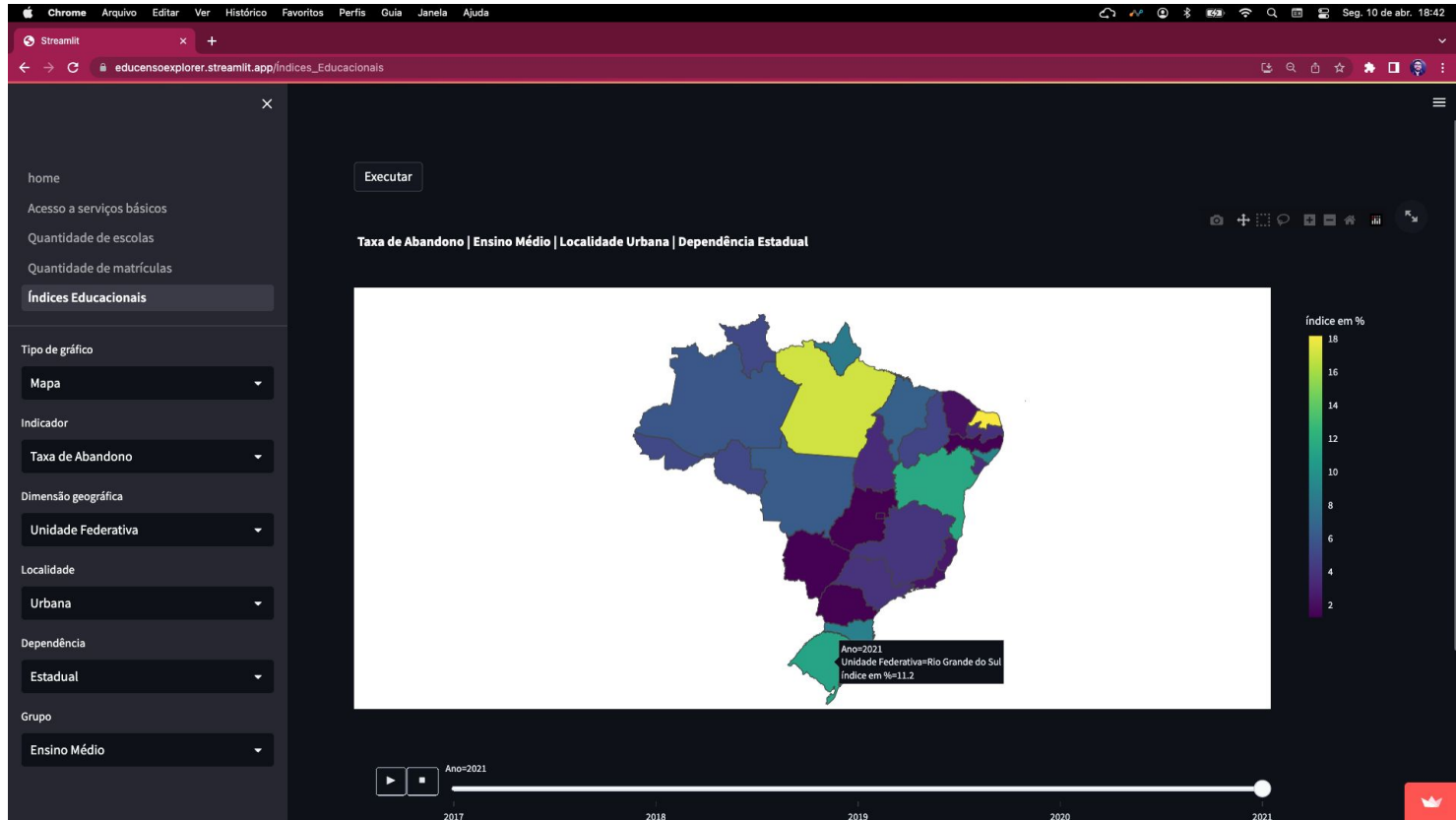
Quantidade de escolas



Quantidade de matrículas



Indicadores Educacionais



Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

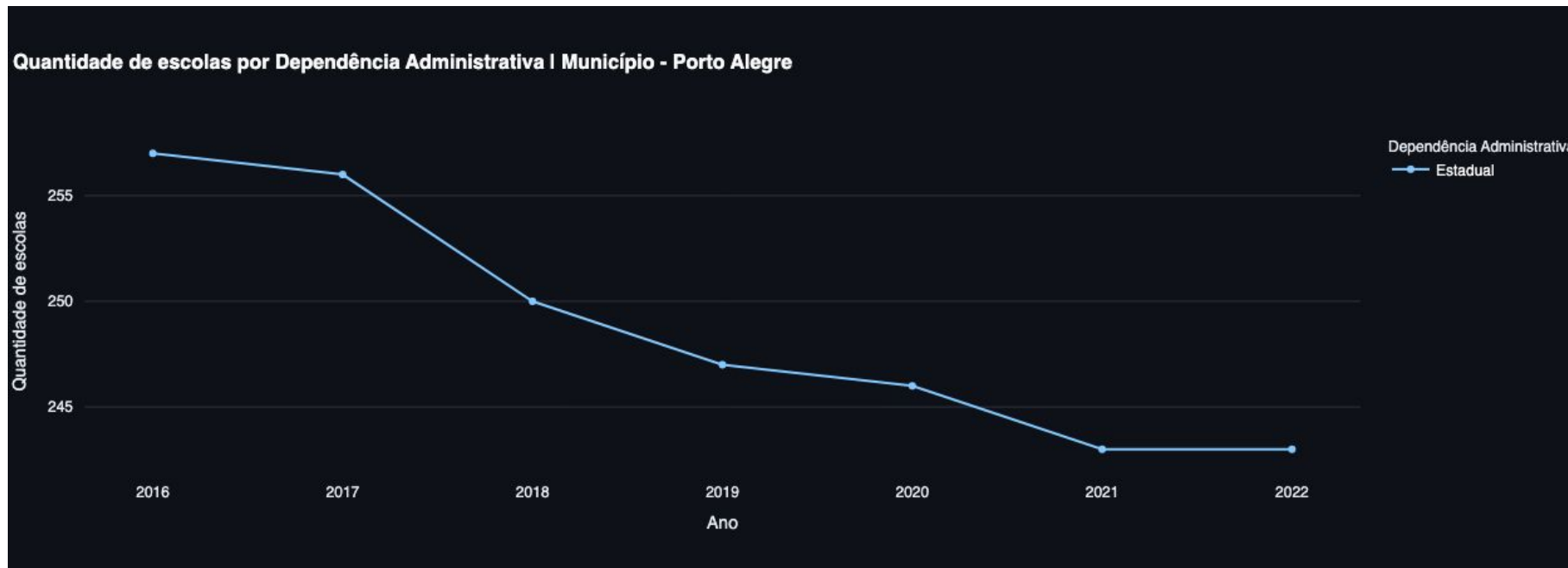
Desenvolvimento

Demonstração

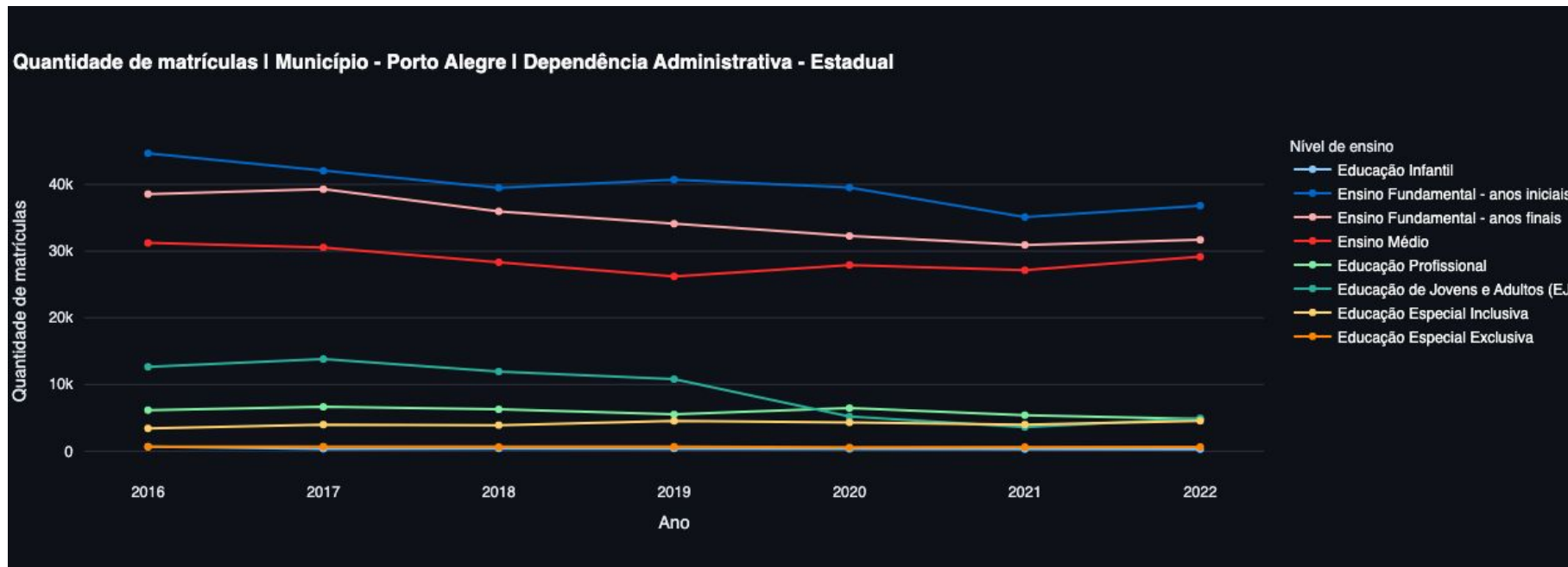
Experimentos

Conclusões e Trabalhos futuros

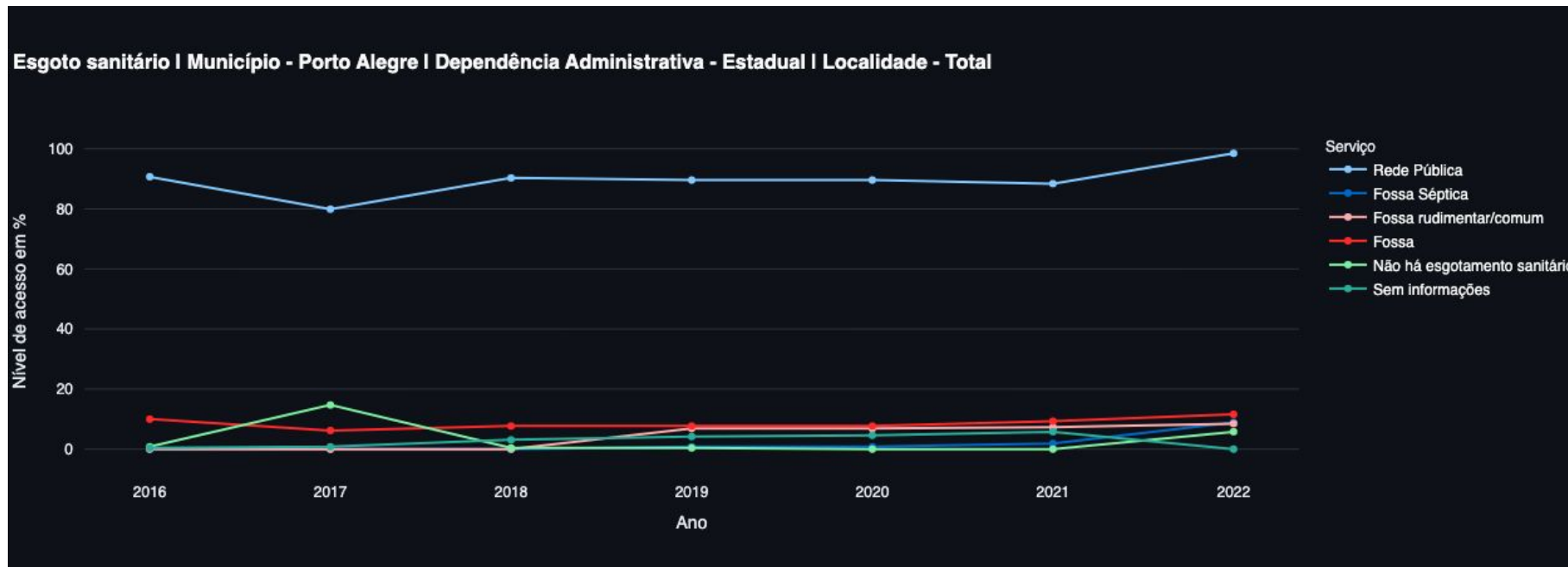
Quantidade de escolas em Porto Alegre



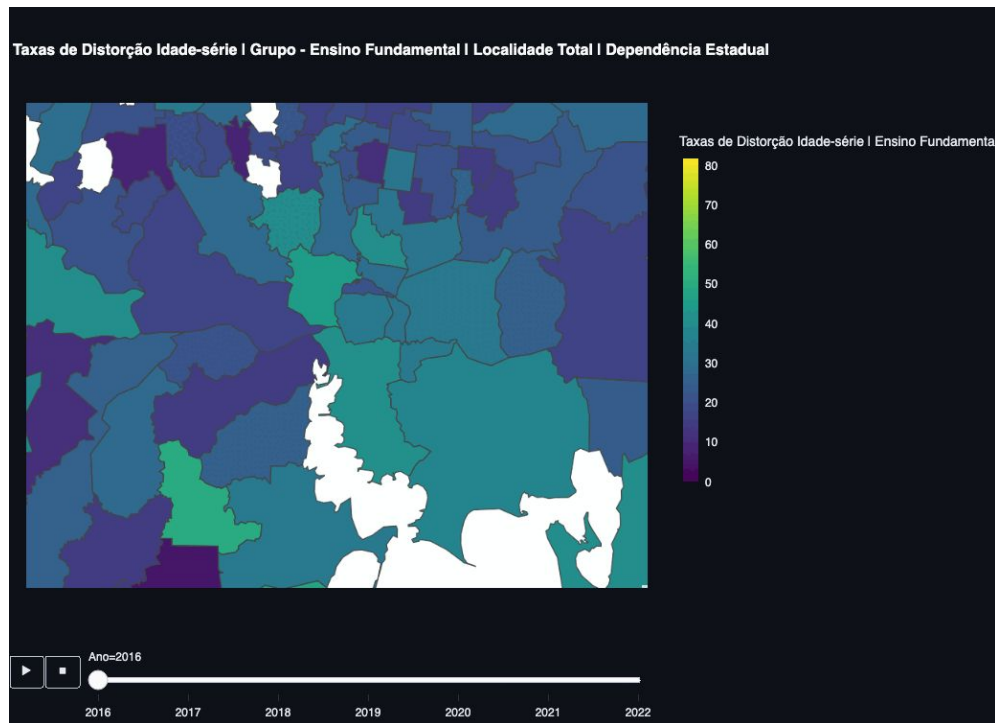
Quantidade de matrículas em Porto Alegre



Acesso a tratamento de esgoto em Porto Alegre



Taxa de distorção idade-série na região metropolitana de Porto Alegre



Sumário

Introdução e Contextualização

Objetivos

Trabalhos Relacionados

Desenvolvimento

Demonstração

Experimentos

Conclusões e Trabalhos futuros

Conclusões

ETL e
visualização
robusta,
econômica,
extensível e de
fácil execução

Conclusões

ETL e
visualização
robusta,
econômica,
extensível e de
fácil execução

Compartilhamento
do projeto e dados
facilitado com
Parquet, Github e
Streamlit

Conclusões

ETL e
visualização
robusta,
econômica,
extensível e de
fácil execução

Compartilhamento
do projeto e dados
facilitado com
Parquet, Github e
Streamlit

Disponibilização
dos dados pelo
INEP precisa
melhorar

Trabalhos futuros

Direcionar a
visualização
para
responder
perguntas
mais
específicas
sobre o Censo
Escolar

Trabalhos futuros

Direcionar a
visualização
para
responder
perguntas
mais
específicas
sobre o Censo
Escolar

Evoluir o ETL,
continuamente
corrigindo falhas e o
atualizando
conforme as
mudanças feitas
pelo INEP e as
necessidade
analíticas

Trabalhos futuros

Direcionar a visualização para responder perguntas mais específicas sobre o Censo Escolar

Evoluir o ETL, continuamente corrigindo falhas e o atualizando conforme as mudanças feitas pelo INEP e as necessidade analíticas

Estender a aplicação para além de visualização dados, implementando recursos de Aprendizado de Máquina

MUITO OBRIGADO!

