

Trabalho Prático 2

Introdução a Banco de Dados

Daniel Nogueira Junqueira - 2021072244

Lucas Lourenço Reis Resende - 2021014767

Rafael Camacho F. T de Oliveira - 2020030106

1. Introdução e Objetivos

O seguinte trabalho tem como objetivo trabalhar com os dados recuperados do governo relacionado a crimes violentos fazendo correlação com os indicadores de educação nos municípios de Minas Gerais.

Objetivos

O objetivo deste trabalho consiste em encontrar relações diretas ou indiretas entre índices de educação e de crimes a partir da análise e correlação dos dados do Censo Escolar e de Crimes Violentos dos municípios de Minas Gerais.

Com relação a isso, a principal hipótese levantada pelo grupo é a da existência de uma relação inversa entre os indicadores, ou seja, municípios com resultados relativamente piores nos dados de educação tendem a ter um número maior de crimes, partindo do pressuposto de que manter uma pessoa nas escolas reduz a chance dela cometer delitos.

Para realizar a análise, buscamos dados do INPE, do portal de dados abertos do Governo Federal e do IBGE, que estão citados a seguir.

- Crimes Violentos:
<https://dados.gov.br/dados/conjuntos-dados/crimes-violentos>
- Censo Escolar 2022:
<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados>
- População Municipal:
<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html>

A tabela abaixo apresenta os metadados sobre os dados obtidos, contendo informações sobre a obtenção destes e informações de cobertura.

Dado	Fonte	Data de Obtenção	Órgão Produtor	Data de Referência	Cobertura
Censo Escolar	https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados	19/06/2023	INEP	08/02/2023	Estados e Municípios do Brasil
Crimes Violentos	https://dados.gov.br/dados/conjuntos-dados/crimes-violentos	19/06/2023	Governo Federal	02/06/2023	Municípios de MG
População Municipal	https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html	19/06/2023	IBGE	12/12/2022	Estados e Municípios do Brasil

Metadados dos conjuntos analisados

2. Manipulação dos dados

Os dados obtidos foram inseridos em um SGBD, mais específico no PostgreSQL de acordo com o esquema relacional obtido através da análise exploratória, após a filtragem e elaboração do modelo relacional, como apresentado no próximo tópico.

3. Análise exploratória

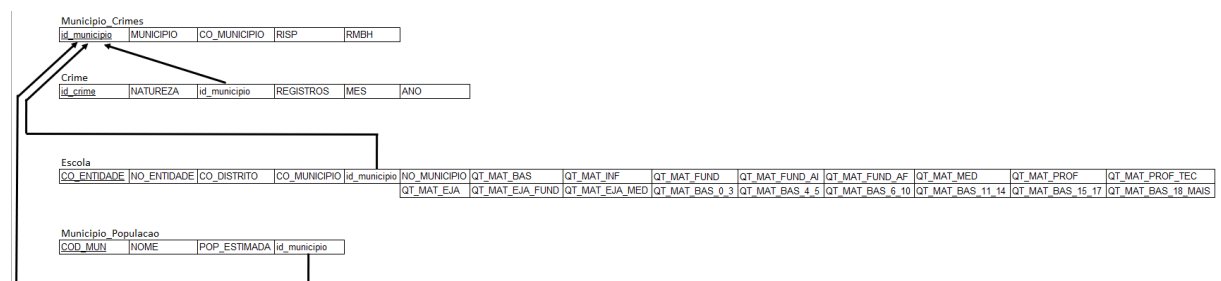
Buscando um melhor aproveitamento dos dados, primeiramente, foi realizada uma filtragem para eliminar indicadores que seriam de pouca relevância para a análise e proposta do trabalho. Mais especificamente, foram removidos indicadores do censo escolar, que continha muitas colunas e campos que não eram interessantes trabalharmos ou não ia fazer diferença na análise que estávamos fazendo. Mesmo realizando essa limpeza em certas colunas nos dados de censo escolar, o conteúdo dos dados ainda ficou bem grande devido a vários aspectos da educação que o arquivo aborda.

No link abaixo estão os dados já tratados.

https://drive.google.com/drive/folders/1MOWbzztH9BIRBuBzar6vdqc4TEyqbSm?usp=drive_link

Modelo Relacional

Para permitir a importação e análise dos dados foi elaborado um modelo relacional que representa o conjunto obtido após a filtragem e adaptação para possibilitar o carregamento para o SGBD. A imagem abaixo apresenta esse diagrama.



Modelo relacional para o conjunto de dados

Dicionário de Dados

Buscando possibilitar o entendimento do significado das variáveis e indicadores utilizados foi elaborado um dicionário de dados para os conjuntos de dados utilizados a partir das informações disponibilizadas pelas fontes dos dados. As tabelas a seguir apresentam o nome das variáveis, uma breve descrição, seu tipo e se se trata de uma chave primária ou estrangeira.

ESCOLA					
N	Nome da Variável	Descrição da Variável	Tipo	Tam	
1	NO_MUNICIPIO	Nome do Município	Char	150	
2	CO_MUNICIPIO	Código do Município	Num	7	
3	CO_DISTrito	Código do Distrito da Escola	Num	9	
4	CO_ENTIDADE	Código da Escola	Num	8	Chave Primária
5	NO_ENTIDADE	Nome da Escola	Char	100	
6	id_municipio	Id de município	int		Chave Estrangeira
Número de Matrículas na Educação Básica					
7	QT_MAT_BAS	Número de Matrículas na Educação Básica	Num	8	
8	QT_MAT_INF	Número de Matrículas na Educação Infantil	Num	8	
9	QT_MAT_FUND	Número de Matrículas no Ensino Fundamental	Num	8	
10	QT_MAT_FUND_AI	Número de Matrículas no Ensino Fundamental - Anos Iniciais	Num	8	
11	QT_MAT_FUND_AF	Número de Matrículas no Ensino Fundamental - Anos Finais	Num	8	
12	QT_MAT_MED	Número de Matrículas no Ensino Médio	Num	8	
13	QT_MAT_PROF	Número de Matrículas na Educação Profissional	Num	8	
14	QT_MAT_PROF_TEC	Número de Matrículas na Educação Profissional Técnica - Modo profissionalizante de ensino correspondente às turmas de cursos técnicos de nível médio nas formas articuladas (integrada ou concomitante), ou subsequente ao ensino médio e de normal/magistério	Num	8	
15	QT_MAT_EJA	Número de Matrículas na Educação de Jovens e Adultos (EJA)	Num	8	
16	QT_MAT_EJA_FUND	Número de Matrículas na Educação de Jovens e Adultos (EJA) - Ensino Fundamental	Num	8	
17	QT_MAT_EJA_MED	Número de Matrículas na Educação de Jovens e Adultos (EJA) - Ensino Médio	Num	8	
Número de Matrículas na Educação Básica - Segundo a Faixa Etária do Aluno					
18	QT_MAT_BAS_0_3	Número de Matrículas na Educação Básica - Até 3 anos de idade	Num	8	
19	QT_MAT_BAS_4_5	Número de Matrículas na Educação Básica - Entre 4 e 5 anos de idade	Num	8	
20	QT_MAT_BAS_6_10	Número de Matrículas na Educação Básica - Entre 6 e 10 anos de idade	Num	8	
21	QT_MAT_BAS_11_14	Número de Matrículas na Educação Básica - Entre 11 e 14 anos de idade	Num	8	
22	QT_MAT_BAS_15_17	Número de Matrículas na Educação Básica - Entre 15 e 17 anos de idade	Num	8	
23	QT_MAT_BAS_18 MAIS	Número de Matrículas na Educação Básica - Com 18 ou mais anos de idade	Num	8	

Dicionário de dados para a tabela “Escola”

Crime					
N	Nome da Variável	Descrição da Variável	Tipo	Tam	
1	id_crime	Identificador de crime	int		Chave Primária
2	NATUREZA	Tipo de crime	Char	150	
3	id_municipio	Id de município	int		Chave estrangeira
4	REGISTROS	Nº de registros	Num	100	
5	MES	Mês de registro	Num	2	
6	ANO	Nome da Escola	Char	100	

Dicionário de dados para a tabela “Crime”

Municipio_Crimes					
N	Nome da Variável	Descrição da Variável	Tipo	Tam	
1	id_municipio	Id de município	int		Chave Primária
2	MUNICIPIO	Nome do Município	Char	150	
3	CO_MUNICIPIO	Código do Município	Num	7	
4	RISP	Indicador de pertencimento à determinada Região Integrada de Segurança Pública	Num	2	
5	RMBH	Indicador de pertencimento à Região Metropolitana de Belo Horizonte	Bool		

Dicionário de dados para a tabela “Municipio_Crimes”

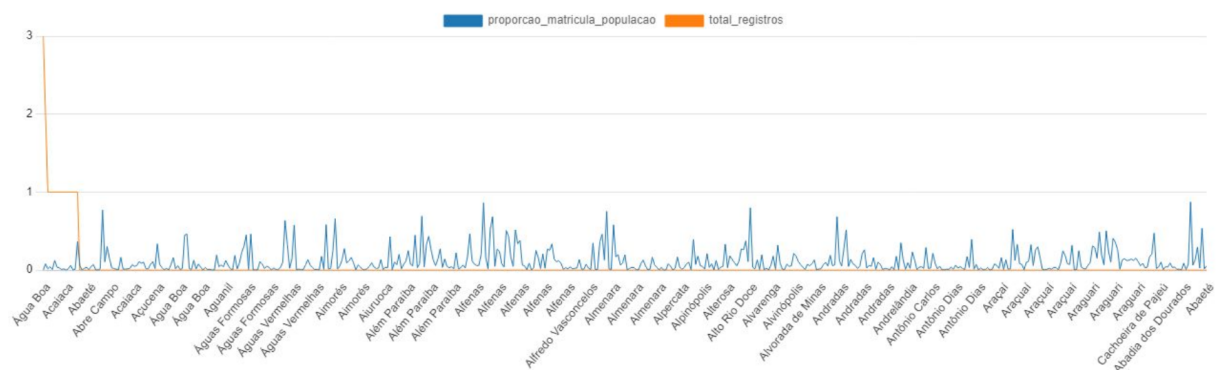
Municipio_Populacao					
N	Nome da Variável	Descrição da Variável	Tipo	Tam	
1	COD_MUN	Código do Município sem o código estadual (31)	Num	5	Chave Primária
2	NOME	Nome do Município	Char	150	
3	POP_ESTIMADA	População estimada	Num	7	
4	id_municipio	Id de município	int		Chave estrangeira

Dicionário de dados para a tabela “Municipio_Populacao”

4. Análise Descritiva e de Correlação

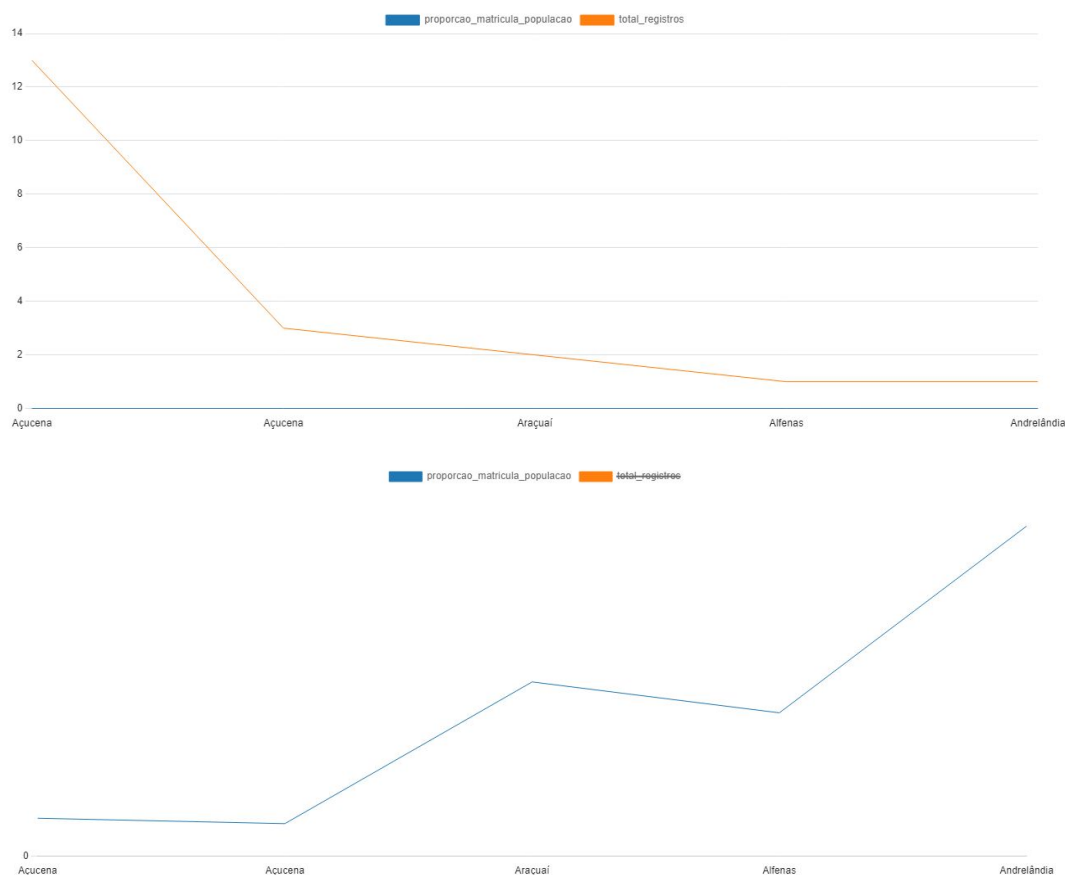
A análise de correlação foi pautada no número de registros de crimes de um município e a quantidade de pessoas matriculadas em institutos de educação de cada município. Porém, após plotar os dois dados em gráficos analisá-los foi possível perceber que municípios com maior número de matrículas não necessariamente teriam maior índice de escolaridade uma vez que não a população de cada área não estava sendo considerada. Logo, foi necessário a adição de uma tabela com a população de cada município em nosso banco de dados.

Em seguida, após as alterações necessárias no modelo, foi dado seguimento à análise de correlação entre a quantidade de registros de crimes e a quantidade de matrículas por população total de cada município. Para a análise foram gerados gráficos de proporção de pessoas matriculadas pela população total x número de registros de crimes, em cada município.



Para melhorar a visualização foram plotados apenas os dados referentes aos municípios com população menor que 50.000 e apesar de inicialmente apresentar um resultado de acordo com o proposto ao mostrar um maior número de registros nas cidades com menor índice de escolaridade, o restante do gráfico deixa claro como os dados do banco de dados apresentam inconsistências com diversos municípios com nenhum registro de crime, principalmente em cidades menores como essas com menor população em que a disseminação desses dados acontece de forma mais precária.

Em uma análise mais focada foram selecionadas 5 cidades com número expressivo de registros para facilitar a comparação. No primeiro gráfico é possível perceber a curva descendente de registros enquanto no segundo foram removidos os registros de crimes para perceber o aumento na escolaridade quase que na mesma proporção do decaimento dos delitos.



5. Conclusão

Para realização deste trabalho foram necessárias diversas decisões tomadas pelo grupo para enfrentar uma série de dificuldades que nos foram apresentadas durante todo o seu decorrer, visando a criação de um banco de dados relacional o mais estável e íntegro possível para fazer cruzamentos de dados e análise dos mesmos. Desde a preparação dos dados considerando a normalização do banco até a consideração de uma tabela populacional para uma análise de correlação mais precisa e otimizada.

O resultado esperado inicialmente seria que o número de crimes de um município seria menor de acordo o número de pessoas matriculadas por população do município aumentasse, criando assim uma relação inversa entre um indicativo educacional e a criminalidade de uma região. Os resultados finais da análise de nossos dados se mostraram incapazes de gerar uma relação entre os dados com precisão, uma vez que vários dados apresentavam registros duvidosos como o número de crimes em diversos municípios igual a 0 e partes das tabelas em branco. Apesar dessa variação dos dados, foi possível observar uma relação entre o índice de educação e a criminalidade dos municípios em alguns dados do banco como demonstrado nos últimos gráficos.