

"Priorização de Auditorias no DF por ranqueamento: Um estudo de Caso usando dados de Saúde, Educação e Segurança Pública das Regiões Administrativas"

Artigo Final
Disciplina Mineração de Dados - 05.2023
Mestrado Profissional - PPCA

Componentes:

Augusto Nunes - augcrn@gmail.com
Ennio Ferreira - ennioferreirab@gmail.com
Mayra Campos - mayra.campos@gmail.com



Por
Mayra Campos

Contexto

Contexto

→ Controladoria-Geral do Distrito Federal (CGDF);

" (...)assegurar a boa e regular aplicação dos recursos públicos; [e] coordenar o sistema de controle interno do DF"

→ Ações de controle divididas 3 linhas de defesa do controle-interno;

→ Garantir efetividade das ações:

◆ Auditoria Baseada em Riscos (Risk Based Auditing):

Nas ações e recomendações foca-se prioritariamente nos contextos que apresentam maior risco, impacto nos objetivos da organização.



Problema

Problema

- Planejamento das auditorias — necessidade de mensurar os riscos inerentes às estas ações da administração pública.
 - ◆ Ter uma ferramenta para indicar fragilidades.
- Garantir que o planejamento alcance os objetivos dadas as escassez de:
 - ◆ auditores,
 - ◆ e ferramentas para agilizar as auditorias.



Proposta

Proposta

- Pesquisar forma de mensurar os riscos na priorização de auditorias;
- Tentar criar um Índice para Priorizar Objetos de Auditoria no contexto da Controladoria-Geral do Distrito Federal, usando dados:
 - educacionais,
 - de saúde,
 - de segurança,
 - e socioeconômicos.



Limitações

Limitações

- O índice aqui sugerido está avaliado com relação às auditorias realizadas em 2022 (planejadas e *ad hoc*).
- RAs (Regiões Administrativas) não são entes de direito equivalentes aos municípios;

Limitações

- Diversidade de fontes
 - ◆ Diferentes metodologias
 - ◆ Diferentes temporalidades
 - ◆ Dados faltantes

Limitações

- Diferentes metodologias
 - ◆ Diferentes escalas
 - Às vezes melhor = maior,
 - Em outras melhor = menor.

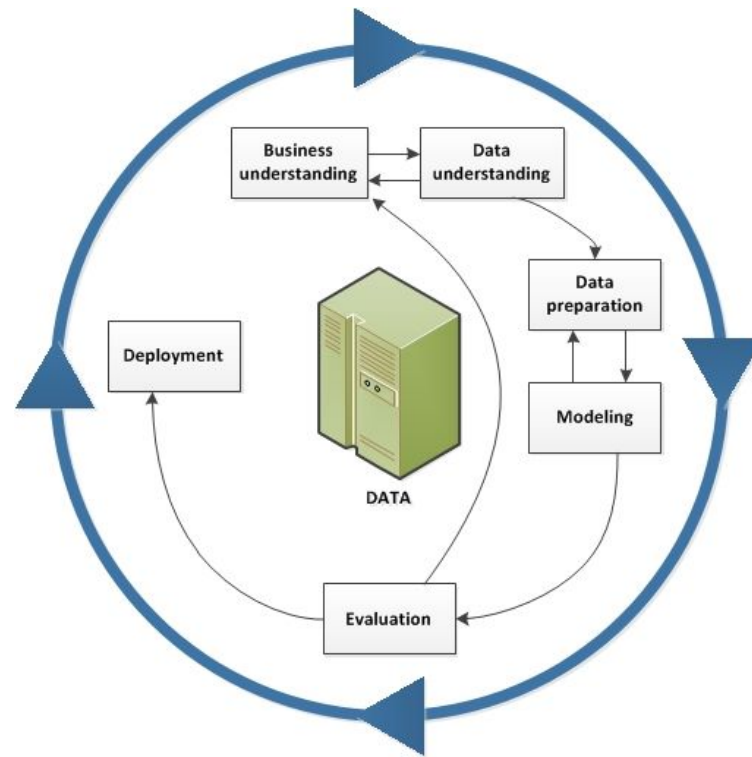


Metodologia

Metodologia

CRISP-DM - *Cross Industry Standard Process for Data Mining* (Processo Padrão de Vários Segmentos de Mercados para Mineração de Dados).

1. Compreensão do domínio (Entendimento do Negócio),
2. Compreensão dos dados,
3. Preparação dos dados,
4. Modelagem,
5. Avaliação e
6. Aplicação (Apresentação dos Dados).





1. Compreensão do Domínio

1. Compreensão do domínio

- Objetivos da aplicação: Cenário escolhido, Objetivos da aplicação, Critérios de sucesso da aplicação.
- Situação a ser avaliada: Inventário de Recurso, Requisitos, suposições e limitações, Custos e Benefícios previstos.
- Determinação das metas de MD: Metas e Critérios de sucesso de MD.
- Produção de projeto: Avaliação inicial de ferramentas e técnicas.



2. Compreensão dos dados

2. Compreensão dos dados

4 dimensões:

- Saúde
- Educação
- Socioeconômica
- Segurança

População por RA (estimativa PDAD)

2. Compreensão dos dados

→ *Saúde*

- ◆ Taxa de mortalidade fetal (por mil nascidos vivos)
- ◆ Taxa de mortalidade por causas evitáveis (por 10.000 habitantes)

2. Compreensão dos dados

→ *Educação*

- ◆ Média de alunos por turma no ensino fundamental
- ◆ Taxa de defasagem idade-série no ensino fundamental
- ◆ Taxa de reprovação no ensino fundamental

2. Compreensão dos dados

→ **Socioeconômica**

- ◆ Grau de instrução dos eleitores
- ◆ Proporção de famílias em situação de pobreza ($R\$ < 200$) que recebe benefícios sociais
- ◆ Proporção de famílias em situação de extrema pobreza ($R\$ < 100$) que recebe benefícios sociais
- ◆ Porcentagem da população com 65 anos ou mais
- ◆ Razão de dependência (população com menos de 15 anos e mais de 60 anos em relação à população de 15 a 60 anos)

2. Compreensão dos dados

→ **Segurança**

- ◆ Taxa de ocorrências de crimes e delitos registrados de CCP (*Crimes Contra o Patrimônio*) (por 1.000 habitantes)
- ◆ Taxa de CVLI (*Crimes Violentos Letais Intencionais*) (por 100.000 habitantes)



Por
Ennio Ferreira

3. Preparação dos Dados

3. Preparação do Dados

- Utilizamos conjuntos de dados abertos disponíveis:
- ◆ PDAD,
 - ◆ Datasus,
 - ◆ e-Gestor AB (Atenção Básica),
 - ◆ SSP-DF (Secretaria de Segurança Pública-DF) e
 - ◆ Censo Escolar DF

3. Preparação do Dados

- Diferentes metodologias
 - ◆ Diferentes escalas
 - Escalonamento MinMax
 - Dados em $[0, 1]$

3. Preparação do Dados

→ Dados faltantes

◆ Imputação usando Regressão
Ridge

4. Modelagem

4. Modelagem

→ Seleção da Técnica

- ◆ Clusterização não-supervisionada
- ◆ K-means
- ◆ Minimizar inércia inter-cluster
- ◆ Máxima diferença entre cluster

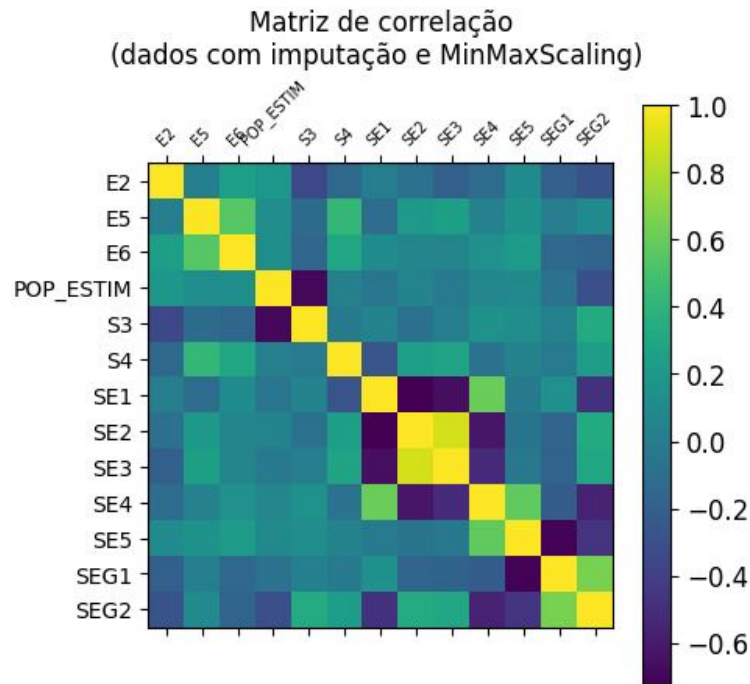
→ Parâmetros do modelo a ser avaliado

- ◆ Número de clusters
- ◆ Elbow.

→ Construção do Modelo

- ◆ Descrição dos modelos:
 - rankings das variáveis escolhidas,
 - 1 intracluster
 - 1 geral.

4. Modelagem - Colineariedade das variáveis



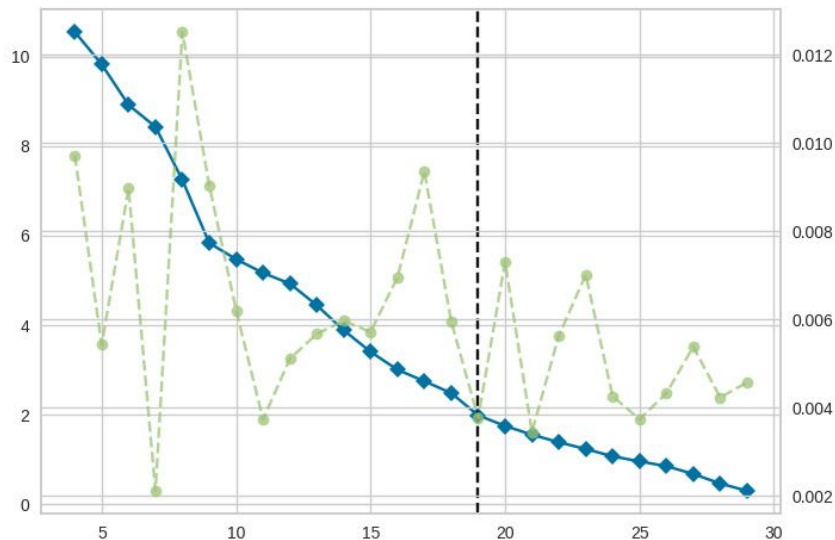
Apresentaram forte correlação, e foram retiradas.

2 variáveis socioeconômicas

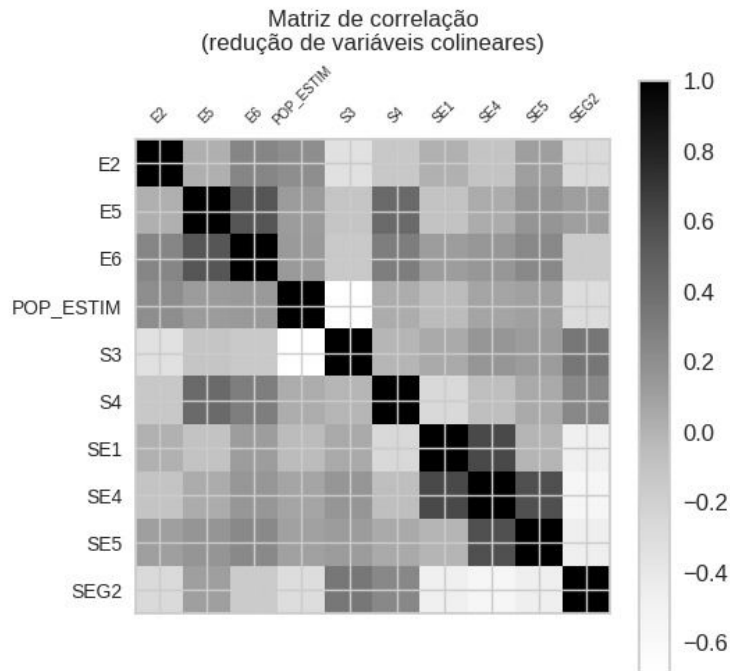
1 variável de segurança

4. Modelagem - Parâmetro do Modelo

▸ KElbowVisualizer
▸ estimator: KMeans
▸ KMeans



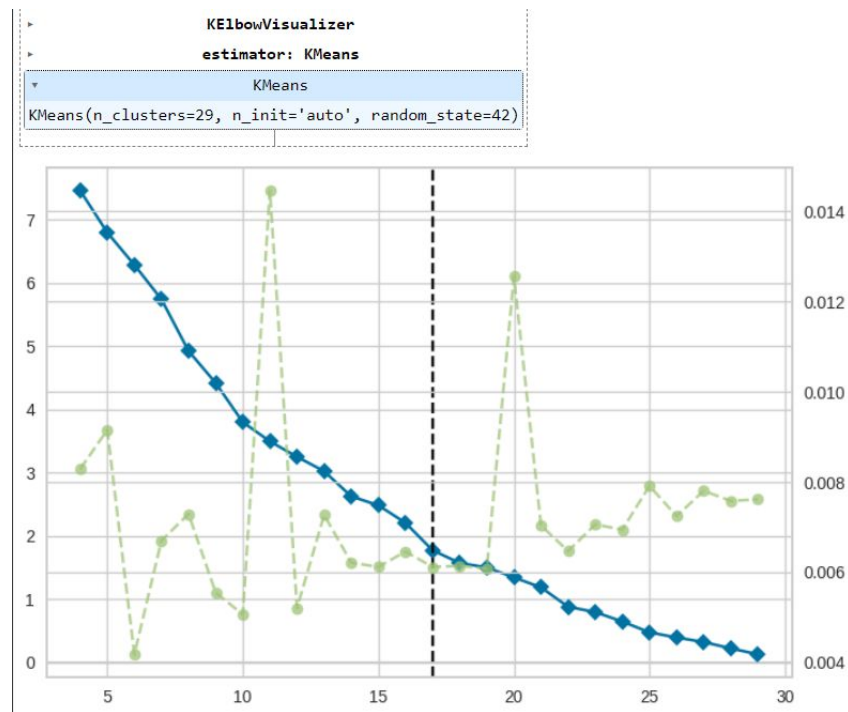
4. Modelagem - Colineariedade das variáveis restantes



Foram utilizadas as variáveis relativas a:

- Educação,
- Saúde e
- Segurança

4. Modelagem - variáveis restantes





Por
Augusto Nunes

5. Avaliação

5. Avaliação

- Avaliação dos Resultados
 - ◆ Concordância dos *rankings* geral e intra-cluster com as auditorias feitas em 2022
- Revisão do Processo
- Determinação dos próximos passos
 - ◆ Lista de possíveis ações
 - ◆ Decisões

[illegible]

5. Avaliação

→ *Ranking* geral

- ◆ As RA foram *rankeadas* de acordo com o *ranking* médio das variáveis
- ◆ Das top 10 RA, 9 sofreram auditoria da CGDF em 2022

5. Avaliação

→ *Ranking* intracluster

◆ Dos 17 *clusters* ajustados,
em 12 tivemos auditoria em
pelo menos uma RA



6. Aplicação

6. Aplicação

- Aplicação do Projeto
 - ◆ Proposta de ferramenta para melhorar o “Planejamento das Auditorias” e assim mensurar melhor os riscos inerentes a atividade.
 - ◆ Apesar das limitações metodológicas, concordância com as auditorias realizadas de fato foi alta

6. Aplicação

- Plano de Monitoramento e Manutenção
 - ◆ O monitoramento feito por “curadorias” iterativas com os auditores de controle interno.
- Revisão do Projeto
 - ◆ Anual, quando da realização do planejamento anual de auditorias (1o semestre)

**Agradecemos
a atenção!**

Augusto Nunes -
augcrn@gmail.com

Ennio Ferreira -
ennioferreirab@gmail.com

Mayra Campos -
mayra.campos@gmail.com

