

Modelos de gestões de cidades médias por perfil de investimento em educação e saúde

allan.sales.89@gmail.com

Abril de 2017

1 Definição

1.1 Contexto

No início de cada nova gestão municipal alguns pontos devem ser levantados pelos gestores, dentre eles: 1) quais são os problemas que afligem a população, 2) quanto deve-se investir para amenizar/resolver os problemas e 3) quais as prioridades de investimento do município. Nesse cenário, pode ser útil encontrar gestões bem sucedidas de cidades com prioridades similares para que possam servir de modelo para novas gestões.

Este experimento foi desenvolvido com foco em cidades médias do Brasil - cidades que apresentam de 100 a 300 mil habitantes -, no período de 2009 a 2012 e o sucesso de uma gestão é mensurada através da métrica Índice FIRJAN de Desenvolvimento Municipal (IFDM) ¹. O IFDM é uma métrica semelhante ao Índice de Desenvolvimento Humano (IDH) ² e procura aferir o desenvolvimento municipal por ano.

1.2 Problema abordado

Tecnicamente o problema pode ser descrito como: agrupar cidades por perfil de investimento em cada área e verificar o crescimento do IFDM nas áreas de interesse. Por exemplo, agrupa-se as cidades pelos investimentos feitos na área de saúde e indica-se como modelos as cidades de cada grupo que tiveram maior crescimento de IFDM na área de saúde.

1.3 Métricas

Para validação dos grupos foi utilizada o coeficiente de Silhouette [1] e para validação das cidades modelos uma análise utilizando boxplots é feita.

¹<http://www.firjan.com.br/ifdm/>

²<http://www.br.undp.org/content/brazil/pt/home/idh0.html>

2 Dados

Os dados de investimentos municipais foram obtidos por meio do banco de dados do Tesouro Nacional ³ enquanto os dados de IFDM foram obtidos no site da FIRJAN ⁴ e a população dos municípios pode ser obtida a partir do IBGE ⁵.

2.1 Construção da base de dados

Com a obtenção dos dados, fez-se necessário alinhá-los de forma a contemplar, para cada cidade e área: 1) a informação de gastos, 2) crescimento/decrescimento do IFDM e 3) a população no período de 2009 a 2012. Para isso algumas medidas foram tomadas:

- Para cada área de investimento somou-se os gastos de cada ano (i.e. o gasto total em saúde da gestão é representado pela soma de gastos em saúde desde 2009 até 2012).
- Devido os gastos de cada cidade serem diferentes, os gastos de cada área são colocados em porcentagem, de forma que o valor do gasto representa quanto do total de investimentos daquela cidade foi feita na área. Por exemplo, 0.4 investidos em saúde representa que 40% dos investimentos daquela cidade foram feitos na área de saúde.
- O IFDM da gestão foi calculado como a diferença do IFDM no final e do início da gestão, ou seja, IFDM do ano 2012 menos IFDM do ano 2008. Neste estudo, utilizamos o IFDM-saúde e o IFDM-educação que medem o desenvolvimento municipal nas áreas de saúde e educação, respectivamente.
- Considerou-se por cidade média qualquer cidade que foi registrado com população entre 100 e 300 mil habitantes no censo de 2010.

2.2 Descrição

A tabela de gastos pronta, utilizando dados de investimentos nas áreas de educação e saúde, apresenta 23 colunas com informações de:

- Município:
1) Nome do município, 2) Unidade Federativa (UF), 3) Região e 4) População.
- IFDM:
1) Crescimento do IFDM-saúde de 2008 a 2012 e 2) Crescimento do IFDM-educação de 2008 a 2012.
- Investimentos em saúde e em sub-áreas de saúde:
1) Total gasto com saúde, 2) Atenção Básica, 3) Assistência Hospitalar, 4) Suporte Profilático, 5) Vigilância Sanitária, 6) Vigilância Epidemiológica, 7) Alimentação e Nutrição e 8) Demais Subfunções de Saúde.

³http://www.tesouro.fazenda.gov.br/pt_PT/contas-anuais

⁴<http://www.firjan.com.br/>

⁵<http://cidades.ibge.gov.br/>

- Investimentos em educação e em sub-áreas de educação:

1) Total gasto com educação, 2) Ensino Fundamental, 3) Ensino Médio, 4) Ensino Profissional, 5) Ensino Superior, 6) Educação Infantil, 7) Educação de Jovens e Adultos, 8) Educação Especial e 9) Demais Subfunções de Educação.

3 Análise

A seguir são descritas algumas informações sobre os dados.

3.1 Análise exploratória

São 184 cidades situadas principalmente no Sudeste do país, sendo São Paulo o estado mais numeroso - como retratado na tabela 1.

Região	Nº Cidades	Estado	Nº Cidades
Sudeste	88	SP	51
		MG	17
		RJ	15
		ES	5
Nordeste	38	BA	11
		PE	8
		CE	6
		MA	6
		RN	2
		PB	2
		SE	1
		PI	1
Sul	36	AL	1
		RS	14
		PR	13
Centro-Oeste	11	SC	9
		GO	5
		MT	3
Norte	11	MS	3
		PA	5
		TO	2
		RR	1
		RO	1
		AP	1
		AM	1
		AC	0

Table 1: Número de cidades por estado e região

A tabela 2 dá uma ideia da distribuição de valores para cada *feature* observada.

Por exemplo, há de se perceber que, em média, os investimentos em educação foram maiores que os investimentos em saúde - 0.2729 e 0.2361, respectivamente - e isso se reflete no IFDM-saúde e IFDM-educação de forma que o IFDM-educação apresenta média de crescimento mais de duas vezes maior que o IFDM-saúde.

Feature	Média	Desvio padrão	Mínimo	25%	50%	75%	Máximo
IFDM-saúde	0.0350	0.0488	-0.0761	0.0012	0.0280	0.0623	0.2117
IFDM-educação	0.0759	0.0348	0.0044	0.0520	0.0742	0.0940	0.2960
Saúde	0.2361	0.0619	0.1004	0.1973	0.2227	0.2703	0.4209
Atenção Básica	0.0999	0.0569	0.0	0.0605	0.0935	0.1354	0.3203
Assistência Hospitalar	0.0893	0.0671	0.0	0.0339	0.0868	0.1242	0.3238
Suporte Profilático	0.0046	0.0092	0.0	0.0	0.0025	0.0059	0.0971
Vigilância Sanitária	0.0020	0.0028	0.0	0.0002	0.0010	0.0027	0.0193
Vigilância Epidemiológica	0.0050	0.0044	0.0	0.0017	0.0037	0.0079	0.0214
Alimentação e Nutrição	0.0021	0.0058	0.0	0.0	0.0	0.0006	0.0332
Demais Subfunções Saúde	0.0333	0.0447	0.0	0.0	0.0162	0.0500	0.2219
Educação	0.2729	0.0624	0.1445	0.2303	0.2704	0.3043	0.4650
Ensino Fundamental	0.1998	0.0680	0.0746	0.1531	0.1916	0.2395	0.4065
Ensino Médio	0.0028	0.0163	0.0	0.0	0.0	0.0009	0.1716
Ensino Profissional	0.0014	0.0054	0.0	0.0	0.0	0.0003	0.0452
Ensino Superior	0.0031	0.0095	0.0	0.0	0.0	0.0015	0.0703
Educação Infantil	0.0450	0.0321	0.0	0.0183	0.0394	0.0712	0.1444
Educação de Jovens e Adultos	0.0021	0.0035	0.0	0.0	0.0007	0.0025	0.0209
Educação Especial	0.0016	0.0029	0.0	0.0	0.0002	0.0023	0.0171
Demais Subfunções Educação	0.0215	0.0324	0.0	0.0019	0.0112	0.0287	0.2205

Table 2: Descrição das características.

3.2 Visualização exploratória

Também é possível notar pela tabela 2 que em algumas *features*, o primeiro quartil ou mesmo a mediana apresentam valores 0, o que significa que diversas cidades não investiram nessas sub-áreas. A figura 1 demonstra a porcentagem de cidades que não fizeram investimento em cada sub-área.

A fim de uma melhor visualização da distribuição de cada *feature*, foi criado uma *scatter matrix*. A imagem não está disponível neste documento por questão de tamanho, mas ele pode ser visualizado através do arquivo *distribuicaoFeatures.pdf* nesta mesma pasta. Nele é possível perceber que grande parte das *features* apresentam viés, sendo o valor zero pico de diversas delas.

3.3 Técnicas e algoritmos

Todas as técnicas utilizadas neste experimento são consolidados na literatura e estão listados a seguir:

- *Principal Component Analysis* (PCA) [2]. Técnica de redução de dimensões que é de importante aplicação nesse contexto principalmente por questão de visualização gráfica dos grupos.
- K-Means [3]. Técnica de agrupamento escolhida para este contexto devido a não haver a possibilidade de alguma cidade encaixar-se em mais de um grupo.

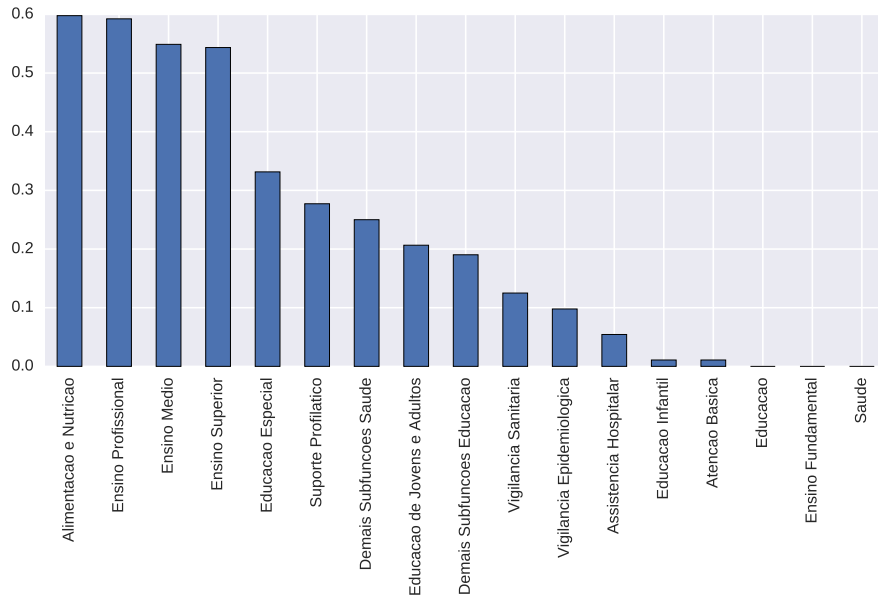


Figure 1: Porcentagem de cidades com zero investimento em cada sub-área.

- Coeficiente de Silhouette. Procura mensurar separação e coesão dos grupos, a técnica pode ser utilizada para escolher o melhor K do K-Means.
- Detecção de *outliers*. As cidades modelos são aquelas que mais desenvolveram o IFDM-área (saúde/educação) mesmo com investimentos nas sub-áreas similares aos de seu grupo. Dessa forma, análise de *boxplots* é interessante por deixar claro o quanto a cidade modelo se desenvolveu comparado ao seu grupo.

3.4 Benchmark

Nas buscas realizadas não foram encontrados estudos semelhantes a esse, de forma que impossibilita-nos a comparar este estudo com algum da literatura. Uma forma alternativa é, inicialmente, comparar a nossa abordagem com uma abordagem *naive*, onde uma gestão pode ser tomada como modelo apenas por examinar o crescimento do IFDM durante o período da gestão.

4 Metodologia

4.1 Pré-processamento

Nessa fase do experimento, os dados foram trabalhados procurando por: 1) transformar os dados das *features* em normal - ou mais próximo de normal - e 2) detecção de

outliers.

- Transformação dos dados. Dado a figura *distribuicaoFeatures.pdf* é possível de se notar a não normalidade de grande parte das *features* disponíveis, sendo assim, foi utilizada a função de raiz cúbica a fim diminuir o viés e transformá-las em algo mais próximo da distribuição normal. Os resultados da transformação podem ser encontrados na figura *distribuicaoFeaturesTransformadas.pdf* (encontrado na mesma pasta desse documento). Após a transformação percebe-se que os diversas *features* estão mais similares a uma curva normal, apesar de alguma parte delas continuarem com forte viés.
- Detecção de *outliers*. Nenhuma gestão de cidade da amostra foi considerada *outlier* e retirada do experimento. Isso devido a possibilidade de que a sua forma de investir possa ser exatamente o fato que a destaca de todas as outras gestões e a conduz a um maior crescimento.

4.2 Implementação

O objetivo deste experimento é descobrir modelos de gestões de cidades por perfil de investimento das cidades e pela área de interesse. Sendo assim, alguns passos devem ser tomados. Em ordem: 1) separar as *features* de cada área e 2) retirar as *features* de gastos totais de cada área. Após essa fase, para cada área: 1) executar o PCA, 2) aplicar o coeficiente de Silhouette, 3) gerar os grupos a partir do K-Means, 4) verificar as cidades com maior crescimento de IFDM em cada grupo e 5) avaliar o quão significativo é o crescimento dessas cidades através da análise de *boxplots*.

4.2.1 Features por área

A primeira medida tomada é a de separar as *features* referentes a educação e as referentes a saúde. Para justificar essa divisão por áreas podemos pensar que uma mesma gestão não necessariamente obterá bons resultados em todas as áreas de investimento. Ela pode, por exemplo, colher bons resultados com os investimentos em saúde mas não ter a mesma competência na educação. Sendo assim, recomendar todo o modelo de investimento dessa gestão de cidade para algum gestor pode não ser a melhor alternativa. Recomendar o modelo de gestão apenas de saúde e buscar uma segunda gestão de cidade para servir de modelo de gestão de educação talvez seja uma melhor saída.

Após a separação, as *features* que indicam o total investido em educação/saúde são retirados da base de dados, já que essa informação é redundante dado que ela pode ser obtida apenas somando os gastos nas sub-áreas de educação/saúde. A tabela 3 apresenta as *features* utilizadas em cada área.

4.2.2 PCA

Após a definição de quais *features* serão utilizadas, o próximo passo é aplicar o PCA nas *features* de cada área separadamente a fim de facilitar a visualização dos grupos gerados. A figura 2 e a figura 3 apresentam a variância explicada de cada componente

<i>Feature Saúde</i>	<i>Feature Educação</i>
Atenção Básica	Ensino Fundamental
Assistência Hospitalar	Ensino Médio
Suporte Profilático	Ensino Profissional
Vigilância Sanitária	Ensino Superior
Vigilância Epidemiológica	Educação Infantil
Alimentação e Nutrição	Educação de Jovens e Adultos
Demais Subfunções de Saúde	Educação Especial
	Demais Subfunções de Educação

Table 3: *Features* por área

gerado para por área. Para decidir o valor de componentes utilizados na análise de grupos, tanto para a área de saúde quanto para a de educação, foi tomado como inspiração o *elbow method* [4]. Assim, decidiu-se que dois componentes para a área de saúde - somando uma variância acumulada de aproximadamente 69% - e dois componentes para a área de educação - culminando numa variância aproximada de 54% - seriam adequados.

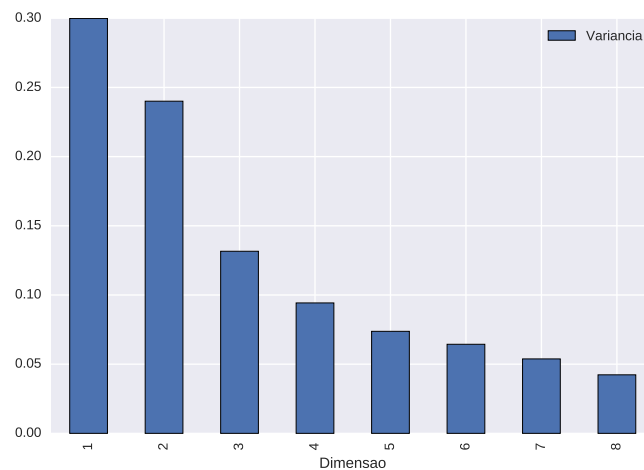


Figure 2: Proporção de variância explicada dos componentes na área de educação.

4.2.3 Silhouette

Com os componentes gerados executa-se o coeficiente de Silhouette a fim de inferir qual a melhor quantidade de grupos e executar o K-Means. Ainda seguindo a ideia do *elbow method* e observando os valores apresentado na tabela 4 e tabela 5, é possível verificar que, de acordo com o coeficiente de Silhouette, o melhor número de grupos

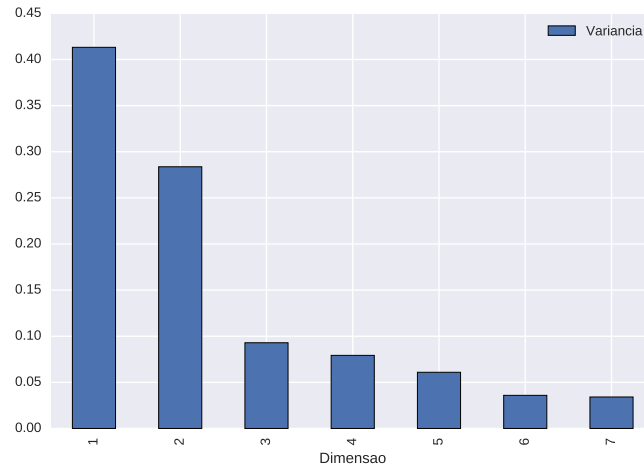


Figure 3: Proporção de variância explicada dos componentes na área de saúde.

para a área de educação são dois e para a área de saúde são três.

Número de grupos	Valor de Silhouette
2	0.3892
3	0.3820
4	0.3818
5	0.3569
6	0.3475
7	0.3640
8	0.3643
9	0.3823

Table 4: Valor de Silhouette pelo número de grupos para educação.

4.2.4 K-Means

Após a definição do número de grupos, executa-se o K-Means. Os grupos podem ser visualizados como apresentados na figura 4 e na figura 5. Uma descrição mais detalhada dos grupos será exposta na seção 5.

4.2.5 Modelos de gestões por perfil

Com os grupos gerados é possível apontar quais são os modelos de gestões por perfil de investimentos a partir do crescimento do IFDM naquela gestão. A tabela 6 apresenta

Número de grupos	Valor de Silhouette
2	0.4234
3	0.4491
4	0.3894
5	0.4148
6	0.4084
7	0.3825
8	0.3828
9	0.3917

Table 5: Valor de Silhouette pelo número de grupos para saúde.

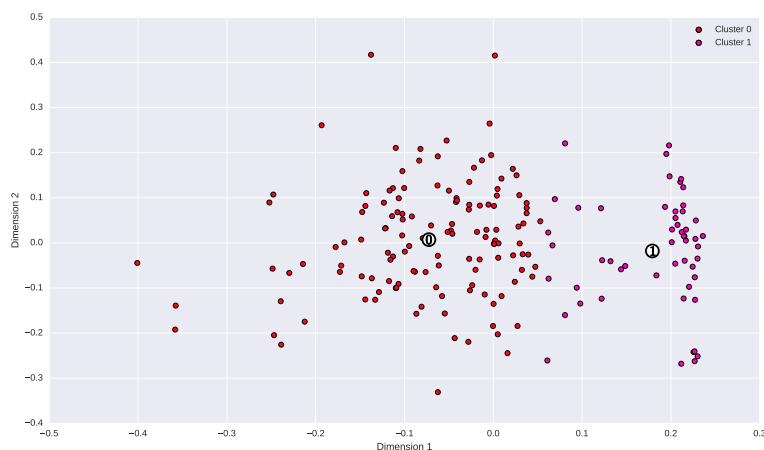


Figure 4: Grupos na área de educação.

as cidades com maior crescimento de IFDM por área. Esses resultados são discutidos na seção 5.

5 Resultados

Como é possível visualizar na tabela 6 e na figura 4 e figura 5, foram encontrados três perfis de gestões municipais na área de saúde e dois perfis na área de educação. A tabela 6 faz um apanhado, de forma resumida, dos grupos e resultados alcançados.

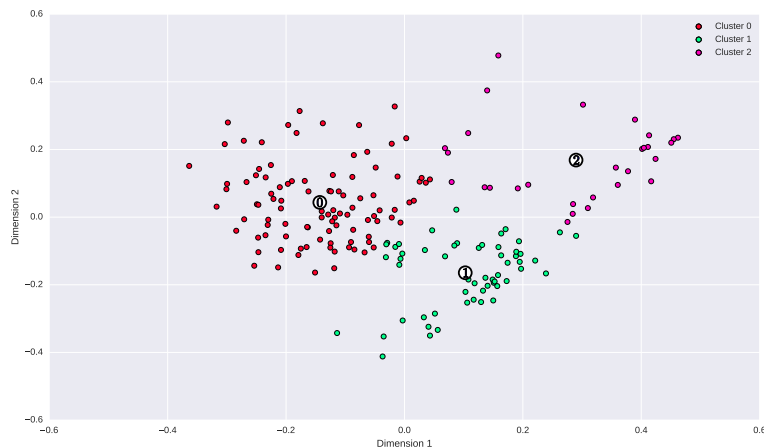


Figure 5: Grupos na área de saúde.

5.1 Avaliação e validação do modelo

5.1.1 Características dos grupos

Considerando os grupos formados baseados nos investimentos em saúde - figura 6 é possível perceber, primeiramente, que os gestores investem pouco - ou nada - em qualquer sub-área que não seja Atenção Básica, Assistência Hospitalar e Demais subfunções de saúde. Para educação - figura 7 -, percebe-se o mesmo comportamento para as *features* de Ensino Fundamental, Educação Infantil e Demais subfunções de educação. Esse comportamento pode ser reflexo da divisão de tarefas e competências municipais, estaduais e federais. Também há de se perceber que existe um padrão para as cidades de cada grupo de acordo com as sub-áreas:

- Grupos baseados em investimento em saúde:

Grupo 0. O grupo destaca-se pelo investimento mais equilibrado nas três áreas. Tem um investimento considerável em Atenção Básica e Assistência Hospitalar e se destaca dos outros dois grupos por ter necessidade de complementar o trabalho com investimentos em Demais subfunções de saúde.

Grupo 1. Apresenta grande investimento em Atenção Básica e Assistência Hospitalar.

Grupo 2. Quase todo o investimento na área de saúde corresponde a Atenção Básica.

- Grupos baseados em investimento em educação:

Grupo 0. Grupo que apresenta investimentos nas três sub-áreas de destaque mas com foco principal em Ensino Fundamental e Educação Infantil.

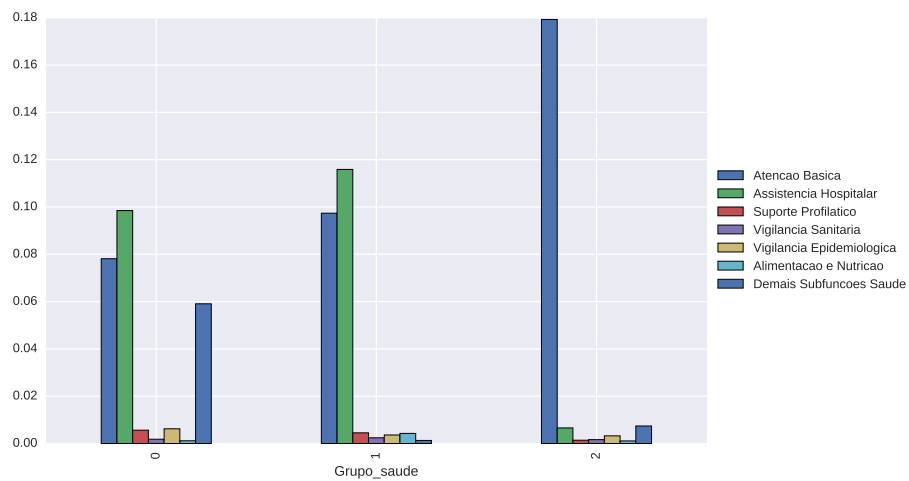


Figure 6: Perfis de grupos de saúde.

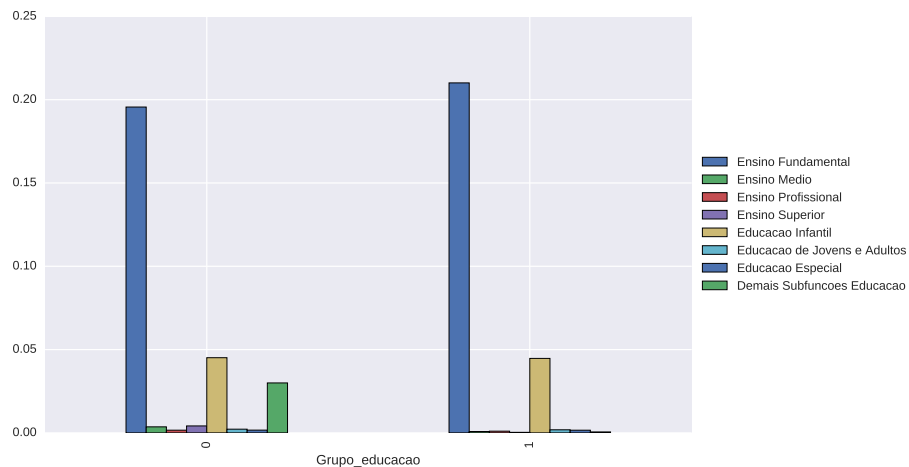


Figure 7: Perfis de grupos de educação.

Grupo 1. Apresenta investimentos similares ao grupo 0 nas sub-áreas de Ensino Fundamental e Educação Infantil e diferencia-se por não ter investimentos consideráveis em Demais funções de educação.

Essas características também são possíveis de serem visualizadas em seus centroides reais, como apresentadas na figura 8 e figura 9.

De forma simples, é possível perceber que nenhum dos grupos se destacou por ter apresentado um maior investimento médio que os outros, aliado a isso, o IFDM-saúde e IFDM-educação médio dos grupos também não apresentam grandes discrepâncias,

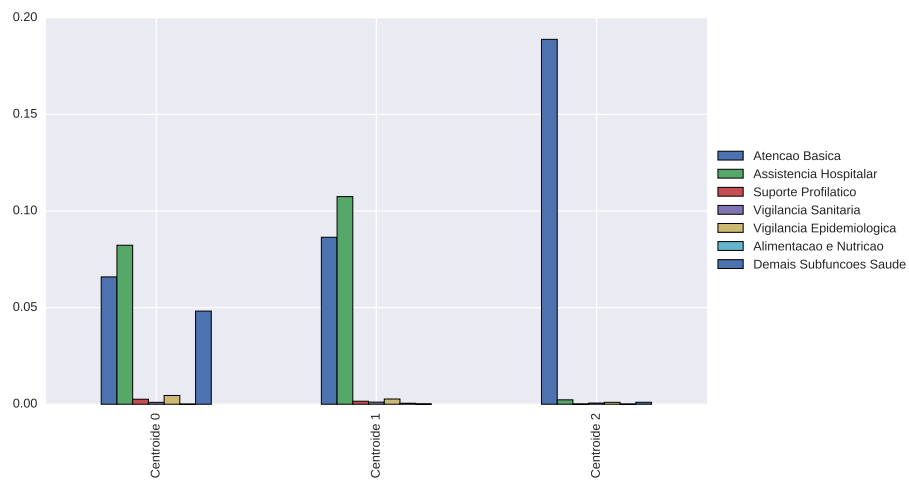


Figure 8: Centroides reais de saúde.

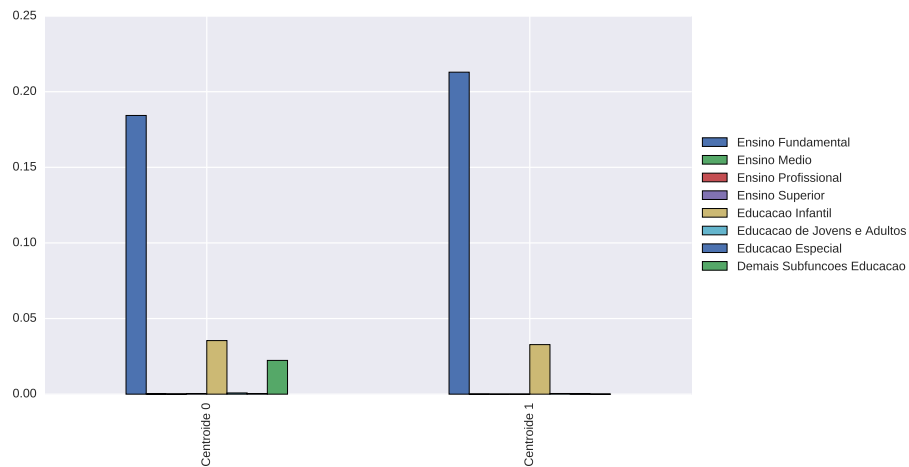


Figure 9: Centroides reais de educação.

como mostra a coluna de Crescimento médio do grupo na tabela 6.

5.1.2 Boxplots

Para apresentar a qualidade das gestões chamadas de modelos, são apresentados dois *boxplots* e a tabela 6 que evidenciam quão grande foi o crescimento dessas cidades em relação às cidades de seus respectivos grupos. Gestões de cidades que obtiveram desenvolvimento tão grande a ponto de serem considerados *outliers* já dão indícios que investiram o dinheiro de forma mais adequada que as de seus grupos.

Na figura 10, nota-se que a distribuição do crescimento do IFDM-educação para os dois grupos são similares. O resultado é esperado dado que a diferença entre os dois grupos são investimentos menores em Demais subfunções de educação. Também é de se perceber que nenhuma cidade teve um decréscimo no valor do IFDM-educação nesse período. Olhando para cada grupo de forma isolada, pode-se evidenciar a gestão de Camaçari - para o Grupo 0 - que apresentou um desenvolvimento vertiginosamente diferenciado em relação a seu grupo, crescendo quase que 4 vezes mais que a média das cidades do grupo. Para o Grupo 1, destaca-se a gestão da cidade de Foz do Iguaçu que apesar de ter menor desenvolvimento nesse tempo, ainda apresenta valor quase que 2 vezes maior que a média do grupo.

Na área de saúde - figura 11 -, é possível perceber que existe alguma diferença nas distribuições de IFDMs-saúde e que quase 25% das cidades de cada grupo obtiveram um decréscimo de IFDM-saúde durante aquela gestão. Além disso, dentre elas, a distribuição dos pontos do Grupo 0 é mais espaça e a do Grupo 2 é a mais concisa. Sobre os modelos de gestões, nota-se no Grupo 0 - que não apresenta alguma cidade como *outlier* no quesito de saúde -, Praia Grande obteve desenvolvimento mais de 4 vezes maior que o desenvolvimento médio do grupo e pode ser apontado como um bom exemplo de gestão.

Área de investimento	Grupo	UF	Município	Crescimento da cidade	Crescimento médio do grupo	Número de pontos
Saúde	0	SP	Praia Grande	0.1517	0.0350	99
	1	BA	Barreiras	0.2117	0.0396	56
	2	MG	Coronel Fabriciano	0.1397	0.0259	29
Educação	0	BA	Camaçari	0.2960	0.0759	53
	1	PR	Foz do Iguaçu	0.1474	0.0756	131

Table 6: Gestões modelos para área de saúde e educação por perfil de gestões.

5.2 Justificativa

Como já mencionado anteriormente, não foram encontrados modelos que indicassem de forma automática bons exemplos de gestões de cidades por perfil de investimentos. Assim, há algumas abordagens que podem ser exploradas para obtermos um *baseline*, como, por exemplo a utilizada neste trabalho: Examinar o IFDM da gestão e recomendar aquelas cidades em que houveram maior crescimento em saúde ou em educação.

Utilizando essa abordagem como limiar, obteríamos como resultado a gestão da cidade de Barreiras como modelo de investimentos em saúde e Camaçari como modelo de investimentos em educação. Para este contexto, podemos afirmar que a abordagem proposta neste experimento tende a ser mais adequada que o *baseline* devido estar sendo considerado as necessidades de investimentos atuais da cidade (o que estamos caracterizando como perfil) e, assim, obtemos exemplos de gestões para cada perfil de investimento encontrado, como exemplificado pela tabela 6.

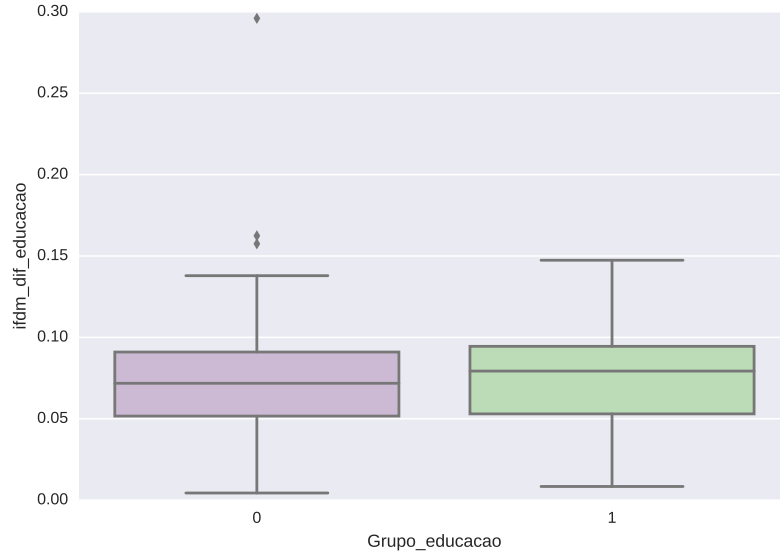


Figure 10: Boxplots dos IFDM-educação.

6 Conclusão

6.1 Aplicação prática

Esse modelo pode facilmente ser aplicado em dados de investimentos públicos de forma a servir como referência a outras gestões de cidades. Na prática, podemos implantar o modelo de forma a dar suporte aos gestores de uma cidade em um novo período de mandato. Os requisitos para que o modelo possa ser utilizado apenas são: 1) fazer o levantamento de quais são as prioridades de investimentos atuais da cidade e 2) usar o levantamento, para poder ter uma ideia de em qual grupo a cidade se encaixaria e tomar o líder de crescimento de IFDM na área de objetivo como parâmetro.

6.2 Discussão

Esse projeto foi desenvolvido seguindo as fases do processo *Knowledge Discovery in Databases* (KDD). Durante o processo foram encontradas algumas dificuldades, ameaças a validade e feitas algumas considerações, tais como:

- Aplicação do dinheiro da melhor forma. Dado que não há uma maneira de julgar quão bem aplicados foram as verbas disponíveis para cada cidade, houve de se considerar que cada gestão municipal fez uso da verba da melhor forma possível.
- Superfaturamento de investimentos e corrupção. Fiscalizar corrupção é e superfaturamento de obras é uma tarefa difícil e não está no escopo desse experimento.

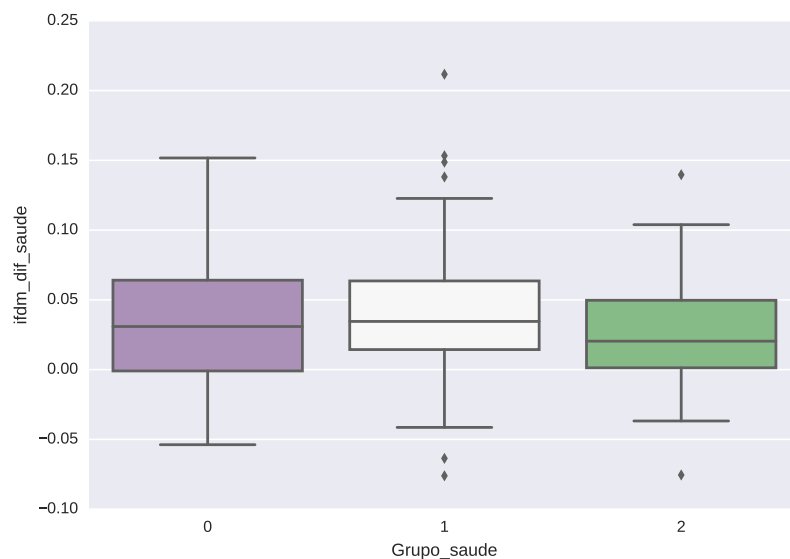


Figure 11: Boxplots dos IFDM-saúde.

Aqui, teve-se de considerar que o valor de todo investimento feito em sub-área correspondia ao valor real e justo a ser pago pelo que foi comprado.

- A variação do IFDM para os anos de 2009 a 2012 é reflexo dos investimentos nesse mesmo período de tempo.

6.3 Trabalhos futuros

Para garantir a consistência dos resultados obtidos, pode-se utilizar dados de outros anos/gestões para reforçar os grupos encontrados. Por exemplo, utilizando dados da gestão de 2013 a 2016 espera-se encontrar grupos similares aos encontrados nas gestões de 2009 a 2012. Além disso, pode-se replicar o experimento para mais áreas, como Emprego e Renda, a fim de encontrar gestões modelos nesse quesito. Também há a possibilidade de aplicar o experimento de forma geral, por exemplo, utilizando todos os investimentos feitos durante a gestão em todas as áreas disponíveis e encontrarmos gestões modelos não apenas em saúde ou educação, mas sim em todos os contextos.

References

- [1] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

- [2] K. Pearson, "On lines and planes of closest fit to systems of point in space," *Philosophical Magazine*, vol. 2, no. 11, pp. 559–572, 1901.
- [3] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA., 1967.
- [4] D. J. Ketchen Jr and C. L. Shook, "The application of cluster analysis in strategic management research: an analysis and critique," *Strategic management journal*, pp. 441–458, 1996.