

Projeto Mineração de Dados Educacionais (MDE)

Inteligência Artificial na Educação

Análise dos dados do IDEP de São Paulo (SP) 2019

Ezequiel França dos Santos

7751964

Abril de 2021

Resumo

O presente relatório tem o intuito de apresentar a análise dos dados do IDEP (Índice de Desenvolvimento da Educação Básica) da rede municipal de ensino de São Paulo. Além de algumas visualizações interessantes do contexto dos dados, ainda procuramos encontrar possíveis correlações que podem estar influenciando os resultados das escolas no IDEP, focando nos aspectos socioeconômicos das escolas e estudantes. Para isto foi utilizado a base de dados fornecida no site da prefeitura de São Paulo, o programa Orange com os addons Associate e Geo.

Introdução

Para este projeto utilizamos a planilha de resultados do IDEP de 2019. O IDEP é um indicador do desenvolvimento da Educação Básica da Rede Municipal de Ensino de São Paulo (RME) calculado pela Secretaria Municipal de Educação (SME) a partir dos resultados das avaliações da Provinha e Prova São Paulo e dos resultados das taxas de aprovação. Sua proposta é similar à proposta nacional do Índice de Desenvolvimento da Educação Básica (IDEB).

O IDEP é calculado para cada escola e varia de acordo com uma escala de 0 a 10, em que a projeção das metas, para cada grupo, levará em consideração o contexto de cada escola pelo Indicador de Nível Socioeconômico (INSE) e pelo Indicador de Complexidade de Gestão (ICG), calculados pelo Inep. Assim, permitirá estabelecer critérios de avanço nos níveis de proficiências e nas taxas de aprovação, de acordo com a realidade de cada grupo. Além do avanço no IDEP ao longo dos anos, também se almeja a diminuição da desigualdade dos resultados, de modo que as diferenças entre os resultados das escolas sejam minimizadas dentro do grupo do qual cada uma faz parte, bem como entre os grupos estabelecidos. O modelo geral do IDEP é dado por:

$$IDEP_{ji} = R_{ji} \cdot F_{ji} \quad 0 \leq R_j \leq 10; 0 \leq F_j \leq 1 \text{ e } 0 \leq IDEP_j \leq 10, \quad (1)$$

em que:

i = ano da aplicação da Prova São Paulo e dos dados de aprovados, reprovados e abandono, consolidados na Escola On-Line (EOL);

R_{ji} = média de proficiência em Ciências, Língua Portuguesa e Matemática, padronizada para um indicador entre 0 e 10, dos estudantes da unidade j , obtida em determinada edição i da Prova São Paulo realizada no 3º e 5º anos para os anos iniciais e 7º e 9º anos para os anos finais;

F_{ji} = indicador do fluxo escolar baseado na taxa de aprovação da etapa de ensino dos estudantes da unidade j ;

INSE

O Indicador de Nível Socioeconômico (Inse) é considerado um constructo latente que sintetiza, de maneira unidimensional, a partir das respostas dos estudantes aos questionários contextuais do Saeb, informações sobre a escolaridade dos pais, renda familiar, posse de bens e contratação de serviços de empregados domésticos. O seu objetivo é contextualizar o desempenho das escolas nas avaliações e exames realizados pelo Inep, bem como seus esforços na realização do trabalho educativo ao caracterizar, de modo geral, o padrão de vida de seu público, relacionado à sua respectiva posição na hierarquia social.

ICG

O Indicador de Complexidade de Gestão (ICG) das escolas resume em uma única medida as informações quanto a seu porte: número de matrícula, número de turnos e número de etapas atendidas. A gestão da escola certamente envolve outros fatores e dimensões não contemplados aqui, entretanto, verifica-se que, mesmo não contemplando outros aspectos na sua construção, o indicador apresenta potencial para contextualização dos resultados das avaliações. O ICG é divulgado em seis níveis, mas para que fosse utilizado na projeção das metas, os níveis foram agrupados em: Grupo 1, formado pelos níveis 1, 2 e 3 e Grupo 2, formado pelos níveis 4, 5 e 6.

Faixa

Agrupamento criado pela prefeitura que cruza ICG e INSE.

Agrupamento de INSE e ICG por Faixas

	INSE	ICG
Faixa 1	3	1
Faixa 2	3	2
Faixa 3	4	1
Faixa 4	4	2
Faixa 5	5	1
Faixa 6	5	2

Fonte: SME/COPED/NTA

Tabela 1: Agrupamento INSE e ICG Prefeitura de SP

Desenvolvimento e Resultados

A base de dados utilizada não dispunha dos dados de geolocalização, apenas os distritos de cada escola. Por este motivo, um script foi desenvolvido para obtenção dos dados de latitude e longitude destes distritos, afim de criar-se visualizações em mapas, conformes veremos a seguir.

Com isto, as colunas da nossa tabela ficaram da seguinte forma:

CodigoEscola	Código interno de cada escola	Texto
NomeEscola	Nome da Escola	Texto
Diretoria	Diretoria de Ensino	Categoria
Distrito	Distrito a qual pertence	Categoria
(INSE)	Indicador de Nível Socioeconômico	Categoria
(ICG)	Indicador de Complexidade de Gestão	Categoria
Faixa	Faixa criada pela prefeitura considerando ICG x INSE	Categoria
2018	Resultado IDEP 2018	Feature
2019	Resultado IDEP 2018	Feature
2019M	Meta de Resultado IDEP 2019	Feature
2020M	Meta de Resultado IDEP 2020	Feature
2021M	Meta de Resultado IDEP 2021	Feature
2022M	Meta de Resultado IDEP 2022	Feature
2023M	Meta de Resultado IDEP 2023	Feature
Latitude	Latitude do Distrito	Feature
Longitude	Longitude do Distrito	Feature

Tabela 2: Colunas do Dataset utilizado.

A primeira análise gráfica é a comparação dos resultados de 2018 e 2019, considerando os distritos e a meta para 2019. No gráfico 1.0 a seguir,

O Tamanho está relacionado a meta de IDEP 2019, Cores estão relacionadas a Distritos e $x = \text{IDEP 2019}$, $y = \text{IDEP 2018}$

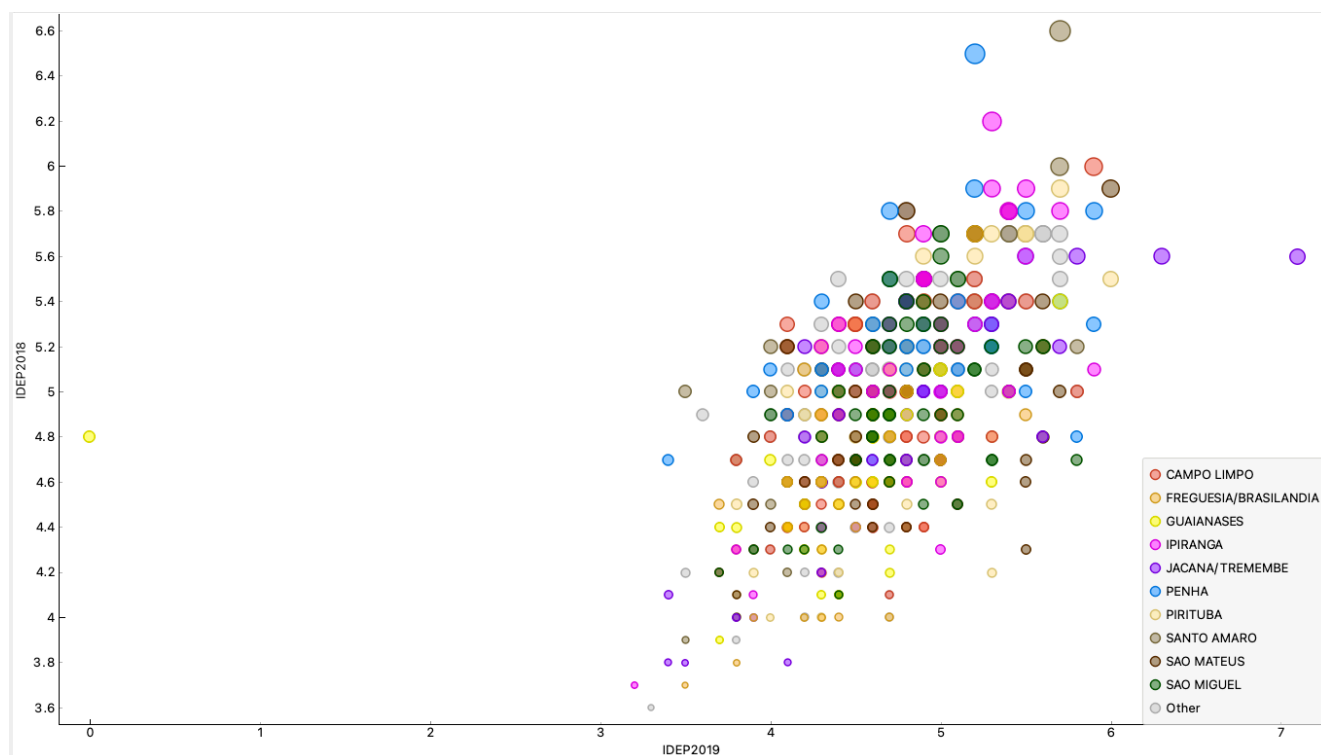


Gráfico 1.0: IDEP 2018 x 2019 + Meta 2019 por distrito

Neste caso, podemos perceber uma discrepância entre os distritos de cores rosa e roxa dos demais quando consideramos as metas 2019.

Ainda no intuito de analisar correlações, como dito anteriormente através da obtenção dos dados de geolocalização, construímos as colunas Latitude e Longitude, pois uma das análises desejadas era olhar o mapa da cidade e perceber se os resultados e a localização das escolas tinham alguma correlação.

Os gráficos a 2a e 2b , em duas versões fundo escuro e claro, mostram a distribuição das notas de 2019 no mapa da cidade de São Paulo.

É interessante observar que, diferente do que imaginávamos, não há uma correlação. Temos escolas com notas mais altas (mais amareladas) tanto no centro (Região Central, Zonas mais ricas da cidade) assim como casos nos extremos (Periferias da Cidade). Existe uma pequena diferença, mas não ao ponto de podermos levantar essa correlação.

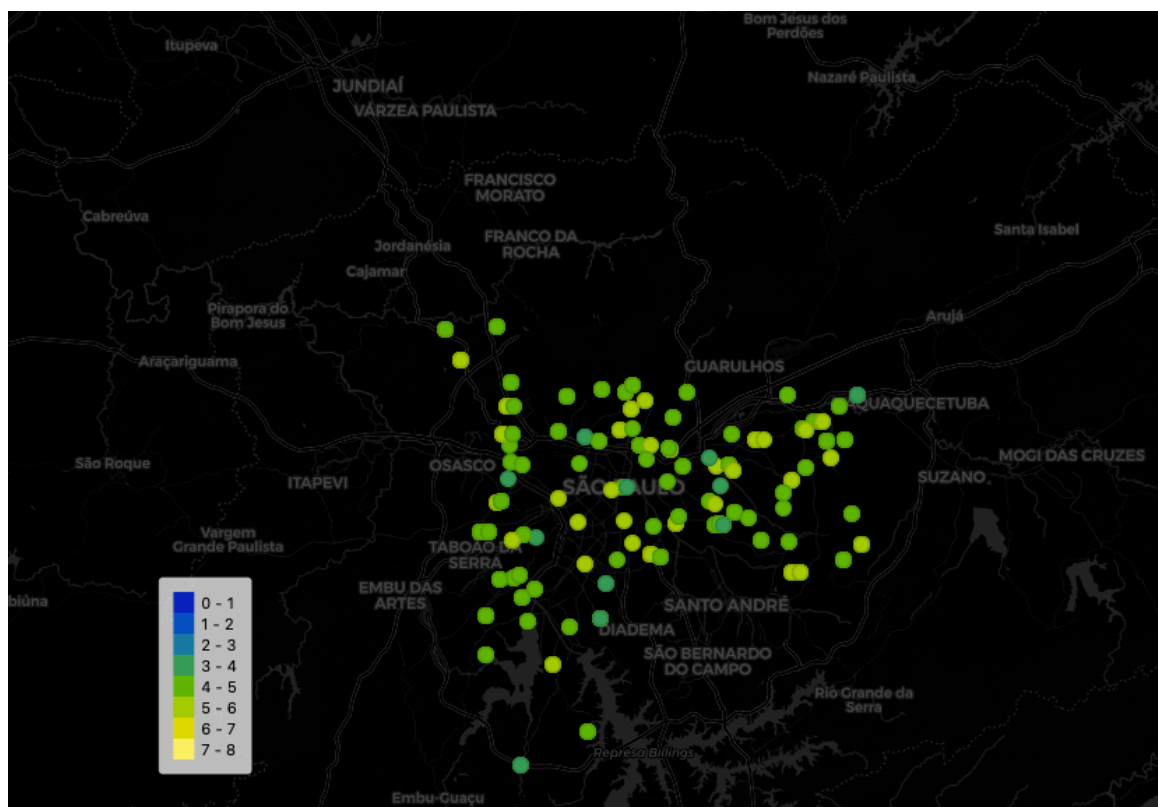


Gráfico 2a: Distribuição da notas (0-10) pelo mapa de São Paulo

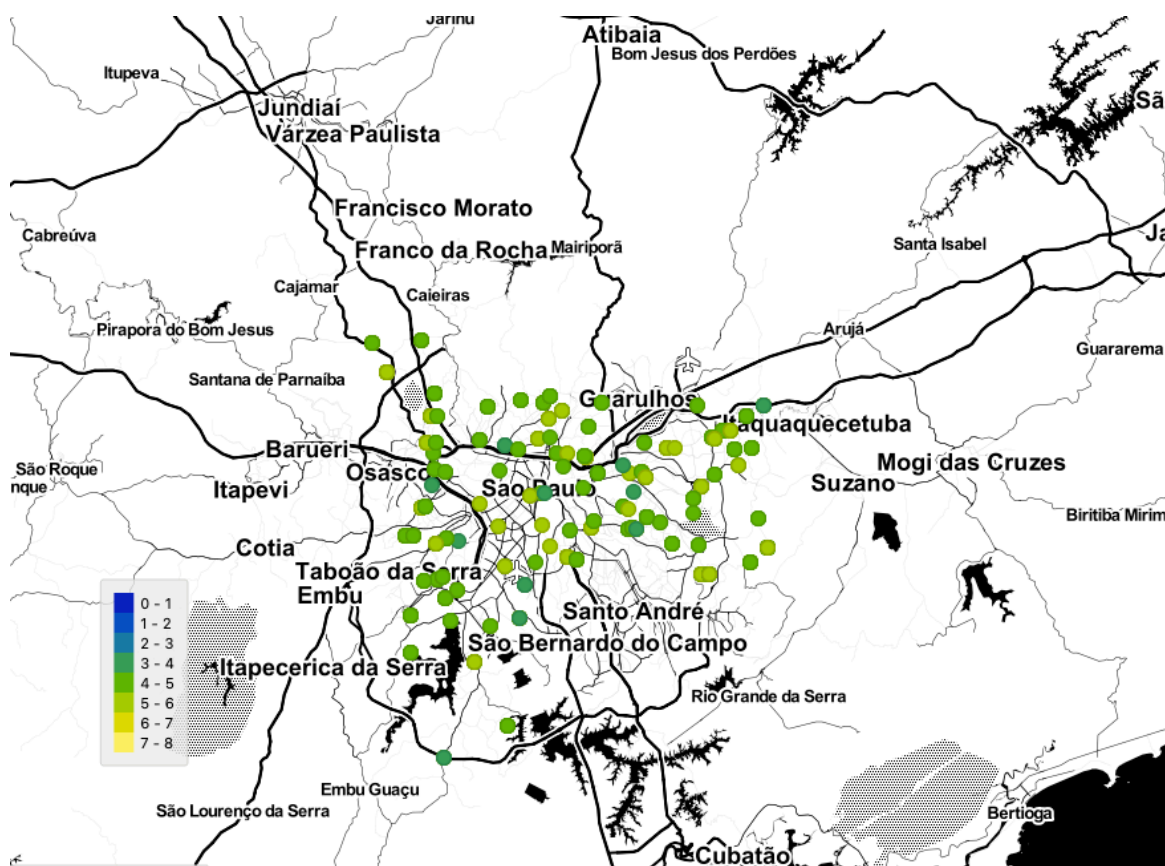


Gráfico 2b: Distribuição da notas (0-10) pelo mapa de São Paulo

Ainda na busca de correlações Socioeconômicas e o desempenho no IDEP, um gráfico de distribuição do Nível Socioeconômico.

Apesar de simples este gráfico 3.0 nos ajuda a entender alguns problemas, quando nos deparamos novamente com os dados percebemos que o INSE é calculado com uma média de todos estudantes, eliminando os extremos, o que é um problema, já que o fator socioeconômico era o ponto central da análise.

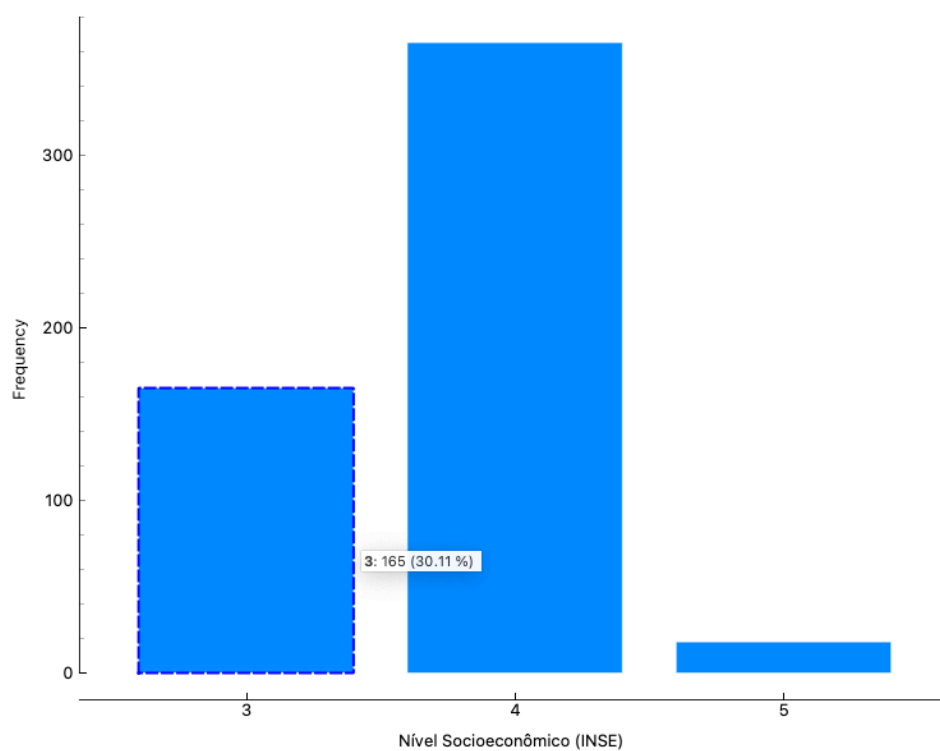


Gráfico 3.0: Distribuição do nível socioeconômico

Confiantes de que os fatores socioeconômicos apontariam para algum caminho, continuamos a fazer diversas associações e combinações, como a Figura 1.0, onde combinamos INSE com ICG e percebemos que o verde (baixa complexidade e índice social mais alto) são a minoria das escolas.

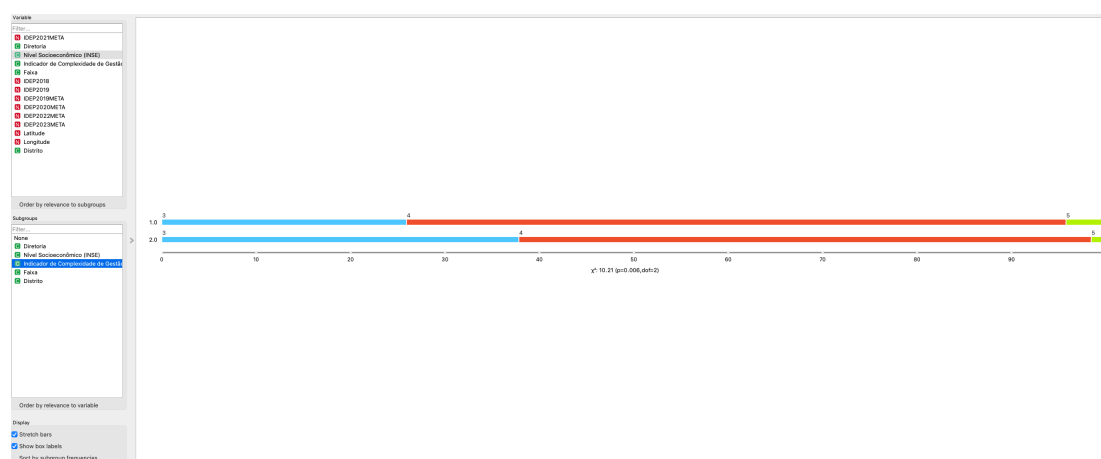


Figura 1: Box Plot INSE x ICG

Diante disto o gráfico 4.0 de violino deu a visão dos resultados do IDEP 2019 e o INSE. Percebe-se claramente o baixo número de resultados no violino verde, que possui maior INSE.

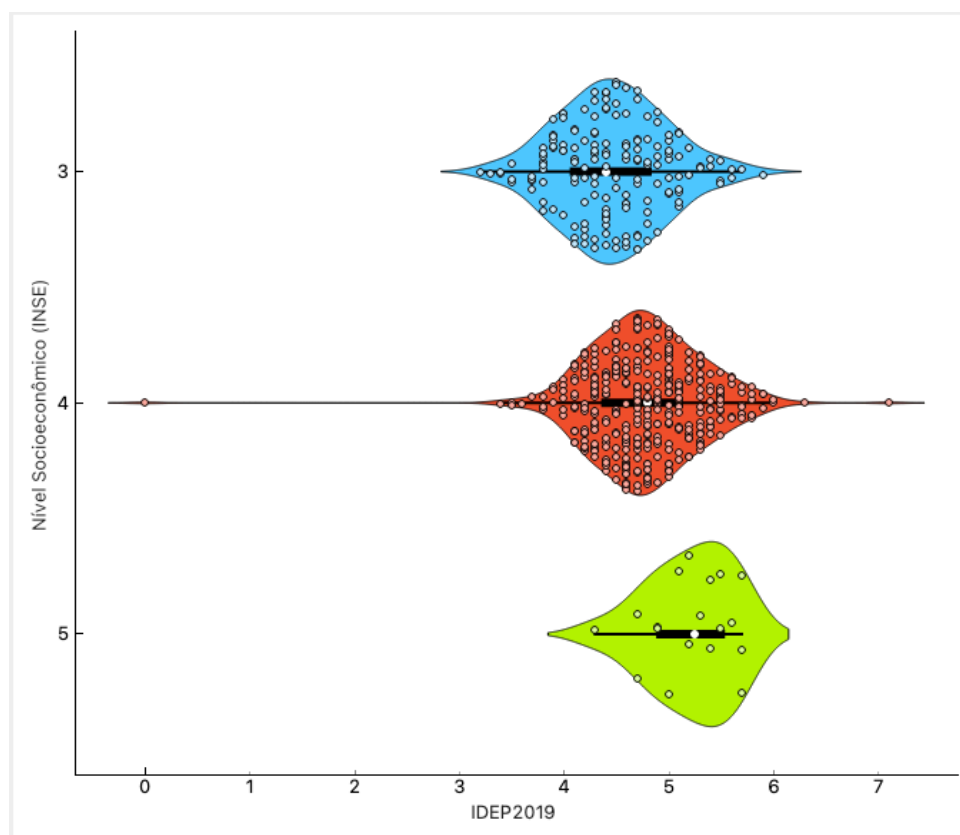


Gráfico 4.0: Gráfico de Violino INSE x IDEP2019

Apesar da sugestão para o gráfico 4.0 ter vindo da figura 1.0, podemos ver na figura 2.0 que o classificador de *Frequent Itemsets* também indicava relação forte entre ICG e INSE.

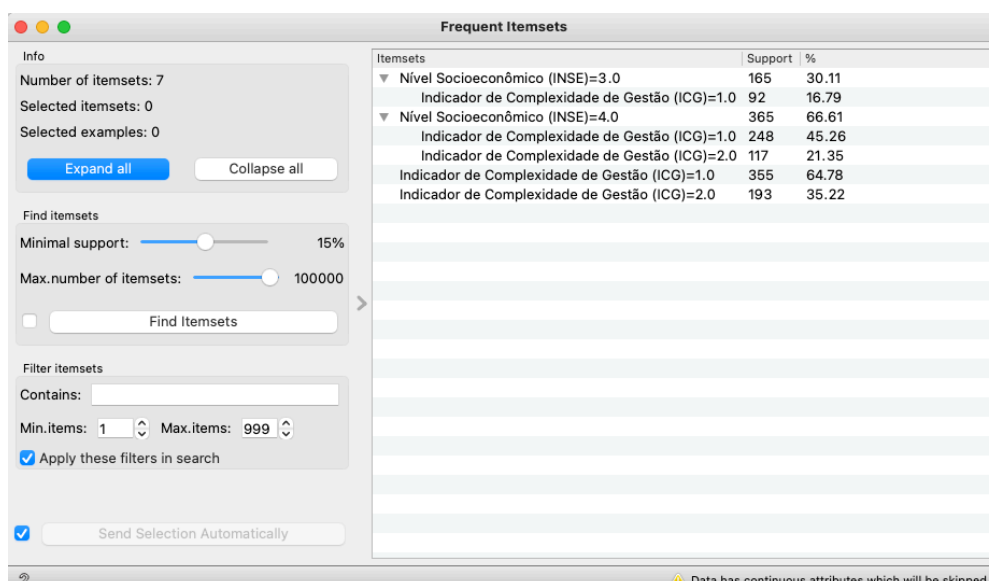


Figura 2.0: Frequent Itemsets - INSE 3 e 4 relacionados com baixo ICG

Conclusões

Apesar da noção clara do quanto os fatores socioeconômicos são impactantes na vida de um estudante, vide o que a pandemia de COVID-19 fez de desigualdades, os dados muitas vezes apontam coisas diferentes e por isso devemos ter cuidado e analisar com o maior número de ângulos possíveis. Podemos supor sim, através da análise destes dados que escolas com IGC maior tendem a ter resultados menores, assim como a combinação de IGC e INSE são fatores que também tendem a levar o valor do IDEP para baixo.

Complexidade de gestão leva a menos cuidado com os alunos, e quando se trata de alunos com vulnerabilidade social, essa soma acaba não sendo linear. Apenas o fator geográfico não é suficiente para determinar os resultados do IDEP, o que é um resultado positivo, todavia, ainda vemos nas periferias os menores resultados.

Bibliografia

Santos, Ezequiel. **Inteligência Artificial na Educação**, 2021. Notas de Aula..

São Paulo (SP). Secretaria Municipal de Educação. IDEP : Índice de Desenvolvimento da Educação Paulistana. – São Paulo : SME, 2019. 20p. : il.

IDEP, Dados obtidos em: <http://dados.prefeitura.sp.gov.br/dataset/idep>

Apêndice

ICG:

GRUPO 1 Nível 1 - Porte inferior a 50 matrículas, operando em único turno e etapa e apresentando a Educação Infantil ou Anos Iniciais como etapa mais elevada. Nível 2 - Porte entre 50 e 300 matrículas, operando em 2 turnos, com oferta de até 2 etapas e apresentando a Educação Infantil ou Anos Iniciais como etapa mais elevada. Nível 3 - Porte entre 50 e 500 matrículas, operando em 2 turnos, com 2 ou 3 etapas e apresentando os Anos Finais como etapa mais elevada.

GRUPO 2 Nível 4 - Porte entre 150 e 1.000 matrículas, operando em 2 ou 3 turnos, com 2 ou 3 etapas, apresentando Ensino Médio/Profissional ou a EJA como etapa mais elevada. Nível 5 - Porte entre 150 e 1.000 matrículas, operando em 3 turnos, com 2 ou 3 etapas, apresentando a EJA como etapa mais elevada. Nível 6 - Porte superior a 500 matrículas, operando em 3 turnos, com 4 ou mais etapas, apresentando a EJA como etapa mais elevada.