**ANÁLISE DA CORRELAÇÃO ENTRE A SITUAÇÃO DE TRABALHO E DESEMPENHO DOS ALUNOS DE SISTEMAS DE INFORMAÇÃO NO ENADE**

**Caio Cézar da Silva Rezende**

**Luiz Antonio Borges Cantarino**

**Phillipe Ferreira de Souza**

**Tatiane Ornelas Martins Alves**

**RESUMO**

**O presente texto apresenta uma análise do Exame Nacional de Desempenho dos Estudantes (ENADE) com o foco em encontrar uma correlação entre o desempenho do avaliado no exame e sua situação de trabalho.**

**Palavras-chave: Educação Superior. ENADE. Exame em larga escala.**

*ANALYSIS OF THE CORRELATION BETWEEN THE SITUATION OF WORK AND THE PERFORMANCE OF THE STUDENTS OF INFORMATION SYSTEMS AT ENADE*

***ABSTRACT***

**This paper presents an analysis of the National Assessment of Students Performance (ENADE) with focus to find a correlation between the assessed student performance in the exam and his work situation.**

***Keywords: Higher Education. ENADE. Large-scale assessment.***

\*Graduandos em Bacharelado em Sistemas de Informação pela Faculdade Metodista Granbery.

**1 - INTRODUÇÃO**

Cada vez mais percebe-se que as organizações vêm buscando definições de como se garantirem no mercado. Um dos principais meios de se destacar no mercado competitivo tem sido através da utilização de sistemas de informação, em destaque aqueles para análise de dados. Diante desse cenário se destaca a necessidade da utilização de ferramentas, que trabalhem os dados de interesse das empresas e auxiliem no processo decisório, identificação dos problemas, diagnósticos, listagem das alternativas de soluções, tomada de decisão e avaliação das decisões, entre outros. De posse dessas ferramentas, gestores podem ter uma tomada de decisão mais segura e precisa.

A importância de se trabalhar dados também é fundamental para instituições de ensino, visto que tais ferramentas podem auxiliar no melhor entendimento sobre o aprendizado de seus discentes, bem como saber sobre o desempenho destes em determinado conteúdo e comparar turmas a fim de encontrar o ponto de dificuldade no aprendizado. No Brasil, o INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) está ciente da importância da análise de dados e é atualmente o maior gerador de dados referentes à educação no país. Através do INEP é possível obter dados sobre a educação através de microdados que representa a menor fração de um dado e pode estar relacionado a uma pesquisa ou avaliação, essas informações contidas nos microdados são muito úteis para as instituições de ensino e servem para diversas tomadas de decisões. No entanto, existe uma grande dificuldade de se obter conhecimento útil devido à complexidade de extração e análise dos dados disponibilizados, o que dificulta uma análise manual, sendo assim é necessário ferramentas de sistema de informação para análise (NAMEN; BORGES; SADALA, 2013).

Dentro desse cenário, o presente trabalho se propõe a desenvolver uma solução de *Business Intelligence*(BI) que utilizará de dados fornecidos pelo INEP referentes ao ENADE (Exame Nacional de Desempenho dos Estudantes) para gerar conhecimento a fim de proporcionar ao usuário informações importantes sobre o tema, como, por exemplo, avaliar a influência de possuir um vínculo empregatício no desempenho dos estudantes, ao se realizar a análise comparativa das notas obtidas através do ENADE entre os que possuem vínculo e os que não possuem, a fim de descobrir o quanto o trabalho impacta na vida dos estudantes.

Pretende-se avaliar o desempenho dos estudantes que possuem algum vínculo empregatício como também o desempenho dos estudantes que estudam em uma IES(Instituição de Ensino Superior) pública ou privada, a distribuição das rendas familiares por IES pública e privada, quem possui o melhor rendimento dentre os alunos que trabalham e estudam em uma IES pública ou privada e a satisfação dos alunos em relação ao IES.

A contribuição trazida por esse trabalho é de grande relevância para as IES, pois com ela é possível que as mesmas melhorem em vários processos. A utilização do BI transforma os dados brutos das planilhas do INEP em informações úteis. Possibilitando aos gestores da educação decisões mais assertivas.

No Capítulo 1 discutiremos sobre o método utilizado para desenvolver este trabalho. No Capítulo 2 abordaremos sobre os conceitos que são necessários para o entendimento deste trabalho. No Capítulo 3 falaremos sobre o esquema que foi utilizado para o armazenamento dos dados e no Capítulo 4 comentaremos sobre os resultados obtidos através de nossa pesquisa.

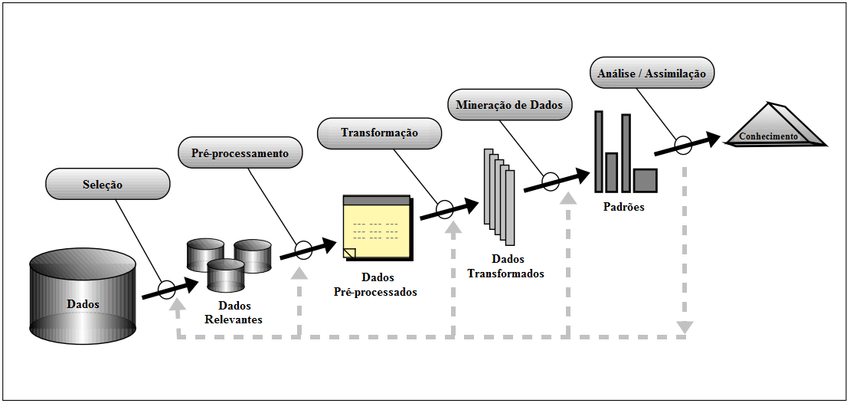
**2 - FUNDAMENTAÇÃO TEÓRICA**

Nessa seção são apresentados os principais conceitos necessários ao entendimento deste trabalho. Utilizamos de etapas do KDD(*Knowledge Discovery in Databases*) para executar uma análise exploratória pelo Microsoft Power BI. Foi modelada uma base de dados através do conceito *star schema*(esquema estrela) com os dados INEP visando melhorar a performance das consultas.

**2.1 - KNOWLEDGE DISCOVERY IN DATABASES – KDD**

O termo KDD significa descoberta de conhecimento em bancos de dados. Refere-se ao amplo procedimento de descoberta de conhecimento em dados, como pode ser observado na Figura 1, e enfatiza as aplicações de alto nível de técnicas específicas de Mineração de Dados. É um campo de interesse para pesquisadores em várias áreas, incluindo inteligência artificial, aprendizado de máquina, reconhecimento de padrões, bancos de dados, estatísticas, aquisição de conhecimento para sistemas especialistas e visualização de dados (PATRICIO, 2018).

**Figura 1.** O ciclo do processo de KDD

****

**Fonte.** Adaptado de Fayyad et al. (1996)

O principal objetivo do processo KDD é extrair informações dos dados no contexto de grandes bancos de dados.

A descoberta de conhecimento em bancos de dados é considerada uma análise e modelagem exploratória programada de vastos repositórios de dados. O KDD é o procedimento organizado de reconhecimento de padrões válidos, úteis e compreensíveis de conjuntos de dados enormes e complexos (TEÓFILO, 2015).

Utilizamos o processo de KDD para sermos capazes de processar, normalizar e analisar os dados disponibilizados pelo INEP para gerar conhecimento referente ao desempenho acadêmico dos alunos de Sistemas de Informação que estudam em escolas públicas e privadas.

**2.1.1 - CONSOLIDAÇÃO**

Foi feita uma pesquisa e aquisição de dados do ENADE disponibilizados pelo INEP entre os anos 2014 e 2018. Os dados disponibilizados estavam salvos em formato de arquivo de texto com a extensão (.txt), foi feita uma conversão para o formato .csv (*comma-separated-values*). Foi utilizada uma ferramenta de administração de banco de dados chamada DBeaver[[1]](#footnote-1) Os dados foram inseridos em um único banco de dados chamado microdados.

Foi feita uma busca pelas Instituições de Ensino Superior (IES) usando a ferramenta do Ministério da Educação e-MEC. Primeiro obtivemos um relatório de todas as IES em situação ativa, logo em seguida foi feita uma segunda consulta em busca de instituições extintas, pois como a consulta foi feita em 2020 algumas instituições que estavam ativas durante o ENADE 2014 não foram encontradas na primeira consulta.

**2.1.2 SELEÇÃO**

Uma vez que agrupamos os dados em uma única fonte, os dados que serão utilizados para o processo de descoberta de conhecimento devem ser determinados. Isso inclui descobrir quais os dados estão acessíveis, obter dados importantes e depois integrar todos os mesmos para a descoberta de conhecimento em um conjunto. Dados que serão usados como base para a criação dos modelos.

Se alguns atributos significativos estiverem faltando, nesse ponto, todo o estudo poderá não ter êxito e mais atributos serão considerados. Por outro lado, organizar, coletar e operar repositórios de dados avançados é caro e existe um acordo, ou seja, uma forma equivalente de tratar todos esses dados faltantes, com a oportunidade de melhorar e entender os fenômenos. Esse arranjo refere-se a um ponto em que os aspectos interativo e iterativo do KDD estão ocorrendo. Isso começa com os melhores conjuntos de dados disponíveis e depois expande e observa o impacto em termos de descoberta e modelagem de conhecimento (RIBEIRO, 2011).

Entendendo o processo de seleção e com acesso aos dados que serão utilizados nas próximas etapas, será definido os critérios de inclusão e exclusão, representados na Tabela 1.

**Tabela 1 – Critérios de inclusão e de exclusão**

|  |  |
| --- | --- |
| **Critérios de inclusão** | **Critérios de exclusão** |
| * Provas em que o curso de sistema de informação esteve presente. * Alunos que responderam à questão referente à situação de trabalho (exceto estágio ou bolsas). * Qual a renda total de sua família, incluindo seus rendimentos. * Quantas horas por semana, aproximadamente, o aluno dedicou aos estudos, excetuando as horas de aula. * Qual o principal motivo para você ter escolhido este curso. * Qual a principal razão para você ter escolhido a sua instituição de educação superior? * As disciplinas cursadas contribuíram para sua formação integral, como cidadão e profissional. | * Código da área de enquadramento do curso no Enade diferentes de 4006 (Sistemas de informação) * Alunos que não possuíam notas brutas e nem outras notas * Alunos que não responderam a questão QE\_I10(Qual alternativa a seguir melhor descreve sua situação de trabalho?) |

**2.1.3 - PRÉ-PROCESSAMENTO**

A confiabilidade nesta etapa se deve pelo fato de que os dados devem ser otimizados. Ela incorpora a limpeza de dados, por exemplo: Manipulação de quantidades ausentes e remoção de ruídos ou *outliers*. Pode incluir técnicas estatísticas complexas ou usar um algoritmo de mineração de dados nesse contexto (GOLDSCHMIDT; PASSOS, 2005).

Por exemplo, quando se suspeita que um atributo específico não possui confiabilidade ou tem muitos dados ausentes, nesse ponto, esse atributo pode se tornar o objetivo do algoritmo supervisionado de *Data Mining*. Um modelo de previsão para esses atributos será criado e, depois disso, os dados ausentes poderão ser previstos. A expansão à qual se presta atenção a esse nível depende de vários fatores. Independentemente disso, estudar os aspectos é significativo e regularmente revela por si só, para estruturas de dados corporativos.

Medir o desempenho acadêmico é um estudo complexo, é necessário analisar vários aspectos de vários grupos de indivíduos, aspectos socioeconômicos, sociodemográficos, institucionais e pedagógicos são apenas alguns exemplos que influenciam o rendimento do discente. O conhecimento adquirido desse estudo influencia o estabelecimento de políticas públicas que visam a melhor qualidade na formação dos estudantes do ensino superior e consequentemente na vida social após a conclusão do curso (Vargas, 2014; Valera et al., 2009; Urbina, 2014).

**2.1.4 - TRANSFORMAÇÃO**

As técnicas incorporam redução de dimensão (por exemplo, seleção e extração de recursos e amostragem de registros), também transformação de atributos (por exemplo, discretização de atributos numéricos e transformação funcional). Esta etapa pode ser essencial para o sucesso de todo o projeto KDD e geralmente é muito específica do projeto.

Esta etapa diferencia-se do KDD, pois não é utilizada a mineração de dados, mas sim a separação dos dados relevantes para responder às questões de investigação propostas por este trabalho.

A base de dados foi reduzida mantendo apenas as seguintes colunas:

1. NU\_ANO: Ano de realização do exame
2. CO\_IES: Código da IES (e-Mec)
3. CO\_GRUPO: Código da área de enquadramento do curso no Enade
4. CO\_MODALIDADE: Código da modalidade de Ensino
5. NT\_GER: Nota bruta da prova - Média ponderada da formação geral (25%) e componente específico (75%). (valor de 0 a 100)
6. CO\_RS\_I01: Qual o grau de dificuldade desta prova na parte de Formação Geral?
7. CO\_RS\_I02: Qual o grau de dificuldade desta prova na parte do Componente Específico?
8. QE\_I06: Onde e com quem você mora atualmente?
9. QE\_I08: Qual a renda total de sua família, incluindo seus rendimentos?
10. QE\_I09: Qual alternativa a seguir, melhor descreve sua situação financeira (incluindo bolsas)?
11. QE\_I10: Qual alternativa a seguir melhor descreve sua situação de trabalho (exceto estágio ou bolsas)?
12. QE\_I23: Quantas horas por semana, aproximadamente, você dedica aos estudos, excetuando as horas de aula?
13. QE\_I25: Qual o principal motivo para você ter escolhido este curso?
14. QE\_I26: Qual a principal razão para você ter escolhido a sua instituição de educação superior?
15. QE\_I27: As disciplinas cursadas contribuíram para sua formação integral, como cidadão e profissional.

Também foi criado uma coluna chamada PERIODO contendo as colunas referentes ao turno do curso. Para a avaliação 2014 as colunas eram IN\_MATUT, IN\_VESPER, IN\_NOTURNO para os períodos matutino, vespertino e noturno. Para o ano 2017 a coluna referente ao turno era a coluna CO\_TURNO\_GRADUACAO.

**2.1.5 - AVALIAÇÃO**

Nesta etapa, os padrões são avaliados e interpretados, regras e confiabilidade extraídas para o objetivo caracterizado na primeira etapa. Neste estágio, consideramos as etapas de pré-processamento quanto ao seu impacto. Inicialmente são avaliados os resultados do algoritmo de mineração de dados. Como o objetivo do nosso projeto é criar uma solução de BI, não utilizamos mineração de dados e sim avaliamos os resultados obtidos pela solução. Esta etapa se concentrou na compreensibilidade e utilidade dos resultados. Assim, o conhecimento identificado também é registrado para uso posterior.

**3 - ESQUEMA ESTRELA**

Idealizado por Ralph Kimball, criador da Modelagem Dimensional, precursor dos sistemas para análise de dados e *Data Warehouse*, o *Star Schema* ou Esquema Estrela, tem esse nome pois o diagrama de entidade-relacionamento simula uma estrela. No diagrama existe uma tabela central de Fatos no centro, ligada por tabelas contendo as dimensões dos dados. É um dos modelos mais utilizados na modelagem dimensional, seu objetivo é melhorar a performance de sistemas voltados para consultas e auxiliar na tomada de decisão (KIMBALL *et al*., 1998).

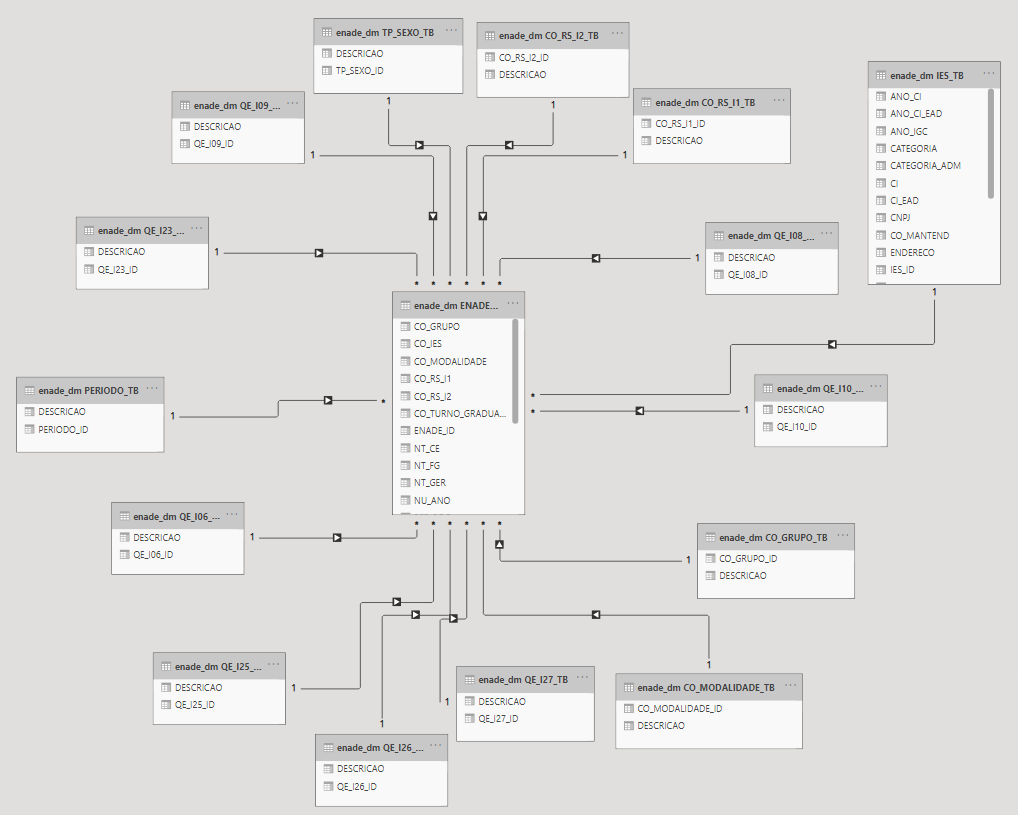
Características do esquema estrela:

* Dados sem redundância em uma única tabela de fatos;
* A cada dimensão uma tabela;
* A chave primária da tabela de fatos geralmente é uma chave composta por todas as suas chaves estrangeiras;
* Cada chave das tabelas de fatos é gerada pelo sistema, que atuam como um identificador exclusivo, pois podem haver casos em que existem dois ou mais fatos e granularidade idênticos;
* Cada dimensão caracteriza apenas uma tabela, altamente desnormalizada;
* Devido a sua simplicidade estrutural há um ganho de tempo durante o carregamento de grandes lotes de registro;
* Ganho em desempenho durante a consulta;
* Fácil de ser compreendido.

Levando em consideração todos esses benefícios citados acima, após a consolidação dos dados, eles foram organizados seguindo o modelo conceitual do esquema estrela.

Na Figura 3 é possível observar o banco de dados feito com base no Esquema Estrela.

**Figura 2.** Estrutura da base de dados

**~~~~**

**3.1 - ARQUITETURA DA SOLUÇÃO DE BI PROPOSTA**

A partir dos dados do INEP é realizado um processo de Extração, Transformação e Carga desses dados para poderem ser armazenados em um formato adequado ao processo decisório, denominado Esquema Estrela, deixando os dados prontos para serem utilizados para extração e análise de dados (ALVES, 2018). O banco de dados utilizado é o MariaDB, que está hospedado no Amazon Web Services (AWS).

Foi utilizado o serviço de análise de negócios da MicroSoft Power BI e dos dados devidamente filtrados para realizar a construção da solução de análise de dados do ENADE. Com a solução devidamente implementada, é dada permissão aos usuários consultarem e verificarem o conhecimento gerado a partir das informações extraídas dos microdados do INEP.

**4 - TRABALHOS RELACIONADOS**

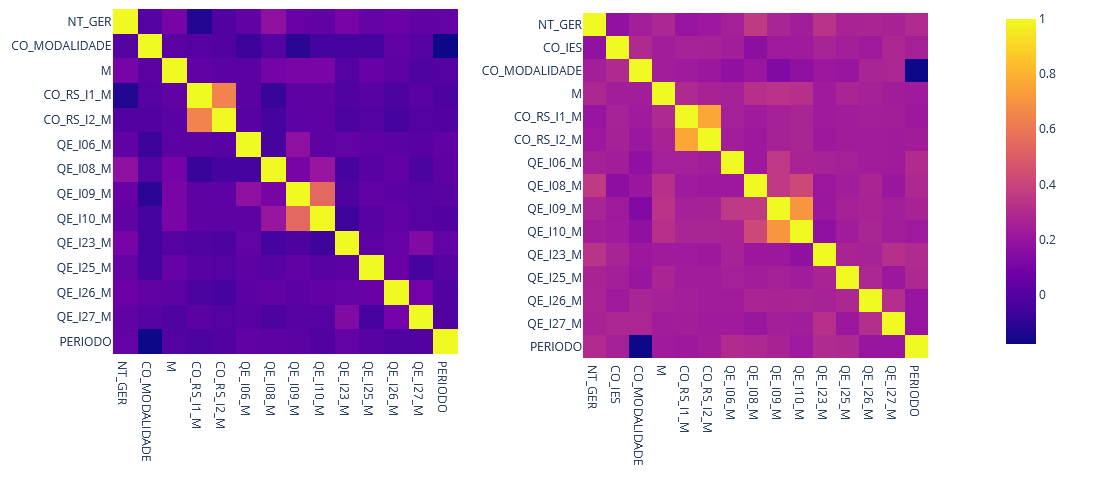
Esta seção visa apresentar alguns trabalhos relacionados com os tópicos abordados anteriormente.  
 O ENEM passou também a utilizar o sistema de *data mining* para observar o nível de desempenho dos alunos na temática de redação através do uso do software Weka que se pode prever 61% das amostras da base de dados de 2016 (ALVES, 2018). Esse projeto se assemelha ao nosso pelo fato de também trabalhar com análise de dados em uma prova de âmbito nacional.

O trabalho teve como objetivo analisar como se encontra o curso de Sistemas de Informação (SI) do Câmpus de Ciências Exatas e Tecnológicas (CCET) da Universidade Estadual de Goiás (UEG) em relação aos cursos de SI do país, correlacionando a situação socioeconômica com o bom desempenho (notas maiores ou igual a 50) no Exame Nacional de Desempenho dos Estudantes (ENADE). (RIBEIRO, 2018). Esse projeto se assemelha muito ao nosso, pois ele também trabalha com análise de dados, utiliza o ENADE e é uma análise de desempenho acadêmico. Seu ponto diferencial é o meio que ele utiliza para analisar o desempenho acadêmico e pelo fato de comparar apenas uma única universidade.

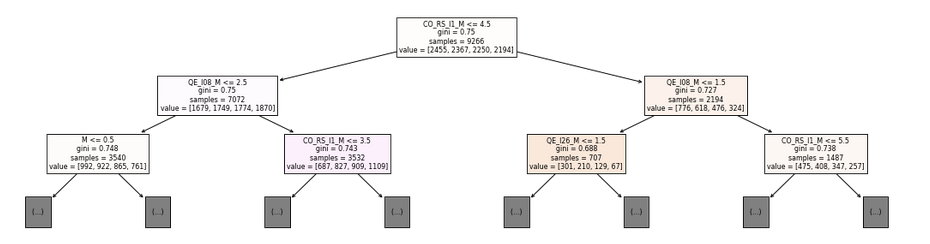
**5 - RESULTADOS OBTIDOS**

**5.1 Mapa de Calor de Correlação e árvore de classificação**

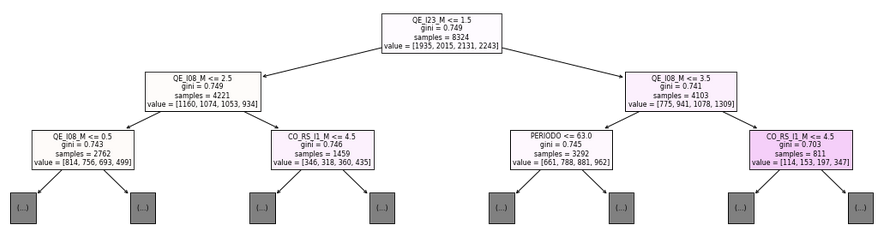
**Figura 3.** Mapas de Calor de Correlação para os anos 2014 e 2017 respectivamente



**Figura 4.** Árvore de classificação para o ano de 2014



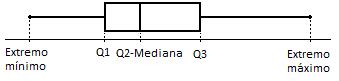
**Figura 5.** Árvore de classificação para o ano de 2017



Foram utilizadas as bibliotecas Pandas e SciKit-Learn da linguagem Python para medir a correlação entre as variáveis selecionadas e a nota geral (NT\_GER), foi possível verificar que os dados possuem uma fraca correlação, como é possível verificar no mapa de calor de correlação (Figura 3). Foi criada uma árvore de classificação para verificar qual variável tinha mais impacto na nota, foi verificado que seria a coluna referente a renda total da família (CO\_RS\_I01) para o ano 2014 (Figura 4) e tempo de estudo (QE\_I26) para o ano 2017 (Figura 5), mas temos que levar em consideração que as árvores tinha uma baixa precisão (0.275 para 2014 e 0.2595 para 2017).

**4.2 - Gráfico de caixa ou Box Plot**

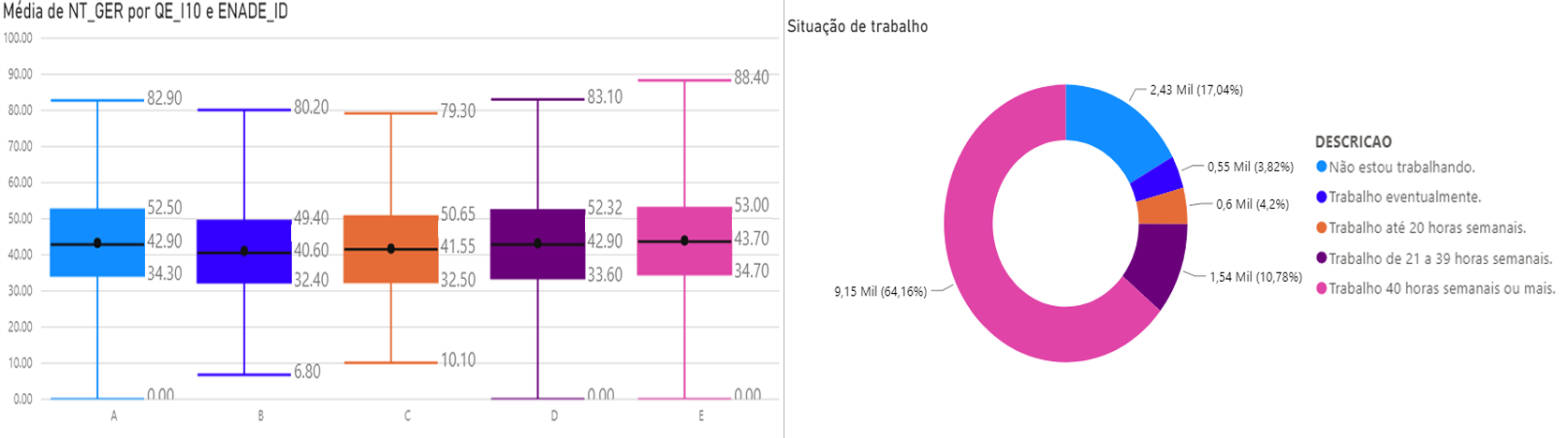
Foi utilizado o gráfico de caixa ou box plot na Figura 6 para avaliar a distribuição dos dados. Ele é composto por 6 pontos. O 1°quartil 25% dos dados são menores que ou iguais a este valor, 2°quartil ou mediana 50% dos dados são menores que ou iguais a este valor e o 3º quartil 75% dos dados são menores que ou iguais a este valor, também possui a máxima e a mínima que é o maior e o menor valor dentre todos valores. Em nossos gráficos foi omitido a média para facilitar a interpretação dos dados.

**Figura.6** Box Plot  


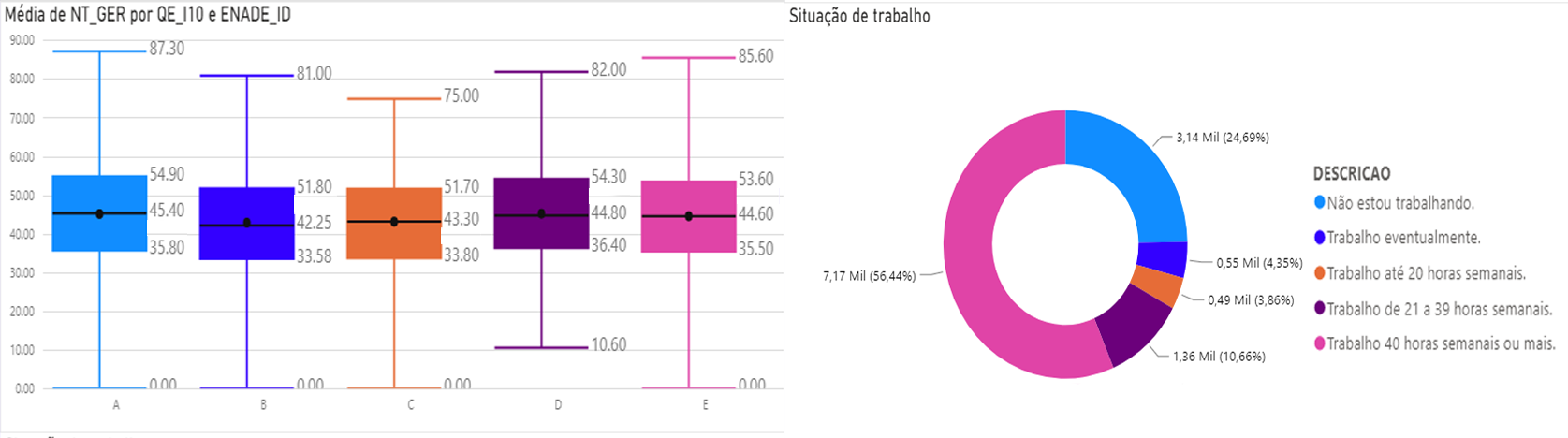
**Fonte.** AEE (Agrupamento De Escolas De Ermesinde)

**4.3 - Existe uma relação entre trabalho e desempenho acadêmico.**

**Figura 7.** Desempenho dos alunos por situação acadêmica em 2014

****

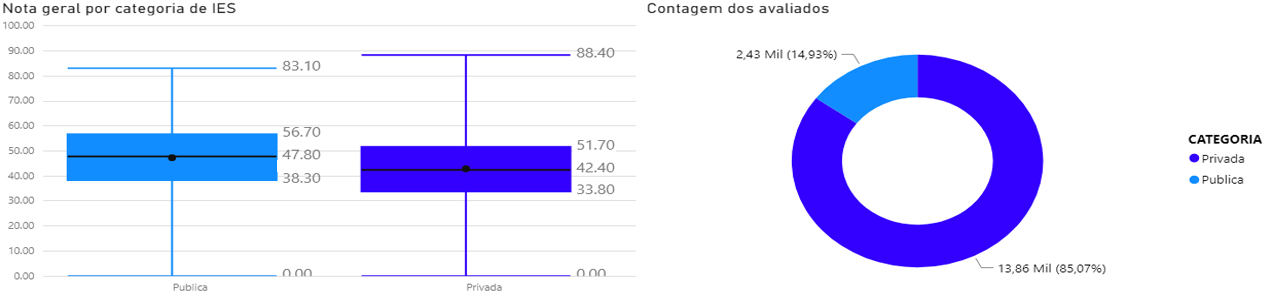
**Figura 8.** Desempenho dos alunos por situação acadêmica em 2017

****

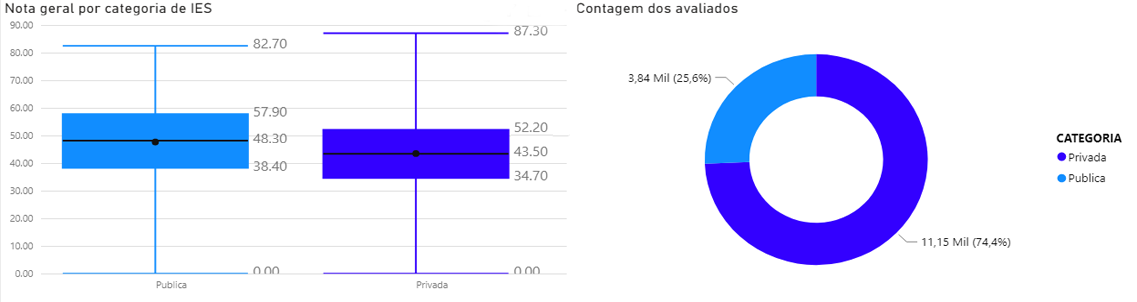
Como pôde ser visto nas figuras 7 e 8, temos uma diferença fraca em relação às notas quando levamos em consideração a condição de trabalho do aluno. Em 2014, foi possível observar que os avaliados que não trabalham tem uma nota um pouco menor dos que não trabalham 40 horas semanalmente, mas em 2017 o cenário se inverte.

**4.4 - Desempenho dos alunos em IES pública e privada.**

**Figura 9.** Desempenho dos alunos em IES pública e privada 2014



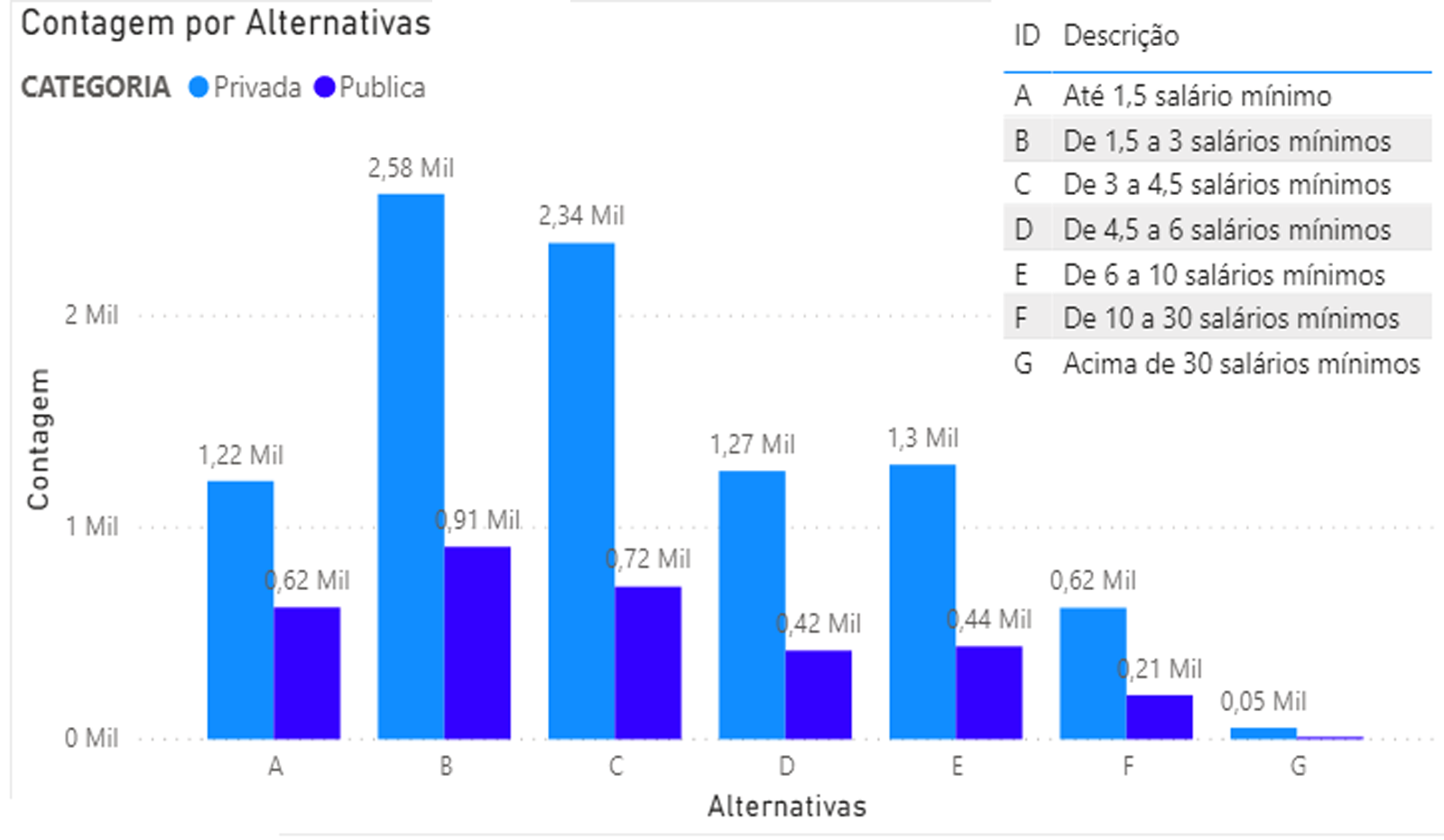
**Figura 10.** Desempenho dos alunos em IES pública e privada2017



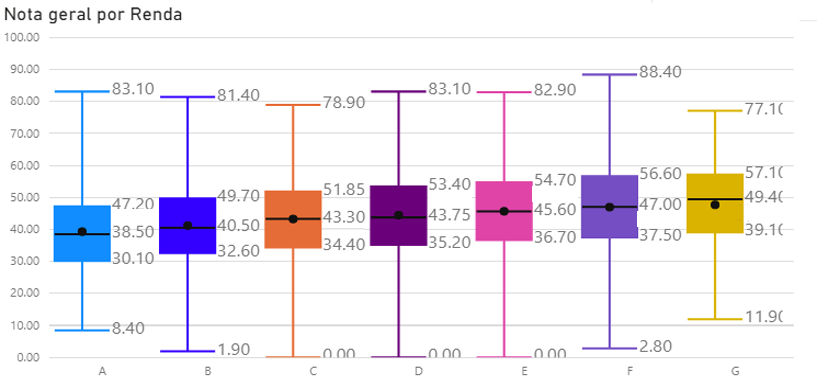
É possível observar nas Figuras 9 e 10 que nos dois anos há uma predominância na quantidade de alunos que estão estudando em instituições privadas, em contrapartida as notas daqueles que estudam em instituições públicas são maiores.

**4.5 - Distribuição das rendas das famílias em IES públicas e privadas**

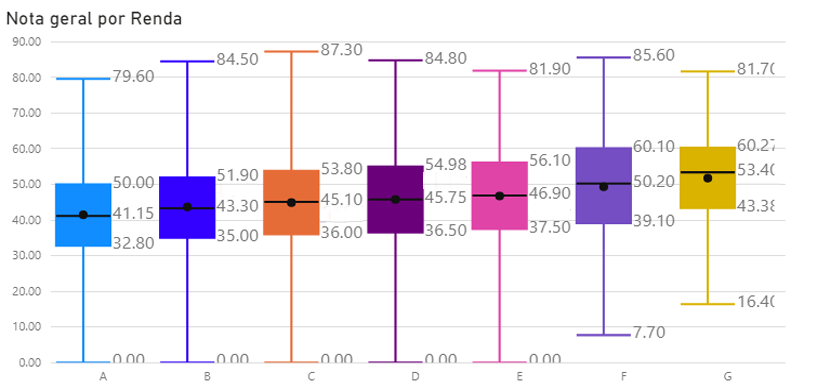
**Figura 11.** Contagem dos alunos por renda e instituição



**Figura 12.** Desempenho do aluno por renda 2014



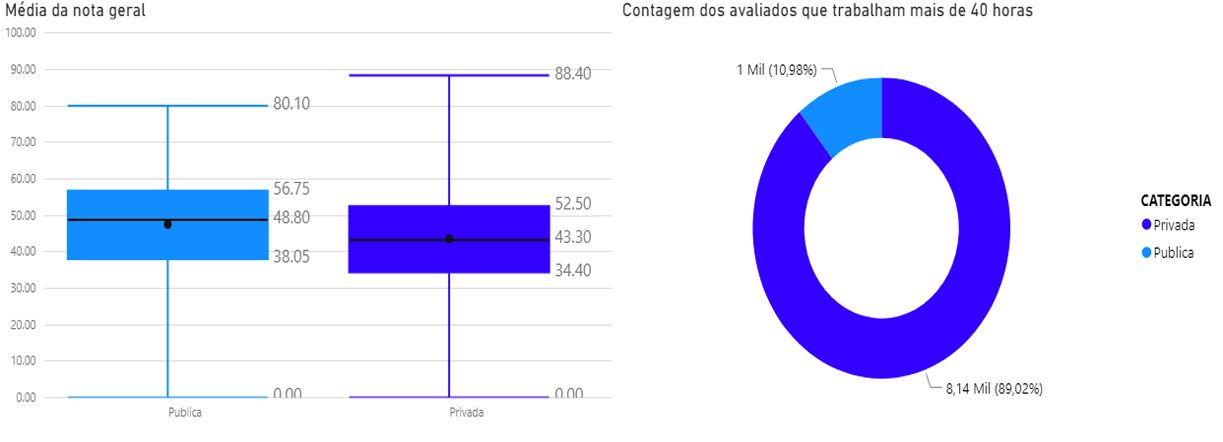
**Figura 13.** Desempenho do aluno por renda 2017

****

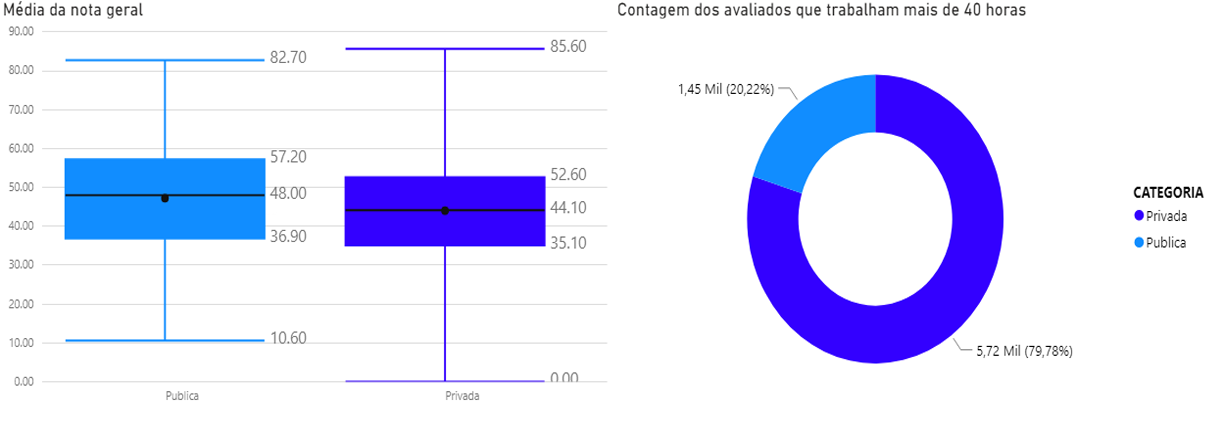
Em relação a renda dos avaliados é possível observar uma leve melhora da nota, a medida que a renda familiar aumenta, a nota da renda A (Até 1,5 salário mínimo) é inferior à renda G (Acima de 30 salários mínimos) como visto nas Figuras 12 e 13 e verificando as legendas na Figura 11. Mas não é possível dizer que isso é uma relação direta com a renda como vimos no mapa de calor de correlação (Figura 4). Além da renda, existem outras variáveis que causam um certo impacto na nota, como o tempo de estudo e a qualidade da educação durante os ensinos médio e fundamental.

**4.6 - Dos alunos que trabalham o rendimento acadêmico é melhor do que os que estudam em IES pública e privada?**

**Figura 14.** Rendimento acadêmico dos alunos que trabalham e estudam instituições Públicas e Privadas no ano 2014

****

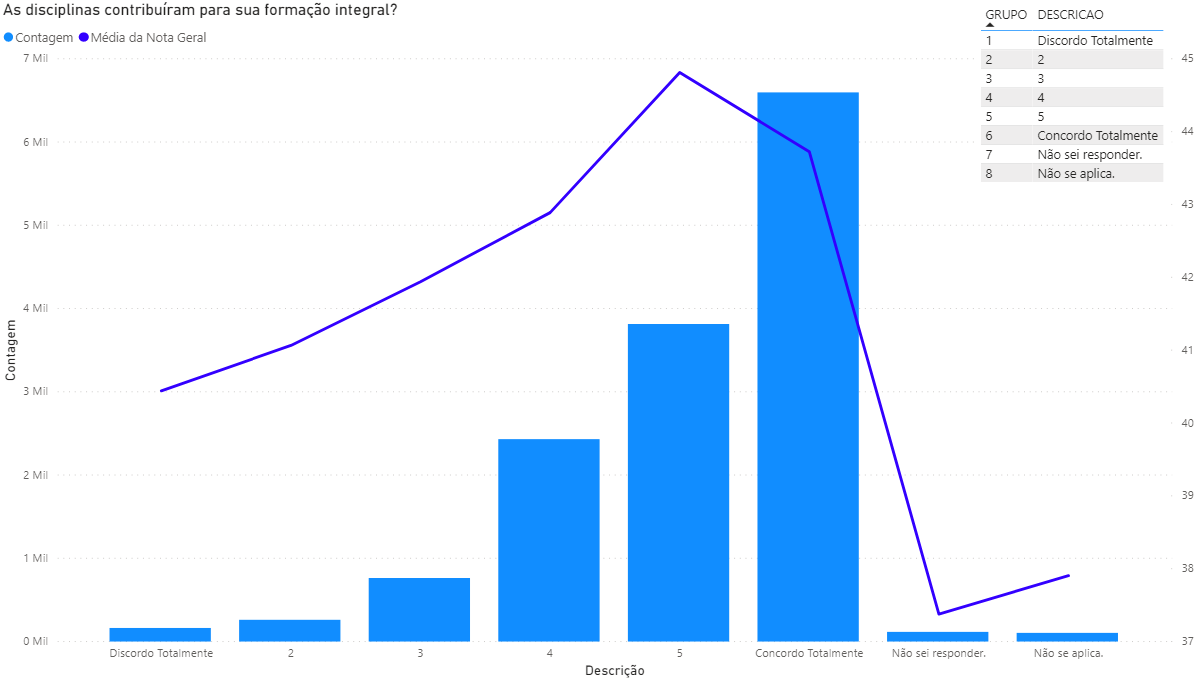
**Figura 15.** Rendimento acadêmico dos alunos que trabalham e estudam instituições Públicas e Privadas no ano 2017

****

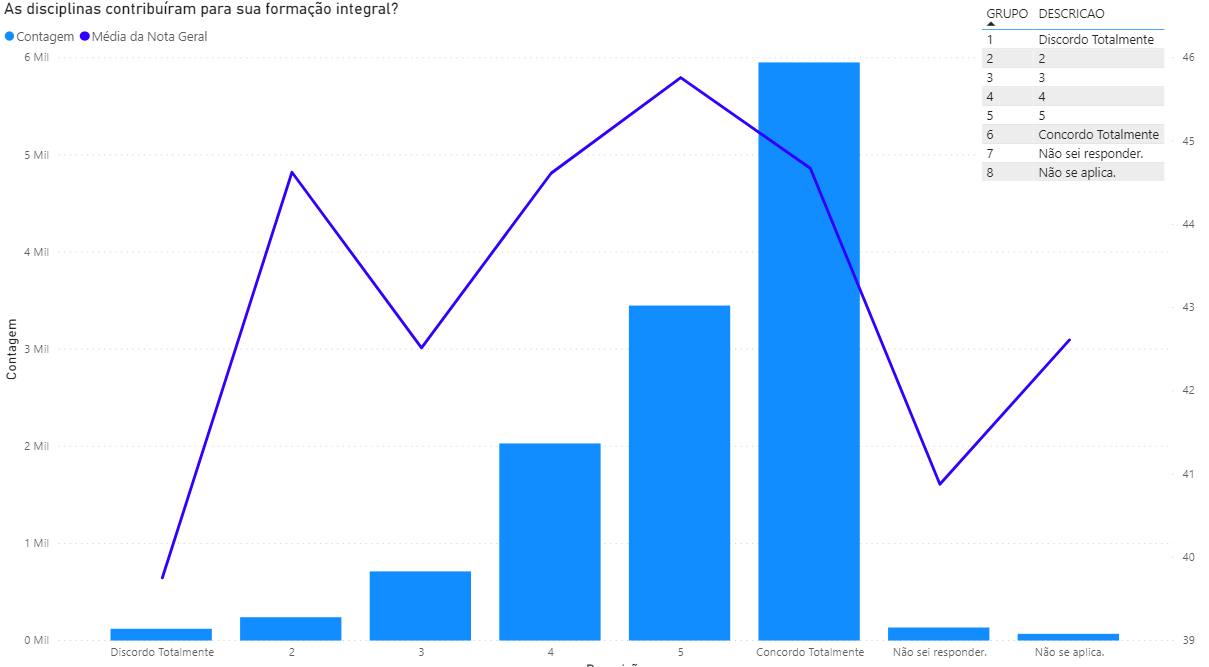
Como foi visto na Seção 4.4, a nota dos alunos das instituições públicas é um pouco maior do que a dos alunos de instituições privadas. Os alunos das instituições públicas possuem um menor extremo mínimo e um maior Q1, Q3 e mediana em relação aos alunos de escolas privadas, mas foi na escola privada que encontramos o maior extremo máximo Figura 14 e 15.

**4.5 - Qual a satisfação do avaliado em relação a contribuição do curso da IES**

**Figura 16.** Contribuição para formação em 2014

****

**Figura 17.** Contribuição para formação em 2017



A partir dos gráficos das Figuras 16 e 17 é possível observar que a maioria dos alunos acredita que o curso de Sistemas de Informação contribuiu para sua formação de forma geral. Em 2014 a nota média dos alunos está crescendo junto e exponencial com o nível de concordância com exceção da anomalia em 2017.

**5 - CONSIDERAÇÕES FINAIS**

O artigo investigou a complexa base de dados do ENADE, com foco nos alunos de Sistemas de Informação nos anos de 2014 e 2017, com o intuito de descobrir se o trabalho impacta na vida do estudante.

Inicialmente a ideia da ferramenta descrita no artigo seria capaz de analisar todos os cursos, em busca de informações relevantes para os *stakeholders* a partir de um formulário de busca. Essa ideia teve que ser descartada, pois seria algo que demandaria uma grande quantidade de tempo e um hardware mais potente. Após discussões e análise dos dados à nossa disposição, os cursos foram filtrados, assim mantendo apenas o curso de Sistemas de Informação e a granularidade das IES ficou no nível jurídico (Faculdade Privada sem fins lucrativos, Pessoa Jurídica de Direito Público - Federal, etc).

Após utilizarmos do Python com auxílio de bibliotecas como pandas e sklearn junto com o Power BI foi possível identificar uma fraca correlação entre a nota geral do estudante e a situação de trabalho, expandimos nossa busca para outras questões de investigação e chegamos na mesma conclusão. A nota do estudante deve-se a diversas variáveis e não pode ser resumida a um único fator, mostrando assim a complexidade da área da educação. As informações disponibilizadas sobre as questões de investigação são de grande valia para futuros estudos focados na área de educação.

**REFERÊNCIAS**

NAMEN, A.; BORGES, S.; SADALA, M. Indicadores de qualidade do ensino fundamental: o uso das tecnologias de mineração de dados e de visões multidimensionais para apoio à análise e definição de políticas públicas. **Revista Brasileira de Estudos Pedagógicos**. Brasília , v. 94, n. 238, p. 677-700, Dezembro. 2013 . Disponível em: http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S2176-66812013000300003&lng=en&nrm=iso. Acesso em: 18 mar 2020.

ALVES, R.D. **Predição do desempenho da redação do enem utilizando técnica de mineração de dados**. Orientador: Cristian Cechinel. 2018. 67 f. Trabalho de conclusão de curso (Bacharel em Tecnologias da Informação e Comunicação) - Universidade Federal de Santa Catarina, Florianópolis, 2018.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining:** um guia prático. 1. ed. Rio de Janeiro: Contexto, 2005.

KANTARDZIC, M. **Data Mining**:Concepts, Models, Methods, and Algorithms. 2. ed. Universidade de Louisville: Wiley–Blackwell, 2011.

KIMBALL, R. *et al.* **The Data Warehouse** **Lifecycle Toolkit**. 2. ed. Safari Books Online: John Wiley & Sons, 1998. 672p.

ACADEMIA, Qedu. **O que são microdados?.** QEdu Academia, 2020. Disponível em: https://academia.qedu.org.br/glossario/o-que-sao-microdados. Acesso em: 18 mar 2020.

RIBEIRO, L.S. **Uma abordagem semântica para seleção de atributos no processo de KDD**. Orientador: Valéria Gonçalves Soares. 2010. 121f. Dissertação (Pós-Graduação em Informática) - Universidade Federal da Paraíba, João Pessoa, 2010.

TEÓFILO, Daniel. **KDD - Knowledge Discovery in Database**. 2015. Disponível em: https://danielteofilo.wordpress.com/2015/02/16/kdd-knowlegde-discovery-in-database/. 2015. Acesso em: 19 mar 2020.

VERSIANE, Daniela. Data Mining. **Blog Edital Concursos Brasil**. [*S. l.*], [20--].Disponível em: https://editalconcursosbrasil.com.br/blog/tecnologia\_data-mining/. Acesso em: 18 mar 2020.

1. https://dbeaver.io/ [↑](#footnote-ref-1)