

Relação entre proficiência do Enem e situação socioeconômica dos candidatos ao ensino superior por meio do algoritmo k-means

Marília Magalhães Maia¹, Luiza Helena Felix de Andrade², Fabricio Figueiredo de Oliveira³.

Resumo: Diante das desigualdades que afetam os estudantes brasileiros, o presente trabalho trata de analisar as circunstâncias que dividem os candidatos as vagas nas instituições de ensino superior através dos seus rendimentos no Exame Nacional do Ensino Médio (Enem), por meio da implementação das notas referentes as 5 áreas avaliadas pelo exame, aplicado em 2018, no algoritmo de agrupamento k-means. Em vista disso, os dados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) que guarda informações sobre os inscritos no exame, são usados para a implementação e análise sobre o agrupamento resultante do algoritmo. Como resultado da divisão de uma amostra dos inscritos no exame em 2 grupos, observa-se a renda, a cor/raça e a instituição que o aluno frequentou durante o ensino médio como sendo a circunstância que os dividem. Diante disso, a relação entre o resultado encontrado e as cotas redigidas pela Lei nº 12.711, apresentando o resultado como argumento da necessidade dessas para fornecer o poder de competitividade à população estudantil mais atingida pelas desigualdades presentes na realidade brasileira.

Palavras-chave: Enem; K-means; Cotas; Análise de dados; Políticas Sociais.

1. INTRODUÇÃO

O Exame Nacional do Ensino Médio (Enem), criado em 1998 pelo Ministério da Educação (MEC) tem como objetivo “avaliar o desempenho do estudante ao fim da escolaridade básica. ”, segundo o próprio MEC[1]. Entretanto, passou a ter finalidades diferentes da idealizada em sua criação, quando, em 2005, tornou-se porta de acesso a bolsas de estudo em instituições de ensino superior privadas, pelo Programa Universidade para Todos (ProUni), no qual as notas dos candidatos são os principais parâmetros para a seleção. Assim como, em 2009, com a criação do Sistema de Seleção Unificada (Sisu), pelo MEC, dando início ao “novo Enem”, se tornando a partir de então, o principal exame seletivo para ingresso ao ensino superior do país [2] [3].

Entretanto, tendo conhecimento das desigualdades que afetam a população brasileira, quanto as diferenças de oportunidades entre raças e rendas e o desnível entre a educação privada e pública, diversas instituições desenvolveram ações afirmativas, fornecendo cotas nas vagas de ingresso, para o público afetado pelas desigualdades [4]. Diante do cenário de diferentes ações afirmativas, no dia 29 de agosto de 2012, foi sancionada a Lei das Cotas (Lei nº 12.711), a qual em seu Art. 1º diz:

“As instituições federais de educação superior vinculadas ao Ministério da Educação reservarão, em cada concurso seletivo para ingresso nos cursos de graduação, por curso e turno, no mínimo 50% (cinquenta por cento) de suas vagas para estudantes que tenham cursado integralmente o ensino médio em escolas públicas.

Parágrafo único. No preenchimento das vagas de que trata o caput deste artigo, 50% (cinquenta por cento) deverão ser reservados aos estudantes oriundos de famílias com renda igual ou inferior a 1,5 salário-mínimo (um salário-mínimo e meio) *per capita*” [5]

Em seguida, no Art. 3º é expresso que dentre as vagas referidas no art. 1º dessa Lei, devem ser preenchidas por turno e curso, por autodeclarados pretos, pardos e indígenas e pessoas portadoras de deficiências seguindo a proposição de vagas às proporções de pretos, pardos e indígenas e pessoas portadoras de deficiências da população da unidade da Federação a qual se apresenta, seguindo o mais recente censo do Instituto Brasileiro de Geografia e Estatística (IBGE).

No entanto, a Lei das Cotas é questionada desde o início do processo de criação, por grupos que a consideram inconstitucional e que diminui a qualidade dos alunos selecionados, já que estes, por sua vez, são oriundos de uma educação de menor qualidade em comparação as instituições de ensino básico particular. Por outro lado, os grupos a favor da Lei, alegam ser uma medida necessária na tentativa de equiparar as oportunidades, fornecendo poder de competitividade justa entre os estudantes, assumindo a desigualdade racial, de renda e educacional como uma realidade nacional e de tendência cíclica, caso o não tenha intervenção do governo [4].

Diante desse debate, o presente trabalho tem como objetivo analisar as características que dividem os candidatos às vagas nas instituições de ensino superior através dos seus rendimentos no Enem, por meio da implementação das notas referentes as 5 áreas avaliadas pelo exame, aplicado em 2018, no algoritmo de agrupamento k-means [6].

Dessa forma, será trabalhado no primeiro momento de uma revisão a literatura, apresentando um breve levantamento histórico do Enem e da Lei das Cotas, além de uma seção apresentando o algoritmo k-means. Em seguida, será apresentado o método de pesquisa, descrevendo quais os parâmetros utilizados no refinamento dos dados e suas justificativas. Como sequência, a apresentação dos resultados encontrados e a discussão sobre o mesmo, em busca de compreender como se dividem os grupos e correlacionar ao contexto do debate sobre as cotas. Ademais, neste momento apresentará, também, os resultados encontrados quando realizado o mesmo procedimento sobre os dados de 2019 e a justificativa do uso dos dados do ano anterior como principal fonte de análise. E então, apresentar as considerações finais e temas para pesquisas futuras.

2. DESENVOLVIMENTO

2.2 Fundamentação Teórica

Durante o governo de Fernando Henrique Cardoso ocorreram diversas mudanças na educação. Tendo como início a criação da Lei de Diretrizes e Bases da Educação Nacional (LDBEN) em 1996, passando a reger a educação básica com a finalidade de “desenvolver o educando, assegurar-lhe a formação comum indispensável para o exercício da cidadania e fornecer-lhe meios para progredir no trabalho e em estudos posteriores” (Lei n 9.394/96, art.22). A reforma do ensino médio, guiada pelas Diretrizes Curriculares Nacionais do Ensino Médio ocorrida em 1998 e as criações, no mesmo ano, das avaliações padronizadas da educação, como o Exame Nacional do Ensino Médio (Enem) foram importantes mudanças ocorridas nesse período [7].

O Enem, desenvolvido pelo Ministério da Educação (MEC), através do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) como ferramenta para “avaliar o desempenho do estudante ao fim da escolaridade básica” [1], teve, também, como finalidade o acesso à educação superior de instituições privadas, que fizeram uso das notas dos candidatos no processo seletivo. Sendo um exame voluntário, teve um considerável aumento na participação quando, em 2001, os alunos da rede pública de ensino passaram a ser isentos a taxa de inscrição. Na primeira aplicação alcançou menos de 160 mil inscrições, aumentando para cerca de 1,6 milhões no primeiro ano de isenção [8]. Ainda segundo [8], o próximo aumento significativo de inscritos no exame ocorreu em 2005, após a criação do Programa Universidade para Todos (ProUni), no ano anterior, quando foi alcançado o marco de 3 milhões de inscritos. O programa “oferece bolsas de estudo integrais ou parciais (50%) em instituições particulares de ensino superior” para estudantes do ensino médio da rede pública ou de escolas particulares com bolsas integrais fornecidas pela própria escola e estudantes portadores de deficiências, com renda familiar bruta mensal de até a 1,5 salários-mínimos, brasileiros e que ainda não possuam diploma de ensino superior, além disso os estudantes devem ter participado da edição mais recente do Enem, obtido nota média de até 450 e não ter zerado a redação [9].

Em 2009 o Enem passou por reformulações, além de torna-se o processo de seleção unificada das universidades públicas federais. A reformulação tinha como proposta a “democratizar as oportunidades de acesso às vagas federais de ensino superior, possibilitar a mobilidade acadêmica e induzir a reestruturação dos currículos do ensino médio [10]”, além de dar autonomia às universidades na escolha do uso do Enem como processo seletivo. Segundo [10], a partir de então, elas podiam escolher se usariam as notas dos alunos como:

- Fase única, com o sistema de seleção unificada, informatizado e on-line;
- Primeira fase;
- Combinação com o vestibular da instituição;
- Fase única para as vagas remanescentes do vestibular.

Além disso, para o processo de seleção, o MEC criou o Sistema de Seleção Unificada (Sisu), o qual une as vagas disponibilizadas pelas instituições de ensino superior públicas oferta para os candidatos que fizeram o Enem, de forma que os candidatos com melhores notas são selecionados e cada instituição estabelece os padrões de peso dado a cada área avaliada pelo exame para cada curso disponibilizado [3].

Diante às mudanças, o “novo Enem”, passou a possuir 180 questões distribuindo-se entre as áreas, linguagens e códigos, matemática, ciências da natureza e ciências humanas, além da redação. Além disso, passou a ser aplicado em dois dias, diferentemente de quando foi criado, quando era aplicado somente em um dia, mas possuía apenas 63 questões [7]. Ainda mais, é válido ressaltar a implementação da Teoria de Resposta ao Item (TRI) como

uma mudança ocorrida durante a reformulação, o que permitiu a comparação das notas com as aplicações de anos futuros, já que esta permite que o exame mantenha o mesmo grau de dificuldade [11]. Dessa forma, passou a ser possível comparar o desempenho dos alunos que concluem o ensino básico ao passar dos anos e traçar os perfis dos estudantes brasileiros.

Sendo a desigualdade uma realidade presente no cotidiano do povo brasileiro, no âmbito da educação ela se apresenta nitidamente. Esta, por sua vez, sendo elitizada desde a colonização portuguesa, período no qual apenas as famílias ricas e consequentemente, constituídas de pessoas brancas, tinham condições de fornecer educação aos jovens, apresenta ainda nos tempos atuais fortes resquícios dessa elitização. Segundo o Censo de 2015 [12], no âmbito da educação superior é onde se nota os números mais alarmantes, sendo a frequência da população branca mais que o dobro dos percentuais em relação a população negra, correspondendo respectivamente a 23,4% e 10,7%. Além disso, segundo [12], [13], os pretos e pardos constituem a maior porcentagem com relação a distribuição de renda e condições de moradia de pessoas abaixo da linha da pobreza, seguindo o mesmo padrão quanto ao analfabetismo entre pessoas com 15 anos ou mais, em todos os casos com percentuais maiores que o dobro da população branca. Já a população indígenas que vive em suas terras apresenta 67,7% de analfabetismo entre os jovens da mesma faixa etária já citada [12]. Além disso, segundo o Censo Escolar de 2015,

“As populações negra e indígena, apesar dos avanços recentes, ainda enfrentam dificuldades em acessar e permanecer nos diversos espaços educacionais. A desigualdade observada nos indicadores educacionais evidencia como a inclusão do campo cor/raça é capaz de garantir maior detalhamento na análise do perfil educacional dos brasileiros.”

Diante desse cenário de desigualdade, em 1999 a então deputada Nice Lobão, na época do partido Democrata (DEM), apresentou o Projeto de Lei (PL nº 73/99) o qual, em maio 2004, o então Presidente Luiz Inácio Lula da Silva apresentou ao Congresso Nacional, como Projeto de Lei nº 3627, que previa a reserva de 50% das vagas das instituições públicas federais de ensino superior para estudantes de escolas públicas, subdividindo-se metade-metade entre aqueles que possuem renda *per capita* até 1,5 salários-mínimos e aqueles que não possuem, internamente a essa subdivisão ainda há uma outra, quanto a cor/raça dos candidatos, entre aqueles que são pardos, pretos ou indígenas e que não são, essa divisão ocorrendo de forma proporcional aos percentuais das etnias nas unidades da Federação, de acordo com o último censo do Instituto Brasileiro de Geografia e Estatística (IBGE) [5]. A fim de facilitar a compreensão do funcionamento das cotas a Figura 1 ilustra a distribuição.

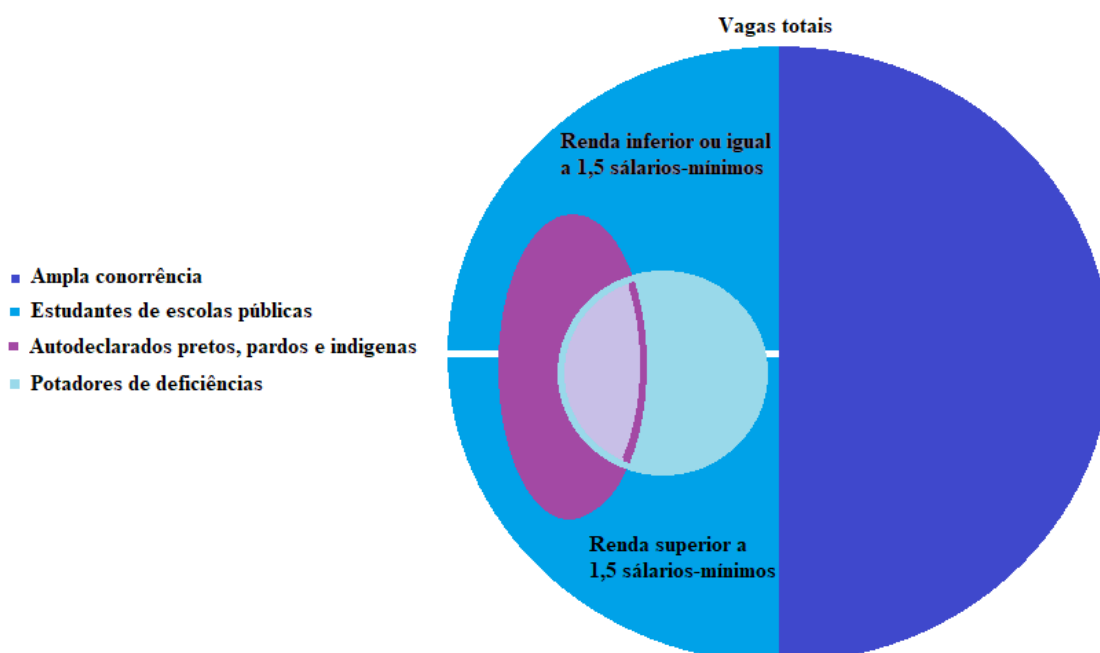


Figura 1 - Distribuição das cotas (autoria própria).

Entretanto, durante o período de tramites do projeto, diversas instituições de ensino superior promoveram ações afirmativas que buscavam minimizar os impactos da desigualdade, fornecendo cotas e auxílios aos grupos afetados por aquela. E então, em 29 de agosto de 2012 a Lei das Cotas (Lei n. 12.711/12) foi sancionada pela então

Presidenta Dilma Rousseff, passando a ser implementada no ano seguinte e com sua totalidade implementada até 2016, seguindo os mesmos parâmetros de seu Projeto de Lei. Entretanto, esta, assim como as demais leis já criadas que disponibilizava cotas nas universidades públicas foram alvo de contestações jurídicas, alegando serem inconstitucionais, porém, descordadas pelo Supremo Tribunal Federal (STF) [14].

Um outro argumento contra a implementação das cotas no ensino superior refere-se à diminuição da qualidade dos alunos selecionado, o que provocaria uma redução da qualidade do ensino superior. Entretanto, estudos comprovam a não relação entre os alunos que ingressam nas instituições pelas cotas e a qualidade da formação, como apresentado em [15], já que as cotas ainda promovem a competição pelas vagas disponibilizadas e apenas aqueles que possuem as melhores notas as conseguem. Além disso, também segundo [15], “apesar de o sistema de cotas retirar alunos com notas maiores, a diferença de notas é marginal ” e há entre os beneficiados pelas cotas, alunos com notas tão boas quanto as dos não beneficiados.

Diante disso, a análise sobre as circunstâncias dos candidatos a vagas nas instituições pela nota do Enem que se dividem em 2 grupos pelo algoritmo k-means, se faz uma discussão necessária.

2.1.1 Algoritmo K-means

Nos últimos anos houve um aumento do volume de dados produzidos diariamente. Diante disso, existe a necessidade da manipulação destes para as mais variadas finalidades, dentre elas a necessidade da agrupar, a qual estimula o processo de aprendizado de máquinas, o qual se trata dos processos computacionais embasados pelos conhecimentos estatísticos para a formulação de algoritmos, através do uso de dados de treinamento ensinando como devem manipular os dados a serem inseridos [16]. O aprendizado de máquinas se dividem em duas classificações dos algoritmos desenvolvidos, as quais se nomeiam aprendizado supervisionado e não-supervisionado, diferenciando-se pelos dados a os quais trabalham, o primeiro manipula aqueles que são rotulados, ou seja, aqueles que suas características de classificação são conhecidas previamente, enquanto o segundo é aplicado a dados não-rotulados.

O processo de agrupamento de dados não-rotulados é necessário quando deseja-se descobrir quais as características e informações que os dividem, além de ser um método ótimo para organizar e resumir os dados por meio de protótipos de grupos [16]. Dentre os algoritmos de aprendizado não-supervisionado de agrupamento, segundo [16], o k-means é o mais simples e popular, apesar de ter mais de 60 anos e apresentar divergências quanto a sua criação, visto que de forma independente, foi desenvolvido por Steinhaus (1956), Lloyd [17] (proposto em 1957, publicado em 1982), Ball e Hall (1965) e MacQueen (1967) [16].

O k-means agrupa os dados que mais se aproximam de cada centróide e tem como parâmetros de funcionamento a inicialização aleatória, durante a determinação dos centróides iniciais, a quantidade de agrupamentos, nomeada de k , e a métrica, a qual a mais utilizada é a distância Euclidiana ao quadrado, devido a existência e unicidade do ponto que a minimiza e este mesmo ponto, sendo o centróide, descrito matematicamente pela média aritmética [16] [17]. Diante disso, é possível descrever seu funcionamento pelos passos apresentados pela Figura 2.

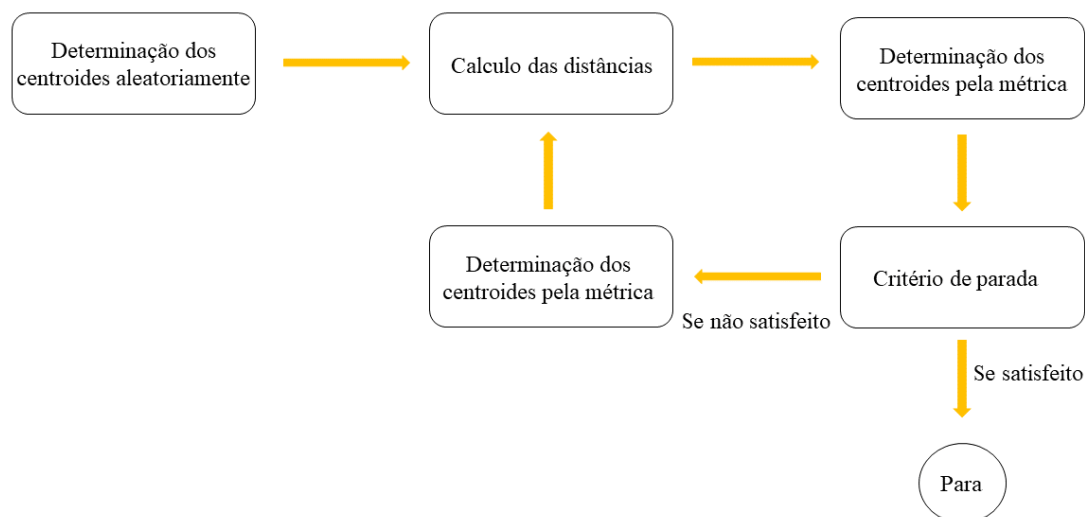


Figura 2 - Funcionamento do algoritmo k-means (autoria própria).

A determinação aleatória dos centróides é o passo inicial do algoritmo, sendo restrito à primeira iteração e se trata da determinação aleatória da localização dos k centróides, onde k é definido de acordo com a quantidade desejada de grupos e como forma de otimização o centróide é escolhido entre os dados de tal modo que não sejam

próximos, evitando que os centroides sejam localizados muito distante dos dados e que não sejam distantes entre si, já o agrupamento dos dados ocorre de acordo com as distancias aos centroides. Dessa forma, a partir da segunda iteração, os centroides passam a ser determinado pela métrica, sendo definidos pelos pontos que minimizam a soma das distancias a n pontos, onde a função distancia variando com a métrica escolhida, a determinação do centroide também varia [6]. Além disso, ainda segundo [6], no passo seguinte, são calculados as distancias de todos os dados a todos os centroides as quais são usadas no passo seguinte na seleção dos dados que apresentam as menores distancias a cada um dos centroides, de tal forma que um cluster (grupo) é composto por um centroide e todos os dados que se aproximam mais a ele. E por fim, o teste de parada do algoritmo realiza a paralização do mesmo ou permite o prosseguimento para a próxima iteração, caso não seja atingida a quantidade máxima de iterações desejada ou ainda houver significativa mudança nos clusters entre uma iteração e a seguinte. Dessa forma, o resultado do algoritmo são os k clusters, sobre os quais é possível a realização de análises quanto as características que os agrupam, além de ser possível a simplificação da coleta das informações que os dados trabalhados transmitem, já que o centroide representa os dados pertencentes ao mesmo cluster que ele, para a coleta basta seleciona-lo.

2.2. Metodologia

A análise das circunstâncias que atingem os alunos que concorrem a vagas nas instituições de ensino superior foi possível através de uma divisão resultada da implementação de suas notas do principal exame nacional de avaliação do ensino médio em um algoritmo de agrupamento, o k -means. O Instituto Nacional de Estudo e Pesquisa Educacional Aniso Teixeira (Inep) disponibiliza o conjunto de dados composto com as notas dos candidatos no Enem de cada ano, além das suas respostas ao questionário aplicado no momento da inscrição, as presenças e se houve desclassificação dos mesmos, sem disponibilizar a identificação.

Dessa forma, foram escolhidos os dados referentes à aplicação do exame nos anos 2018 e 2019. Entretanto, devido à grande quantidade de inscritos em ambos os anos, 5513662 e 5095308 respectivamente, foi necessária uma amostragem, a qual realizada com nível de confiança de 95% e erro amostra de 2% para ambas as quantidades, resultaram em 1537 como o valor que representa de forma satisfatória os dados para cada ano. Diante disso, foram selecionados 1537 alunos, de forma aleatoriamente entre os inscritos. Em seguida, foi realizado um refinamento selecionar os alunos que fossem candidatos ao ensino superior e que possuem um valor numérico para as proficiências das áreas avaliadas. Dessa forma, os alunos selecionados satisfaziam as seguintes características:

- Presentes e não desclassificados em todas as provas do exame
- Concluído o ensino médio no ano de aplicação do exame, ou em anos anteriores
- Não terem realizado o exame com o intuito de treinamento.

Portanto, com a seleção e refinamento dos proprietários das notas, estas foram implementadas no algoritmo k -means. Sendo este seguindo os paços apresentados anteriormente, foi posto em operação com o valor de k igual a 2, resultando 2 agrupamentos, e restrito a quantidade de 100 iterações máximas.

Diante do resultado do agrupamento, foi realizada uma análise sobre quais as características que apresentaram significativas diferenças entre os grupos. As tais características foram obtidas entre as respostas dos candidatos ao questionário já citado e foram selecionadas aquelas que se referem a cor/raça que os alunos se autodeclaram, variando as respostas entre, não declarado, branca, preta, parda, amarela e indígena, a renda familiar mensal, informando intervalos de valores proporcionais ao salário mínimo do ano, iniciando com 0 reais, até rendas superiores a 20 salários mínimos. Além disso, as questões que se referem à quantidade de pessoas que moram na residência do candidato, variando a respostas de 1 à 20, e em relação aos dados de 2018, a questão quanto ao tipo de escola que frequentou o ensino médio com alternativas a serem escolhidas entre “somente em escola pública”, “parte em escola pública e parte em escola privada sem bolsa de estudo integral”, “parte em escola pública e parte em escola privada com bolsa de estudo integral”, “somente em escola privada sem bolsa de estudo integral” e “somente em escola privada com bolsa de estudo integral”. No entanto, para a obtenção da renda *per capita* para ser avaliada como uma característica que diferencia os grupos, foi realizada a divisão do valor médio do intervalo informado por candidato como renda mensal familiar pela quantidade de pessoas que moram na mesma casa também informada por ele.

Como sequência das semelhanças entre as características analisadas e os critérios avaliados para o beneficiamento das cotas, onde o critério inicial é em relação a instituição de ensino médio que o candidato frequentou, não foi possível realizar a análise completa sobre as características e as relacionar as cotas com os dados referentes a 2019, pois a pergunta que fazia referência à instituição que foi frequentada não existe no conjunto de dados disponibilizados. Dessa forma, a análise principal foi realizada sobre os dados referentes a 2018 e enfatizados, nas características da renda e cor/raça, pelos dados de 2019. Portanto, ao relacionar os critérios para as cotas e características selecionadas, foi possível o refinamento entre as respostas quanto a cor/raça, sendo descartadas da análise as respostas “não declarado”, “branca” e “amarela”, além de não ter sido levada em consideração respostas diferentes de “somente em escola pública” em relação a pergunta quanto a instituição frequentada e as rendas *per capita* inferior ou igual a um salário mínimo e meio, do ano de aplicação do exame.

Todos os cálculos para a obtenção da renda *per capita* e quantificação processos de avaliação foram realizados pelo programa Python, através de um jupyter notebook.

2.3. Resultados e Discursões

Quando efetuada a implementação dos dados do Enem de 2018 no k-means, para quantidade de grupos igual a 2, foi obtido o cluster 0, com 471 integrantes e o cluster 1, com 448. Quanto a distribuição, é possível observar no gráfico a seguir:

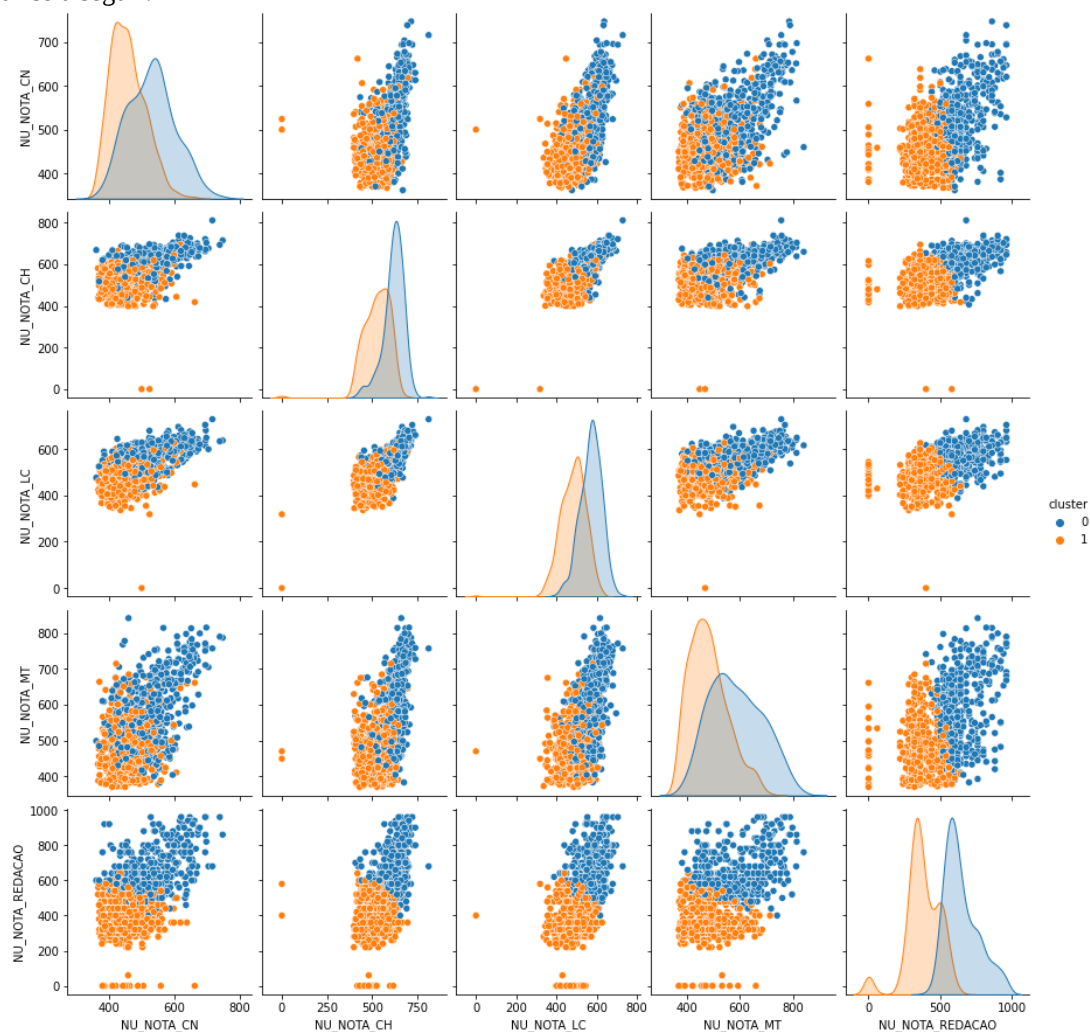


Figura 3- Agrupamento da amostra de dados do Enem de 2018 implementado no algoritmo K-means (autoria própria).

A Figura 3 correlaciona duas a duas as coordenadas dos dados, as quais correspondem às proficiências das provas de matemática, linguagens e códigos, ciências da natureza e ciências humanas, e expressa na diagonal principal a distribuição da coordenada por ela mesmo. As proficiências variam nos valores de 0 a 1000 e ao observar a distribuição dos integrantes de cada cluster, observa-se que o cluster 1, representado pela cor laranja, encontrando-se majoritariamente a esquerda e próximo aos eixos, o que permite classificar seus integrantes como aquele que possuem desempenho menor, em relação aos demais estudantes, os quais se encontram no cluster 0, localizado mais distante dos eixos e a direita dos gráficos.

Diante dos grupos formados, a análise sobre as informações que os separam além das notas, encontra-se o estudo integral do ensino médio em escolas públicas como um fator que apresenta significativa relevância. O grupo que possui os integrantes com menores notas, apresenta as proporções dos alunos que estudaram somente em escola pública durante todo o ensino médio demonstrada na Figura 4:

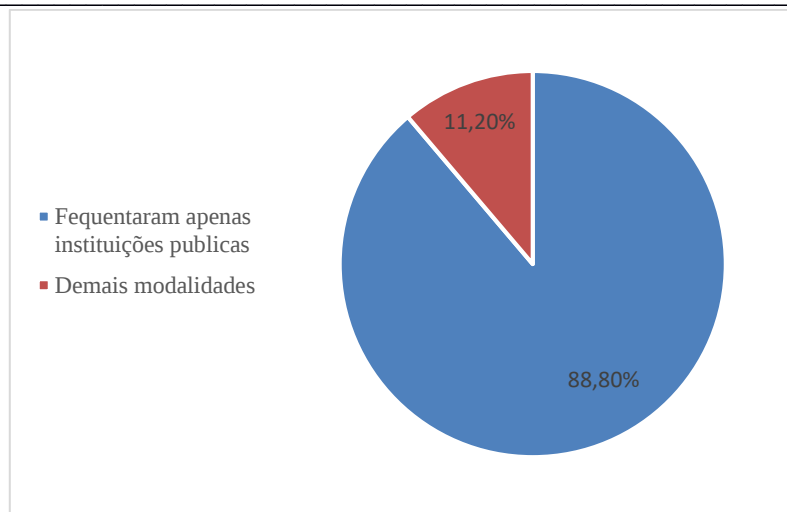


Figura 4- Proporção dos integrantes do cluster 1 quanto as instituições que frequentou todo o ensino médio (autoria própria).

Já, em relação ao outro cluster, apresenta-se a figura 5:

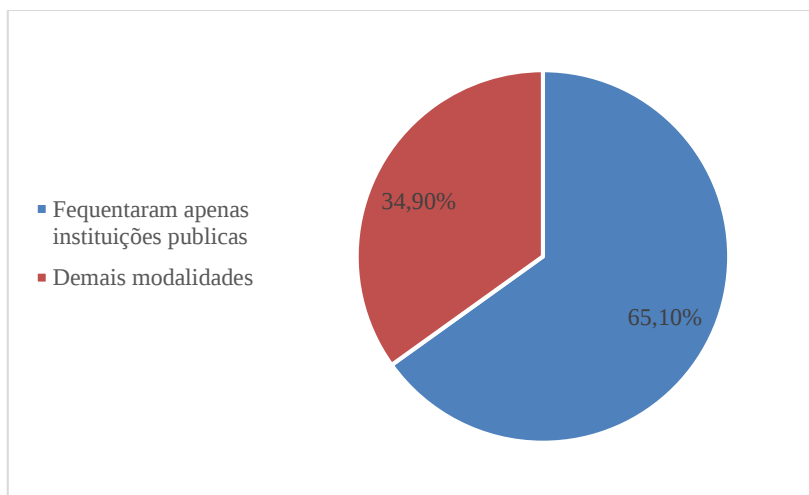


Figura 5- Proporção dos integrantes do cluster 0 quanto as instituições que frequentou todo o ensino médio (autoria própria).

Portanto, é possível notar que, os alunos que estudam em escolas públicas possuem geralmente, notas menores, o que demonstra a precariedade do ensino público e o desnivelamento quanto o ensino particular. Dessa forma, alunos de que fizeram todo o ensino médio em escolas públicas possuem poder competitivo menor quando competem por uma vaga no ensino superior com um aluno de outra modalidade escolar. Resultado semelhante foi encontrado por [19], o qual apresenta como resultado a afirmação da existência do agrupamento natural entre dos alunos por modalidade educacional frequentada.

Analisando os agrupamentos quanto a renda familiar *per capita* de até 1,5 salários-mínimos, o que equivale para o ano de 2018 a R\$ 1431,00, pode ser visto na Figura 6 e Figura 7 as representações das proporções dos integrantes de cada cluster.

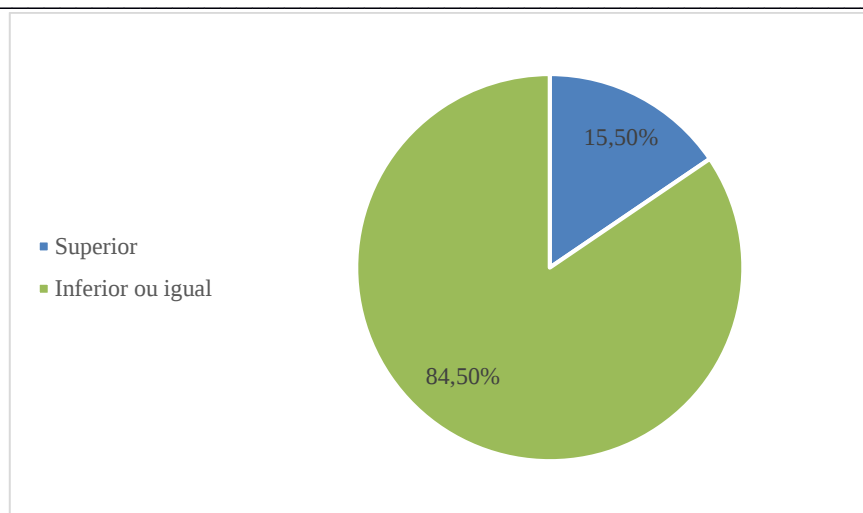


Figura 6- Proporção dos integrantes do cluster 0 quanto a renda per capita de 1,5 salários-mínimos (autoria própria).

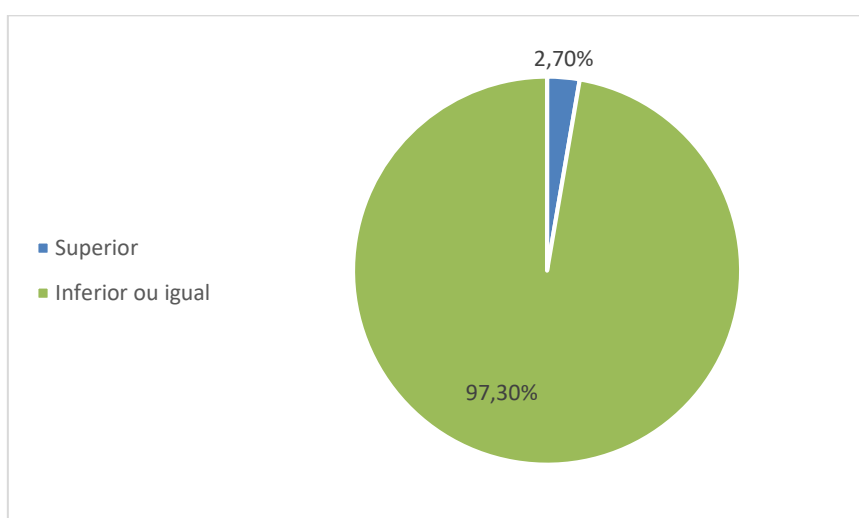


Figura 7- Proporção dos integrantes do cluster 1 quanto a renda per capita de 1,5 salários-mínimos (autoria própria).

Diante do exposto, é possível notar que o cluster constituído dos integrantes com desempenho menores apresentou a maior concentração daqueles com rendas superiores ou iguais a 1,5 salários-mínimos, apresentando relação quanto desempenho no exame e a renda familiar.

Outra informação que se apresenta como característica de agrupamento, ou seja, que é possível diferenciar os dois grupos por esta informação é quanto a raça. Os agrupamentos resultantes revelaram a proporcionalidade entre os pretos, pardos e indígenas em relação a população de cada cluster de acordo com a Figura 8 e a Figura 9.

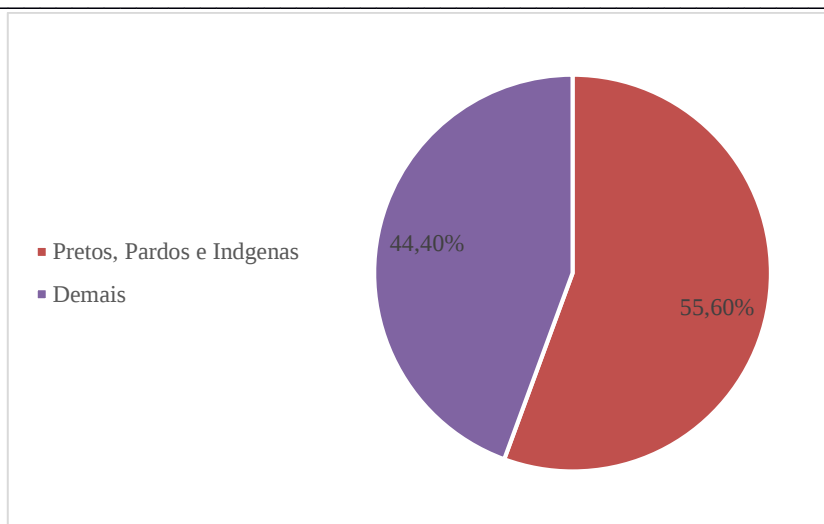


Figura 8- Proporção dos integrantes do cluster 0 quanto a cor/raça (autoria própria).

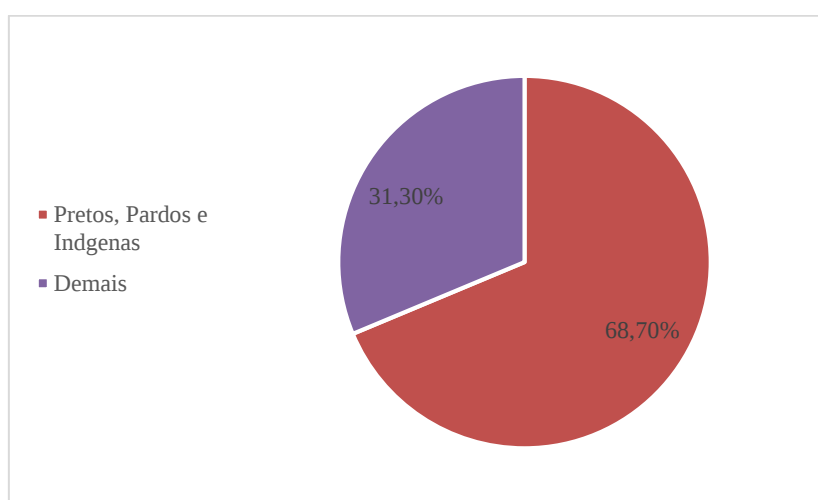


Figura 9- Proporção dos integrantes do cluster 1 quanto a cor/raça (autoria própria).

Os gráficos apresentam a relação entre a cor/raça dos candidatos e as notas no exame. A relação apresenta que os alunos que se autodeclararam pretos, pardos e indígenas possuem majoritariamente notas menores. Entretanto, é válido ressaltar que os autodeclarados pretos, pardos e indígenas constituem a população com taxas de analfabetismo mais que o dobro do que a população branca, segundo [13], além de apresentarem a mesma relação entre as taxas quanto à distribuição de renda e condições de moradia das pessoas a baixo das linhas da pobreza.

Diante da relação entre as características analisadas sobre os agrupamentos e os critérios para o beneficiamento das cotas nas instituições federais de ensino superior, quando avaliadas as proposições nos clusters sobre os alunos que satisfazem os critérios máximos, ou seja, que estudaram todo o ensino médio na rede pública, possuem renda *per capita* de até 1,5 salários-mínimos e se autodeclararam pretos, pardos ou indígenas, encontra-se os seguintes resultados:

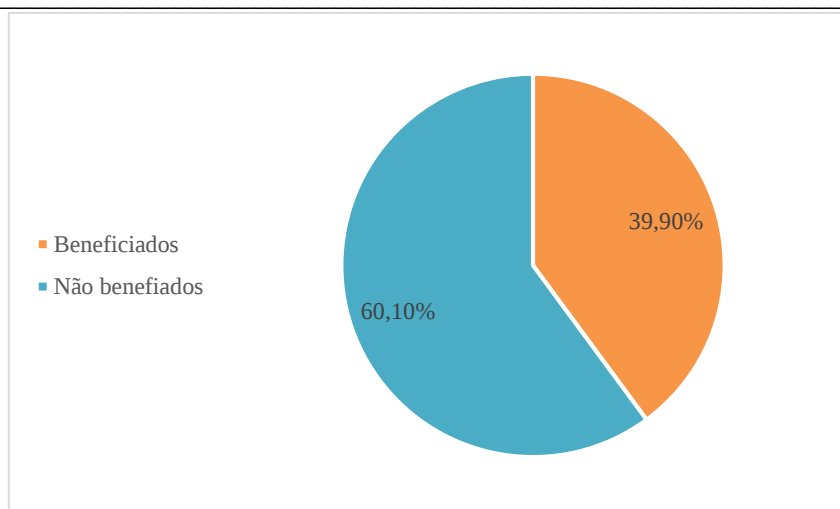


Figura 10- Proporção dos integrantes do cluster 0 quanto o beneficiamento das cotas no nível máximo (própria autoria).

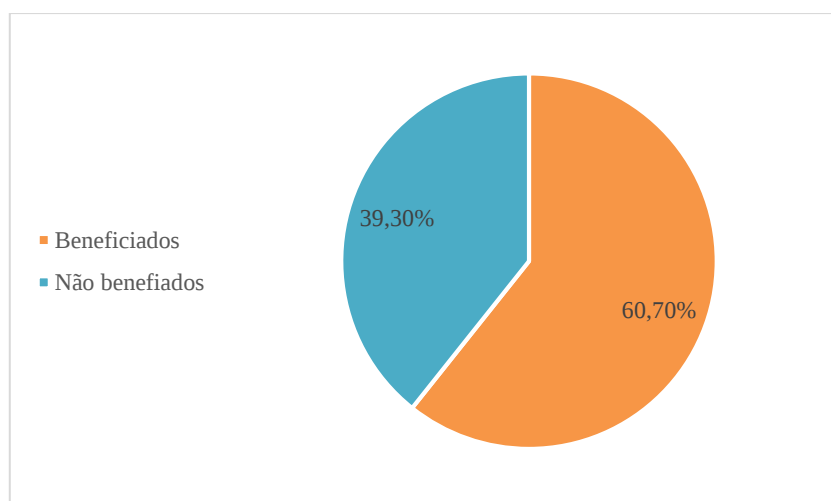


Figura 11- Proporção dos integrantes do cluster 1 quanto o beneficiamento das cotas no nível máximo (própria autoria).

Dessa forma, é possível observar a relação entre as notas dos alunos e satisfação dos critérios das cotas, de modo que o cluster 1, com estudantes de menores notas possui percentuais maiores que o outro.

A fim de enfatizar os resultados encontrados na análise realizada até este momento, foi efetuado a mesma sobre os dados correspondente ao Enem aplicado em 2019, também implementados no mesmo algoritmo, encontramos o agrupamento representado na Figura 12:

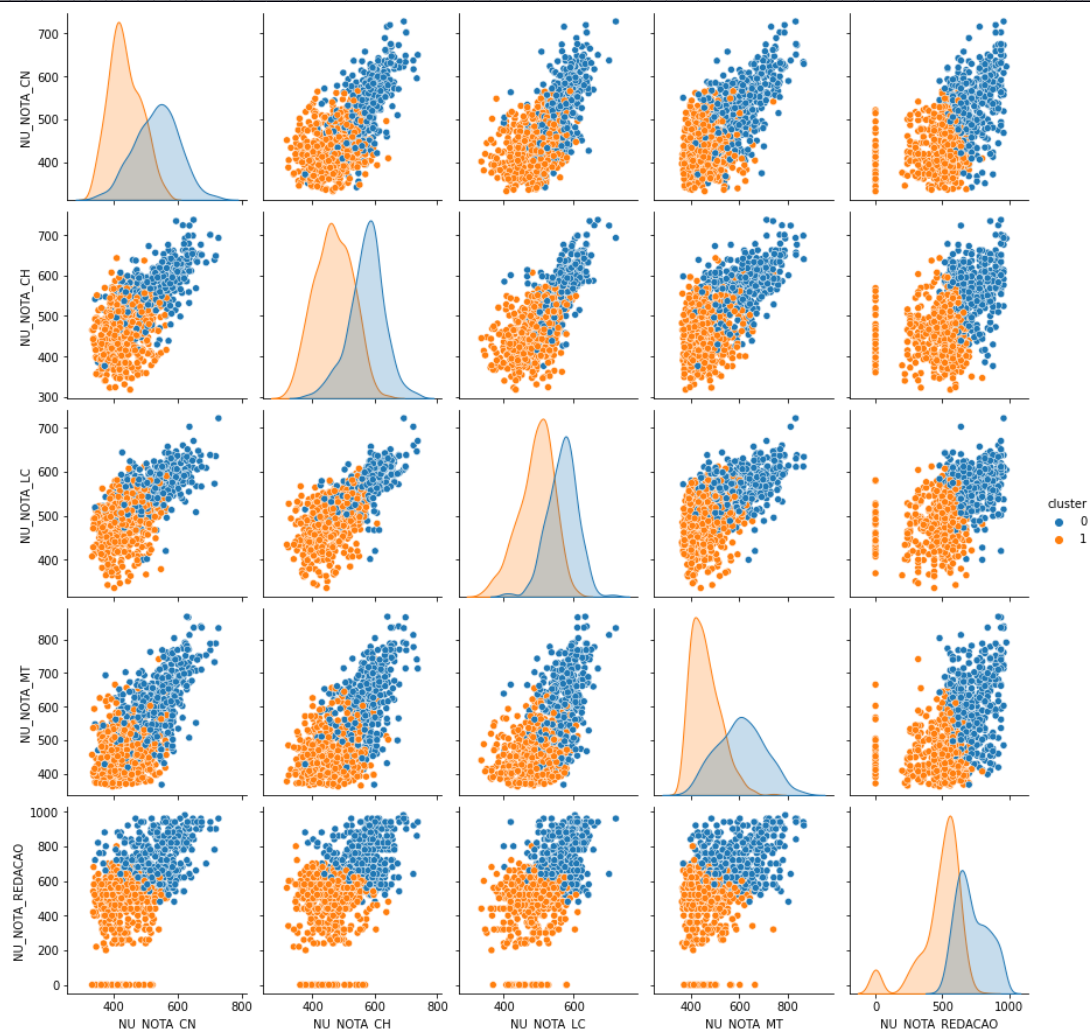


Figura 12- Agrupamento da amostra de dados do Enem de 2019 implementado no algoritmo k-means (autoria própria).

Encontra-se de forma semelhante aos dados da análise anterior, o cluster 0 constituído dos alunos com as melhores notas e o cluster 1 com os demais. Entretanto a semelhança se estende também sobre a análise das informações que dividem os grupos. O cluster 0 é constituído por 426 e o cluster 1 por 551 integrantes, apresentam respectivamente, as seguintes proporções quanto a renda e a cor/raça dos integrantes:

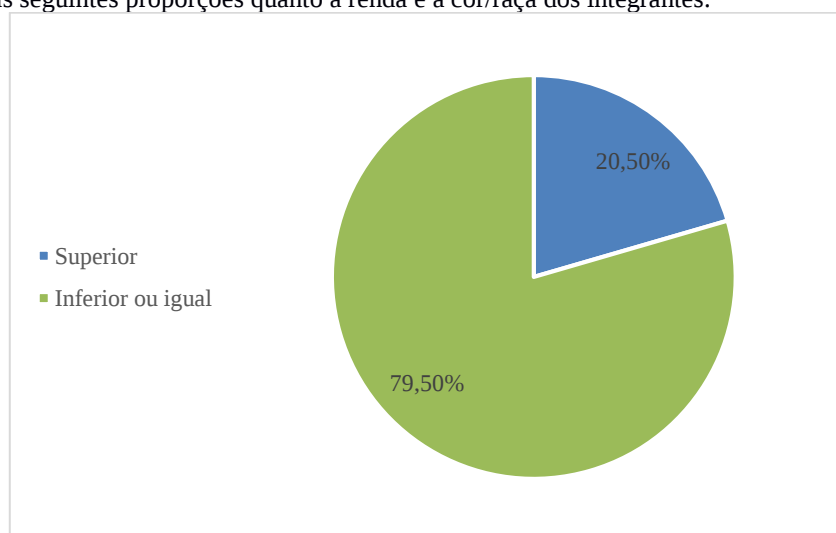


Figura 13- Proporção dos integrantes do cluster 0 quanto a renda per capita de até 1,5 salários-mínimos (autoria própria).

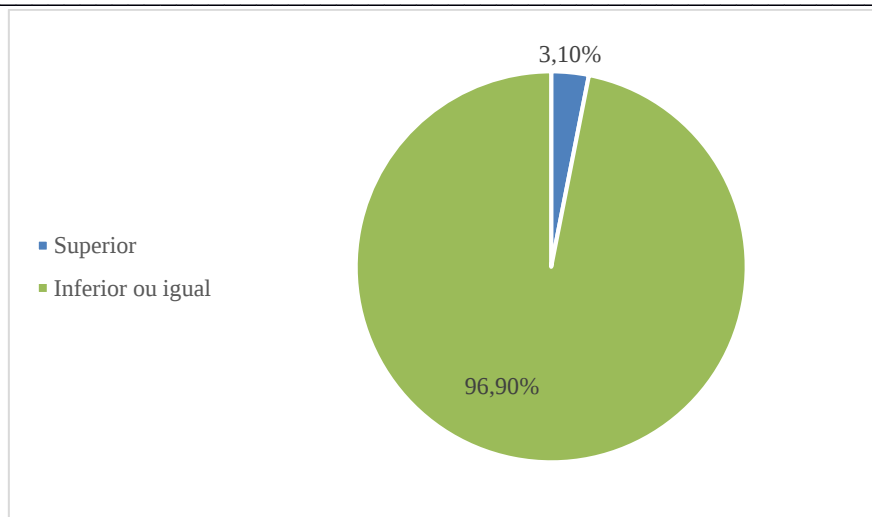


Figura 14- Proporção dos integrantes do cluster 1 quanto a renda per capita de até 1,5 salários-mínimos (autoria própria).

Analisando a Figura 13 e a Figura 14, é possível observar a semelhança entre os resultados encontrados, tornando-se um padrão os percentuais acima de 95% referentes aos alunos com renda *per capita* familiar mensal inferior ou igual a 1,5 salários-mínimos pertencentes ao cluster com integrantes de desempenho menor.

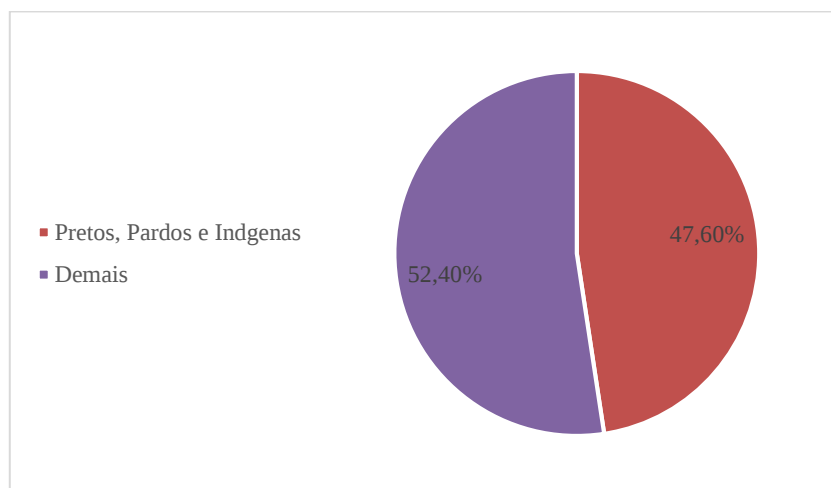


Figura 15- Proporção dos integrantes do cluster 0 quanto a cor/raça (autoria própria).

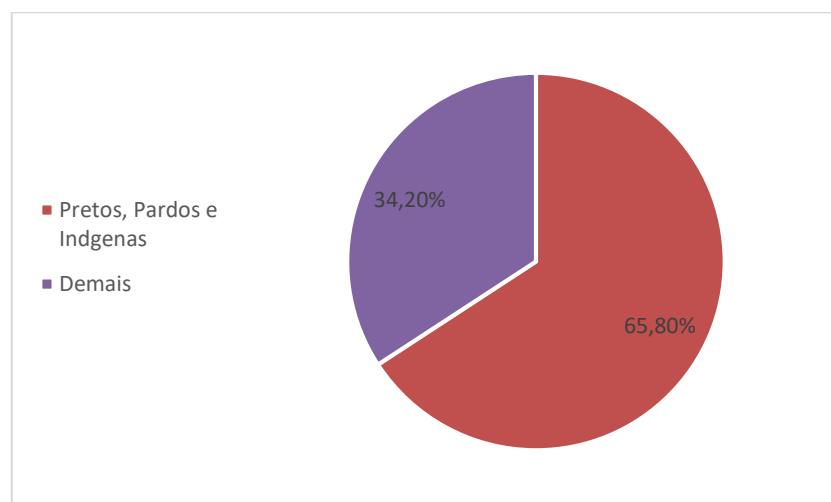


Figura 16- Proporção dos integrantes do cluster 1 quanto a cor/raça (autoria própria).

Dessa forma, é possível observar a repetição da proporção sobre os percentuais quanto a cor/raça dos integrantes de cada cluster quanto a análise realizada sobre os dados de 2018. Portanto, é possível estabelecer o padrão quanto relação da cor/raça dos candidatos e suas notas.

As relações entre as situações socioeconômicas dos candidatos e as notas podem ser relacionada entre si, quando se observa a população preta ou parda predominando a população com renda baixa [13], e esta por sua vez, é majoritária entre os estudantes das escolas públicas, além disso, a população preta, parda e indígena é apresentada no censo educacional de 2015 [12], com frequências inferiores em todos os níveis educacionais e com mais elevados índices de analfabetismo, também apresentado por [13]. Entretanto, ausência da pergunta quanto à modalidade de escola que o aluno frequentou durante todo o ensino médio no questionário aplicado no momento da inscrição no Enem em 2019, não é possível a verificação do padrão como realizado na análise em relação ao ano anterior, desenvolvida nesse mesmo trabalho.

Contudo quando se analisa os grupos, resultados da implementação dos dados de 2018, no algoritmo, em relação aos critérios para o beneficiamento das cotas, encontram-se 39,9% dos integrantes do cluster 0 que concorrem às vagas destinadas apenas para os candidatos que estudaram o ensino médio integralmente em escolas públicas, que possuem renda *per capita* inferior a 1,5 salários mínimo e se autodeclararam pretos, pardos ou indígenas. Entretanto, satisfazendo os mesmos critérios para concorrer às vagas citadas, o cluster 1 possui 60,7%. Portanto, o cluster com os alunos de pior desempenho possui mais da metade dos seus componentes classificados no nível mais abrangente das cotas, demonstrando o quanto este grupo apresenta-se potencialmente menos favorável a ingressar no ensino superior devido ao conjunto de fatores já apresentados.

3. CONCLUSÕES

Apresenta-se nesse artigo uma análise sobre as características que dividem os candidatos a ingressarem em instituições de ensino superior através dos seus desempenhos no Enem, divididos em 2 grupos pelo algoritmo k-means, baseando-se apenas pelas notas nas 5 competências avaliadas pelo exame. O método aplicado apresenta os dois grupos que se diferem por serem constituídos por integrantes com desempenho melhor do que do outro. Além disso, diante dessa diferença, a análise sobre as características informadas pelos candidatos apresenta o grupo com menor desempenho com percentuais significativamente maiores quanto aos candidatos que estudaram todo o ensino médio em instituições públicas, além dos candidatos que se declararam pretos, pardos e indígenas e aqueles que possuem renda familiar de até 1,5 salários-mínimos, expondo a relação entre as características analisadas e notas dos candidatos.

Portanto, visto o fato das cotas para as vagas nas instituições de ensino superior serem restritas por níveis seguindo as 3 características já citadas, o grupo de menor rendimento apresenta mais de 60% dos integrantes classificados simultaneamente pelas tais características, expõem a importância e a necessidade das cotas como medida provisória para equiparar o potencial competitivo entre os privilegiados e os afetados pelas desigualdades. Dessa forma, recomenda-se como trabalhos futuros a análise crítica quanto a formas de diminuir o impacto das desigualdades da realidade brasileira no desempenho dos estudantes do ensino básico.

3. REFERÊNCIAS BIBLIOGRAFICAS

- [1] MINISTERIO DA EDUCAÇÃO (MEC). Enem. Disponível em: <<http://portal.mec.gov.br/ultimas-noticias/418-enem-946573306>>. Acesso em: 02 nov 2020.
- [2] SILVEIRA, F. L.; BARBOSA, M. C. B.; SILVA, R. Exame nacional do ensino Médio (Enem): Uma análise crítica. Rev. Bras. Ensino Fis., Rio Grande do Sul, v. 37, n. 1, p. 1–5, 2015. Disponível em: <https://www.scielo.br/scielo.php?pid=S1806-11172015000101101&script=sci_arttext&tlng=pt>. Acesso em: 19 out 2020
- [3] MINISTERIO DA EDUCAÇÃO (MEC). SISU. Disponível em: <<http://portal.mec.gov.br/SISU>>. Acesso em: 02 nov 2020.
- [4] ANHAIA, B. C. A Lei de Cotas e a universidade pública brasileira em debate. Revista Contraponto, Rio Grande do Sul. v.4, n.2, 2017. Disponível em: <<https://www.seer.ufrgs.br/contraponto/article/view/78915>>. Acesso em: 19 out 2020.
- [5] BRASIL, PL 3627/2004. Disponível em: <<https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=254614>>. Acesso em: 10 nov 2020.
- [6] MATTE, M. K. Impacto do uso da desigualdade triangular para acelerar o algoritmo k-means, Centro Universitario Campo Limpo Paulista, São Paulo. Unifaccamp, 2020. Disponível em: <www.cc.faccamp.br/Dissertacoes/MarceloKucharMatte.pdf>. Acesso em: 20 out 2020
- [7] KLITZKE, M. K. Exame nacional do ensino médio (Enem): uma tentativa de democratização do acesso ao ensino superior. Res. Policy, v. 9, n.2, p. 155–162, 2014. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/196325>>. Acesso em: 04 nov 2020
- [8] GARCIA, V. C. M. et al. Enem: um demonstrativo das mudanças socioeconômicas no perfil dos participantes, Rev. Meta Avaliação. v. 1, n. 1, p. 104, 2009.
- [9] MINISTERIO DA EDUCAÇÃO (MEC). ProUni. Disponível em: < Prouni - Home (mec.gov.br)>. Acesso em: 02 nov 2020
- [10] MINISTERIO DA EDUCAÇÃO (MEC). Novo Enem. Disponível em: < Novo Enem - Ministério da

-
- Educação (mec.gov.br)>. Acesso em: 15 nov 2020
- [11] MINISTÉRIO DA EDUCAÇÃO (MEC). MEC vai esclarecer sobre TRI à Justiça Federal do Ceará. Disponível em: < <http://portal.mec.gov.br/ultimas-noticias/222-537011943/16018-mec-vai-esclarecer-sobre-tri-a-justica-federal-do-ceara> >. Acesso em 15 nov 2020
- [12] INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. O item cor ou raça no Censo Escolar da educação básica. p. 10, 2015. Disponível em: <http://download.inep.gov.br/educacao_basica/educacenso/documentos/2015/cor_raca.pdf >. Acesso em: 4 nov 2020.
- [13] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Desigualdades sociais por cor ou raça no Brasil, Estud. e Pesqui. Informações Demográficas e Socioeconômicas. v. 41, p. 1–12, 2019. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/25844-desigualdades-sociais-por-cor-ou-raca.html%0AEstudos>>. Acesso em: 14 nov 2020.
- [14] ANHAIA, B. C. Educação superior e inclusão social: um estudo comparado de duas ações afirmativas no Brasil: dos debates à prática. Repositório digital - UFRGS, Rio Grande do Sul. 2013. Disponível em: <Educação superior e inclusão social : um estudo comparado de duas ações afirmativas no Brasil : dos debates à prática (ufrgs.br)> . Acesso em: 19 out 2020
- [15] VILELA, L.; MENEZES-FILHO, N.; TACHIBANA, T. Y. As Cotas nas Universidades Públicas Diminuem a Qualidade dos Alunos Selecionados ? Simulações com dados do Enem. Instituto de Ensino e Pesquisa (Insper), São Paulo v. 17, 2016. Disponível em: < Insper: Ensino Superior em Negócios, Direito e Engenharia > . Acesso em: 19 out 2020
- [16] JAIN, A. K. Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. Michigan, USA, v. 31, n. 8, p. 651–666, 2010. Disponível em: <Data clustering: 50 years beyond K-means - ScienceDirect>. Acesso em: 27 jul 2020
- [17] LLOYD, S. P. Least squares quantization in {PCM}.: Special issue on quantization, IEEE Trans. Inf. Theory, vol. 28, no 2, p. 129–137, 1982. Disponível em: <Least squares quantization in PCM - IEEE Journals & Magazine>. Acesso em: 27 jul 2020
- [18] PINELE, J. Geometria do modelo estatístico das distribuições normais multivariadas, Repositório - UNICAMP, 2017
- [19] LEONI, R. C.; SAMPAIO, N. A. de S. Desempenho das escolas públicas e privadas da região do Vale do Paraíba: uma aplicação da técnica de agrupamentos kmeans com base nas variáveis do Enem 2015, Cad. do IME - Série Estatística, v. 42, n. 0, 2017. Disponível em: <DESEMPENHO DAS ESCOLAS PÚBLICAS E PRIVADAS DA REGIÃO DO VALE DO PARAÍBA: UMA APLICAÇÃO DA TÉCNICA DE AGRUPAMENTOS KMEANS COM BASE NAS VARIÁVEIS DO ENEM 2015 | Leoni | Cadernos do IME - Série Estatística (uerj.br)> . Acesso em: 19 out 2020
-