

Projeto Final: Análise Exploratória de Dados
Alexia Naara da Silva Cardoso
Iago de Sousa Campelo Matos
Curso de Inverno: Ciência de Dados
Grupo Katie - UFAL

Análise de Médias de Estudantes Norte Americanos

Data source: [Students Data Analysis](#)



1 - Introdução

O relatório que será abordado a seguir tem a finalidade em obter resultados quanto às médias obtidas de alunos, através de uma análise de um arquivo disponibilizado pelo Kaggle. Este trabalho foi desenvolvido devido ao curso de inverno Ciência de Dados do grupo Katie da Universidade Federal de Alagoas.

2 - DataSet

O dataset escolhido para esse projeto refere-se à notas particulares de alunos e suas respectivas médias, relacionadas à aprovação ou não-aprovação dos mesmos. O motivo da escolha foi o interesse pelo conjunto de dados e por ser menor, fazendo com que tivéssemos mais facilidade ao manuseá-lo, ainda como iniciantes.

2.1 - Volumetria

Instâncias: 105 linhas

Atributos: 17 colunas

Tamanho: Compactado (2 kB). Descompactado(5,71 kB)

Formato: csv (valores separados por vírgula)

2.2 - Variáveis de interesse

Apesar da larga quantidade de atributos, apenas utilizamos os IDs para identificação dos alunos, suas notas para cálculo de médias e os seus tipos de aprovação.

2.2.1 - ID do Aluno

Encontra-se na coluna **ID**. Esta coluna apresenta um número sequencial que identifica unicamente um aluno.

Trata-se de um valor inteiro.

2.2.2 - Notas dos Alunos

Encontram-se nas colunas subsequentes à coluna **ID** após o tratamento. Apresentam notas de disciplinas diferentes, sendo uma coluna para cada uma.

Tratam-se de números inteiros.

2.2.3 - Tipo de aprovação

Encontram-se na coluna **y**. Esta coluna apresenta três tipos de aprovação: 0 em caso de 'failed', 1 em caso de 'inland application' e 2 em caso de 'onboard application', onde apenas 0 apresenta reprovação.

Trata-se de um inteiro [0, 1, 2]

3 - Objetivo

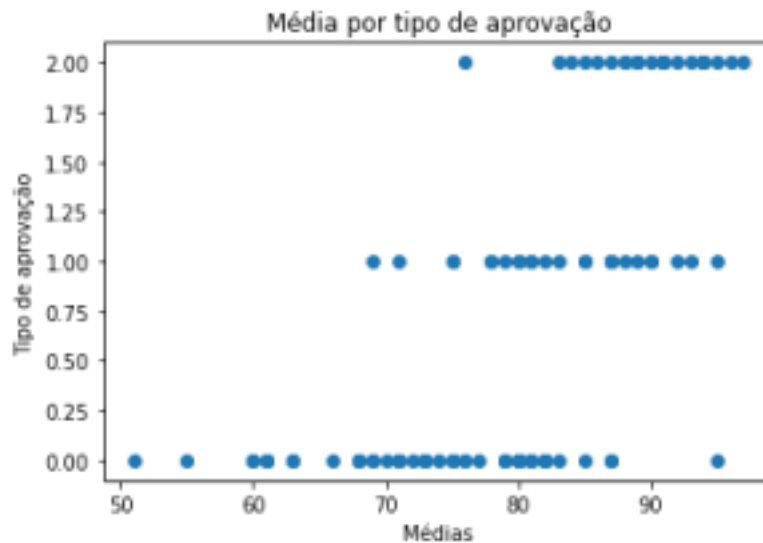
O projeto consiste em analisar quantos alunos foram aprovados ou reprovados, ainda que suas médias fossem grandes ou pequenas.

4 - Entendimento dos Dados

Com o dataset já pronto, ao analisar os dados decidimos que os mais importantes seriam os IDs, as notas e os tipos de aprovação, de acordo com o nosso objetivo.

5 - Modelagem

Analizando os dados obtidos, escolhemos o gráfico de dispersão para visualização dos resultados, obtendo assim a seguinte imagem:



6 - Análise Exploratória dos Dados

Percebemos que, ainda que certos alunos tirassem notas acima da média, não foi o suficiente para que ocorresse a aprovação dos mesmos, o que nos levou à conclusão de que as notas não são o único critério de avaliação, o que pode puxar outra avaliação de acordo com os outros dados.

Ainda é possível perceber que os que foram aprovados por 'onboard application' tiveram médias acima de 80, enquanto os aprovados por 'inline application' tiveram médias muito mais variadas, mas sempre acima de 65. Dos reprovados, todos tiveram notas bastante variadas.