

Gustavo Lopes dos Santos

CPD: 53158

Prof. Me. Aldo Henrique Dias Mendes

Mineração de Dados

07 de maio de 2023

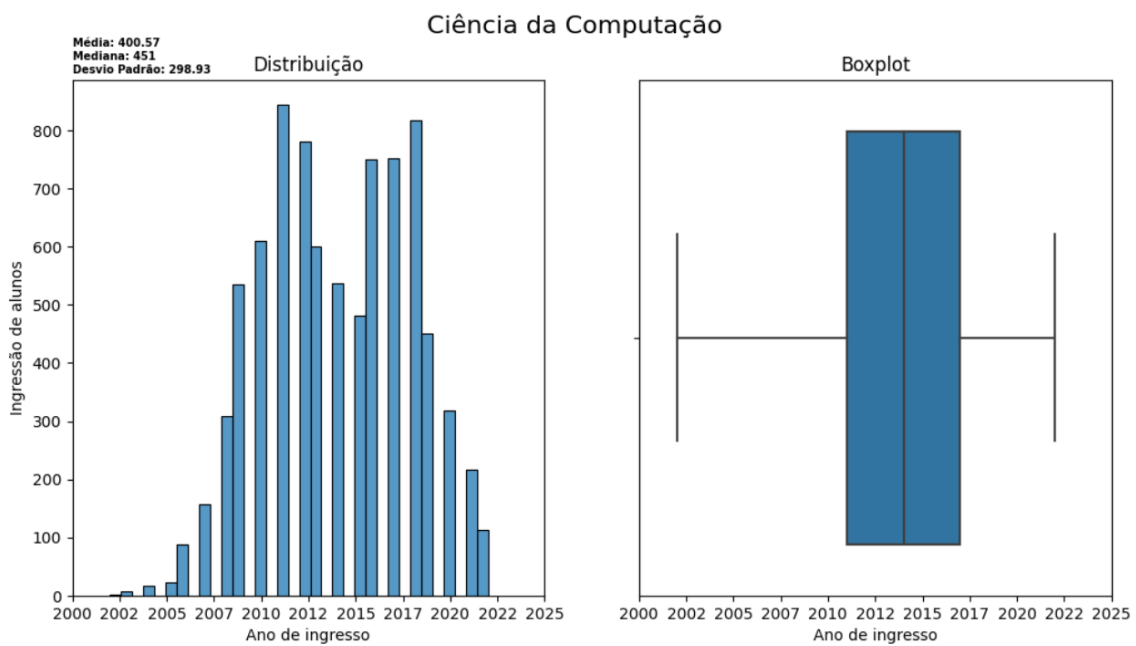
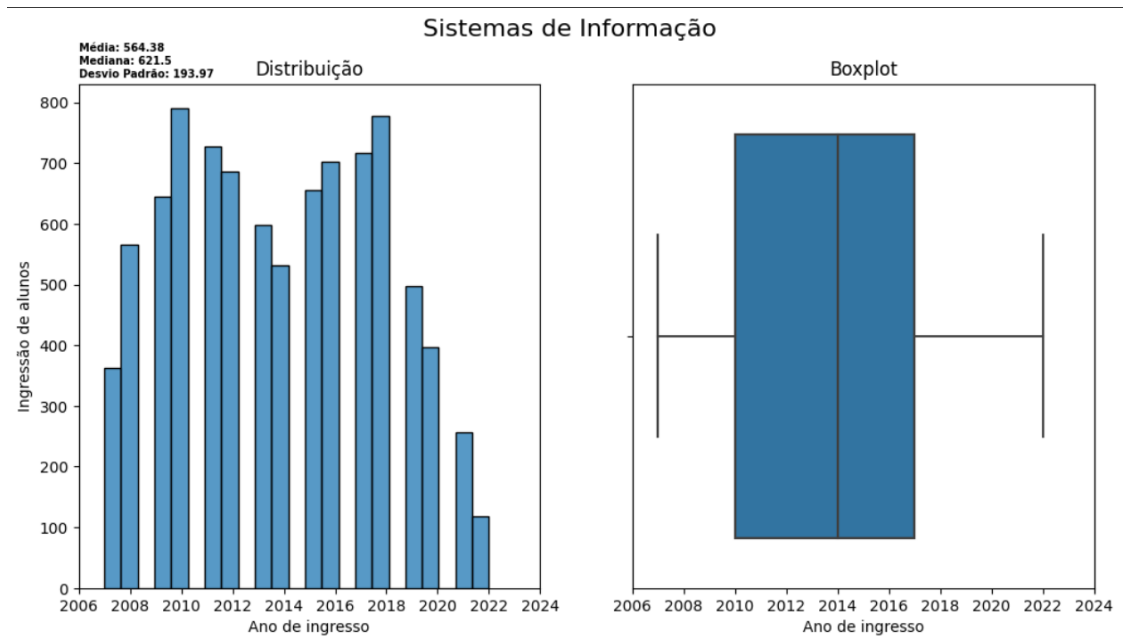
## Projeto 1

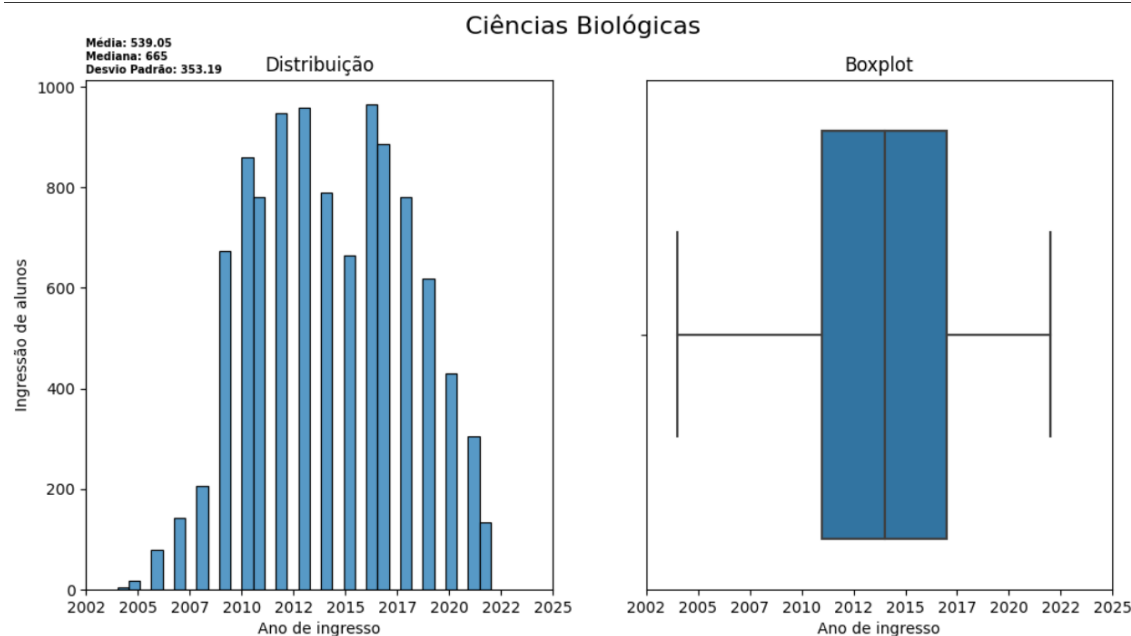
Nesse projeto, foram realizadas a Mineração de Dados de duas API's de dados abertos, uma no formato CSV dos Estudantes de Graduação da Universidade Federal de Lavras (Ufla), e outra no formato JSON das Moedas Comemorativas do Banco Central do Brasil (CBC), processando, analisando e explorando os conjuntos de dados, foi possível encontrar alguns padrões interessantes caso não fossem analisados, não seria possível obter.

### Parte 1

A base de dados obtida dos Estudantes de Graduação da Universidade Federal de Lavras era uma base de dados muito grande, então a primeira coisa feita foi a extração das colunas que não iriam ser necessárias para a análise. Após salvar um novo arquivo no qual usaremos para trabalhar sem as colunas indesejadas, fazemos o tratamento de alguns dados, a data de ingresso na universidade na base de dados se dá por “ano/semestre”, entretanto só é necessário o ano para a análise, então fazemos a formatação em uma nova variável do ano de ingresso da coluna ['semestre\_letivo\_ingresso']. Cada curso, possui anos distintos que tiveram alunos matriculados. Então para cada curso, foi gerado um gráfico de distribuição com a ingressão de alunos por ano. Também foi gerado um gráfico Boxplot pelo ano de ingresso. Além disso, foi calculado a média geral dos anos que tiveram alunos matriculados, mediana e desvio padrão.

Alguns dos gráficos gerados:



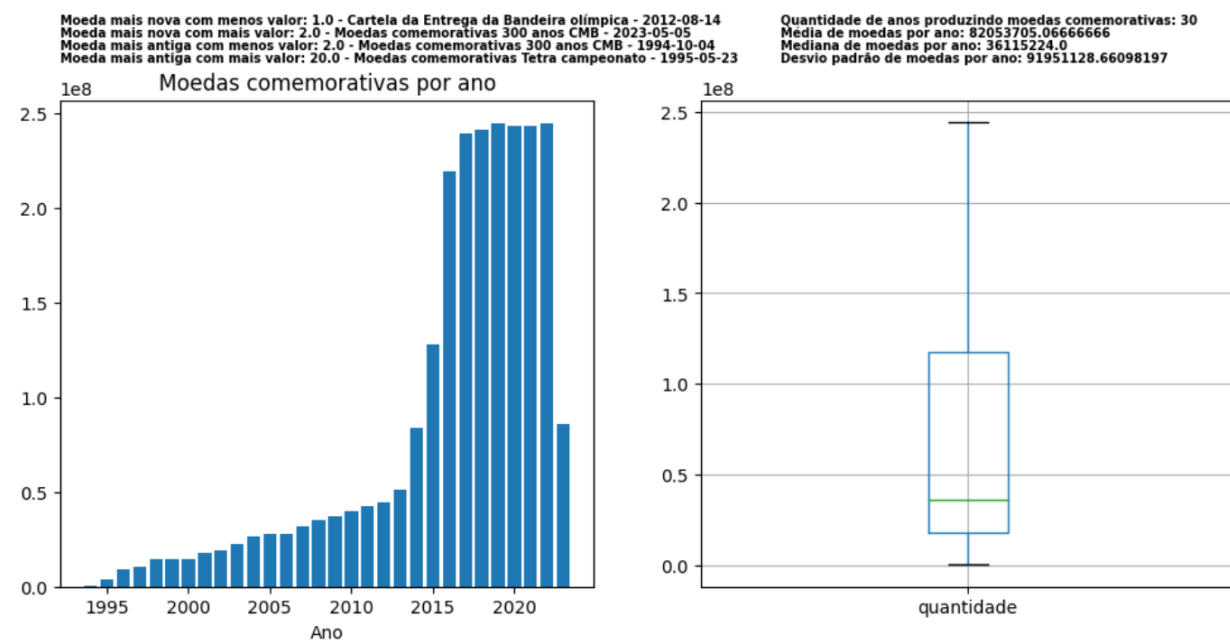


Como pode-se perceber fazendo uma análise no gráfico Boxplot dos últimos gráficos, 50% dos alunos se matricularam entre 2010 e 2017. Outra coisa que se pode observar, é que a partir do ano de 2020, a quantidade de alunos matriculados apenas decaiu, isso pode-se atribuir a vários fatores, conforme a Folha de São Paulo (apud Censo da Educação Superior, 2022), “Além da queda dos concluintes, as universidades públicas tiveram, redução de ingressantes pelo terceiro ano consecutivo. Em 2020 elas receberam 527.006 novos alunos, uma queda de 5,8% em relação a 2019, quando foram 559.293. Desde 2017, a queda acumulada é de 10,7% no número de ingressantes”.

## Parte 2

A base obtida das moedas comemorativas, também era uma base de dados muito grande, então dos dados retornados da API, foi feito um filtro para remover um deles. Foi criado um

dicionário que possui alguns tratamentos de dados, A data da criação da moeda (ano, mês, dia), ano, quantidade de moedas criadas, categoria da moeda e valor da moeda. Dessa forma, foi possível pesquisar os tipos de moedas mais novas e mais antigas com menores e maiores valores, também foi possível calcular a média, mediana e desvio padrão de moedas comemorativas já produzidas. Com os filtros aplicas, é possível fazer outros tipos de pesquisas.



Alguns fatos interessantes, é que além das moedas tradicionais, também foi possível observar que o Banco Central também produz moedas comemorativas com valor acima de R\$1,00. O gráfico apresenta as moedas em 1e8, ou seja, 100.000.000, para interpretar a quantidade de um ano, é só multiplicar 2,0 por 100.000.000, por exemplo. Algo muito interessante nesse gráfico, é que houve uma disparada na produção de moedas no ano de 2014 em relação ao de 2013 por conta da copa do mundo, e em 2016 há um crescimento de quase duas vezes em relação à 2015 por conta dos jogos olímpicos do rio. Enquanto este documento está sendo escrito, já

foram produzidas mais moedas comemorativas que no ano de 2014 e ainda faltam 7 meses para o ano acabar.

## CONCLUSÃO

A mineração de dados é uma ferramenta valiosa para identificar padrões e tendências em grandes conjuntos de dados. É importante ter uma estratégia clara e definir objetivos específicos antes de iniciar o projeto. A coleta e o pré-processamento de dados são fatores críticos para o sucesso do projeto, pois garantem que os dados sejam limpos e organizados para análise.

Por fim, é importante lembrar que a mineração de dados é uma ferramenta poderosa, mas que deve ser usada com responsabilidade e ética. A privacidade dos dados deve ser protegida e os dados devem ser usados apenas para fins legítimos. Todos os dados nesse projeto foram utilizados pelo portal de dados aberto do gov.br, os links e referências estão citados nos trabalhos citados, caso queira, o código-fonte do trabalho juntamente das bases estão em um repositório aberto do GitHub.

### Trabalhos citados

Código-fonte: <https://github.com/HannigUsT/Projeto-1-DataMining>

Folha de São Paulo. Universidades Públicas tiveram queda de 18,8% no número de concluintes.

2022. Disponível em:

<https://www1.folha.uol.com.br/educacao/2022/02/universidades-publicas-tiveram-queda-de-188-no-numero-de-concluintes.shtml>. Acesso em: 07 de maio de 2023.

Gov.br. <https://dados.gov.br/dados/conjuntos-dados>