

# Trabalho final Mineração de Dados

Bruno Perussatto e  
Gabriel Vinicius Schmitt Caetano





## **Etapas seguidas:**

- Seleção dos dados
- Pré-processamento
- Transformação
- Análises gráficas e definição de hipóteses
- Mineração



# Seleção dos dados e Pré-processamento

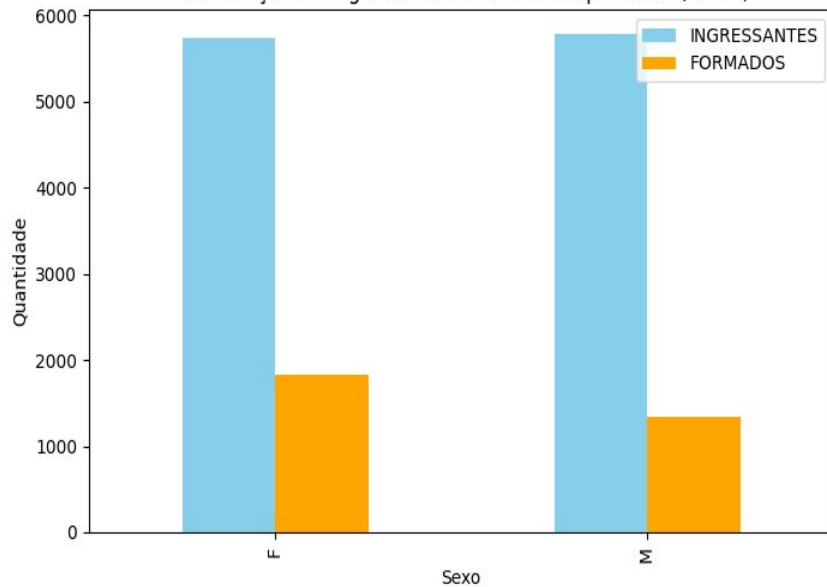
- Consideramos 2 conjuntos disponibilizados pelo professor e exploramos cada um deles individualmente no decorrer do trabalho
- Um conjunto de dados referentes a aprovações e reprovações em disciplinas do CT
- Outro de dados referentes a ingressantes e formados nos centros da UFSM
- Dados foram todos convertidos para CSV
- Um arquivo duplicado CE.csv foi removido
- Os conjuntos foram unidos em arquivos únicos para cada tipo
- No conjunto de ingressantes e formados foi adicionada uma coluna “CENTRO”



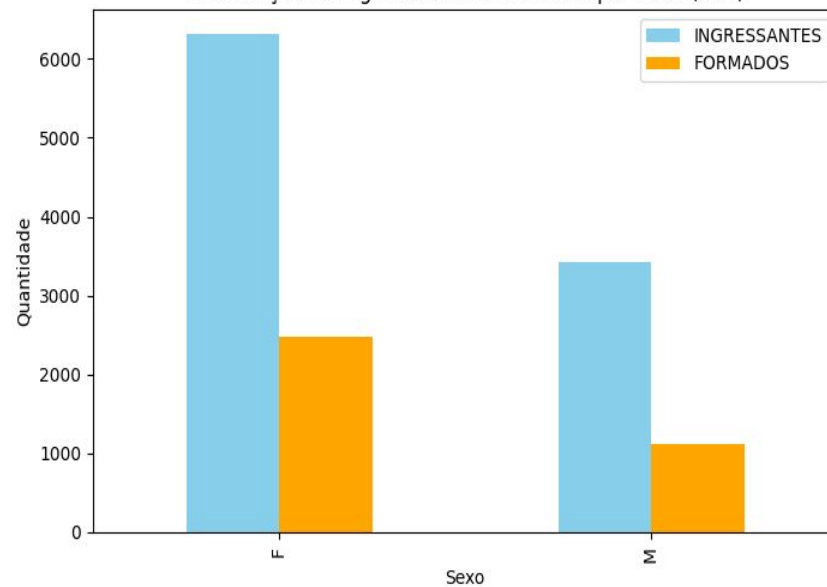
# Transformação dos dados

- Coluna Situação na tabela das disciplinas pivotada
- Coluna SEXO na tabela de ingressantes e formados pivotada
- Remoção dos registros sem 100%
- Nova coluna Ano (ano+semestre)
- Nova coluna Situação Outros

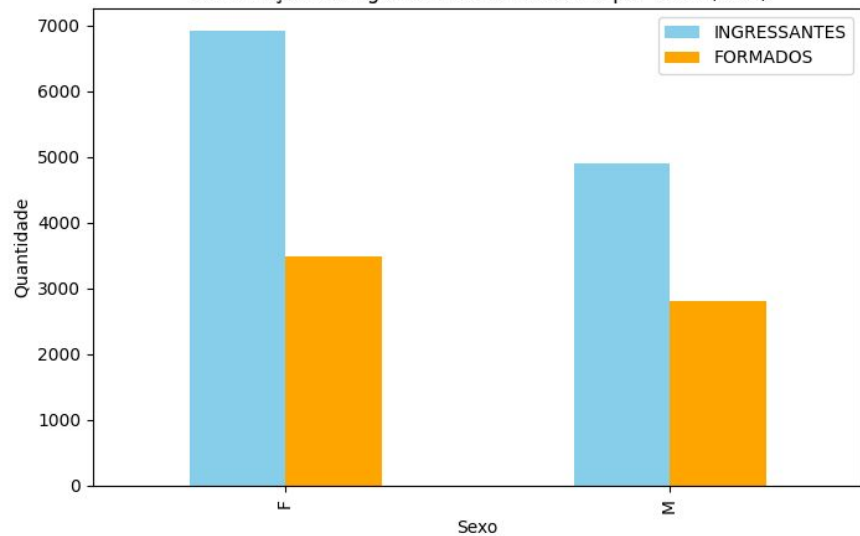
Distribuição de Ingressantes e Formados por Sexo (CCNE)



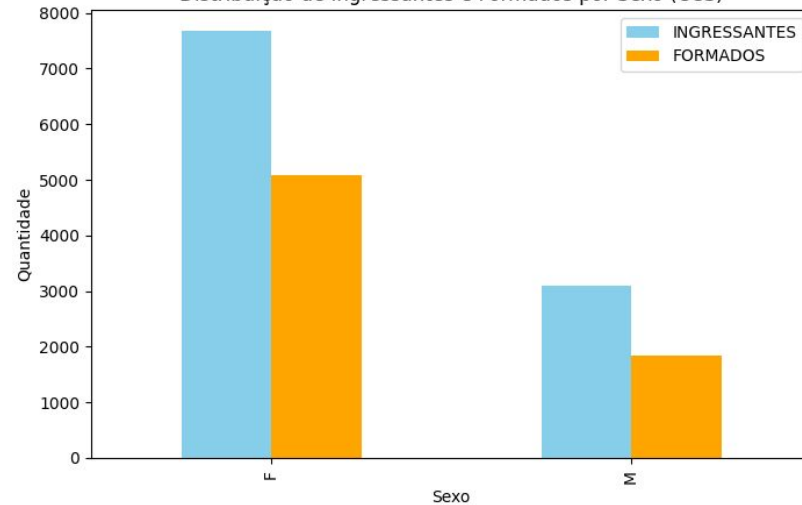
Distribuição de Ingressantes e Formados por Sexo (CAL)

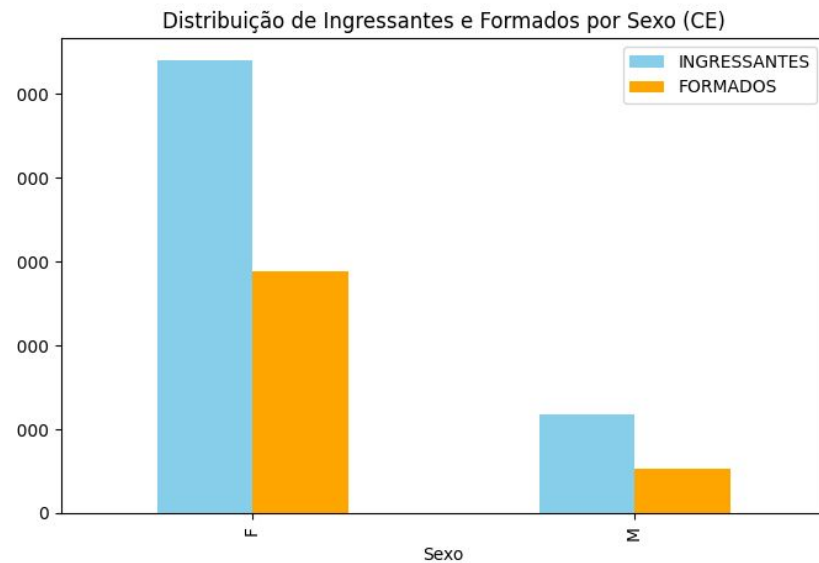
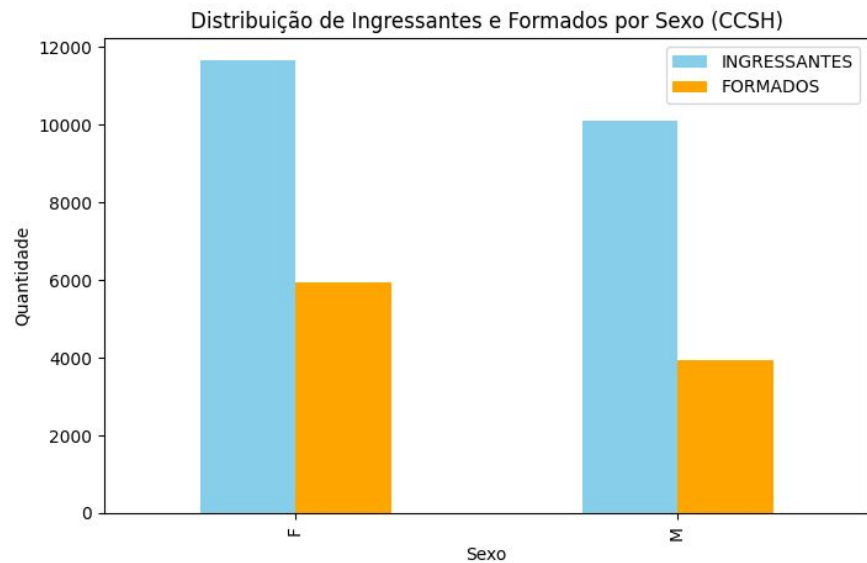


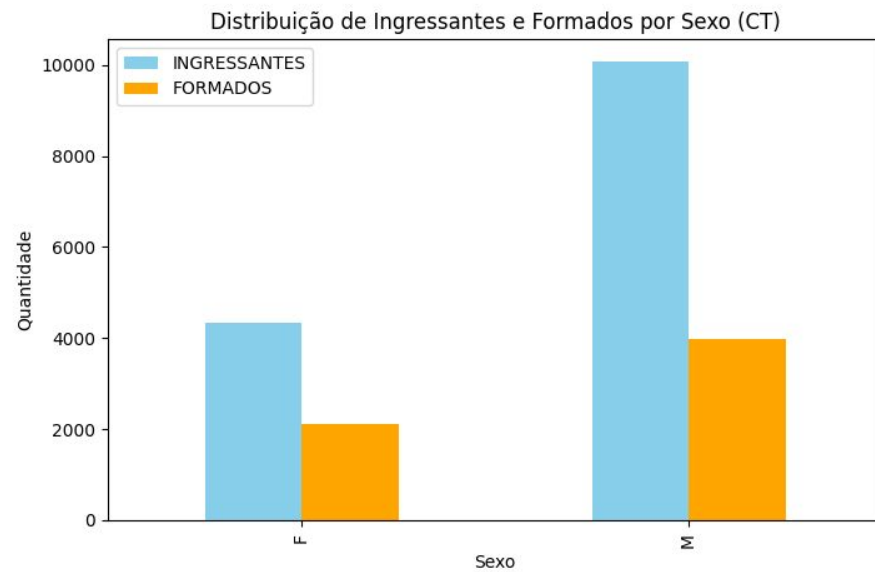
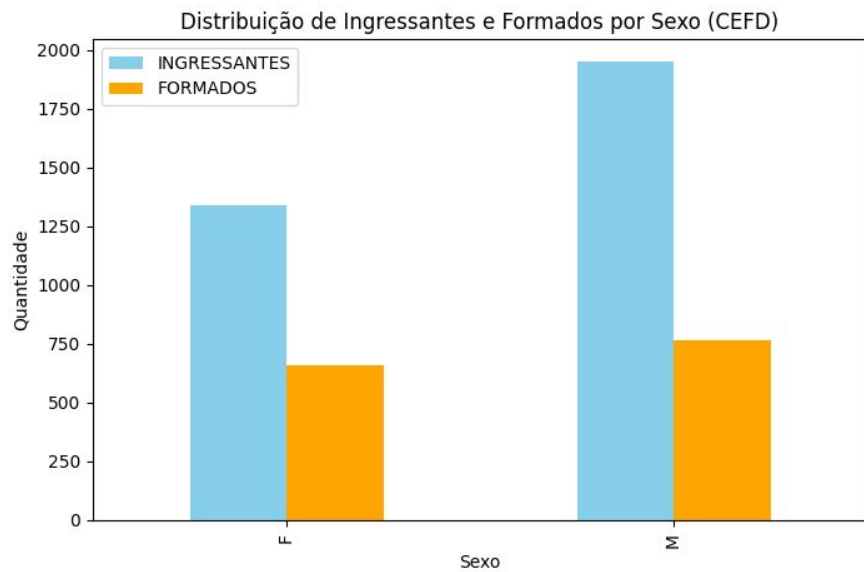
Distribuição de Ingressantes e Formados por Sexo (CCR)



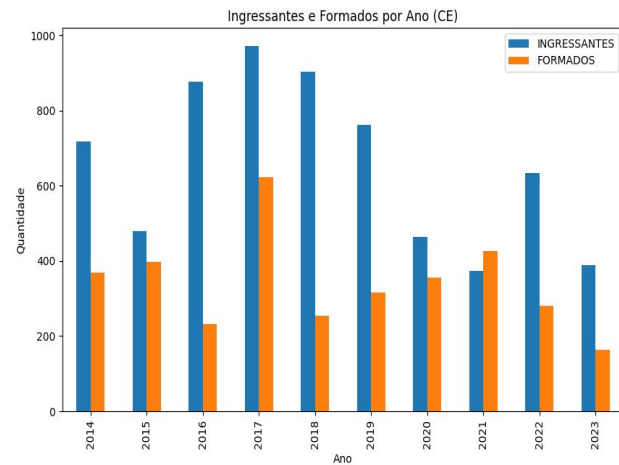
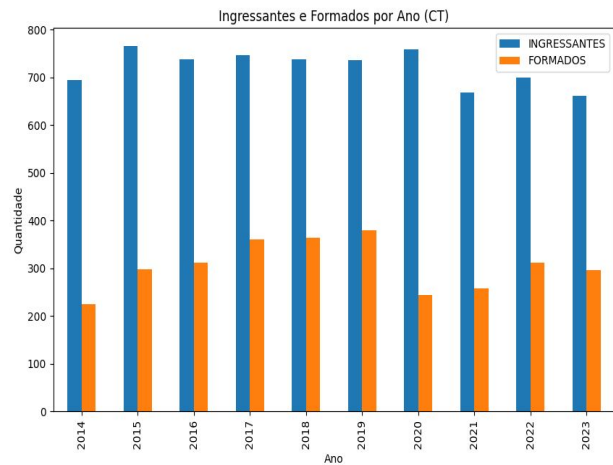
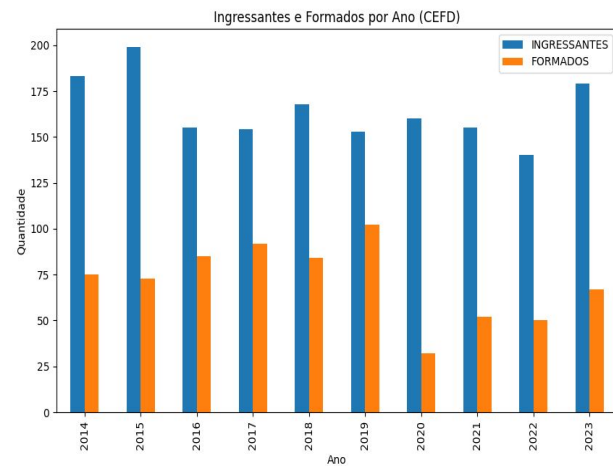
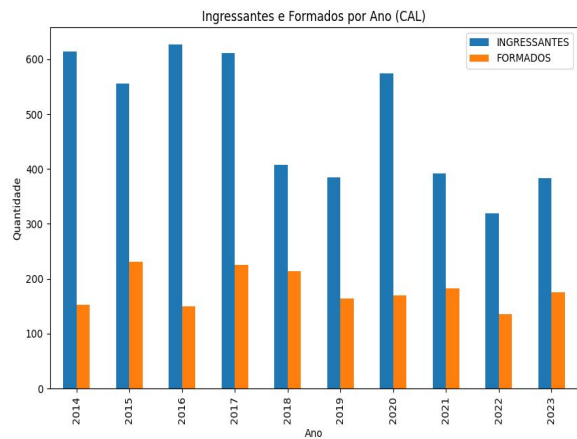
Distribuição de Ingressantes e Formados por Sexo (CCS)

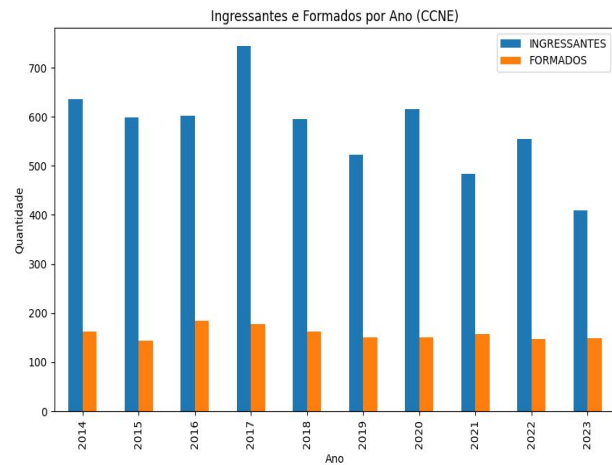
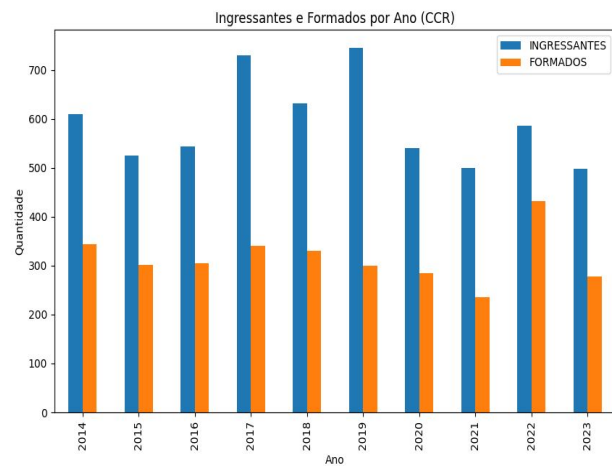
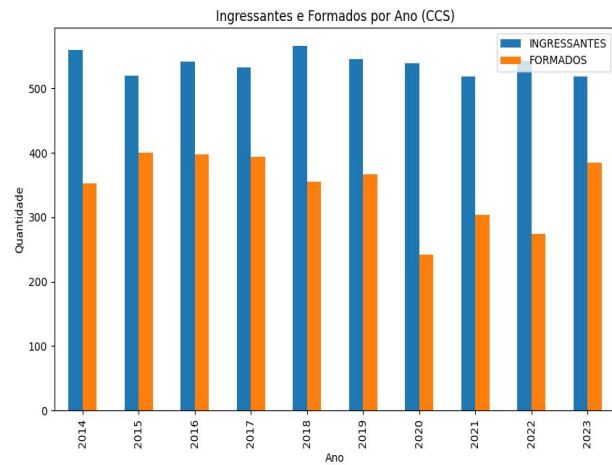
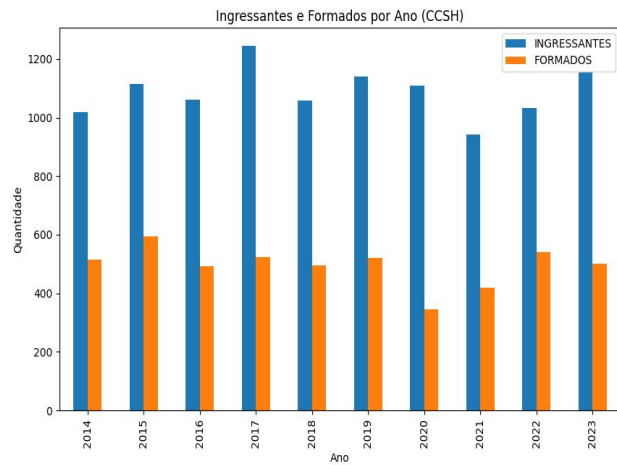




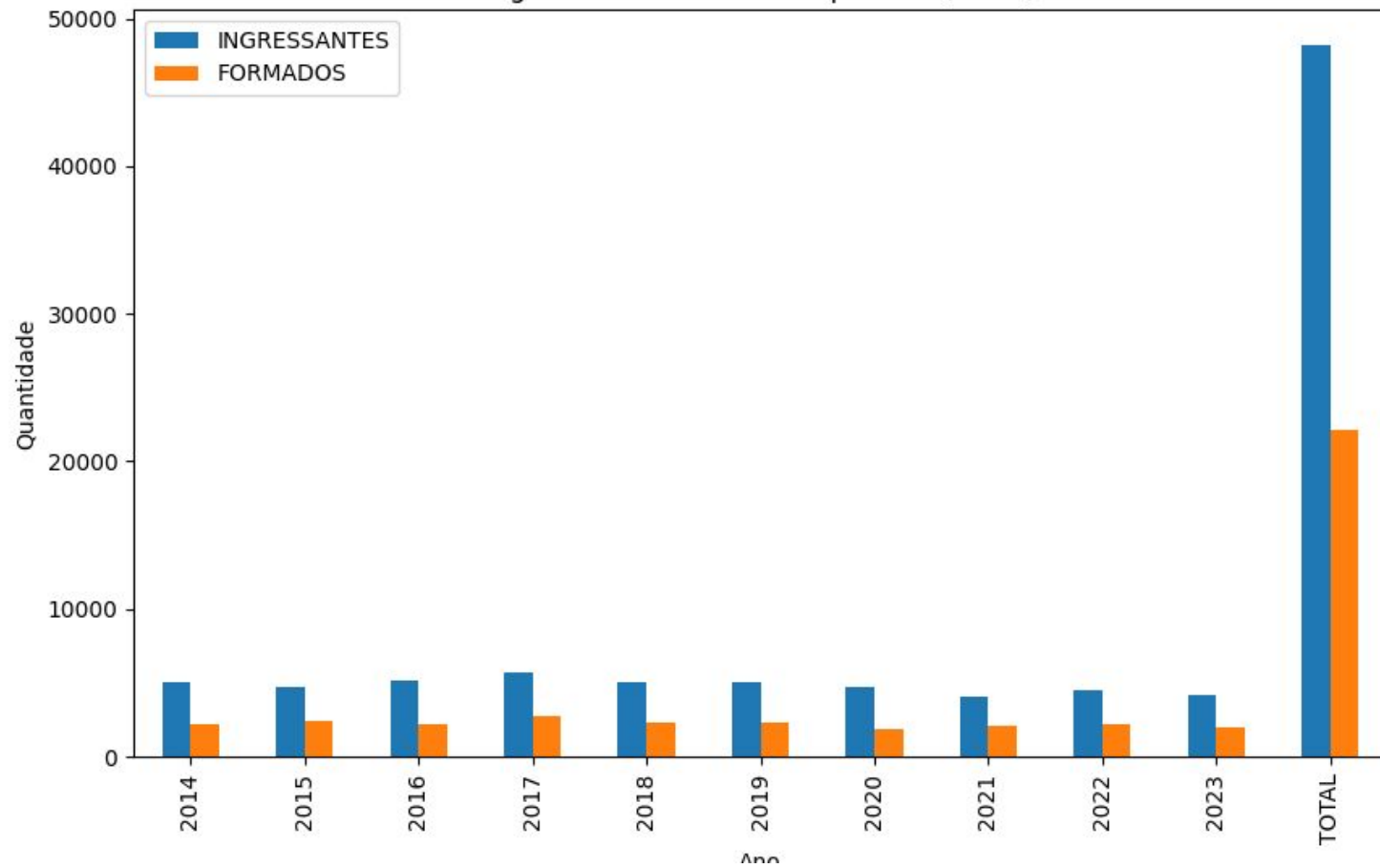




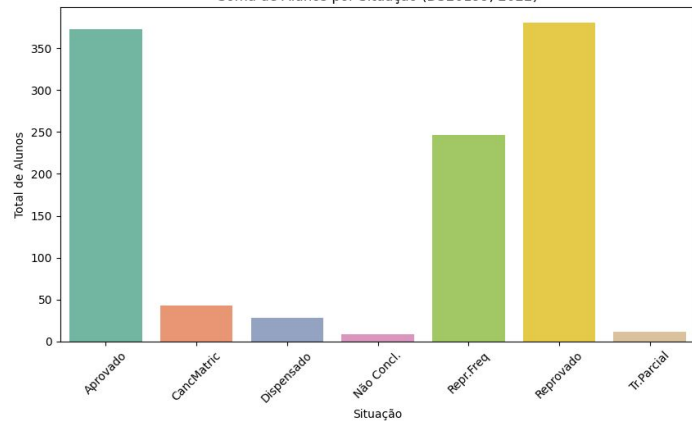




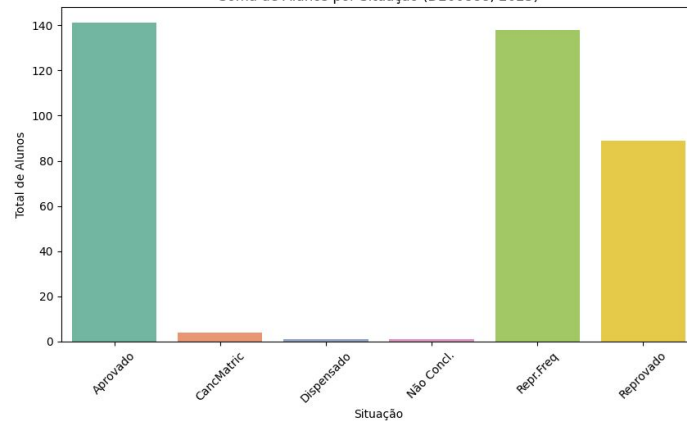
Ingressantes e Formados por Ano (Geral)



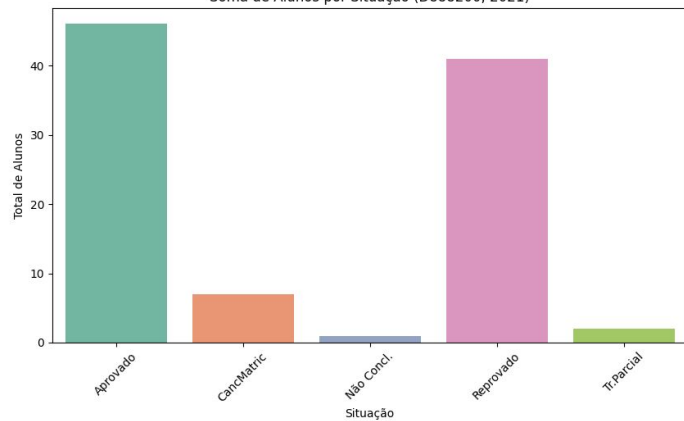
Soma de Alunos por Situação (D320199, 2022)



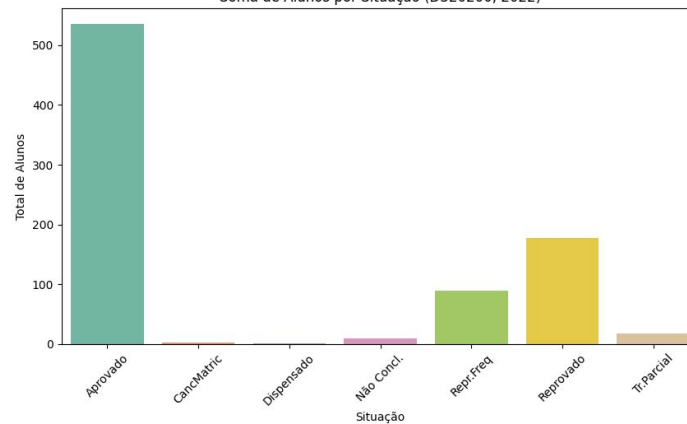
Soma de Alunos por Situação (D200888, 2023)



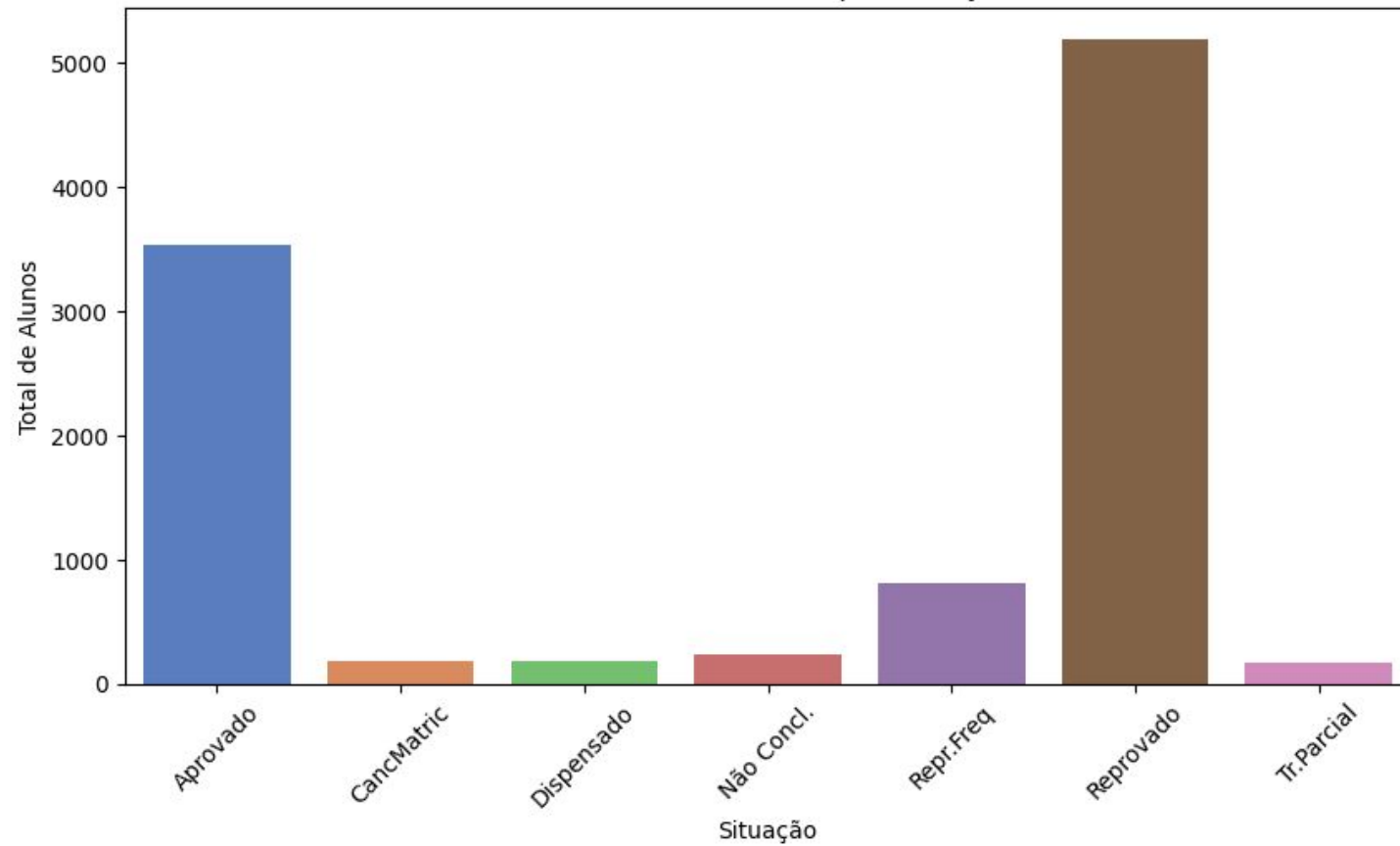
Soma de Alunos por Situação (D888200, 2021)



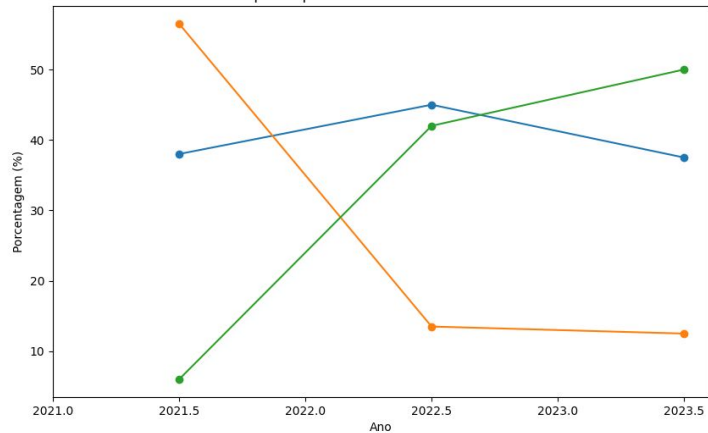
Soma de Alunos por Situação (D320200, 2022)



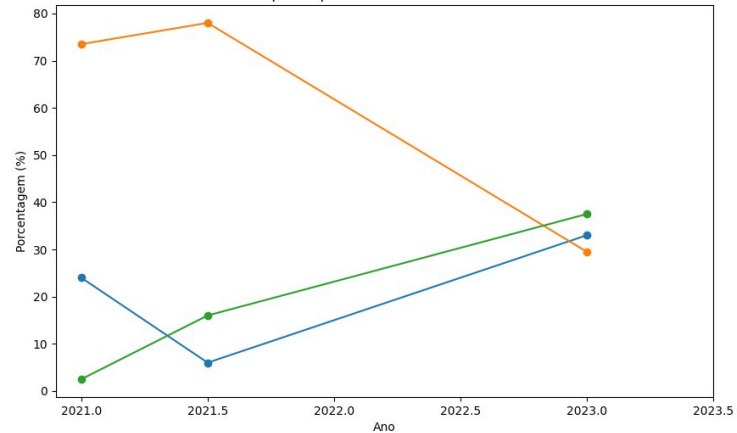
Soma Geral de Alunos por Situação



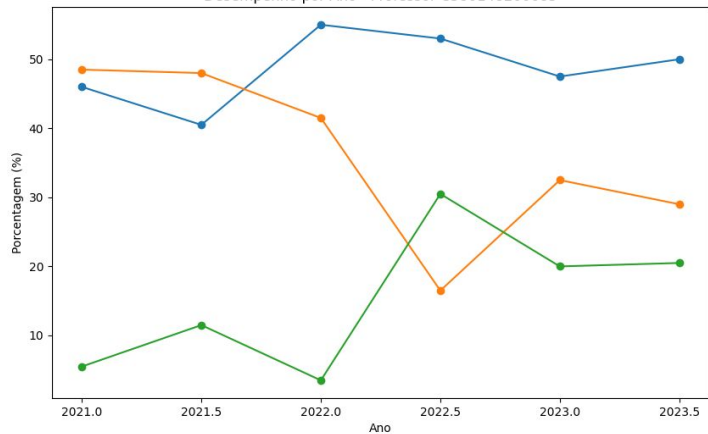
Desempenho por Ano - Professor C218200683888



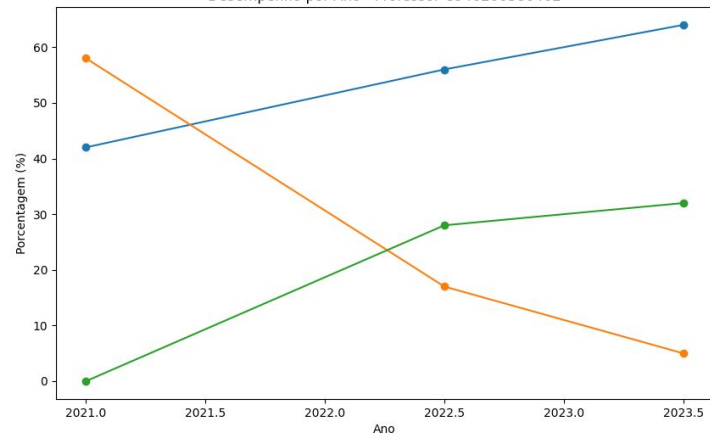
Desempenho por Ano - Professor P888402148



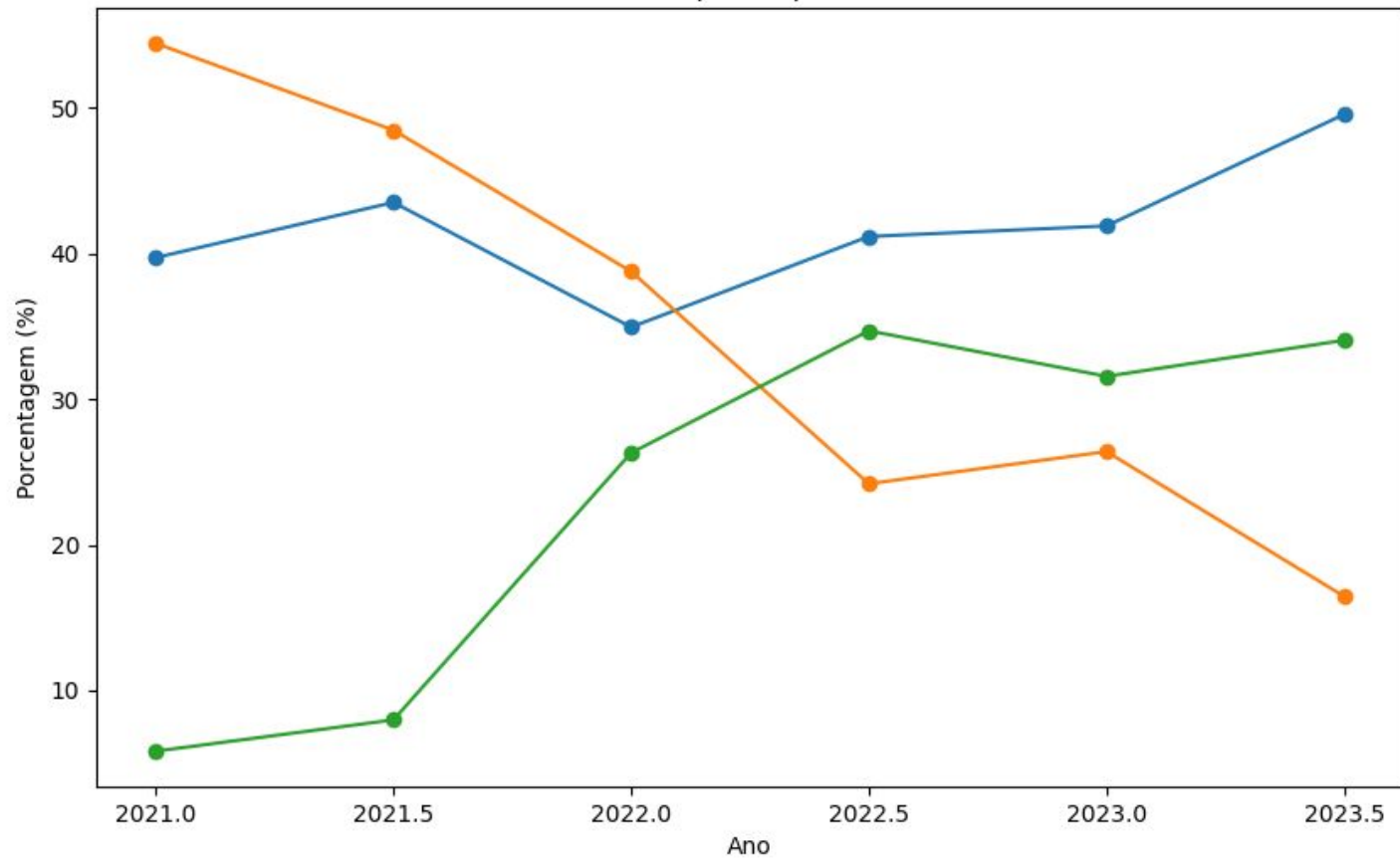
Desempenho por Ano - Professor C380148200685



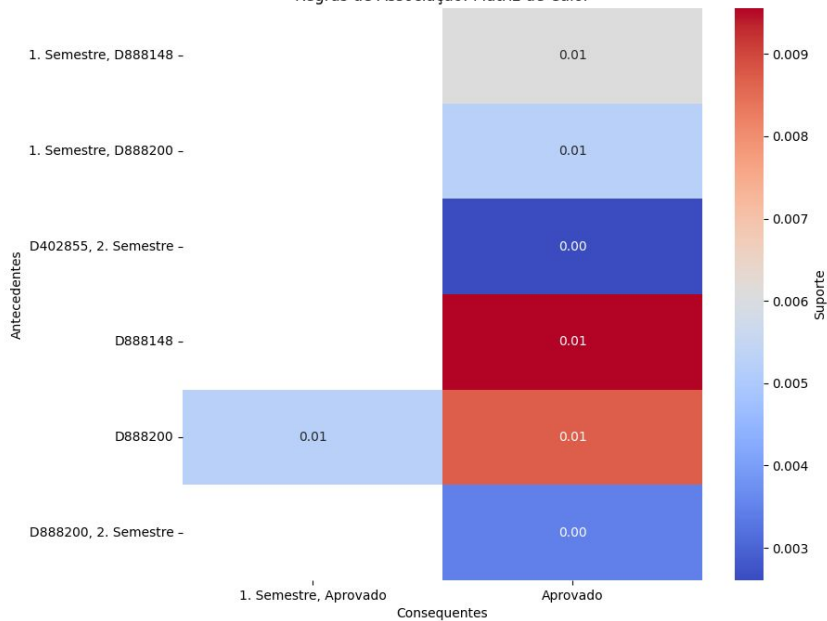
Desempenho por Ano - Professor C946200380402



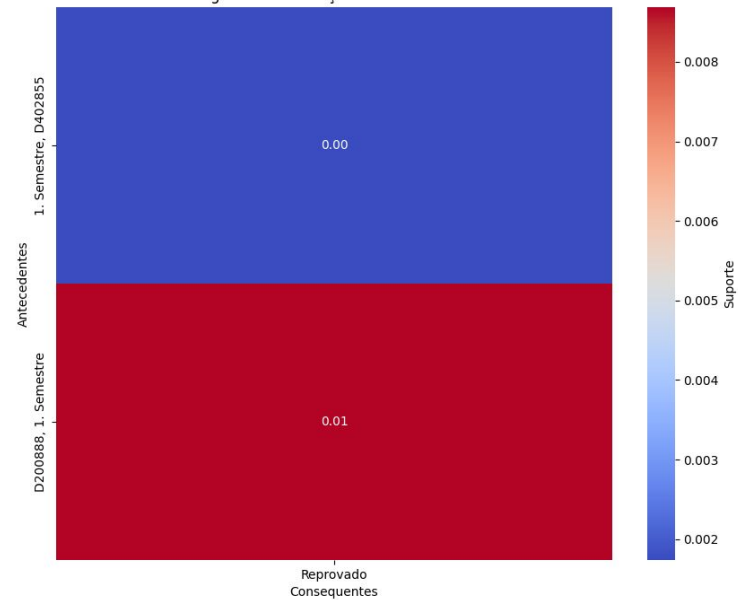
Desempenho por Ano



Regras de Associação: Matriz de Calor

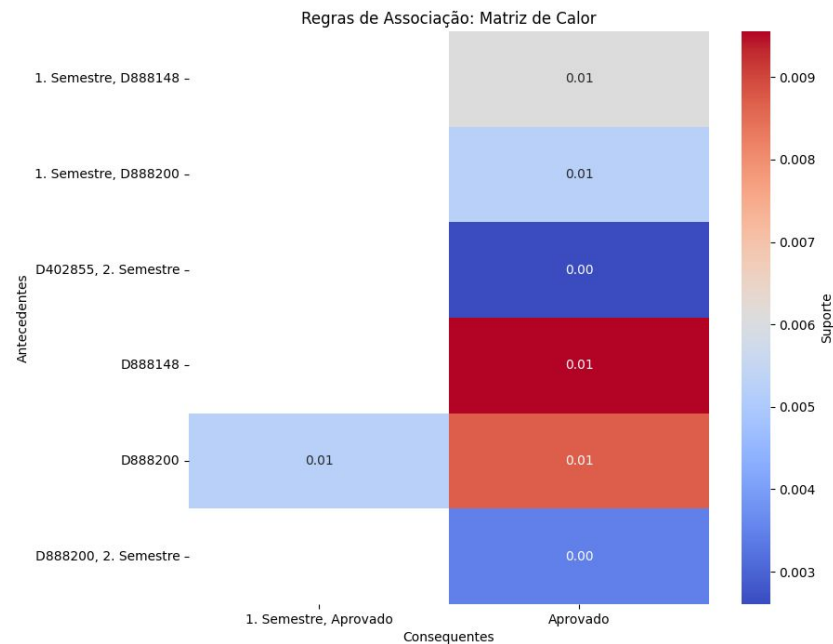


Regras de Associação: Matriz de Calor

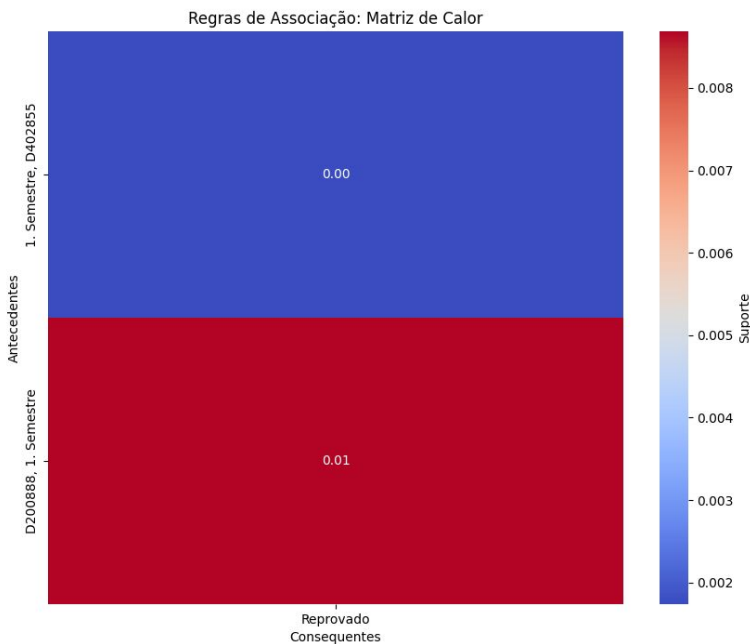




# Aprovação



# Reprovação





## Limpeza e filtragem dos dados para o conjunto de ingressantes e formados

```
RangeIndex: 80 entries, 0 to 79
Data columns (total 4 columns):
#   Column                      Non-Null Count  Dtype
---  -
0   CENTRO                      80 non-null     object
1   ANO                         80 non-null     float64
2   TOTAL_INGRESSANTES          80 non-null     float64
3   PERCENTUAL_MASCULINO        80 non-null     float64
dtypes: float64(3), object(1)
```



# Mineração do conjunto de ingressantes e formados via árvore de decisão

Métricas geradas pelo modelo:

Acurácia do modelo de Árvore de Decisão: 87.50%				
	precision	recall	f1-score	support
0	1.00	0.67	0.80	3
1	0.80	1.00	0.89	4
2	0.80	0.80	0.80	5
3	0.75	1.00	0.86	3
4	1.00	1.00	1.00	1
5	1.00	0.75	0.86	4
6	1.00	1.00	1.00	3
7	1.00	1.00	1.00	1
accuracy			0.88	24
macro avg	0.92	0.90	0.90	24
weighted avg	0.89	0.88	0.87	24



## Resultados testados

```
novo_ingressante = pd.DataFrame({  
    'ANO': [2016],  
    'TOTAL_INGRESSANTES': [600],  
    'PERCENTUAL_MASCULINO': [70]  
})
```

```
novo_ingressante = pd.DataFrame({  
    'ANO': [2016],  
    'TOTAL_INGRESSANTES': [900],  
    'PERCENTUAL_MASCULINO': [20]  
})
```

```
novo_ingressante = pd.DataFrame({  
    'ANO': [2016],  
    'TOTAL_INGRESSANTES': [600],  
    'PERCENTUAL_MASCULINO': [50]  
})
```

```
0 centro previsto para a entrada 1 é: CT  
0 centro previsto para a entrada 2 é: CE  
0 centro previsto para a entrada 3 é: CCNE
```