

# **TRABALHO FINAL DE MINERAÇÃO DE DADOS**

**Utilização de técnicas de mineração de dados para  
a extração de informações a partir de dados  
referentes a disciplinas e cursos da UFSM.**

**Augusto Kist Lunardi, Davi de Castro Machado, João Pedro  
Azenha Righi, Jonathan Weber Nogueira**

# ETAPAS DA KDD

- Seleção dos Dados
- Pré-Processamento
- Transformação
- Mineração
- Avaliação



# SELEÇÃO, PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS



COD_CURSO	NOME_UNIDA	ANO	SEXO	NGRESSANTES	FORMADOS	Arquivo_Origem
724	Artes Cenicas	2014	M	9	0	CAL.xlsx
724	Artes Cenicas	2014	F	12	0	CAL.xlsx
724	Artes Cenicas	2015	M	10	0	CAL.xlsx
724	Artes Cenicas	2015	F	17	0	CAL.xlsx
724	Artes Cenicas	2016	M	7	0	CAL.xlsx
724	Artes Cenicas	2016	F	16	0	CAL.xlsx
724	Artes Cenicas	2017	M	6	0	CAL.xlsx
724	Artes Cenicas	2017	F	17	0	CAL.xlsx
724	Artes Cenicas	2018	M	8	0	CAL.xlsx
724	Artes Cenicas	2018	F	12	0	CAL.xlsx
724	Artes Cenicas	2019	M	7	0	CAL.xlsx
724	Artes Cenicas	2019	F	17	0	CAL.xlsx
724	Artes Cenicas	2020	M	8	0	CAL.xlsx
724	Artes Cenicas	2020	F	13	0	CAL.xlsx
724	Artes Cenicas	2021	M	7	0	CAL.xlsx
724	Artes Cenicas	2021	F	16	0	CAL.xlsx
724	Artes Cenicas	2022	M	8	0	CAL.xlsx
724	Artes Cenicas	2022	F	14	0	CAL.xlsx
724	Artes Cenicas	2023	M	6	0	CAL.xlsx
724	Artes Cenicas	2023	F	14	0	CAL.xlsx
724	Artes Cenicas - TOTAL	M		76	0	CAL.xlsx
724	Artes Cenicas - TOTAL	F		148	0	CAL.xlsx

**ANÁLISE E  
INTERPRETAÇÃO  
DAS INFORMAÇÕES**

# TRATAMENTO DE ERROS E PADRONIZAÇÕES

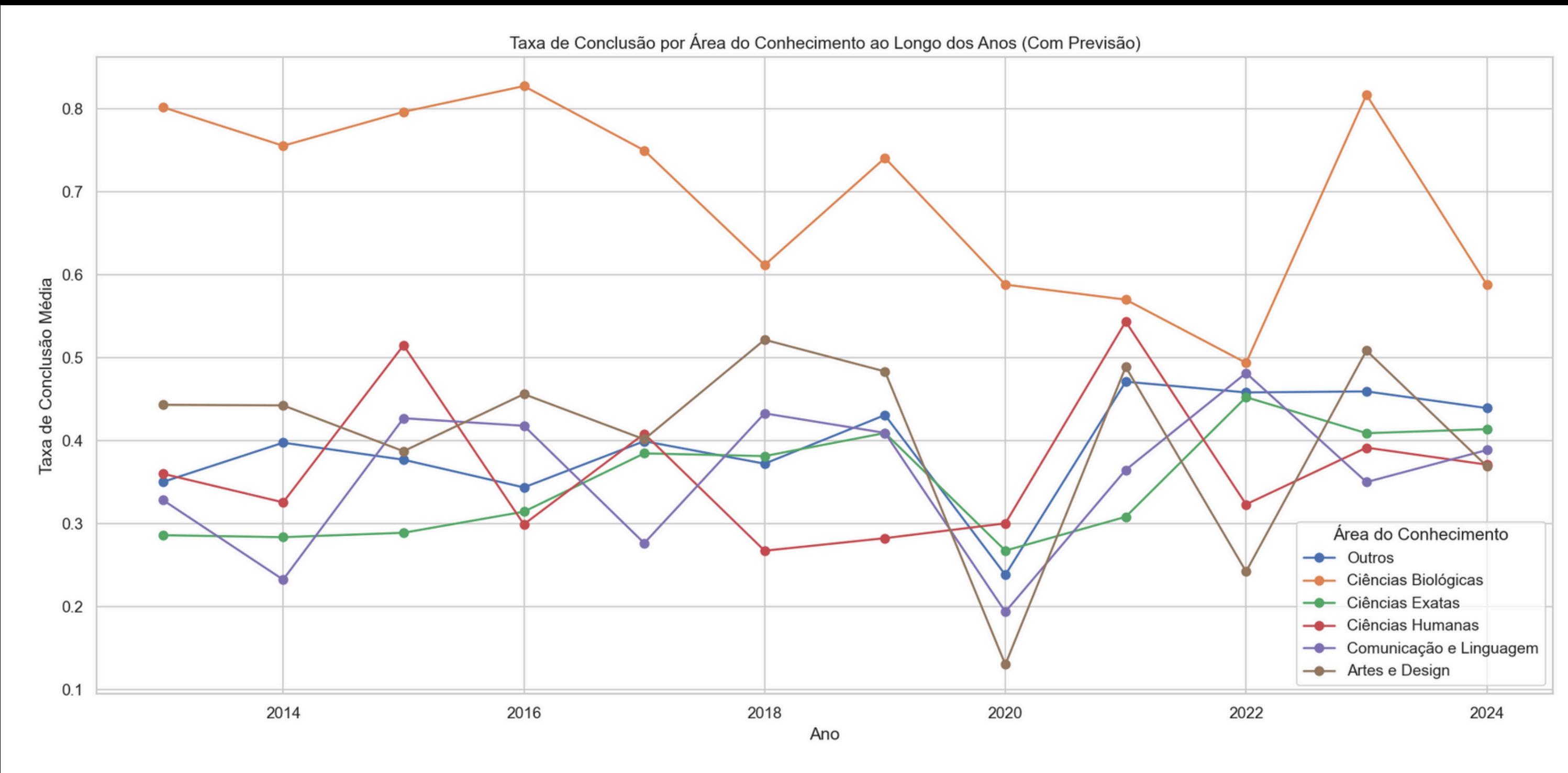
Ano	Semestre	Cód..Disciplina	Cód..Turma	Situação	%	Alunos	Professor	Cód..Curso	Curso	Arquivo_Origem
2021	1. Semestre	D200888	10	Aprovado	15	5	P888402148	139	Curso de Bach	D200888_2021.xlsx
2021	1. Semestre	D200888	10	Dispensado	3	1	P888402148	139	Curso de Bach	D200888_2021.xlsx
2021	1. Semestre	D200888	10	Reprovado	82	27	P888402148	139	Curso de Bach	D200888_2021.xlsx
2021	1. Semestre	D200888	11	Aprovado	31	17	P148855320	521	Ciências Econô	D200888_2021.xlsx
2021	1. Semestre	D200888	11	Reprovado	67	37	P148855320	521	Ciências Econô	D200888_2021.xlsx
2021	1. Semestre	D200888	11	Tr.Parcial	2	1	P148855320	521	Ciências Econô	D200888_2021.xlsx
2021	1. Semestre	D200888	11	Aprovado	54	20	P394200218	402	Engenharia Flo	D200888_2021.xlsx
2021	1. Semestre	D200888	11	Reprovado	46	17	P394200218	402	Engenharia Flo	D200888_2021.xlsx
2021	1. Semestre	D200888	12	Aprovado	20	3	P512200642	133	Química - Bach	D200888_2021.xlsx
2021	1. Semestre	D200888	12	Não Concl.	7	1	P512200642	133	Química - Bach	D200888_2021.xlsx
2021	1. Semestre	D200888	12	Reprovado	60	9	P512200642	133	Química - Bach	D200888_2021.xlsx
2021	1. Semestre	D200888	12	Tr.Parcial	13	2	P512200642	133	Química - Bach	D200888_2021.xlsx
2021	1. Semestre	D200888	12	Não Concl.	7	1	P512200642	110	Química Indust	D200888_2021.xlsx
2021	1. Semestre	D200888	12	Reprovado	93	13	P512200642	110	Química Indust	D200888_2021.xlsx
2021	1. Semestre	D200888	13	Aprovado	68	34	P394200218	302	Engenharia Civ	D200888_2021.xlsx
2021	1. Semestre	D200888	13	CancMatric	2	1	P394200218	302	Engenharia Civ	D200888_2021.xlsx
2021	1. Semestre	D200888	13	Dispensado	2	1	P394200218	302	Engenharia Civ	D200888_2021.xlsx
2021	1. Semestre	D200888	13	Não Concl.	10	5	P394200218	302	Engenharia Civ	D200888_2021.xlsx
2021	1. Semestre	D200888	13	Reprovado	18	9	P394200218	302	Engenharia Civ	D200888_2021.xlsx
2021	1. Semestre	D200888	14	Aprovado	33	16	P888402148	305	Engenharia Qu	D200888_2021.xlsx
2021	1. Semestre	D200888	14	Não Concl.	2	1	P888402148	305	Engenharia Qu	D200888_2021.xlsx
2021	1. Semestre	D200888	14	Reprovado	65	31	P888402148	305	Engenharia Qu	D200888_2021.xlsx

# MINERAÇÃO DE DADOS

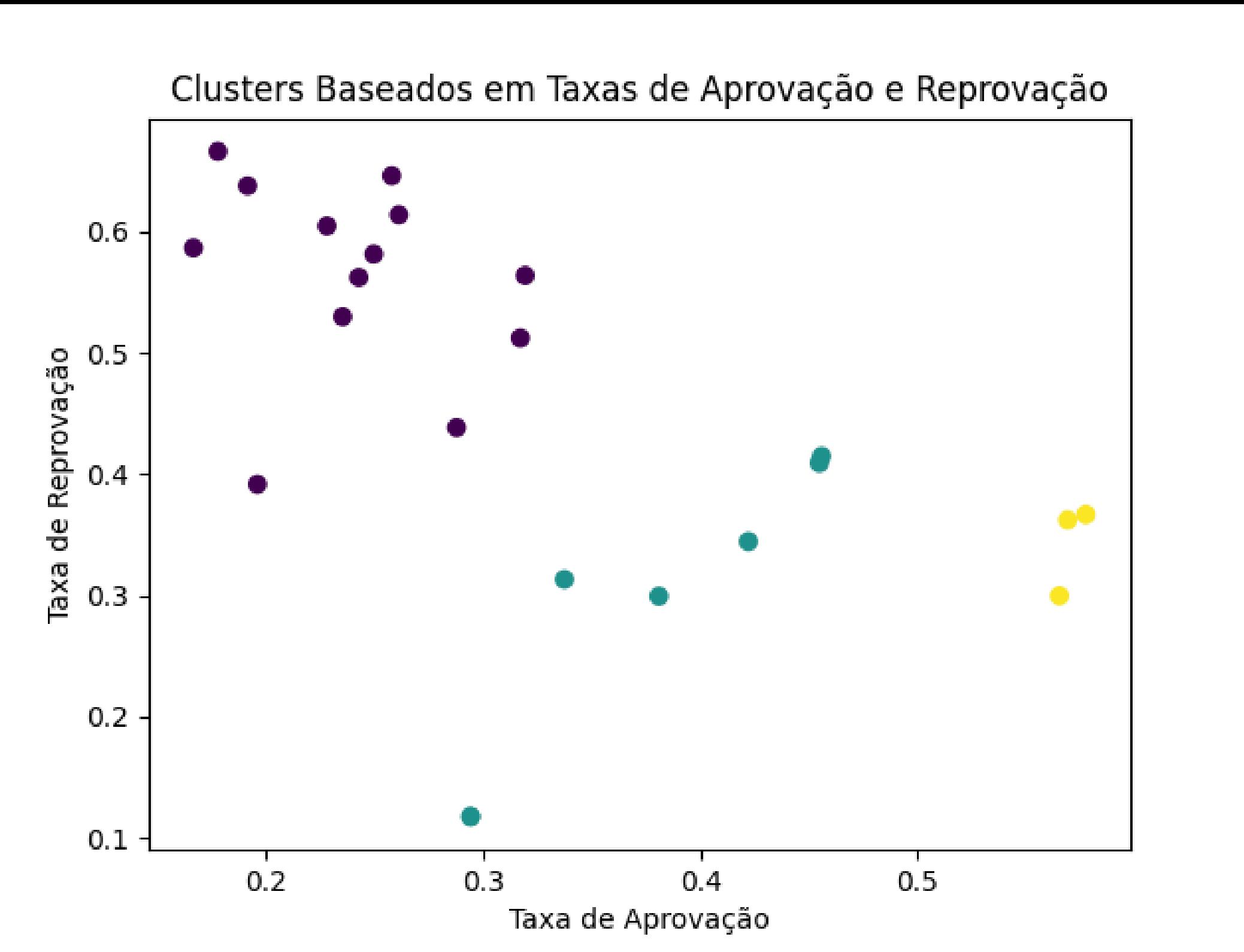
- Regressão Linear
- K-means
- Apriori



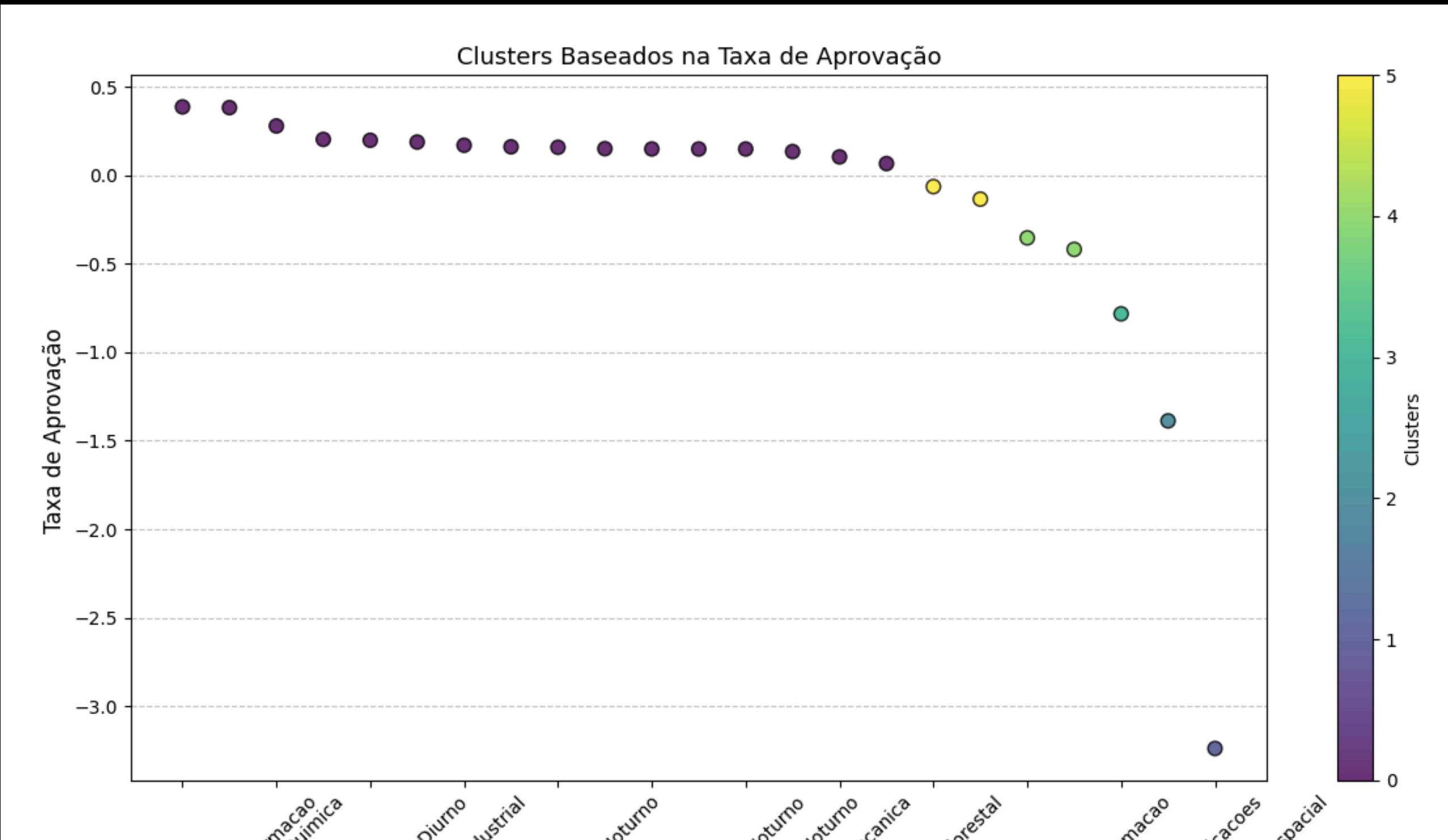
# REGRESSÃO LINEAR



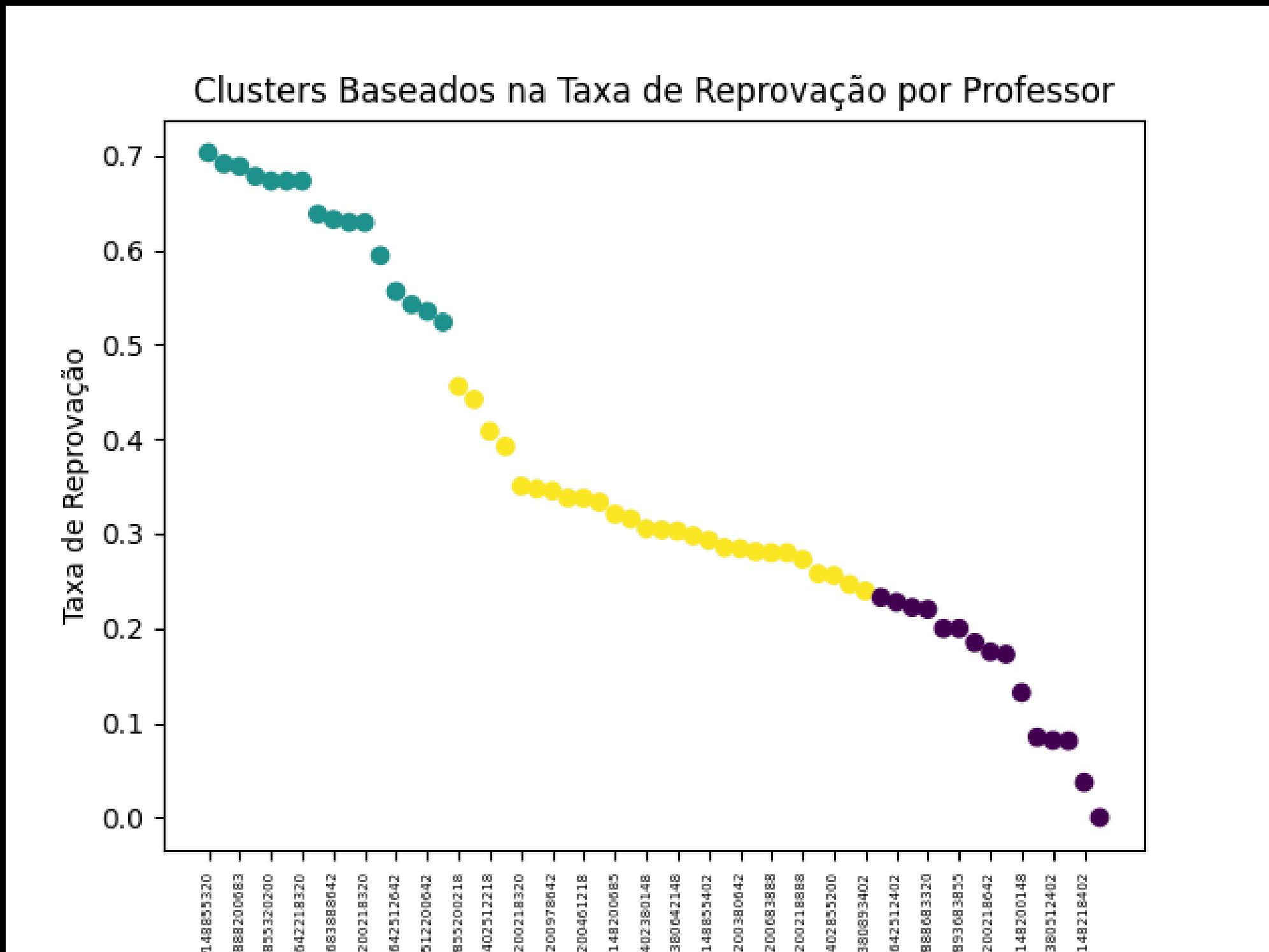
# K-MEANS



# K-MEANS



# K-MEANS



# REGRAS DE ASSOCIAÇÃO USANDO APRIORI (COM MAIOR CONFIANÇA)

Tabela_1	antecedents	consequents	support	confidence
	frozense({Curso de Engenharia Aeroespacial - A292200218320})	frozense({Aprovado})	0.0017376194613379669	0.6666666666666666
	frozense({Engenharia Química - A292200218320'})	frozense({Reprovado})	0.0017376194613379669	0.6666666666666666
	frozense({Engenharia Química - P380148200'})	frozense({Reprovado})	0.0017376194613379669	0.6666666666666666

Aqui podemos ver que a regra "Engenharia Química - Prof. A29220... → Reprovado" tem uma confiança de 66.67%, sugerindo uma alta taxa de reprovação nesse curso com este professor. Similarmente, a regra "Engenharia Aeroespacial - Prof. A292200... → Aprovado" também tem alta confiança, indicando uma boa taxa de aprovação com o mesmo professor em um curso diferente. Essas informações podem servir para investigar o ensino deste professor perante estes dois cursos.

# REGRAS DE ASSOCIAÇÃO USANDO APRIORI (COM MAIOR SUPORTE)

Tabela_1	antecedents	consequents	support	confidence
frozense({Ciéncia da Computaçáo - Bacharelado - C380148200685})	frozense({Aprovado})	0.005212858384013901	0.4000000000000001	
frozense({Bacharelado em Sistemas de Informaçáo - C380148200685})	frozense({Aprovado})	0.005212858384013901	0.31578947368421056	
frozense({Bacharelado em Sistemas de Informaçáo - C380148200685})	frozense({Reprovado})	0.005212858384013901	0.31578947368421056	

As regras com maior suporte indicam padrões recorrentes. A regra "Sistemas de Informação - Prof. C38014... → Aprovado" possui alto suporte, assim como a regra "Ciéncia da Computaçáo - Prof. C38014 → Aprovado". Isso significa que essas são combinações frequentes no conjunto de dados analisados. Essas regras ajudam a identificar tendências de desempenho entre os cursos.

# AVALIAÇÃO E CONCLUSÕES

