



Universidade Federal do Rio Grande do Norte
Instituto Metrópole Digital



RELATÓRIO DO PROJETO 1: Cursos de Graduação no Brasil

Francisco Willem Romão Moreira
20220036966

Natal-RN
2024

RESUMO

Este relatório apresenta uma análise detalhada dos cursos de graduação no Brasil, utilizando dados extraídos da página de Dados Abertos do Ministério da Educação. O processo incluiu etapas de pré-processamento, como a verificação e tratamento de dados nulos e identificação e tratamento de outliers e ruídos. A análise abrangeu a distribuição dos cursos por região, perfil por modalidade de ensino e situação dos cursos. Os resultados mostram uma maior concentração de cursos na região Sudeste e uma predominância de cursos a distância. A maioria dos cursos está em atividade, com uma menor proporção em extinção ou extintos. A análise final foi focada nos cursos de Tecnologia da Informação, refletindo a importância da tecnologia nos dias atuais. Este estudo fornece uma visão abrangente da oferta de cursos de graduação no Brasil e destaca para futuras pesquisas, como a identificação de clusters de cursos com características semelhantes e a classificação do risco de extinção.

Palavras-chave: ciência de dados, pré-processamento, análise de dados, cursos de graduação brasil, cursos de tecnologia da informação.

LISTA DE FIGURAS

1	Informações gerais do dataset	9
2	Informações sobre valores nulos nas colunas do dataset	9
3	Replace para padronizar nome dos cursos de tecnologia.	10
4	Ruídos em ordem decrescente QT_VAGAS_AUTORIZADAS.	11
5	Ruídos em ordem crescente QT_VAGAS_AUTORIZADAS.	11
6	Limpeza QT_VAGAS_AUTORIZADAS.	11
7	Resultado limpeza QT_VAGAS_AUTORIZADAS.	12
8	Ruídos em ordem decrescente CARGA_HORARIA.	12
9	Ruídos em ordem crescente CARGA_HORARIA.	12
10	Limpeza CARGA_HORARIA.	13
11	Resultado limpeza CARGA_HORARIA.	13
12	Tipos de dados de cada coluna.	13
13	Os 10 Cursos mais Frequentes no Brasil.	14
14	Regiões com mais Cursos Em Atividade.	15
15	Regiões com mais Cursos Em Atividade.	15
16	Regiões com mais Cursos Em Atividade.	15
17	Participação das regiões no PIB.	16
18	Estados com mais Cursos Em Atividade.	16
19	Os 15 Municípios do RN com mais Cursos Em Atividade.	17
20	Distribuição dos Cursos Em Atividade por Modalidade de Ensino.	18
21	Quantidade de Cursos EaD e Presencial por Região.	18
22	Quantidade de Cursos EaD e Presencial por Estado.	19
23	Quantidade de Cursos Presencial e EaD por Municípios do RN.	19
24	Distribuição de Cursos Presenciais por Grau.	20
25	Distribuição de Cursos EaD por Grau.	20
26	Distribuição de Cursos Presenciais por Grau no RN.	21
27	Distribuição de Cursos EaD por Grau no RN.	21
28	Distribuição por Categoria Administrativa (Educação Presencial).	22
29	Distribuição por Categoria Administrativa (EaD).	22
30	Situação dos Cursos.	25
31	Situação dos Cursos Gráfico de Barras.	25
32	Situação dos Cursos do RN.	26
33	Distribuição da Situação dos Cursos de Educação Presencial.	27
34	Distribuição da Situação dos Cursos de Educação a Distância.	27
35	Distribuição de Cursos por Categoria Administrativa e Situação (Em atividade)	28
36	Distribuição de Cursos por Categoria Administrativa e Situação (Em extinção)	28
37	Distribuição de Cursos por Categoria Administrativa e Situação (Extinto)	29
38	Distribuição das Situações de Curso por Região	29
39	BTI UFRN	30
40	Distribuição dos cursos de TI no Brasil	31
41	Distribuição dos Cursos de TI no Brasil por Região	31
42	Distribuição dos Cursos de TI no Brasil por Estado	32
43	Os 15 Municípios do RN com mais Cursos de TI	33
44	Distribuição dos Cursos de TI Em Atividade por Modalidade de Ensino	34

		3
45	Distribuição dos Cursos de TI Modalidade Presencial e a Distância	34
46	Situação dos Cursos de TI	35
47	Cursos TI Extintos e Em Extinção	35
48	Total de Vagas Autorizadas por Curso	36
49	Média da Carga Horária por Curso	37

LISTA DE TABELAS

1	Quantidades únicas das variáveis do dataset	7
2	Distribuição de Cursos por Região	14
3	Distribuição de Cursos por Município	17
4	Distribuição de Cursos por Grau Presencial	19
5	Distribuição de Cursos por Grau EaD	20
6	Distribuição de Cursos por Grau no RN	21
7	Modalidade Presencial	23
8	Modalidade EaD	24
9	Distribuição de Cursos por Situação	25
10	Distribuição de Cursos por Situação RN	26
11	Distribuição de Cursos por Modalidade e Situação	26
12	Distribuição de Cursos por Categoria Administrativa e Situação	27
13	Distribuição de Cursos por Região e Situação	29
14	Quantidade de Cursos na Área de Tecnologia da Informação	30
15	Quantidade de Cursos de TI no Rio Grande do Norte	32
16	Distribuição de Cursos de TI por Modalidade	33
17	Média de Carga Horária por Curso	36

SUMÁRIO

1	INTRODUÇÃO	6
1.1	Contexto	6
1.2	Objetivo	6
1.3	Importância	6
2	METODOLOGIA	7
2.1	Descrição dos Dados.....	7
2.2	Pré-processamento dos Dados - Limpeza e Transformação	7
2.2.1	<i>Exploração do dataset e seleção de variáveis</i>	<i>7</i>
2.2.2	<i>Tratamento de valores ausentes</i>	<i>9</i>
2.2.3	<i>Outliers e ruídos</i>	<i>10</i>
2.2.4	<i>Transformação.....</i>	<i>13</i>
3	ANÁLISE E RESULTADOS	14
3.1	Análise Exploratória de Dados (EDA)	14
3.2	Distribuição dos cursos por região	14
3.3	Perfil dos cursos por modalidade de ensino.....	17
3.4	Situação dos cursos.....	24
3.5	Cursos de tecnologia.....	29
4	DISCUSSÃO	37
4.1	Interpretação dos resultados	37
4.2	Limitações	37
4.3	Implicações.....	37
5	CONCLUSÃO.....	38
5.1	Resumo dos achados.....	38
5.2	Trabalhos futuros.....	38
	REFERÊNCIAS.....	39

1 INTRODUÇÃO

1.1 Contexto

O ensino superior no Brasil desempenha um papel de extrema importância no desenvolvimento econômico, social e cultural do país. Com uma vasta diversidade de instituições e cursos, o setor enfrenta desafios que necessitam de análises para promover melhorias. Este projeto se propõe a realizar uma análise abrangente dos cursos de graduação no Brasil, focando em aspectos como a distribuição geográfica dos cursos, o perfil por modalidade de ensino, a situação atual dos cursos e uma análise específica dos principais cursos de tecnologia.

1.2 Objetivo

O objetivo principal deste estudo é fornecer uma visão ampla sobre a estrutura e a situação dos cursos de graduação no Brasil. A análise será conduzida com base nos dados disponíveis na página de Dados Abertos do Ministério da Educação (MEC, 2022). As áreas de foco incluem:

1. **Distribuição dos Cursos por Região:** Explorar a distribuição geográfica dos cursos de graduação no Brasil por estado. Identificar quais estados possuem mais cursos e quais têm menos. Analisar se existe alguma correlação entre a disponibilidade de cursos e indicadores socioeconômicos das regiões.
2. **Perfil dos Cursos por Modalidade de Ensino:** Comparar a quantidade e a distribuição dos cursos de graduação (Em Atividade), Presenciais e EaD. Analisar se existem diferenças nas áreas de estudo mais comuns entre as duas modalidades, bem como no grau dos cursos e na carga horária.
3. **Situação dos Cursos:** Avaliar a proporção de cursos ativos e inativos. Analisar se existe alguma relação entre a situação dos cursos e a categoria da instituição de ensino superior e a modalidade.
4. **Cursos de Tecnologia:** Nesta seção, decidi seguir a mesma linha das análises anteriores, porém, com um filtro para os principais cursos de Tecnologia da Informação no Brasil.

1.3 Importância

Este estudo é de extrema importância para diversos stakeholders, incluindo formuladores de políticas, administradores educacionais e pesquisadores. A análise feita pode fornecer insights sobre a distribuição dos cursos de graduação, ajudar a identificar áreas que necessitam de intervenção e melhorias, e promover uma melhor compreensão da evolução e dinâmica dos cursos de tecnologia no Brasil. Em última análise, os resultados podem orientar estratégias para a expansão e aprimoramento do ensino superior no país, promovendo maior equidade e qualidade educacional.

2 METODOLOGIA

2.1 Descrição dos Dados

A origem dos dados é a página de dados abertos do Ministério da Educação (MEC, 2022). O dataset foi coletado diretamente dessa página por meio do download do arquivo CSV. As principais características do dataset incluem o detalhamento quantitativo dos Cursos de Graduação (Licenciatura, Bacharelado, Tecnológico, Sequencial e ABI - Área Básica de Ingresso) no Brasil, segmentados por diversas variáveis: código da Instituição de Educação Superior (IES), nome da IES, categoria da IES, organização acadêmica, código do curso, nome do curso, grau, área OCDE, modalidade de ensino (presencial ou EaD), situação do curso (ativo ou inativo), vagas autorizadas, carga horária, código do município (IBGE), município, UF e região.

2.2 Pré-processamento dos Dados - Limpeza e Transformação

2.2.1 Exploração do dataset e seleção de variáveis

Inicialmente o dataset tinha 902.676 linhas e 18 colunas, as colunas eram formada pelas variáveis: CODIGO_IES, NOME_IES, CATEGORIA_ADMINISTRATIVA, ORGANIZACAO_ACADEMICA, CODIGO_CURSO, NOME_CURSO, GRAU, AREA_OCDE, MODALIDADE, SITUACAO_CURSO, QT_VAGAS_AUTORIZADAS, CARGA_HORARIA, CODIGO_AREA_OCDE_CINE, AREA_OCDE_CINE, CODIGO_MUNICIPIO, MUNICIPIO, UF e REGIAO.

Ao aplicar um `nunique()` obtive o seguinte resultado:

Tabela 1 – Quantidades únicas das variáveis do dataset

Coluna	Quantidade Única
CODIGO_IES	3706
NOME_IES	3672
CATEGORIA_ADMINISTRATIVA	6
ORGANIZACAO_ACADEMICA	6
CODIGO_CURSO	86239
NOME_CURSO	2206
GRAU	5
AREA_OCDE	408
MODALIDADE	2
SITUACAO_CURSO	3
QT_VAGAS_AUTORIZADAS	857
CARGA_HORARIA	3435
CODIGO_AREA_OCDE_CINE	423
AREA_OCDE_CINE	519
CODIGO_MUNICIPIO	3439
MUNICIPIO	3325
UF	28
REGIAO	6

Fonte: elaborado pelo autor (2024)

Note a diferença que existe na quantidade de dados únicos entre a coluna CODIGO_IES e a coluna NOME_IES, explorando o dataset percebi que a coluna CODIGO_IES,

algumas vezes mostrava códigos diferentes quando se tratava de uma mesma instituição. Dado essa situação, e também pela natureza do trabalho não vi o porque de utilizar a coluna CODIGO_IES.

A mesma situação acontece com as colunas CODIGO_MUNICIPIO e MUNICIPIO. CODIGO_CURSO e NOME_CURSO. Indicando possíveis erros no preenchimento do dataset.

A coluna AREA_OCDE e AREA_OCDE_CINE consistia em vários dados muito diversos e que não representava com abrangência a área no sentido geral, e sim em sentido mais específico. Exemplo: ao invés da área ser educação, eles colocam Pedagogia, assim como, Educação Física ao invés de ser da área saúde, é a própria educação física. Então para fins de avaliação não me ajudará a responder as perguntas de minhas análises futuras.

Assim, posso fazer a remoção das seguintes colunas: CODIGO_IES, CODIGO_CURSO, AREA_OCDE, CODIGO_AREA_OCDE_CINE, AREA_OCDE_CINE, CODIGO_MUNICIPIO.

Com isso tenho minha seleção de features, ou seja, as colunas/variáveis com quem eu quero trabalhar, são: NOME_IES, CATEGORIA_ADMINISTRATIVA, ORGANIZACAO_ACADEMICA, NOME_CURSO, GRAU, MODALIDADE, SITUACAO_CURSO, QT_VAGAS_AUTORIZADAS, CARGA_HORARIA, MUNICIPIO, UF e REGIAO. Saindo de 18 colunas para 12, com uma redução de 6 colunas.

A seguir, veremos do que se tratam algumas colunas no que tange a seus valores únicos.

- **NOME_CURSO:**

Exemplos de valores únicos: ENGENHARIA AGRONÔMICA, LETRAS, GEOGRAFIA, GESTÃO DO SOCIAL, INTERDISCIPLINAR EM CULTURA, LINGUAGENS E TECNOLOGIAS APLICADAS, FÍSICA COM HABILITAÇÃO EM PESQUISA BÁSICA.

- **CATEGORIA_ADMINISTRATIVA:**

Valores únicos: Privada com fins lucrativos, Privada sem fins lucrativos, Pública Municipal, Pública Federal, Pública Estadual, Especial.

- **ORGANIZACAO_ACADEMICA:**

Valores únicos: Centro Universitário, Universidade, Faculdade, Instituto Federal de Educação, Ciência e Tecnologia, Centro Federal de Educação Tecnológica, Instituição Especialmente Credenciada para oferta de cursos lato sensu.

- **GRAU:**

Valores únicos: Bacharelado, Licenciatura, Tecnológico, Área Básica de Ingresso (ABI), Sequencial.

- **MODALIDADE:**

Valores únicos: Educação a Distância, Educação Presencial.

- **SITUACAO_CURSO:**

Valores únicos: Em atividade, Em extinção, Extinto.

- **UF:**

Valores únicos: SP, PB, SE, PR, PE, RN, SC, RJ, RS, GO, BA, MG, MT, CE, PA, RO, TO, AM, ES, AC, MA, AL, DF, ZZ, MS, PI, RR, AP.

- **REGIAO:**

Valores únicos: SUDESTE, NORDESTE, SUL, CENTRO-OESTE, NORTE, IGNORADO/EXTERIOR.

Sabendo que não existe estado do Brasil com as siglas ZZ, então vamos analisar de qual município e região se trata. Quantos cursos existem nessa categoria? Existem pouco menos de 200 cursos na categoria ZZ. Dado que a categoria ZZ da coluna UF e a categoria IGNORADO/EXTERIOR da coluna REGIAO tem a mesma quantidade de cursos (191), então ambas correspondem ao mesmo perfil. E como eu quero analisar apenas o Brasil resolvi fazer uma remoção.

```
df = df.drop(df[(df['UF'] == 'ZZ') & (df['REGIAO'] == 'IGNORADO/EXTERIOR')].index)
```

2.2.2 Tratamento de valores ausentes

Figura 1 – Informações gerais do dataset

```
[ ] df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 902068 entries, 0 to 902675
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   NOME_IES                             902068 non-null object
1   CATEGORIA_ADMINISTRATIVA             902068 non-null object
2   ORGANIZACAO_ACADEMICA                902068 non-null object
3   NOME_CURSO                           902068 non-null object
4   GRAU                                 902068 non-null object
5   MODALIDADE                           902068 non-null object
6   SITUACAO_CURSO                       902068 non-null object
7   QT_VAGAS_AUTORIZADAS                 902068 non-null int64
8   CARGA_HORARIA                       902068 non-null int64
9   MUNICIPIO                           902068 non-null object
10  UF                                   902068 non-null object
11  REGIAO                              902068 non-null object
dtypes: int64(2), object(10)
memory usage: 89.5+ MB
```

Figura 2 – Informações sobre valores nulos nas colunas do dataset

```
df.isnull().sum()

NOME_IES                             0
CATEGORIA_ADMINISTRATIVA             0
ORGANIZACAO_ACADEMICA                0
NOME_CURSO                           0
GRAU                                 0
MODALIDADE                           0
SITUACAO_CURSO                       0
QT_VAGAS_AUTORIZADAS                 0
CARGA_HORARIA                       0
MUNICIPIO                           0
UF                                   0
REGIAO                              0
dtype: int64
```

As figuras acima mostram que não existe valores nulos ou ausentes no dataset.

2.2.3 Outliers e ruídos

A limpeza que feita servirá apenas para análises futuras dos cursos de TI a despeito da quantidade de vagas e da carga horária. Com base em (INSPER, 2023), filtro os cursos mais procurados e com maior relevância no Brasil. São eles:

- Tecnologia da Informação
- Ciência da Computação
- Ciências da Computação
- Engenharia da Computação
- Engenharia de Computação
- Engenharia de Software
- Análise e Desenvolvimento de Sistemas
- Sistemas de Informação
- Gestão da Tecnologia da Informação
- Engenharia de Telecomunicações
- Redes de Computadores
- Engenharia Mecatrônica
- Engenharia de Controle e Automação
- Sistemas para Internet
- Banco de Dados
- Jogos Digitais

Analizando com unique todos os cursos, percebi que alguns cursos como o de Ciência da Computação e Engenharia da Computação divergia no nome, então apliquei um replace para todos esses casos estabelecendo o padrão abaixo.

Figura 3 – Replace para padronizar nome dos cursos de tecnologia.

```
dfti['NOME_CURSO'] = dfti['NOME_CURSO'].replace('CIÊNCIAS DA COMPUTAÇÃO', 'CIÊNCIA DA COMPUTAÇÃO')
dfti['NOME_CURSO'] = dfti['NOME_CURSO'].replace('CIÊNCIAS DE COMPUTAÇÃO', 'CIÊNCIA DA COMPUTAÇÃO')

dfti['NOME_CURSO'] = dfti['NOME_CURSO'].replace('ENGENHARIA DA COMPUTAÇÃO', 'ENGENHARIA DE COMPUTAÇÃO')

dfti['NOME_CURSO'] = dfti['NOME_CURSO'].replace('GESTÃO EM TECNOLOGIA DA INFORMAÇÃO', 'GESTÃO DA TECNOLOGIA DA INFORMAÇÃO')
dfti['NOME_CURSO'] = dfti['NOME_CURSO'].replace('GESTÃO DE TECNOLOGIA DA INFORMAÇÃO', 'GESTÃO DA TECNOLOGIA DA INFORMAÇÃO')
```

Com isso, primeiro analisei possíveis ruídos e outliers na coluna de quantidade de vagas autorizadas.

Figura 4 – Ruídos em ordem decrescente QT_VAGAS_AUTORIZADAS.

```
dfti[['QT_VAGAS_AUTORIZADAS', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['QT_VAGAS_AUTORIZADAS'], ascending=False).head(10)
```

	QT_VAGAS_AUTORIZADAS	NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
489344	73260	REDES DE COMPUTADORES	UNIVERSIDADE PAULISTA	Educação a Distância	Itapetininga
801603	73260	REDES DE COMPUTADORES	UNIVERSIDADE PAULISTA	Educação a Distância	Niterói
81190	73260	GESTÃO DA TECNOLOGIA DA INFORMAÇÃO	UNIVERSIDADE PAULISTA	Educação a Distância	Tailândia
364854	73260	REDES DE COMPUTADORES	UNIVERSIDADE PAULISTA	Educação a Distância	Guarantã do Norte
550990	73260	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE PAULISTA	Educação a Distância	Cruzeiro do Sul

Figura 5 – Ruídos em ordem crescente QT_VAGAS_AUTORIZADAS.

```
[ ] dfti[['QT_VAGAS_AUTORIZADAS', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['QT_VAGAS_AUTORIZADAS'], ascending=True).head(10)
```

	QT_VAGAS_AUTORIZADAS	NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
896478	0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	Universidade do Estado de Mato Grosso Carlos A...	Educação Presencial	Nova Mutum
508215	0	SISTEMAS DE INFORMAÇÃO	UNIVERSIDADE DE SANTA CRUZ DO SUL	Educação Presencial	Sobradinho
440217	0	ENGENHARIA DE COMPUTAÇÃO	UNIVERSIDADE DO ESTADO DO AMAZONAS	Educação Presencial	Manaus
127965	0	SISTEMAS DE INFORMAÇÃO	UNIVERSIDADE DE SANTA CRUZ DO SUL	Educação Presencial	Venâncio Aires
213479	0	REDES DE COMPUTADORES	Fatec Cruzeiro - Prof. Waldomiro May	Educação Presencial	Cruzeiro

Pesquisando sobre a distribuição da quantidade de vagas em cursos no Brasil, não achei nenhuma informação. Existem muitos outliers conforme o gráfico de boxplot aplicado no notebook e das análises do dataframe, com cursos individuais com mais de 10.000 vagas e até mesmo cursos com 73.000 e 0 vagas! Apesar de serem cursos EaD, não posso confirmar se isso é possível. De qualquer forma vamos fazer o tratamento desses possíveis outliers/ruídos.

Figura 6 – Limpeza QT_VAGAS_AUTORIZADAS.

```
[ ] dfti_limpo_qtv = dfti[
    (dfti['QT_VAGAS_AUTORIZADAS'] > dfti['QT_VAGAS_AUTORIZADAS'].quantile(0.001)) &
    (dfti['QT_VAGAS_AUTORIZADAS'] < dfti['QT_VAGAS_AUTORIZADAS'].quantile(0.600))]

# dfti_limpo_qtv = dataframe ti em atividade limpo para quantidade de vagas

[ ] dfti_limpo_qtv['QT_VAGAS_AUTORIZADAS'] = np.where(
    dfti_limpo_qtv['QT_VAGAS_AUTORIZADAS'] >= dfti_limpo_qtv['QT_VAGAS_AUTORIZADAS'].quantile(0.600),
    dfti_limpo_qtv['QT_VAGAS_AUTORIZADAS'],
    dfti_limpo_qtv['QT_VAGAS_AUTORIZADAS'].quantile(0.001))
```

Figura 7 – Resultado limpeza QT_VAGAS_AUTORIZADAS.

```
[ ] dfti_limpo_qtv[['QT_VAGAS_AUTORIZADAS', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['QT_VAGAS_AUTORIZADAS'], ascending=False)
```

	QT_VAGAS_AUTORIZADAS		NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
740049	990.0		SISTEMAS DE INFORMAÇÃO	UNIVERSIDADE DE TAUBATÉ	Educação a Distância	São José dos Campos
265250	990.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE DE TAUBATÉ	Educação a Distância		Recife
252289	990.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE DE TAUBATÉ	Educação a Distância		Paraisópolis
104473	990.0		SISTEMAS DE INFORMAÇÃO	UNIVERSIDADE DE TAUBATÉ	Educação a Distância	Piquete
483939	990.0		SISTEMAS DE INFORMAÇÃO	UNIVERSIDADE DE TAUBATÉ	Educação a Distância	São Bento do Sapucaí


```
[ ] dfti_limpo_qtv[['QT_VAGAS_AUTORIZADAS', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['QT_VAGAS_AUTORIZADAS'], ascending=True)
```

	QT_VAGAS_AUTORIZADAS		NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
320254	32.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOL...	Educação Presencial		Venâncio Aires
413964	32.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE DO ESTADO DO AMAZONAS	Educação Presencial		Maués
732415	32.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE DO ESTADO DO AMAZONAS	Educação Presencial		Boca do Acre
145728	32.0		SISTEMAS DE INFORMAÇÃO	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOL...	Educação Presencial	Colatina
412958	32.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE DO ESTADO DO AMAZONAS	Educação Presencial		Humaitá

Agora vamos analisar possíveis ruídos e outliers em CARGA_HORARIA.

Figura 8 – Ruídos em ordem decrescente CARGA_HORARIA.

```
dfti[['CARGA_HORARIA', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['CARGA_HORARIA'], ascending=False).head()
```

	CARGA_HORARIA		NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
311124	8600	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	FACULDADE DE TECNOLOGIA PENTÁGONO	Educação Presencial		Santo André
517409	8000	ENGENHARIA DE COMPUTAÇÃO	CENTRO UNIVERSITÁRIO MAURÍCIO DE NASSAU DE BAR...	Educação Presencial		Barreiras
677004	7760	CIÊNCIA DA COMPUTAÇÃO	Centro Universitário Vértice	Educação Presencial		Matipó
62208	7560	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	CENTRO UNIVERSITÁRIO RITTER DOS REIS	Educação Presencial		Canoas
848008	7560	ENGENHARIA DE COMPUTAÇÃO	CENTRO UNIVERSITÁRIO RITTER DOS REIS	Educação Presencial		Porto Alegre

Figura 9 – Ruídos em ordem crescente CARGA_HORARIA.

```
dfti[['CARGA_HORARIA', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['CARGA_HORARIA'], ascending=True).head()
```

	CARGA_HORARIA		NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
453746	0	SISTEMAS PARA INTERNET	Faculdade de Tecnologia Senac Curitiba Portão	Educação Presencial		Curitiba
58041	0	BANCO DE DADOS	Faculdade de Tecnologia Senac Curitiba Portão	Educação Presencial		Curitiba
5134	0	CIÊNCIA DA COMPUTAÇÃO	UNIVERSIDADE ESTADUAL DO CEARÁ	Educação Presencial		Iguatu
48389	0	REDES DE COMPUTADORES	Faculdade de Tecnologia Senac Curitiba Portão	Educação Presencial		Curitiba
262468	1700	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	FACULDADE VISCONDE DE CAIRÚ	Educação a Distância		Salvador

Nos sites das respectivas faculdades acima não achei a carga horária. Existem ruídos nesses dados, conforme o site (UNOPAR, 2022) a carga horária de uma faculdade, em geral, tem duração mínima exigida pelo Ministério da Educação de 2.400 horas para bacharelados, para tecnólogos entre 1.600 a 2.000 horas, enquanto graduações mais longas como Medicina e Odontologia, podem chegar a 7.200 horas. Levando em conta que os cursos mais longos de TI são as Engenharias, e elas tem em média 3.800 horas, então vamos estabelecer esse padrão.

Figura 10 – Limpeza CARGA_HORARIA.

```
[ ] dfti_limpo_ch = dfti[
    (dfti['CARGA_HORARIA'] > dfti['CARGA_HORARIA'].quantile(0.001)) &
    (dfti['CARGA_HORARIA'] < dfti['CARGA_HORARIA'].quantile(0.985))]

# dfti_limpo_ch = dataframe ti em atividade limpo para carga horária

[ ] dfti_limpo_ch['CARGA_HORARIA'] = np.where(
    dfti_limpo_ch['CARGA_HORARIA'] >= dfti_limpo_ch['CARGA_HORARIA'].quantile(0.985),
    dfti_limpo_ch[(dfti_limpo_ch['CARGA_HORARIA'] > dfti_limpo_ch['CARGA_HORARIA'].qu
    dfti_limpo_ch['CARGA_HORARIA'])]
```

Figura 11 – Resultado limpeza CARGA_HORARIA.

```
[ ] dfti_limpo_ch[['CARGA_HORARIA', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['CARGA_HORARIA'], ascending=False).head()
```

	CARGA_HORARIA	NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
664823	3894.0	ENGENHARIA DE COMPUTAÇÃO	UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL	Educação Presencial	Campo Grande
895535	3885.0	ENGENHARIA DE COMPUTAÇÃO	UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	Educação Presencial	Natal
174548	3885.0	ENGENHARIA DE TELECOMUNICAÇÕES	UNIVERSIDADE FEDERAL DE SANTA MARIA	Educação Presencial	Santa Maria
849511	3885.0	ENGENHARIA DE COMPUTAÇÃO	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JAN...	Educação Presencial	Rio de Janeiro
197260	3884.0	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNO...	Educação Presencial	Itumbiara


```
dfti_limpo_ch[['CARGA_HORARIA', 'NOME_CURSO', 'NOME_IES', 'MODALIDADE', 'MUNICIPIO']].sort_values(by=['CARGA_HORARIA'], ascending=True).head()
```

	CARGA_HORARIA	NOME_CURSO	NOME_IES	MODALIDADE	MUNICIPIO
233590	2001.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE PRESBITERIANA MACKENZIE	Educação a Distância	Bauru
338841	2001.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE PRESBITERIANA MACKENZIE	Educação a Distância	Ribeirão Preto
86236	2001.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE PRESBITERIANA MACKENZIE	Educação a Distância	São Luís
573130	2001.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE PRESBITERIANA MACKENZIE	Educação a Distância	São Paulo
311331	2001.0	ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	UNIVERSIDADE PRESBITERIANA MACKENZIE	Educação a Distância	Aracaju

2.2.4 Transformação

Verificando se os tipos de dados estão adequados para representar cada variável do dataset, vemos que estão adequados para o que sua respectiva coluna representa.

Figura 12 – Tipos de dados de cada coluna.

```
[ ] df.dtypes
```

NOME_IES	object
CATEGORIA ADMINISTRATIVA	object
ORGANIZACAO_ACADEMICA	object
NOME_CURSO	object
GRAU	object
MODALIDADE	object
SITUACAO_CURSO	object
QT_VAGAS_AUTORIZADAS	int64
CARGA_HORARIA	int64
MUNICIPIO	object
UF	object
REGIAO	object
dtype:	object

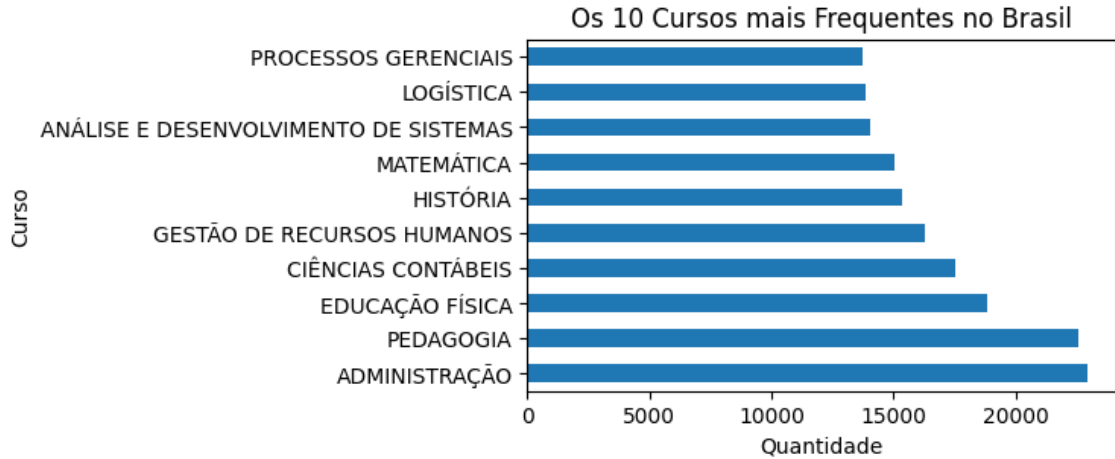
Sobre técnicas de transformação em si como discretização, binarização, codificação, normalização e etc... até o momento não vi como necessário. A depender do algoritmo de Machine Learning que irei aplicar no futuro, é muito provável que farei uso de alguma técnica de transformação.

3 ANÁLISE E RESULTADOS

3.1 Análise Exploratória de Dados (EDA)

Antes de tudo analisei os 10 cursos mais frequentes no Brasil levando em conta que o dado inclui todos os cursos tanto **em atividade** como os **extintos**.

Figura 13 – Os 10 Cursos mais Frequentes no Brasil.



Foi criado outros dataframes, `df_ea` (apenas com cursos em atividade) e `df_earn` (cursos do Rio Grande do Norte em atividade). Uma vez que o dataframe geral (`df`) possui 902.068 linhas, com isso o `df_ea` tem uma redução de 49.740 linhas.

3.2 Distribuição dos cursos por região

Esta seção pretende:

- Explorar a distribuição geográfica dos cursos de graduação no Brasil por estado.
- Identificar quais estados possuem mais cursos e quais têm menos.
- Analisar se existe alguma correlação entre a disponibilidade de cursos e indicadores socioeconômicos das regiões.

Primeiro vamos analisar a quantidade de cursos em atividade por região.

Tabela 2 – Distribuição de Cursos por Região

Região	Quantidade
SUDESTE	324727
NORDESTE	194187
SUL	178700
NORTE	79383
CENTRO-OESTE	75331

Fonte: elaborado pelo autor (2024)

Figura 14 – Regiões com mais Cursos Em Atividade.

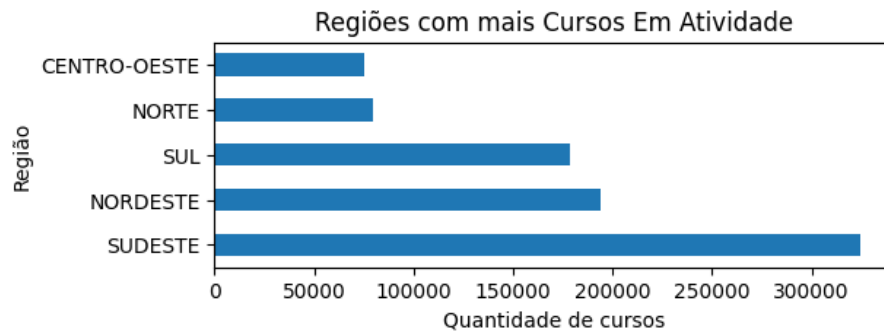
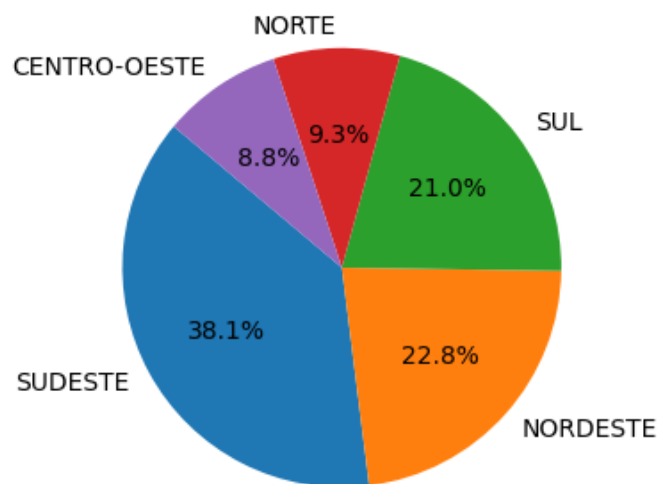


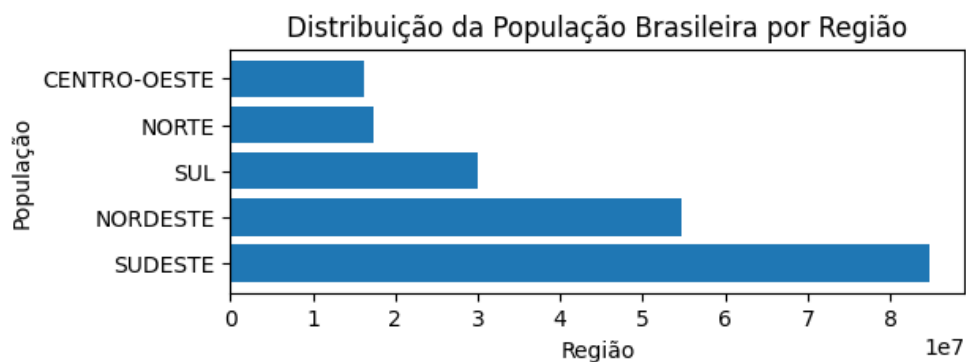
Figura 15 – Regiões com mais Cursos Em Atividade.

Distribuição dos Cursos Em Atividade por Região



Os dados abaixo são do IBGE e foram retirados do site (EDUCACAO, 2022) e fazem parte do censo demográfico de 2022.

Figura 16 – Regiões com mais Cursos Em Atividade.



Percebemos que as regiões que possuem mais habitantes (Figura 16) são as que tem mais cursos em atividade (Figura 14).

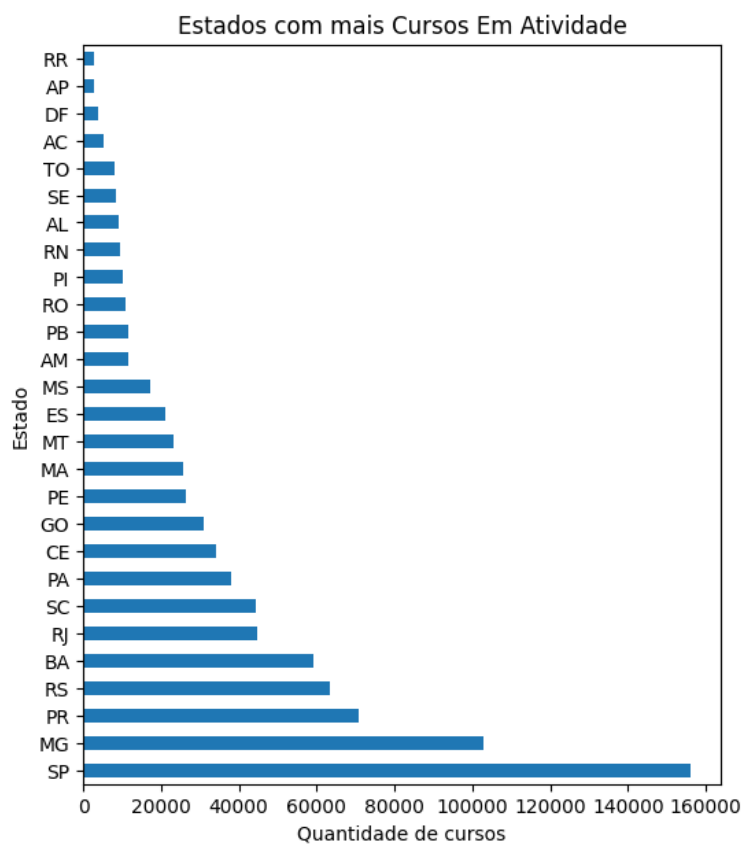
Por outro lado, não existe uma correlação entre o PIB das regiões e quantidade de cursos.

Figura 17 – Participação das regiões no PIB.



A seguir vamos analisar a quantidade de cursos por estado.

Figura 18 – Estados com mais Cursos Em Atividade.



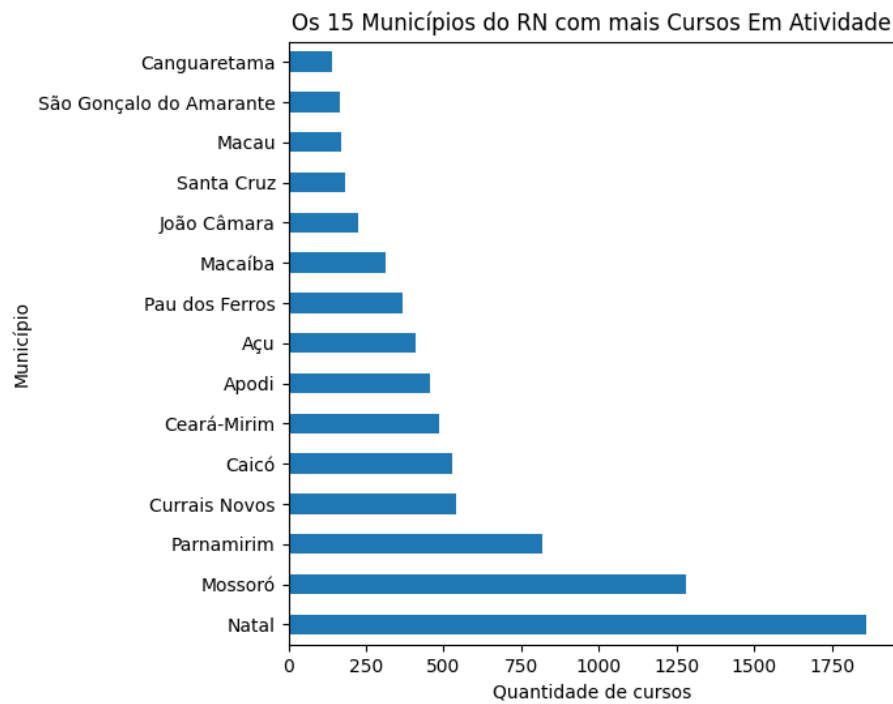
Agora uma análise da quantidade de cursos nos municípios do Rio Grande do Norte.

Tabela 3 – Distribuição de Cursos por Município

Município	Quantidade
Natal	1863
Mossoró	1280
Parnamirim	817
Currais Novos	542
Caicó	528
Ceará-Mirim	485
Apodi	456
Açu	411
Pau dos Ferros	367
Macaíba	314
João Câmara	225
Santa Cruz	184
Macau	172
São Gonçalo do Amarante	165
Canguaretama	142

Fonte: elaborado pelo autor (2024)

Figura 19 – Os 15 Municípios do RN com mais Cursos Em Atividade.



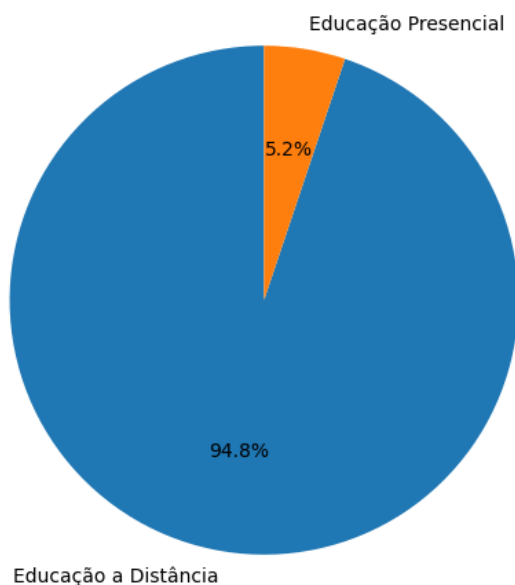
3.3 Perfil dos cursos por modalidade de ensino

Esta seção pretende:

- Comparar a quantidade e a distribuição dos cursos de graduação **Em Atividade**, **Presenciais** e **EaD**.
- Analisar se existe diferenças no grau dos cursos e na carga horária de acordo com a modalidade.

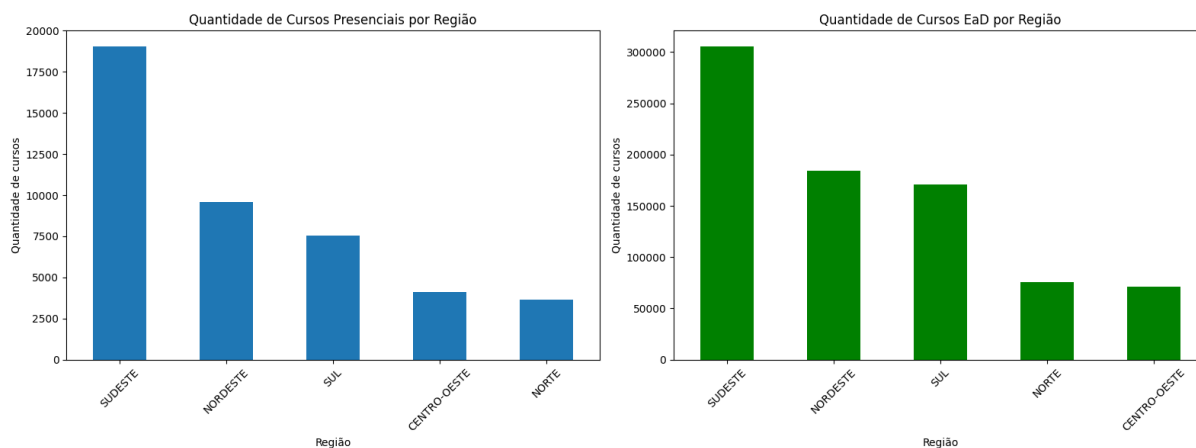
Figura 20 – Distribuição dos Cursos Em Atividade por Modalidade de Ensino.

Distribuição dos Cursos Em Atividade por Modalidade de Ensino



Como mostrado em (Figura 20), existem mais cursos EaD do que presenciais.

Figura 21 – Quantidade de Cursos EaD e Presencial por Região.



Os gráficos tem uma tendência bem semelhante proporcionalmente, exceto que o Centro-Oeste possui mais cursos presenciais que o Norte, e em compensação, o Norte possui mais cursos EaD.

Agora vamos ver como se comportam essa análise para cada Unidade Federativa e para os Municípios do Rio Grande do Norte.

Figura 22 – Quantidade de Cursos EaD e Presencial por Estado.

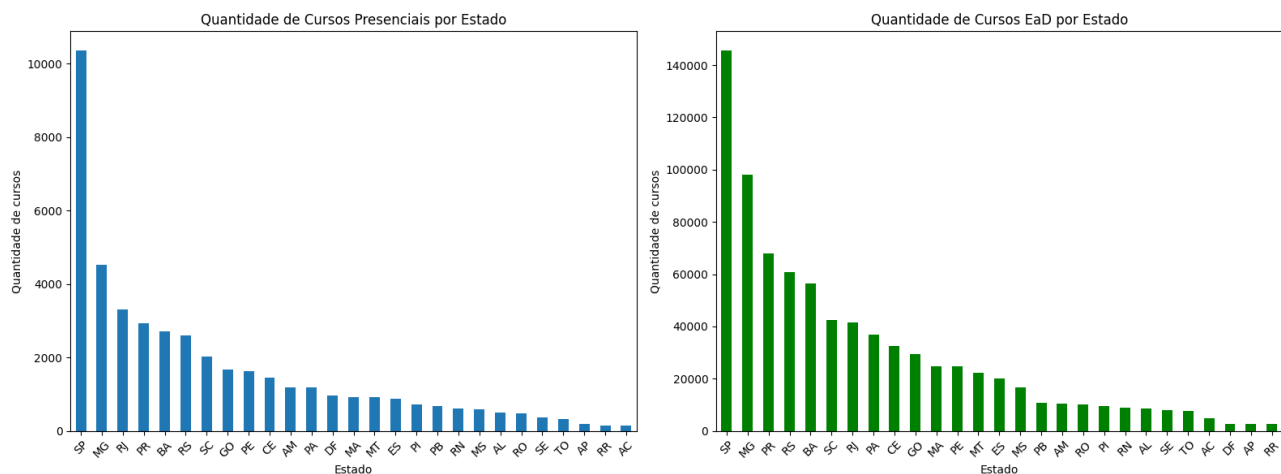
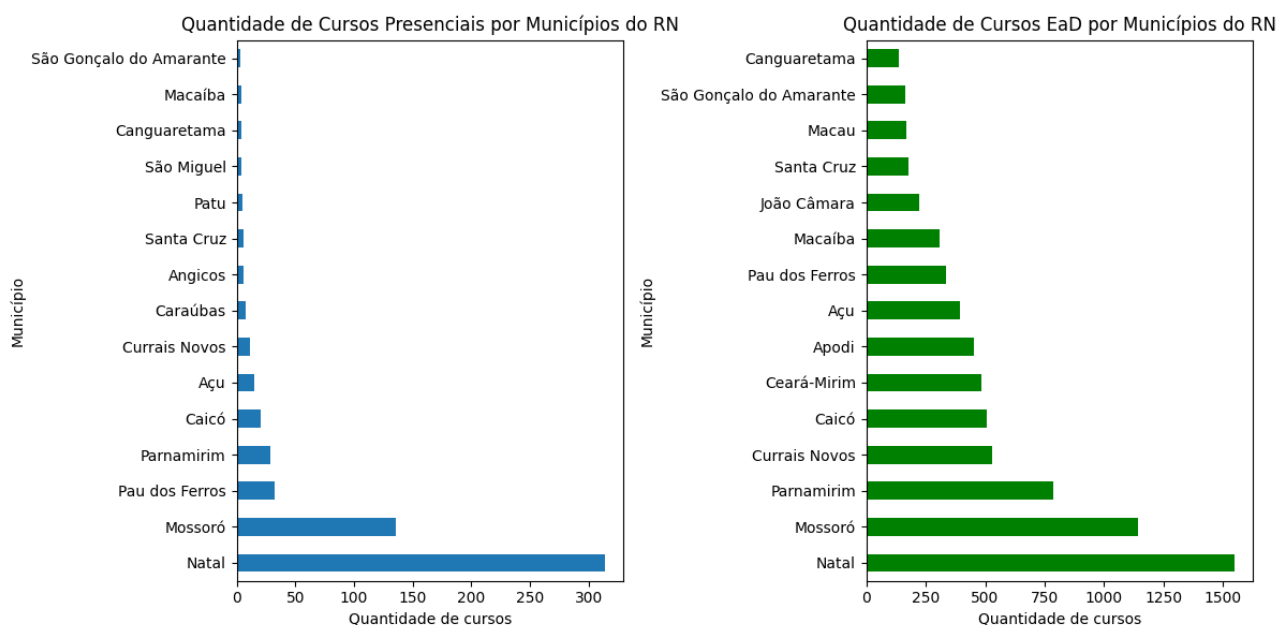


Figura 23 – Quantidade de Cursos Presencial e EaD por Municípios do RN.



Agora vamos analisar a quantidade de cursos por modalidade em seus respectivos Graus.

Tabela 4 – Distribuição de Cursos por Grau Presencial

Grau	Quantidade
Bacharelado	27114
Tecnológico	8443
Licenciatura	7995
Área Básica de Ingresso (ABI)	335
Sequencial	81

Fonte: elaborado pelo autor (2024)

Tabela 5 – Distribuição de Cursos por Grau EaD

Grau	Quantidade
Tecnológico	437131
Bacharelado	215848
Licenciatura	155343
Área Básica de Ingresso (ABI)	36
Sequencial	2

Fonte: elaborado pelo autor (2024)

Figura 24 – Distribuição de Cursos Presenciais por Grau.

Distribuição de Cursos Presenciais por Grau

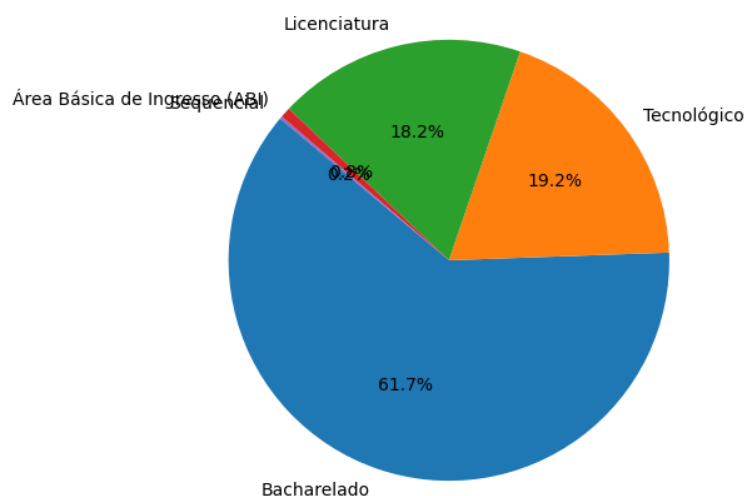
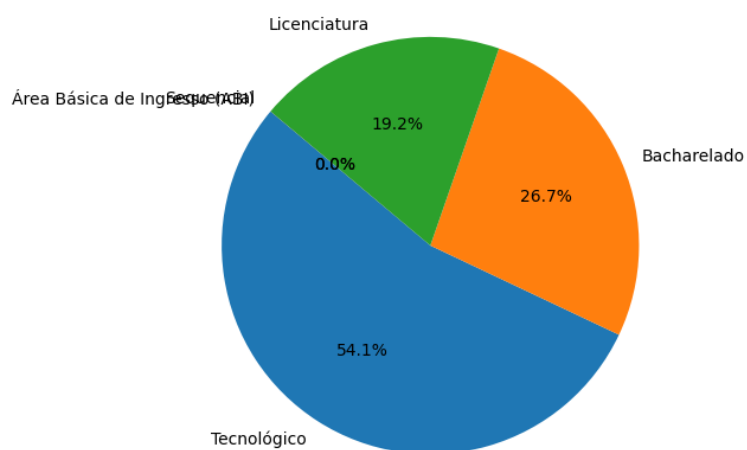


Figura 25 – Distribuição de Cursos EaD por Grau.

Distribuição de Cursos EaD por Grau



Agora vamos construir a mesma análise para o Rio Grande do Norte.

Tabela 6 – Distribuição de Cursos por Grau no RN

Grau	Quantidade
Bacharelado	385
Licenciatura	124
Tecnológico	103
Área Básica de Ingresso (ABI)	2

Fonte: elaborado pelo autor (2024)

Figura 26 – Distribuição de Cursos Presenciais por Grau no RN.

Distribuição de Cursos Presenciais por Grau no RN

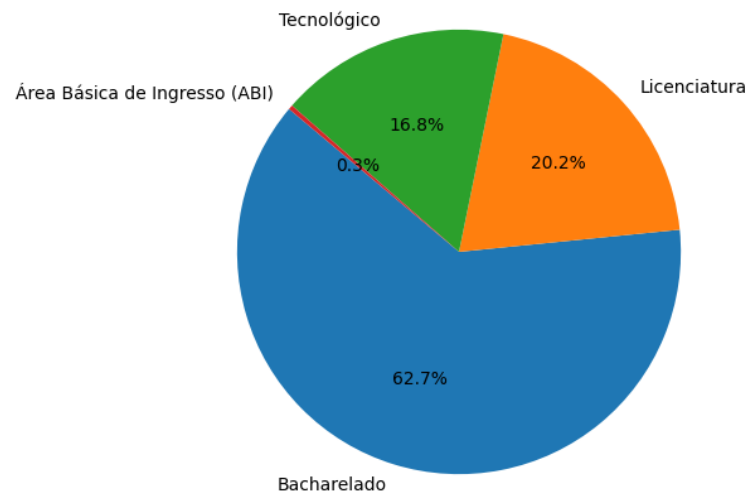
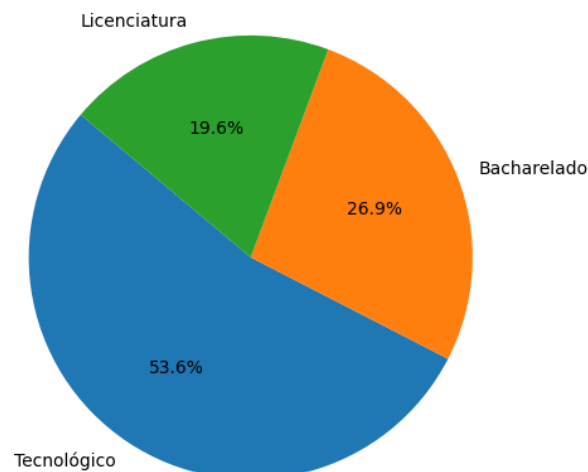


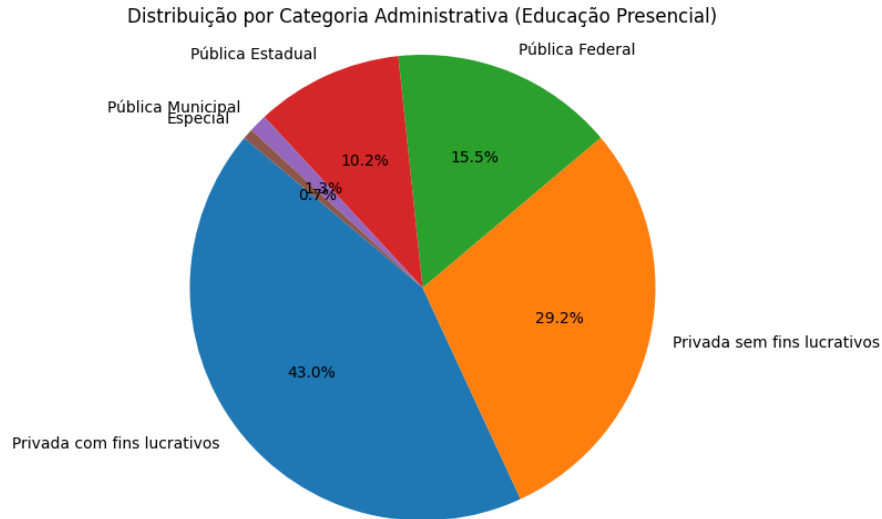
Figura 27 – Distribuição de Cursos EaD por Grau no RN.

Distribuição de Cursos EaD por Grau no RN



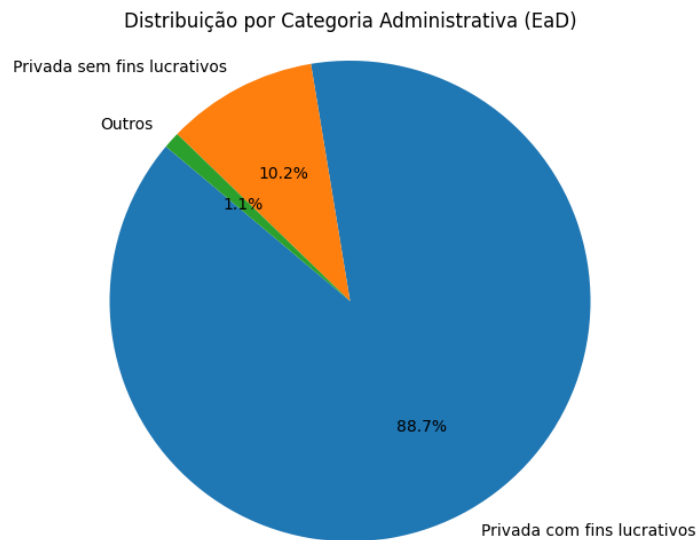
Seguindo a mesma ideia, vamos analisar as categorias administrativas para cada modalidade e quais são mais frequentes que outras.

Figura 28 – Distribuição por Categoria Administrativa (Educação Presencial).



Como visto em (Figura 28) as faculdade privadas continuam detendo a maior quantidade de cursos presenciais.

Figura 29 – Distribuição por Categoria Administrativa (EaD).



Assim como, a maioria dos cursos EaD são de faculdades privadas (Figura 29).

O que seria a Categoria Administrativa Especial? Enquadra-se nessa categoria a instituição de educação superior criada por lei, estadual ou municipal, e existente na data da promulgação da Constituição Federal de 1988, que não seja total ou preponderantemente mantida com recursos públicos, portanto, não gratuita (INEP, 2016).

Agora vamos analisar quais cursos mais tem em cada modalidade.

Tabela 7 – Modalidade Presencial

Nome do Curso	Quantidade
ADMINISTRAÇÃO	2275
DIREITO	1965
PEDAGOGIA	1884
EDUCAÇÃO FÍSICA	1650
CIÊNCIAS CONTÁBEIS	1494
ENFERMAGEM	1376
PSICOLOGIA	1260
ENGENHARIA CIVIL	1181
FISIOTERAPIA	983
CIÊNCIAS BIOLÓGICAS	896
ENGENHARIA DE PRODUÇÃO	879
NUTRIÇÃO	812
FARMÁCIA	793
GESTÃO DE RECURSOS HUMANOS	783
ARQUITETURA E URBANISMO	752
ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	709
BIOMEDICINA	696
MATEMÁTICA	685
ODONTOLOGIA	659
ENGENHARIA MECÂNICA	638
ENGENHARIA ELÉTRICA	608
LOGÍSTICA	558
MEDICINA VETERINÁRIA	533
HISTÓRIA	507
SISTEMAS DE INFORMAÇÃO	475
SERVIÇO SOCIAL	459
ESTÉTICA E COSMÉTICA	436
CIÊNCIA DA COMPUTAÇÃO	434
QUÍMICA	432
GEOGRAFIA	427

Fonte: elaborado pelo autor (2024)

Tabela 8 – Modalidade EaD

Nome do Curso	Quantidade
ADMINISTRAÇÃO	17426
PEDAGOGIA	17303
EDUCAÇÃO FÍSICA	16363
CIÊNCIAS CONTÁBEIS	15092
GESTÃO DE RECURSOS HUMANOS	14884
HISTÓRIA	14035
MATEMÁTICA	13346
ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	13087
GESTÃO PÚBLICA	12943
LOGÍSTICA	12861
PROCESSOS GERENCIAIS	12851
GESTÃO FINANCEIRA	12762
GEOGRAFIA	12134
GESTÃO COMERCIAL	12105
CIÊNCIAS BIOLÓGICAS	11759
MARKETING	11510
GESTÃO AMBIENTAL	11418
SERVIÇO SOCIAL	11281
FILOSOFIA	10167
ENGENHARIA DE PRODUÇÃO	9992
GESTÃO DA TECNOLOGIA DA INFORMAÇÃO	9868
GESTÃO HOSPITALAR	9589
ARTES VISUAIS	8917
COMÉRCIO EXTERIOR	8663
GESTÃO DA QUALIDADE	8341
CIÊNCIAS ECONÔMICAS	8253
QUÍMICA	7295
DESIGN DE INTERIORES	7248
SEGURANÇA PÚBLICA	7193
NEGÓCIOS IMOBILIÁRIOS	7181

Fonte: elaborado pelo autor (2024)

3.4 Situação dos cursos

Esta seção pretende:

- Avaliar a proporção de cursos ativos e inativos.
- Analisar se existe alguma relação entre a situação dos cursos e a categoria da instituição de ensino superior e a modalidade.

Existem 3 categorias para a coluna SITUAÇÃO_CURSO:

- Em atividade
- Em extinção

Tabela 9 – Distribuição de Cursos por Situação

Situação do Curso	Quantidade
Em atividade	852328
Extinto	39479
Em extinção	10261

Fonte: elaborado pelo autor (2024)

Figura 30 – Situação dos Cursos.

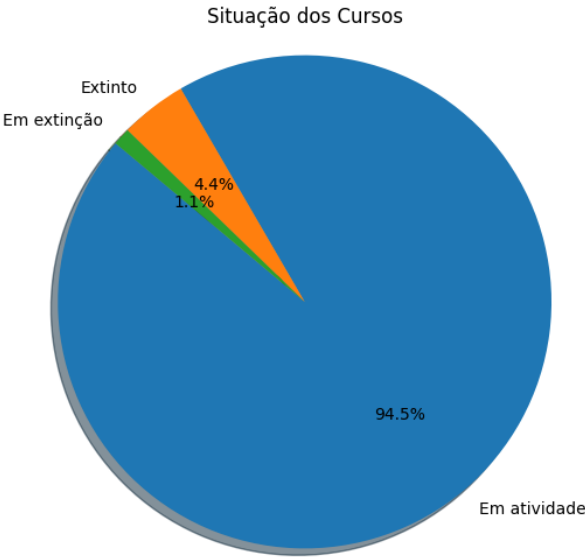
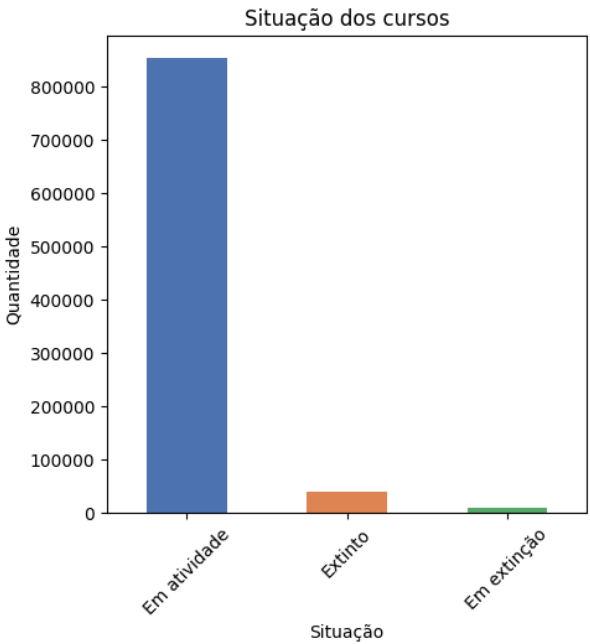


Figura 31 – Situação dos Cursos Gráfico de Barras.



Como visto, a maioria dos cursos criados ainda estão em atividade.

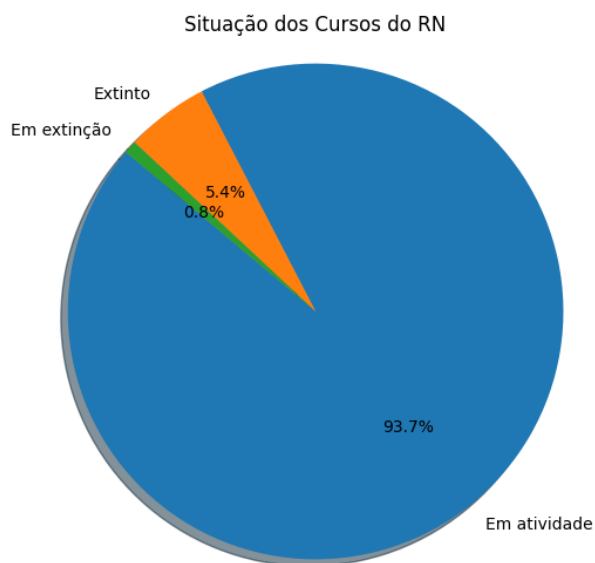
Agora vamos fazer essas análises para os cursos no Rio Grande do Norte.

Tabela 10 – Distribuição de Cursos por Situação RN

Situação do Curso	Quantidade
Em atividade	9499
Extinto	551
Em extinção	85

Fonte: elaborado pelo autor (2024)

Figura 32 – Situação dos Cursos do RN.



Agora uma análise da situação pela modalidade.

Tabela 11 – Distribuição de Cursos por Modalidade e Situação

Situação do Curso	Educação Presencial	Educação a Distância
Em atividade	43968	808360
Em extinção	6306	3955
Extinto	22363	17116

Fonte: elaborado pelo autor (2024)

Figura 33 – Distribuição da Situação dos Cursos de Educação Presencial.

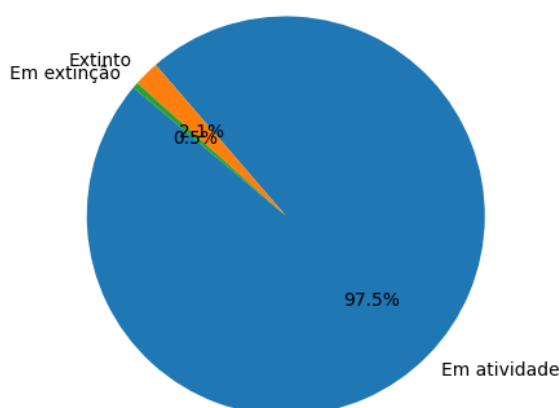
Distribuição da Situação dos Cursos de Educação Presencial



Um pouco mais da metade dos cursos presenciais estão em atividade (Figura 33).

Figura 34 – Distribuição da Situação dos Cursos de Educação a Distância.

Distribuição da Situação dos Cursos de Educação a Distância



A maioria dos cursos EaD estão Em Atividade (Figura 34).

A seguir uma análise das situações dos cursos por categoria administrativa.

Tabela 12 – Distribuição de Cursos por Categoria Administrativa e Situação

Situação do Curso	Especial	Privada com fins lucrativos	Privada sem fins lucrativos	Pública Estadual	Pública Federal	Pública Municipal
Em atividade	625	736063	95055	6455	10758	3372
Em extinção	24	3721	4979	771	716	50
Extinto	134	21248	11803	2754	3328	212

Fonte: elaborado pelo autor (2024)

Figura 35 – Distribuição de Cursos por Categoria Administrativa e Situação (Em atividade)

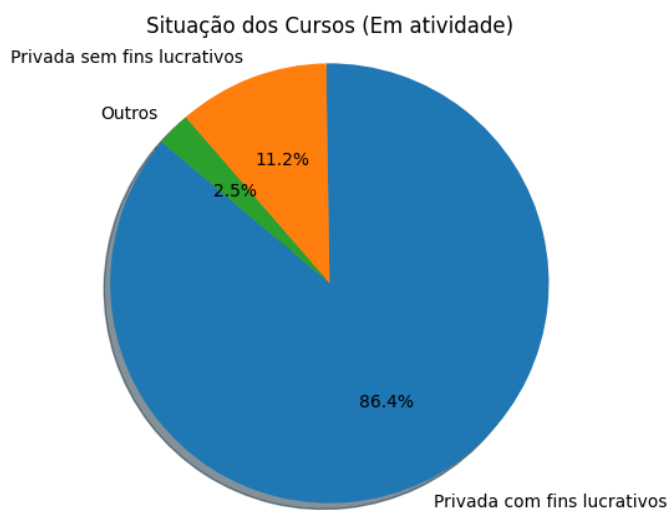


Figura 36 – Distribuição de Cursos por Categoria Administrativa e Situação (Em extinção)

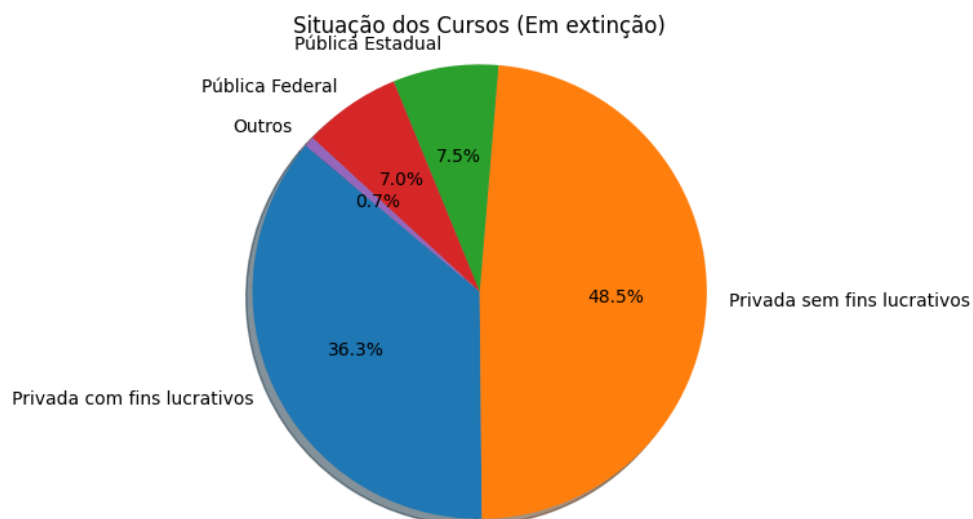
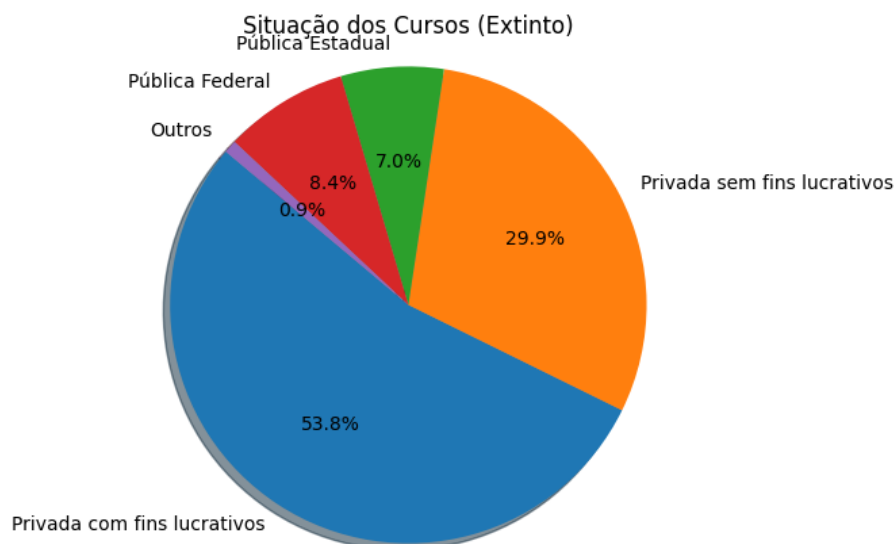


Figura 37 – Distribuição de Cursos por Categoria Administrativa e Situação (Extinto)



Algo interessante de ser observado é a alta porcentagem de cursos em extinção (Figura 36) e extintos (Figuras 37) para a categoria administrativa Privada sem fins lucrativos.

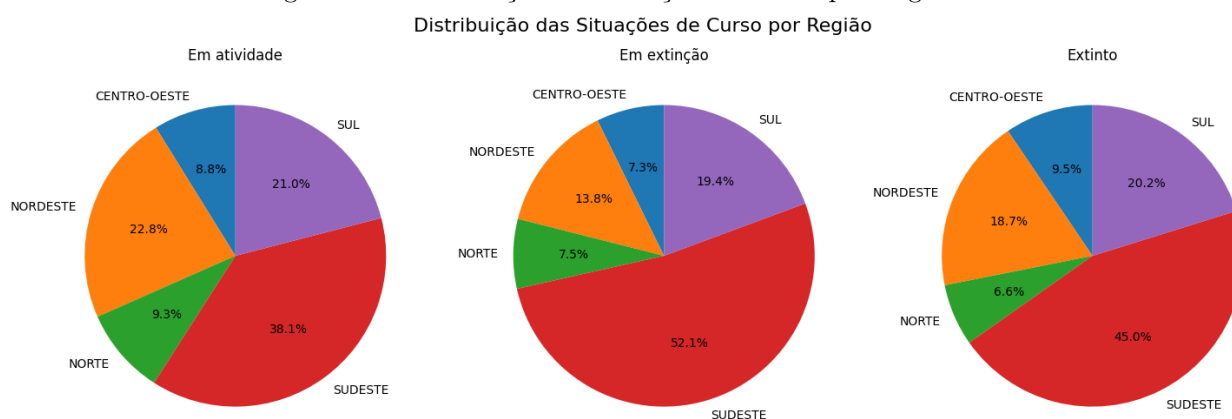
Importante lembrar que as regiões Centro-Oeste e Norte são as que possuem menos cursos. Isso ajudará a ter uma visão menos enviesada dessa análise seguinte.

Tabela 13 – Distribuição de Cursos por Região e Situação

Situação do Curso	CENTRO-OESTE	NORDESTE	NORTE	SUDESTE	SUL
Em atividade	75331	194187	79383	324727	178700
Em extinção	745	1416	765	5349	1986
Extinto	3742	7365	2622	17779	7971

Fonte: elaborado pelo autor (2024)

Figura 38 – Distribuição das Situações de Curso por Região



3.5 Cursos de tecnologia

Nesta seção, decidi seguir a mesma linha das análises anteriores, porém, com um filtro para os principais cursos de Tecnologia da Informação no Brasil.

Uma das minhas primeiras curiosidades foi procurar o curso Bacharelado em Tecnologia da Informação na UFRN. E como visto (Figura 39), ele foi preenchido corretamente.

Figura 39 – BTI UFRN

NOME_IES	CATEGORIA_ADMINISTRATIVA	ORGANIZACAO_ACADEMICA	NOME_CURSO	GRAU	MODALIDADE	SITUACAO_CURSO	QT_VAGAS_AUTORIZADAS	CARGA_HORARIA	MUNICIPIO	UF	REGIAO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	Pública Federal	Universidade	TECNOLOGIA DA INFORMACAO	Bacharelado	Educacao Presencial	Em atividade	300	2600	Natal	RN	NORDESTE

Existem 14 cursos cujo nome é Tecnologia da Informação (Tabela 14), desses, o da UFRN é o único curso de grau Bacharelado e modalidade Presencial.

Tabela 14 – Quantidade de Cursos na Área de Tecnologia da Informação

Nome do Curso	Quantidade
ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	13796
GESTÃO DA TECNOLOGIA DA INFORMACÃO	10096
ENGENHARIA DE SOFTWARE	6635
REDES DE COMPUTADORES	6399
JOGOS DIGITAIS	6385
SISTEMAS PARA INTERNET	6324
SISTEMAS DE INFORMACÃO	5140
CIÊNCIA DA COMPUTAÇÃO	4136
BANCO DE DADOS	3164
ENGENHARIA DE COMPUTAÇÃO	2625
ENGENHARIA DE CONTROLE E AUTOMACÃO	457
ENGENHARIA MECATRÔNICA	433
ENGENHARIA DE TELECOMUNICAÇÕES	25
TECNOLOGIA DA INFORMACÃO	14

Fonte: elaborado pelo autor (2024)

Figura 40 – Distribuição dos cursos de TI no Brasil

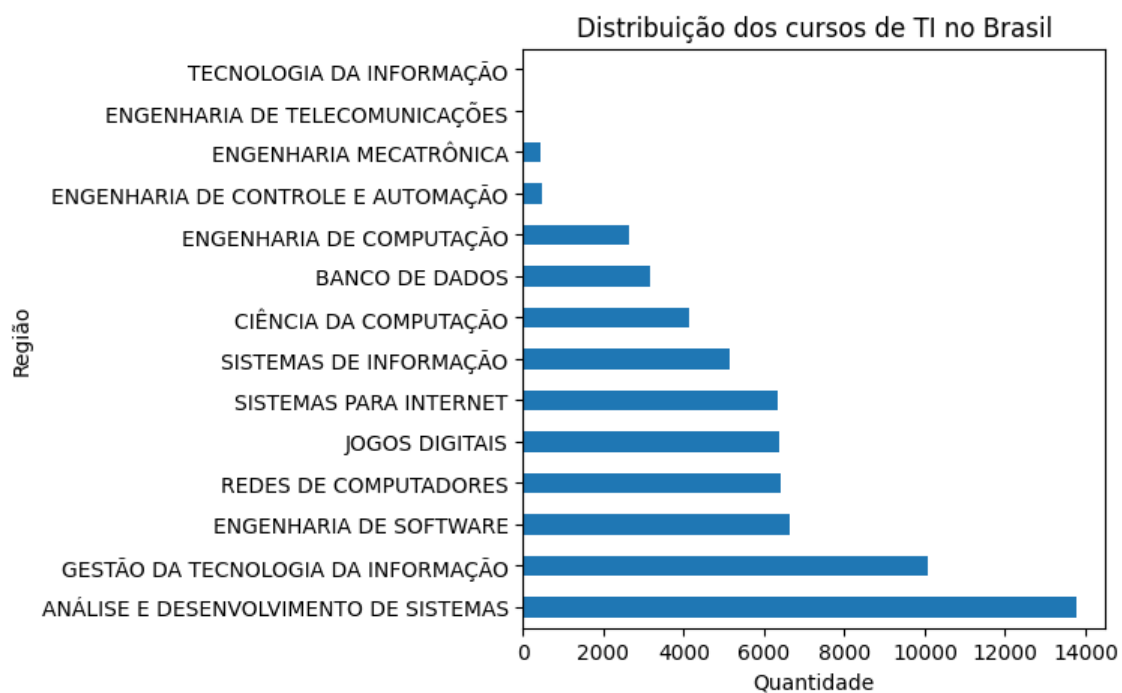


Figura 41 – Distribuição dos Cursos de TI no Brasil por Região

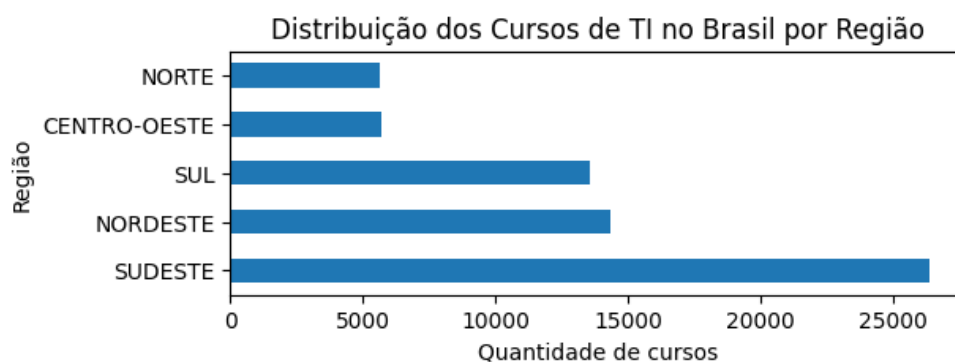


Figura 42 – Distribuição dos Cursos de TI no Brasil por Estado

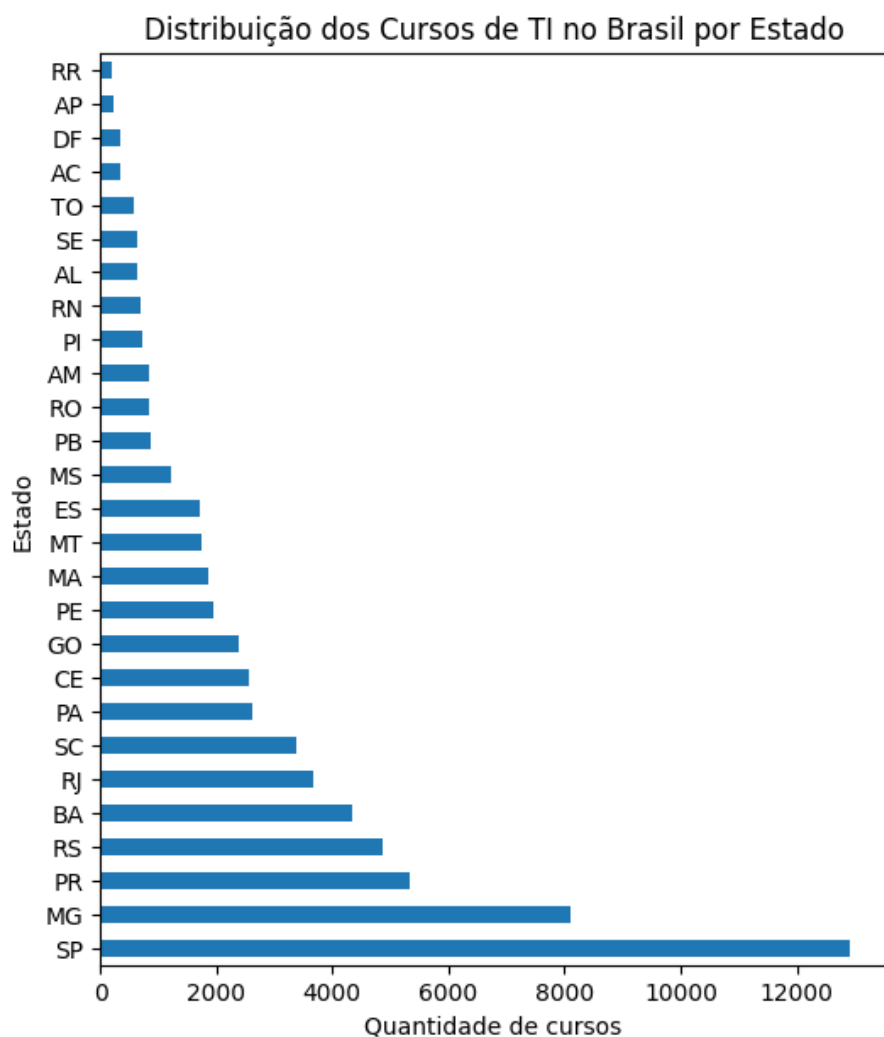
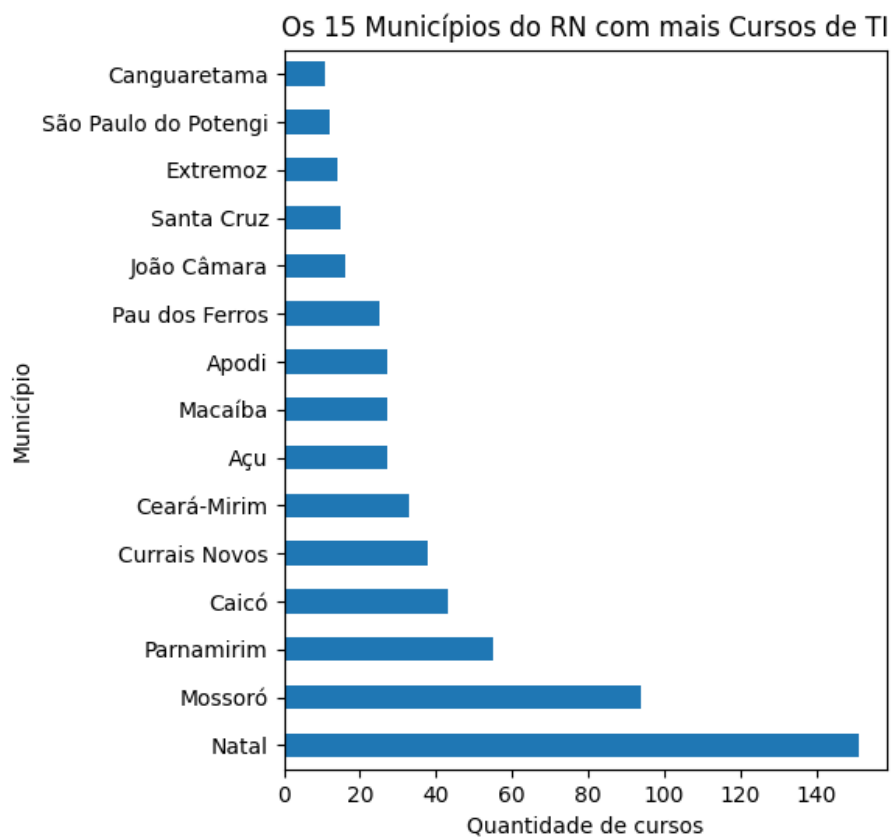


Tabela 15 – Quantidade de Cursos de TI no Rio Grande do Norte

Nome do Curso	Quantidade
ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	155
GESTÃO DA TECNOLOGIA DA INFORMAÇÃO	98
REDES DE COMPUTADORES	75
ENGENHARIA DE SOFTWARE	73
JOGOS DIGITAIS	66
SISTEMAS PARA INTERNET	62
SISTEMAS DE INFORMAÇÃO	50
CIÊNCIA DA COMPUTAÇÃO	44
BANCO DE DADOS	34
ENGENHARIA DE COMPUTAÇÃO	33
ENGENHARIA DE CONTROLE E AUTOMAÇÃO	5
ENGENHARIA MECATRÔNICA	2
ENGENHARIA DE TELECOMUNICAÇÕES	1
TECNOLOGIA DA INFORMAÇÃO	1

Fonte: elaborado pelo autor (2024)

Figura 43 – Os 15 Municípios do RN com mais Cursos de TI



Dando continuidade, veremos a distribuição dos cursos por modalidade de ensino.

Tabela 16 – Distribuição de Cursos de TI por Modalidade

Modalidade	Quantidade
Educação a Distância	62483
Educação Presencial	3146

Fonte: elaborado pelo autor (2024)

Figura 44 – Distribuição dos Cursos de TI Em Atividade por Modalidade de Ensino

Distribuição dos Cursos de TI Em Atividade por Modalidade de Ensino

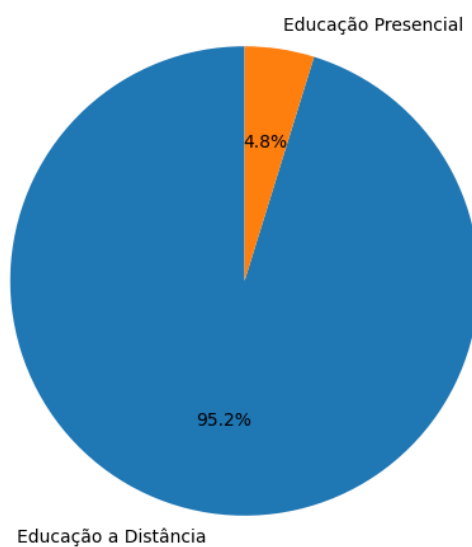


Figura 45 – Distribuição dos Cursos de TI Modalidade Presencial e a Distância

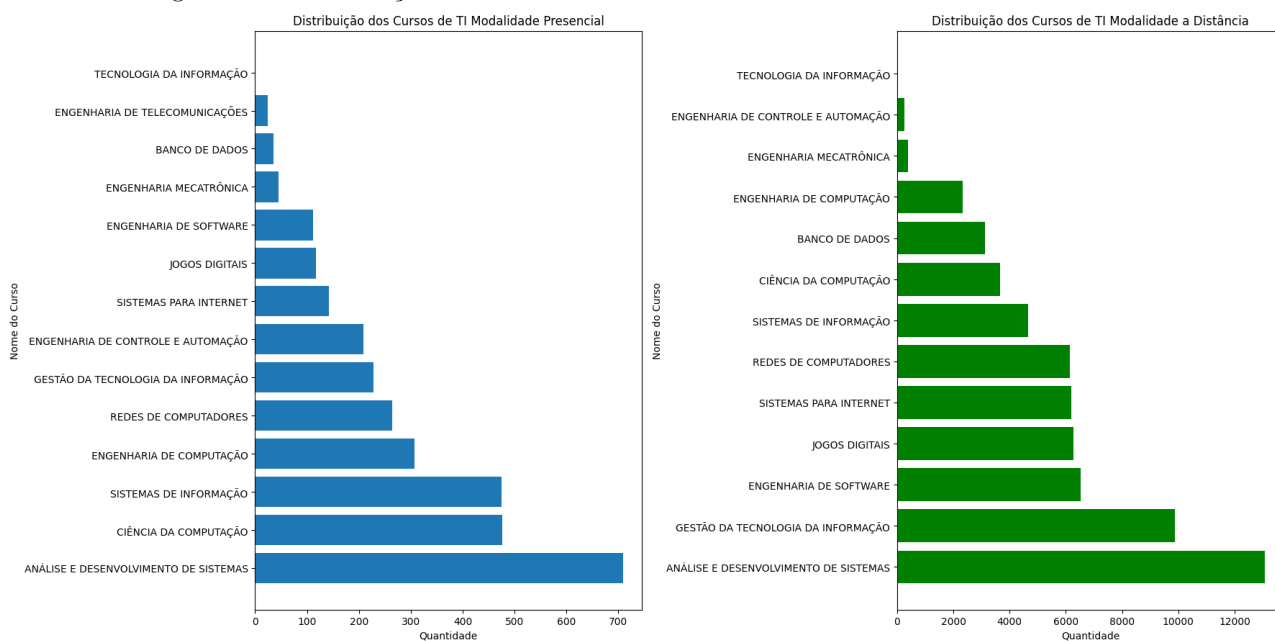


Figura 46 – Situação dos Cursos de TI

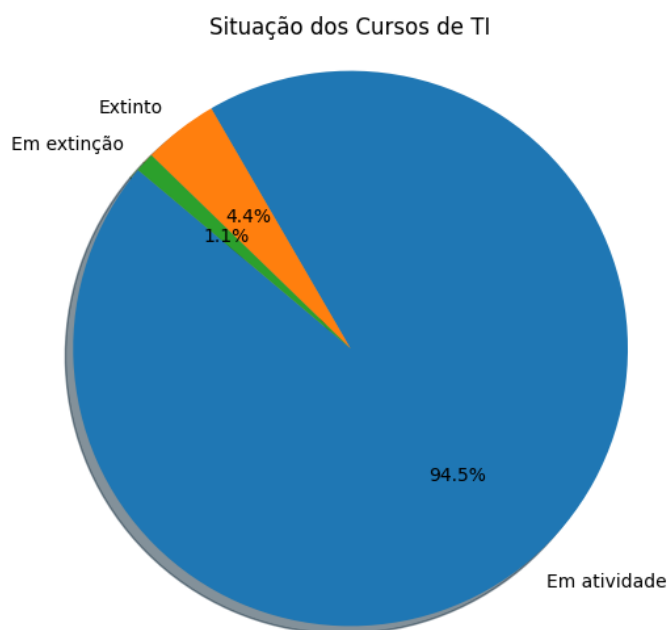


Figura 47 – Cursos TI Extintos e Em Extinção

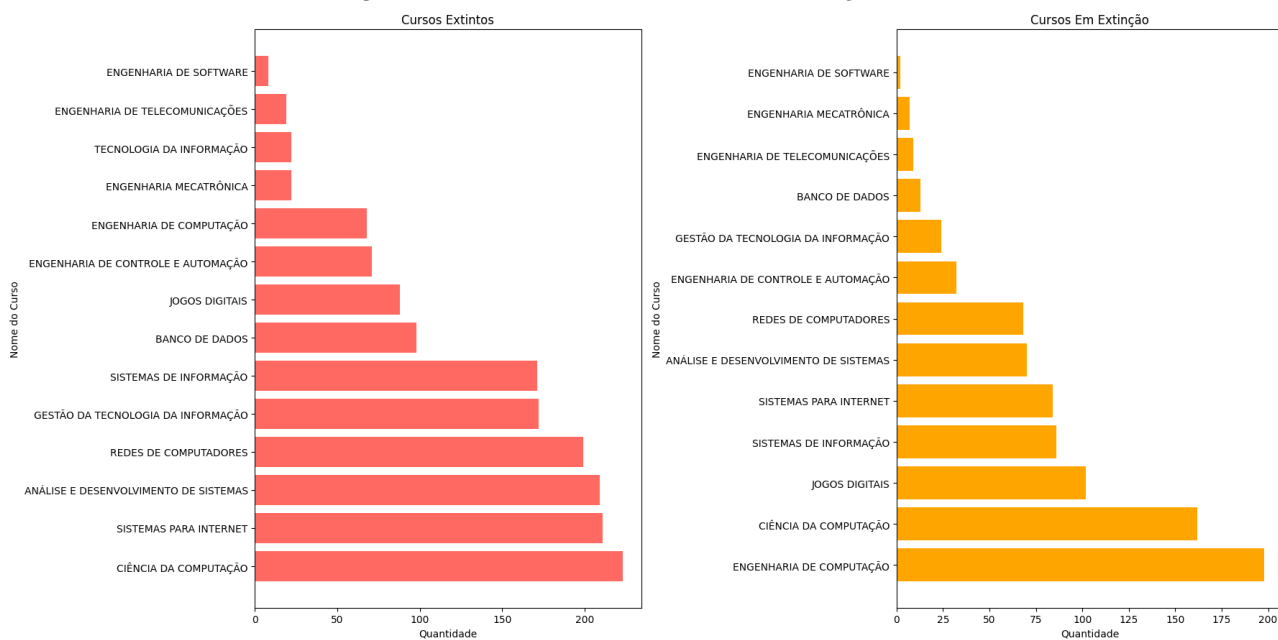
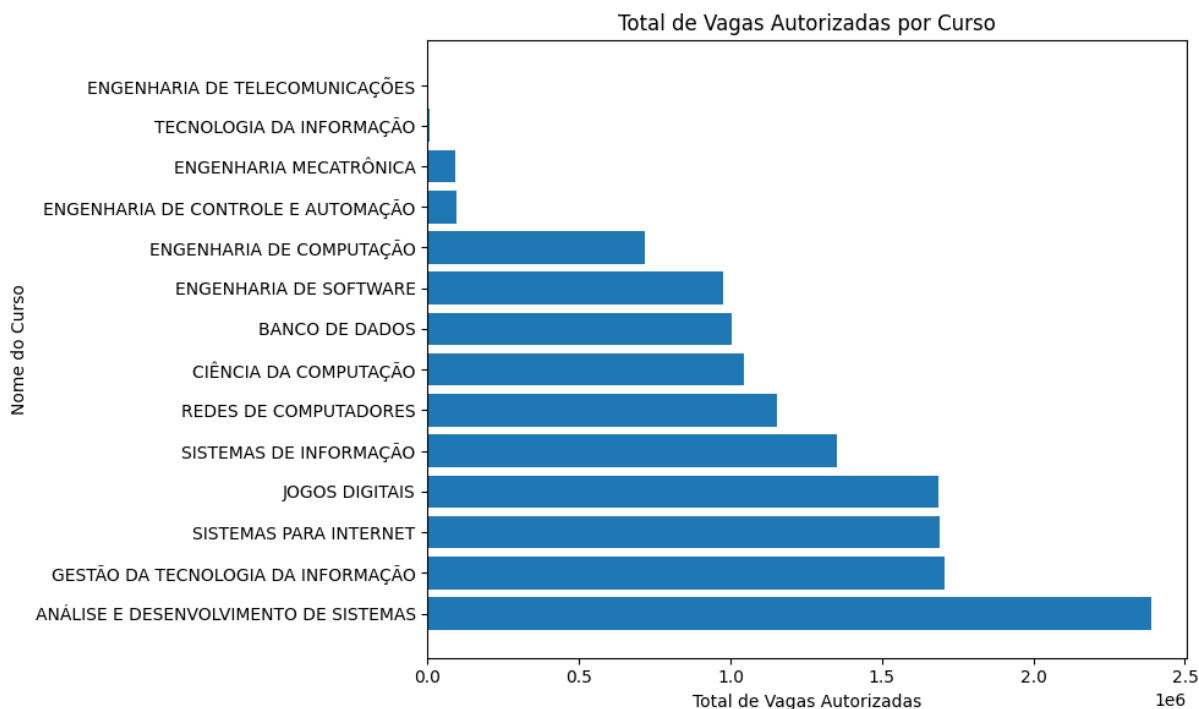


Figura 48 – Total de Vagas Autorizadas por Curso



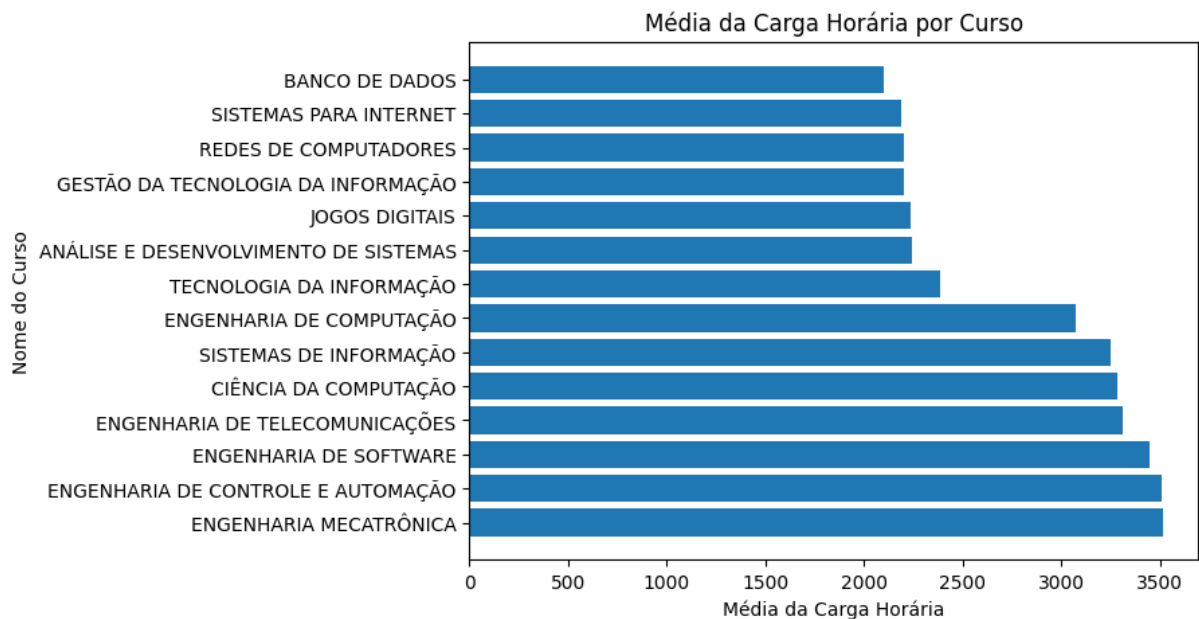
Agora vamos analisar a distribuição de carga horária por curso.

Tabela 17 – Média de Carga Horária por Curso

Nome do Curso	Carga Horária (horas)
ENGENHARIA MECATRÔNICA	3515
ENGENHARIA DE CONTROLE E AUTOMAÇÃO	3506
ENGENHARIA DE SOFTWARE	3449
ENGENHARIA DE TELECOMUNICAÇÕES	3310
CIÊNCIA DA COMPUTAÇÃO	3280
SISTEMAS DE INFORMAÇÃO	3251
ENGENHARIA DE COMPUTAÇÃO	3071
TECNOLOGIA DA INFORMAÇÃO	2382
ANÁLISE E DESENVOLVIMENTO DE SISTEMAS	2243
JOGOS DIGITAIS	2233
GESTÃO DA TECNOLOGIA DA INFORMAÇÃO	2205
REDES DE COMPUTADORES	2204
SISTEMAS PARA INTERNET	2190
BANCO DE DADOS	2101

Fonte: elaborado pelo autor (2024)

Figura 49 – Média da Carga Horária por Curso



4 DISCUSSÃO

4.1 Interpretação dos resultados

O objetivo central do projeto foi concretizado, onde eu queria fornecer uma visão ampla sobre a estrutura e a situação dos cursos no Brasil. Muita das minhas hipóteses iniciais foram confirmadas, como por exemplo: de que existem mais cursos EaD do que Presenciais (Figura 20). Outra é que existem mais cursos EaD em Instituições de Ensino Superior Privada do que Pública (Figura 29). Já para os cursos de TI eu queria verificar a hipótese sobre o curso de Análise e Desenvolvimento de Sistemas ser o mais recorrente no Brasil, e isso se confirmou (Figura 40).

4.2 Limitações

Este estudo se limita a verificar a distribuição geral dos cursos. Ele poderia ser mais proveitoso se incluísse dados temporais, como o ano em que cada curso foi inaugurado em cada universidade, com isso poderíamos acompanhar as tendências.

4.3 Implicações

Os resultados deste estudo, que analisam a distribuição geral dos cursos universitários, têm várias implicações práticas e teóricas. Praticamente, as universidades podem usar essas informações para alinhar melhor suas ofertas de cursos com as demandas do mercado de trabalho, especialmente em áreas emergentes como Tecnologia e Negócios. Políticas educacionais podem ser formuladas para incentivar o desenvolvimento de cursos em áreas sub-representadas, melhorando a diversidade e a adaptabilidade das instituições de ensino. Teoricamente, este estudo amplia o entendimento sobre a oferta educacional, sugerindo que fatores institucionais e econômicos desempenham um grande papel na criação de novos cursos. Isso abre caminho para novas pesquisas que explorem a relação

entre políticas educacionais e a criação de cursos, bem como a influência de tendências socioeconômicas na educação superior.

5 CONCLUSÃO

5.1 Resumo dos achados

Este relatório apresentou uma análise detalhada dos cursos de graduação no Brasil, utilizando dados extraídos da página de Dados Abertos do Ministério da Educação.

A despeito de dados nulos, o dataset apresentou uma consistência impressionante, sem valores nulos ou ausentes (Figura 2). Em termos de transformação, as colunas estavam corretamente tipificadas de acordo com o tipo de dado, facilitando as análises subsequentes (Figura 12). Contudo, ao examinar outliers e ruídos, identificamos que as colunas de vagas autorizadas e carga horária continham muitos valores discrepantes, exigindo uma atenção especial para garantir a precisão das análises (Figuras 5, 4, 9 e 8).

Observamos uma concentração significativa de cursos de graduação na região Sudeste, que possui a maior quantidade de cursos entre todas as regiões do Brasil, seguida pelas regiões Nordeste e Sul. As regiões Norte e Centro-Oeste apresentaram uma menor oferta de cursos de graduação (Figura 14). No Rio Grande do Norte, as cidades de Natal e Mossoró têm a maior concentração de cursos (Figura 19).

A análise revelou que a maioria dos cursos está em atividade, com uma menor proporção de cursos em extinção ou extintos (Figura 30). A situação dos cursos está correlacionada com a categoria administrativa, por exemplo, IES privadas sem fins lucrativos tem uma alta porcentagem de cursos em extinção e extintos (Figuras 36 e 37).

A maioria dos cursos são de grau Bacharelado para modalidade presencial (Figura 24) e para Tecnólogos a modalidade EaD (Figura 25).

Cursos presenciais tem uma proporção maior no quesito extinto e em extinção comparado a cursos EaD (Figuras 33 e 34).

Focamos em cursos de Tecnologia da Informação e áreas afins, eles são amplamente oferecidos e mostram uma distribuição mais homogênea entre as modalidades EaD (Figura 44). Pouco menos de 6% desses cursos estão extintos ou em extinção (Figura 46).

5.2 Trabalhos futuros

Possíveis direções para pesquisas e análises futuras incluem agrupar os cursos de graduação em clusters. Isso permitiria explorar padrões comuns entre os cursos dentro de cada cluster e identificar grupos de cursos com características semelhantes. Além disso, outra linha de investigação poderia ser a classificação do grau de risco de um determinado curso ser extinto, utilizando os atributos do dataset como base.

REFERÊNCIAS

- EDUCACAO, M. População do brasil. 2022. Disponível em: <https://mundoeducacao.uol.com.br/geografia/populacao-brasileira.htm>.
- INEP. Diretoria de estatísticas educacionais. 2016. Disponível em: https://www.uff.br/sites/default/files/paginas-internas-orgaos/modulo_ies_2016.pdf.
- INSPER. Explorando os cursos de graduação na Área de tecnologia: Conceitos e opções. 2023. Disponível em: <https://www.insper.edu.br/noticias/cursos-de-graduacao-na-area-de-tecnologia/>.
- MEC. Cursos de graduação do brasil. 2022. Disponível em: <https://dadosabertos.mec.gov.br/indicadores-sobre-ensino-superior/item/183-cursos-de-graduacao-do-brasil>.
- UNOPAR. Carga horária da faculdade: como funciona e o que o mec exige? 2022. Disponível em: <https://blog.unopar.com.br/carga-horaria-faculdade/>.