



UNIVERSIDADE PRESBITERIANA MACKENZIE
FACULDADE DE COMPUTAÇÃO E INFORMÁTICA
TECNOLOGIA EM CIÊNCIAS DE DADOS

PROJETO APLICADO I

PROFESSOR:

Ismar Frango

GRUPO:

David Wildson Gonçalves Santos – 10744094 – 10744094@mackenzista.com.br
Henrique Machado Barbosa – 10747413 – 10747413@mackenzista.com.br
Tomaz Teles Carneiro – 10746600 – 10746600@mackenzista.com.br
Vanessa Cordeiro Gonçalves – 10415118 – 10415118@mackenzista.com.br

São Paulo
2025

Sumário

1. Contexto do Estudo.....	3
a. Nome da empresa.....	3
b. Missão.....	3
c. Visão.....	3
d. Valores.....	3
e. Quantidade de colaboradores.....	3
f. Apresentação da Empresa e Área de Atuação.....	3
g. Premissas do Projeto.....	4
h. Objetivos.....	4
i. Metas.....	4
j. Cronograma de Atividades.....	5
k. Pensamento Computacional em Contextos Organizacionais.....	5
2. Referências de Aquisição do Dataset.....	5
a. Dataset Escolhido.....	5
b. Origem dos Dados.....	5
c. Descrição da Base.....	5
d. Limitações de Uso.....	6
e. Informações Sobre a Base:.....	6
3. Descrição da Origem.....	6
4. Descrição do Dataset.....	6
a. Proposta do Dataset.....	6
b. Conteúdo do Dataset.....	7
c. Descrição das colunas.....	7
d. Observações (Problemas e Fenômenos Registrados).....	8
5. Proposta Analítica.....	8
a. Objetivo.....	8
b. Principais Tópicos	9
c. Proposta Analítica.....	9
6. Análise Exploratória.....	10
a. Análise Temporal.....	10
b. Análise Distribuição	11
c. Análise Geográfica.....	11
d. Análise Comparativa.....	12
e. Scripts.....	12
f. Resultados.....	13
7. Repositório no GitHub.....	15
8. Conclusão.....	15
9. Referências Bibliográficas.....	15

1. Contexto do Estudo

a. Nome da empresa

Agência Nacional de Telecomunicações - Anatel

b. Missão

Regular o setor de telecomunicações para contribuir com o desenvolvimento do Brasil.

c. Visão

Ser reconhecida como instituição de excelência que promove um ambiente favorável para as comunicações no Brasil, em benefício da sociedade brasileira.

d. Valores

Capacitação Institucional, Segurança Regulatória, Transparência e Participação social

e. Quantidade de colaboradores

Atualmente a ANATEL tem em seu quadro um total de 1260 servidores públicos efetivos.

f. Apresentação da Empresa e Área de Atuação

A Agência Nacional de Telecomunicações (Anatel) é uma autarquia especial vinculada ao Ministério das Comunicações, criada pela Lei nº 9.472/1997. Trata-se do órgão regulador do setor de telecomunicações no Brasil, com autonomia administrativa e financeira.

Sua área de atuação abrange a regulação, fiscalização e promoção da competição nos serviços de telecomunicações, incluindo telefonia fixa e móvel, internet banda larga, TV por assinatura e serviços de comunicação multimídia. Além disso, a Anatel é responsável por zelar pelos direitos dos consumidores, garantindo a qualidade, continuidade e acessibilidade dos serviços prestados pelas operadoras.

Por meio de iniciativas de monitoramento e atendimento, a agência coleta e organiza dados de reclamações, denúncias, pedidos de informação e sugestões feitas pelos usuários, funcionando como um elo essencial entre consumidores, empresas prestadoras de serviços e o governo. Dessa forma, a Anatel desempenha papel estratégico no desenvolvimento e modernização das telecomunicações brasileiras, fundamentais para a inclusão digital, a inovação tecnológica e a infraestrutura de comunicação do país.

g. Premissas do Projeto

Desenvolvimento de um estudo prático sobre os dados de uma empresa escolhida pelo grupo, visando trabalhar aspectos práticos e reais do cotidiano de um cientista de dados.

h. Objetivos

- Compreender o comportamento das manifestações registradas na Anatel, identificando padrões de volume, evolução temporal e distribuição geográfica das solicitações.
- Analisar a qualidade percebida dos serviços de telecomunicações, a partir da categorização por serviço (telefonia móvel, fixa, internet, TV por assinatura) e pelas marcas/operadoras.
- Identificar os principais problemas relatados pelos consumidores, destacando assuntos recorrentes e possíveis gargalos no setor.
- Avaliar os canais de entrada utilizados pelos cidadãos, verificando quais são mais acessados e como contribuem para a representatividade das manifestações.
- Produzir insights que possam apoiar políticas públicas e decisões regulatórias da Anatel e subsidiar ações de melhoria pelas operadoras.

i. Metas

- Criar séries históricas que indiquem a evolução das reclamações por período (Ano, Mês, Ano/Mês).
- Mapear a distribuição geográfica das manifestações por UF e município.
- Elaborar dashboards ou relatórios visuais que apresentem a situação por serviço e por marca.
- Identificar o ranking de problemas mais frequentes relatados pelos consumidores.
- Calcular indicadores comparativos, como quantidade de reclamações por serviço e por operadora.
- Propor recomendações baseadas em evidências para melhorar a transparência, a regulação e o atendimento ao consumidor.

j. Cronograma de Atividades

	AGOSTO	SETEMBRO	OUTUBRO	NOVEMBR	RESPONSÁVEIS
	S4	S1 S2 S3 S4	S1 S2 S3 S4	S1 S2 S3	
ETAPA 1					
FORMAÇÃO DE GRUPOS		█			TODOS
DEFINIÇÃO DA PREMISSAS DO PROJETO		█			HENRIQUE E TOMAZ
OBJETIVOS E METAS					HENRIQUE E TOMAZ
ORGANIZAÇÃO DO DOCUMENTO		█			TODOS
CRIAÇÃO DO REPOSITÓRIO		█			VANESSA E DAVID
CRIAÇÃO DO CRONOGRAMA DE ATIVIDADES		█			VANESSA E DAVID
ETAPA 2					
ELABORAÇÃO DE UMA PROPOSTA ANALÍTICA		█			TODOS
AJUSTES NOS OBJETIVOS E CRONOGRAMA			█		VANESSA
PREPARAÇÃO, TRATAMENTO E TRANSFORMAÇÕES DA BASE DE DADOS			█		TOMAZ E DAVID
ELABORAÇÃO DA ANÁLISE EXPLORATÓRIA EM PYTHON			█		HENRIQUE E VANESSA
ESTRUTURAÇÃO DO REPOSITÓRIO			█		TODOS
DESCRIÇÃO DA BASE TEÓRICA DO MÉTODO ANALÍTICO			█		TODOS
ETAPA 3					
AJUSTES NO CÓDIGO DE ANÁLISE EXPLORATÓRIA			█	█	TODOS
CONSOLIDAÇÃO DOS RESULTADOS ANALÍTICOS			█	█	TODOS
ELABORAÇÃO DO ESBOÇO DA APRESENTAÇÃO			█	█	TODOS
ETAPA 4					
FINALIZAÇÃO DA ANÁLISE EXPLORATÓRIA			█	█	TODOS
CONCLUSÃO DA APRESENTAÇÃO			█	█	TODOS
ELABORAÇÃO DO DOCUMENTO FINAL			█	█	TODOS

TAREFAS REALIZADAS
TAREFAS A REALIZAR

⊕ Cronograma - Projeto Aplicado I

k. Pensamento Computacional em Contextos Organizacionais

Utilização da Análise Exploratória de Dados, Conhecimentos voltados ao contexto de Bases de Dados, Utilização de Linguagens de Programação (R e Python), Utilização de conceitos de Analytics.

2. Referências de Aquisição do Dataset

a. Dataset Escolhido

Reclamações Registradas na Anatel (link: <https://dados.gov.br/dados/conjuntos-dados/solicitacoesregistradasnaanatel>)

b. Origem dos Dados

Dados provenientes do GOV.BR pela estrutura “Dados Abertos”. Operacionalizado pela Gerência de Tratamento de Solicitações de Consumidores da Agência Nacional de Telecomunicações (Anatel).

c. Descrição da Base

Quantidade de registros de reclamações, denúncias, pedidos de informação e sugestões em relação às operadoras e à Anatel.

d. Limitações de Uso

- Detalhamento limitado: não há informações sobre perfil do usuário, status da solicitação, tempo de resposta ou resolução.
- Padronização: campos como Assunto e Problema podem conter inconsistências textuais, exigindo tratamento prévio.

e. Informações Sobre a Base:

Período de Coleta: A partir de Janeiro/2015.

Data de atualização do arquivo: 19/08/2025

3. Descrição da Origem

Origem e Contexto: A Agência Nacional de Telecomunicações (Anatel) é a autarquia federal responsável pela regulação, fiscalização e acompanhamento do setor de telecomunicações no Brasil. A agência tem como missão garantir que os serviços de telecomunicações sejam prestados de forma adequada, eficiente e em conformidade com a legislação vigente, assegurando a defesa dos direitos dos consumidores.

A Anatel mantém canais oficiais de atendimento ao cidadão, como telefone, aplicativo, site e ouvidoria, por meio dos quais são registradas reclamações, denúncias, pedidos de informação e sugestões relacionadas tanto às operadoras de telecomunicações (telefonia fixa, móvel, internet, TV por assinatura) quanto à própria agência.

A base de dados observada reflete esse processo de atendimento e mediação entre consumidores e prestadoras de serviços, sendo composta pela quantidade de registros efetuados ao longo do tempo. Os dados são coletados diretamente a partir das manifestações dos usuários e processados pela Anatel com o objetivo de:

- Monitorar a qualidade e a transparência do atendimento das operadoras;
- Identificar problemas recorrentes e tendências no setor;
- Apoiar políticas públicas e decisões regulatórias;
- Fortalecer a proteção dos consumidores.

Assim, o contexto em que os dados foram gerados está diretamente ligado ao papel regulador da Anatel e ao exercício do direito dos cidadãos de demandar serviços de telecomunicações mais justos, acessíveis e de qualidade.

4. Descrição do Dataset

a. Proposta do Dataset

A proposta do dataset é registrar e organizar as manifestações feitas pelos consumidores à Anatel (reclamações, denúncias, pedidos de informação e

sugestões) relacionadas aos serviços e operadoras de telecomunicações no Brasil, permitindo o acompanhamento do volume, tipo, localidade e natureza dos problemas reportados para fins de monitoramento, regulação e análise da qualidade do setor.

b. Conteúdo do Dataset

Para uma análise preliminar do conteúdo do dataset, foi utilizada a ferramenta Jupyter Notebook para extrair as colunas dos arquivos do dataset das reclamações registradas na Anatel:

```
import pandas as pd

reclamacoes = pd.read_csv("reclamacoes.csv", encoding='utf-8',
on_bad_lines='skip', delimiter=';')
contexto = pd.read_csv("reclamacoes_contexto.csv", encoding='utf-8',
on_bad_lines='skip', delimiter=';')

print(' '.join(reclamacoes.columns.values))

print(' '.join(contexto.columns.values))
```

Os resultados obtidos estão descritos abaixo.

reclamacoes.csv: DataExtracao, SOLICITAÇÕES, Ano, Mês, AnoMês, UF, Cidade, CO_MUNICIPIO, CanalEntrada, Condição, TipoAtendimento, Serviço, Marca, Assunto, Problema

reclamacoes_contexto.csv: DataExtracao, SOLICITAÇÕES, Ano, Mês, AnoMês, UF, Cidade, CO_MUNICIPIO, CanalEntrada, Condição, TipoAtendimento, Serviço, Marca, Assunto, Problema, Linha

c. Descrição das colunas

DataExtracao – Data em que os dados foram extraídos.

SOLICITAÇÕES – Quantidade de registros por linha (reclamações, denúncias, pedidos de informação ou sugestões).

Ano, Mês, AnoMês – Referência temporal do registro.

UF, Cidade, CO_MUNICIPIO – Localização geográfica da manifestação.

CanalEntrada – Canal utilizado pelo consumidor (ex.: Call Center, internet, aplicativo).

Condição – Situação da solicitação, conforme classificação da Anatel (Novo, Reencaminhado).

TipoAtendimento – Categoria de atendimento registrada.

Serviço – Tipo de serviço de telecomunicações relacionado (telefonia fixa, móvel, internet, TV por assinatura etc.).

Marca – Operadora ou prestadora do serviço.

Assunto – Tema principal da manifestação.

Problema – Detalhamento específico da questão relatada.

d. Observações (Problemas e Fenômenos Registrados)

Para a preparação para a futura análise do dataset, utilizou-se comandos comuns da biblioteca Pandas do Python, e foram observadas as seguintes características.

A tabela de reclamações tem 15,952,407 linhas. Essa é uma grande quantidade de informações, correspondente a reclamações pelo Brasil inteiro à Anatel. Devido ao tamanho da amostra, é possível uma análise refinada sobre quaisquer propriedades estatísticas e correlações entre variáveis do dataset, com baixas incertezas. Assim, é possível determinar a existência de correlações fracas e eventos raros, o que só é possível com amostras desse tamanho.

Analizando os tipos de dados das colunas, notou-se que a maior parte dessas é de natureza categórica:

#	Column	Dtype
0	DataExtracao	object
1	SOLICITAÇÕES	int64
2	Ano	int64
3	Mês	int64
4	AnoMês	object
5	UF	object
6	Cidade	object
7	CO_MUNICIPIO	object
8	CanalEntrada	object
9	Condição	object
10	TipoAtendimento	object
11	Serviço	object
12	Marca	object
13	Assunto	object
14	Problema	object

Isso significa que pode ser feita uma análise extensiva sobre correlações entre categorias, como, por exemplo: quais marcas têm a maior quantidade de problemas, como a condição do aparelho afeta as reclamações, entre outras.

5. Proposta Analítica

a. Objetivo:

O objetivo da análise é identificar e quantificar as principais tendências, padrões geográficos e a distribuição temporal das manifestações dos consumidores no setor de

telecomunicações brasileiro registradas na Anatel, buscando fornecer insights para a atuação de regulação e a avaliação da qualidade dos serviços oferecidos pelas operadoras.

b. Principais Tópicos:

Quais são os principais motivadores de manifestações?

Quais são os serviços (telefonia fixa, móvel, internet, TV por assinatura) e os assuntos/problemas (cobrança, qualidade, cancelamento, etc.) que geram o maior volume de manifestações?

Qual a performance comparando diferentes operadoras?

Quais operadoras de telecomunicações concentram o maior volume de manifestações, em termos absolutos e relativos (se possível, estimando o volume por cliente/mercado)?

Existem padrões geográficos nas manifestações?

Há concentração de manifestações em determinadas Unidades da Federação (UF) ou Regiões? Os problemas variam significativamente entre os estados?

Como as manifestações evoluíram ao longo do tempo?

Houve aumento ou redução no volume total de manifestações por ano/mês desde 2015? Essa variação pode ser correlacionada com eventos regulatórios, sazonalidade ou mudanças no mercado?

Qual o perfil da demanda por tipo de manifestação?

Qual a proporção entre reclamações, denúncias, pedidos de informação e sugestões? A proporção mudou ao longo do tempo? Existe sazonalidade?

c. Proposta Analítica:

i. Limpeza e Preparação dos Dados:

Tratamento de Inconsistências: Padronizar e normalizar os campos textuais críticos, como Assunto e Problema, por meio de técnicas de tratamento de strings ou agrupamento para lidar com variações e inconsistências.

Hierarquização: Criar categorias de alto nível para os campos normalizados (ex: agrupar "Cobrança indevida", "Valor divergente" em "Problemas de Cobrança").

Temporalidade: Garantir que a coluna de data (período de coleta a partir de Janeiro/2015) esteja em formato adequado para análises de série temporal.

Estruturação Geográfica: Garantir a correta identificação e padronização das informações geográficas (UF, Região).

ii. Análise Exploratória e Descritiva:

Análise de Volume: Calcular o volume total e mensal de manifestações por ano.

Distribuição de Frequências: Frequência de manifestações por Operadora, Serviço e UF.

Cruzamento das variáveis:

Operadora vs. Assunto/Problema (para identificar os principais ofensores de cada empresa).

UF vs. Serviço (para identificar prioridades regulatórias regionais).

Visualização: Criação de gráficos de barras para rankings, mapas de calor para distribuição geográfica e gráficos de linha para a evolução temporal.

iii. Análise Preditiva e de Tendências

Modelagem de Tendências: Utilização de modelos de série temporal para prever o volume futuro de manifestações totais ou por tipo de serviço, detectando outliers (picos anômalos).

Análise de Variação: Estudar a taxa de crescimento/declínio das manifestações ao longo do tempo, identificando se o problema está sendo resolvido.

6. Análise Exploratória

a. Análise Temporal

i. Volume total ao longo do tempo:

Agregação: Conte o número total de manifestações por Mês/Ano.

Visualização: Gráfico de Linha (Série Temporal) do volume total de manifestações ao longo do tempo (desde Jan/2015).

Insights Chave: Identificação de tendência, Detecção de sazonalidade (picos em meses específicos) e Identificação de eventos anômalos (outliers, como picos repentinos que podem indicar um grande problema de serviço ou uma mudança regulatória/de plataforma).

ii. Volume por tipo de manifestação ao longo do tempo:

Agregação: Conte o número de manifestações por Mês/Ano, segregando por Tipo de Manifestação (Reclamação, Denúncia, etc.).

Visualização: Gráfico de Linhas Múltiplas (uma para cada tipo) ou Gráfico de Área Empilhada.

Insights Chave: Verificar se a proporção entre Reclamações e Pedidos de Informação mudou. Isso pode indicar uma alteração na forma como o consumidor interage com a Anatel.

b. Análise de Frequência e Distribuição

i. Ranking de manifestações:

Agregação: Contagem do número de ocorrências para as categorias de Assunto e, em seguida, Problema (usando as categorias normalizadas).

Visualização: Gráfico de Barras para mostrar o Top 10 de Assuntos e Problemas mais frequentes.

Insights Chave: Identificar o principal Assunto (ex: "Cobrança") e o principal Problema específico (ex: "Cobrança após cancelamento"). Estes são os pontos de atenção regulatória primários.

ii. Ranking de operadoras:

Agregação: Conte o número de manifestações por Operadora.

Visualização: Gráfico de Barras para mostrar o Top 5 ou Top 10 de operadoras com maior volume.

Insights Chave: Quais operadoras dominam o volume de manifestações.

c. Análise Geográfica

i. Distribuição por UF:

Agregação: Conte o número de manifestações por UF.

Visualização: Mapa Coroplético (Heatmap Geográfico) do Brasil, colorindo as UFs de acordo com o volume de manifestações.

Insights Chave: Identificar UFs com alto volume que não são necessariamente as mais populosas

ii. Problemas por Região:

Cruzamento de Informação: Cruzar UF (ou Região) com as categorias normalizadas de Assunto.

Visualização: Gráficos de Barras Agrupadas (mostrando a distribuição do Top 3 Assuntos dentro de cada Região).

Insights Chave: Identificação de perfis de reclamação regionais.

d. Análise Comparativa:

i. Assuntos e Problemas por operadora:

Cruzamento de Informação: Cruzar Operadora com as categorias de Assunto.

Visualização: Gráfico de Barras Empilhadas ou uma Tabela Dinâmica de Calor (heatmap).

Insights Chave: Identificação do calcanhar de Aquiles de cada operadora)

ii. Ranking Operadoras por serviço:

Cruzamento de Informação: Cruzar Operadora com Serviço (Telefonia Móvel, Internet Banda Larga, etc.).

Visualização: Gráfico de Barras Agrupadas.

Insights Chave: Identificação de problemas de performance em serviços específicos.

e. Scripts

```
import pandas as pd

from zipfile import ZipFile

import matplotlib.pyplot as plt

# Carregar o dataset

with ZipFile('base/consumidor_reclamacoes/consumidor_reclamacoes.zip', 'r') as arquivo:

    reclamacoes = pd.read_csv(arquivo.open("reclamacoes.csv"), encoding='utf-8',
    on_bad_lines='skip', delimiter=';')

    contexto = pd.read_csv(arquivo.open("reclamacoes_contexto.csv"), encoding='utf-8',
    on_bad_lines='skip', delimiter=';')

reclamacoes
```

```

contexto

reclamacoes.columns

contexto.columns

contexto.drop("Linha", axis=1, inplace=True)

reclamacoes.drop("Ano", axis=1, inplace=True)

reclamacoes.drop("Mês", axis=1, inplace=True)

reclamacoes["DataExtracao"] = pd.to_datetime(reclamacoes["DataExtracao"])

reclamacoes["AnoMês"] = pd.to_datetime(reclamacoes["AnoMês"])

reclamacoes.nunique()

reclamacoes.dtypes

for col in reclamacoes.columns:

    if reclamacoes.dtypes[col] == object:

        reclamacoes[col] = pd.Categorical(reclamacoes[col])

reclamacoes.dtypes

reclamacoes.groupby("Marca")["Marca"].count().sort_values(ascending=False)

reclamacoes.groupby("Marca")["Marca"].count().sort_values(ascending=False).plot.pie(autopct='%.1f%%', figsize=(8, 8), labeldistance=1.1)

plt.ylabel("")

reclamacoes.groupby("TipoAtendimento")["TipoAtendimento"].count().plot.pie(autopct='%.1f%%', figsize=(6, 6))

plt.ylabel("")

reclamacoes['Count'] = 1

pivot = reclamacoes.pivot_table(index='AnoMês', columns='Marca', values='Count',
aggfunc='sum', fill_value=0, )

cumulativo = pivot.cumsum()

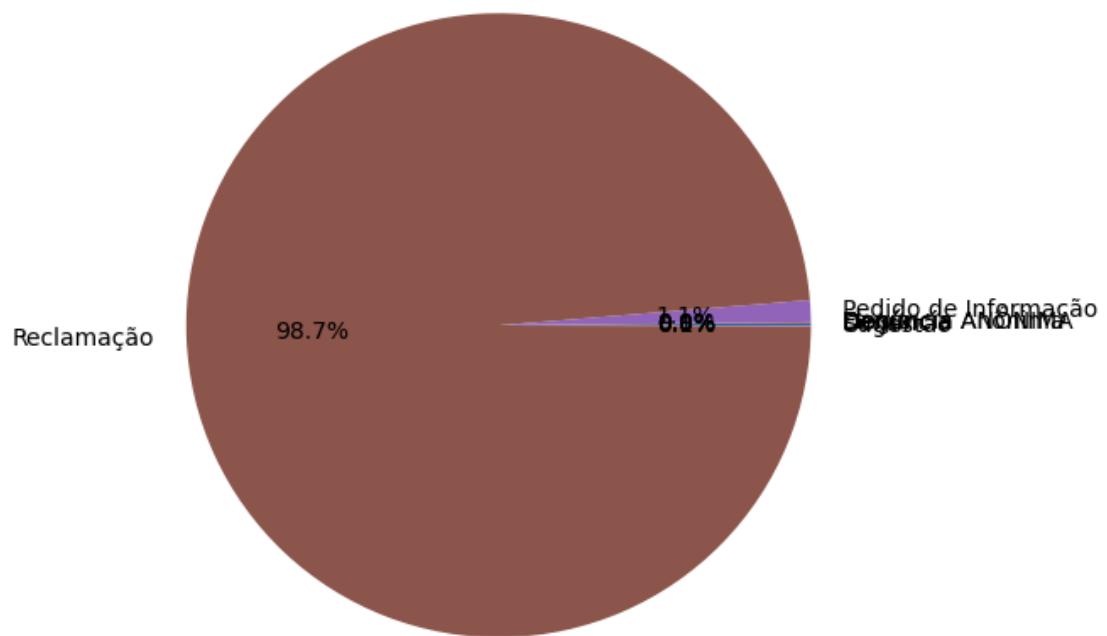
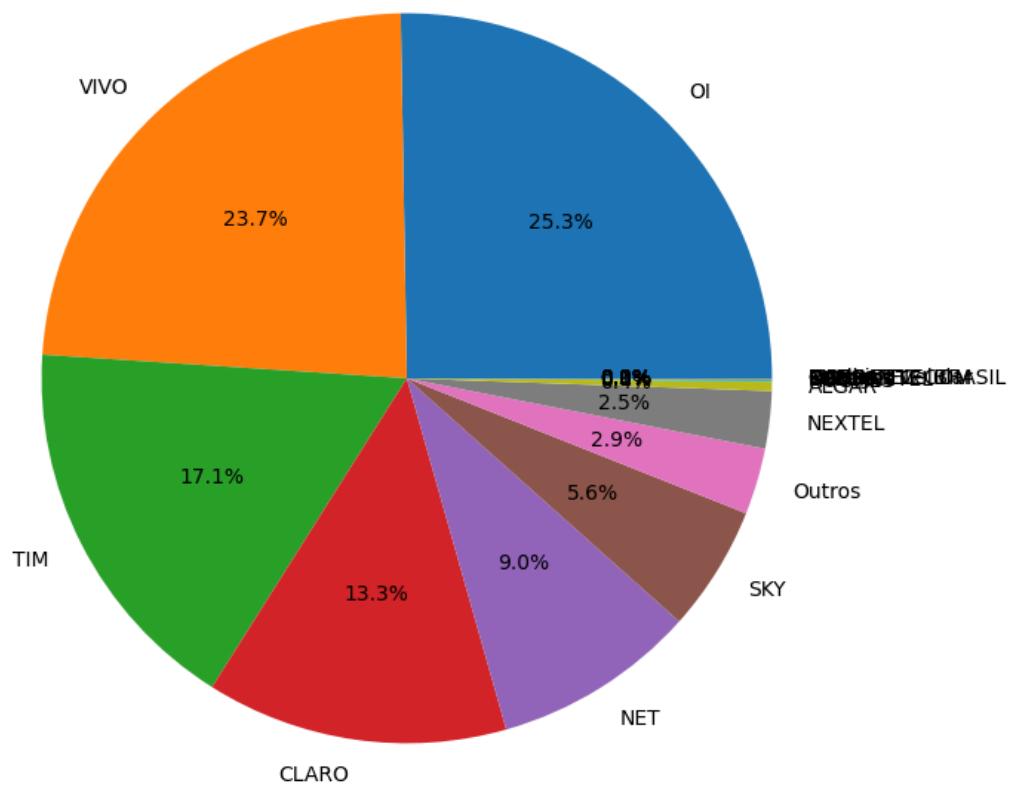
ax = cumulativo.plot(figsize=(10,6))

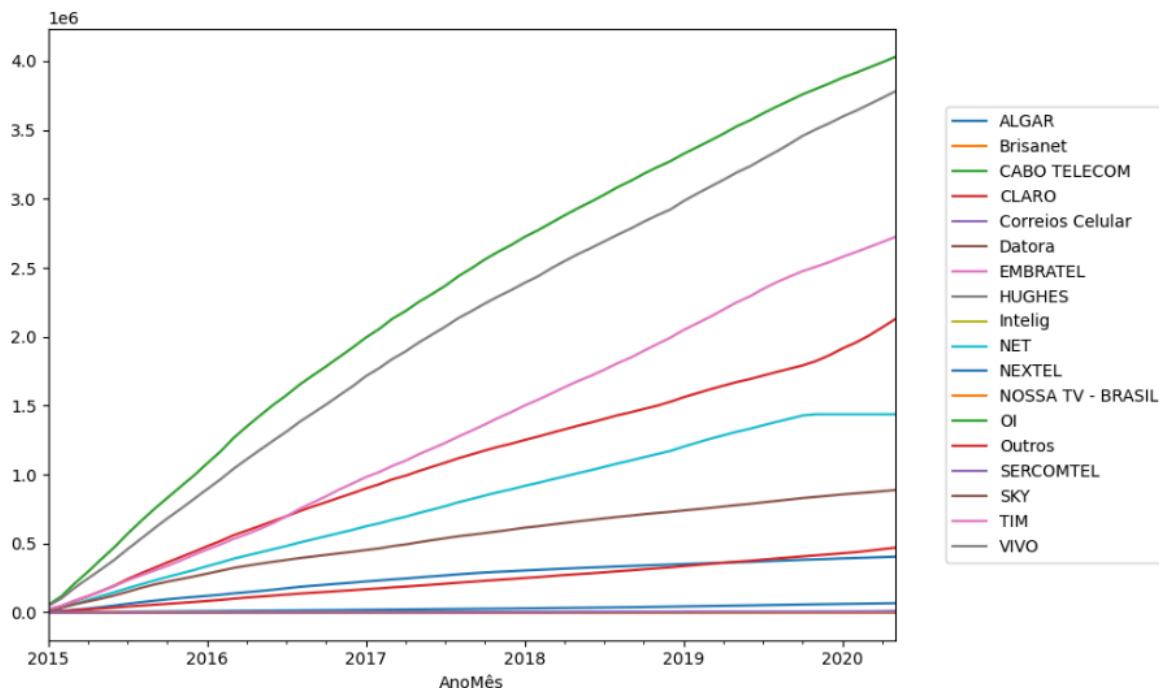
ax.legend(bbox_to_anchor=(1.05, 0.5), loc='center left')

plt.tight_layout()

```

f. Resultados





7. Repositório no GitHub

<https://github.com/vanessacordeiro/projeto-aplicado-I-mackenzie/>

8. Conclusão

Em decorrência disso, espera-se que seja possível, ao longo do projeto, uma análise extensiva das propriedades do dataset de reclamações à Anatel. Em particular, é possível determinar várias correlações entre cada categoria e as reclamações. Para atingir as metas propostas no projeto, serão empregadas Análise Exploratória de Dados (AED), limpeza de dados, engenharia de dados, Python, Pandas, Jupyter Notebooks, e possivelmente outras ferramentas. Dessa forma, esse projeto possibilita uma aplicação real da Ciência de Dados, com possíveis novas descobertas significativas sobre reclamações da Anatel.

9. Referências Bibliográficas

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES – ANATEL. Sobre a Anatel. Disponível em: <https://www.gov.br/anatel/pt-br/acesso-a-informacao/sobre-a-anatel>. Acesso em: 10 set. 2025.

Agência Nacional de Telecomunicações – Dados Abertos. Disponível em: <https://dados.gov.br/dados/organizacoes/visualizar/agencia-nacional-de-telecomunicacoes>. Acesso em: 10 set. 2025.

O que é a Anatel e qual sua função?. Disponível em: <https://vcx.solutions/anatel/>. Acesso em: 10 set. 2025.

Dados abertos - Sobre - ANATEL.

<https://dados.gov.br/dados/organizacoes/visualizar/agencia-nacional-de-telecomunicacoes>. Acesso em: 16 out 2025.

Portal da transparência - Consulta de servidores por órgão - ANATEL

<https://portaldatransparencia.gov.br/servidores/orgao?paginacaoSimples=true&tamanhoPagina=&offset=&direcaoOrdenacao=asc&orgaosServidorLotacao=OR41231&colunasSelecionadas=linkDetalhamento%2CorgaoSuperiorLotacaoSIAPE%2CorgaoLotacaoSIAPE%2CquantidadeAtivos%2CquantidadeInativos%2CquantidadePensionistas&ordenarPor=orgaoSuperiorLotacaoSIAPE&direcao=asc>. Acesso em 16 out 2025.