



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Analytics e Business Intelligence

RELATÓRIO TÉCNICO

ANÁLISE AVANÇADA PARA CONTENÇÃO DO NÚMERO CRESCENTE DE
RECLAMAÇÕES NO PROCON

Lennon Brandão Freitas do Nascimento

Belo Horizonte

2022

SUMÁRIO

- 1. Introdução4**
 - 1.1. Contexto4
 - 1.2. Objetivos.....5
 - 1.3. Público alvo5
- 2. Modelo de Dados5**
 - 2.1. Modelo Dimensional5
 - 2.2. Fatos e Dimensões.....6
- 3. Integração, Tratamento e Carga de Dados 10**
 - 3.1. Fontes de Dados 10
 - 3.2. Processos de Integração e Carga (ETL) 10
- 6. Links14**
- REFERÊNCIAS 15**

1. Introdução

1.1. Contexto

Diariamente, os órgãos públicos de defesa do consumidor atendem milhares de cidadãos. Para que as demandas sejam atendidas com maior velocidade, os Programas de Proteção e Defesa do Consumido (PROCONS) estaduais tratam a maior parte das demandas recebidas com procedimentos locais. Entretanto, existem casos onde há a necessidade de abrir um processo administrativo para a resolução do problema. Esse processo administrativo para tratamento da demanda individual do consumidor é denominado Reclamação [1].

As Reclamações são cadastradas em uma base de dados denominada Cadastros de Reclamações Fundamentadas. Essa base é formada pelas Reclamações encerradas pelo Programa de Proteção e Defesa do Consumidor (PROCON) de todos os estados do país no período de doze meses, que indicam se houve ou não acordo entre as partes [2]. Essas informações são consolidadas por meio do Sistema Nacional de Informações de Defesa do Consumidor (SINDEC) que consiste em um sistema informatizado que integra processos e procedimentos relativos ao atendimento aos consumidores nos PROCONS, visando proporcionar um instrumento de gestão adequado ao dinamismo que é típico de seus setores de atendimento [3].

O SINDEC integra hoje vinte e seis PROCONS estaduais e seiscentos e sessenta e sete PROCONS municipais. Como vários desses PROCONS contam com mais de uma unidade, o Sistema opera em novecentas e sessenta e oito unidades espalhadas por mais de setecentas cidades brasileiras. Esses PROCONS atendem a uma média mensal de duzentos e dezesseis mil consumidores [4]. Assim, com essa integração de informações, o SINDEC permitiu o aumento da transparência, perante a sociedade, dos problemas que ocorrem no mercado de consumo, evidenciando a existência de uma grande demanda por parte dos consumidores brasileiros para resolução de problemas, por meio dos registros disponibilizados. Tendo isso em vista, há uma necessidade por parte desses órgãos de uma análise direcionada a fim que campanhas e intervenções sejam planejadas no intuito de diminuir a frequência de ocorrências de Reclamações.

1.2. Objetivos

Este trabalho visa a realização de uma análise avançada empregando o uso de estatística com a base de dados de Cadastros de Reclamações Fundamentadas, consolidada pela SINDEC no período de 2018 até 2020. Durante o desenvolvimento haverá a etapa de processos de preparação, transformação, e carregamentos dos dados. Essa etapa favorecerá a criação de *Dashboards* contendo informações essenciais acerca das Reclamações realizadas por cada consumidor registrado. E, por fim, será realizada a análise citada no decorrer dos anos disponibilizados para que sejam propostas intervenções para a diminuição das ocorrências de Reclamações por parte dos consumidores brasileiros e assim melhorar o serviço do PROCON.

1.3. Público alvo

O público-alvo deste projeto serão os analistas de órgãos que compõem o Sistema Nacional de Defesa do Consumidor. Mais especificamente são: PROCONS, Ministério Público e Defensoria Pública. Esses órgãos podem realizar campanhas a partir da análise avançada sem que haja analistas com grande conhecimento na área de TI, a fim de intervir no crescimento da quantidade de Reclamações bem como agilizar o processo de conciliação entre as partes envolvidas.

2. Modelo de Dados

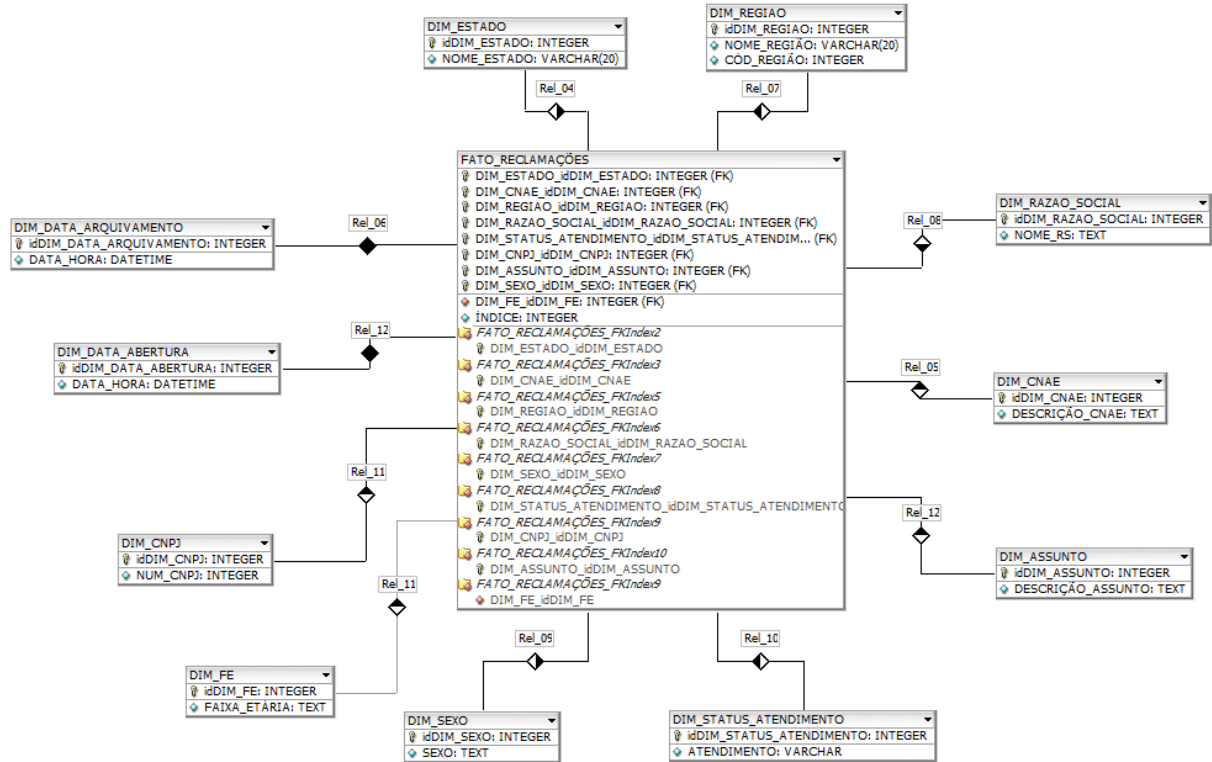
2.1. Modelo Dimensional

O conceito de Modelagem Dimensional foi desenvolvido por Ralph Kimball e é constituído essencialmente por dois conceitos: fatos e dimensões. Os Fatos são eventos do processo de negócios, enquanto a dimensão fornece o contexto em torno desse fato. Este modelo foi desenvolvido a fim de possibilitar o armazenamento dados de forma a facilitar o armazenamento e recuperação os dados em um *data warehouse* [5].

Utilizando-se dos conceitos da modelagem dimensional, neste projeto foram empregadas as informações da base de dados de Reclamações Fundamentadas disponibilizada em arquivos de formato .csv com vinte e três colunas para construir a tabelas fatos e dimensões. A modelagem realizada foi baseada no esquema estrela, do inglês, *star schema*, o que possibilitou uma melhor visão acerca da disposição dos

dados e como será o processo de relação entre eles. Na figura 1 é apresentado o modelo dimensional gerado no *software DBDesigner*.

Figura 1 - Modelagem Dimensional para base selecionada.



2.2. Fatos e Dimensões

A partir da análise exploratória dos dados na base de dados selecionada, foi possível criar uma tabela fato e onze tabelas dimensão conforme preconizado na modelagem dimensional. Na tabela 1, serão apresentadas todas as tabelas geradas bem como sua classificação como fato ou dimensão.

Tabela 1 - Tabelas geradas na modelagem dimensional.

Nome da tabela	Natureza da tabela
FATO_RECLAMAÇÕES	Fato
DIM_DATA_ABERTURA	Dimensão
DIM_DATA_ARQUIVAMENTO	Dimensão
DIM_ESTADO	Dimensão
DIM_REGIÃO	Dimensão
DIM_RAZÃO_SOCIAL	Dimensão
DIM_CNAE	Dimensão
DIM_ASSUNTO	Dimensão
DIM_STATUS_ATENDIMENTO	Dimensão

DIM_SEXO	Dimensão
DIM_FE	Dimensão
DIM_CNPJ	Dimensão

Ao visualizar a Tabela 1, é possível observar de forma clara e resumida o resultado da modelagem dimensional realizada. Para um detalhamento maior do conteúdo das dimensões, é possível observar as Tabelas 2 a 12. Da mesma forma, para detalhamento do conteúdo da Fato, é possível observar a Tabela 13.

Tabela 2 - Descrição da tabela DIM_DATA_ABERTURA

TABELA: DIM_DATA_ABERTURA		
Descrição da tabela:	Contém a data de abertura do processo de Reclamação	
Coluna	Tipo	Descrição
IdDIM_DATA_ABERTURA	INTEGER	Identificação única do campo
DATA_HORA	DATETIME	Data de abertura das reclamações (Formato/exemplo: 2018-08-13 12:20:39.00)

Tabela 3 - Descrição da tabela DIM_DATA_ARQUIVAMENTO

TABELA: DIM_DATA_ARQUIVAMENTO		
Descrição da tabela:	Contém a data de arquivamento do processo de Reclamação	
Coluna	Tipo	Descrição
IdDIM_DATA_ARQUIVAMENTO	INTEGER	Identificação única do campo
DATA_HORA	DATETIME	Data de arquivamento das reclamações (Formato/exemplo: 2018-08-13 12:20:39.000)

Tabela 4 - Descrição da tabela DIM_ESTADO

TABELA: DIM_ESTADO		
Descrição da tabela:	Contém a informação do estado de origem da Reclamação	
Coluna	Tipo	Descrição
IdDIM_ESTADO	INTEGER	Identificação única do campo
NOME_ESTADO	VARCHAR (20)	Sigla com dois caracteres representando Unidade da Federação do Procon

Tabela 5 - Descrição da tabela DIM_REGIÃO

TABELA: DIM_REGIÃO		
Descrição da tabela:	Contém a informação acerca da região de origem da Reclamação	
Coluna	Tipo	Descrição
IdDIM_REGIÃO	INTEGER	Identificação única do campo
NOME_REGIÃO	VARCHAR (20)	Região do Procon
CÓD_REGIÃO	INTEGER	Código identificador da região do Procon: 01 - Norte, 02 - Nordeste, 03 - Sudeste, 04 - Sul e 05 - Centro-Oeste

Tabela 6 - Descrição da tabela DIM_RAZÃO_SOCIAL

TABELA: DIM_RAZÃO_SOCIAL		
Descrição da tabela:	Contém a razão social da empresa que recebeu a Reclamação	
Coluna	Tipo	Descrição
IdDIM_RAZÃO_SOCIAL	INTEGER	Identificação única do campo
NOME_RS	TEXT	Razão social do fornecedor (empresa) na base de dados do Sindec

Tabela 7 - Descrição da tabela DIM_CNAE

TABELA: DIM_CNAE		
Descrição da tabela:	Contém descrição do CNAE da empresa que recebeu a Reclamação	
Coluna	Tipo	Descrição
IdDIM_CNAE	INTEGER	Identificação única do campo
DESCRIÇÃO_CNAE	TEXT	Descrição da Classificação Nacional de Atividades Econômicas principal do fornecedor.

Tabela 8 - Descrição da tabela DIM_ASSUNTO

TABELA: DIM_ASSUNTO		
Descrição da tabela:	Contém o assunto da Reclamação aberta	
Coluna	Tipo	Descrição
IdDIM_ASSUNTO	INTEGER	Identificação única do campo
DESCRIÇÃO_ASSUNTO	TEXT	Descrição dos assuntos do Sindec (produto ou serviço objeto da reclamação)
CÓD_ASSUNTO	INTEGER	Código identificador do assunto

Tabela 9 - Descrição da tabela DIM_STATUS_ATENDIMENTO

TABELA: DIM_STATUS_ATENDIMENTO		
Descrição da tabela:	Contém a informação se Reclamação foi atendida ou não	
Coluna	Tipo	Descrição
IdDIM_STATUS_ATENDIMENTO	INTEGER	Identificação única do campo
ATENDIMENTO	VARCHAR	Código identificador da reclamação fundamentada atendida ou não pela empresa/fornecedor: S – Atendida N – Não Atendida

Tabela 10 - Descrição da tabela DIM_SEXO

TABELA: DIM_SEXO		
Descrição da tabela:	Contém informações acerca do sexo do consumidor que realizou a Reclamação	
Coluna	Tipo	Descrição
IdDIM_SEXO	DATETIME	Identificação única do campo
SEXO	TEXT	Código identificador do sexo do consumidor: M – Masculino F – Feminino N – Não se aplica (são as reclamações (de ofício) em que o Procon é o reclamante)

Tabela 11 - Descrição da tabela DIM_FE

TABELA: DIM_FE		
Descrição da tabela:	Contém informações acerca do faixa etária do consumidor que realizou a Reclamação	
Coluna	Tipo	Descrição
IdDIM_FE	DATETIME	Identificação única do campo
FAIXA_ETÁRIA	TEXT	Faixa etária do consumidor: “até 20 anos”, “entre 21 e 30 anos”, “entre 31 e 40 anos”, “entre 41 e 50 anos”, “entre 51 e 60 anos”, “entre 61 e 70 anos”, “mais de 70 anos”, “Não Informada” (data de nascimento não informada no cadastro do consumidor) e “Não se aplica” (são as reclamações (de ofício) em que o Procon é o reclamante)

Tabela 12 - Descrição da tabela DIM_CNPJ

TABELA: DIM_CNPJ		
Descrição da tabela:	Contém a informação de CNPJ da empresa ou CPF da pessoa que recebeu Reclamação.	
Coluna	Tipo	Descrição
IdDIM_DATA_ARQUIVAMENTO	DATETIME	Identificação única do campo
NUM_CNPJ	INTEGER	Número do CNPJ - Cadastro Nacional de Pessoa Jurídica ou CPF - Cadastro de Pessoa Física

Tabela 13 – Descrição da tabela FATO_RECLAMAÇÕES

TABELA: FATO_RECLAMAÇÕES		
Natureza da tabela:	FATO	
Coluna	Tipo	Descrição
ÍNDICE	INTEGER	Identificação única do campo
idDIM_ESTADO	INTEGER	Chave estrangeira da tabela DIM_ESTADO
idDIM_CNAE	INTEGER	Chave estrangeira da tabela DIM_CNAE
idDIM_REGIAO	INTEGER	Chave estrangeira da tabela DIM_REGIAO
idDIM_RAZAO_SOCIAL	INTEGER	Chave estrangeira da tabela DIM_RAZAO_SOCIAL
idDIM_SEXO	INTEGER	Chave estrangeira da tabela DIM_SEXO
idDIM_FE	INTEGER	Chave estrangeira da tabela DIM_FE
idDIM_STATUS_ATENDIMENTO	INTEGER	Chave estrangeira da tabela DIM_STATUS_ATENDIMENTO
idDIM_CNPJ	INTEGER	Chave estrangeira da tabela DIM_CNPJ
idDIM_ASSUNTO	INTEGER	Chave estrangeira da tabela DIM_ASSUNTO
idDIM_DATA_ARQUIVAMENTO	INTEGER	Chave estrangeira da tabela DIM_DATA_ARQUIVAMENTO
idDIM_DATA_ABERTURA	INTEGER	Chave estrangeira da tabela DIM_DATA_ABERTURA

3. Integração, Tratamento e Carga de Dados

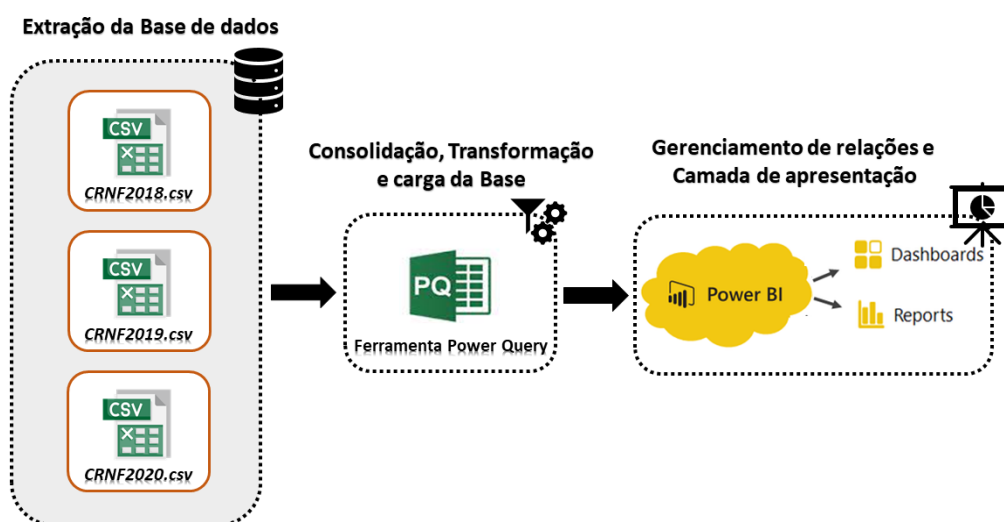
3.1. Fontes de Dados

O início do processo de análise exploratória, é a disponibilização dos dados. Assim, a base de dados utilizadas neste trabalho foi adquirida através do site <https://dados.gov.br/> do Governo Federal. Essa base de dados foi criada a partir do SINDEC que consiste em um sistema informatizado que integra processos e procedimentos relativos ao atendimento aos consumidores nos PROCONS. Inicialmente era composta por três arquivos no formato CSV (*comma separated values*), contendo as informações de cada processo de Reclamação que são: *CNRF2018.csv*, *CNRF2019.csv*, *CNRF2020.csv*. Assim, ao conhecer as informações contidas nesses arquivos, foi possível realizar a modelagem, conforme já apresentado na Figura 1.

3.2. Processos de Integração e Carga (ETL)

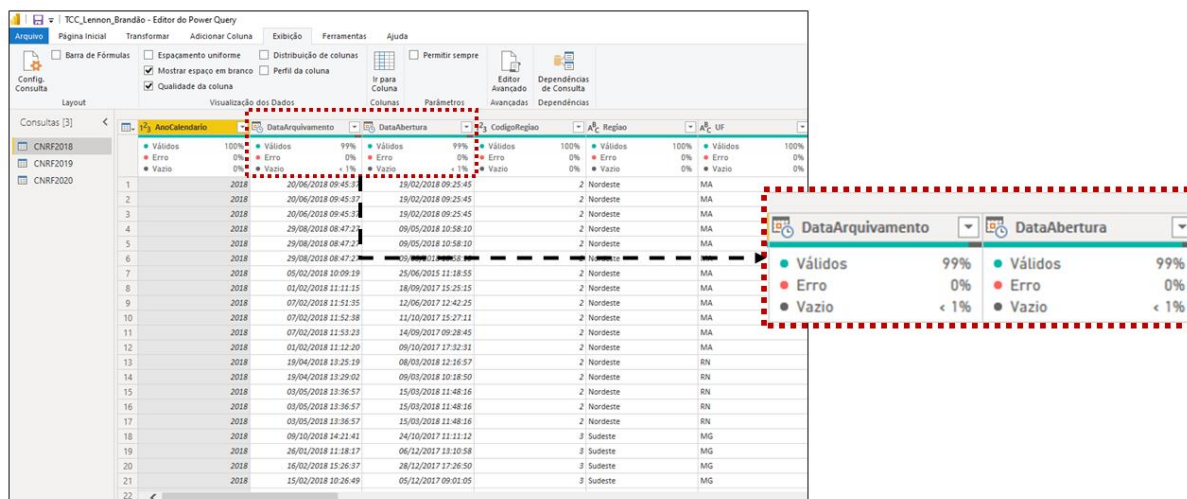
A metodologia utilizada no processo de extração, transformação e carga contém como ferramentas utilizadas, o Power Query Editor e Power BI. O Power Query atua como ferramenta de consulta, onde é possível localizar fontes de dados, fazer conexões e moldar esses dados. O Power BI, por sua vez, possibilitou gerenciar as relações criadas na base e servirá para a realização da camada de apresentação do projeto. Essa metodologia é apresentada na Figura 2.

Figura 2 - Metodologia para o ETL



Conforme já citado anteriormente, a extração da base de dados foi realizada no site do Governo Federal brasileiro e gerada a partir do SINDEC. Em posse dos arquivos, eles são importados via Power BI para o *software* Power Query Editor. É possível visualizar o resultado desse processo de carregamento dos três arquivos, na Figura 3.

Figura 3 - Arquivos importados no ambiente do Power Query.



Além de apresentar o ambiente do Power Query, a Figura 3 também apresenta uma análise realizada automaticamente por esse *software* que indicou que em algumas colunas havia presença de campos vazios. Essa análise já mostra a necessidade de tratamento da base de dados antes da realização da carga.

Na etapa de transformação, foram realizados os seguintes passos para o tratamento inicial da base:

- Consolidação da base de dados: dado que os arquivos no formato csv de cada ano são separados (*CNRF2018.csv*, *CNRF2019.csv*, *CNRF2020.csv*), a primeira etapa foi a consolidação em uma tabela só;
- Limpeza dos dados: Essa etapa consistiu na realização da Remoção de colunas não utilizadas, de células vazias, preenchidas com *null* ou com tipagem errada. As colunas retiradas constam na Tabela 14;

Tabela 14 - Colunas retiradas da base de dados.

Nome da Coluna Retirada	Descrição
RadicalCNPJ	Aplica-se para pessoa jurídica e serve para agrupar as informações de um mesmo fornecedor (matriz e

	filiais), sendo os oito primeiros dígitos do número do CNPJ.
RazaoSocialRFB	Razão social do fornecedor na base de dados da RFB – Receita Federal do Brasil.
NomeFantasiaRFB	Nome fantasia do fornecedor na base de dados da RFB – Receita Federal do Brasil.
CNAEPrincipal	Código identificador da Classificação Nacional de Atividades Econômicas principal do fornecedor.
CodigoProblema	Código identificador do problema
DescricaoProblema	Descrição dos problemas do SINDEC.
CEPConsumidor	Código identificador do CEP do consumidor.

- c) Mudança de tipo: Durante essa etapa, foi realizada a alteração do tipo dos dados referente a seis colunas;
- d) Ajuste no conteúdo: Alteração de conteúdo da coluna “Atendida” que indica o status de atendimento da Reclamação. Os caracteres “S” e “N” foram substituídos pela palavra “Sim” e “Não”, respectivamente.

Após esse tratamento, foi possível verificar que houve mudança no resultado da análise de qualidade da coluna realizada automaticamente pelo Power Query Editor. Todas as colunas ainda presentes no banco de dados apresentavam boa qualidade e continham as informações de qualidade de dados igual o mostrado na Figura 4.

Figura 4 - Análise de qualidade da coluna do Power Query.

1 ² 3 AnoCalendario	DataArquivamento	DataAbertura	A ^B C CodigoRegiao
● Válidos 100%	● Válidos 100%	● Válidos 100%	● Válidos 100%
● Erro 0%	● Erro 0%	● Erro 0%	● Erro 0%
● Vazio 0%	● Vazio 0%	● Vazio 0%	● Vazio 0%

A realização desses passos citados, podem ser visualizados na Figura através da linguagem M que foi gerada.

Figura 5 - Código em linguagem M para geração da base de dados tratada

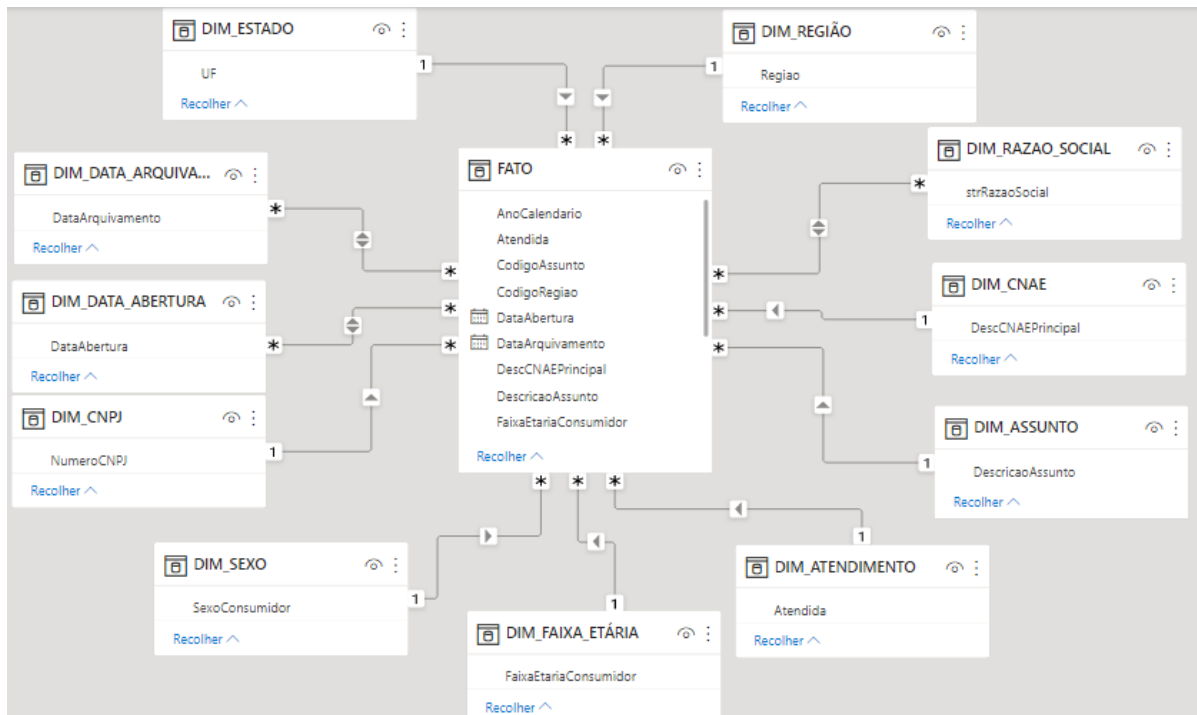
```

1 let
2     Fonte = Table.Combine({CNRF2018, CNRF2019, CNRF2020}),
3     #"Colunas Removidas" = Table.RemoveColumns(Fonte,{"", "_1"}),
4     #"Linhas Filtradas" = Table.SelectRows(#"Colunas Removidas", each [DataAbertura] <> null and [DataAbertura] <> ""),
5     #"Colunas Removidas1" = Table.RemoveColumns(#"Linhas Filtradas",{"strNomeFantasia", "RadicalCNPJ", "RazaoSocialRFB", "NomeFantasiaRFB",
6         "CNAEPrincipal", "CodigoProblema", "DescricaoProblema", "CEPConsumidor"}),
7     #"Linhas Filtradas1" = Table.SelectRows(#"Colunas Removidas1", each [DescCNAEPrincipal] <> null and [DescCNAEPrincipal] <> ""),
8     #"Tipo Alterado" = Table.TransformColumnTypes(#"Linhas Filtradas1",{("AnoCalendario", Int64.Type), ("DataArquivamento", type datetime),
9         ("DataAbertura", type datetime), ("CodigoRegiao", Int64.Type), ("Tipo", type text)}),
10    #"Valor Substituido" = Table.ReplaceValue(#"Tipo Alterado", "1", "CNPJ", Replacer.ReplaceText, {"Tipo"}),
11    #"Valor Substituido1" = Table.ReplaceValue(#"Valor Substituido", "0", "CPF", Replacer.ReplaceText, {"Tipo"}),
12    #"Linhas Filtradas2" = Table.SelectRows(#"Valor Substituido1", each [NumeroCNPJ] <> null and [NumeroCNPJ] <> ""),
13    #"Tipo Alterado2" = Table.TransformColumnTypes(#"Linhas Filtradas2",{("NumeroCNPJ", type text)}),
14    #"Valor Substituido2" = Table.ReplaceValue(#"Tipo Alterado2", "S", "Sim", Replacer.ReplaceText, {"Atendida"}),
15    #"Valor Substituido3" = Table.ReplaceValue(#"Valor Substituido2", "N", "Não", Replacer.ReplaceText, {"Atendida"}),
16    #"Tipo Alterado3" = Table.TransformColumnTypes(#"Valor Substituido3",{("CodigoAssunto", Int64.Type)}),
17    #"Linhas Filtradas3" = Table.SelectRows(#"Tipo Alterado3", each not Text.Contains([DescCNAEPrincipal], "NULL")),
18    #"Linhas Filtradas4" = Table.SelectRows(#"Linhas Filtradas3", each ([Atendida] = "Não" or [Atendida] = "Sim")),
19    #"Tipo Alterado1" = Table.TransformColumnTypes(#"Linhas Filtradas4",{("CodigoRegiao", type text), ("CodigoAssunto", type text)})
20 in
21     #"Tipo Alterado1"

```

Em seguida, utilizando a base consolidada e tratada, foram extraídas todas as dimensões da base de dados. Vale ressaltar que para a criação das tabelas dimensões, foi requerida a retirada de elementos duplicados, a fim de que se cumprisse o objetivo de otimização no processo de consulta proposto pela modelagem dimensional utilizada. Todavia, algumas tabelas dimensões não tiveram suas duplicidades retiradas devido ao contexto em que estava inserido. Após a geração dessas tabelas de dimensões, foi realizada a carga para o ambiente do Power BI conforme visto na Figura 6.

Figura 6 - Relacionamento gerado no Power BI



6. Links

https://github.com/lennonbfn/Projetos-Academico/tree/main/TCC_PucMinas

REFERÊNCIAS

- [1] lei 8.078/90 e no Decreto 2.181/97
- [2] Relatório do Cadastro Nacional de Reclamações Fundamentadas 2012
- [3] <https://dados.gov.br/dataset/cadastro-nacional-de-reclamacoes-fundamentadas-procons-sindec1>
- [4] <https://sindecnacional.mj.gov.br/report/Mapa>
- [5] <https://hevodata.com/learn/dimensional-data-modelling/>