

# **IMPLEMENTANDO O PROCESSO DE ETL (*EXTRACT, TRANSFORM AND LOAD*) PARA A ANÁLISE DE VARIÁVEIS PERTINENTES A UM *DATASET* ORIUNDO DA PLATAFORMA DE DADOS ABERTOS DO GOVERNO FEDERAL NO QUE TANGE AS RECLAMAÇÕES REALIZADAS ENTRE OS ANOS DE 2017 E 2018**

**Rodrigo Silva Rocha<sup>1</sup>, Geraldo Henrique Neto<sup>1</sup>,**

<sup>1</sup>Faculdade de Tecnologia de FATEC Ribeirão Preto (FATEC)  
Ribeirão Preto, SP – Brasil

silva.rodrigo31@gmail.com, Geraldo.henriqueneto@gmail.com

***Resumo.** Dados abertos do governo são publicações e divulgações de informações do setor público na web, compartilhadas em formatos bruto e aberto, com o objetivo de promover a interpretação por aplicações digitais desenvolvidas. Este trabalho apresenta como objeto, elucidar a evolução do quadro de reclamações, de construção de dashboard, visando a fácil elucidação de resultados.*

***Abstract.** General government data are publications and disclosures of public sector information on the web, shared in raw and open formats, with the aim of promoting interpretation by developed applications. This work has as its object, to elucidate the evolution of the complaints framework, through the construction of a technical panel and the easy elucidation of results.*

## **1 Introdução**

Dentro da segmentação de mercado, para a avaliação do consumidor, quando não é atendido, ele procura os órgãos oficiais para ajuda, sabendo disso, as empresas procuram analisar esses dados, tendo em vista que, se tiver o índice alto, pode ocorrer a imagem ficar como péssima e, conseqüentemente, o nível de atração ser baixo.

De acordo com Veiga e Guimarães (2015), dados abertos do governo são publicações e divulgações de informações do setor público na web, compartilhadas em formatos bruto e aberto, com o objetivo de promover a interpretação por aplicações digitais desenvolvidas. Outra definição apresentada, são dados em que a sociedade pode utilizar sem restrição, dando os créditos ao autor, compartilhando a mesma licença.

O acesso aos dados abertos do governo, está cada vez mais disponível para a população em geral, contudo, mesmo com o grau de facilidade de obtenção, não são dados simples de serem interpretados, como por exemplo a estrutura de como está organizado de acordo com a nomenclatura, que pode se tornar de difícil compreensão no primeiro contato (PIZANI; SIQUEIRA, 2018).

O termo *Business Intelligence* (BI), ou Inteligência de Negócio pode ser definido como a utilização de diversas fontes de informação para a definição de estratégias de competição nos negócios da empresa. O objetivo principal de BI, consiste na definição

de regras e de ferramentas para a limpeza e transformação do volume de dados, para transformar em depósitos estruturados, independentemente de sua origem (REIS, 2018). Neste trabalho será analisado o índice de reclamações, com base no seu setor e estado de origem, identificando o percentual de reclamações atendidas.

Os *datasets*, em formato *csv* (*comma-separated-values*), foram obtidos do Cadastro Nacional de Reclamações Fundamentadas (PROCONS - SINDEC). O SINDEC é um sistema informatizado que integra processos e procedimentos, oriundo de atendimento aos consumidores dos Procons, com o objetivo de proporcionar uma ferramenta de gestão adequado ao dinamismo, que é típico de seus setores de atendimento (SENACON/MJ, 2020).

Os *datasets*, são resultados de um trabalho integrado, com o objetivo de ampliar a voz de milhões de consumidores em todo o Brasil. O SINDEC integra 26 Procons estaduais e 351 Procons municipais. Como vários desses Procons contam com mais de uma unidade, o sistema opera em 675 unidades espalhadas por 448 cidades brasileiras. A média de atendimento dos Procons é de 216 mil consumidores (SENACON/MJ, 2020).

De acordo com SINDEC, demanda é definida como os tipos de atendimentos realizados pelo Procon. Estas demandas são classificadas como procedimentos mais céleres de atendimento, realizados por telefone, carta encaminhada ao fornecedor, desses processos administrativos instaurados, na definição do SINDEC, é classificado como Reclamação (SENACON/MJ, 2020).

O objetivo desse trabalho é fazer um levantamento de reclamações, usando as entidades empresas, estado, cnae, para identificar a evolução das reclamações para algum tipo de ação emergencial.

Este trabalho é estruturado da seguinte forma: na seção 2, serão apresentados a fundamentação teórica, a seção 3, a apresentação da metodologia, a seção 4 os resultados e discursões, será apresentado a carga de dados e os dashboards e a seção 5, a conclusão desse trabalho.

## **2 Referencial Teórico**

### **2.1 ETL**

No processo de ETL (*extract-transform-load*), as etapas de extração e transformação dos dados compreendem os procedimentos de limpeza, conversões de formatos, extração de atributos (DOS SANTOS PINTO, 2015). De Oliveira (2015) ressalta que o ETL consome cerca de um terço do orçamento em um projeto de desenvolvimento de um *Data Warehouse*.

A extração de dados ocorre usando regras ou tabelas de referência, ou então, realizando combinação de dados com outros dados. As funções de base de dados são integradas em uma ferramenta para capturar dados de uma ou mais base de dados, inserindo em uma base proposta ou em um *data warehouse* (SHARDA; DELEN; TURBAN, 2019).

Após a extração de dados, é iniciada a transformação de dados, podendo destacar cálculo de métricas derivadas, realização de limpeza de dados, unificação, seleção das colunas necessárias e realização de *split* de dados. Após, é iniciada a carga de dados, inicialmente em uma *staging* área e, posteriormente, no modelo dimensional (PITON, 2018).

Os dados usados em processo de ETL podem vir de qualquer fonte, dentre elas: um aplicativo de *mainframe*, um aplicativo de ERP (*Enterprise Resource Planning*), uma ferramenta de CRM (*Customer Relationship Management*), um arquivo plano, uma planilha do Excel ou até mesmo de uma fila de mensagens (SHARDA; DELEN; TURBAN, 2019).

Mesmo que seja possível para programadores desenvolver *software* para ETL, é mais simples utilizar ferramentas já disponíveis. Alguns critérios que devem ser levados em conta para a escolha são: Capacidade de ler e transcrever entre um número ilimitado de arquiteturas de fontes de dados; Captura e entrega automática de metadados, um histórico de conformidade com padrões abertos e uma interface fácil de usar para o desenvolvedor e o usuário funcional (SHARDA; DELEN; TURBAN, 2019).

## 2.2 *Staging Area*

*Staging Area* é um banco de dados relacional sem nenhum relacionamento, onde são guardados os dados não tratados, evitando usar o servidor de origem, com o risco de ficar lento (PITON, 2018). O *Staging Area* é uma camada intermediária entre a fonte de dados e do *Data Warehouse*, não sendo ligado ao aplicativo final. A estruturação de dados simples e definida é requerida para um desempenho aceitável, reduzindo o risco do processo de se tornar caótico e, posteriormente, atrapalhar o desenvolvedor responsável (REIS, 2018).

Quando os dados chegam na *Staging Area*, são processadas informações como filtragem de dados (correção de erros de digitação, solução de conflitos de domínio, tratamento de elementos ausentes ou a divisão de formatos padrão, combinação de dados de várias origens, cancelamento de dados duplicados e atribuição de chaves de *warehouse*). Na maioria das situações, ela não se baseia na tecnologia relacional, podendo ser formada por um sistema de arquivos simples. Antes da etapa de extração, análise sobre os dados operacionais deve ser realizada a fim de identificar as necessidades de transformações (MONTEIRO; PINTO; ROSA MARIA E. MOREIRA DA, 2004).

Após a etapa de transformação ser concluída, os dados são carregados no *Data Warehouse* em um processo chamado de carga de dados. Esta é a última operação sobre os dados da *Staging Area*, no entanto, é importante destacar que os dados devem estar pronto, para o modelo de *Data Warehouse* proposto (OGRODOVSKI, 2013).

## 2.3 *Data Warehouse*

*Data Warehouse* (DW), ou armazém de dados, é definido como uma coleção de bancos de dados projetada para fornecer informações necessárias e apoiar as tomadas de decisão. A integração baseada em DW surgiu para tentar solucionar as deficiências da integração ponto a ponto. No modelo, os dados de determinada fonte são extraídos, transformados e

inserindo no DW, depois são colocados nos repositórios de dados (NETO, 2014).

Esse tipo de organização deve destacar as métricas importantes e relevantes dos dados corporativos, sendo necessário o provimento da possibilidade e flexibilidade de como os dados serão interpretados e analisados posteriormente, produzindo uma fonte estável de informações históricas, de uma forma consistente e confiável para qualquer consumidor autorizado (PIZANI; SIQUEIRA, 2018).

O conteúdo de um DW deve ser de fácil entendimento, intuitivo e obvio para o usuário de negócios, facilitando o acesso à informação. Também deve ser crível, montado, organizado e limpo, apresentando consistência com as definições realizadas pela organização (BRITO; PITON, 2017).

O *Data Warehouse* oferece uma solução de armazenamento de dados centralizados, possuindo grande volume de informações históricas, otimizado para operações de leitura e confiável para a tomada de decisão. As necessidades do negócio deverão ser a maior prioridade no projeto, captando as regras de negócio em um modelagem adequada (BRITO; PITON, 2017).

## **2.4 Modelagem Dimensional**

Modelagem dimensional é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios. É utilizada para sumarizar e reestruturar dados e apresenta-los em visões que suportem a análise dos valores desses dados (REIS, 2018).

O processo de modelagem dimensional deve se iniciar com a definição do processo do negócio. Para a definição do processo, será necessário ouvir os usuários do sistema atual, verificar quais indicadores de desempenho são analisados, quais relatórios são necessários e os objetivos nos quais são utilizados esses indicadores. Um dos exemplos de processos são compra de materiais, pedidos e estoque (VIEIRA, 2012).

Em um modelo dimensional, a tabela fato representa uma medição de negócio, é também, a principal tabela de um modelo dimensional em que as medições numéricas de desempenho da empresa estão armazenadas. A fato representa um item, ou um evento de negócio que é utilizado para analisar o processo de negócio da empresa (REIS, 2018). Essa tabela correspondem a cerca de 90% do volume total de um DW, suas chaves primárias, são compostas de todas as chaves estrangeiras, ou de atributos que garantem a unicidade do registro (BRITO; PITON, 2017).

As tabelas referentes às dimensões possuem elementos que participam da tabela fato, determinando um contexto do assunto, contendo as características do evento (PIZANI; SIQUEIRA, 2018). As dimensões possuem os atributos descritivos usados pelo BI, filtrando e agrupando os fatos. As dimensões podem ser consideradas a alma do DW, pois só com elas é possível dar sentido aos dados, criando análises importantes para o negócio, sendo impulsores da experiência do BI com um todo para os seus usuários, independente de que sejam (PIZANI; SIQUEIRA, 2018).

### 3 Metodologia

Este trabalho visa explorar os arquivos CSV obtido no site de dados abertos do governo (SENA/MJ, 2020) para a análise de reclamações do período de 2017-2018, com o uso de ferramentas de BI. Os principais fatores a serem investigados são (foram levantados 30 maiores registros para cada item pertinente):

- Grupo de setores com maior reclamação;
- Contagem de reclamações atendidas por empresas;
- Evolução do quadro de reclamações por trimestre por empresas;
- Estado com maior índice de reclamação e o assunto reclamado.
- Estado e número de assuntos reclamado.

As seguintes tecnologias foram adotadas para o desenvolvimento proposto neste trabalho:

- *SQL Power Architect* - para o desenho do diagrama dimensional;
- *Pentaho Data Integration* - para o desenvolvimento de processo de ETL;
- PostgreSQL - para o gerenciamento do banco de dados relacional;
- *Microsoft Power BI Desktop* - para a elaboração dos *dashboards*.

*SQL Power Architect* foi criado por *designers* de DW e tem muitos recursos únicos a serem usados pelos *designers*. A ferramenta permite que os usuários façam engenharia reversa de banco de dados (BI, 2020).

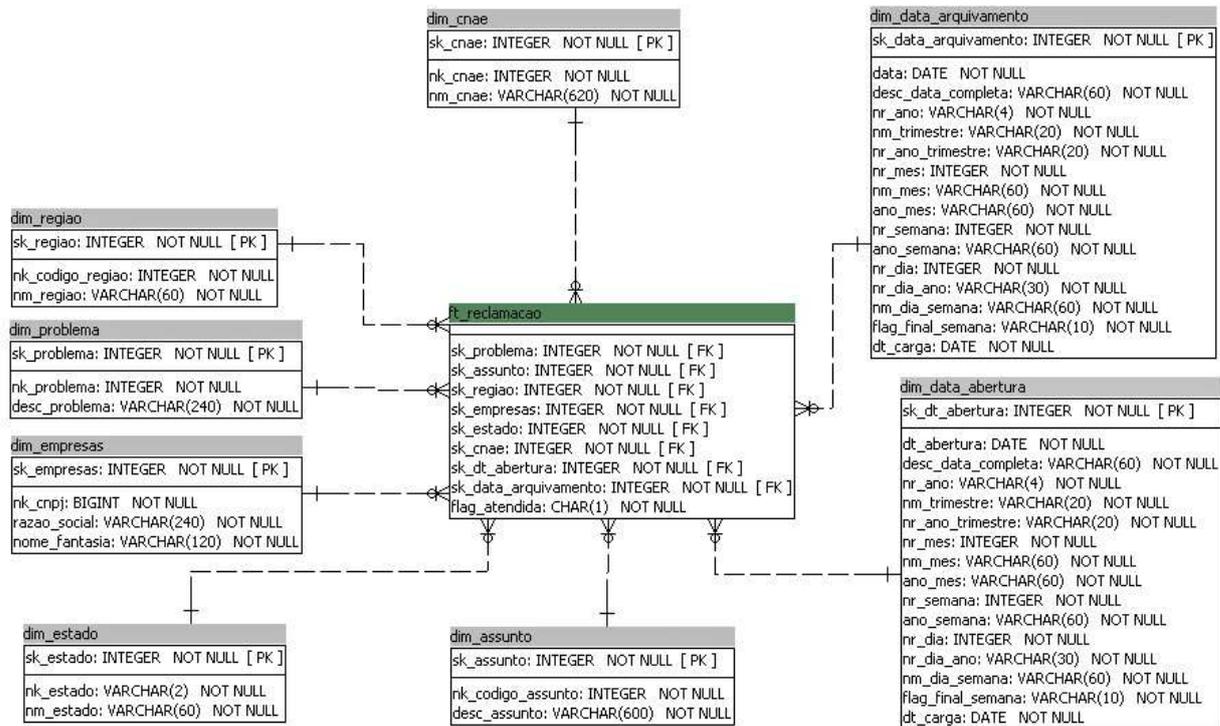
O PostgreSQL é um sistema gerenciador de banco de dados objeto relacional de código aberto, que usa a linguagem SQL combinada com muitos recursos que armazenam e escalam com segurança. Ganhou uma forte reputação por sua arquitetura comprovada, confiabilidade, integridade de dados, conjuntos robustos de recursos, extensibilidade e dedicação da comunidade de código aberto para fornecer soluções inovadoras e de alto desempenho (POSTRGRESQL: ABOUT, 2020).

O *Pentaho Data Integration (Kettle)* é uma ferramenta para transformações de dados. Destacam-se as seguintes funcionalidades: migração de dados, conversões de sistemas ou carga de banco de dados, transformação de dados, além de tarefas avançadas como transferência de arquivos, envio de e-mails, integrações com *webservices* e programações com *scripts* com *schedules* (VANTARA, 2019).

O *Power BI Desktop* é um aplicativo gratuito que pode ser instalado no computador local e que permite conexão aos dados. Permite também a criação de modelos visuais como relatórios para a realização de diversas análises (POWER BI, 2019).

### 3.1 Regras de Negócio e Modelagem Dimensional

Regras de negócios são leis que ditam o negócio, e mostram como devem funcionar seus processos. É a forma o qual o comportamento do negócio é seguido mediante a itens como legislação, regulamentações e padrões de mercado. Foram definidas as seguintes dimensões e tabela fato para o problema analisado, conforme pode ser visualizado na Figura 1:



**Figura 1: Modelagem Dimensional proposta**  
Fonte: autoria própria

- Dimensão Região - Representada pela tabela dm\_regiao, armazena dados da região (nm\_regiao) e a chave nk\_codigo\_regiao que identifica o código da região do procon.
- Dimensão Problema - Representada pela tabela dim\_problema, armazena dados das reclamações, e possui as colunas: nk\_problema (é o código da descrição do problema) e desc\_problema (é a descrição dos problemas do SINDEC).
- Dimensão Empresas - Representada pela tabela dim\_empresas, armazena os dados das empresas, e possui as colunas: nk\_cnpj (é o número do CNPJ e CPF); razao\_social (é a razão social do fornecedor da base de dados da receita federal do Brasil) e nome\_fantasia (é a identificação do fornecedor da base de dados).
- Dimensão Estado - Representada pela tabela dim\_estado, armazena os dados dos estados. Possui as colunas: nk\_estados (são as siglas dos estados) e nm\_estado (são os nomes dos estados).
- Dimensão Assunto - Representada pela tabela dim\_estado, armazena os dados dos

assuntos do SINDEC. Possui as colunas: nk\_codigo\_assunto (esse campo irá ser representado pelos códigos representados nas respectivas bases) e desc\_assunto (é o assunto listado pelo SINDEC).

- Dimensão Data de Abertura - Representada pela tabela dim\_data\_abertura, armazena a data de abertura das reclamações. Possui as colunas: dt\_abertura (é data de abertura do processo, ex: 01/01/2017); desc\_data\_completa (é a sintaxe da data completa ex: 01 de Janeiro de 2017); nr\_ano (é o ano da abertura da reclamação, ex: 2017); o número do trimestre, representa do trimestre do ano (ex: 1T); nr\_ano\_trimestre (representa o trimestre pertencente aquele ano, ex: 2017/1T); nr\_mes (representa o número do mês, ex: 01); nm\_mes (representa o mês por extenso, ex: Janeiro); ano\_mes (representa o ano e o mês juntos, ex: 2017/01); nr\_semana (representa o número na semana no ano, ex: 34); ano\_semana (representa o número do ano e semana juntos, ex: 2017/34); nr\_dia (representa o número do dia do mês, ex: 31); nr\_dia\_ano (representa o número do dia junto com o ano em questão, ex: 2017/31); nm\_dia\_semana (representa o dia da semana por extenso, ex: quinta-feira); flag\_final\_semana (recebe se o dia é final de semana, ex: sim ou não) e dt\_carga (representa a data de entrada dos dados no DW).
- Dimensão Data de Arquivamento - Representada pela tabela dim\_data\_arquivamento, armazena a data de arquivamento das reclamações. Possui as colunas: dt\_arquivamento (é a data de arquivamento do processo, ex: 01/01/2017); desc\_data\_completa (é a sintaxe a data completa ex: 01 de Janeiro de 2017); nr\_ano (é o ano da abertura da reclamação, ex: 2017); o número do trimestre (representa o trimestre do ano, ex: 1T); nr\_ano\_trimestre (representa o trimestre pertencente aquele ano, ex: 2017/1T); nr\_mes (representa o número do mês, ex: 01); nm\_mes (representa o mês por extenso, ex: Janeiro); ano\_mes (representa o ano e o mês juntos, ex: 2017/01); nr\_semana (representa o número na semana no ano, ex: 34); ano\_semana (representa o número do ano e semana juntos, ex: 2017/34); nr\_dia (representa o número do dia do mês, ex: 31); nr\_dia\_ano (representa o número do dia junto com o ano em questão, ex: 2017/31); nm\_dia\_semana (representa o dia da semana por extenso, ex: quinta-feira); flag\_final\_semana (recebe se o dia é final de semana, ex: sim ou não); e dt\_carga (representa a data de entrada dos dados no DW).
- Dimensão CNAE - Representada pela tabela dim\_cnae, armazena os dados da classificação nacional das atividades econômicas principal do fornecedor. possui as colunas: nk\_cnae, (é código cadastrado) e nm\_cnae (recebe a descrição da classificação). Será utilizada a API CNAE - Cadastro Nacional de Atividades Econômicas (2017) para complemento desses dados.
- Fato Reclamação - Representada pela tabela ft\_reclamação, faz ligação com todas as dimensões do modelo, utilizando a *surrogate key*, se comportando como chave estrangeira e possui a medida do valor da reclamação (flag\_atendida).

### 3.2 Requisitos Funcionais

Requisitos funcionais são as funcionalidades que devem ser criadas para permitir que os usuários realizem suas tarefas ou para satisfazer requisitos de negócio, definindo o comportamento do sistema, expresso na forma de serviços ou funções. Foram levantados os seguintes requisitos funcionais:

- RF 1: O sistema deve importar 2 arquivos CSV do portal de dados abertos;
- RF 2: O sistema deve converter os arquivos CSV em tabelas;
- RF 3: O sistema deve acrescentar, na dimensão tempo, os campos de semestre, trimestre e mês;

### 3.3 RF 4: O sistema não deve permitir inserção de dados pelo usuário final. Requisitos não-funcionais

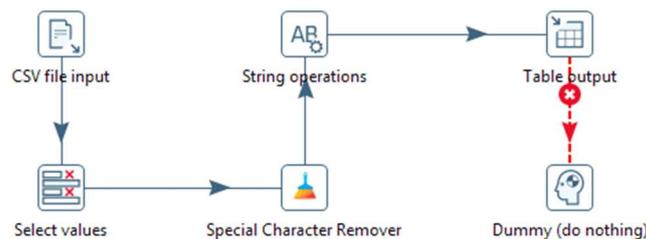
Os requisitos não funcionais tratam de restrições, enfrentadas pelos sistemas, podendo ser de *timing*, processo de desenvolvimento e padrões. Foram levantados os seguintes requisitos não-funcionais:

- RNF 1: O Sistema deve utilizar, para a carga dos dados, o *Pentaho Integration 8.3*;
- RNF 2: O sistema deve utilizar o PostgreSQL como serviço de banco de dados;
- RNF 3: O sistema deve utilizar como ferramenta visual o Power B.I.

## 4 Resultados

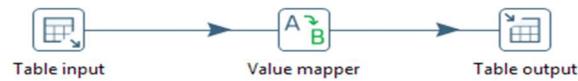
### 4.1 Carga de dados na *Staging Area*

A tabela armazém foi criada para armazenar as informações dos arquivos CNRF2018.csv e CNRF\_2017.csv, disponibilizados pelo site Cadastro Nacional de Reclamações Fundamentadas. Na Figura 2 está representado o início do processo: a extração dos dados dos arquivos csv's para a tabela armazém.



**Figura 2: Início da carga para a tabela Armazém**  
Fonte: autoria própria

Após a migração de dados para a tabela armazém, foi realizada a carga para suas respectivas tabelas no *schema staging*. A transformação nesse caso constituiu-se de um *input* e *output* para o banco *staging* da entidade estado, como pode ser observado na Figura 3.



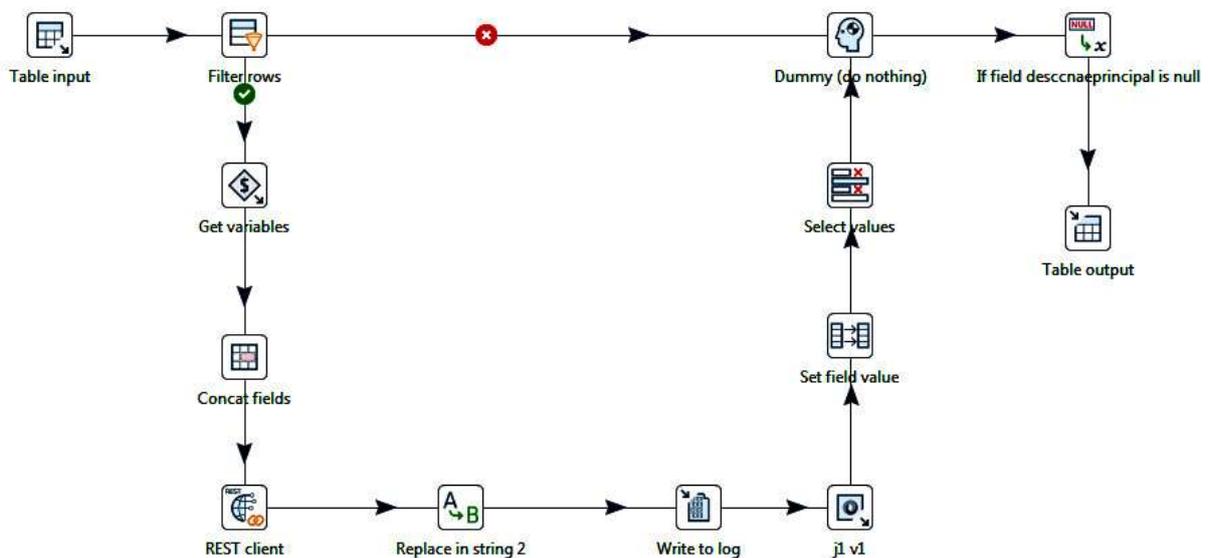
**Figura 3: Transformação staging para a entidade estado**  
**Fonte: autoria própria**

## 4.2 Carga de dados para o DW

Após a preparação da modelagem dimensional, foi executada a carga para o esquema desenvolvido. As transformações das entidades região, estado, assunto, data de abertura, data de fechamento e problema foram constituídas, também, de um *input* e *output*.

### 4.2.1. Carga da dimensão CNAE

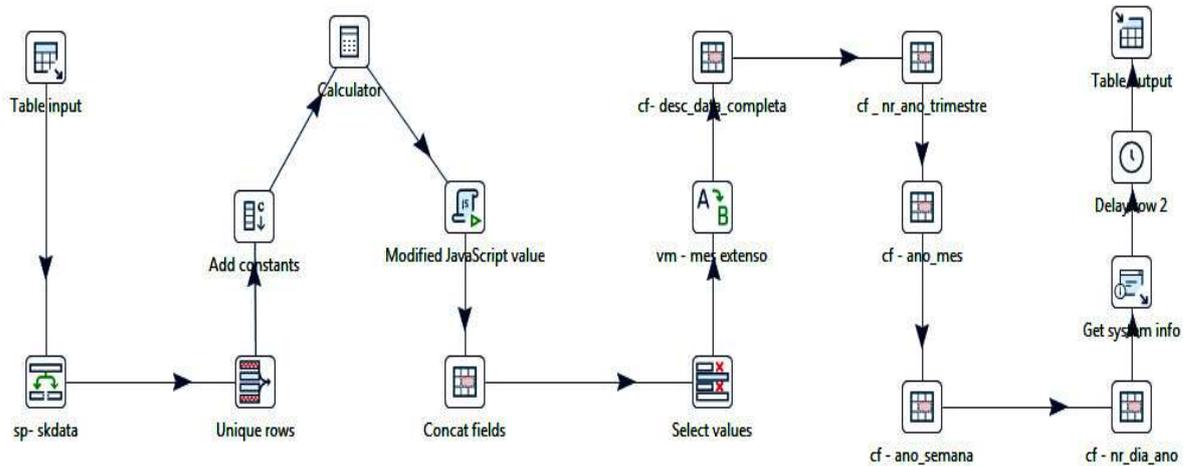
A carga da dimensão CNAE, como demonstrado na Figura 4, foi executada da seguinte forma: caso não encontrasse a descrição “descnaeprincipal” na tabela de *staging*, o preenchimento seria por meio da chamada da api, pelo método GET, pelo o endereço <https://servicodados.ibge.gov.br/api/v2/cnae/subclasses/>, concatenado com o respectivo id, retornando um json, e, por meio desse json seria extraída a descrição do CNAE utilizando o *step json input* (j1 v1), e então, ocorreria a inserção no DW.



**Figura 4: Carga da dimensão CNAE**  
**Fonte: autoria própria**

### 4.2.2. Carga da dimensão data de abertura e data de fechamento

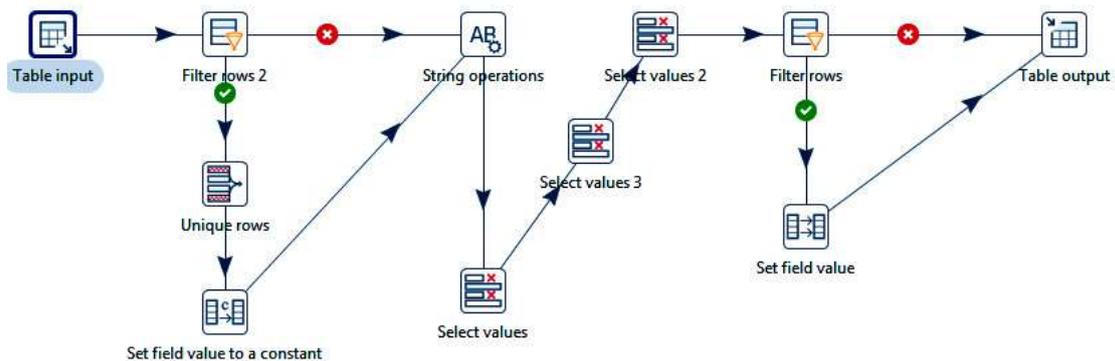
A carga da dimensão de data, mostrada na Figura 5, consistiu em transformar a data em vários atributos, como por exemplo, ano, ano/mês, ano/trimestre, mês por extenso, destacando-se para a formação da *surrogate key*, formada por exemplo: 2010/01/01.



**Figura 5: Transformação de base para as dimensões data de abertura e fechamento**  
 Fonte: autoria própria

#### 4.2.3. Carga da dimensão empresas

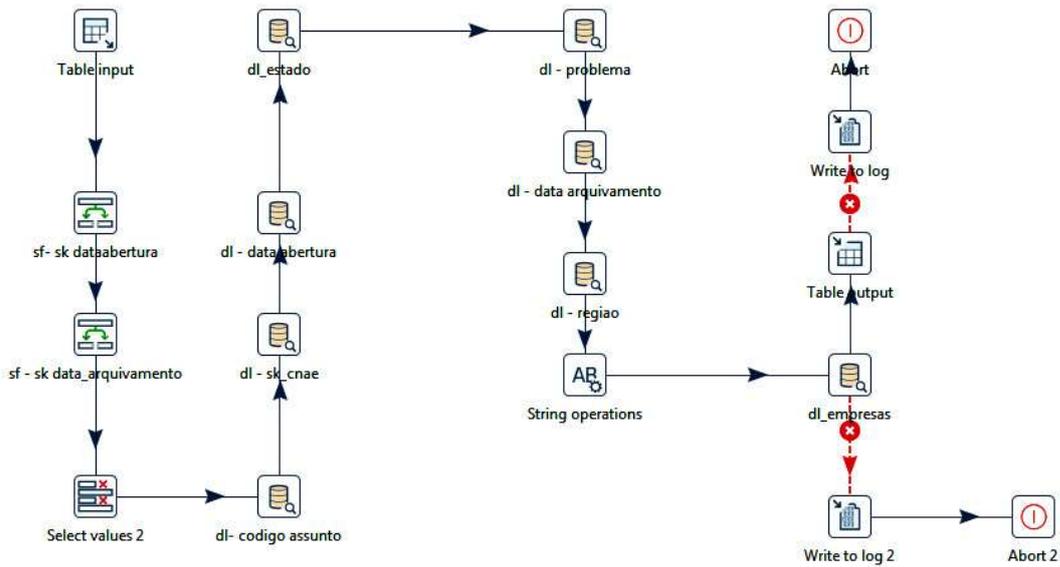
A transformação representada pela Figura 6 foi realizada da seguinte forma: caso uma empresa não tivesse o CNPJ cadastrado, receberia um valor padrão de CNPJ não cadastrado juntamente com o nome fantasia e, caso esse não estivesse preenchido, receberia também um valor padrão.



**Figura 6: Transformação base para a entidade empresas**  
 Fonte: autoria própria

#### 4.2.4. Carga da fato reclamação

Após a finalização da carga dimensões, foi iniciada a carga na tabela fato, como demonstrado na Figura 7. A carga foi feita da seguinte forma: foi passada a chave primária no *step database lookup*, após o retorno na sua respectiva tabela dimensional, foi executada a inserção na tabela fato.

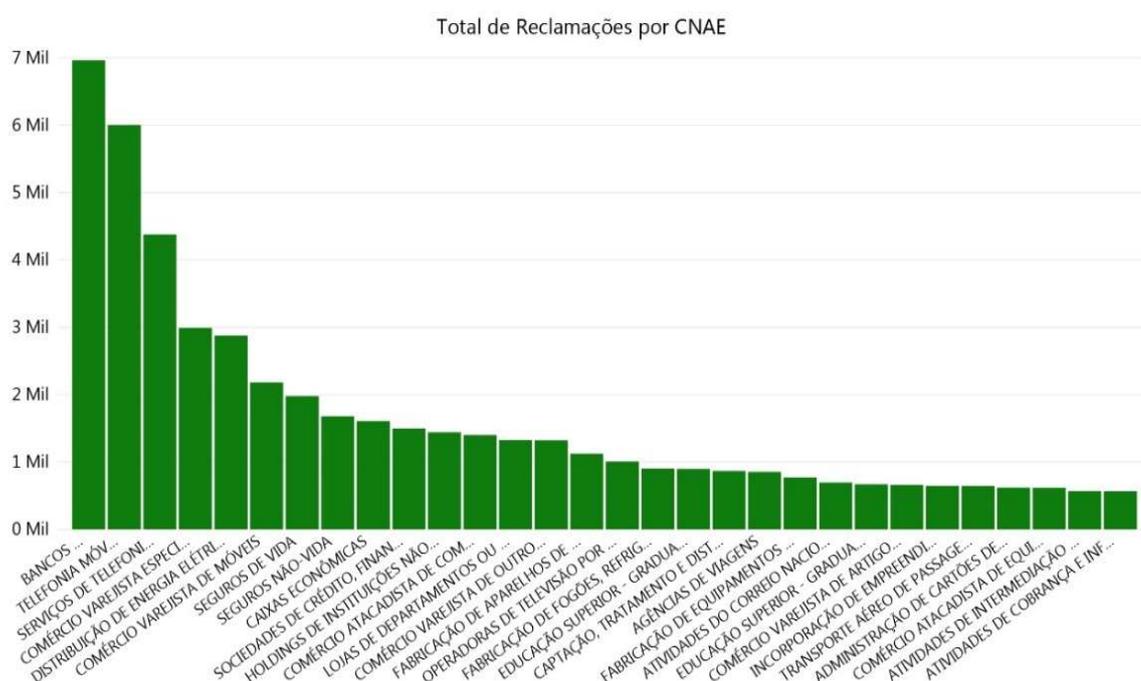


**Figura 7: Fato Reclamação**  
 Fonte: autoria própria

### 4.3 Interpretação dos *dashboards*

#### 4.3.1. Gráfico de CNAES com o maior índice de reclamações

De acordo com o gráfico de colunas clusterizados (Figura 8), foi possível verificar que as CNAES com o maior índice de reclamações são as relacionadas aos setores bancários e telefonia móvel. Isso é demonstrado como por exemplo, alguma solicitação que não é atendida no tempo determinado pela lei, da mesma forma que, essas reclamações, acabam acumulando nas suas respectivas agências reguladoras.



**Figura 8: Total de reclamações por CNAE**

Fonte: autoria própria

### 4.3.2. Tabela de total de reclamações atendidas

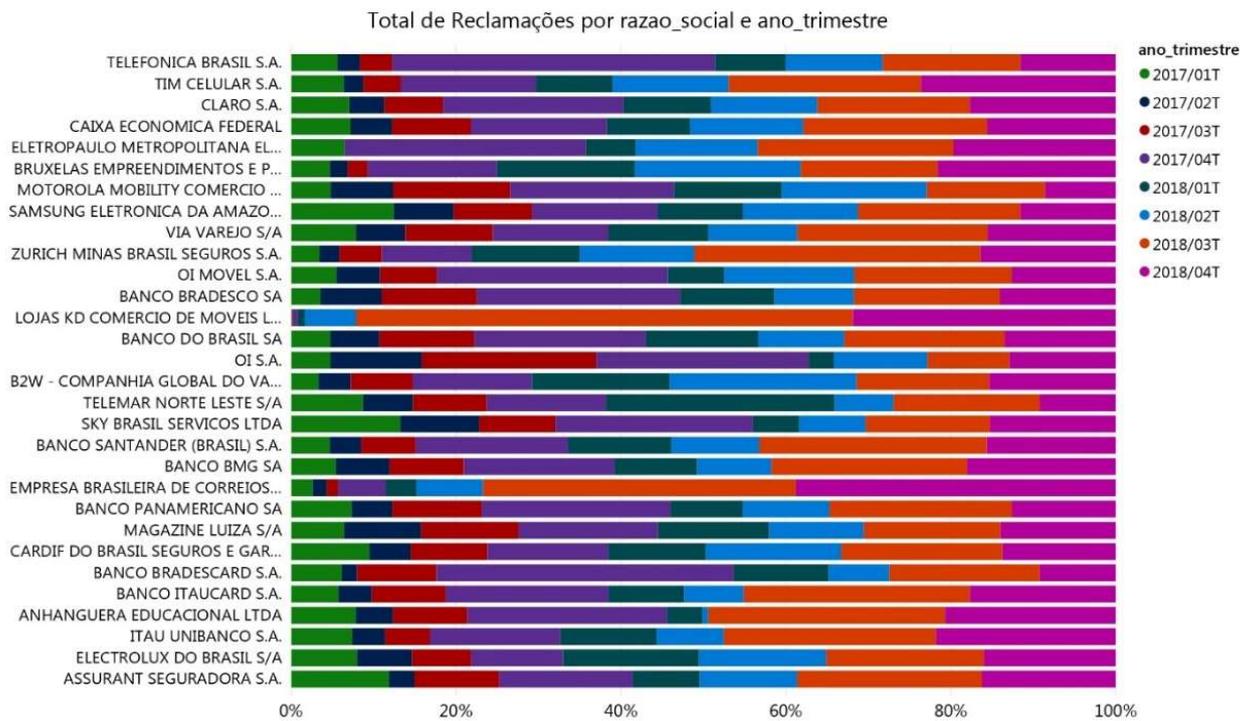
De acordo com a Tabela 1, as empresas, que atenderam a mais reclamações foram as dos setores de telefonia, seguidas de empresas de setores bancários. As empresas que tiveram com o menor desempenho, são do setor bancário e elétricos, juntamente com os serviços de entrega.

razao_social	Total de Reclamações Atendidas
TIM CELULAR S.A.	1872
CLARO S.A.	1474
TELEFONICA BRASIL S.A.	1391
BRUXELAS EMPREENDIMENTOS E PARTICIPACOES S/A	1059
CAIXA ECONOMICA FEDERAL	918
MOTOROLA MOBILITY COMERCIO DE PRODUTOS ELETRONICOS LTDA	849
OI MOVEL S.A.	845
SAMSUNG ELETRONICA DA AMAZONIA LTDA	784
OI S.A.	744
VIA VAREJO S/A	688
TELEMAR NORTE LESTE S/A	668
BANCO BRADESCO SA	579
B2W - COMPANHIA GLOBAL DO VAREJO	544
SKY BRASIL SERVICOS LTDA	487
MAGAZINE LUIZA S/A	446
BANCO DO BRASIL SA	375
ZURICH MINAS BRASIL SEGUROS S.A.	361
ELETRIPAULO METROPOLITANA ELETRICIDADE DE SAO PAULO S.A.	359
BANCO BRADESCARD S.A.	348
BANCO SANTANDER (BRASIL) S.A.	330
ELECTROLUX DO BRASIL S/A	321
CARDIF DO BRASIL SEGUROS E GARANTIAS S/A	313
BANCO BMG SA	292
CENTRAIS ELETRICAS DO PARA S.A. - CELPA	249
BANCO ITAUCARD S.A.	248
WHIRLPOOL S.A	242
EMPRESA BRASILEIRA DE CORREIOS E TELEGRAFOS	235
<b>Total</b>	<b>17675</b>

**Tabela 1: Tabela com o total de reclamações atendidas**  
**Fonte: autoria própria**

### 4.3.3. Gráfico de evolução de total de recebimentos de reclamação por razão social e empresas por trimestre

Na Figura 10 está mostrado o gráfico de barras que mostra a evolução de total de recebimentos de reclamação por razão social e empresas. Observa-se que o 2º trimestre de 2018 foi o que apresentou o menor índice de reclamações, comparado ao quatro trimestre deste mesmo ano. É importante destacar que as empresas dos setores de telefonia e bancários foram as que apresentaram os maiores índices de reclamações no 4º trimestre de 2017.



**Figura 10: Gráfico de barras para a evolução de total de reclamações por trimestre**  
**Fonte: autoria própria**

#### 4.3.4. Gráfico com o total de reclamações por estado

De acordo com a Figura 11, na qual está apresentado o gráfico com o total de reclamações por estado, observa-se que o estado com o maior índice de reclamação foi São Paulo e o menor índice, o estado de Tocantins.

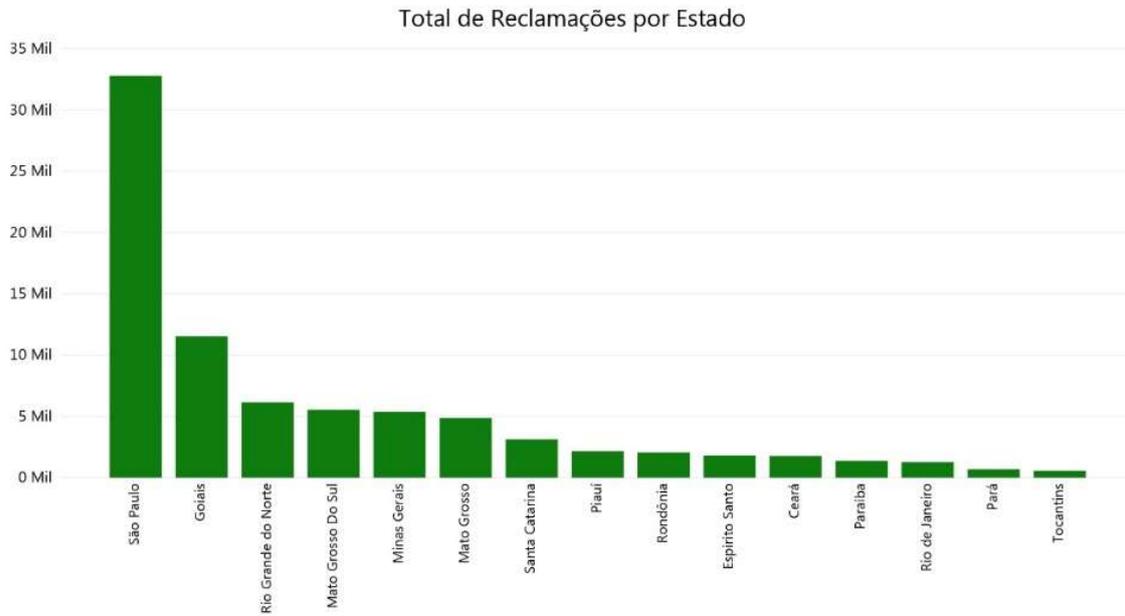
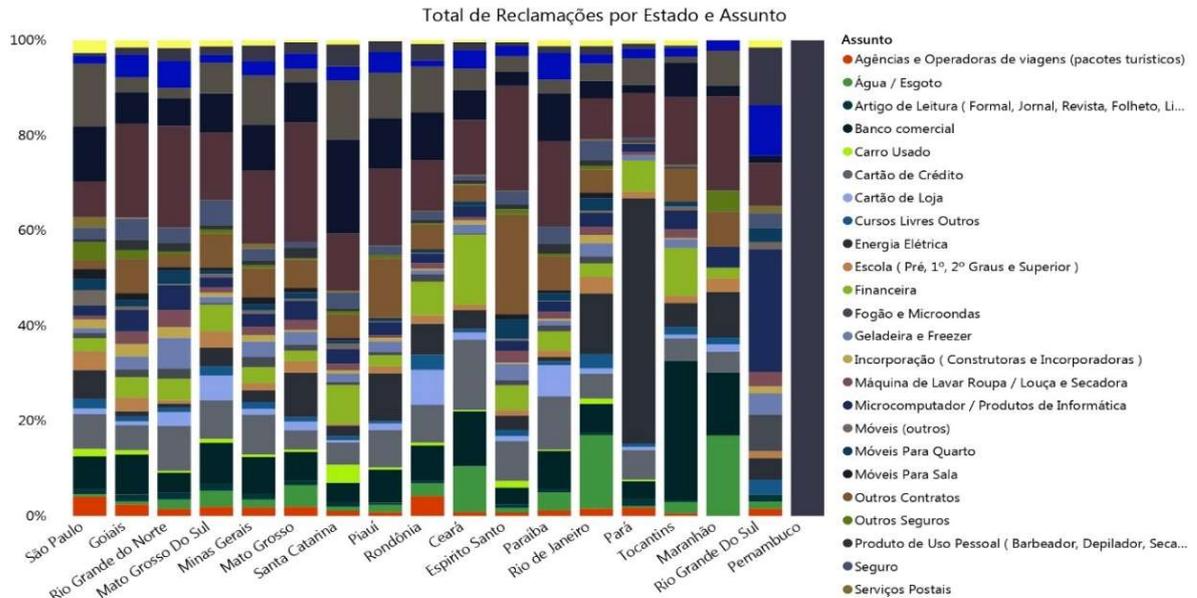


Figura 11: Total de reclamações por estado  
Fonte: autoria própria

#### 4.3.5. Total de reclamações por assunto

De acordo com a Figura 12, na qual está apresentado o total de reclamações por assunto, foi verificado que para o estado de São Paulo o assunto com o maior índice de reclamação foi Cartão de crédito, como também para o Rio Grande do Norte e Paraíba; para Goiás e Mato Grosso do Sul, o assunto foi Banco Comercial; e para os Estados de Piauí e Pará, foi de energia elétrica.



**Figura 12: Gráfico para análise de reclamações por estado e assunto**  
**Fonte: autoria própria**

## 5 Conclusão

O sucesso de um projeto de *Business Intelligence* depende de muito fatores, como estratégia bem definida, engajamento da equipe, tecnologias utilizadas, entre outros. Dentro do contexto de tratamento de dados, os dados públicos, neste caso, apresentados utilizando arquivos csv's, não é de fácil interpretação ao cidadão, que é a parte interessada do projeto, sendo necessário a construção de *dashboards*, de acordo com a especificação das regras de negócio.

É importante destacar a checagem de dados, verificando a integridade deles, para não interferir em alguma interpretação na construção de *dashboards*, corrigindo os pontos necessários, seja dentro das transformações de ETL ou nos próprios csv's.

Para a execução do projeto foram encontradas dificuldades tecnológicas, como por exemplo, a escolha de um banco de dados com o menor consumo de memória. Isso foi resolvido com um banco de dados visando equilibrar o uso entre a transformação.

É importante destacar a escolha dos *dashboard* para cada regra de negócio levantada, dentro do *layout* e jogo de cores, de forma a se obter vários tipos de interpretações.

E, finalmente, podemos publicar o *dashboard* em um domínio público, alimentado por meio da rotina apresentada, melhorando a interface apresenta para fácil uso do usuário.

Dentro da proposta do trabalho, foi possível realizar o processamento de dados e, apesar da dificuldade de manipulação das imagens, foi possível a interpretação de dados, dentro de uma amostra de trinta registros. Uma das propostas de melhoria, seria implementar o sistema via web, sendo alimentado pela atualização de dados, com filtros definidos como, recuperar dados por estado, CNAE e nome da empresa, ano, trimestre, mês.

## 6 Referências

- API CNAE - **Cadastro Nacional de Atividades Econômicas**. (2020). Disponível em: <<https://servicodados.ibge.gov.br/api/docs/cnae?versao=2>>. Acesso em: 2 jun. 2020.
- BI, B. OF. **Business Intelligence Productivity Tools | SQL Power Software**. Disponível em: <<http://www.bestofbi.com/page/architect>>. Acesso em: 21 abr. 2020.
- BRITO, T. D. S. PITON, R. **Solução de Business Intelligence utilizando a plataforma Microsoft na área da Segurança Pública**. (2017) Projeto de Graduação—Rio de Janeiro: Universidade Federal do Estado do Rio de Janeiro (UNIRIO).
- OLIVEIRA, I. S. **TESTE BASEADO EM DEFEITOS PARA AMBIENTES DE DATA WAREHOUSE**. (2015) Dissertação de Mestrado—Curitiba: Universidade Tecnologia Federal do Paraná.
- DOS SANTOS P, C. (2015) **Aplicação de Técnicas de ETL para a Integração de Dados com Ênfase em Big Data na Área de Saúde Pública**. Dissertação de Mestrado—Salvador: Universidade Federal da Bahia
- MONTEIRO, A. V. G. AND PINTO, M. P. O. AND ROSA MARIA E. MOREIRA DA, C. (2004) Uma aplicação de Data Warehouse para apoiar negócios. **Cardenos Do IME: Série Informática**, v. 16, p. 11.
- NETO FREITAS, P. D. F. (2014) **BUSINESS INTELLIGENCE APLICADA NO APOIO À TOMADA DE DECISÕES EM SEGURANÇA PÚBLICA**. Dissertação de Mestrado—Mossoró: Universidade Federal Rural do Semi-Árido.
- OGRODOVSKI, A. **CONSTRUÇÃO DE SPATIAL DATA WAREHOUSES UTILIZANDO ~ SOFTWARES LIVRES** (2013). Trabalho de diplomação de curso—MEDIANEIRA: UTFPR.
- PITON, R. **Data Warehouse Passo a Passo**. Porto Alegre: Raizzer, 2018.
- PIZANI, A. L. AND SIQUEIRA, H. B. **GERAÇÃO DE INDICADORES DA JUSTIÇA: UTILIZANDO UMA SOLUÇÃO DE BUSINESS INTELLIGENCE COM DADOS ABERTOS** (2018). Trabalho de Conclusão de Curso—Florianópolis: UNIVERSIDADE DO SUL DE SANTA CATARINA (UNISUL).
- POSTGRESQL: About**. Disponível em: <<https://www.postgresql.org/about/>>. Acesso em: 2 jun. 2020.
- POWER BI, M. **O que é o Power BI Desktop? - Power BI | Microsoft Docs**. Disponível em: <<https://docs.microsoft.com/pt-br/power-bi/desktop-what-is-desktop>>. Acesso em: 21 abr. 2020.
- REIS, C. **UMA ARQUITETURA DE BI: PORTAL DE ACESSO À INFORMAÇÃO E TRANSPARÊNCIA DO TJSC – GESTÃO DE PESSOAS**. Trabalho de Conclusão de Curso—Florianópolis: UNIVERSIDADE DO SUL DE SANTA CATARINA (UNISUL), 2018.
- SENACON/MJ, S. N. DO C. **Cadastro Nacional de Reclamações Fundamentadas (PROCONS - SINDEC)**. Disponível em: <<http://www.dados.gov.br/dataset/cadastro-nacional-de-reclamacoes-fundamentadas>>

procons-SINDEC1>. Acesso em: 19 abr. 2020.

SHARDA, R.; DELEN, D. AND TURBAN, E (,2019). **BUSINESS INTELLIGENCE E ANÁLISE DE DADOS PARA A GESTÃO DO NEGÓCIO**. 4. ed. Porto Alegre: Boo.

VANTARA, H. **Data Integration - Kettle**. Disponível em: <<https://community.hitachivantara.com/s/article/data-integration-kettle>>. Acesso em: 21 abr. 2020.

VEIGA, J. C. DE A. AND GIMARÃES, J. C. B (2015). **Análise de Dados Abertos Governamentais usando Técnicas de Business Intelligence: um Estudo de Caso das Eleições 2014**. Projeto de Graduação—Rio de Janeiro: Universidade Federal do Estado do Rio de Janeiro (Centro de Ciências Exatas e Tecnologia) - (Escola de Informática Aplicada).

VIEIRA, R. DE S (2012). **IMPLANTAÇÃO DE BUSINESS INTELLIGENCE NO SISTEMA NL GESTÃO**. Trabalho de Conclusão de Curso—Caxias do sul: Universidade de Caxias do Sul.