



INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE PERNAMBUCO

Campus Recife

Departamento Acadêmico de Sistemas, Processos e Controles Eletro-Eletrônicos

Curso Tecnologia em Análise e Desenvolvimento de Sistemas

SIANDRO RODRIGO DE CARVALHO

**ANÁLISE DAS OCORRÊNCIAS DE ACIDENTES DE TRÂNSITO NA CIDADE DO
RECIFE**

Recife

2021

SIANDRO RODRIGO DE CARVALHO

**ANÁLISE DAS OCORRÊNCIAS DE ACIDENTES DE TRÂNSITO NA CIDADE DO
RECIFE**

Trabalho de conclusão de curso apresentado ao Programa de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Ciência e Tecnologia de Pernambuco, como requisito para obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador: Prof^a. Dr^a. Aida Araújo Ferreira

Recife

2021

C331a
2021

Carvalho, Siandro Rodrigo.

Análise das ocorrências de acidentes de trânsito na cidade do Recife / Siandro Rodrigo de Carvalho. --- Recife: O autor, 2021.
44f. il. Color.

TCC (Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas) – Instituto Federal de Pernambuco, Departamento Acadêmico de Sistemas, Processos e Controle Eletro Eletrônicos, 2021.

Inclui Referências.

Orientadora: Professora Dra. Aida Araújo Ferreira

1. Análise de dados. 2. Dados Abertos Governamentais. 3. Business Intelligence. 4. Acidente de trânsito. I. Título. II. Ferreira, Aida Araújo. (orientadora). III. Instituto Federal de Pernambuco.

CDD 005.1 (21ed.)

Trabalho de conclusão de Curso apresentado pelo estudante Siandro Rodrigo de Carvalho à coordenação de Análise e Desenvolvimento de Sistemas, do Instituto Federal de Pernambuco, sob o título de “Análise das Ocorrências de Acidentes de Trânsito na cidade do Recife”, orientado pela Prof. Dr. (a) Aida Araújo Ferreira e aprovado pela banca examinadora formada pelos professores :

Recife, 12 de Novembro de 2021.

Prof. Dr. Aida Araújo Ferreira
CTADS/DACS/IFPE

Prof. Dr. Rafael Roque Aschoff
CTADS/DACS/IFPE

Prof.(a) Dr. Vânia Soares de Carvalho

Recife

2021

AGRADECIMENTOS

Agradeço primeiramente a Deus por tudo que me concedeu até aqui, a saúde e força para superar todas as dificuldades que enfrentei nessa jornada. Aos meus familiares, principalmente à minha mãe por todo esforço para me auxiliar em vários momentos dessa caminhada e à minha, agora, esposa por todo incentivo e fé para que eu chegasse ao fim dessa jornada. E não poderia esquecer de todos os professores que compartilharam comigo seus conhecimentos e experiências ao longo do curso, e em especial à minha orientadora Prof^ª Dr^a Aida Araújo Ferreira pelo compromisso e paciência.

Faça ou não faça. Tentativa não há. (Mestre Yoda)

RESUMO

O presente trabalho apresenta o desenvolvimento de um ambiente de *Business Intelligence*, o qual proporciona a geração de conhecimento a partir da análise das ocorrências de acidentes de trânsito registrados na cidade do Recife, entre os anos de 2015 a 2019 do conjunto de dados de Acidentes de Trânsito com e sem vítimas, os quais estão dispostos na base de dados abertos da prefeitura da referida cidade. Nesse trabalho também foram apresentadas as ferramentas utilizadas no desenvolvimento do ambiente de *business intelligence*, bem como auxiliaram nas análises dos dados e resultados, e tornaram o processo de construção de conhecimento mais ágil; onde pode-se destacar a ferramenta *Spoon*, disponível na *suite Pentaho Data Integration*, a qual fora utilizada no processo de ETL da base de dados; e o plugin *Saiku Analytics*, que auxiliou na visualização dos resultados, a partir de gráficos gerados.

Palavras-chaves: *Business Intelligence*. Dados Abertos Governamentais. Análise de Dados. Acidente de Trânsito.

ABSTRACT

The present work presents the development of a Business Intelligence environment, which provides the generation of knowledge from the analysis of the occurrences of traffic accidents registered in the city of Recife, between the years 2015 to 2019 of the Accidents from Traffic with and without victims, which are available in the open database of the city hall in that city. In this work, the tools used in the development of the business intelligence environment were also presented, as well as helped in the analysis of data and results, and made the knowledge construction process more agile; where we can highlight the Spoon tool, available in the Pentaho Data Integration suite, which was used in the ETL process of the database; and the Saiku Analytics plugin, which helped in viewing the results from generated graphs.

Keywords: Business Intelligence. Open Government Data. Data Analytics. Accidents from Traffic.

LISTA DE FIGURAS

Figura 1 – Diagrama de processos utilizados no estudo.....	20
Figura 2 – Visualização prévia dos dados dos acidentes de 2015.....	22
Figura 3 – Modelo Estrela (<i>Star Schema</i>).....	24
Figura 4 – Rotina para carga dos dados na base Acidentes.....	25
Figura 5 - <i>Script</i> para remoção de espaçamento excedente.....	26
Figura 6 – Visualização do processo ETL da Dimensão Local.....	26
Figura 7 – Visualização do passo Separar_Rua.....	28
Figura 8 – Visualização do processo de ETL da Dimensão Acidente.....	28
Figura 9 – <i>Script</i> para substituição de valores do campo natureza.....	29
Figura 10 – Visualização do processo de ETL da Dimensão Tempo.....	30
Figura 11 – Visualização da função Extrair dia/mes/ano.....	30
Figura 12 – Visualização do passo Separa Hora/Minuto.....	31
Figura 13 – Visualização do processo de criação e carga da tabela Fato.....	31
Figura 14 – Visualização da construção do Cubo OLAP.....	33
Figura 15 – Visualização da criação das dimensões no cubo OLAP.....	33
Figura 16– Visualização da definição da métrica Quantidade.....	34
Figura 17 – Visualização dos parâmetros utilizados na consulta.....	36
Figura 18 – Resultado da consulta de bairros com mais registros por ano.....	36
Figura 19 – Resultado da pesquisa de ruas com maiores índices de acidentes.....	37
Figura 20 – Resultado do horário com maior número de acidentes por bairro.....	38
Figura 21 – Resultado do horário com maior número de acidentes por rua.....	38
Figura 22 – Resultado dos meses com maior índice de acidentes por ano.....	39
Figura 23 – Resultado dos dias da semana com mais registros por ano.....	40
Figura 24 – Resultado do tipo de acidentes com mais registros por ano.....	40

LISTA DE TABELAS

Tabela 1 - Dicionário de dados dos acidentes de trânsito.....	21
Tabela 2 - Quantidade de registros por ano e total desses.....	35

LISTA DE ABREVIATURAS E SIGLAS

BI	<i>Business Intelligence</i>
CSV	<i>Comma Separated Values</i>
CTTU	Companhia de Trânsito e Transporte Urbano
DML	<i>Data Manipulation Language</i>
DW	<i>Data Warehouse</i>
ETL	<i>Extract Transform Load</i>
PUC	<i>Pentaho User Console</i>
PDI	<i>Pentaho Data Intregation</i>
OLAP	<i>On-Line Analytical Processing</i>
OMS	Organização Mundial de Saúde
SQL	<i>Structured Query Language</i>
SW	<i>Schema-Workbench</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Motivação	13
1.2	Justificativa	13
2	OBJETIVO	16
2.1	Objetivo geral	16
2.2	Objetivos específicos	16
3	FUNDAMENTAÇÃO TEÓRICA	17
3.1	Business Intelligence	17
3.2	Portal de dados abertos da prefeitura do Recife	18
4	METODOLOGIA	20
4.1	Levantamento de requisitos	20
4.2	Análise da fonte dos dados	21
4.3	Processo de modelagem dos dados	23
4.4	Processo de Extração, Transformação e Carga dos dados (ETL)	24
4.4.1	<i>Base de dados inicial</i>	24
4.4.2	<i>Processo de criação e ETL da Dimensão Local</i>	26
4.4.3	<i>Processo de criação e ETL da Dimensão Acidente</i>	28
4.4.4	<i>Processo de criação e ETL da Dimensão Tempo</i>	30
4.4.5	<i>Criação e carga da Tabela Fato</i>	31
4.5	Construção do cubo OLAP	32
5	RESULTADOS	35
6	CONCLUSÃO	41
	REFERÊNCIAS	43

1 INTRODUÇÃO

1.1 Motivação

Os acidentes de trânsito são umas das principais causas de mortes no Brasil, e segundo a OMS (Organização Mundial de Saúde) o país é considerado o que possui o trânsito mais violento do mundo, devido ao aumento, de pouco mais, de 3% no número total de óbitos em acidentes de trânsito em um ano, segundo dados mais recentes, referentes ao ano de 2014. Sendo que entre 2012 e 2013, essa porcentagem teve uma queda considerável de 5,7%; revertendo uma elevada tendência de aumento do número de casos, observados nos quatro anos anteriores a esse período.

E entre as regiões mais violentas, em número de óbitos, a região Nordeste se encontra em segundo lugar, ficando atrás somente da região Sudeste. E em relação aos estados daquela região com o maior número de vítimas fatais, Pernambuco se encontra em terceiro lugar, com Recife como a cidade com o maior número de óbitos, no estado, segundo esses dados.

Então, com a grande disponibilidade de dados abertos na “rede”, nos dias atuais; poderíamos buscar formas ou alternativas de alterar esse cenário preocupante, em que se encontra a capital do estado, e o próprio estado de Pernambuco, a cair no *ranking* dos estados brasileiros com índice elevado de violência no trânsito?

1.2 Justificativa

A ideia de Dados Abertos Governamentais, tem como base a disponibilização de um grande número de bases de dados oficiais, de forma a abranger todas as particularidades das atividades do governo, permitindo que a população em geral, além de acesso a esses dados, possam compartilhá-los conforme sua necessidade de informação (RIBEIRO; ALMEIDA, 2011).

Dessa forma, em posse dessas informações governamentais, disponibilizadas ao grande público, permitirá, aos cidadãos, o acesso aos serviços públicos, além de

permitir que os mesmos possam gerar novas informações e conhecimentos, como também de outros novos serviços concebidos, a partir, dessa interação entre governo e sociedade por meio da exploração de dados abertos governamentais (CARLOS; RIBEIRO; MATHEUS, 2010).

Países como Estados Unidos, Austrália, Nova Zelândia, Reino Unido, entre outros, já disponibilizam, há anos, por meio de políticas de acesso a dados públicos através da *web* e de portais que disponibilizam informações públicas em diferentes formatos, para que todo cidadão interessado possa utilizar esses dados. Além disso, com essas informações disponíveis a qualquer indivíduo, os governos incentivam os cidadãos a criarem novos meios, ou ferramentas, de utilização dessas informações disponibilizadas em seus portais.

Já aqui no Brasil, desde 2009, o governo brasileiro vem realizando ações de desenvolvimento de políticas de disseminação de dados governamentais para o livre uso da sociedade, alinhando-se assim, ao grande crescimento do movimento global para a democratização do acesso a dados e informações abertos. Mas, a partir de 2011, o Brasil passou a possuir uma lei com o propósito de regulamentar o direito constitucional do cidadão de acesso às informações públicas (VARGAS; FROZZA, 2015), a chamada Lei de Acesso à Informação (12.527/2011).

Com essa lei sancionada e em vigor no país, muitos órgãos governamentais passaram a disponibilizar informações públicas aos cidadãos, com isso salvaguardando ao cidadão o direito de acesso a espaços de colaboração e participação na administração pública, possibilitando a conversa entre governo e cidadãos com a finalidade de entender as necessidades e unir governo e cidadão em busca do bem comum (OLIVERIO, 2011).

Entre tantos órgãos governamentais, que disponibilizam dados abertos ao cidadão, podemos citar a Prefeitura da Cidade do Recife, a qual lançou em julho de 2013, um portal com uma base de dados abertos, contendo informações relativas a diferentes setores da administração, como saúde, educação e finanças. E dentre essas áreas, ou bases, disponíveis, estão os dados relativos a acidentes ocorridos na cidade do Recife, a qual se encontra entre as principais cidades com o trânsito mais violento do país.

Portanto, com base nessas informações e dados disponíveis no portal da prefeitura, podemos de alguma maneira utilizá-los de modo a encontrar formas de combater a violência no trânsito da cidade do Recife, através do levantamento dos dados de acidentes, e com isso determinar padrões e localidades com grande número de ocorrências, por exemplo; e em posse dessas novas informações, poderemos gerar meios que auxiliem na redução da taxa de acidentes na capital pernambucana.

2 OBJETIVO

2.1 Objetivo geral

Utilizar a base de dados de acidentes de trânsito ocorridos na cidade do Recife, modelando-os de forma a encontrar e determinar padrões de acidentes e locais de ocorrência, para que com isso se possa facilitar a tomada de ações de combate a acidentes de trânsito na capital pernambucana.

2.2 Objetivos específicos

- Modelar os dados relativos aos acidentes;
- Identificar os principais tipos de acidentes e o número de ocorrências;
- Identificar os locais de ocorrência de acidentes e quais os tipos de acidentes que mais ocorreram em um determinado local;

3 FUNDAMENTAÇÃO TEÓRICA

3.1 *Business Intelligence*

Em 1980 foi introduzido o conceito abrangente de *Business Intelligence* (BI) por Howard Dresner, analista do *Gartner Research Group* (POWER, 2003; BUCHANAN; O'CONNELL, 2006). Atualmente, Silva et al. (2016), traz uma definição do BI como uma ferramenta para processamento de um grande volume de dados para o apoio à tomada de decisões gerenciais.

Este processo percorrer a coleta, armazenamento e gestão desses dados, onde há a oportuna transformação de dados em informações, que culminam em decisões e finalmente em ações. Desta forma, ocorre um acesso interativo aos dados, proporcionando a manipulação dos dados, de modo a conceder aos gerentes e analistas de negócios a aptidão em realizar uma análise mais assertiva (NEGASH, 2015).

O BI compõe um processo mais robusto que é a Análise de Dados ou “*Data Analytics*”, como é chamado em inglês. Possui um objetivo de auxiliar o usuário a compreender e analisar o passado, viabilizando a análise das interações entre os dados, para assim criar métodos de resolução para determinadas problemáticas.

No que tange as capacidades dos sistemas de BI, as mesmas podem ser divididas em dois grupos: tecnológico e organizacional. Acerca das capacidades tecnológicas, as mesmas versam sobre as ferramentas, produtos, tecnologias e *softwares* utilizados para coletar dados de fontes heterogêneas dispersas e, posteriormente, consolidá-los num único banco de dados (FEENEY; WILLCOCKS, 1998).

Nessa perspectiva, são incluídos: Ferramentas para extração, transformação e carregamento de dados, que são as principais responsáveis pela transmissão dos dados para o *Data Warehouse* (DW). Esse momento compõe o processo mais crítico na construção de um DW: O *Data Warehouse*, ferramenta utilizada no armazenamento e gerenciamento de diversos registros de corporações, com o objetivo de integrar, limpar, agregar e viabilizar a consulta; Ferramentas analíticas OLAP (*On-Line Analytical Processing*) que ajudam os usuários acessarem,

analisarem e compartilhar informações que estão contidas no DW; Ferramentas de *Data Mining*, que viabilizam a identificação de padrões diferentes, generalizações, regularidades e regras em dados; as Ferramentas para relatórios e *ad-hoc*, que participam da criação e a utilização de diversos relatórios sintéticos; Camada de apresentação, que utilizam interfaces gráficas e multimídia, que objetivam conceder ao usuário um acesso fácil e claro das informações (FISHER; MAHAJAN, 2006).

No que se refere às capacidades organizacionais, são aquelas que desvelam como o formato do BI é utilizado em determinados serviços, no sentido de trabalhar com a informação e o conhecimento, fomentando uma comunicação acertada e uma abordagem analítica e holística dos processos de negócios de organizações. Essa realidade aponta que o BI visa apoiar as atividades de gestão (OLSZAK; ZIEMBA, 2012).

3.2 Portal de dados abertos da prefeitura do Recife

O Portal de Dados Abertos da Prefeitura do Recife publica os dados provenientes de diversas secretarias relacionadas à gestão pública. Essa transparência permite que qualquer cidadão consiga visualizar e/ou analisar os dados, facilitando assim a aplicação dessas informações a diversos objetivos.

O portal está dividido em dez grupos, sendo eles: COVID, Cultura, Educação, Finanças, Governo e Política, Mobilidade, Saúde, Meio Ambiente, Turismo e Urbanismo. Dentre as organizações que compõem tais grupos, têm-se: Controladoria Geral do Município (CGM), Empresa Municipal de Informática (EMPREL), Empresa de Manutenção e Limpeza Urbana (EMLURB), Gabinete de Projetos Especiais - Administração Direta – GABPE, Instituto da Cidade Pelópidas Silveira, Fundação de Cultura Cidade do Recife (FCCR), Ouvidoria, Secretaria de Administração e Gestão de Pessoas (SADGP), Secretaria de Cultura, Secretaria Executiva de Defesa Civil (SEDEC - Recife), Secretaria de Desenvolvimento Social, Juventude, Política Sobre Drogas e Direitos Humanos, Secretaria de Educação, Secretaria de Finanças, Secretaria de Governo e Participação Social, Secretaria de Infraestrutura e Serviços Urbanos, Secretaria de Segurança Urbana, Secretaria de Saúde, Secretaria de Trabalho e Qualificação Profissional, Secretaria de Turismo e

Autarquia de Trânsito e Transporte Urbano do Recife (CTTU), a qual este trabalho debruçou-se

A Autarquia de Trânsito e Transporte Urbano do Recife – CTTU realiza o controle e fiscalização do trânsito, do sistema de transporte e das ações de estudos e implementação de projetos de melhoria da mobilidade urbana e educação para o trânsito da capital de Pernambuco. É uma sociedade de economia mista, onde sua administração indireta é realizada pelo município.

4 METODOLOGIA

A metodologia adotada nesse estudo foi a Observação Direta Intensiva Sistemática, que consiste em examinar os fatos ou fenômenos que se deseja estudar, através da utilização de instrumentos para a coleta dos dados ou de fenômenos observados. Assim, com apoio de informações colhidas a partir de livros, documentações e/ou manuais das ferramentas utilizadas, artigos científicos, como também de materiais disponíveis na *Internet*, através de fóruns de discussão; foi possível à realização desse estudo.

Os dados de acidentes de trânsito ocorridos na cidade do Recife e registrados pela CTTU, foram obtidos através do Portal de Dados Abertos da Cidade do Recife¹; referentes aos anos de 2015 a 2019, totalizando 53.767 registros.

Podemos definir o processo desse estudo de acordo com a imagem que se segue:

Figura 1 – Diagrama de processos utilizados no estudo



Fonte: Autoria própria (2021)

4.1 Levantamento de requisitos

A partir da realização da coleta dos dados, deu-se início ao levantamento dos requisitos necessários à realização do projeto, através da análise, planejamento e modelagem. Assim, nessa etapa, do projeto, foram elencadas e definidas quais demandas que esse estudo se proporia a responder ao se analisar tais dados.

Assim chegou-se às seguintes questões a serem respondidas:

1. Qual local com maior incidência de acidentes na cidade?

¹ <http://dados.recife.pe.gov.br/dataset/acidentes-de-transito-com-e-sem-vitimas>

2. Qual período do dia se verifica mais acidentes?
3. Qual período do ano há maior incidência de acidentes?
4. Qual tipo de ocorrência de acidentes ocorre com maior frequência?

Dessa forma, após serem identificadas as questões que necessitam de respostas, permitiu-se que o DW, no qual seriam dispostas todas as informações brutas, referentes aos acidentes, obtidas no portal de dados; pudesse ser construído, e a partir daí todas as etapas seguintes pudessem ser alcançadas.

4.2 Análise da fonte dos dados

Ao serem definidas as necessidades do projeto, foi necessária uma análise aprofundada dos dados, dispostos nos arquivos obtidos; a qual fez-se necessária para se verificar se os dados possuíam as informações necessárias para responder as questões levantadas.

Para essa análise foi utilizado como suporte o Dicionário dos Dados, o qual descreve as associações dos metadados ao conjunto de dados dos registros de acidentes. Como mostrado na tabela que se segue.

Tabela 1 - Dicionário de dados dos acidentes de trânsito

Campo	Descrição	Tipo	Tamanho
tipo	Tipo de acidente (com ou sem vítima)	Char	100
data	Data do atendimento	Date	10
hora	Hora do atendimento	Numeric	20
bairro	Bairro do acidente	Char	50
endereço	Endereço do acidente	Char	50
natureza	Tipo da ocorrência	Char	50

Os arquivos foram obtidos, como dito anteriormente, a partir do Portal de Dados Abertos da Cidade do Recife, no formato *.csv (*Comma Separated Values*), os quais são atualizados trimestral, e são divididos por ano.

Já com o conhecimento adquirido através da análise do dicionário dos dados, deu-se início à análise dos dados em si, e daí se extrair os dados fundamentais ao projeto. Para isso, foi utilizada a ferramenta *Spoon*, disponível no *Pentaho Data Intregation* (PDI), a qual auxilia na construção do *Data Warehouse*, como também

auxilia na análise dos dados e no processo de ETL (Extração, Transformação e Carga) dos dados.

Na análise dos arquivos, um a um, com ajuda da ferramenta, utilizou-se uma amostra de 1000 linhas de registros, de cada arquivo, para a partir daí serem definidos quais dados seriam necessários para a construção do DW. Nessa análise pôde-se aferir padrões nas estruturas dos dados, como também identificar a ausência de informações, como identificadores primários dos registros e também linha de registros com campos não preenchidos, como mostra a figura 2.

Figura 2 – Visualização prévia dos dados dos acidentes de 2015

#	tipo	situacao	data	hora	bairro	endereco
1	SEM VÍTIMA	CANCELADA	2015-12-03	11:36:00.000	CASA AMARELA	AV NORTE
2	SEM VÍTIMA	FINALIZADA	2015-06-01	06:35:00.000	ENCRUZILHADA	EST DE BELEM
3	SEM VÍTIMA	FINALIZADA	2015-06-01	08:04:00.000	BOA VIAGEM	RUA ERNESTO DE PAULA SANTOS
4	SEM VÍTIMA	FINALIZADA	2015-06-01	08:06:00.000	AFOGADOS	AV SUL
5	SEM VÍTIMA	CANCELADA	2015-06-01	08:13:00.000	MADALENA	RUA BENFICA
6	SEM VÍTIMA	FINALIZADA	2015-06-01	08:40:00.000	TORRE	RUA REAL DA TORRE
7	SEM VÍTIMA	FINALIZADA	2015-06-01	09:30:00.000	SÃO JOSÉ	C SANTA RITA
8	COM VÍTIMA	FINALIZADA	2015-06-01	09:10:00.000	PINA	AV HERCULANO BANDEIRA
9	SEM VÍTIMA	CANCELADA	2015-06-01	09:38:00.000	PINA	AV ANTONIO DE GOES
10	SEM VÍTIMA	CANCELADA	2015-06-01	10:11:00.000	ILHA DO LEITE	AV BEIRA RIO
11	COM VÍTIMA	FINALIZADA	2015-06-01	10:30:00.000	IPSEP	RUA JEAN EMILE FAVRE
12	SEM VÍTIMA	FINALIZADA	2015-06-01	10:56:00.000	TAMARINEIRA	RUA PADRE ROMA
13	SEM VÍTIMA	FINALIZADA	2015-06-01	11:08:00.000	IMBIRIBEIRA	AV DOM JOAO VI
14	SEM VÍTIMA	FINALIZADA	2015-06-01	11:14:00.000	BOA VISTA	RUA DO HOSPICIO
15	COM VÍTIMA	CANCELADA	2015-06-01	11:37:00.000	ILHA DO RETIRO	RUA JOAO ELISIO RAMOS
16	SEM VÍTIMA	FINALIZADA	2015-06-01	12:21:00.000	TORRE	RUA REAL DA TORRE
17	SEM VÍTIMA	FINALIZADA	2015-06-01	12:38:00.000	BOA VIAGEM	AV BOA VIAGEM
18	SEM VÍTIMA	CANCELADA	2015-06-01	12:48:00.000	IMBIRIBEIRA	AV SUL
19	SEM VÍTIMA	FINALIZADA	2015-06-01	13:03:00.000	IMBIRIBEIRA	RUA ENGENHEIRO JOSE APOLINARIO
20	SEM VÍTIMA	FINALIZADA	2015-06-01	13:46:00.000	ARRUDA	RUA ZEFERINO AGRA
21	SEM VÍTIMA	FINALIZADA	2015-06-01	14:15:00.000	ÁGUA FRIA	EST VELHA DE AGUA FRIA
22	SEM VÍTIMA	FINALIZADA	2015-06-01	14:34:00.000	BOA VIAGEM	RUA RIBEIRO DE BRITO
23	SEM VÍTIMA	FINALIZADA	2015-06-01	14:34:00.000	PRADO	RUA PANDIA CALOGERAS
24	SEM VÍTIMA	FINALIZADA	2015-06-01	14:43:00.000	BOA VISTA	RUA JOAO FERNANDES VIEIRA
25	SEM VÍTIMA	FINALIZADA	2015-06-01	15:13:00.000	BOA VIAGEM	AV BOA VIAGEM

Fonte: Autoria própria (2020) – Extraído da ferramenta *Spoon*

Nos arquivos dos anos 2015 e 2016 constatou-se a existência de 20 campos de informações, e nos anos de 2017 a 2019, 21 campos; dos quais foram escolhidos 6 campos, os quais seriam necessários para se obter as repostas necessárias, como também, para a construção do DW.

Os campos escolhidos foram os seguintes:

- Tipo e Natureza: contém os dados necessários para a construção da dimensão Acidente.

- Endereço e Bairro: Possuem os dados fundamentais para a construção da dimensão Local
- Data e Hora: Os quais possuem os dados necessários para a construção da dimensão Tempo.

Agora com as informações necessárias indicadas, partiu-se para a criação de um modelo multidimensional necessário para a construção do DW.

4.3 Processo de modelagem dos dados

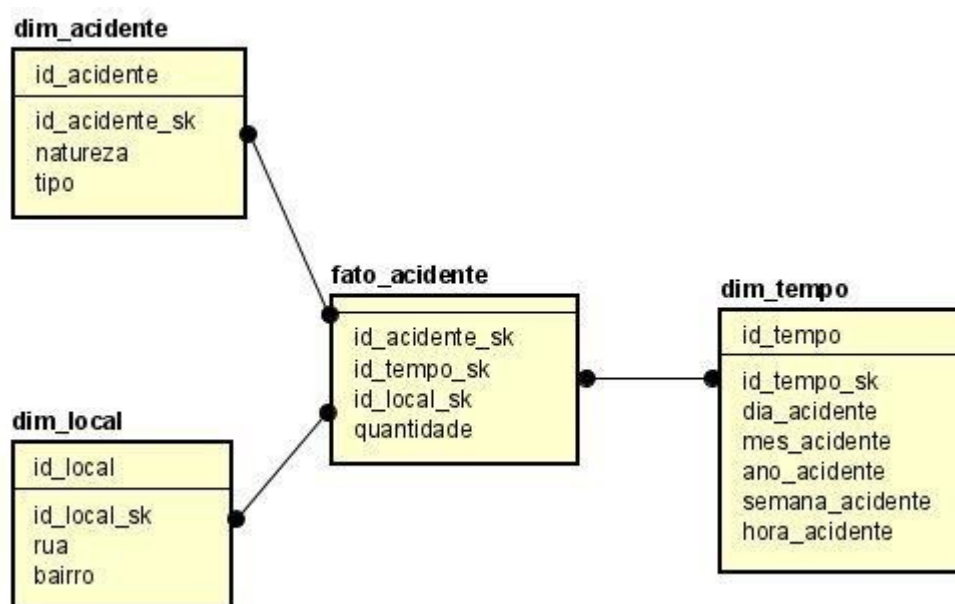
Para Kimball e Ross (2011), a Modelagem Dimensional tem por finalidade abordar os problemas de esquemas demasiadamente complexos, observados no âmbito da apresentação; assim, tornando mais simples o entendimento dos dados, por parte do usuário, como também entregar desempenho, nas consultas; além de proporcionar resiliência a mudanças.

A partir desse conceito, foi escolhido, com base nos modelos de esquemas disponíveis, o modelo estrela, que segundo Machado (2006), esse modelo se mostra como a estrutura básica de um modelo de dados multidimensional. De maneira que sua formação é gerada a partir de uma tabela fato, no centro do desenho, e um conjunto de tabelas dimensões, combinadas ao redor do fato; dando assim, um aspecto de estrela ao diagrama.

Ao final desse processo o modelo gerado apresentou as seguintes características (figura 3):

- 3 tabelas dimensão: Dimensão Tempo, Dimensão Local, Dimensão Acidente;
- Uma tabela fato: Fato.

Figura 3 – Modelo Estrela (Star Schema)



Fonte: Autoria própria (2021)

4.4 Processo de Extração, Transformação e Carga dos dados (ETL)

Esse processo é considerado essencial para a criação das estruturas de Dimensões e Fatos, de um ambiente de DW, pois, é ele quem faz o elo entre o operacional e o DW, tratando da sistematização da transformação e limpeza dos dados provindos de diferentes sistemas organizacionais para a inserção em um DW/*Data Mart*.

4.4.1 Base de dados inicial

Inicialmente foi necessário a criação de uma base de dados preliminar, denominada “**acidentes**”, para armazenar as informações de registros, dos acidentes de trânsito, como também a unificação dos registros, em um único repositório, dos arquivos utilizados nesse estudo; e a partir disso eliminar incorreções dos registros, como por exemplo campos vazios e a falta de um campo de identificação - chaves primárias – (Figura 2) para cada registro informado, o qual se faz necessário para realizar-se, através da separação das colunas, a criação das dimensões necessárias. Esse processo foi realizado de forma gradativa, ou seja,

para cada arquivo de acidente foi realizada uma rotina, resultando no total de três rotinas.

A seguir (Figura 4), podemos observar como ficou cada rotina, para a carga da base de dados “**acidentes**”, as quais resultaram em dois *steps* (passos) cada rotina; e o detalhamento de como cada passo foi executado (para todos os arquivos foram seguidas as mesmas rotinas).

Figura 4 – Rotina para carga dos dados na base Acidentes



Fonte: Autoria própria (2020) – Extraído da ferramenta *Spoon*

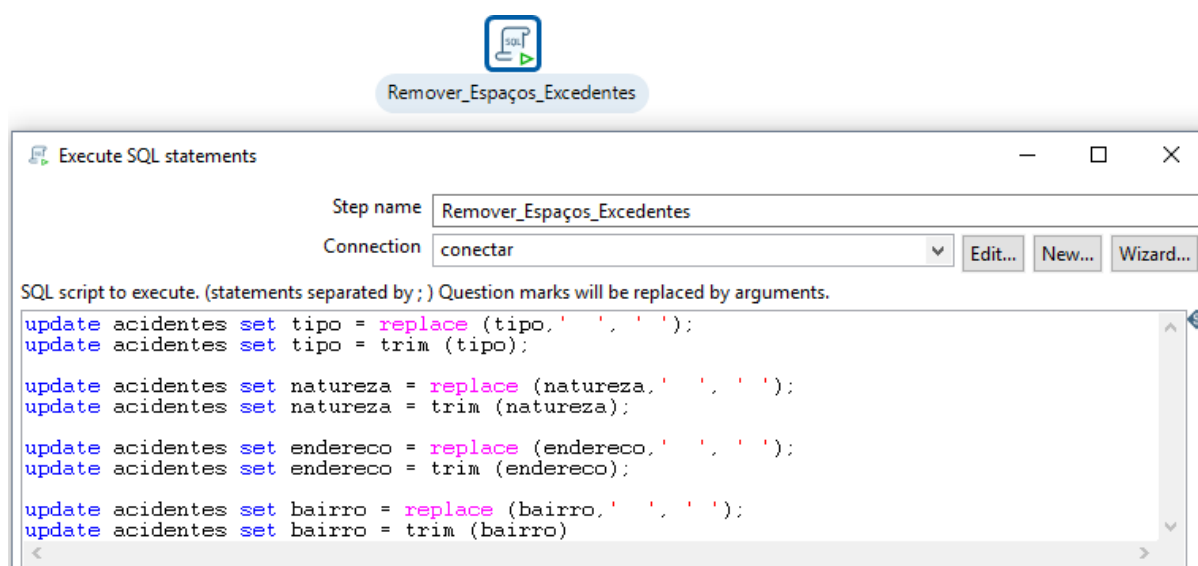
Passo 1 – Extrair dados do arquivo de origem: Foi utilizado um *step*, do tipo *CSV file input*, o qual recebeu o nome “**acidentes_2015**” (2016, 2017, 2018 e 2019), e que tem por função a capacidade de processar dados de um arquivo *.csv, com delimitadores (nesse caso o ponto e vírgula), de forma rápida. Assim, esse passo permite que os dados sejam obtidos de uma fonte, onde estão armazenados, e os guarda em uma tabela na aplicação, de forma temporária.

Passo 2 – Alimentar a base com os dados: O segundo *step* utilizado, é do tipo *Table Output*, o qual permite que os dados do arquivo *.csv possam ser transferidos e armazenados em uma tabela no banco de dados, tendo como saída da tabela um operador DML **INSERT**. Esse passo permite ainda muitas opções de configuração para a tabela, como a opção de criação da tabela, caso a mesma ainda não exista, ou ainda atribuir uma chave de identificação aos dados (chave primária), nesse caso foram utilizadas essas duas opções.

Após a carga inicial dos dados, fez-se necessária a realização da primeira transformação dos dados, pois constatou-se que havia espaçamento excedente nas informações, tanto no início ou final dos registros, como também no meio de registros compostos, por exemplo, o bairro de Boa Viagem, poderia ser encontrado escrito de diversas maneiras como: “Boa Viagem”, “Boa Viagem”, “Boa Viagem ”– espaço excedente no início, no meio ou no final do nome do bairro, respectivamente.

Assim foi necessário utilizar um *script* SQL (Figura 5), o qual pudesse retirar todos os espaçamentos excedentes dos registros, da base de dados, em sua totalidade. De modo que a aplicação oferece um passo para a execução de *scripts* SQL, do tipo *Execute SQL script*.

Figura 5 - Script para remoção de espaçamento excedente



Fonte: Autoria própria (2020) – Extraído da ferramenta *Spoon*

Além do procedimento de remoção de espaços excedentes, foi feito previamente a inserção de mais uma coluna na base de dados, denominada “**quantidade**”, a qual será necessária no processo de criação do cubo OLAP, para ser utilizada como métrica.

Ao final do processo de carga de todos os dados, dos arquivos CSV, para a base de dados de apoio, iniciou-se o processo de criação e carga das Dimensões.

4.4.2 Processo de criação e ETL da Dimensão Local

Para a criação e para o processo de ETL dessa dimensão, foi necessária a utilização de 5 passos, como mostra a figura 6, os quais são descritos logo após.

Figura 6 – Visualização do processo ETL da Dimensão Local



Fonte: Autoria própria (2021) – Extraído da ferramenta *Spoon*

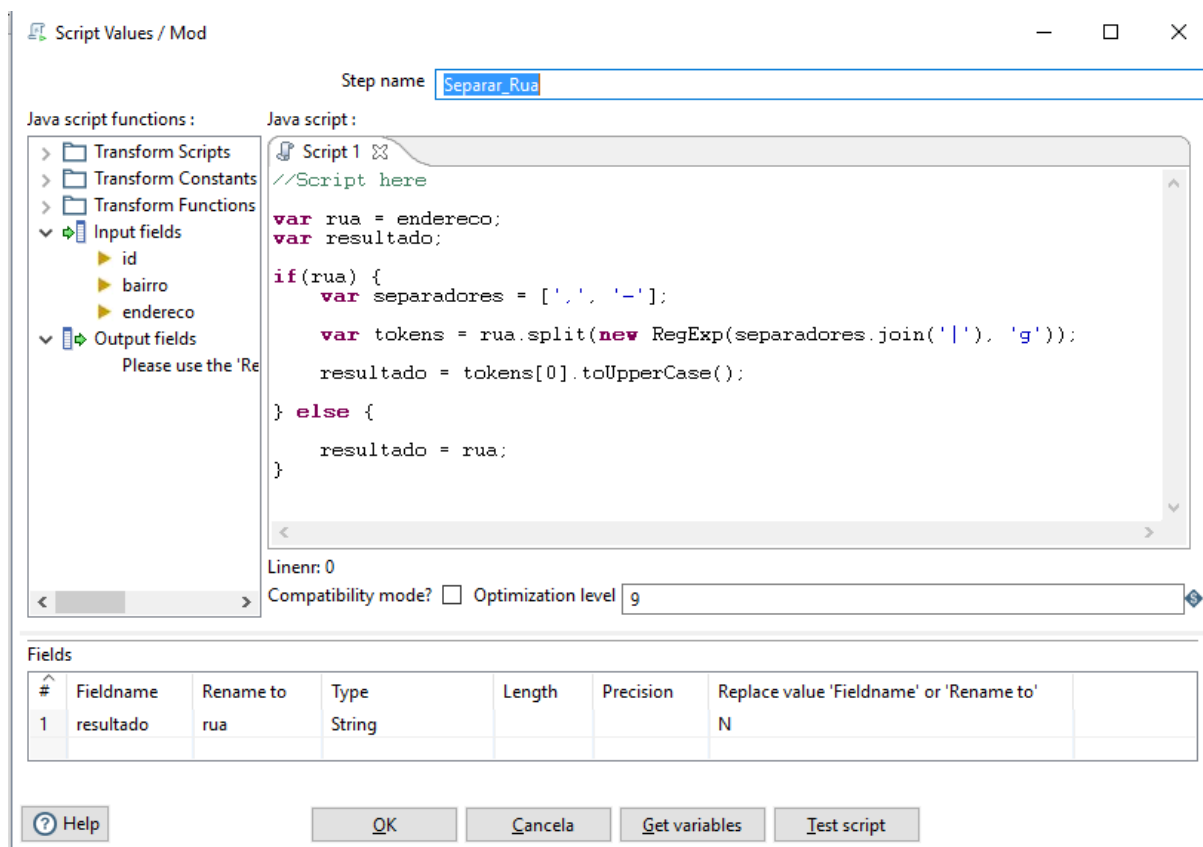
Passo 1 – Extração dos dados da base de dados: O passo “**Acidentes**”, do tipo *Table Input*, permite que os dados sejam lidos diretamente da base de dados “**acidentes**”, através de uma conexão com a mesma.

Passo 2 – Selecionar o nome da rua: O passo “**Separar Rua**”, do tipo *Modified Java Script Value*, permite a utilização de códigos *Javascript*, que auxiliem no tratamentos dos dados. Esse *script* foi utilizado para que somente fossem extraídos dos registros, os nomes das ruas, eliminando qualquer tipo de separador, do campo “**rua**”. Como ilustra a figura 7

Passo 3 – Seleção das colunas pertinentes: Para que fossem extraídas somente as colunas relevantes para essa dimensão, foi utilizado um passo do tipo *Select Values*, o qual foi nomeado por “**Seleciona Valores**”, e que permite que seja possível selecionar e renomear – dentre outras opções – os dados pertinentes à dimensão.

Passo 4 – Carregar os dados na Dimensão Local: O passo subsequente, “**Dimensão Local**”, do tipo *Dimension Lookup/Update*, realiza uma conexão com o DW, e efetua, assim, a criação da Dimensão Local e a carga dos dados selecionados, na mesma.

Figura 7 – Visualização do passo Separar_Rua



Fonte: Autoria própria (2020) – Extraído da ferramenta Spoon

4.4.3 Processo de criação e ETL da Dimensão Acidente

Para a criação e para o processo de ETL dessa dimensão, foi necessária a utilização de 4 passos, os quais são descritos logo após.

Figura 8 – Visualização do processo de ETL da Dimensão Acidente



Fonte: Autoria própria (2020) – Extraído da ferramenta Spoon

Passo 1 – Extração dos dados necessários, a partir da base de dados: O passo “**Acidentes**”, do tipo *Table Output*, realiza, através de uma conexão direta com a base dados, a obtenção dos dados advindos da base de dados “**acidentes**”.

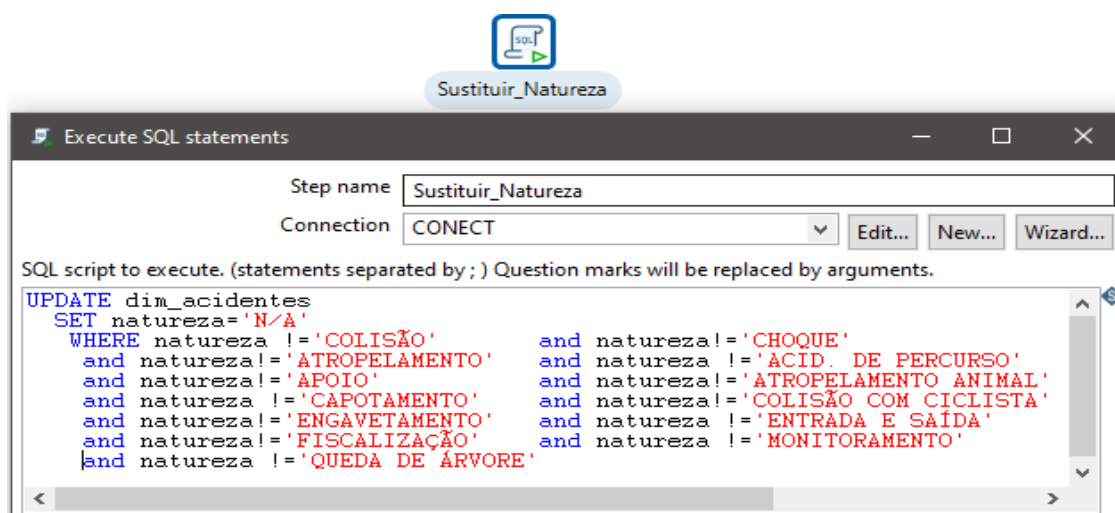
Passo 2 – Seleção dos dados pertinentes à dimensão: O passo “**Seleciona Valores**”, do tipo *Select Value*, seleciona e renomeia os dados pertinentes à Dimensão Acidente.

Passo 3 – Gerar e carregar a Dimensão Acidentes: O passo “**Dimensão Acidente**”, do tipo *Dimension Lookup/Update*, realiza uma conexão com o DW, e efetua, assim, a criação da Dimensão Acidente e a carga dos dados selecionados na mesma.

Ao final do processo de criação e carga dessa dimensão, constatou-se que, no campo “**natureza**”, existiam registros com informações divergentes, pois nesse campo deveriam ser informadas a natureza do acidente, por exemplo, colisão ou atropelamento, porém foram informadas em alguns casos referências do local da ocorrência, ou alguma outra informação, por exemplo: “em frente à praça do Derby”.

Dessa forma fez-se necessário a execução de um *script SQL*, para eliminar essas informações divergentes. Então utilizou-se o passo “**Substituir_Natureza**” (Figura 9), do tipo *Execute SQL script*, para realizar a substituição dessas divergências nas informações, por um valor determinado, nesse caso utilizou-se o valor “**N/A**”.

Figura 9 – Script para substituição de valores do campo natureza



Fonte: Autoria própria (2020) – Extraído da ferramenta Spoon

4.4.4 Processo de criação e ETL da Dimensão Tempo

Para a criação e a rotina de ETL dessa dimensão, foram necessários 6 passos, descritos a seguir, como mostra a figura 10.

Figura 10 – Visualização do processo de ETL da Dimensão Tempo



Fonte: Autoria própria (2020) – Extraído da ferramenta *Sponn*

Passo 1 – Extrair dados relativo ao tempo da base de dados: O passo “**Acidentes**”, do tipo *Table Input*, realiza a conexão com a base de dados e obtém os dados pertinentes, a partir da base “**acidentes**”.

Passo 2 – Obter dados relativos a datas, do campo data: O passo “**Extrair dia/mes/ano**”, do tipo *Calculator*, fornece funções pré-definidas que podem ser aplicadas nos valores dos campos de entrada. Dessa maneira, foram aplicadas as funções *Dia do mês da data A*, *Dia da semana da data A*, *Mês da data A*, *Ano da data A*, que realizam a extração do dia do mês, dia da semana, mês do ano, e ano, respectivamente; e, ao extrair essas informações, armazená-las em um novo campo. Como mostra a figura 11.

Figura 11 – Visualização da função Extrair dia/mes/ano

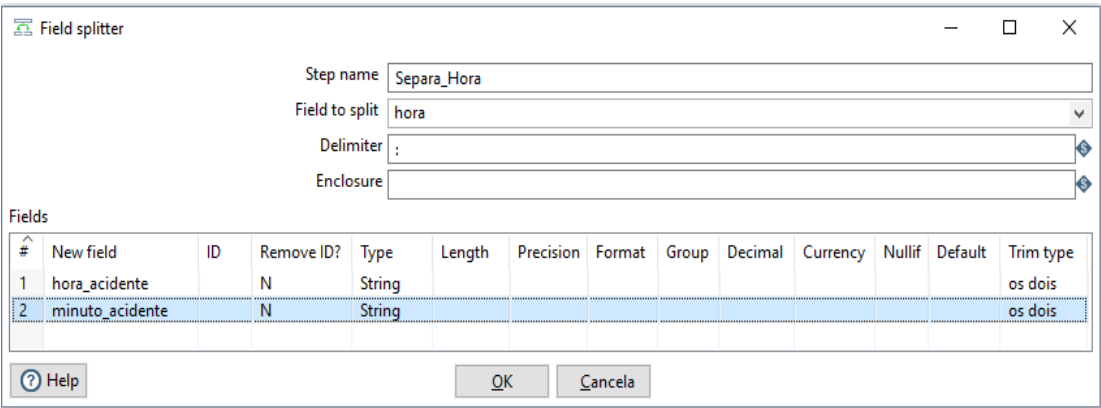
#	Novo campo	Cálculo	Campo A	Campo B	Campo C	Tipo do valor	Tamanho
1	dia_acidente	Dia do mês da data A	data			Integer	
2	mes_acidente	Mês da data A	data			Integer	
3	ano_acidente	Ano da data A	data			Integer	
4	semana_acidente	Dia da semana da data A	data			Integer	

Fonte: Autoria própria (2020) – Extraído da ferramenta *Sponn*

Passo 3 – Extrair informação das horas: O passo “**Separa Hora/Minuto**”, do tipo *Field Splitter*, permite que campos sejam divididos com base nas informações de

um delimitador. Nesse caso, o delimitador utilizado foi o “:” (dois-pontos), que determinará os limites de separação do campo, que foram as horas e os minutos. Como apresenta a figura 12.

Figura 12 – Visualização do passo Separa Hora/Minuto



Fonte: Autoria própria (2020) – Extraído da ferramenta *Spinn*

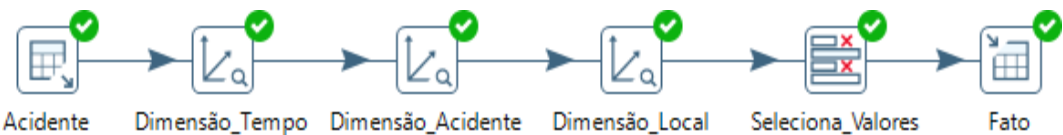
Passo 4 – Seleção dos valores pertinentes: O passo “**Seleciona Valores**”, do tipo *Select Value*, seleciona os campos pertinentes à Dimensão Tempo.

Passo 5 – Criar e carregar a Dimensão Tempo: O passo “**Dimensão Tempo**”, do tipo *Dimension Lookup/Update*, faz uma conexão com o DW e realiza a criação da Dimensão Tempo e posteriormente a carga das informações na mesma.

4.4.5 Criação e carga da Tabela Fato

Para a criação e o processo de ETL da tabela Fato, foram necessários a execução de 6 passos, ilustrados e descritos a seguir.

Figura 13 – Visualização do processo de criação e carga da tabela Fato



Fonte: Autoria própria (2020) – Extraído da ferramenta *Spinn*

Passo 1 – Extração da métrica: O passo “**Acidente**”, do tipo *Table Input*, conecta-se à base de dados para obter a informação da métrica, nomeada

“**quantidade**” necessária à criação do cubo OLAP, a qual foi criada na fase de criação da base de dados de apoio.

Passos 2 a 4 – Extração das chaves primárias das dimensões: Os passos referentes às dimensões, do tipo *Dimension Lookup/Update*, têm por função obter as chaves primárias de cada dimensão, para que as mesmas sejam armazenadas na tabela Fato, como chaves substitutas (*Surrogate Key*), as quais permitem um controle histórico dos dados. Assim sendo, essas chaves, garantirão a exatidão da relação dos registros com as métricas, na tabela Fato.

Passo 5 – Seleção dos dados pertinentes: O passo “**Seleciona_Valores**”, do tipo *Select Values*, seleciona os campos pertinentes à tabela Fato.

Passo 6 – Criação e carga do Fato: O passo “**Fato**”, do tipo *Table Output*, cria a tabela Fato, e realiza a carga das chaves substitutas e as métricas, na mesma.

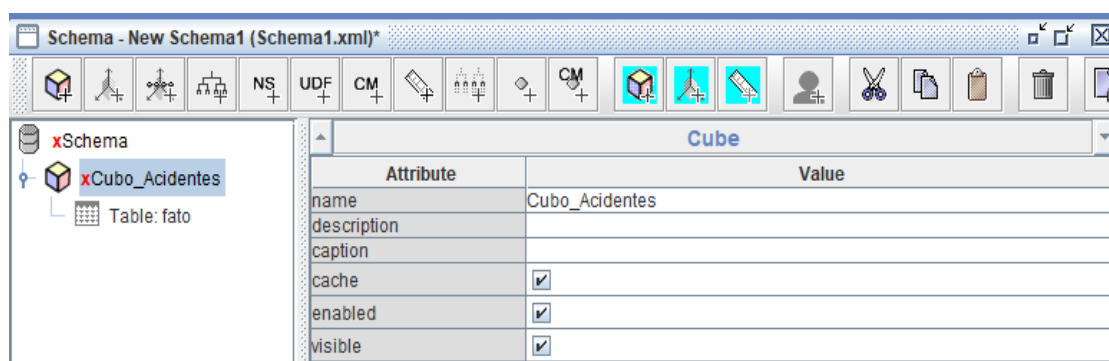
4.5. Construção do cubo OLAP

Ao finalizar a criação e carga das informações no DW, iniciou-se a construção do cubo OLAP, a qual foi necessária a utilização de uma outra ferramenta, da *suite Pentaho*.

A ferramenta utilizada foi a *Schema-Workbench* (SW), a qual permite a criação e o teste de esquemas de cubo OLAP, de forma visual, ou seja, através de uma interface gráfica; necessitando somente a existência de tabelas Fatos e Dimensões. Ao fim da criação dos esquemas, os mesmos são mapeados, de forma lógica, e arquivados em um arquivo XML (*Extensible Markup Language*), contendo assim todas as informações do DW. Além dessas funcionalidades mencionadas anteriormente, essa ferramenta permite que os esquemas sejam publicados, através de um servidor *Pentaho*, e assim, utilizados por alguma ferramenta de análise de dados, da *suite Pentaho*.

O primeiro passo para a construção de um cubo multidimensional, na ferramenta, após a criação de um novo esquema, é definir a tabela fato que foi construída no tópico anterior, como mostra a figura 14.

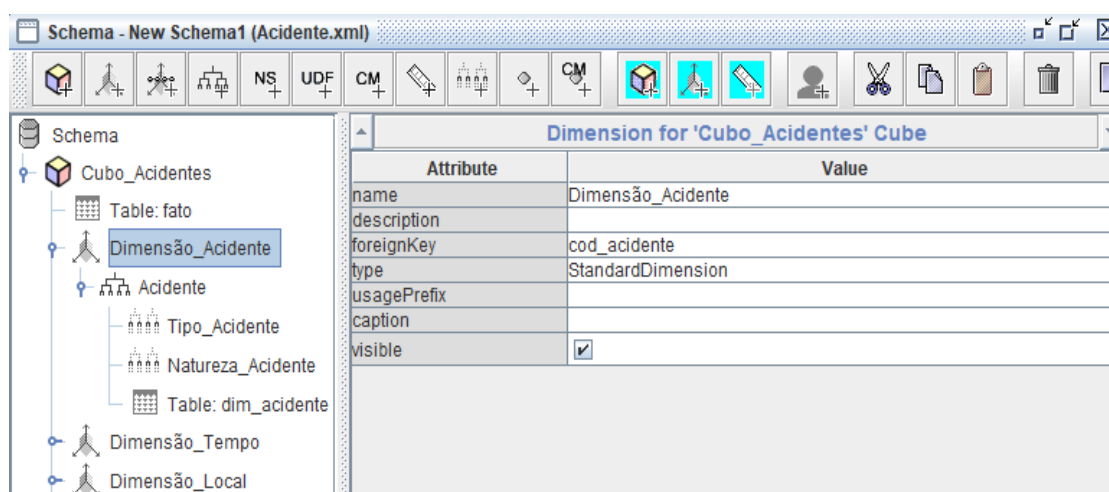
Figura 14 – Visualização da construção do Cubo OLAP



Fonte: Autoria própria (2020) – Extraído da ferramenta *Schema-Workbench*

Após a definição da tabela fato, deu-se início a criação, na ferramenta, das dimensões necessárias ao nosso cubo. Sendo assim, foram definidas as dimensões construídas em tópicos anteriores, em que seus dados serão reproduzidos nas dimensões, aqui construídas. À essas dimensões também foram definidos suas hierarquias e níveis necessários, como ilustra a figura 15.

Figura 15 – Visualização da criação das dimensões no cubo OLAP

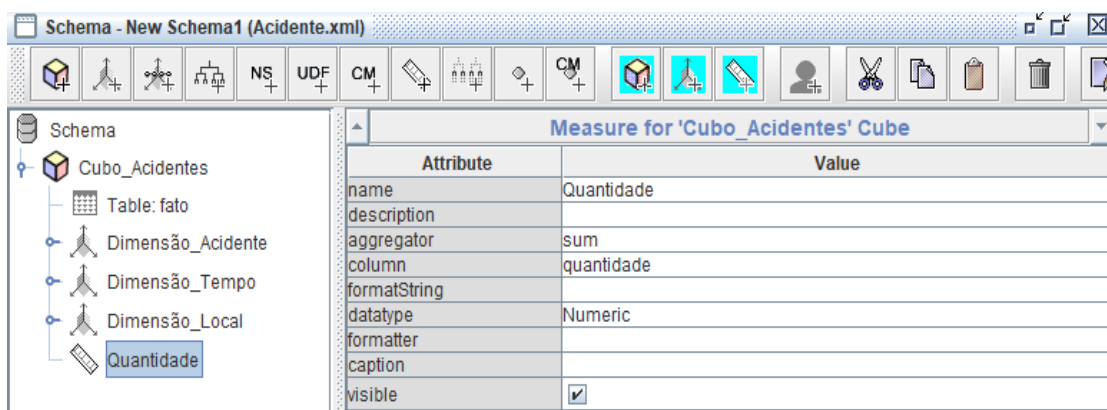


Fonte: Autoria própria (2020) – Extraído da ferramenta *Schema-Workbench*

Agora, após a definição das etapas anteriores, definiu-se a métrica necessária, que é responsável pela contabilização dos acidentes registrados, nesse caso, foi definida a métrica “**quantidade**”, construída no tópico 4.4.1. À métrica, é necessária a definição de um agregador (*aggregator*), que tem por função determinar como as métricas serão calculadas, nesse caso foi definido o agregador “**sum**”, o qual realizará a soma dos acidentes ocorridos em um determinado local e data, ou um determinado tipo de acidente e nos locais de ocorrência. Já no campo

datatype, foi escolhido o tipo *Numeric*, que vem definido por padrão pela ferramenta, como mostra a figura 16.

Figura 16– Visualização da definição da métrica Quantidade



Fonte: Autoria própria (2021) – Extraído da ferramenta *Schema-Workbench*

Para finalizar a criação do cubo OLAP e torná-lo apto a realizar análises, é necessário que o mesmo seja publicado no servidor *Pentaho*. Dessa forma, a ferramenta SW proporciona a validação e publicação, do esquema construído, no servidor e assim, tornando possível que o mesmo seja utilizado para as análises desejadas e a obtenção de resultados, por meio do uso de um ambiente gráfico, baseado na *web*, chamado *Pentaho User Console* (PUC).

Esse ambiente permite a análise de dados, a criação de relatórios interativos, relatórios de painel e a construção de painéis integrados para o compartilhamento de soluções de BI. É também possível integrar outras ferramentas ao PUC, além das ofertadas por padrão, através de extensões adicionadas ao ambiente.

5 RESULTADOS

Esta seção tem por finalidade responder as perguntas levantadas no capítulo 4, desse trabalho, a partir da análise multidimensional dos dados utilizados para essa pesquisa.

Como descrito no capítulo anterior, os dados utilizados, nesse estudo, se referem aos registros de acidentes de trânsito com e sem vítimas, da CTTU, obtidos a partir do porta de Dados Abertos da Prefeitura da Cidade do Recife. Sendo utilizadas as bases referentes aos anos de 2015, 2016, 2017, 2018 e 2019, as quais têm suas estruturas descritas na tabela 2.

Tabela 2 - Quantidade de registros por ano e total desses

Ano	Quantidade de registros
2015	7.273
2016	11.263
2017	11.758
2018	11.411
2019	12.062
Total	53.767

Para a realização dessas análises, fora utilizada a interface *Pentaho User Console* juntamente com o *plugin Saiku Analytics*, a partir do servidor *Pentaho Business Analytics*.

A primeira análise realizada, corresponde ao local com maior incidência de acidentes. Para tal, fora selecionada como métrica o campo “quantidade”, da tabela fato; os campos “bairro acidentes”, da dimensão local, no qual foi aplicado o filtro para os 10 (dez) maiores resultados por quantidade; e o campo “ano infração”, da dimensão tempo, para observarmos as maiores incidências por ano. Como mostra a figura 17.

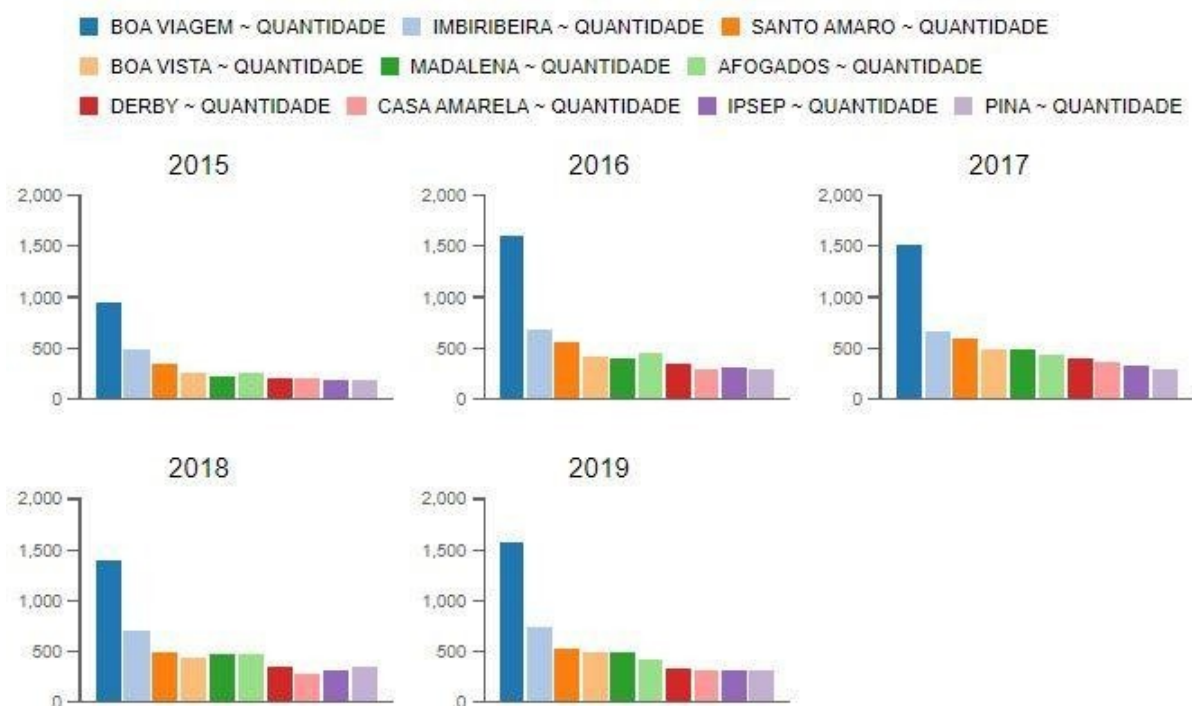
Figura 17 – Visualização dos parâmetros utilizados na consulta

Medidas	
QUANTIDADE	
Colunas	
BAIRRO	LOCAL_BAIRRO
Linhas	
ANO	TEMPO_ANO

Fonte: Autoria própria (2021) – Extraída da ferramenta *Pentaho Business Analytics*

Os resultados obtidos nessa consulta estão representados na figura 18, onde foram retornados os 10 bairros com maior incidência de acidentes. Esses resultados estão compostos pelos anos do acometimento dos acidentes (gráficos) e pelos bairros onde foram registrados (colunas), como disposto nas legendas dos gráficos.

Figura 18 – Resultado da consulta de bairros com mais registros por ano

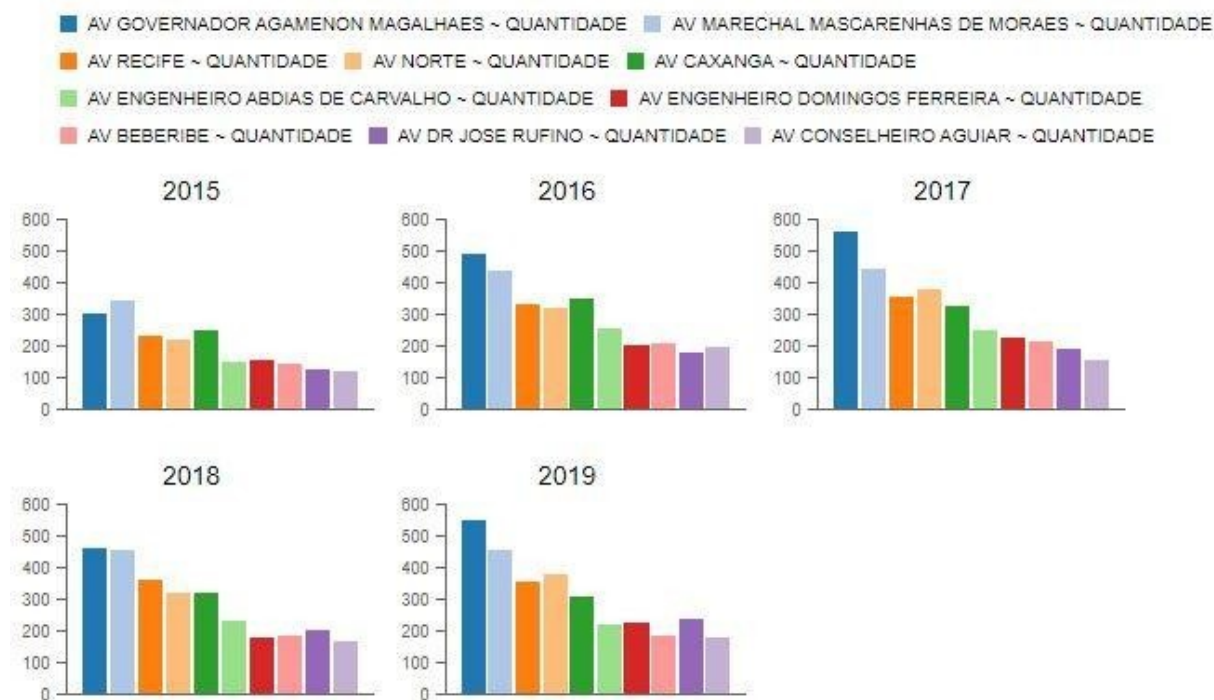


Fonte: Autoria Própria (2021) – Extraída da ferramenta *Pentaho Business Analytics*

Ao analisarmos esses dados, podemos observar que o bairro que possui maior número de registros de acidentes é o bairro de Boa Viagem, em todos os anos utilizados nesse estudo.

Para complementarmos a análise anterior, foi realizada a consulta para se identificar em qual rua registrou-se o maior número de acidentes por ano, como podemos observar na figura que se segue (Figura 19), podemos inferir que a Av. Agamenon Magalhães registra o maior número de acidentes, por 4 anos consecutivos (2016 a 2019), seguida da Av. Mascarenhas de Moraes, a qual registrou o maior índice de registros de acidentes no ano de 2015.

Figura 19 – Resultado da pesquisa de ruas com maiores índices de acidentes



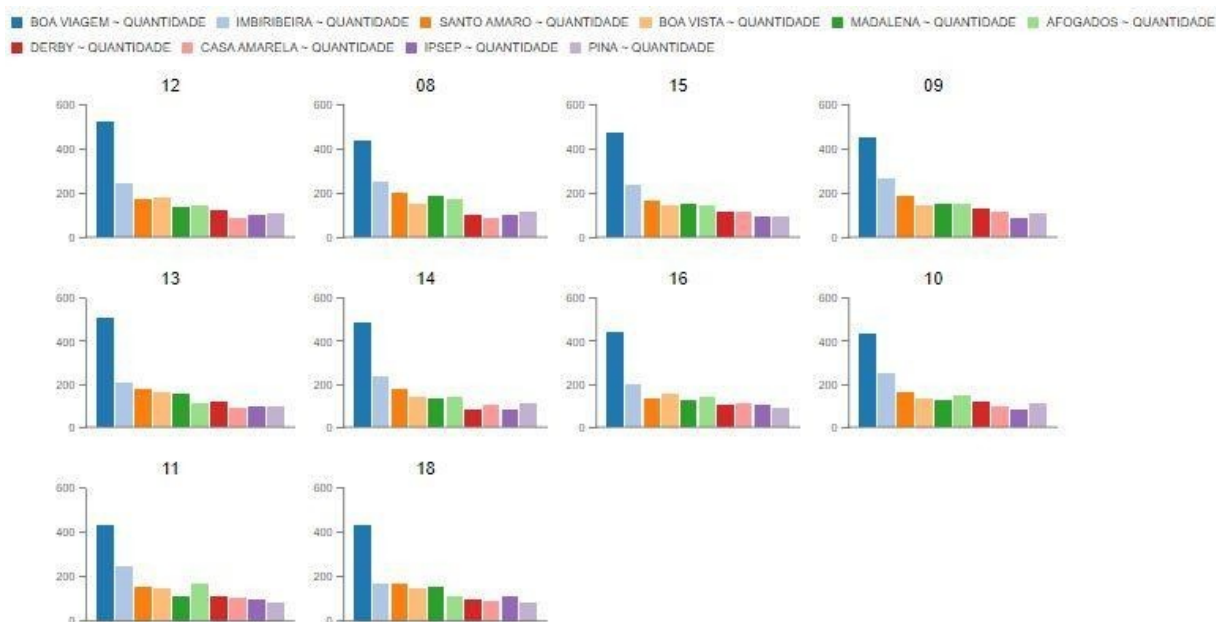
Fonte: Autoria própria (2021) – Extraída da ferramenta *Pentaho Business Analytics*

A segunda análise realizada, se refere ao período do dia em que há maior índice de registros de acidentes, a qual foram utilizados os parâmetros “hora”, da dimensão tempo, e os parâmetros “bairro” e “rua”, da dimensão local, os quais foram consultados separadamente.

Na figura 20 é exibido o gráfico resultante dessa consulta, o qual expõe o horário do dia e o bairro com maior número de registros de acidentes de trânsito. A partir dessa análise, podemos concluir que o horário com maior número de acidentes

registrados por bairro é entre às 12h e 15h, com o bairro de Boa Viagem com o maior índice de incidência.

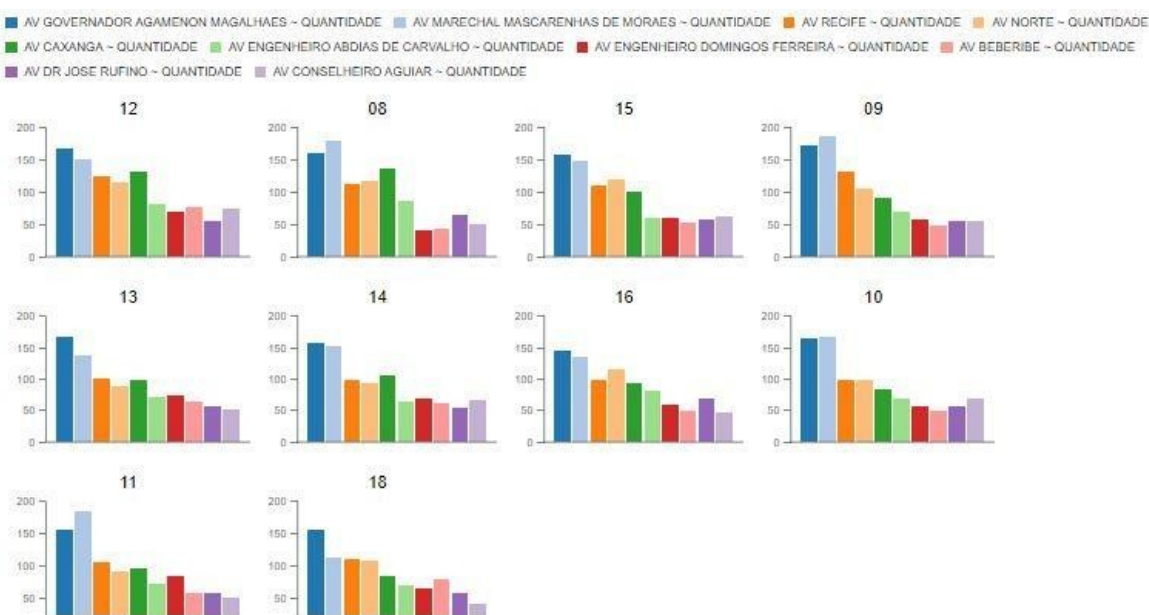
Figura 20 – Resultado do horário com maior número de acidentes por bairro



Fonte: Autoria própria (2021) – Extraída da ferramenta *Pentaho Business Analytics*

No gráfico a seguir apresenta a consulta de registros de acidentes de acordo com maior incidência por hora e rua.

Figura 21 – Resultado do horário com maior número de acidentes por rua



Fonte: Autoria própria (2021) – Extraída da ferramenta *Pentaho Business Analytics*

Como resultado dessa consulta podemos observar que as avenidas Agamenon Magalhães e Mascarenhas de Moraes, possuem os maiores índices de acidentes nos horários, entre às 12h e 15h, respectivamente. Assim, podemos inferir que nesse horários há um maior fluxo de veículos nas ruas.

A próxima análise se refere ao período do ano com maior incidência de acidentes. Para essa análise foram utilizados os parâmetros mês e ano da dimensão tempo, e a métrica quantidade.

A figura a seguir (Figura 22) nos mostra o resultado dessa consulta, revelando que o mês com o maior índice de acidentes é o mês de Dezembro, seguido pelos meses de Outubro, Novembro e Agosto, respectivamente.

Figura 22 – Resultado dos meses com maior índice de acidentes por ano



Fonte: Autoria própria (2021) – Extraída da ferramenta *Pentaho Business Analytics*

Para complementar essa consulta, fora realizada uma consulta para determinar os dias da semana que mais ocorrem registros de acidentes na cidade.

A figura que se segue (Figura 23) revela o resultado dessa consulta e que os registros de acidentes concentram-se entre quarta-feira e sexta-feira, sendo esse último o dia da semana com o maior número de registros de acidentes.

Figura 23 – Resultado dos dias da semana com mais registros por ano



Fonte: Autoria própria (2021) - Extraída da ferramenta *Pentaho Business Analytics*

A quarta e última análise, refere-se aos tipos de ocorrências que mais ocorreram no período analisado.

Na figura a seguir traz o resultado dessa consulta, a qual nos revela que o tipo de acidente com maior índice de registros, ao longo do período estudado, é o tipo colisão, que foi registrado com maior frequência no ano de 2017.

Figura 24 – Resultado do tipo de acidentes com mais registros por ano



Fonte: Autoria própria (2021) – Extraída da ferramenta *Pentaho Business Analytics*

6 CONCLUSÃO

Esse trabalho possibilitou desde o desenvolvimento à aplicação de um ambiente de *Business Intelligence* para a análise de periodicidade de acidentes constantes no conjunto de dados dos Registros de Acidentes com e sem vítimas da cidade do Recife.

Podemos dividir o trabalho em duas etapas, sendo a primeira etapa reservada à introdução dos conceitos do BI, e a segunda reservada à aplicação desses conceitos em um ambiente de BI, com a análise dos dados, transformando-os em informações visuais, através dos gráficos gerados.

A aplicação desses conceitos inicia-se com a análise dos dados na fase de planejamento e estendendo-se durante todo o processo de desenvolvimento do ambiente de *Business Intelligence*, através das ferramentas de BI, baseadas na suíte *Pentaho*; as quais foram de suma importância para o desenvolvimento do ambiente, por facilitarem o trabalho em cada fase; e possibilitando que os conceitos pudessem ser aplicados, auxiliando no tratamento da massa de dados, até o atingimento dos resultados esperados, através da visualização das informações obtidas por meio das consultas realizada.

Após o desenvolvimento do ambiente de BI, foram realizadas as análises necessárias, a partir do cruzamento das informações geradas, as quais resultaram os comportamentos dos acidentes ocorridos na cidade do Recife, entre os anos de 2015 a 2019; revelando os locais com maiores incidências de acidentes, os períodos com os maiores índices de acidentes, desde as horas até os meses, e por fim os tipos de acidentes mais registrados, nesse período.

Uma deficiência encontrada, ao longo do desenvolvimento do projeto, se refere ao arquivo de dados, do ano de 2015, o qual só possui registros a partir do segundo semestre do referido ano. Outras limitações encontradas foram o elevado número de ausência de registros de informações acerca do acidente e a falta de padronização no registro dessas informações (quando informadas), como pudemos observar algumas das maneiras que as informações foram registradas, as quais foram tratadas ao longo do processo de ETL.

Sugere-se a utilização desse trabalho, como base para futuros trabalhos, que utilizem outros conjuntos de dados extraídos do portal de dados abertos da cidade do Recife, como também a construção de um mapa de calor, o qual poderá auxiliar na exibição dos locais com maiores índices de acidentes.

REFERÊNCIAS

- ANDRADE, Eduardo Filipe Ferreira. **Desenvolvimento de motor de BI (Business Intelligence)**. 2007. Disponível em: <https://repositorio.ul.pt/bitstream/10451/4662/1/ulfc096105_tm_Eduardo_Andrade.pdf>. Acesso em: 11 out. 2021.
- BOUMAN, Roland; DONGEN, Jos Van. **Pentaho Solutions: Business Intelligence e Data Warehousing com Pentaho e Mysql**. [S.l: s.n.], 2009. .9780470484326.
- CARLOS, José; RIBEIRO, Manuella Maia; MATHEUS, Ricardo. **Dados Governamentais Abertos E Seus Impactos Sobre Os Conceitos E Práticas De Transparência No Brasil**. [S.l: s.n.], 2010. 9 v. Disponível em: <<https://www.portalseer.ufba.br/index.php/ppgau/article/view/5111/3700>>. Acesso em: 21 mar. 2017.
- DA SILVA, Erivania Pereira. **Bi: Business Intelligence Como Metodologia Na Organização Das Informações E No Alinhamento Da Gestão Do Conhecimento**. 2012. Disponível em: <<https://dspace.mackenzie.br/bitstream/handle/10899/206/ERIVANIA PEREIRA DA SILVA1.pdf?sequence=1&isAllowed=y>>. Acesso em: 11 out. 2021.
- FISHER, Erik; MAHAJAN, Roop L. **Contradictory Intent? US Federal Legislation On Integrating Societal Concerns Into Nanotechnology Research And Development**. Science and Public Policy v. 33, n. 1, p. 5–16 , 2006.
- KIMBALL, Ralph; ROSS, Margy. **The data warehouse toolkit: the complete guide to dimensional modelling**. [S.l: s.n.], 2011. 1–447 p.
- LIRA, Márcio Angelo Bezerra. **Uma Abordagem Para Enriquecimento Semântico De Metadados Para Publicação De Dados Abertos**. [S.l: s.n.], 2014. Disponível em: <www.cin.ufpe.br/~posgraduacao>. Acesso em: 13 jun. 2018.
- LUSTOSA, Rodrigo Bastos. **Processo De Desenvolvimento Participativo De Sistema De Data Warehouse: Uma Aplicação No PROGER**. [S.d.]. Disponível em: <<http://livros01.livrosgratis.com.br/cp118027.pdf>>. Acesso em: 30 maio 2018.
- MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de Data Warehouse: uma visão multidimensional**. 2. ed. São Paulo: Érica, 2006.
- MAZZOLA, Irany Salgado. **Projeto Para Modelagem De Data Warehouse Dimensional**. , [S.d.]. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/83465/184713.pdf?sequence=1>>. Acesso em: 30 maio 2018.
- NEGASH, Solomon. **Business Intelligence**. n. January 2003 , 2015.
- NOVAIS, Ramon Ramos de Castro. **Modelagem Dimensional**. , 2012. Disponível em: <<http://www.fatecsp.br/dti/tcc/tcc00071.pdf>>. Acesso em: 30 maio 2018.

OLIVERIO, Marcio Araujo. **Governo Aberto como Ferramenta de Comunicação Entre o Governo e o Cidadão.** , 2011. Disponível em: <http://www.inpecc.pro.br/media/uploads/pesquisas/governo_aberto_como_ferramenta_de_comunicacao_entre_o_governo_e_o_cidadao.pdf>. Acesso em: 25 abr. 2017.

OLSZAK, Celina M; ZIEMBA, Ewa. **Critical Success Factors for Implementing Business Intelligence Systems in Small and Medium Enterprises on the Example of Upper Silesia, Poland.** Interdisciplinary Journal of Information, Knowledge, and Management v. 7 , 2012. Disponível em: <<http://www.ijikm.org/Volume7/IJIKMv7p129-150Olszak634.pdf>>.

PEDROSO, Louise; TANAKA, Asterio; CAPELLI, Claudia. **A Lei De Acesso À Informação Brasileira E Os Desafios Tecnológicos Dos Dados Abertos Governamentais. IX Simpósio Brasileiro de Sistemas de Informação** v. 0, p. 523–528 , 2013. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2013/0048.pdf>>.

RIBEIRO, Claudio Jose Silva; ALMEIDA, Reinaldo Figueiredo. **Dados Abertos Governamentais (Open Government Data): Instrumento Para Exercício De Cidadania Pela Sociedade.** , 2011. Disponível em: <<http://arq.3rengtt.com.br/wp-content/uploads/2015/09/EnancibXII-RibeiroAlmeida.pdf>>. Acesso em: 21 mar. 2017.

RIBEIRO, Claudio Jose Silva; PEREIRA, Durval Vieira. **Em Busca De Uma Interoperabilidade Semântica Na Publicação De Dados Abertos Governamentais: A Experiência Sobre Os Dados Da Previdência Social.** , 2013. Disponível em: <<http://enancib.ibict.br/index.php/enancib/xivenancib/paper/viewFile/4378/3501>>. Acesso em: 1 abr. 2017.

SHARDA, Ramesh *et al.* **Business Intelligence And Analytics: Systems For Decision Support.** Tenth Edit ed. [S.l: s.n.], 2015. Disponível em: <https://d1wqtxts1xzle7.cloudfront.net/62386890/business_intelligence20200316-50198-128rsmo-with-cover-page-v2.pdf?Expires=1633533664&Signature=YYfq2~~CTRgh9nGc-bkC42g3Q2nJiLHgLI2aVQf3zzdyDSggH4sYXdZ~DYrq96zXK0L9m-OQcRTxci9HvZThOyAm-L3XawK4e3BKoo5NjObyk3gq>. Acesso em: 6 out. 2021.

SILVA, A.C. *et al.* **Uso Da Ferramenta Pentaho BiServer: Analisando A Violencia Contra Mulher, No Estado Do Rio Grande Do Sul.** Disponível em: <<http://oro.open.ac.uk/45861/1/paper.pdf>>. Acesso em: 11 out. 2021.

SILVA, Victor Luiz de Oliveira. **Aplicação De Uma Ferramenta De B.I. Para Analisar O Investimento Das Universidades Federais.** UNIVERSIDADE FEDERAL DE GOIÁS, GOIÂNIA, 2019. Disponível em: <[https://repositorio.bc.ufg.br/bitstream/ri/17996/5/TCCG - Engenharia da Computação - Victor Luiz de Oliveira e Silva - 2019.pdf](https://repositorio.bc.ufg.br/bitstream/ri/17996/5/TCCG_-_Engenharia_da_Computação_-_Victor_Luiz_de_Oliveira_e_Silva_-_2019.pdf)>. Acesso em: 11 out. 2021.

VARGAS, Dimi Antoni; FROZZA, Angelo Augusto. **Como Vai A Câmara? Uma Proposta De Uso De Dados Abertos.** , 2015.