PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Pós-Graduação em Analytics e Business Inteligence

André Felipe Oliveira Moraes

BOLETIM DE OCORRÊNCIA DE ACIDENTE DE TRÂNSITO COM VITIMA

Belo Horizonte

2023

MÓDULO A – DISCOVERY E PROJETO DE SOLUÇÃO

Contexto do Projeto

Contexto Organizacional: A prefeitura de Belo Horizonte traz em sua base de dados um dataset dos boletins de ocorrência que foram criados durante os acidentes entre veículos automotores que tiveram vítimas em toda cidade de Belo Horizonte.

Motivação: Tais dados podem ser utilizados pela própria prefeitura para melhorar os indicies de acidentes ou por uma companhia de seguros para determinar os lugares onde mais ocorrem acidentes e, posteriormente, fazer uma análise de risco com essas informações no campo das ciências atuárias.

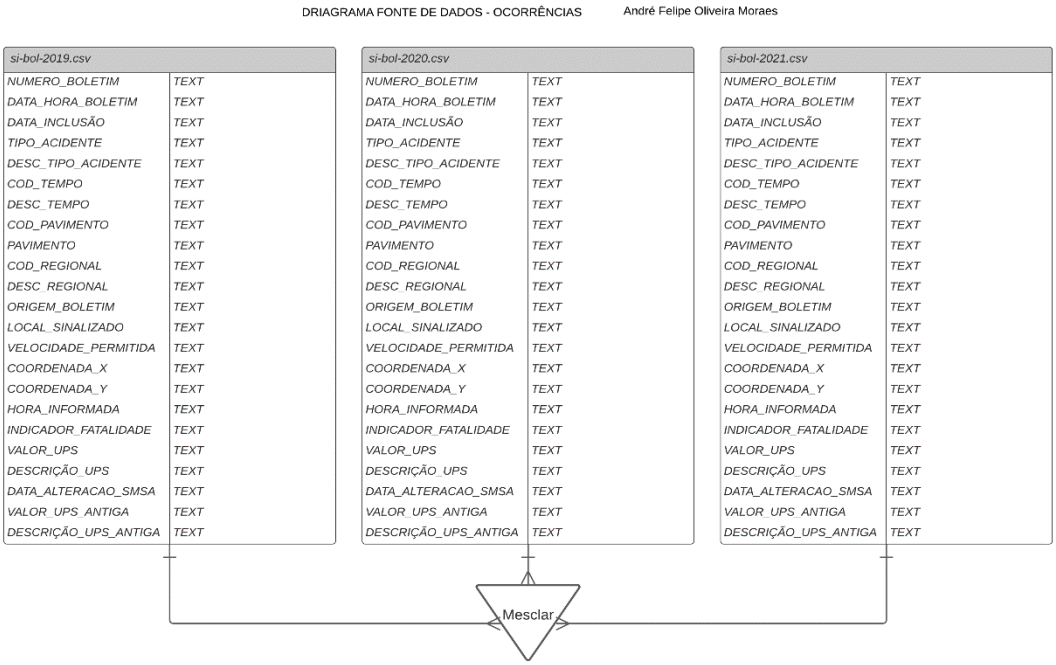
Objetivos estratégicos: Identificar as principais vias que possuem mais acidentes, ocorrências por ano, os bairros com maior número de acidentes, os dias do ano que temos mais acidentes, identificar o tipo de acidente mais comum.

Stakeholders: Coordenadores, gerente e diretores de uma determinada empresa de seguros que possuem clientes em Belo Horizonte.

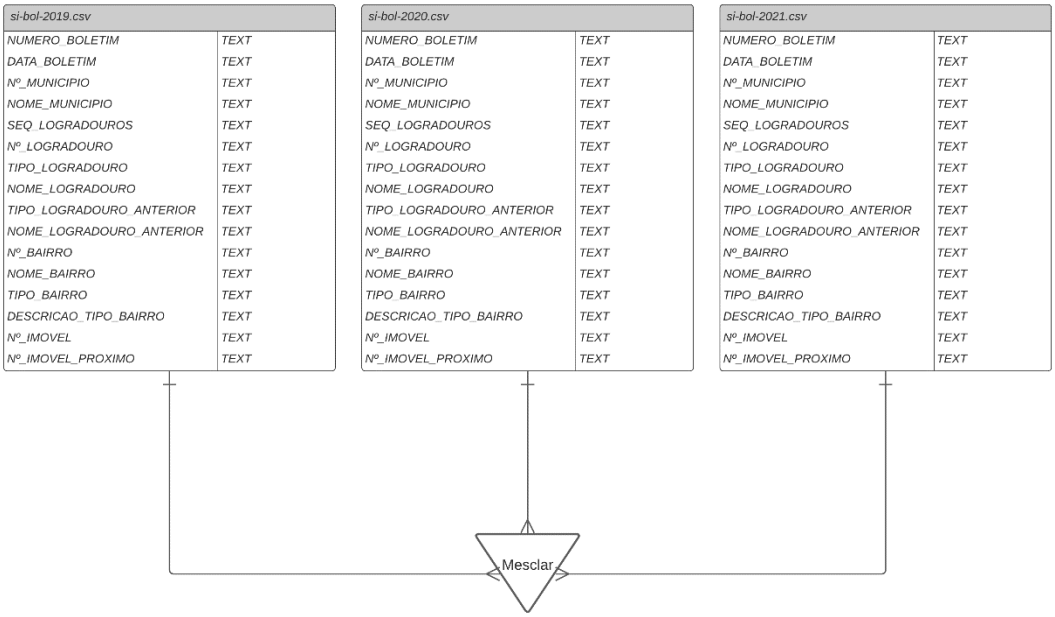
Fonte de dados: A fonte de dados foi toda extraída do site da prefeitura de Belo Horizonte (https://dados.pbh.gov.br/dataset/relacao-de-ocorrencias-de-acidentes-de-transito-com-vitima; https://dados.pbh.gov.br/dataset/relacao-dos-logradouros-dos-locais-de-acidentes-de-transito-com-vitima; https://dados.pbh.gov.br/dataset/relacao-dos-veiculos-envolvidos-nos-acidentes-de-transito-com-vitima) e contempla os anos de 2019, 2020 e 2021.

Modelo de dados

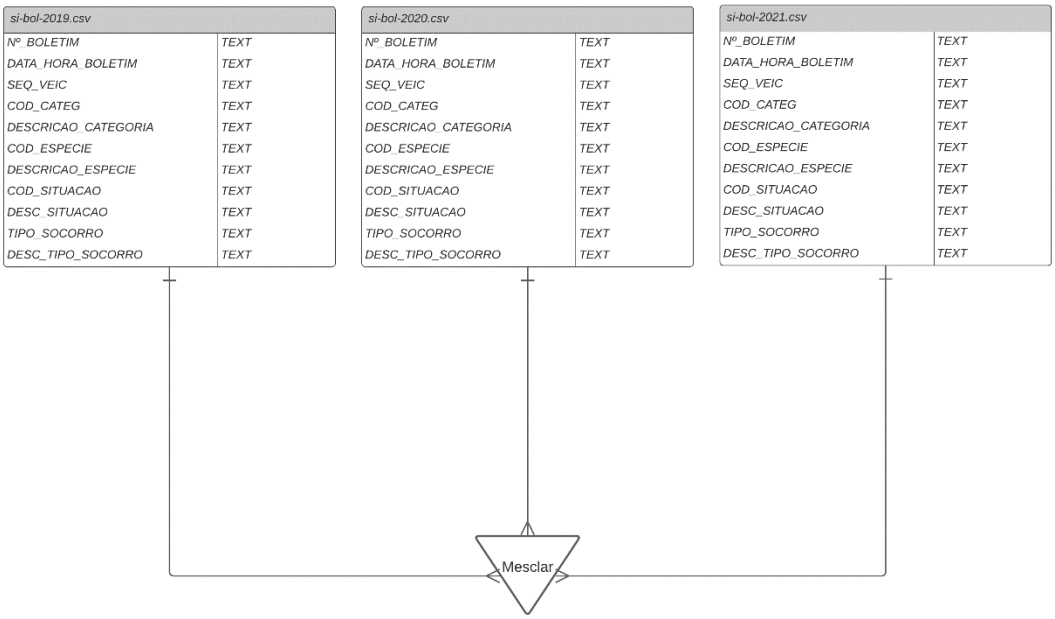
Fonte de Dados:



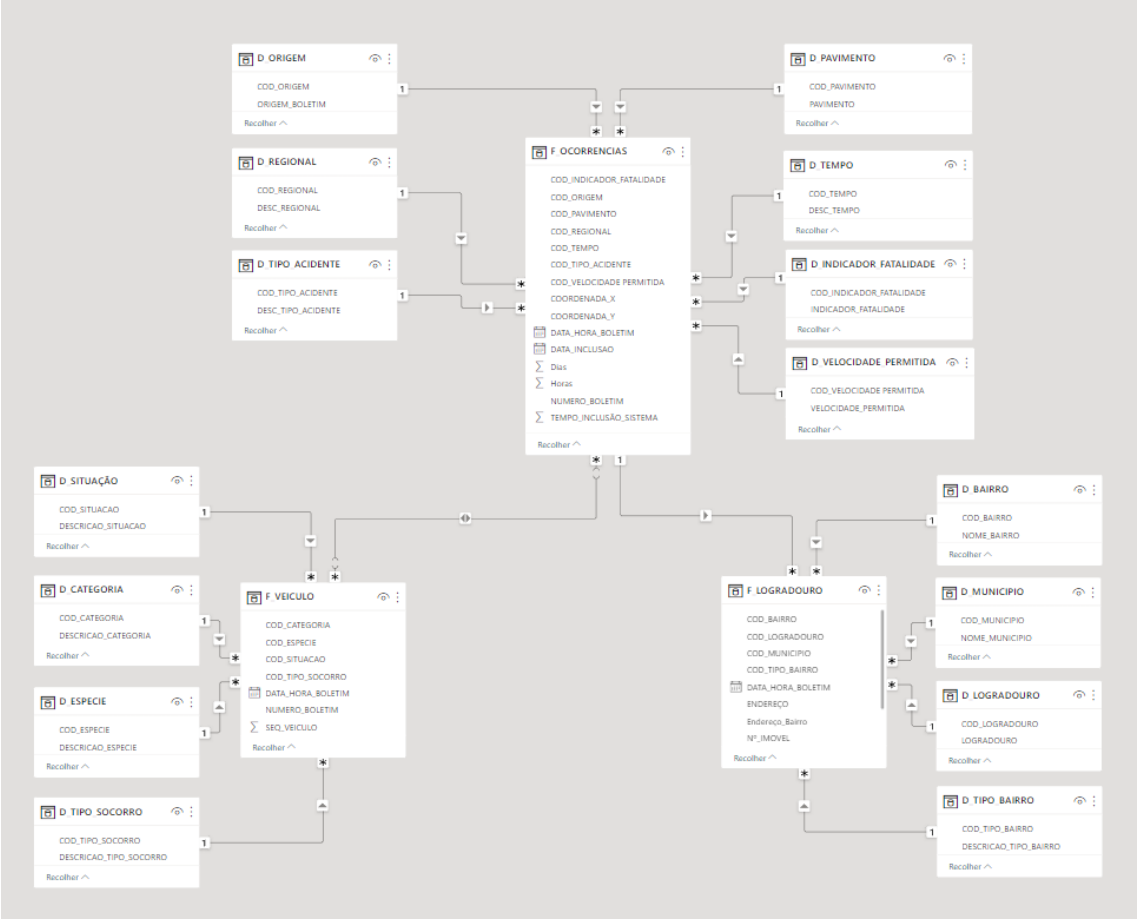
DRIAGRAMA FONTE DE DADOS - LOGRADOURO André Felipe Oliveira Moraes



DRIAGRAMA FONTE DE DADOS - VEICULOS André Felipe Oliveira Moraes



Base Dimensional:



As tabelas fato no projeto são as tabelas:

F_OCORRENCIAS, F_VEICULO e F_LOGRADOURO.

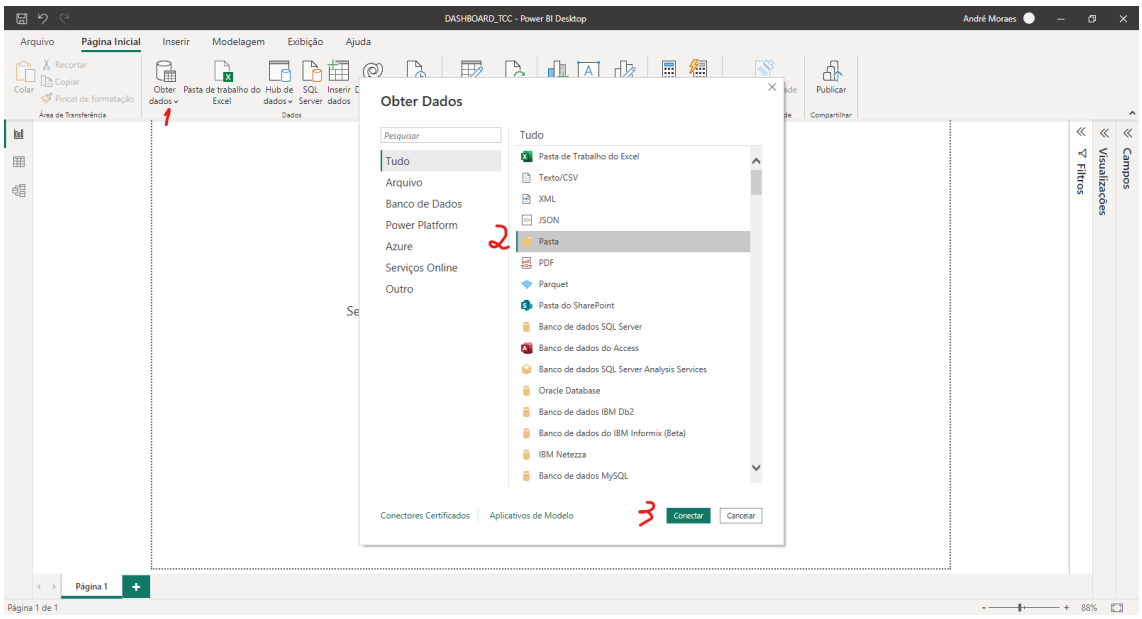
As dimensões são as tabelas:

D_ORIGEM, D_PAVIMENTO, D_REGIONAL, D_TEMPO, D_TIPO_ACIDENTE, D_INDICADOR_FATALIDADE, D_VELOCIDADE_PERMITIDA, D_SITUAÇÃO, D_CATEGORIA, D_ESPECIE, D_TIPO_SOCORRO, D_BAIRRO, D_MUNICIPIO, D_LOGRADOURO e D_TIPO_BAIRRO.

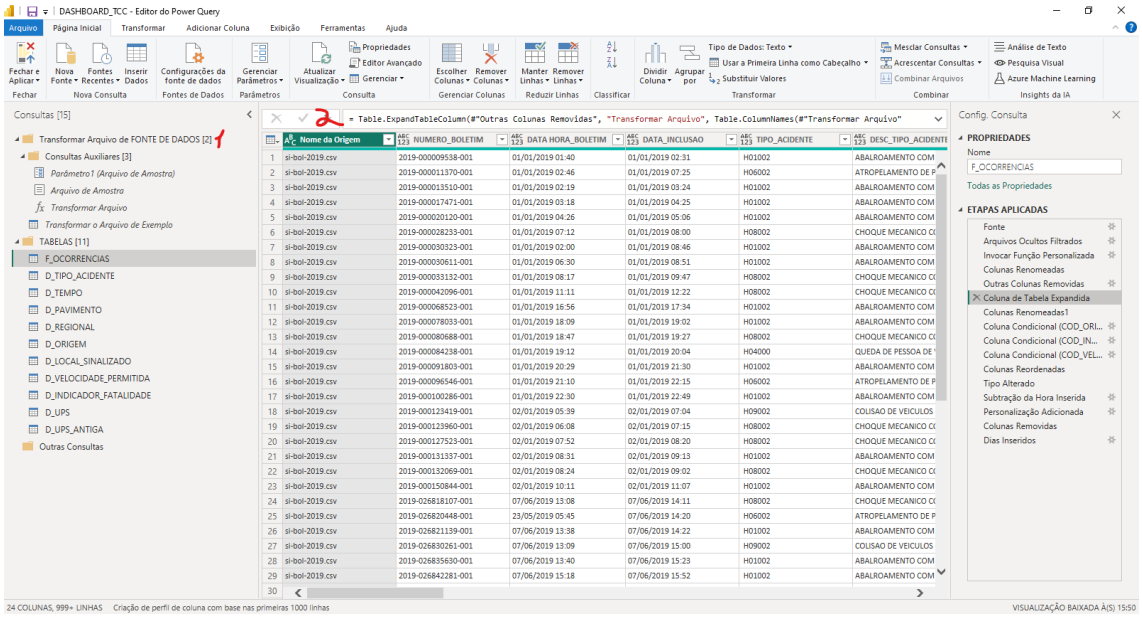
A tabela fato possui a sigla **F** no começo do nome e as dimensões a sigla **D** no começo do nome para facilitar a sua identificação, sendo que o modelo dimensional é o Star Schema ou esquema de estrela.

Processos de Integração, Tratamento e Carga de Dados

Ingestão de Dados e Processos ETL – A ingestão de dados e os processos ETL ocorreram diretamente na ferramenta Microsoft Power BI no qual foi possível o tratamento e carga dos dados utilizando o Power Query. A ingestão de dados foi feita utilizando a opção de pasta, do qual faz a leitura dos arquivos que estão na pasta na qual estão salvas a fonte de dados que são os arquivos sl-bol-2019.csv, sl-bol-2020.csv e sl-bol-2021.csv, conforme a seguir:



Com isso, o Power BI gerou uma transformação de arquivos que conseguiram ler os arquivos da pasta no formato de csv e os agrupou cada base de dados em uma única tabela que foram denominadas como **F_OCORRENCIAS, F_VEICULO e F_LOGRADOURO** que gerou a coluna “Nome da Origem” que é referente ao arquivo da origem e a cada linha presente no arquivo, conforme a seguir:



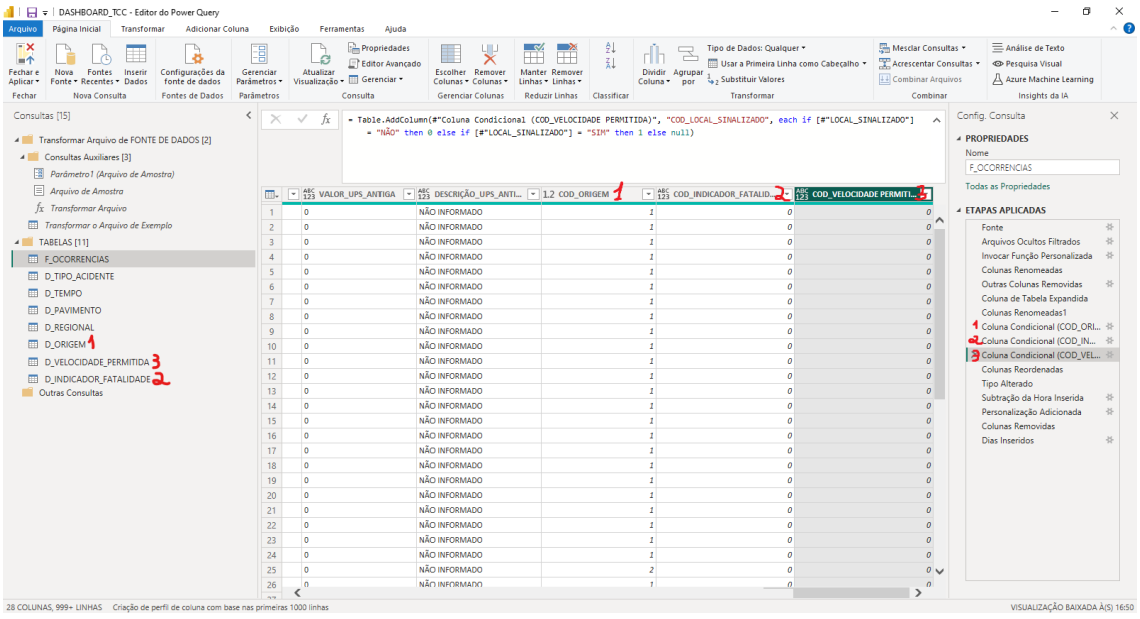
ETL F_OCORRÊNCIAS:

Em seguida, foram renomeadas algumas colunas que apresentavam um espaçamento em branco na frente do nome e que poderiam atrapalhar no desenvolvimento de medidas e análises ao decorrer do trabalho. Em suma, todas as colunas tiveram seu nome alterado para retirar esse espaço, com exceção da coluna Nome da Origem que não tinha esse problema e da coluna " TIPO_ACIDENTE", que passou a ter no seu inicio o sufixo “COD_” para manter a padronização da base de dados.

Foi adicionado uma coluna condicional denominada como “COD_ORIGEM”, tal coluna foi criada para atribuir ID a coluna “ORIGEM_BOLETIM” atribuindo os códigos 0, 1 e 2 para cada situação, sendo o 1 para “POLÍCIA MILITAR”, 2 para “POLÍCIA CIVIL” e 0 para “NI”, ou seja, não informado. Com isso, foi possível criar a dimensão D_ORIGEM.

Também, foi adicionado outra coluna condicional denominada como “COD_INDICADOR_FATALIDADE”, do qual atribuiu ID a coluna “INDICADOR_FATALIDADE” com os códigos 0 e 1, sendo o 0 para “NÃO” e o 1 para “SIM”. Isso possibilitou a criação da dimensão D_INDICADOR_FATALIDADE.

Por fim, foi adicionado uma nova coluna condicional nomeada como “COD_VELOCIDADE_PERMITIDA”, da qual criou ID para a coluna “VELOCIDADE_PERMITIDA” que atribuiu os códigos 0,1,2,3,4,5,6,7 e 8, sendo o 0 de “0”, 1 de “20”, 2 de “30”, 3 de “40”, 4 de “50”, 5 de “60”, 6 de “70”, 7 de “80” e 8 de “110”. Portanto, essas seriam as medidas dos agentes de trânsito da velocidade da via, porém, cabe ressaltar que não existe via com velocidade permitida igual a 0 KM/H e, portanto, o dado pode ser tido como não informado pelo agente. Cabe ressaltar que com essa coluna foi possível criar a dimensão D_VELOCIDADE_PERMITIDA.



Em seguida, as colunas da tabela F_OCORRENCIAS foram reordenadas para que as colunas condicionais que foram acrescentadas na base ficassem perto da coluna correspondente para facilitar as etapas seguintes.

Foram alterados os tipos dos dados, visto que todos estavam como genérico que é o “ABC123” que o Power BI atribui automaticamente e passaram a ser do tipo Texto para todos que são Ids como “NUMERO_BOLETIM” e as colunas que começavam como “COD_”, além de alterar para o tipo DateTime os campos “DATA_HORA_BOLETIM” e “DATA_INCLUSAO”.

Em seguida, foi criado uma nova coluna denominada como TEMPO_INCLUSAO_SISTEMA do qual fazia a subtração da DATA_INCLUSAO com a DATA_HORA_BOLETIM e gerava o tempo de duração entre as duas datas.

Também, houve um tratamento na duração para caso a mesma apresentação tempo negativo, ou seja, menor que 0 e, portanto, ele passou a representar 0, visto que se a ocorrência foi incluída antes do seu acontecimento se trata de um erro de preenchimento.

Na etapa seguinte as colunas de Descrição das dimensões foram excluídas, além da coluna “Nome da Origem” que não era mais necessário, visto que os dados já estavam mesclados entre si e o número da ocorrência já conta com o ano do dado no seu início e as colunas VALOR_UPS e VALOR_UPS_ANTIGA que apresentavam valores igual a 0, sendo assim, sem nenhuma atribuição prática para o projeto. O campo “DATA_ALTERACAO_SMSA” que estava com o valor de “00/00/0000” para todos os registros e, portanto, presume-se que seja um dado para saber quando a base de dados passou por uma alteração e por não haver registros válidos entende-se que não sofreu alteração, logo é um dado irrelevante para a análise.

Por fim, foi acrescentado a coluna de Dias que é justamente o número de dias que cada ocorrência demorou para entrar no sistema.

ETL D_TIPO_ACIDENTE, D_TEMPO, D_PAVIMENTO e D_REGIONAL,

Essas tabelas tiveram um ETL praticamente igual, com isso resolvi reuni-los aqui para explicar os pontos em comum entre eles. Essas dimensões foram criadas a partir da duplicação da tabela fato F_OCORRENCIAS, com isso as etapas iniciais são as mesmas da tabela de origem.

Tal opção foi feita para que os dados não se perdessem na hora de excluir as colunas de descrição da tabela fato, visto que o Power BI tem a opção de “Duplicar” e “Referenciar”, sendo que em “Duplicar” ele faz a duplicação da tabela selecionada com todas as etapas que ela sofreu e na opção de “Referenciar” ela apenas utiliza a tabela depois de todo o tratamento sem as etapas de tratamento que ela sofreu.

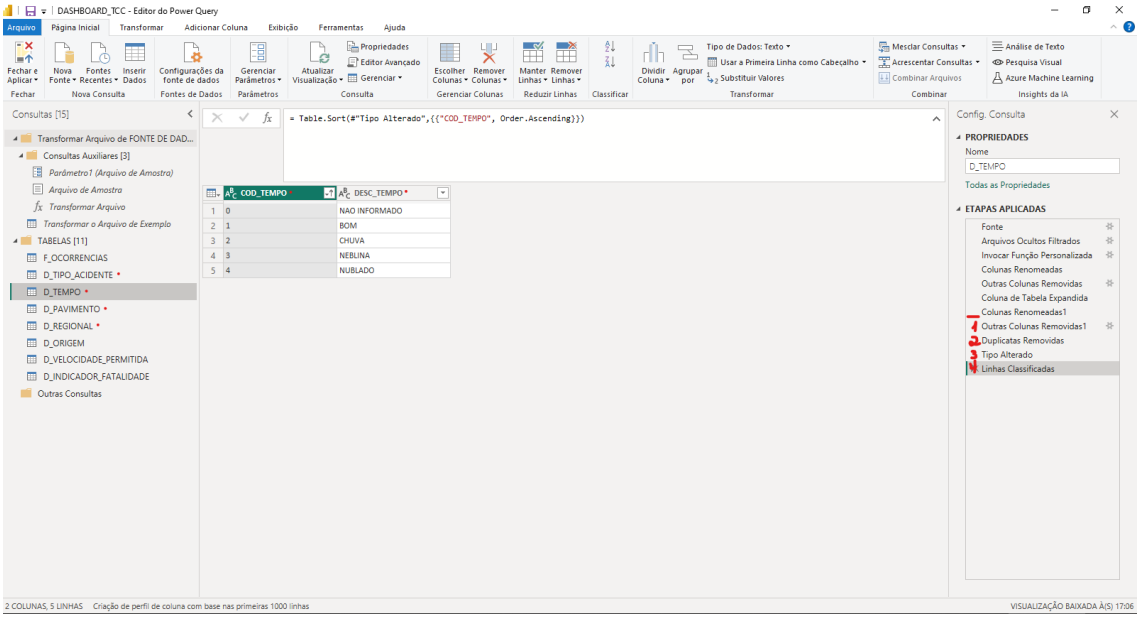
Portanto, até a etapa “Colunas Renomeadas1” as dimensões tem o mesmo ETL que a tabela fato, ou seja, temos todo o trabalho da leitura dos arquivos da pasta e a renomeação para adequação das informações. Em seguida, são excluídas todas as colunas que não fazem parte da dimensão e são mantidos a coluna de ID que tem como inicio “COD_” e a sua respectiva descrição.

Por conseguinte, temos a remoção das duplicatas dos registros o que mantem apenas o ID e a descrição distinto da base inteira e possibilidade fazer a conexão dimensional de um para muitos.

Em seguida, as colunas tem os seus tipos alterados de genérico “ABC123” para o tipo texto “ABC”, apesar dos Ids contarem com números em sua maioria, estes não sofrerão qualquer tipo de operação matemática e, por isso, passam a serem texto.

Na tabela D_REGIONAL tem um passo a mais que as outras tabelas, pois o valor que estava em branco foi substituído por “NI”, ou seja, não informado.

Por fim, foi feito a classificação das linhas em ordem crescente para facilitar a identificação dos registros e de seus Ids na hora da leitura e manuseio dos dados.



ELT D_ORIGEM, D_LOCAL_SINALIZADO, D_VELOCIDADE_PERMITIDA E D_INDICADOR_FATALIDADE

Assim como os ETLs que foram feitos nas outras dimensões, essas tabelas foram criadas a partir da duplicação da tabela fato F_OCORRENCIAS, com isso as etapas iniciais são as mesmas da tabela de origem.

Tal opção foi feita para que os dados não se perdessem na hora de excluir as colunas de descrição da tabela fato, visto que o Power BI tem a opção de “Duplicar” e “Referenciar”, sendo que em “Duplicar” ele faz a duplicação da tabela selecionada com todas as etapas que ela sofreu e na opção de “Referenciar” ela apenas utiliza a tabela depois de todo o tratamento sem as etapas de tratamento que ela sofreu.

Nessas tabelas todas possuem a etapa de criação de coluna condicional, visto que na base de dados original não havia as colunas de Id para esses assuntos, portanto, todos os ETLs dessas tabelas possuem esse passo a mais que é apenas para a criação da coluna de ID.

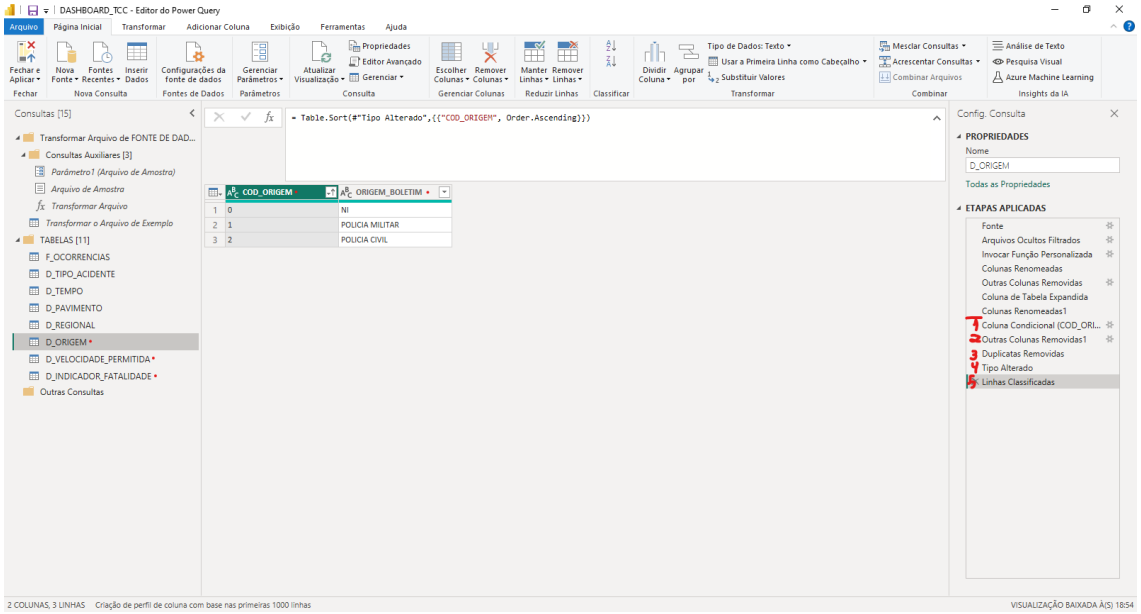
Portanto, até a etapa “Colunas Renomeadas1” as dimensões tem o mesmo ETL que a tabela fato, ou seja, temos todo o trabalho da leitura dos arquivos da pasta e a renomeação para adequação das informações. Em seguida, são excluídas todas as colunas que não fazem parte da dimensão e são mantidos a coluna de ID que tem como início “COD_” e a sua respectiva descrição.

Por conseguinte, temos a remoção das duplicatas dos registros o que mantem apenas o ID e a descrição distinto da base inteira e possibilidade fazer a conexão dimensional de um para muitos.

Em seguida, as colunas tem os seus tipos alterados de genérico “ABC123” para o tipo texto “ABC”, apesar dos Ids contarem com números em sua maioria, estes não sofrerão qualquer tipo de operação matemática e, por isso, passam a serem texto.

Na tabela D_REGIONAL tem um passo a mais que as outras tabelas, pois o valor que estava em branco foi substituído por “NI”, ou seja, não informado.

Por fim, foi feito a classificação das linhas em ordem crescente para facilitar a identificação dos registros e de seus Ids na hora da leitura e manuseio dos dados.

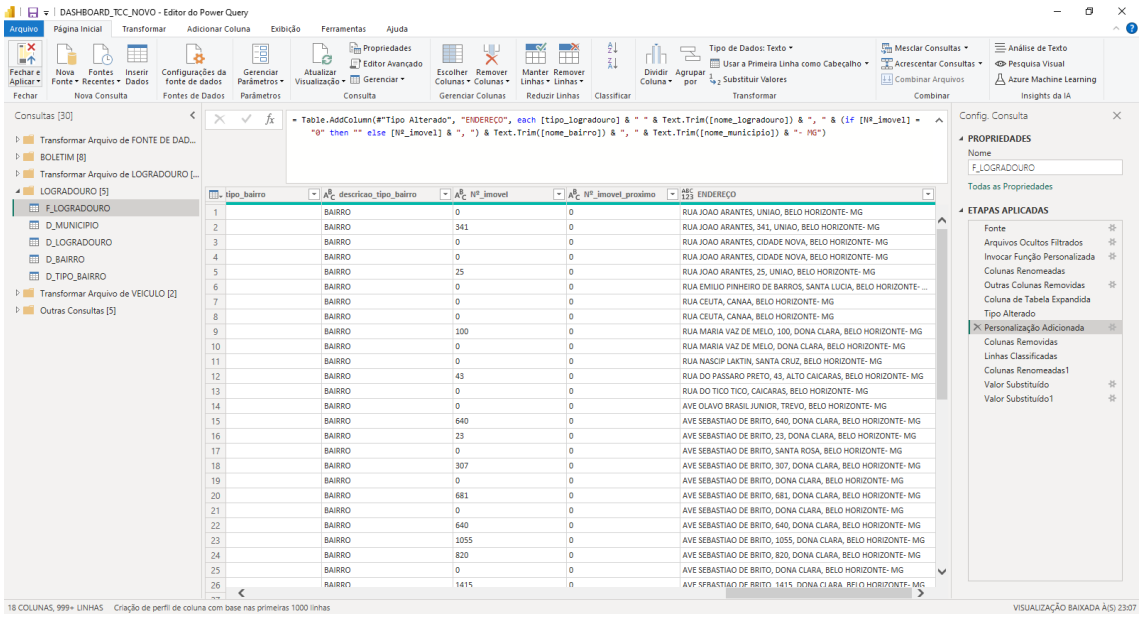


ETL F_LOGRADOURO:

Assim como F_OCORRENCIAS, a tabela fato F_LOGRADOURO passou pelas mesmas transformações para a leitura da base de dados na pasta de logradouro.

Em seguida, o tipo de todas as colunas foi alterado, visto que todos estavam como genérico que é o “ABC123” que o Power BI atribui automaticamente e passaram a ser do tipo Texto para todos com exceção para data_boletim que passou a ser datetime e seq_logradouros que passou a ser do tipo int.

Foi adicionado a coluna denominada como “ENDEREÇO”, tal coluna foi criada para formar o endereço dos locais de acidente combinando os dados do logradouro e removendo o espaçamento em branco nas células.



As colunas de descrição das dimensões foram excluídas, bem como as colunas Nome da Origem, tipo_logradouro_anterior, nome_logradouro_anterior, nº_imovel_proximo e tipo_logradouro, pois estes não serão usados na análise.

Em seguida, as linhas foram classificadas de forma ascendente pelo Nº_boletim.

As colunas foram renomeadas para seguir o padrão da tabela F_OCORRENCIAS, tais como acréscimo de COD_, bem como colocar as colunas em maiúsculo e padronizar as colunas número boletim e data boletim.

Por fim, o endereço que começava com PCA SETE DE SETEMBRO foi substituído por AVE AFONSO PENA, visto que não existe um logradouro com o endereço do primeiro e, portanto, foi necessário fazer a correção para que o mapa lesse corretamente.

ELT D_MUNICIPIO, D_LOGRADOURO, D_BAIRRO E D_TIPO_BAIRRO

Essas dimensões foram criadas a partir da duplicação da tabela fato F_LOGRADOURO, com isso as etapas iniciais são as mesmas da tabela de origem.

Tal opção foi feita para que os dados não se perdessem na hora de excluir as colunas de descrição da tabela fato, visto que o Power BI tem a opção de “Duplicar” e “Referenciar”, sendo que em “Duplicar” ele faz a duplicação da tabela selecionada com todas as etapas que ela sofreu e na opção de “Referenciar” ela apenas utiliza a tabela depois de todo o tratamento sem as etapas de tratamento que ela sofreu.

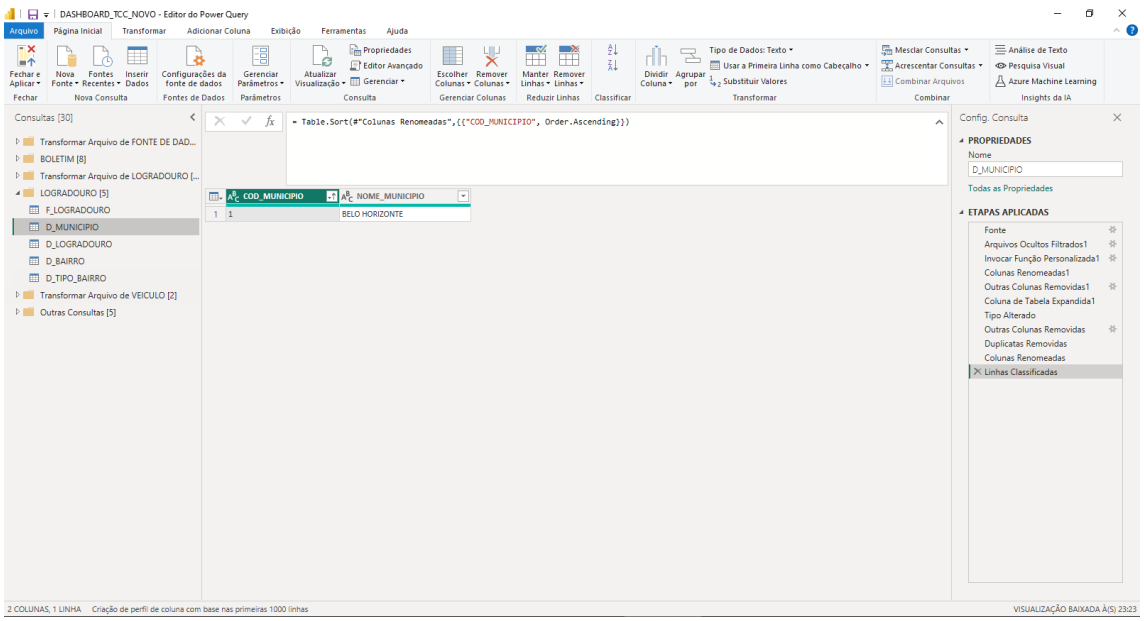
Portanto, até a etapa “Tipo Alterado” as dimensões tem o mesmo ETL que a tabela fato, ou seja, temos todo o trabalho da leitura dos arquivos da pasta e a renomeação para adequação das informações. Em seguida, são excluídas todas as colunas que não fazem parte da dimensão e são mantidos a coluna de ID e a sua respectiva descrição.

Por conseguinte, temos a remoção das duplicatas dos registros o que mantem apenas o ID e a descrição distinto da base inteira e possibilidade fazer a conexão dimensional de um para muitos.

Em seguida, as colunas são renomeadas para ficarem maiúsculo e adequarem ao ID da tabela fato.

Na tabela D_LOGRADOURO tem um passo a mais que as outras tabelas, pois as colunas tipo_logradouro e nome_logradouro são combinados pra formarem o logradouro.

Por fim, foi feito a classificação das linhas em ordem crescente para facilitar a identificação dos registros e de seus Ids na hora da leitura e manuseio dos dados.



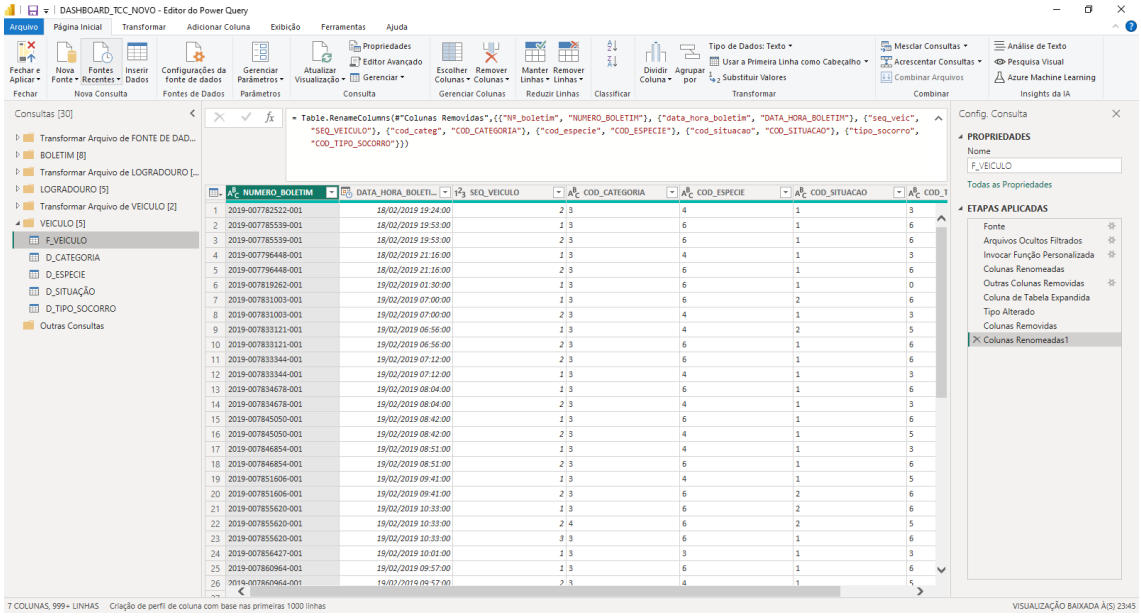
ETL F_VEICULO:

Assim como F_OCORRENCIAS, a tabela fato F_VEICULO passou pelas mesmas transformações para a leitura da base de dados na pasta de logradouro.

Em seguida, o tipo de todas as colunas foi alterado, visto que todos estavam como genérico que é o “ABC123” que o Power BI atribui automaticamente e passaram a ser do tipo Texto para todos com exceção para data_hora_boletim que passou a ser datetime e seq_veic que passou a ser do tipo int.

As colunas de descrição das dimensões foram excluídas, bem como a coluna Nome da Origem, pois esta não será usada na análise.

Por fim, as colunas foram renomeadas para seguir o padrão da tabela F_OCORRENCIAS, tais como acréscimo de COD_, bem como colocar as colunas em maiúsculo e padronizar as colunas número boletim e data boletim.



ELT D_CATEGORIA, D_ESPECIE, D_SITUACAO E D_TIPO_SOCORRO

Essas dimensões foram criadas a partir da duplicação da tabela fato F_VEICULO, com isso as etapas iniciais são as mesmas da tabela de origem.

Tal opção foi feita para que os dados não se perdessem na hora de excluir as colunas de descrição da tabela fato, visto que o Power BI tem a opção de “Duplicar” e “Referenciar”, sendo que em “Duplicar” ele faz a duplicação da tabela selecionada com todas as etapas que ela sofreu e na opção de “Referenciar” ela apenas utiliza a tabela depois de todo o tratamento sem as etapas de tratamento que ela sofreu.

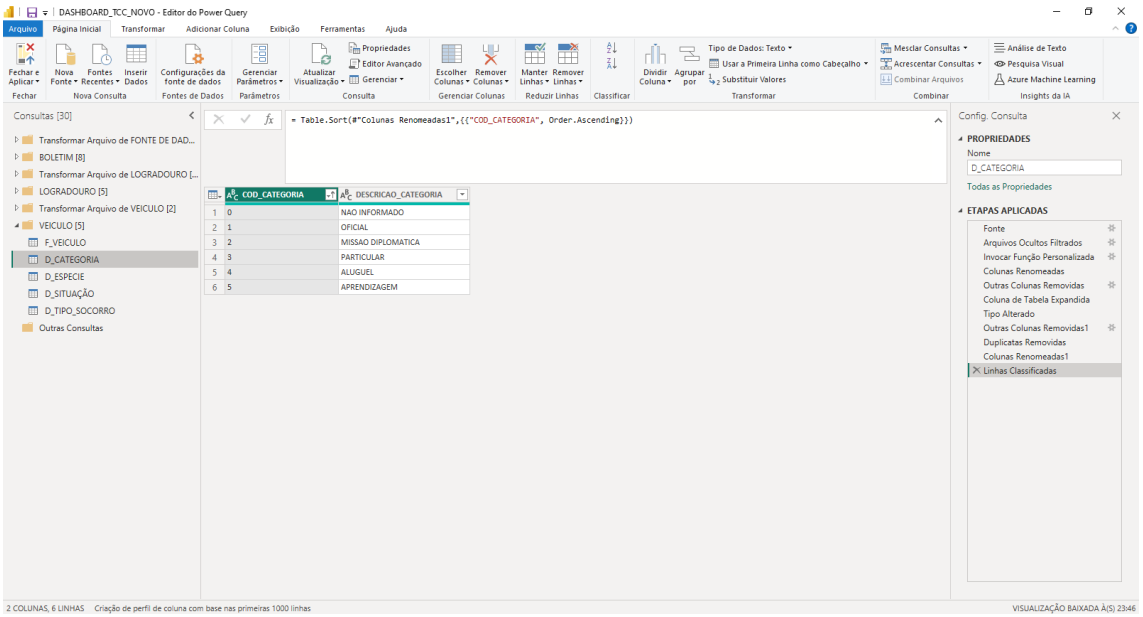
Portanto, até a etapa “Tipo Alterado” as dimensões tem o mesmo ETL que a tabela fato, ou seja, temos todo o trabalho da leitura dos arquivos da pasta e a renomeação para adequação das informações. Em seguida, são excluídas todas as colunas que não fazem parte da dimensão e são mantidos a coluna de ID e a sua respectiva descrição.

Por conseguinte, temos a remoção das duplicatas dos registros o que mantem apenas o ID e a descrição distinto da base inteira e possibilidade fazer a conexão dimensional de um para muitos.

Em seguida, as colunas são renomeadas para ficarem maiúsculo e adequarem ao ID da tabela fato.

Na tabela D_SITUAÇÃO tem um passo a mais que as outras tabelas, pois os valores do campo de DESCRICAO_SITUACAO são substituídos no caso de ESTACIONADO e vazio para apenas “ESTACIONADO”.

Por fim, foi feito a classificação das linhas em ordem crescente para facilitar a identificação dos registros e de seus Ids na hora da leitura e manuseio dos dados.



Códigos Fonte (Link para repositório externo)

Repositório da Prefeitura de Belo Horizonte com os boletins de ocorrência:

https://dados.pbh.gov.br/dataset/relacao-de-ocorrencias-de-acidentes-de-transito-com-vitima

Repositório da Prefeitura de Belo Horizonte com os logradouros dos boletins de ocorrência:

https://dados.pbh.gov.br/dataset/relacao-dos-logradouros-dos-locais-de-acidentes-de-transito-com-vitima

Repositório da Prefeitura de Belo Horizonte com os veículos dos boletins de ocorrência:

https://dados.pbh.gov.br/dataset/relacao-dos-veiculos-envolvidos-nos-acidentes-de-transito-com-vitima

Dashboard com o tratamento de dados no Google Drive:

https://drive.google.com/drive/folders/1AFe6HQDRMgajhu6v_ul3YGLTu7FLlkIi?usp=sharing

Fonte de dados no Google Drive:

https://drive.google.com/drive/folders/19bc7xSXAjYxTzn60nErgN8JBDjBo3O8X?usp=sharing

Módulo B - Painel de Controle (Dashboard)

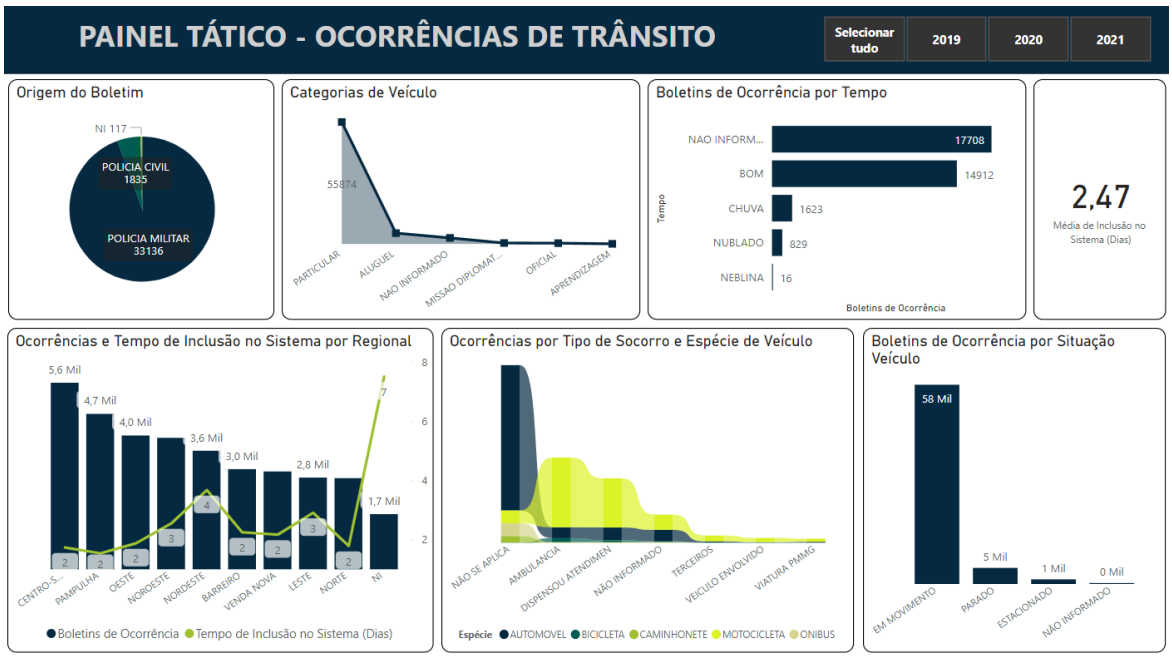
Visualização de Dados

Painel Estratégico – Público alvo: Âmbito público: Stakeholders sendo eles o Prefeito e Vice-Prefeito de Belo Horizonte, o presidente da BHTrans, Secretários da Prefeitura, Secretários da BHTrans.

Âmbito privado: CEO Diretores e Gerentes da companhia de seguros.

Painel Tático – Público alvo: Âmbito público: Além dos anteriores, coordenadores e supervisores da BHTrans.

Âmbito privado: Além dos anteriores, os coordenadores e superiores da companhia de seguros.



Painel Operacional – Público alvo: Âmbito público: Além dos anteriores, analistas de trânsito e estagiários da BHTrans.

Âmbito privado: Além dos anteriores, analistas, trainee e estagiários da companhia de seguros.

Teste de Homologação

O teste de homologação foi desenvolvido no Python utilizando o Jupyter Notebook e tem como intuito acessar os dados dos sistemas fontes diretamente utilizando a biblioteca Pandas para manipulação de dataframes. Sendo assim, os dados são acessados diretamente do sistema fonte e não sofrem qualquer perda ou manipulação. A seguir serão acrescentados os prints da tela do Jupyter Notebook comparando os dados manipulados via Python diretamente do sistema fonte em comparação as visualizações do Power BI. Cabe ressaltar que o código fonte do teste de homologação estará na pasta de homologação para acesso e teste.

Início da montagem do ambiente de homologação no Jupyter Notebook:

TESTE DE HOMOLOGAÇÃO

In [1]:
```
#Importação da biblioteca de manipulação de dataframs o pandas
import pandas as pd
```

In [2]:
```
#Carrega os dados da Fonte de Dados que é utilizada o Power BI
do1 = pd.read_csv('../../FONTE DE DADOS/BOLETIM/si-bol-2019.csv', encoding="ANSI", delimiter=';')
do2 = pd.read_csv('../../FONTE DE DADOS/BOLETIM/si-bol-2020.csv', encoding="ANSI", delimiter=';')
do3 = pd.read_csv('../../FONTE DE DADOS/BOLETIM/si-bol-2021.csv', encoding="ANSI", delimiter=';')
dl1 = pd.read_csv('../../FONTE DE DADOS/LOGRADOURO/si-log-2019.csv', encoding="ANSI", delimiter=';')
dl2 = pd.read_csv('../../FONTE DE DADOS/LOGRADOURO/si-log-2020.csv', encoding="ANSI", delimiter=';')
dl3 = pd.read_csv('../../FONTE DE DADOS/LOGRADOURO/si-log-2021.csv', encoding="ANSI", delimiter=';')
dv1 = pd.read_csv('../../FONTE DE DADOS/VEICULO/si-veic-2019.csv', encoding="ANSI", delimiter=';')
dv2 = pd.read_csv('../../FONTE DE DADOS/VEICULO/si-veic-2020.csv', encoding="ANSI", delimiter=';')
dv3 = pd.read_csv('../../FONTE DE DADOS/VEICULO/si-veic-2021.csv', encoding="ANSI", delimiter=';')
```

In [3]:
```
#Unifica as bases de dados de 2019, 2020 e 2021 dividido pelo assunto.
dados_ocorrencia = pd.concat([do1, do2, do3], axis=0)
dados_logradouro = pd.concat([dl1, dl2, dl3], axis=0)
dados_veiculo = pd.concat([dv1, dv2, dv3], axis=0)
```

In [4]:
```
#Verificação do tamanho dos datasets unidos
print(dados_ocorrencia.shape)
print(dados_logradouro.shape)
print(dados_veiculo.shape)
```

```
(35088, 23)
(46194, 16)
(64469, 11)
```

In [5]:
```
#retirando o espaço em branco nas colunas
colunas = dados_ocorrencia.columns.tolist()
colunas_sem_espaco = [coluna.strip() for coluna in colunas]
dados_ocorrencia.columns = colunas_sem_espaco
colunas = dados_logradouro.columns.tolist()
colunas_sem_espaco = [coluna.strip() for coluna in colunas]
dados_logradouro.columns = colunas_sem_espaco
colunas = dados_veiculo.columns.tolist()
colunas_sem_espaco = [coluna.strip() for coluna in colunas]
dados_veiculo.columns = colunas_sem_espaco
```

In [6]:
```
#Função para tirar o espaçamento em branco dos dados do tipo string
dados_ocorrencia = dados_ocorrencia.apply(lambda x: x.str.strip() if x.dtype == "object" else x)
dados_logradouro = dados_logradouro.apply(lambda x: x.str.strip() if x.dtype == "object" else x)
dados_veiculo = dados_veiculo.apply(lambda x: x.str.strip() if x.dtype == "object" else x)
```

Comparação do campo DATA HORA_BOLETIM com a visualização do Power BI Boletins por Ano (Painel Estratégico):

- Jupyter Notebook

In [7]:
```
#Verificando o tipo de campo que está a coluna DATA HORA_BOLETIM
dados_ocorrencia['DATA HORA_BOLETIM'].head()
```

Out[7]:
```
0    01/01/2019 01:40
1    01/01/2019 02:46
2    01/01/2019 02:19
3    01/01/2019 03:18
4    01/01/2019 04:26
Name: DATA HORA_BOLETIM, dtype: object
```

In [8]:
```
#Convertendo o campo de DATA HORA_BOLETIM do tipo object para datetime
dados_ocorrencia['DATA HORA_BOLETIM'] = pd.to_datetime(dados_ocorrencia['DATA HORA_BOLETIM'])
```

In [9]:
```
#Contagem de quantas linhas tem no dataset de ocorrências no ano de 2019
print(len(dados_ocorrencia[dados_ocorrencia['DATA HORA_BOLETIM'].dt.year == 2019 ]))
```

```
13342
```

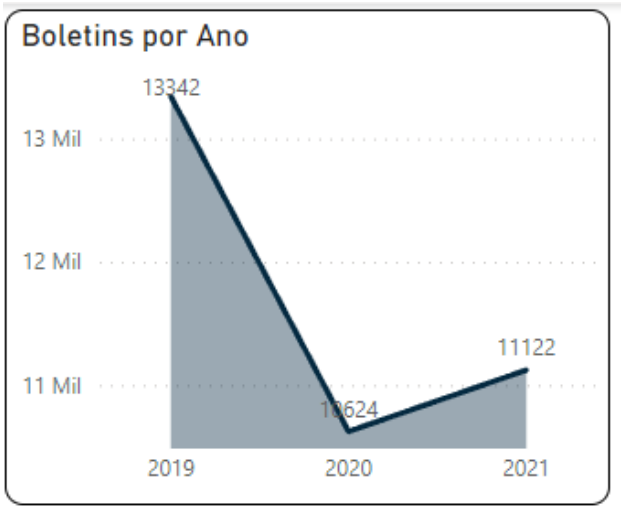
In [10]:
```
#Contagem de quantas linhas tem no dataset de ocorrências no ano de 2020
print(len(dados_ocorrencia[dados_ocorrencia['DATA HORA_BOLETIM'].dt.year == 2020 ]))
```

```
10624
```

In [11]:
```
#Contagem de quantas linhas tem no dataset de ocorrências no ano de 2021
print(len(dados_ocorrencia[dados_ocorrencia['DATA HORA_BOLETIM'].dt.year == 2021 ]))
```

```
11122
```

- Power BI



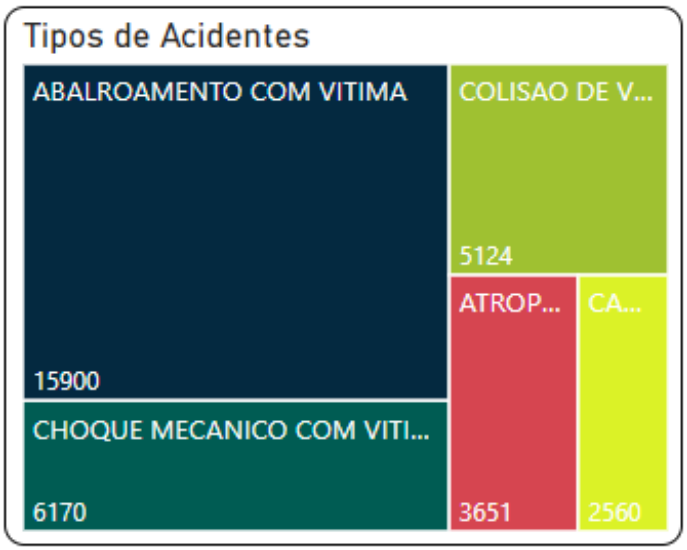
Comparação do campo DESC_TIPO_ACIDENTE com a visualização do Power BI Tipos de Acidentes (Painel Estratégico):

- Jupyter Notebook

```
In [12]: #Contagem de tipos de acidentes por ocorrências
         print(dados_ocorrencia['DESC_TIPO_ACIDENTE'].value_counts())
```

```
ABALROAMENTO COM VITIMA                              15900
CHOQUE MECANICO COM VITIMA                            6170
COLISAO DE VEICULOS COM VITIMA                        5124
ATROPELAMENTO DE PESSOA SEM VITIMA FATAL              3651
CAPOTAMENTO/TOMBAMENTO COM VITIMA                     2560
QUEDA DE PESSOA DE VEICULO                            1015
OUTROS COM VITIMA                                      376
ATROPELAMENTO DE ANIMAL COM VITIMA                     134
ATROPELAMENTO DE PESSOA COM VITIMA FATAL               112
QUEDA DE VEICULO COM VITIMA                             34
QUEDA E/OU VAZAMENTO DE CARGA DE VEICULO C/ VITIMA      11
CAPOTAMENTO/TOMBAMENTO SEM VITIMA                        1
Name: DESC_TIPO_ACIDENTE, dtype: int64
```

- Power BI



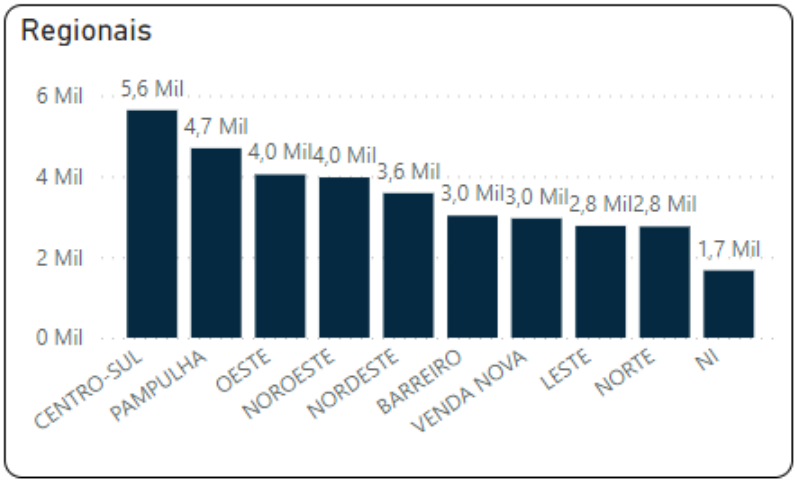
Comparação do campo DESC_REGIONAL com a visualização do Power BI Regionais (Painel Estratégico):

- Jupyter Notebook

```
In [13]: #Contagem de regionais por ocorrências
         print(dados_ocorrencia['DESC_REGIONAL'].value_counts())
```

```
CENTRO-SUL    5634
PAMPULHA      4693
OESTE         4041
NOROESTE      3971
NORDESTE      3581
BARREIRO      3024
VENDA NOVA    2955
LESTE         2770
NORTE         2753
              1666
Name: DESC_REGIONAL, dtype: int64
```

- Power BI



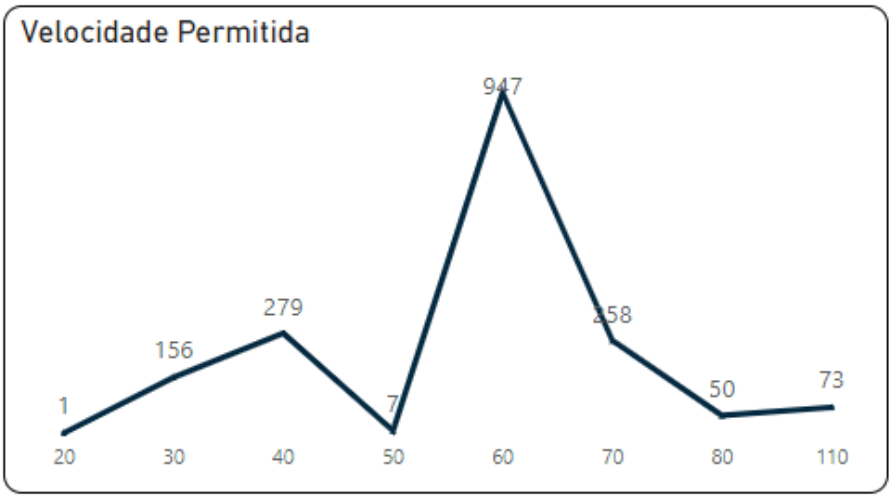
Comparação do campo VELOCIDADE_PERMITIDA com a visualização do Power BI Velocidade Permitida (Painel Estratégico):

- Jupyter Notebook

```
In [14]: #Contagem de velocidade permitida por ocorrências
         print(dados_ocorrencia['VELOCIDADE_PERMITIDA'].value_counts())
```

```
0      33317
60       947
40       279
70       258
30       156
110       73
80        50
50         7
20         1
Name: VELOCIDADE_PERMITIDA, dtype: int64
```

- Power BI



Comparação do campo INDICADOR_FATALIDADE com a visualização do Power BI Acidentes Fatais e Acidentes Sem Fatalidade (Painel Estratégico):

- Jupyter Notebook

```
In [15]: #Contagem de indicador de fatalidade por ocorrências
         print(dados_ocorrencia['INDICADOR_FATALIDADE'].value_counts())

         NÃO    34792
         SIM      296
         Name: INDICADOR_FATALIDADE, dtype: int64
```

-Power BI

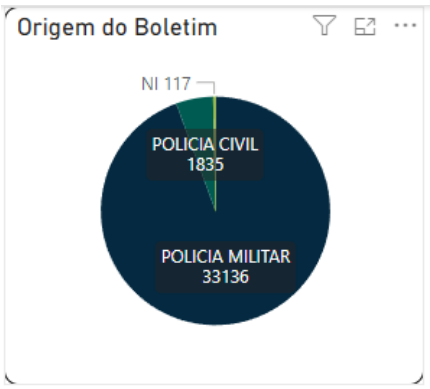
Comparação do campo ORIGEM_BOLETIM com a visualização do Power BI Origem do Boletim (Painel Tático):

- Jupyter Notebook

```
In [16]: #Contagem de origem por ocorrências
         print(dados_ocorrencia['ORIGEM_BOLETIM'].value_counts())

         POLICIA MILITAR    33136
         POLICIA CIVIL       1835
         NI                   117
         Name: ORIGEM_BOLETIM, dtype: int64
```

- Power BI



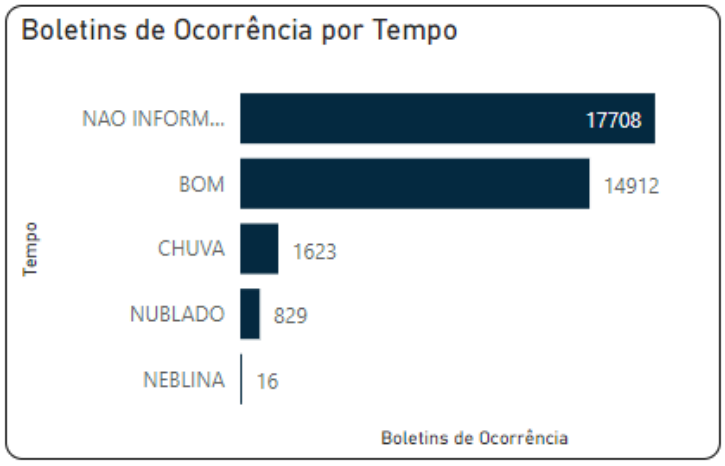
Comparação do campo DESC_TEMPO com a visualização do Power BI Boletins de Ocorrência por Tempo (Painel Tático):

- Jupyter Notebook

```
In [17]: #Contagem de tempo por ocorrências
         print(dados_ocorrencia['DESC_TEMPO'].value_counts())

         NAO INFORMADO    17708
         BOM              14912
         CHUVA             1623
         NUBLADO            829
         NEBLINA             16
         Name: DESC_TEMPO, dtype: int64
```

- Power BI



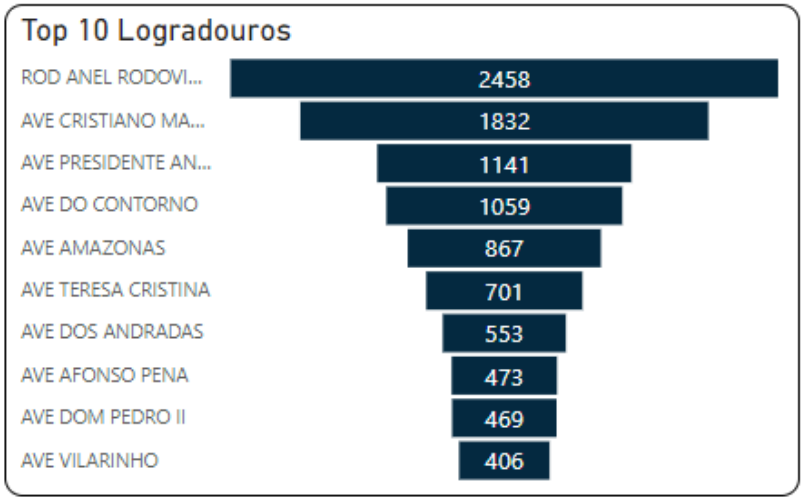
Comparação do campo nome_logradouro com a visualização do Power BI Top 10 Logradouros (Painel Estratégico):

- Jupyter Notebook

```
In [18]: #Contagem dos top 10 logradouros sem o tipo de logradouro (ex. ROD, AVE, RUA)
         print(dados_logradouro['nome_logradouro'].value_counts().nlargest(10))
```

```
ANEL RODOVIARIO CELSO MELLO AZEVEDO    2458
CRISTIANO MACHADO                      1832
PRESIDENTE ANTONIO CARLOS              1141
DO CONTORNO                            1059
AMAZONAS                                867
TERESA CRISTINA                         701
DOS ANDRADAS                            553
AFONSO PENA                             473
DOM PEDRO II                            469
VILARINHO                               406
Name: nome_logradouro, dtype: int64
```

- Power BI



Comparação dos campos DESC_REGIONAL, Nº_bairro e Nº_logradouro com a visualização do Power BI Regionais, Bairros e Logradouros (Painel Estratégico):

- Jupyter Notebook

```
In [19]: #Contagem distinta das quantidade de regionais considerando as não informadas
         print(dados_ocorrencia['DESC_REGIONAL'].nunique())

         10

In [20]: #Contagem distinta das quantidade de regionais excluindo as não informadas
         dados_ocorrencia_regional = dados_ocorrencia[dados_ocorrencia['COD_REGIONAL'] != 0]
         print(dados_ocorrencia_regional['DESC_REGIONAL'].nunique())

         9

In [21]: #Contagem distinta das quantidade de bairros
         print(dados_logradouro['Nº_bairro'].nunique())

         454

In [22]: #Contagem distinta das quantidade de logradouros
         print(dados_logradouro['Nº_logradouro'].nunique())

         5303
```

- Power BI

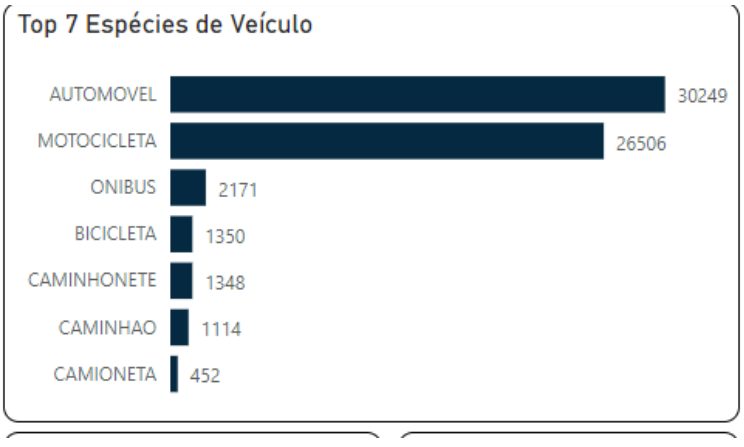
Comparação dos campos descricao_especie com a visualização do Power BI Top 7 Espécies de Veículo (Painel Estratégico):

- Jupyter Notebook

```
In [23]: #Contagem do top 7 espécies de veiculo
         print(dados_veiculo['descricao_especie'].value_counts().nlargest(7))

         AUTOMOVEL      30249
         MOTOCICLETA    26506
         ONIBUS          2171
         BICICLETA       1350
         CAMINHONETE     1348
         CAMINHAO        1114
         CAMIONETA        452
         Name: descricao_especie, dtype: int64
```

- Power BI



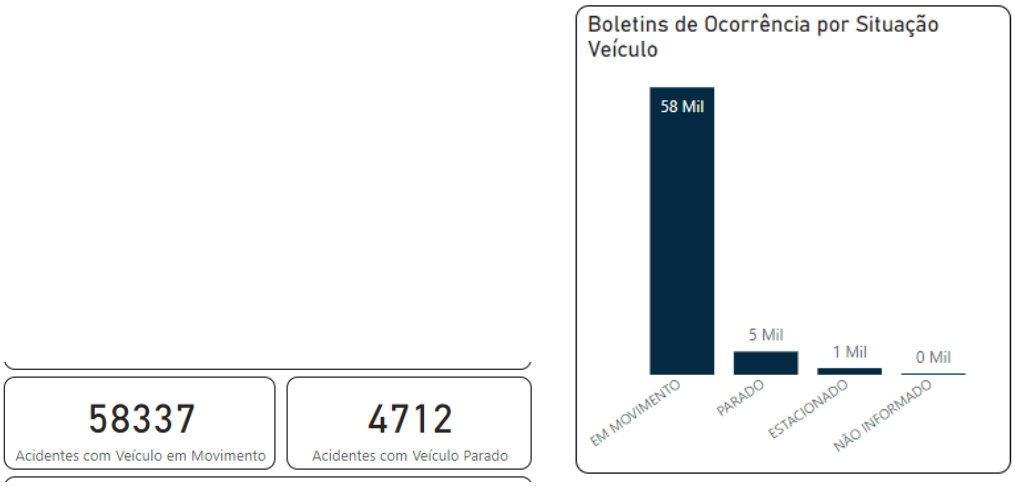
Comparação dos campos desc_situacao com a visualização do Power BI Acidentes Com Veículo em Movimento, Acidentes com Veículo Parado e Boletins de Ocorrência por Situação Veículo (Painel Estratégico e Tático):

- Jupyter Notebook

```
In [24]: #Contagem de situação por veiculos
         print(dados_veiculo['desc_situacao'].value_counts())

EM MOVIMENTO     58337
PARADO            4712
ESTACIONADO        863
                   497
NÃO INFORMADO       60
Name: desc_situacao, dtype: int64
```

- Power BI



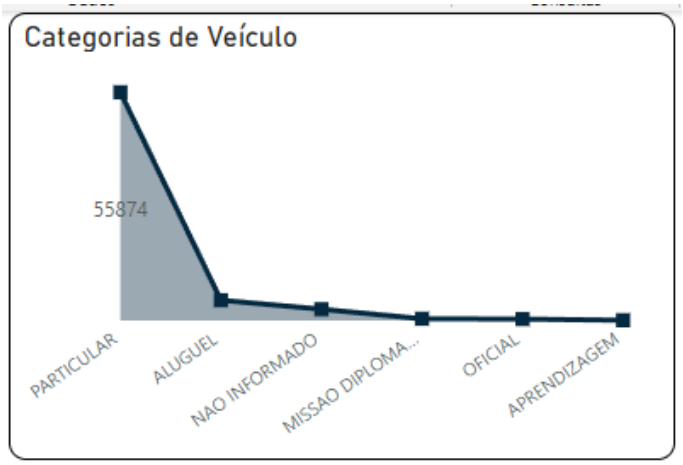
Comparação dos campos descricao_categoria com a visualização do Power BI Categorias do Veículo (Painel Tático):

- Jupyter Notebook

```
In [25]: #Contagem de categoria por veiculos
         print(dados_veiculo['descricao_categoria'].value_counts())

PARTICULAR            55874
ALUGUEL                4973
NAO INFORMADO          2767
MISSAO DIPLOMATICA      432
OFICIAL                 357
APRENDIZAGEM             66
Name: descricao_categoria, dtype: int64
```

- Power BI



Comparação dos campos desc_tipo_socorro e descrição_especie com a visualização do Power BI Ocorrências por Tipo de Socorro e Espécie de Veículo (Painel Tático):

- Jupyter Notebook

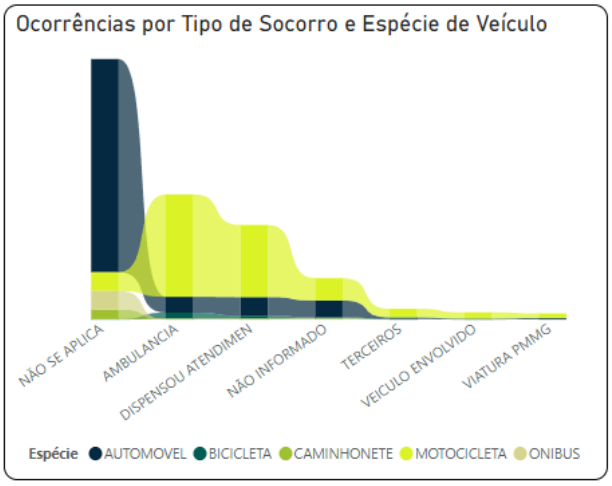
In [26]:
```
#Contagem de tipo de socorro por veiculos
print(dados_veiculo['desc_tipo_socorro'].value_counts())
```

```
NÃO SE APLICA          31294
AMBULANCIA             14432
DISPENSOU ATENDIMEN    10869
NÃO INFORMADO           5103
TERCEIROS               1232
VEICULO ENVOLVIDO        831
VIATURA PMMG             708
Name: desc_tipo_socorro, dtype: int64
```

In [27]:
```
#Aqui temos a combinação das colunas de tipo de socorro com a espécie de veiculos ordenado pela frequência de vezes que
#aparece na visualização.
dados_filtrados = dados_veiculo.groupby(['desc_tipo_socorro','descricao_especie']).size().reset_index(name = 'count')
print(dados_filtrados.sort_values(by='count', ascending=False).head(10))
```

```
      desc_tipo_socorro descricao_especie  count
60        NÃO SE APLICA         AUTOMOVEL  24077
12           AMBULANCIA       MOTOCICLETA  11561
32  DISPENSOU ATENDIMEN       MOTOCICLETA   8137
51        NÃO INFORMADO       MOTOCICLETA   2512
74        NÃO SE APLICA       MOTOCICLETA   2122
21  DISPENSOU ATENDIMEN         AUTOMOVEL   2099
77        NÃO SE APLICA            ONIBUS   2091
38        NÃO INFORMADO         AUTOMOVEL   1922
0            AMBULANCIA         AUTOMOVEL   1802
65        NÃO SE APLICA       CAMINHONETE   1099
```

- Power BI



Códigos Fonte

Repositório da Prefeitura de Belo Horizonte com os boletins de ocorrência:

https://dados.pbh.gov.br/dataset/relacao-de-ocorrencias-de-acidentes-de-transito-com-vitima

Repositório da Prefeitura de Belo Horizonte com os logradouros dos boletins de ocorrência:

https://dados.pbh.gov.br/dataset/relacao-dos-logradouros-dos-locais-de-acidentes-de-transito-com-vitima

Repositório da Prefeitura de Belo Horizonte com os veículos dos boletins de ocorrência:

https://dados.pbh.gov.br/dataset/relacao-dos-veiculos-envolvidos-nos-acidentes-de-transito-com-vitima

Dashboard com o tratamento de dados no Google Drive:

https://drive.google.com/drive/folders/1AFe6HQDRMgajhu6v_ul3YGLTu7FLlkIi?usp=sharing

Fonte de dados no Google Drive:

https://drive.google.com/drive/folders/19bc7xSXAjYxTzn60nErqN8JBDjBo3O8X?usp=sharing

Dashboard e vídeo de apresentação no Google Drive:
https://drive.google.com/drive/folders/1ARPf4n4LN-Mv9xiksaSKa9MFEoHyTFq5?usp=sharing

Vídeo de apresentação publicado no Youtube:
https://www.youtube.com/watch?v=lXAUhefgUtY

Teste e Homologação:
https://drive.google.com/drive/folders/1E9CG2PLt-UpIgfW1lNyRMRNXPFDJYDRR?usp=sharing