

# Análise sobre Dados de Acidentes e Infrações de Trânsito ocorridos na Cidade do Recife

Artur da Nova Franco<sup>1</sup>, Lucas Glasner Régis<sup>2</sup>

*Centro de Informática - Universidade Federal de Pernambuco (CIn/UFPE)*

---

## Abstract

Este documento tem o objetivo de trazer informações relevantes sobre o trânsito da cidade do Recife, mais especificamente sobre acidentes e infrações de trânsito, para um melhor entendimento deste ambiente, e também propor melhorias nessa área. Para conseguir realizar este objetivo, foi feita uma análise sobre dados de acidentes e infrações de trânsito coletados a partir do portal de dados abertos da prefeitura do Recife. Esses dados foram pré-processados, explorados e após a análise como um todo, foi possível obter respostas para algumas perguntas que foram feitas previamente sobre os dados. Com essas respostas em mãos, pode-se partir para o planejamento de melhorias para o trânsito em diversos aspectos.

*Keywords:* Análise de Dados, Trânsito, Acidentes, Infrações, Cidade, Recife, Ciência dos Dados

---

## 1. Introdução

O trânsito, no geral, é uma situação pela qual grande parte das pessoas passa todos os dias, tornando-o um local de ocorrência de diversos tipos de acidentes e cometimento de infrações, e isso quando refletido no contexto geral de uma cidade, torna-se mais um ambiente de insegurança no meio de tantos outros. Quanto mais evitarmos cometer infrações ou acidentes no trânsito, melhoramos o mesmo, melhorando assim, o dia-dia em geral de uma cidade. Pensando nessa

---

<sup>1</sup>Graduando em Engenharia da Computação - anf@cin.ufpe.br

<sup>2</sup>Graduando em Sistemas de Informação - lgr3@cin.ufpe.br

melhora, propomos uma análise sobre dados de acidentes e infrações de trânsito ocorridos nos anos de 2015 e 2016 e registrados pela CTTU (Companhia de Trânsito e Transporte Urbano), sendo estes dados relativos à cidade do Recife. O objetivo dessa análise é entender melhor esse ambiente, podendo ao final do projeto, propor algum método que possa oferecer melhorias para o trânsito, inicialmente na cidade do Recife, mas podendo também ser estendido à outras cidades. Os dados serão coletados, pré-processados, analisados e visualizados, como será explicado no decorrer deste documento.

## 2. Metodologia

Nesta seção, serão explicados como foram realizados cada etapa de desenvolvimento do projeto. Tais etapas são importantes para uma boa análise de dados. São elas:

### 2.1. Coleta dos dados

Neste projeto, foram utilizados dados sobre acidentes de trânsito (com e sem vítimas) ocorridos nos anos de 2015 e 2016, e dados das infrações de trânsito registradas pela CTTU também nos anos de 2015 e 2016, sendo todos os conjuntos de dados relativos à cidade do Recife. Os dados foram coletados, em formato CSV (*Comma-separated values* ou valores separados por vírgula), a partir do portal de dados abertos da Prefeitura do Recife (<http://dados.recife.pe.gov.br>).

### 2.2. Pré-processamento dos dados

O pr-processamento dos dados coletados foi feito utilizando principalmente a biblioteca *pandas* disponível para *Python*, a qual fornece diversas ferramentas para tratar e manipular conjuntos de dados. Ao realizar a coleta dos dados, foram obtidos os conjuntos de dados brutos e após realizar o pr-processamento dos mesmos, foram obtidos conjuntos de dados tratados, ou seja, mais amigáveis para trabalhar. O pré-processamento consistiu basicamente em:

- Excluir dos *datasets* as colunas que não consideramos irrelevantes para o desenvolvimento do projeto;

- Tratar linhas com valores ausentes apenas deletando-as do *dataset*, uma vez que a quantidade de linhas com valores ausentes era bem menor quando comparada ao número total de linhas do *dataset*;
- Converter colunas relativas à datas para o tipo *datetime*, no padrão *yyyy-mm-dd*, com o objetivo de facilitar na manipulação dos dados. Por exemplo, 2016-02-15 (15 de fevereiro de 2016);
- Colocar um padrão em algumas colunas, também com o objetivo de facilitar a manipulação dos dados. Por exemplo, algumas colunas relativas à horas estavam em padrões diferentes para cada conjunto de dados, visto isso, todas as colunas relativas à horas foram colocadas no formato *hh:mm*, por exemplo, 8 : 57 (8 horas e 57 minutos).

### 2.3. Análise exploratória dos dados

Realizamos uma análise exploratória dos dados com o objetivo de entender melhor a respeito da natureza dos dados, detectar hipóteses, ou seja, descobrir coisas novas a partir dos dados, e também confirmar hipóteses, ou seja, responder algumas perguntas que fizemos sobre os dados. Buscamos responder perguntas como:

- É mais provável ocorrerem acidentes de trânsito em bairros de classe mais baixa?
- É mais provável ocorrerem acidentes de trânsito no horário da noite?
- É mais provável ocorrerem infrações de trânsito no horário da noite?
- Para a maioria dos acidentes foi registrada uma infração?
- O registro de infrações ajuda a diminuir a quantidade de acidentes?

Para responder as perguntas acima, buscamos utilizar estatísticas descritivas disponíveis na biblioteca *pandas* e recursos gráficos disponíveis no pacote *pyplot* da biblioteca *matplotlib*, ambas as bibliotecas disponíveis para *Python*.

Analisando os conjuntos de dados sobre acidentes, percebemos que para o ano de 2016 s haviam dados de 10 meses (Janeiro à Outubro) e para o ano de 2015 só haviam dados de 7 meses (Junho à Dezembro). Para fazer uma  
65 estimativa da quantidade total de acidentes nos anos de 2015 e 2016, os meses inexistentes nos *datasets* foram completados com a média da quantidade de acidentes em cada ano, calculada apenas para os meses existentes nos *datasets*. Diante disso tambm, para outras partes da análise, foram juntados em um s conjunto de dados, os dados sobre acidentes de 2015 e 2016, resultando em um  
70 *dataset* contendo dados sobre acidentes de trânsito ocorridos em Recife, durante o intervalo entre Junho de 2015 e Outubro de 2016. Isso foi feito também para os conjuntos de dados sobre registros de infrações de trânsito, juntando os *datasets* relativos à 2015 e 2016 em um só.

Na análise exploratória foram descobertas coisas como:

- 75 • De 2015 para 2016, houve um decaimento de cerca de 15% na quantidade de acidentes de trânsito ocorridos na cidade do Recife.
- O bairro do Recife onde ocorreram mais acidentes de trânsito entre Junho de 2015 e Outubro de 2016 Boa Viagem (13.74% dos acidentes).
- O logradouro do Recife onde ocorreram mais acidentes de trânsito entre  
80 Junho de 2015 e Outubro de 2016 a Avenida Marechal Mascarenhas de Moraes (4.36% dos acidentes).
- O período do dia no qual ocorreram mais acidentes de trânsito entre Junho de 2015 e Outubro de 2016 foi o período da tarde, ou seja, entre 13h e 18h.
- 85 • Os intervalos de horas do dia nos quais ocorreram mais acidentes de trânsito entre Junho de 2015 e Outubro de 2016 foram os intervalos entre 8h e 9h e entre 12h e 13h.
- De 2015 para 2016, houve um decaimento de cerca de 9.28% na quantidade de registros de infrações de trânsito ocorridos na cidade do Recife.

- De todas as infrações registradas no Recife em 2015 e 2016, quase metade (46.85%) foram registradas pelas lombadas eletrônicas.
- O período do dia no qual ocorreram mais registros de infrações de trânsito em 2015 e 2016 foi o período da tarde, ou seja, entre 13h e 18h.
- Os intervalos de horas do dia nos quais ocorreram mais registros de infrações de trânsito em 2015 e 2016 foram os intervalos entre 10h e 11h e entre 15h e 16h.

#### 2.4. Técnica de aprendizagem

Apenas com análise exploratória não foi possível compreender a relação entre a proporção de acidentes e a proporção de infrações ocorridos na cidade do Recife em 2016. Então aplicamos uma técnica de aprendizagem chamada de regressão linear para conseguir entender melhor e visualizar essa relação entre essas duas variáveis. Foi aplicada ainda a matriz de correlação e obtivemos um coeficiente de correlação de aproximadamente 0.52, o que indica uma correlação moderada entre as duas variáveis.

### 3. Resultados e visualização dos dados

Nesta seção, veremos os resultados gráficos da nossa análise.

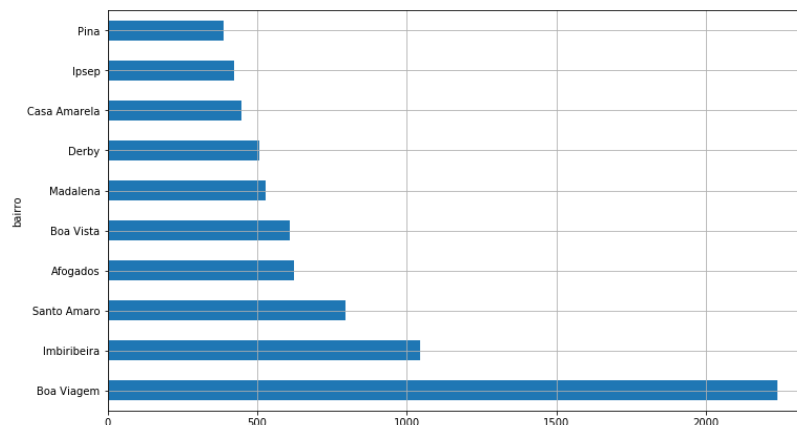


Figure 1: Os 10 bairros do Recife onde ocorreram mais acidentes de trânsito (Jun/15 à Out/16)

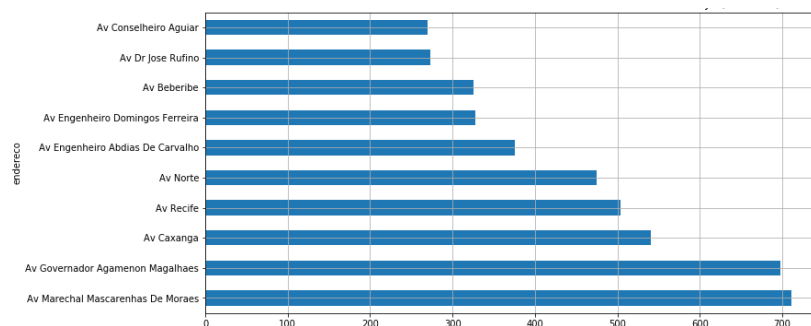


Figure 2: Logradouros do Recife onde ocorreram mais acidentes de trânsito (Jun/15 à Out/16)

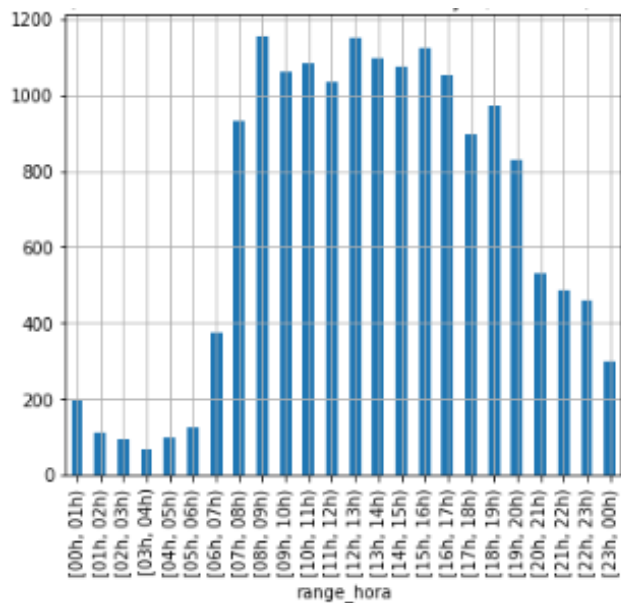


Figure 3: Quantidade de acidentes de trânsito por hora do dia (Jun/15 à Out/16)

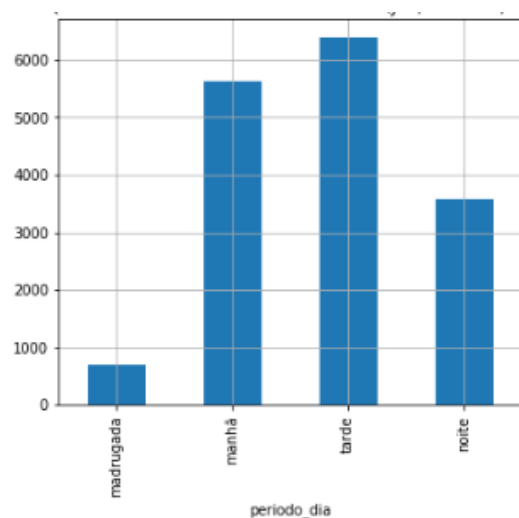


Figure 4: Quantidade de acidentes de trânsito por período do dia(Jun/15 à Out/16)

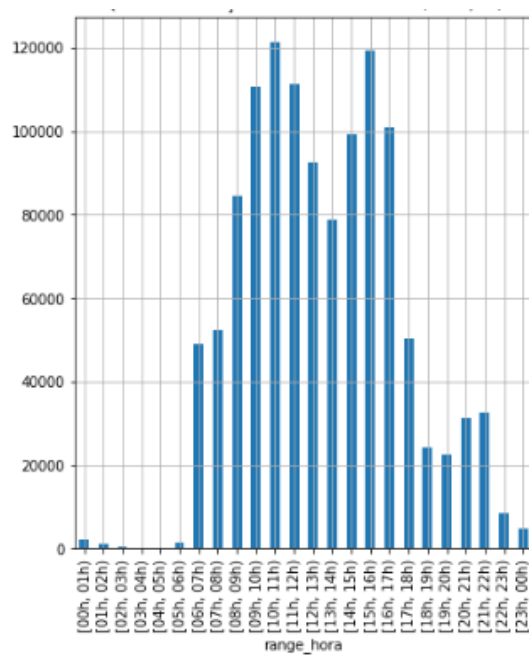


Figure 5: Quantidade de infrações de trânsito por hora do dia (2015/16)

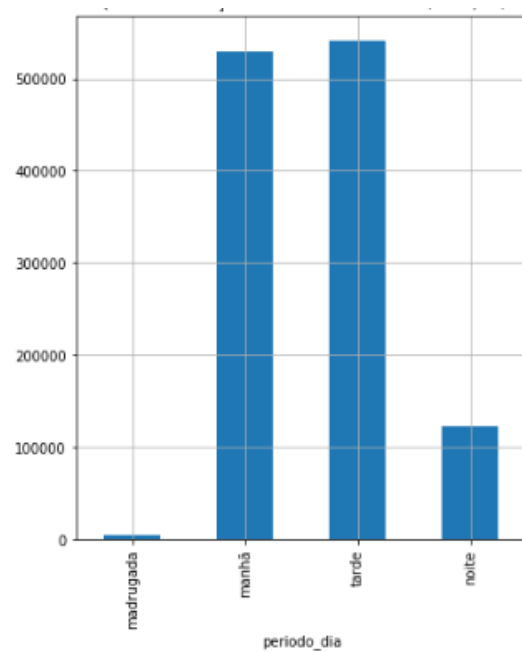


Figure 6: Quantidade de infrações de trânsito por período do dia (2015/16)

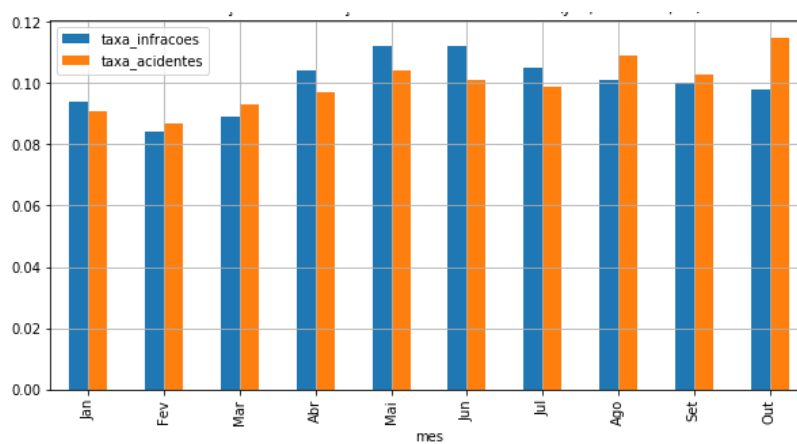


Figure 7: Proporção de infrações e acidentes de trânsito por mês (Jan/16 à Out/16)



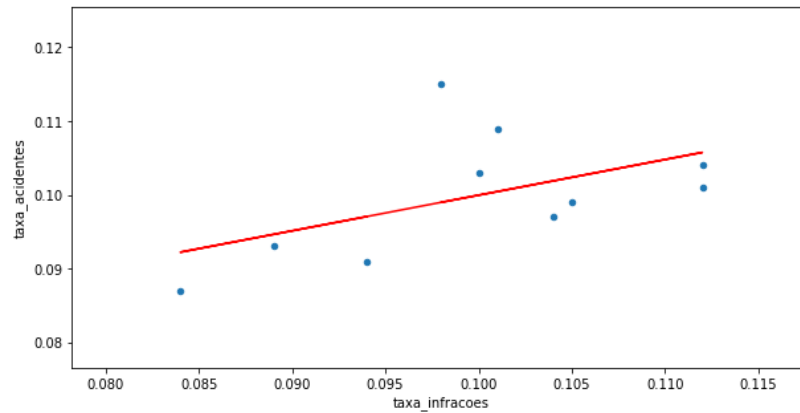


Figure 8: Regressão linear entre a taxa de acidentes e a taxa de infrações

	taxa_infracoes	taxa_acidentes
taxa_infracoes	1.000000	0.521628
taxa_acidentes	0.521628	1.000000

Figure 9: Matriz de correlação entre as duas variáveis especificadas

#### 4. Conclusões

É mais provável ocorrerem acidentes de trânsito em bairros de classe mais baixa? Não necessariamente. Na Figura 1 pudemos descobrir que o bairro onde ocorrem mais acidentes de trânsito é Boa Viagem (com mais do que o dobro da quantidade de acidentes do segundo colocado), que é um bairro de classe média-alta. Isso se deve também à Boa Viagem ser disparado o bairro do Recife com maior número de habitantes, além de possuir três avenidas grandes e movimentadas.

É mais provável ocorrerem acidentes de trânsito no horário da noite? Não. Podemos observar na Figura 4 que a quantidade maior de acidentes ocorreram de tarde, seguido pelo período da manhã, com uma certa superioridade com relação a quantidade de acidentes que ocorreram de noite.

É mais provável ocorrerem infrações de trânsito no horário da noite? Não.  
120 Podemos observar na Figura 6 que a quantidade maior de infrações ocorreram  
de tarde, seguido pelo período da manhã, com uma grande superioridade com  
relação a quantidade de infrações que ocorreram de noite. Isto se deve pois no  
período da noite em muitos lugares o trânsito é lento, impedindo que alguém  
125 ultrapasse o limite de velocidade de uma via, por exemplo. Além de que a partir  
de um certo horário da noite, as lombadas eletrônicas são desligadas e já se pode  
estacionar em alguns lugares que não poderia se fosse durante o dia.

O registro de infrações ajuda a diminuir a quantidade de acidentes? Não nec-  
essariamente. Analisando a Figura 8, observamos que existe uma certa tendência  
de que quando a proporção de infrações cresce, a proporção de acidentes também  
130 cresce. Logo, não é verdade que quanto mais infrações são registradas, menos  
acidentes acontecem. Analisando a Figura 9, vemos que o coeficiente de cor-  
relação entre as duas variáveis especificadas é de aproximadamente 0.52, que  
indica uma correlação moderada. Para que a resposta desta pergunta fosse sim,  
esse coeficiente deveria ser negativo, que indica que quando o valor de uma  
135 variável tende a crescer, o da outra tende a diminuir.

Para a maioria dos acidentes foi registrada uma infração? Nós não obtivemos  
uma resposta clara para esta pergunta, uma vez que pela natureza do problema  
real e dos próprios conjuntos de dados, não encontramos uma maneira para tal.

Com essas perguntas respondidas, podemos partir para uma próxima etapa,  
140 ou um próximo trabalho, não necessariamente ligado à tecnologia da informação,  
para planejar melhorias para o trânsito do Recife com base nessas respostas.  
Um exemplo raso: sabendo que Boa Viagem ganha com folga em quantidade  
de acidentes, planejaria-se melhorias com o foco lá, ou sabendo que a maior  
quantidade de infrações ocorre no intervalo entre 10h e 11h, aumentaria-se a  
145 fiscalização nas ruas durante esse horário.