

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

Artur da Nova Franco, Lucas Glasner Régis

Centro de Informática, Universidade Federal de Pernambuco - CIn/UFPE

Abstract

This template helps you to create a properly formatted L^AT_EX manuscript.

Keywords: `elsarticle.cls`, L^AT_EX, Elsevier, template

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 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:

20 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
25 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
30 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 nós consideramos irrelevantes para o
35 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 às 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);
40
- Colocar um padrão em algumas colunas, também com o objetivo de facilitar a manipulação dos dados. Por exemplo, algumas colunas relativas às horas estavam em padrões diferentes para cada conjunto de dados, visto
45 isso, todas as colunas relativas às 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
50 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?
- 55 • 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
60 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.
- 90 • 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 in-
95 frações de trânsito em 2015 e 2016 foram os intervalos entre 10h e 11h e entre 15h e 16h.

2.4. Técnica de aprendizagem

2.5. Visualização dos dados

3. Resultados

100 **4. Conclusões**