

# Envirocar visualization

## exploring an environmental and traffic data set

Rodrigo Claro Zembruski  
Programa de Pos Graduacao em Ciencia da Computacao  
Universidade Federal do Rio Grande do Sul - UFRGS  
Porto Alegre, Brasil  
Web page: [inf.ufrgs.br/~rczembruski](http://inf.ufrgs.br/~rczembruski)

**Abstract**—Envirocar eh um projeto que coleta e armazena dados ambientais e dados de trafego. Sao dados de sensores implantados dentro dos automoveis que trazem informacoes precisas sobre metricas dos automoveis e dados relevantes ao meio-ambiente, como emissao de gas-carbonico e consumo do automovel.

Esse artigo tem por objetivo apresentar uma serie de visualizacoes para analise exploratoria desse banco de dados. Com isso, teremos uma intuicao sobre se esse banco de dados pode servir para alguma coisa, ou nao.

**Keywords**—envirocar; smartcars; kibana; geolocation; data visualization;

### I. INTRODUCTION

Envirocar eh um projeto da Universidade de Munster. O objetivo do projeto eh oferecer uma plataforma simples para coleta, armazenamento e distribuicao de dados ambientais e dados de logistica de trafego [ref].

A plataforma une dados gerados por sensores comuns, que vem na maioria dos automoveis atuais com sensores de geolocalizacao, presente na maioria dos smartphones. Com isso, cria-se uma base de dados relevante para se medir o fluxo de automoveis nas avenidas e quantidade de emissao de poluentes nas cidades.

#### A. Motivacao

Um dos grandes desafios da sociedade nos nossos dias eh preservar e desenvolver a mobilidade urbana de maneira sustentavel, isto eh, mitigar o impacto do transporte de passageiros meio-ambiente [ref].

Alem disso, observar padroes de comportamento nos motoristas e nas cidades ao longo do mundo pode ser interessante tanto para ajuda-los em alguma coisa, quanto para ver diferencas culturais de paises e culturas diferentes evidenciadas no modo em que dirigem.

#### B. Objetivos

O objetivo desse trabalho eh fazer uma exploracao basica pela base dados para ganhar intuicao sobre ela. Para observar se padroes basicos de comportamento dos usuarios podem ser detectados conforme ideias basicas que se tem sobre isso e, por fim, para saber se a base de dados pode ser utilizada como base para experimentos mais complexos, como treinamento de modelos de aprendizagem para um ganho no mundo

relevante, como melhorias no trafego, detecao de anomalias em comportamentos e outras coisas mais.

**Contributions:** Acredito que esse trabalho possa dar um insight interessante sobre o quao relevante esse dataset pode ser para a observacao de comportamento dos motoristas.

### C. Resultados esperados

In order to produce this application, we start with this processing, followed by this technique. In order to cope with this challenge, we introduce this formulation to produce this intermediate result. The formulation leads to this type of system, which is efficiently solved by adapting this technique. The final result is produced by this transform. The whole process is schematized in Fig. ??.

## II. CARACTERIZACAO DOS DADOS

Aqui sera listada, primeiramente, uma descricao sobre a estrutura dos dados apresentados no data set e sem seguida sera apresentado o problema e questoes que as visualizacoes pretendem responder.

### A. Caracterizacao dos dados

Cada item do dataset eh composto por dois atributos: 'properties' e 'features'.

- properties: possui as caracteristicas gerais do veiculo em questao. Esta subdividido nos seguintes itens:
  - type – atributo categorico – car or truck
  - constructionYear – discreto
  - model – categorico
  - fuelType – categorico
  - engineDisplacement – discreto
  - manufacturer – categorico
- features: possui as caracteristicas do deslocamento – ou da 'viagem' em si. A viagem eh caracterizada pela composicao, a cada dois segundos, do seguinte conjunto de dados:
  - coordinates – geoespacial – lat/long
  - speed – continuo – km/h
  - rpm – continuo – u/min
  - gps accuracy – nao sei – nao sei
  - maf – nao sei – nao sei
  - engine load – nao sei – nao sei
  - gps pdop – nao sei – nao sei

- o2 lambda voltage – nao sei – nao sei
- throttle position – nao sei – nao sei
- consumption – continuo – l/h
- gps vdop – nao sei – nao sei
- gps speed – continuo – km/h
- gps bearing – nao sei – nao sei
- intake pressure – nao sei – nao sei
- co2 – atributo continuo – kg/h
- time – atributo temporal – timestamp

### B. Questoes a serem respondidas

Existe uma serie de questoes que me vem rah cabeca nesse instante que, inclusive, ja tomei notas.

- quais marcas/modelos consomem mais gasolina?
- quais marcas/modelos poluem mais o meio ambiente?
- qual o perfil de pilotagem de cada automovel?
- muita gente usa esse dataset?
- de que lugares sao as pessoas que utilizam esse sistema? no Brasil, no mundo?
- existe muita gente utilizando esse sistema hoje em dia?
- quais os modelos e fabricantes mais comuns?
- os modelos e fabricantes mais comum na Europa sao os mesmos do Brasil?
- podemos descobrir quais sao as regioes mais poluidas e regioes menos poluidas?

## III. PREPARACAO DOS DADOS PARA VISUALICAO

Aqui eu vou descrever tudo o que fiz para baixar os dados. Aqui vou explicar por que usar o Elasticsearch tb.

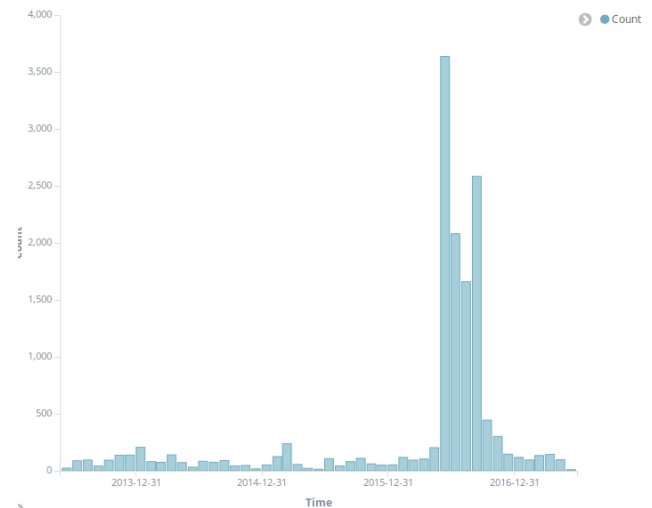
## IV. TECNICA DE VISUALIZAO DESENVOLVIDA

A exploracao dos dados foi feito com a ferramenta Kibana.

Ela, tipicamente, eh utilizada para analise por inspeo manual e visualizao de informaes presentes no Elasticsearch (MARINO, 2015; VAARANDI; PIHELIGAS, 2014). Dessa forma, eh ela que ira apresentar os dados armazenados pelo Logstash no Elasticsearch, em uma interface, via browser, altamente customizvel com histogramas e outros painis que propiciam uma viso geral sobre os dados. O Kibana possibilita transformar os logs em informaes teis (valor) atravs de Dashboards, pois permite realizar correlao de eventos, filtrar logs por origem, hospedeiros, entre outras combinaes (VAARANDI; NIZINSKI, 2013)

### A. Histograma

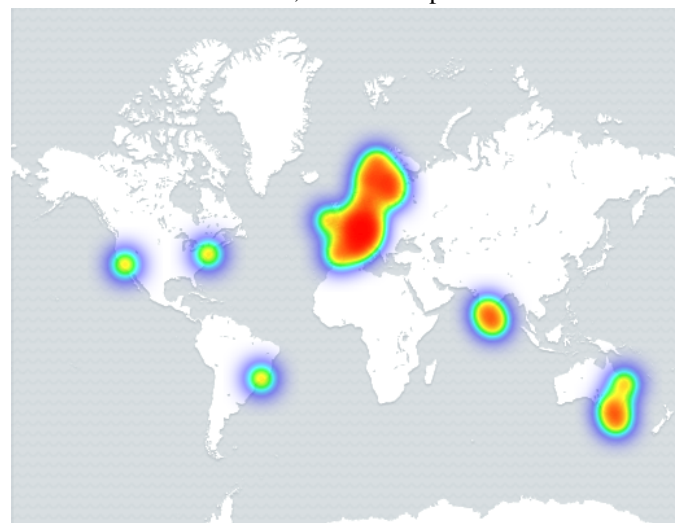
Com o intuito de saber se muita gente ou se pouca gente utiliza a plataforma enviocar, foi plotado um histograma. Com ele, alem de saber se muita gente ou se pouca gente o utiliza, podemos ver a evolucao de sua utilizacao ao longo do tempo.



No eixo x, observamos a quantidade de uso da plataforma ao longo do tempo.

### B. Heatmaps

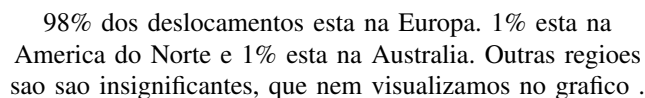
Outra duvida que eu tinha no inicio era a distribuicao geografica das pessoas que utilizam a plataforma. Como a plataforma foi desenvolvida na Alemanha, supus que a maioria das pessoas estavam utilizando por l. Estava realmente curioso para saber se existem pessoas em outros lugares do mundo a utilizando tambem. Para isso, um heatmap veio a calhar:



Aqui, podemos observar que a maioria dos usuarios, de fato, se encontra, na Europa. Contudo, existem pessoas utilizando em diversos outros pontos do globo tambm.

### C. Grafico em pizza

Baseado na visualizacao do heatmap, montei o globo dividido em 6 regioes: America do Norte, America do Sul, Australia, Europa, India. Com isso, criei um grafico de pizza, que mostra que, de fato, a maioria esmagadora das pessoas esta na Europa. Existe alguma representatividade tambem na America do Norte e na Australia.



Uma duvida que eu tive foi quais sao os fabricantes de automovel mais presentes no banco de dados em cada regio do globo. Achei que a divisao por sunburst ficava um pouco complicada de visualizar; entao, fiz n graficos em pizza, em que cada grafico corresponde a uma regio do globo e eh exibida a fatia de marcas por regio.

Botar uma frase bacana aqui.

Para responder a pergunta de quais são os modelos mais utilizados por fabricante, achei bem interessante de fazer um sunburst. No primeiro nível, observamos quais são as 5 marcas mais utilizadas e no segundo nível se observam os 3 modelos mais utilizados por cada fabricante. Achei prudente colocar somente 3 modelos no segundo nível, pois eu tentei colocar mais e ficou uma coisa muito confusa.

Botar uma frase bacana aqui.

[illegible][illegible]

#### ACKNOWLEDGMENT

The authors would like to thank this colleague and this financing institute.