

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: <http://inf.ufrgs.br/~rczembruski>

Abstract—Envirocar eh um projeto que coleta e armazena dados ambientais e dados de trafego de automoveis. Sao dados de sensores implantados dentro dos veiculos que trazem informacoes precisas sobre navegacao dos carros e dados relativos a conservacao do meio-ambiente, como emissao de gas-carbonico e consumo do veiculo.

Esse artigo tem por objetivo caracterizar os dados e apresentar uma serie de visualizacoes para analise exploratoria desse banco de dados. Com isso, sera obtida uma intuicao sobre o tipo de dados apresentado. Alem disso, serao obtidas informacoes acuradas sobre a viabilidade do uso desse banco de dados para modelagem de algoritmos de aprendizado de maquina.

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 supostamente relevante para que seja medido 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 e no bem-estar geral das pessoas [ref].

Uma base de dados consistente eh crucial para que se confirme – ou refute – hipoteses relativas a comportamento no transito em diferentes partes do globo. Alem disso, a base consistente pode ser muito importante para a inferencia de padroes comportamento e de pilotagem dos motoristas de automovel.

Uma analise exploratoria eh um passo necessario para que se verifique se a base realmente pode ser utilizada com os propósitos citados ou se ela deve ser descartada.

B. Objetivos

O objetivo desse trabalho eh fazer uma exploracao basica pela base dados para ganhar intuicao sobre ela. Serao obser-

vados padroes basicos de comportamento em diferentes partes do globo e tambem padroes individuais de pilotagem para determinados motoristas.

Com isso, sera determinado se o uso dessa base pode ser aconselhavel em experimentos mais complexos, como treinamento de modelos de aprendizagem .

Contribuicoes: O trabalho dara uma sugestao sobre a possibilidade de uso da base de dados envirocar em projetos de aprendizagem de maquina. Tambem sera feita uma analise comparativa entre diferentes tecnicas de visualizacao dos dados coletados.

II. CARACTERIZACAO DOS DADOS

Os dados sao obtidos atraves de sensores comuns em diversos automoveis (OBD-II) e enriquecidos com informacoes de geolocalizacao que estao presentes em smartphones. Para preservar a identidade dos usuarios, suas informacoes pessoais nao sao disponibilizadas e tambem sao retiradas as informacoes dos primeiros e ultimos 200m de deslocamento, com objetivo que o usuario nao seja identificado atraves das informacoes de origem e de destino.

Uma vez que os automoveis estao em deslocamento, suas informacoes mudam rapidamente. Desse modo, informacoes populadas no dataset a cada 2 segundos.

Os dados originais do servidor sao persistidos num banco de dados NoSQL – MongoDB – e oferecidos abertos ao publico atraves de um RESTful web service. Os dados sao oferecidos em um formato JSON, que eh simples de entender e processar.

Para realizacao desse trabalho, foram acessadas as apis rest e persistidas as informacoes de cada deslocamento no banco de dados Elasticsearch, por motivos que serao esclarecidos em seguida.

A. Caracterizao geral dos dados

Cada item do dataset eh composto por dois atributos principais, que sao subdivididos em diversas partes e serao mais detalhados na sequencia: 'properties' e 'features'.

- **properties:** possui as caracteristicas gerais do veiculo em questao. Esta subdividido nos itens apresentados na tabela 1. Veja um exemplo de propriedades extraido do banco de dados.

```
"properties": {
```

TABLE I
PROPERTIES: PROPRIEDADES GERAIS DE CADA VEICULO

Variavel	Descricao
type	categorica
constructionYear	discreta
model	categorica
fuelType	categorica
engineDisplacement	discreta
manufacturer	categorica

TABLE II
FEATURES: INFORMACOES DE CADA TIMESTAMP DO DESLOCAMENTO

Variavel	Unidade de medida
coordinates	geoespacial
speed	continuo
rpm	continuo
gps accuracy	continuo
maf	continuo
engine load	continuo
gps pdop	continuo
o2 lambda voltage	continuo
throttle position	continuo
consumption	continuo
gps vdop	continuo
gps speed	continuo
gps bearing	continuo
intake pressure	continuo
co2	continuo
time	temporal

```

"sensor": {
  "type": "car",
  "properties": {
    "constructionYear": 2011,
    "model": "Avensis",
    "fuelType": "gasoline",
    "id": "574e78cbe4b09078f97bbb4a",
    "engineDisplacement": 1800,
    "manufacturer": "Toyota"
  }
}

```

- features: possui as características do deslocamento – ou da 'viagem' em si. A viagem é caracterizada pela composição, a cada dois segundos, do seguinte conjunto de dados:

```

{
  "geometry": {
    "coordinates": [
      6.443663779195363,
      51.20348336793408
    ],
    "type": "Point"
  },
  "type": "Feature",
  "properties": {
    "phenomenons": {
      "Speed": {
        "value": 34.647194623947144,
        "unit": "km/h"
      },
      "Rpm": {
        "value": 1584.3050694465637,
        "unit": "u/min"
      }
    }
  }
}

```

```

"GPS Accuracy": {
  "value": 2.999999910593033,
  "unit": "%"
},
"MAF": {
  "value": 9.437765815629945,
  "unit": "l/s"
},
"Engine Load": {
  "value": 43.96216858346017,
  "unit": "%"
},
"GPS PDOP": {
  "value": 1.4999999776482582,
  "unit": "precision"
},
"O2 Lambda Voltage": {
  "value": 3.2762881521175586,
  "unit": "v"
},
"Throttle Position": {
  "value": 21,
  "unit": "%"
},
"Consumption": {
  "value": 3.102402130874109,
  "unit": "l/h"
},
"GPS VDOP": {
  "value": 1.1779116287827491,
  "unit": "precision"
},
"GPS Speed": {
  "value": 33.47147411240462,
  "unit": "km/h"
},
"GPS HDOP": {
  "value": 0.899999986588955,
  "unit": "precision"
},
"Intake Pressure": {
  "value": 44.14216932654381,
  "unit": "kPa"
},
"GPS Bearing": {
  "value": 304.7174273121018,
  "unit": "deg"
},
"Intake Temperature": {
  "value": 13.000000387430191,
  "unit": "c"
},
"CO2": {
  "value": 7.290645007554156,
  "unit": "kg/h"
},
"O2 Lambda Voltage ER": {
  "value": 0.9988191702782387,
  "unit": "ratio"
},
"GPS Altitude": {
  "value": 104.90115790988267,
  "unit": "m"
}
},
"id": "590ad752268d1b08a47f18d4",
"time": "2017-03-27T04:51:05Z"

```

}
}

B. Questões a serem respondidas

A ideia básica é responder questões relativas a (a) padrões de motoristas em diferentes partes do globo, (b) padrões de comportamento individual no volante e (c) possibilidade de uso da base em modelos de aprendizagem de máquina.

- Existem muitos registros nesse banco de dados?
- Em que período de tempo esse banco de dados foi utilizado? Ainda hoje ele é bastante utilizado?
- Em que regiões do globo estão os usuários desse sistema?
- Os modelos e fabricantes mais comuns na Europa são os mesmos do Brasil?
- Quais marcas e modelos consomem mais combustível?
- Quais os modelos e fabricantes mais comuns?
- Quais são as marcas e modelos que mais poluem? quais são as marcas e modelos que menos poluem?
- Podemos descobrir quais são as regiões mais poluídas e regiões menos poluídas?
- É possível utilizar este dataset para extrair padrões de comportamento de um determinado usuário?

III. PREPARAÇÃO DOS DADOS PARA VISUALIZAÇÃO

O *envirocar* é uma base aberta de dados. Entretanto, não existe uma forma simples para fazer download da base inteira. A entrega dos dados é feita via serviços REST que possibilitam acesso a determinadas partes do banco de dados.

Para que se realizasse a obtenção da base inteira para que pudessem ser feitas as visualizações, utilizou-se um script capaz de varrer a base de maneira sequencial. Esse mesmo script é responsável pela persistência dos dados no *Elasticsearch*.

A base de dados possui 15 mil viagens. O número bruto não é tão grande quanto se supunha, mas, para cada viagem, é enviado ao banco de dados uma série de informações a cada dois segundos. Dessa maneira, embora o número de viagens seja relativamente pequeno, o tamanho de cada viagem é grande. A indexação no *Elasticsearch* ocupou 6 Gb de memória em disco.

IV. TÉCNICA DE VISUALIZAÇÃO DESENVOLVIDA

A exploração dos dados foi feita com três abordagens. Uso de Kibana, linguagem R e API de mapas do Google.

Kibana é uma ferramenta utilizada para análise por inspeção manual e visualização de informações que funciona de maneira natural com o *ElasticSearch*. Dessa forma, é ela que irá apresentar os dados armazenados no *ElasticSearch*, em uma interface, via browser, altamente customizável com histogramas, mapas e outros painéis que propiciam uma visão geral sobre os dados. O Kibana possibilita transformar os logs em informações (valor) através de Dashboards, pois permite realizar correlação de eventos, filtrar logs por origem, hospedeiros, entre outras combinações (VAARANDI; NIZIN-SKI, 2013).

Foi utilizado o *ElasticSearch* como base primária dos dados, pois além do suporte à persistência, vem junto com uma série

de mecanismos e algoritmos de recuperação de informação. A ferramenta permite combinar geolocalização com outras técnicas de recuperação de informações baseadas em texto.

Outra ferramenta para visualização de informações utilizada foi a linguagem R. A linguagem também possui uma variada gama de técnicas para visualização de dados, além de possuir um suporte muito grande a questões de estatística e de probabilidade.

O R não é uma ferramenta de tão alto nível de abstração quanto o Kibana em que, uma vez que os dados estão persistidos, se montam visualizações de maneira simplificada. Em vez disso, é necessário que se escreva código para que as visualizações apareçam. Os gráficos surgem da maneira mais crua do que no Kibana, mas possui muito mais flexibilidade, maleabilidade e extensibilidade. Além disso, existe um conjunto muito mais vasto de visualizações que o Kibana, com vieses mais estatísticos, como *boxplots* e matrizes de correlações.

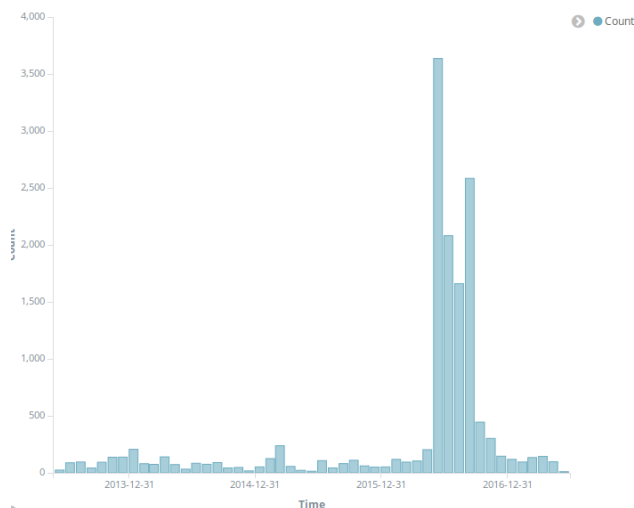
Por fim, para detectar padrões de comportamento individuais, foi implementada utilizando a API do Google Maps uma ferramenta simples para visualizar por quais lugares um determinado motorista frequenta com maior frequência.

A. Histograma

Com o intuito de saber se muita gente utiliza a plataforma *envirocar*, foi desenvolvido um histograma em que o eixo x corresponde à variável temporal e o eixo y corresponde à quantidade de viagens naquele período de tempo. Com isso, poderá ser inferido se bastante gente vem utilizando a ferramenta e, principalmente, se seu uso vem crescendo ou decrescendo com o passar do tempo. É um estudo de tendências, portanto.

Pode-se observar o gráfico abaixo, que não existe uma tendência de aumento ou de diminuição ao longo do tempo. O que se observa é um 'boom' de uso no período em volta do ano de 2016, mas que não se mantém no restante do tempo. No período de novembro de 2016, por exemplo, tivemos 3600 viagens naquele mês. Mas no início de 2017, não passavam de 100 viagens por mês.

Não se pode observar, portanto, um uso massivo da plataforma. Entretanto, temos uma amostra que parece razoável para que se possa tentar responder as outras perguntas do artigo.



No eixo x, observamos a quantidade de uso da plataforma ao longo dos meses.

B. Heatmaps

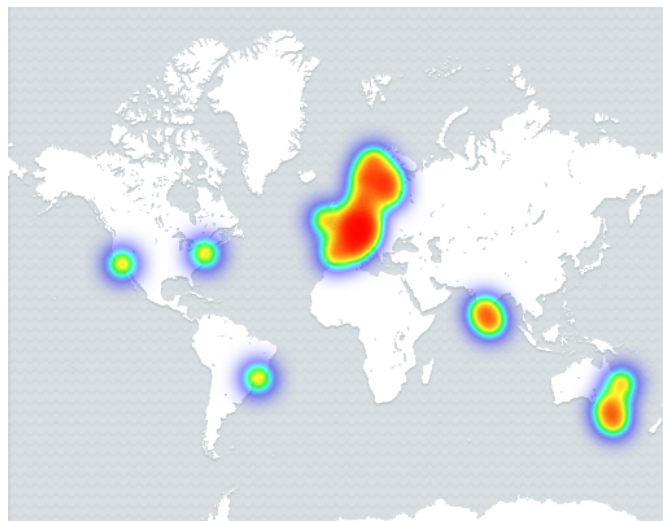
Uma vez visualizada a distribuicao temporal da utilizacao da plataforma, ira ser observada a distribuicao geografica dos usuarios. Como ja se sabe que eh uma plataforma desenvolvida na Alemanha, pode-se supor que muita gente na Europa ira utilizar a ferramenta, em relacao as outras regioes do globo.

A visualizacao por calor, utilizando como metrica a quantidade de viagens dispostas no globo terrestre, dara um senso de quais regioes sao mais utilizadas a ferramenta.

Observando o grafico a seguir, observamos aquilo que ja era suposto de antemao: de fato, na Europa, a plataforma eh mais utilizada. Alem disso, conseguimos obter alguns dados que nao eram imaginados. Vemos pontos isolados de uso nos Estados Unidos, Brasil, India e Australia.

Com as informacoes geoespaciais, ganha-se confianca de que poderao ser respondidas outras questoes, como o comportamento dos motoristas em diferentes regioes do globo. Alem disso, pode-se utilizar esse grafico para clusterizar de maneira macro o globo terrestre de acordo com essa base de dados: America do Norte, America do Sul, Europa, India e Australia.

A visualizacao de heatmap traz uma ideia muito clara sobre os pontos que mais utilizam a ferramenta. Traz-nos a ideia de que em diversos pontos do globo ela esta sendo utilizada, com enfase na Europa. Traz confianca de que poderemos observar padroes de comportamento em diversos pontos do globo. Ela traz uma ideia basica, mas nao definitiva, da quantidade de pessoas que utiliza a ferramenta ao longo do globo.



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

O heatmap deu uma ideia basica sobre a quantidade de pessoas que utilizam a plataforma ao redor do globo. Entretanto, nao trouxe uma ideia definitiva sobre as proporcoes de cada regioao.

Para isso, sera utilizado um grafico em pizza, que traz claramente a ideia de proporcoes ao usuario, ao dispor em fatias cada segmento analisado.

Nesse grafico, sera utilizado o cluster inferido visualmente do heatmap com as seguintes regioes: Europa, America do Norte, America do Sul, India e Australia.

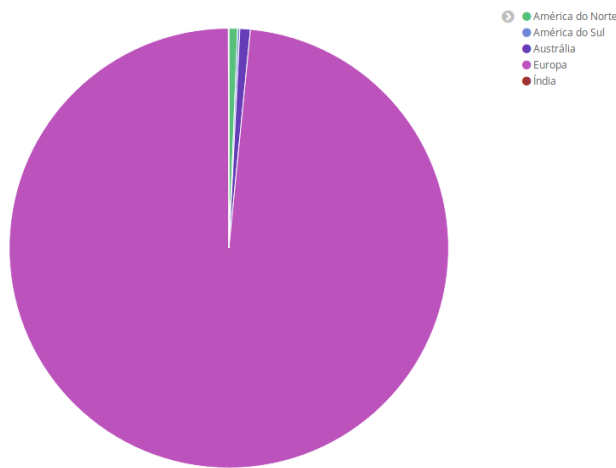
Aqui, pode-se observar a imensa maioria, de fato, utilizando na Europa: 97% dos usuarios esta la. Temos alguma representatividade, com 1% dos Usuarios na America do Norte e 1% na Australia.

Na America do Sul e na India, nao chega a 1% de usuarios. Parece ter sido realmente um usuario curioso que comecou a utilizar a plataforma em sua terra natal.

Essa informacao podera ser utilizada de uma maneira que achei incrivel: para comparacao de dados globais, como estatisticas e tentativas de detecao de comportamento em massa, devera ser dado foco nos usuarios da Europa.

Entretanto, essa falta de representatividade em outras partes do globo podera favorecer um olhar mais individualista. Por exemplo, pode-se acreditar que na America do Sul e na India, eh sempre a mesma pessoa que utiliza a ferramenta. Desse modo, poderemos observar algum comportamento individual. Do mesmo modo pode ser feito na America do Norte e na Australia, com um pouco mais de criterio, pois la houve muito mais uso.

Entretanto, na Europa, nao podemos fazer isso, pois eh muita gente usando. Na Europa, serao feitas analises mais genericas.



98% dos deslocamentos esta na Europa. 1% esta na America do Norte e 1% esta na Australia. Outras regioes sao insignificantes, que nem visualizamos no grafico .

D. Graficos em pizza

Observadas algumas questoes temporais e espaciais, irao ser analisados alguns padroes de comportamento em diferentes regioes do globo. Sera utilizado aquele cluster inferido do primeiro heatmap para fazer uma analise sobre quais sao os fabricantes mais utilizados em diferentes partes do globo.

Sobre padroes de comportamento, espera-se verificar se supostos padroes culturais podem de uma determinada regio do globo pode ser inferida atraves deste banco de dados. Por exemplo, eh verdade que americanos gostam de carros espacosos? Eh verdade que na Europa e Estados Unidos os carros sao muito melhores do que os da America do Sul?

Para responder essa pergunta, algumas visualizacoes foram feitas algumas suposicoes de tecnicas de visualizacao: sunburst e nugget.

A tecnica de sunburst nao pareceu muito adequada devido a quantidade de informacoes – muito grande para um espaco pequeno – e o biscoito nao transpareceu de maneira clara o que estava sendo perguntado.

Para isso, foi utilizada novamente a tecnica de visualizacao de graficos em pizza. Entretanto, agora, em vez de plotar somente um grafico, o foram plotados 5 graficos em pizza, em que cada pizza corresponde a uma regio do cluster. Essa visualizacao foi escolhida em detrimento do sunburst pois a segmentacao trouxe mais evidencia para responder aquilo que estava sendo perguntado. O sunburst trouxe muita informacao em muito pouco espaco fisico, o que pareceu meio confuso.

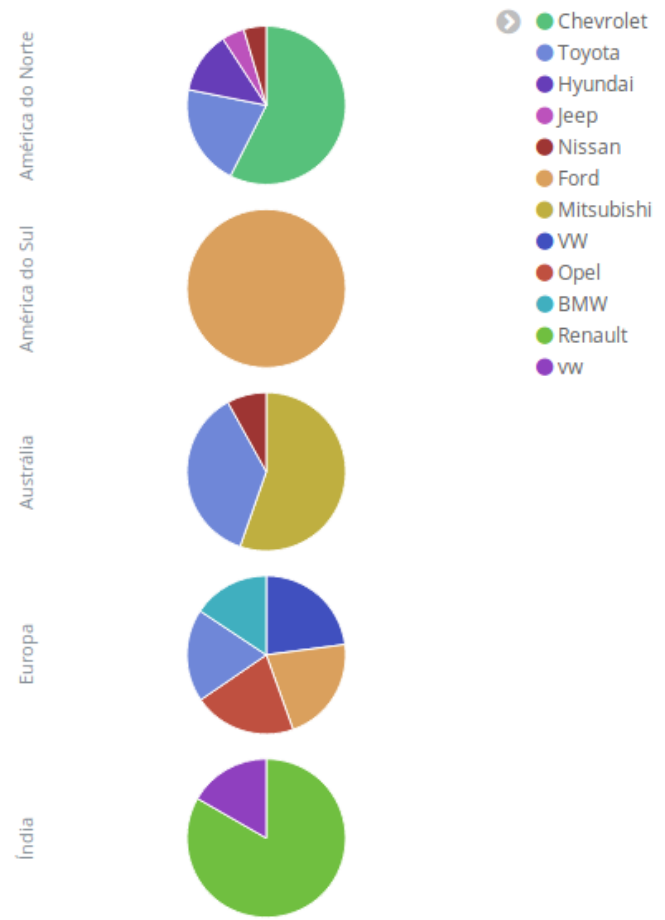
Observando os fabricantes, vemos que as principais marcas na America do Norte, Australia e Europa sao as mesmas.

- America do Norte: Chevrolet, Toyota e Hyundai
- Australia: Ford, Toyota e Opel
- Europa: Toyota, Ford, Opel, VW, BMW

Ja na America do Sul e India, parecem ser regioes mais humildes tambem, pois aqui nao observamos Hyundai e BMW.

- America do Sul: Ford
- India: Renault e VW

Baseado nessa amostra de dados, nao se pode chegar a uma conclusao definitiva para as questoes levantadas. A distribuicao entre Europa, Estados Unidos e Australia ficaram razoavelmente parecidas. Ja na America do Sul e India, realmente parecem regioes menos desenvolvidas. Entretanto, a amostra dos dados nao parece ser significativa o suficiente para uma conclusao assertiva. A observacao do grafico parece corroborar a ideia inicial, utilizada para fazer as perguntas, mas nao ha evidencias que, de fato, elas acontecem.



Estados Unidos e Europa apresentam resultados parecidos. America do Sul supoe que somente um cara utilizou a plataforma.

E. Sunburst

A Sunburst visualization is a radial space-filling visualization technique for displaying tree like structures. There are other space-filling visualization methods that use other visual encodings for describing hierarchies.

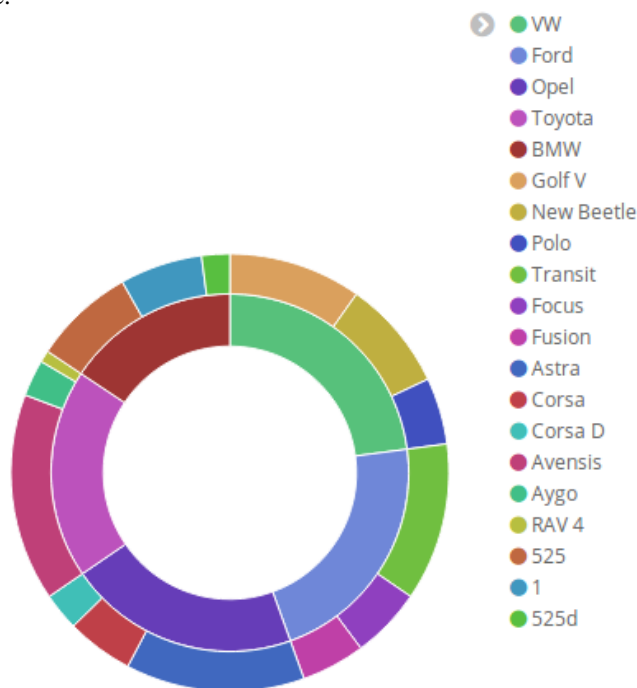
Essa tecnica mostra-se muito adequada para condensar informacoes hierarquicas, como para responder a pergunta: para cada fabricante, qual eh o modelo de automovel mais utilizado.

Como na Europa existe uma amostra muito mais significativa do que em outras regioes do globo, a pergunta fica limitada a regio da Europa. Alem disso, para que a visualizacao nao acontecesse de maneira muito confusa, para que ela ficasse clara, foram escolhidos os 5 fabricantes mais utilizados e os 3

modelos mais utilizados por cada fabricante. Se não houvesse essa limitação, o gráfico ficaria muito poluído, o que gera um desconforto ao usuário.

Esse gráfico, novamente, corrobora a intuição de que na Europa o nível dos automóveis é maior do que no Brasil. Observamos que os fabricantes mais comuns lá são os mesmos que os presentes no Brasil. Exceto pela BMW, que muito pouco se observa por aqui, vemos que VW, Ford, Opel (GM) e Toyota são os mais presentes por lá.

A diferença de padrão de veículos se observa nos modelos mais comuns, uma vez que os fabricantes são os mesmos. Por exemplo, os modelos mais comuns da VW são New Beetle, Golf e Polo. Notavelmente, são veículos mais sofisticados que os modelos comuns brasileiros. O mesmo se observa nos veículos da Ford, cujos modelos mais populares são Transit, Fusion e Focus. Aqui, esses automóveis são automóveis de elite.



Observamos que as marcas mais comuns na Europa são as mesmas do Brasil, exceto pela BMW. Nunca vemos BMWs pelo Brasil.

F. Radar

Para que se possa identificar quais são as marcas e modelos que mais poluem e que menos poluem de uma forma bastante rica, optou-se por fazer uma visualização em forma de radar. Dessa forma, foi possível ver, para cada fabricante, uma série de atributos que parecem ser correlacionados. São eles: rotação do motor (rpm), consumo, velocidade, duração da viagem, tempo de viagem e quantidade de gás carbônico emitido.

Para que isso pudesse ser sintetizado adequadamente na forma de um radar, foram feitas uma série de computações em cada um dos itens. Lembrando que diversos desses atributos mencionados, como CO_2 , rpm, consumo e velocidade mudam periodicamente – a cada dois segundos –, foi feita

uma média para cada um desses atributos em cada viagem. Posteriormente, foi feita uma normalização desses dados, pois a ferramenta de visualização que o Kibana oferece não permite que se coloque diferentes escalas em cada um dos eixos.

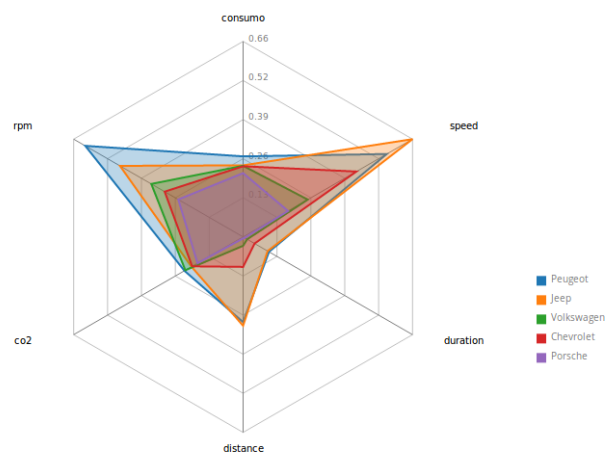
Computadas essas médias e normalizados os dados, podemos observar o gráfico de radar. E ele nos leva a algumas informações bastante curiosas:

Peugeot é o fabricante que normalmente se anda com a rpm mais alta e é o cara que tem maior consumo, mas não é o fabricante que anda mais rápido.

O fabricante que anda mais rápido é o Jeep, e ele consome menos que os Peugeot, normalmente.

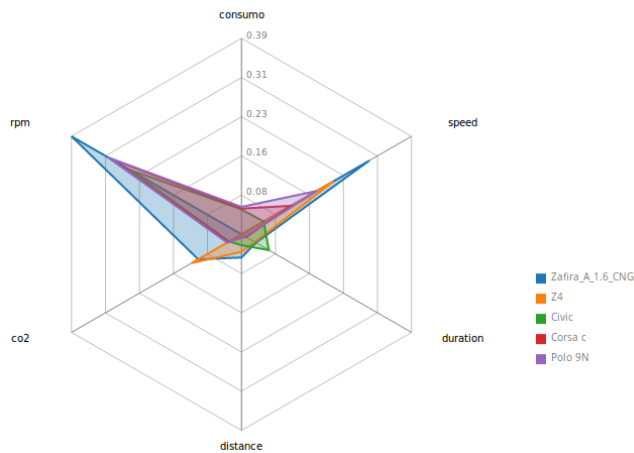
Curioso observar também, que o Porsche é a marca que se anda com rpm mais baixa – ao contrário do que normalmente se imagina – e cuja velocidade é mais baixa também. Acredito que isso se deve porque o Porsche deve ter uma representatividade muito baixa na amostra. Portanto, se você quiser andar rápido, não compre um Porsche!

Para quem deseja comprar um automóvel que consuma pouco e que ande com a rotação do motor baixa, recomenda-se, de acordo com esse dataset, o uso de um Chevrolet.

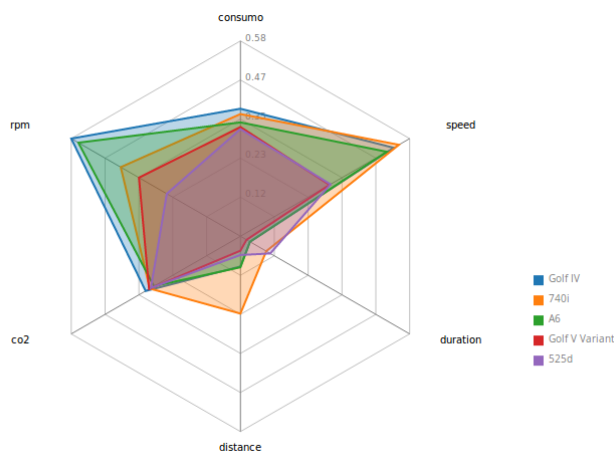


Diversos atributos seccionado por fabricante.

O mesmo tipo de visualização pode ser aplicado, não aos fabricantes, mas aos modelos dos automóveis em si. Na figura X, observam-se os modelos que menos poluem – observe que dos 5 apresentados, 2 são Chevrolet, o que está de acordo com o gráfico de radar apresentado por fabricantes – e observe que os que mais poluem são modelos esportivos, cuja rotação do motor normalmente é alta.



Modelos que menos poluem.



Modelos que mais poluem.

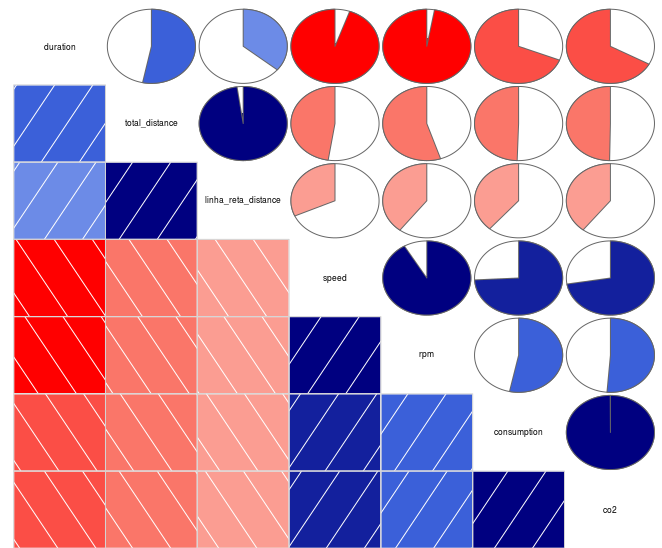
G. Matriz de correlacao

Como foi observado no grafico de radar que a rotacao do motor esta intimamente ligada ao consumo dos automoveis, decidi por fazer uma apresentacao de uma matriz de correlacao. O intuito, com essa visualizacao eh verificar quais atributos estao mais ligados um ao outros. Por exemplo, para ter certeza de que, se o automovel esta com rotacao maior, que ele esta poluindo mais. Alem disso, para poder observar se faz sentido associar a velocidade ah quantidade de emissao de poluentes.

Nessa visualizacao, ao contrario das anteriores, foi utilizada linguagem R. Ela possui mais suporte a visualizacoes e trabalhos estatisticos do que o Kibana.

Nessa representacao, quanto mais azul e quanto mais cheio esta o circulo superior, maior eh a correlacao entre os atributos. Observamos que quanto maior o consumo de combustivel, maior eh a quantidade de gas carbonico emitido. Observamos tambem que existe grande correlacao entre rpm e velocidade, o que me surpreendeu um pouco, ate. Outra coisa que me surpreendeu um pouco foi que parece nao haver correlacao entre velocidade e duracao de uma viagem. Me surpreendeu porque normalmente eu ando em velocidade mais alta quando

faco viagens mais longas. Mas acho que na Europa nao eh assim que acontece. Existem outras coisas que se pode observar por ai tb.



Diversos atributos seccionado por fabricante.

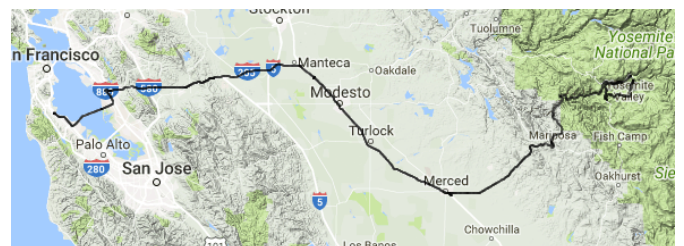
H. Mais um heatmap

Com intuito de fazer uma analise micro, isto eh, em vez de analisar uma populacao inteira em determinada regioao, decidi por fazer umas visualizacoes micro. Isto eh, para cara conjunto de viagens que eu tenho razoes suficientes para crer que sejam da mesma pessoa, fiz um heatmap das viagens que ela faz.

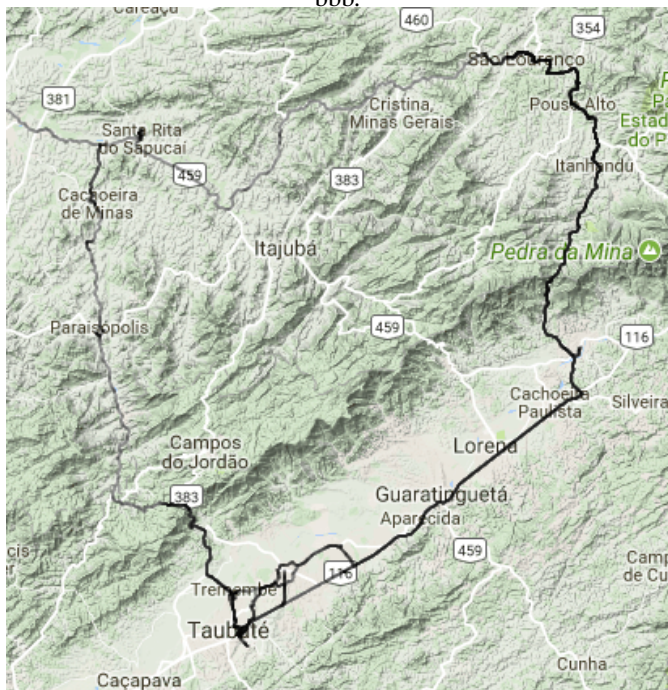
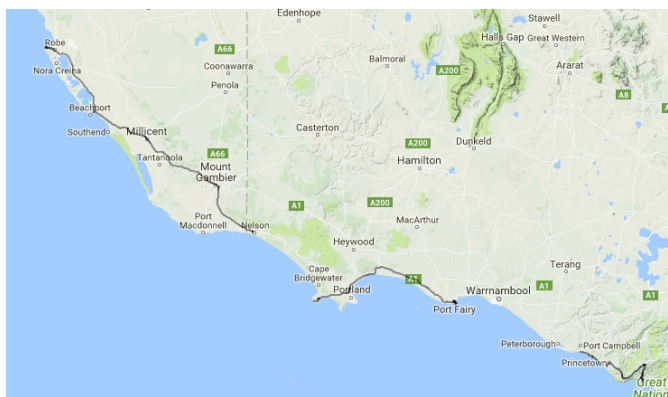
Com isso, observamos que em muitos casos, as pessoas realmente costumam fazer o mesmo deslocamento periodicamente. Observe as figuras X e Y, em que os pontos muito escuros indicam que o usuario fez aquele caminho muitas vezes.

Com isso, observamos que esse dataset pode ser utilizado, sim, para treinar algum algoritmo que seja capaz de prever, para um determinado usuario, onde ele estara em determinado periodo de tempo.

Alem disso, pode-se fazer uma analise em real time para saber o quao longe o usuario esta dos lugares que ele costuma ir.



aaa.



V. RESULTADOS

As visualizacoes deram uma visao geral sobre a base de dados, e praticamente todas as perguntas que puderam ser respondidas.

- observamos que, de fato, a maioria dos usuários esta na Europa. Alem disso, observamos que por um curto periodo de tempo, houve muita gente utilizando a plataforma, mas que essa quantidade de gente a utilizando nao se sustentou por muito tempo.
- pode-se observar alguns padroes de comportamento utilizando esse dataset: na Europa, os fabricantes mais comuns sao praticamente mesmos fabricantes do Brasil. Entretanto, os carros comercializados por esses fabricantes por la sao superiores aos brasileiros.
- observamos de maneira clara, atraves da visualizacao de radar, quais sao as marcas e fabricantes cujos veiculos poluem mais e quais poluem menos. Vimos, nesse mesmo grafico de radar, quais sao os que gastam mais e os que gastam menos combustivel.

- observamos ainda que esse dataset, embora não seja tão rico quanto se supunha no início, ainda assim ele pode ser utilizado para se detectar padrões de comportamento individuais nas pilotagens.
- COLOCAR QUE, DE FATO, AS VIAGENS SE REPETEM PERIODICAMENTE.

VI. CONCLUSAO

Foi atingido objetivo de se fazer uma exploracao basica na base de dados oferecida pelo envirocar. Obteve-se intuicao sobre essa base de dados e padroes de comportamento dos motoristas de automovel.

Padrões culturais, como o tipo de veículos e fabricantes utilizados na Europa em comparação com o Brasil foram identificados. Além disso, padrões individuais de pilotagem de automóveis foram identificados.

Alem disso, foi observadas as distribuicoes temporais e espaciais dos registros de veiculos.

A base de dados nao eh tao utilizada quanto se supunha – o autor imaginava que mais gente utilizasse a plataforma –, mas possui-se uma amostra significativa de pessoas para que se possa levar o trabalho mais adiante e utilizar essa base para treinar algum modelo de machine learning em trabalhos futuros.

Alem disso, pode-se fazer uma analise basica sobre o uso de ferramentas como Kibana e como R para visualizacao de informacoes. O Kibana, mais simples, apresenta graficos mais bonitos. O R, entretanto, apresenta graficos mais cruz e uma linguagem de baixo nivel. O R, portanto, vai ser mais aproveitado por usuarios mais experientes – como programadores –, enquanto o Kibana pode ser utilizado por usuarios mais leigos.

REFERENCES

- XXXX
- XXXX
- XXXX
- XXXX
- XXXX
- XXXX
- XXXX