

# ***Vehicle Energy Dataset - um dataset enriquecido, equilibrando integridade e anonimização dos dados***

**João Batista Firmino Júnior<sup>1</sup>, Dr. Francisco Dantas Nobre Neto<sup>1</sup>**

<sup>1</sup>Instituto Federal da Paraíba

Avenida Primeiro de Maio, 720

batista.firmino@academico.ifpb.edu.br, dantas.nobre@ifpb.edu.br

**Abstract.** *Obtaining displacement databases for research in the area of destination and trajectory prediction is a challenging task. In this sense, the study proposes to describe and mine the Vehicle Energy Dataset (VED), suitable for the evaluation of clustering techniques, with 384 vehicles and their displacements, concentrated in the city of Ann Harbor, Michigan (USA). This work combines the application of the ST-DBSCAN algorithms, as a space-time clustering technique, with that of map-matching, through an adequate Coordinate Reference System, in addition to counting the repetition of the origin-destination cells of the trajectories and, in the end, 231 vehicles were obtained with at least 1 trajectory with an origin-destination repetition of at least 2 times.*

**Resumo.** *Obter bases de dados de deslocamentos para pesquisas na área de predição de destinos e de trajetórias é uma tarefa desafiadora. Neste sentido, o estudo propõe-se a descrever e minerar o Vehicle Energy Dataset (VED), adequada para a avaliação de técnicas de agrupamento, com 384 veículos e seus deslocamentos, concentrados na cidade de Ann Harbor, em Michigan (EUA). Este trabalho combina a aplicação dos algoritmos ST-DBSCAN, como técnica de agrupamento espaço-temporal, com a do map-matching, mediante um Sistema de Referência de Coordenadas adequado, além de realizar contagem da repetição das células de origem-destino das trajetórias e, ao final, obtiveram-se 231 veículos com, no mínimo, 1 trajetória com uma repetição origem-destino de, no mínimo, 2 vezes.*

## **1. Introdução**

Diante de um contexto tecnológico apto à captação, armazenamento e disponibilização de bases de dados, bem como à existência de bibliotecas computacionais suficientemente robustas para manipulá-los e visualizá-los, há a necessidade de tratar a qualidade desses dados de diferentes origens, ao mesmo tempo que equilibrando o poder de atividades como predição de trajetórias com questões éticas de anonimização. Um exemplo que preenche esse cenário é o *Vehicle Energy Dataset* (VED)<sup>1</sup>, cujos dados vêm sendo utilizados para análises sobre economia de energia de carros elétricos ou híbridos, como é o caso do artigo de Oh et al. (2022), mas que se relacionam ao entendimento dos sistemas de mobilidade urbana em geral – são projetos que necessitam de base de dados reais para a validação de modelos computacionais, a exemplo de modelos de predição. Mas a “realidade” retratada por esses dados, por vezes, pode comprometer certos direitos

---

<sup>1</sup> Disponível em: <https://github.com/gsoh/VED>. Acessado em: 09 de agosto de 2023.

fundamentais. Por isso, há necessidade de certas técnicas de anonimização, que foram utilizadas pelos pesquisadores que conduziram o VED. Ao mesmo tempo, há a dificuldade em relação a encontrar uma base de dados apropriada para análise e validação de modelos computacionais em pesquisas relacionadas com mobilidade humana.

Outra questão importante diz respeito ao ineditismo dessa base de dados em relação à facilidade de obtenção mediante um repositório público (até a presente data) no *github*. É mais recente que outras bases de dados públicas, como a do *Geolife*, da Microsoft. O que não significa dizer que seja inédita em relação às bases privadas.

O foco deste artigo consiste não apenas no detalhamento do VED, uma base de dados relativamente recente. Mas também na modificação em relação à original, ganhando em poder de geração de conhecimento, preservando as questões éticas (como na anonimização das reais residências e/ou locais de trabalho dos usuários que colaboraram com a pesquisa). Assim, essa investigação da base de dados pode ser representada pela seguinte Questão de Pesquisa: “*Como melhor tratar os dados do VED de forma a encontrar padrões de agrupamento espaço-temporal, harmonizando com a anonimização aplicada no conjunto de dados original?*”. As contribuições do artigo são, de forma estruturada:

- Gerar dados no formato *datetime* a partir dos campos temporais fornecidos, mantendo o anonimato dos usuários;
- Encontrar as trajetórias percorridas mais de uma vez;
- Uma vez encontradas as trajetórias percorridas mais de uma vez, como gerar agrupamentos espaço-temporais.

Inicialmente, é necessário definir o principal termo de trabalho deste artigo. O termo “trajetória” (ou, especificamente, o conceito de *raw trajectory*) refere-se a uma sequência de pontos, no qual há a posição do objeto em movimento no espaço e no timestamp em que o ponto foi coletado (LEITE et al., 2019). Para Li et al. (2021), considerando um objeto em movimento, trajetória é uma sequência finita, uma ordenação sequencial de pontos espaço-temporais, que é amostrada de um rastreamento mediante algum dispositivo de localização. Para discretizar os dados, foi utilizado o conceito de “tesselação”, que, em tradução aproximada de Gold (2016), refere-se a “quadriculados discretos, com seus próprios atributos e, como eles são conectados, com alguma representação de vizinhança”, sendo bastante úteis como modelos para entender e processar o espaço geográfico (GOLD, 2016).

## 2. Trabalhos Relacionados

Em relação a trabalhos existentes, esta seção descreve brevemente trabalhos sobre: o conjunto de dados original (VED), anonimização de usuários, mineração de trajetórias e agrupamento espaço-temporal. Há o trabalho que apresentou o VED (OH et al., 2022), utilizando-o no contexto da análise de dados de consumo de energia, que ressalta haver uma anonimização sobre onde cada usuário vive, localizações específicas em tempos específicos e pontos de interesse pessoais, sem uma perda muito acentuada da resolução dos dados, mas também sem uma perda da privacidade dos usuários. Os métodos para a anonimização, foram o *random-fogging*, o *geofencing* e o *major intersections bounding* (delimitação de intersecção principal). Ainda um diferencial importante, uma maior

elaboração dos dados de dois campos: *Timestamp (ms)* e *DayNum*. Além disso, encontra as trajetórias percorridas mais de uma vez e a geração de agrupamentos espaço-temporais, úteis para atividades como predição de trajetórias, sem se concentrar no *eco-routing*, tema do artigo de origem.

Por outro lado, também houve o cuidado para se evitar o que é ressaltado em Farzanehfar et al. (2021), que consiste na possibilidade de re-identificação de usuários com base em conjuntos de dados de grande escala. Os autores conduziram pesquisa em que demonstraram a diminuição desse risco diminuir levemente com o tamanho do conjunto de dados. Mais especificamente, avalia-se que 93% das pessoas poderiam ser identificadas num conjunto de dados de 60 milhões de pessoas usando quatro pontos de informação auxiliar (a distribuição do ciclo circadiano; a distribuição da frequência de atividades; o número de registros por trajetória ao longo de uma distribuição; e lugares em que o usuário pode ter potencialmente estado) a partir de uma pesquisa com dados de um milhão de usuários, por três meses, analisando os resultados por análise de sensibilidade. Mesmo com essas adversidades, o tratamento dos dados para este trabalho não revelou essas informações, nem em termos individuais nem coletivos - exceto o número de registros por trajetória, coletivamente.

Sobre uma mineração de dados de trajetórias, necessária para o trabalho, Zheng et al. (2015), ao identificar, na época, uma ausência de uma revisão sistemática que situe a mineração desse tipo de dado, evidencia toda uma sistematização de onde e em relação a quem se situa esse conjunto de práticas, de técnicas e suas correlações e diferenças. Atluri et al. (2018) são mais enfáticos no uso de um termo como mineração de dados “espaço-temporais” ou *spatio-temporal data mining* (STDM), mas tanto Zheng et al. (2015) como esses autores consideram a abrangência desses conceitos e processos como além da mobilidade humana, incluindo, os primeiros, pessoas, veículos, animais e fenômenos naturais; e os últimos, por campos do conhecimento humano, e não propriamente por agentes, tais como Ciência Climática, Neurociência, Ciência Ambiental, Agricultura de Precisão, Epidemiologia, Mídias Sociais, Dinâmica de Tráfego (Transportes), Física do comportamento solar e Segurança.

Zheng et al (2015) consideram, em geral, o processo de mineração de dados de trajetórias como: filtragem de dados destoantes (de ruído), detecção de pontos de paradas, compressão de trajetórias, segmentação de trajetórias, *map-matching*, tudo para reduzir a incerteza contida nos dados e permitir melhor riqueza de informações. Enquanto Atluri et al. (2018) consideram, a partir das propriedades e tipos de dados de espaço-temporais, problemas e métodos como agrupamento, predição, mineração de padrões frequentes, desafios como detecção de anomalias, detecção de mudanças e mineração de relacionamentos.

## 2.1. Sobre agrupamento espaço-temporal

De acordo com Birant e Kut (2007), o ST-DBSCAN envolve quatro parâmetros: dois limites de distância (distância para atributos espaciais e distância para atributos não-espaciais), número mínimo de pontos por grupo e um último parâmetro utilizado para prevenir a descoberta de grupos combinados devido à existência de pequenas diferenças em valores não-espaciais vizinhos. Na prática, o trabalho foi feito com um *software* para atividades de Sistema de Informações Geográficas, que permite manipulações espaciais,

tanto com vetores como com imagens (*rasters*) e uso de banco de dados geográfico – o QGIS, que é *open source*, antes da utilização de *scripts* em Python. Entretanto, só é necessário especificar, no QGIS, os dados de entrada, um campo *datetime* (que pode ser entendido como o “limite de distância não-espacial” a que os autores se referem), o limite de distância espacial e, no máximo (por ser opcional), tratar os pontos das bordas como ruído ou não. Evidencia-se que é recomendável, nesse contexto, trabalhar com uma projeção *Universal Transverse Mercator* (UTM), em metros, para evitar discrepâncias em relação a coordenadas geográficas, que ocorrem em ângulos. Por isso, antes, a transformação entre sistemas e projeções.

Feng & Zhu (2016) também apresentam os procedimentos de pré-processamento, gerenciamento de dados e consultas. E só depois chegam à mineração de dados de trajetórias mediante mineração de padrões, agrupamentos, classificação e descoberta do conhecimento, e proteção de privacidade. Esses autores consideram a mineração de padrões como a análise de padrões de mobilidade de um ou múltiplos objetos em movimento. Mais precisamente, na análise do comportamento dos movimentos, inclui-se o comportamento emergente de seres humanos em relação a algum evento súbito recente. Porém, é a partir dessa análise do comportamento dos movimentos humanos que é possível explorar uma das questões em aberto, de acordo com esses autores, que é a complementação a partir de outros dados.

De modo geral, todos esses trabalhos permitem encontrar um caminho metodológico minimamente consistente na mineração dos dados de trajetórias, com suas etapas, sequenciais ou paralelas, que pode gerar a resolução de várias questões de pesquisa que venham a ser feitas.

### 3. Metodologia

Explorando os dados, os algoritmos e, principalmente, a relação e interação entre os dados de trajetórias e os algoritmos, pretende-se explicar cada etapa da condução da pesquisa a partir de alguns *scripts* preparados para a mineração da base de dados do VED. As subseções descrevem as etapas conduzidas.

#### 3.1. Entendendo o VED

O *Vehicle Energy Dataset*, de acordo com Oh et al. (2020), possui dados de trajetórias de 383 veículos (mais precisamente, 384, uma vez que 1 registro foi acrescentado, correspondendo ao veículo com o identificador “439”). Os atributos encontrados originalmente, sem nenhuma filtragem, são: *DayNum*, *VehId*, *Trip*, *Timestamp (ms)*, *Latitude [deg]*, *Longitude [deg]*, *Vehicle Speed[km/h]*, *MAF[g/sec]*, *Engine RPM[RPM]*, *Absolute Load[%]*, *OAT[DegC]*, *Fuel Rate[L/hr]*, *Air Conditioning Power[kW]*, *Air Conditioning Power[Watts]*, *Heater Power[Watts]*, *HV Battery Current[A]*, *HV Battery SOC[%]*, *HV Battery Voltage[V]*, *Short Term Fuel Trim Bank 1[%]*, *Short Term Fuel Trim Bank 2[%]*, *Long Term Fuel Trim Bank 1[%]* e *Long Term Fuel Trim Bank 2[%]*. Todas as especificações (tipo de dado e significado de cada atributo) constam nos arquivos estáticos do VED, no *github*, conforme referenciado anteriormente.

Na Figura 1, uma visão geral com alguns dos campos do VED, de um dos inúmeros arquivos em “csv”:

	DayNum	VehId	Trip	Timestamp(ms)	Latitude[deg]	Longitude[deg]	Vehicle Speed[km/h]	MAF[g/sec]	Engine RPM[RPM]	Absolute Load[%]	...
0	1.586651	8	706	0	42.277558	-83.698750	40.0	22.129999	2285.0	49.019608	...
1	1.586651	8	706	200	42.277558	-83.698750	40.0	22.129999	2285.0	67.450981	...
2	1.586651	8	706	1100	42.277558	-83.698750	45.0	22.129999	2285.0	67.450981	...
3	1.586651	8	706	2100	42.277558	-83.698750	47.0	6.150000	2744.0	67.450981	...
4	1.586651	8	706	4200	42.277558	-83.698750	48.0	21.440001	1982.0	67.450981	...
...	...	...	...	...	...	...	...	...	...	...	...
489409	7.992231	601	128	896900	42.288847	-83.753957	12.0	24.150000	2042.0	18.431374	...
489410	7.992231	601	128	897100	42.288847	-83.753957	18.0	24.150000	2042.0	18.431374	...
489411	7.992231	601	128	897900	42.288847	-83.753957	18.0	30.350000	2152.0	18.431374	...
489412	7.992231	601	128	898100	42.288847	-83.753957	24.0	30.350000	2152.0	18.431374	...
489413	7.992231	601	128	899500	42.288847	-83.753957	29.0	30.350000	2152.0	18.431374	...

489414 rows x 22 columns

**Figura 1. Visão geral do conteúdo do VED**

Conforme o artigo de Oh et al. (2020), os dados cobrem 374000 milhas e o trabalho se destaca pelo cuidado com a privacidade dos participantes (tanto que a identificação é por veículo, genérica, e não por usuário), pelo enfoque dado ao trânsito urbano e pela utilidade tanto para pesquisas sobre economia/consumo de energia de veículos urbanos até se configurar como um material útil para mineração e possível predição de trajetórias e destinos.

Os dados de contexto vêm como documento no formato “.xlsx” contendo informações como tipo de veículo, classe do veículo, configuração do motor, potência do motor, transmissão, drive de rodas e peso do veículo. Quanto aos dados principais, quando descompactados em valores separados por vírgula, por sua vez, possuem coordenadas geográficas (em graus decimais), o identificador de cada trajetória e o identificador de cada veículo, uma série de atributos sobre consumo de combustível, velocidade, temperatura e outros sobre consumo de energia, e os atributos de tempo. Como o artigo citado detalha melhor esses atributos, tem-se aqui a concentração nos atributos mais sensíveis: o de longitude, o de latitude, os fundamentais para o entendimento do fator “tempo”: *DayNum* e *Timestamp(ms)*. A partir deles, um conjunto de etapas com cada uma das bibliotecas *Python* como a *MovingPandas*<sup>2</sup> e a *Scikit-Mobility*<sup>3</sup>, a ser explicado na subseção 3.3.

No Quadro 1, há a descrição dos campos que foram considerados após filtragem, com a Ordem (um identificador), os nomes dos Atributos e a Descrição desses atributos. As demais colunas envolviam dados específicos de consumo de energia, não sendo úteis para este trabalho.

<sup>2</sup> *MovingPandas*. Disponível em: <https://github.com/movingpandas/>. Acesso em: 1 abr. 2023.

<sup>3</sup> *Scikit-mobility* — *scikit-mobility* documentation. Disponível em: <https://scikit-mobility.github.io/scikit-mobility/>. Acesso em: 1 abr. 2023.

Geograficamente, são trajetórias situadas na cidade de Ann Arbor, Michigan, Estados Unidos, captadas de novembro de 2017 a novembro de 2018, sendo relativamente recentes em termos de dados abertos dessa natureza.

**Quadro 1. Relação das colunas filtradas para este experimento**

Ordem	Atributos	Descrição
1	DayNum	Tempo em dias, fracionados
2	VehId	Identificador de cada veículo
3	Trip	Identificador de cada trajetória
4	Timestamp (ms)	Tempo em microssegundos
5	Latitude[deg]	Latitude em graus decimais
6	Longitude[deg]	Longitude em graus decimais
7	Vehicle Speed[km/h]	Velocidade em quilômetro por hora

Uma questão pertinente sobre os dados do VED é a modificação dos dados originais por necessidades de privacidade, mas que tornou muito nebulosa o tratamento em *Python* de um tipo de dados fundamental: o *datetime*. Nessa reestruturação, realiza-se uma reprodução sintética, aproximada, dos dados de data e hora - algo que neste trabalho também será mostrado através do compartilhamento de *scripts*.

### 3.2. Criação do novo conjunto de dados

A metodologia responde às Questões de Pesquisa: “*Como melhor tratar os dados do VED de forma a encontrar padrões de agrupamento espaço-temporal, harmonizando com a anonimização aplicada no conjunto de dados original?*”, e segue as seguintes etapas:

- a) Criação de um *datetime* sintético mediante o uso da função do Pandas “*to\_datetime()*” fixado em “D” e com um *timestamp* de origem e o uso do “*date\_range()*” com o número de Períodos igual a quantidade de registros;
- b) Obtenção das repetições Origem-Destino: primeiro foi feita a filtragem pelo identificador de cada veículo e concatenação em um único *DataFrame*, e a ordenação pelas colunas do identificador de viagem e *timestamp*. Em seguida, após a técnica de determinação de *datetime* sintético (pois o VED oculta o *datetime* original):

- b.i) Criação da malha de 100m a diagonal (a partir da junção dos limites de Ann Arbor com os de mais 4 localidades vizinhas, considerando o limite maior);
- b.ii) Obtenção das posições (registros) iniciais e finais de cada viagem;
- b.iii) Geração da tabela com os pontos iniciais, finais e o identificador de cada viagem;
- b.iv) Geração da tabela com os *grids* de origem e de destino;
- b.v) Somatório de cada repetição dentro de cada *grid* de cada trajetória, verificando se a soma das repetições de cada veículo é realmente igual ao total de trajetórias.

c) Preparação com ST-DBSCAN e comparação:

- c.i) Adequação à coleção de trajetórias do *MovingPandas*;
- c.ii) Exploração dos dados a partir de um *DataFrame* dessa biblioteca;
- c.iii) Geração de dados no formato *Geopackage*, de trajetórias em polilinhas e de pontos de parada para posterior uso na visualização dos dados com o *software* QGIS. Parte-se, assim, para uma verificação dos *N* veículos que possuem ao menos *x* trajetórias com ao menos *y* repetições, para resultar dados de pontos de parada que permitam uma futura predição de trajetórias com mais acurácia.
- c.iv) Utilização do ST-DBSCAN do QGIS com os parâmetros de 100 metros o limite espacial, vinte minutos o limite temporal, e um mínimo de cinco pontos por grupo. São 100 metros devido ao tamanho das grades, cuja granularidade foi previamente testada em função da qualidade dos resultados; quanto aos vinte minutos e aos cinco pontos, foram parâmetros escolhidos em função da redução dos dados resultantes, em termos de processamento computacional;
- c.v) No *Scikit-Mobility*, para a modelagem via agrupamento (a partir dos pontos de parada) com visualização, de uma agenda com características comuns entre diferentes horários, para o veículo com maior número de repetições de pontos de origem e de destino, após as filtragens. É feito o renomeamento do “CLUSTER\_ID” para “cluster”, fazendo com que o algoritmo de agrupamentos do Scikit-Mobility entenda como se fosse o seu próprio DBSCAN (a partir dos pontos de parada). As agendas (os gráficos), ao final, são visualmente comparadas. Uma observação final é que o *user* sempre foi o identificador de cada viagem, e não o identificador do veículo.

#### 4. Exemplificação dos resultados e Discussão

Os resultados se condensam a partir da Figura 2, contendo, como diferencial, uma coluna com esses novos dados (sintéticos):

```
In [7]: df['datetime'] = pd.to_datetime(df['datetime']).round('s')
df.datetime

Out[7]: 293386 2017-11-05 18:15:56
293387 2017-11-05 18:17:54
293388 2017-11-05 18:19:51
293389 2017-11-05 18:21:49
293390 2017-11-05 18:23:47
...
38575 2018-11-08 17:33:08
38576 2018-11-08 17:35:06
38577 2018-11-08 17:37:04
38578 2018-11-08 17:39:01
38579 2018-11-08 17:40:59
Name: datetime, Length: 270330, dtype: datetime64[ns]
```

**Figura 2. Geração da coluna *datetime***

Com essa nova informação, estão sendo considerados no mínimo uma trajetória com, no mínimo, duas repetições. Mais precisamente: o caso de haver ao menos duas repetições em ao menos uma trajetória gerou 231 veículos do total. A coluna “duplicados” corresponde ao número de trajetórias com repetições - presente na Figura 3. As etapas foram: a partir de um agrupamento de agrupamento do somatório de “VehId”, “grid\_oridem” e “grid\_destino”, gerou-se a coluna “repetições”; com ela, utilizou-se um FOR para a eliminação dos resultados com a coluna “repetições” menor ou igual a dois; e, em seguida, o que aparece na Figura 3, em que são filtrados apenas os “duplicados” maiores ou iguais a um.

```
[69] 1 table[table['duplicados'] >= 1]
✓ 0.0s
...


| duplicados |     |
|------------|-----|
| VehId_x    |     |
| 5          | 1   |
| 8          | 2   |
| 10         | 9   |
| 11         | 5   |
| 12         | 7   |
| ...        | ... |
| 607        | 2   |
| 608        | 5   |
| 609        | 5   |
| 610        | 6   |
| 624        | 2   |


231 rows x 1 columns
```

**Figura 3. Trecho do resultado que consolida os 231 veículos com ao menos uma trajetória com duas repetições**

A Figura 4, um histograma com a contagem de duplicados, gerando 30 duplicados com até mais de 80 repetições. Isso significa 30 valores únicos de trajetórias duplicadas dos 231 veículos por até 80 rotas (repetições, gerando itinerários únicos). Por exemplo: há cinco trajetórias que se repetem mais de 40 vezes.



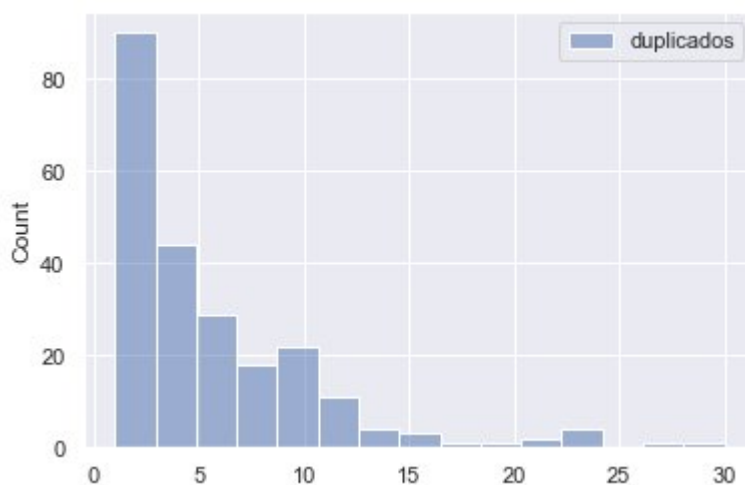


Figura 4. Histograma do resultado

Por fim, na Figura 5, o resultado da modelagem com ST-DBSCAN somado ao DBSCAN para esse veículo aleatório “276”.

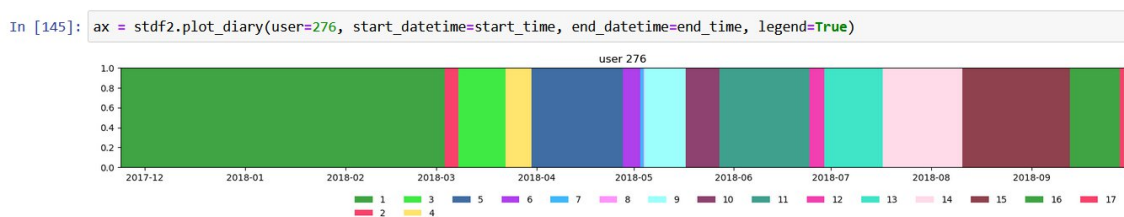


Figura 5. Modelagem com ST-DBSCAN e DBSCAN

Dessa forma, pode-se discutir: o que efetivamente foi ao final produzido com o VED corresponde a um novo campo (o *datetime* utilizado na Figura 2), um conjunto de procedimentos para a obtenção de veículos com ao menos uma trajetória, cada uma delas com, pelo menos, duas repetições (úteis para, com os pontos de parada, obterem-se predições de trajetórias) e o próprio resultado do agrupamento espaço-temporal, com 17 grupos. Os parâmetros foram escolhidos combinando o ambiente do QGIS com o do DBSCAN “tradicional” (esse último com um mínimo de uma amostra e considerando 100 m de raio do agrupamento).

## 5. Conclusão

Este trabalho reuniu um conjunto de testes com os dados, de forma a manter a anonimização e a demonstrar as possibilidades de uma pesquisa capaz de manipular dados com o intuito de, futuramente, gerar predições de trajetórias com mais acurácia.

Ainda há, porém, muito a ser feito. Embora a base de dados do VED apresente viabilidade na avaliação de modelos na área de predição de trajetórias, ele também pode ser utilizado em investigações científicas que abordem a questão da anonimização dos dados – anonimização esta mantida neste trabalho.

Dessa forma, o trabalho trouxe formas de utilização dos dados do VED sem quebrar a privacidade dos usuários, mesmo não tendo trazido nenhuma estratégia nova de anonimização. Porém, trouxe mais informações, a partir das trajetórias com repetições: o quadro consolidado foi o da Figura 3, que pôde derivar gráficos como o da Figura 4.

Porém a própria anonimização traz desafios na ponderação entre tornar os dados visíveis para pesquisas ao mesmo tempo que não descaracterizar o anonimato. Um exemplo disso foi durante a necessidade de criar o *datetime* a partir dos atributos temporais existentes, conforme salientado anteriormente, e outro pode surgir em futuras pesquisas envolvendo dados semânticos – além, evidentemente, de ser necessário ter o cuidado de não revelar determinados destinos ou não descaracterizar o processo de anonimização realizado pelos pesquisadores que elaboraram o VED.

Todavia, certas Questões de Pesquisa podem ser abordadas com esses dados, como a predição de acidentes de trânsito para posterior melhoramento de vias e sinais de trânsito; planejamento de destinos e de rotas; e até predições para inferior o aumento ou diminuição de casos de determinada doença infecto-contagiosa em setores de um município – nesse último caso, desde que acrescentados atributos relacionados à doença e atendimentos em unidades de saúde próximas.

## Referências

- Atluri, G.; Karpatne, A.; Kumar, V. (2018). Spatio-Temporal Data Mining. In: ACM Computing Surveys, v. 51, n. 4, p. 1–41, 22 ago. 2018.
- Birant, D.; Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. In: Data & Knowledge Engineering, v. 60, n. 1, p. 208–221, jan. 2007.
- Feng, Z.; Zhu, Y. (2016). A Survey on Trajectory Data Mining: Techniques and Applications. In: IEEE Access, v. 4, p. 2056–2067, 2016.
- Farzanehfar, A.; Houssiau, F.; De Montjoye, Y.-A. The risk of re-identification remains high even in country-scale location datasets. In: Patterns, v. 2, n. 3, p. 100204, mar. 2021.
- Gold, C. Tessellations in GIS: Part I—putting it all together. In: Geo-spatial Information Science, v. 19, n. 1, p. 9–25, 2 jan. 2016.
- Leite da Silva, C.; May Petry, L.; Bogorny, V. In: 2019 8TH Brazilian Conference on Intelligent Systems (BRACIS). *A survey and comparison of trajectory classification methods*. (Anais...). Salvador, Brazil: IEEE, 2019. p.788–793. DOI: <https://doi.org/10.1109/BRACIS.2019.00141>.
- Li, K.; Rao, X.; Pang, X.; Chen, L.; Fan, S. *Route search and planning: A survey*. Big data research, v. 26, n. 100246, p. 100246, 2021. DOI: <https://doi.org/10.1016/j.bdr.2021.100246>.
- Marketos, G. (2009). Mobility Data Warehousing and Mining. In: Proceedings of the VLDB 2009 PhD Workshop. Co-located with the 35th International Conference on Very Large Data Bases (VLDB 2009). Lyon, France, August 24, 2009.
- Oh, G.; Leblanc, D. J.; Peng, H. (2020). Vehicle Energy Dataset (VED), A Large-Scale Dataset for Vehicle Energy Consumption Research. In: IEEE Transactions on Intelligent Transportation Systems, p. 1–11, 2020.

OH, G. S. VED (Vehicle Energy Dataset). Disponível em: <https://github.com/gsoh/VED>. Acesso em: 1 abr. 2023.

Yu, L.; Zhang, Z.; Ding, R. *Map-Matching on Low Sampling Rate Trajectories through Frequent Pattern Mining*. Scientific Programming, v. 2022, p. 1–15, 21 mar. 2022.

Zheng, Y. (2015). Trajectory Data Mining. In: ACM Transactions on Intelligent Systems and Technology, v. 6, n. 3, p. 1–41, 12 maio. 2015.