

## Identificación de patrones de congestión vehicular utilizando algoritmos de agrupamiento de trayectorias basados en densidad

### Identification of vehicular traffic patterns using density-based trajectory clustering algorithms

Reyes, Gary; Roldán, José; Macias, Angélica; Cordova, Francisco; León, Oscar

**Gary Reyes** gary.reyesz@ug.edu.ec  
Universidad de Guayaquil, Ecuador  
**José Roldán** jose.roldanv@ug.edu.ec  
Universidad de Guayaquil, Ecuador  
**Angélica Macias** angelica.maciasu@ug.edu.ec  
Universidad de Guayaquil, Ecuador  
**Francisco Cordova** francisco.cordovar@ug.edu.ec  
Universidad de Guayaquil, Ecuador  
**Oscar León** oscar.leong@ug.edu.ec  
Universidad de Guayaquil, Ecuador

**Investigación, Tecnología e Innovación**  
Universidad de Guayaquil, Ecuador  
ISSN: 1390-5147  
ISSN-e: 2661-6548  
Periodicidad: Cuatrimestral  
vol. 14, núm. 17, 2022  
revistaiti@ug.edu.ec

Recepción: 10 Julio 2022  
Aprobación: 13 Noviembre 2022

URL: <http://portal.amelica.org/ameli/journal/593/5933639001/>

DOI: <https://doi.org/10.53591/iti.v14i17.1473>

Los autores que publican en Investigación, Tecnología e Innovación conocen y aceptan las siguientes condiciones: Los autores retienen los derechos de copia (copyright) sobre los trabajos, y ceden a Investigación, Tecnología e Innovación el derecho de la primera publicación del trabajo, bajo licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0 que permite a terceros compartir la obra siempre que se indique su autor y su primera publicación esta revista. Los autores conservan los derechos de autor y garantizan a Investigación, Tecnología e Innovación el derecho de publicar el trabajo a través de los canales que considere adecuados. Los autores son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la versión del trabajo publicado en Investigación, Tecnología e Innovación, haciendo reconocimiento a su publicación en esta revista. Se autoriza a los autores a difundir electrónicamente sus trabajos una vez que sean aceptados para publicación.

**Resumen:** **Contexto:** Este trabajo se centra en identificar patrones de congestión. Para ello se utilizan 3 conjuntos de datos de las ciudades de Beijing, Guayaquil y Roma. **Método:** Se efectúa la implementación del algoritmo Dyclee, modificado para agrupar celdas de trayectorias en base a velocidades, con el cual se realiza experimentos en los que se calculó adecuadamente los patrones de volumen de servicio e índice de operatividad en base a sus resultados. La modalidad de la investigación predominante es la bibliográfica. Sin embargo, el estudio también incluye características de una investigación de campo, debido a que los algoritmos fueron ejecutados con datos de trayectorias GPS de tres conjuntos de datos diferentes. **Resultados:** Para validar esta investigación se ejecutaron dos experimentos. En el primer experimento se determinó que el algoritmo Dyclee calculó correctamente los patrones de congestión. En el segundo experimento se encontró que TRADBSCAN obtuvo los mejores resultados con respecto a métricas de validación tomando en cuenta los parámetros establecidos. **Conclusiones:** Se concluyó que el algoritmo de clustering basado en densidad Dyclee, es capaz de identificar patrones de congestión vehicular.

**Palabras clave:** Congestionamiento, Clustering, Trayectorias, Tráfico, Densidad.

**Abstract:** **Context:** This paper focuses on identifying congestion patterns. Three data sets from the cities of Beijing, Guayaquil and Rome are used for this purpose. **Method:** The implementation of the Dyclee algorithm, modified to group cells of trajectories based on speeds, is carried out, with which experiments are performed in which the patterns of service volume and operability index are adequately calculated based on their results. The predominant research modality is bibliographic. However, the study also includes characteristics of a field research, because the algorithms were executed with GPS trajectory data from three different data sets. **Results:** To validate this research, two experiments were run. In the first experiment it was determined that the Dyclee algorithm correctly calculated the congestion patterns. In the second experiment it was found that TRADBSCAN obtained the best results with respect to validation metrics taking into account the established parameters. **Conclusions:** It was concluded that



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.

Como citar:: Reyes, G., Roldán, J., Macias, A., Cordova, F., & León, O. (2022). Identificación de patrones de congestión vehicular utilizando algoritmos de agrupamiento de trayectorias basados en densidad. *Investigación, Tecnología E Innovación*, 14(17), 1–15. <https://doi.org/10.53591/iti.v14i17.1473>

the Dylee density-based clustering algorithm is able to identify vehicular congestion patterns.

**Keywords:** Congestion, Clustering, Trajectories, Traffic, Density.

## INTRODUCCIÓN

El congestionamiento es un problema creciente en muchas ciudades debido a la creciente demanda de movilización en las vías, algo relacionado al aumento poblacional en todo el mundo. Éste es un factor que influye en la vida cotidiana de las personas debido al estrés que conlleva estar en una vía congestionada, además del consumo aumentado de gasolina, algo que además de afectar a la economía, afecta al medio ambiente, debido a los gases emitidos por los vehículos.

Las ciudades y países abordan el congestionamiento de maneras diferentes, las cuales, por lo general, involucran realizar estudios de campo, con el equipamiento necesario para medir el congestionamiento en una vía. Dichos estudios pueden llegar a ser costosos, y los resultados de estos son por lo general confidenciales. Si existiera una metodología, con la cual, partiendo de esos datos, facilitaría el análisis e identificación de los niveles de congestión, podría atraer beneficios considerables a los encargados de dicho análisis, así como a la comunidad científica, algo que es una de las principales motivaciones para realizar este proyecto.

Como fue establecido por (Reyes, Lanzarini, Hasperué, et al., 2021), el análisis de trayectorias vehiculares con el objetivo de generar clústeres e identificar los patrones que emergen contribuyen considerablemente al análisis y caracterización del flujo de tránsito en diferentes partes del mundo. Por lo que el presente proyecto busca, mediante el análisis de trayectorias utilizando algoritmos de clustering basados en densidad, lograr la identificación de patrones, también conocidos como indicadores, de congestión. La investigación consistirá principalmente en una investigación bibliográfica con la cual se tendrá una base teórica, para realizar una selección de algoritmos, métricas y patrones a estudiar, las cuales serán adaptados para el procesamiento de trayectorias, y con las cuales se realizarán experimentaciones con datasets de repositorios públicos para conocer si se puede realizar la identificación de dichos patrones con estos algoritmos.

Como trabajos previos, se tiene el estudio de (Cedeño & Piña, 2021), el cual consistió en la implementación de un algoritmo de clustering dinámico en lenguaje Python, con las debidas modificaciones para procesar trayectorias GPS, estos encontraron que este algoritmo tuvo resultados favorables utilizando al coeficiente Silhouette como técnica para la validación de los resultados. El estudio presentado por (Arias & Zamora, 2020), el cual se basó en la implementación de un algoritmo de clustering dinámico de flujos de datos para identificar las evoluciones de patrones de movimiento, obtuvo buenos resultados. Por último se encuentra el estudio realizado por (Bastidas & Burgos, 2021), en el cual se

utilizaron los indicadores de congestión vehicular resultantes del análisis de agrupamientos de trayectorias con el fin de identificar los lugares en donde existe un mayor y menor flujo vehicular, en base a la agrupación de trayectorias en base a velocidades.

### *GPS*

De acuerdo con (Wang et al., 2020), el GPS consiste en un sistema de navegación con el uso de ondas de radio en las cuales, se puede proporcionar datos georreferenciales y de tiempo a un receptor GPS en todo el planeta con el uso de satélites.

### *Trayectoria GPS*

De acuerdo con (Reyes Zambrano, 2019) una trayectoria GPS consiste en una secuencia de puntos de coordenadas geográficas discretas. Lo cual da a entender que una trayectoria es un conjunto de puntos de coordenadas los cuales tienen algunas

### *Congestión Vehicular*

En base a lo planteado (Thomson & Bull, 2002), el congestionamiento vehicular consiste en una situación que prevalece cuando en un flujo de tránsito, se incrementa el tiempo de circulación de los vehículos en este debido al ingreso de un nuevo automóvil en dicho flujo.

### *Algoritmos de Clustering*

Según (Reyes et al., 2020), un algoritmo de clustering busca crear grupos con elementos que poseen características compartidas, ya sea en base a un centroide, del cual se pueden encontrar una cierta cantidad de puntos a cierta medida de distancia, con la cual se encuentra la similitud entre objetos; así como en base a un descriptor, que permitirá identificar las características únicas de cada grupo.

### *Dycee*

Según Barbosa Roa et al. (2019) Dycee es: “un novedoso algoritmo de clustering en-línea para transmitir datos en base a una etapa basada en distancia y otra etapa basada en densidad, las cuales pueden ejecutarse paralelamente y a una frecuencia mayor y menor, respectivamente.”

Dycee utiliza un aprendizaje incremental, no supervisado, en el cual, este no realiza una suposición sobre la estructura de datos que está analizando, sino que la descubre progresivamente, como el aprendizaje humano, cambiando la estructura del agrupador y los conceptos para describir los datos que se recibieron. (Barbosa Roa et al., 2019)

La primera etapa opera a la tasa del flujo de datos y crea micro clústeres juntando muestras de datos que se encuentran cerca en base a la distancia de Manhattan; mientras que la segunda etapa toma lugar en una frecuencia más baja y analiza la distribución de los micro clústeres, la densidad del micro clúster puede ser baja, media o alta, y permite crear los clústeres finales con un enfoque basado en densidad; definiendo un clúster como una serie de micro clústeres conectados en donde cada uno presenta una alta densidad y cada micro clúster borde presenta una densidad media o alta. Estas etapas funcionan de manera paralela e intercambian información para mejorar el rendimiento mutuamente; lo cual ha permitido que tenga un alto rendimiento en entornos envolventes de altos espacios dimensionales y buenas capacidades de eliminación de ruido. (Barbosa Roa et al., 2019)

### *TRADBSCAN*

TRADBSCAN es un entorno de trabajo basado en DBSCAN que es un algoritmo de agrupación especial basada en densidad. En este entorno de trabajo cada trayectoria es dividida en sub trayectorias y luego en sub trayectorias de agrupamiento. Luego, las sub trayectorias son agrupadas por DBSCAN de acuerdo a su tiempo de separación, su patrón de trayectoria es definida como la conexión de sub trayectorias agrupadas durante todo el tiempo de separación (Liu et al., 2011).

Liu et al. (2011) explica que la implementación de TRADBSCAN empieza con un conjunto de trayectorias, las cuales son particionadas en segmentos de líneas, y luego, se le aplica el algoritmo DBSCAN para agrupar los segmentos de línea que compartan una marca de tiempo; luego se forman patrones de grupos de trayectorias al conectar los clústeres que tengan la mayor cantidad de elementos parecidos y líneas de tiempo continuas. Liu et al. Recalca que la razón de utilizar DBSCAN es que es capaz de encontrar grupos irregulares, su forma y número de incertidumbre, no sensible al ruido.

### *Coefficiente Silhouette*

El ancho silhouette se calcula como el valor promedio de silhouette para cada punto. Este valor mide el nivel de confianza en una particular asignación de clúster para cada elemento individual; y es calculado dividiendo la diferencia entre la distancia promedio entre cada punto de dato a todos los otros puntos de su clúster y la distancia promedio entre dicho punto de dato a todos los otros puntos del clúster vecino más cercano, este último tiende hacia la distancia mínima en la que se puede usar una medida de distancia como euclidiana o de manhattan como medida de disimilitud (Scherl, 2010)

El rango del coeficiente silhouette abarca desde el -1 hasta el 1, con lo que, mientras más cercano sea al 1, el punto está asignado a un clúster correcto; si está más cerca de 0, el dato puede ser asignado a otro clúster más cercano dado es equidistante a ambos clústeres cercanos; y si está más cerca de -1, el dato está clasificado de forma errónea porque se encuentra en algún lugar entre los dos clústeres; cada uno de los coeficientes de los puntos se puede promediar

para encontrar la calidad de agrupamiento del cluster en general. (Ansari et al., 2015) En sí, mientras más alto el coeficiente silhouette mejor es la calidad del agrupamiento realizado.

#### *Coeficiente Davies Bouldin*

El coeficiente Davies Bouldin busca calcular una medida de similitud entre dos clústeres, un clúster, y el más similar a este, mediante el cálculo de la razón entre la dispersión de los clústeres analizados, y la medida de disimilitud entre los dos clústeres, algo que tiene que cumplir ciertas condiciones, como que dicha razón sea mayor o igual a 0, es decir, que no sea un número negativo; se busca minimizar el promedio del valor de similitud calculado para cada clúster y su más similar, es decir, la similitud mínima, ya que esta representa una buena separación entre clústeres. (Scherl, 2010)

#### *Volumen o Tasa de Servicio*

Según el (Transportation Research Board, 2000), el volumen o tasa de servicio es: “la tasa equivalente horaria en la cual vehículos, bicicletas, o personas, pasan un punto en una vía. Es calculada como el número de dichos elementos que pasan el punto, dividido para el intervalo de tiempo que tardan en pasar.” (p. 53)

Para el cálculo del volumen o tasa de servicio del período el (Transportation Research Board, 2000) plantea la ecuación (1)

$v_i$ : es el volumen o tasa de servicio (veh/h) del período  $i$  de la hora pico  
 $v$ : es el volumen de la hora pico (veh/h)  
 $F_i$ : es el factor del período  $i$  de la hora pico.

#### *Índice de operatividad*

De acuerdo con el (Transportation Research Board, 2000), el índice de operatividad, es: “la relación de la tasa o volumen de servicio con la capacidad; también es conocido como grado de saturación” (p. 363)

Para el cálculo del índice de operatividad para cada período, el (Transportation Research Board, 2000) plantea la ecuación detallada en la ecuación (2)

**Índice de Operatividad de periodo  $i = \frac{v_i}{c}$**   
 $v_i$  : volumen o tasa de servicio (veh/h) del período  $i$   
 $c$  : capacidad de la vía.

## MATERIALES Y MÉTODOS

La investigación realizó dos experimentos, el primero, determina si la identificación de patrones de congestión es posible con el agrupamiento realizado con el algoritmo de Dyclee, se medirán características y variables

estadísticas de los grupos generados, con lo cual se procederá a medir la estimación de los patrones utilizando la medida de tendencia central de la media aritmética, que pretenderá obtener una vista general del nivel de demanda y saturación de las vías de los datasets seleccionados. El segundo experimento permitirá validar los resultados de los agrupamientos de Dylee y TRADBSCAN con las métricas seleccionadas, Silhouette y Davies-Bouldin además de comparar la calidad de los agrupamientos por velocidad de ambos algoritmos.

Para realizar los experimentos, se necesita realizar la carga y el procesamiento de los datos de trayectorias GPS de los conjuntos de datos seleccionados (Beijing, Guayaquil y Roma), por lo que se va a utilizar el Sistema Gestor de Base de Datos de PostgreSQL, con el fin de realizar los experimentos planteados utilizando los algoritmos, los cuales fueron implementados previamente en R, Dylee, modificado para procesar trayectorias, por (Reyes, Lanzarini, Estrebou, et al., 2021) y TRADBSCAN por (Hasperué et al., 2021). Se busca que estos datasets tengan por lo menos cierta cantidad de datos, la latitud y la longitud, para el agrupamiento de trayectorias por celdas de Dylee, y una marca de tiempo, la cual es indispensable para el desarrollo de estos experimentos, ya que estos permiten realizar el cálculo de las velocidades que toman las trayectorias de los diferentes autos, algo fundamental para la agrupación mediante las velocidades. Además, estas marcas de tiempo se utilizan para dividir las trayectorias en intervalos de tiempo de 15 minutos, debido a que los patrones se calculan con diferentes datos de que fluctúan a través de los períodos, algo que permitirá que el algoritmo de Dylee calcule los patrones en base a diferentes períodos de los datasets. Para lo cual, se deberá seleccionar la hora de volumen pico, conocida también como hora de mayor demanda, la cual se identificará en los diferentes datasets y se utilizará como muestra de estos; debido a su uso en los cálculos de los patrones seleccionados.

### *Población*

La población del estudio consistirá en un conjunto de puntos o registros GPS, que van a ser el objeto de análisis, provenientes de los 3 datasets de repositorios públicos de Beijing, Guayaquil y Roma. En la Tabla 1 se pueden apreciar la cantidad total de registros o puntos en los datasets escogidos.

**Tabla 1**

Cantidad de registros de los datasets seleccionados.

Población	Cantidad de registros
Beijing	62138
Guayaquil	30557
Roma	708097

Fuente: Autores

Autores



### *Muestra*

Para el análisis de los patrones de congestión, es recomendado el uso de la hora de mayor demanda de un conjunto de datos, para calcular con mayor exactitud dichos patrones. Por lo que, se procedió a calcular el total de la hora de volumen pico de los 3 datasets, identificando primero el año con mayor volumen, luego el mes en ese año, seguido por el día en dicho mes, y terminando con la hora en tal día, con lo que se obtuvieron los resultados de la Tabla 2. En esta se puede apreciar, que para la hora de volumen pico, el cual fue identificado como el período entre las 2 y 3 de la tarde el 02 de febrero del 2008 en el dataset de Beijing se obtuvieron 38144 datos, un 61,39% comparado con el valor original. Mientras que, en Guayaquil, el cual abarcó desde las 5 hasta las 6 de la tarde del día 17 de octubre del 2017, en el cual se obtuvo una muestra de 25495 datos, un 83,43% del total de datos original. Por último, en Roma, se obtuvo que la hora de volumen pico, que consistió en el período desde las 1 y 2 de la tarde del 12 de febrero del 2014, el cual abarcó 35729 registros, lo cual representa un 6.67% con respecto al total original.

**Tabla 2**  
Cantidad de registros de la hora de volumen pico

Periodo	Cantidad registros general	Cantidad registros en hora pico	Porcentaje
Beijing	62138	38144	61,39%
Guayaquil	30557	25495	83,43%
Roma	535949	35729	6,67%

Fuente: Autores

Autores

## RESULTADOS

En esta sección se detallan los resultados de los dos experimentos realizados sobre los 3 datasets escogidos, el primer experimento abarcando los resultados de los agrupamientos de Dylee, la identificación de patrones de congestión de Dylee, y los resultados de la herramienta visual; y el segundo experimento, abarcando los resultados de la comparación de las métricas calculadas en base a los agrupamientos realizados por Dylee y por DBSCAN.

### *Primera experimentación*

En la Tabla 3 se encuentran los resultados de los experimentos realizados sobre el dataset de Beijing a lo largo de los 4 períodos, agrupados por promedio. En la primera columna se tiene la cantidad de puntos promedio del período, algo que permite saber cuáles fueron los períodos con mayor cantidad de registros, en este caso, este fue el período 2, con 1894 y el de menor cantidad, el 4, con 1385 puntos promedios. La velocidad mínima promedio más alta, 21,13 km/h, le pertenece al período 4, mientras que la más baja, 16,12 km/h, al período 3. La velocidad

máxima promedio más alta se encuentra en el período 1, con un valor de 36,22 km/h y la más baja, en el período 3, con un valor de 31,13 km/h. La velocidad promedio más alta de los períodos promediados, es de 27,92 km/h, perteneciente al período 1, y la más baja, la del período 3, es de 22,78 km/h. El rango de velocidad promedio de cada período fue el más bajo, con 12,91 km/h en el período 4, y el más alto en el período 2, con 16,63 km/h. Por último, la desviación estándar de velocidad más alta de los períodos fue calculada en el período 1, es decir, 4,34 y la más baja fue calculada en el período 4, con 3,34.

**Tabla 3**

Resultados del agrupamiento del algoritmo Dyclee para los cuatro períodos del dataset de Beijing

Grupo	Cantidad de Puntos	Velocidad Mínima	Velocidad Máxima	Velocidad Promedio	Rango de Velocidad	Desviación Estándar de Velocidad
Promedio del Período 1	1865,00	20,88	36,22	27,92	15,35	4,34
Promedio del Período 2	1894,00	19,36	35,99	26,97	16,63	4,19
Promedio del Período 3	1744,00	16,12	31,13	22,78	15,01	3,73
Promedio del Período 4	1385,00	21,13	34,04	26,51	12,91	3,34
Promedio General	1722,00	19,37	34,34	26,04	14,97	3,90

Fuente: Autores

Autores

En la Tabla 4 se aprecian los resultados de los promedios de patrones de congestión por período en el dataset de Beijing. En la primera columna, se denota el período específico del cual se encontraron los promedios de indicadores, obteniendo 4 períodos en total. La segunda columna, el total de celdas que conforman dicho período, obteniendo un promedio entre los 4 períodos de 2966 celdas. La tercera columna, el volumen de servicio, se obtuvo un promedio entre los períodos de 436,35 vehículos por hora; algo que denota, que hay un bajo volumen de autos circulando en las calles a lo largo de los períodos, siendo el cuarto período el de menor demanda y el primero el de mayor demanda. Por último, la columna de índice de operatividad muestra un promedio de 0,18 entre los 4 períodos, lo que indica que las vías no están saturadas en los respectivos períodos, teniendo al cuarto período con menor índice de operatividad y al primero con el mayor índice.



**Tabla 4**

Resultados de los patrones de congestamiento vehicular promediados por periodo para Beijing

Periodo	Total de Celdas	Volumen de servicio	Índice de operatividad
1	3075	441,23968	0,1838645
2	3055	438,13056	0,1825684
3	2822	438,85168	0,1828693
4	2910	427,19489	0,1780128
Promedios	2966	436,3542	0,1818287

Fuente: Autores

Autores

En la Tabla 5 se encuentran los resultados de los experimentos realizados sobre el dataset de Roma a lo largo de los 4 períodos, agrupados por promedio. En la primera columna se tiene la cantidad de puntos promedio del período, algo que permite saber cuáles fueron los períodos con mayor cantidad de registros, en este caso, este fue el período 2, con 1265 y el de menor cantidad, el 4, con 1184 puntos promedios. La velocidad mínima promedio más alta, 20,53 km/h, le pertenece al período 3, mientras que la más baja, 16,25 km/h, al período 4. La velocidad máxima promedio más alta se encuentra en el período 3, con un valor de 35,38 km/h y la más baja, en el período 4, con un valor de 30,35 km/h. La velocidad promedio más alta de los períodos promediados, es de 27,05 km/h, perteneciente al período 3, y la más baja, la del período 1, es de 23,01 km/h. El rango de velocidad promedio de cada período fue el más bajo, con 12,83 km/h en el período 2, y el más alto en el período 3, con 14,85 km/h. Por último, la desviación estándar de velocidad más alta de los períodos fue calculada en el período 3, es decir, 4,12 y la más baja fue calculada en el período 1, con 3,58.

**Tabla 5**

Resultados del agrupamiento del algoritmo Dyclee para los cuatro períodos del dataset de Roma

Grupo	Cantidad de Puntos	Velocidad Mínima	Velocidad Máxima	Velocidad Promedio	Rango de Velocidad	Desviación Estándar de Velocidad
Promedio del Periodo 1	1202,00	16,78	31,00	23,01	14,22	3,58
Promedio del Periodo 2	1265,00	17,93	30,75	23,87	12,83	3,66
Promedio del Periodo 3	1205,00	20,53	35,38	27,05	14,85	4,12
Promedio del Periodo 4	1184,00	16,25	30,35	23,44	14,10	3,63
Promedio General	1214,00	17,87	31,87	24,34	14,00	3,75

Fuente: Autores

Autores

En la Tabla 6 se aprecian los resultados de los promedios de patrones de congestionamiento por período en el dataset de Roma. En la primera columna, se denota el período específico del cual se encontraron los promedios de indicadores, obteniendo 4 períodos en total. La segunda columna, el total de celdas que conforman dicho período, obteniendo un promedio entre los 4 períodos de 1184 celdas. La tercera columna, el volumen de servicio, se obtuvo un promedio entre los períodos de 613,17 vehículos por hora; algo que denota, que hay un relativamente bajo volumen de autos circulando en las calles a lo largo de los períodos, siendo el cuarto período el de menor demanda y el segundo el de mayor demanda. Por último, la columna de índice de operatividad muestra un promedio de 0,26 entre los 4 períodos, lo que indica que las vías no están muy saturadas en los respectivos períodos, teniendo al cuarto período con menor índice de operatividad y al segundo con el mayor índice.

**Tabla 6**

Resultados de los patrones de congestionamiento vehicular promediados por periodo para Roma

Periodo	Total de Celdas	Volumen de servicio	Índice de operatividad
1	1279	601,10734	0,2504674
2	1226	652,4	0,2718445
3	1178	612,5117	0,2552254
4	1051	586,67061	0,2444598
Promedios	1184	613,17241	0,2554993

Fuente: Autores

Autores

En la Tabla 7 se encuentran los resultados de los experimentos realizados sobre el dataset de Guayaquil a lo largo de los 4 períodos, agrupados por promedio. En la primera columna se tiene la cantidad de puntos promedio del período, algo que permite saber cuáles fueron los períodos con mayor cantidad de registros, en este caso, este fue el período 1, con 1281 y el de menor cantidad, el 4, con 687 puntos promedios. La velocidad mínima promedio más alta, 32,81 km/h, le pertenece al período 1, mientras que la más baja, 17,72 km/h, al período 4. La velocidad máxima promedio más alta se encuentra en el período 1, con un valor de 49,57 km/h y la más baja, en el período 4, con un valor de 32 km/h. La velocidad promedio más alta de los períodos promediados, es de 41,45 km/h, perteneciente al período 1, y la más baja, la del período 4, es de 24,56 km/h. El rango de velocidad promedio de cada período fue el más bajo, con 14,28 km/h en el período 4, y el más alto en el período 1, con 16,76 km/h. Por último, la desviación estándar de velocidad más alta de los períodos fue calculada en el período 1, es decir, 4,69 y la más baja fue calculada en el período 4, con 3,76.

**Tabla 7**

Resultados del agrupamiento del algoritmo Dylee para los cuatro períodos del dataset de Guayaquil

Grupo	Cantidad de Puntos	Velocidad Mínima	Velocidad Máxima	Velocidad Promedio	Rango de Velocidad	Desviación Estándar de Velocidad
Promedio del Periodo 1	1281,00	32,81	49,57	41,45	16,76	4,69
Promedio del Periodo 2	1111,00	31,47	45,84	38,19	14,36	3,82
Promedio del Periodo 3	1269,00	31,23	47,06	38,66	15,83	4,20
Promedio del Periodo 4	687,00	17,72	32,00	24,56	14,28	3,76
Promedio General	1087,00	28,31	43,62	35,71	15,31	4,12

Fuente: Autores

Autores

En la Tabla 8 se aprecian los resultados de los promedios de patrones de congestamiento por período en el dataset de Guayaquil. En la primera columna, se denota el período específico del cual se encontraron los promedios de indicadores, obteniendo 4 períodos en total. La segunda columna, el total de celdas que conforman dicho período, obteniendo un promedio entre los 4 períodos de 297 puntos. La tercera columna, el volumen de servicio, se obtuvo un promedio entre los períodos de 1693,35 vehículos por hora; algo que denota, que hay un moderado volumen de autos circulando en las calles a lo largo de los períodos, siendo el cuarto período el de menor demanda y el primero el de mayor demanda. Por último, la columna de índice de operatividad muestra un promedio de 0,71 entre los 4 períodos, lo que indica que las vías están moderadamente saturadas en los respectivos períodos, teniendo al cuarto período con menor índice de operatividad y al primero con el mayor índice de operatividad.

**Tabla 8**

Resultados de los patrones de congestamiento vehicular promediados por periodo para Guayaquil

Periodo	Total de Celdas	Volumen de servicio	Índice de operatividad
1	323	2437,9355	1,0158052
2	369	1617,9787	0,6741612
3	342	1784,6012	0,7435902
4	153	935,32	0,389724
Promedios	297	1693,9588	0,7058201

Fuente: Autores

Autores

*Segunda experimentación*

En la Tabla 9 se detallan los resultados generales de la métrica de Davies-Bouldin promediados de cada dataset en base a los dos algoritmos. En esta, se puede observar, que en el dataset de Beijing, según la métrica, el algoritmo de TRADBSCAN obtuvo un mejor resultado, 0,345, a comparación de Dyclee con 0,372. En el dataset de Roma, se obtuvo que TRADBSCAN tuvo un coeficiente menor, 0,328, mientras que Dyclee, 0,378, denotando una mejor calidad de agrupamiento del primer algoritmo. En el dataset de Guayaquil, se obtuvo un resultado de 0,334 para TRADBSCAN y un 0,311, para Dyclee, siendo este último el que mejor resultado obtuvo. De forma general, se obtiene que TRADBSCAN tuvo un promedio de 0,336, y Dyclee uno de 0,354, siendo el mejor resultado el de TRADBSCAN, con lo que se denota que este tuvo una mejor calidad de agrupamientos.

**Tabla 9**

Resultados promediados de la métrica Davies-Bouldin para los agrupamientos de cada dataset

Dataset	TRADBSCAN	Dyclee
Beijing	0,343875226	0,37188027
Roma	0,328336911	0,377648683
Guayaquil	0,334501949	0,310935356
Promedios	0,335571362	0,353488103

Fuente: Autores

Autores

En la Tabla 10 se detallan los resultados generales de la métrica Silhouette promediados de cada dataset en base a los dos algoritmos. En esta, se puede observar que, en el dataset de Beijing, según la métrica, el algoritmo de TRADBSCAN obtuvo un mejor resultado, 0,574, a comparación de Dyclee con 0,543. En el dataset de Roma, se obtuvo que TRADBSCAN tuvo un coeficiente mayor, 0,701, mientras que Dyclee, 0,529, denotando una considerable mejor calidad de agrupamiento por parte del primer algoritmo. En el dataset de Guayaquil, se obtuvo un resultado de 0,619 para TRADBSCAN y un 0,614, para Dyclee, siendo el primero el que mejor resultado obtuvo. De forma general, se obtiene que TRADBSCAN tuvo un promedio de 0,632, y Dyclee uno de 0,562, siendo el mejor resultado el de TRADBSCAN, con lo que se denota que este tuvo una mejor calidad de agrupamientos.

**Tabla 10**

Resultados promediados de la métrica Silhouette para los agrupamientos de cada dataset

Dataset	TRADBSCAN	Dyclee
Beijing	0,574288748	0,542676102
Roma	0,701211347	0,529589026
Guayaquil	0,619930674	0,614259991
Promedios	0,631810256	0,56217504

Fuente: Autores

## DISCUSIÓN

Se encontró que, en base al primer experimento, el algoritmo de Dylee logra una buena calidad de agrupamiento a lo largo de los 4 períodos de los 3 datasets; algo que permite el correcto cálculo de los indicadores de congestión vehicular en base a celdas.

En base al segundo experimento, las métricas de validación de Davies-Bouldin y Silhouette permitieron determinar cuál de los dos algoritmos presentaba mejores resultados de agrupamiento, en ambos criterios, TRADBSCAN obtuvo mejores resultados que Dylee.

Una de las principales limitaciones encontradas en el estudio es la utilización de datasets antiguos, debido a que la mayoría de repositorios públicos poseen datos de la primera década del año 2000, dado el reciente auge de la tecnología GPS en los dispositivos móviles, lo cual impide que los resultados obtenidos sean representativos de la situación actual del tráfico en las ciudades que serán seleccionadas.

Otra limitación importante consiste en que los datasets cuentan con una cantidad de datos poco significativa, debido a que consisten en datos de trayectorias GPS registrados por una pequeña cantidad de taxis que participaron en estudios de movilidad, algo que impide que los datos utilizados sean una representación total del estado real del tráfico, sino un porcentaje de este, que no toma en cuenta el impacto que tienen otros tipos de vehículos, como los particulares, o el transporte público y otros vehículos pesados.

Por último, una limitación a considerar, es que debido a como los datos disponibles en los datasets son limitados, no contienen los datos para tomar en cuenta factores como signos de pare, semáforos, y otros tipos de señales de tránsito los cuales son necesarios para el análisis del flujo vehicular interrumpido, por lo que los patrones de congestión a identificar concretamente serán en base al flujo vehicular ininterrumpido.

## CONCLUSIONES

En el presente estudio se realizaron dos experimentos diferentes, el primero, con el fin de identificar los patrones de congestión con los resultados del agrupamiento de Dylee en los datasets de Beijing, Guayaquil y Roma; y el segundo, con el fin de validar los resultados de los agrupamientos de Dylee y TRADBSCAN, y encontrar cuál de los dos posee mejor calidad de agrupamiento para dichos datasets.

En el primer experimento, se encontró que hubo una satisfactoria agrupación de trayectorias en base a velocidades. Estas se generaban en orden de velocidades promedio más bajas a más altas y presentaban una desviación estándar baja, algo que denota una baja presencia de ruido. Además, tenían un conjunto determinado de rangos de velocidades máximos y mínimos. También se realizó el correcto cálculo de los patrones de congestión de las celdas generadas por el algoritmo, las cuales mostraron una consistencia en resultados, lo que permitió una adecuada identificación de la demanda de las vías, es decir, el volumen o tasa de servicio de estas, y de la saturación de estas, el índice de operatividad en las vías de los datasets.

En el segundo experimento, se encontró que hubo una correcta validación de los grupos producidos por los algoritmos de Dyclee y DBSCAN, ya que cumplían los criterios apropiados para la validación mediante la métrica de Davies-Bouldin, ambos estando cerca del 0, así como los de Silhouette, en el cual ambos algoritmos presentaron valores que se acercaban al 1. Sin embargo, de manera general, los resultados de agrupamiento de TRADBSCAN, tenían una mejor calidad de agrupamiento en base a las métricas, especialmente en el dataset de Roma y Beijing, y en menor medida, el de Guayaquil.

## REFERENCIAS BIBLIOGRÁFICAS

- Ansari, Z., Azeem, M. F., Ahmed, W., & Babu, A. (2015). Quantitative Evaluation of Performance and Validity Indices for Clustering the Web Navigational Sessions. *World of Computer Science and Information Technology Journal*, 1, 217–226. <https://arxiv.org/ftp/arxiv/papers/1507/1507.03340.pdf>
- Arias, B., & Zamora, B. (2020). *Algoritmo de Clustering dinámico para trayectoria GPS*. [http://repositorio.ug.edu.ec/bitstream/redug/52739/1/B-CISC-PTG-1940-2021\\_Arias\\_Martinez\\_Bryan\\_Jose\\_-\\_Zamora\\_Litardo\\_Bryan\\_Steven.pdf%0A](http://repositorio.ug.edu.ec/bitstream/redug/52739/1/B-CISC-PTG-1940-2021_Arias_Martinez_Bryan_Jose_-_Zamora_Litardo_Bryan_Steven.pdf%0A)
- Barbosa Roa, N., Travé-Massuyès, L., & Grisales-Palacio, V. H. (2019). DyClee: Dynamic clustering for tracking evolving environments. *Pattern Recognition*, 94, 162–186. <https://doi.org/https://doi.org/10.1016/j.patcog.2019.05.024>
- Bastidas, M., & Burgos, D. (2021). *Identificación de congestamiento vehicular a través del análisis de agrupamiento de trayectorias GPS*. <http://repositorio.ug.edu.ec/handle/redug/57168%0A>
- Cedeño, P., & Piña, Á. (2021). *Adaptación del algoritmo de clustering dinámico Pyclee para el procesamiento y análisis de trayectorias GPS*. [http://repositorio.ug.edu.ec/bitstream/redug/57094/1/B-CISC-PTG-1996-2021\\_Cedeño\\_Núñez\\_Pedro\\_Xavier\\_-\\_Piña\\_Naranjo\\_Ángel\\_Ricardo.pdf](http://repositorio.ug.edu.ec/bitstream/redug/57094/1/B-CISC-PTG-1996-2021_Cedeño_Núñez_Pedro_Xavier_-_Piña_Naranjo_Ángel_Ricardo.pdf)
- Hasperué, W., Estrebou, C. A., Camele, G., López, P., Jimbo Santana, P. R., Reyes Zambrano, G., Lanzarini, L. C., & Fernández Bariviera, A. (2021). Procesamiento inteligente de grandes volúmenes de información y de flujos de datos. *XXIII Workshop de Investigadores En Ciencias de La Computación (WICC 2021, Chilecito, La Rioja)*. <http://sedici.unlp.edu.ar/handle/10915/120089>
- Liu, L., Song, J., Guan, B., Wu, Z., & He, K. (2011). Tra-DBScan: A Algorithm of Clustering Trajectories. *Applied Mechanics and Materials*, 121–126. <https://doi.org/10.4028/www.scientific.net/AMM.121-126.4875>
- Reyes, G., Lanzarini, L. C., Estrebou, C. A., & Maquilón, V. (2021). Vehicular Flow Analysis Using Clusters. *XXVII Congreso Argentino de Ciencias de La Computación (CACIC)(Modalidad Virtual, 4 al 8 de Octubre de 2021)*. <http://sedici.unlp.edu.ar/handle/10915/130341>
- Reyes, G., Lanzarini, L., Hasperué, W., & Bariviera, A. F. (2020). GPS trajectory clustering method for decision making on intelligent transportation systems. *Journal of Intelligent & Fuzzy Systems*, 38, 5529–5535. <https://doi.org/10.3233/JIFS-179644>
- Reyes, G., Lanzarini, L., Hasperué, W., & Bariviera, A. F. (2021). Proposal for a Pivot-Based Vehicle Trajectory Clustering Method. *Transportation Research Record*, 03611981211058429. <https://doi.org/10.1177/03611981211058429>



- Reyes Zambrano, G. (2019). GPS Trajectory Compression Algorithm. In M. Botto-Tobar, J. Barzola-Monteses, E. Santos-Baquerizo, M. Espinoza-Andaluz, & W. Yáñez-Pazmiño (Eds.), *Computer and Communication Engineering* (pp. 57–69). Springer International Publishing.
- Scherl, M. (2010). *Benchmarking of Cluster Indices*. [https://epub.ub.uni-muenchen.de/12797/1/DA\\_Scherl.pdf](https://epub.ub.uni-muenchen.de/12797/1/DA_Scherl.pdf)
- Thomson, I., & Bull, A. (2002). La congestión del tránsito urbano: Causas y consecuencias económicas y sociales. *Revista de La CEPAL*, 2002(76), 109–121. <https://doi.org/10.18356/fd4a1f83-es>
- Transportation Research Board. (2000). *Highway Capacity Manual*. Transportation Research Board, National Research Council.
- Wang, S., Ding, S., & Xiong, L. (2020). A New System for Surveillance and Digital Contact Tracing for COVID-19: Spatiotemporal Reporting Over Network and GPS. *JMIR MHealth and UHealth*, 8(6), 2–2. <https://doi.org/10.2196/19457>