

## Aplicación de la minería de datos espacial basado en técnicas de agrupamiento para detectar los niveles de congestión del tráfico vehicular en la ciudad de Trujillo

Application of spatial data mining based on aggregation techniques to detect traffic congestion levels in the city of Trujillo

José Arturo Díaz-Pulido\* 

Facultad de Ciencias Físicas y Matemáticas, Universidad Nacional de Trujillo, Av. Juan Pablo II s/n – Ciudad Universitaria, Trujillo, Perú.

\* Autor correspondiente: [jdiazpulido@unitru.edu.pe](mailto:jdiazpulido@unitru.edu.pe) (J. Díaz-Pulido)

DOI: [10.17268/rev.cyt.2024.02.01](https://doi.org/10.17268/rev.cyt.2024.02.01)

### RESUMEN

El análisis del tráfico vehicular en la red vial de la ciudad de Trujillo, detecto puntos críticos con mayor afluencia de vehículos, determinándose un modelo computacional (figura 8) usándose valores aleatorios para detectar la congestión del tráfico vehicular en la red vial de la ciudad de Trujillo. Para la gestión y análisis de minería de datos espacial se usó la metodología CRISP-DM, aplicando aprendizaje no supervisado, debido a que los datos son aleatorios y simulados en tiempo real, lo que permitió el uso de técnicas de agrupamiento de clasificación con los algoritmos k-means y dbscan, y como resultado se obtuvo la ubicación espacial de los nodos con un nivel de congestión, deduciendo las posibles incidencias o eventualidades que ocurren en ellos a nivel vehicular, como es el caso de la velocidad, el tiempo, la capacidad en cada nodo. Se hizo la comparación de los algoritmos de agrupamiento k-means y dbscan para validar el modelo computacional, utilizando las técnicas del acodamiento y del promedio Silhouette respectivamente.

**Palabras clave:** densidad del tráfico; minería de datos espacial; clustering; k-means; dbscan.

### ABSTRACT

The analysis of vehicular traffic in the road network of the city of Trujillo detected critical points with the greatest influx of vehicles, determining a computational model (Figure 8) using random values to detect vehicular traffic congestion in the road network of the city of Trujillo. For the management and analysis of spatial data mining, the CRISP-DM methodology was used, applying unsupervised learning, because the data are random and simulated in real time, which allowed the use of classification clustering techniques with k-means and dbscan algorithms, and as a result the spatial location of the nodes with a level of congestion was obtained, deducing the possible incidences or eventualities that occur in them at vehicular level, as is the case of speed, time, capacity at each node. The k-means and dbscan clustering algorithms were compared to validate the computational model, using the k-means and Silhouette averaging techniques, respectively.

**Keywords:** traffic density, spatial data mining, clustering, k-means, dbscan.

### 1. INTRODUCCIÓN

La congestión del tráfico vehicular es un desafío importante en el área de planificación del transporte vehicular en las redes viales de las diferentes ciudades del mundo, tornándose complejo cuando se produce la congestión vehicular.

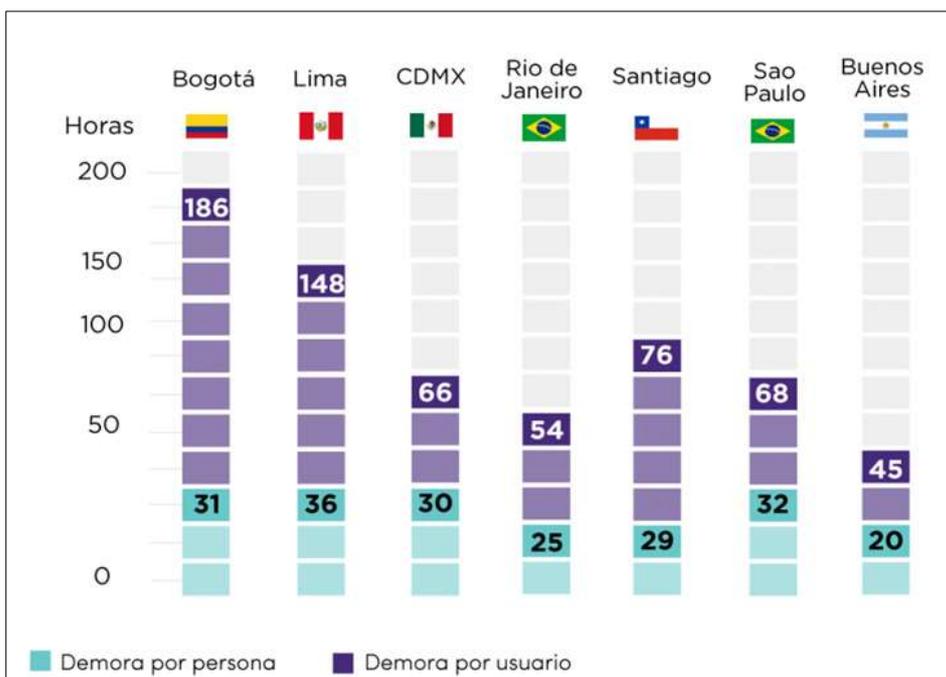
La congestión vehicular generalmente se relaciona con un exceso de vehículos en una parte de la carretera en un momento determinado lo que resulta en velocidades más lentas, a veces mucho más lentas que las velocidades normales o de flujo libre. El tiempo de viaje y la densidad del tráfico son las medidas de tráfico más utilizadas para cuantificar la congestión en las carreteras. La densidad del tráfico se considera como la medida principal para cuantificar la congestión de carreteras que no sean intersecciones señalizadas. Como la densidad es difícil de medir, generalmente se adoptan métodos indirectos para estimar la densidad a partir de otros parámetros, como el flujo, la velocidad o la ocupación (Alvarenga-Rodríguez & Alvarenga-Rodríguez, 2021).



La congestión vehicular es un problema que afecta a muchas ciudades del mundo, generando pérdidas económicas, sociales y medioambientales. Para tener una proporción de lo que implican estas pérdidas, a modo de ejemplo, a Buenos Aires y la Ciudad de México la congestión les cuesta 2 y 2,3 veces lo que el gobierno local invierte en educación. La inversión de Sao Paulo en salud equivale a lo que le cuesta la congestión.

El aumento de la tasa de la urbanización y la ausencia de un plan eficiente del uso del suelo, han generado desafíos importantes para la movilidad urbana.

Otra determinante es la infraestructura vial y la asignación de prioridades en su uso que ha favorecido el transporte individual. Esto repercute en el incremento de la tasa de motorización y la reducción en el uso del transporte público. La tasa de crecimiento por cada 1.000 habitantes fue 4,7% en Latinoamérica los últimos 10 años (BID, 2019).



**Figura 1.** Demora por persona y usuario de vehículo particular en ciudades seleccionadas de Latinoamérica.

Según el Ranking de Congestión de Tráfico - Informe Anual 2023, indica que ciudades como Bogotá, Lima, Ciudad de México y Río de Janeiro se encuentran entre las más congestionadas del mundo, descrito en el portal web <https://trafficindex.org/reports/annual-report-2023/>. Tal como lo señala la figura 1, Lima ha sido catalogada como la ciudad con mayor congestión vehicular de América Latina y la octava de todo el mundo.

En varias ciudades de nuestro país, se presenta congestión vehicular por la informalidad y falta de cultura vial, a nivel de peatones, transporte público y privado, paraderos informales que se ubican en lugares públicos y privados, como: centros comerciales, colegios, hospitales, etc.

En este artículo se consideró la información existente acerca de los 66 puntos críticos de rutas detectadas por la dirección metropolitana de Transporte de la ciudad de Trujillo. Luego, se seleccionaron y procesaron las fuentes de datos para el estudio de forma automática, almacenando las variables más relevantes, tales como: la velocidad que disminuye y el tiempo que aumenta en determinados momentos, cuando los vehículos se trasladan por las diferentes rutas de la ciudad de Trujillo. Las técnicas de “clustering” para obtener agrupamientos de puntos críticos o rutas congestionadas cuyas características fueran similares, se determinaron por valor de la medida de la velocidad.

Para lograr el objetivo principal de determinar el nivel de congestión del tráfico vehicular a través de técnicas de agrupamiento en los diferentes nodos de la ciudad de Trujillo, se diseñó la base de datos espacial, partiendo de la abstracción de la información real de la problemática que se observa a diario en diferentes lugares o vías de la red vial de la ciudad de Trujillo, logrando con esto, la implementación un sistema de información basado en lenguaje de programación Python para simular el control y gestión de flujo del transporte vehicular en la ciudad de Trujillo, con datos extraídos de la información pública acerca de la congestión vehicular de la ciudad de Trujillo brindados por la oficina de TMT (Transporte Metropolitano de Trujillo). Se usó el software aplicativo RapidMiner para comprobar el modelo de caracterización de patrones de tráfico vehicular a nivel de k-means y dbscan.

Se aplicó Minería de datos espaciales basada en técnicas de agrupamiento o “clustering”. Determinando la implementación de los algoritmos k-means y dbscan respectivamente; con el algoritmo k-means se logró obtener la mayor precisión de agrupamientos y con el algoritmo dbscan se consiguió tendencias de densidad en los diferentes puntos críticos simulados.

## 2. METODOLOGÍA

### 2.1. Fundamentos teóricos de la investigación

#### a. Congestión vehicular

Según (Moreira et al., 2020) define la congestión vehicular como: “la condición en que existen muchos vehículos circulando y cada uno de ellos avanza lenta e irregularmente”. Ellos indican que tanto su definición como la del diccionario son de carácter subjetivo y no conllevan a una precisión suficiente. Hay que recalcar que es claro que ellos la definen de la manera en que las personas observan el tráfico vehicular.

#### b. Modelo Computacional

Un modelo computacional es una representación matemática y/o lógica de un sistema o proceso que se implementa y ejecuta en un computador. Estos modelos se diseñan para simular y analizar sistemas complejos, predecir comportamientos y realizar optimizaciones, entre otras (*Introduction to Computational Science: Modeling and Simulation for the ... - Angela B. Shiflet, George W. Shiflet - Google Libros*, s. f.).

En general, un modelo computacional consta de:

- **Entradas (inputs):** Datos o condiciones iniciales que se alimentan al modelo.
- **Procesos:** Conjunto de reglas, algoritmos, ecuaciones y lógica que el modelo utiliza para transformar las entradas en salidas.
- **Salidas (outputs):** Resultados generados por el modelo después de procesar las entradas.

Algunas características y aspectos clave de los modelos computacionales incluyen:

- **Abstracción:** Un modelo computacional no representa todos los detalles de un sistema real, sino que se enfoca en aquellos aspectos que son relevantes para el propósito del modelo.
- **Validación y verificación:** Es crucial validar y verificar un modelo computacional para asegurarse de que es una representación precisa del sistema real y que los resultados que produce son confiables.
- **Flexibilidad:** Dado que se basan en código y algoritmos, los modelos computacionales pueden ser adaptados y modificados según sea necesario para explorar diferentes escenarios o hipótesis.
- **Aplicaciones:** Se utilizan en una amplia variedad de campos, desde la física y la ingeniería hasta la biología, la economía y las ciencias sociales. Por ejemplo, en la meteorología, se usan modelos computacionales para predecir el clima; en la biología, para simular la interacción de proteínas; y en la economía, para modelar el comportamiento de los mercados.
- **Limitaciones:** A pesar de su utilidad, estos modelos tienen limitaciones inherentes debido a la simplificación de sistemas reales y la precisión de los datos de entrada. Por lo tanto, siempre es importante interpretar los resultados con precaución.

#### c. Modelo de caracterización de patrones de tránsito vehicular

La caracterización de patrones de tráfico vehicular se refiere al estudio y análisis de los flujos de tráfico en una red vial, con el objetivo de entender, predecir y gestionar el comportamiento y la demanda del tráfico. Un modelo de caracterización puede basarse en diversas métricas y técnicas dependiendo del objetivo del estudio (Moreira et al., 2020).

Dentro de un modelo de caracterización de patrones de tráfico vehicular, se pueden considerar varios aspectos:

- **Intensidad del tráfico:** Se refiere al número de vehículos que pasan por un punto o segmento de la red en un periodo determinado, usualmente expresado en vehículos por hora (vph).
- **Distribución por tipo de vehículo:** Identifica la proporción de diferentes tipos de vehículos (automóviles, camiones, motocicletas, autobuses, etc.).
- **Velocidad y tiempo de viaje:** Estudia la velocidad a la que se desplazan los vehículos y el tiempo que tardan en recorrer un segmento.
- **Patrones temporales:** Analiza cómo varía el tráfico a lo largo del día, entre días laborables y fines de semana, o entre distintas estaciones del año.
- **Origen y destino:** Investigación sobre los puntos de inicio y término de los viajes, para entender las principales rutas y demandas de desplazamiento.

- **Condiciones de congestión:** Analiza los puntos o periodos donde el tráfico se vuelve más denso y se reduce la velocidad de circulación, lo que puede llevar a atascos.
- **Comportamiento del conductor:** Estudio sobre cómo se comportan los conductores en diferentes situaciones, como en zonas escolares, intersecciones, zonas de obras, entre otras.
- **Impacto de eventos o situaciones especiales:** Analiza cómo eventos específicos (como un concierto, un partido de fútbol o una construcción) pueden alterar los patrones normales de tráfico.

Estos modelos se utilizan en la planificación urbana y de transporte, diseño de infraestructuras, establecimiento de políticas de tráfico, entre otros. Por ejemplo, pueden ayudar a determinar dónde sería más efectivo construir un nuevo puente, ampliar una carretera o establecer una nueva línea de transporte público. Además, estos modelos también son cruciales para desarrollar soluciones inteligentes de transporte, como sistemas de semáforos adaptativos o aplicaciones de navegación en tiempo real.

#### d. Algoritmo de agrupamiento K-means

Se basa en el agrupamiento dividido, que intenta encontrar un número de clústeres (K) especificado por el usuario, representado por sus centroides. El algoritmo básico se describe a continuación.

Primero, se seleccionan K centroides iniciales. donde K es un parámetro especificado por el usuario y corresponde al número deseado de clústeres.

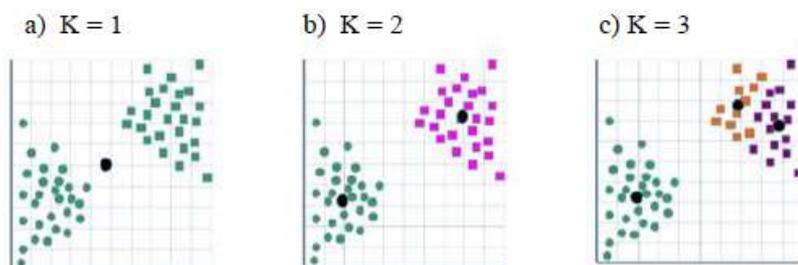
Cada punto está asociado con el centroide más cercano, y cada colección de puntos asociados con un centroide representa un grupo. El centroide de cada clúster se actualiza en función de la asignación de puntos a los clústeres. El procedimiento de asignación y actualización se repite hasta que los puntos en los clusters no cambien o los centroides no cambien (Yang et al., 2019). (Figura 2).

El algoritmo básico de K-means consta de los siguientes pasos:

1. Seleccionar K puntos iniciales como centroides
2. Repetir
3. Formar K cluster asignando cada punto a su centroide más cercano
4. Recalcular el centroide de cada cluster
5. hasta que los centroides no cambien.

Para ejecutar este algoritmo se requiere elegir a priori el valor K (no se sabe cuántos grupos puede haber), esto se puede hacer de dos formas:

- Se puede utilizar un método jerárquico sobre una muestra de los datos (por eficiencia) para estimar el valor de K.
- Usar un valor de K alto, ver los resultados y ajustar.
  - Siempre que se aumente el valor de K disminuirá el valor de la suma de los cuadrados dentro de cada grupo (WCSS).
  - Lo normal es ir probando con varios valores de K y comprobar cuanto no hay de una mejora significativa en SSE.



**Figura 2.** Algoritmo K-means. Tomado de (Tan et al., 2019).

Ejemplo: para encontrar tres clusters en datos de prueba, a partir de tres centroides definidos inicialmente, los clústeres finales se encuentran en cuatro iteraciones de asignación-actualización

Las limitaciones de K-means según (Tan et al., 2019) son:

- K-means tiene problemas cuando los clústeres son de diferente tamaño, densidades, que no tengan forma esférica.
- K-means tiene problemas cuando los datos contienen outliers.

En el **algoritmo k-means, la actualización de los centroides** se realiza en base a los puntos asignados a cada cluster. La fórmula para actualizar un centroide en el paso  $t+1$  es la siguiente:

$$C_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x \in S_i^{(t)}} x$$

Donde:

- $C_i^{(t+1)}$  :es el nuevo centroide del cluster  $i$  en la iteración  $t+1$ .
- $|S_i^{(t)}|$  :es el número de puntos asignados al cluster  $i$  en la iteración  $t$ .
- $x$ : son los puntos asignados al cluster  $i$  en la iteración  $t$ .

Esta fórmula calcula el nuevo centroide como el promedio de todos los puntos asignados al cluster en la iteración actual. Es decir, se suma la posición de cada punto y se divide por el número total de puntos asignados al cluster.

La ventaja de K-means es ser un algoritmo simple, efectivo para pequeñas y medianas cantidades de datos. Utiliza el promedio para representar los centros de los clusters.

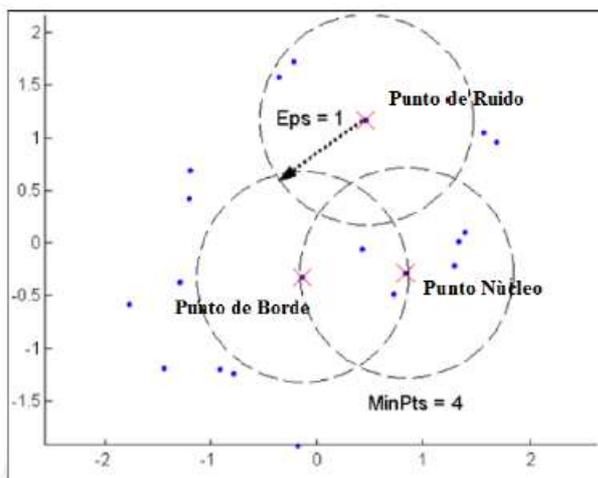
#### e. Algoritmo de agrupamiento DBSCAN

Es un método de clustering basado en densidad. La idea es hacer crecer un clúster siempre y cuando la densidad en el entorno del objeto exceda de un umbral. Este tipo de método permite la detección de clústeres de forma arbitraria, sirviendo además para filtrar datos ruidosos (Yang et al., 2019).

DBSCAN (agrupación espacial basada en la densidad de aplicaciones con ruido) detecta objetos centrales. H. Objetos adyacentes al alcance de la mano. Un objeto central y sus vecinos están conectados para formar regiones densas similares a grupos (Han et al. 2019). Este es un algoritmo de agrupamiento basado en la densidad que produce agrupamientos particionados en los que el algoritmo determina automáticamente el número de agrupamientos. El algoritmo no produce un agrupamiento perfecto, ya que los puntos dispersos se clasifican como ruido y se omiten (Tan et al., 2019).

Este algoritmo utiliza un enfoque de densidad basado en el centro. La densidad se calcula para un punto determinado contando el número mínimo de puntos (MinPts) necesarios para formar un grupo que se ajuste al radio de vecindad máximo ( $\epsilon$ ). Esto nos permite clasificar los puntos como aquellos que se encuentran dentro de la región densa (puntos centrales), aquellos que se encuentran en el borde de la región densa (puntos de borde) y aquellos que están dispersos (puntos de ruido o de fondo) (Tan et al., 2019).

- **Puntos de núcleo (core):** Puntos que tienen más de MinPts vecinos dentro de su vecindario de Radio Eps.
- **Puntos de borde (border):** Son los puntos que tienen menos de MinPts vecinos dentro de su vecindario de radio Eps, pero están en la vecindad de un punto de núcleo.
- **Puntos de ruido (noise):** Son aquellos puntos que no caen en ninguna de las dos categorías anteriores.



**Figura 3.** Puntos de Núcleo, Borde y Ruido (Considerando un Eps de valor 1 y un MinPts de valor 4). Tomado de (Tan et al., 2019).

El algoritmo comienza eligiendo un punto arbitrario  $p$ . Si  $p$  es un punto central, entonces comienza la formación de grupos y todos los objetos que se encuentran cerca de  $p$  se colocan en ese grupo. Si  $p$  no es el punto central, se visita otro objeto en el conjunto de datos. El procesamiento continúa hasta que se hayan procesado todos los objetos. Los puntos fuera del grupo formado se denominan puntos de ruido, y los puntos que no son ni ruido ni centro se denominan puntos de borde.

Todos los pares de puntos centrales que están lo suficientemente cerca entre sí (menos de la distancia Eps) se asignan al mismo grupo. De manera similar, los puntos de borde lo suficientemente cerca del punto central se colocan en el mismo grupo que el punto central y los puntos de ruido se descartan (Tan et al., 2019).

El algoritmo DBSCAN se describe como sigue:

Algoritmo DBSCAN (Geppino Pucci Lorenzo Cazzador & Andrea Pietracaprina, 2023)

1. Etiquetar todos los puntos como núcleo, borde o ruido
2. Eliminar puntos de ruido.
3. Poner un borde entre todos los puntos de núcleo que están dentro de Eps de cada uno de otros
4. Convierta cada grupo de puntos centrales conectados en un clúster separado.
5. Asignar cada punto de borde a uno de los clústeres de sus puntos de núcleo asociados.

DBSCAN puede encontrar grupos de forma arbitraria. Sin embargo, es posible que no produzca un agrupamiento perfecto porque los puntos dispersos se consideran ruido y se filtran.

## 2.2. Desarrollo de la investigación

Se desarrolló un modelo computacional (figura 6) que explica el proceso del análisis de la congestión vehicular permitiendo ingresar características del flujo de vehículos, explicando su comportamiento por ubicación, fecha y hora de ocurrencia, grado de congestión vehicular en determinadas horas pico, etc. Para llevar a cabo la ejecución de este modelo se recopilan datos y luego, mediante la aplicación de técnicas de minería de datos, se analizan y se encuentran patrones de comportamiento del tráfico vehicular.

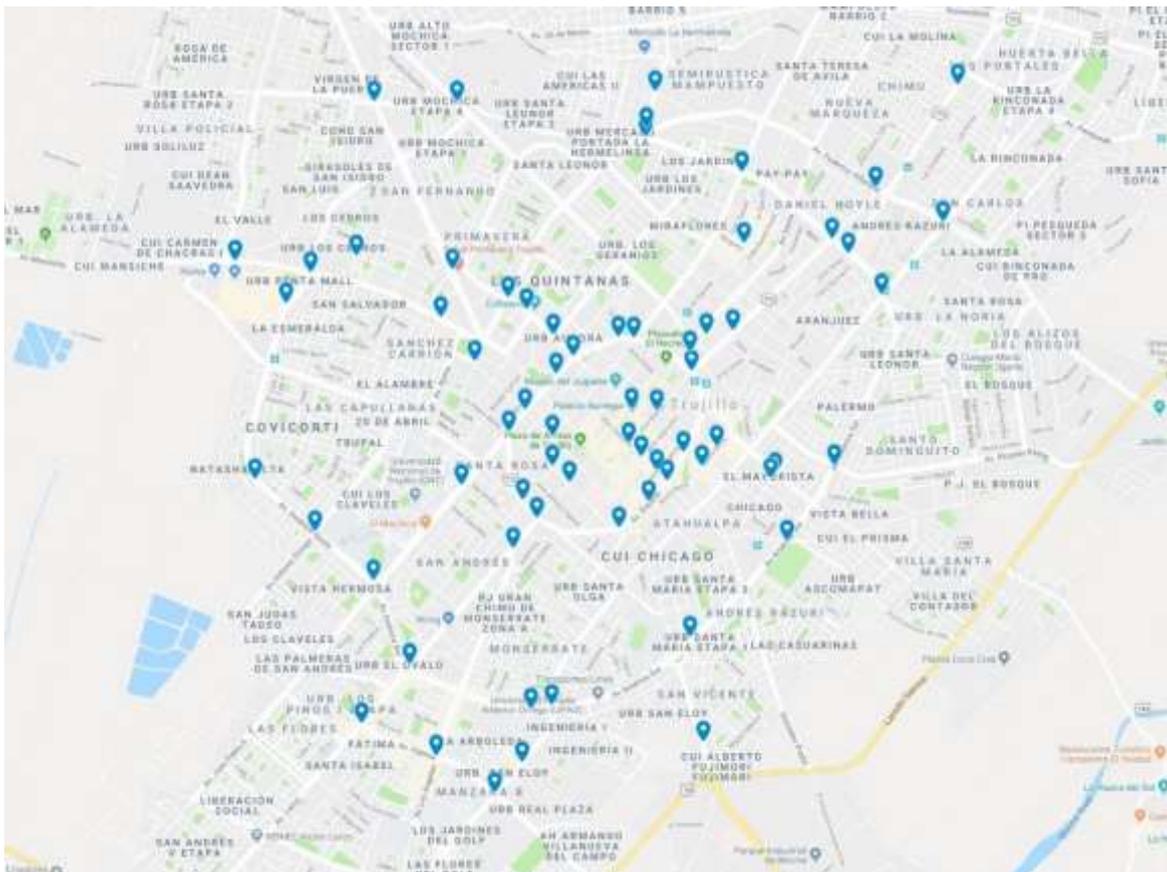
### 2.1. Población y Muestra

#### 2.1.1. Población

Para este estudio se consideró como población todas las vías de la ciudad de Trujillo.

#### 2.1.2. Muestra

Debido a la falta de información de las vías de la ciudad de Trujillo, se consideró la muestra por conveniencia (Hernández González, s. f.) de los 66 puntos (lugares) de congestionamiento proporcionados por la dirección de Transporte Metropolitano de la ciudad de Trujillo (ver Figura 4 y Anexo A).



**Figura 4.** Puntos críticos detectados por la dirección metropolitana de transporte de la ciudad de Trujillo.

## 2.2. Aplicación de la metodología CRISP-DM (Espinosa-Zúñiga & Espinosa-Zúñiga, 2020)

### 2.2.1. Comprensión del negocio

- **Determinar los objetivos del negocio**
- ✓ **Determinar mediante el aforo/ volumen simulado de vehículos, el nivel del congestionamiento, tal como lo indica en su información la oficina Metropolitana de transportes de la ciudad de Trujillo (ver Anexo A).**
- ✓ **Extraer conocimiento, el cual puede ser en forma de relaciones, patrones o reglas inferidas de los datos y previamente desconocidos, o bien en forma de una descripción más concisa. Estas relaciones o resúmenes constituyen el modelo de los datos analizados.**
- ✓ **Diagnosticar a través de un modelo descriptivo, las rutas congestionadas utilizando técnicas de agrupamiento de datos espaciales basadas en los algoritmos K-means y dbscan.**
- **Evaluar la situación**

Debido a que se cuenta con información descriptiva, se elabora una base de datos. Tal como se muestra en la figura 5.

### Análisis de la Información

De la figura 4 se determina a cada punto de referencia (calle, jirón, avenida, esquina, etc.) como un nodo (el cual tiene espacialmente una ubicación geográfica a través de su longitud y latitud), siendo así reconocido hacia adelante en los datos de la figura 5.

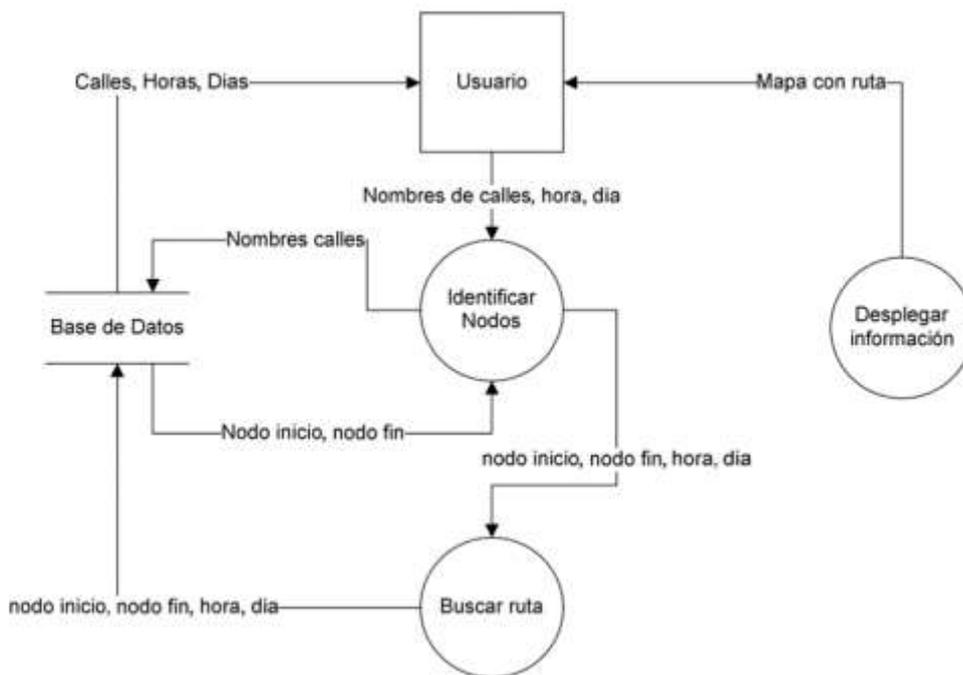


Figura 5. Diagrama de flujo de datos de la situación actual del tráfico en la ciudad de Trujillo.

### Diseño de la Información

De acuerdo a la información proporcionada por la Oficina Metropolitana de Transporte de la ciudad de Trujillo, se hace el diseño de una base de datos relacional (figura 6) para almacenar los datos identificados en la congestión en un determinado punto de referencia o nodo.

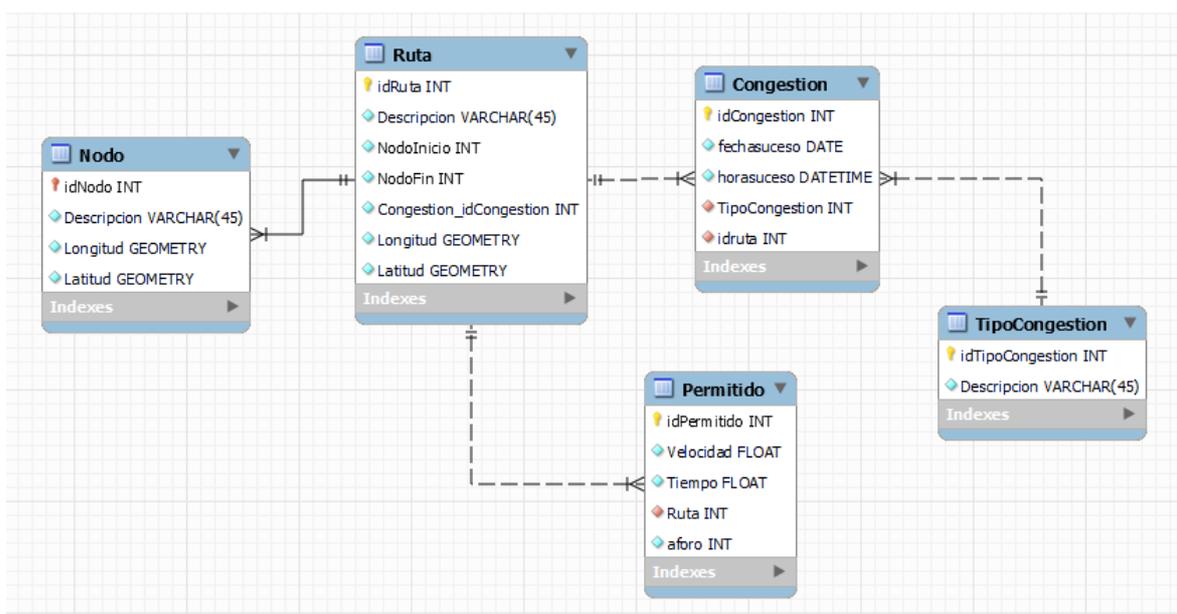


Figura 6. Modelo relacional de la base de datos espacial, para almacenar y determinar la congestión y sus niveles en la ciudad de Trujillo.

**Descripción de las Variables:**

De acuerdo al análisis elaborado de la información brindada por la oficina Metropolitana de Transporte de la ciudad de Trujillo, se determinó el modelo relacional de la base de datos de la figura 6.

**Tabla 1.** Tabla Nodo que almacena los puntos de referencias del sistema vial del tráfico vehicular.

<b>idNodo</b>	Clave primaria.
<b>Descripción</b>	Descripción o nombre de un punto geográfico (ovaló, esquina de calles, etc.)
<b>Longitud</b>	Determina la ubicación de la dirección a coordenada geográfica, perteneciente a la ruta.
<b>Latitud</b>	

**Tabla 2.** Tabla Ruta, almacena la descripción del nodo.

<b>idRuta</b>	Clave primaria.
<b>Descripción</b>	Descripción o nombre de la ruta (calle, avenida, jirón, pasaje, etc.)
<b>NodoInicio</b>	Identifica el nodo inicial (lugar de origen) desde donde inicia una ruta.
<b>NodoFin</b>	Identifica el nodo final (lugar destino) donde finaliza una ruta.
<b>Longitud</b>	Determina la ubicación de la dirección o coordenada geográfica, perteneciente a la ruta.
<b>Latitud</b>	

**Tabla 3.** Tabla Congestión, almacena los momentos de las ocurrencias del fenómeno.

<b>idCongestion</b>	Clave primaria.
<b>FechaSuceso</b>	Descripción de la fecha (día/mes/año), cuando ocurre la congestión.
<b>Horasuceso</b>	Descripción de la hora (hora:minuto:segundo), cuando ocurre la congestión.
<b>TipoCongestion</b>	Identifica el nivel de congestión.

**Tabla 4.** Tabla Tipocongestion, almacena los grados de congestión.

<b>idTipoCongestion</b>	Clave primaria.
<b>Descripción</b>	Descripción del nivel de congestión.

**Tabla 5.** Tabla Permitido, almacena los datos asignados por la oficina metropolitana de transporte.

<b>idPermitido</b>	Clave primaria.
<b>Velocidad</b>	Cantidad de velocidad permitida (según norma establecida por la municipalidad provincial de Trujillo)
<b>Tiempo</b>	Cantidad de tiempo permitido (según norma establecida por la municipalidad provincial de Trujillo)
<b>Aforo</b>	Cantidad de vehículos permitidos para cada ruta (según norma establecida por la municipalidad provincial de Trujillo)

- **Determinar objetivos de minería de datos**

Los objetivos de minería de datos en nuestra investigación son las siguientes:

- Procesar toda la información recopilada, para que así pueda ser manejada por los métodos clustering, estos dos métodos solo trabajan con valores cuantitativos.
- Hallar la distancia de los datos, en cuestión de acercamiento entre unidades vehiculares, esto medirá cuanta es la diferencia del nivel de congestamiento en cada punto de la ciudad donde se produzca la aglomeración.

- **Producir plan de proyecto**

Para producir el plan de proyecto de esta investigación se procedió a estimar el tiempo de realización:

- Etapa 1: Datos analizados, estructurados y recopilación de la información. Estimación: 1 mes.
- Etapa 2: Realizar consultas para adquirir muestras que puedan representar los datos. Estimación: 2 semanas.
- Etapa 3: Fase 3 Preparación de los datos, seleccionar, limpiar, conversión y el formateo si en caso fuera necesario, para facilitar el análisis de estos datos. Estimación: 1 mes.
- Etapa 4: Seleccionar la técnica que se utilizara para la ejecución de los modelos sobre los datos procesados. Estimación: 2 semanas.
- Etapa 5: Realizar un análisis general de cómo se trabajó la metodología incluyendo la etapa anterior. Estimación: 2 semanas.
- Etapa 6: Generar informes acerca de los avances hechos, para una futura referencia si se tiene que replantear algún criterio de evaluación. Estimación: 2 semanas.
- Etapa 7: Mostrar los resultados. Estimación: 2 semanas.

### **Comprensión de los datos**

En esta fase de CRISP-DM se procederá a recolectar todos los datos iniciales para poder establecer un primer contacto con el problema, conocer los datos y averiguar su calidad.

- ✓ **Recolectar datos Iniciales**

Los datos utilizados en este tema de investigación son datos asociados a las velocidades, tiempos, distancias de rutas, semaforizaciones, latitudes y longitudes de lugares de referencia a puntos geográficos estos datos serán recolectados de manera ordenada en una hoja de cálculo en MS Excel, el cual se procedió a ordenar de manera que cada se convierta en una variable numérica.

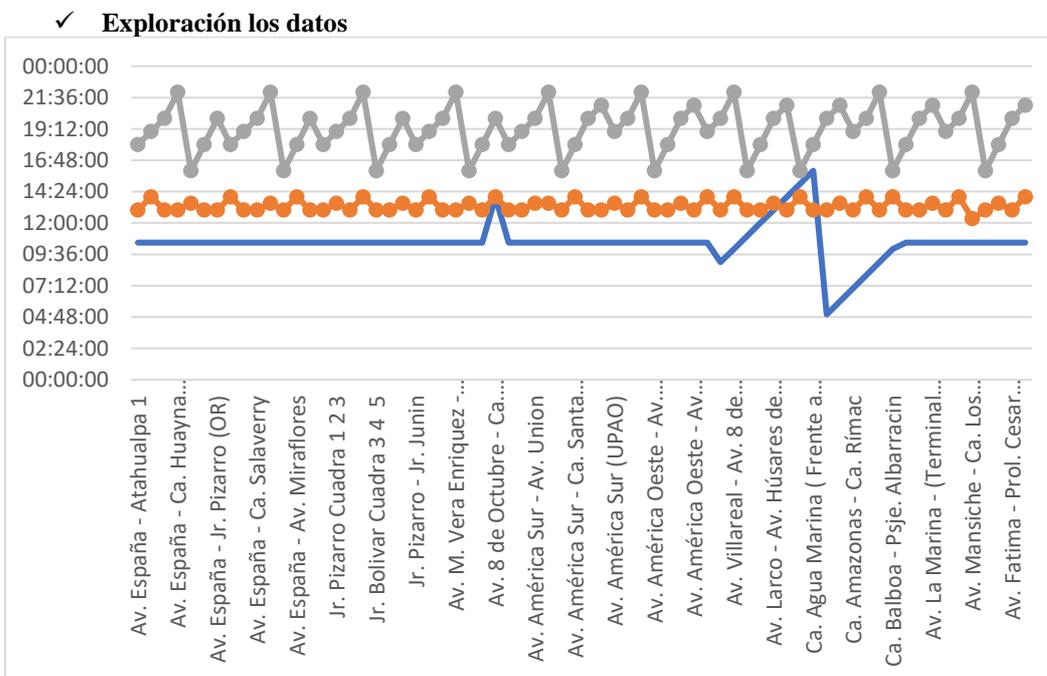
- ✓ **Describir los datos**

Se trabajó con 66 puntos y con los algoritmos que comparen estos datos. Cada punto crítico hace referencia a:

Los campos seleccionados para cada punto crítico de referencia son:

- Latitud: viene hacer la distancia en grados, minutos y segundos con respecto al ecuador (0°).

- Longitud: Es la distancia en grados, minutos y segundos que hay con respecto al meridiano de Greenwich (0°).
- Semáforo: determina si en el punto crítico de referencia puede existir un semáforo.
- HoraPuntaM: determina la hora punta por la mañana en que se produce congestión vehicular en un determinado punto crítico.
- HoraPuntaT: determina la hora punta por la tarde en que se produce congestión vehicular en un determinado punto crítico.
- HoraPuntaN: determina la hora punta por la noche en que se produce congestión vehicular en un determinado punto crítico.
- Cantidadvehiculos: determina el número de vehículos que actualmente se encuentren en el punto crítico.
- Tope: se refiere al número limite que debe existir en cualquier punto crítico.
- VelocidadPermtida: se refiere a la velocidad impuesta según ordenanza municipal que debe existir para cruzar la ruta o punto crítico.
- TiempoPermitido: se refiere al tiempo impuesto según ordenanza municipal que debe existir para cruzar la ruta o punto crítico.
- VelocidadAprox: se refiere a la velocidad actual con la que se encuentra una unidad móvil al cruzar la ruta o punto crítico.
- TiempoAprox: se refiere al tiempo actual con la que se encuentra una unidad móvil al cruzar la ruta o punto crítico.



**Figura 7.** Exploración de los datos en general

En la Figura 7 se muestra los diferentes puntos críticos (ubicaciones geográficas) con relación a las diferentes horas punta. Se puede apreciar que depende mucho de las horas punta en que se encuentra congestionado los puntos críticos para que la relación varié, no todos los puntos cuentan con la misma hora punta.

✓ **Verificar la calidad de los datos**

Después de hacer la exploración inicial de los datos se puede afirmar que estos son completos. Los datos cubren los casos requeridos para la obtención de los resultados necesarios para poder cumplir los objetivos del proyecto. Los datos no contienen errores y tampoco se encuentran valores fuera de rango ya que son datos generados y controlados respectivamente de manera automática por el software simulador, por lo que no hay riesgo de ruido en el proceso de la minería de datos.

### 2.2.2. Preparación de los datos

En esta fase de la metodología CRISP – DM, se procederá a mostrar los pasos para alistar la data, previamente a ser procesada.

#### ✓ **Seleccionar datos**

Se utilizó solo los datos numéricos, ya que se utilizarán los métodos de clustering, estos dos métodos tienen como requisito utilizar datos numéricos para ser comparados. Sin embargo, hay datos que se pueden prescindir.

Los campos seleccionados para cada punto crítico de referencia son:

- Latitud: viene hacer la distancia en grados, minutos y segundos con respecto al ecuador (0°).
- Longitud: Es la distancia en grados, minutos y segundos que hay con respecto al meridiano de Greenwich (0°).
- Semáforo: determina si en el punto crítico de referencia puede existir un semáforo.
- HoraPuntaM: determina la hora punta por la mañana en que se produce congestión vehicular en un determinado punto crítico.
- HoraPuntaT: determina la hora punta por la tarde en que se produce congestión vehicular en un determinado punto crítico.
- HoraPuntaN: determina la hora punta por la noche en que se produce congestión vehicular en un determinado punto crítico.
- Cantidadvehiculos: determina el número de vehículo que actualmente se encuentren en el punto crítico.
- Tope: se refiere al número limite que debe existir en cualquier punto crítico.
- VelocidadPermtida: se refiere a la velocidad impuesta según ordenanza municipal que debe existir para cruzar la ruta o punto crítico.
- TiempoPermitido: se refiere al tiempo impuesto según ordenanza municipal que debe existir para cruzar la ruta o punto crítico.
- VelocidadAprox: se refiere a la velocidad actual con la que se encuentra una unidad móvil al cruzar la ruta o punto crítico.
- TiempoAprox: se refiere al tiempo actual con la que se encuentra una unidad móvil al cruzar la ruta o punto crítico.

El motivo de inclusión o exclusión de algunos campos, depende de la importancia de los datos para llegar al objetivo de la investigación.

#### ✓ **Limpiar datos**

Nuestra base de datos no necesita una limpieza a profundidad, porque se manejó los criterios básicos para construir la información, no permitir datos nulos, no permitir datos textuales, como también la redundancia de los datos, esto provocaría una pérdida de tiempo al momento de analizar los datos de forma ordenada.

#### ✓ **Construir datos**

##### **Atributos derivados**

En esta parte solo se puede destacar la transformación de los datos numéricos de los diferentes puntos críticos porque al momento de utilizar los métodos de clustering para que puedan ser comparados.

##### **Registros generados**

Todo registro generado será un gráfico, para entender mejor la data, adicionalmente si el grafico es demasiado grande para ser entendible, se procedió a desmenuzar este.

#### ✓ **Integrar datos**

Mediante el programa python, la información fue procesada desde MS Excel.

##### **Formatear datos**

No se formatearon datos para el método de clustering ya que todos son numéricos.

### 2.2.3. Modelado

En esta fase de la metodología CRISP-DM se procedió a generar los modelos a realizar, porque la data obtenida debe ser representada de manera gráfica para un mejor entendimiento.

#### 2.2.3.1. Seleccionar técnicas de modelado

Se utilizó el algoritmo k-means, conocido como un algoritmo de clasificación no supervisada, este agrupa objetos en k grupos basándose en sus características, consta de tres pasos:

- Inicialización: Escogido el número de grupos, se establece en K centroides en el espacio de los datos.

- Asignación de objetos a los centroides: cada objeto es asignado a su centroide más cercano.
- Actualización de centroides: se actualiza dependiendo de cada grupo tomado como nuevo centroide.

Se utilizó el algoritmo dbscan, conocido como un algoritmo de agrupación en clústeres dbscan que se basa en la aglomeración de agrupaciones determinada por la densidad que está diseñado para descubrir agrupaciones de forma arbitraria. Dbscan requiere solo un parámetro de entrada y ayuda al usuario a establecer un valor apropiado para él. En los puntos de un mismo clúster, su k-ésimo vecino debería estar más o menos a la misma distancia. En los puntos de ruido, su k-ésimo vecino debería estar más lejos. dbscan está diseñado para descubrir los clústeres y el ruido en una base de datos espacial (Han et al 2012). Idealmente, se deberían conocer los parámetros apropiados Eps (radio de vecindad) y MinPts (como vecinos mínimos para considerar un punto como punto central) de cada grupo y al menos un punto del grupo respectivo. Luego, se podrían recuperar todos los puntos que son de densidad alcanzable desde el punto dado usando los parámetros correctos. Pero no hay una manera fácil de obtener esta información por adelantado para todos los grupos de la base de datos. Sin embargo, hay una forma simple y efectiva: la heurística para determinar los parámetros Eps y MinPts. Por tanto, dbscan usa valores globales para Eps y MinPts, es decir, los mismos para todos los grupos. Los parámetros de densidad del grupo “más delgado” son buenos candidatos para estos valores de parámetros globales que especifican la densidad más baja que no se considera ruido.

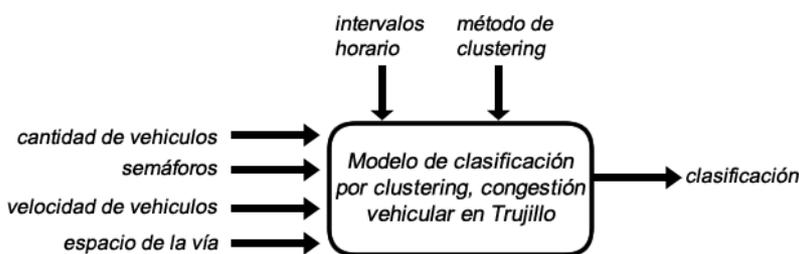
El conjunto de datos de este estudio está compuesto, en parte, por datos geográficos, lo que influyó en la selección del algoritmo a aplicar en la etapa de minería de datos, por lo cual se escogió dbscan como el más apropiado. El algoritmo de clustering dbscan requiere solo un parámetro de entrada y ayuda al usuario a determinar un valor apropiado para ello. dbscan es eficiente incluso para grandes bases de datos espaciales. Este algoritmo está diseñado para descubrir los clústeres y el ruido en un espacio espacial (Yang et al., 2019).

### 2.2.3.2. Generar el plan de prueba

El procedimiento que se empleó para probar la calidad y validez del modelo será la comparación de los algoritmos k-means y dbscan. Estas medidas determinaran el modelo más óptimo.

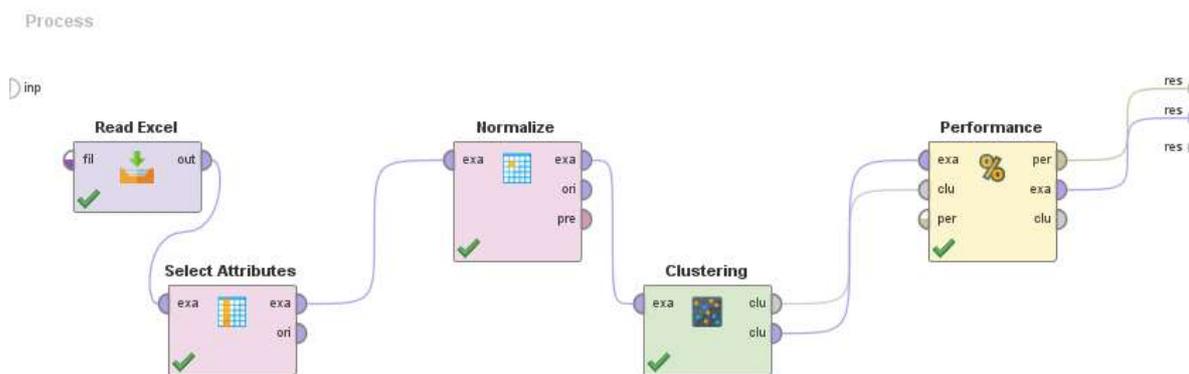
### 2.2.3.3. Construir el modelo

Para el presente trabajo, según lo expuesto por (Epstein, 2020), los modelos desarrollados son modelos explicativos, no modelos predictivos, sino modelos explícitos cuyos resultados son evaluados por otros después de la implementación. La intención de este modelo es explicar, no predecir. El grado de congestión puede explicarse a partir del lugar de ocurrencia y la fecha y hora de ocurrencia, pero es imposible predecir la hora y el lugar de ocurrencia. Los datos primero se recopilan y luego se analizan mediante la aplicación de técnicas de minería de datos para descubrir patrones de comportamiento de la congestión. La información oculta se puede encontrar aplicando técnicas de minería de datos y puede generar nuevos problemas para desarrollar aún más el modelo.



**Figura 8.** Prototipo del modelo de caracterización de congestión vehicular.

Para ejecutar el modelo (figura 8) elegido sobre los datos de entrenamiento se utilizó el software Rapid-Miner, donde se pudo definir y construir el tipo de modelado. El análisis se realizó utilizando la técnica de minería de datos, clustering. Tal como se muestra en la figura 8.



**Figura 9.** Diseño del modelo en software RapidMiner

Se implementaron 2 algoritmos de minería de datos: k-means y dbscan. Estos algoritmos son de tipo no supervisado lo que significa que no necesitan de una variable de clasificación y pueden ser utilizados en grandes cantidades de datos (Chakraborty et al., 2019).

### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. Resultados Computacionales

Los resultados obtenidos en este estudio permitieron comprender cómo se organiza y gestiona la información espacial en las bases de datos espaciales según sus propiedades (representación, relaciones y operaciones). De igual forma, sobre estas bases de datos se puede realizar el proceso de preparación y transformación de conjuntos de datos de entrada para su posterior aplicación de técnicas de minería de datos.

- Evaluación del modelo.
- Determinar el nivel de Congestionamiento.

##### 3.1.1. Evaluación del modelo

Los algoritmos de agrupamiento k-means y dbscan funcionan de manera diferente, pero k-means calcula aleatoriamente la cantidad de grupos o particiones para recuperar y la distancia entre los objetos dentro de cada grupo. En dbscan, el radio de distancia entre objetos y el número mínimo de objetos en cada grupo. Allí se entendió la utilidad, funcionalidad y aporte a esta investigación de los algoritmos de agrupamiento. Porque permiten un análisis espacial preciso de elementos aleatorios sin un patrón común.

Para detectar la mayor cantidad de concentraciones y ubicar aquellos puntos de la ciudad de mayor tráfico recurrente, se tomaron en cuenta un dataset identificando la información de los puntos de la ciudad brindados por la oficina metropolitana de Transporte de Trujillo (Syakur et al., 2021).

##### 3.1.1.1. Análisis de datos

Se aplicó para los métodos de clustering k-means y dbscan el mismo dataset respectivamente.

Para el algoritmo k-means, se realizó la técnica de acomodamiento para encontrar el valor más adecuado para el número de agrupamientos (clúster) en una primera instancia.

Para el algoritmo K-means se aplicó el método del coeficiente Silhouette para medir la calidad de los agrupamientos (clusters) encontrados. Este método va a permitir determinar qué tan bien cada objeto se encuentra dentro de su agrupación. El número óptimo de clusters k es el que maximiza la silueta promedio en un rango de posibles valores de k.

Para el algoritmo dbscan, se utilizó la técnica KNN (k-nearest neighbors), para hallar el valor más adecuado de la distancia epsilon. El objetivo de su aplicación es calcular el promedio de las distancias de todos los puntos a sus k vecinos más cercanos. El valor de k será incremental en un ciclo y se corresponde con MinPts.

A continuación, con la distancia óptima calculada con la técnica KNN, se realizaron varias iteraciones para calcular los agrupamientos obtenidos con el algoritmo dbscan, con un valor de minPts de 2 a 8. El valor de minPts óptimo se lo asociara con número máximo de clusters determinado, luego de la ejecución de todas las iteraciones.

El algoritmo dbscan provee una clusterización comparativamente mejor con otros métodos, además es apropiado para el tratamiento de datos geográficos, con lo cual es posible establecer gran cantidad de clústeres en la zona de estudio.

Se comparó los resultados del modelo computacional usando los algoritmos k-means y dbscan. Con la ejecución del algoritmo K-means se obtuvieron resultados más precisos, mientras que con el algoritmo dbscan se obtuvo información de tipo aglomerativo.

### 3.1.1.2. Ejecución del modelo computación usando el Algoritmo K-means

Para encontrar un número óptimo de clústeres, que no involucren agrupaciones heterogéneas (pocos clústeres); o datos que siendo muy similares unos a otros los agrupemos en clústeres diferentes (muchos clústeres). Por lo que, se consideraron 2 métodos para dicha elección: la técnica de acodamiento y el método del coeficiente Silhouette. En la figura 10, se ejecutó la técnica de acodamiento, donde se puede observar que el número sugerido de clústeres K, como parámetro para el algoritmo k-means está en 3.

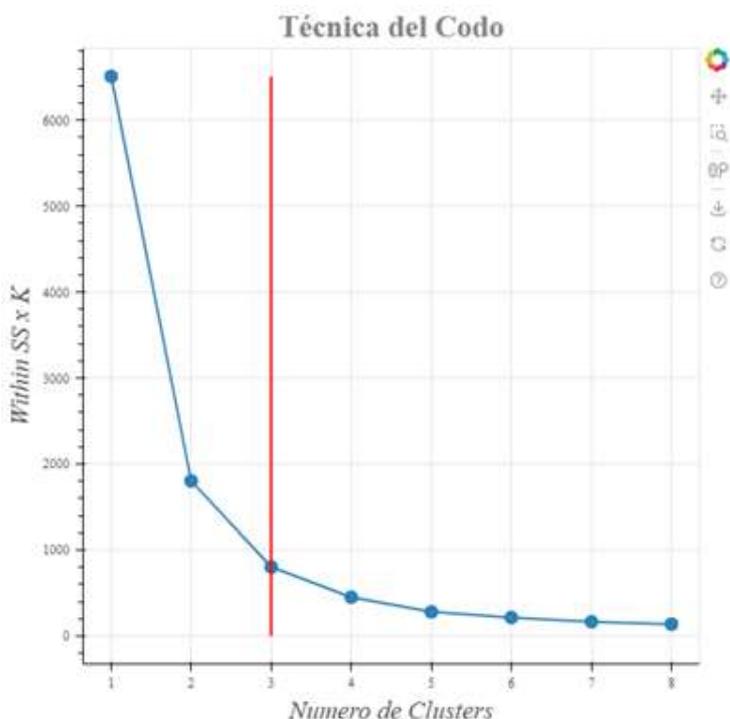


Figura 10. Técnica del codo en K-means

#### Explicación:

Para comprender el cálculo de WCSS, hay que considerar en que para un punto del eje X de la gráfica (cuales indican la cantidad de clúster) se hagan cálculos sobre 3 puntos (A, B y C), siendo B el punto analizado, A el punto que está inmediatamente atrás y C el punto inmediatamente adelante, ya que lo que se busca es ver el punto donde se forma el codo para lo cual necesitamos al menos 3 puntos de referencia. Entonces:

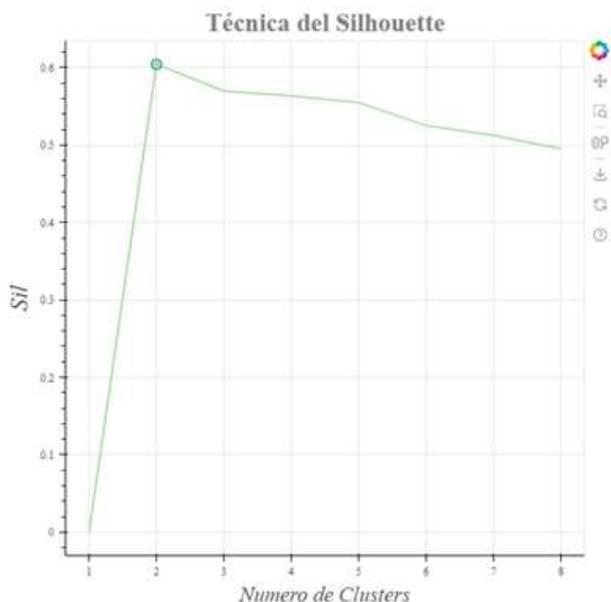
Calcular la diferencia de WCSS (eje Y) entre A y B (D1)

Calcular la diferencia de WCSS (eje Y) entre B y C (D2)

Dividir D1/D2

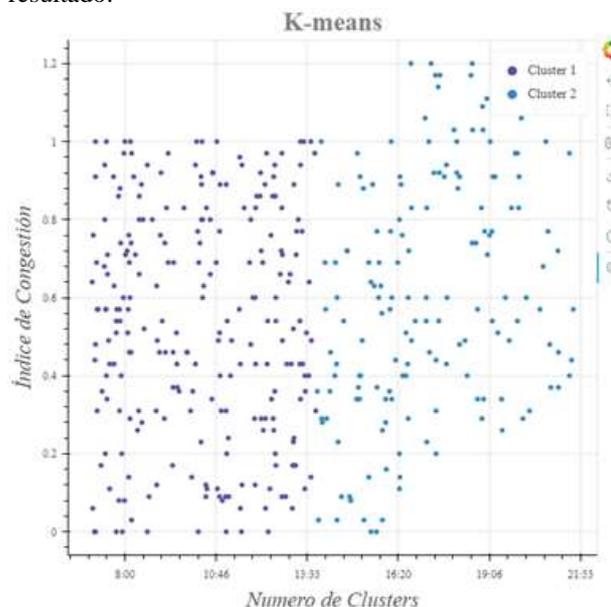
El cociente indica cuantas veces entra D2 en D1, esto permite el puntaje óptimo del codo. Para nuestro caso, se calcula la suma de errores cuadráticos dentro del clúster para diferentes valores de K y se elige la K para la cual la suma de errores cuadráticos comienza a disminuir. Esto es visible como un codo.

En la figura 11, se ejecutó el método del coeficiente Silhouette, en un rango de entre 2 a 8 particiones, el resultado sugiere que el número óptimo de agrupamientos (clusters), para el dataset es 2.



**Figura 11.** Técnica Promedio Silhouette para algoritmo K-means.

A continuación, en la figura 12 se procesa el algoritmo K-means, con parámetro K =2, y se obtiene el siguiente resultado:

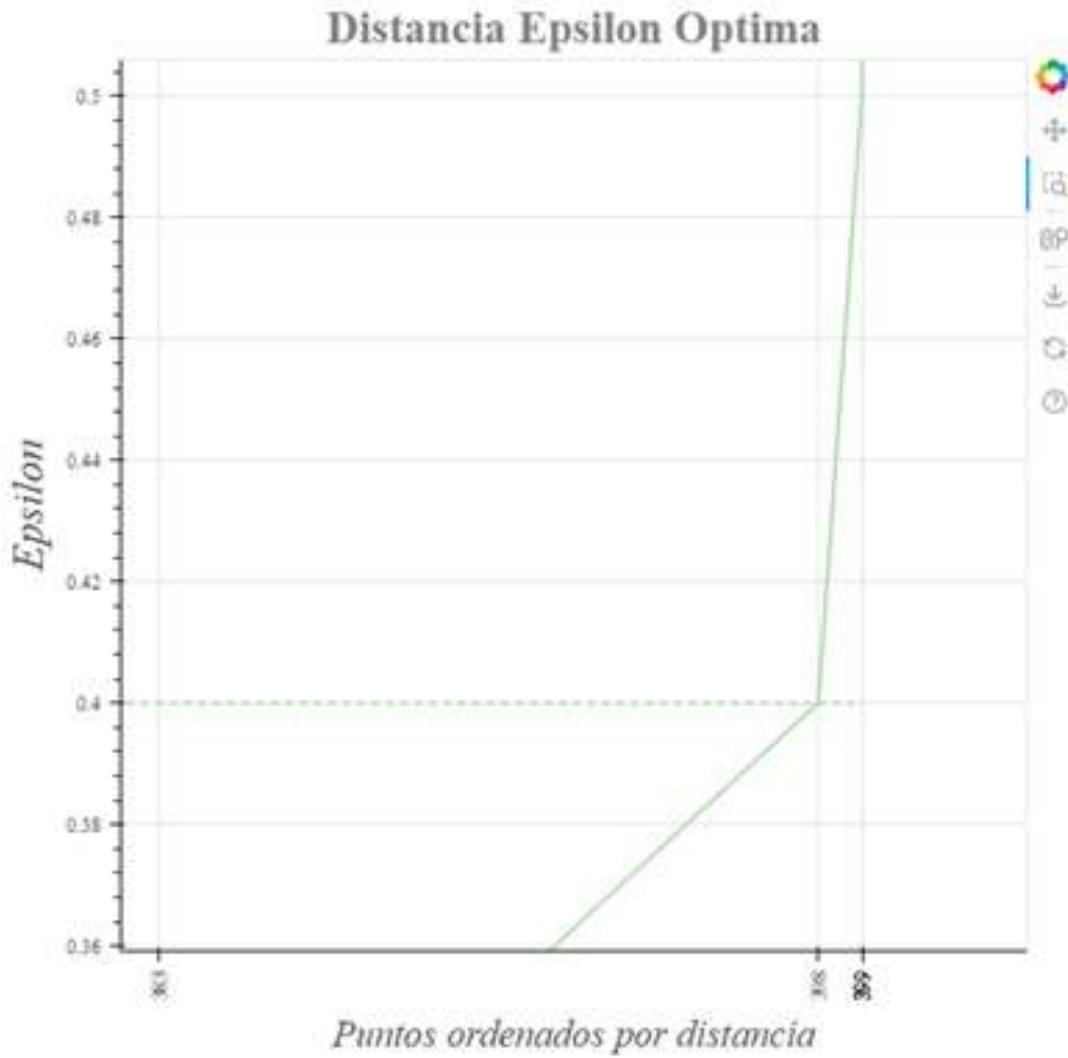


**Figura 12.** Aplicación del algoritmo K-means. Modelo k=2.

El objetivo principal es particionar datos en K clústeres (para un K dado). Se contrasta el modelo comprobando las particiones de clústeres, por medio de k. Ejemplo: k=2.

**3.1.1.3. Ejecución del modelo computación usando el Algoritmo DBSCAN**

En la figura 13, se muestra la aplicación de la técnica KNN para evaluar distancias de 0 a 1.5, para el dataset en cuestión y se determinó que con un número de MinPts = 37, la distancia épsilon óptima es de 0.4.



**Figura 13.** Determinación de distancia épsilon optima

A continuación, en la figura 14, luego de las iteraciones con un valor del parámetro MinPts de 3 a 8, se puede observar el número común de clústeres calculado es de 2.

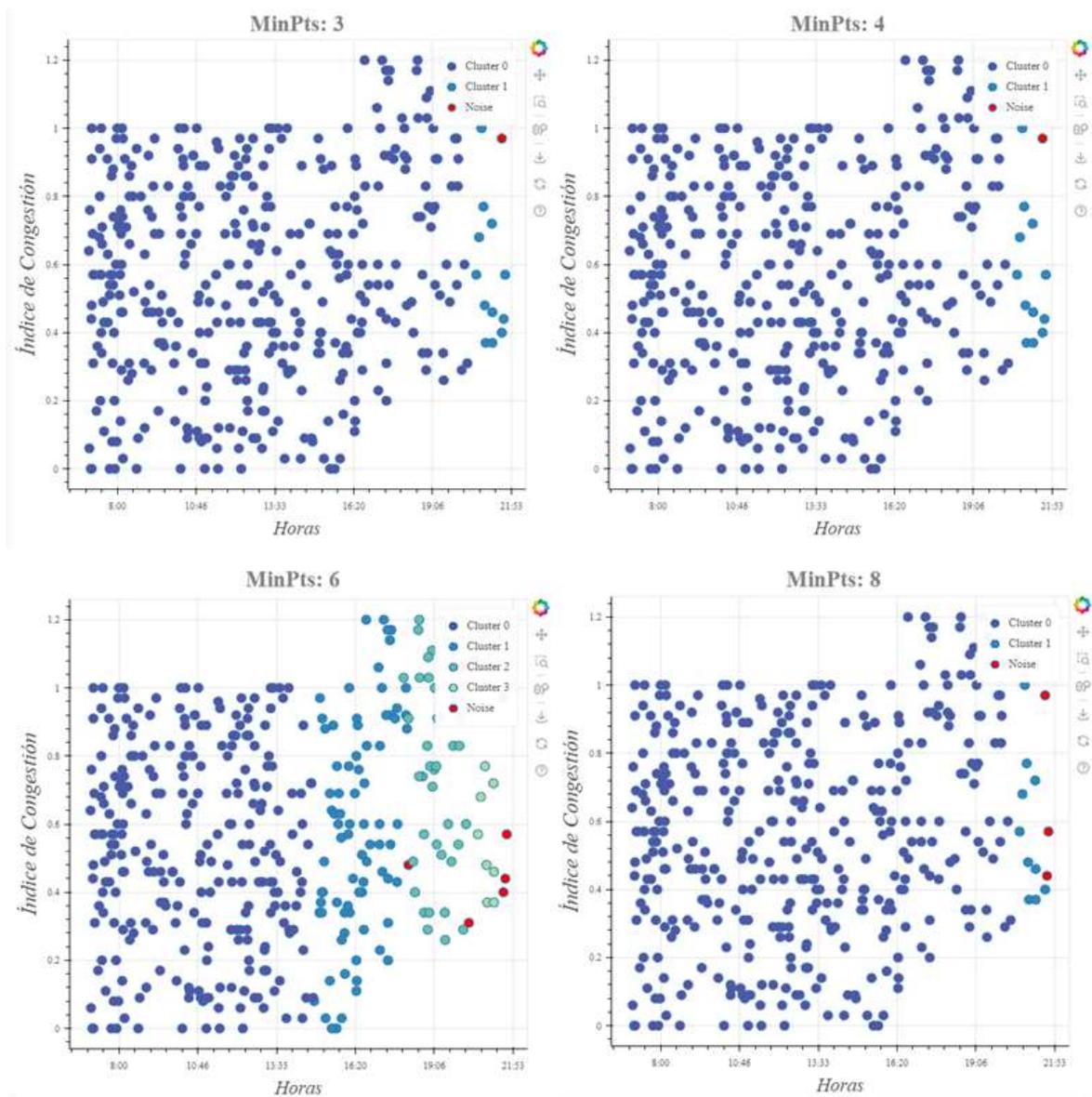


Figura 14. Comparación de Clústeres en relación al número de puntos mínimos en dbscan.

#### 4. CONCLUSIONES

##### Se contrastó el modelo computacional

Utilizando el algoritmo k-means, el valor del promedio Silhouette, se maximiza con 0.6043 con 2 clústers.

Tabla 6. Resultados de evaluación de algoritmos de k-means y dbscan.

Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6	Clúster 7
0,6043	0,5697	0,5635	0,5548	0,5251	0,5125

Con respecto al algoritmo dbscan, el número común de clústers calculado con una distancia  $\epsilon = 0.4$ , es de 2.

El número de clases (grupos) para el dataset evaluado es de 3 (número de horarios punta por turnos: HPM, HPT, HPN), por lo que se puede concluir que los algoritmos k-means y dbscan, han presentado alto grado de bondad y una calidad aceptable en lo que a medidas de similitud se refiere, con respecto al conjunto de datos en lo que fueron aplicados.

**Se logró determinar los índices / niveles de congestión**

El modelo computacional propuesto (figura 6) permitió mostrar los índices / niveles de congestión considerados en la tabla 7. Estos índices son generados como resultados de las condiciones de las variables como: velocidad permitida, tiempo permitido, semáforo, hora punta mañana, hora punta tarde, hora punta noche, cantidad permitida en cada nodo.

Com el modelo computacional se gestiona y controla el flujo vehicular reportando por cada nodo un nivel de congestión, usando las técnicas de agrupamiento k-means y dbscan.

**Tabla 7.** Descripción y valores de los índices de congestión.

Definición Circulación / Demora	Índice de Congestión
Fluida	<0,60
Estable / Ligera	0,61-0,70
Estable / Aceptable	0,71-0,80
Pre-Inestable / Tolerable	0,81-0,90
Inestable, Congestionada / Intolerable	0,91-1,00
Forzada, Congestión Total	>1,00

Fuente: Dirección Metropolitana de Transporte de la ciudad de Trujillo.

- Con el uso del modelo computacional se logró medir el nivel de congestión. Para lo cual, según reporte de las figuras 17 y 18, se obtiene el punto de referencia 18, tal como indica la figura 15.



**Figura 15.** Punto de referencia 18 (Jr. Gamarra – Jr. Grau) del sistema vial de la ciudad de Trujillo, situado por el sistema de software.

Fuente: Google Maps.

Hora	Dataset					
	Etiqueta	Latitud	Longitud	Semaforo	Hora Punta M	Hora Punta T
18	Jr. Gamarra - Jr. Grau	-8,112989	-79,024259	Si	07:00-08:30	11:30-13:30

Hora Punta N	Tope	Velocidad Permitida	Tiempo Permitido	Espacio Permitido	Cant Vehiculos
16:30-21:16	25	40	10	400	23

Velocidad Aprox	Tiempo Aprox	Ind. Congestion
25,0	16	Inestable, Congestionada/Intolerable

nivel de congestión

**Figura 16.** Nivel de congestión detectado en el punto de referencia 18 (Jr. Gamarra – Jr. Grau) en dataset analizado por el sistema de software.

**Explicación de la figura 15:**

Para determinar en qué nivel de congestión se encuentra el nodo 18, se usó los datos de la tabla 17, cuyo índice se calcula mediante la cantidad de vehículos que están ocupando la ruta sobre la cantidad máxima permitida (Tope). En el caso de la figura 15, se obtuvo el nivel de congestión:

$$nivel\ de\ congestion = \frac{Cantidad\ vehiculos}{Tope} = \frac{23}{25} = 0,92$$

Ubicándose este valor en la tabla 11, en el siguiente ítem:

Inestable, Congestionada / Intolerable	0,91-1,00
--	-----------

- La aplicación de modelo computacional en machine learning, a través de los algoritmos k-means y dbscan permitió la clasificación de los agrupamientos o clústeres para determinar el nivel de índices de congestio-namiento vehicular en los diferentes nodos de la ciudad de Trujillo. tal como se demuestra en las figuras 16 y 17 respectivamente, y se especifica en la figura 14.

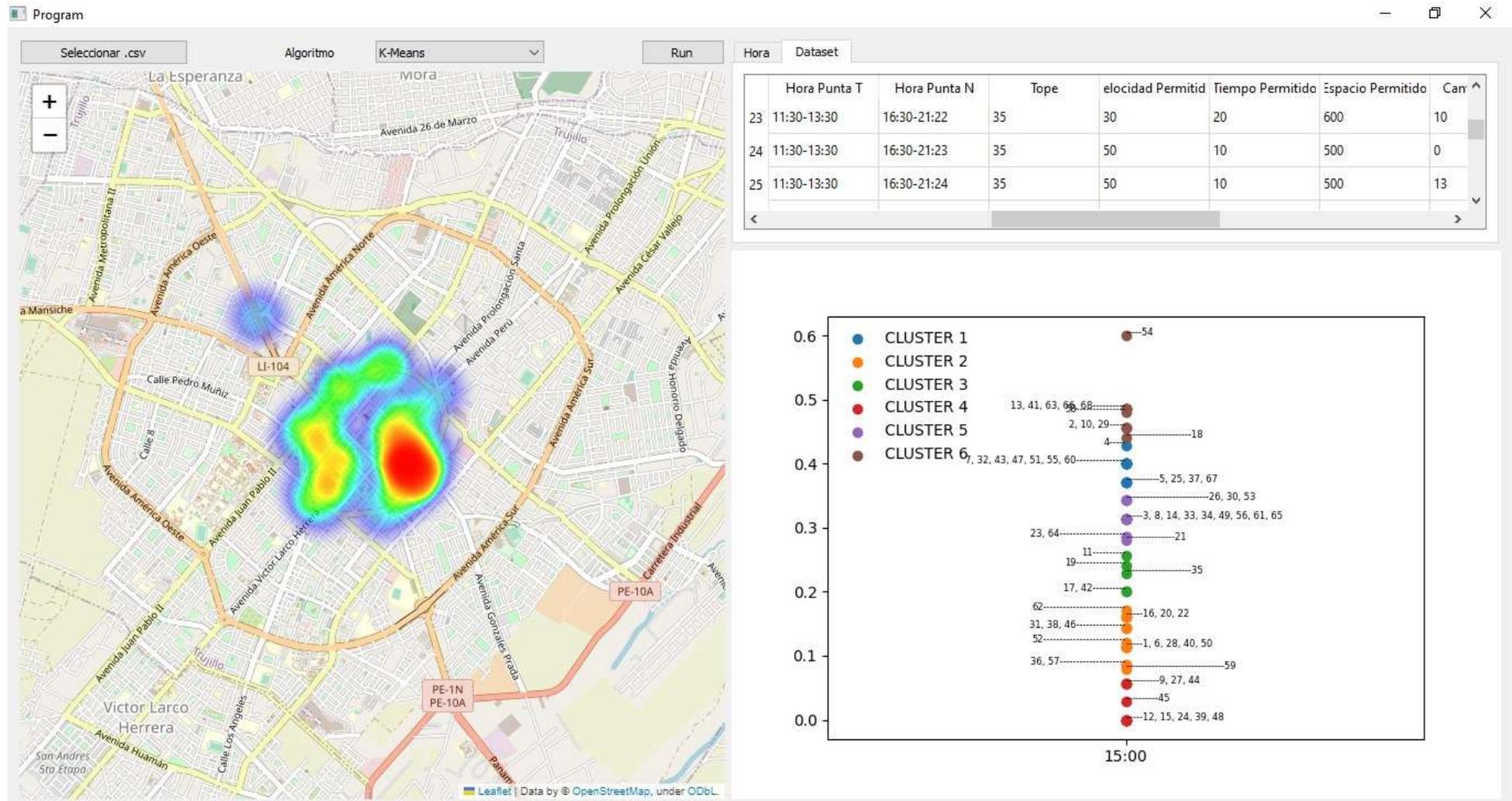


Figura 17. Clasificación de agrupamientos mediante el algoritmo k-means.

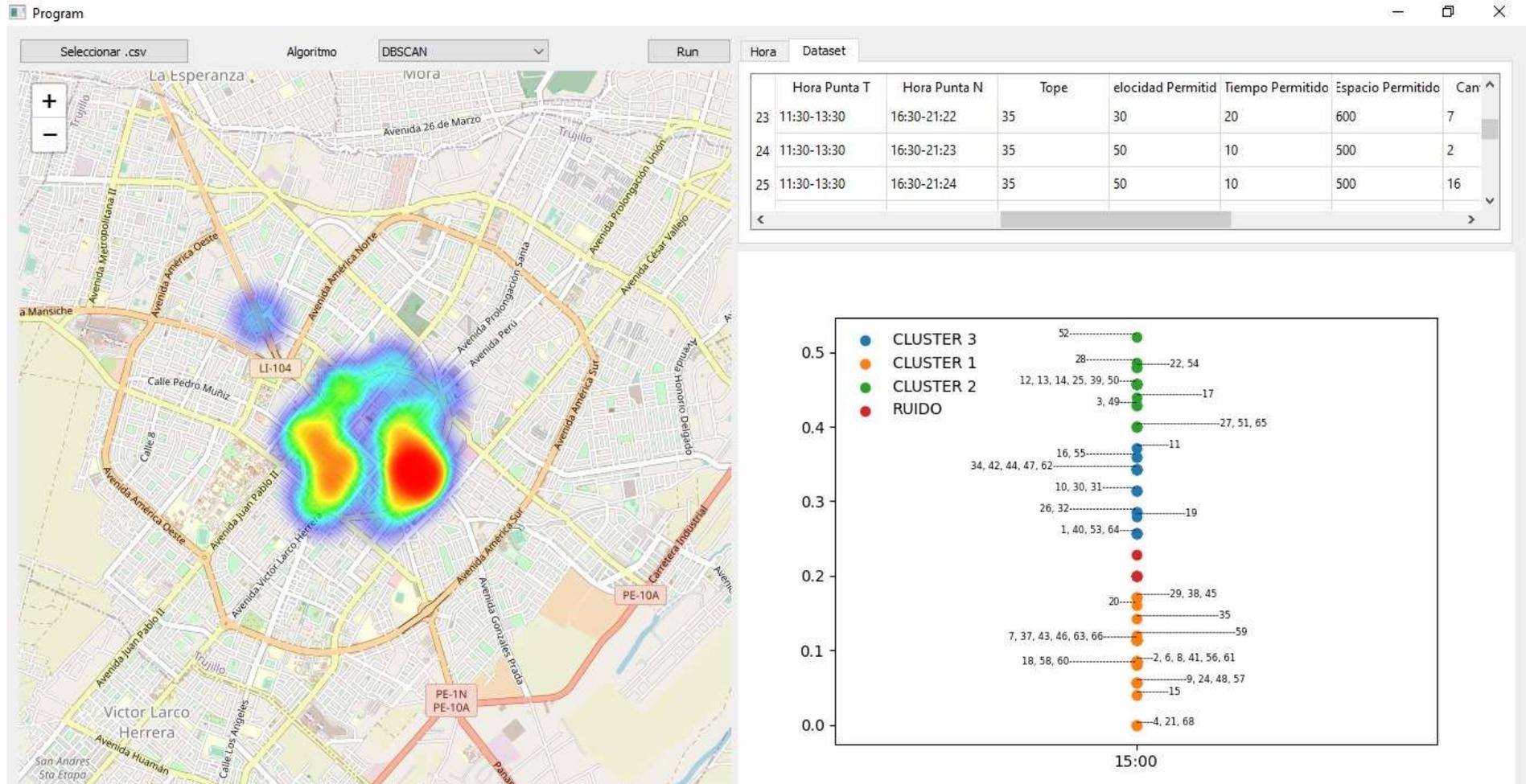


Figura 18. Clasificación de agrupamientos mediante el algoritmo dbscan.

- Se hizo la comparación respectiva del modelo computacional ejecutado con los algoritmos k-means y dbscan, respectivamente, tal como lo indican los resultados de la tabla 6. demostrándose que ambos algoritmos permiten un grado de similitud en sus resultados, notando que ambos algoritmos k-means y dbscan, coinciden cuando tienen en común 2 clústers.

## REFERENCIAS BIBLIOGRÁFICAS

- Alvarenga-Rodríguez, J. M., & Alvarenga-Rodríguez, J. M. (2021). Estudio comparativo entre la velocidad y la densidad en modelos macroscópicos del tráfico vehicular incorporando la variable espacial en la velocidad. *Revista de Matemática Teoría y Aplicaciones*, 28(1), 125-142.  
<https://doi.org/10.15517/RMTA.V28I1.41890>
- Chakraborty, S., Nagwani, N. K., & Dey, L. (2014). Performance Comparison of Incremental K-means and Incremental DBSCAN Algorithms. *International Journal of Computer Applications*, 27(11), 14-18.  
<https://doi.org/10.5120/3346-4611>
- Epstein, J. M. (2020). *Why Model?*
- Espinosa-Zúñiga, J. J., & Espinosa-Zúñiga, J. J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería, investigación y tecnología*, 21(1), 1-13.  
<https://doi.org/10.22201/II.25940732E.2020.21N1.008>
- Geppino Pucci Lorenzo Cazzador, C., & Andrea Pietracaprina, C. (2023). *Studio e implementazione di density clustering DBSCAN approssimato*. <https://thesis.unipd.it/handle/20.500.12608/53312>
- Hernández González, O. (s. f.). *Aproximación a los distintos tipos de muestreo no probabilístico que existen*. Recuperado 2 de abril de 2024, de <http://www.revngi.sld.cu/index.php/mgi/article/view/907>
- Introduction to Computational Science: Modeling and Simulation for the ... - Angela B. Shiflet, George W. Shiflet - Google Libros*. (s. f.). Recuperado 2 de abril de 2024, de <https://books.google.es/books?hl=es&lr=&id=UZApAgAAQBAJ&oi=fnd&pg=PP1&dq=model+computational+computer+science&ots=ktLqVhWSH&sig=vHTJUCRRCZwYCHDKVuj-VApiYHk#v=onepage&q=model%20computational%20computer%20science&f=false>
- Moreira, J. S., León, C. C., Zambrano, G. R., Mercado, C., & Joel, J. (2020). Parámetros que influyen en el congestionamiento vehicular [ Parameters influencing in the vehicular overcrowding ]. *International Journal of Innovation and Applied Studies*, 24(4), 1440-1455. <http://www.ijias.issr-journals.org/>
- Syakur, M. A., Khotimah, B. K., S Rochman, E. M., -, al, Jian, S., Li, D., & Yu, Y. (2021). Research on Taxi Operation Characteristics by Improved DBSCAN Density Clustering Algorithm and K-means Clustering Algorithm. *Journal of Physics: Conference Series*, 1952(4), 042103.  
<https://doi.org/10.1088/1742-6596/1952/4/042103>
- Tan, P.-N., Steinbach, M., & Pearson, V. K. (2019). *Introduction to Data Mining*. [www.pearsoned.co.uk](http://www.pearsoned.co.uk)
- Yang, Q., Yue, Z., Chen, R., Zhang, J., Hu, X., & Zhou, Y. (2019). Real-time detection of traffic congestion based on trajectory data. *The Journal of Engineering*, 2019(11), 8251-8256.  
<https://doi.org/10.1049/JOE.2019.0872>

**ANEXO****PUNTOS DE CONGESTIONAMIENTO EN EL TRÁNSITO VEHICULAR E INTERSECCIONES SEMAFORIZADAS.**

PUNTO	INTERSECCIÓN CON CONGESTIONAMIENTO VEHICULAR	INTERSECCIÓN SEMAFORIZADA	HORA PUNTA CONGESTIONAMIENTO		
			MAÑANA	TARDE	NOCHE
0	Av. España - Atahualpa	Av. España - Atahualpa	07:00 A 08:30	11:30 A 13:30	16:30 A 21:012
1	Av. España - Jr. Estete	Av. España - Jr. Estete	07:00 A 08:30	11:30 A 13:30	16:30 A 21:012
2	Av. España - Ca. Zela	Av. España - Ca. Zela	7:00 A 08:30	11:30 A 13:30	16:30 A 21:02
3	Av. España - Ca. Huayna Cápac	Av. España - Ca. Huayna Cápac	07:00 A 08:30	11:30 A 13:30	16:30 A 21:03
4	Av. España - Atahualpa	Av. España - Atahualpa	07:00 A 08:30	11:30 A 13:30	16:30 A 21:04
5	Av. España - Av. 29 de diciembre	Av. España - Av. 29 de diciembre	07:00 A 08:30	11:30 A 13:30	16:30 A 21:05
6	Av. España - Jr. Pizarro (OR)	Av. España - Jr. Pizarro (OR)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:06
7	Av. España - Independencia (Janos)	Av. España - Independencia (Janos)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:07
8	Av. España - Jr. Bolognesi - Av. Pedro Muñiz	Av. España - Jr. Bolognesi, Av. Pedro Muñiz	07:00 A 08:30	11:30 A 11:30	16:30 A 21:08
9	Av. España - Ca. Salaverry		07:00 A 08:30	11:30 A 13:30	16:30 A 21:09
10	Av. España - Av. Mansiche	Av. España - Av. Mansiche	07:00 A 08:30	11:30 A 13:30	16:30 A 21:010
11	Av. España - Av. M. Vera Enriquez	Av. España - Av. M. Vera Enriquez	07:00 A 08:30	11:30 A 13:30	16:30 A 21:11
12	Av. España - Av. Miraflores	Av. España - Av. Miraflores	07:00 A 08:30	11:30 A 13:30	16:30 A 21:12
13	Av. España - Av. Perú	Av. España - Av. Perú	07:00 A 08:30	11:30 A 13:30	16:30 A 21:13
14	Jr. San Martín - Jr. Almagro	Jr. San Martín - Jr. Almagro	07:00 A 08:30	11:30 A 13:30	16:30 A 21:14
15	Jr. Pizarro Cuadra 1, 2 y 3	Jr. Pizarro Cuadra 1, 2 y 3	07:00 A 08:30	11:30 A 13:30	16:30 A 21:15
16	Jr. Independencia - Jr. Bolognesi	Jr. Independencia - Jr. Bolognesi	07:00 A 08:30	11:30 A 13:30	16:30 A 21:16
17	Jr. Gamarra - Jr. Grau	Jr. Gamarra - Jr. Grau	07:00 A 08:30	11:30 A 13:30	16:30 A 21:17
18	Jr. Bolívar Cuadra 3, 4 y 5		07:00 A 08:30	11:30 A 13:30	16:30 A 21:18
19	Jr. Gamarra - Jr. Bolívar	Jr. Gamarra - Jr. Bolívar	07:00 A 08:30	11:30 A 13:30	16:30 A 21:19

PUNTO	INTERSECCIÓN CON CONGESTIONAMIENTO VEHICULAR	INTERSECCIÓN SEMAFORIZADA	HORA PUNTA CONGESTIONAMIENTO		
			MAÑANA	TARDE	NOCHE
20	Jr. Gamarra - Jr. Ayacucho	Jr. Gamarra - Jr. Ayacucho	07:00 A 08:30	11:30 A 13:30	16:30 A 21:20
21	Jr. Pizarro - Jr. Junín		07:00 A 08:30	11:30 A 13:30	16:30 A 21:21
22	Av. Nicolás de Piérola - Av. Valcárcel	Av. Nicolás de Piérola - Av. Valcárcel	07:00 A 08:30	11:30 A 13:30	16:30 A 21:22
23	Av. América Norte - Av. M. Vera Enríquez	Av. América Norte - Av. M. Vera Enríquez	07:00 A 08:30	11:30 A 13:30	16:30 A 21:23
24	Av. M. Vera Enríquez - (Banco de la Nación)		07:00 A 08:30	11:30 A 13:30	16:30 A 21:24
25	Av. M. Vera Enríquez - Av. 9 de octubre	Av. M. Vera Enríquez - Av. 9 de Octubre	07:00 A 08:30	11:30 A 13:30	16:30 A 21:25
26	Av. América Norte - 8 de octubre (La Hermelinda)	Av. America Norte - 8 de Octubre (las Hermelindas)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:26
27	Av. 8 de octubre - Ca. Pucara (La Hermelinda)		06:30 A 12:30	12:30 A 14:30	
28	Av. América Norte - Av. Miraflores	Av. America Norte - Av. Miraflores	07:00 A 08:30	11:30 A 13:30	16:30 A 21:27
29	Av. América Norte - Av. Perú	Av. América Norte - Av. Perú	07:00 A 08:30	11:30 A 13:30	16:30 A 21:28
30	Av. América Sur - Av. Union	Av. América Sur - Av. Union	07:00 A 08:30	11:30 A 13:30	16:30 A 21:29
31	Av. America Sur - Av. Vallejo	Av. America Sur - Av. Vallejo	07:00 A 08:30	11:30 A 13:30	16:30 A 21:30
32	Av. América Sur - Av. Ricardo Palma	Av. América Sur - Av. Ricardo Palma	07:00 A 08:30	11:30 A 13:30	16:30 A 21:31
33	Av. America Sur - Ca. Santa Cruz	Av. America Sur - Ca. Santa Cruz	07:00 A 08:30	11:30 A 13:30	16:30 A 21:32
34	Av. América Sur - Ca Túpac Yupanqui		07:00 A 08:30	11:30 A 13:30	16:30 A 21:33
35	Av. América Sur - Av. Costa Rica	Av. América Sur - Av. Costa Rica	07:00 A 08:30	11:30 A 13:30	16:30 A 21:34
36	Av. América Sur (UPAO)	Av. América Sur (UPAO)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:35
37	Av. America Sur (Ovalo Larco)		07:00 A 08:30	11:30 A 13:30	16:30 A 21:36
38	Av. América Oeste - Av. Juan Pablo II		07:00 A 08:30	11:30 A 13:30	16:30 A 21:37
39	Av. América Oeste - Av. Antenor Orrego	Av. América Oeste - Av. Antenor Orrego	07:00 A 08:30	11:30 A 13:30	16:30 A 21:38
40	Av. América Oeste (Frente a la Corte)	Av. América Oeste (Frente a la Corte)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:39

PUNTO	INTERSECCIÓN CON CONGESTIONAMIENTO VEHICULAR	INTERSECCIÓN SEMAFORIZADA	HORA PUNTA CONGESTIONAMIENTO		
			MAÑANA	TARDE	NOCHE
41	Av. América Oeste ((Frente al Mall Saga Falabella))		07:00 A 08:30	11:30 A 13:30	16:30 A 21:40
42	Av. América Oeste - Av Mansiche	Av. América Oeste - Av Mansiche	07:00 A 08:30	11:30 A 13:30	16:30 A 21:41
43	Av. Túpac Amaru - Av. Indoamérica		07:00 A 08:30	11:30 A 13:30	16:30 A 21:42
44	Av. Nicolás de Pierola - Av. Indoamérica	Av. Nicolás de Pierola - Av. Indoamérica	07:00 A 08:30	11:30 A 13:30	16:30 A 21:43
45	Av. Villareal - Av. 8 de octubre (Mercado las Hermelindas)	Av. Villareal - Av. 8 de octubre (Mercado las Hermelindas)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:44
46	Av. Villareal - Av. Peru		07:00 A 08:30	11:30 A 13:30	16:30 A 21:46
47	Av. Villareal - Av. Vallejo	Av. Villareal Av. Vallejo	07:00 A 08:30	11:30 A 13:30	16 30 A 21:47
48	Av. Larco - Av. Húsares de Junín	Av. Larco - Av. Húsares de Junín	07:00 A 08:30	11:30 A 13:30	16:30 A 21:48
49	Av. Fátima - Av. Húsares de Junín	Av. Fátima - Av. Húsares de Junín	07:00 A 08:30	11:30 A 13:30	-16:10 A 21:49
50	Av. Fátima - Av. Larco	Av. Fátima - Av. Larco	07:00 A 08:30	11:30 A 13:30	16.30 A 21 50
51	Ca. Agua Marina (Frente a la ULADECH)		07:00 A 08:30	11:30 A 13:30	16:30 A 21:51
52	Av. Perú Cdra. 1, 2, 3 y 4	Av. Perú Cdra. 1, 2, 3 y 4	07:00 A 08:30	11:30 A 13:30	16:30 A 21 52
53	Ca. Rimac (Mercado Unión)		07:00 A 08:30	11:30 A 13:30	16:30 A 21:53
54	Ca. Amazonas - Ca. Rimac		07:00 A 08:30	11:30 A 13:30	16:30 A 21:54
55	Av. Ejército (Frente a la UPN)		07:00 A 08:30	11:30 A 13:30	16:30 A 21:55
56	Av. José María Eugeren - Ca. Balboa		07:00 A 08:30	11:30 A 13:30	16:30 A 21:56
57	Ca. Balboa - Psje. Albarra-cin		07:00 A 08:30	11:30 A 13:30	16:30 A 21:57
58	Ca. Sinchi Roca (Mercado Mayorista)		07:00 A 08:30	11:30 A 13:30	16:30 A 21:58
59	Ca. Sinchi Roca - Ca. Zela		07:00 A 08:30	11:30 A 13:30	16:30 A 21:59
60	Av. La Marina - (Terminal América Expresa)		07:00 A 08:30	11:30 A 13:30	16:30 A 21.60
61	Av. Prolog. Unión (Plaza Veá)		07:00 A 08:30	11:30 A 13:30	16:30 A 21:61
62	Av. Mansiche Mall Aventura Plaza	Av. Mansiche Mall Aventura Plaza	07:00 A 08:30	11:30 A 13:30	16:30 A 21:61

PUNTO	INTERSECCIÓN CON CONGESTIONAMIENTO VEHICULAR	INTERSECCIÓN SEMAFORIZADA	HORA PUNTA CONGESTIONAMIENTO		
			MAÑANA	TARDE	NOCHE
63	Av. Mansiche - Ca. Los Zafiros (Hospital Regional)	Av. Mansiche Ca. Los Zafiros (Hospital Regional)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:63
64	Av. Nápoles - Ca. Salaverry	Av. Nápoles - Ca. Salaverry	07:00 A 08:30	11:30 A 13:30	16:30 A 21:64
65	Av. Juan Pablo II (Puerta Principal UNT)	Av. Juan Pablo II (Puerta Principal UNT)	07:00 A 08:30	11:30 A 13:30	16:30 A 21:65

Fuente: Observatorio de la Movilidad de Transporte de Trujillo