

Mobility Patterns in Telecommunication Operator Customers

Gary Reyes, Msc^{1,2} , Roberto Tolozano-Benites, PhD ¹ ,

Johnny Israel Rodriguez Aviles² , and Leonel Efraín Vega Salazar² ,

¹Carrera de Sistemas Inteligentes, Universidad Bolivariana del Ecuador, Campus Durán Km 5.5 vía Durán Yaguachi, Durán 092405, Ecuador gxreyesz@ube.edu.ec , rtolozano@ube.edu.ec

²Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil, Cda. Universitaria Salvador Allende, Guayaquil 090514, Ecuador, gary.reyesz@ug.edu.ec , johnny.rodriguezav@ug.edu.ec , leonel.vegas@ug.edu.ec

ABSTRACT – *The purpose of this article is to identify urban mobility patterns in the city of Quito, Ecuador, applying clustering algorithms to geographic location and mobile service consumption data of customers during the month of May for the years 2019, 2020, and 2021. The study integrates geolocation data with demographic and consumption information to examine variations in mobility patterns among different population groups and geographical areas. Using a non-experimental and cross-sectional research design, clustering models were developed to classify users with common mobility behaviors and mobile data usage. The results reveal significant differences in mobility patterns influenced by socioeconomic factors and the COVID-19 pandemic, offering valuable insights for telecommunications operators in infrastructure and service planning. The findings underline the importance of flexibility and adaptability in telecommunications services and the crucial role of data analysis in strategic decision-making.*

Keywords: *Urban mobility, Geospatial clustering, Data usage patterns, Big data analysis, Algorithm optimization.*

PATRONES DE MOVILIDAD EN CLIENTES DE OPERADORAS DE TELECOMUNICACIONES

Gary Reyes, Msc^{1,2}, Roberto Tolozano-Benites, PhD¹,

Johnny Israel Rodríguez Aviles², and Leonel Efraín Vega Salazar²,

¹Carrera de Sistemas Inteligentes, Universidad Bolivariana del Ecuador, Campus Durán Km 5.5 vía Durán Yaguachi, Durán 092405, Ecuador gxreyesz@ube.edu.ec, rtolozano@ube.edu.ec

²Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil, Cda. Universitaria Salvador Allende, Guayaquil 090514, Ecuador, gary.reyesz@ug.edu.ec, johnny.rodriguezav@ug.edu.ec, leonel.vegas@ug.edu.ec

ABSTRACT – *The purpose of this article is to identify urban mobility patterns in the city of Quito, Ecuador, applying clustering algorithms to geographic location and mobile service consumption data of customers during the month of May for the years 2019, 2020, and 2021. The study integrates geolocation data with demographic and consumption information to examine variations in mobility patterns among different population groups and geographical areas. Using a non-experimental and cross-sectional research design, clustering models were developed to classify users with common mobility behaviors and mobile data usage. The results reveal significant differences in mobility patterns influenced by socioeconomic factors and the COVID-19 pandemic, offering valuable insights for telecommunications operators in infrastructure and service planning. The findings underline the importance of flexibility and adaptability in telecommunications services and the crucial role of data analysis in strategic decision-making.*

Keywords: *Urban mobility, Geospatial clustering, Data usage patterns, Big data analysis, Algorithm optimization.*

RESUMEN– *El propósito de este artículo es identificar patrones de movilidad urbana en la ciudad de Quito-Ecuador, aplicando algoritmos de clustering a datos de ubicación geográfica y consumo de servicios móviles de los clientes durante el mes de mayo para los años 2019, 2020 y 2021. El estudio integra datos de geolocalización con información demográfica y de consumo para examinar las variaciones en los patrones de movilidad entre diferentes grupos poblacionales y áreas geográficas. Utilizando un diseño de investigación no experimental y transversal, se desarrollaron modelos de clustering para clasificar usuarios con comportamientos comunes en movilidad y uso de datos móviles. Los resultados revelan diferencias significativas en los patrones de movilidad influenciados por factores socioeconómicos y la pandemia de COVID-19, ofreciendo información valiosa para las operadoras de telecomunicaciones en la planificación de infraestructuras y servicios. Los hallazgos subrayan la importancia de la flexibilidad y adaptabilidad en los servicios de telecomunicaciones y el papel crucial del análisis de datos en la toma de decisiones estratégicas.*

Palabras clave: *Movilidad urbana, Agrupación geoespacial, Patrones de uso de datos, Análisis de big data, Optimización de algoritmos.*

I. INTRODUCCIÓN

En la era actual, caracterizada por la generación constante y masiva de datos campos como el Big Data, Machine

Learning y el campo de la minería de datos se han consolidado como pilares fundamentales para comprender y navegar en este vasto océano de información. En este contexto, el presente artículo se adentra en el análisis de datos geoespaciales, una rama de estudio que cobra especial relevancia dada la creciente urbanización y los desafíos que esto conlleva. La investigación se enfoca específicamente en la aplicación y adaptación de los algoritmos de agrupamiento K-Means, DBSCAN y PyClee para el procesamiento de los conjuntos de datos de ubicación y movilidad de los clientes de una Operadora de telecomunicaciones de Ecuador, una metodología emergente dentro del aprendizaje automático no supervisado [15], destinada a desentrañar los complejos patrones de tráfico en zonas urbanas.

A nivel internacional, se ha realizado un estudio sobre patrones de movilidad urbana en la población de Wuhan, China tal como podemos ver en la Referencia [1], La detección de patrones anómalos de movilidad urbana es un tema importante en los estudios de movilidad urbana. En este documento, se propuso un marco para detectar patrones anómalos de movilidad urbana basados en trayectorias GPS [3,4], y datos de Puntos de Interés (POI). Se utilizaron el algoritmo DBSCAN y la técnica de peso TF-IDF ponderado para identificar regiones funcionales; los patrones anómalos de movilidad urbana se diferenciaron de los patrones normales después del agrupamiento de vectores de movilidad.

El análisis de trayectorias GPS ha cobrado una importancia significativa en la era actual debido al volumen masivo de datos generados por dispositivos GPS, ofreciendo una oportunidad única para comprender y optimizar la movilidad urbana. Los algoritmos de clustering, juegan un papel crucial en este ámbito al permitir la identificación de patrones vehiculares, congestionamientos y flujos de tráfico comunes, tal como podemos verlo en las referencias [9,10,11].

En Ecuador también existen estudios sobre los patrones de movilidad más específico en la ciudad de Cuenca como lo podemos evidenciar en la referencia [2]. La distribución territorial condiciona la movilidad en transporte público de las mujeres que residen en la Ciudadela Jaime Roldós, ya que ocasiona que sus viajes sean extensos y sin condiciones óptimas, generando inequidad. Su comprobación se da al

evidenciar que en Cuenca se percibe a la segregación como un problema latente; pues a raíz de tomarlo de forma superficial, se ha invisibilizado la triple vulneración que sufren las mujeres en cuanto a género, territorio y movilidad.

La trascendencia de este estudio radica en su contribución tanto a nivel académico como en su aplicabilidad práctica. En un mundo donde la expansión urbana es una constante, y los retos en términos de movilidad y gestión urbana se multiplican, el análisis exhaustivo de la movilidad de las personas puede proveer información relevante y crítica para el diseño de políticas públicas de movilidad más efectivas, eficientes y sostenibles.

El estudio adopta una metodología científica rigurosa y cuantitativa, valiéndose de datos cuantitativos derivados de observaciones meticulosas para evaluar y explorar los resultados obtenidos, y así abordar la hipótesis planteada. El corazón de esta investigación reside en identificar patrones de movilidad y uso de datos en clientes de una operadora de telecomunicaciones, para lo cual se ha elegido un diseño de investigación no experimental, transversal y descriptivo, específicamente adaptado al ámbito de la movilidad de los clientes y el uso de sus datos móviles. La metodología se caracteriza por su enfoque creativo e inductivo, esencial en la exploración de comportamientos humanos en este contexto. A través de la experimentación y observación de diferentes patrones de comportamiento de clientes de dispositivos móviles, se busca comprender y analizar cómo se generan y utilizan los datos de movilidad, llegando a conclusiones lógicas y fundamentadas científicamente.

El enfoque del estudio está claramente delimitado al análisis de patrones de movilidad en la ciudad de Quito-Ecuador, enfocándose en clientes de una operadora de telecomunicaciones durante el mes de mayo de 2021. El ámbito espacial (Quito), temporal (mayo, 2021) y poblacional (clientes móviles) define la dimensión del estudio. La necesidad de este estudio se hace evidente al considerar indicadores como el crecimiento desplazamiento de las personas, el incremento en tiempos de traslado, la congestión y la planificación vial deficiente. La originalidad del enfoque radica en la aplicación de técnicas de clustering como se puede ver en la referencia [12]. para segmentar y analizar patrones diferenciados entre grupos de usuarios móviles, una perspectiva que, aunque innovadora, es completamente viable dada la disponibilidad de datos y la capacidad analítica requerida.

En cuanto a la metodología específica, se emplearán algoritmos avanzados como K-Means [12], DBSCAN [5]. y PyClee [6], cada uno aportando una perspectiva única en la identificación y análisis de patrones de movilidad:

Patrón 1: Evolución de la Movilidad Urbana Durante la Pandemia COVID-19

Este patrón se enfoca en la evolución de la movilidad urbana en Quito a lo largo de fases clave: antes, durante y después de la pandemia de COVID-19, aplicando algoritmos como K-Means y DBSCAN para agrupar a los residentes según sus comportamientos de movilidad.

Patrón 2: Perfiles de Desplazamiento por Valor de Suscripción de Datos

Se aborda la complejidad de un conjunto de datos multidimensional utilizando K-Means y DBSCAN para agrupar los datos y posteriormente desglosar los resultados en visualizaciones por clúster, destacando la influencia de variables como latitud, longitud, costo y distancia.

Patrón 3: Horarios Pico de Movilidad por Zona Geográfica

Se emplea el algoritmo PyClee, una implementación que carga datos desde un archivo CSV y, tras un cuidadoso preprocesamiento, identifica clústeres durante horas pico específicas, ofreciendo una visión detallada de la movilidad en horarios.

II. METODOS Y MATERIALES

En el vasto campo del análisis de datos, el estudio de patrones emerge como una herramienta crucial, ofreciendo percepciones profundas y acciones concretas derivadas del comportamiento y las tendencias subyacentes en conjuntos de datos complejos. Los patrones, configuraciones reconocibles y recurrentes, son fundamentales para comprender y predecir fenómenos en una variedad de dominios, desde movilidad urbana hasta comportamiento de consumidores. Al aplicar algoritmos de clustering [14] como K-Means y DBSCAN, así como técnicas innovadoras como PyClee, es posible segmentar y analizar datos de manera eficiente, revelando conexiones ocultas y dinámicas que de otra manera permanecerían invisibles. Estos patrones no solo sirven para comprender el presente, sino que también equipan a tomadores de decisiones con la capacidad de anticipar y moldear futuros escenarios, optimizando recursos y mejorando intervenciones en áreas clave como la planificación urbana y la gestión de la movilidad.

La experimentación realizada en el estudio abarca una exploración detallada de patrones de movilidad utilizando algoritmos de clustering como K-Means y DBSCAN, y una innovadora aplicación de PyClee para la detección de horarios pico de movilidad. La metodología consistió en la recolección, preparación y análisis de conjuntos de datos de trayectorias GPS, datos demográficos y costos de planes de datos.

Experimentación con K-Means y DBSCAN:

PATRON 1: Evolución de la movilidad urbana antes, durante y después de la pandemia COVID-19

Se enfocó en la integración de conjuntos de datos que abarcan costos de planes, datos demográficos y trayectorias

GPS [7,8]. Tras una meticulosa limpieza y preparación, se aplicaron algoritmos de clustering para segmentar y analizar la movilidad.

Los pasos de la metodología utilizada para experimentar con el algoritmo K-Means fueron:

-Preparación de Datos: Se consolidaron y limpiaron datos de múltiples fuentes.

-Filtrado Temporal: Se focalizó en las trayectorias de mayo para cada año seleccionado.

-Aplicación y Visualización: K-Means se aplicó a coordenadas geográficas, segmentando en clústeres y visualizando los resultados con marcadores demográficos y límites claros.

-Análisis Descriptivo: Se contabilizaron registros, clasificados por género, y se calcularon porcentajes para ilustrar patrones de movilidad.

Los pasos de la metodología utilizada para experimentar con el algoritmo DBSCAN fueron:

-Configuración y Aplicación: Se aplicó DBSCAN a conjuntos de datos anuales, identificando clústeres y descartando el ruido.

-Visualización y Análisis: Los resultados se visualizaron y compararon con los de K-Means, ofreciendo información importante sobre agrupaciones naturales y patrones de movilidad.

PATRON 2: Segmentación por Valor de Suscripción de Datos y Edad

Se exploró cómo los patrones de movilidad varían según el valor de suscripción de datos y grupos demográficos.

Los pasos de la metodología utilizada para experimentar con el algoritmo K-Means fueron:

-Preprocesamiento y Preparación: Se prepararon los datos, incluyendo segmentación por edad.

-Selección y Escalamiento de Características: Se seleccionaron y normalizaron características relevantes.

-Implementación y Análisis: Se aplicó K-Means, visualizando y analizando los clústeres según la edad y el consumo de datos.

Los pasos de la metodología utilizada para experimentar con el algoritmo DBSCAN fueron:

-Preprocesamiento y Configuración: Se replicó el proceso de K-Means, ajustando DBSCAN a los datos escalados.

-Visualización y Análisis Estadístico: Se presentaron visualmente los clústeres y se realizó un análisis detallado, proporcionando una visión clara de la distribución por grupos de edad y consumo de datos.

PATRON 3: Horarios Pico de Movilidad por Zona Geográfica con PyClee

La experimentación con PyClee ofreció una perspectiva única sobre cómo la movilidad urbana se concentra durante los horarios pico.

-Preparación y Filtrado Temporal y Espacial: Se enfocó en periodos y zonas específicas, filtrando por fechas y horarios pico.

-Preparación para y Aplicación de DyClee: Se ajustaron los datos para DyClee [6], aplicando el algoritmo y visualizando la evolución dinámica de los clústeres.

-Extracción de Centroides y Análisis Cuantitativo: Se extrajeron los centroides para análisis cuantitativos, permitiendo integrar estos datos con otras fuentes para una comprensión más profunda de la movilidad urbana.

Este estudio demuestra un enfoque integral y multifacético para analizar la movilidad urbana, utilizando técnicas avanzadas de clustering y visualización para extraer patrones significativos e información destacada que son prácticos de conjuntos de datos complejos y multidimensionales. La adaptabilidad de estos métodos a diferentes contextos y tipos de datos demuestra su potencial para informar políticas de movilidad y estrategias urbanas.

III. RESULTADOS

Este estudio examina meticulosamente la movilidad urbana en Quito, analizando datos de movilidad de clientes de una operadora de telecomunicaciones, demográficos y de suscripción a servicios móviles para analizar patrones de movilidad en fases distintas del ciclo urbano, incluyendo la era pre, durante y postpandemia del COVID-19. A través de técnicas de clustering como K-Means, DBSCAN y PyClee, se revelan tendencias clave en la movilidad, perfiles de desplazamiento por valor de suscripción de datos y variaciones de movilidad en horarios pico por zonas geográficas. Estos análisis profundos no solo aportan una comprensión detallada de la dinámica urbana y el comportamiento social en Quito, sino que también ofrecen perspectivas críticas para la formulación de políticas públicas adaptativas y basadas en datos. En este estudio, para determinar el número óptimo de clústeres en los algoritmos de clustering, se empleó el método de la silueta tanto para el patrón 1 y 2.

PATRON 1: Evolución de la movilidad urbana antes, durante y después de la pandemia COVID-19

El conjunto de datos para el Patrón 1 abarcó tres años críticos que reflejan la evolución de la movilidad urbana en respuesta a la pandemia de COVID-19. En 2019, previo a la pandemia, se analizaron 1,318,734 registros, proporcionando una línea base de movilidad normal. En 2020, durante el pico de la pandemia, se procesaron 898,248 registros, lo que refleja el impacto directo de las medidas restrictivas en los patrones de desplazamiento. Finalmente, en 2021, se observaron 629,359 registros, ilustrando un período de transición hacia la nueva normalidad o adaptación postpandemia.

El Patrón 1 del estudio aborda la evolución de la movilidad urbana en la ciudad de Quito del año 2021 durante tres fases críticas: prepandemia, durante la pandemia y postpandemia de COVID-19.

Para el análisis prepandemia en 2019, como se muestra en la figura 1 los mapas de K-Means muestran una distribución densa y uniforme de la movilidad, sin restricciones notables, reflejando una situación de normalidad urbana. En contraste, durante la pandemia en 2020, en la figura 2 se observa una reducción significativa en la movilidad, evidenciada por la disminución de puntos y la dispersión de clústeres en ambos mapas, ilustrando el impacto de las medidas de confinamiento y las restricciones de movilidad.

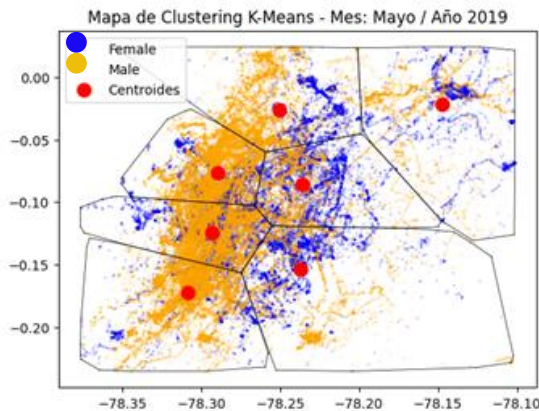


Figura 1 Distribución de Género en Clústeres de Movilidad Urbana: K-Means en Mayo 2019

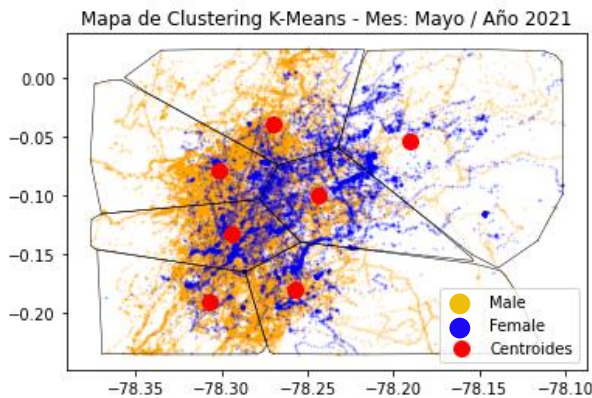


Figura 2 Distribución de Género en Clústeres de Movilidad Urbana: K-Means en Mayo 2020

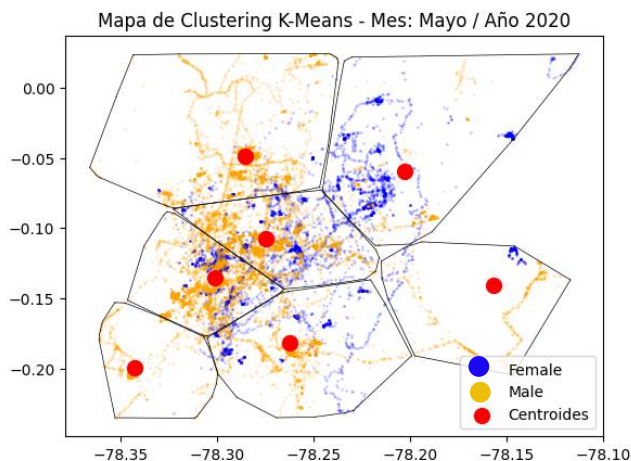


Figura 3 Distribución de Género en Clústeres de Movilidad Urbana: K-Means en Mayo 2021

Para la fase postpandemia o de adaptación en 2021, se percibe en la figura 3 un aumento en la movilidad comparado con 2020, aunque aún no se alcanzan los niveles de 2019. Esto se atribuye a la adaptación a nuevas formas de trabajo y a la reconfiguración de la actividad urbana, lo cual es evidente en la distribución y densidad de los clústeres en los mapas correspondientes.

En resumen, el análisis de este patrón revela una clara evolución de la movilidad urbana en Quito, desde una situación de normalidad prepandemia, pasando por una etapa de restricción significativa durante la pandemia, hasta un periodo de recuperación y adaptación en la fase postpandemia. Los mapas y análisis descriptivos proporcionan una comprensión visual y cuantitativa de estos cambios, ofreciendo una base sólida para futuras investigaciones y la toma de decisiones en planificación urbana y políticas públicas.

ALGORITMO K-MEANS

Tabla 1 Comparativa de Resultados de Clustering algoritmo K-Means en el Análisis de Patrones de Movilidad de Género Durante 2019-2021

Año	Total Registro	Porcentaje Hombres	Porcentaje Mujeres	Porcentaje Movilidad(%)	Cantidad de clusters
2019	1318734	59.51%	40.49%	24.18	7
2020	898248	67.90%	32.10%	18.47	7
2021	629359	75.15%	24.85%	11.54	7

A continuación, se observa en la Tabla 1. Un resumen de los resultados alcanzados con el algoritmo K-means en donde se observa como se determinó que la cantidad óptima de clústeres para cada una de las fases fue de siete, lo que refleja una agrupación consistente de patrones de desplazamiento a pesar de las variaciones en la conducta de movilidad

De la misma manera, se observa en la Tabla 2. Un resumen de los resultados alcanzados con el algoritmo DBSCAN en donde se observa como con el algoritmo DBSCAN resultó en una cantidad variable de clústeres, con once en 2019, veinte en 2020 y doce en 2021, lo que sugiere una fluctuación mayor en la densidad y distribución de la movilidad urbana, probablemente debido a la respuesta

adaptativa de los ciudadanos a las restricciones de movimiento impuestas durante la pandemia.

ALGORITMO DBSCAN

Tabla 2 Comparativa de Resultados de Clustering algoritmo DBSCAN en el Análisis de Patrones de Movilidad de Género Durante 2019-2021

Año	Total Registro	Porcentaje Hombres	Porcentaje Mujeres	Porcentaje Movilidad(%)	Cantidad de clusters
2019	1318734	59.51%	40.49%	24.18	11
2020	898248	67.90%	32.10%	18.47	10
2021	629359	75.15%	24.85%	11.54	12

El estudio sobre la movilidad urbana en Quito ha revelado patrones distintivos asociados con las fases de la pandemia de COVID-19. A través del uso de técnicas de clustering como K-Means y DBSCAN, se obtuvieron información importante sobre cómo la pandemia alteró el flujo y la estructura de la movilidad en la ciudad.

PATRÓN 2: Perfiles de desplazamiento en función del valor de suscripción de datos

En la sección de resultados para el Patrón 2, se destaca el análisis detallado de los patrones de desplazamiento basados en el valor de suscripción de datos, donde se aplicaron dos métodos de clustering para interpretar los conjuntos de datos, los cuales tienen un total de 98082. Con el método K-Means, se procesó un total de 13 clúster, estos variaron en tamaño desde el más pequeño con solo 5 registros hasta el más grande con 44,095 registros. Este enfoque permitió identificar agrupaciones claras y diferenciadas en base a la movilidad y el comportamiento de suscripción.

Por otro lado, utilizando el algoritmo DBSCAN, que es particularmente eficiente para detectar la densidad y aislar datos atípicos, se identificaron 21 clusters, incluyendo uno que agrupa los datos no asignados a ningún otro cluster, marcado con el índice -1 y conteniendo 28,553 registros. Los tamaños de los clusters con DBSCAN mostraron una variabilidad considerable, reflejando la diversidad en los patrones de desplazamiento y suscripción dentro de la población estudiada.

El análisis incorporó múltiples dimensiones de datos, incluyendo latitud y longitud para localización geográfica, coste y distancia de viaje para reflejar el valor de suscripción, así como grupos de edad para contextualizar los datos demográficamente. Esta combinación de variables proporciona una comprensión rica y matizada de cómo los diferentes segmentos de la población se desplazan y suscriben a servicios, ofreciendo perspectivas valiosas para las empresas de servicios móviles y los planificadores urbanos.

El Patrón 2 del estudio aborda los perfiles de desplazamiento en función del valor de suscripción de datos, destacando la complejidad de los datos geospaciales. El

análisis utiliza algoritmos K-Means y DbSCAN para agrupar datos según latitud, longitud, costo y distancia, y se enfoca en los patrones de movilidad de diferentes grupos de edad en Quito. Los datos se visualizan en gráficos individuales por clúster, permitiendo una exploración detallada de las relaciones entre las variables.

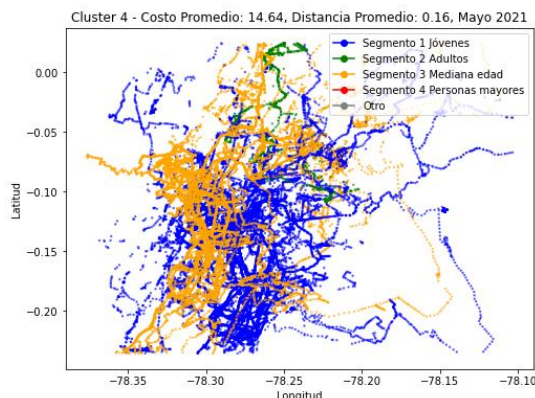


Figura 4 Análisis de Perfiles de Desplazamiento por Grupos de Edad: Cluster 4 en Mayo 2021

Como podemos ver en la figura 4 el algoritmo K-Means revela que el Clúster 4, con la mayor cantidad de registros y una predominancia de adultos jóvenes y adultos de mediana edad, muestra un costo promedio moderado y una distancia promedio baja, indicando desplazamientos locales o uso de datos en lugares fijos. El Clúster 1, con un costo promedio bajo, sugiere sensibilidad al precio o preferencia por planes económicos, y el Clúster 0, con un costo y una distancia promedio superiores, podría incluir usuarios que requieren más datos o prefieren planes con mejores características.

Por otro lado, DbSCAN identifica clústeres dominantes como la figura 5 también para los clústeres 1 y 20, con distribuciones de edad y costos promedio variados, destacando patrones de suscripción y movilidad específicos. En particular, los clústeres 18 y 19, con costos y distancias promedio más altos, sugieren que suscriptores con planes más costosos tienden a desplazarse más.

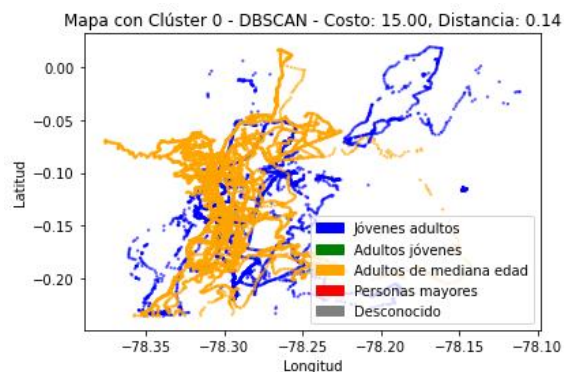


Figura 5 Mapa de Desplazamiento por Edad con DBSCAN: Costos y Distancias en el Cluster 0

En resumen, el Patrón 2 proporciona una comprensión detallada de cómo los perfiles de desplazamiento varían según el valor de suscripción y el grupo de edad en Quito, ofreciendo información valiosa para la toma de decisiones en el sector de servicios móviles y la planificación urbana.

ALGORITMO K-MEANS

Tabla 3 Comparativa de Resultados de Clustering algoritmo K-Means el cual se interpretan los diferentes perfiles de desplazamiento por valor de suscripción de datos

Cluster	Total Registros	% Registros	Costo Promedio	Distancia Promedio (km)
0	14767	0.1506	19.37	0.22
1	19467	0.1985	8.43	0.21
2	210	0.0021	15.84	8.35
3	4542	0.0463	39.68	0.22
4	44095	0.4496	14.64	0.16
5	3662	0.0373	8.48	1.08
6	30	0.0003	16.79	21.22
7	447	0.0046	14.64	5.21
8	6948	0.0708	16.24	0.80
9	2337	0.0238	16.95	1.68
10	5	0.0001	13.16	50.54
11	73	0.0007	17.37	13.35
12	1269	0.0129	13.96	3.07
13	230	0.0023	39.74	2.29

ALGORITMO DBSCAN

Tabla 4 Comparativa de Resultados de Clustering algoritmo DBSCAN el cual se interpretan los diferentes perfiles de desplazamiento por valor de suscripción de datos

Cluster	T. Reg	% Registros	Costo Promedio	Distancia Promedio(km)
-1	28553	0.2911	17.36	0.99
0	18003	0.1836	15.00	0.14
1	18046	0.184	8.40	0.19
2	265	0.0027	8.40	0.55
3	195	0.002	8.40	0.61
4	293	0.003	8.40	0.58
5	112	0.0011	8.40	1.55
6	3525	0.0359	16.00	0.12
7	103	0.0011	16.00	0.22
8	4131	0.0421	18.00	0.14
9	2623	0.0267	20.00	0.11
10	1178	0.012	19.59	0.07
11	121	0.0012	19.59	0.16
12	125	0.0013	19.59	0.12
13	153	0.0016	20.00	0.19
14	157	0.0016	13.44	0.10
15	298	0.003	13.44	0.13
16	299	0.003	13.44	0.07
17	64	0.0007	13.44	0.06
18	2650	0.027	40.00	0.09
19	183	0.0019	40.00	0.22
20	17005	0.1734	14.00	0.15

En el análisis comparativo entre los algoritmos K-Means y DBSCAN tal como podemos ver en la tabla 3 y 4, aplicados al estudio de perfiles de desplazamiento en función del valor de suscripción de datos, se observan diferencias significativas en la segmentación de la movilidad. El algoritmo K-Means, conocido por su simplicidad y eficacia en conjuntos de datos bien definidos y separados, identificó clúster con un número relativamente menor de agrupaciones y una distribución más uniforme en cuanto a la cantidad de registros. Por ejemplo, el clúster más grande representaba aproximadamente el 44.96% de los datos, con un costo promedio de suscripción de \$14.64 y una corta distancia promedio de desplazamiento de 0.16 Km, indicando posibles áreas de alta densidad con una preferencia por suscripciones moderadas y desplazamientos frecuentes en zonas locales.

Por otro lado, DBSCAN, que destaca por su habilidad para identificar clúster basados en densidad y aislar puntos atípicos, produjo una cantidad mayor de clúster con una variabilidad más alta en cuanto a la proporción de registros. El clúster mayor en DBSCAN abarcaba el 29.11% de los registros con un costo promedio de \$17.36 y una distancia promedio de 0.99 Km, lo que sugiere una detección más detallada de grupos con diferentes patrones de movilidad, probablemente influenciados por la variabilidad en el valor de suscripción de datos y la necesidad de desplazamientos.

PATRON 3:

Para el Patrón 3, el análisis reveló la existencia de 6 clúster centrales que encapsulan la movilidad y las tendencias de suscripción dentro de un conjunto de 16,566 registros. La identificación de estos clústeres centrales es fundamental para comprender las dinámicas subyacentes de los patrones de desplazamiento, ya que cada uno representa un núcleo distintivo de actividad y comportamiento de los usuarios

El Patrón 3 del estudio se centra en los horarios pico, (los cuales son desde las 5 p.m. y las 7 p.m.) de movilidad por zonas geográficas, utilizando para ello una implementación detallada y metodológica. La primera fase consiste en la carga y preparación de los datos desde un archivo CSV, seguido por el renombramiento de columnas, la conversión de formatos y la eliminación de filas con datos nulos. Posteriormente, los datos son filtrados por fechas y horas específicas para enfocarse en los horarios pico, una etapa crucial para capturar los patrones más significativos de movilidad urbana.

El algoritmo DyClee se configura con precisión para trabajar con dos dimensiones y un tamaño de hiper caja relativo. Esta configuración permite la identificación efectiva de clústeres de actividad sin la necesidad de normalización previa, lo que garantiza la integridad y representatividad de los datos. Cada clúster detectado por el algoritmo es analizado minuciosamente para extraer los centroides de los

microclusters, puntos clave que representan el 'corazón' de las agrupaciones y reflejan áreas críticas dentro del contexto urbano.

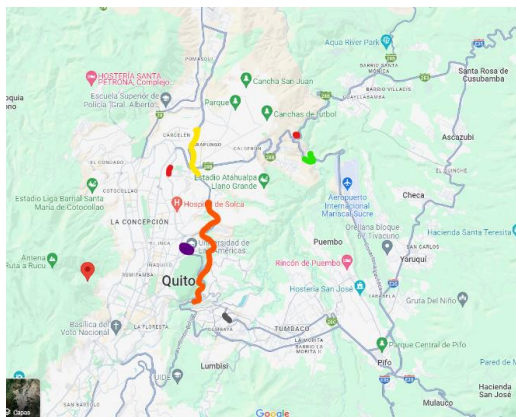


Figura 6. Mapa señalando los clústeres encontrados en la ciudad de Quito.

Los resultados de la implementación revelan patrones intrigantes y áreas de actividad concentrada en la ciudad de Quito. Por ejemplo, se observan en la figura 6 que los clústeres a lo largo de vías principales como la Avenida Simón Bolívar, sugiriendo altas densidades de actividad que podrían corresponder a tráfico vehicular intenso o zonas comerciales activas. Otros clústeres en puntos estratégicos como el Portal Shopping, Carcelén Bajo, o la Universidad de las Américas destacan por su posible significado en términos de flujo de tráfico, interés comercial, residencial, o actividad académica.

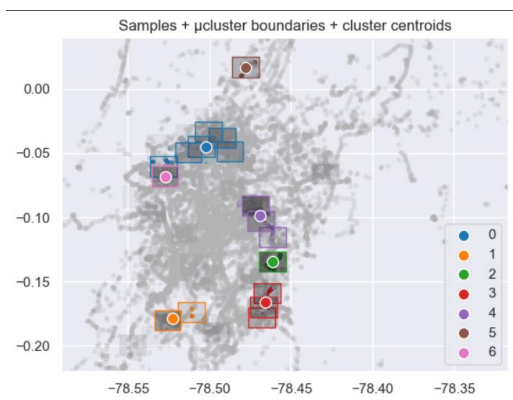


Figura 7. Mapa de Quito con los clústeres centroides señalados.

La representación visual de estos clústeres como se puede observar en la figura 7, con cada uno identificado por un color único y marcado con un símbolo que señala el centroide, proporciona una interpretación clara de las zonas de mayor concentración de datos. Estos centroides son de especial interés, ya que indican áreas de actividad intensa o de importancia estratégica, ofreciendo percepciones valiosas para

el análisis de tráfico, la planificación urbana, o el estudio de comportamientos sociales.

IV. CONCLUSIONES

La integración de datos de geolocalización con información demográfica reveló diferencias significativas en los patrones de movilidad entre distintas zonas de Quito en los periodos antes, durante y después de la pandemia, lo que indica una clara correlación entre la movilidad urbana y factores socioeconómicos.

Los modelos de clustering demostraron ser eficaces para agrupar a los usuarios en función de sus hábitos de movilidad y consumo de datos, facilitando la identificación de patrones específicos de comportamiento en diferentes ubicaciones de la ciudad

Los patrones de desplazamiento identificados ofrecen información crucial para el sector de la telefonía, permitiendo a las operadoras adaptar su infraestructura y servicios a las necesidades cambiantes de movilidad en Quito. Esto resulta en una mejor cobertura, satisfacción del cliente y eficiencia operativa en áreas clave de la ciudad.

Transformación Significativa en Patrones de Movilidad y Consumo de Datos: El estudio revela cómo la pandemia ha transformado de manera integral los patrones de movilidad y consumo de datos en Quito.

Necesidad de Flexibilidad y Adaptabilidad en Servicios de Telecomunicaciones: La variabilidad en los patrones de movilidad y consumo de datos resalta la necesidad de que las operadoras de telefonía móvil sean flexibles y adaptables en sus ofertas de servicio.

Importancia del Análisis de Datos en la Toma de Decisiones: El uso efectivo del análisis de datos es crucial para entender y responder a las cambiantes necesidades de los consumidores, especialmente en un contexto postpandemia.

Conclusiones del Patrón 1: Evolución de la Movilidad Urbana en Quito

Impacto Profundo de la Pandemia en la Movilidad: El estudio revela cómo la pandemia de COVID-19 alteró significativamente los patrones de movilidad en Quito, con una disminución general que refleja las restricciones de movilidad y el cambio hacia el teletrabajo.

Divergencia en la Movilidad Basada en Género: Se observó un incremento en la movilidad entre los hombres y una disminución en las mujeres. Esta tendencia puede estar influenciada por factores socioeconómicos y culturales, donde los hombres posiblemente continuaron en roles laborales que requerían presencia física, mientras que las mujeres pudieron

haber asumido más responsabilidades domésticas o roles que permiten el teletrabajo.

Conclusiones del Patrón 2: Perfiles de Desplazamiento por Valor de Suscripción de Datos

Variabilidad en Preferencias de Consumo de Datos: Existe una clara diferencia en las preferencias de consumo de datos entre varios grupos demográficos. Los jóvenes tienden a preferir planes más económicos, mientras que los usuarios con necesidades de movilidad más altas optan por planes más costosos.

Relación entre Movilidad y Consumo de Datos: El análisis muestra una correlación directa entre la movilidad de los usuarios y sus patrones de consumo de datos, lo que sugiere que los patrones de movilidad influyen en las necesidades de consumo de datos.

Conclusiones del Patrón 3: Horarios pico de movilidad por zona geográfica

El estudio de los patrones de movilidad durante el mes de mayo de 2021, específicamente entre las 5 p.m. y las 7 p.m., ha identificado zonas clave de concentración de usuarios móviles. Los clústeres representados en la visualización demuestran una actividad significativa en ciertas áreas, con núcleos de alta densidad indicativos de movilidad intensa en estos horarios pico. Estos resultados sugieren que hay zonas específicas dentro de la ciudad que experimentan mayores volúmenes de tráfico y, por ende, podrían requerir atención prioritaria en la planificación de infraestructuras y servicios urbanos.

REFERENCIAS

- [1] Z. Xu, G. Cui, M. Zhong, and X. Wang, "Anomalous Urban Mobility Pattern Detection Based on GPS Trajectories and POI Data," *ISPRS International Journal of Geo-Information* 2019, Vol. 8, Page 308, vol. 8, no. 7, p. 308, Jul. 2019, doi: 10.3390/IJGI8070308.
- [2] K. M. Ulloa-Chacha, "INCIDENCIA DE LA DISTRIBUCIÓN TERRITORIAL EN LA MOVILIDAD COTIDIANA DE MUJERES," *Universidad-Verdad*, vol. 2, no. 79, pp. 60–77, Dec. 2021, doi: 10.33324/UV.VI79.454.
- [3] B. J. Arias Martínez and B. S. Zamora Litardo, "ALGORITMO DE CLUSTERING DINAMICO PARA TRAYECTORIA GPS," Universidad de Guayaquil, Guayaquil, 2020.
- [4] V. Llaque and Y. Jenniffer, "ANÁLISIS DE TRAYECTORIAS VEHICULARES GPS PARA EVALUAR SU CALIDAD DE AGRUPAMIENTO UTILIZANDO ALGORITMOS CLUSTERING DE MINERÍA DE DATOS 2018.
- [5] M. Hahsler, M. Piekenbrock, and D. Doran, "dbscan: Fast Density-Based Clustering with R," *J Stat Softw*, vol. 91, pp. 1–30, Oct. 2019, doi: 10.18637/JSS.V091.I01.
- [6] N. Barbosa Roa, L. Travé-Massuyès, and V. H. Grisales-Palacio, "DyClee: Dynamic clustering for tracking evolving environments," *Pattern Recognit*, vol. 94, pp. 162–186, Oct. 2019, doi: 10.1016/J.PATCOG.2019.05.024.
- [7] L. Bazzana Tanevitch et al., "Vehicular Flow Analysis Using Clusters," *Memorias del Congreso Argentino en Ciencias de la Computación - CACIC* 2021, vol. 1, no. 1, pp. 31–40, 2021, Accessed: Jan. 20, 2024. [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/130341>
- [8] F. Villegas, "Relatividad y el Sistema de Posicionamiento Global (GPS) Palabras clave: Relatividad Especial, Relatividad General, Sistema de Posicionamiento Global. Relativity and the Global Positioning System (GPS)," 2020.
- [9] G. Reyes, R. Tolozano-Benites, L. Lanzarini, C. Estrebow, A. F. Bariviera, and J. Barzola-Monteses, "Methodology for the Identification of Vehicle Congestion Based on Dynamic Clustering," *Sustainability*, vol. 15, no. 24, p. 16575, Dec. 2023, doi: 10.3390/su152416575.
- [10] G. Reyes Zambrano, E. S. Z. Meza, O. León-Granizo, and C. C. León, "Evaluation of a data flow clustering algorithm for GPS trajectory analysis," *Ecuadorian Science Journal*, vol. 6, no. 2, pp. 33–45, Sep. 2022, doi: 10.46480/esj.6.2.99.
- [11] G. Reyes, V. Estrada, R. Tolozano-Benites, and V. Maquilón, "Batch Simplification Algorithm for Trajectories over Road Networks," *ISPRS International Journal of Geo-Information* 2023, Vol. 12, Page 399, vol. 12, no. 10, p. 399, Sep. 2023, doi: 10.3390/IJGI12100399.
- [12] Daniel Heras, C. *Biblioteca para la Evaluación Sistemática de Algoritmos de Clúster* 2023 [Universidad Politécnica de Madrid]. https://oa.upm.es/75555/1/TFG_DANIEL_HERAS_CALVO.pdf
- [13] Bobadilla, J. 2020. *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. In J. Bobadilla (Ed.), *RA-MA Editorial* (1st ed., Vol. 0). RA-MA Editorial. https://elibro.net/es/lc/uguayaquil/titulos/222698?fs_q=Machine_Learnin_g&prev=fs
- [14] Lozano, C., Antonio, R., la Fuente Aparicio, D., Jesús Sainz Palmero, M., & Ismael, G. 2020. *Clasificación de fallos con métodos no lineales y algoritmos de agrupación basados en densidad*.
- [15] IBM. (2023). *¿Qué es el aprendizaje no supervisado?* | IBM. <https://www.ibm.com/es-es/topics/unsupervised-learning>