Facultad de Ciencias Económicas y Empresariales
ICADE

ANÁLISIS DE SERVICIOS DE BIKESHARING EN ESTADOS UNIDOS

Autor: Ana Fernández-Valmayor Gallardo
Director: Carlos Miguel Vallez

MADRID | Abril 2025

Resumen

El presente Trabajo de Fin de Grado analiza los patrones de viaje en servicios de *Bikesharing* en Estados Unidos, centrándose en las ciudades de Boston, San Francisco - San José y Chicago. A partir de seis bases de datos públicas correspondientes a los años 2019 y 2023, antes y después de la pandemia del COVID-19, se ha aplicado una metodología ETL para unificar la información en una tabla canónica con el mismo formato y estructura. Posteriormente, se ha calculado la variable velocidad a partir de la distancia y la duración de cada trayecto, con el objetivo de explorar el comportamiento de los usuarios según la ciudad, el año y el tipo de cliente (suscrito u ocasional). El análisis incluye visualizaciones descriptivas de la velocidad media, los trayectos más frecuentes y la distribución horaria de los viajes. Además, se ha desarrollado un *dashboard* interactivo en Power BI que permite explorar los datos de forma visual y dinámica. Este trabajo pone en valor la capacidad de la analítica de datos para obtener información útil a partir de grandes volúmenes de datos.

Abstract

This Final Degree Project analyzes travel patterns in Bikesharing services in the United States, focusing on the cities of Boston, San Francisco - San José, and Chicago. Using six public datasets from the years 2019 and 2023—before and after the COVID-19 pandemic—an ETL process was applied to unify the data into a common table with consistent format and structure. A new variable—speed—was calculated based on trip distance and duration, allowing for a deeper analysis of user behavior by city, year, and customer type (subscriber or customer). The project includes descriptive visualizations of average speed, most frequent routes, and hourly travel distribution. A dynamic dashboard in Power BI was also developed to enable an interactive and visual exploration of the data. This work highlights the power of data analytics to extract valuable insights from large datasets.

Keywords

Bikesharing, BSS, urban mobility, speed, travel patterns, AWS, COVID-19, ETL, Power BI, dashboard, Boston Blue Bikes, Chicago Divvy, Bay Area, San Francisco y San José.

Índice

Índice de tablas, ilustraciones y ecuaciones

Anexo clausula ChatGPT

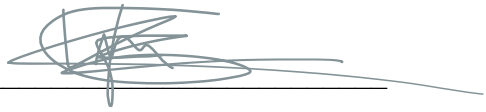
Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Ana Fernández-Valmayor, estudiante de E2-Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Análisis de servicios de Bikesharing en Estados Unidos", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación [el alumno debe mantener solo aquellas en las que se ha usado ChatGPT o similares y borrar el resto. Si no se ha usado ninguna, borrar todas y escribir “no he usado ninguna”]:

1. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
2. **Interpretador de código:** Para realizar análisis de datos preliminares.
3. **Estudios multidisciplinares:** Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
6. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 6 de abril de 2025

Firma: 

1. Motivación para el trabajo de Fin de Grado

En la última década, la sostenibilidad ha tomado un rol crucial en la economía, la cultura y la forma de vida de las personas. La sociedad orienta sus preferencias hacia productos ecológicos, formas de transporte eléctricas e incluso invierten en fondos verdes para impulsar la sostenibilidad.

En el año 2015, los Estados Miembros de las Naciones Unidas adoptaron la Agenda 2030 que establece 17 objetivos llamados Objetivos de Desarrollo Sostenible (ODS), que deben ser perseguidos por los países firmantes para el año 2030 (Russo, 2022). Según Francesco Russo (2022) un aspecto importante para la consecución de estos objetivos es determinar políticas de movilidad para el transporte de personas y mercancías, especialmente en el contexto urbano. La movilidad es un factor fundamental para el desarrollo social y económico de las ciudades, aunque actualmente también representa un desafío importante en términos de impacto ambiental. “En este contexto, el ámbito más prometedor es el de MaaS: Movilidad como Servicio” (p.2).

El MaaS se entiende como la integración de diferentes modos de transporte para ofrecer opciones accesibles y convenientes al usuario, facilitado por tecnologías avanzadas de TIC (Tecnologías de la Información y Comunicación) (Russo, 2022).

Los servicios de *Bikesharing* actuales cumplen varias de estas características por lo que pueden ser considerados un componente del MaaS. Además, reducen el impacto ambiental en las ciudades promoviendo los objetivos de la agenda 2030 y refuerzan la salud y el bienestar de los usuarios.

Por otro lado, este tema ha sido elegido para el TFG desde un punto de vista más personal. La decisión se basa en mi pasión por el análisis de grandes cantidades de datos con el fin de obtener conclusiones útiles para la toma de decisiones. A primera vista, estos conjuntos de datos son poco informativos. Sin embargo, tras un proceso de limpieza y un análisis correcto se puede obtener información muy valiosa acerca de los hábitos de los clientes y sus trayectos.

2. Estado del arte

Para realizar una revisión efectiva de la literatura sobre los servicios de *Bikesharing* o BSS (*BikeSharing Services*) dividiré esta sección en tres partes: la primera, será una revisión histórica de los BSS; la segunda, un análisis de la literatura global publicada sobre BSS desde 2010 y; por último, una revisión específica de los estudios de patrones de viajes en ciudades de Estados Unidos (tema central del trabajo de fin de grado).

Para empezar, merece la pena hacer un breve recorrido en el tiempo para comprender la historia de los BSS.

Según DeMaio (2009) existen 4 generaciones de BSS, sin embargo, otros autores como Chen et al. (2018), introducen una 5ª generación que se caracteriza por uso de bicicletas sin anclaje.

Según DeMaio (2009), el nacimiento de los servicios de *Bikesharing* surge en 1965 en Ámsterdam con las "Witte Fietsen" o bicicletas blancas. Este servicio de 1ª generación se caracterizaba por ofrecer bicicletas blancas al público, permitiendo que cualquiera hiciese uso de ellas para llegar a su destino y luego las dejara disponibles para el siguiente usuario. Sin embargo, el sistema colapsó en pocos días debido a su uso indebido: muchas bicicletas fueron robadas para uso personal o incluso, terminaron en los canales de la ciudad.

Tres décadas más tarde surgieron los servicios de 2ª generación en Dinamarca que ponían a disposición del público 26 bicicletas y 4 estaciones. Este sistema también experimentó problemas de robo por lo que, se determinó que el principal inconveniente era la falta de rastreo. Los servicios de 3ª generación nacieron en Inglaterra en 1996 y contaban con numerosos avances tecnológicos, entre ellos, el uso de una tarjeta magnética para alquilar las bicicletas que hacía las veces de rastreador. Por último, los servicios de 4ª generación se caracterizan por la mejora en la distribución de las bicicletas, la facilidad de instalación de estaciones, mejora de los servicios de rastreo y la asistencia al pedaleo (DeMaio, 2009).

La 5ª generación de BSS introducida por Chen et al. (2018) se caracteriza por el uso de bicicletas sin anclaje y la incorporación de gestión avanzada mediante Big Data. Más

recientemente, Salah IH, et al. (2024) han propuesto una posible 6ª generación de BSS, caracterizada por bicicletas autónomas que serían capaces de redistribuirse automáticamente en función de la demanda. Aunque estos estudios están aún en etapas iniciales, representan un avance prometedor para el futuro de estos servicios.

La segunda sección de la revisión de la literatura se centrará en las principales etapas de la investigación realizada sobre los BSS a partir del año 2010. Este análisis permitirá identificar los temas de estudio más frecuentes, así como aquellos que han recibido menos atención en la literatura.

Según Vallez C.M., et al. (2021) los estudios sobre los sistemas de *Bikesharing* (BSS) han evolucionado a lo largo de cuatro etapas. Entre 2010 y 2012, la investigación se centró en aspectos como la seguridad, las políticas y los modelos de operación. Posteriormente, entre 2013 y 2014, se exploraron los beneficios y el impacto potencial de estos sistemas (Fishman et al., 2013). En 2015, los estudios se enfocaron en la gobernanza y administración de los BSS, comparando modelos público-privados, municipales y privados, además de abordar desafíos clave como la redistribución eficiente de bicicletas y la inclusión de usuarios provenientes del automóvil (Ricci, 2015).

A partir de 2016, la investigación se diversificó, abordando temas como el impacto climático, los patrones de viaje y el comportamiento de los usuarios. En los años más recientes (2019-2020), los estudios han analizado el efecto del COVID-19 en la movilidad urbana, la expansión de los BSS en Asia y el desarrollo de la quinta generación de estos servicios, caracterizada por bicicletas sin anclaje (*Dockless Bikes*), así como los retos en la redistribución de bicicletas para equilibrar la oferta en distintas zonas.

Finalmente, para concluir esta revisión de la literatura, se analizarán los estudios específicos sobre patrones de viaje. Dado que este trabajo de fin de grado se enfoca en el análisis de dichos patrones en tres ciudades de Estados Unidos, se examinará la literatura relevante dentro de este contexto. El objetivo es establecer una línea de investigación clara y diferenciada, evitando la duplicación de estudios previos.

Un estudio significativo para contextualizar este análisis es el de Kou, Z., & Cai, H. (2019), que explora los patrones de viajes en BSS de ocho ciudades de Estados Unidos.

Este trabajo analiza cómo las distribuciones de distancia y duración de los trayectos varían según el tamaño del sistema, la ciudad y el propósito del viaje, distinguiendo entre desplazamientos de usuarios miembros y turistas. Los hallazgos muestran cómo el diseño y la distribución de las estaciones puede influir en la distribución de los viajes, proporcionando información valiosa para el diseño urbanístico y la planificación de BSS. Este estudio analiza variables como la distancia y el tiempo de viaje para posteriormente compararlos.

En el presente trabajo de fin de grado, además del análisis de las variables de distancia y duración, se incorporará la variable de velocidad, lo que permitirá obtener otro punto de vista de los patrones de viaje. El cálculo de esta variable podría facilitar y predecir la identificación de bicicletas que puedan estar más expuestas a un mayor desgaste debido a velocidades elevadas. Asimismo, se examinará el impacto del COVID-19 en los patrones de viaje, utilizando bases de datos de los años 2019 y 2023 para comparar posibles cambios en los comportamientos de los usuarios.

3. Objetivos y alcance

El principal objetivo de este trabajo de fin de grado es analizar los patrones de viaje de servicios de *Bikesharing* en tres ciudades de Estados Unidos antes y después del Covid-19. En particular, se analizarán las ciudades de Boston, San Francisco - San José y Chicago. Los objetivos que se quieren alcanzar son entre otros:

1) Crear una tabla canónica y unificada con las mismas columnas y el mismo formato para las 6 bases de datos que se analizarán.
2) Calcular y estudiar la variable velocidad a partir de la distancia recorrida y la duración del viaje.
3) Comprender los hábitos, comportamientos y necesidades de la población de cada ciudad elegida.
4) Comparar estos resultados antes y después de la pandemia para estudiar su efecto en los patrones de viaje.

Las tres ciudades han sido elegidas por su diferencia geográfica: una ciudad se encuentra en la costa este, otra en la costa oeste y la tercera en el centro del país. Por otro lado, el

alcance temporal abarca los años completos 2019 y 2023 con el objetivo de estudiar el efecto de la pandemia.

4. Estudio demográfico y social de Boston, San Francisco y Chicago

Para entender adecuadamente las diferencias en los patrones de viajes entre las bases de datos, es fundamental comprender la demografía y el clima de cada ciudad.

Ilustración 1: Mapa Boston, San Francisco y Chicago

Fuente: elaboración propia a partir de (Wikipedia, 2007)

La primera ciudad por analizar es Boston. Boston es la capital del estado de Massachussets y se encuentra en la costa este del país, al norte del estado de Connecticut. La zona metropolitana, donde está instalado el servicio de *Bikesharing Blue Bike*s (*metropolitan area),* comprende unos 232,2 kilómetros cuadrados de extensión y una población de 1,032 millones de personas (US Census Bureau, 2023). La media de edad de su población es de 39,7 años y cabe destacar que el 14,4% son jóvenes (personas de entre 20 y 30 años). De hecho, Boston es considerada una ciudad universitaria debido a que hay más de 60 centros universitarios/superiores en la zona metro entre los que destacan *Harvard, MIT* y *Boston University* (Thomson, 2024).

El porcentaje de empleabilidad en Boston ronda el 66% y el tiempo medio de desplazamiento al trabajo es de 31 minutos (US Census Bureau, 2023).

En cuanto a la meteorología, Boston se caracteriza por un clima continental con inviernos fríos y veranos calurosos. La media de temperaturas anuales oscila entre -5ºC y 28ºC. Además, en la ciudad llueve con abundancia y de forma continuada durante el año. Cada año caen aproximadamente 1.162 mm de precipitaciones (Climate Data, 2025).

La segunda zona que se analizará es la que abarca las ciudades de San Francisco y San José. Ambas ciudades se encuentran dentro del estado de California, en la costa oeste de Estados Unidos. San Francisco, en concreto, se caracteriza por su relieve y topografía únicos, con más de 40 colinas que conforman un paisaje urbano caracterizado por calles empinadas y pronunciadas pendientes (Polo, 2024).

La zona en la que está instalado el servicio de *Bikesharing Bay Wheel* conforma las ciudades de San Francisco y San José. Estas áreas abarcan un total de 185 kilómetros cuadrados y tienen una población de 1,89 millones de personas. La media de edad de la población es de 39,4 años y el porcentaje de empleabilidad es de 65,6%. Los trabajadores normalmente tardan en desplazarse a sus lugares de trabajo 29 minutos y la mayoría acuden en su coche particular, aunque el 3,5% lo hacen en bicicleta (US Census Bureau, 2023).

El clima en la zona de San Francisco y San José se caracteriza por temperaturas suaves durante todo el año con poca diferencia entre estaciones. La media de temperaturas anuales oscila entre 5ºC y 21ºC lo que resulta en un clima ni demasiado frío ni demasiado caluroso. La precipitación anual alcanza los 581 mm al año concentrándose en los meses de enero, febrero, marzo, noviembre y diciembre (Climate Data, 2025).

Por último, se analizará la ciudad de Chicago. Chicago es la capital del estado de Illinois y se encuentra en el centro-oeste del país. Chicago es una de las ciudades más importantes de Estados Unidos con una extensión de 265 kilómetros cuadrados y una población de 2,75 millones de habitantes (US Census Bureau, 2023). El servicio de *Bikesharing* de Chicago se llama Divvy.

La media de edad de la población de Chicago es de 36,4 años, de los cuales un 17,4% son jóvenes de entre 20 y 30, además el 45.7% de su población tiene un título igual o superior al grado universitario. La población de Chicago tarda una media de 33 minutos en

desplazarse a sus trabajos y normalmente utilizan sus coches propios (45.5% de la población), mientras que solo el 1.6% emplea la bicicleta según el censo estadounidense (US Census Bureau, 2023).

Al encontrarse al norte del país, Chicago tiene un clima frío con rangos anuales de -5ºC a 20ºC. Las precipitaciones son constantes durante el año y alcanzan los 1.075 mm al año (Climate Data, 2025).

En la siguiente tabla se presenta una comparativa de la información recogida para las tres ciudades.

Tabla 1: Resumen ciudades analizadas

	Boston Metro Area	*San Francisco y San José*	*Chicago*
Nombre BSS	Blue Bikes	Bay Wheels	Divvy
Extensión	232 km^2	185 km^2	265 km^2
Población	1,03 M	1,89 M	2,75 M
Edad media	39,7 años	39,4 años	36,4 años
Educación (grado o superior)	52,2%	60,4%	45.7%
Tiempo desplazamiento	31 min.	29 min	33 min.
Desplazamiento en bicicleta	N/A	3,5%	1,6%
Temperaturas	-5ºC / 28ºC	5ºC / 21ºC	-5ºC / 20ºC
Precipitaciones anuales	1.162 mm	581 mm	1.075 mm

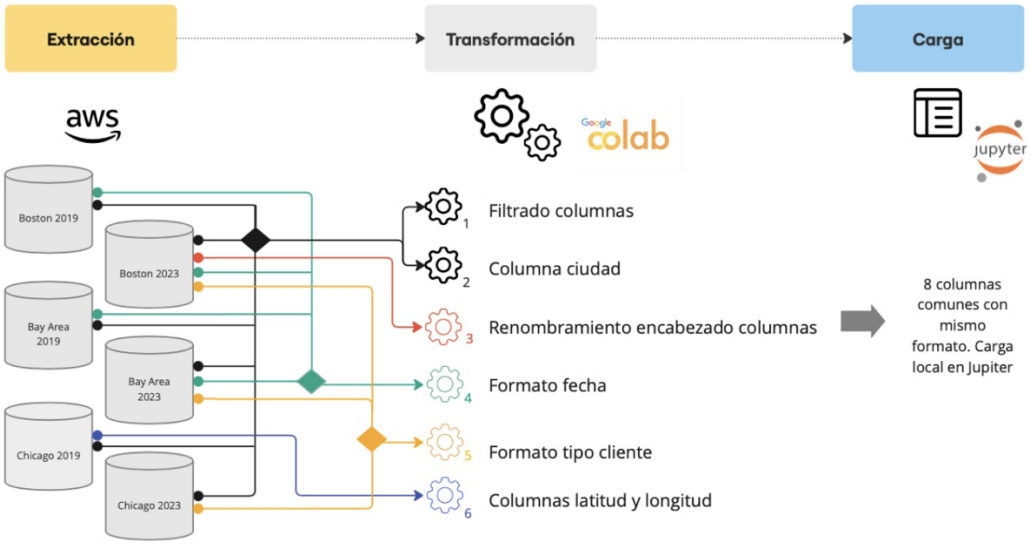
Fuente: elaboración propia a partir de US Census 2023 y Climate Data

5. Metodología: ETL

Este trabajo sigue la metodología ETL que corresponde a las fases extracción, transformación y carga (por sus siglas en inglés: *extraction, transformation y load*). Para integrar datos de distintas fuentes es necesario aplicar una serie de reglas que permitan limpiar y organizar los datos en bruto. Este proceso de ETL facilita la estructuración y unificación de los datos (Amazon Web Services, s.f.).

En primer lugar, se extraerán los datos de los tres servicios de *Bikesharing*. Luego, se llevarán a cabo las transformaciones pertinentes para conformar un modelo canónico que estandarice la información procedente de las distintas bases de datos. Finalmente, los datos organizados se cargarán en el sistema, permitiendo desarrollar visualizaciones y obtener conclusiones. El siguiente mapa conceptual representa los pasos que se llevarán a cabo:

Ilustración 2: ETL Bikesharing



Fuente: elaboración propia

Este proceso será explicado en detalle en cada una de las fases del proceso ETL durante los apartados posteriores.

6. Implementación ETL:

Se implementará la metodología ETL utilizando el lenguaje de programación de Python. Se comenzará utilizando la herramienta de Google Colab para la extracción y las seis transformaciones básicas. Posteriormente, para el cálculo de variables y la comparación de patrones de viaje, el análisis se trasladará a un entorno local en Jupyter Notebook. Este cambio responde a la necesidad de una mayor eficiencia en el manejo y procesamiento de datos, ya que trabajar de forma local ofrece mayor estabilidad y seguridad al tratar con grandes volúmenes de información.

En cuanto a la visualización de los resultados, inicialmente se utilizará la librería Matplotlib de Python para la representación gráfica de los datos. Además, se desarrollará un cuadro de mandos en Power BI, lo que permitirá una exploración más dinámica e interactiva de la información.

6.1. Extracción

En este apartado se explicará primeramente la estructura original o *raw* de las bases de datos encontradas en el servidor de AWS (*Amazon Web Service*), indicando el número de columnas y la tipología del dato que conforma cada una. Seguidamente se explicará el código de Python que extrae los datos del repositorio web.

Todos los datos son de acceso público y se encuentran en repositorios abiertos de AWS. Por lo tanto, el proceso de extracción es común para todas las bases de datos, y el código correspondiente se explicará una única vez al final de esta sección.

6.1.1. Extracción datos Blue Bikes Boston

Blue Bikes es un servicio de *Bikesharing* instalado en la zona metro de la ciudad de Boston. Su funcionamiento se gestiona a través de una aplicación que permite escanear el código de una bicicleta, utilizarla y estacionarla en otro punto de la ciudad. Blue Bikes ofrece varios planes y tarifas entre los que destacan: el pase diario, mensual o anual (Blue Bikes, s.f.).

Los datos de esta plataforma de *Bikesharing* son de acceso público y se almacenan en un repositorio de AWS al que se ha tenido acceso a través del siguiente enlace: https://s3.amazonaws.com/hubway-data/index.html

En esta sección se detallarán el nombre de las columnas y la tipología de datos que conforman los *raw datasets* o conjunto de datos sin procesar encontrados en el repositorio de *Blue Bikes* para los años 2019 y 2023.

Tabla 2: Datos sin procesar Blue Bikes Boston 2019

Tripduration	Starttime	Stoptime	Start station id	Start station name	Start station latitud	Start station longitude	
Float (medido en segundos)	DateTime: aaaa-MM-dd hh:mm:ss. SSSS	DateTime: aaaa-MM-dd hh:mm:ss. SSSS	Int	String	Float	Float	
End station id	End station name	End station lattitud	End station longitude	bikeid	Usertype	Birth year	Gender
Int	String	Float	Float	Int	String (customer o subscriber)	Int	Int (0 o 1)

Fuente: Elaboración propia

Tabla 3: Datos sin procesar Blue Bikes Boston 2023

Tripduration	Starttime	Stoptime	Start station id	Start station name	Start station lattitud	Start station longitude
Float (medido en segundos)	DateTime: aaaa-MM-dd hh:mm:ss. SSSS ó aaaa-MM-dd hh:mm:ss	DateTime: aaaa-MM-dd hh:mm:ss. SSSS ó aaaa-MM-dd hh:mm:ss	Int	String	Float	Float
End station id	End station name	End station lattitud	End station longitude	bikeid	Usertype	Postal code
Int	String	Float	Float	Int	String (customer o subscriber) (casual o member)	Int

Fuente: Elaboración propia

A grandes rasgos ambas tablas son parecidas, sin embargo, se puede observar que la tabla correspondiente al año 2023 omite información sobre el género y el año de nacimiento del usuario. Esto se debe a la ley de protección de datos estadounidense, que prohíbe a los negocios compartir información personal y sensible de los clientes (Department of Justice, 2020). Esta misma omisión se repite en las tablas analizadas posteriormente.

Por otro lado, llama la atención las diferencias en formato de las columnas de fecha: *starttime* y *stoptime*. En algunas observaciones, los datos incluyen los milisegundos, mientras que en otras únicamente se registran los segundos. Esta inconsistencia será abordada más adelante en una transformación específica de corrección del formato fecha (transformación 4) en la que se establecerá un formato de fecha estándar para homogeneizar todos los datos y garantizar la consistencia de las tablas.

Por último, cabe destacar que existe una particularidad a la hora de recolectar los datos durante el año 2023. A partir del mes de mayo de 2023 los encabezados de la tabla cambian su nombre. Esto supone un trabajo extra de identificación y renombramiento de

columnas que se abordará en la transformación de renombramiento de encabezados de columnas (transformación 3).

Además del cambio en los encabezados de las columnas, los valores de la columna "tipo de cliente" también varía a partir de mayor del 2023. Pasan de ser *subscriber/customer* a *member/casual.* Esto probablemente se deba a que se cambiaron los estándares de recolección de datos, lo que provocó el renombramiento de columnas y valores. Esto supondrá otra transformación adicional para corregir el formato de la columna de tipo de cliente (transformación 5). Todas las transformaciones se explicarán en detalle más adelante.

6.1.2. Extracción datos Lyft Bay Area

Lyft es una famosa plataforma estadounidense que ofrece diversos servicios de movilidad, siendo su servicio de vehículos con conductor (VTC) uno de los más populares y el principal competidor de Uber en Estados Unidos (Fernández, 2019). En 2008, la empresa lanzó *Urban Solution* en Canadá, una propuesta que apuesta por las bicicletas de uso compartido como solución a la movilidad en las grandes ciudades. En 2010 este servicio llegó a EE. UU. y tres años más tarde a las costas de California (Lyft Urban Solutions, 2025).

Bay Wheels es el servicio de *Bikesharing* que ofrece Lyft para la zona de San Francisco y San José en el estado de California. Los viajes que se realizan en estas zonas se recogen en un repositorio web AWS de acceso público al que se accede a través del siguiente enlace: https://s3.amazonaws.com/baywheels-data/index.html

Las siguientes tablas detallan las columnas y la tipología de datos que conforman la base de datos sin procesar de *Bay Wheels* para los años 2019 y 2023:

Tabla 4: Datos sin procesar Bay Wheels Bay area 2019

Duration_sec	Start_time	End_time	Start_station_id	Start_station_name	Start_station_latitude	Start_station_longitude
Float (medido en segundos)	DateTime: aaaa-MM-dd hh:mm:ss. SSSS ó aaaa-MM-dd hh:mm:ss	DateTime: aaaa-MM-dd hh:mm:ss. SSSS ó aaaa-MM-dd hh:mm:ss	Int	String	Float	Float
End_station_id	End_station_name	End_station_latitude	End_station_longitude	Bike_id	User_type	Bike_share_for_all_trip
Int	String	Float	Float	Int	String (customer o subscriber)	String (Yes or No)

Fuente: Elaboración propia

Tabla 5: Datos sin procesar Bay Wheels Bay area 2023

ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name
String	String (classic bike o electric bike)	DateTime: aaaa-MM-dd hh:mm:ss. SSSS ó aaaa-MM-dd hh:mm:ss	DateTime: aaaa-MM-dd hh:mm:ss. SSSS ó aaaa-MM-dd hh:mm:ss	String	String	String
end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual	
String	Float	Float	Float	Float	String (member o casual)	

Fuente: Elaboración propia

Ambas tablas son similares, sin embargo, se vuelve a omitir información sensible de los usuarios en la tabla del año 2023 con el fin de garantizar la protección de datos de los clientes.

Por otro lado, al igual que en los datos de Boston, existe una inconsistencia en el formato de las fechas que requerirá una homogeneización en etapas de transformación posteriores (transformación 4).

Por último, también será necesario homogeneizar los valores de los tipos de clientes. Como se puede comprobar en el año 2019, los clientes se subdividen en *customer* o *subscribers,* mientras que, en el 2023, se desglosan en *member* o *casual.* En transformaciones posteriores se homogeneizará los tipos de clientes para que sean comparables (transformación 5).

6.1.3. Extracción datos Divvy Lyft Chicago

Divvy Bikes, también propiedad de la plataforma Lyft, es un servicio de *Bikesharing* establecido en la zona centro de la ciudad de Chicago, estado de Illinois. Los datos de los viajes realizados se encuentran en un repositorio AWS público extraído a partir de este enlace: https://divvy-tripdata.s3.amazonaws.com/index.html

Las siguientes tablas detallan las columnas y la tipología de datos que conforman la base de datos *raw* de *Divvy Bikes* para los años 2019 y 2023:

Tabla 6: Datos sin procesar Divvy Bikes Chicago 2019

trip_id	start_time	end_time	bikeid	tripduration	from_station_id
Int	DateTime: aaaa-MM-dd hh:mm:ss	DateTime: aaaa-MM-dd hh:mm:ss	Int	Float	Float
from_station_name	**to_station_id**	**to_station_name**	**usertype**	**Gender**	**Birthyear**
String	Int	String	String (customer o subscriber)	String (Male o Female)	Int

Fuente: Elaboración propia

Tabla 7: Datos sin procesar Divvy Bikes Chicago 2023

ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name
String	String (classic bike o electric bike)	DateTime: aaaa-MM-dd hh:mm:ss	DateTime: aaaa-MM-dd hh:mm:ss	String	String	String
end_station_id	**start_lat**	**start_lng**	**end_lat**	**end_lng**	**member_casual**	
String	Float	Float	Float	Float	String (member o casual)	

Fuente: Elaboración propia

Igual que en las tablas anteriores, en el año 2023 se omite información sensible sobre los clientes por motivos de regulación en la privacidad de los datos sensibles. Además, como en la ciudad San Francisco, se repite la inconsistencia en los nombres de los tipos de

clientes. Esta homogeneización se abordará en secciones de transformación posteriores (transformación 5).

Por el contrario que las bases de datos anteriores, en el caso de Chicago, sí que existe una consistencia en el formato de las fechas por lo que no se tendrá que corregir el formato de estas columnas.

Sin embargo, al analizar en detalle la información se identifica un problema mayor. En el año 2019 no se recolectó información sobre la latitud y longitud de las estaciones. Solo se recoge el identificador de la estación inicial y final. Esto supone un inconveniente ya que más adelante se calculará la distancia Haversine de cada viaje y será imposible si no existen datos de posición geográfica. En las secciones siguientes de transformación se detallará cómo se ha logrado recuperar esta información mediante la extracción de una tabla relacional que vincula el identificador de cada estación con sus respectivas coordenadas geográficas, permitiendo así su utilización posterior (transformación 6).

6.1.4. Proceso de extracción de datos de un web service

Dado que todas las bases de datos se encuentran en un repositorio de AWS (*Amazon Web Service*), en este apartado se explicará el proceso de extracción común a todos los repositorios. El código utilizado será el mismo para las seis bases de datos.

El proceso de extracción desde un servidor en línea como es AWS comienza con la definición de la URL del AWS dónde se encuentran los datos (URL definida en cada sección de extracción). Posteriormente, se obtiene el contenido Web y se convierte en información estructurada XML gracias a la librería de *BeautifulSoup.*

El contenido estructurado se carga en una variable que se filtra para obtener los nombres de los archivos con terminación ".ZIP". Se extraen los nombres de los archivos con terminación ".ZIP" disponibles en el servidor y se seleccionan solo los nombres de los archivos que contienen el año seleccionado. En el caso de querer analizar el año 2019, se obtienen todos los archivos que contengan esta fecha en su nombre. Lo mismo ocurre para el año 2023.

Aquí finaliza la sección de extracción. En la siguiente sección se describirán en profundidad las transformaciones pertinentes para obtener una tabla canónica con el mismo número de columnas y el mismo formato.

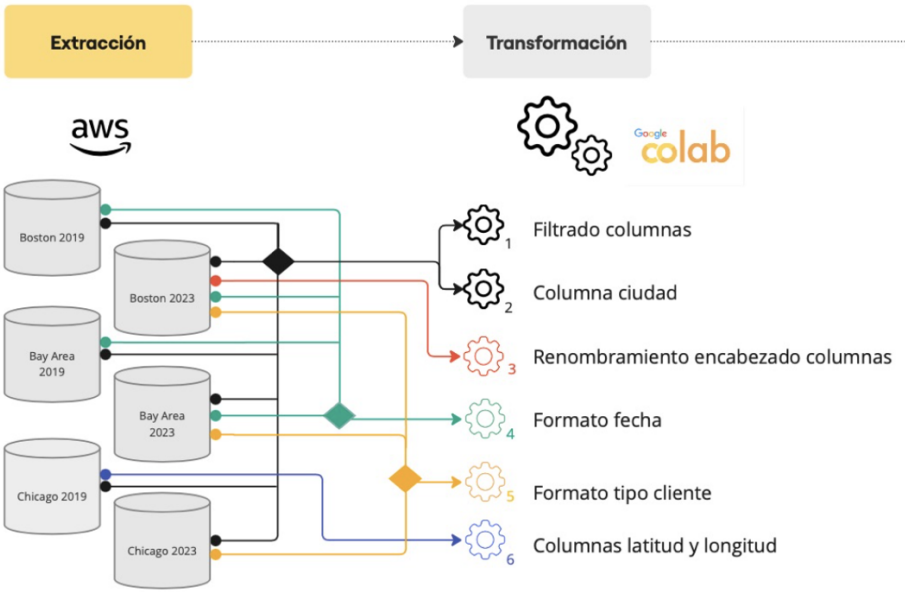
6.2. Transformación

La sección de transformación es la segunda en la metodología ETL y la más extensa. Es el momento de realizar las transformaciones necesarias a las bases de datos sin procesar para homogeneizar la información.

El objetivo es obtener una tabla con información común para los seis repositorios de datos. Para ello es necesario que cada tabla se transforme para llegar a una versión común. Es importante recordar que no todas las bases de datos necesitan las mismas transformaciones. Como se identificó en la sección de extracción, hay tablas que ya tienen un formato correcto, por lo que, se mantendrán.

El siguiente diagrama indica las seis transformaciones que se llevarán a cabo en el apartado de transformación y qué repositorios experimentarán tales transformaciones:

Ilustración 3: Transformaciones pertinentes a cada base de datos



Fuente: Elaboración propia

6.2.1. Transformación 1: Filtrado de columnas

La primera transformación que se llevará a cabo es el filtrado de columnas relevantes a partir de los *datasets* sin procesar. En este caso, el código se encarga de procesar los datos en bruto extraídos de AWS para crear tablas consolidadas con las columnas necesarias para el análisis. El filtrado de columnas tiene como objetivo reducir la complejidad de los datos, eliminando las columnas irrelevantes. Esta transformación se realiza para las seis bases de datos.

El proceso comienza definiendo una lista de las columnas que se conservarán del *dataset* original extraído previamente. Estas columnas han sido seleccionadas cuidadosamente siendo las esenciales para la consecución de los objetivos de este trabajo. Las columnas seleccionadas son siete (aunque en la transformación posterior se añade una octava en relación con la ciudad del servicio):

1. Hora de inicio del viaje
2. Hora de fin del viaje
3. Latitud de inicio de viaje
4. Longitud de inicio de viaje
5. Latitud de fin de viaje
6. Longitud de fin de viaje
7. Tipo de usuario.

La selección de estas columnas responde a la necesidad de capturar únicamente la información que permite estudiar los patrones de viaje, el comportamiento de los usuarios y el cálculo de la velocidad.

Cabe resaltar que cada base de datos sin procesar tiene un nombre concreto para cada una de estas siete columnas seleccionadas. Por ejemplo, en los datos originales de Boston se le llama “stoptime” a la columna que indica la hora de fin de viaje y, sin embargo, en Bay Area la misma columna recibe el nombre de “end_time”. Por este motivo, para tener claro el nombre de las columnas a mantener, se ha explicado en profundidad la nomenclatura y el tipo de dato en el apartado de extracción.

Debido a que los datos se encuentran archivados mensualmente en AWS, el código realiza doce iteraciones en las que procesa cada archivo mensual correspondiente al año seleccionado. El archivo extraído, que en su formato original contiene numerosas columnas, es filtrado para conservar únicamente las 7 columnas seleccionadas en la lista anterior. Este enfoque, aparte de descargar la información esencial, reduce el tiempo de ejecución y de procesamiento.

6.2.2. Transformación 2: Creación de campo ciudad

En la segunda transformación se creará una columna denominada *city*. Esta transformación tiene como objetivo incluir la ciudad a la que pertenece cada registro. La columna *city* será especialmente significativa al unir los seis *datasets* en una única tabla, ya que permitirá diferenciar los datos de distintas ciudades garantizando que la información de los viajes no se mezcle. Esta transformación afecta también a las seis bases de datos.

La implementación de la creación de un nuevo campo se lleva a cabo añadiendo al *dataframe* un nuevo campo llamado *city*, que puede recibir los siguientes valores: Boston, San Francisco & San José o Chicago. Esto se realiza mediante la siguiente instrucción: df ['city'] = "Nombre ciudad". Este código inserta una columna adicional llamada *city* en la tabla para que cada fila incluya de forma explícita el nombre de la ciudad asociada al conjunto de datos.

Incluir esta columna será fundamental para el análisis posterior, ya que permitirá consolidar la información en una única tabla sin que se mezcle la ubicación de los patrones de viaje.

6.2.3. Transformación 3: Renombramiento encabezado columnas

Como se mencionó al describir los datos sin procesar de *Boston Blue Bikes*, este conjunto de datos presenta una inconsistencia en los nombres de las columnas: a partir de mayo de 2023, los encabezados de las columnas cambian su nombre, aunque se mantiene la información. Esta variación podría deberse a la implementación de nuevos estándares en

la recolección de datos a mediados de 2023, lo que habría llevado a *Blue Bikes* a modificar la forma en que registra la información.

Si se aplicase el código de filtrado de columnas, aparecería un error a partir de mayo debido a que no identificaría los nombres de los encabezados de las columnas por el cambio de nombre. A continuación, se muestra una tabla que ilustra cómo cambian los nombres de los encabezados a partir de mayo de 2023:

Tabla 8: Renombramiento de columnas Boston 2023

De enero a abril 2023	De mayo a diciembre 2023
starttime	started_at
stoptime	ended_at
start station latitude	start_lat
start station longitude	start_lng
end station latitude	end_lat
end station longitude	end_lng
usertype	member_casual

Fuente: Elaboración propia

Para resolver esta disparidad, se modificará el código para detectar automáticamente si un archivo pertenece al período "enero-abril" o "mayo-diciembre" y filtrar por los nombres correctos según corresponda.

De esta manera, si el archivo pertenece a un mes entre enero y abril, el código extraerá las columnas con los nombres definidos en la lista *columns_early_2023*, que incluye *starttime, stoptime, start station latitude*, etc. Si, por el contrario, el mes se encuentra entre mayo y diciembre se extraerá la información de las columnas definidas en la lista *columns_late_2023,* que corresponde a los encabezados de los archivos en esos meses.

De esta forma, el código filtrará y extraerá la información de las columnas requeridas correctamente.

6.2.4. Transformación 4: Corrección formato fecha

En esta sección se transformarán los datos de las columnas de inicio y fin de viaje con el fin de homogeneizar el formato de fecha. Esta transformación afecta a las siguientes bases de datos: Boston 2019, Bay Area 2019 y Bay Area 2023.

Tal como se introdujo en la sección de extracción, existen bases de datos que registran el formato tiempo hasta los milisegundos. Esta información es demasiado precisa y no es necesaria para el análisis. Por este motivo y con el fin de homogeneizar el formato fecha, se procederá a eliminar la información relacionada a los milisegundos.

La implementación de esta transformación se llevará a cabo estableciendo el siguiente formato de fecha estándar: aaaa-MM-dd hh:mm:ss, eliminando cualquier información adicional como los milisegundos.

Se ha podido comprobar en la sección de extracción que existen inconsistencias en los formatos de fecha de varias tablas escogidas. Dado que en este trabajo es suficiente registrar el tiempo hasta las unidades de segundos, se procesarán aquellos datos que incluyan información adicional, ajustándolos al formato estándar seleccionado: aaaa-MM-dd hh:mm:ss.

El código utilizado para esta transformación identifica cadenas en las columnas de fecha que contengan un punto (“.”) seguido de números, ya que esta estructura representa los milisegundos. Por ejemplo, en el formato: “2019-01-31 17:57:44.6130”, el código reemplaza todo lo que aparece después del punto por un vacío convirtiéndolo en “2019-01-31 17:57:44”. Si no se encuentra un punto, no se realiza ninguna modificación.

De esta manera se consigue estandarizar el formato de las fechas y garantizar la uniformidad de los datos.

6.2.5. Transformación 5: Corrección formato tipo de cliente

Como se adelantó en la sección de extracción, varios *datasets* categorizan a los usuarios en términos de "member" y "casual". Sin embargo, otros utilizan los términos "Subscriber" y "Customer" para representar la misma información. Ambos nombres aportan el mismo significado para el análisis: "member" o "Subscriber" se refiere a un usuario que está suscrito al sistema mientras que "casual" o "Customer" se refiere a un usuario que utiliza el servicio de forma ocasional sin estar suscrito.

Para la correcta identificación de los datos es necesario transformar los valores a un único formato estándar. Se elegirá "Subscriber" y "Customer". Esta transformación afecta a Boston 2023, Bay Area 2023 y Chicago 2023.

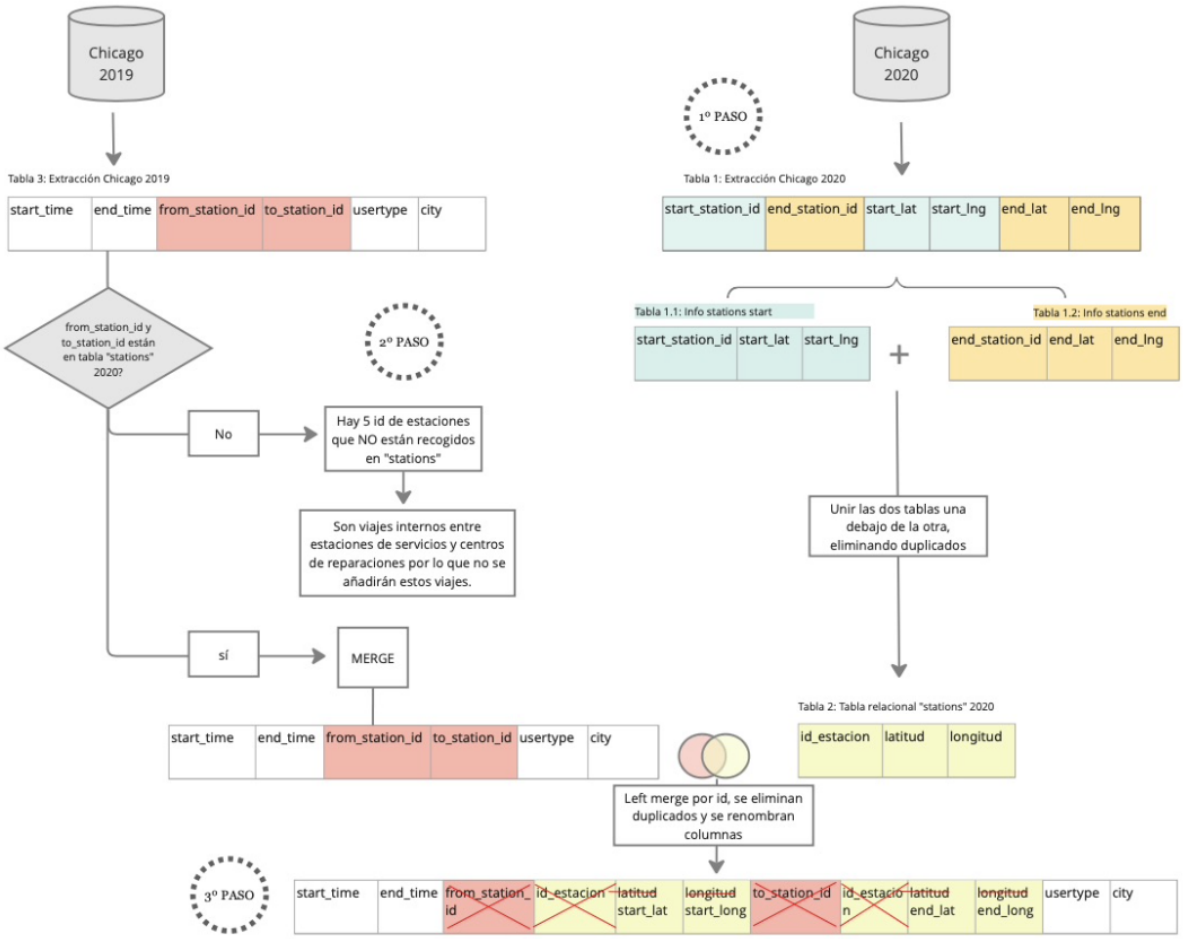
Esta corrección se implementará sustituyendo "member" por "Subscriber" y "casual" por "Customer" a partir del método *replace* de la biblioteca de pandas. Estas instrucciones permiten reemplazar cada aparición de un valor por otro. De esta manera, la columna de tipo de usuario queda estandarizada en todas sus filas.

Está estandarización es fundamental para proporcionar integridad a las tablas, además permitirá desglosar y filtrar la información por tipo de usuario.

6.2.6. Transformación 6: Creación latitud y longitud

La última transformación se realizó solo sobre la base de datos de Chicago 2019. Al analizar los datos se descubrió una limitación: la base de datos para el año 2019 no recoge información geográfica de las estaciones, en su lugar, los datos solo incluyen la información sobre los identificadores de las estaciones. Para solucionar este problema se adoptó el siguiente enfoque:

Ilustración 4: Proceso adición de latitud y longitud para Chicago 2019



Fuente: Elaboración propia

Primeramente, se identificó que la base de datos del año 2020 (año posterior al analizado) contenía información geográfica de cada estación. A partir de esta, se extrajeron los datos correspondientes a las estaciones de inicio y fin de los viajes, incluyendo su identificador, latitud y longitud. Esa información se consolidó en una única tabla relacional en la que cada identificador correspondía a una única latitud y longitud para todas las estaciones registradas en 2020. Esta tabla relacional se denominó “stations” y recoge la información geográfica de todas las estaciones utilizadas en el año 2020. “Stations” fue seleccionada como fuente de referencia para recuperar la información faltante en los datos de años anteriores.

A continuación, se analizaron los identificadores de las estaciones de 2019 con el objetivo de detectar posibles ausencias en la tabla relacional. Se identificaron cinco estaciones

cuyos *IDs* no estaban registrados en la tabla relacional "stations" 2020. Estos 5 *IDs* corresponden a estaciones de uso interno del servicio, por ejemplo: *divvy cassette repair mobile station* o *special event.* Al ser estaciones de uso interno y poco recurrentes, no se incluirán en el análisis de trayectos.

Para finalizar el proceso se hizo un *merge* por la izquierda por "*id_stations*" entre la tabla de viajes 2019 y la tabla relacional "stations" 2020. El objetivo es asignar a cada estación del año 2019 sus coordenadas geográficas correspondientes, ya que esta información no aparece en la tabla original. El proceso consistió en realizar un *left merge*, en el que a cada ID de estación (tanto de inicio como de fin) se le añadieron su correspondiente ID, latitud y longitud procedente de "stations" 2020. Posteriormente, se eliminaron las columnas duplicadas para mantener un formato limpio y estructurado. Como resultado se obtuvo una tabla con 8 columnas que se adaptará al formato de la tabla canónica.

6.3. Carga

Una vez se ha terminados los procesos de extracción y transformación comienza la última fase de la metodología ETL: la carga (*load*). Los procesos previos tenían como objetivo estandarizar los datos para obtener seis tablas con un formato unificado y el mismo número de columnas. Como resultado, estas seis tablas contienen el mismo formato estructurado en las siguientes ocho columnas. Esta tabla muestra el nombre de las 8 columnas y el formato unificado y estándar de los valores de cada variable:

Tabla 9: Información tabla canónica

	Columna 1	**Columna 2**	**Columna 3**	**Columna 4**	**Columna 5**	**Columna 6**	**Columna 7**	**Columna 8**
Nombre	starttime	stoptime	startlat	startlong	endlat	endlong	usertype	city
Descripción	Inicio de viaje	Fin de viaje	Latitud estación inicio	Longitud estación inicio	Latitud estación fin	Longitud estación fin	Tipo usuario	Ciudad
Formato de la variable	DateTime: aaaa-MM-dd hh:mm:ss	DateTime: aaaa-MM-dd hh:mm:ss	Float	Float	Float	Float	String: Subscriber o Customer	String: Boston, San Francisco & San Jose o Chicago

Fuente: Elaboración propia

Dado el gran volumen de datos a procesar, esta fase y las posteriores se llevarán a cabo en un entorno local utilizando Jupyter Notebook, lo que permitirá una manipulación más eficiente de los datos.

El proceso comienza con la descarga de los seis conjuntos de datos trabajados en las fases anteriores, que presentan el mismo formato de ocho columnas con idéntico formato de variables. Los archivos CSVs descargados son los siguientes:

- Boston 2019
- Boston 2023
- Bay Area 2019
- Bay Area 2023
- Chicago 2019
- Chicago 2023

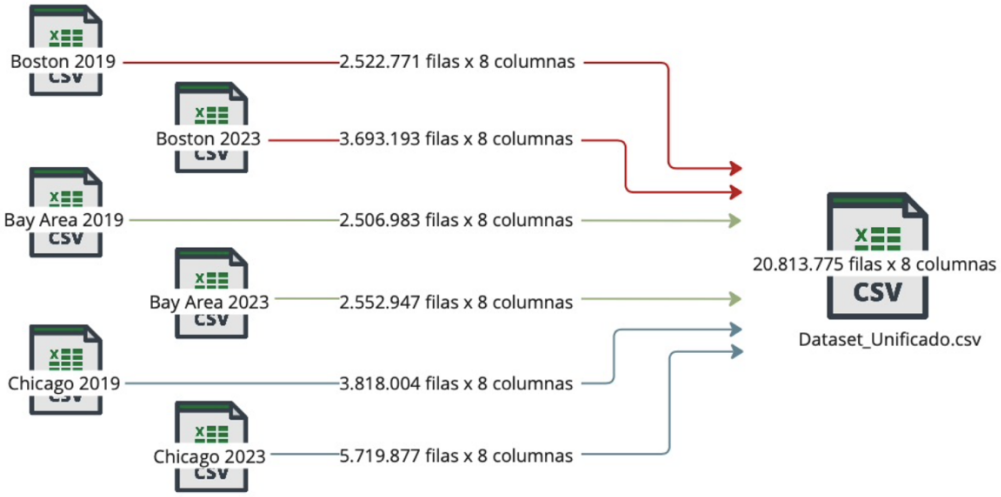
Una vez descargados, se procede a la unificación de los seis archivos en un único *dataset* mediante el uso de pandas en Jupyter Notebook. El resultado es un archivo consolidado con 8 columnas, que integra la información de los seis conjuntos de datos originales.

Para garantizar la correcta ejecución del proceso de unificación, se lleva a cabo una verificación: primero, se suman las filas de cada conjunto de datos por separado y, posteriormente, se compara el total con el número de filas del archivo final. Tras cerciorarnos de que el resultado es correcto se obtiene como resultado un CSV denominado "Dataset_Unificado", que cuenta con 8 columnas y 20.813.775 filas, abarcando la información de todas las ciudades y períodos analizados.

Debido al alto consumo de tiempo y memoria que requiere este proceso, la carga y unificación de datos solo se ejecutará una única vez. Una vez generado el archivo unificado, este se descarga en el escritorio, permitiendo su acceso directo sin necesidad de repetir el procedimiento completo de unificación cada vez que se quiera correr el código.

Con este paso, se da por cumplimentado el primer objetivo: la creación de una tabla canónica con la información unificada de seis conjuntos de datos distintos. Este resultado facilita el análisis y visualización posterior de los patrones de viaje.

Ilustración 5: unificación y carga



Fuente: Elaboración propia

7. Cálculo de la variable velocidad

Una vez finalizado el proceso de extracción, transformación y carga (ETL), y obtenida la tabla canónica que unifica los seis conjuntos de datos en un formato común, se da paso a la siguiente fase del trabajo. En esta nueva etapa se abordará otro de los objetivos principales del trabajo: el cálculo de la variable velocidad en km/h. Incorporar esta variable permitirá enriquecer el análisis, ofreciendo una visión distinta y más precisa sobre los patrones de movilidad de los usuarios antes y después de la pandemia.

La creación de una tabla canónica facilita este proceso, ya que el cálculo de la velocidad se reduce a una simple operación: dividir la distancia recorrida (en kilómetros) entre la duración del viaje (en horas), siguiendo la fórmula estándar de la velocidad:

Ecuación 1: Fórmula de la velocidad

$$\frac{Distancia\ recorrida\ en\ el\ viaje\ (km) \rightarrow procedente\ de\ startlat, startlong, endlat\ y\ endlong}{Duración\ del\ viaje\ (h) \rightarrow procedente\ de\ starttime\ y\ stoptime}$$

Recordemos que la tabla canónica recoge información clave para este cálculo. Registra, por un lado, la posición geográfica de la estación de inicio y fin del viaje, así como la hora de inicio y fin. Para determinar la distancia recorrida en cada trayecto, se aplicará la fórmula de Haversine, que permite calcular la distancia en línea recta entre dos puntos geográficos a partir de sus coordenadas de latitud y longitud. Por otro lado, la duración del viaje se obtendrá calculando la diferencia entre la hora inicio y fin, expresada en horas. Estos dos procesos serán analizados en detalle en los apartados siguientes.

7.1. Distancia Haversine

Empezaremos calculando el numerador de la fórmula de la velocidad: la distancia en kilómetros. Como se venía anunciando anteriormente, la tabla canónica recoge la latitud y longitud de las estaciones de inicio y fin por cada trayecto. La manera más sencilla de calcular la distancia a partir de esta información es a través de la fórmula de Haversine.

La distancia Haversine calcula la distancia en línea recta entre dos coordenadas geográficas en una esfera, es decir calcula la distancia más corta entre dos puntos situados sobre un objeto esférico, como la Tierra. Aunque la Tierra no es una esfera perfecta, esta aproximación se usa por su simpleza y su similitud a la realidad (Nathan, 2016).

La fórmula se descompone en los siguientes cálculos:

Ecuación 2: Fórmula Haversine

$$a = sin^2\left(\frac{\Delta\emptyset}{2}\right) + \cos(\emptyset_1) \cdot \cos(\emptyset_2) \cdot sin^2\left(\frac{\Delta\lambda}{2}\right)$$

$$c = 2 \cdot arctan2\left(\sqrt{a}, \sqrt{1-a}\right)$$

$$d = \mathrm{R} \cdot \mathrm{c}$$

Donde:

$\Delta\emptyset = \text{latitud final} - \text{latitud inicial}$

$\Delta\lambda = logitud\ final - longitud\ inicial$

$R = 6.373\ (radio\ de\ la\ Tierra\ en\ km)$

Esta fórmula tiene como salida la distancia en kilómetros entre la estación de inicio y fin. Esta variable llamada "*Distance*" se añadirá como la novena columna de la tabla canónica para recoger la información sobre la distancia recorrida por viaje medida en kilómetros.

Antes de finalizar este paso, es fundamental asegurarnos de que la fórmula no genere valores negativos, ya que una distancia negativa carece de sentido y significaría que la fórmula de Haversine no funciona. Para ello, se realizará una prueba que identifique y muestre los registros con distancias negativas. Dado que el resultado no muestra ningún registro con valores negativos, podemos continuar calculando el denominador de la fórmula de la velocidad.

7.2. Duración de viaje

Para calcular la duración de los viajes, se ha utilizado la información de las columnas starttime (hora de inicio de viaje) y stoptime (hora de finalización). La diferencia entre estos dos valores permite obtener la duración del trayecto.

Dado que ambas columnas están registradas en el formato aaaa-MM-dd hh:mm:ss, ha sido necesario transformar la diferencia entre ellas en una unidad de tiempo manejable. Para ello, se ha empleado la *función .dt.total_seconds()*, que convierte el resultado en segundos. Posteriormente, para expresar la duración en horas, se divide el valor obtenido entre 3.600, lo que permite un futuro cálculo de la velocidad en unidades de kilómetros por hora (*km/h*).

Antes de finalizar, como en el paso anterior, es importante comprobar que la fórmula no genera valores negativos, ya que el tiempo no puede expresarse en valores inferiores a cero. Para ello, se identifican e imprimen los registros con una duración negativa. En este caso, se detectan 451 registros con un tiempo menor que cero. Dado que un viaje no puede finalizar antes de haber comenzado, estos registros serán eliminados. Probablemente estas anomalías se deban a errores en el sistema de recolección de datos.

Tras cerciorarnos de que el cálculo de la duración es correcto, se almacena como la décima columna llamada "*Duration*" en nuestra tabla canónica.

7.3. Análisis descriptivo de variable velocidad

Una vez calculados el numerador y denominador, se genera la undécima columna de la tabla canónica llamada "*Speed_kmh*" que recoge el cociente de la distancia entre la duración del viaje, es decir la velocidad medida en kilómetros por hora.

Hagamos un resumen de cómo quedaría nuestra tabla canónica con nombre Dataset_Unificado después de estos cálculos.

Tabla 10: Información tabla canónica tras cálculo velocidad

	Columna 1	Columna 2	Columna 3	Columna 4	Columna 5	Columna 6	Columna 7	Columna 8	Columna 9	Columna 10	Columna 11
Nombre	starttime	stoptime	startlat	startlong	endlat	endlong	usertype	city	Distance	Duration	Speed_kmh
Descripción	Inicio de viaje	Fin de viaje	Latitud estación inicio	Longitud estación inicio	Latitud estación fin	Longitud estación fin	Tipo usuario	Ciudad	Distancia en km	Duración en horas	Velocidad en km/h
Formato de la variable	DateTime: aaaa-MM-dd hh:mm:ss	DateTime: aaaa-MM-dd hh:mm:ss	Float	Float	Float	Float	String: Subscriber o Customer	String: Boston, San Francisco & San Jose o Chicago	Float	Float	Float

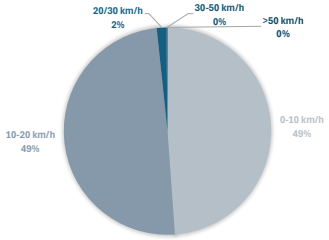
Fuente: Elaboración propia

Antes de comenzar a analizar los resultados y representar gráficamente las conclusiones, se revisará la variable velocidad para identificar posibles *outliers* o errores.

En primer lugar, se realizará un conteo de registros según rangos de velocidad, clasificando las filas en los siguientes intervalos: 0 a 10 km/h, 10 a 20 km/h, 20 a 30 km/h, 30 a 50 km/h y más de 50 km/h. El objetivo de este análisis es identificar los rangos más representativos y frecuentes en la muestra, y así depurar aquellos valores atípicos o extremos que no aportan valor significativo al estudio. Los resultados obtenidos se presentan en la siguiente tabla.

Tabla 11: Número de registros por rangos de velocidad

	Núm. de registros	% del total
0-10 km/h	10156894	48.8%
10-20 km/h	10283702	49.4%
20/30 km/h	308776	1.5%
30-50 km/h	16366	0.1%
>50 km/h	21724	0.1%
	20813775	**100.0%**



Fuente: Elaboración propia

Gracias a este análisis, se concluye que el 99,8% de los registros se encuentran en los rangos de velocidad comprendidos entre 0 y 30 km/h, lo que indica que estos valores explican prácticamente la totalidad de los datos. Por ello, se procederá a filtrar el *dataset* y eliminar aquellos registros con velocidades superiores a 30 km/h, ya que representan tan solo el 0,2% del total y corresponden a valores poco razonables dentro del contexto analizado. Esto genera un total de 20.749.372 de viajes con velocidades menores a 30 km/h.

Teniendo el *dataset* filtrado según este criterio, se realiza un análisis descriptivo de la variable velocidad obteniendo una velocidad media de 9.8 km/h con una desviación típica de 4.5 km/h. Estos resultados parecen razonables al tratarse de velocidades alcanzadas por bicicletas de uso urbano.

Durante los siguientes apartados se analizarán distintas variables del *dataset* ya depurado: *dataset* filtrado, para incluir únicamente registros con velocidades menores a 30 km/h.

8. Análisis de resultados y visualizaciones

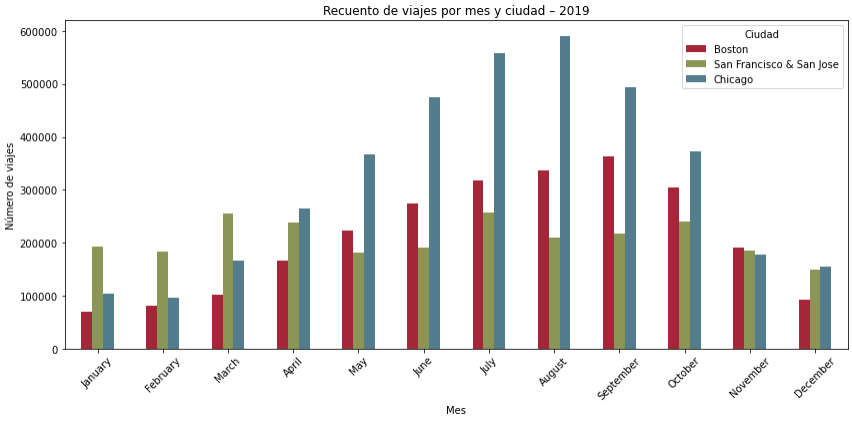
Para comenzar el análisis y contextualizar los resultados empecemos recordando que se han registrado un total de 20.749.372 trayectos con velocidades menores a 30 km/h. Como primer paso, se representará la evolución mensual del número de viajes por ciudad, con el objetivo de identificar qué ciudad concentra un mayor volumen de trayectos a lo largo del año y cuál es su distribución mensual. Por otro lado, con el fin de obtener una primera visión sobre el perfil de los usuarios, se analizará la distribución porcentual entre usuarios suscritos y ocasionales en cada ciudad. Esto permitirá identificar qué ciudades

registran una mayor proporción de usuarios fidelizados frente a aquellos que utilizan el servicio de forma esporádica.

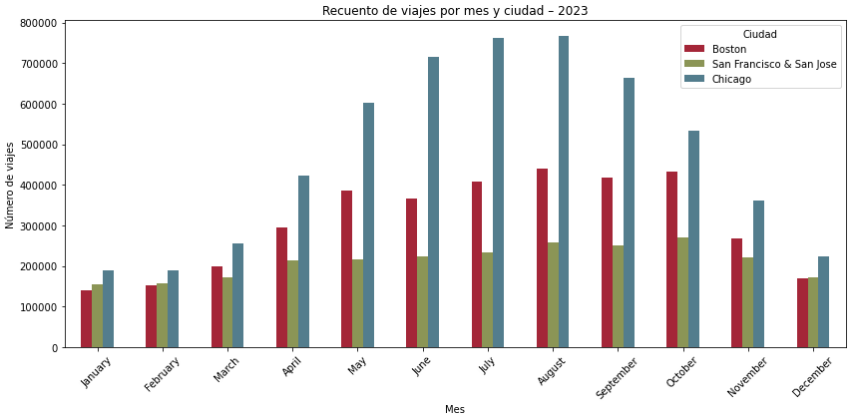
En un análisis preliminar, destaca el notable incremento en el número total de viajes, que pasa de 8,8 millones en 2019 a casi 12 millones en 2023, lo que representa un aumento del 34% en cuatro años. Este crecimiento puede estar influenciado, al menos en parte, por el impacto de la pandemia del COVID-19. Tras las restricciones de movilidad y confinamientos vividos en 2020, muchos ciudadanos comenzaron a valorar alternativas de transporte más saludables, sostenibles y, sobre todo, más seguras para reducir el riesgo de contagio. La necesidad de mantener la distancia social impulsó a una parte significativa de la población urbana a sustituir el transporte público por el uso de la bicicleta.

Además, durante los meses más críticos de la pandemia, cuando la mayoría de las opciones para hacer ejercicio permanecían cerradas, los sistemas de BSS ofrecieron una vía accesible para mantenerse activo físicamente (Filipe Teixeira, Silva et al., 2023). Este contexto no solo pudo consolidar el uso de la bicicleta como medio de transporte, sino que posiblemente permitió que muchas personas descubrieran su utilidad y practicidad en entornos urbanos. Como resultado, el nivel de adopción y la evolución de los BSS podría haberse visto acelerado en los años posteriores a la pandemia.

Ilustración 6: Números de viajes al mes por ciudad



Fuente: Elaboración propia



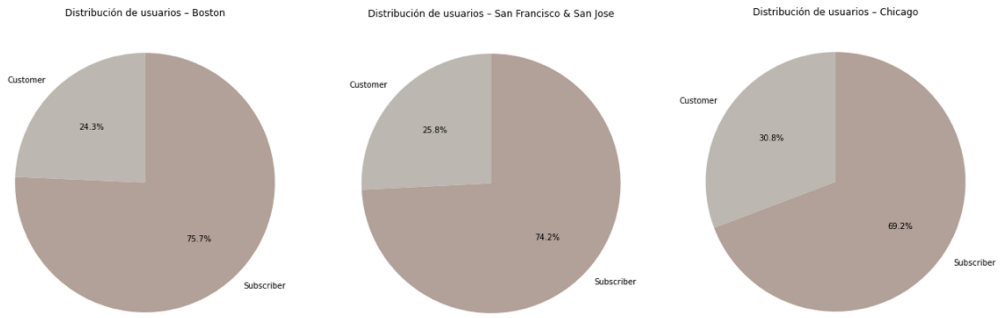
Fuente: Elaboración propia

Si nos centramos en los resultados por ciudad, se puede confirmar que Chicago es la ciudad que concentra el mayor número de trayectos, destacándose con una diferencia considerable, especialmente durante los meses de primavera y verano. Esta tendencia podría explicarse debido a que Chicago cuenta con la mayor población urbana entre las tres ciudades analizadas, con aproximadamente 2,75 M de personas, lo que implica una mayor cobertura del servicio de *bikesharing* y, en consecuencia, un mayor número de trayectos mensuales.

Además, tanto en Chicago como en Boston, se aprecia un fuerte aumento del número de trayectos a partir de mayo, lo cual está estrechamente relacionado con las condiciones meteorológicas. Recordemos que ambas ciudades experimentan inviernos con temperaturas muy bajas, lo que reduce significativamente el uso de bicicletas en los primeros meses del año. Por el contrario, San Francisco muestra una tendencia mucho más estable a lo largo del año, lo que puede atribuirse a su clima templado y a la ausencia de inviernos extremos, lo que permite un uso más constante del servicio independientemente de la estación del año.

Por otro lado, al analizar la distribución de los tipos de usuarios, se observa claramente que Chicago presenta un mayor porcentaje de usuarios ocasionales en comparación con el resto de las ciudades (aproximadamente 5 puntos de diferencia). Este dato podría ser indicativo de una mayor capacidad de atracción del servicio hacia usuarios no recurrentes, posiblemente como resultado de campañas de marketing efectivas, o bien por una infraestructura ciclista más desarrollada, que convierte el uso de la bicicleta en una alternativa atractiva para explorar la ciudad de forma ocasional o turística.

Ilustración 7: Distribución de usuarios



Fuente: Elaboración propia

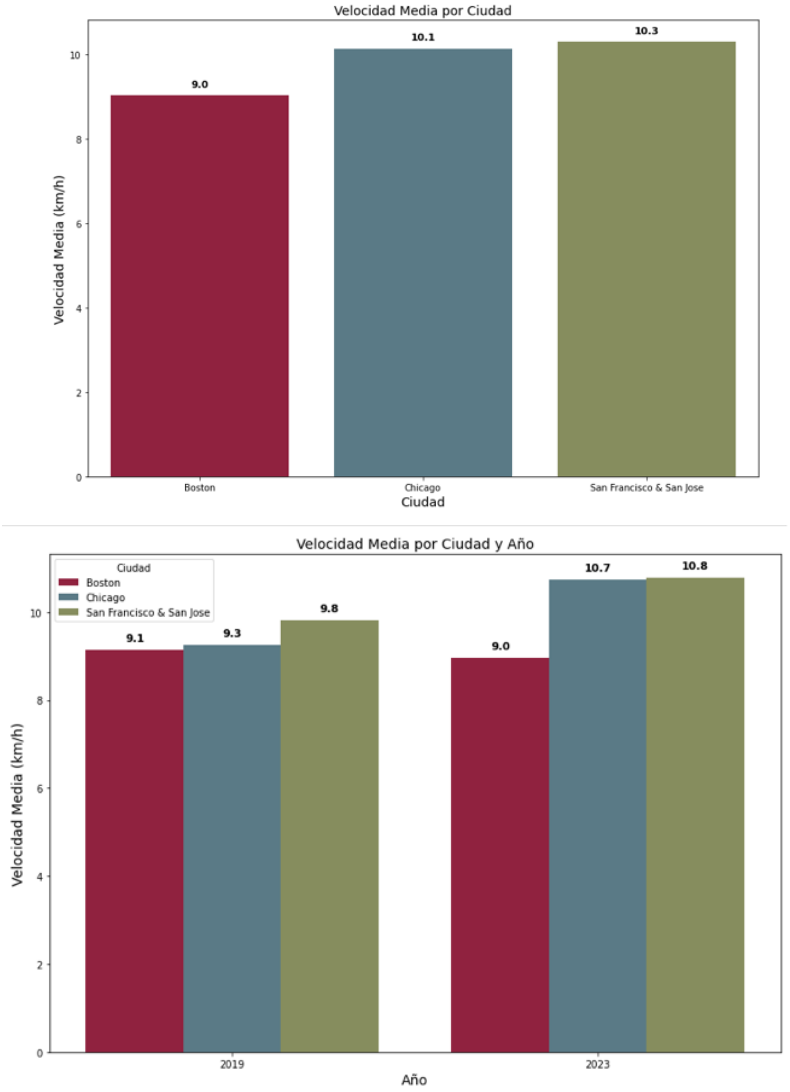
Con estas premisas establecidas, en las siguientes secciones se abordará el análisis tanto de la variable calculada de velocidad como de los trayectos realizados por los usuarios.

8.1. Análisis de la variable velocidad

En primer lugar, se ha decidido poner especial énfasis en el análisis de la variable velocidad ya que es la variable que se ha venido calculando durante este trabajo y resulta especialmente relevante para identificar qué ciudad y año presentan las velocidades más altas, así como para analizar los factores que podrían influir en dichas diferencias. El cálculo de la velocidad podría aportar información útil desde el punto de vista operativo, ya que podría ayudar a detectar qué bicicletas están sometidas a un mayor desgaste. Aquellas que circulan habitualmente a velocidades más altas podrían, por ejemplo, presentar mayores necesidades de mantenimiento, especialmente en frenos o ruedas.

La variable velocidad ha sido calculada dividiendo la distancia entre la duración. Como se adelantaba en secciones anteriores, la media de la variable velocidad es aproximadamente 10km/h lo que indica que la mayoría de los viajes se acercan a esta velocidad. Sin embargo, las velocidades pueden variar en promedio unos 4.5 km/h por encima o por debajo de esta media. Para profundizar más en este análisis e identificar en qué ciudades y años se alcanzan mayores velocidades, se ha representado la velocidad media por ciudad y por ciudad/año.

Ilustración 8: Resultados de la variable velocidad por ciudad y año



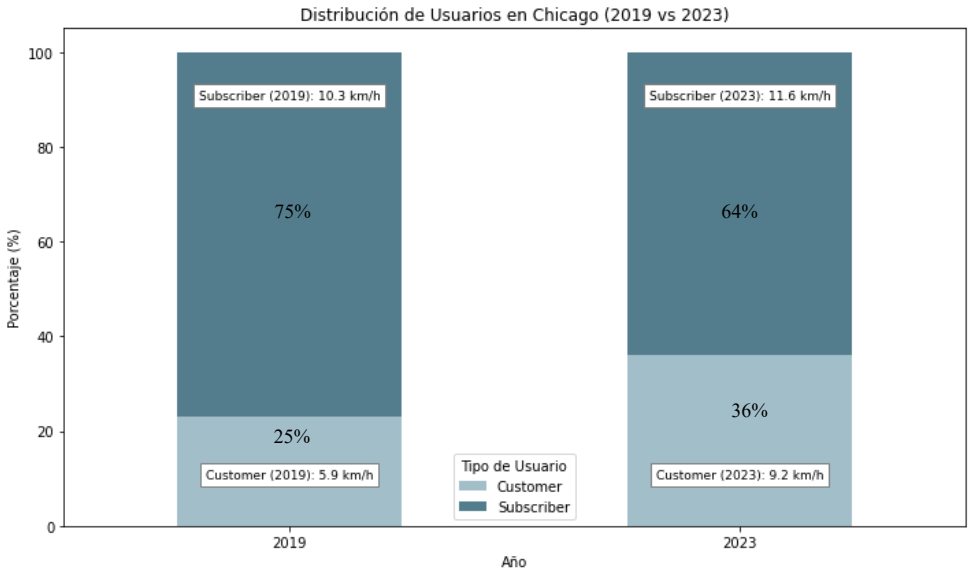
Fuente: Elaboración propia

La ciudad que alcanza mayor velocidad media es San Francisco y San José. Se podría pensar que este dato tiene bastante sentido debido a la topografía del territorio. San francisco cuenta con calles de hasta 30% o 40% de inclinación y fue construida sobre 40 colinas (Polo, 2024), lo que significa que los viajes con bajadas pueden llegar a alcanzar velocidades muy superiores a la media. Sin embargo, desde un punto de vista contrario, también se podría argumentar que las subidas ralentizan los trayectos. Por lo tanto, sería necesario un análisis más detallado que reflejase la altitud de cada viaje para confirmar esta hipótesis.

Por otro lado, llama la atención el incremento en velocidad media en la ciudad de Chicago. De 2019 a 2023 la velocidad media aumenta en 1.4 km/h. Una primera hipótesis

podría ser que este incremento se deba a un aumento en el número de viajes realizados por usuarios suscritos (*subscribers*), quienes tienden a utilizar las bicicletas como medio de transporte habitual y suelen realizar trayectos más rápidos que los usuarios ocasionales (*customers*). Sin embargo, un análisis más detallado revela que la proporción entre *subscribers* y *customers* no ha variado significativamente entre ambos años. De hecho, en 2023 se observa un aumento en el número de viajes realizados por *customers*, y lo más sorprendente es que la velocidad media de este grupo crece un 56%.

Ilustración 9: Chicago velocidad y proporción *customers - subscribers*



Fuente: Elaboración propia

Esto sugiere que el aumento general de la velocidad media se debe, en realidad, a un cambio en el comportamiento de los clientes ocasionales. Una posible explicación es un incremento en las tarifas por minuto: al no estar suscritos a un plan anual, estos usuarios pagan por uso, por lo que un aumento en las tarifas podría motivarlos a reducir el tiempo de duración, completando los mismos trayectos en menos tiempo para ahorrar costes. Otra hipótesis podría ser una mejora en la red ciclista de la ciudad, debido a que una ampliación y mejora de los carriles bici facilitarían la fluidez del tráfico sobre ruedas. De hecho, según el departamento de transporte de Chicago (2023), la red ciclista de la ciudad ha crecido a una media de 30 millas al año desde 2020, doblando el ratio anual de expansión anterior a ese mismo año. Una mejora de infraestructura podría contribuir positivamente a un aumento de la eficiencia de los desplazamientos y traducirse en un aumento de la velocidad.

Estas explicaciones son hipótesis que ayudan a interpretar este comportamiento, pero que podrían no ser del todo correctas. Sería recomendable profundizar en esta cuestión en trabajos futuros, incorporando variables adicionales como los precios exactos de las tarifas en Chicago o información sobre la inversión en construcción de carriles bici.

8.2 Análisis de trayectos

En segundo lugar, se decide también profundizar en el análisis y la representación de conclusiones de los trayectos realizados debido a que pueden aportar información valiosa sobre los hábitos de movilidad, tanto de usuarios suscritos como ocasionales. Este tipo de análisis puede tener aplicaciones prácticas como el diseño de una segmentación más precisa de clientes, o incluso la redistribución estratégica de bicicletas en función de los trayectos más frecuentes y los horarios de mayor demanda, contribuyendo a una mejora del servicio.

Se comenzará representando los trayectos más frecuentes por ciudad para entender mejor las necesidades y demanda de los usuarios. Y, después, se analizarán los horarios más comunes de uso por ciudad y tipo de usuario para entender a qué hora se realizan más recorridos.

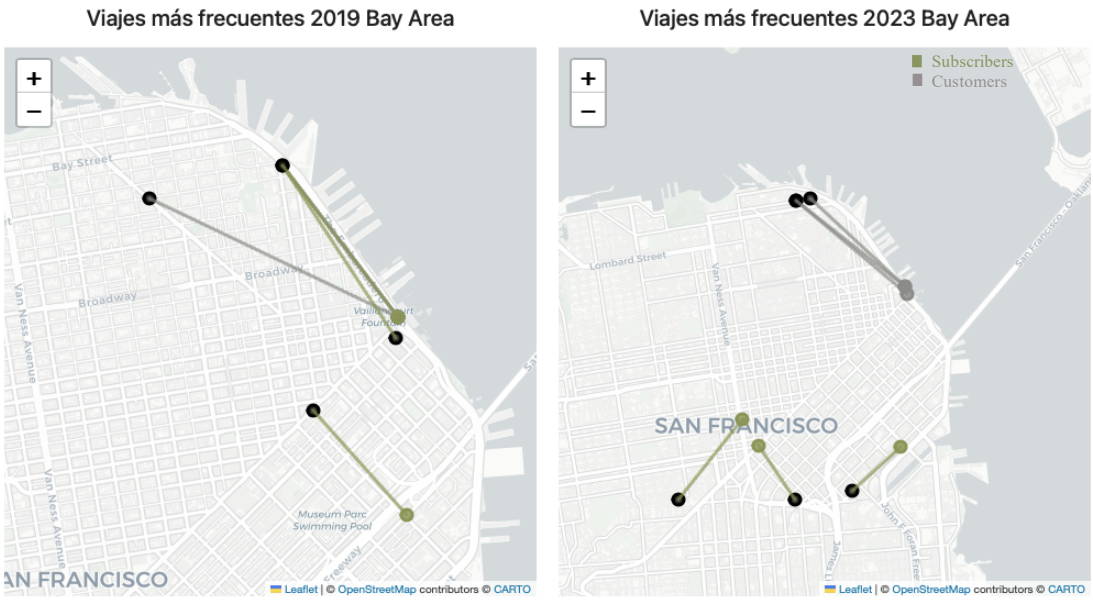
Para analizar los trayectos más frecuentes, se han elaborado mapas gracias a la librería *folium* capaz de crear geo representaciones web interactivas con Python. Estos mapas muestran y comparan los recorridos más habituales en los años 2019 y 2023, diferenciando entre usuarios suscritos y ocasionales.

Los recorridos más frecuentes se han calculado agrupando el *dataset* filtrado por las coordenadas de inicio y fin de viaje. Por ejemplo, si dos trayectos comienzan en el mismo punto A y terminan en el mismo punto B, se contabilizan como dos ocurrencias del mismo trayecto. Este proceso se repite para cada par de coordenadas de inicio y fin, permitiendo ordenar los trayectos de mayor a menor ocurrencia para obtener los trayectos más frecuentes. Este análisis es interesante, ya que permite observar cómo varían los patrones

de viaje según el perfil del usuario y cuáles son las necesidades de movilidad de cada tipo de usuarios.

Se esperaría que los usuarios suscritos utilizasen el servicio principalmente para desplazamientos cotidianos, como ir o volver del trabajo/universidad/colegio, por lo que es probable que sus trayectos se concentren en zonas residenciales, universitarias y empresariales. Por otro lado, se espera que los usuarios ocasionales realicen trayectos relacionados con el ocio o el turismo, siendo más habituales en zonas de interés turístico.

Ilustración 10: Mapas trayectos más frecuentes por ciudad, año y tipo de usuario

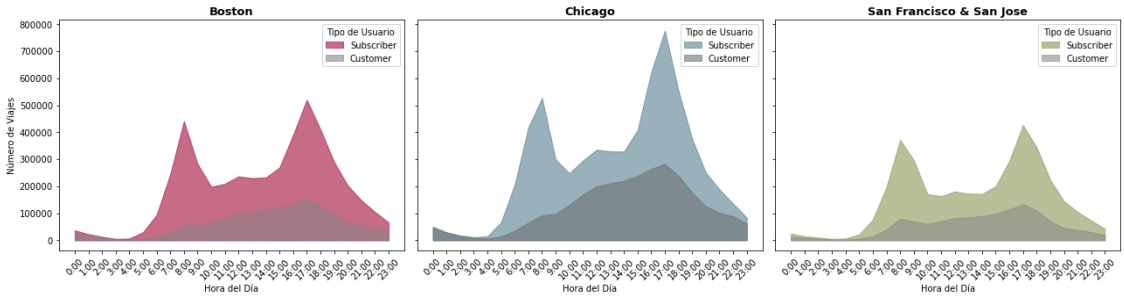


Fuente: Elaboración propia

Un claro ejemplo del comportamiento esperado se observa en la ciudad de Chicago. En el caso de los usuarios ocasionales (*customers*), se observa una mayor concentración de trayectos en zonas turísticas con amplios carriles bici, ideales para el ocio y turismo. Estas áreas incluyen los alrededores del lago Michigan, el *Riverwalk* y *Oak Street Beach*. En contraste, los usuarios suscritos (*subscribers*) tienden a desplazarse por zonas universitarias y empresariales. Un ejemplo destacado es *Hyde Park*, ubicado al sur de la ciudad, conocido por ser una zona residencial que alberga la Universidad de Chicago.

También resulta interesante investigar cuándo se realizan estos desplazamientos y con qué frecuencia a lo largo del día. Con este objetivo, se ha analizado la distribución horaria de los viajes, diferenciando entre usuarios suscritos y ocasionales, con el fin de identificar posibles patrones temporales en el uso del servicio.

Ilustración 11: Distribución horaria de trayectos por ciudad, año y tipo de cliente



Fuente: Elaboración propia

Los resultados muestran diferencias claras entre ambos perfiles: en el caso de los usuarios suscritos, se observan dos picos muy marcados en torno a las 8:00 de la mañana y las 17:00 horas, coincidiendo con los horarios de entrada y salida del trabajo o centros educativos. En cambio, los usuarios ocasionales (gris) tienden a utilizar las bicicletas de forma más dispersa a lo largo del día, aunque con una mayor concentración hacia la tarde, especialmente alrededor de las 17:00 horas. Estos patrones refuerzan la idea de que los usuarios suscritos hacen un uso funcional mientras que los ocasionales uno más recreativo o turístico.

Tras el análisis de los trayectos se puede confirmar que los usuarios suscritos suelen realizan desplazamientos cotidianos por zonas universitarias, residenciales y empresariales sobre todo a las ocho de la mañana y a las cinco de la tarde. Mientras que los usuarios ocasionales realizan recorridos por zonas turísticas sobre todo a media tarde.

9. Visualización dinámica e interactiva en Power BI

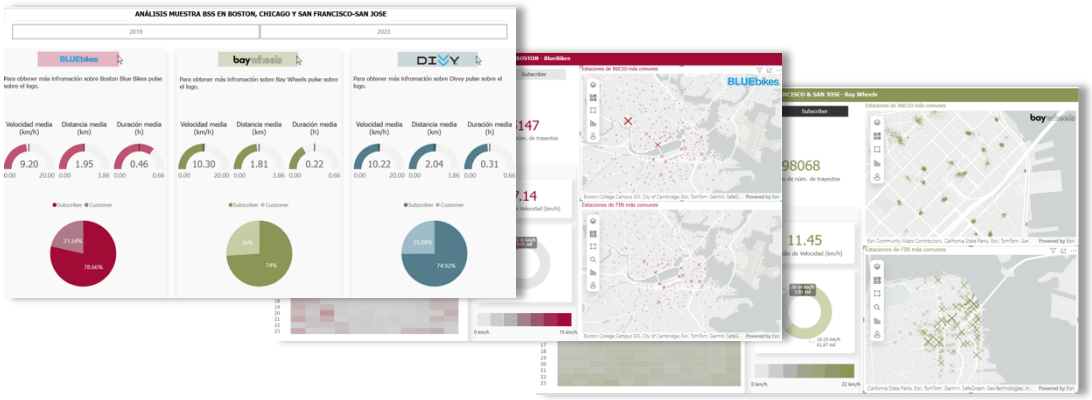
Como apartado final del Trabajo de Fin de Grado, he querido incluir una sección realizada con Power BI, ya que es una herramienta muy útil en el entorno empresarial y permite crear visualizaciones interactivas y dinámicas. A diferencia de los gráficos hechos en Python, Power BI permite aplicar filtros y explorar los datos de forma más flexible, lo que aporta un valor añadido al análisis.

Para construir el *dashboard* se ha trabajado con el *dataset* filtrado por velocidades menores de 30 km/h, es decir, el mismo *dataset* utilizado para las visualizaciones de Pyhton hechas con Matplotlib.

El diseño del *dashboard* es sencillo y claro. Cuenta con una vista general comparativa entre ciudades, donde se resumen algunas de las métricas principales estudiadas durante el trabajo. Además, el usuario puede acceder a una interfaz específica para cada ciudad, donde encontrará un análisis más detallado de la velocidad media, las estaciones más frecuentes y el número total de viajes registrados.

Esta herramienta facilita la comprensión de los datos permitiendo al usuario explorar diferentes aspectos como la velocidad, distancia y duración media por ciudad, así como identificar los meses y horarios con mayores velocidades o las estaciones de inicio y fin más frecuentes. Todo esto puede consultarse de forma personalizada según el año o el tipo de usuario.

Ilustración 12: Interfaz Power Bi



Fuente: Elaboración propia enlace al Dashboard en Anexos

10.Conclusiones

Una vez expuesto el trabajo de fin de grado se analizarán ahora las conclusiones que se han alcanzado tras el análisis.

El primer objetivo pretendía crear una tabla canónica con las mismas columnas y formato para 6 bases de datos distintas. Tras un proceso de ETL con una gran cantidad de transformaciones se puede afirmar que se ha conseguido crear una tabla común con 8 columnas que permite aunar la información relevante las siguientes bases de datos: *Boston Blue Bikes* 2019*, Boston Blue Bikes* 2023, *San Francisco & San José Bay Area* 2019, *San Francisco & San José Bay Area* 2023, *Chicago Divvy* 2019 y *Chicago Divvy* 2023. La limpieza, transformación y unificación de la información ha permitido la consecución del resto de objetivos.

En cuanto al segundo objetivo, que trataba de calcular la velocidad dividiendo la distancia recorrida entre la duración, se ha llegado a la conclusión de que en torno al 98% de los

viajes se encuentran en un rango de velocidad comprendido entre 0 y 20 km/h con una media de 10 km/h. Además, la ciudad con mayor velocidad media es San Francisco & San José alcanzando un promedio de 10,3 km/h por trayecto. Esto se puede deber a su topografía con pronunciadas pendientes, sin embargo, sería necesario confirmarlo mediante un análisis en profundidad.

La ciudad de Chicago tampoco se queda atrás en términos de velocidad media sobre todo durante el 2023. Este incremento de velocidad en Chicago se debe al incremento en la velocidad media de los clientes ocasionales, que realizan sus viajes en 2023 un 56% más rápido que en 2019. Las altas velocidades probablemente desgasten más las bicicletas afectando a su mantenimiento. Con este análisis los servicios de *Bikesharing* podrían centrar sus esfuerzos en identificar aquellas bicicletas con mayor desgaste y prever las reparaciones para ofrecer servicios de calidad priorizando sus recursos.

El tercer objetivo, consistía en comprender los hábitos, comportamientos y necesidades de la población para cada ciudad. Tras el análisis de trayectos y la información demográfica y social de cada ciudad se puede concluir que, los usuarios de los servicios de *Bikesharing* presentan comportamientos muy distintos según su tipología. Se ha demostrado que los usuarios suscritos suelen usar las bicicletas como medio de transporte cotidiano, es decir para ir y volver del trabajo/centro educativo. Sus trayectos se caracterizan por distancias y duraciones generalmente cortas con velocidades altas. Por otro lado, los usuarios ocasionales aprovechan este tipo de transporte para realizar trayectos turísticos no cotidianos durante horas menos puntuales. Estos trayectos se caracterizan por abarcar distancias y duraciones más largas y alcanzar velocidades bajas al consistir en trayectos recreativos y de ocio.

En cuarto lugar, se eligieron dos años de análisis con el objetivo de estudiar el efecto del Covid-19 en los patrones de viaje antes y después del Covid-19. Aunque no se puede atribuir de forma directa la diferencia entre ambos años exclusivamente al impacto de la pandemia, sí se ha observado un aumento significativo del 34% en el número total de trayectos en 2023 respecto a 2019. Este incremento podría estar relacionado con un cambio en las preferencias de movilidad de la población tras la pandemia, reflejando una mayor adopción de medios de transporte individuales y al aire libre como el *Bikesharing*.

En definitiva, este trabajo ha supuesto una oportunidad para aplicar de forma práctica la analítica de datos a un caso real. A lo largo del proceso, he podido trabajar con grandes volúmenes de información, aplicar técnicas de limpieza y transformación de datos, desarrollar visualizaciones interactivas y, sobre todo, extraer conclusiones a partir de datos inicialmente sin información. El estudio de los servicios de *Bikesharing* en tres grandes ciudades de Estados Unidos ha permitido identificar patrones de comportamiento diferenciados entre tipos de usuarios, así como observar cambios en los hábitos de movilidad tras la pandemia. Además, el cálculo y análisis de la variable velocidad ha aportado una nueva perspectiva sobre la eficiencia de los trayectos y su posible impacto en el mantenimiento de las bicicletas.

Toda la información con respecto al código y al *Dashboard* Power BI se encuentra en los enlaces del anexo.

11.Limitaciones y trabajos futuros

En cuanto a las limitaciones del trabajo, cabe mencionar que el análisis se ha centrado solo en los datos disponibles de tres ciudades (Boston, Chicago y San Francisco & San José) y en los años 2019 y 2023. Aunque esto ha permitido comparar lo que ocurría antes y después del COVID-19, habría sido interesante poder incluir también los años intermedios, especialmente 2020, para ver con más detalle cómo afectó la pandemia en tiempo real.

Por otro lado, al calcular la variable velocidad, se ha encontrado una limitación en aquellos viajes que empiezan y terminan en la misma estación. Recordemos que la distancia en km se ha calculado a partir de la fórmula de Haversine que calcula la distancia en línea recta entre dos coordenadas de una esfera. Dado que el punto de inicio y final en los viajes mencionados coincide, la distancia obtenida es cero, lo que implica que la velocidad calculada también sería igual a cero.

Con respecto a posibles trabajos futuros, sería interesante investigar con más profundidad el caso de Chicago, ya que es la ciudad que más ha incrementado tanto su número de trayectos como su velocidad media. Se podría analizar si este crecimiento se debe a algún

plan de inversión en infraestructura ciclista o si, simplemente, ha habido un cambio de comportamiento más marcado tras la pandemia.

Por otro lado, en temas relacionados con la velocidad, sería interesante usar los IDs de las bicicletas para hacer un seguimiento más técnico y detectar cuáles alcanzan velocidades más altas. Las altas velocidades pueden tener un efecto negativo en el desgaste de las bicicletas. Esto podría servir para hacer un mantenimiento más eficiente y anticiparse a posibles averías o problemas. Sería una herramienta muy útil para mejorar la calidad del servicio y asegurar que todas las bicicletas se encuentren en buen estado.

12.Bibliografía

Amazon Web Services. (s.f.). *¿Qué es ETL?* Obtenido de https://aws.amazon.com/es/what-is/etl/

Blue Bikes. (s.f.). *Boston Blue Bikes*. Obtenido de https://bluebikes.com

Chen, F., Turon, K., Kłos, M., Czech, P., Pamuła, W., & Sierpinski, G. (2018). Fith-generation bikesharing system: Examples from Poland and Chica. *Scientifics Journal of Silesian University of Technology, Series Transport*, 99, 05-13 ISSN: 0209-3324.

Chicago Department of Transportation (CDoT). (Spring de 2023). Obtenido de https://www.chicago.gov/content/dam/city/depts/cdot/bike/2023/2023_Chicago%20Cycling%20Update.pdf.

Climate Data. (2025). *Climate Data: Boston USA*. Obtenido de https://es.climate-data.org/america-del-norte/estados-unidos-de-america/massachusetts/boston-1722/

Climate Data. (2025). *Climate Data: Chicago*. Obtenido de https://es.climate-data.org/america-del-norte/estados-unidos-de-america/illinois/chicago-1574/

Climate Data. (2025). *Climate Data: San Francisco*. Obtenido de https://es.climate-data.org/america-del-norte/estados-unidos-de-america/california/san-francisco-385/

DeMaio, P. (2009). Bike-Sharing: History, Impacts, Model of Provision and Future. *J. Public Transp.*, 12, 41-56.

Department of Justice. (2020). *California Consumer Privacy Act (CCPA)*. Obtenido de https://oag.ca.gov/privacy/ccpa

Fernández, J. (30 de Marzo de 2019). *Expansión*. Obtenido de Lyft: el rival amable de Uber que ha conquistado EEUU: https://www.expansion.com/economia-digital/companias/2019/03/30/5c9a18c322601df3528b45f2.html

Filipe Teixeira, J., Silva , C., & Moura e Sá, F. (abril de 2023). *National library of medicine*. Obtenido de Potential of Bike Sharing During Disruptive Public Health Crises: A Review of COVID-19 Impacts: https://journals.sagepub.com/doi/10.1177/03611981231160537

Fishman, E., Washington, S., & Haworth, N. (2013). Bike Share: A Synthesis of the Literature. *Transport Reviews*, 33(2), 148–165. https://doi.org/10.1080/01441647.2013.775612.

Kou, Z., & Cai, H. (2019). Understanding bike sharing travel patterns: An analysis of trip data from eight cities. . *Physica A: Statiscal Mechanics and its Applications*, 515, 785-797.

Lyft Urban Solutions. (2025). *Lyft*. Obtenido de Sobre Nosotros: https://lyfturbansolutions.com/es/sobre-nosotros

Nathan. (Septiembre de 2016). Obtenido de Nathan Fun: https://nathan.fun/posts/2016-09-07/haversine-with-python/

Polo, C. (2024). *Grand Voyage*. Obtenido de https://blog.grandvoyage.com/que-ver-en-san-francisco/

Ricci, M. (2015). Bike sharing: A review of evidence on impacts and processes of implementation and operation. *Reasearch in Transportation Business & Management*, 28-38.

Russo, F. (13(8), 355 de 2022). *Sustainable Mobility as a Service: Dynamic Model for Agenda 2030 Policies*. Obtenido de https://www.mdpi.com/2078-2489/13/8/355

Salah, I., Mukku, V., Kania, M., Assmann, T., & Zadek, H. (2024). Could the next generation of bike-sharing with autonomous bikes be financually sustainable? *Joournal of Urban Mobility, 6*.

Thomson, E. (Enero de 2024). *A complete guide to colleges in Boston*. Obtenido de The Best Schools: https://thebestschools.org/local/ma/boston/

US Census Bureau. (2023). *Chicago City*. Obtenido de https://data.census.gov/profile/Chicago_city,_Illinois?g=160XX00US1714000

US Census Bureau. (2023). *San Francisco County, California*. Obtenido de https://data.census.gov/profile/San_Francisco_County,_California?g=050XX00US06075

US Census Bureau. (2023). *US Census. Boston-Cambridge-Newton, MA-NH Metro Area*. Obtenido de https://data.census.gov/profile/Boston-Cambridge-Newton,_MA-NH_Metro_Area?g=310XX00US14460

Vallez, C., Castro, M., & Contreras, D. (2021). Challenges and Opportunities in Dock-Based Bike-Sharing Rebalancing: A Systemativ Review. *Sustainability*, 13(4):1829.

Wikipedia. (2007). *Wikipedia*. Obtenido de https://es.wikipedia.org/wiki/Archivo:Blank_US_Map,_Mainland_with_no_States.svg

12.1. Anexos

Anexo 1: Link al repositorio de GitHub con el código:
https://github.com/anafdezvalmayor/BSS-ANALISIS-EEUU.git

Anexo 2: Link al Power Bi:
https://app.powerbi.com/view?r=eyJrIjoiYTU2ODllNmQtODY5MC00YjU4LTlkYTItZjI5ZTUyYzJkNmI3IiwidCI6ImJjZDI3MDFjLWFhOWItNGQxMi1iYTIwLWYzZTNiODMwNzBjMSIsImMiOjh9