

SISTEMA DE ECOBICI EN CABA: Funcionamiento y Mejoras Potenciales

Docentes: Leonardo Lo Presti, Jennifer Goldfeld
Alumno: Mariano Cipponeri

21-09-2025

Índice

PRIMERA PARTE DEL TRABAJO	2
Introducción	2
Objetivos	3
Herramientas a utilizar en el proyecto	3
Primeras tablas	3
Tablas a presentar en la primera entrega	3
Resultados esperados	7
Diagrama de Entidad Relación	7
SEGUNDA PARTE	8
Las bases de datos utilizadas	8
Descripción de las Vistas	8
Funciones	9
Store Procedures	10
Triggers	10
TERCERA PARTE	11
Transacciones	11
Back Up del esquema	11
Informes gráficos	11

PRIMERA PARTE DEL TRABAJO

Introducción

El servicio de *ECOBICIS* es un sistema de bicicletas públicas ofrecidas por el Gobierno de la Ciudad de Buenos Aires (**GCBA**) que funciona desde el año 2010.

Desde el momento en que comenzó a funcionar el sistema, el mismo fue presentando mutaciones en diferentes aspectos:

- Se pasó de un sistema de manual y presencial a uno completamente automatizado (en la actualidad el retiro de las bicicletas requiere de la utilización de la app de Ecobici).
- Se fue aumentando el stock y las prestaciones de las bicicletas (hoy en día todas poseen GPS)
- En 2018 se concesionó el servicio, siendo la empresa brasilera **TEAMBICI** la que ganó la licitación (y la ostenta hasta la actualidad).
- Finalmente, uno de los cambios más relevantes se produjo en 2020, cuando se aranceló parcialmente el servicio, teniendo que pagar los no residentes un importe por la utilización de las bicicletas.

Si bien el servicio de *ECOBICI* viene demostrando un exitoso desempeño (si se toma como parámetro la evolución de unidades y estaciones disponibles, los usuarios y viajes realizados), tanto desde el GCBA como desde TEAMBICI se busca realizar modificaciones en el servicio, en pos de mejorar la experiencia para los usuarios, incrementar la penetración del servicio y mejorar los márgenes de ganancia del Gobierno y la empresa a cargo de la concesión.

Establecido dicho norte, se buscará realizar un estudio que de cuenta tanto del volumen de la utilización de ECOBICI, como del perfil de los usuarios/viajes realizados. Esto permitirá llevar a cabo acciones orientadas a satisfacer los objetivos planteados en el párrafo anterior.

Si bien este trabajo **parte de una situación hipotética**, los principales insumos provendrán de bases de datos de ECOBICI en el sitio de datos abiertos del GCBA.

El grueso de los campos y registros procesados provendrán de las bases del dataset de datos abiertos GCBA, sin embargo, se incluirán campos ficticios, con el fin de poder profundizar el uso de SQL. Cuando esto suceda, será debidamente aclarado.

Objetivos

A través del uso del herramental ofrecido por SQL en las bases de datos mencionadas, se buscará:

- Describir el perfil de los usuarios de ECOBICI (sexo, edad, con la posibilidad de incluir el campo *residencia*).
- Calcular el volumen de usuarios / recorridos realizados y su evolución en el tiempo
- Caracterizar los recorridos realizados (barrio y estación de origen, momento en el que se realizó el viaje, duración temporal, distancia recorrida).
- Evaluar las calificaciones otorgadas al recorrido realizado (el campo de calificación del viaje será ficticio), de manera tal de buscar relaciones entre la calificación y variables como: modelo de bicicleta utilizado o estación de origen.
- Relacionar perfiles de usuario con el volumen de uso de ECOBICI, con el fin de desarrollar estrategias de penetración del servicio.
- Estimar cuál habría sido el volumen de ingresos generados por la ECOBICI en caso de haber sido un servicio tarifado para toda la población en 2024.

Herramientas a utilizar en el proyecto

Con el fin de cumplir con los objetivos propuestos, se utilizarán las siguientes herramientas:

- Las bases de *Ecobici* disponibles en datos abiertos del GCBA serán descargadas y modeladas vía *RStudio*. Como resultado, se generarán dos bases de muestra. También se utilizará *RStudio* para elaborar los gráficos del informe final. Los paquetes utilizados serán los siguientes:
 - dplyr, readr, utils, tidyverse, sf, readxls, openxlsx, janitor, scale y ggplot2.
- Para la elaboración del esquema *ecobici*, con sus tablas y relaciones, se utilizará MySQL Workbench, que también se aplicará para cargar los datos de las bases de muestra generadas con R, realizar funciones, triggers, store procedures, transacciones, generar un back up de todos los archivos generados, y exportar los resultados provenientes de las vistas en formato .csv (que serán de insumo para la elaboración de gráficos en R).

Primeras tablas

Tablas a presentar en la primera entrega

En esta primera entrega, se generaran las siguientes tablas:

GENERO:

Esta tabla se utilizará para etiquetar al campo *genero* en las futuras consultas. Se establece como Primary Key *id_genero* y se relacionará con el *id_genero* de la tabla **usuarios**.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id del genero	id_genero	INT	PK
genero	genero_usuario	VARCHAR (10)	-

USUARIOS:

Esta tabla se utilizará para aportar información del usuario en la tabla de **recorridos**, se vinculará a través de la primary key *id_usuario*.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id del usuario	id_usuario	INT	PK
id del genero	genero_usuario	INT	FK
edad	edad_usuario	INT	-
fecha de inscripción en ECOBICI	fecha_alta	DATE	-
hora de la fecha del alta	hora_alta	TIME	-

COMUNA:

Esta tabla se utilizará para etiquetar al campo *comuna* en las futuras consultas. Se establece como Primary Key *id_comuna* y se relacionará con el *id_comuna* de la tabla **barrios**.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id de la comuna	id_comuna	INT	PK
nombre de la comuna	nombre_comuna	VARCHAR (10)	-

BARRIO:

Esta tabla se utilizará para etiquetar al campo *barrio* y aportar información de los barrios en las futuras consultas. Se establece como Primary Key *id_barrio* y se relacionará con el *id_barrio* de la tabla **recorridos**.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id del barrio	id_barrio	INT	PK
nombre del barrio	nombre_barrio	VARCHAR (40)	-
id de la comuna	id_comuna	INT	FK
cantidad de habitantes del barrio	poblacion	INT	-
cantidad de habitantes mujeres del barrio	poblacion_fem	INT	-
cantidad de habitantes varones del barrio	poblacion_masc	INT	-

ESTACIONES:

Esta tabla se utilizará para etiquetar al campo *estacion* en las futuras consultas y para aportar información de las estaciones. Se establece como Primary Key *id_estacion* y se relacionará con el *id_estacion* de la tabla **recorridos**.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id de la estacion de bici	id_estacion	INT	PK
nombre de la estación	nombre	VARCHAR (40)	-
dirección de la estación	direccion	VARCHAR (60)	-
id del barrio	id_barrio	INT	FK
latitud de la estación	latitud	DECIMAL(9,6)	-
longitud de la estación	longitud	DECIMAL(9,6)	-

MODELO:

Esta tabla se utilizará para etiquetar al campo *modelo* (de la bicicleta) en las futuras consultas. Se establece como Primary Key *id_modelo* y se relacionará con el *id_modelo* de la tabla **recorridos**.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id del modelo de la bici	id_modelo	INT	PK
nombre del modelo de la bici	modelo	VARCHAR (10)	-

MESES:

Esta tabla se utilizará para etiquetar al campo *mes* en las futuras consultas. Se establece como Primary Key *id_mes* y se relacionará con el *id_mes* de la tabla **recorridos**.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id del mes	id_mes	INT	PK
nombre del mes del año	mes	VARCHAR (15)	-

PRECIOS:

Esta tabla se utilizará para etiquetar al campo *precio* en las futuras consultas. Se establece como Primary Key *id_mes* y se relacionará con el *id_mes* de la tabla **recorridos**.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id del precio	id_precio	INT	PK
precio sugerido por el cobro del servicio	precio	INT	-

RECORRIDOS:

Esta tabla aportará información del *recorrido*, relacionándose con las tablas anteriormente descritas.

Nombre Campo	Abreviatura	Tipo de datos	Tipo de clave
id del recorrido realizado	id_recorrido	INT	PK
id del usuario	id_usuario	INT	FK
id de la estacion de origen	id_estacion_orig	INT	FK
id del mes en que se produce el recorrido	id_mes	INT	FK
fecha en que se retira la bici	fecha_origen	DATETIME	-
id de la estacion de destino	id_estacion_dest	INT	FK
fecha y hora en que se produce el retiro	fecha_dest	DATETIME	-
id del modelo de la bici utilizada	id_modelo	INT	FK
calificación de la experiencia de uso de la bici	calificacion	INT	-
id del precio que se debería haber pagado por la bici	id_precio	INT	FK

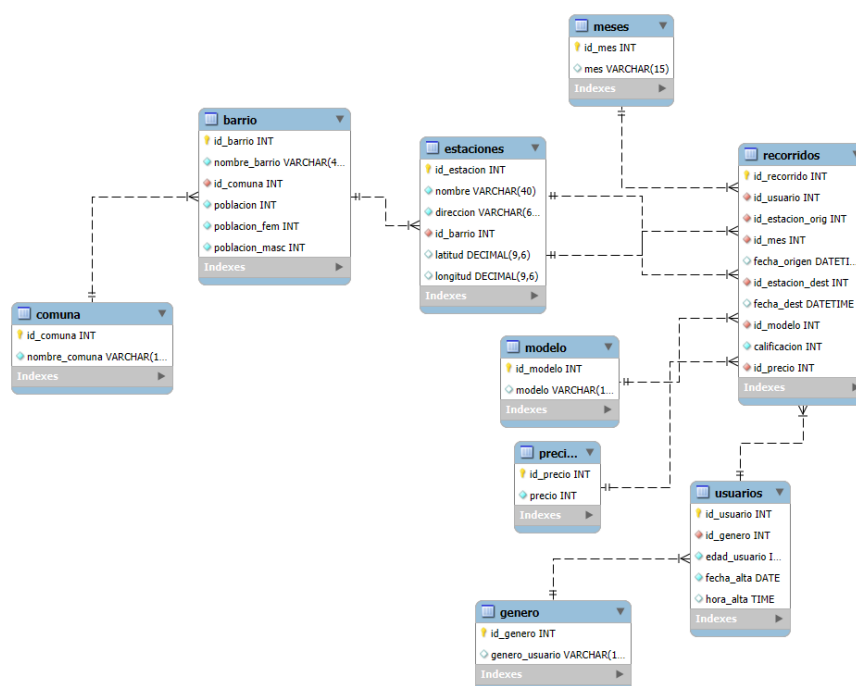
Resultados esperados

A través de la utilización de las bases mencionadas, a priori (ya que a futuro se irán modificando los objetivos o sumando nuevas propuestas), se espera realizar vía SQL las siguientes consultas / vistas:

- Cantidad de usuarios por sexo
- Cantidad de usuarios por edad
- Recorridos por sexo
- Recorridos por edad
- Cantidad de recorridos por estación
- Cantidad de recorridos por barrio
- Calificación del recorrido por estación
- Recorridos por tipo de modelo de bicicleta utilizada
- Calificación promedio otorgada por modelo de bicicleta
- Ingresos que se podrían haber generado por la utilización del servicio por barrio
- Ingresos que se podrían haber generado por la utilización del servicio por estación
- Usuarios que hubiesen gastado más en el servicio (se podrá seleccionar a los quintiles más altos)
- Kilómetros recorridos por estación
- Kilómetros por sexo
- Kilómetros por edad
- Kilómetros por estación
- Minutos de recorrido por estación de origen
- Minutos de recorrido por sexo
- Minutos de recorrido por edad
- Minutos de recorrido por estación de origen

A futuro, se incorporarán consultas que implicarán el cruce de una mayor cantidad de variables, como por ejemplo cantidad de recorridos por sexo, según estación.

Diagrama de Entidad Relación



SEGUNDA PARTE

En esta segunda entrega se presentarán las primeras **Vistas, Funciones, Store Procedures y Triggers**; los cuales fueron elaborados con el fin de satisfacer los objetivos propuestos en la primera entrega.

Las bases de datos utilizadas

El query generado para la primera entrega (query *1_tabla_relaciones*) permitía generar las tablas del esquema junto con las relaciones entre las mismas .

Luego de correr ese query, se deben cargar las dos bases comprimidas de recorridos y usuarios (**en la carpeta bases_2da_entrega del repositorio**).

Con respecto a esas bases, se debe aclarar que no son bases completas, dada la cantidad de registros de las mismas (la base completa de usuarios tiene **un total de 247 mil casos**, mientras que el total de registros en la base de recorridos del 2024 **asciende a 3.31 millones**) se realizaron dos bases de muestra, que serán de utilidad para comprobar el buen funcionamiento de las tablas, relaciones, vistas, funciones, store procedures y triggers propuestos.

Para importar las dos tablas se **utilizó el asistente gráfico de MySQL Workbench** (se debe hacer click derecho sobre la tabla usuarios/recorridos; poner “table data import wizard”; abrir las bases descomprimidas; poner “use existing table” y seleccionar ecobici.usuarios/ecobici.recorridos; clickear “next” en las dos ventanas que emergen).

Descripción de las Vistas

En esta segunda entrega, se presentan 14 vistas:

1) **vista_recorridos_x_edad:**

Objetivo: presentar una frecuencia de recorridos según edad de los usuarios

Tablas involucradas: recorridos; usuarios

2) **vista_recorridos_porc_getario:**

Objetivo: distribuir los recorridos por grupo etario, con absolutos y porcentajes

Tablas involucradas: recorridos; usuarios.

3) **vista_recorridos_porc_genero:**

Objetivo: distribuir los recorridos por genero, con absolutos y porcentajes

Tablas involucradas: recorridos; usuarios; genero.

4) **recorridos_porc_genero_edad:**

Objetivo: distribuir los recorridos por genero y edad, con absolutos y porcentajes

Tablas involucradas: recorridos; usuarios; genero.

5) **vista_calificaciones_modelo:**

Objetivo: obtener la calificación promedio por cada tipo de modelo de bicicleta

Tablas involucradas: recorridos; modelo

6) **vista_calificacion_barrio:**

Objetivo: obtener la calificación promedio por barrio de origen del recorrido

Tablas involucradas: recorridos; estaciones; barrio

7) vista_recorridos_x_est:

Objetivo: obtener la cantidad de recorridos por estacion

Tablas involucradas: recorridos; estaciones

8) vista_recorridos_x_est:

Objetivo: obtener la cantidad de recorridos por estacion

Tablas involucradas: recorridos; estaciones

9) vista_top_estacion:

Objetivo: obtener la estación con mayor cantidad de recorridos

Tablas involucradas: recorridos; estaciones

10) vista_usuarios_genero:

Objetivo: conocer la distribución de usuarios por genero

Tablas involucradas: usuarios; genero

11) vista_usuarios_edad:

Objetivo: conocer la distribución de usuarios por edad

Tablas involucradas: usuarios; genero

12) vista_ganancias:

Objetivo: conocer los ingresos totales que podría haber generado el servicio

Tablas involucradas: recorridos; precio

13) vista_ingresos_est:

Objetivo: conocer los ingresos totales que podría haber generado el servicio, según estación

Tablas involucradas: recorridos; estaciones; precio

14) vista_km_x_reco:

Objetivo: conocer la cantidad de kilómetros por recorrido (utilizando el dato de latitud y longitud de las estaciones de origen y destino).

Tablas involucradas: recorridos; estaciones

Funciones

En esta etapa del trabajo se generaron **dos funciones:**

1) fn_dif_tiempo_minutos:

Objetivo: La tabla de recorridos ofrecida por el portal de datos abiertos de CABA ofrece una fecha de origen y una de destino (ambos campos con una precisión de fecha, hora, minutos y segundos). La función generada permite **conocer la duración en minutos de cada recorrido.**

Tablas involucradas: recorridos

2) fn_dif_tiempo_minutos:

Objetivo: Al calcular el promedio de calificaciones otorgadas a cada barrio, esta función **permite realizar un monitoreo de la calidad del servicio ofrecido por cada barrio.**

Tablas involucradas: recorridos; estaciones; barrios

Store Procedures

Se presentan, por el momento, dos Store Procedures:

1) estaciones_por_barrio:

Objetivo: Ofrecer un listado de estaciones por cada barrio, seleccionando a cada barrio por su `id_barrio` devuelve el nombre del barrio, la comuna a la que pertenece, el nombre de cada estación y la dirección.

Tablas involucradas: estaciones; barrios; comunas

2) recorridos_usuario:

Objetivo: Listar los recorridos realizados por cada usuario. La selección de cada usuario (a partir de su `id`) arroja una tabla con las principales características de los recorridos realizados (fecha de inicio de viaje, estación de origen, barrio, comuna, dirección y calificación otorgada).

Tablas involucradas: recorridos; usuarios; genero; estaciones; barrio; comuna

Triggers

El sistema de Ecobicis es dinámico, constantemente se van sumando nuevos usuarios al sistema (y se van descartando los usuarios que no respetan las reglas de uso de las bicicletas -tiempos de viaje excesivos o daños a las unidades).

Por estos motivos, es necesario generar triggers que registren las altas y bajas de usuarios del sistema.

Con ese fin se generaron dos triggers:

-tr_auditoria_usuarios

-tr_auditoria_usuarios eliminados

Estos triggers hacen que se registren en la tabla *auditoria_usuarios* el `id_usuario`, `id_genero`, edad del usuario, fecha y hora de modificacion (del alta o baja), el usuario (data entry) que realizó el alta/baja y un comentario en el que se aclara si se realizó un alta o una baja.

Algo similar sucede con las estaciones, el GCBA prevé la incorporación de nuevas estaciones y la baja de las estaciones que no registren una cantidad significativa de recorridos. Entonces, se agregaron al esquema los siguientes triggers. . .

-tr_auditoria_estaciones

-tr_auditoria_estaciones eliminadas

... que registran en la tabla *auditoria_estaciones* el `id_estacion`, nombre, dirección, fecha y hora de alta/baja, el data entry que realizó la acción y el comentario que especifica si se trata de un alta o una baja.

TERCERA PARTE

Transacciones

Se incluyeron tres ejemplos para poner en práctica el uso de transacciones en el esquema realizado (las mismas figuran en el query *6_transacciones* de la carpeta *queries* del proyecto).

1) Commit:

En la tabla *usuarios* se insertó un nuevo usuario utilizando el comando *commit* (con la previa desactualización del *autocommit*)

Se agregó un ejemplo similar con la actualización de la edad del nuevo usuario generado (utilizando el comando *-update_*).

2) Rollback:

En la tabla *estaciones* se inició una transacción para insertar una nueva estación. Luego se realizó otra transacción en la que nueva estación se eliminaba, pero fue restituída a través de un *rollback*.

3) Savepoint:

En la tabla de *recorridos* se modificaron las calificaciones otorgadas a dos recorridos realizados.

Se aplicó un savepoint al modificar la primera calificación y luego de modificar la segunda calificación se aplicó el comando *rollback to* el savepoint realizado. De esta manera, la calificación del primer caso fue modificada y la del segundo mantuvo el valor original.

Luego de realizar los tres ejercicios, **se seteo el autocommit para que quede activado.**

Back Up del esquema

Con el fin de garantizar la seguridad del esquema trabajado, se realizó un *back up* de los archivos involucrados.

Para eso, se realizaron los siguientes pasos: en *Administration* se seleccionó *data export* y, luego, el esquema *ecobici* y se optó por generar un *self contained file* (un único archivo sql con toda la información seleccionada). El back up de nombre **Dump_ecobici_20250921** fue guardado en la carpeta *back_up* del proyecto.

Informes gráficos

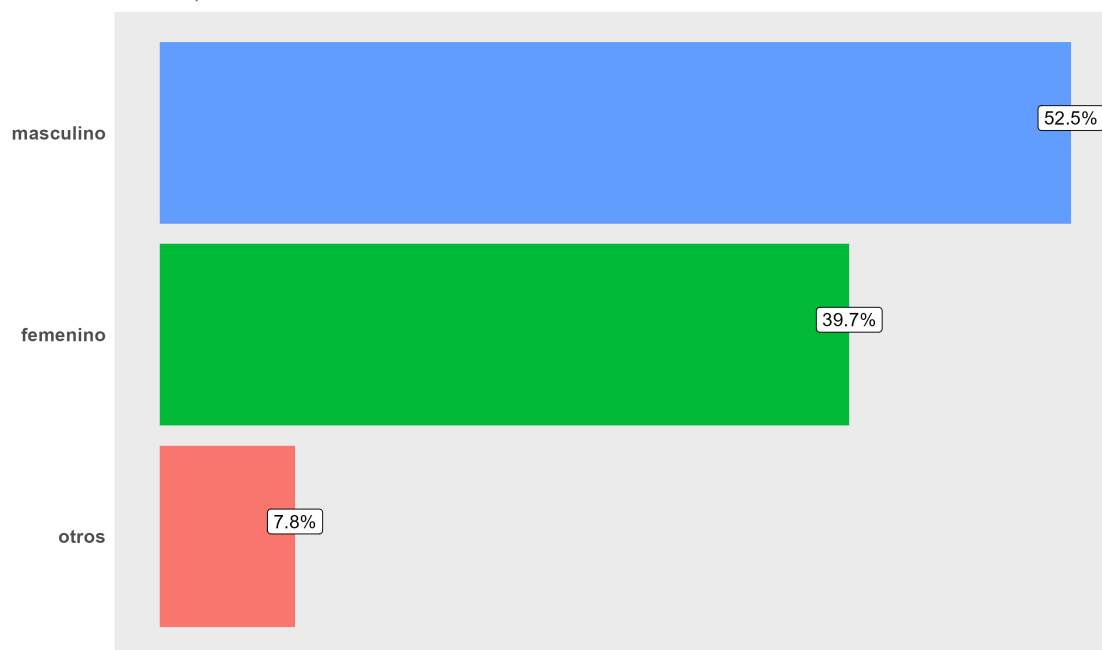
Como punto final del trabajo, se presentan algunos gráficos con resultados que surgieron de las vistas generadas.

Las tablas fueron exportadas desde **MySQL Workbench** en formato .csv, mientras que los gráficos se levantaron en R y fueron elaborados utilizando el paquete *ggplot2* (el script “3_generacion_graficos” con las sentencias para realizar los gráficos se encuentra en la carpeta *procesamientos_r*).

En el primer gráfico se puede apreciar cómo en la muestra de usuarios seleccionados, predominan los de género masculino. Otro punto a destacar es que el 7,8% de la muestra completó la categoría “otros” a la hora de registrarse.

Distribución de los usuarios según género

ECOBICI; año 2024

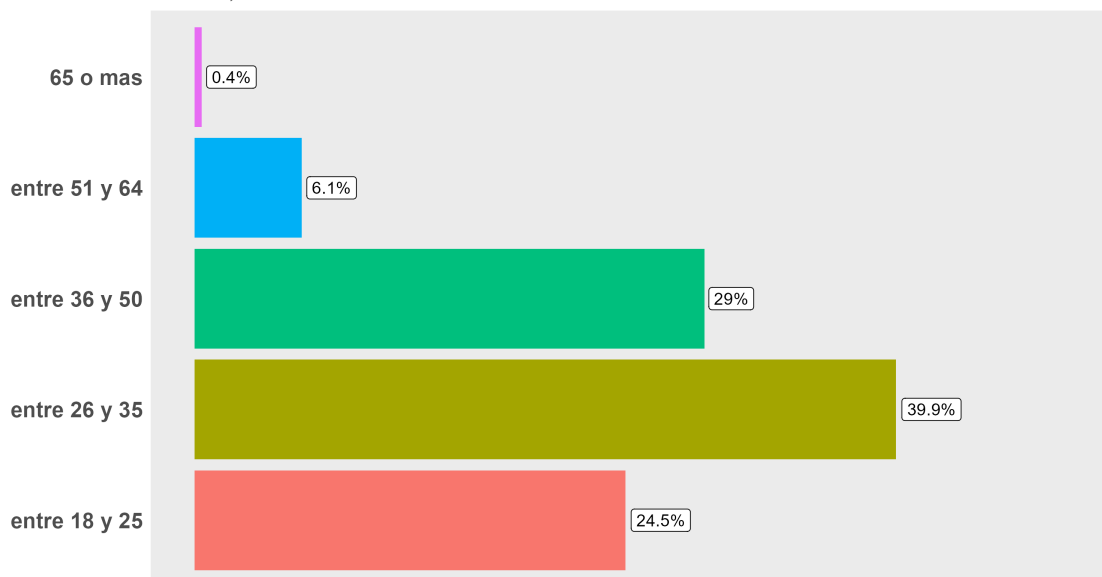


Fuente: Elaboración propia en base a Datos Abiertos GCBA

Tal como se puede apreciar a continuación, 4 de cada 10 usuarios tienen entre 26 y 35 años, seguidos en peso por los que tienen entre 36 y 50 años.

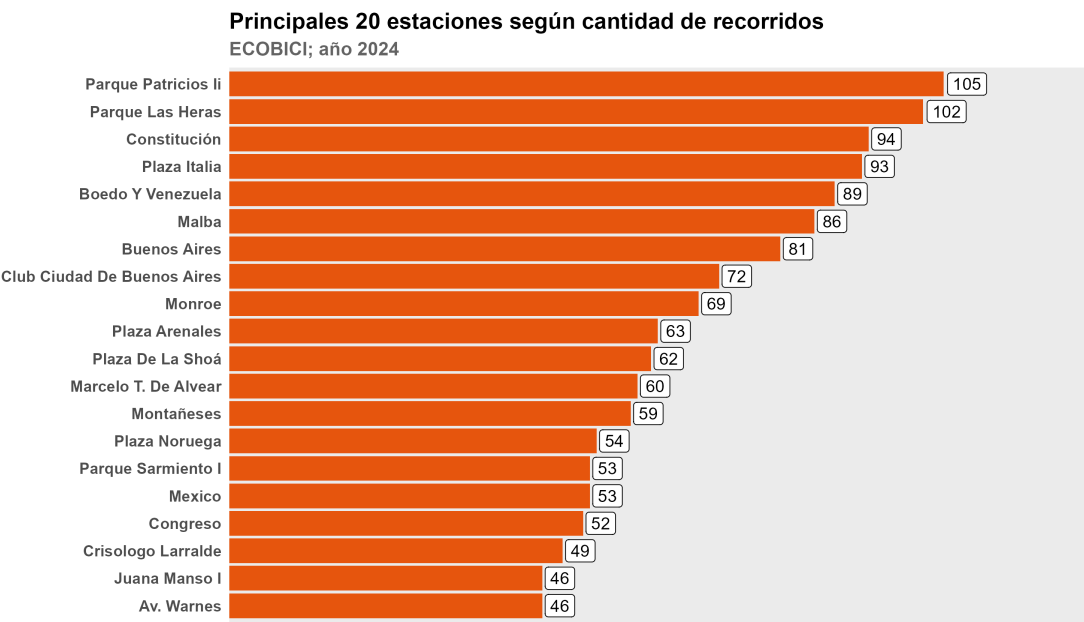
Distribución de los usuarios según grupo etario

ECOBICI; año 2024



Fuente: Elaboración propia en base a Datos Abiertos GCBA

En el top 20 de estaciones, se destacan “Parque Patricios II”, “Parque Las Heras”, “Constitución” y “Plaza Italia” como las que mayor cantidad de recorridos concentraron en el año.



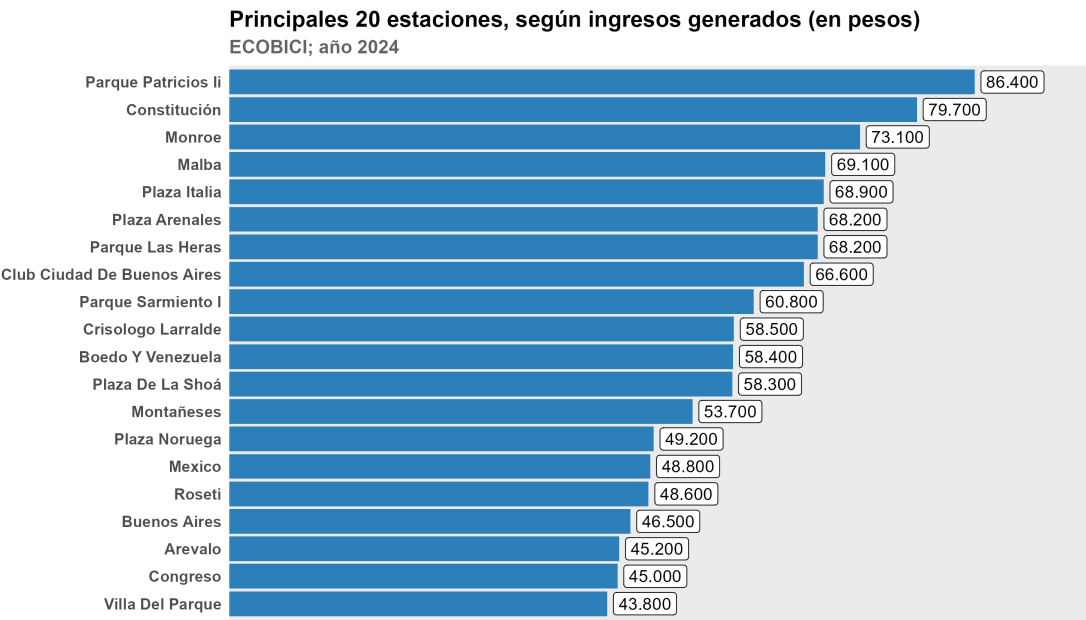
Fuente: Elaboración propia en base a Datos Abiertos GCBA

La estaciones Venancio Flores, Soldado de La Frontera, Nicasio Oroño, Huaura, Coghlan, Balbín 2 y Av. Triunvirato **se destacaron por tener calificaciones perfectas.**



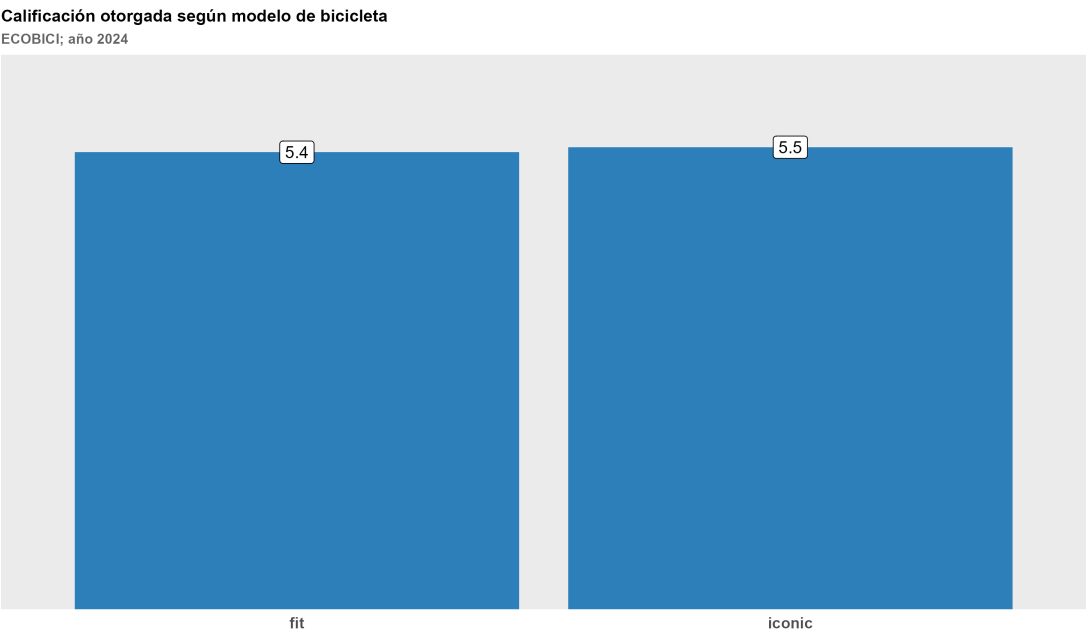
Fuente: Elaboración propia en base a Datos Abiertos GCBA

Al analizar los ingresos generados por estación, Parque Patricios, Constitución y Monroe fueron las que mayor aporte registraron.



Fuente: Elaboración propia en base a Datos Abiertos GCBA

Finalmente, no se observan diferencias a la hora de comparar las calificaciones otorgadas a los recorridos según modelo de bicicleta.



Fuente: Elaboración propia en base a Datos Abiertos GCBA