



Ejercicio de análisis de datos

26/05/2021
CABA, Argentina

Subsecretaría de Gobierno Abierto y Rendición de Cuentas
Secretaría General y Relaciones Internacionales
Gobierno de la Ciudad de Buenos Aires

En este informe se plantea la resolución del siguiente ejercicio:

Consigna

Generar un reporte con Análisis Exploratorio de Datos, mostrando como máximo 6 charts (gráficos o tablas), para describir los datos. Sacar insights y/o conclusiones relevantes de los gráficos.

Datos

Información sobre los recorridos realizados con el sistema público de bicicletas por año. Cuenta con datos de origen, destino, horario, género y edad de los usuarios a lo largo de los años.

Introducción

El presente informe tiene como objetivo formular un análisis utilizando los datos de usuarios del sistema de bicicletas públicas de la Ciudad Autónoma de Buenos Aires y de los recorridos realizados por los mismos (ambos grupos de *datasets* disponibles en <https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>)

El resultado del trabajo será expuesto en 2 partes:

1. Limpieza: se resumirá la labor hecha para que las bases de datos puedan ser utilizadas.
2. EDA: se utilizará conjuntamente visualizaciones y texto para exponer la información que se obtuvo analizando los datos limpios, justificando la relevancia de la misma según corresponda.

El trabajo de ciencia de datos fue realizado en base al lenguaje Python 3 en el sistema Jupyter Notebook. Debido a la naturaleza y el objetivo del informe, no se insertaron fragmentos de programación. Para analizarlos, los códigos y el presente informe se encuentran disponibles en el siguiente repositorio GitHub:

[https://github.com/Agustin-Bulzomi/Projects/tree/main/Programming/Gobierno%20Abierto%20\(Python\)](https://github.com/Agustin-Bulzomi/Projects/tree/main/Programming/Gobierno%20Abierto%20(Python))

Se encuentran allí dos Jupyter Notebooks, dividiendo el proceso en las mismas dos instancias que se presentan en este informe: una para la limpieza y otra para el EDA. Para reproducir los códigos, hay *datasets* que fueron subidos al repositorio online y cuyos links están actualizados en el código para que se los utilice remotamente. Las bases de datos más pesadas (los 7 años de recorridos) no pueden ser ubicadas en el repositorio, por lo que deben ser descargadas vía la primera página web mencionada (el link vía URL para la conexión remota al .csv estaba caído al momento de la creación del informe, por lo que la descarga para utilización local fue la solución

Limpieza

Dataset: recorridos realizados por año

El primer paso para trabajar los datos consistió en analizar la estructura de los *datasets*, chequear inconsistencias y definir qué sería utilizado a futuro.

1) Se observó que el género solo estaba disponible en las bases de datos del 2015 al 2018. Considerando la riqueza de análisis que puede proveer como variable, se decidió imputar la información para usar todos los años.

2021

Se comenzó con el único año cuyo *dataset* presentaba el nombre del usuario: descargando una base de datos de nombres y géneros se aplicó un *merge* que logró imputar un 94% de las filas, sin asignar ningún género erróneo a los usuarios. El 6% restante se completó de manera aleatoria, lo cual asignó inevitablemente mal algunos géneros, pero no produjo un detrimento de la validez del *dataset* pues se realizó respetando las proporciones de género presentes en el 94% válido (0.656 “masculino”, 0.344 “femenino”). Esta decisión es estratégica ya que, al fin y al cabo, la información que se quiere recolectar no es el género de cada individuo sino las proporciones de género de los usuarios.

2020 - 2019

Gracias a que los *datasets* del 2019, 2020 y 2021 tenían como punto en común ser los únicos con *id_usuario*, se generó un *dataframe* en base al 2021: cada *id_usuario* con su respectivo género. El mismo fue luego aplicado sobre el 2019 y el 2020 para que cada fila con *id_usuario* recibiera el género encontrado en el 2021 (de validez casi absoluta: 94%). En el 2020 se logró imputar un 67% de los casos, por lo que se completó el 33% restante de misma manera que en el 2021: asignando aleatoriamente, pero respetando las proporciones. Lamentablemente, en el 2019 solo se logró imputar un 20% lo cual, sumado al 3% ya presente, no alcanzó al cuarto de la información válida. Debido a la baja representatividad de la muestra respecto del universo de unidades del 2019, se decidió entonces imputar de otra manera: se calculó un promedio entre las proporciones del 23% válido del 2019 (0.74 “masculino, 0.26 “femenino”) y el promedio entre las proporciones del año anterior (2018) y el posterior (2020) (0.691 “masculino, 0.309 “femenino”). Asumiendo una tendencia relativamente estable de las proporciones de género a lo largo de los años (que luego se graficó y corroboró), este sistema de imputación provee de robustez al resultado.

Los años anteriores estaban completos. Así, se logró el 100% de las +14.000.000 de filas con un riesgo bajo de pérdida de validez

2) Había una diferencia en el formato de la duración de los recorridos: del 2015 al 2019 estaban en segundos y del 2020 al 2021 en formato “dd:mm:ss”. Se procedió con un *split* del texto de los últimos años para sustraer los minutos y ubicarlos en una nueva columna. Para los casos del 2015 al 2019 se dividió de manera vectorizada para transformar los segundos en minutos, redondeando para que no haya una infinidad de diferentes valores según el resto (segundos) resultantes de la división.

3) Se encontró una diferencia en los nombres de las *features* según el año de la información, lo cual suele ser arreglado prontamente unificando los nombres de las columnas. Sin embargo, es menester denotar un caso particular: la diferencia entre *id_estacion* y el número adjunto en los nombres de las estaciones parecía ser más que un problema de unificación. Se decidió remitir a las fuentes, pesquizando las dos bases de datos de estaciones de bicicletas públicas de la Ciudad Autónoma de Buenos Aires (disponibles en <https://data.buenosaires.gob.ar/dataset/estaciones-bicicletas-publicas>).

Al observar la información allí presente se encontró un cambio en la nomenclatura cuando se sustituyó el sistema utilizado para las bicicletas a partir del 2019: el número que antes era *id_estacion* ahora es el código y se creó un nuevo *id_estacion* (en la mayoría de los casos, similar pero distinto al código). En las bases de datos, la única forma de mantener homogéneo el sistema de identificación de estaciones era utilizando el código (*id_estacion*) de los años 2015 a 2018 y obteniendo el código que aparecía antepuesto al nombre en los años 2019 a 2021:

	duracion_recorrido	id_estacion_origen	nombre_estacion_origen	id_estacion_destino	nombre_estacion_destino
0	447	323	240 - ECHEVERRIA	289	255 - BARRANCAS DE BELGRANO
1	438	167	275 - PLAZA 24 DE SEPTIEMBRE	273	223 - GAINZA
2	414	247	282 - Tronador	400	313 - De Los Incas
3	38	158	158 - VILLARROEL	158	158 - VILLARROEL
4	484	29	029 - Parque Centenario	99	099 - Malabia
...
1144217	690	277	292 - PLAZA BOLIVIA	44	044 - Ecoparque
1144218	1360	79	079 - AZUCENA VILLAFLO	168	168 - Estados Unidos
1144219	1169	79	079 - AZUCENA VILLAFLO	8	008 - Congreso
1144220	1031	79	079 - AZUCENA VILLAFLO	75	075 - Plaza Primero de Mayo
1144221	1795	79	079 - AZUCENA VILLAFLO	207	123 - Armenia

Fig. 1: impresión de pantalla con una muestra de 10 filas en los extremos del dataset 2021.

Fuente: elaboración propia

Se procedió, entonces, creando dos nuevas columnas llamadas “*codigo_origen*” y “*codigo_destino*” y completándolas con el *integer* del texto con números obtenido al dividir en partes el nombre de las estaciones. Antes, se debió superar un par de obstáculos: a) no todas las filas tenían el código antepuesto, por lo que en casi todas las estaciones se duplicaban los nombres (uno con el código y otro sin); b) no todos los nombres tenían el mismo formato (algunos tenían tildes y otros no, algunos estaban en mayúsculas y otros no, algunos estaban capitalizados y otros no, algunos tenían “&” y otros “y”, etc.); c) algunos tenían errores de tipeo (“Fitz Roy y Gorrtiti”) o de espaciado (“F.J.Santamaría de Oro”), etc. Una vez resuelta la unificación se obtuvo un *dataframe* con cada uno de los 396 códigos con sus respectivos nombres e *id_estacion* viejos. Con él, se procedió a imputar todos los casos necesarios (código erróneo/faltante, nombre erróneo/faltante, latitudes y longitudes erróneas/faltantes).

	minutos	id_usuario	año	genero_usuario	codigo_origen	nombre_origen	codigo_destino	nombre_destino
0	15	666202	2020	F	116	hospital aleman	142	armenia y gorriti
1	11	666202	2020	F	14	pacifico	142	armenia y gorriti
2	17	666202	2020	F	134	don bosco	142	armenia y gorriti
4	12	666202	2020	F	194	peron y acuna de figueroa	142	armenia y gorriti
5	27	666202	2020	F	142	armenia y gorriti	142	armenia y gorriti
...
22202	26	279101	2020	M	150	rodrigo bueno	111	macacha guemes
23398	15	690136	2020	M	393	barrio 31	3	aduana
26597	51	1827	2020	M	373	jorgelina de simone	240	echeverria
30309	19	92175	2020	F	196	hospital argerich	126	ministerio de justicia y seguridad
34447	31	588772	2020	M	108	usina del arte	199	estados unidos y boedo

Fig. 2: impresión de pantalla con una muestra de 10 filas en los extremos del dataset 2020.


Fuente: elaboración propia

4) Al final de la *notebook*, se eliminaron los *outliers*, siendo necesario borrar filas solo en la columna numérica de duración. Utilizando los límites 0 y 60 minutos provistos por el sistema, se quitaron 176.121 filas (1,24%).

Resultado

El dato *id_usuario* fue el único imposible de obtener, pues incluso con el listado de las bases de datos de usuarios del sistema no podría asignarse el identificador a cada recorrido sin un punto que los relacione.

El resultado de los 4 pasos anteriores llevó del total inicial de 14.282.666 filas y 12 columnas con nulos en todas salvo en “año” a un *dataset* limpio de 13.969.250 filas (97.8%) y 12 columnas sin nulos



año	0	año	0
genero_usuario	9310370	genero_usuario	0
id_estacion_origen	48340	codigo_origen	0
nombre_estacion_origen	4	nombre_origen	0
long_estacion_origen	1197873	long_estacion_origen	0
lat_estacion_origen	1197873	lat_estacion_origen	0
duracion_recorrido	43872	minutos	0
id_estacion_destino	48922	codigo_destino	0
nombre_estacion_destino	6	nombre_destino	0
long_estacion_destino	1198771	long_estacion_destino	0
lat_estacion_destino	1198771	lat_estacion_destino	0
id_usuario	2148217	id_usuario	2141938

Figs. 3 y 4: impresiones de pantalla con sumatoria de nulos por columna en el *dataset* inicial y en el final.

Fuente: elaboración propia

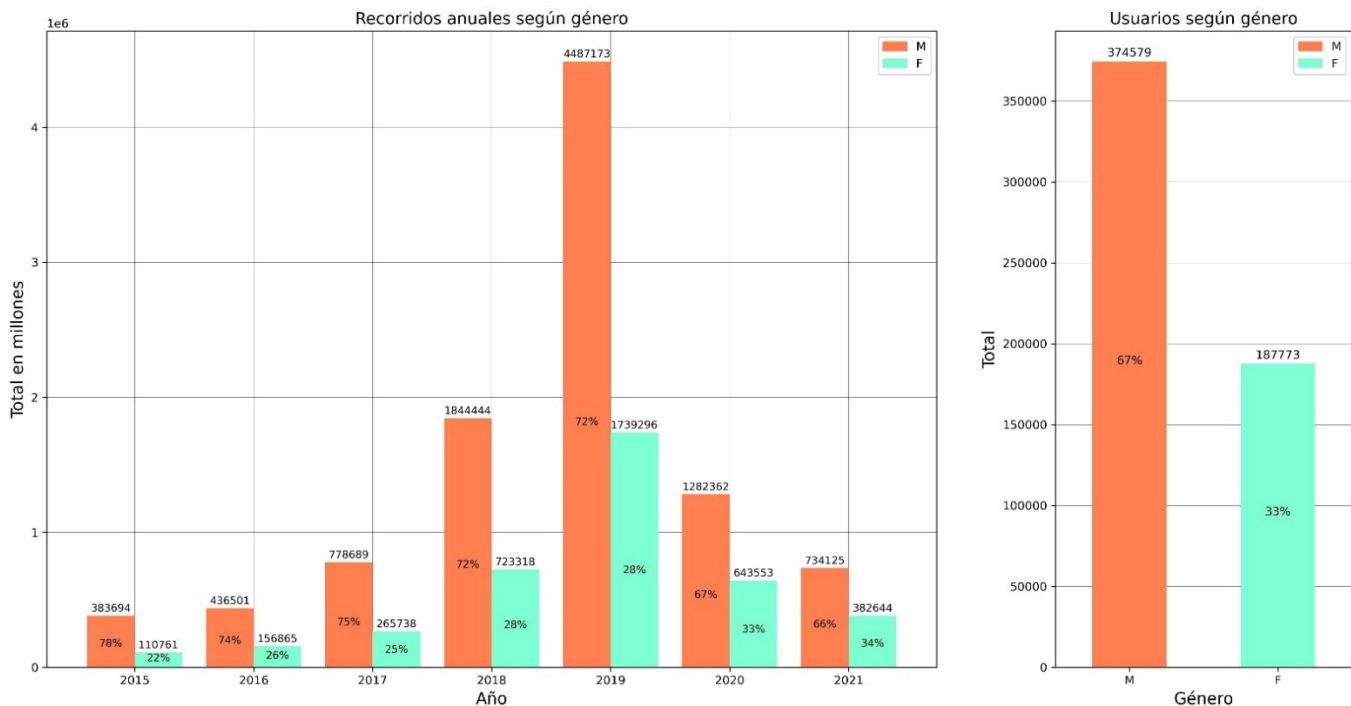
El *dataset* de usuarios no requería de una limpieza tan profunda, y pudo ser utilizado casi directamente en el análisis exploratorio, sobre el cual se explayará a continuación.

EDA

El Análisis Exploratorio de Datos no procura generar muchas métricas estadísticas pues la intención del informe es presentar de manera visual y clara la información que se puede obtener de las bases de datos. En ese sentido, y considerando que el factor clave en esta labor es producir conclusiones aplicables para el sector público y privado, se deja de lado el análisis más numérico y se enfoca en visualizaciones e inferencias de las mismas. Para un enfoque más computacional/estadístico, hay más información en las Jupyter Notebooks.

Gráfico 1: análisis de género - barras

Se comenzó agrupando los recorridos según año y calculando los totales de recorridos por cada género. Para este caso, el valor “N” / “No informado” fue quitado pues su baja representatividad (menor al 0.00%) generaría una columna vacía. Debido a la enorme diferencia en proporciones de recorridos realizados por personas masculinas vs recorridos realizados por personas femeninas, se procedió también a calcular la proporción de usuarios según su género. Se graficaron ambos datos utilizando la visualización vía barras, ubicando totales encima de cada bloque y porcentaje por sobre el total anual en el centro de cada bloque:



Figs. 5 y 6: recorridos anuales según género y usuarios según género. Fuente: elaboración propia

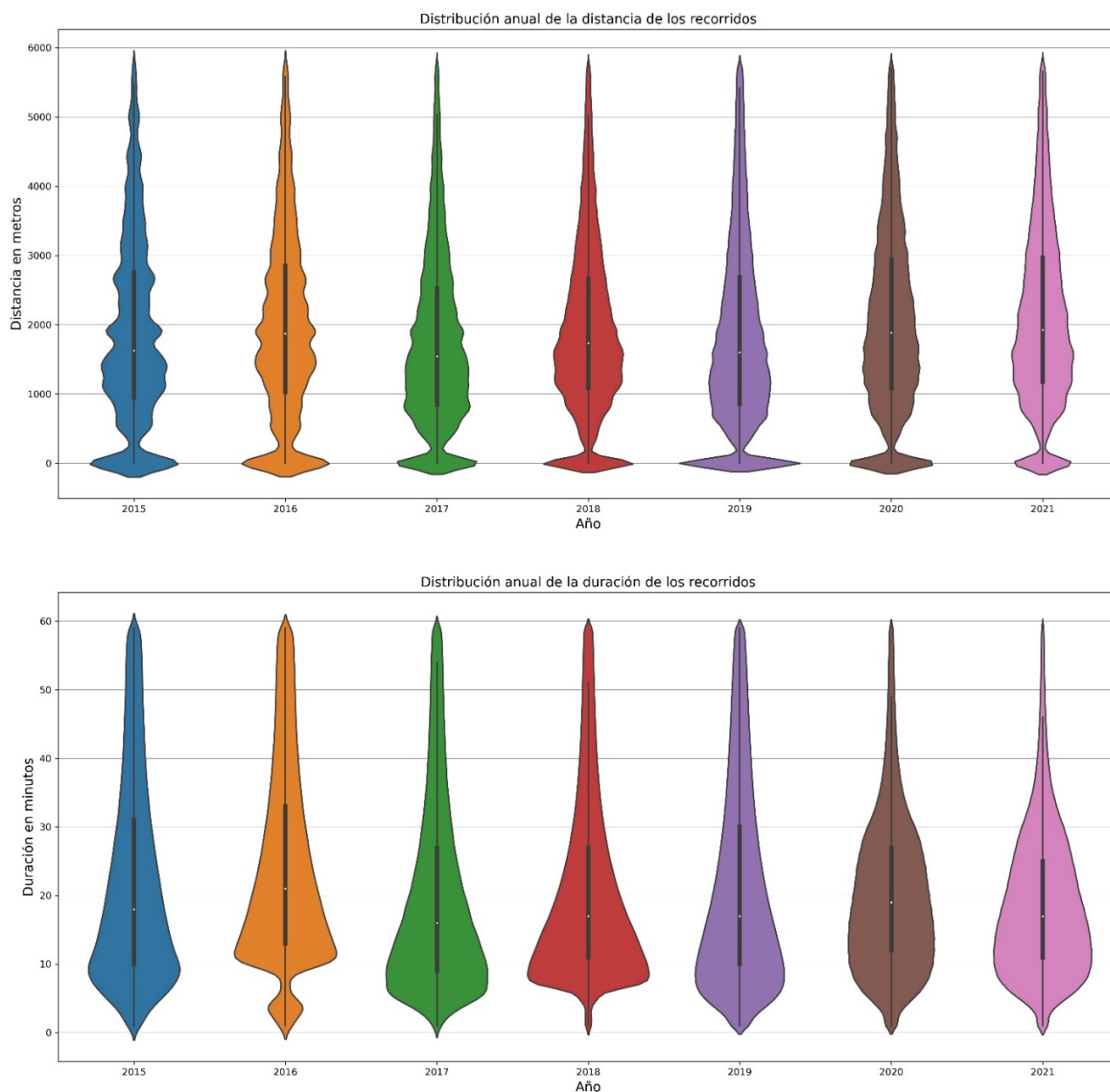
Algunas inferencias importantes a realizar son:

- 1) Si bien es sutil, en el primer gráfico se ve un aumento en los recorridos hechos por mujeres que es proporcionalmente mayor al aumento en los recorridos hechos por hombres, como puede observarse en el 22% de recorridos en el 2015 vs. el 34% en el 2021. Sin embargo, la proporción de recorridos al 2021 según género sigue siendo sujeto de análisis: solo 1/3 de los mismos es realizado por una mujer. Tanto para el sector público como para el sector privado esto puede ser importante a la hora de definir estrategias futuras que capten nuevos usuarios: ¿conviene apelar a los hombres porque usan más las bicicletas o conviene apelar a las mujeres porque todavía no incursionaron tanto en el sistema?
- 2) El segundo gráfico demuestra que lo susodicho no es causado por una diferencia en la cantidad de recorridos por usuario según su género, sino que hay más hombres en el sistema. Para el sector público este dato es importante pues se podría plantear una política de género: ¿existe algún obstáculo cultural, laboral o de algún ámbito por el cual esto se da? ¿Hay menor interés por parte de las mujeres a utilizar el sistema de bicicletas públicas?
- 3) El primer gráfico muestra que la cantidad de recorridos se redujo drásticamente en el 2020 por la pandemia. A su vez, por el ritmo actual parece que este año habrá más recorridos que en el anterior. ¿Se puede inferir que la movilidad ciudadana es mayor en comparación al año anterior? Estos datos en clave mensual podrían ser útiles para que el Estado calcule el efecto de la cuarentena y el grado de cumplimiento de las medidas de aislamiento.
- 4) A su vez, se observa una tendencia exponencialmente alcista en los primeros años: en tan solo 4 años la cantidad de recorridos se multiplicó por casi 13 (un aumento del 1185% aproximadamente). Para el sector privado este dato es fundamental pues demuestra un potencial de lucro muy grande: asumiendo que el descenso en la cantidad no hubiera sucedido de no haber pandemia, se puede plantear una investigación de mercado para calcular si hay todavía en el mercado una demanda de bicicletas sin satisfacer que permita un crecimiento de las mismas dimensiones. A la hora de adentrarse en este análisis, es importante definir los límites estructurales y culturales: los Países Bajos no llegaron a los impresionantes números de utilización de bicicletas sin un rediseño de todo el sistema

transporte de la ciudad y una concientización general. Es decir, no solo habría que estudiar la posible demanda por satisfacer sino también si la misma puede ser resuelta solo con una mayor oferta de bicicletas o si habrá un tope externo que imposibilite el aumento exponencial que se observó hasta ahora. Para el sector público, esto significa analizar si solo se deben ubicar más bicicletas y estaciones en el sistema o si se debe renovar el sistema de transporte urbano, la dirección de las calles, las señalizaciones, reeducar, etc.

Gráfico 2: análisis de duración y distancia de recorridos - *violinplots*

Se calculó la distancia geodésica entre ambas estaciones del recorrido. Para ello fue necesario generar valores de tipo *point* utilizando las latitudes y longitudes de las estaciones. Luego de quitar los *outliers*, se graficó un conjunto de *violinplots*, los cuales, además de mostrar el rango intercuartílico, la mediana y los extremos como lo hacen los *boxplots*, exponen la distribución de los datos a lo largo de todos los valores. En este caso, ambos gráficos expresan con la variación de los anchos la distribución de la duración y de la distancia de los recorridos:



Figs. 7 y 8: distribución anual de la distancia y de la duración de los recorridos. Fuente: elaboración propia

Algunas inferencias importantes a realizar son:

- 1) La distancia y la duración se mantuvo relativamente estable a lo largo de los años. Es más, la distancia se tornó más homogénea a lo largo del tiempo (como puede verse por la menor variación en el ancho de los *violinplots*). Esto puede ser importante para analizar el efecto de una mayor disposición de estaciones en el sistema: si luego de un tiempo las distancias y las duraciones de los recorridos no aumentan, es posible inferir que los máximos y los promedios de distancias y duraciones no cambien al proveer de más estaciones.
- 2) Se observa una considerable proporción de valores 0 en la distancia, que deben representar recorridos que finalizaron en la estación de origen. Esto puede tener implicancias importantes que se analizarán en otro gráfico.
- 3) A lo largo de los años la duración de los recorridos se fue acortando. Como puede verse en el segundo gráfico, hay una tendencia de los *violinplots* a volverse más anchos en los valores de 5 a 30 minutos y a volverse considerablemente más finos a partir de los 30 minutos para arriba. Esto quiere decir que cada vez es menor la proporción de recorridos duraderos: los usuarios usan las bicicletas menos tiempo que antes o hay más usuarios que ahora usan las bicicletas, en promedio, menos tiempo que el año anterior.
- 4) Para el sector público, estos datos son importantes a la hora de definir las estrategias para proseguir con el sistema: no se requiere extender el tiempo a más de 60 minutos, al contrario: la gente que utiliza tanto las bicicletas no pareciera ser significativa. A su vez, en términos absolutos, el 75% de los usuarios recorren una distancia menor a 3000 metros y utilizan la bicicleta un tiempo menor a 25 minutos. Para el año 2021, las medianas rondan, respectivamente, en torno a los 1900 metros y 18 minutos. Esto es fundamental a la hora de definir qué estaciones podrían ser utilizadas como destino partiendo de una determinada estación de origen: es muy improbable que dejen la bicicleta en una estación que esté a más de 3000 metros. En consecuencia, se puede también analizar qué transporte público podría ser el que reemplaza el uso de la bicicleta en dichos usuarios: no sería probable que el 50% que viaja menos de 1900 metros requiera de un viaje de 2 conexiones entre colectivo y subte. Estos valores ayudan a planificar la óptima disposición de estaciones y bicicletas para las distancias y duraciones de trayectos demandados.

Gráfico 3: análisis de tipo de viaje - torta

Como se mencionó en el análisis anterior, los cálculos dieron como resultado una importante cantidad de recorridos que finalizan en la estación de origen. Se procedió a graficar una torta con las proporciones: 1 de cada 10 usuarios deja la bicicleta en donde la tomó.

Proporción de recorridos que finalizan en la estación de origen

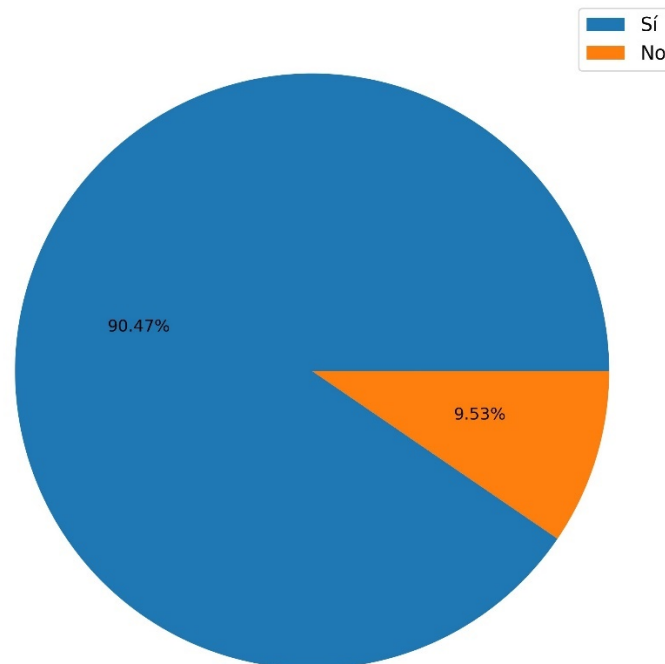


Fig. 9: proporción de viajes que finalizan en la estación de origen. Fuente: elaboración propia

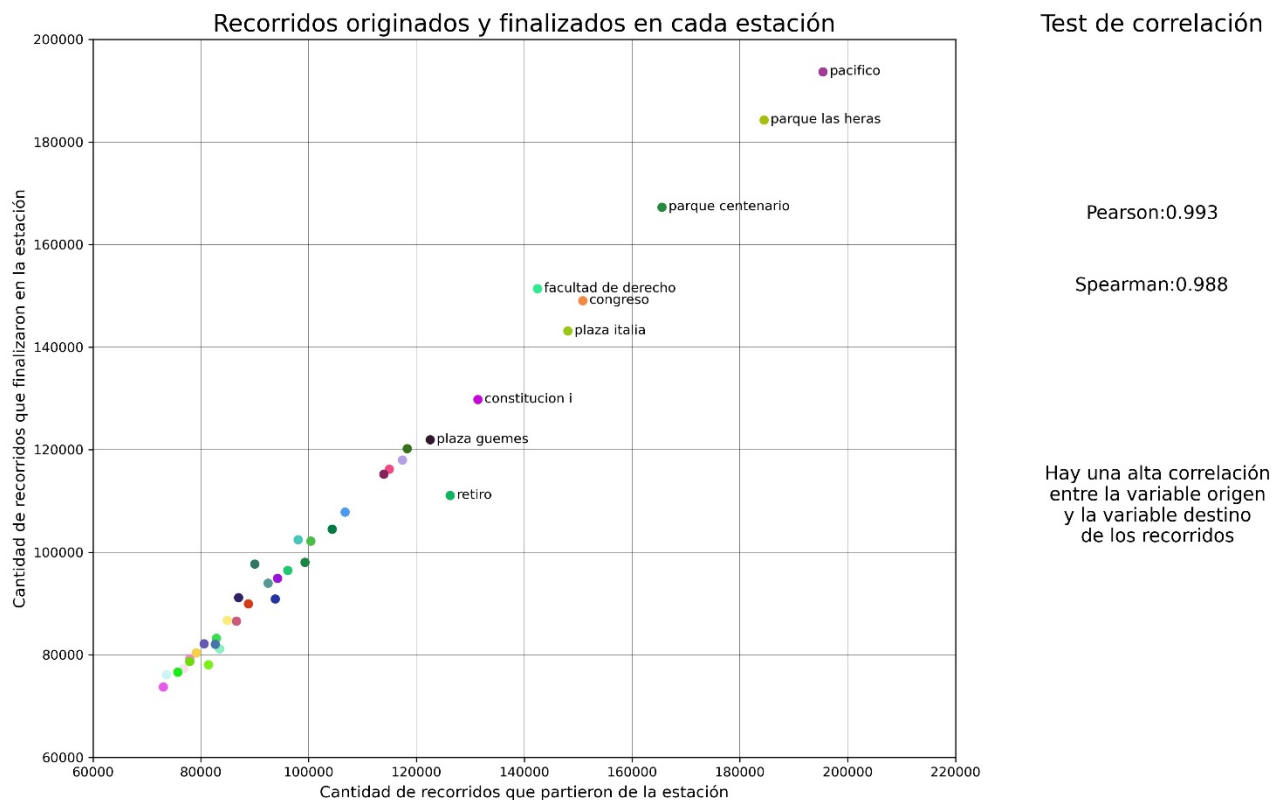
Algunas inferencias importantes a realizar son:

- 1) Podría distinguirse dos tipos de viajes en base al fenómeno presente: a) quienes dejan la bicicleta en otra estación requieren del sistema para transporte; b) quienes dejan la bicicleta en la estación de origen requieren del sistema para pasear/para hacer un trámite corto y volver. Esto es importante para el sector privado para considerar qué demanda podría alcanzar: a) los sistemas de *delivery* como Glovo, Rappi, PedidosYa o Wabi hace ya un tiempo ofrecen entrega de objetos personales, productos fuera del sistema de compras interno a la app y demás. ¿Hay una población que pagaría para no realizar ese trámite en bicicleta y que se lo realice un tercero vía la app? b) las empresas que ofrecen servicios de transporte individuales como los *scooters* Movo, Grin, Lime, las empresas que venden *scooters*, *rollers*, bicicletas o incluso calzado/ropa deportiva, ¿podrían apelar a ese 10% de recorridos posiblemente realizados por placer/ejercicio? Si se encuentra una demanda de medios de transporte para pasear/vestimenta para ejercitarse podría haber una oportunidad de mercado.
- 2) Para el sector público se podría analizar el caso inverso: 9/10 recorridos se usan para moverse de un lugar a otro. ¿Cuál es el flujo? El hecho de que el 90% de los recorridos sean para transporte y no para placer, ¿indica que hay una demanda de mejores condiciones de transporte en otros ámbitos? ¿Es posible que aquellos que utilizan el sistema de bicicleta para viajar lo hagan por oposición a un colectivo/subte muy atestado? ¿Si mejoran las condiciones en otros sistemas de transporte, dejarían las bicicletas? Para ese 1/10 de personas que viaja por placer/trámite, difícilmente la solución sea el subte, el colectivo o el taxi.

Gráfico 4: análisis de correlación entre estaciones (*scatter plot*)

Se procedió a analizar las estaciones más utilizadas. El cálculo reporta que el 30% de los recorridos parten de y/o finalizan en tan solo 40 estaciones (un 10% del total). Esto es un

indicador sumamente importante para que el sector público analice dónde podrían requerirse más estaciones y bicicletas. Luego se comparó las mismas para corroborar si había diferencias en cuántos recorridos partían de ellas y cuántos finalizaban en ellas. El gráfico ofrece respuestas:



Figs. 10 y 11: recorridos originados y finalizados en cada estación y test de correlación de variables. Fuente: elaboración propia

No se graficó la totalidad de los nombres de las 40 estaciones más usadas para dejar clara la visualización. Algunas inferencias importantes a realizar son:

- 1) Hay una enorme correlación entre cantidad de recorridos originados y cantidad de recorridos finalizados en una estación. La métrica informa, entonces, que es raro que una estación reciba muchas bicicletas, pero no sea requerida para partir o viceversa. Esto indica que el flujo de tránsito es relativamente equilibrado y no hay sectores de la ciudad en donde se demanden bicicletas para ir a otro lado sin que haya personas que quieran ir de otro lado a esa estación. Para el sector público esto es importante por el planeamiento del tránsito y de las futuras estaciones: no pareciera haber una demanda localizada que muestre un flujo unidireccional de tránsito. Al menos, eso pareciera surgir de la cantidad de recorridos, pero es menester aclarar que no se consideran horarios por lo que quizás la demanda de origen es para viajar al trabajo a la mañana y de destino es cuando se vuelve a la tarde. En ese sentido, entonces sí hay un flujo unidireccional temporalmente que podría explicar posibles desabastecimientos de bicicletas en horarios específicos.
- 2) Para el sector privado, estas estaciones tan requeridas representan un enorme caudal de tránsito en la zona, podrían ser excelentes ubicaciones para ofrecer servicios de transporte como trayectos nuevos de colectivos, búsqueda de clientes de taxis, ubicación de los *scooters* mencionados anteriormente, etc.

Gráfico 5: análisis de usuarios según edad (histograma)

Como el *dataset* de recorridos no ofrecía la variable “edad”, se utilizó el conjunto de bases de datos de usuarios disponible en la misma página. El resultado, entonces, no explaya la cantidad de recorridos hechos sino la cantidad de usuarios según la edad:

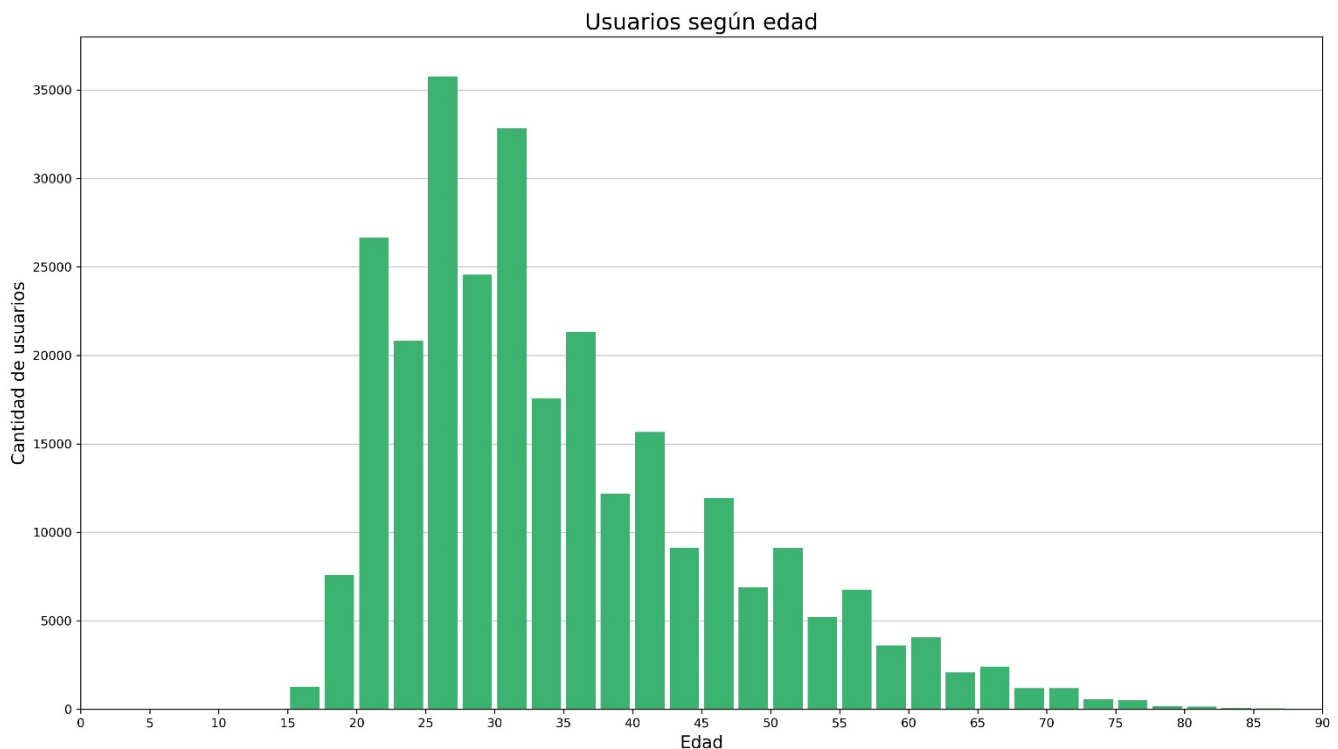


Fig. 12: usuarios según edad. Fuente: elaboración propia

La inexistencia de valores para las barras de histograma menor a 15 se explican por el límite de edad para usar el sistema. De ahí en adelante, se pueden hacer unas inferencias importantes:

- 1) Para el sector privado, si una empresa ofrece servicios de transporte puede encontrar aquí el rango etario más demandado para el uso del sistema de bicicletas. En caso de querer vender bicicletas u ofrecer un servicio de transporte alternativo que compita con las mismas, esta información es clave para determinar la estrategia de mercado y de publicidad. Siendo la enorme mayoría usuarios menores a 37.5 años, es posible que sea conveniente adoptar medidas de acercamiento más modernas para apelar a los jóvenes: redes sociales o ciertos canales de televisión, pero no diarios o radio.
- 2) Para el sector público, es importante analizar por qué sucede esto. Es claro que hay menor población en los rangos etarios más grandes y, más aún, es poco probable que usen la bicicleta seguido por las limitaciones de salud. Sin embargo, es fundamental analizar si las diferencias en uso se dan solo por ese factor o si hay otras variables a considerar: ¿es accesible el sistema para la gente mayor? ¿es intuitivo para personas no acostumbradas al uso de los celulares? Contrario al análisis anterior, para el sector público podría ser una política sanitaria importante no apelar a la enorme proporción de usuarios jóvenes que ya demuestran un enorme interés en el sistema de bicicletas, sino a la población de 40 a 70 años que lo usa poco y podría ser la que más beneficios de salud y calidad de vida podría encontrar si comenzara a movilizarse con bicicletas. Sería hasta económicamente estratégico invertir para acercar el sistema a

gente mayor pues las externalidades positivas que resulten de ello compensarían la inversión al requerir, quizás, menores gastos en medicamentos u hospitales públicos para tratar enfermedades cada vez más presentes por el sedentarismo contemporáneo.

Gráfico 6: análisis de usuarios según cantidad de viajes - histograma

Finalmente, se analizó el total de recorridos según cantidad de usuarios para observar el uso el sistema, ¿1000 recorridos son 1000 usuarios que usan el sistema una vez o 10 que lo usan 100 veces? Debido a la enorme diferencia que se observa en los valores del primer gráfico, se expuso nuevamente debajo usando la escala logarítmica, disminuyendo así diferencias y, en consecuencia, comparando mejor los valores bajos. Al aparecer visibles más resultados, se gira para agrandar el rango del eje con los valores de la variable “cantidad de usuarios”:

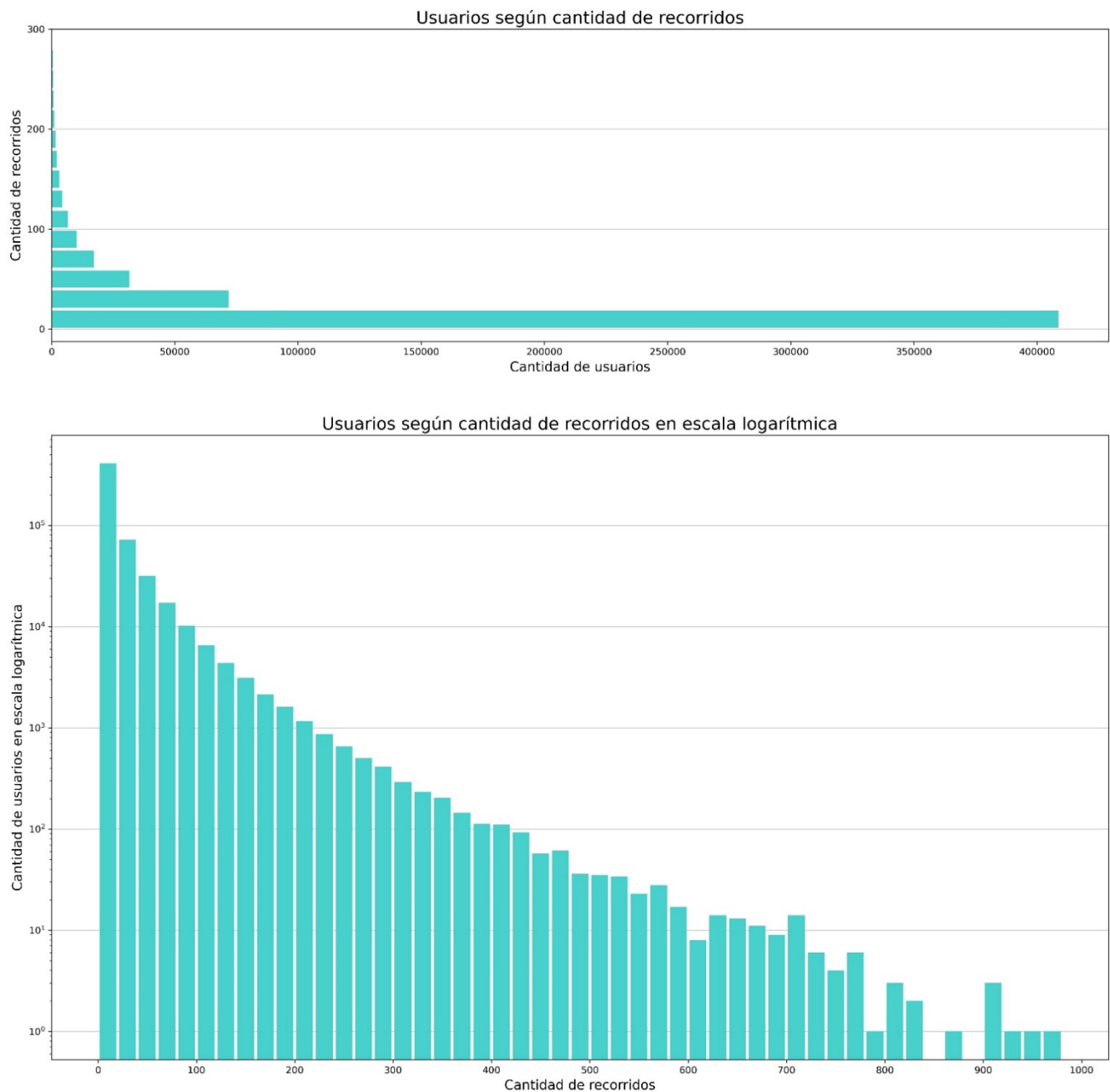


Fig. 13: usuarios según cantidad de viajes. Fuente: elaboración propia

Una inferencia importante a realizar es: la enorme mayoría de los usuarios realizó menos de 20 viajes. Esto es decisivo para el análisis del sistema de bicicletas: ¿por qué los usuarios entran al sistema e igualmente lo usan poco? ¿No les interesa tanto usar las bicicletas como para que sea un uso rutinario o es que hubo algún motivo por el cual abandonaron el uso? Este aspecto es crítico a la hora de alcanzar a la ciudadanía: si los usuarios no usan las bicicletas muchas veces, es posible que invertir el dinero en otros medios de transporte genere mayores utilidades a la población; por el contrario, si los usuarios quieren usar las bicicletas, pero desisten por algún obstáculo/algún disgusto, es posible que invertir el dinero para solucionarlo genere un aumento considerable en cantidad de recorridos por usuario que ya ingresó al sistema. ¿Habrà que mejorar la app? ¿Habrà que mejorar las condiciones de las bicicletas? En cualquier caso, es un dato importante el hecho de que la gente pase por el trámite de generar su usuario y pero use tan pocas veces el sistema.