

TP1

Índice

Alumnos	1
Introducción	1
Objetivos	1
Análisis	2
Análisis univariado	2
Edad	2
Género	5
Día de la semana	6
Duración	8
Distancia	9
Dirección de origen y dirección de destino	10
Análisis Bivariado	13
Conclusiones	14

Alumnos

Lucas Bachur, Sebastián Mestre, Santiago Coronel.

Introducción

EcoBici es un sistema de bicicletas compartidas que funciona en la Ciudad de Buenos Aires, cuenta con más de 200 estaciones y 1.200 rodados. Actualmente el sistema está presente en 30 de los 48 barrios de la Ciudad Autónoma de Buenos Aires, funcionando los 365 días del año 24 horas. Existen 2 formas de usar el sistema: mediante una app o mediante una tarjeta. Mediante la app, el usuario debe registrarse y, mediante su nombre de usuario y password, ingresa al sistema, elige la estación en la que se encuentra y luego elige la bici que quiere. Mediante tarjeta, el usuario debe apoyarla en un lector y el sistema le asigna una bicicleta. El usuario se moviliza por la ciudad durante un tiempo limitado (30 minutos) y luego debe devolver la bicicleta en alguna de las estaciones habilitadas. El sistema le asigna al usuario un id y registra datos tales como la edad, el género, estación de origen, estación de destino... etc.

Objetivos

El objetivo de este informe es reflejar las características más notables acerca de los datos brindados. Para llevar esto a cabo se va a utilizar un software estadístico que permita efectuar un estudio descriptivo que incluya: tablas de distribución de frecuencias, gráficos y medidas descriptivas.

Análisis

Análisis univariado

Edad

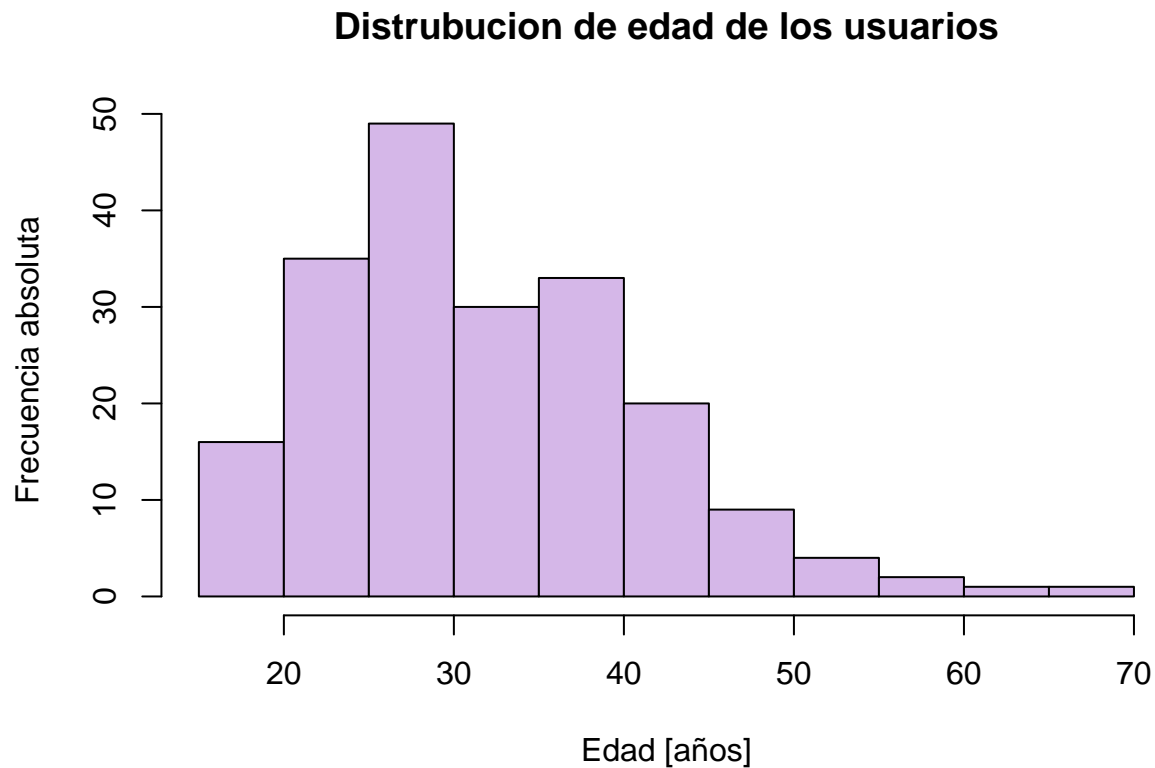


Figura 1: Edad Usuarios

Se observa en el histograma de la Figura ?? que la mayoría de los usuarios estan en las franjas etarias más bajas. Más específicamente, se ve una concentracion de usuarios en la franja de 20 a 40 años, con un pico en el intervalo [25, 30]. Se detecta que usuarios con una edad mayor a 40 años hace un menor uso del servicio. Una aclaración importante es que el número de personas que figuran entre 15 y 20 años es baja, ya que sólo estamos teniendo en cuenta los mayores a 18.

Cuadro 1: Frecuencias Edad Usuarios

	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
[19,24)	30	30	0.15	0.15
[24,29)	56	86	0.28	0.43
[29,33)	32	118	0.16	0.59
[33,38)	29	147	0.145	0.735
[38,43)	21	168	0.105	0.84
[43,48)	19	187	0.095	0.935
[48,53)	7	194	0.035	0.97
[53,57)	3	197	0.015	0.985

	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
[57,62)	1	198	0.005	0.99
[62,67)	2	200	0.01	1
Total	200	1525	1	

En la tabla ?? podemos ver incluso mejor como llegado a los 33 años, la frecuencia acumulada es casi del 60%.

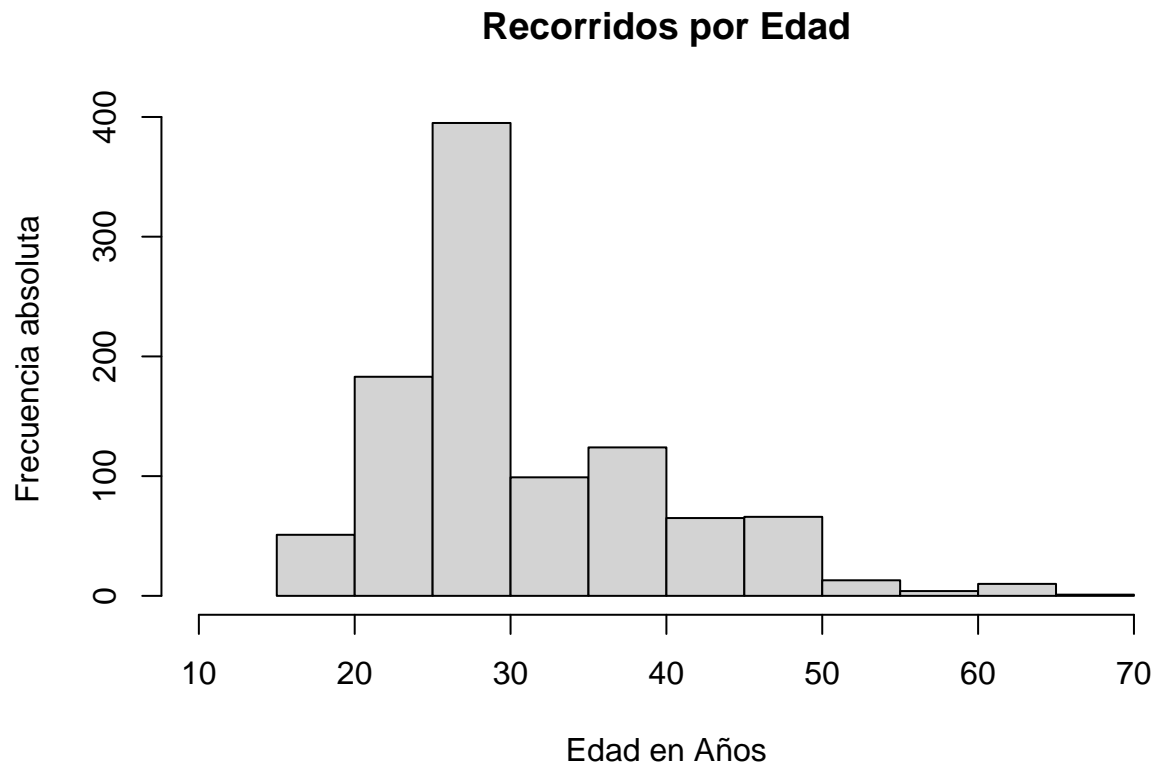


Figura 2: Edad Recorridos

El gráfico ?? de distribución de recorridos por edad nos muestra resultados similares a lo visto en el análisis de usuarios por edad. Sin embargo, una cosa interesante sucede cuando se analiza los viajes realizados por usuario, repartiendolos en cada edad. Se puede ver que si bien hay una leve tendencia a que baje este número, no es nada determinante. O sea, lo que podemos decir es que el hecho de que haya menos viajes realizados por gente mayor a 40 años, no tiene que ver tan directamente con que cada usuario haga menos viajes. Sino que se puede asociar también con el hecho de que son menos los usuarios en esas edades. Esto es lo que termina causando el tamaño de la disparidad.

Género

Se analizará ahora la distrución del género. Algo interesante para acotar es que si un usuario no proporciona esta información, por defecto su género asignado será “Otro”. Esto significa que los datos no representan exactamente la distribución de género real.

Usuarios por Genero

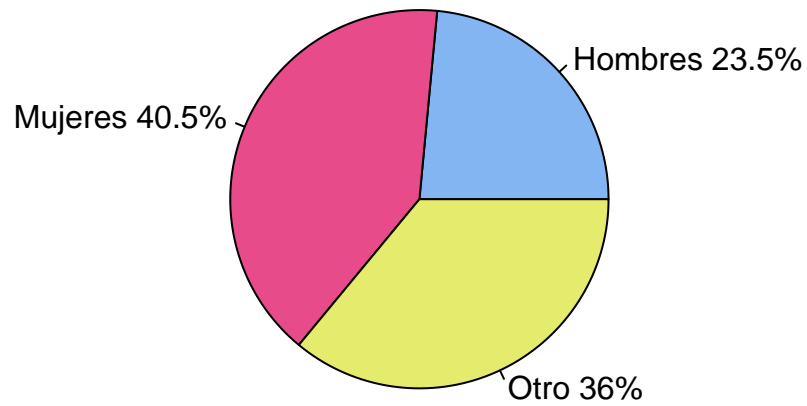
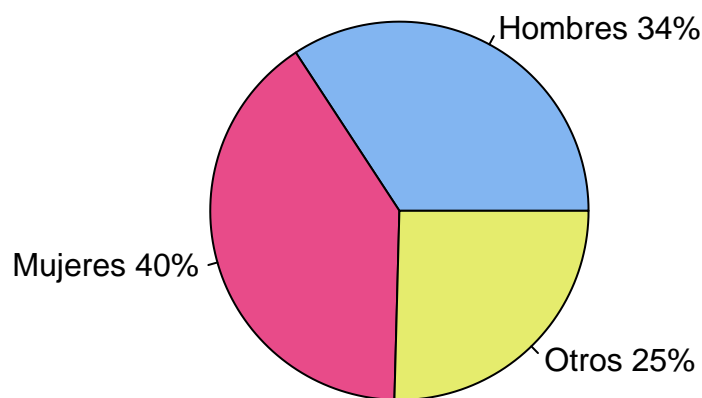


Figura 3: Genero Usuarios

Viajes por Genero



De los gráficos ?? y (??) se puede observar que:

- la mayoría de usuarios son mujeres
- la mayoría de viajes lo realizan mujeres

Estos datos tienen una relación directa y se puede decir que la razón por la que más viajes los realizan personas de género femenino es que estos son la mayoría.

Día de la semana

Vamos a analizar ahora la distribución de viajes que se realizan por día de la semana.

La distribución muestra que el día en que se realizan más viajes es el lunes, mientras que los días con menor frecuencia son viernes y domingo. La disminución de viajes en los 2 días recién mencionados puede atribuirse al hecho de que las personas pueden tender a quedarse en sus hogares a descansar durante el fin de semana, o bien realizan actividades fuera de la casa, pero se transportan de otra manera.

Cuadro 2: Frecuencias Genero por Edad

	Frecuencia Abs	Frecuencia Rel	Frecuencia Acum Abs	Frecuencia Acum Rel
Lunes	161	0.1592483	161	0.1592483
Martes	145	0.1434224	306	0.3026706
Miercoles	146	0.1444115	452	0.4470821
Jueves	146	0.1444115	598	0.5914936
Viernes	137	0.1355094	735	0.7270030
Sabado	141	0.1394659	876	0.8664688
Domingo	135	0.1335312	1011	1.0000000

En esta tabla se observa con mayor precisión la diferencia en frecuencia entre cada día.

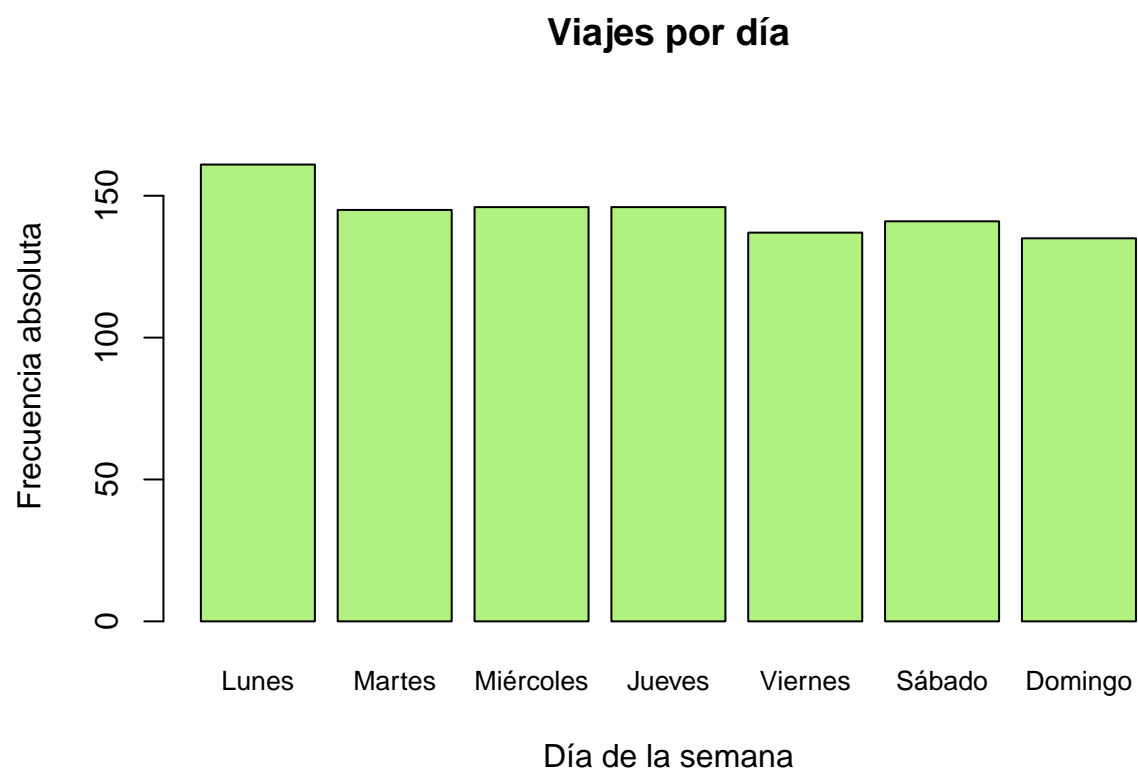


Figura 4: Días de la semana

Duración

Se hace primero el análisis del gráfico (??) de la distribución de duración:

Distribución de la duración de los recorridos (en minutos).

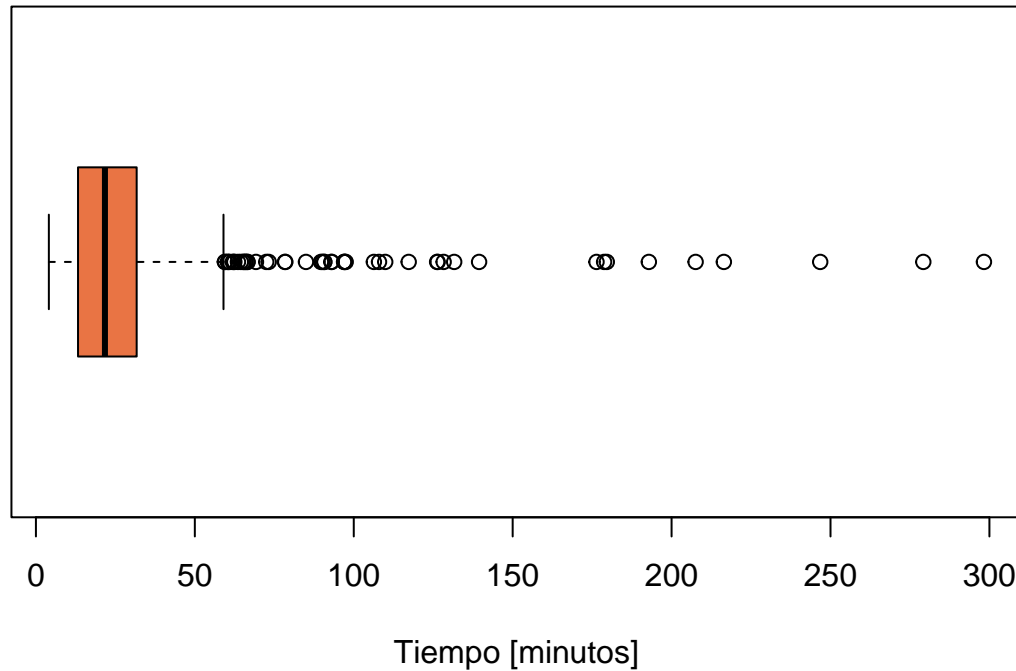


Figura 5: Duracion Recorridos

Al observarlo, se nota que hay muchos valores alejados de la media. Estas son algunas de las medidas descriptivas de la variable (en minutos):

- Mínimo: 4
- Primer Cuartil: 13
- Mediana: 22
- Promedio: 27
- Tercer Cuartil: 32
- Máximo: 298

Se observa que hay algunos valores outliers muy por encima de la media. Podría considerarse la posibilidad de que sean errores de registro o del sistema; aunque en este caso también podrían adjudicarse a un error/eventualidad en relación al usuario.

Por otro lado, interpretamos que la mayor concentración de outliers que se encuentran en el intervalo $[60, 100]$ son resultados lógicos del factor humano, ya que representan un intervalo de tiempo razonablemente menor.

Distancia

En este análisis, sucede lo mismo que en el anterior: es conveniente mostrar la gráfica sin outliers. El análisis va a ser directamente de esta.

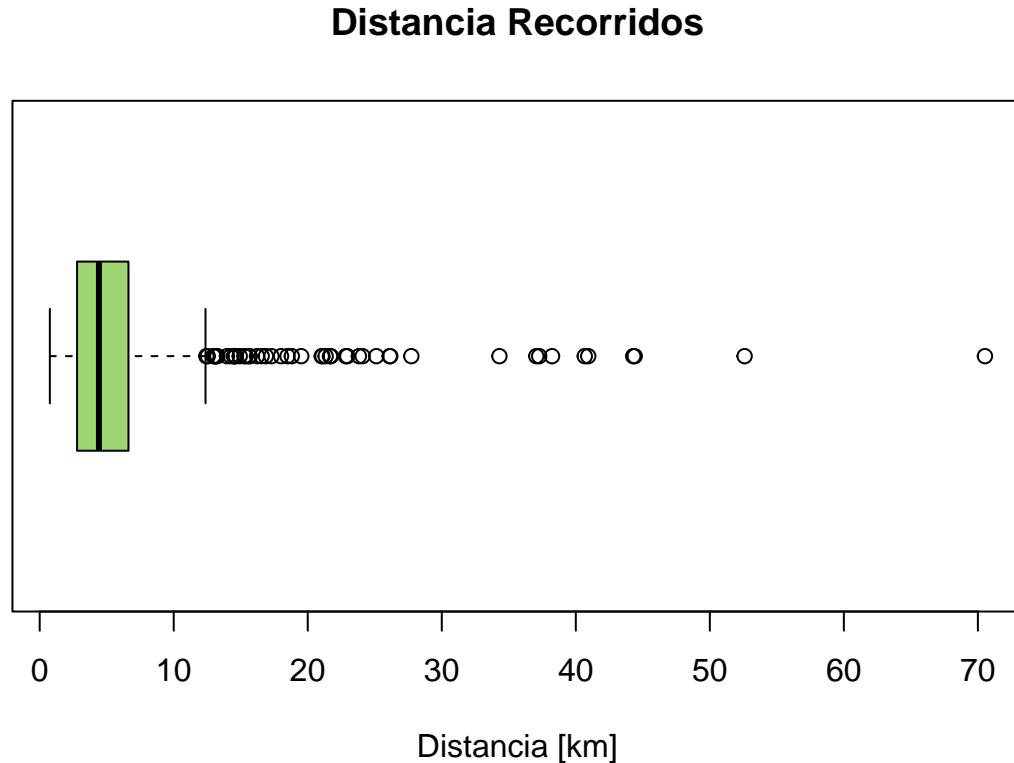


Figura 6: Distancia Recorridos

Medidas descriptivas de la variable (en metros):

- Mínimo: 758
- Primer Cuartil: 2788
- Mediana: 4422
- Promedio: 5567
- Tercer Cuartil: 6625
- Máximo: 70528

La conclusión que se puede sacar es que la mayoría de los viajes son de media distancia, variando principalmente entre los 3 y 6 kilómetros. No hay gran cantidad de viajes muy cortos ni muy largos, aunque efectivamente existen.

Recordando el análisis anterior (*referencia*) sobre la duración de los recorridos, notamos que existe cierta correspondencia entre la distribución de los outliers. Esto puede deberse a que los usuarios que debían viajar largas distancias no cumplieron con el tiempo debido, o bien simplemente decidieron no respetarlo. Concluimos que existe evidencia a favor de extender el tiempo de uso de las bicicletas, o al menos considerar una política de uso acondicionada a los viajes de mayor distancia .

Dirección de origen y dirección de destino

Para esta variable, el análisis va a ser alrededor de **cuántas** estaciones fueron salida o llegada de cada cantidad de viajes:

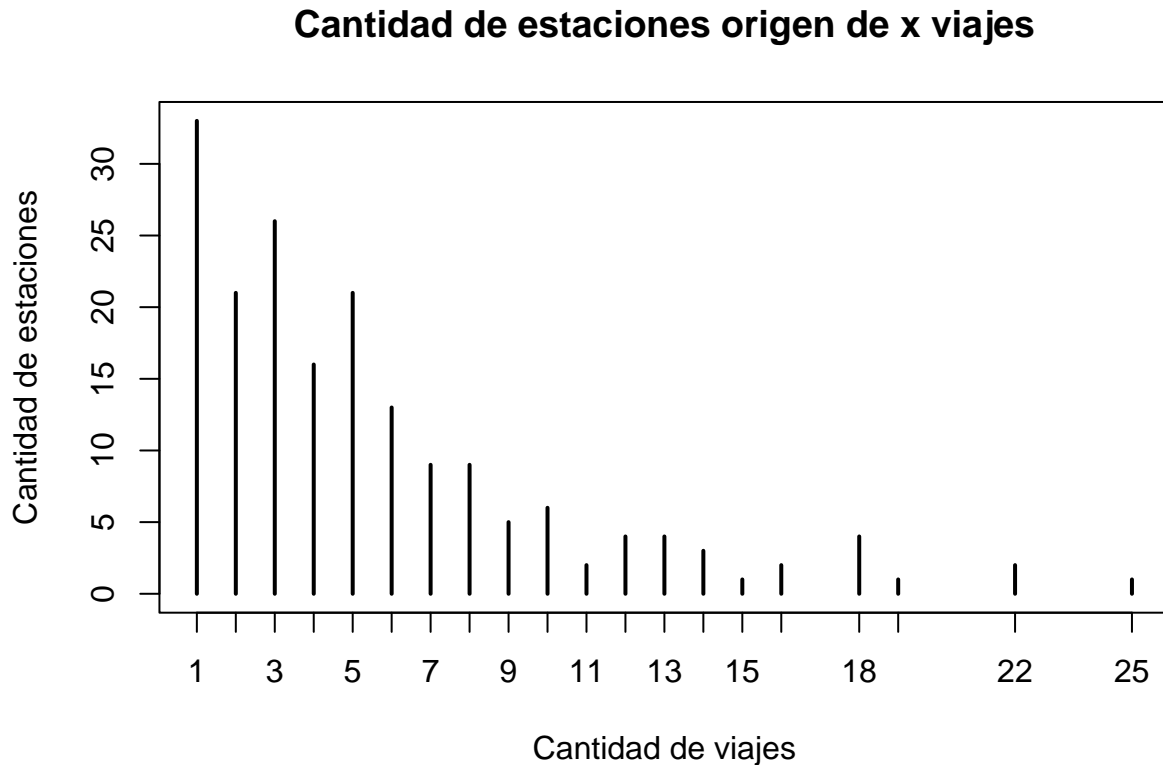


Figura 7: Estaciones Origen

Como se puede ver en (??) y (??), hay unos pocos “outliers” en estos gráficos también, pero la información se puede ver bien igualmente. Lo que se puede notar de estos gráficos es que no hay muchas estaciones que concentren gran cantidad de tráfico. Si hay lugar para mejorar el sistema va a ser en las estaciones que tienen estos valores extremos.

El otro foco de análisis con estas variables es cuántos viajes empiezan y terminan en la misma estación. En el gráfico ?? se ve que la mayoría de los recorridos comienzan y acaban en diferentes lugares. Si un viaje cumple con esta condición, lo más probable es que sea del tipo recreativo. Así que la conclusión es que la mayoría de la gente no hace este tipo de viajes.

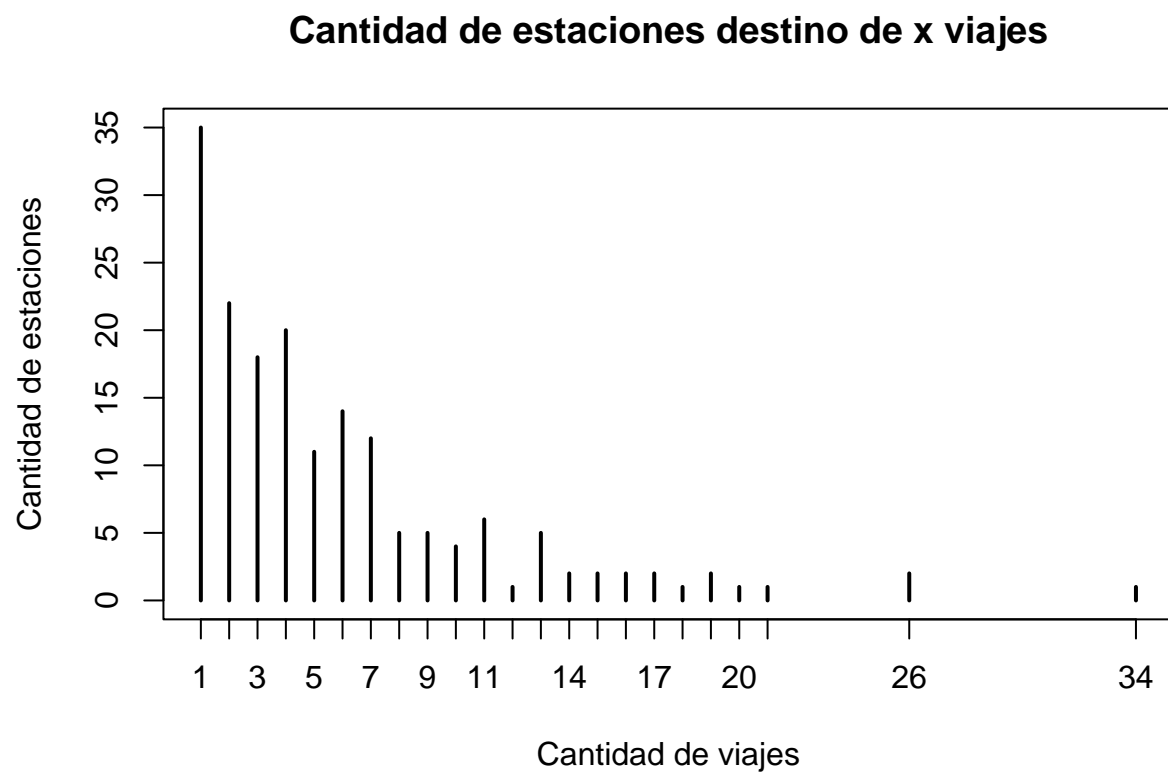


Figura 8: Estaciones Destino

Distribución de viajes que empiezan y terminan en el mismo lugar

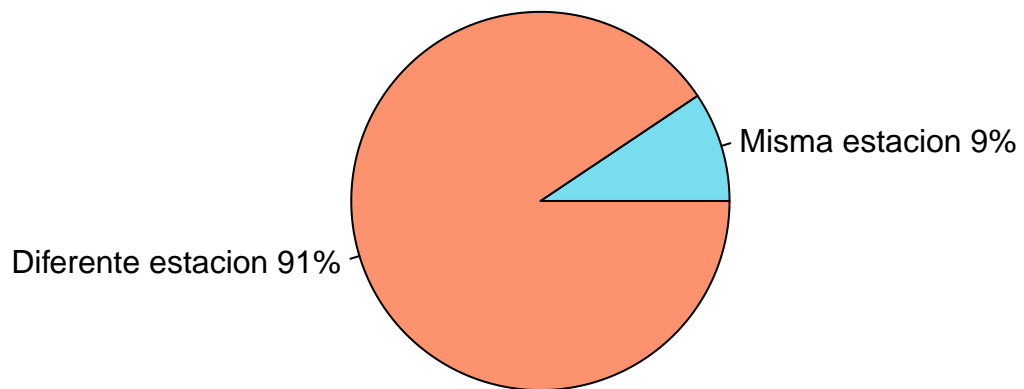


Figura 9: Viajes por Estacion

Análisis Bivariado

El estudio se realizó relacionando la edad de los usuarios con su género.

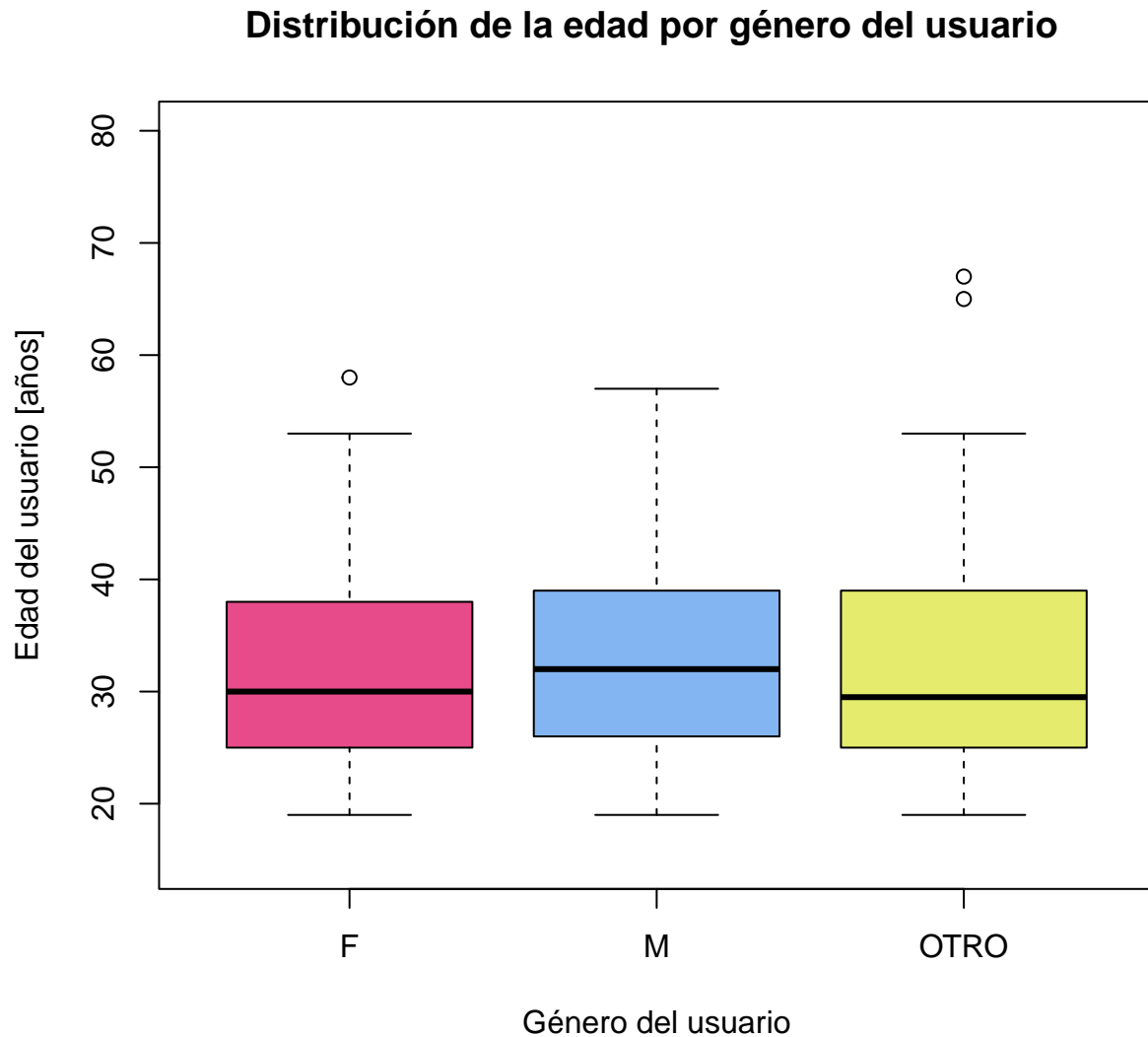


Figura 10: Genero por Edad

Al observar la gráfica ?? se puede notar que los análisis por edad de cada género tienen distribuciones bastante parecidas. Para los 3 casos la media está cerca de los 30 años y el valor mínimo va a ser 18 siempre. También se puede ver que para la gran mayoría de los usuarios, la edad no supera los 60 años para ninguno de los géneros.

Cuadro 3: Frecuencias Genero por Edad

	[19,24)	[24,29)	[29,33)	[33,38)	[38,43)	[43,48)	[48,53)	[53,57)	[57,62)	[62,67)
F	13	37	48	62	69	78	79	80	81	81
M	6	16	26	33	39	44	46	47	47	47
Otro	11	33	44	52	60	65	69	70	70	72

Y con la tabla ?? se pueden ver bien las distribuciones acumuladas.

Conclusiones

Luego de que se analizaron los distintos gráficos y tablas que se crearon para este informe se puede concluir que el sistema de EcoBicis brinda un medio de transporte muy utilizado por una gran variedad de ciudadanos de la Ciudad de Buenos Aires. La mayoría de estos se encuentran en el rango etario de entre 25 y 30 años y el género predominante entre los usuarios es el femenino. El servicio se suele utilizar para viajes de entre 10 y 30 minutos, y se recorre un promedio de aproximadamente 4500 metros. Una posible mejora al sistema es agregar más estaciones en las zonas más concurridas. Nuestro análisis denotó que, si bien la mayoría de las estaciones no parecen sufrir de sobrecarga, hay unas pocas que tienen un número particularmente alto de concurrencia, alejándose estadísticamente del resto.