



UNR - FCEIA

PROBABILIDAD Y ESTADÍSTICA

Trabajo Práctico - Estadística Descriptiva

Antuña Pablo
Regolo Alesandro
Regolo Giuliano

Índice

1. Análisis Descriptivo	2
1.1. Análisis Univariado	2
1.1.1. Variable Género	2
1.1.2. Variable Edad	3
1.1.3. Variable Día	5
1.1.4. Variable Distancia	6
1.1.5. Variable Duración	8
1.1.6. Análisis en base a la Dirección de origen	10
1.1.7. Análisis en base a la Dirección de salida	12
1.1.8. Análisis en base a la cantidad de usuarios	14
1.2. Análisis Bivariado	16
2. Conclusiones	17
3. Anexo	18

Introducción

El objetivo de este estudio es presentar un análisis descriptivo del uso del sistema de bicicletas públicas que dispone la ciudad autónoma de Buenos Aires. Esto se llevará a cabo a través de la utilización de una serie de medidas resúmenes, gráficos y tablas, se caracteriza a un conjunto de unidades, intentando descubrir regularidades y singularidades de los mismos. Para obtener los resultados se utiliza el paquete estadístico R con el cual, mediante elaboración propia, se generarán las tablas y gráficos del informe. Todos los datos y valores que se utilizaran en el estudio son provenientes de la muestra provista de la base de datos de la municipalidad de la ciudad Autónoma de Buenos Aires (<https://data.buenosaires.gob.ar/dataset/estaciones-bicicletas-publicas>).

1. Análisis Descriptivo

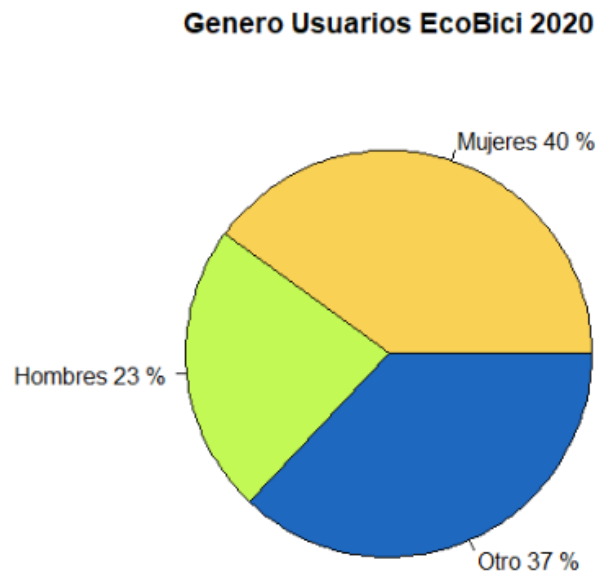
1.1. Análisis Univariado

1.1.1. Variable Género

En este gráfico de sectores se analiza el porcentaje de personas de cada género que utilizan el sistema de EcoBicis. Como se observa en el gráfico, los usuarios se encuentran divididos en 3 géneros distintos:

- Hombres un 23 %
- Otros un 37 %
- Mujeres un 40 %

A su vez, este análisis no es realmente representativo de la distribución del género de los usuarios ya que el valor “otros” es caso especial en el que se representa no solo a aquellos individuos que no se sienten representados por los géneros masculinos o femenino, sino también a aquellos usuarios que no hayan proporcionado esta información. Dejando esto como consecuencia que las cosas que se infieren pueden no ser enteramente correctas.



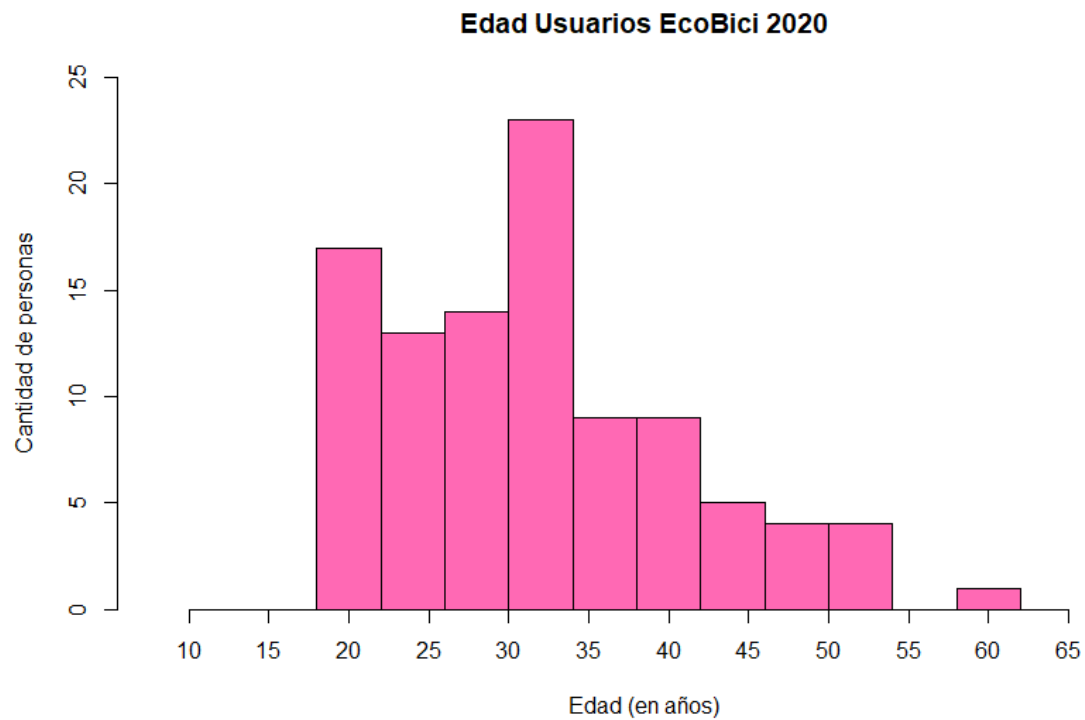
A continuación se resume la información provista en el gráfico representada en una tabla de distribución de frecuencias

Genero Usuarios EcoBici 2020

Género ▾	Frecuencia Absoluta ▴	Frecuencia Relativa ▴	Porcentaje ▴
Mujeres	40	0.40	40
Hombres	23	0.23	23
Otro	37	0.37	37
Total	100	1.00	100

1.1.2. Variable Edad

En el siguiente histograma se representa la edad de los usuarios que utilizan el sistema de EcoBicis. Como se puede observar en el gráfico la mayor parte de los usuarios que utilizan el servicio de bicicletas públicas son jóvenes que no superan los 35.



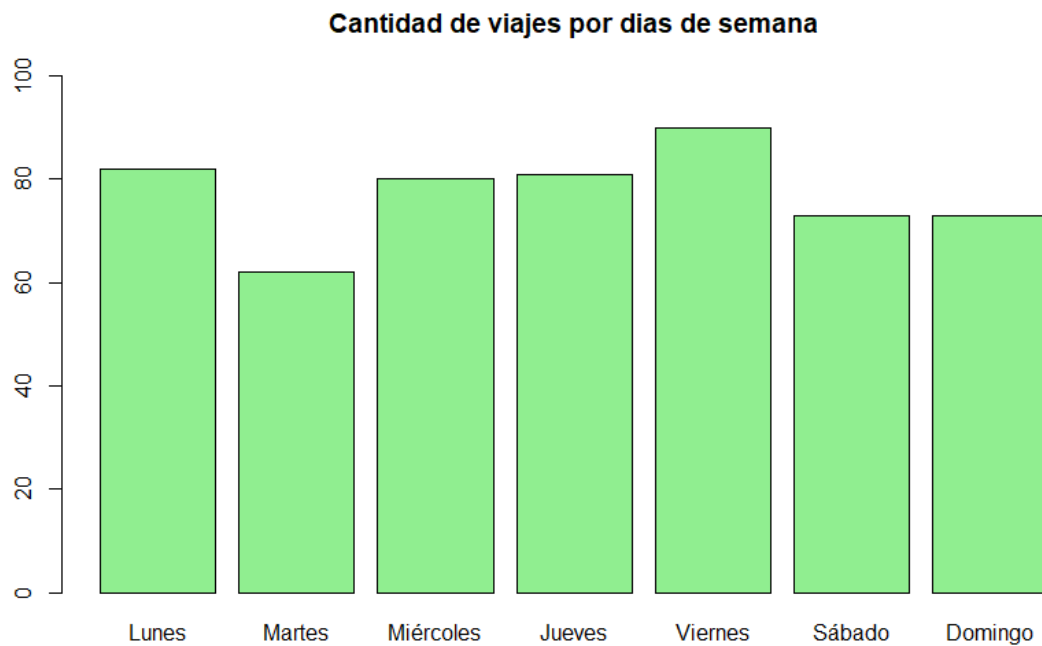
Al observar detenidamente la tabla de frecuencias se puede notar que la mayor frecuencia absoluta se encuentra en el intervalo de las personas que poseen entre 30 y 34 años con un 23 % de de los usuarios.

Edad Usuarios EcoBici 2020

Intervalos	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
[18,22)	12	12	0.121	0.121
[22,26)	13	25	0.131	0.252
[26,30)	12	37	0.121	0.373
[30,34)	23	60	0.232	0.605
[34,38)	14	74	0.141	0.746
[38,42)	11	85	0.111	0.857
[42,46)	3	88	0.030	0.887
[46,50)	5	93	0.051	0.938
[50,54)	5	98	0.051	0.989
[58,62)	1	99	0.010	1

1.1.3. Variable Día

Como se puede observar en la gráfica, el uso del sistema de EcoBicis se encuentra distribuido de manera equitativa aunque cabe recalcar que se realiza un mayor uso de las bicicletas públicas en el transcurso de la semana, con un leve incremento en el uso los días Viernes.



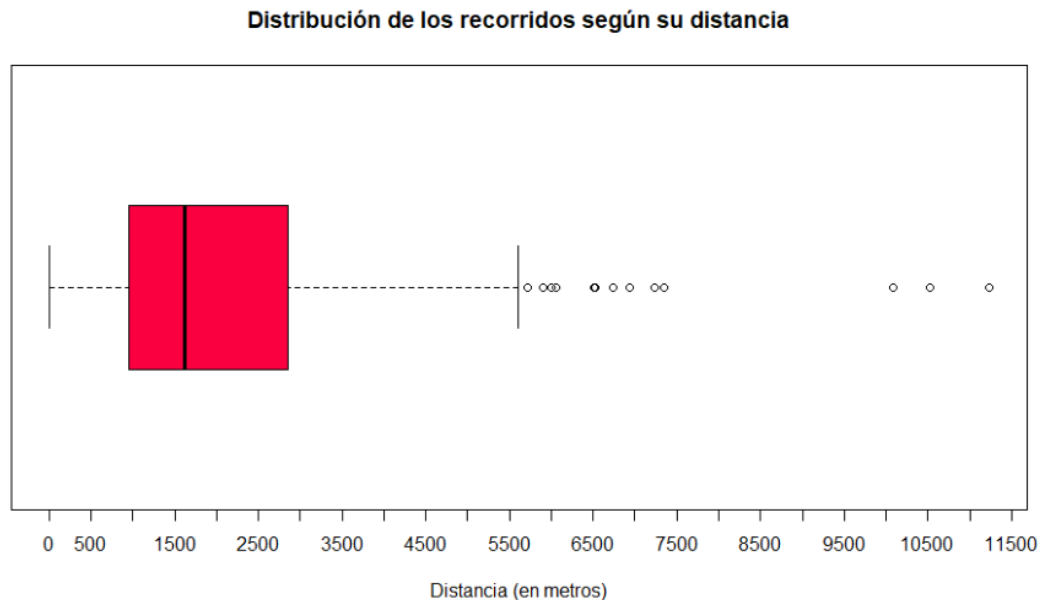
Se resume la información obtenida en una tabla de frecuencia. En esta se puede observar que el 73 % de los viajes se realizan a lo largo de la semana mientras que el 27 % restante se realizan en el fin de semana.

Cantidad de viajes por días de semana

Día	Frecuencia Absoluta	Frecuencia Relativa	Porcentaje
Lunes	82	0.152	15.2
Martes	62	0.115	11.5
Miercoles	80	0.148	14.8
Jueves	81	0.150	15.0
Viernes	90	0.166	16.6
Sabado	73	0.135	13.5
Domingo	73	0.135	13.5
Total	541	1	100

1.1.4. Variable Distancia

En el siguiente gráfico, se puede observar, cómo la distribución de las distancias recorridas se encuentra centralizada alrededor de 1000m a 3000m, con algunos valores extremos rondando los 12000m.



A la hora de realizar un análisis de las medidas resúmenes de distancia como se presentan valores extremos muy distantes de los demás valores la distancia media, este valor no será realmente representativo de los valores recolectados, por eso se optó por trabajar con la mediana, la cual es 1618 m. Esto nos lleva a la conclusión de que un 50 % de los viajes recorren menos de 1618m, a su vez sólo un 25 % de los recorridos llegan a distancias mayores a 2848m.

- Mínimo: 0m
- 1° Cuartil: 951m
- Mediana: 1618m
- 3° Cuartil: 2848m
- Máximo: 11240m

No obstante, todos los análisis que se realizaron no presentan gran relevancia ya que la distancia es medida entre parada y parada, por lo que si un usuario toma una bici, realiza su recorrido y luego vuelve a la estación de origen, es decir si la

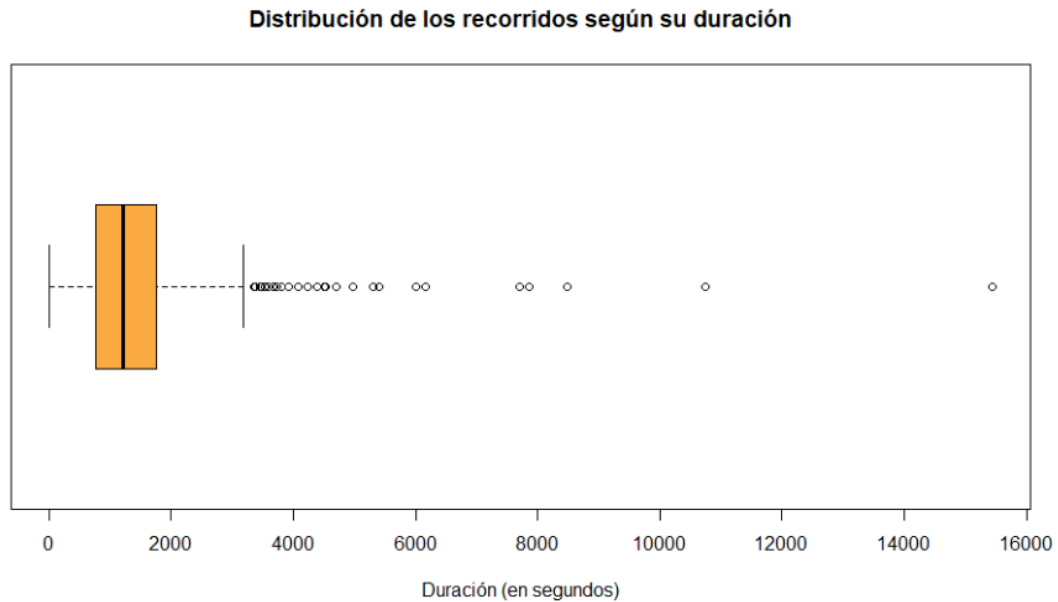
estación de origen y la estación destino coinciden, dicha distancia será considerada nula. Por lo que se debe tener en cuenta esto a la hora de inferir con este análisis.

Distribucion de los recorridos segun su distancia

Intervalos	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
[0,483)	81	81	0.149722736	0.1497227
[483,966)	56	137	0.103512015	0.2532348
[966,1.45e+03)	99	236	0.182994455	0.4362292
[1.45e+03,1.93e+03)	76	312	0.140480591	0.5767098
[1.93e+03,2.42e+03)	46	358	0.085027726	0.6617375
[2.42e+03,2.9e+03)	62	420	0.114602588	0.7763401
[2.9e+03,3.38e+03)	39	459	0.072088725	0.8484288
[3.38e+03,3.86e+03)	14	473	0.025878004	0.8743068
[3.86e+03,4.35e+03)	24	497	0.044362292	0.9186691
[4.35e+03,4.83e+03)	15	512	0.027726433	0.9463956
[4.83e+03,5.31e+03)	8	520	0.014787431	0.9611830
[5.31e+03,5.8e+03)	4	524	0.007393715	0.9685767
[5.8e+03,6.28e+03)	7	531	0.012939002	0.9815157
[6.28e+03,6.76e+03)	3	534	0.005545287	0.9870610
[6.76e+03,7.24e+03)	3	537	0.005545287	0.9926063
[7.24e+03,7.73e+03)	1	538	0.001848429	0.9944547
[9.66e+03,1.01e+04)	1	539	0.001848429	0.9963031
[1.01e+04,1.06e+04)	1	540	0.001848429	0.9981516
[1.11e+04,1.16e+04)	1	541	0.001848429	1.0000000

1.1.5. Variable Duración

En el siguiente gráfico, se puede observar la asimetría de la distribución de los datos observados. Si bien la gran mayoría de los datos observados se encuentra entre los valores 0s y 2000s, existen algunos valores que sobrepasan con creces estos valores.



A la hora de realizar un análisis de las medidas resúmenes de duración como se presentan valores extremos muy distantes de los demás valores, a la hora de analizar la duración media, este valor no será realmente representativo de los valores recolectados, por eso se optó por trabajar con la mediana, la cual es 1215s. Esto nos lleva a la conclusión de que un 50 % de los viajes duran menos de 1215s, a su vez sólo un 25 % de los recorridos tienen una duración mayor a 15453s.

- Mínimo: 169s
- 1°Cuartil: 768.8s
- Mediana: 1217s
- 3° Cuartil: 1761.2s
- Máximo: 15453s

Distribucion de los recorridos segun su duraci3n

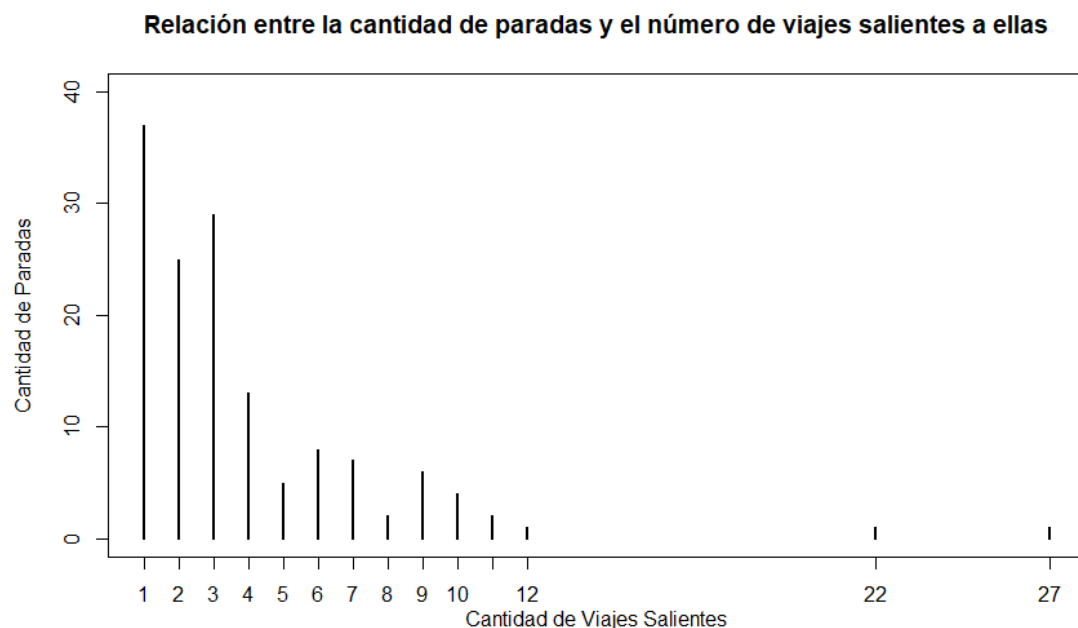
Intervalos	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
[169,826)	158	158	0.292	0.292
[826,1.48e+03)	182	340	0.337	0.629
[1.48e+03,2.14e+03)	124	464	0.230	0.859
[2.14e+03,2.8e+03)	34	498	0.063	0.922
[2.8e+03,3.45e+03)	13	511	0.024	0.946
[3.45e+03,4.11e+03)	14	525	0.026	0.972
[4.11e+03,4.77e+03)	5	530	0.009	0.981
[4.77e+03,5.42e+03)	3	533	0.005	0.986
[5.42e+03,6.08e+03)	1	534	0.002	0.988
[6.08e+03,6.74e+03)	1	535	0.002	0.990
[7.4e+03,8.05e+03)	2	537	0.004	0.994
[8.05e+03,8.71e+03)	1	538	0.002	0.996
[1.07e+04,1.13e+04)	1	539	0.002	0.998
[1.53e+04,1.59e+04)	1	540	0.002	1.000

1.1.6. Análisis en base a la Dirección de origen

A la hora de realizar un análisis de los datos brindados por la variable **dirección de origen**, ya que la información que brinda resulta difícil de representar en una tabla o gráfico, se decidió contar la cantidad de viajes salientes realizados en cada estación.

Esta nueva variable (cantidad de viajes salientes realizados en cada estación) es de tipo cuantitativo. Teniendo esto en cuenta, se procede a realizar el análisis.

En la siguiente gráfica se representa cuántas paradas son utilizadas como punto de partida en una determinada cantidad de viajes realizados. Como se puede observar la gran mayoría de las paradas se utilizan con poca frecuencia, es decir, 1(una), 2(dos) o 3(tres) veces. Esto produce una distribución asimétrica de los valores observados. A su vez, se puede ver que existen paradas las cuales presentan una gran frecuencia de uso como salida.



A continuación en la tabla se puede observar que un 26 % de las paradas son utilizadas una única vez, un 17 % de las paradas son utilizadas 2 veces y un 20 % de las paradas son utilizadas 3 veces. Además, la parada más frecuentada es utilizada 27 veces y representa menos del 1 % del total de paradas.

Relación entre la cantidad de paradas y el numero de viajes salientes a ellas

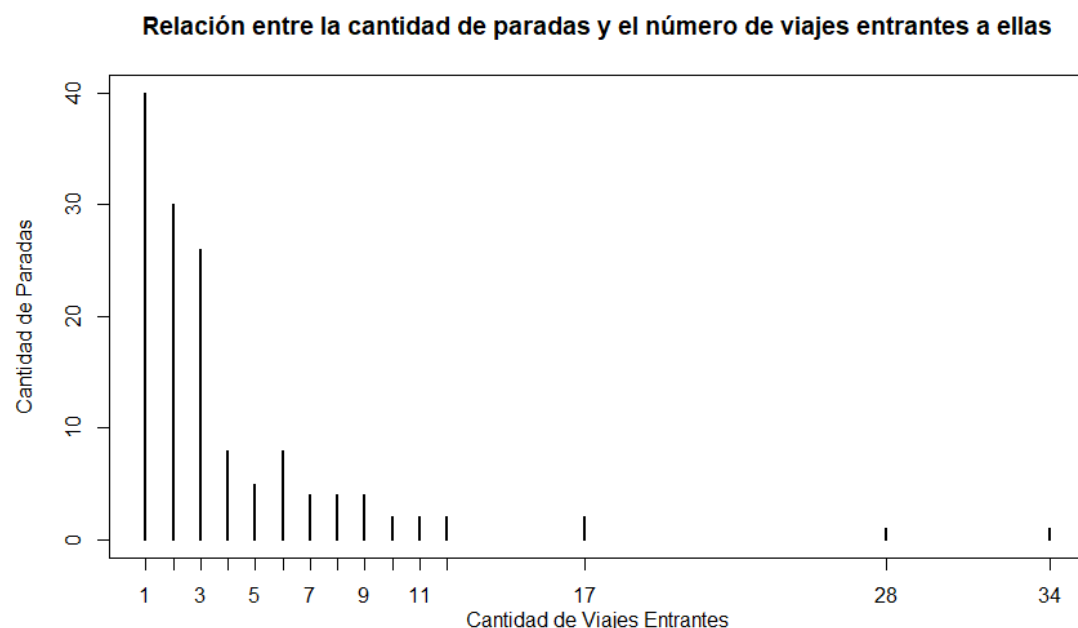
Cantidad de Viajes ▲	Frecuencia Absoluta ▼	Frecuencia Relativa ▼	Porcentaje ▼
1	37	0.262	26.2
2	25	0.177	17.7
3	29	0.206	20.6
4	13	0.092	9.2
5	5	0.035	3.5
6	8	0.057	5.7
7	7	0.050	5.0
8	2	0.014	1.4
9	6	0.043	4.3
10	4	0.028	2.8
11	2	0.014	1.4
12	1	0.007	0.7
22	1	0.007	0.7
27	1	0.007	0.7
Total	141	1.000	100

1.1.7. Análisis en base a la Dirección de salida

Al igual que en el análisis anterior los datos brindados por la variable **dirección de salida**, la información que brinda resulta difícil de representar en una tabla o gráfico y se decidió contar la cantidad de viajes entrantes realizados en cada estación.

Esta nueva variable (cantidad de viajes entrantes realizados en cada estación) es de tipo cuantitativo. Teniendo esto en cuenta, se procede a realizar el análisis.

En la siguiente gráfica se representa cuántas paradas son utilizadas como punto de llegada en una determinada cantidad de viajes realizados. Como se puede observar la gran mayoría de las paradas se utilizan con poca frecuencia, es decir, 1(una), 2(dos) o 3(tres) veces. Esto produce una distribución asimétrica de los valores observados. A su vez, se puede ver que existen paradas las cuales presentan una gran frecuencia de uso como destino.



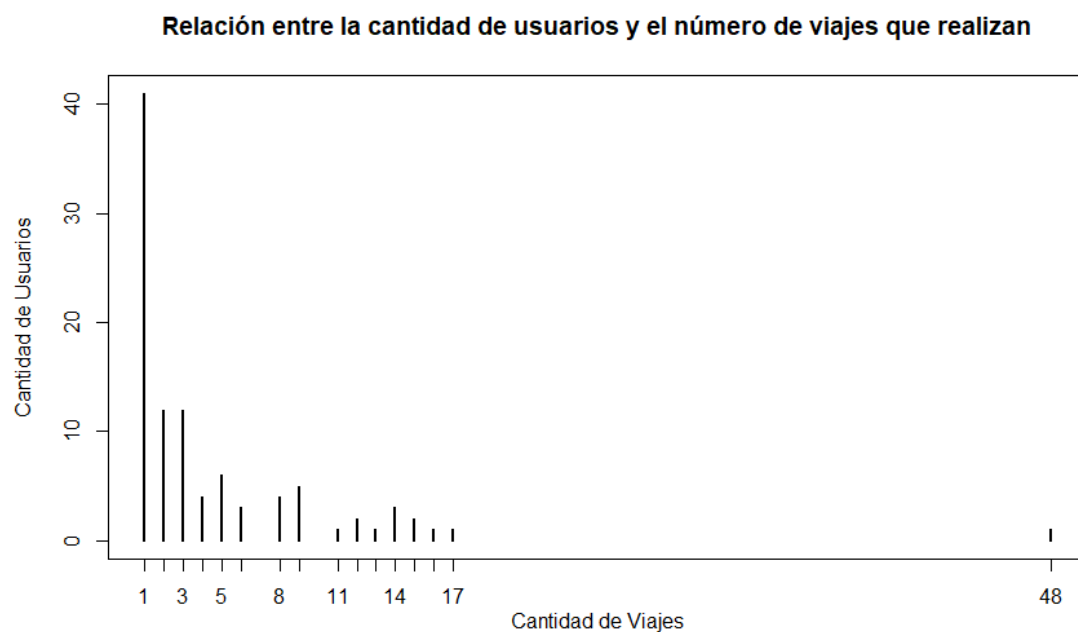
Como se puede notar en la tabla, el 28 % de las paradas únicamente reciben 1 viaje y el 21 % de las paradas reciben 2 viajes y por último un 18 % de las paradas reciben 3 viajes. Y la parada que es la más frecuentada como destino de los usuarios se utilizó un total de 34 veces.

Relación entre la cantidad de paradas y el numero de viajes entrantes a ellas

Cantidad de Viajes ▲	Frecuencia Absoluta ▼	Frecuencia Relativa ▼	Porcentaje ▼
1	40	0.288	28.8
2	30	0.216	21.6
3	26	0.187	18.7
4	8	0.058	5.8
5	5	0.036	3.6
6	8	0.058	5.8
7	4	0.029	2.9
8	4	0.029	2.9
9	4	0.029	2.9
10	2	0.014	1.4
11	2	0.014	1.4
12	2	0.014	1.4
17	2	0.014	1.4
28	1	0.007	0.7
34	1	0.007	0.7
Total	139	1.000	100.0

1.1.8. Análisis en base a la cantidad de usuarios

Para este análisis se decidió trabajar con una nueva variable, cantidad de viajes hechos por usuario, creada a partir de la cantidad de veces que un usuario utilizó el servicio. Teniendo esto en cuenta se realizará un análisis sobre la relación entre la cantidad de usuarios y los viajes que estos realizan. Al tratar de realizar la gráfica, uno de los usuarios realizó una cantidad desmesurada de viajes (98 viajes) y por motivos de fácil visualización e interpretación se decidió no incluirlo en la gráfica. No obstante en el análisis de esta variable si se lo tuvo en cuenta. Al observar la gráfica se puede notar a simple vista que la mayor parte de los usuarios que utilizan este sistema realizan pocos viajes.



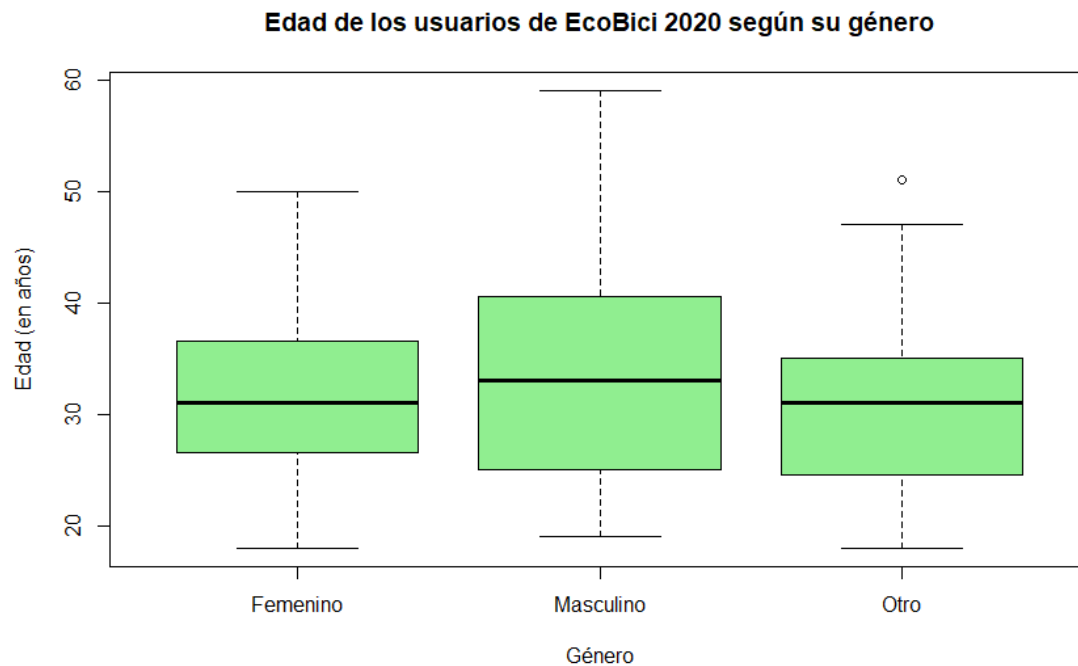
En la siguiente tabla se puede observar que el 41 % de los usuarios realizaron un único viaje, un 12 % utilizaron el sistema para viajar solo 2 veces y otro 12 % lo utilizó para realizar 3 viajes. A su vez, se registró que la mayor cantidad de viajes realizados es de 98 y podemos ver en la tabla que dicha cantidad fue alcanzada por un único usuario.

Relación entre la cantidad de usuarios y el número de viajes que realizan

Cantidad de Viajes	Frecuencia Absoluta	Frecuencia Relativa	Porcentaje
1	41	0.41	41
2	12	0.12	12
3	12	0.12	12
4	4	0.04	4
5	6	0.06	6
6	3	0.03	3
8	4	0.04	4
9	5	0.05	5
11	1	0.01	1
12	2	0.02	2
13	1	0.01	1
14	3	0.03	3
15	2	0.02	2
16	1	0.01	1
17	1	0.01	1
48	1	0.01	1
98	1	0.01	1
Total	100	1.00	100

1.2. Análisis Bivariado

Se realizó un estudio relacionando la edad de los usuarios con su género el cual se puede ver representado en el siguiente gráfico:



Al observar la gráfica detenidamente se puede concluir que, mientras que los usuarios que se sienten representados por el género femenino o por “otros” se encuentran en un espectro más reducido de la tabla, los que se representan con el género masculino poseen valores más dispersos para el valor edad. A su vez, también poseen los usuarios con mayor edad registrados en el estudio.

Medidas de Interés Edad de los usuarios según su género

Medidas de Interés	Femenino	Masculino	Otro
Mínimo	18.000	19.000	18.000
1er Cuartil	26.750	25.000	24.750
Mediana	31.000	33.000	31.000
Media	31.725	33.217	31.944
3er Cuartil	36.250	40.500	35.000
Máximo	50.000	59.000	51.000

Gracias a la tabla, se puede observar cómo efectivamente la mediana del género masculino, se encuentra unos 2 años por encima de las otras categorías y a su vez también presentan un mayor valor máximo con una diferencia de casi 9 años con las otras categorías. Además, tanto el género femenino como otros, poseen una mediana de 31, es decir el 50 % de los usuarios son menores a 31 años.

2. 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. El 74 % de los usuarios se encuentran en el rango etario de los 18 a 38 años. Estos mantienen una proporción muy pareja entre los géneros aunque posee un rango más variado de edades para el género masculino. El 75 % de los viajes realizados por los usuarios no superan los 3km de distancia entre estaciones y el 75 % de los viajes duran menos de 30 minutos. A su vez, del análisis de salida y entrada de usuarios en las distintas estaciones de bicicletas del sistema, se puede observar cómo la gran mayoría de las paradas son utilizadas con muy poca frecuencia (1,2 o 3 veces), por otra parte 2 o 3 paradas son utilizadas con una gran frecuencia llegando a números como 34 veces, en el caso de viajes entrantes, o 27, en el caso de viajes salientes. No obstante el uso del sistema a lo largo de la semana se encuentra bien distribuido con una leve creciente los viernes.

3. Anexo

Durante el análisis de la duración de los recorridos fue necesario ignorar un dato registrado ya que se tomó como un valor erróneo que de no ser removido alteraría la disposición del gráfico y los valores que se registraron en la tabla de frecuencia. Este valor fue considerado erróneo ya que el valor 2 representa la duración en segundos del recorrido y esta duración está asociada a un valor de distancia de 1405,56 metros. Realizar un viaje de tal distancia en dicho tiempo fue considerado como un hecho no posible y no representativo de la muestra.

3822 Guemes	2516 Pereyra Lucena	2	1405.56	Martes
-------------	---------------------	---	---------	--------

Al analizar y realizar las respectivas tablas y gráficas que involucran los datos referentes a la edad de los usuarios, no se tuvieron en cuenta los datos con valor “N/A”. Esto se debe a que, como se mencionó anteriormente en el caso de la duración, este valor se tomó como una entrada errónea la cual no se puede representar como dato.

22490	OTRO	NA
-------	------	----