

INFORME ESTADISTICO:

ANALISIS SOBRE EL SISTEMA DE BICIS COMPARTIDAS “EcoBici”

AUTORES:

Cainelli Ignacio

Arroyo Joaquin

AÑO 2021

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Licenciatura en Ciencias de la Computación

### 1. OBJETIVO

Resumir la información que brinda el gobierno de la ciudad de Buenos Aires sobre el servicio “EcoBici” para analizar el género y la edad de los usuarios, la duración y distancia de los recorridos, los días en los que se realizan dichos recorridos y las estaciones más utilizadas.

### 2. MATERIALES

**Población de estudio:** Corresponde a personas que residen en CABA, las cuales utilizaron el servicio.

**Muestra del estudio:** Corresponde a dos bases de datos, la primera tiene información sobre 100 usuarios que utilizaron el servicio en el año 2020, y la segunda sobre recorridos que se hicieron con el transporte que provee dicho servicio.

En resumen,

#### **Base de datos N.º 1:**

**Unidad de análisis:** Usuarios del servicio

**Muestra:** 100 usuarios (Tamaño de la muestra  $n = 100$ )

**Variables:**

Género: genero del usuario (Cualitativa)

Edad: edad del usuario (Cuantitativa continua)

#### **Base de datos N.º 2:**

**Unidad de análisis:** Recorridos del servicio

**Muestra:** 428 recorridos (Tamaño de la muestra  $n = 428$ )

**Variables:**

Estación de salida: estación donde el usuario retira la bicicleta (Cualitativa)

Estación de llegada: estación donde el usuario devuelve la bicicleta (Cualitativa)

Duración: duración del recorrido (Cuantitativa continua)

Distancia: distancia del recorrido (Cuantitativa continua)

Día: día en el cual se realizó el recorrido (Cualitativa)

Las bases utilizadas fueron obtenidas en <https://data.buenosaires.gob.ar/dataset>

**Software:** Se utilizó Excel para las bases de datos, RStudio para su análisis.

### 3. RESULTADOS

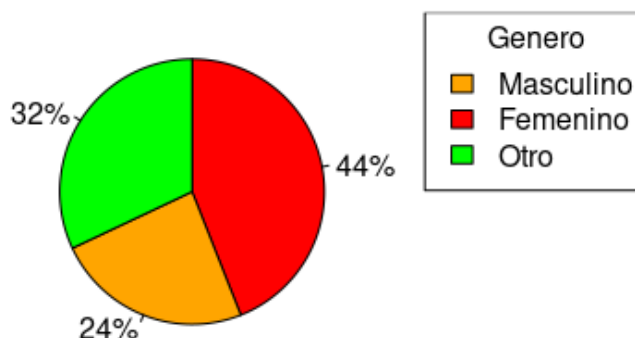
En el **cuadro N.º 1** se presenta la tabla de frecuencia del género de los usuarios.

**Cuadro N.º 1:**

Genero	Frecuencia Absoluta	Frecuencia Relativa	Porcentaje
Femenino	44	0.44	44
Masculino	24	0.24	24
Otro	32	0.32	32
<b>Total</b>	100	1.00	100

En la **figura N.º 1** se presenta el grafico de sectores del género de los usuarios.

**Figura N.º 1:**



#### **Análisis cuadro N.º 1 y figura N.º 1: (Género)**

##### Medida de interés:

Moda = femenino

Interpretación:

El género femenino es el que más utilizó el servicio.

Se observa que el segundo género que más utilizó el servicio es OTRO.

A partir de este último resultado, se puede ver que quizás muchos usuarios no indicaron su género, por lo que automáticamente la aplicación les asignó el género OTRO; por lo cual, la aplicación debería ser más estricta en el momento en el cual un usuario genera su perfil, es decir que, si no indica su género, no lo deje seguir al siguiente paso. Esto serviría para tener información más acertada con respecto al género de quienes utilizan el servicio.

En el **cuadro N.º 2** se presenta la tabla de frecuencia de la edad en años de las personas que utilizaron el servicio. Notar que hay 99 personas ya que en la información utilizada una de ellas no indicó su edad.

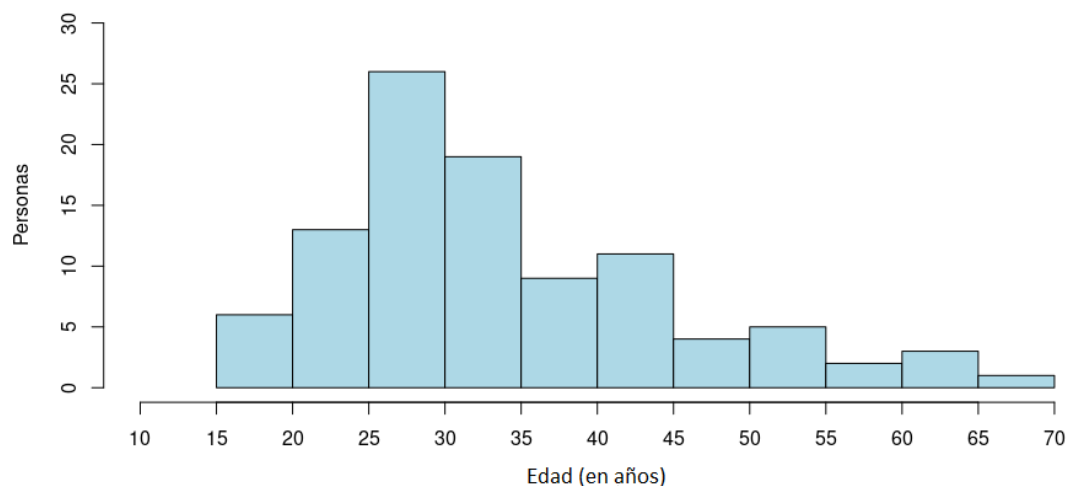
**Cuadro N.º 2: (Edad en años)**

Intervalos	Frecuencia absoluta	Frecuencia relativa	Porcentaje	Frecuencia absoluta acumulada	Frecuencia relativa acumulada	Porcentaje acumulado
(15,20]	6	0,06	6,06	6	0,06	6,06
(20,25]	13	0,13	13,13	19	0,19	19,19
(25,30]	26	0,26	26,26	45	0,45	45,45
(30,35]	19	0,19	19,19	64	0,64	64,64
(35,40]	9	0,09	9,09	73	0,73	73,73
(40,45]	11	0,11	11,11	84	0,84	84,84
(45,50]	4	0,04	4,04	88	0,88	88,88
(50,55]	5	0,05	5,05	93	0,93	93,93
(55,60]	2	0,02	2,02	95	0,95	95,95
(60,65]	3	0,04	3,04	98	0,99	98,99

(65,70]	1	0,01	1,01	99	1,00	100,00
<b>Total</b>	99	1,00	100,00			

En la **figura N.º 2** se presenta el histograma con la distribución de la edad en años de los usuarios.

**Figura N.º 2:**



#### **Análisis cuadro N.º 2 y figura N.º 2: (Edad)**

Todos los valores de esta sección representan una edad en años.

##### Medidas de tendencia central:

Media = 34,45

Mediana = 31

Interpretación:

El promedio de la edad de los usuarios es de 34,45 años

El 50% de los usuarios tiene 31 años o menos

##### Medidas de posición:

Mínimo = 18

Cuartil 1 = 27

Cuartil 3 = 41

Máximo = 70

Interpretación:

El edad mínima de un usuario es de 18

El 25% de los usuarios tiene 27 años o menos

El 75% de los usuarios tiene 41 años o menos

La edad máxima de un usuario es de 70 años

Con estos datos se observa que el servicio es utilizado en su mayoría por personas menores a las 41 años, es decir, personas relativamente jóvenes.

Medidas de dispersión:

Desvío estándar = 11,19

Rango = 52

Rango intercuartílico = 14

Varianza = 125,21

Interpretación:

El desvío estándar indica que tan dispersos están los datos con respecto a la media, a partir de esto se observa que aproximadamente el 68% de los datos se encuentran a no más de la edad de 45,64 y el 95% de los datos se encuentran a no más de la edad de 56,83. Con esta información se puede ver efectivamente que el servicio es utilizado por gente medianamente joven.

El rango intercuartílico nos indica que hay una distancia de 14 entre el primer cuartil y el tercer cuartil, es decir, hay un 50% de usuarios entre los 27 y 41 años. Esto indica que la mitad de los usuarios son adultos.

El valor de la varianza se ve reflejado en el cuadro N.º 2 donde se observa que la mayoría de las edades están concentradas a una distancia baja con respecto a la media.

En el **cuadro N.º 3** se presenta la tabla de frecuencia de la cantidad de recorridos por día.

**Cuadro N.º 3:**

Días	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Lunes	74	0,17	17
Martes	54	0,13	13
Miércoles	54	0,13	13
Jueves	67	0,15	15
Viernes	63	0,15	15
Sábado	68	0,16	16
Domingo	48	0,11	11
<b>Total</b>	428	1,00	100

En la **figura N.º 3** se presenta el grafico de barras con el porcentaje de los recorridos que se hicieron en cada día.

**Figura N.º 3:**



#### **Análisis cuadro N.º 3 y figura N.º 3: (Días)**

##### Medida de interés:

Moda = lunes

Interpretación:

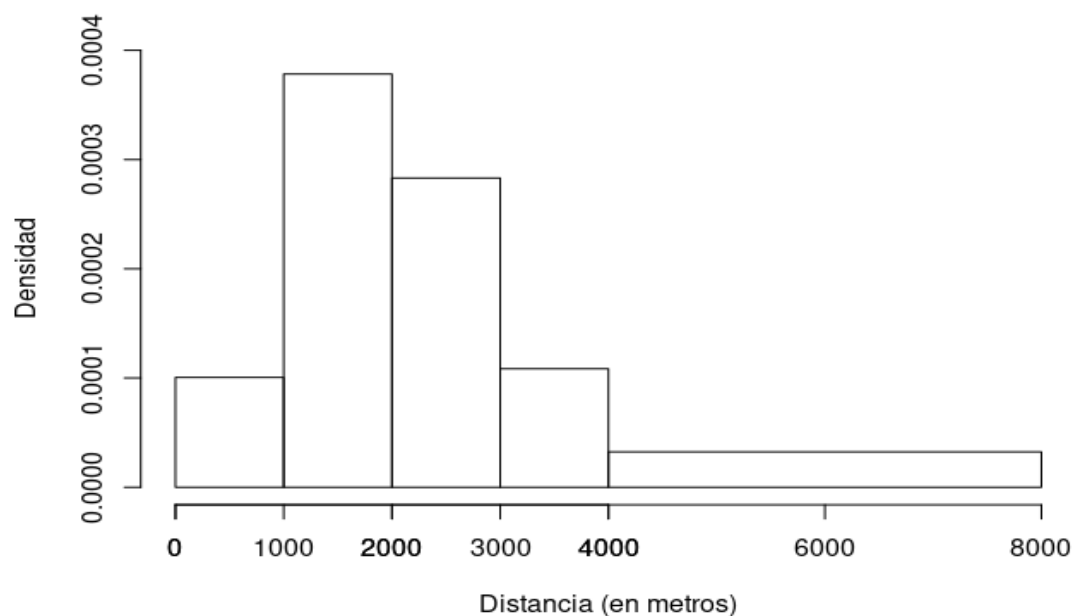
El día en el que más se utilizó el servicio fue el día lunes.

También se observa que el día Sábado también tiene altos porcentajes, es decir que este servicio no solo se utiliza para ir al trabajo, escuela, etc., sino que también se utiliza para moverse hacia actividades más recreativas.

Se observa que el día domingo se reduce la utilización del servicio.

En la **figura N.º 4** se presenta el histograma con la distancia de los recorridos. Notar que en el grafico no se incluyeron los recorridos con distancia nula.

**figura N.º 4:**



#### **Análisis figura N.º 4: (Distancia)**

Se analizan los recorridos sin incluir aquellos con distancia nula.

Todos los valores de esta sección representan distancia del recorrido en metros.

##### Medidas de tendencia central:

Media = 2414

Mediana = 2090

Interpretación:

El promedio de distancia recorrida es 2414 metros

El 50% de los usuarios recorrió 2090 metros o menos

##### Medidas de posición:

Mínimo = 242

Cuartil 1 = 1566

Cuartil 3 = 2888

Máximo = 7662

Interpretación:

La distancia mínima recorrida fue de 242 metros

El 25% de los usuarios recorrió 1566 metros o menos

El 75% de los usuarios recorrió 2888 metros o menos

La distancia máxima recorrida fue de 7662 metros

##### Medidas de dispersión:

Desvío estándar = 1335,534

Rango = 7419,05

Rango intercuartílico = 1322,53

Varianza = 1783650

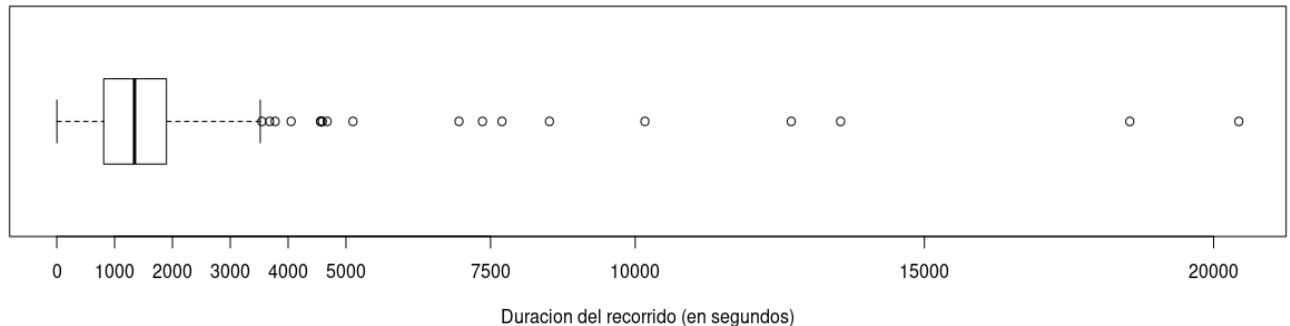
Interpretación:

El desvío estándar indica que tan dispersos están los datos con respecto a la media, a partir de esto se observa que aproximadamente el 68% de los datos se encuentran a no más de la distancia de 3749,5 metros y el 95% de los datos se encuentran a no más de la distancia de 5085 metros. Con esta información se concluye en que la gran mayoría de los recorridos tuvieron una distancia menor a aproximadamente 5km.

El rango intercuartílico indica que hay una distancia de 1322,53 entre el primer cuartil y el tercer cuartil, es decir, hay un 50% de los recorridos que tienen una distancia entre 1566 y 2888 metros.

En la **figura N.º 5** se presenta el boxplot con la duración de los recorridos. Notar que el máximo antes de los resultados atípicos es 3517 segundos y hay 19 resultados atípicos.

**figura N.º 5:**



#### **Análisis figura N.º 5: (Duración)**

Todos los valores de esta sección representan duración del recorrido en segundos.

##### Medidas de tendencia central:

Media = 1651

Mediana = 1344

Interpretación:

El promedio de duración de los recorridos es de 1651 segundos

El 50% de los recorridos duraron 1344 segundos o menos

##### Medidas de posición:

Mínimo = 2

Cuartil 1 = 812

Cuartil 3 = 1894

Máximo = 20436

Interpretación:

La duración mínima de un recorrido fue de 2 segundos. Podemos suponer que claramente el usuario no hizo ningún recorrido, si no que tomó la bicicleta y la volvió a dejar, o fue un error del sistema

El 25% de los recorridos duró 812 segundos o menos

El 75% de los recorridos duró 1894 segundos o menos

La duración máxima de un recorrido fue de 20436 segundos

##### Medidas de dispersión:

Desvío estándar = 1817,777

Rango = 20434

Rango intercuartílico = 1082,5

Varianza = 3304315



Interpretación:

El desvío estándar indica que tan dispersos están los datos con respecto a la media, a partir de esto se observa que aproximadamente el 68% de los datos se encuentran a no más de la duración 3468,7 segundos y el 95% de los datos se encuentran a no más de la duración de 5286,4 segundos. Con esta información se concluye en que la gran mayoría de los recorridos tuvieron una duración menor a aproximadamente 88 minutos.

El rango intercuartílico nos indica que hay una distancia de 1082,5 entre el primer cuartil y el tercer cuartil, es decir, hay un 50% de los recorridos que tienen una duración entre 812 y 1894 segundos.

Que la varianza sea tan grande se ve reflejado en la figura N.º 5 donde vemos que hay una gran cantidad de casos atípicos, que además tienen valores alejados al valor medio de duración de los recorridos.

En el **cuadro N.º 4** se presenta la tabla de frecuencia de las estaciones de salida. Notar que se incluyen las estaciones con una frecuencia mayor a 5. En la sección otros se incluyen todas las estaciones con frecuencia igual a 5 o menor.

**Cuadro N.º 4:**

Estación	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Av. Juan de Garay 1050	26	0,0607	6,07
Olleros Av. y Del Libertador Av.	15	0,0350	3,50
García del Río y Pinto	14	0,0327	3,27
Gainza, Martín de, Gral. 494	12	0,0280	2,80
Av. Patricias Argentinas y Estivao	12	0,0280	2,80
Güemes 3330 y coronel Díaz	11	0,0257	2,57
Santa Fe Av. y Bullrich Av.	11	0,0257	2,57
Blanes, Juan Manuel 383	10	0,0233	2,33
Carcova 3508	9	0,0210	2,10
Salguero, Jerónimo y Mansilla, Lucio Norberto Gral.	8	0,0186	1,86
Pereyra Lucena 2516	7	0,0163	1,63
Sucre, Antonio José de, Mcal. Y Vertiz virrey Av.	7	0,0163	1,63
Matienzo, Benjamín, teniente 1798	6	0,0140	1,40
Díaz, cnel. Av. 2300 y French	6	0,0140	1,40
Guardia Vieja 3302 y Agüero	6	0,0140	1,40
Godoy Cruz 3279 y Libertador	6	0,0140	1,40
Maure 1819	6	0,0140	1,40
OTROS	256	0,5981	59,81
<b>Total</b>	<b>428</b>	<b>1,0000</b>	<b>100</b>

#### **Análisis cuadro N.º 4: (Estaciones de salida)**

##### Medida de interés:

Moda = Av. Juan de Garay 1050

Interpretación:

La estación más concurrida para tomar una bicicleta es la que se encuentra en Av. Juan de Garay al 1050.

También se observa que hay 252 estaciones que tienen una frecuencia menor a 5, es decir, más del 50% de las estaciones tienen una baja frecuencia de usuarios que toman su bicicleta en ellas, con respecto a las demás.

En el **cuadro N.º 5** se presenta la frecuencia de las estaciones de llegada. Notar que se incluyen las estaciones con una frecuencia mayor a 5. En la sección otros se incluyen todas las estaciones con frecuencia igual a 5 o menor.

**Cuadro N.º 5:**

<b>Estación</b>	<b>Frecuencia absoluta</b>	<b>Frecuencia relativa</b>	<b>Porcentaje</b>
Av. Juan de Garay 1050	20	0,0467	4,67
Blanes, Juan Manuel 383	19	0,0443	4,43
García del Río y Pinto	14	0,0327	3,27
Olleros Av. y Del Libertador Av.	14	0,0327	3,27
Santa Fe Av. y Bullrich Av.	14	0,0327	3,27
Maure 1819	11	0,0257	2,57
Pereyra Lucena 2516	10	0,0233	2,33
Vicente López 2000	9	0,0210	2,10
Superi 2519	9	0,0210	2,10
Güemes 3330 y coronel Díaz	9	0,0210	2,10
1226 Ecuador Y Mansilla, Lucio Norberto, General	8	0,0186	1,86
Brandsen 1619	8	0,0186	1,86
Espinosa 1785	8	0,0186	1,86
Carcova 3508	8	0,0186	1,86
Bulnes 441 y Perón, Juan Domingo	8	0,0186	1,86
Godoy Cruz 3279 y Libertador	7	0,0163	1,63
J. Ramírez de Velasco 285	7	0,0163	1,63
Av. Sarmiento 2601	6	0,0140	1,40
Miguel Sánchez y Av. Libertadores	6	0,0140	1,40
Salguero, Jerónimo y Mansilla, Lucio Norberto Gral.	6	0,0140	1,40
OTROS	227	0,5303	53,04
<b>Total</b>	<b>428</b>	<b>1,0000</b>	<b>100</b>

### Análisis cuadro N.º 5: (Estaciones de llegada)

#### Medida de interés:

Moda = Av. Juan de Garay 1050

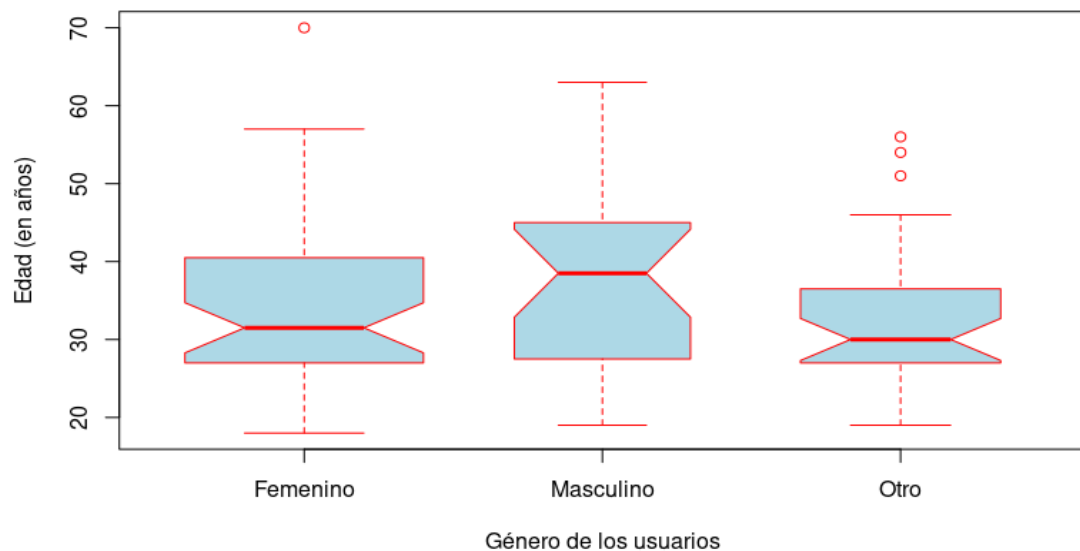
Interpretación:

La estación más concurrida para dejar una bicicleta es la que se encuentra en Av. Juan de Garay al 1050.

También se observa que hay 227 estaciones que tienen una frecuencia menor a 5, es decir, más del 50% de las estaciones tienen una baja frecuencia de usuarios que dejan su bicicleta en ellas, con respecto a las demás, lo mismo que pasa con las estaciones de salida.

En la **figura N.º 6** se presenta un boxplot comparativo entre el género y la edad de los usuarios.

Figura N.º 6:



### Análisis figura N.º 6: (Genero vs Edad)

#### Masculino:

#### Medidas de tendencia central:

Media = 38,12

Mediana = 38,50

Interpretación:

El promedio de la edad de los usuarios masculinos es de 38,12 años

El 50% de los usuarios masculinos tienen 38,5 años o menos

#### Medidas de posición:

Mínimo = 19

Cuartil 1 = 27,75

Cuartil 3 = 45

Máximo = 63

Interpretación:

El edad mínima de un usuario masculino es de 19

El 25% de los usuarios masculinos tiene 27,75 años o menos

El 75% de los usuarios masculinos tiene 45 años o menos

La edad máxima de un usuario masculino es de 63 años

Con estos datos se observa que la mayoría de usuarios masculinos que utilizan este servicio tienen 45 años o menos, por lo tanto, podemos decir que una gran cantidad de estos usuarios son personas relativamente jóvenes.

Observamos en el grafico que no hay ningún caso atípico.

### Femenino:

#### Medidas de tendencia central:

Media = 33,86

Mediana = 31,50

Interpretación:

El promedio de la edad de los usuarios femeninos es de 33,86 años

El 50% de los usuarios femeninos tienen 31,50 años o menos

#### Medidas de posición:

Mínimo = 18

Cuartil 1 = 27

Cuartil 3 = 40,25

Máximo = 70

Interpretación:

El edad mínima de un usuario femenino es de 18

El 25% de los usuarios femeninos tiene 27 años o menos

El 75% de los usuarios femeninos tiene 40,25 años o menos

La edad máxima de un usuario femenino es de 70 años

Con estos datos se observa que la mayoría de usuarios femeninos que utilizan este servicio tienen 40,25 años o menos, por lo tanto, podemos decir que una gran cantidad de estos usuarios son personas jóvenes.

Observamos en el grafico que solo hay un caso atípico.

### Otro:

#### Medidas de tendencia central:

Media = 32,45

Mediana = 30

Interpretación:

El promedio de la edad de los usuarios con género OTRO es de 38,12 años

El 50% de los usuarios con género OTRO tienen 38,5 años o menos

#### Medidas de posición:

Mínimo = 19

Cuartil 1 = 27

Cuartil 3 = 36,5

Máximo = 56

Interpretación:

El edad mínima de un usuario con género OTRO es de 19

El 25% de los usuarios con género OTRO tiene 27 años o menos

El 75% de los usuarios con género OTRO tiene 36,5 años o menos

La edad máxima de un usuario con género OTRO es de 56 años

Con estos datos se observa que la mayoría de usuarios con género OTRO que utilizan este servicio tienen 45 años o menos, por lo tanto, podemos decir que una gran cantidad de usuarios con genero distinto a masculino y femenino, o que no indicaron su género, son personas jóvenes.