

# ESTUDIO BICIMAD

Marina Mas Villar  
Pablo Muñoz Vicente  
Arturo Tarazona Pérez

## 1. INTRODUCCIÓN

En este trabajo hemos realizado un estudio sobre el uso de las bicicletas públicas de la empresa BiciMAD distribuidas por la Comunidad de Madrid.

BiciMAD es uno de los medios de transporte público de la ciudad de Madrid, un servicio prestado al 100% con bicicletas eléctricas, práctico, sencillo y ecológico. Proyecto que lleva en marcha desde junio de 2014. El objetivo de BiciMAD es proporcionar un elemento alternativo de transporte limpio y saludable al ciudadano y fomentar el uso de la bicicleta en la ciudad.

Para realizar el estudio hemos cogido de la base de datos publicada en su página web ([https://opendata.emtmadrid.es/Datos-estaticos/Datos-generales-\(1\)](https://opendata.emtmadrid.es/Datos-estaticos/Datos-generales-(1))) varios archivos del tipo *.json*. En cada uno ellos encontramos información relativa a:

- **Tipo de usuario que realiza el viaje** (1: Usuario anual, 2: Usuario ocasional, 3: Trabajador de la empresa).
- **Rango de edad del usuario** (0: Desconocido, 1: entre 0 y 16 años, 2: entre 17 y 18 años, 3: entre 19 y 26 años, 4: entre 27 y 40 años, 5: entre 41 y 65 años, 6: más de 66 años)
- **Tiempo de viaje**
- **Fecha y hora estimada del viaje**
- **Estación de salida**
- **Estación de llegada**

Todos los estudios se realizarán sobre el mes de diciembre de 2019, utilizando los datos del archivo *"201912\_movements.json"*. Sin embargo, nos ha parecido interesante comparar el uso de las bicicletas en función del tipo de usuario del mes de diciembre con algún mes de periodo vacacional, para lo que hemos seleccionado el mes de julio de ese mismo año, cuyos datos encontramos en el archivo *"201907\_movements.json"*. o Puesto que el grupo de edad 0 es de "Rango desconocido" no lo tendremos en cuenta para ningún estudio, ya que no nos aporta ningún tipo de información.

Para poder ejecutar nuestro programa deberás tener los archivos con los datos anteriormente citados en la misma carpeta en la que se encuentre el archivo *.py* con el programa. Los resultados aparecerán automáticamente una vez que éste se ejecute.

En nuestro programa hemos implementado una función *main*, que lee, haciendo uso de la biblioteca *pyspark*, los datos contenidos en los archivos *.json*, añadiéndolos en las variables *rdd\_diciembre* y *rdd\_julio*, que son pasadas como parámetros de entrada a la función *estudios*.

Para entender el funcionamiento de *estudio* es necesario conocer cómo funciona *get\_data*. Ésta recibe como parámetro de entrada los datos relativos a un único viaje y haciendo uso de la biblioteca *json*, esta línea de texto es transformada en un diccionario del que se extrae la información relativa a los diversos factores que vamos a estudiar, devolviéndolos en la siguiente tupla: (*semana*, *tiempo\_viaje*, *tipo\_usuario*, *rango\_edad*, (*estacion\_salida*, *estacion\_llegada*))

En la función principal *estudios* aplicamos con un *map* la función *get\_data* a cada viaje, obteniendo así las listas de tuplas *datos\_diciembre* y *datos\_julio*.

## 2. MOTIVACIÓN Y ESTUDIOS REALIZADOS

BiciMAD quiere llegar a todo el mundo, hemos decidido centrar nuestro estudio en aspectos como la frecuencia de uso de las bicicletas o el tiempo medio de uso en función del rango de edad con el objetivo de motivar a aquellos grupos que menos consuman sus servicios a hacerlo, en lo que posiblemente también pueda tenerse en cuenta el uso por tipo de usuario, aspecto que también hemos decidido estudiar.

Con el objetivo de optimizar los recursos de la empresa, hemos decidido estudiar las 5 estaciones más y menos frecuentadas por los usuarios, para que BiciMAD valore si fuera beneficioso aumentar o disminuir el número de bases de éstas.

### USO EN FUNCIÓN DEL RANGO DE EDAD

Nuestra función *estudios* se encarga de analizar cómo influye el rango de edad del usuario en el uso de las bicicletas. Para ello, nos quedamos con el dato del rango de edad de cada viajero con la función *map*, y haciendo uso de *Counter* contamos cuántos hay de cada tipo. En primer lugar, empezamos distinguiendo si los viajes se hacían entre semana o en fin de semana (para esto hicimos uso de un *filter* que se quedaba con las tuplas correspondientes) y obtuvimos los siguientes resultados:

Uso en función del rango de edad entre semana:

{0: 77607, 4: 49352, 5: 41194, 3: 9144, 1: 1913, 6: 1061, 2: 770}

Uso en función del rango de edad en fin de semana:

{0: 25129, 4: 13157, 5: 8653, 3: 2885, 1: 488, 6: 307, 2: 305}

Observamos que no hay diferencias significativas entre el nivel de uso para los diferentes rangos de edad, por lo que no es relevante hacer esta distinción. Por tanto, al hacer el estudio global obtuvimos los siguientes resultados:

Uso en función del rango de edad:

{0: 102736, 4: 62509, 5: 49847, 3: 12029, 1: 2401, 6: 1368, 2: 1075}

De modo que los usuarios que más utilizan las bicicletas se encuentran en un rango de edad entre los 27 y 40 años, mientras que el rango que menos las utiliza es el correspondiente a 17 y 18 años.

### TIEMPO MEDIO DE USO EN FUNCIÓN DE LA EDAD

Queremos también ver la relación existente entre el tiempo medio de uso y la edad del usuario. Haremos el estudio distinguiendo entre los días de entre semana y los de fin de

semana. Al igual que en los casos anteriores, la función *estudios* es la encargada de realizar esto. En primer lugar, filtramos las tuplas para quedarnos con aquellas que recojan los datos de viajes realizados entre semana y haciendo uso de la función *map* nos quedamos únicamente con la información referida a la edad de los usuarios y el tiempo de uso de la bicicleta en el viaje. A continuación, agrupamos todos los tiempos por rango de edad haciendo uso de la función *groupByKey* y finalmente, calculamos la media de los tiempos correspondientes a los distintos rangos de edad. Hemos obtenido los siguientes resultados:

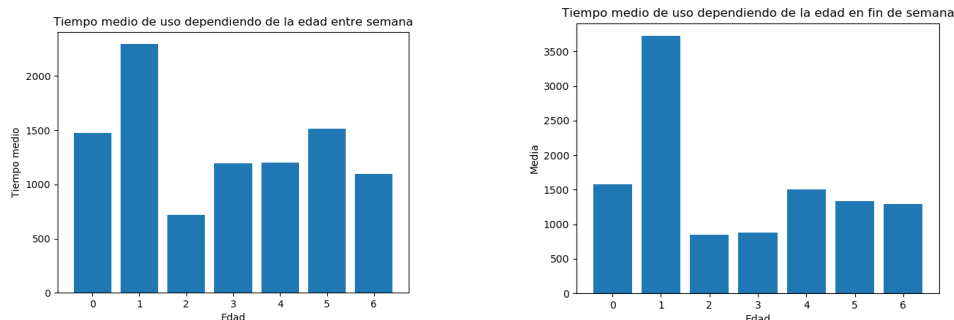
Lista de tiempo medio de uso según la edad entre semana:

```
[(0, 1477.772430322007), (1, 2294.96863565081), (2, 719.8779220779221), (3, 1195.2050524934384), (4, 1200.9450680823472), (5, 1514.5585279409622), (6, 1093.9877474081056)]
```

Lista de tiempo medio de uso según la edad en fin de semana:

```
[(0, 1577.7591627203628), (1, 3723.4180327868853), (2, 847.8852459016393), (3, 875.0329289428076), (4, 1499.141445618302), (5, 1335.442852189992), (6, 1290.4820846905538)]
```

Contrastamos estos resultados gráficamente:



Como podemos observar tanto gráfica como analíticamente el grupo de edad que invierte más tiempo en los viajes tanto entre semana como los fines de semana es el 1, entre 0 y 16 años, además el tiempo de uso por parte de este grupo aumenta los fines de semana. Por otra parte, encontramos diferencias significativas entre el uso que hacen el resto de grupos entre semana y los fines. Mientras que los grupos de edad entre 17-18, 27-40 y los mayores de 66 años aumentan el tiempo de sus viajes durante el fin de semana, los grupos de edad entre 19-26 y 41-65 hacen viajes más largos entre semana.

### USO EN FUNCIÓN DEL TIPO DE USUARIO POR ÉPOCAS DEL AÑO

Como ya hemos comentado, para este estudio hemos decidido comparar los datos con los que estamos trabajando en el trabajo, pertenecientes a diciembre de 2019, con los del mes de julio, por tratarse de un periodo vacacional en el que aparentemente el tipo de usuarios que harán uso de BiciMAD será distinto, pues tendría sentido que aparecieran más conductores ocasionales debido al turismo.

Para analizar el uso de BiciMAD en relación con el tipo de usuario, en la función *estudios*, almacenaremos haciendo uso de *filter* únicamente el tipo de usuario de cada viaje, tanto para *datos\_diciembre* como para *datos\_julio*, por separado. Tras esto, de nuevo usaremos *Counter* para contar el uso en los 3 tipos.

Los resultados que hemos obtenido son los siguientes:

Uso en función del tipo de usuario en verano: {1: 128757, 3: 7802, 2: 1295}  
Uso en función del tipo de usuario en invierno: {1: 214867, 3: 16476, 2: 622}

Observando los datos recogidos, sí es cierto que en general el uso de las bicicletas en diciembre de los usuarios con abono anual es más elevado que en julio. Esto puede deberse probablemente a que este grupo suele estar mayoritariamente ocupado por ciudadanos madrileños que en gran parte abandonan la comunidad debido al periodo vacacional, al igual que ocurre con los trabajadores de la empresa, que mantienen esta misma proporción. Sin embargo, lo que resulta más interesante de este estudio es que en el caso de los usuarios ocasionales ocurre lo contrario a los otros grupos. Los usuarios que deciden utilizar las bicicletas sin contratar abono anual, es decir, personas que probablemente no estén en la comunidad el resto del año, superan el doble de los que lo hacen en verano, lo cual puede deberse al turismo, ya sea extranjero u ocasional.

### **ESTACIONES MÁS Y MENOS TRANSITADAS**

Finalmente, hemos estudiado cuáles han sido las estaciones más y menos transitadas por los usuarios de BiciMAD a lo largo del mes de diciembre de 2019. Para realizar el estudio no hacemos distinción en estaciones de salida y estaciones de llegada, únicamente nos interesan cuáles han sido las más transitadas. Para ello nuestra función *estudios* haciendo uso de *Counter* cuenta el número de veces que se pasa por cada parada, ya sea de salida como de llegada. Hemos obtenido que las estaciones más transitadas son las siguientes: [(175, 6532), (43, 6342), (57, 5725), (83, 5342), (90, 4949)]. Las estaciones menos transitadas son las siguientes: [(193, 493), (144, 474), (29, 223), (1, 151), (28, 126)]

La estación 175, la más transitada, está en la calle Jaime el Conquistador n30, muy próximo al Parque de Madrid Río, zona de un gran interés turístico. La número 43 en la calle Lavapiés y la 83 en la calle Pío Baroja, cerca del Retiro. No nos sorprende que ésta sea una de las estaciones más transitadas ya que el paseo en bici por el Retiro es un plan típico de Madrid. La número 57 la calle Santa Cruz del Marcenado, entre Arguelles y San Bernardo y, por último, la 90 en la calle Goya.

Por otra parte entre las menos transitadas se encuentran las paradas número 28, 1 y 29, muy próximas a la Puerta del Sol, zona bastante peatonal. La parada 193 se corresponde con la zona de las Ventas y finalmente, la 144 en Diego de León.