**Fuentes de Datos**

**Analista de datos:** Luis Fernando Martínez

**Fecha de creación:** 2024/06/26

**Fecha de actualización:** 2024/07/02

**Fuente**

Los datos serán proporcionados por Motivate International Inc. en el enlace [Index of bucket "divvy-tripdata"](https://divvy-tripdata.s3.amazonaws.com/index.html) .

En el sitio de descarga, se buscan los archivos comprimidos con la convención de nombre “aaaamm-divvy-tripdata.zip”, cada comprimido contiene un archivo .csv y una carpeta con archivos para macOS, en este caso se usarán los .csv, cada comprimido corresponde a la información de viajes de un solo mes.

**Licencia**

[Data License Agreement | Divvy Bikes](https://divvybikes.com/data-license-agreement)

**Alcance**

Los datos proporcionados son datos públicos con restricciones de privacidad que prohíben usar información de identificación personal de los ciclistas.

Se utilizarán los registros de los últimos 12 meses, que, para efectos de este proyecto, serán los registros desde mayo de 2023 hasta mayo de 2024.

**Ubicación de los datos en el proyecto**

Los archivos fuente estarán en la ruta *Capstone Project/Data/Original*

Los archivos fuente originales no se deben modificar, en su lugar, se debe hacer una copia de estos para su preparación, proceso y análisis.

Estos son:

* 202305-divvy-tripdata.csv
* 202306-divvy-tripdata.csv
* 202307-divvy-tripdata.csv
* 202308-divvy-tripdata.csv
* 202309-divvy-tripdata.csv
* 202310-divvy-tripdata.csv
* 202311-divvy-tripdata.csv
* 202312-divvy-tripdata.csv
* 202401-divvy-tripdata.csv
* 202402-divvy-tripdata.csv
* 202403-divvy-tripdata.csv
* 202404-divvy-tripdata.csv
* 202405-divvy-tripdata.csv

**Estructura de los datos fuente**

Cada archivo contiene 13 columnas.

* ride\_id (texto): Identificador único del viaje.
* rideable\_type (texto): Tipo de bicicleta.
* started\_at (fecha tiempo): Fecha y tiempo del inicio del viaje.
* ended\_at (fecha tiempo): Fecha y tiempo del fin del viaje.
* start\_station\_name (texto): Nombre de la estación donde inició el viaje.
* start\_station\_id (texto): Identificador único de la estación donde inició el viaje.
* end\_station\_name (texto): Nombre de la estación donde terminó el viaje.
* end\_station\_id (texto): Identificador único de la estación donde terminó el viaje.
* start\_lat (decimal): Latitud de la ubicación donde inició el viaje.
* start\_lng (decimal): Longitud de la ubicación donde inició el viaje.
* end\_lat (decimal): Latitud de la ubicación donde terminó el viaje.
* end\_lng (decimal): Longitud de la ubicación donde terminó el viaje
* member\_casual (texto): Estado de membresía del usuario.

**Metadatos**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre Archivo** | **Extensión** | **Tamaño [MB]** | **Filas** |
| 202305-divvy-tripdata.csv | csv | 116 | 604828 |
| 202306-divvy-tripdata.csv | csv | 137 | 719619 |
| 202307-divvy-tripdata.csv | csv | 146 | 767651 |
| 202308-divvy-tripdata.csv | csv | 147 | 771694 |
| 202309-divvy-tripdata.csv | csv | 127 | 666372 |
| 202310-divvy-tripdata.csv | csv | 102 | 537114 |
| 202311-divvy-tripdata.csv | csv | 69,6 | 362519 |
| 202312-divvy-tripdata.csv | csv | 42,9 | 224074 |
| 202401-divvy-tripdata.csv | csv | 28 | 144874 |
| 202402-divvy-tripdata.csv | csv | 43,7 | 223165 |
| 202403-divvy-tripdata.csv | csv | 57,9 | 301688 |
| 202404-divvy-tripdata.csv | csv | 78,3 | 415026 |
| 202405-divvy-tripdata.csv | csv | 115 | 609494 |
| Total |  |  | 6348118 |

**Datos Preparados**

Los datos preparados se almacenarán en la ruta *Capstone Project/Data/Preparados*

Sobre cada uno de los archivos, se ordenarán los datos por los campos member\_casual de manera ascendente y sarted\_at de manera ascendente, luego se filtrarán filas que no tengan fechas de inicio o fin de viaje (started\_at y ended\_at respectivamente), y se filtrarán las filas que no tengan member\_casual.