# Description of data

1. **Identify Data Sources:**

The data is provided by Divvy, which is a program of the [Chicago Department of Transportation](http://www.cityofchicago.org/city/en/depts/cdot.html) (CDOT).

1. **Provide General Information:**

Divvy Data

* Divvy provides monthly data of its bike-share program to public
* Monthly data from 2022-07 to 2023-06 is being used in this project

1. **Describe Data Content:**

|  |  |
| --- | --- |
| Column | Description |
| ride\_id | The id of ride |
| rideable\_type | Type of bike |
| started\_at | Start time |
| ended\_at | End time |
| start\_station\_name | Start station name |
| start\_station\_id | Start station id |
| end\_station\_name | End station name |
| end\_station\_id | End station id |
| start\_lat | Start station latitude |
| start\_lng | Start station longitude |
| end\_lat | End station latitude |
| end\_lng | End station longitude |
| member\_casual | If rider buys annual memberships |

1. **Data Collection Method:**

Downloaded under the website:

<https://divvy-tripdata.s3.amazonaws.com/index.html>

1. **Data Cleaning and Preprocessing:**

Outline any cleaning and preprocessing steps performed on the data:

* Missing value handling.
* Data transformation (normalization, scaling, etc.).
* Outlier removal.
* Aggregation or summarization.

1. **Data Quality and Validity:**

Provide information about the data's quality and validity:

* Any known data quality issues.
* Steps taken to address or mitigate data quality issues.
* Data validation methods used.

1. **Data Updates and Frequency:**

Specify how often the data is updated or refreshed:

* Is it a one-time snapshot or regularly updated?
* Frequency of updates (daily, weekly, monthly, etc.).

1. **Data Ownership and Permissions:**

Mention any permissions or restrictions associated with the data:

* Data ownership (source, organization, etc.).
* Legal or ethical considerations (privacy, usage restrictions, etc.).

1. **Relevant Documentation:**

Provide links or references to any relevant documentation about the data source:

* Data dictionaries.
* API documentation.
* Methodology reports.

1. **Citations and References:**

If you're using publicly available data, provide citations or references to the original sources to give credit and ensure transparency.

1. **Data Source Relationships:**

Explain how different data sources are related, if applicable, and any data integration or joining processes performed.

1. **Visual Representation:**

Consider creating a visual representation of the data flow, showing how data is collected, processed, and integrated from various sources.