
Description of data

1. Identify Data Sources:

The data is provided by Divvy, which is a program of the [Chicago Department of Transportation](#) (CDOT).

2. 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

3. 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

4. Data Collection Method:

Downloaded under the website:

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

5. 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.

6. 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.

7. 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.).

8. 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.).

9. Relevant Documentation:

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

- Data dictionaries.
- API documentation.
- Methodology reports.

10. Citations and References:

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

11. Data Source Relationships:

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

12. Visual Representation:

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