Data Information

## Source :

Cyclistic is a fictional company. This is a historical trip data for the city of Chicago provided by Motivate International Inc downloaded from the following [link](https://divvybikes-marketing-staging.lyft.net/system-data), under this [license](https://divvybikes.com/data-license-agreement). Only data for the years 2021 and 2022 has been downloaded.

Data has good credibility as it is a second party's data from an existing well known bike share company.

## Data Organization :

Data is organized into csv files for each month, for a total of 24 files (two years). Each record represents a ride with the following attributes.

## Attributes :

| **Attribute** | **Type** | **Constraints** | **Description** |
| --- | --- | --- | --- |
| ride\_id | String | Should be Unique | A unique reference assigned to each ride |
| rideable\_type | String | Should follow a defined list of values | Bikes Types |
| started\_at | Date | - Should be formatted as date/Time  - Shouldn't be Blank  - Shouldn't be greater than end Date | The date & time the ride starts |
| ended\_at | Date | - Should be formatted as date/Time  - Shouldn't be Blank  - Shouldn't be Before start date | The date & time the ride Ends |
| start\_station\_name | String | - Should be a list of values of unique names  - Each name should have a unique id  - Shouldn't be blank | The name of the station where the bike has been picked up and the ride started |
| start\_station\_id | String | - Should be a list of values of unique ids  - Each id should have a unique name  - Shouldn't be blank | The id of the station where the bike has been picked up and the ride started |
| end\_station\_name | String | - Should be a list of values of unique names  - Each name should have a unique id  - Shouldn't be blank | The name of the station where the bike has been parked and the ride ended |
| end\_station\_id | String | - Should be a list of values of unique ids  - Each id should have a unique name  - Shouldn't be blank | The id of the station where the bike has been parked and the ride ended |
| start\_lat | Numeric/Float | - Should be between -90 & 90  - Decimals should be limited to 6 | Latitude where the ride started |
| start\_lng | Numeric/Float | - Should be between -180 & 180  - Decimals should be limited to 6 | Longitude where the ride started |
| end\_lat | Numeric/Float | - Should be between -90 & 90  - Decimals should be limited to 6 | Latitude where the ride ended |
| end\_lng | Numeric/Float | - Should be between -180 & 180  - Decimals should be limited to 6 | Longitude where the ride ended |
| member\_casual | String | - Should be "member" or "casual"  - Shouldn't be blank | Specify if the rider has an annual membership or not |

## Schema to answer the questions :

| - What are the members’ top five routes ?  - Are there areas where members are concentrated ? If yes, what are the top two Areas ? | Pairs of “start\_station\_id” / “end\_station\_id”  Pairs of “start\_lat-start\_lng” / end\_lat-end\_lng” |
| --- | --- |
| - What are the two most frequent times of the day for bike trips for members ?  - Do members ride during weekends or only on business days ? | “Started\_at”  “Ended\_at” |
| - How many times do members ride in a day ? | No available data per rider, only ride ids are available |

## Records Summary :

|  | Count\_Duplicates\_Id | Count\_NULL\_Id |
| --- | --- | --- |
| Data\_2021 + 2022 | 0 | 0 |

| **Year** | **2021** | **2022** |
| --- | --- | --- |
| Number\_Of\_Rides | 5595063 | 5667717 |
| Duplicates | 0 | 0 |
| Missing\_Start\_Time | 0 | 0 |
| Missing\_End\_Time | 0 | 0 |
| Missing\_Start\_Station\_Id | 690806 | 833064 |
| Missing\_End\_Station\_Id | 739170 | 892742 |
| Missing\_Start\_Coordinates | 0 | 0 |
| Missing\_End\_Coordinates | 9542 | 11716 |
| Missing\_Membership\_Type | 0 | 0 |
| Invalid\_Latitudes | 0 | 0 |
| Invalid\_Longitudes | 0 | 0 |

| **Data\_Check** | **Total** |
| --- | --- |
| Invalid\_DateTimes | 1184 |
| Missing\_Start\_Date | 0 |
| Missing\_End\_Date | 0 |