Here's an overview of the first few rows of the dataset:

| ride\_id | rideable\_type | started\_at | ended\_at | start\_station\_name | start\_station\_id | end\_station\_name | end\_station\_id | start\_lat | start\_lng | end\_lat | end\_lng | member\_casual |

|-------------------|---------------|---------------------|---------------------|---------------------------|------------------|-----------------------------|----------------|-----------|------------|----------|-----------|---------------|

| 011C1903BF4E2E28 | classic\_bike | 2023-09-23 00:27:50 | 2023-09-23 00:33:27 | Halsted St & Wrightwood Ave | TA1309000061 | Sheffield Ave & Wellington Ave | TA1307000052 | 41.929143 | -87.649077 | 41.936253 | -87.652662 | member |

| 87DB80E048A1BF9F | classic\_bike | 2023-09-02 09:26:43 | 2023-09-02 09:38:19 | Clark St & Drummond Pl | TA1307000142 | Racine Ave & Fullerton Ave | TA1306000026 | 41.931248 | -87.644336 | 41.925566 | -87.658416 | member |

| 7C2EB7AF669066E3 | electric\_bike | 2023-09-25 18:30:11 | 2023-09-25 18:41:39 | Financial Pl & Ida B Wells Dr | SL-010 | Racine Ave & 15th St | 13304 | 41.875061 | -87.633135 | 41.861267 | -87.656625 | member |

| 57D197B010269CE3 | classic\_bike | 2023-09-13 15:30:49 | 2023-09-13 15:39:18 | Clark St & Drummond Pl | TA1307000142 | Racine Ave & Belmont Ave | TA1308000019 | 41.931248 | -87.644336 | 41.939743 | -87.658865 | member |

| 8A2CEA7C8C8074D8 | classic\_bike | 2023-09-18 15:58:58 | 2023-09-18 16:05:04 | Halsted St & Wrightwood Ave | TA1309000061 | Racine Ave & Fullerton Ave | TA1306000026 | 41.929143 | -87.649077 | 41.925566 | -87.658416 | member |

The dataset contains the following columns:

- `ride\_id`: Unique identifier for each ride

- `rideable\_type`: Type of bike used (classic\_bike, electric\_bike, etc.)

- `started\_at`: Start time of the ride

- `ended\_at`: End time of the ride

- `start\_station\_name`: Name of the start station

- `start\_station\_id`: ID of the start station

- `end\_station\_name`: Name of the end station

- `end\_station\_id`: ID of the end station

- `start\_lat`: Latitude of the start station

- `start\_lng`: Longitude of the start station

- `end\_lat`: Latitude of the end station

- `end\_lng`: Longitude of the end station

- `member\_casual`: Type of user (member or casual)

Is there any specific analysis you would like to perform on this data?

### Data Cleaning Steps

\*\*Missing Values:\*\*

- `start\_station\_name`: 101,312 missing values

- `start\_station\_id`: 101,312 missing values

- `end\_station\_name`: 107,291 missing values

- `end\_station\_id`: 107,291 missing values

- `end\_lat`: 838 missing values

- `end\_lng`: 838 missing values

\*\*Data Types:\*\*

- `started\_at` and `ended\_at` should be converted to datetime

- Other columns have the appropriate data types

### Next Steps:

1. Handle missing values:

- For `start\_station\_name`, `start\_station\_id`, `end\_station\_name`, `end\_station\_id`: These might be missing due to rides starting or ending at unspecified locations.

- For `end\_lat` and `end\_lng`: We might drop these rows or fill them with mean/median values if appropriate.

2. Convert `started\_at` and `ended\_at` to datetime.

Let's start with these steps:

#### 1. Handle Missing Values

- Dropping rows with missing `end\_lat` and `end\_lng`

- For missing station names and IDs, fill with 'Unknown' or similar placeholder

#### 2. Convert `started\_at` and `ended\_at` to datetime

I'll proceed with these steps.

### Data Cleaning Completed

- All missing values have been addressed.

- The `started\_at` and `ended\_at` columns have been successfully converted to datetime format.

The dataset is now clean and ready for further analysis. What would you like to do next?

Let's perform the following analyses:

1. \*\*Summary Statistics\*\*: Get basic descriptive statistics for the dataset.

2. \*\*Ride Duration Analysis\*\*: Calculate the duration of each ride and analyze the distribution.

3. \*\*Trends Over Time\*\*: Analyze the number of rides over time (daily, weekly, monthly trends).

4. \*\*Rideable Type Analysis\*\*: Distribution of different types of bikes used.

5. \*\*User Type Analysis\*\*: Distribution of members vs. casual users.

### Summary Statistics

Let's start with the basic summary statistics for numerical columns.

### Summary Statistics

- \*\*Latitude and Longitude:\*\*

- `start\_lat`, `start\_lng`, `end\_lat`, `end\_lng`: Provide the geographical coordinates of the start and end points.

- Mean, standard deviation, and range (min to max) for latitude and longitude are provided, indicating the general area covered by the rides.

- \*\*Ride Duration (in minutes):\*\*

- Count: 665,533 rides

- Mean: 16.01 minutes

- Standard Deviation: 31.40 minutes

- Minimum: -81.68 minutes (negative value likely indicates data errors)

- 25th Percentile: 5.83 minutes

- Median (50th Percentile): 10.17 minutes

- 75th Percentile: 17.88 minutes

- Maximum: 1499.92 minutes

### Next Steps

1. \*\*Trends Over Time\*\*: Analyze the number of rides over time (daily, weekly, monthly trends).

2. \*\*Rideable Type Analysis\*\*: Distribution of different types of bikes used.

3. \*\*User Type Analysis\*\*: Distribution of members vs. casual users.

Let's start with analyzing trends over time.

### Trends Over Time

#### Daily Trends

- Number of rides per day (first few days in September 2023):

- 2023-09-01: 24,083 rides

- 2023-09-02: 26,435 rides

- 2023-09-03: 25,973 rides

- 2023-09-04: 23,069 rides

- 2023-09-05: 22,068 rides

#### Monthly Trends

- Total rides in September 2023: 665,533 rides

#### Weekly Trends

- Average number of rides by day of the week:

- Monday: 69,577 rides

- Tuesday: 75,875 rides

- Wednesday: 86,933 rides

- Thursday: 90,062 rides

- Friday: 115,988 rides

- Saturday: 137,587 rides

- Sunday: 89,511 rides

#### Hourly Trends

- Average number of rides by hour of the day:

- Peak hours: 17:00 (67,011 rides), 16:00 (58,430 rides)

- Morning peak: 08:00 (32,720 rides)

- Evening hours (20:00 to 23:00): Decrease in number of rides

### Next Steps

1. \*\*Rideable Type Analysis\*\*: Analyze the distribution of different types of bikes used.

2. \*\*User Type Analysis\*\*: Analyze the distribution of members vs. casual users.

Let's proceed with the rideable type analysis.

### Rideable Type Analysis

- \*\*Distribution of Different Types of Bikes:\*\*

- Classic Bikes: 346,534 rides

- Electric Bikes: 318,999 rides

### User Type Analysis

- \*\*Distribution of Members vs. Casual Users:\*\*

- Members: 404,596 rides

- Casual Users: 260,937 rides

### Summary

- The dataset consists of rides predominantly using classic bikes, followed closely by electric bikes.

- There are more rides by members compared to casual users.

Would you like to perform any additional analysis or visualizations on this data?