

Data profile

Data: 2022 citi bike ride information

Source: <https://citibikenyc.com/system-data>

Limitations: This data comes directly from the company and its database which makes this a very reliable data set, but it does not come without limitations. The data that I have chosen is only from the year of 2022 which offers a limitation on historical data. We are also limited by not having secondary data that may affect our own. I.e. weather data, event data, etc.

Ethics: I do not see any ethical limitations in the collection process of this data.

Variable	Description	quantitative/ qualitative	Structured/Unstructured	Time variant/invariant	Nominal/Ordinal/Discrete/Continuous
ridable_type	Type of bicycle	qualitative	structured	Time-invariant	Binary
start_station_name	name of start station	qualitative	structured	Time-invariant	Nominal
start_station_id	id of start station	quantitative	unstructured	Time-invariant	Discrete
end_station_name	name of end station	qualitative	structured	Time-invariant	Nominal
end_station_id	id of end station	quantitative	unstructured	Time-invariant	Discrete
start_latitude	starting latitude	quantitative	structured	Time-invariant	Discrete
start_longitude	starting longitude	quantitative	structured	Time-invariant	Discrete
end_latitude	ending latitude	quantitative	structured	Time-invariant	Discrete
end_longitude	ending longitude	quantitative	structured	Time-invariant	Discrete
member_type	member classification	qualitative	structured	Time-invariant	Binary
start_date	date of start	quantitative	structured	Time-invariant	Discrete
start_time	time ride started	quantitative	structured	Time-invariant	Discrete
end_date	date of the end of the ride	quantitative	structured	Time-invariant	Discrete
end_time	time ride ended	quantitative	structured	Time-invariant	Discrete

1. This data set contains 895485 rows and 14 columns.
2. Missing values have been noted but will not affect the analysis therefore have been left alone.
3. Data is free of duplicate values.

Questions:

What are the busiest days?

What are the busiest times?

How long do riders ride on average?

What are the most popular stations for riders?