

Annual database relationships

Cyclistic 2022-2023+

Duplicate

Cyclistic 2022-2023.csv1

Cyclistic 2022-2023.csv

Zip_Code_Lat_Long_Yea...

How do relationships differ from joins? [Learn more](#)

Cyclistic 2022-202...	Operator	Cyclistic 2022-202...
Abc Neighborhood S ▼	= ▼	Abc neighborhood ei ▼
# Zip Code Start ▼	= ▼	# zip code end (C) ▼

Cyclistic 2022-2023+

Duplicate

Cyclistic 2022-2023.csv1

Cyclistic 2022-2023.csv

Zip_Code_Lat_Long_Yea...

How do relationships differ from joins? [Learn more](#)

Cyclistic 2022-202...	Operator	Zip_Code_Lat_Lon...
# Zip Code End ▼	= ▼	# zip code end (Zi ▼

The first relationship involves the CSV file for annual trends generated with BigQuery, which is matched (related) with a duplicate file. These datasets were related by matching the initial neighborhood and zip code of each trip with the final neighborhood and zip code of each trip. This was done to calculate congestion, or the difference between bikes leaving and arriving at each neighborhood.

The second relationship involves the CSV file for annual trends generated with BigQuery, matched with a zip code and coordinates CSV file. These datasets were related by matching the zip codes of the final locations.

The second

Calculating total trips per station and congestion

Calculated Fields

- Total trips per initial neighborhood, using variables from the original CSV file.

Cyclistic 2022-2023+

```
{ FIXED [Neighborhood Start]: SUM([Trip Count]) }
```

- Total trips per final neighborhood, using variables from the duplicate CSV file.

Cyclistic 2022-2023+

×

```
{ FIXED [Neighborhood End ]: SUM([Trip Count ]) }
```

- Congestion

Cyclistic 2022-2023+

```
[Neigh_end_count]-[Neigh_start_count]
```

Summer database relationships

🗑️ Cyclistic 07-08-09 2023+

Duplicate



How do relationships differ from joins? [Learn more](#)

Cyclistic 07-08-09 ...	Operator	Cyclistic 07-08-09 ...
Abc STA Start ▼	= ▼	Abc end station nam ▼

The relationship involves the CSV file for summer trends generated with BigQuery, which is matched (related) with a duplicate file. These datasets were related by matching the initial Station (STA) of each trip with the final station of each trip. This was done to calculate congestion, or the difference between bikes leaving and arriving at each neighborhood.

Calculating total trips per station and congestion



Calculated Fields

- Total trips per initial station, using variables from the original CSV file.

Station_start_count 


```
{ FIXED [STA Start]: COUNT([STA Start]) }
```

- Total trips per final station, using variables from the duplicate CSV file.

End Station Count  

```
{ FIXED [end station name (Cyclistic 07-08-09 2023.csv1)]  
: COUNT([end station name (Cyclistic 07-08-09 2023.csv1)] }
```

- Congestion

Congestion 

```
[End Station Count] - [Station_start_count]
```