# Data Preparation strategies

## Data source

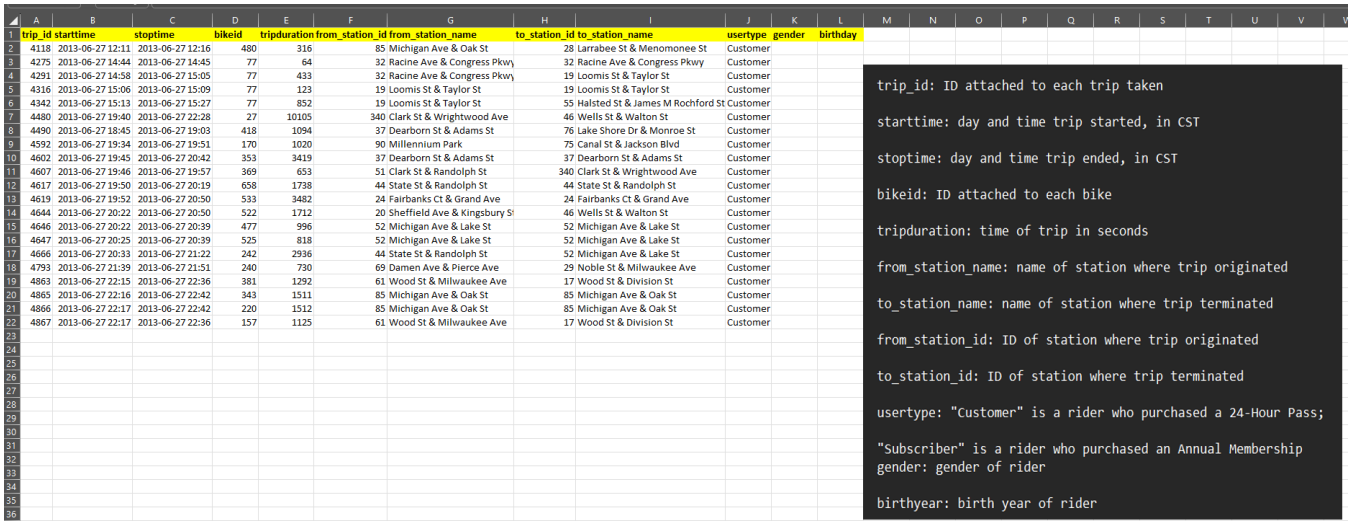
I will use Cyclistic's historical trip data to analyze and identify trends. Click here. This is a real public dataset suitable for exploring how different customer types are using Cyclistic bikes. But note that data-privacy issues prohibit you from using riders' personally identifiable information. The data has been made available by Motivate International Inc. under this license.)

## How data are organized

- There are various csv files that are in Zip folders.
- Data size ranges from 3 MB up to 57 MB
- Data are arranged in time periods (in years and quarters)
- Data are stored in long format.

To quickly capture the columns and shape of data, I manually snapshot some top rows for each table in Excel and compare similarities.

**The first form of data table Is from 2013 – 2019. These data tables include key features such as Trip duration and gender and birthday. These tables can go up to 1+ millions of rows.**



| trip_id | starttime | stoptime | bikeid | tripduration | from_station_id | from_station_name | to_station_id | to_station_name | usertype | gender | birthday |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4118 | 2013-06-27 12:11 | 2013-06-27 12:16 | 480 | 316 | 85 | Michigan Ave & Oak St | 28 | Larrabee St & Menomonee St | Customer | | |
| 4275 | 2013-06-27 14:44 | 2013-06-27 14:45 | 77 | 64 | 32 | Racine Ave & Congress Pkwy | 32 | Racine Ave & Congress Pkwy | Customer | | |
| 4291 | 2013-06-27 14:58 | 2013-06-27 15:05 | 77 | 433 | 32 | Racine Ave & Congress Pkwy | 19 | Loomis St & Taylor St | Customer | | |
| 4316 | 2013-06-27 15:06 | 2013-06-27 15:09 | 77 | 123 | 19 | Loomis St & Taylor St | 19 | Loomis St & Taylor St | Customer | | |
| 4342 | 2013-06-27 15:13 | 2013-06-27 15:27 | 77 | 852 | 19 | Loomis St & Taylor St | 55 | Halsted St & James M Rochford St | Customer | | |
| 4480 | 2013-06-27 19:40 | 2013-06-27 22:28 | 27 | 10105 | 340 | Clark St & Wrightwood Ave | 46 | Wells St & Walton St | Customer | | |
| 4490 | 2013-06-27 18:45 | 2013-06-27 19:03 | 418 | 1094 | 37 | Dearborn St & Adams St | 76 | Lake Shore Dr & Monroe St | Customer | | |
| 4592 | 2013-06-27 19:34 | 2013-06-27 19:51 | 170 | 1020 | 90 | Millennium Park | 75 | Canal St & Jackson Blvd | Customer | | |
| 4602 | 2013-06-27 19:45 | 2013-06-27 20:42 | 353 | 3419 | 37 | Dearborn St & Adams St | 37 | Dearborn St & Adams St | Customer | | |
| 4607 | 2013-06-27 19:46 | 2013-06-27 19:57 | 369 | 653 | 51 | Clark St & Randolph St | 340 | Clark St & Wrightwood Ave | Customer | | |
| 4617 | 2013-06-27 19:50 | 2013-06-27 20:19 | 658 | 1738 | 44 | State St & Randolph St | 44 | State St & Randolph St | Customer | | |
| 4619 | 2013-06-27 19:52 | 2013-06-27 20:50 | 533 | 3482 | 24 | Fairbanks Ct & Grand Ave | 24 | Fairbanks Ct & Grand Ave | Customer | | |
| 4644 | 2013-06-27 20:22 | 2013-06-27 20:50 | 522 | 1712 | 20 | Sheffield Ave & Kingsbury St | 46 | Wells St & Walton St | Customer | | |
| 4646 | 2013-06-27 20:22 | 2013-06-27 20:39 | 477 | 996 | 52 | Michigan Ave & Lake St | 52 | Michigan Ave & Lake St | Customer | | |
| 4647 | 2013-06-27 20:25 | 2013-06-27 20:39 | 525 | 818 | 52 | Michigan Ave & Lake St | 52 | Michigan Ave & Lake St | Customer | | |
| 4666 | 2013-06-27 20:33 | 2013-06-27 21:22 | 242 | 2936 | 44 | State St & Randolph St | 52 | Michigan Ave & Lake St | Customer | | |
| 4793 | 2013-06-27 21:39 | 2013-06-27 21:51 | 240 | 730 | 69 | Damen Ave & Pierce Ave | 29 | Noble St & Milwaukee Ave | Customer | | |
| 4863 | 2013-06-27 22:15 | 2013-06-27 22:36 | 381 | 1292 | 61 | Wood St & Milwaukee Ave | 17 | Wood St & Division St | Customer | | |
| 4865 | 2013-06-27 22:16 | 2013-06-27 22:42 | 343 | 1511 | 85 | Michigan Ave & Oak St | 85 | Michigan Ave & Oak St | Customer | | |
| 4866 | 2013-06-27 22:17 | 2013-06-27 22:42 | 220 | 1512 | 85 | Michigan Ave & Oak St | 85 | Michigan Ave & Oak St | Customer | | |
| 4867 | 2013-06-27 22:17 | 2013-06-27 22:36 | 157 | 1125 | 61 | Wood St & Milwaukee Ave | 17 | Wood St & Division St | Customer | | |

trip_id: ID attached to each trip taken

starttime: day and time trip started, in CST

stoptime: day and time trip ended, in CST

bikeid: ID attached to each bike

tripduration: time of trip in seconds

from_station_name: name of station where trip originated

to_station_name: name of station where trip terminated

from_station_id: ID of station where trip originated

to_station_id: ID of station where trip terminated

usertype: "Customer" is a rider who purchased a 24-Hour Pass;

"Subscriber" is a rider who purchased an Annual Membership
gender: gender of rider

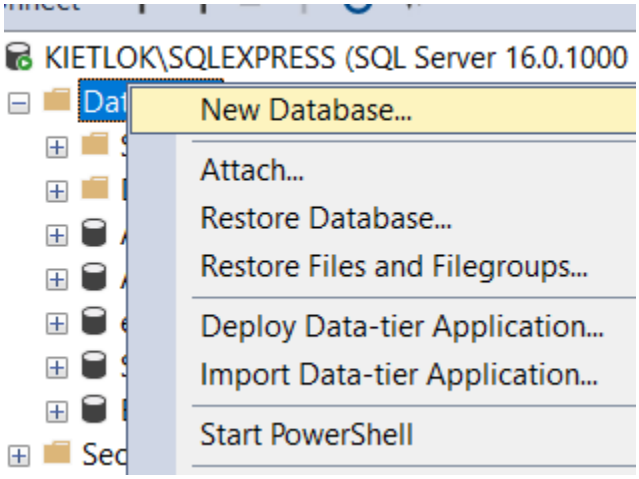
birthyear: birth year of rider

**The second form of data is from 2020 – August 2024. These tables include key features such as start and end time, member type but with no Gender and Birthday. These tables can go up to 1+ millions of rows.**

| 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BD0A6FF6FFF9B921 | electric_bike | 2020-11-01 13:36 | 2020-11-01 13:45 | Dearborn St & Erie St | 110 | St. Clair St & Erie St | 211 | 41.8941765 | -87.62912733 | 41.89443417 | -87.62337917 | casual |
| 96A7A7A4BDE4F82D | electric_bike | 2020-11-01 10:03 | 2020-11-01 10:14 | Franklin St & Illinois St | 672 | Noble St & Milwaukee Ave | 29 | 41.89095867 | -87.63534283 | 41.900675 | -87.66248033 | casual |
| C61526D06582BDC5 | electric_bike | 2020-11-01 0:34 | 2020-11-01 1:03 | Lake Shore Dr & Monroe St | 76 | Federal St & Polk St | 41 | 41.88098283 | -87.61675417 | 41.8720545 | -87.62955033 | casual |
| E533E89C32080B9E | electric_bike | 2020-11-01 0:45 | 2020-11-01 0:54 | Leavitt St & Chicago Ave | 659 | Stave St & Armitage Ave | 185 | 41.89549917 | -87.682013 | 41.9177445 | -87.69139183 | casual |
| 1C9F4EF18C168C60 | electric_bike | 2020-11-01 15:43 | 2020-11-01 16:16 | Buckingham Fountain | 2 | Buckingham Fountain | 2 | 41.87649733 | -87.620358 | 41.87644833 | -87.620338 | casual |
| 7259585D8276D338 | electric_bike | 2020-11-14 15:55 | 2020-11-14 16:44 | Wabash Ave & 16th St | 72 | Lake Shore Dr & Monroe St | 76 | 41.86028883 | -87.625806 | 41.880985 | -87.6167735 | casual |
| 91FE5C8F8A676594 | electric_bike | 2020-11-14 16:47 | 2020-11-14 17:03 | Lake Shore Dr & Monroe St | 76 | Wabash Ave & 16th St | 72 | 41.88100567 | -87.61677617 | 41.86047367 | -87.62584233 | casual |
| 9E7A79ADA90C2695 | electric_bike | 2020-11-14 16:04 | 2020-11-14 16:19 | | | | | 41.91 | -87.62 | 41.91 | -87.62 | casual |
| A5B02C0D41DBCDAF | electric_bike | 2020-11-14 16:24 | 2020-11-14 16:51 | Marshfield Ave & Cortland St | 58 | Larrabee St & Armitage Ave | 288 | 41.91606667 | -87.6690415 | 41.91814933 | -87.643875 | casual |
| 8234407C29FE41DC | electric_bike | 2020-11-14 1:24 | 2020-11-14 1:31 | Clark St & 9th St (AMLI) | 394 | Michigan Ave & 18th St | 273 | 41.87085383 | -87.63116867 | 41.8579115 | -87.62466683 | casual |
| 3D2F9317211E3350 | electric_bike | 2020-11-14 12:05 | 2020-11-14 12:09 | Michigan Ave & 8th St | 623 | Buckingham Fountain | 2 | 41.87260267 | -87.624212 | 41.876464 | -87.620367 | casual |
| 586E86B0E03B75CE | electric_bike | 2020-11-14 9:10 | 2020-11-14 9:23 | | | Spaulding Ave & Armitage Ave | 506 | 41.95 | -87.71 | 41.9171835 | -87.71024533 | casual |
| 1587848E23E6703D | electric_bike | 2020-11-14 15:03 | 2020-11-14 15:09 | Lakeview Ave & Fullerton Pkwy | 313 | Lakeview Ave & Fullerton Pkwy | 313 | 41.92578567 | -87.63902483 | 41.92589483 | -87.63909383 | casual |
| 953E715FCB27D545 | electric_bike | 2020-11-14 13:04 | 2020-11-14 13:10 | | | | | 41.79 | -87.59 | 41.8 | -87.59 | casual |
| 67359E807ED2BA0D | electric_bike | 2020-11-14 7:26 | 2020-11-14 7:34 | Mies van der Rohe Way & Chicago Ave | 173 | Clark St & Schiller St | 301 | 41.8970255 | -87.621665 | 41.907886 | -87.63142783 | casual |

## Load Data

In this project, I will use SQL Server Studio to manage and query data. First, I will create a new database named Bike_Share.




Then I will Load data csv from 2013 to 2023

While loading the data to SQL Server Database, I encountered some tables have wrong columns name

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 - Renta | 01 - Renta | 01 - Renta | 01 - Renta | 01 - Renta | 03 - Renta | 03 - Renta | 02 - Renta | 02 - Renta | User Type | Member ( | 05 - Member Details | Member Birthday Year | | |
| 22178529 | ######## | ######## | 6251 | 446 | 81 | Daley Cen | 56 | Desplaine | Subscribe | Male | 1975 | | | |
| 22178530 | ######## | ######## | 6226 | 1,048.00 | 317 | Wood St & | 59 | Wabash A | Subscribe | Female | 1984 | | | |
| 22178531 | ######## | ######## | 5649 | 252 | 283 | LaSalle St | 174 | Canal St & | Subscribe | Male | 1990 | | | |

To resolve this, I simply change the columns name to align with the other tables.

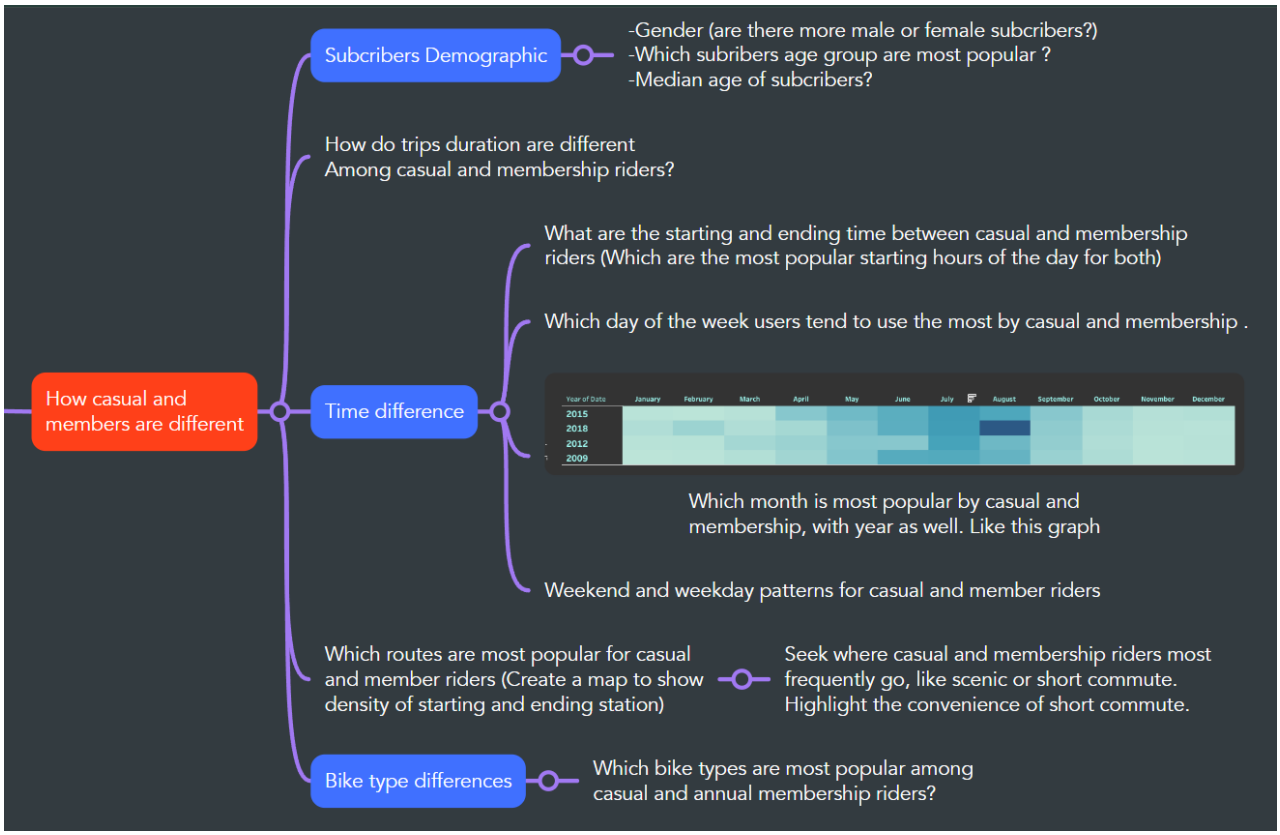
**After that, I backup the database (Bike_Share.bak)**

Before we proceed to any data modeling phase, I selected top 100 rows of all the form of data and visualize any potential valuable insights that I can extract from this data and then start the data modeling process.

| | trip_id | starttime | stoptime | bikeid | tripduration | from_station_id | from_station_name | to_station_id | to_station_name | usertype | gender | birthyear |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2886259 | 2014-07-31 23:56:00 | 2014-08-01 00:03:00 | 2602 | 386 | 291 | Wells St & Evergreen Ave | 53 | Wells St & Erie St | Subscriber | Female | 1979 |
| 2 | 2886258 | 2014-07-31 23:58:00 | 2014-08-01 00:07:00 | 2403 | 495 | 98 | LaSalle St & Washington St | 106 | State St & Pearson St | Subscriber | Male | 1974 |
| 3 | 2886257 | 2014-07-31 23:58:00 | 2014-08-01 02:10:00 | 669 | 7947 | 240 | Sheridan Rd & Irving Park Rd | 240 | Sheridan Rd & Irving Park Rd | Customer | NULL | NULL |
| 4 | 2886256 | 2014-07-31 23:58:00 | 2014-08-01 00:19:00 | 2431 | 1282 | 47 | State St & Kinzie St | 14 | Morgan St & 18th St | Customer | NULL | NULL |
| 5 | 2886255 | 2014-07-31 23:57:00 | 2014-08-01 02:10:00 | 2885 | 7972 | 240 | Sheridan Rd & Irving Park Rd | 240 | Sheridan Rd & Irving Park Rd | Customer | NULL | NULL |
| 6 | 2886254 | 2014-07-31 23:57:00 | 2014-08-01 02:28:00 | 2307 | 9067 | 28 | Larrabee St & Menomonee St | 28 | Larrabee St & Menomonee St | Customer | NULL | NULL |
| 7 | 2886253 | 2014-07-31 23:57:00 | 2014-08-01 01:37:00 | 192 | 6006 | 255 | Indiana Ave & Roosevelt Rd | 181 | LaSalle St & Illinois St | Customer | NULL | NULL |
| 8 | 2886252 | 2014-07-31 23:56:00 | 2014-08-01 00:16:00 | 1543 | 1161 | 288 | Larrabee St & Armitage Ave | 17 | Wood St & Division St | Customer | NULL | NULL |
| 9 | 2886251 | 2014-07-31 23:56:00 | 2014-08-01 00:10:00 | 2801 | 824 | 29 | Noble St & Milwaukee Ave | 47 | State St & Kinzie St | Customer | NULL | NULL |
| 10 | 2886250 | 2014-07-31 23:56:00 | 2014-08-01 02:29:00 | 3030 | 9142 | 28 | Larrabee St & Menomonee St | 28 | Larrabee St & Menomonee St | Customer | NULL | NULL |
| 11 | 2886249 | 2014-07-31 23:56:00 | 2014-08-01 00:04:00 | 1422 | 505 | 246 | Ashland Ave & Belle Plaine Ave | 326 | Clark St & Leland Ave | Subscriber | Male | 1988 |
| 12 | 2886248 | 2014-07-31 23:56:00 | 2014-08-01 00:04:00 | 1329 | 464 | 291 | Wells St & Evergreen Ave | 106 | State St & Pearson St | Subscriber | Female | 1988 |
| 13 | 2886247 | 2014-07-31 23:56:00 | 2014-08-01 00:10:00 | 444 | 843 | 29 | Noble St & Milwaukee Ave | 47 | State St & Kinzie St | Customer | NULL | NULL |
| 14 | 2886246 | 2014-07-31 23:56:00 | 2014-08-01 00:14:00 | 2920 | 1121 | 113 | Bissell St & Armitage Ave | 289 | Wells St & Concord Ln | Subscriber | Male | 1986 |

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 9340B064F0AEE130 | electric_bike | 2023-07-23 20:06:00 | 2023-07-23 20:23:00 | Kedzie Ave & 110th St | 20204 | Public Rack - Racine Ave & 109th Pl | 877 | 41.6924057006836 | -87.7009048461914 | 41.6948356628418 | -87.6530380249023 | member |
| 2 | D1460EE3CE0D8AF8 | classic_bike | 2023-07-23 17:05:00 | 2023-07-23 17:19:00 | Western Ave & Walton St | KA1504000103 | Milwaukee Ave & Grand Ave | 13033 | 41.8984184265137 | -87.6865997314453 | 41.8915786743164 | -87.6483840942383 | member |
| 3 | DF41BE31B895A25E | classic_bike | 2023-07-23 10:15:00 | 2023-07-23 10:24:00 | Western Ave & Walton St | KA1504000103 | Damen Ave & Pierce Ave | TA1305000041 | 41.8984184265137 | -87.6865997314453 | 41.9093971252441 | -87.6776885986328 | member |
| 4 | 9624A293749EF703 | electric_bike | 2023-07-21 08:28:00 | 2023-07-21 08:33:00 | Racine Ave & Randolph St | 13155 | Clinton St & Madison St | TA1305000032 | 41.8841133117676 | -87.6569442749023 | 41.8827514648438 | -87.6411895751953 | member |
| 5 | 2F68A6A4CDB4C99A | classic_bike | 2023-07-08 15:47:00 | 2023-07-08 15:58:00 | Clark St & Leland Ave | TA1309000014 | Montrose Harbor | TA1308000012 | 41.9670867919922 | -87.6672897338867 | 41.963981628418 | -87.63818359375 | member |
| 6 | 9AEE973E6B941A9C | classic_bike | 2023-07-10 08:45:00 | 2023-07-10 08:50:00 | Racine Ave & Randolph St | 13155 | Sangamon St & Lake St | TA1306000015 | 41.8840675354004 | -87.656852722168 | 41.8857803344727 | -87.6510238647461 | member |
| 7 | E366E997FDA1582B | classic_bike | 2023-07-25 14:31:00 | 2023-07-25 14:38:00 | Clark St & Leland Ave | TA1309000014 | Sheridan Rd & Montrose Ave | TA1307000107 | 41.9670867919922 | -87.6672897338867 | 41.961669921875 | -87.6546401977539 | member |
| 8 | 1BB3E73851E6C2C1 | classic_bike | 2023-07-07 10:12:00 | 2023-07-07 10:18:00 | Clark St & Leland Ave | TA1309000014 | Ravenswood Ave & Berteau Ave | TA1309000018 | 41.9670867919922 | -87.6672897338867 | 41.9579200744629 | -87.6735687255859 | member |
| 9 | DA1E1D0866E6566E | electric_bike | 2023-07-04 21:57:00 | 2023-07-04 22:08:00 | Clark St & Leland Ave | TA1309000014 | Sheffield Ave & Wellington Ave | TA1307000052 | 41.9670906066895 | -87.6674880981445 | 41.9362525939941 | -87.6526641845703 | member |
| 10 | 39BF4A73A704CA85 | classic_bike | 2023-07-29 10:51:00 | 2023-07-29 11:03:00 | Warren Park East | RP-002 | Sheridan Rd & Loyola Ave | RP-009 | 42.0045509338379 | -87.6806640625 | 42.0010452270508 | -87.6612014770508 | member |
| 11 | F036C3340FF1EE0F | classic_bike | 2023-07-22 13:28:00 | 2023-07-22 13:47:00 | Jefferson St & Monroe St | WL-011 | Halsted St & Maxwell St | TA1309000001 | 41.8803291320801 | -87.6427459716797 | 41.8648834228516 | -87.6470718383789 | member |
| 12 | 253D62F6B36B340C | electric_bike | 2023-07-03 12:36:00 | 2023-07-03 12:44:00 | Public Rack - Kostner Av... | 1019 | Kedzie Ave & Milwaukee Ave | 13085 | 41.9317512512207 | -87.7371520996094 | 41.9295654296875 | -87.7078552246094 | member |
| 13 | 1AE843E603ED53A2 | electric_bike | 2023-07-11 12:43:00 | 2023-07-11 12:51:00 | Public Rack - Kostner Av... | 1019 | Kedzie Ave & Milwaukee Ave | 13085 | 41.9317665100098 | -87.7371215820313 | 41.9295654296875 | -87.7078552246094 | member |
| 14 | 398A58DC66DFAD8A | electric_bike | 2023-07-16 08:10:00 | 2023-07-16 08:16:00 | Western Ave & Walton St | KA1504000103 | Wood St & Milwaukee Ave | 13221 | 41.8984222412109 | -87.6864929199219 | 41.9076538085938 | -87.6725540161133 | member |
| 15 | C87CB0A25C9682DF | classic_bike | 2023-07-24 21:35:00 | 2023-07-24 21:47:00 | Racine Ave & Randolph St | 13155 | Wood St & Milwaukee Ave | 13221 | 41.8840675354004 | -87.656852722168 | 41.9076538085938 | -87.6725540161133 | member |
| 16 | 70FFA226EBD6721F | classic_bike | 2023-07-16 10:58:00 | 2023-07-16 11:05:00 | Racine Ave & Randolph St | 13155 | Clinton St & Lake St | 13021 | 41.8840675354004 | -87.656852722168 | 41.8856353759766 | -87.6418228149414 | member |

Here are my initial insights. We might discover more in the EDA phase.



## Data Integration

Because of the uniqueness of this data, tables from 2013 to 2019 contain Gender and Birthday columns, but starting from 2020, these two columns are dropped and a new Bike Type column as well as the longitude and latitude are introduced. So, I will merge 2013-2019 and 2020-2023.

Note that when populating the temp table, I won't need some Id columns to use less computer resources.

```
USE Bike_Share;

/* DATA INTEGRATION: MERGING TABLES */

-- Create a table called Trips_2013_2019
CREATE TABLE Trips_2013_2019
(
    Start_Time smalldatetime,
    Stop_Time smalldatetime,
    Trip_Duration int,
    From_Station_Name nvarchar(70),
    To_Station_Name nvarchar(70),
    User_Type nvarchar(50),
    Gender char(10),
    Birth_Year smallint,
    From_Station_Id smallint,
    To_Station_Id smallint
);

-- Insert data from Trips_2013
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthday, from_station_id, to_station_id
FROM [Trips_2013];


--Insert data from Trips_2014_Q1_Q2
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [Trips_2014_Q1_Q2];


--Repeat the insertions for the remaining tables
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [Trips_2014_Q3_07];
```

```
-- Insert [dbo].[Trips_2014_Q3_0809]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [Trips_2014_Q3_0809];


-- Insert [dbo].[Trips_2014_Q4]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2014_Q4]

-- Insert [dbo].[Trips_2015_Q1]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2015_Q1]

-- Insert [dbo].[Trips_2015_Q2]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2015_Q2]

-- Insert [dbo].[Trips_2015_Q3_07]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2015_Q3_07]
```

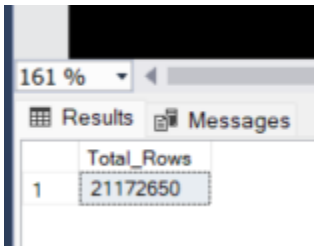
```
-- Insert [dbo].[Trips_2015_Q3_08]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2015_Q3_08]

-- Insert [dbo].[Trips_2015_Q3_09]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2015_Q3_09]

-- Insert [dbo].[Trips_2015_Q4]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2015_Q4]

-- Insert [dbo].[Trips_2016_Q1]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2016_Q1]

-- Insert [dbo].[Trips_2016_Q2_04]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2016_Q2_04]

-- Insert [dbo].[Trips_2016_Q2_05]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2016_Q2_05]

-- Insert [dbo].[Trips_2016_Q2_06]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2016_Q2_06]
```

```
-- Insert [dbo].[Trips_2016_Q3]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2016_Q3]

-- Insert [dbo].[Trips_2016_Q4]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2016_Q4]

-- Insert [dbo].[Trips_2017_Q1]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2017_Q1]

-- Insert [dbo].[Trips_2017_Q2]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2017_Q2]

-- Insert [dbo].[Trips_2017_Q3]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2017_Q3]
```

```
-- Insert [dbo].[Trips_2017_Q4]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2017_Q4]


-- Insert [dbo].[Trips_2018_Q1]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthday, from_station_id, to_station_id
FROM [dbo].[Trips_2018_Q1]

-- Insert [dbo].[Trips_2018_Q2]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2018_Q2]

-- Insert [dbo].[Trips_2018_Q3]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2018_Q3]

-- Insert [dbo].[Trips_2018_Q4]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2018_Q4]
```

```sql
-- Insert [dbo].[Trips_2019_Q1]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2019_Q1]

-- Insert [dbo].[Trips_2019_Q2]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    starttime, stoptime, tripduration, from_station_name, to_station_name, usertype, gender,
birthday, from_station_id, to_station_id
FROM [dbo].[Trips_2019_Q2]

-- Insert [dbo].[Trips_2019_Q3]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2019_Q3]

-- Insert [dbo].[Trips_2019_Q4]
INSERT INTO Trips_2013_2019
(Start_Time, Stop_Time, Trip_Duration, From_Station_Name, To_Station_Name, User_Type, Gender,
Birth_Year, From_Station_Id, To_Station_Id)
SELECT
    start_time, end_time, tripduration, from_station_name, to_station_name, usertype, gender,
birthyear, from_station_id, to_station_id
FROM [dbo].[Trips_2019_Q4]

-- Finally, check the total rows for the completed table
SELECT COUNT(*) AS Total_Rows FROM Trips_2013_2019;
```

Our Temp table Trips_2013_2019 will have 21,172,650 rows. That's huge!!




Next, we will repeat the merging process for tables from 2020 to 2023.

```
----Create a temp table merging data from 2020 to 2023
CREATE TABLE Trips_2020_2023 (
       Ride_Type nvarchar(50),
       Start_Time smalldatetime,
       Stop_Time smalldatetime,
       From_Station_Name nvarchar(70),
    To_Station_Name nvarchar(70),
       Start_Lat float,
       Start_lng float,
       End_Lat float,
       End_Lng float,
       User_Type nvarchar(50)
       )

--Insert [dbo].[Trips_2020_Q1]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q1]

--Insert [dbo].[Trips_2020_Q2_04]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q2_04]


--Insert [dbo].[Trips_2020_Q2_05]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q2_05]
```

```
--Insert [dbo].[Trips_2020_Q2_06]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q2_06]

--Insert [dbo].[Trips_2020_Q3_07]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q3_07]

--Insert [dbo].[Trips_2020_Q3_08]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q3_08]

--Insert [dbo].[Trips_2020_Q3_09]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q3_09]

--Insert [dbo].[Trips_2020_Q4_10]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q4_10]
```

```
--Insert [dbo].[Trips_2020_Q4_11]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q4_11]


--Insert [dbo].[Trips_2020_Q4_12]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2020_Q4_12]

--Insert [dbo].[Trips_2021_Q1_01]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q1_01]

--Insert [dbo].[Trips_2021_Q1_02]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q1_02]

--Insert [dbo].[Trips_2021_Q1_03]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q1_03]
```

```
--Insert [dbo].[Trips_2021_Q2_04]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q2_04]

--Insert [dbo].[Trips_2021_Q2_05]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q2_05]


--Insert [dbo].[Trips_2021_Q2_06]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q2_06]

--Insert [dbo].[Trips_2021_Q3_07]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q3_07]

--Insert [dbo].[Trips_2021_Q3_08]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
                                    FROM  [dbo].[Trips_2021_Q3_08]
```

```
--Insert [dbo].[Trips_2021_Q3_09]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q3_09]


--Insert [dbo].[Trips_2021_Q4_10]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q4_10]

--Insert [dbo].[Trips_2021_Q4_11]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q4_11]


--Insert [dbo].[Trips_2021_Q4_12]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2021_Q4_12]


--Insert [dbo].[Trips_2022_Q1_01]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q1_01]
```

```
--Insert [dbo].[Trips_2022_Q1_02]
INSERT INTO Trips_2020_2023
	(Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
	rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q1_02]

--Insert [dbo].[Trips_2022_Q1_03]
INSERT INTO Trips_2020_2023
	(Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
	rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q1_03]


--Insert [dbo].[Trips_2022_Q2_04]
INSERT INTO Trips_2020_2023
	(Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
	rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q2_04]

--Insert [dbo].[Trips_2022_Q2_05]
INSERT INTO Trips_2020_2023
	(Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
	rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q2_05]

--Insert [dbo].[Trips_2022_Q2_06]
INSERT INTO Trips_2020_2023
	(Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat, Start_lng,
End_Lat, End_Lng, User_Type)
SELECT
	rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q2_06]
```

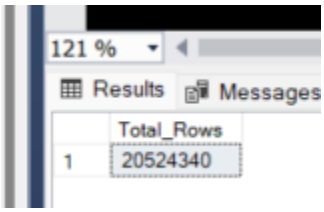
```
--Insert [dbo].[Trips_2022_Q3_07]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q3_07]

--Insert [dbo].[Trips_2022_Q3_08]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q3_08]


--Insert [dbo].[Trips_2022_Q3_09]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q3_09]

--Insert [dbo].[Trips_2022_Q4_10]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM [dbo].[Trips_2022_Q4_10]


--Insert [dbo].[Trips_2022_Q4_11]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q4_11]
```

```
--Insert [dbo].[Trips_2022_Q4_12]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2022_Q4_12]

--Insert [dbo].[Trips_2023_Q1_01]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q1_01]


--Insert [dbo].[Trips_2023_Q1_02]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q1_02]

--Insert [dbo].[Trips_2023_Q1_03]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q1_03]

--Insert [dbo].[Trips_2023_Q2_04]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q2_04]

--Insert [dbo].[Trips_2023_Q2_05]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q2_05]
```

```
--Insert [dbo].[Trips_2023_Q2_06]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q2_06]

--Insert [dbo].[Trips_2023_Q3_07]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q3_07]

--Insert [dbo].[Trips_2023_Q3_08]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q3_08]

--Insert [dbo].[Trips_2023_Q3_9]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q3_9]

--Insert [dbo].[Trips_2023_Q4_10]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q4_10]

--Insert [dbo].[Trips_2023_Q4_11]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q4_11]
```

```
--Insert [dbo].[Trips_2023_Q4_12]
INSERT INTO Trips_2020_2023
       (Ride_Type, Start_Time, Stop_Time, From_Station_Name, To_Station_Name, Start_Lat,
Start_lng, End_Lat, End_Lng, User_Type)
SELECT
       rideable_type, started_at, ended_at, start_station_name, end_station_name, start_lat,
start_lng, end_lat, end_lng, member_casual
FROM  [dbo].[Trips_2023_Q4_12]


-- Finally, check the total rows for the completed table
SELECT COUNT(*) AS Total_Rows FROM Trips_2020_2023;
```

Our Trips_2020_2023 table contain 20,524,340 rows.

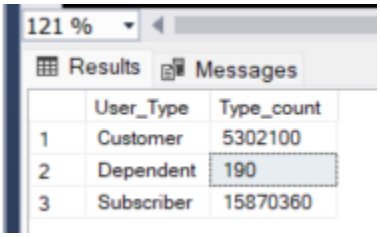



After merging to these 2 new tables, I consider drop the individual tables since we have them all in the backup file.

## Data cleaning and validation

### For table Trips_2013_2019

```
--See Total user type
SELECT User_Type,
       COUNT (User_Type) AS Type_count
FROM Trips_2013_2019
GROUP BY User_Type
```




| | User_Type | Type_count |
|---|---|---|
| 1 | Customer | 5302100 |
| 2 | Dependent | 190 |
| 3 | Subscriber | 15870360 |

We have

- 5302100 customers (Day-pass, short rides)
- 15870360 Subscribers (Annual membership).
- 190 Dependent

Since in this scope of project we only focus on customer and membership riders, I will drop the dependent values.

```
--Drop Dependent Users
DELETE FROM Trips_2013_2019
WHERE User_Type = 'Dependent'
```

Next, I will check for from and to station id nulls for Trips_2013_2019 table. Note that for Customer types, there will be no data for Gender and birth year.

```
--Check nulls for From and To station Id
SELECT *
FROM Trips_2013_2019
WHERE From_Station_Id IS NULL OR To_Station_Id IS NULL;
```

We don't have any nulls in these 2 columns.

With regard to the Birth_Year columns, I consider removing any birthyear that less than 1920 as 100+ years old membership would be likely abnormal entries. However, some other data values are still valuable so I will just let it be for now.

```
--Identify any Birth year that < 1920
SELECT*
FROM Trips_2013_2019
WHERE Birth_Year < 1920
```

We have 5611 records.

## For table Trips_2020_2023

Firstly, I ensure we only have 2 types of customers, member and casual

```
--See Total user type for 2020_2023
SELECT User_Type,
       COUNT (User_Type) AS Type_count
FROM Trips_2020_2023
GROUP BY User_Type
```

We have 12247549 members and 8276791 casuals

Next, I will identify bicycle types

```
--identify Total bicycle types for Trips_2020_2023
SELECT Ride_Type,
       COUNT (Ride_Type) AS _count
FROM Trips_2020_2023
GROUP BY  Ride_Type
```

We have 3 types of bicycles: electric, classic, docked

Next, I will check null values for from station and to station name columns

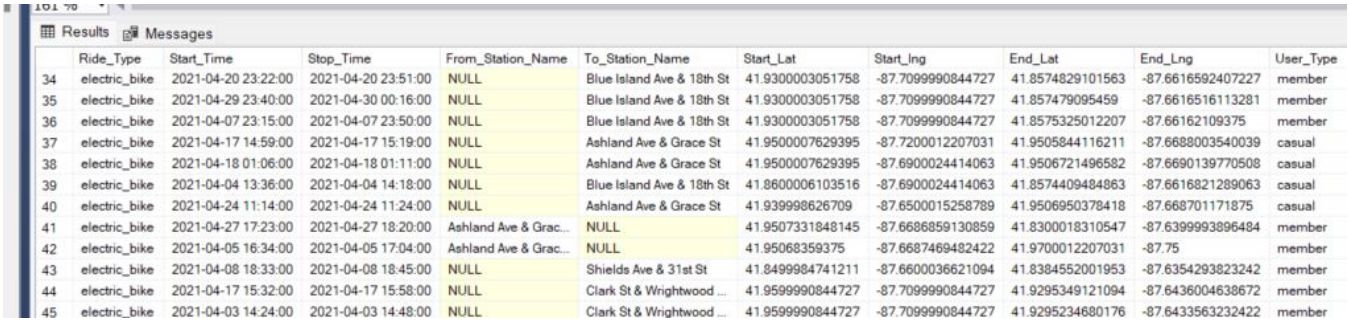
```
--Check nulls for From_Station_Name to To_Station_Name for Trips_2020_2023
SELECT*
FROM Trips_2020_2023
WHERE From_Station_Name is null or To_Station_Name is null
```

We have 3,844, 356 rows that contain nulls either in From_Station_Name or To_Station_Name.

```
--Check nulls for End_Lng for Trips_2020_2023
SELECT*
FROM Trips_2020_2023
WHERE End_Lng is null
```

We have 21885 for Ending Latitude

For our data, given we still have many other valuable data in other columns and the longitude and latitude is adequate, so I decided to leave the nulls value untouched.



Results Messages

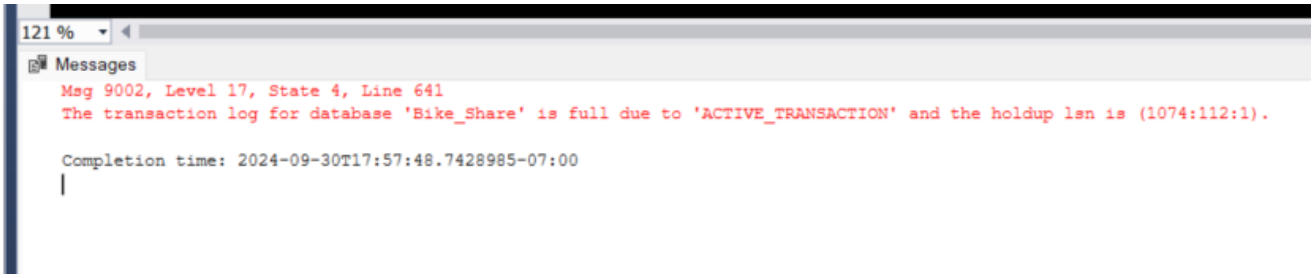
| | Ride_Type | Start_Time | Stop_Time | From_Station_Name | To_Station_Name | Start_Lat | Start_lng | End_Lat | End_Lng | User_Type |
|---|---|---|---|---|---|---|---|---|---|---|
| 34 | electric_bike | 2021-04-20 23:22:00 | 2021-04-20 23:51:00 | NULL | Blue Island Ave & 18th St | 41.9300003051758 | -87.7099990844727 | 41.8574829101563 | -87.6616592407227 | member |
| 35 | electric_bike | 2021-04-29 23:40:00 | 2021-04-30 00:16:00 | NULL | Blue Island Ave & 18th St | 41.9300003051758 | -87.7099990844727 | 41.857479095459 | -87.6616516113281 | member |
| 36 | electric_bike | 2021-04-07 23:15:00 | 2021-04-07 23:50:00 | NULL | Blue Island Ave & 18th St | 41.9300003051758 | -87.7099990844727 | 41.8575325012207 | -87.66162109375 | member |
| 37 | electric_bike | 2021-04-17 14:59:00 | 2021-04-17 15:19:00 | NULL | Ashland Ave & Grace St | 41.9500007629395 | -87.7200012207031 | 41.9505844116211 | -87.6688003540039 | casual |
| 38 | electric_bike | 2021-04-18 01:06:00 | 2021-04-18 01:11:00 | NULL | Ashland Ave & Grace St | 41.9500007629395 | -87.6900024414063 | 41.9506721496582 | -87.6690139770508 | casual |
| 39 | electric_bike | 2021-04-04 13:36:00 | 2021-04-04 14:18:00 | NULL | Blue Island Ave & 18th St | 41.8600006103516 | -87.6900024414063 | 41.8574409484863 | -87.6616821289063 | casual |
| 40 | electric_bike | 2021-04-24 11:14:00 | 2021-04-24 11:24:00 | NULL | Ashland Ave & Grace St | 41.939998626709 | -87.6500015258789 | 41.9506950378418 | -87.668701171875 | casual |
| 41 | electric_bike | 2021-04-27 17:23:00 | 2021-04-27 18:20:00 | Ashland Ave & Grac... | NULL | 41.9507331848145 | -87.6686859130859 | 41.8300018310547 | -87.6399993896484 | member |
| 42 | electric_bike | 2021-04-05 16:34:00 | 2021-04-05 17:04:00 | Ashland Ave & Grac... | NULL | 41.95068359375 | -87.6687469482422 | 41.9700012207031 | -87.75 | member |
| 43 | electric_bike | 2021-04-08 18:33:00 | 2021-04-08 18:45:00 | NULL | Shields Ave & 31st St | 41.8499984741211 | -87.6600036621094 | 41.8384552001953 | -87.6354293823242 | member |
| 44 | electric_bike | 2021-04-17 15:32:00 | 2021-04-17 15:58:00 | NULL | Clark St & Wrightwood ... | 41.9599990844727 | -87.7099990844727 | 41.9295349121094 | -87.6436004638672 | member |
| 45 | electric_bike | 2021-04-03 14:24:00 | 2021-04-03 14:48:00 | NULL | Clark St & Wrightwood ... | 41.9599990844727 | -87.7099990844727 | 41.9295234680176 | -87.6433563232422 | member |

## Data normalization

I noticed for Trips_2013_2019 the customer types are 'Customer' and 'Subscriber" while for Trips_2020_2023, they are 'member' and 'casual'. Therefore, to keep it consistent, I decided to keep **'Customer' and 'Subscriber'.**

```
--Change user types to 'Casual' and 'Subscriber' for Trips_2020_2023
AlTER TABLE Trips_2020_2023
ADD User_Type2 nvarchar(50);

UPDATE Trips_2020_2023
SET User_Type2 =     CASE
                            WHEN User_Type = 'casual' THEN 'Customer'
                            ELSE 'Subscriber'
END;

ALTER TABLE Trips_2020_2023
DROP COLUMN User_Type;
```

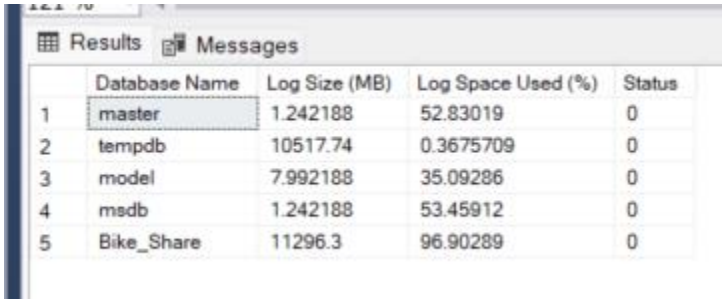
However, when I try to perform the above queries, I encountered the transaction log error.




So, to deal with this, I check the disk log transaction space

```
DBCC SQLPERF(LOGSPACE);
```

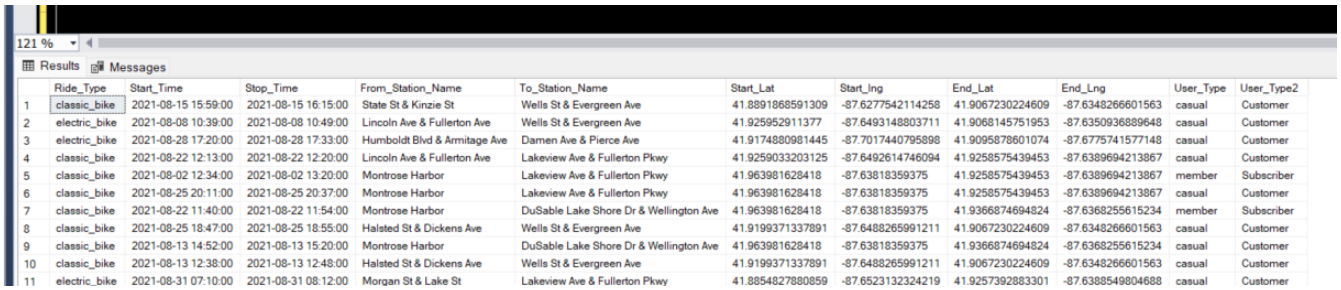
I can see that the Bike_Share database takes nearly 97% of log space. So instead of directly shrink the log space which could slow the queries over time, I decided to use temp table to achieve this query.




| | Database Name | Log Size (MB) | Log Space Used (%) | Status |
|---|---|---|---|---|
| 1 | master | 1.242188 | 52.83019 | 0 |
| 2 | tempdb | 10517.74 | 0.3675709 | 0 |
| 3 | model | 7.992188 | 35.09286 | 0 |
| 4 | msdb | 1.242188 | 53.45912 | 0 |
| 5 | Bike_Share | 11296.3 | 96.90289 | 0 |

```sql
--Try temp table
SELECT *,
    CASE
        WHEN User_Type = 'casual' THEN 'Customer'
        ELSE 'Subscriber'
    END AS User_Type2
INTO #TempTrips_2020_2023
FROM Trips_2020_2023;


--View top 100 of the #tempTrips
SELECT top 100* from #TempTrips
```

It works!!



121 %

Results | Messages

| | Ride_Type | Start_Time | Stop_Time | From_Station_Name | To_Station_Name | Start_Lat | Start_lng | End_Lat | End_Lng | User_Type | User_Type2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | classic_bike | 2021-08-15 15:59:00 | 2021-08-15 16:15:00 | State St & Kinzie St | Wells St & Evergreen Ave | 41.889186859130 9 | -87.6277542114258 | 41.9067230224609 | -87.6348266601563 | casual | Customer |
| 2 | electric_bike | 2021-08-08 10:39:00 | 2021-08-08 10:49:00 | Lincoln Ave & Fullerton Ave | Wells St & Evergreen Ave | 41.925952911377 | -87.6493148803711 | 41.9068145751953 | -87.6350936889648 | casual | Customer |
| 3 | electric_bike | 2021-08-28 17:20:00 | 2021-08-28 17:33:00 | Humboldt Blvd & Armitage Ave | Damen Ave & Pierce Ave | 41.9174880981445 | -87.7017440795898 | 41.9095878601074 | -87.6775741577148 | casual | Customer |
| 4 | classic_bike | 2021-08-22 12:13:00 | 2021-08-22 12:20:00 | Lincoln Ave & Fullerton Ave | Lakeview Ave & Fullerton Pkwy | 41.9259033203125 | -87.6492614746094 | 41.9258575439453 | -87.6389694213867 | casual | Customer |
| 5 | classic_bike | 2021-08-02 12:34:00 | 2021-08-02 13:20:00 | Montrose Harbor | Lakeview Ave & Fullerton Pkwy | 41.963981628418 | -87.63818359375 | 41.9258575439453 | -87.6389694213867 | member | Subscriber |
| 6 | classic_bike | 2021-08-25 20:11:00 | 2021-08-25 20:37:00 | Montrose Harbor | Lakeview Ave & Fullerton Pkwy | 41.963981628418 | -87.63818359375 | 41.9258575439453 | -87.6389694213867 | casual | Customer |
| 7 | classic_bike | 2021-08-22 11:40:00 | 2021-08-22 11:54:00 | Montrose Harbor | DuSable Lake Shore Dr & Wellington Ave | 41.963981628418 | -87.63818359375 | 41.9366874694824 | -87.6368255615234 | member | Subscriber |
| 8 | classic_bike | 2021-08-25 18:47:00 | 2021-08-25 18:55:00 | Halsted St & Dickens Ave | Wells St & Evergreen Ave | 41.9199371337891 | -87.6488265991211 | 41.9067230224609 | -87.6348266601563 | casual | Customer |
| 9 | classic_bike | 2021-08-13 14:52:00 | 2021-08-13 15:20:00 | Montrose Harbor | DuSable Lake Shore Dr & Wellington Ave | 41.963981628418 | -87.63818359375 | 41.9366874694824 | -87.6368255615234 | casual | Customer |
| 10 | classic_bike | 2021-08-13 12:38:00 | 2021-08-13 12:48:00 | Halsted St & Dickens Ave | Wells St & Evergreen Ave | 41.9199371337891 | -87.6488265991211 | 41.9067230224609 | -87.6348266601563 | casual | Customer |
| 11 | electric_bike | 2021-08-31 07:10:00 | 2021-08-31 08:12:00 | Morgan St & Lake St | Lakeview Ave & Fullerton Pkwy | 41.8854827880859 | -87.6523132324219 | 41.9257392883301 | -87.6388549804688 | casual | Customer |

And this is the final step for our data models. In summary, we would have the following tables:

- Station_name (585 rows)
- Trips_2013_2019 (21,172,460 rows)
- Trips_2020_2023 (20,524,340 rows)
- A temp table #TempTrips (a copy of Trips_2020_2023 with normalized User_Type).

**We will continue the Analysis in the Analysis section**