| Date | Tables Affected | Records Affected | Change Type | Description of Change | Reason for Change | Additional Details |
|---|---|---|---|---|---|---|
| 2/25/2022 | All Spreadsheets | 4771 | Removal | Remove null entries for "end_lng" and "end_lat" | Accuracy - data in this column is necessary for this analysis and cannot be blank. | |
| 2/25/2022 | All Spreadsheets | 311705 | Update | Replace all instances of "docked_bike" with "classic_bike" | Consistency - "docked" and "classic" refer to the same style of bike. | |
| 2/25/2022 | All Spreadsheets | 5590292 | Addition | Create column "ride_length" | Perform calculations that will be used to gain further insight into data. | |
| 2/25/2022 | All Spreadsheets | 5590292 | Addition | Create column "day_of_week" | Perform calculations that will be used to gain further insight into data. | Calculated using the WEEKDAY function in Excel. |
| 2/25/2022 | All Spreadsheets | 5590292 | Addition | Create column "distance_traveled" | Perform calculations that will be used to gain further insight into data. | Formula for converting latitude and longitude coordinates to distance in miles=CONVERT((ACOS(COS(RADIANS(90-I2)) *COS(RADIANS(90-K2)) +SIN(RADIANS(90-I2)) *SIN(RADIANS(90-K2)) *COS(RADIANS(J2-L2))) *6371), "km", "mi") |
| 3/2/2022 | All Spreadsheets | 13030 | Update | Update #NUM! errors in "distance_traveled" column to 0.00 | Accuracy - this error message occurs when there is no difference between start and end coordinates. | |
| 3/2/2022 | All Spreadsheets | 5590292 | Update | Round "start_lat", "start_lng", "end_lat", and "end_lng" columns to 6 decimal places. | Consistency - 6 digits is precise enough for the purpose of this analysis. | |
| 3/2/2022 | All Spreadsheets | 146 | Removal | Remove records where "ended_at" was earlier than "started_at" | Accuracy - a glitch occurred when some start and end times were encoded, causing errors. | |
| 3/29/2022 | All Spreadsheets | 5590146 | Update | Convert "day_of_week" column from single digit to text label for each day | Clarity - the names will be more meaningful to users than the digits when performing queries. | Formula for converting Day of Week values to Names in Excel: =IFS(O2=1, "Sunday", O2=2, "Monday", O2=3, "Tuesday", O2=4, "Wednesday", O2=5, "Thursday", O2=6, "Friday", O2=7, "Saturday") |
| 4/3/2022 | SQL - bikes_2021 table | 5590146 | Addition | Create column "season"; update records based on time of year (Spring, Summer, Autumn, Winter) when trip took place. | Categorize data to aid with analysis. | Spring took place between 3/20/2021 and 06/19/2021. Summer took place from 06/20/2021-09/21/2021. Autumn took place from 09/22/2021-12/20/2021. Winter took place both from 01/01/2021-03/19/2021 and 12/21/2021-12/31/2021. Seasons were calculated based on the date in the "started_at" column. |
| 4/3/2022 | SQL - bikes_2021 table | 5590146 | Addition | Create column "trip_type"; update records based on type of trip recorded (Canceled, Out-and-Back, Point-to-Point). | Categorize data to aid with analysis. | "Canceled" trips refer to those with distance_traveled = 0 and ride_length of less than 1 minute. "Out-and-Back" trips refer to those that started and ended at the same station, with distance_traveled = 0 and ride_length greater than 1 minute. "Point-to-Point" trips refer to those that started and ended at different stations and had distance_traveled > 0. |
| 4/6/2022 | SQL - bikes_2021 table | 811 | Removal | Remove records where "ended_at" was greater than 1 day after "started_at" | Accuracy - it is not typical for users to rent a bike for longer than a 24 hour period and this could be indicative of a glitch in how data was encoded. | |
| 4/6/2022 | SQL - bikes_2021 table | 1 | Removal | Removed an outlier which placed the end point of a particular ride ~70 miles from the start point in less than an hour. | Accuracy - this appears to be an error with how this ride data was encoded. | |