| Version | Date | Change Type | Change |
|---|---|---|---|
| 1 | 6/26/2024 | Fixed | Adjusted column names for uniformity across sheets. Same names for columns with same data across sheets. Everything in a consistent format. (start_time, end_time, trip_duration, start_station_id, end_station_id, start_station_name, end_station_name, user_type, birth_year) |
| 1 | 6/26/2024 | New | Created trip_duration column for 2020 Q1 and calculated data. |
| 1 | 6/26/2024 | New | Added day of week column with WEEKDAY function (Sunday=1) |
| 1 | 6/26/2024 | Changed | Changed trip duration from 2019-Q4 from seconds to minutes. |
| 1 | 6/26/2024 | Changed | Changed user_type in 2019_Q4 data to match format from 2020_Q1. |
| 1 | 6/27/2024 | Fixed | Used filter function and conditional formatting to look for null values: No null values to fix in 2020_Q1 data. The only data in 2019_Q4 with significant nulls are the gender and birth_year columns. For now, I am keeping all the data, because the lack of demographic data does not impede the assessment of other factors like trip_duration. I may separate out data with valid gender and birth_year for separate analysis at a later stage. |
| 1 | 6/27/2024 | Fixed | Used filter function to look for errors in trip_duration: Found data with negative and zero values in 2020_Q1 data. Because there was no possibility of inquiring for adjusted/correct start and end times for these entries, this data was removed from the analysis to avoid incorrect conclusions. There were no similar issues with 2019_Q4 data. |
| 1 | 6/27/2024 | Fixed | No indications while filtering of any misspelled words, mistyped numbers, or extra spaces/characters. |
| 1 | 6/27/2024 | Fixed | No duplicate values were found using remove duplicates tool. |
| 1 | 6/27/2024 | Fixed | No mismatched data types found. |
| 1 | 6/27/2024 | Fixed | No indication of truncated data. Data which was deemed incorrect due to having no duration or negative duration was already removed in previous step of cleaning. |
| 1 | 6/27/2024 | Fixed | Overview of data confirms that it is reasonable given my knowledge of the bike share business and relevant to the case study goals. |

| | | | |
|---|---|---|---|
| 1 | 6/27/2024 | New | Both quarters have trip id, day_of_week, start_time, end_time, trip _duration, start_staion_id, start_station_name, end_station_id, end_station_name, and user_type data. I will be creating sheets with this data to facilitate analysis, but keeping the current clean versions of the data for each quarter as well so that analysis can be done on the gender and birth_year data in 2019_Q4 or the longitude/latitude data in 2020-Q1 if deemed relevant. |
| 1 | 6/27/2024 | New | Create Pivot tables (Summary 2020_Q1 and Summary 2019_Q4) for each quarter to help investigate trends. |
| 1 | 6/27/2024 | New | Calculate AVG trip duration by day of week for all, casual, and member. |
| 1 | 6/27/2024 | New | Calculate number of trips by day of week for all, casual, and member. |
| 1 | 6/27/2024 | New | Create new sheet (Analysis Insights) for viewing tables with aggregate data for analysis |
| 1 | 6/27/2024 | New | Create separate sheets of full clean 2019_Q4 data. Name one for gender and filter out blanks; name the other for birthyear and filter out blanks. |
| 2 | 6/28/2024 | New | Create pivot table and generate avg trip_duration and trip count for both user types divided by gender. |
| 2 | 6/28/2024 | New | Add a column for age and convert birthyear column to age for easier analysis. Then divide into age group. |
| 2 | 6/28/2024 | New | Add pivot table and generate AVG trip duration and trip count for both user types divided by age groups. |
| 2 | 6/28/2024 | New | Go over the Analysis Insights sheet with all aggregated data tables from previous steps to determine different trends by user type. |
| 2 | 6/28/2024 | New | Create bar chart graphs for simple visualization of aggregate data, place in Graphs sheet. Use with exploratory analysis. |
| 3 | 6/29/2024 | Fixed | Realized 1-7 integer representation of the day of the week is cumbersome, changing to text weekday in all sheets instead. |
| 4 | 6/29/2024 | Changed | Realized that I needed a column for age group in order to properly set up visualizations. |