Task A3

**Cycle\_stations**

58 Unique IDs but in the Seattle cycle trip there are 61 unique station IDs. These can still be joined though.

Modification\_date (could include who, how or why it was changed) don’t seem to be particularly useful

**Cycle\_trips**

Trip\_ID column has duplicate rows in it – remove duplicate values so that it can be a primary key.

From and t station name is unnecessary as we already have the IDs

Startime and stoptime in cycle trips is not the same data type, length, precision or scale as the date primary key in seattle weather conditions. Change them to 2 date columns and 2 time columns so can be linked to weather\_conditions.

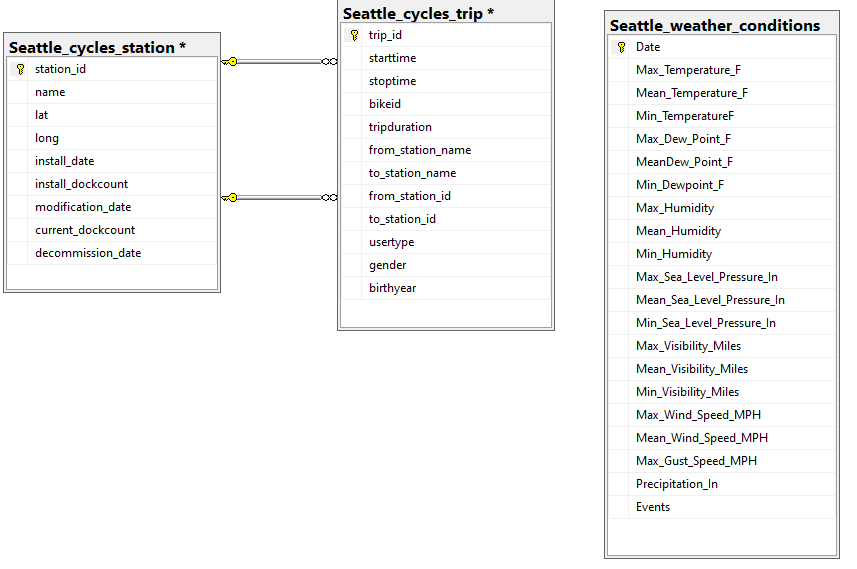
Trip duration may be an unnecessary column because this info is available from the starttime/stoptime columns – this could lead to unreconcilable differences

User IDs would be useful – maybe add to separate table and include dob, home address, work address

Birthyear column might be more useful if it contained the full birthdate

**Weather\_conditions**

Events column is poorly formatted



|  |  |  |  |
| --- | --- | --- | --- |
| Table Name | Column Name | Definition | Sample Values |
| Seattle\_cycles\_station | station\_id | A unique ID for each bike station.  The ID appears to be based on the location but it is not clear exactly how  There are no nulls and no duplicated IDs | BT-01  BT-03 |
| Seattle\_cycles\_station | name | The name of the bike station based on the intersection of 2 roads or a building  There are no nulls or duplicated values | 3rd Ave & Broad St  15th Ave E & E Thomas St  UW Engineering Library / E Stevens Way NE & Jefferson Rd  UW Intramural Activities Building |
| Seattle\_cycles\_station | lat | The latitude of the station location captured as a float | 47.6582870483398  47.614315032959 |
| Seattle\_cycles\_station | long | The longitude of the station location captured as a float | -122.325248718262  -122.322082519531 |
| Seattle\_cycles\_station | install\_date | The date the cycle station was installed in the format YYYY-MM-DD  Most dates are the same (2014-10-13) | 2014-10-13  2016-08-09 |
| Seattle\_cycles\_station | install\_dockcount | Number of bike docks per station at install date as an integer | 16  20  18 |
| Seattle\_cycles\_station | modification\_date | Date the number of docks was changed in the format YYYY-MM-DD  Has a lot of NULL values | 2015-03-13  NULL  2015-03-04 |
| Seattle\_cycles\_station | current\_dockcount | Number of existing docks per station as an integer | 20  0  14 |
| Seattle\_cycles\_station | decommission\_date | If the station was decommissioned then a date will be in this column in the format YYYY-MM-DD  Most values are null except for 4 stations | 2015-10-29  2016-03-18  NULL |

|  |  |  |  |
| --- | --- | --- | --- |
| Table Name | Column Name | Definition | Sample Values |
| Seattle\_cyclehire | trip\_id | Supposed to be a unique ID for each trip as an integer starting from 431  The same ID appears more than once (example ID 431 appears 2 times) | 431  756  1075 |
| Seattle\_cyclehire | starttime |  |  |
| Seattle\_cyclehire | stoptime |  |  |
| Seattle\_cyclehire | bikeid |  |  |
| Seattle\_cyclehire | tripduration |  |  |
| Seattle\_cyclehire | from\_station\_name |  |  |
| Seattle\_cyclehire | to\_station\_name |  |  |
| Seattle\_cyclehire | from\_station\_id |  |  |
| Seattle\_cyclehire | to\_station\_id |  |  |
| Seattle\_cyclehire | usertype |  |  |
| Seattle\_cyclehire | gender |  |  |
| Seattle\_cyclehire | from\_station\_name |  |  |