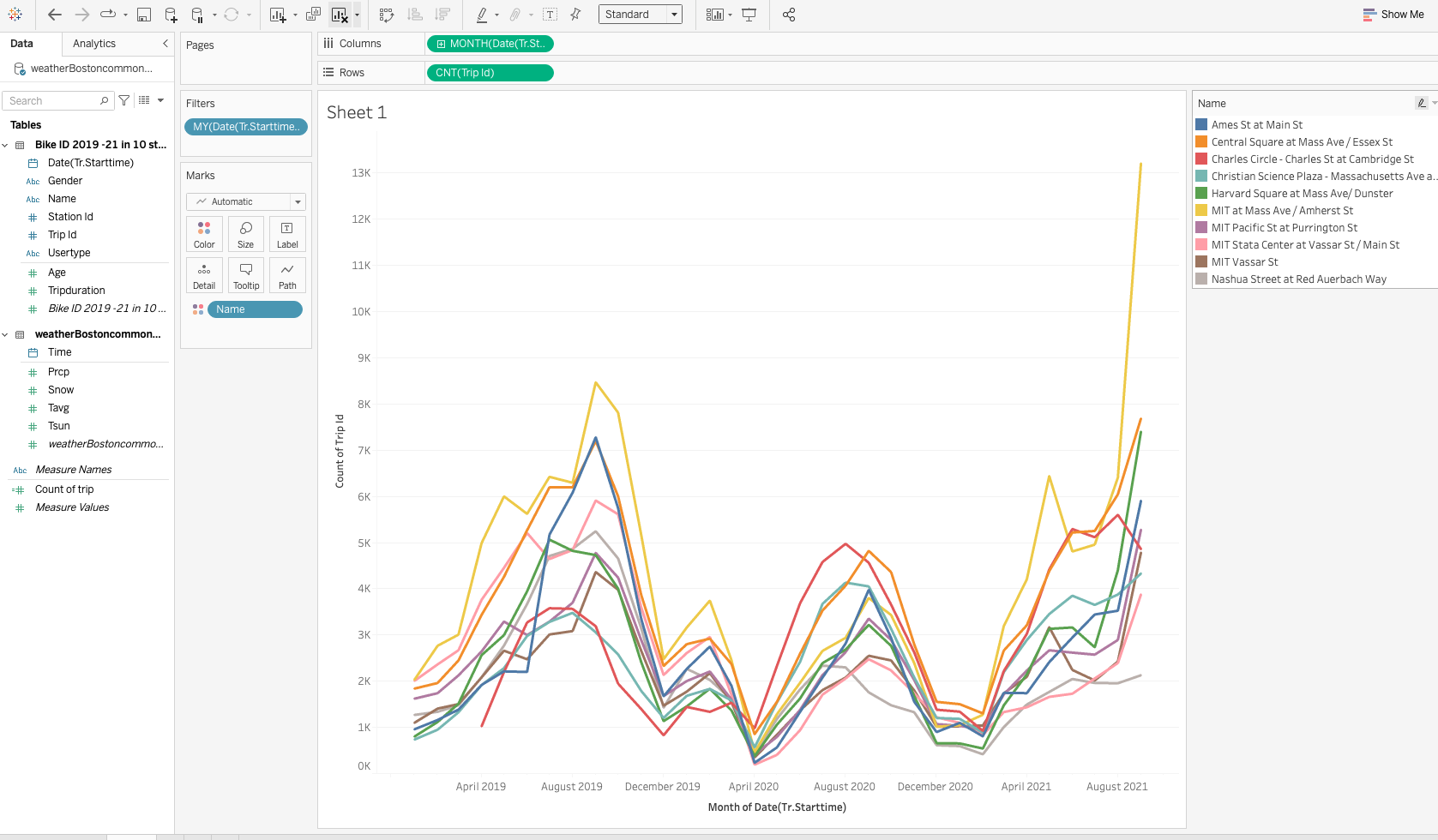
* Business Question: How has seasonality historically affected bike usage trends?
  + Do the most and least popular bike stations change with different weather conditions?
    - SQL
      * SQL Query 1
        + SELECT \*
        + FROM trip tr INNER JOIN station st ON tr.startStationId = st.stationId
      * SELECT st.name, COUNT(st.name), date(tr.starttime)
      * FROM trip tr INNER JOIN station st ON tr.startStationId = st.stationId
      * WHERE YEAR(tr.starttime) >= 2019
      * GROUP BY st.name, date(tr.starttime)
      * ORDER by COUNT(st.name) DESC
    - 25 stations

10 stations

[SELECT](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/select.html) tr.tripId, st.stationId, st.name, 2022-tr.birthyear as age, tr.gender, tr.usertype, [date](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/date-and-time-types.html)(tr.starttime), tr.tripduration FROM trip tr INNER JOIN station st ON tr.startStationId = st.stationId WHERE [year](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/date-and-time-types.html)(tr.starttime) >= 2019 [AND](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_and) (st.stationId = 67 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 68 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 60 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 107 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 80 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 74 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 178 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 46 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 179 [or](https://phpmyadmin.it.umich.edu/url.php?url=https://dev.mysql.com/doc/refman/5.7/en/logical-operators.html%23operator_or) st.stationId = 190) GROUP BY tr.tripID

Close



* + - * Other queries:
        + most and least used stations overall
        + most and least used stations in January, April, August, October (or we can choose specific days with various levels of sun/precipitation in each season)
    - Tableau
      * Tableau 1
        + Map of most and least used stations by season / weather conditions

Concern with this - doesn’t Tableau only go by city level?

* + - * + Bar chart - top 3 most/least popular stations and a bar with demand in each season
        + Stacked bar for each most/least popular station and total annual usage broken down by season
        + Pre-weather conditions: heat map of what the most/least popular stations are (we’d have to combine start/end location)
  + Do different age demographics or gender have different bike usage behavior in different weather conditions?
    - SQL
      * SQL Query 1
    - Tableau
      * Tableau 1
        + Scatterplot with age demographics and average minutes biked each season
        + Side by side bar chart with each gender and average minutes biked per season (or average number of rides per season)
        + Could also break down age cohorts within gender
  + Do subscribers use the bike more frequently in wintery weather compared to pay-as-you-go users?
    - SQL
      * SQL Query 1
        + Pull total rides by subscribers and total rides by pay-as-you-go users (in december thru february? Or on days with snow?)
        + When > X mm of snow or precipitation, bar chart showing usage at a particular station
        + Can also look at minutes ridden for both categories
    - Tableau
      * Tableau 1
        + Side by side bar chart comparing total rides taken in winter months
        + Side by side bar chart comparing average minutes ridden in winter months
        + Donut chart of total rides taken in winter months broken down by subscribers vs pay as you go users
  + How did COVID cases in Boston affect Blue Bike ridership? Were there changes before/after vaccines were available? How did those changes affect typical seasonality trends?
    - SQL
      * SQL Query 1
    - Tableau
      * Tableau 1
        + Bar chart showing total Jan vs March vs May 2020 rides
        + If we have vaccine data, stacked bar chart showing vaccines administered vs rides taken or minutes
  + Is there a “low performing” share of docking stations that have consistently underperformed across different weather conditions and COVID that the company should consider eliminating?
    - SQL
      * Lowest used stations for start and end points
      * Number of rides across each station
    - Tableau
      * Distribution of rides among all stations
      * Percentage of stations that receive <10% of all rides start/end
* Boston Common Coordinates: 42.356683349699290, -71.061666458845140
* Auburndale Station Coordinates: 42.348055414933460, -71.247759461402900
* API Documentation: <https://github.com/meteostat/meteostat-python>
* Formats&Units: <https://dev.meteostat.net/formats.html#time-format>
* [**Covid Data API**](https://apidocs.covidactnow.org/#historic-data-for-all-states-counties-or-metros)**:** 
  + Key: 51d786a7cdb14bb88acd98fe7148ee1e
  + CBSA code for Boston: 14460
  + [https://api.covidactnow.org/v2/cbsa/{cbsa\_code}.timeseries.json?apiKey=YOUR\_KEY\_HERE](https://api.covidactnow.org/v2/cbsa/%7Bcbsa_code%7D.timeseries.json?apiKey=YOUR_KEY_HERE)
* [Tableau COVID cases](https://drive.google.com/file/d/1MrMETMAdZpoYYr6PMOAjqUMAcpAP44HM/view?usp=sharing)

**SQL**

Most and least used stations

**Plan**

Pre-Lab on Wednesday:

* Think through SQL queries, note on google doc under sub business questions
* Think through Tableau visualizations

Lab on Wednesday:

* SQL queries
* Insert everything into Tableau

Appendix

* Variation of weather information in different part of Boston

**Weather data in Boston Common**

| time | tsun | snow | prcp | tavg |
| --- | --- | --- | --- | --- |
| 1/1/20 | 980 | 0 | 0 | 3.7 |
| 1/2/20 | 269 | 0 | 0 | 3.8 |
| 1/3/20 | 0 | 0 | 0 | 7.8 |
| 1/4/20 | 0 | 0 | 3.3 | 7.7 |
| 1/5/20 | 116 | 0 | 2.8 | 3.5 |
| 1/6/20 | 133 | 0 | 0 | 0 |
| 1/7/20 | 161 | 0 | 0 | 2.8 |
| 1/8/20 | 350 | 30 | 0.5 | 3 |
| 1/9/20 | 540 | 0 | 0 | -2 |
| 1/10/20 | 106 | 0 | 0 | 2.8 |

**Weather data in Auburndale Station**

| time | tsun | snow | prcp | tavg |
| --- | --- | --- | --- | --- |
| 1/1/20 | 980 | 0 | 0 | 1.9 |
| 1/2/20 | 269 | 0 | 0 | 2.8 |
| 1/3/20 | 0 | 0 | 0 | 6.1 |
| 1/4/20 | 0 | 0 | 2.3 | 4.3 |
| 1/5/20 | 116 | 0 | 2.3 | 1.1 |
| 1/6/20 | 133 | 0 | 0 | -1.9 |
| 1/7/20 | 161 | 0 | 0 | 1.1 |
| 1/8/20 | 350 | 30 | 0 | 0.7 |
| 1/9/20 | 540 | 0 | 0 | -4.6 |
| 1/10/20 | 106 | 0 | 0 | 3.6 |