

Description of the Question

Our overarching business question is to understand how seasonality has historically affected Blue Bike usage trends in Boston, Massachusetts. In order to do so, we created the following sub-questions:

- How does weather affect general ridership trends?
- Does seasonality affect the breakdown of usage between subscribers and one-off riders?
- Did COVID equally affect ridership of subscribers and one-off users?
- How did COVID vaccination affect Blue Bike ridership?

Description of Data Sources

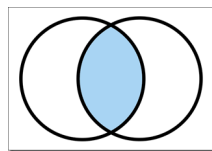
We used the following data sources to explore our business question and subquestions:

- **Boston Blue Bike Data (Provided in SQL / Tableau):** We used Boston Blue Bike historic data to access from/to location, date/time of ride, number of rides taken, and user age demographics. We focused on the data provided from 2019-2021. For from/to location, we focused on the top 10 stations used; these stations have the most data to analyze trends across years and seasons.
- **Weather Data (Meteostat API):** We used this historical weather data set to pull high and low temperatures, precipitation, and minutes of sunshine for Boston.
- **COVID-19 Data Hub (Tableau Dashboard):** We pulled COVID case rates in Suffolk County (where Boston is located) to evaluate changes in bike ridership amidst COVID seasonality.

How Data Sources Were Combined

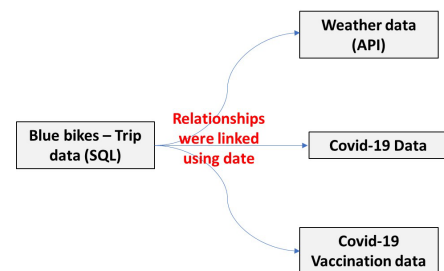
The top 10 bike stations based on the number of trip counts were selected for our business break down analysis. The Station ID of the Station table is linked with the Start station ID of the Trip table using the INNER JOIN function in SQL. Adding to the INNER JOIN, Conditional statements were used to filter out the Top 10 stations and for the Years 2019-2021. Then the raw data with the station IDs, station names, trip duration, age, gender, user types and year were extracted for our tableau analysis.

Blue bikes Station table (station ID)



Blue bikes Trip table (Start station ID)

The weather data was extracted from the meteostat by using Python's Advanced Programming Interface. We found the center coordinates of Boston and extracted the weather data using the respective latitude and longitude positions. The attributes like min temperature, max temperature, average temperature, snow fall (mm), precipitation (mm), and sunshine (minutes) data were considered for our business analysis.



Boston Covid cases and vaccine data were extracted from the [Github - Covid-19 US data](#) and [US vaccine timeline](#). The relationships between the three data sources were linked to the base bike data using dates.

Results

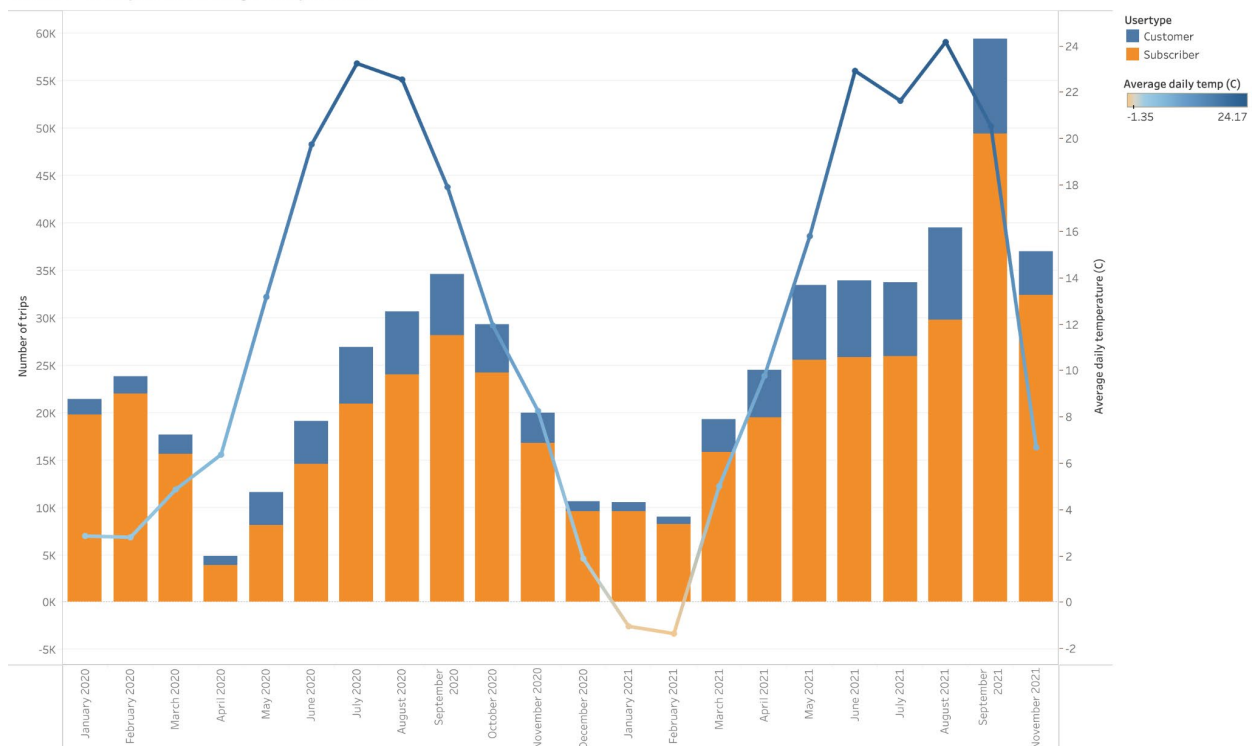
Our overarching question has a clear answer; seasonality strongly influences bike usage trends. We hypothesized it would; seasons of weather, COVID cases, and vaccination data all contribute to bike riders' decisions to hit the road. In terms of our sub-questions, here is what we found:

How does weather affect general ridership trends?

While we hypothesized that temperature would certainly impact bike rentals, we anticipated that the earlier warmer months, such as May and June, would represent the peak of bicycle rentals as Bostonians and tourists celebrated the end of wintry weather and spent more time outside. However, we noted that the number of trips consistently peaks in September, likely reflecting the cooler end of summer temperatures and the return to school for university students. Therefore, Blue Bikes management needs to prepare for heaviest bike usage and repair needs in August and September.

Exhibit I in the Appendix shows ridership and temperature along with day of the week. That visualization reflects greater ridership on weekdays, suggesting that riders may primarily use bikes for commuting to school or work; this relationship will need to be explored through survey data of bike users.

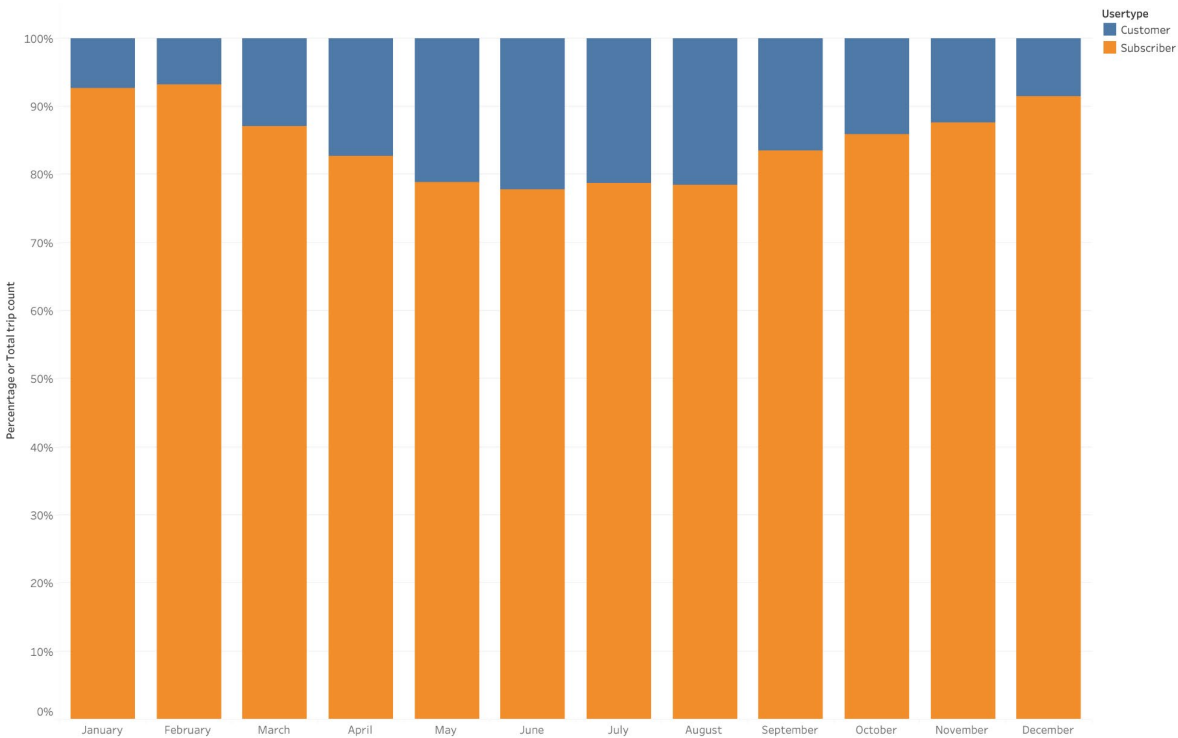
Number of trips and Average temperature



Does seasonality affect the breakdown of usage between subscribers and one-off riders?

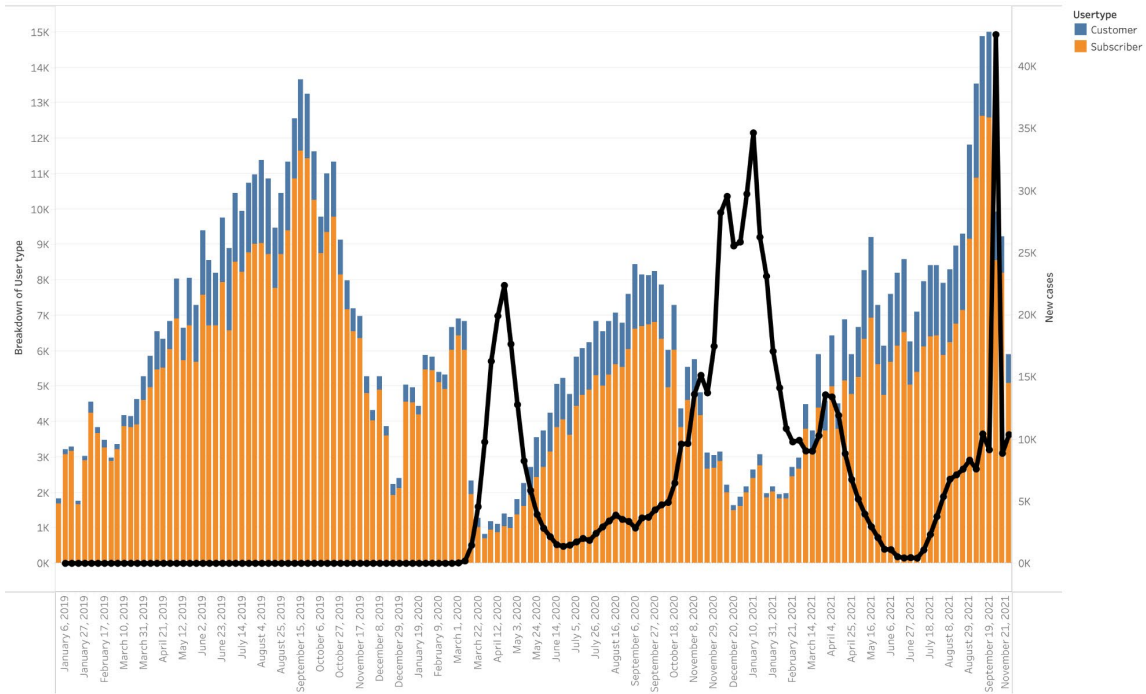
We found that the distribution of subscribers and one-off riders does show seasonal trends. In colder, wintry months, one-off riders represent a lower share of total Blue Bike ridership—in January and February, one-off riders represent less than 10 percent of total trips taken. In warmer months, particularly May through August, one-off riders increase their share of total trips taken to ~20 percent each month. We suspect much of this increase is driven by summertime tourists, but it may also reflect Bostonians seeking to spend more time outside in the summer. Additional survey data would help illuminate this shift in one-off rider demand, but the trend does suggest that Blue Bikes should consider running marketing campaigns targeting one-off riders during the warmer months to further capitalize on this trend.

Breakdown of customer type by month

*Did COVID equally affect ridership of subscribers and one-off users?*

No; we saw a significant increase in the share of one-off riders in 2021, especially as new COVID cases dipped to the lowest rates of the pandemic thus far. We predict these were new riders seeking to celebrate their newly-established freedom from staying at home. Another hypothesis is that this influx of one-off riders was driven by former subscribers who had previously used Blue Bikes to commute to school or work and let their subscriptions lapse in the pandemic. Now that schools and workplaces are starting to reopen, perhaps former subscribers are not regularly traveling, but taking more one-off trips as needed. Blue Bikes should evaluate whether this is true; if so, the company should engage former subscribers and offer a discounted renewal rate to attract them back into the fold long-term.

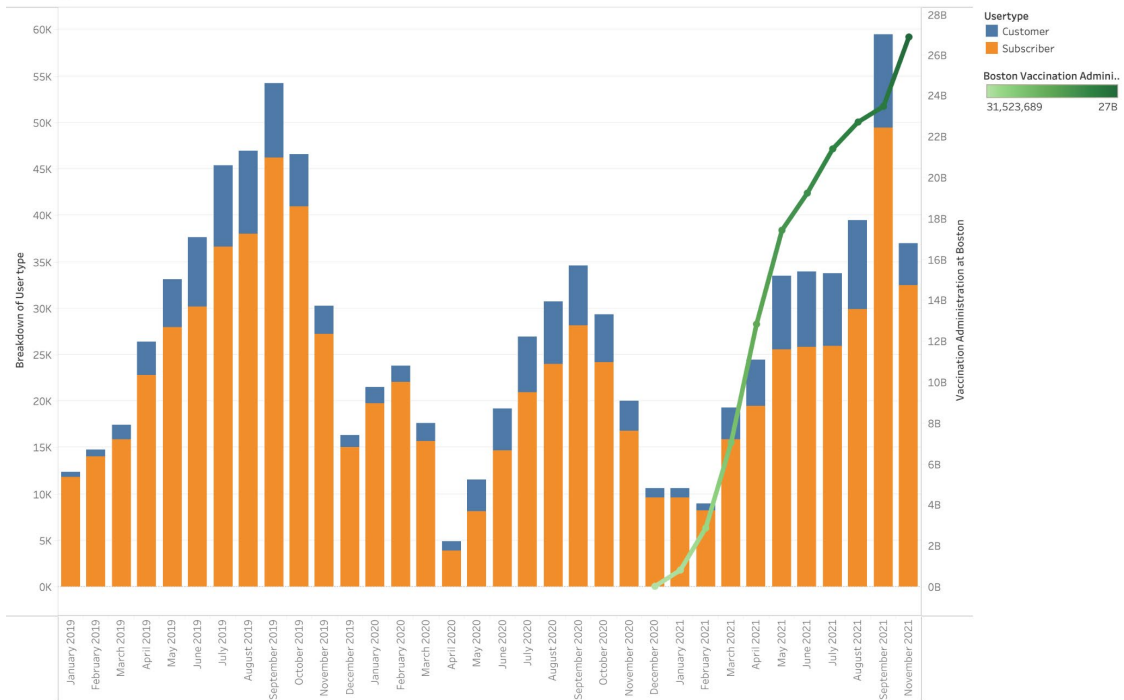
User types and COVID-19 new cases



How did COVID vaccination affect Blue Bike ridership?

COVID vaccination rates rose as temperatures rose, so it is difficult to determine how much each factor drove the increase in rides. However, after rides diminished in 2020 as COVID proliferated, the rise of vaccines in 2021 corresponded with the greatest number of rides taken in three years in September 2021.

User types and Vaccination cases



Appendix

1. Weather Data (Meteostat API)

We selected Boston Common area to scrape weather data (Coordinates: 42.356683349699290, -71.061666458845140)

```
import meteostat #import weather data
import pandas as pd

from datetime import datetime
s = datetime(2020,1,1) #start date
e = datetime(2021,12,31) #end date
loc = meteostat.Point(42.356683349699290, -71.061666458845140, 75) #location coordinates
data = meteostat.Daily(loc, s, e) #pull weather data from loc (detroit) between start and end date
data = data.fetch() #update the data
measures = ["tsun", "snow", "prcp", "tavg"] #defining measure variable
table = data[measures] #grab min and max temperature from weather api

df = pd.DataFrame(table)
df.to_csv("weatherBostoncommon20-21.csv")
```

2. Boston Blue Bike Data (Provided in Tableau)

2.1 Ranking Top 10 stations where users use most after 2019

```
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
```

2.2 Selecting data from only Top 10 stations

```
SELECT tr.tripId, st.stationId, st.name, 2022-tr.birthyear as age, tr.gender, tr.usertype, date(tr.starttime),
tr.tripduration
FROM trip tr INNER JOIN station st ON tr.startStationId = st.stationId
WHERE year(tr.starttime) >= 2019 AND (st.stationId = 67 or st.stationId = 68 or st.stationId = 60 or
st.stationId = 107 or st.stationId = 80 or st.stationId = 74 or st.stationId = 178 or st.stationId = 46 or
st.stationId = 179 or st.stationId = 190)
GROUP BY tr.tripID
```

2.3 Categorize generation by age By creating new calculated field

Exhibit II

Number of trips by month compared to generation



Exhibit III

User types and COVID-19 Confirmed cases

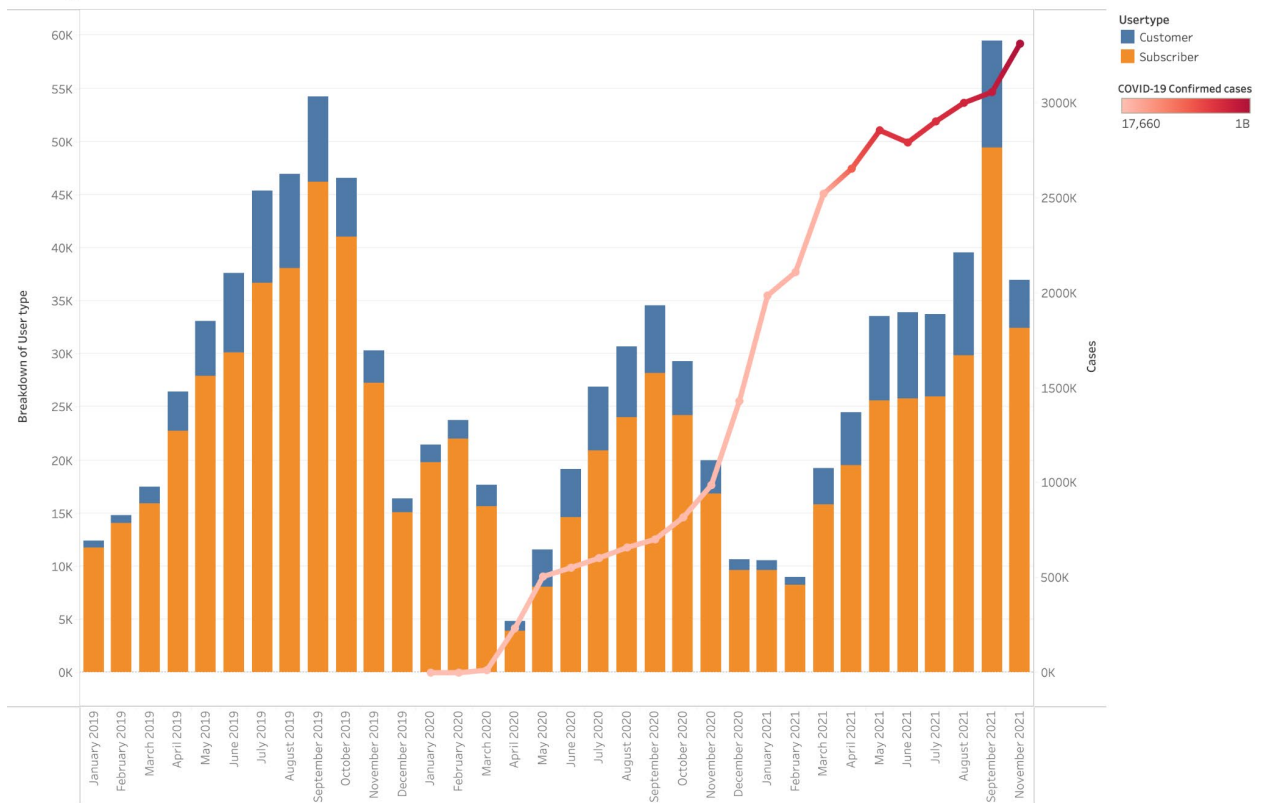


Exhibit IV

Breakdown cumulative and daily new COVID-10 cases

