Bay Wheels User Analysis

– Data Set **Descriptions**

The **lyft.baywheels** dataset reports information about rentals made on theBay Wheels bike share system. Each row represents a single rental; we will be making use of the following fields in this project:

- **started_date –** Date for start of rental
- **started_at –** Timestamp for start of rental
- **ended_at –** Timestamp for end of rental
- **start_station_name –** For rentals that started from a bike dock,the name of the dock.
- **end_station_name –** For rentals that ended at a bike dock,the name of the dock.
- **start_lat, start_lng –** Latitude and longitude, respectively, of the start of the rental.
- **end_lat, end_lng –** Latitude and longitude, respectively, of the end of the rental.
- **member_casual –** String indicating whether the rental was made by a system "member", who has a monthly subscription with the bikeshare system, or by a "casual" user, who is making a one-time rental.

The **ford.gobike** dataset has information very similar to the **lyft.baywheels** table, but reports rides prior to Lyft's takeover of the bikeshare system. One major distinction between the two tables is different field names. The field names in the ford.gobike dataset will be explained through the course of the project tasks.

- **date –** Date of weather recordings
- **temperature_avg –** Average temperature in Fahrenheit
- **precipitation –** Recorded precipitation in inches

– **Task 1:** Top User Engagement

Below is a table of equivalent columns between the two datasets, detailing which columns in the lyft.baywheels data set match which columns in the ford.gobike data table.

Lyft Bay Wheels	**Ford GoBike**
started_date	start-date
started_at	start_time
ended_at	end_time
start_station_name	start_station_name
end_station_name	end_station_name
start_lat	start_station_latitude
start_lng	start_station_longitude
end_lat	end_station_latitude
end_lng	end_station_longitude
member_casual	user_type

The first step is to write a query that filters the **`ford.gobike`** data to include only records from the year 2020. This will allow me to focus the analysis specifically on data from that year.

```
SELECT
  *
FROM
  ford.gobike
WHERE
  date_part('year', start_date) = 2020
```

The next step is to write a query that unions the **`ford.gobike`** dataset with the **`lyft.baywheels`** dataset, ensuring that the column names from the Lyft data are used as the standard for the analysis. I'll be filtering the Ford data to include only records from the year 2020. Another good idea would be to include a new column called **`data_source`** that indicates whether each row comes from the Lyft or Ford dataset.

```
SELECT
  started_at,
  ended_at,
  start_station_name,
  end_station_name,
  start_lat,
  start_lng,
  end_lat,
  end_lng,
  member_casual,
  'Lyft' AS data_source
FROM
  lyft.baywheels
UNION
SELECT
  start_time,
  end_time,
  start_station_name,
  end_station_name,
  start_station_latitude,
  start_station_longitude,
  end_station_latitude,
  end_station_longitude,
  user_type,
  'Ford' AS data_source
FROM
  ford.gobike
WHERE
  date_part('year', start_date) = 2020
```

started_at	ended_at	start_station_name	end_station_name	start_lat	start_lng	end_lat	end_lng	member_casual	data_source
00:00:00	00:06:06	Downtown Berkeley BART	Parker St at Fulton St	37.870139	-122.268422	37.8624644	-122.2647910	member	Lyft
00:00:00	00:08:14	null	null	37.79	-122.42	37.79	-122.44	casual	Lyft
00:00:00	00:08:37	null	24th St at Chattanooga St	37.7667976667	-122.4165145	37.7518194	-122.4266139	Subscriber	Ford
00:00:00	00:10:10	null	null	37.79	-122.42	37.81	-122.43	casual	Lyft
00:00:00	00:11:45	Powell St BART Station (Market St at 5th St)	11th St at Natoma St	37.7838993571	-122.4084448	37.7735069	-122.4160402	Customer	Ford
00:00:01	00:02:35	null	null	37.7918	-122.4106	37.7918	-122.4106	member	Lyft
00:00:01	00:06:55	Divisadero St at Clay St	Turk St at Fillmore St	37.7895771666	-122.4407014	37.7805013333	-122.4320023	casual	Lyft
00:00:01	00:10:52	Market St at Dolores St	null	37.769264	-122.4268961	37.78	-122.44	member	Lyft
00:00:01	00:18:57	null	null	37.7513796667	-122.4340651	37.7834913333	-122.409397	Subscriber	Ford
00:00:02	00:02:39	null	null	37.3248	-121.8669	37.3291	-121.8715	casual	Lyft
00:00:02	00:07:36	null	null	37.76	-122.43	37.76	-122.44	casual	Lyft
00:00:02	00:13:09	Jackson Playground	Broadway at Battery St	37.7650276666	-122.3990861	37.7986538333	-122.4007551	casual	Lyft
00:00:02	00:18:07	Terry Francois Blvd at Warriors Way	null	37.7690948871	-122.3863327	37.7943423333	-122.3950636	Subscriber	Ford

— **Task 2:** Preparing the Data and Creating New Features

To create the **`member_type`** column based on the **`member_casual`** values and ensure they match the Lyft classifications, I will write a query that maps the Ford data classifications to the corresponding Lyft terms. The following query will add a new column **`member_type`** that standardizes the values to either **`member`** or **`casual`** based on the original **`member_casual`** field. Rows where **`member_casual`** is 'Subscriber' will be classified as 'member', and rows where it is 'Customer' will be classified as 'casual'. All other values will remain unchanged.

```
SELECT
  *,
  case
    when member_casual = 'Customer' then 'casual'
    when member_casual = 'Subscriber' then 'member'
    else member_casual
  end as member_type
FROM
  project.ford_lyft_analysis
```

started_date	started_at	ended_at	start_station_name	end_station_name	start_lat	start_lng	end_lat	end_lng	member_casual	data_source	member_type
2020-01-01	00:05:19	00:09:54	null	null	37.76552233	-122.4197163	37.766134	-122.429531	Subscriber	Ford	member
2020-01-01	00:06:30	00:27:07	Jones St at Post St	Valencia St at 22nd St	37.78732676	-122.4132782	37.7552126	-122.4209752	Subscriber	Ford	member
2020-01-01	00:08:36	00:10:48	null	null	37.76522716	-122.3965486	37.764923	-122.4009416	Subscriber	Ford	member
2020-01-01	00:09:23	00:14:57	null	Hyde St at Post St	37.77667783	-122.4157261	37.78734902	-122.4166510	Subscriber	Ford	member
2020-01-01	00:10:02	00:27:58	The Embarcadero at Steuart St	San Francisco Caltrain Station (King St at 4	37.79146400	-122.391034	37.77632881	-122.3944383	Customer	Ford	casual
2020-01-01	00:10:02	00:27:59	The Embarcadero at Steuart St	San Francisco Caltrain Station (King St at 4	37.79146400	-122.391034	37.77632881	-122.3944383	Customer	Ford	casual
2020-01-01	00:10:38	00:30:36	Steuart St at Market St	null	37.79410633	-122.3943653	37.79080683	-122.4175	Subscriber	Ford	member
2020-01-01	00:10:41	00:27:01	null	null	37.78916083	-122.4247645	37.75268483	-122.4112281	Customer	Ford	casual
2020-01-01	00:11:07	00:22:59	49th St at Telegraph Ave	Telegraph Ave at Carleton St	37.8359455	-122.2623663	37.8623199	-122.258801	Customer	Ford	casual
2020-01-01	00:12:04	00:20:30	14th St at Mission St	Folsom St at 9th St	37.7682646	-122.4201102	37.7737172	-122.4116467	Customer	Ford	casual
2020-01-01	00:12:11	00:20:28	14th St at Mission St	Folsom St at 9th St	37.7682646	-122.4201102	37.7737172	-122.4116467	Customer	Ford	casual
2020-01-01	00:13:51	00:19:11	Bancroft Way at Telegraph Ave	Haste St at Telegraph Ave	37.8688126	-122.258764	37.86604310	-122.2588044	Customer	Ford	casual
2020-01-01	00:14:09	00:45:48	Market St at Franklin St	16th St at Prosper St	37.77379320	-122.4212390	37.764285	-122.4318042	Customer	Ford	casual

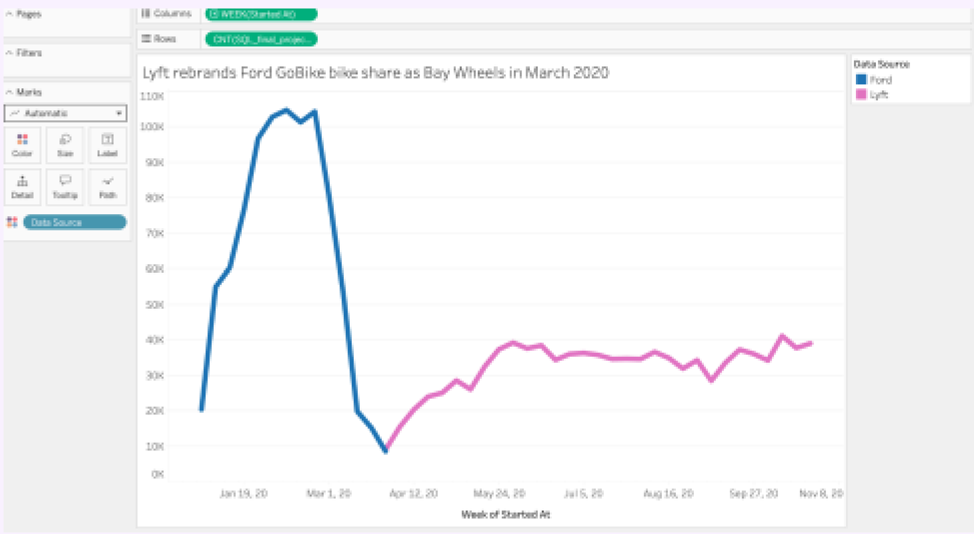
To incorporate the weather data into the analysis, I will modify the query to join the **`sf.weather`** table with the **`project.ford_lyft_analysis`** table based on the **`started_date`** field. The query will also return the average daily temperature and the amount of precipitation. This will allow us to analyze how weather patterns may influence user behavior.

```
SELECT
  date,
  temperature_avg,
  precipitation,
  fla.*,
  case
    when member_casual = 'Customer' then 'casual'
    when member_casual = 'Subscriber' then 'member'
    else member_casual
  end as member_type
FROM
  project.ford_lyft_analysis AS fla
  INNER JOIN sf.weather ON sf.weather.date = fla.started_date
```

date	temperature_avg	precipitation	started_date	started_at	ended_at	start_station_name	end_station_name	start_lat	start_lng	end_lat	end_lng	member_casual	data_source	member_type
2020-01-01	53	0	2020-01-01	00:05:19	00:09:54	null	null	37.76552233	-122.4197163	37.766134	-122.429531	Subscriber	Ford	member
2020-01-01	53	0	2020-01-01	00:06:30	00:27:07	Jones St at Post St	Valencia St at 22nd St	37.78732676	-122.4132782	37.7552126	-122.4209752	Subscriber	Ford	member
2020-01-01	53	0	2020-01-01	00:08:36	00:10:48	null	null	37.76522716	-122.3965486	37.764923	-122.4009416	Subscriber	Ford	member
2020-01-01	53	0	2020-01-01	00:09:23	00:14:57	null	Hyde St at Post St	37.77667783	-122.4157261	37.78734902	-122.4166510	Subscriber	Ford	member
2020-01-01	53	0	2020-01-01	00:10:02	00:27:58	The Embarcadero at Steuart St	San Francisco Caltrain Station (King St at 4	37.79146400	-122.391034	37.77632881	-122.3944383	Customer	Ford	casual
2020-01-01	53	0	2020-01-01	00:10:02	00:27:59	The Embarcadero at Steuart St	San Francisco Caltrain Station (King St at 4	37.79146400	-122.391034	37.77632881	-122.3944383	Customer	Ford	casual
2020-01-01	53	0	2020-01-01	00:10:38	00:30:36	Steuart St at Market St	null	37.79410633	-122.3943653	37.79080683	-122.4175	Subscriber	Ford	member
2020-01-01	53	0	2020-01-01	00:10:41	00:27:01	null	null	37.78916083	-122.4247645	37.75268483	-122.4112281	Customer	Ford	casual
2020-01-01	53	0	2020-01-01	00:11:07	00:22:59	49th St at Telegraph Ave	Telegraph Ave at Carleton St	37.8359455	-122.2623663	37.8623199	-122.258801	Customer	Ford	casual
2020-01-01	53	0	2020-01-01	00:12:04	00:20:30	14th St at Mission St	Folsom St at 9th St	37.7682646	-122.4201102	37.7737172	-122.4116467	Customer	Ford	casual
2020-01-01	53	0	2020-01-01	00:12:11	00:20:28	14th St at Mission St	Folsom St at 9th St	37.7682646	-122.4201102	37.7737172	-122.4116467	Customer	Ford	casual
2020-01-01	53	0	2020-01-01	00:13:51	00:19:11	Bancroft Way at Telegraph Ave	Haste St at Telegraph Ave	37.8688126	-122.258764	37.86604310	-122.2588044	Customer	Ford	casual
2020-01-01	53	0	2020-01-01	00:14:09	00:45:48	Market St at Franklin St	16th St at Prosper St	37.77379320	-122.4212390	37.764285	-122.4318042	Customer	Ford	casual

Task 3: Visualizing and Analyzing Using Tableau

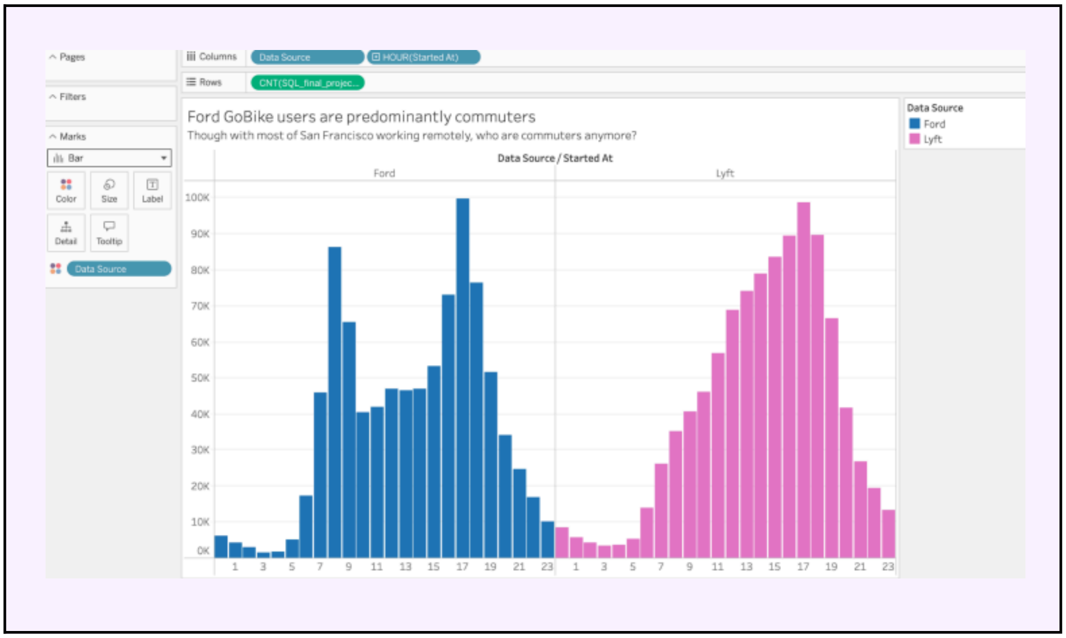
The next step is to start exploring the data by plotting the number of rentals made each week, using the `Started Date` field to determine each rental's week. I'll also add color to the chart to clearly distinguish between the data sources, highlighting when the transition from Ford to Lyft occurred. This visualization will help identify trends over time and any changes in rental patterns associated with the switch in data sources.



> The transition happened at the end of March 2020. There is a huge difference in the volume of rentals, but this is likely caused by the shutdown in San Francisco that occurred at the same time due to the Coronavirus pandemic and not the transition from Ford GoBike to LyftBaywheels.

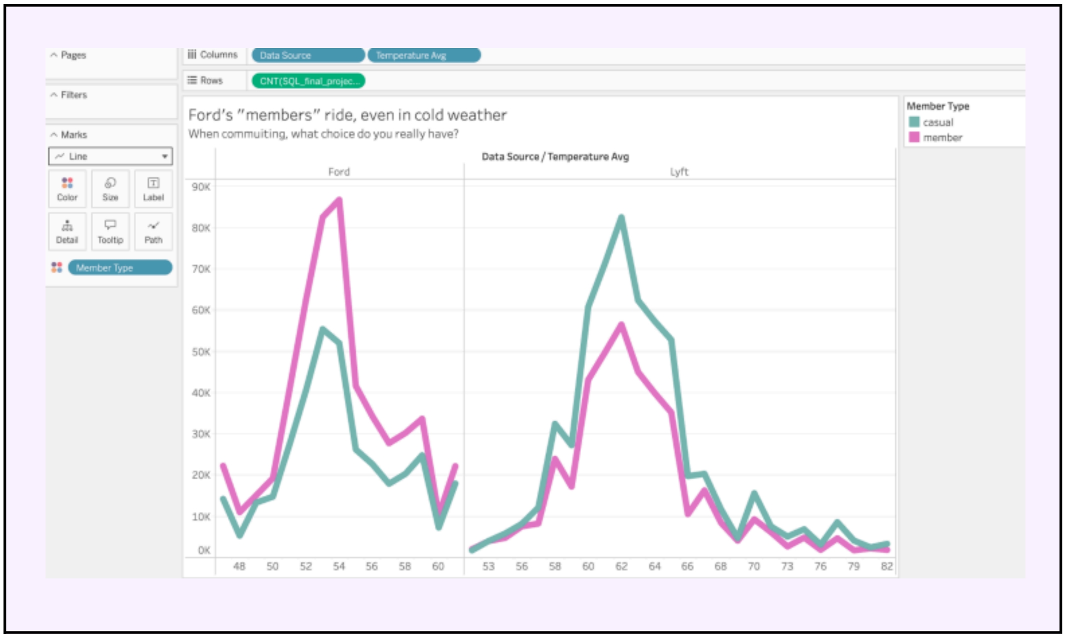
The next step is to create a bar chart that depicts the total number of rides during each hour of the day. To break down the hourly usage patterns by data source, I will create two side-by-side bar charts: one illustrating the total number of rides for

Ford GoBike data, and the other for Lyft Baywheels data. This will allow for a clear comparison of hourly ride patterns between the two data sources.



Ford's GoBike product appears to be used mostly by commuters, with heavy traffic around morning and evening rush hours and a lull in usage during the middle of the day. Lyft seems to see a steady increase in riders from mid-morning on. It's worth remembering that due to the Coronavirus shut down, most San Franciscans did not have a morning commute in 2020 and the late-afternoon-to-early-evening spike in traffic might be due to post work-day exercise or errands.

The next step is to create a line plot with the average temperature on the horizontal axis and the number of rides taken on the vertical axis. I will plot one line for each `Member Type` to distinguish between members and casual users. I'll also add `Data Source` to the columns to compare Ford ridership with Lyft ridership.



Member riders are more likely to ride in the colder days (likely due to using it for commuting to work) while casual riders use the opportunity to ride when the weather is nicer.

– **Task 4:** Communicating Results

It's clear from both visualizations that Ford's loyal customers were commuters. The largest number of riders are riding bikes between 7a - 10a (a morning commute to the office) and again between 4p - 7p (an evening commute home). Lyft's current Baywheels users generally ride throughout the day, with stillthe busiesttime in the evening (4p - 7p). While it may be the case that many San Franciscans are not commuting to work, due to the pandemic, it would still be worthwhile for Lyft to build off Ford's successful customer base with a membership model that perhaps offered benefits to commuters. Lyft might offer a membership discount for commuters, or even a free “Lyft” home for members on a particularly cold or rainy day.