

Lyft – Bay Wheels

Business Analysis

GIJS VAN PAMELEN for UDACITY

Introduction



The business

In select locations, LYFT offers bikes that can be rented by users to travel around certain areas of a city¹.

There are two ways to use the bikes²:

- Subscriber (member): \$15p/m or \$149p/y for unlimited trips up to 45min trip duration.
- Customer (one-off): \$2 per trip up to 30min
- Trips over 45/30 mins are \$3 per extra 15 mins

The data³

LYFT makes anonymized data on trips taken in the Bay Area publicly available for analytics purposes. For this analysis, we've limited ourselves to the data from Jan-2020.

This data consists of basic trip data, such as start location & start time, end-location and end-time as well as trip metadata, most notable user-type (member or not) and rental method (LYFT app or Clipper public transport card).

1. <https://www.lyft.com/bikes/>

2. <https://www.lyft.com/bikes/bay-wheels/pricing>

3. <https://www.lyft.com/bikes/bay-wheels/system-data>

Objective & Outline

To assess LYFT's Bay Wheels business, we develop a high-level assessment of revenue as main impact metric and investigate its relationship with available factors.

Revenue, depending on¹:

- Number of trips
- Trip duration

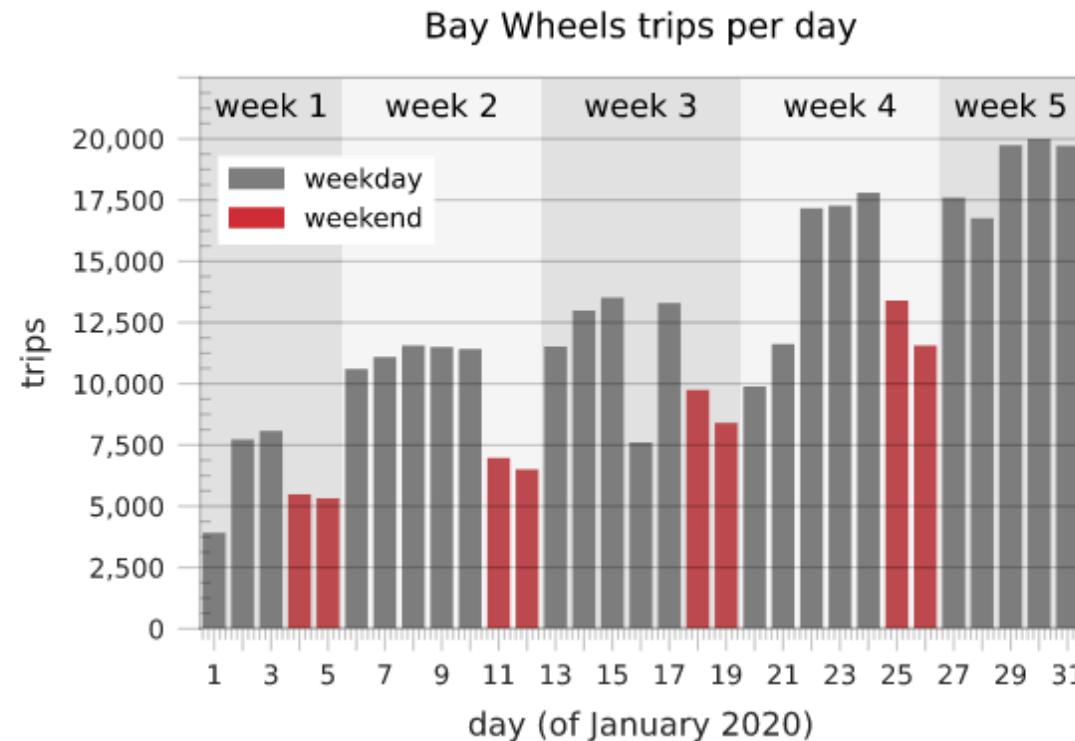
Underlying factors (metadata):

- Customer type ('Customer' or 'Subscriber')
- Rental access method ('App', 'Clipper card' or 'Undefined')
- Location ('San Francisco', 'Berkeley' or 'San Jose')
- Distance covered

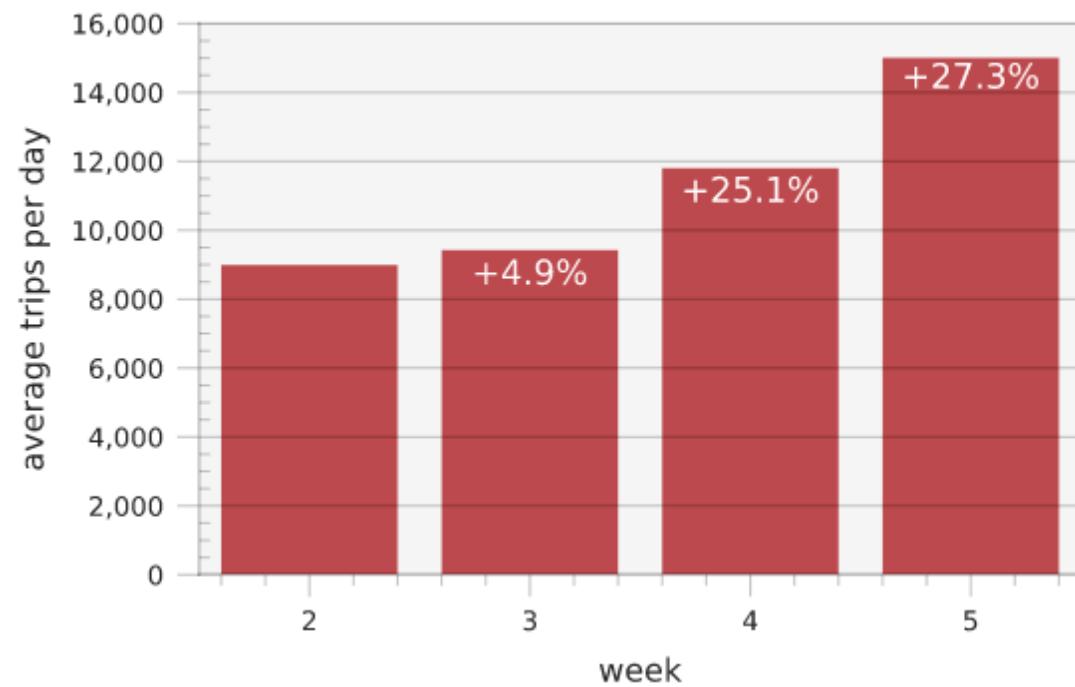


1. <https://www.lyft.com/bikes/bay-wheels/pricing>

Growth of bike-usage



average Bay Wheels trips per week-day



Highlights

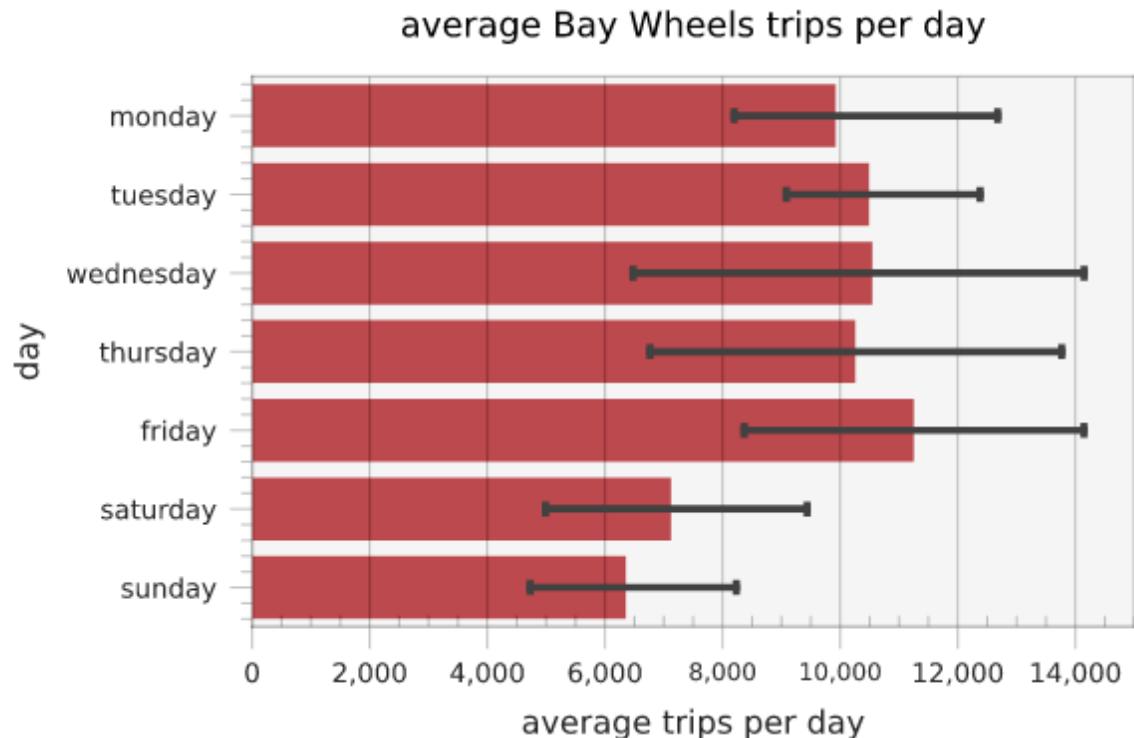
- Week-on-week growth of trips of 19.1% for weeks 3-5.
- Less trips during the weekend compared to weekdays.
- We do not have data on the same amount of days for each week, hence we only look at weekdays in week 2-5 to get a more accurate view on the weekly growth.

Insights

- Lower number of trips in week-1 could be caused by holidays (people are likely to take time off around the holidays).
- Drop in trips taken during weekend could mean that the bikes are used predominantly for commuting traffic.
- There is no data on potential marketing campaigns ran in January that could explain the growth. We did find that Lyft would bring back e-bikes in the start of 2020.

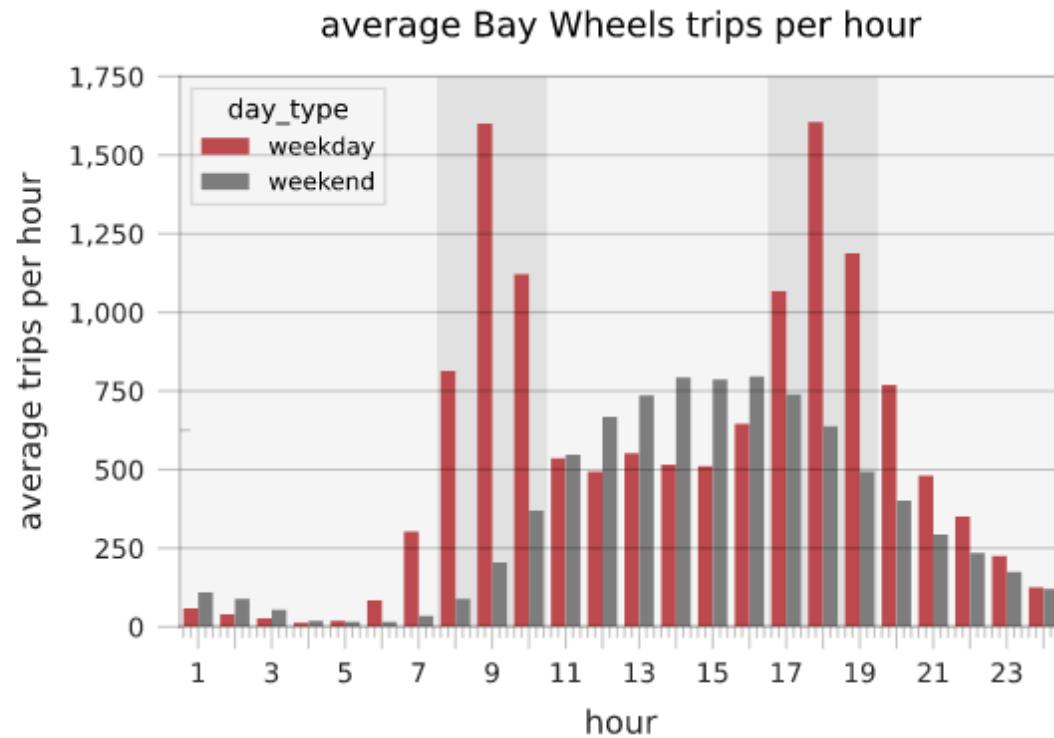
1. <https://www.sfchronicle.com/business/article/Lyft-s-Bay-Wheels-to-resume-e-bike-rentals-in-SF-14821170.php>

Trips during the Weekend vs. Weekdays



Highlights

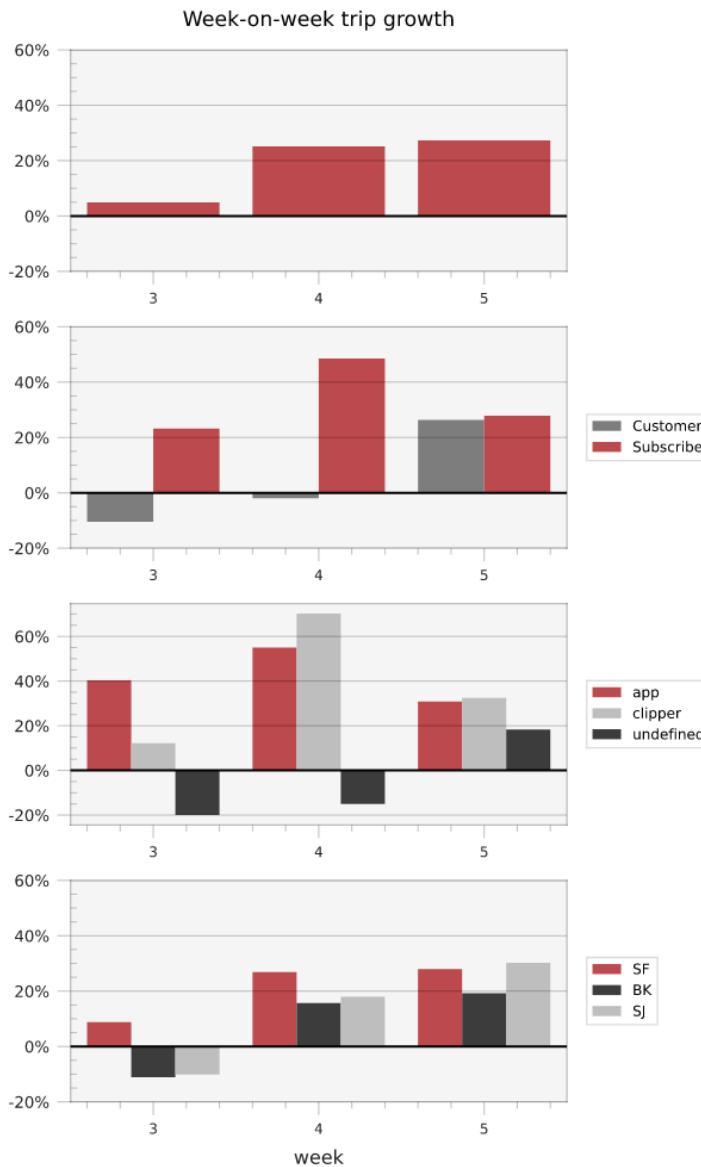
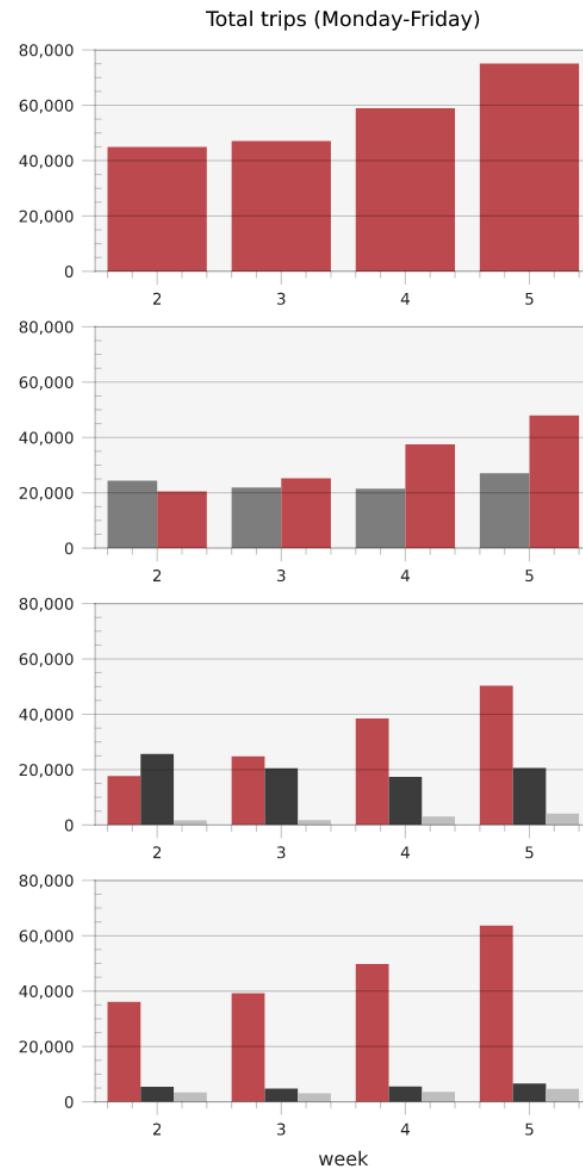
- We see 56% more trips on weekdays compared to weekend-days. This has been established with limited data, where the weekend in the busiest week (5) was not included. For an improved estimate, more weeks would need to be included.
- Effect of commuting is clear in the hourly graph. On weekdays, the hours 8-11 am and 17-20pm account for 51% of daily trips, whereas these hours only account for 31% of daily trips on weekends.



Insights

- Bay Wheels seems to be a popular service for commuters traveling to and from work on weekdays. On weekends, it seems to be used more for trips during the afternoon and for trips just after midnight.
- There is no distinctive difference between the type of user or the location for weekend or weekday use.
- We do see that 61% of travelers use the app to rent a bike in the weekend, versus only 54% using the app during the peak hours on weekdays

Demographic breakdown of weekly growth



Highlights

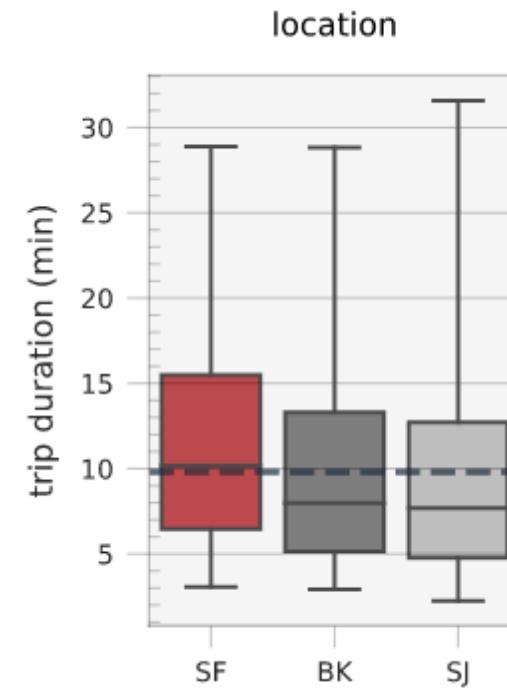
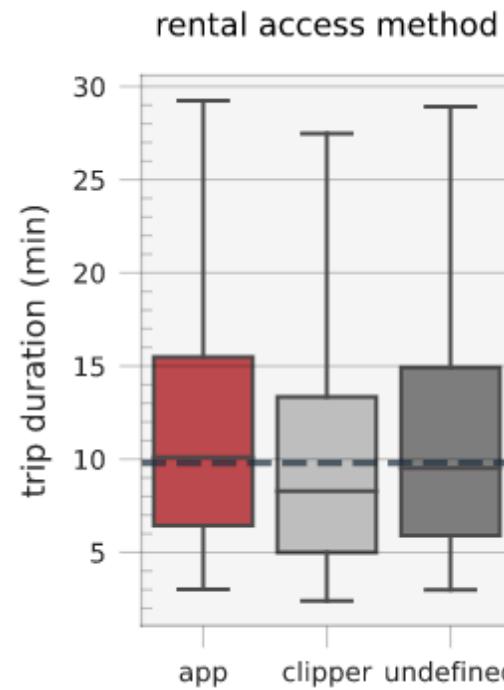
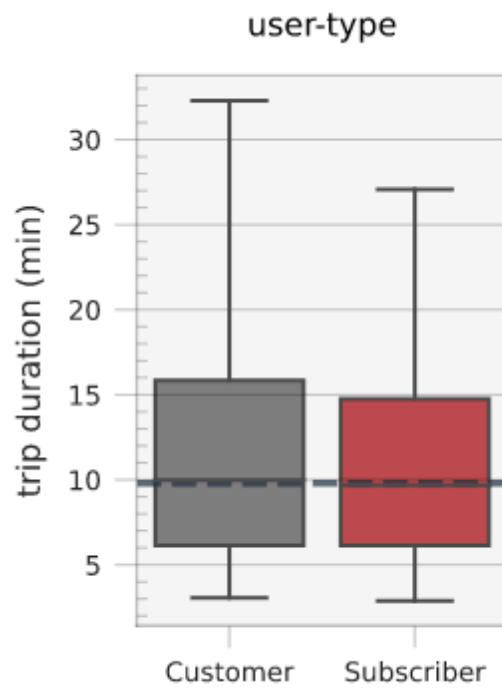
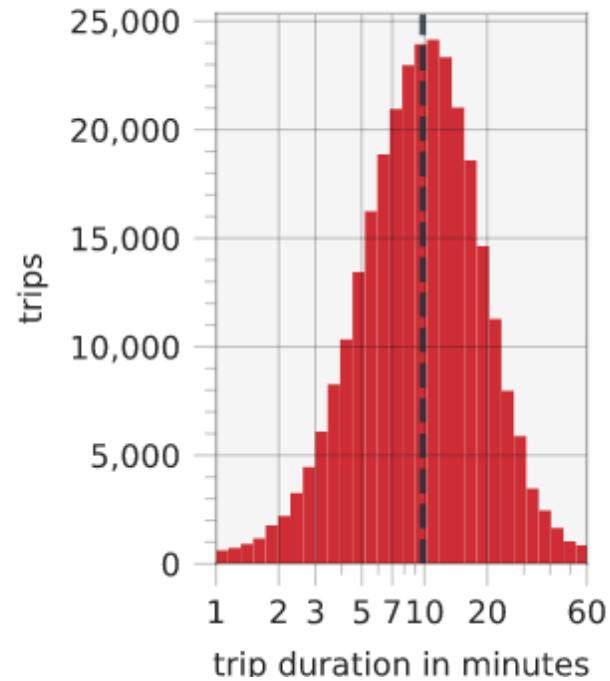
- The week-on-week growth rate for trips made by subscribers is 33%, substantially higher compared to the overall growth rate of 19%.
- Rental Access Method ‘undefined’ (i.e. no data available) is decreasing over time. In week 2, 57% of all trips were logged as ‘undefined’, in week 5 this has dropped to 27%.

Insights

- Since this data is anonymized, we cannot see if a single user makes multiple trips here nor can we see if a user changes from customer to subscriber, which limits our interpretation of the observed breakdown of the growth.
- One possible explanation for the ‘undefined’ access method could be that this is a new parameter. It is not yet present in the data dictionary provided¹ and the rapid decline of trips with this flag might be caused by a slow introduction of a new logging method (in app or backend) for this parameter. Looking at data from more recent months might provide more insight to this.

1. <https://www.lyft.com/bikes/bay-wheels/system-data>

Trip Duration



Highlights

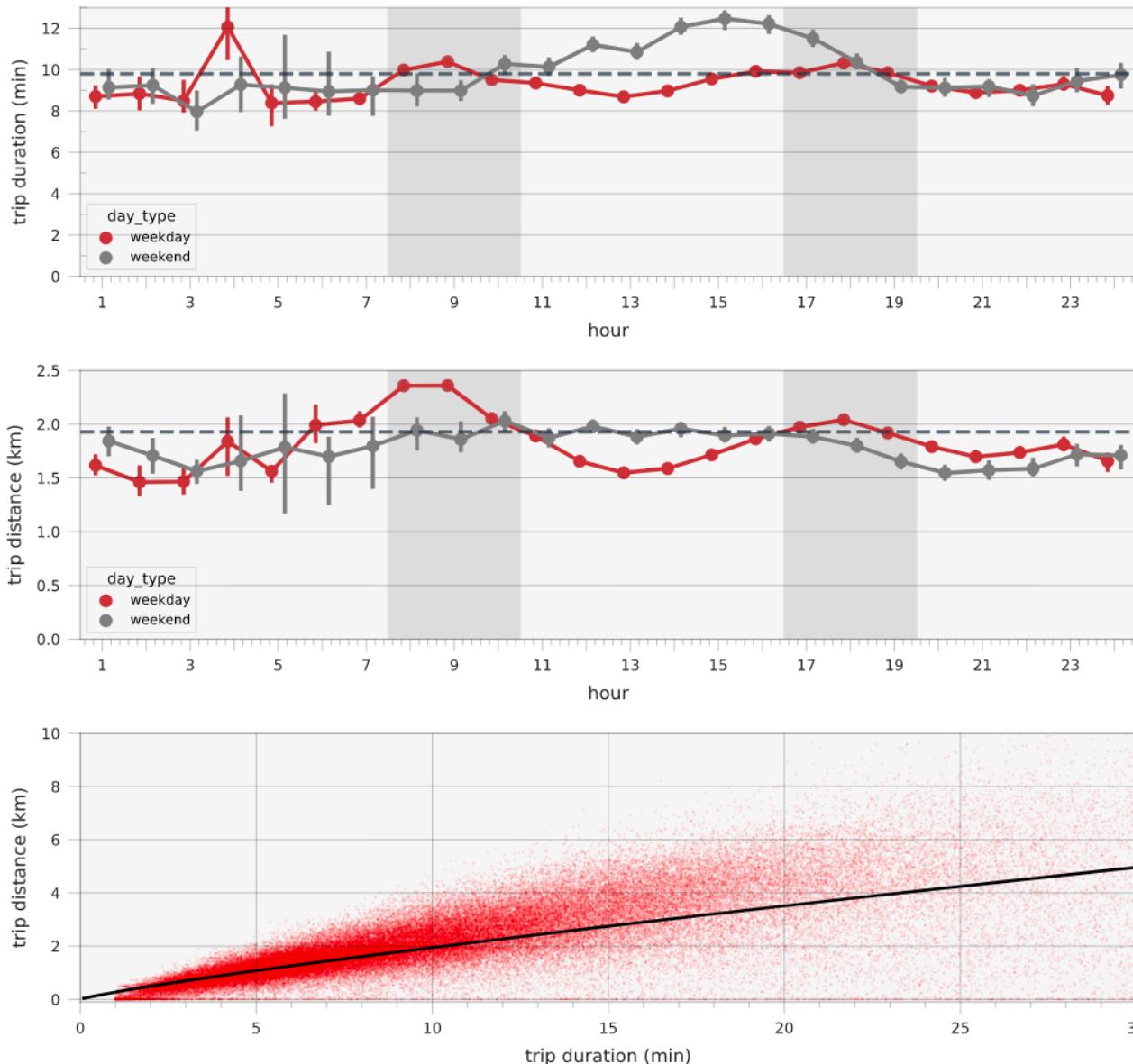
- Trip duration follows a right-skewed distribution that is approximately log-normal with a median trip duration of 9:48 minutes
- 95% of trips take less than 30 minutes, and ~99% of trips take less than 1 hour.
- The median duration of trips in San Francisco is substantially longer (10:11 min) compared to the other locations (Berkeley 7:59 min, San Jose 7:41 min)

Insights

- There is little difference between user-types (median trip duration differs by only 18s), however the distribution of trip durations for customers is skewed more upwards.
- ‘Undefined’ and ‘App’ access methods again show similar patterns. Combined with the earlier seen growth of the individual groups, this supports the hypothesis that we are witnessing a shift from trips logged as ‘undefined’ to ‘app’.
- The difference in trip duration per location could be related to a restriction to a specific area¹ which differs in size per city.

1. <https://member.baywheels.com/map/>

Trip Distance & Duration



Highlights

- The distance distribution of trips shows a similar distribution as seen for the duration (right-skewed); with a median distance of 1.9km
- Comparing distance and duration per hour, some interesting differences are revealed:
 - The trip duration on weekend-afternoons is higher than average (~12 min), but the distance of these trips stays around the median (1.9km)
 - On peak commuting hours on weekdays, the duration stays around its median (~10 min), whereas the trip distance sees its peak (2.4km around 8-9am)

Insights

- We see ~500 trips where the duration > 0 but the distance is zero. This could be caused by the fact distance is measured between start and end-point; hence it's zero if the user returns to the starting position.
- Overall, we see that there is a strong positive correlation (correlation coefficient of 0.67¹) between distance and duration. Applying regression², reveals an average speed of bikers of 8 km/h.

1. Calculated on a subset with outliers removed (duration > 0 and < 60 mins, distance > 0 km)

2. Value based on regression based on subset mentioned in (1). In the plot on the left, a regression line is plotted where the regression was applied on the log-transformed distance and duration, which proved to have a superior RMSE of 119 vs. 185 for trips in the lower left quadrant (by median trip duration and median trip distance).

Conclusion

Strong weekly growth of trips (19.1%)

- This growth correlates mainly with a strong growth of 'Subscribers' making trips (+33% week-on-week).
- Most trips occur in San Francisco (~85% of total), however, the growth is present across all locations in week 4 and 5 (trip growth from week 4-5 is > 18% for all locations).

Distinct patterns of bikes usage for weekends and weekdays

- On weekdays, there are 56% more trips compared to weekends.
- 51% of all trips on weekdays occurring during the 6 morning and evening rush hours (8-11 am, 5-8 pm).
- During the morning rush-hours, we also observe an up-to 25% increase in trip-distance compared to the median distance (1.9km).
- On weekend, most trips occur in the afternoon. In the afternoon, we see an increase in trip duration of up-to 20% compared to the median (~10 min), without a notable increase in trip duration.

Understanding bike usage patterns could improve LYFT's business

- Marketing & differential pricing could help drive business towards hours or days where less trips are recorded.
- Bikes could be optimized for the observed usage patterns in terms of i.e. the battery-pack required for e-bikes.

Discussion & Outlook

Not covered in this report

Whereas data was provided on start and end station, we did not investigate any patterns related to this.

Limitation of data to January 2020

Only data from January 2020 was used to create this report, which limits the interpretation of the growth patterns and weekend vs. weekday patterns. Gathering data on more months would provide more accuracy on the statistics gathered as well as context.

Valuable additional data-sources

- As we are dealing with bike-trips, collecting weather data could provide an extra dimension towards a more diagnostic analysis.
- Having a (separate) dataset on customer behavior (i.e. per customer, without locations for privacy) would provide more context as to whether we're observing a growing userbase, or a growing usage pf bikes by a constant userbase.