Bike Sharing Service in Boston, MA

Daren Smith - Data Analyst

Preface

This project uses data from a bike rental service in Boston, MA during the years of 2012 and 2013.

Only relevant results are shown in this presentation. To see the code and process, please visit the project page on my GitHub profile at:

https://github.com/smithdaren/Bike-Sharing-Boston/blob/main/BikeShareProject.ipynb

All analysis was done in Python.

Table of Contents

01

Overview

Bike Sharing Services

To meet city growth and public transit demands, bike rental services are available.

Users pay for usage through the amount of time they rent bikes.

Bike services assist in increased community access and provide clean transportation.

02

Data Preparation & Exploratory Data Analysis

Rentals in Boston, MA

Data was gathered from a bike rental service in Boston, MA. Rentals between the beginning of October 2012 to the end of November 2013 are recorded.

Over **1 million rentals** are recorded and included in the analysis.

Additionally, information on **131 stations** installed in Boston are available.

Image Source: ButteBag, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Example of Data

Rental Data:

rental...	status	durati...	start_...	start_s...	end_d...	end_st...	bike_nr	subscr...	zip_co...	birth_...	gender	durati...
708191	Closed	240	2012-11-...	53	2012-11-...	67	B00412	Registered	02139	nan	Male	240
708190	Closed	840	2012-11-...	36	2012-11-...	115	B00491	Registered	02127	nan	Female	840
708189	Closed	420	2012-11-...	139	2012-11-...	140	B00101	Registered	02145	nan	Female	420
708187	Closed	120	2012-11-...	22	2012-11-...	81	B00092	Registered	02116	nan	Male	120
708186	Closed	420	2012-11-...	32	2012-11-...	55	B00509	Registered	02115	nan	Female	420

Station Data:

id	terminal	station	munic...	lat	lng	nb_do...	install...	last_day
3	B32006	Colleges ...	Boston	42.340021	-71.100812	15	2013-04-...	2013-11-...
4	C32000	Tremont ...	Boston	42.345392	-71.069616	15	2013-04-...	2013-11-...
5	B32012	Northeas...	Boston	42.341814	-71.090179	15	2013-04-...	2013-11-...
6	D32000	Cambrid...	Boston	42.361285	-71.06514	15	2013-04-...	2013-11-...
7	A32000	Fan Pier	Boston	42.353412	-71.044624	15	2013-04-...	2013-11-...

Rental Data Available

	Data Type	Description
Rental ID	ID	Unique ID of each rental
Status	Categorical	Whether the trip is in progress or finished
Duration	Numerical	Number of seconds the trip took
Start Date	DateTime	The date and time a rental was started
Start Station	ID	Unique ID of station the bike was rented from
End Date	DateTime	The date and time a rental was ended
End Station	ID	Unique ID of station the bike was returned to
Bike Number	ID	Unique ID of the bike used during the rental
Subscription Type	Categorical	Indicates if the renter is a Registered or Casual user
Zip Code	Categorical	Zip code of user (if known)
Gender	Categorical	The gender of user (if known)

Station Data Available

	Data Type	Description
ID	ID	Unique ID of each station
Terminal	ID	Terminal code of each station
Station	ID	Readable name of each station
Municipality	Categorical	Municipality or district where the station is located
Latitude	Numerical	Latitude coordinates of station location
Longitude	Numerical	Longitude coordinates of station location
Number of Docks	Numerical	Number of bike docks installed in station
Install Date	DateTime	Date of station installation
Last Day	DateTime	Date of last rental from station

Notes

Missing Dates

One station (Overland St at Brookline Ave with ID number of 34) is missing installation and last day dates. Data for this station was still used for analysis.

Negative Trip Durations

49 trips showed negative trip durations. This could be due to issues with time records during Daylight Savings Time. These trips were removed from analysis.

Rental Duration Time

15.57 minutes

Average

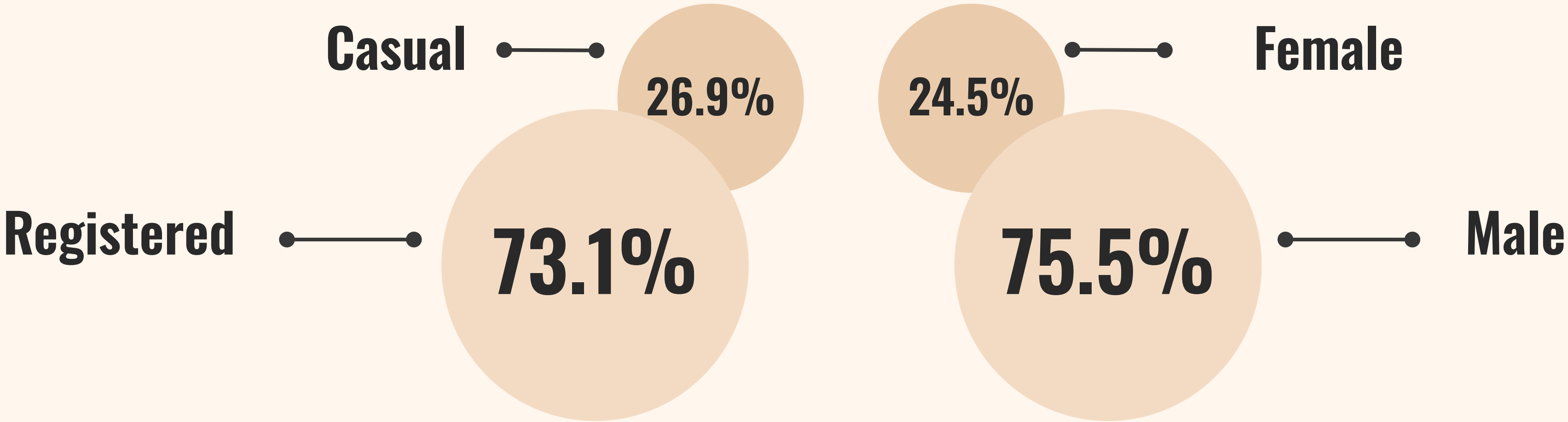
11.0 minutes

Median

52288.0 minutes

Max

Riders by Subscription Type & Gender

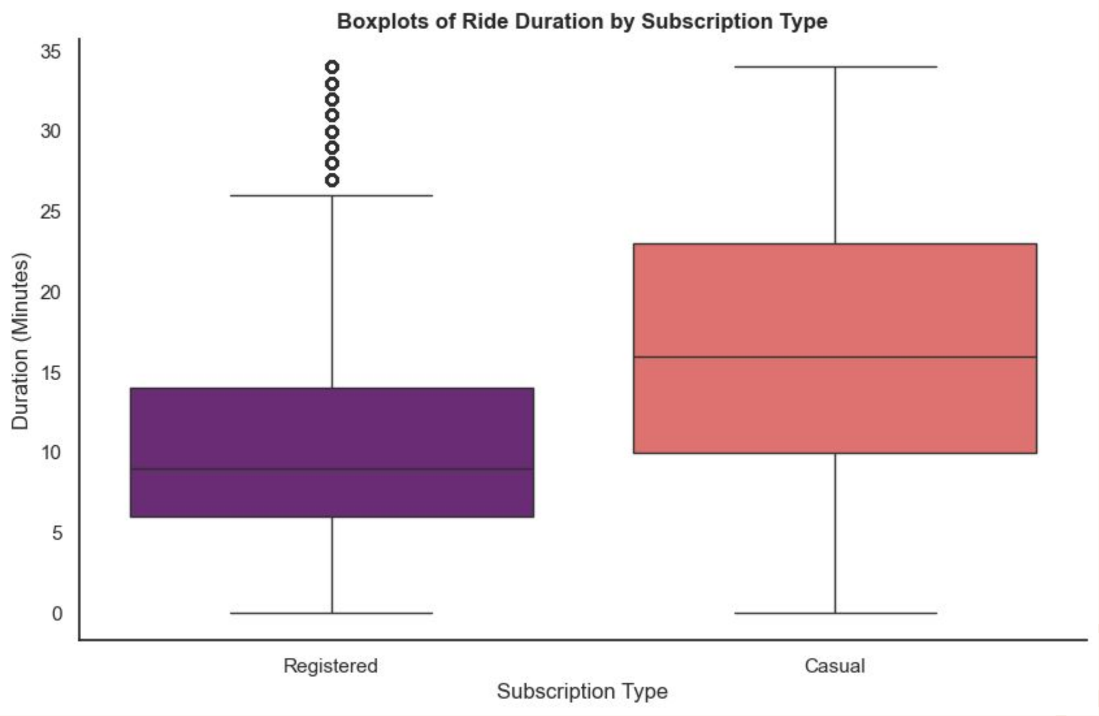


Note: Only registered riders have their gender data present, so these proportions are only for registered members.

Ride Duration by Subscription Type

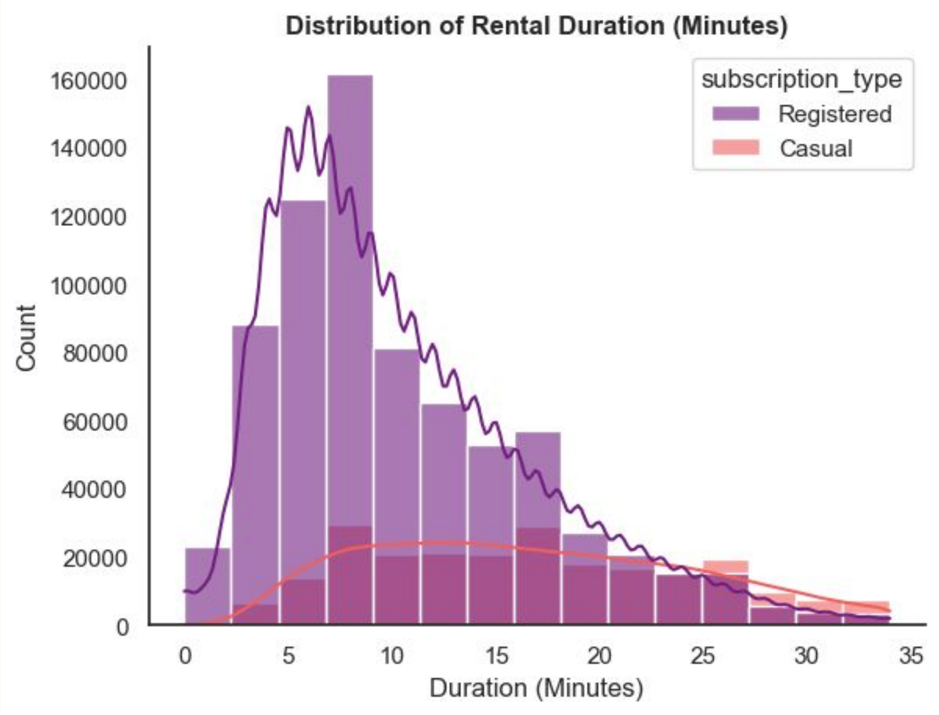
After removing high rental times, we see the distribution of rental times for each subscription type.

While they aren’t too different, the Casual users tend to have a higher variance in their ride times.



Ride Duration by Subscription Type

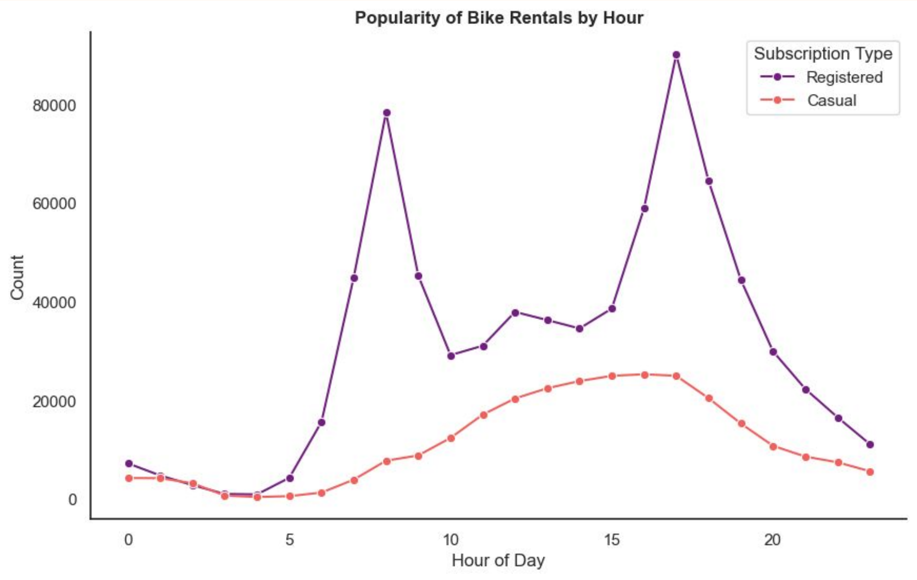
- Registered riders have a peak riding time of around 7.5 minutes
- Casual riders don't have a certain preferred riding length and their time might vary.



03

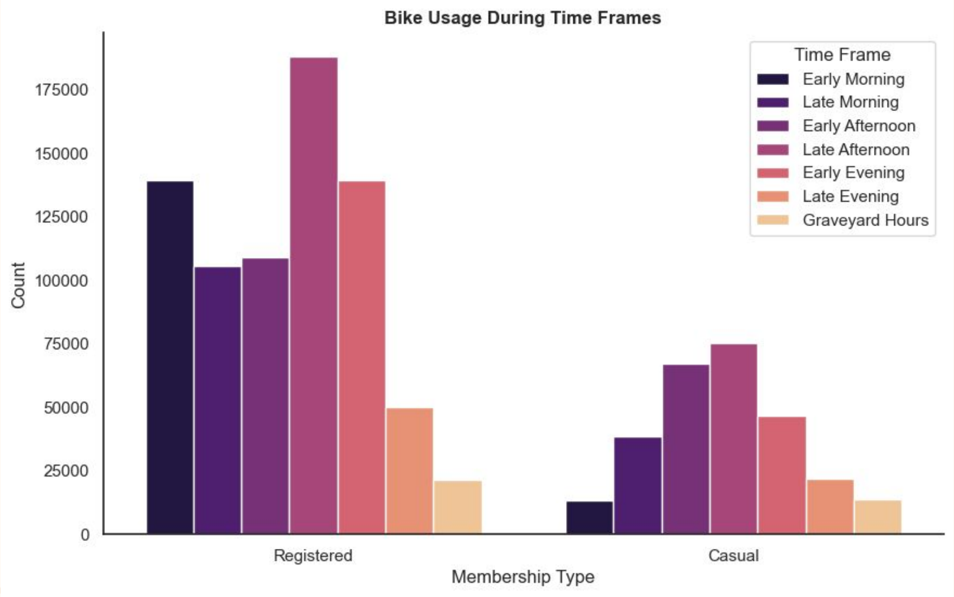
Rentals by Time

Total Rentals by Hour



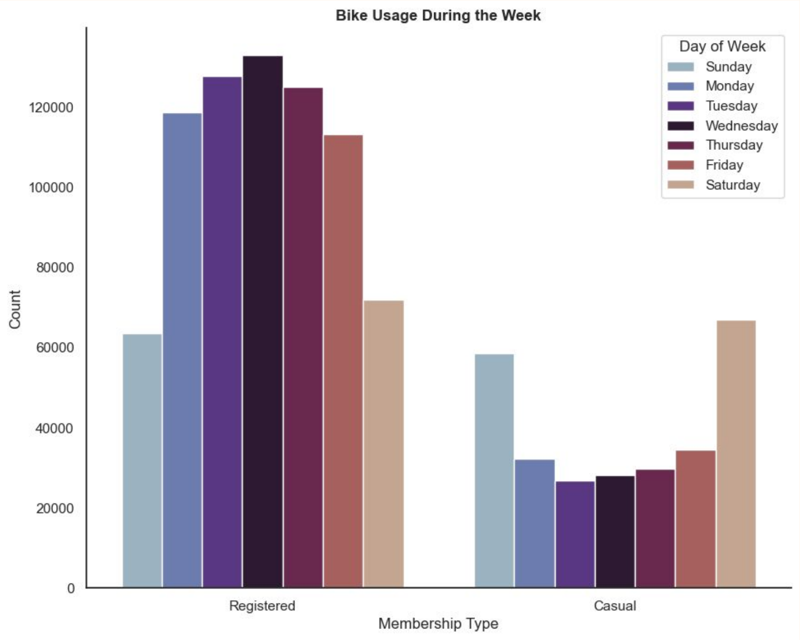
- Peak hours for Registered users are between 7 and 9 am, and between 4 and 6 pm. Likely commuters to school or work.
- Casual users rent throughout the afternoon until they start to taper off around 5 pm.

Total Rentals by Hour



To reinforce the previous slides, we see that ridership peaks for Registered at the highest point in the late afternoon, while it peaks in the early afternoon for casual users.

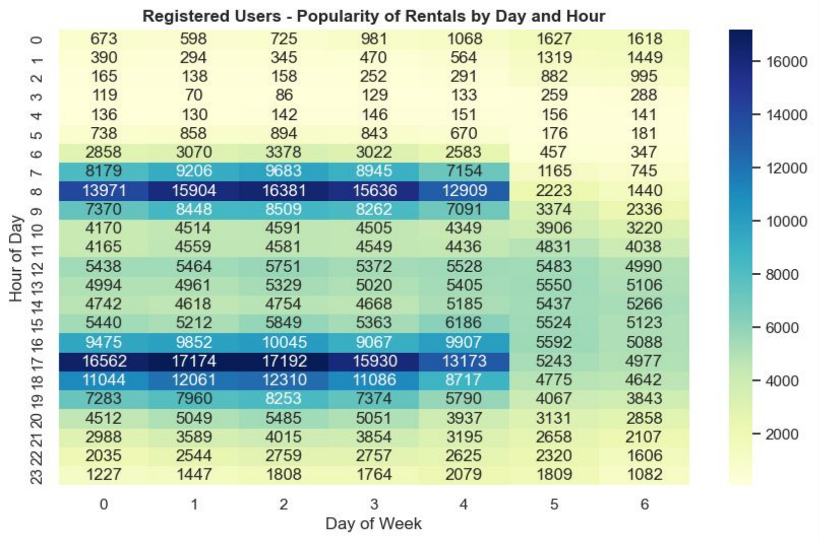
Rentals by Weekday

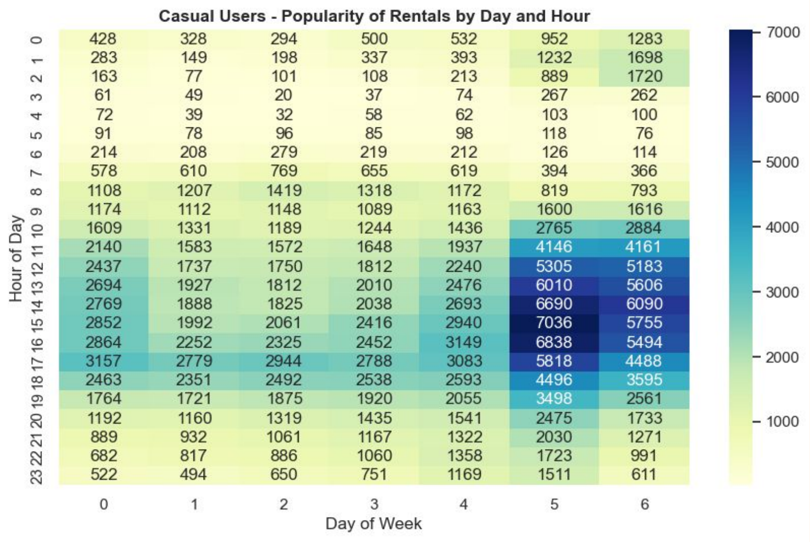


- Premium users rent from us primarily during the week
- Casual users are more active during the weekend

Peak Days and Hours

Note: 0 = Monday, 6 = Sunday

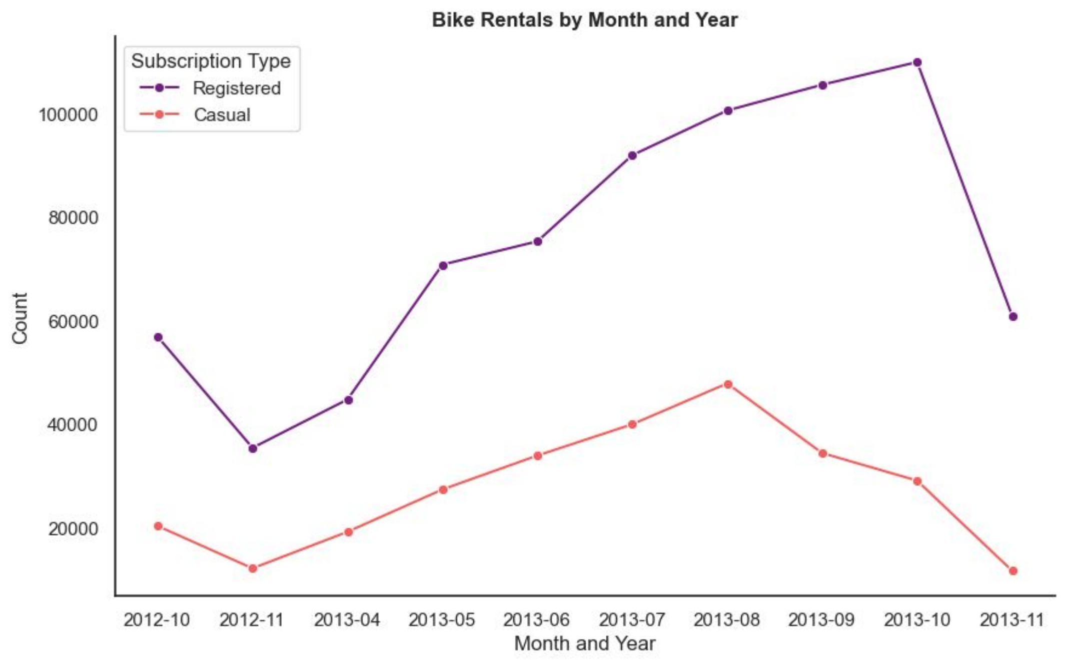




Highest counts are 5 pm on Tuesday and Wednesday. There is a slight drop off in usage overall on Friday. Weekend usage is small but tends to be between 12 pm to 5 pm

Medium usage in late morning and afternoon on Monday and Friday. We can see Saturday and Sunday both show high usage counts, especially during the afternoon and early evening.

Rentals Per Month



- Registered users ride often throughout the Fall and drop off in October
- Casual users peak in August
- Overall, seems as though rentals are increasing over time

Rentals Statistics

3410

Total Rentals Per Day (All Stations)

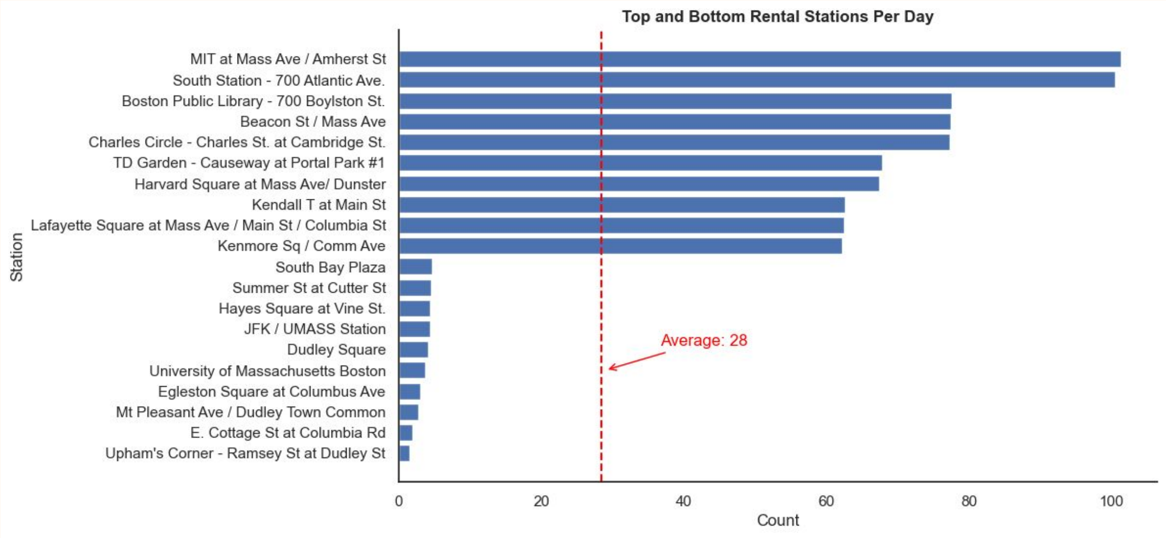
Average Rentals Per Station Per Day

28

17.6

Average Bike Docks Per Station

Stations With Highest and Lowest Average Rentals

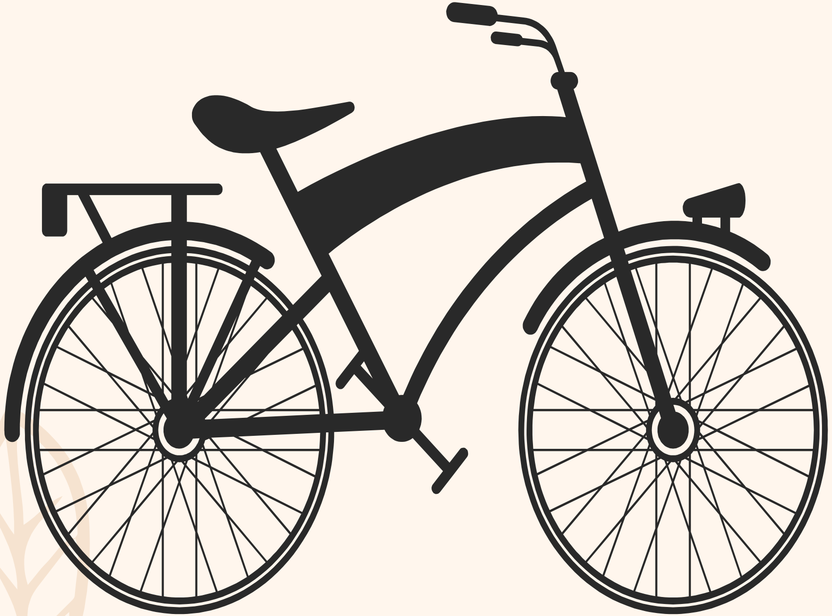


Certain stations (MIT and South Station) have high rentals per day. Others (Upham’s Corner and E. Cottage) have low usage.

Stations with low rentals can be understocked and bikes here can be shifted into high density stations.

Most Popular Routes

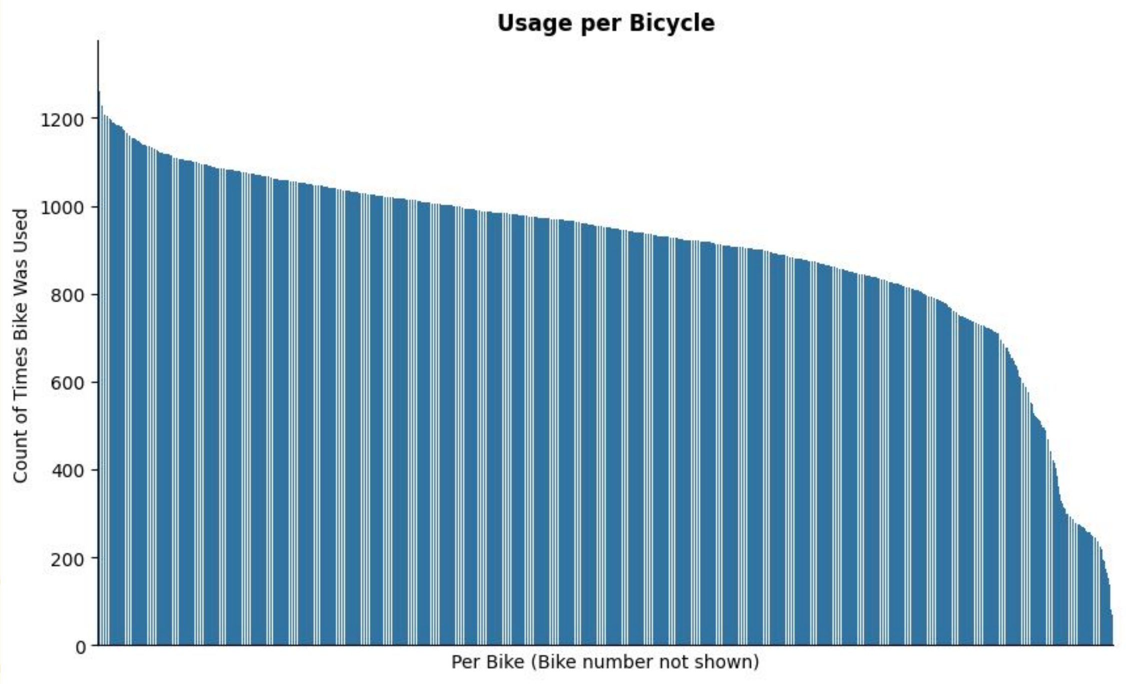
	Start Station	End Station	Total Trips
Route 01	Beacon St / Mass Ave	MIT at Mass Ave / Amherst St	4747
Route 02	MIT at Mass Ave / Amherst St	Beacon St / Mass Ave	4610
Route 03	Kenmore Sq / Comm Ave	MIT at Mass Ave / Amherst St	2539
Route 04	MIT at Mass Ave / Amherst St	Kenmore Sq / Comm Ave	2370
Route 05	Lewis Wharf - Atlantic Ave.	South Station - 700 Atlantic Ave.	2025



04

Bike Information

Total Rentals Per Bike

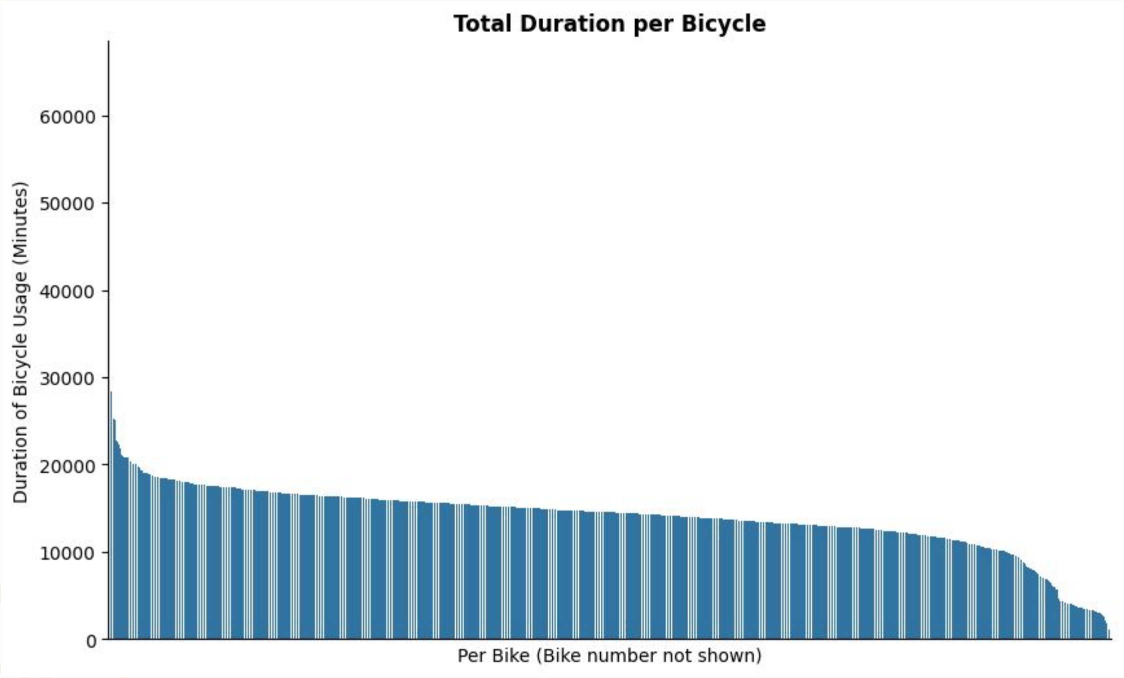


Top Rented Bikes

Bike #	Total Trips
T01307	1310
T01296	1261
T01035	1259
T01416	1247
T01193	1227

Bikes that are rented often can be identified and circulated out of inventory for maintenance.

Total Rentals Per Bike



Top Ridden Bikes

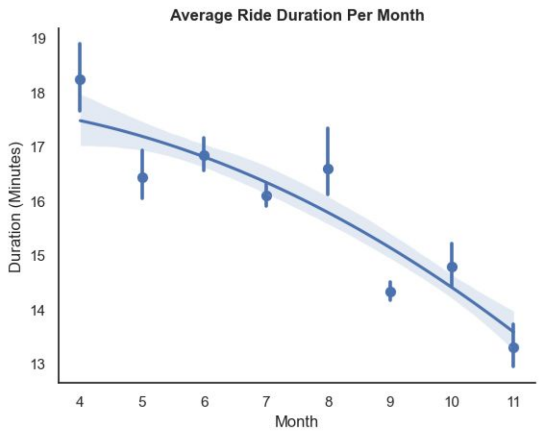
Bike #	Total Minutes Ridden
B00050	65206
B00411	48085
B00132	36223
T01186	28304
T01033	28186

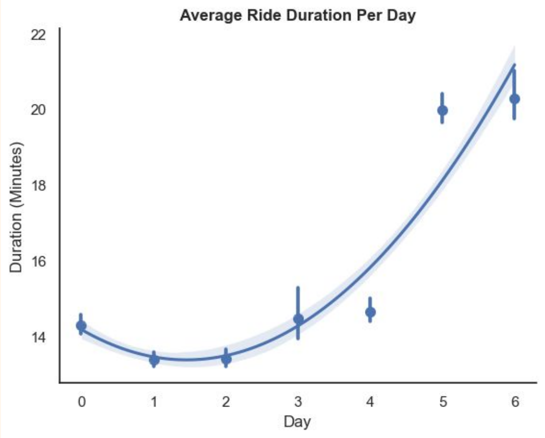
These bikes have had high total riding times. We can also identify these for maintenance and rotate in fresher bikes.

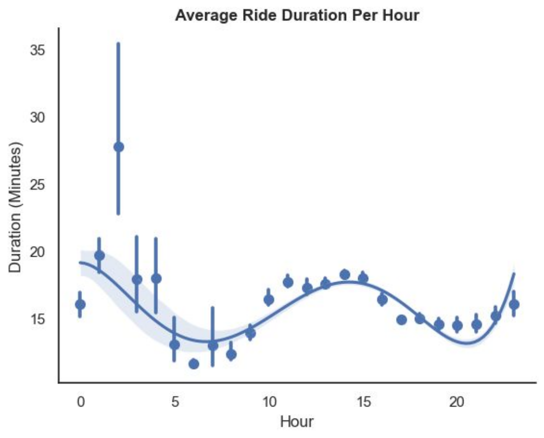
05

Predicting Riding Times

Average Ride Duration Over Time







The relationships between time and duration is not completely normal, but we can see some distinct average ride times depending on the month, the day, and the hour.

Model Development

Our predictor variables are not numerical, so we cannot do a correlation matrix. So instead, we will hot-encode our categorical variables and run our regression model.
The dependent variable is:

- Duration (in minutes)

The independent variables included are:

- Subscription Type
- Gender
- Hour
- Day of Week
- Month
- Starting Station (only used in Model 1)

The data was split into 80% training and 20% testing samples for model prediction.

Model Output

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       duration_minutes   R-squared:                       0.208
Model:                            OLS   Adj. R-squared:                  0.208
Method:                 Least Squares   F-statistic:                     1263.
Date:                Mon, 08 Jan 2024   Prob (F-statistic):               0.00
Time:                        21:53:13   Log-Likelihood:            -2.8406e+06
No. Observations:              809329   AIC:                         5.682e+06
Df Residuals:                  809160   BIC:                         5.684e+06
Df Model:                         168
Covariance Type:            nonrobust
```

Model 1 (with starting station included):

- R-Squared of 20.8%, but highly complex due to large number of starting stations.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       duration_minutes   R-squared:                       0.185
Model:                            OLS   Adj. R-squared:                  0.185
Method:                 Least Squares   F-statistic:                     4387.
Date:                Mon, 08 Jan 2024   Prob (F-statistic):               0.00
Time:                        21:53:39   Log-Likelihood:            -2.8519e+06
No. Observations:              809329   AIC:                         5.704e+06
Df Residuals:                  809286   BIC:                         5.704e+06
Df Model:                          42
Covariance Type:            nonrobust
```

Model 2 (with starting station excluded):

- R-Squared of 18.5%, but less complex with starting stations omitted.

06

Recommendations & Next Steps

Recommendations

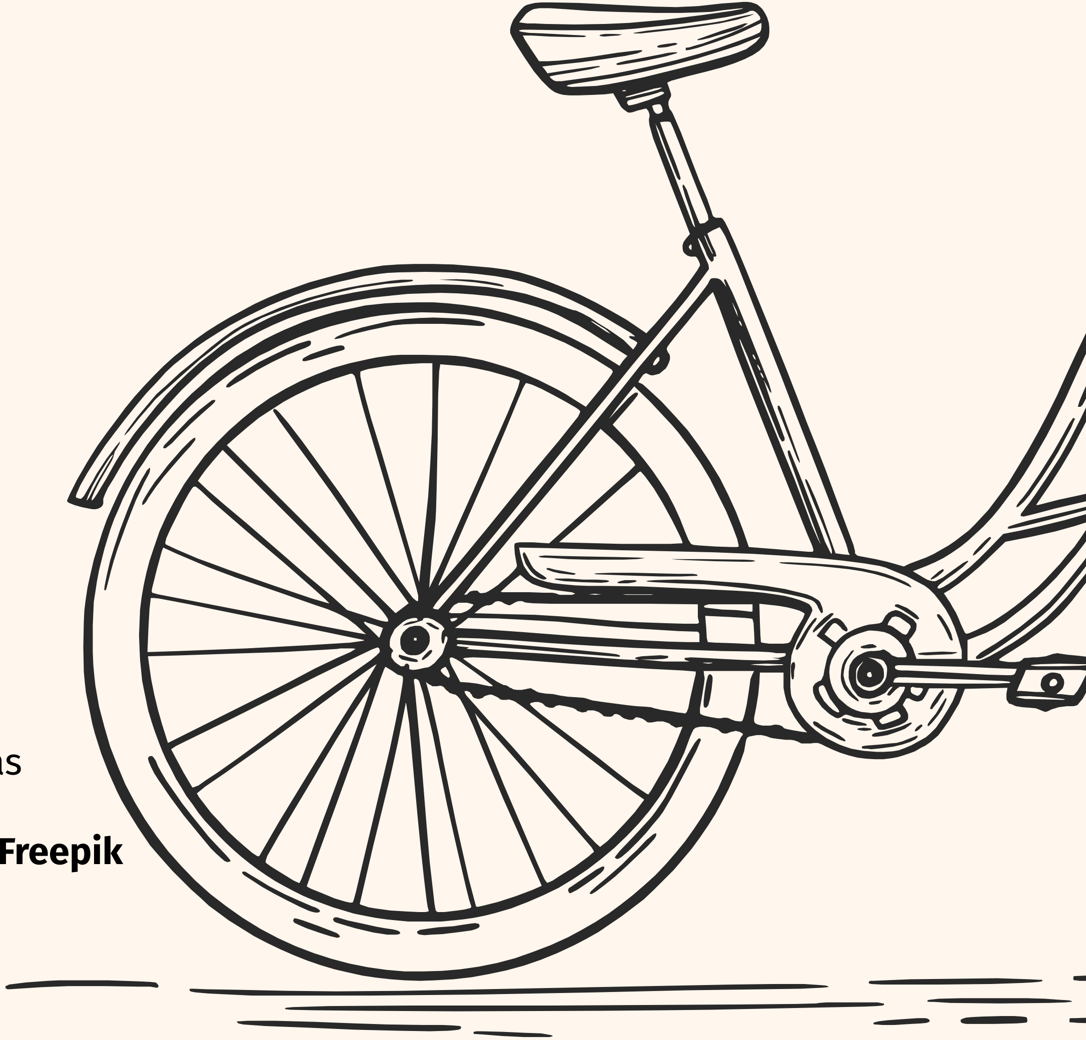
1. Implement rider IDs into our Data so we can track consistent usage for riders.
2. Improve timestamp collection (especially around daylight savings time), and record all timestamps with proper timezone information.
3. Identify bikes that are heavily used and rotate them out of service for maintenance. Swap in less used bikes to handle the extra load.
4. Identify stations with high rental rates and truck in extra bikes to these stations throughout busy hours.
5. Stock up bikes throughout the early afternoon (when bike usage drops) in order to handle high volume demands in the late afternoon and early evening.
6. Take advantage of casual rentals on weekends. As registered users don't use this service as much on the weekends, find ways to encourage socialization or exploration on weekends for our casual members to make up for that difference.

Next Steps

1. Place station locations on a geolocated map, such as Google Maps, to see the density and dispersion of the stations. Take this a step further and identify the most common routes in order to understand how riders use the service and plan future station installments.
2. Identify more data to include in our model to help improve predictive accuracy. This includes rider demographics (income, neighborhood of residence, etc.) or behaviours (work status, student status, socialization habits, etc.). Also get meteorological data (temperature, precipitation, wind, humidity, etc.) to help with the predicted usage and duration of rentals.

Thanks

Presentation by:
Daren Smith
Data Analyst



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution