

Bike Rental Analysis

Executive Summary

- 70 bike stations are available, the majority of which are in San Francisco and San Jose.
- 324,652 bike trips occurred in 2014, most which were in San Francisco and San Jose.
- The average trip duration was ~15 minutes, though this varies slightly by city.
- Weather in the Bay Area is suitable for biking year-round.
- Weekday peak hours are from 7:30 AM to 10:30 AM and 4:00 PM to 8:00 PM.
- On average, each bike is rented out for only 1.49% of the month.
- Higher temperatures and visibility are associated with an increased number of bike trips, while precipitation and cloud cover are associated with fewer bike trips.
- Link to GitHub Repo: https://github.com/ibrahimemam7/CodingInR_midterm
- Link to Code Book: <https://www.kaggle.com/datasets/benhamner/sf-bay-area-bike-share/discussion/23165>

Exploratory Data Analysis and Data Cleaning

EDA Round 1

Exploratory data analysis (EDA) was conducted for the three datasets provided (stations, trips, weather). The EDA provided insight into the type of information contained within each dataset. Most importantly, the first-pass EDA helped to clarify the next-steps needed for data cleaning and preparation.

Data Cleaning and Preparation

The stations dataset had no duplicates or unusual observations. The installation date column was converted to POSIX format for ease of manipulation in downstream analysis.

The trips dataset had no duplicates. Trip duration was converted to minutes (originally reported in seconds) for ease of interpretation. Trips that were shorter than 3 minutes and had the same start/end station were cancelled trips. Additionally, trips longer than 12 hours were removed since these cases were likely due to the bike being improperly returned to the dock. The ID numbers of all excluded trips are noted in a CSV file in the repository linked above. Additionally, the start and end times of trips were converted to POSIX format for easier downstream manipulation.

The weather dataset had no duplicates. No notable outliers were found in the dataset. The dates for each observation were converted to POSIX format for consistency with the other two datasets. According to the codebook linked above, trace amounts of precipitation (<0.01 inches) were indicated with "T". To keep this column in numeric format for downstream analysis, the value 0.005 inches was imputed in place of "T."

EDA Round 2 - Summary of Clean Data

I – Bike Stations

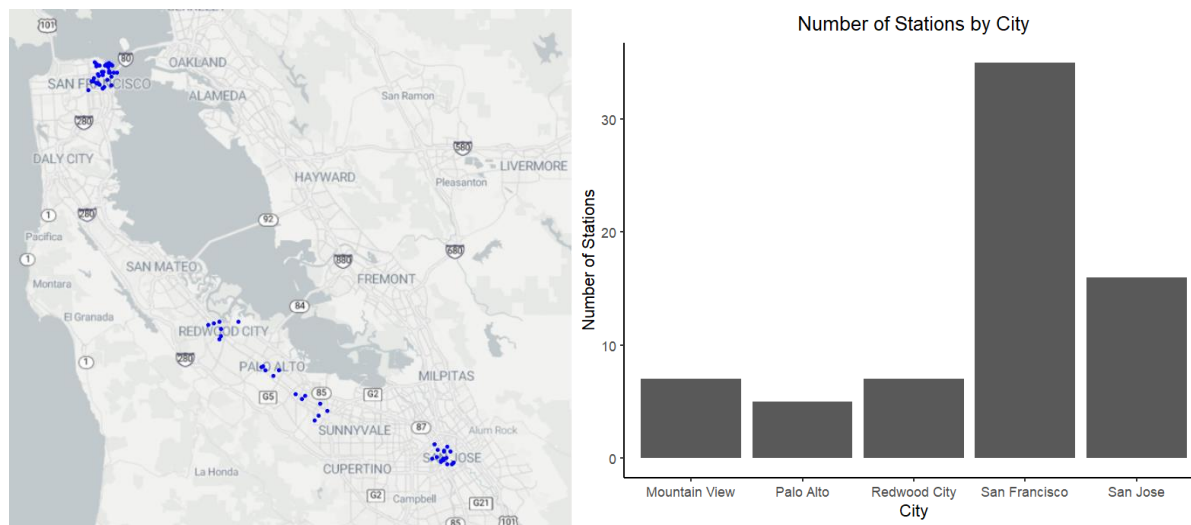


Figure 1. Left: map of bike rental stations across the bay area. Right: number of bike rental stations by city. Total of 70 stations.

A total of 70 bike rental stations are available throughout the bay area, with the majority in San Francisco and San Jose.

II – Bike Usage

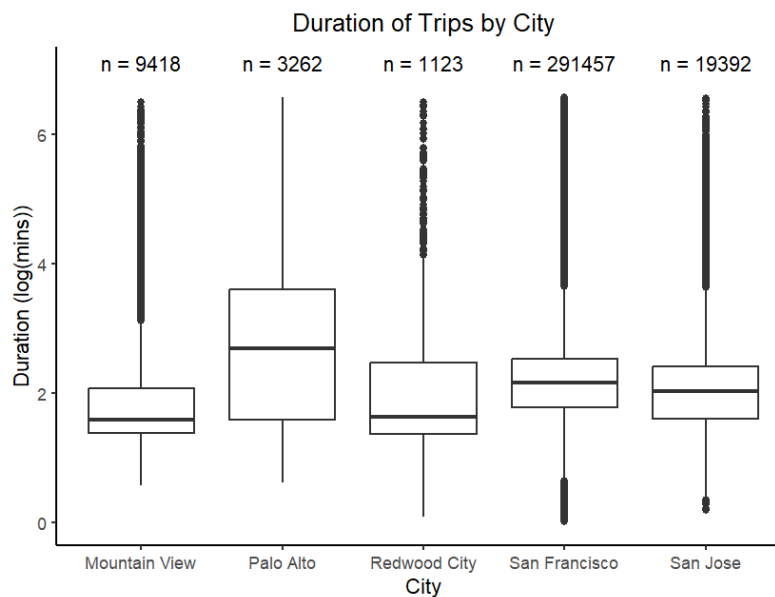


Figure 2. Average duration of trips and number of trips for each city.

A total of 324,652 bike trips occurred in 2014. Most of the trips occurred in San Francisco and San Jose. This is not surprising since those are the two most populated cities represented in the dataset. Duration was log-transformed due to the large variation in trip duration. The average duration was 15.07 minutes.

III – Weather Patterns

Weather observations were made for each day of the year 2014 in 5 cities across the Bay Area. Key data is summarized in the figures below.

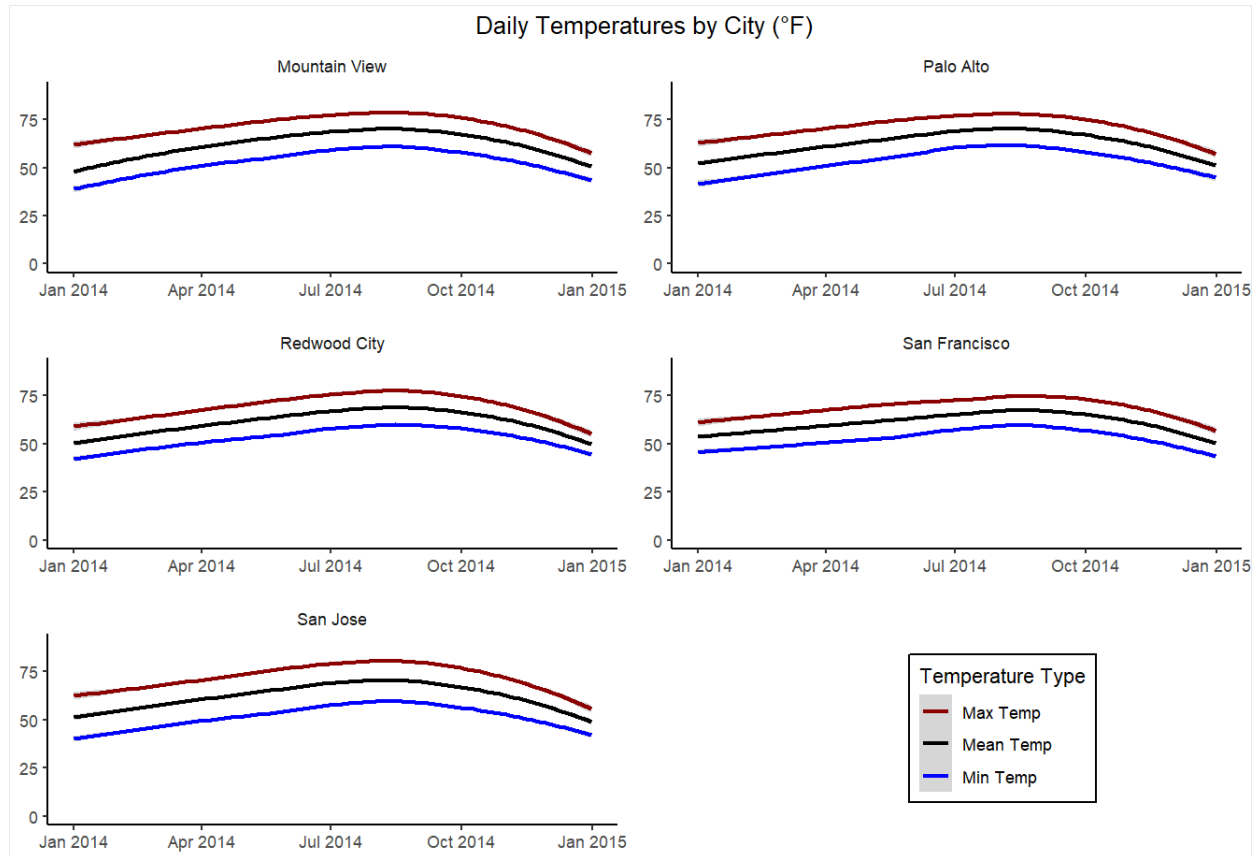


Figure 3. Average, minimum, and maximum temperature throughout 2014 in the Bay Area.

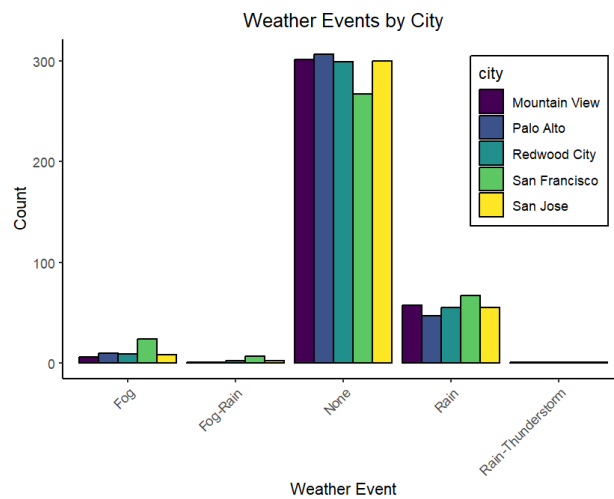


Figure 4. Number of weather events in 2014 for 5 cities in the Bay Area.

The average temperature in the Bay Area is approximately 63.03 °F. The temperature is typically suitable for biking all year round. The data shows no meaningful differences in temperature across the 5 cities. For all 5 cities, most days had no weather events (e.g. rain, fog, or thunderstorms). However, San Francisco had more foggy days than other nearby cities. Overall, all 5 cities have weather than is highly conducive to biking.

Rush Hours Analysis

To determine the rush hour period during weekdays, the midpoint time of each trip was calculated. This method prevents any bias towards earlier hours or later hours by using trip start time or trip end time, respectively. Weekday trip volume was visualized using the histogram below.

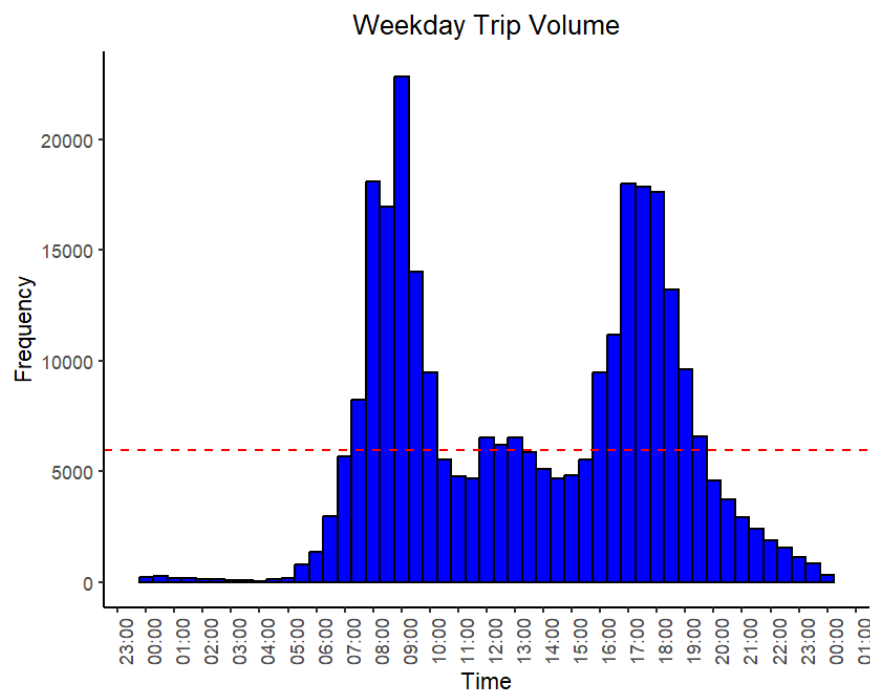


Figure 5. Weekday trip volume (0.5-hour increments). Red dashed line represents average frequency.

Based on the histogram above, peak hours seem to be from 7:30 AM to 10:30 AM and 4:00 PM to 8:00 PM. This result is not surprising since these times coincide with the typical times in which people would commute to or from work. There is also a slight increase in trip volume from 12:00 PM to 1:30 PM, likely due to lunch time. However, this time was much less busy than the previously mentioned time intervals, so it is not considered as part of the rush hours.

High Traffic Stations Analysis

The most frequent starting stations and ending stations during the weekday rush hours are shown below.

Table 1. Highest Traffic Start and End Stations During Weekday Rush Hours

Rank	Start Stations		End Stations	
	<i>Name</i>	<i>Number of Trips</i>	<i>Name</i>	<i>Number of Trips</i>
1	Market at 4th	1990	Market at 4th	2023
2	Harry Bridges Plaza (Ferry Building)	1975	Market at Sansome	1984
3	Embarcadero at Sansome	1934	Harry Bridges Plaza (Ferry Building)	1977
4	Market at Sansome	1930	Embarcadero at Sansome	1892
5	2nd at Townsend	1691	2nd at Townsend	1771
6	Powell Street BART	1677	Powell Street BART	1737
7	2nd at South Park	1378	Steuart at Market	1387
8	Steuart at Market	1363	Townsend at 7th	1386
9	Townsend at 7th	1335	2nd at South Park	1385
10	Commercial at Montgomery	1246	San Francisco Caltrain (Townsend at 4th)	1284

The same analysis was conducted for weekends, though not for specific hours of the day. The results are summarized in the table below.

Table 2. Highest Traffic Start and End Stations During the Weekend

Rank	Start Stations		End Stations	
	<i>Name</i>	<i>Number of Trips</i>	<i>Name</i>	<i>Number of Trips</i>
1	Harry Bridges Plaza (Ferry Building)	3162	Embarcadero at Sansome	3364
2	Embarcadero at Sansome	3108	Harry Bridges Plaza (Ferry Building)	3173
3	Market at 4th	1655	Market at 4th	1866
4	Embarcadero at Bryant	1600	Powell Street BART	1673
5	2nd at Townsend	1536	San Francisco Caltrain (Townsend at 4th)	1646
6	Powell Street BART	1485	2nd at Townsend	1587
7	San Francisco Caltrain (Townsend at 4th)	1360	Embarcadero at Bryant	1380
8	Grant Avenue at Columbus Avenue	1296	Steuart at Market	1215
9	Market at Sansome	1091	Market at Sansome	1106
10	Powell at Post (Union Square)	1080	Grant Avenue at Columbus Avenue	1099

Bike Utilization Analysis

To determine the monthly utilization per bike, the number of bikes available during each month was determined. To achieve this, the number of unique bike IDs was counted for each month. This approach assumes that each functional bike will be used at least once throughout the month. Any bikes that were not used throughout the month were considered out of service. The percent utilization was calculated for each month and visualized in the figure below.

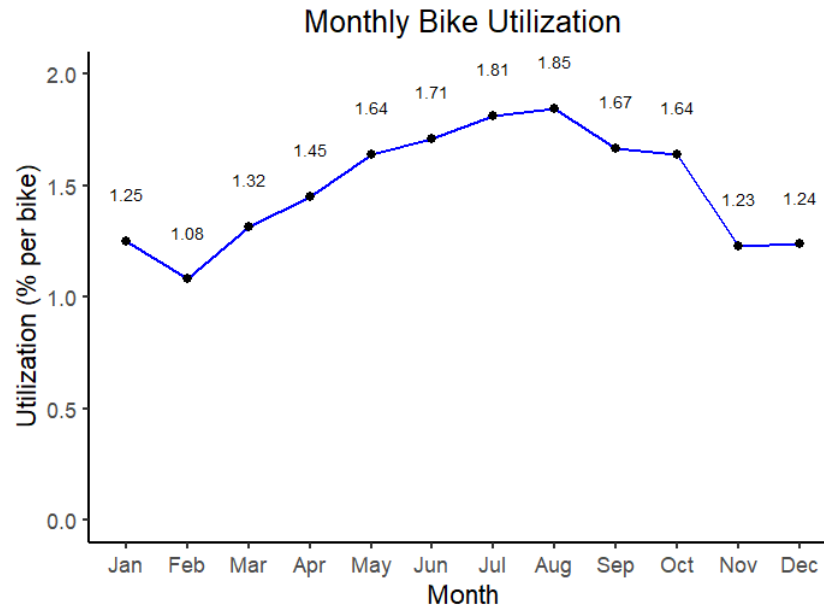


Figure 6. Percent utilization of bikes for each month in 2014.

Figure 6 shows that overall utilization of the bikes is very low (average year-round utilization is 1.49%). This means that on average, each bike is rented out for only 1.49% of the month. The utilization is at its highest in the summer months, and at its lowest during the winter months. The existing fleet of bikes is likely sufficient for meeting trip demand.

Weather Impact Analysis

Correlation analysis was conducted to understand the impact of weather on the number of bike trips occurring and the duration of the trips. The results are summarized below.

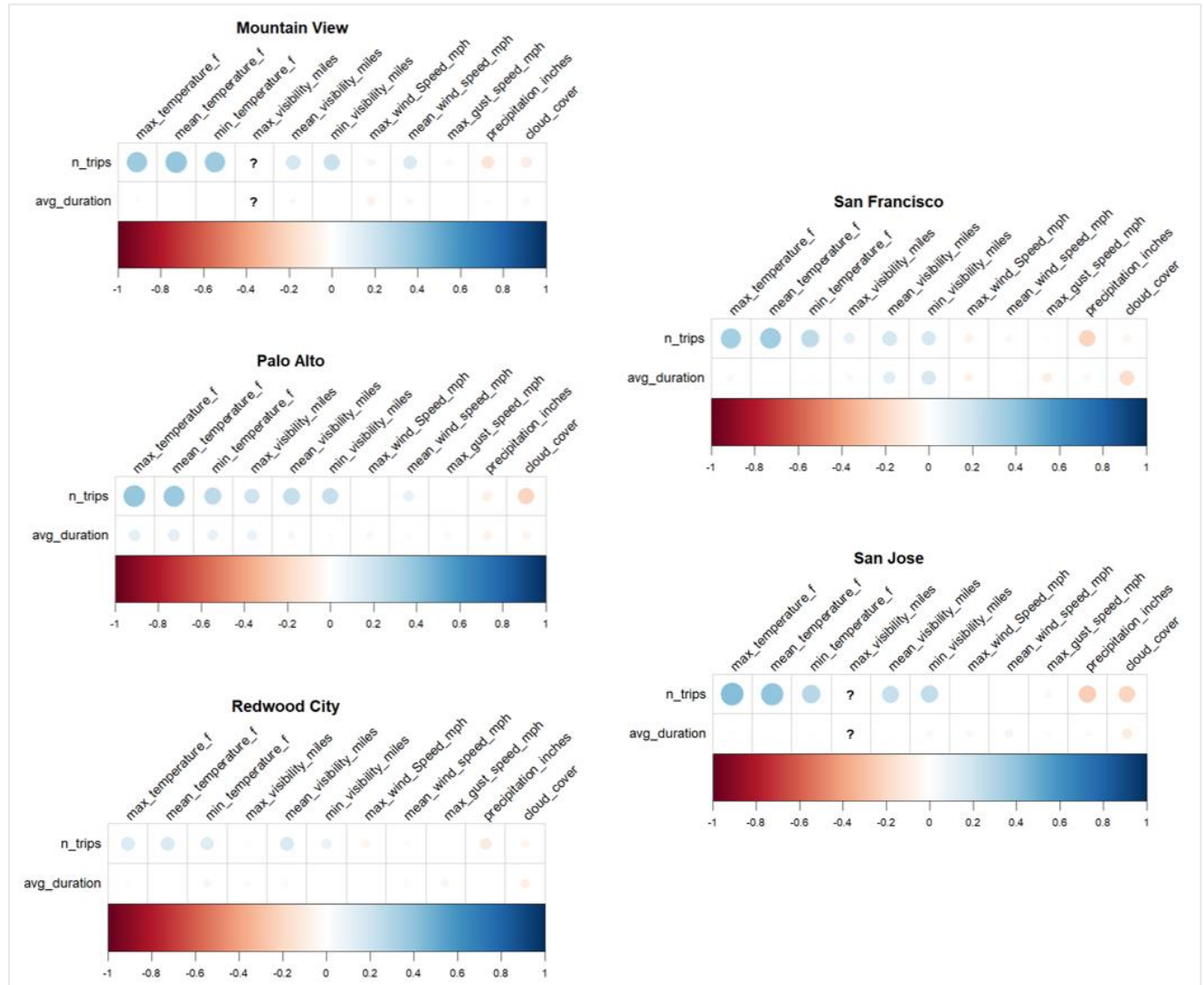


Figure 7. Correlations between bike usage and weather in 5 cities across the Bay Area. One indicates a strong positive correlation, negative one indicates a strong negative correlation, and zero represents no correlation.

Note: question marks (?) indicate cases where the correlation could not be determined because the standard deviation of the variable was equal to zero.

A similar trend is observed across all cities. Higher temperatures and visibility are associated with an increased number of bike trips, while precipitation and cloud cover are associated with a decreased number of bike trips. Interestingly, the weather does not seem to be strongly correlated to the average duration of trips in most cases. Therefore, warmer and clearer weather influence how many trips occur but not the duration of the trips.

Proposed Future Analyses

1. *Number of bike trips in each city per capita*: identify if residents of each city are equally likely to rent bikes.
2. *Weekend rush hours*: identify which hours of the weekend have the highest demand.
3. *Bike utilization for each bike ID*: identify which bikes are rarely used (may need to be serviced).
4. *Weather/trip correlations on weekdays vs weekends*: determine if recreational bikers are more sensitive to weather conditions than commuters.