How Does the Average Temperature and Month Affect Bike Trips From San Francisco Bay Area?

Ruiqian Chen

ECO225 H1: Big-Data Tools for Economists

Professor Nazanin Khazra

April 16, 2022

Introduction

Summer is the most favourable season for outdoor activities, such as swimming, kayaking, and biking. San Francisco Bay Area is undoubtedly one of the hottest places in America to spend your vacations. This study aims to examine how the number of bike trips per day varies according to the temperature and month of the year using the San Francisco Bay Area bike share data from August 2013 to August 2015. We selected the number of trips per day as the response variable to answer our research question. Predictor variables to study concerning the number of trips per day are the average temperature (measured in Fahrenheit) and the month of the start day of the trip. Since warmer temperatures accompany summer and winter generally lead to severe weather conditions, month and temperature can be said to be closely related. From a previous study conducted by Mark Martinez, he focused on examining the effect of weather on daily bike rentals from Citi Bike, a bike-share system in New York City. He ran four multiple regressions, using temperature, precipitation, wind speed, snowfall, and month as explanatory variables. While fixing other variables fixed, every unit higher in precipitation, windspeed, or snowfall adversely impacts the ridership. The temperature exhibited a quadratic relationship with ridership. When the temperature is low, ridership rises as the temperature is one Fahrenheit higher, and when the temperature is high, ridership slowly decreases as the temperature gets higher. The highest ridership occurred from August to October, around summertime, and it is much higher than in January. All models have an R-squared of more than 0.7, indicating that predictors explain 70 percent of the variation in ridership in the model (2017). Another research about bike share in Washington, DC, also focused on the weather impact on daily trips. The temperature was used as a categorical variable with a 10°F increment starting from 10°F to 100°F. The number of trips rises from 10°F (-12.22°C) to less than 90°F (32.22°C) and

diminishes when reaching 90°F (Gebhart & Noland, 2014). Two studies have revealed a non-linear relationship between the temperature and the number of trips, and it seems that a moderate temperature is more suitable for a bike trip. A study of the effect of introducing a single trip bike share fare of $2 in Washington, DC, discovered an increase of 41% in the ridership. This large percentage informed that the single trip fare option significantly elevated the bike share usage (Kaviti et al., 2018).

We will also build linear regression models using temperature and month as the explanatory variables, including weekday and weekend subgroups and monthly oil prices, and examine the distinction between years. The weather dataset we have ranged from 39.2°F (4°C) to 80.6°F (27°C), and it is a continuous variable in the model. In contrast to investigating the impact of global warming or weather conditions, our motivation is to seek how the variation in month and temperature affect the popularity of bike-share usage. We will generate two maps of the average of the response variable for 2013 and 2015 by San Francisco Bay Area zip codes. To find out where the frequency of bike share usage is the most. Visualizations of the variables' distribution and relation will be presented, either grouped by year, weekdays, weekends, or simply ungrouped. Monthly oil price data between 2013 to 2015 in California were gathered through HTML-based web scraping from the Energy Information Administration's official website. We analyzed the relationship between the dependent variable and the newly added data. The conjecture is that the months with higher gasoline price has higher bike share usage. The response and predictor variables are obtained by grouping and merging the raw datasets to record information on a daily trip, weather, and stations.

Data

The data records for daily bike shares rental activities from 2013 to 2015 were obtained from the Kaggle website. A total of four datasets are available: station, status, trip, and weather. Station and status datasets provide information about each station in the bay area, including its name, unique ID, location, and the availability of bikes. There are 70 stations administrated by the San Francisco Bay Area bike share system. The trip dataset holds around 670k observations, a relatively large sample, and records the start and end date of a trip for a particular user. The weather dataset provides weather conditions such as minimum, maximum, and mean of the temperature, dew point, and humidity on a specific date from August 29, 2013, to August 30, 2015. After grouping by date and merging the weather dataset with the trip dataset, we obtained a final dataset with the date as a unit of observation with the number of daily trips, mean temperature, and month as column variables.

Visualization and Summary Statistics

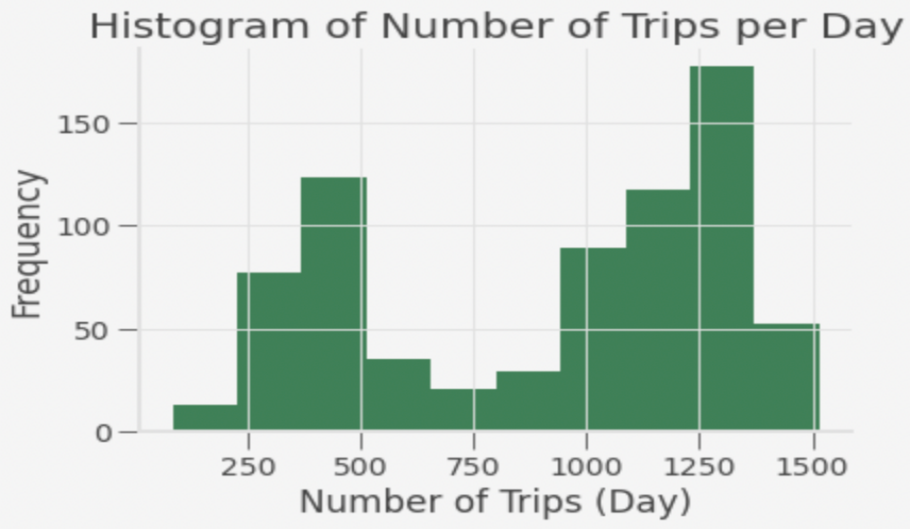
The only dependent variable we want to investigate is the number of trips per day from the San Francisco Bay Area bike share system. It is calculated by grouping the trip dataset by the trip's start date and counting the number of observations for each day. Therefore, the dataset has changed from panel to time series data, date as a unit of observation. As shown in Table 1, the number of trips per day has a mean of around 914, with a range from 81 to 1516. The other two variables are the predictors temperature and month, and each has the same number of observations as the number of trips per day. The mean average temperature is about 61.4°F, with a day reaching the maximum of 80.4 °F. The month has a mean of 6.53, which indicates that more people rode a bike after June.

Table 1. Summary table of number of daily trips, average temperature, and month

	Number of trips	Average temperature in Fahrenheit	Month
count	733.000000	733.000000	733.000000
mean	913.995907	61.351978	6.532060
std	402.875099	7.057318	3.444425
min	81.000000	39.200000	1.000000
25%	451.000000	56.600000	4.000000
50%	1046.000000	61.200000	7.000000
75%	1266.000000	67.200000	10.000000
max	1516.000000	80.400000	12.000000

Based on the histogram in Figure 1, the bike share has been used approximately 1300 times daily for 175 days. It is bimodal, which implies there are two classes with different characteristics. The values are in the range from about 100 to somewhere around 1500.

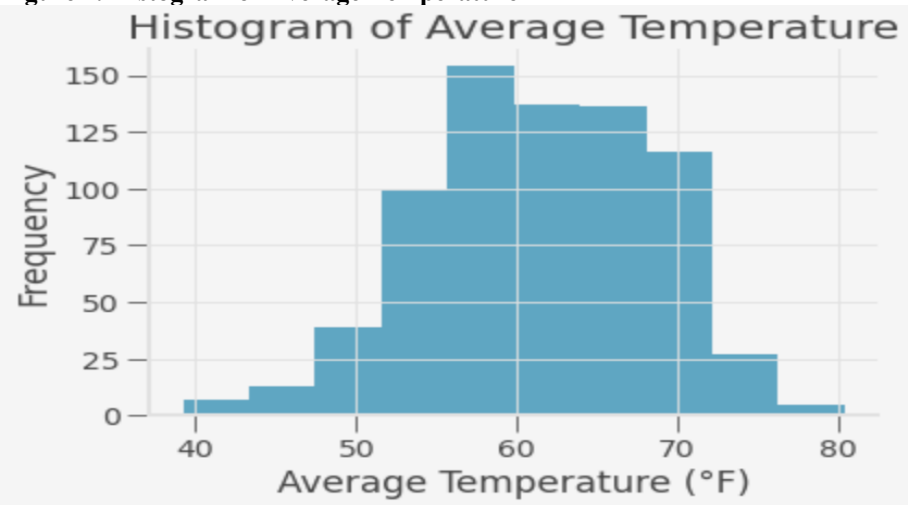
Figure 1. Histogram of Number of Trips per Day



The average temperature distribution is unimodal and approximately symmetric, with the highest frequency occurring around 55°F. Most of the time, the San Francisco Bay Area, from

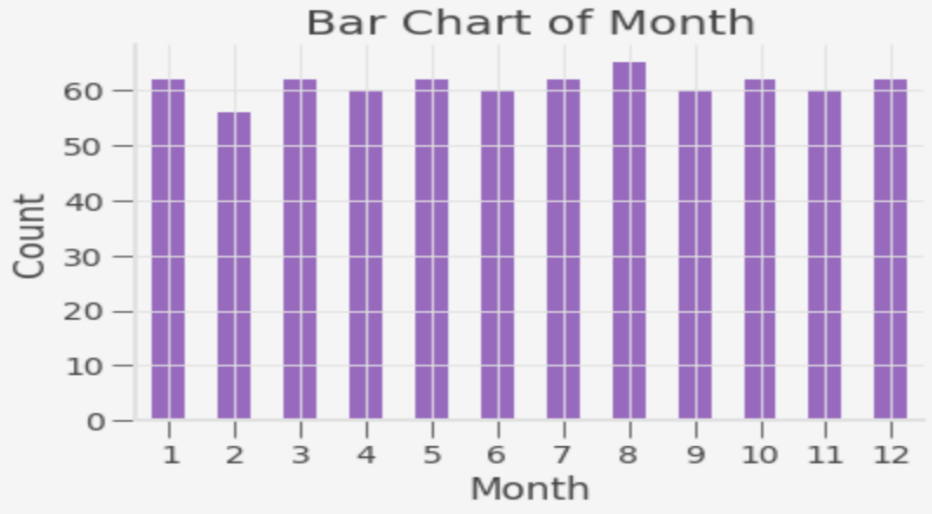
2013 to 2015, had a temperature above 55°F and below 70°F, which was a relatively mild climate.

Figure 2. Histogram of Average Temperature



As we can see in the bar chart from Figure 3, the frequency of each month is roughly the same. Some months have less value for the count, and this may be due to no rental activities for bike share or the absence of data for some specific days.

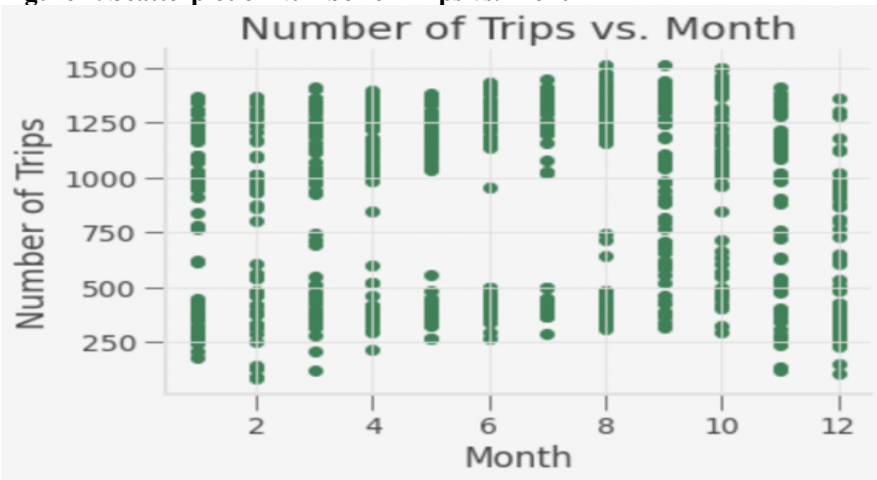
Figure 3. Bar Chart of Month



From Figure 4, There is no clear relationship between the number of trips and the twelve months. Rentals of bikes were, on average, more likely from August to October than from

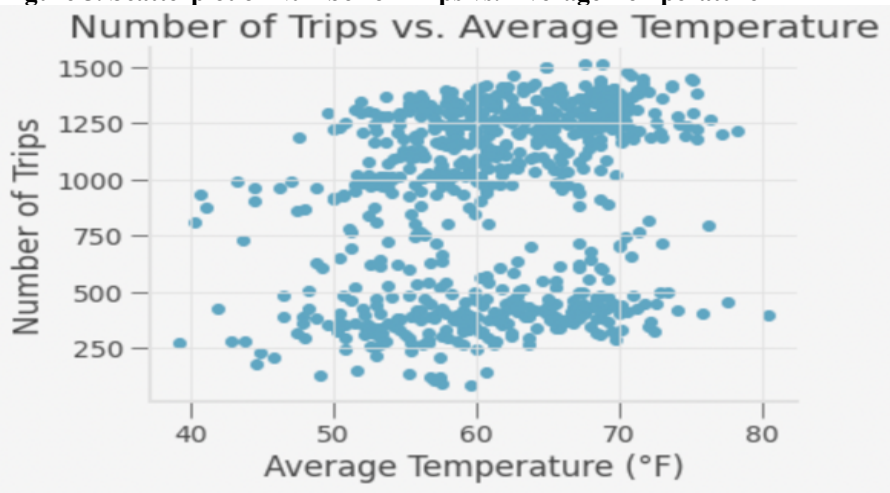
November to February. The month and number of bike trips have a quadratic relationship, but we cannot conclude since the month is classified as a categorical variable.

Figure 4. Scatterplot of Number of Trips vs. Month



The scatterplot of the number of trips versus average temperature in Figure 5 shows two clusters, each exhibiting a linear trend. The clusters imply that there may be some confounding variables causing a significant difference between the number of daily trips of the two groups. People generally prefer bike trips on days with moderate or warmer weather over days with cold weather.

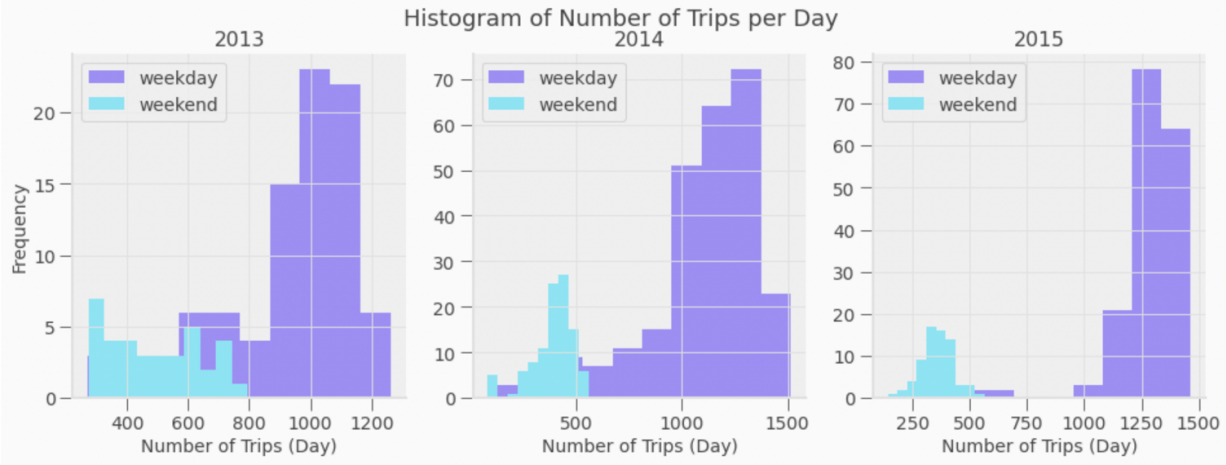
Figure 5. Scatterplot of Number of Trips vs. Average Temperature



People travel much more on weekdays than on weekends, observing the histogram in Figure 6. Especially in 2015, there was a significant difference in the number of trips per day

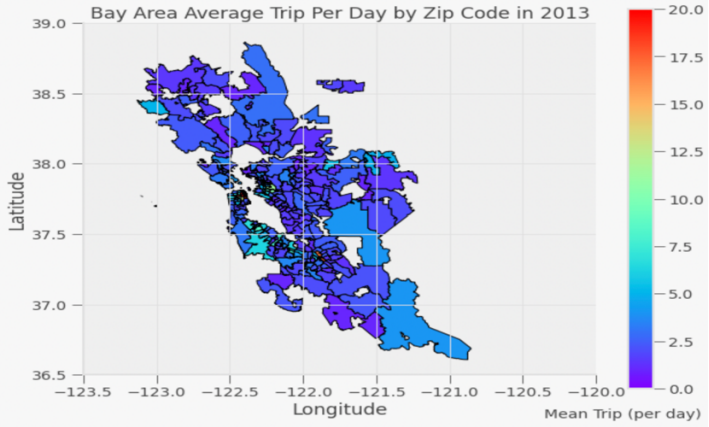
between the two groups. In addition, bike share usage increased gradually from 2013 to 2015. Observations indicate weekday is in purple and weekend is in blue.

Figure 6. Histograms of Number of Trips per Day from 2013-2015



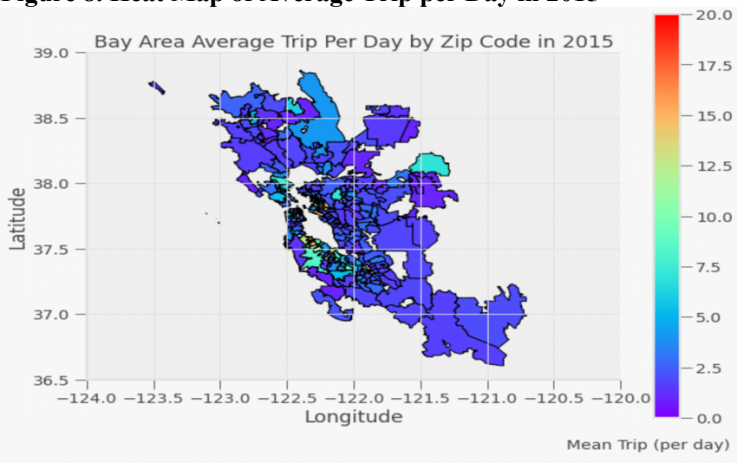
The heat map in Figure 7. display the average number of trips per day grouped by zip code in the San Francisco Bay Area in 2013. Seven observations with mean trips per day higher than 20 are excluded for us to focus on the more common values. Most bike-share usage is in the small bright red portion of the map, which has a value of 18.1 and a zip code of 94117, located in San Francisco. In contrast, dark purple areas depict the least amount of bike rentals. In this case, the minimum value is 1. A total of 18 zip codes have a value of 1, including 94037, located in Montara. This map is created using the Bay Area zip codes data downloaded from San Francisco Open Data website. Even if we exclude the average trip above 20, the map still appears blue or purple. Among all zip codes, the mean daily trip in 2013 was less than 5. Large areas tend to contain fewer records, while the location shown in red is smaller but has an average of around 20 bike share usage. The highest observation from the excluded rows is 88.3, with a zip code of 94107, located in San Francisco.

Figure 7. Heat Map of Average Trip per Day in 2013



On the Bay Area map for 2015, the average bike trip per day appears more purple in areas with fewer bike rental activities in 2013. A slight increase in bike share usage occurred in areas with mean trips over five per day in 2013. A limit of 20 excludes six observations having an average bike trip per day higher than 20, and the highest observation is also in San Francisco with zip code 94107, which has a mean of 130.1.

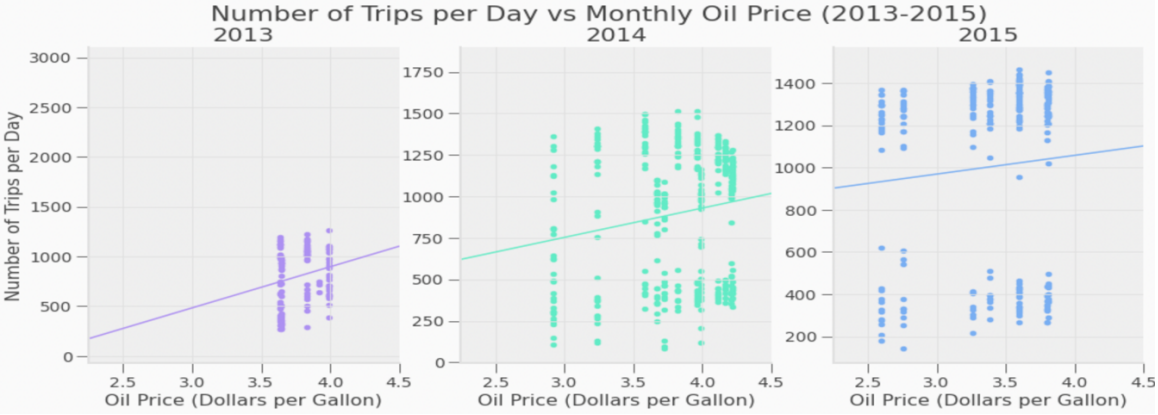
Figure 8. Heat Map of Average Trip per Day in 2015



The three scatterplots exhibit relationship between trips per day and oil prices by month from 2013 to 2015. In 2013, the oil prices were high for each month on average, ranging from 3.60 to 4.00 U.S. dollars per gallon. In contrast, in 2014, oil prices were in a broader range from about $2.75 to $4.25. Observations of bike share usage clustered when the oil price was around $4.00, and in 2015 the cost for some months reached below $3.00. Thus on average, oil prices

were lower compared to other years. The effect of oil price was reflected in a flatter slope of the regression line, indicating that the oil price fewer influences bike share use.

Figure 9. Scatterplots of Number of Trips per Day vs. Monthly Oil Price from 2013-2015



Regression Results

According to Table 2, including four regression models, each has a different set of predictors. Hence, we need to compare them using the adjusted R-squared. Their adjusted R-squared is around 0.75, so the model explains the response well. For example, the first model has a 73% variation in bike trips per day explained by temperature and weekday/weekend. Model 1 has the largest F-statistic of 969.03, which indicates that this model is highly statistically significant overall. All four models have AIC and BIC above 9000, almost 10,000. Lower AIC and BIC suggest that it is preferable to the other three models. With variables oil price added to the fourth model, BIC does penalize the model. Therefore, model 3 is preferable among the four since it has a higher adjusted R-squared, a large enough F-statistic, smallest AIC and BIC. Most of the slope coefficients are statistically significant at a 1% significance level. In February, the count for bike trips does not statistically differ from December if holding temperature, year, weekday, and oil price fixed. On the other hand, after controlling temperature, month, year, and weekday, the oil price is not statistically significant in explaining the response.

Days with one Fahrenheit higher temperature have daily bike trips on average eight more, controlling for the month, year, and week of the day. Summer months like June have the highest number of bike trips per day, an average of 190 more than in December. All positive coefficients indicate that December is the least common month for a bike trip, after controlling whether travellers use the bike share on weekdays or weekends, year, and temperature. After fixing the year, month, and temperature, weekday bike share usage was approximately 726 per day, more on average than on the weekend. As we observed from the summary statistic table, the maximum number of trips per day is 1516. Thus, this slope coefficient is highly statistically significant and highly economically significant. The year also contributes to explaining the response. The bike-share system becomes more popular with each passing year, holding temperature, month, and weekday fixed.

Table 2. Ordinary Least Squares (OLS) Regressions

OLS Regressions

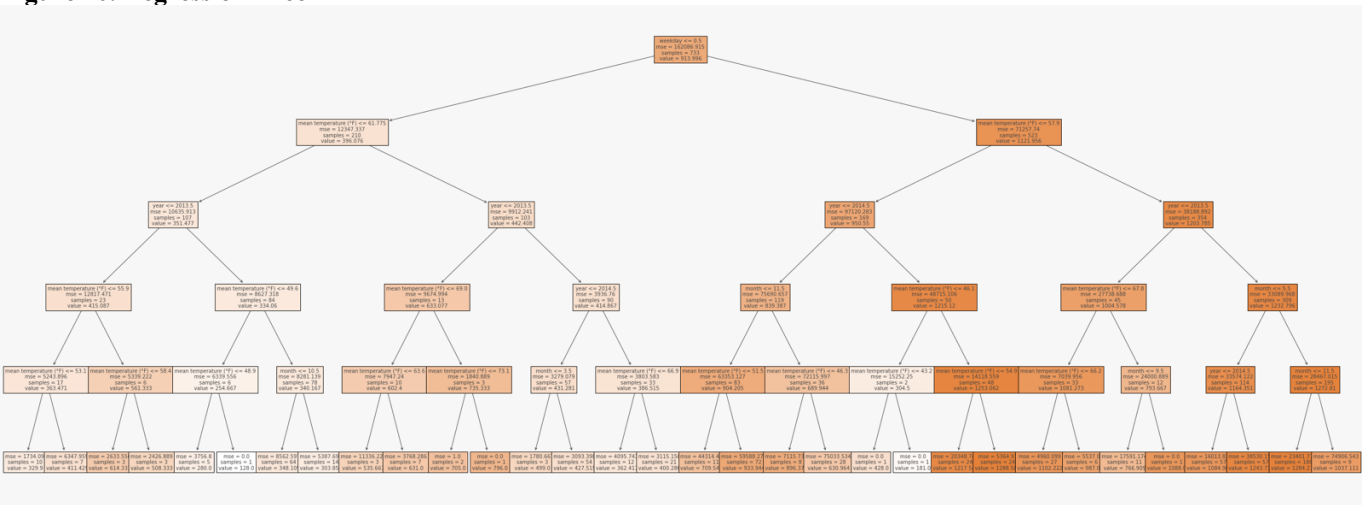
	Model 1	Model 2	Model 3	Model 4
const	-474.62***	-342.26***	-317.07***	-201.68
	(69.30)	(105.77)	(103.27)	(223.52)
mean_temperature_f	14.20***	8.79***	7.55***	7.42***
	(1.11)	(1.98)	(1.98)	(2.00)
Jan		194.41***	97.58**	108.71**
		(36.69)	(38.44)	(42.94)
Feb		141.67***	48.96	63.98
		(38.63)	(39.71)	(47.37)
Mar		220.81***	132.54***	162.05**
		(39.96)	(40.54)	(64.91)
Apr		243.21***	155.19***	186.28***
		(40.39)	(40.91)	(67.28)
May		227.96***	142.83***	182.93**
		(42.21)	(42.40)	(80.90)
Jun		270.51***	190.46***	226.90***
		(47.00)	(46.64)	(78.08)
Jul		223.58***	147.11***	186.50**
		(50.51)	(49.85)	(84.05)
Aug		243.05***	173.31***	206.18***
		(50.66)	(49.85)	(75.34)
Sep		175.62***	195.58***	217.36***
		(48.74)	(47.86)	(60.76)
Oct		269.46***	283.85***	298.81***
		(43.20)	(42.17)	(49.41)
Nov		174.81***	181.39***	187.06***
		(38.39)	(37.17)	(38.44)
year14			77.85***	66.87**
			(25.61)	(31.81)
year15			196.59***	165.30***
			(30.81)	(61.97)
weekday	724.98***	725.67***	725.52***	725.56***
	(17.24)	(16.68)	(16.10)	(16.10)
oil price				-31.53
				(54.15)
R-squared	0.73	0.75	0.77	0.77
R-squared Adj.	0.73	0.74	0.76	0.76
F-statistic	969.03	164.54	156.71	146.80
AIC	9929.13	9889.60	9839.67	9841.32
BIC	9942.93	9953.96	9913.22	9919.47
No. observations	733	733	733	733

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

The regression tree includes four variables, temperature, weekday, year, and month. The maximum depth of the tree is five. Thus there are 32 (2^5) terminal nodes. Most terminal nodes

include relatively small sample sizes. Increasing the tree size may encounter the issue of over-fitting. Weekday observations are predicted to have more bike trips per day, having a sample of 523. Suppose an observation is recorded on the weekend, having a temperature lower or equal to 53.1°F in 2013. In that case, it is predicted to be approximately 330 bike rental activities on that day with a mean square error o 1734.09. Many high mean square error above 1,000 indicates that the bias in the tree is significant. Since the larger the sample size, the more considerable the variation of the predicted value to the actual measurement. The tree predicts that the most number of daily trips to be on the weekday, with an average temperature between 54.9°F and 57.9°F in 2015, which we cannot observe from the linear regression.

Figure 10. Regression Tree



Conclusion

From August 2013 to August 2015, the average number of bike trips per day in the San Francisco Bay Area was 914. The number of bike rentals increased after June. We found that, on average, bike rental is higher from August to October. November to February are the coldest months, so bike rental activities are less frequent. The average temperature illustrates the number of trips differently compared to the month. When the weather is pleasant, people have high

incentives to go for a bike ride. Much fewer individuals rent a bike in extreme weather such as below 50°F (10°C) or above 80°F (27°C). These two independent variables used to explain bike trips are correlated, which is reasonable since the day's temperature depends on the seasons. Separating weekday and weekend observations, we observe that the use of bikes shares significantly increased during weekdays. Bicycles are commonly used as a form of transportation on weekdays.

The bike-share usage on weekdays and weekends may differ by the habit of weekday travellers and weekend travellers. The weekend is generally a rest time, so people rarely go out on weekends but prefer to go to work during the week, resulting in increased usage of bike shares. Despite the significant difference in bike share usage between weekdays and weekends, the number of daily trips on weekdays increased in 2015. Both groups observed a positive linear relationship between our dependent variable and the average temperature.

When we look more in-depth, observing the distribution of bike-sharing activities in 2013 and 2015 by zip codes informed us about the changes in the use of bikes and other details such as population. Most of the areas have an average annual bike rental of less than five, which indicates that there are zip code areas that add up to give us a maximum daily trip of 1516, from the summary statistics. Indeed, even some places are small but show a significantly high demand for bike shares. The most frequent bike rentals occur in San Francisco, implying that the bike share system is a more popular and well-known transportation method. Even though commute time for people in the Bay Area has been long since 2010, the driving rate in San Francisco has declined prominently (Castañeda, 2020). From 2013 to 2014, the San Francisco Bay Area

population rose each year. The population was 3,298,000 in 2013, 3,304,000 in 2014, and 3,309,000. The annual increase in population was equal to or greater than 0.15%, a relatively small number but could contribute slightly to the rise of bike rentals (San Francisco Metro Area Population 1950-2022, n.d.).

Oil price could be a variable outside our data that influences bike-share usage. According to the gasoline price data from the U.S. Energy Information Administration, average oil prices in U.S. dollars per gallon decreased from 2013 to 2015. As the oil price rises, bike-share systems are used more often. In 2013, it was shown as a steeper slope than in other years between the relationship between daily trips and oil prices. The slope became flattered encountered by the decline in oil prices in 2015. Therefore, prices drove bike rentals less when cheaper, and vice versa.

The most suitable model to explain the variation in daily bike trips includes month, year, temperature, and weekday as predictors. After controlling for the month, year, and week of the day, a temperature of one Fahrenheit higher is associated with approximately eight extra bike trips per day. This is reasonable since people often go out for outdoor activities during warmer, mild climates rather than colder or extreme weather. Month and temperature are correlated, but after fixing the temperature, September and October have the highest number of bike trips per day, about 200 or more compared to December. People might have more vacations during these two months. The popularity of each year is also distinct. In 2015 the number of bike trips was, on average, 197 higher per day than in 2013. It is both statistically significant and economically significant. Weekday and weekend subgroups significantly differ on average weekday bike share

usage was 726 per day more than on the weekend. We can infer that people from San Francisco Bay Area prefer to stay at home during weekends and are more active on weekdays.

We discovered that weekday bike rental activities were significantly higher than on weekends. They may have different properties and can be observed as two distinguished subgroups. In the future, we want to add more variables to explain better the frequency of bike trips in the San Francisco Bay Area. Our current models have some lurking variables such as the air quality, the number of bus stations, and whether there is a shopping mall beside the bike share station. Adding these additional variables helps to explain the response better. Also, regression models can include some interaction terms, for example, year and temperature or month and temperature.

In conclusion to our research, how often people use bike-share at the San Francisco Bay Area bike share does influence by the month and weather of the day. Bike-share rental activities are greatly affected by the day of the week and moderately influenced by the month or the temperature. San Francisco County is the most popular bike-share system, having a zip code area that reached an average number of trips per day of 130 in 2015. The month variable is also related to the gasoline price, and their relationship further induced our dependent variable to fluctuate.

References

- Bay Area ZIP codes: Datasf: City and county of San Francisco. San Francisco Data. (n.d.). Retrieved March 22, 2022, from https://data.sfgov.org/Geographic-Locations-and-Boundaries/Bay-Area-ZIP-Codes/u5j3-svi6

- California all grades all formulations retail gasoline prices (dollars per gallon). (n.d.). Retrieved March 21, 2022, from https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=emm_epm0_pte_sca_dpg&f=m
- Martinez, Mark. "The impact weather has on NYC Citi Bike share company activity." Journal of Environmental and Resource Economics at Colby 4, no. 1 (2017): 12.
- Gebhart, Kyle, and Robert B. Noland. "The impact of weather conditions on bikeshare trips in Washington, DC." Transportation 41, no. 6 (2014): 1205-1225.
- Kaviti, Shruthi, Mohan M. Venigalla, Shanjiang Zhu, Kimberly Lucas, and Stefanie Brodie. "Impact of pricing and transit disruptions on bikeshare ridership and revenue." Transportation 47, no. 2 (2020): 641-662.
- San Francisco Metro Area Population 1950-2022. MacroTrends. (n.d.). Retrieved March 24, 2022, from https://www.macrotrends.net/cities/23130/san-francisco/population#:~:text=The%20current%20metro%20area%20population,a%200.15%25%20increase%20from%202021.
- Castañeda, L. (2020, January 11). How the Bay Area has changed since 2010. The Mercury News. Retrieved March 23, 2022, from https://www.mercurynews.com/2020/01/09/how-bay-area-demographics-have-changed-since-2010/