Capital Bikeshare Analysis:
Bike Rental Demand from 2011 through 2012

By:
Chandini Ramesh

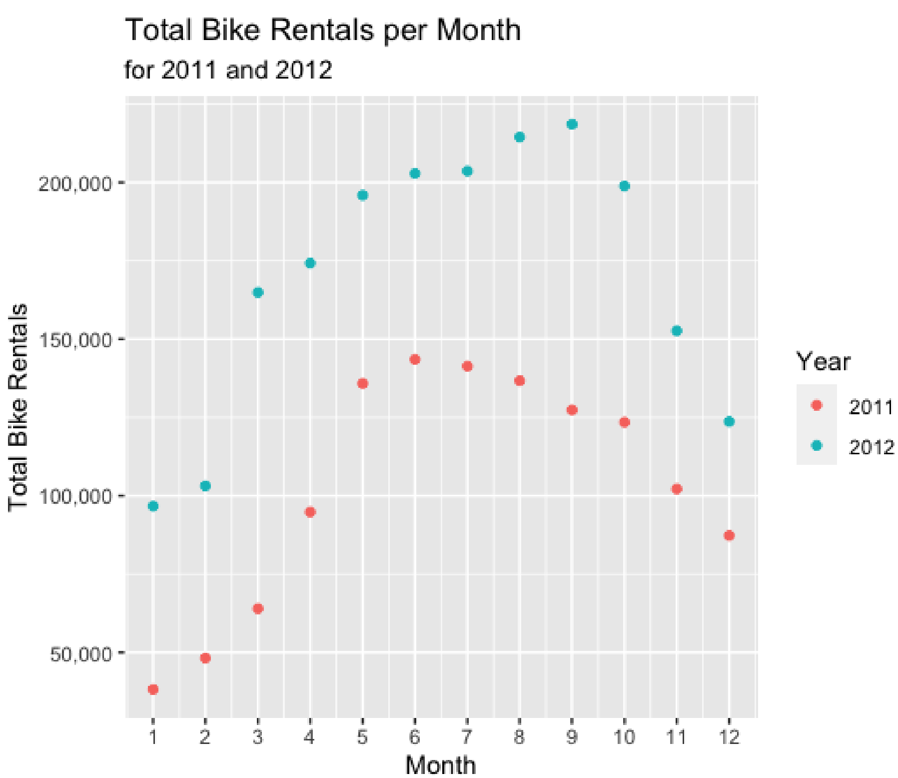
October 7, 2022
Georgetown University
Business Analytics
OPIM 601

Capital Bikeshare Analysis

The objective of this analysis is to use data related to Washington DC's bikeshare program to derive key insights and help Capital Bikeshare understand what factors impact the demand for bikes. The data provided shows bikeshare activity from January 2011 through December 2012, and details key factors like when the bike was rented, what the weather was like, and whether the person renting the bike was a registered or casual user. Using the dataset provided, we can gather key insights about when the demand for bikes increases and when it wanes.

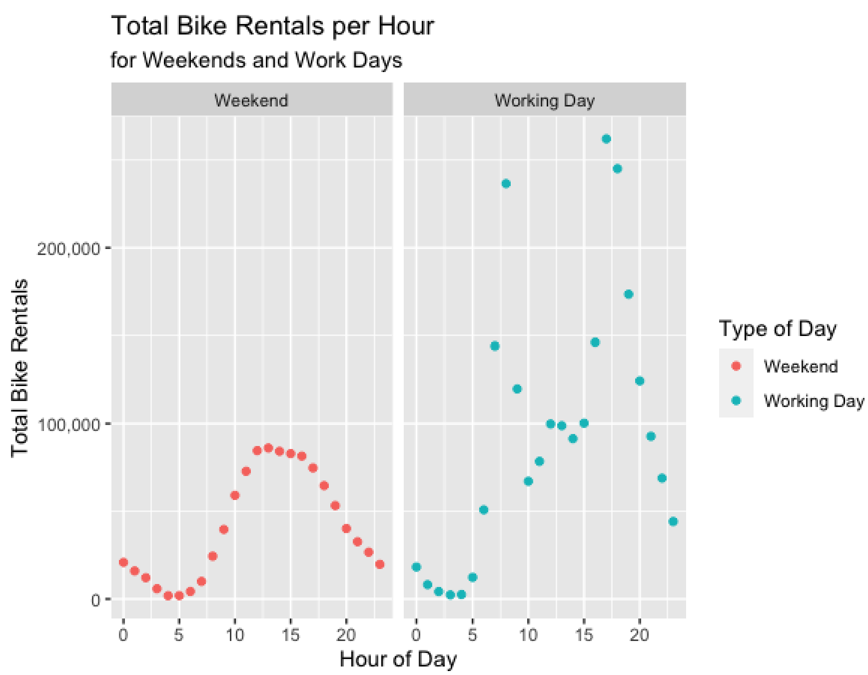
I started by doing an exploratory analysis of the data, to understand the dataset better and clean up any anomalies. One of the first items I checked for was if there was any missing data. In this case, there was none, so we could proceed without addressing this concern. Then, to see if there was any noticeable trend in bike demand over months in the year (Figure 1), I plotted the months versus the total number of bike rentals for both 2011 and 2012. From this plot we can see that for both years, there is a peak in the demand for bikes particularly between May and September. There was also many more bike rentals in 2012 compared to 2011. One reason we could be seeing this trend in the months is due to the warmer weather. But we will dig deeper to see if that is in fact the case. If this is true, one way Capital Bikeshare could use this information is to invest in more bikes to be available during the summer months. This would help increase revenue during summer and decrease costs in the winter.

Figure 1: <u>Total Bike Rentals per Month, for 2011 and 2012</u>



Months in the year have an impact on the number of bikes being rented. But what other factors can affect demand? Many people commute to work using bikes, and the team at Capital bikeshare are hypothesizing that bike rentals are higher during these office commute times. By plotting this we can see that this hypothesis is true (Figure 2). Bike rentals peak from 5-10 am and then again from 3-8 pm. These timings coincide well with the typical work or school commute timings. I confirmed this further by comparing the number of bike rentals on the weekday vs weekends. This showed that the number of bike rentals on working days was greater than that of weekends during these commute times, confirming Capital Bikeshare's assumption that people use these bicycles to commute to work. Using this data, Capital Bikeshare could offer special incentives to encourage people to use bike rentals programs more on the weekends. They could create special weekend events, or offer discounts in times of less demand, as a way to encourage more rentals.

Figure 2: Total Bike Rentals per Hour for Weekends and Weekdays



Now that we have examined how bike rentals are affected by month and time of day, let's look at season to further confirm our hypothesis regarding people being more likely to bike in warmer weather. In Figure 3, we can see that most bike rentals took place in the Summer and Spring over Fall and Winter. This is further evidence of the theory that weather impacts bike trends. We also have a temperature variable indicator. Plotting this variable against the average number of bike rentals (Figure 4), we see that the number of bike rentals increases as the temperature increases for both registered and casual users. Using this visualization, as well as the season and month visualizations we saw earlier, we can see that weather does seem to influence the bike rentals. But we can go deeper by running a linear regression.

Figure 3: Average Bike Rentals per Season

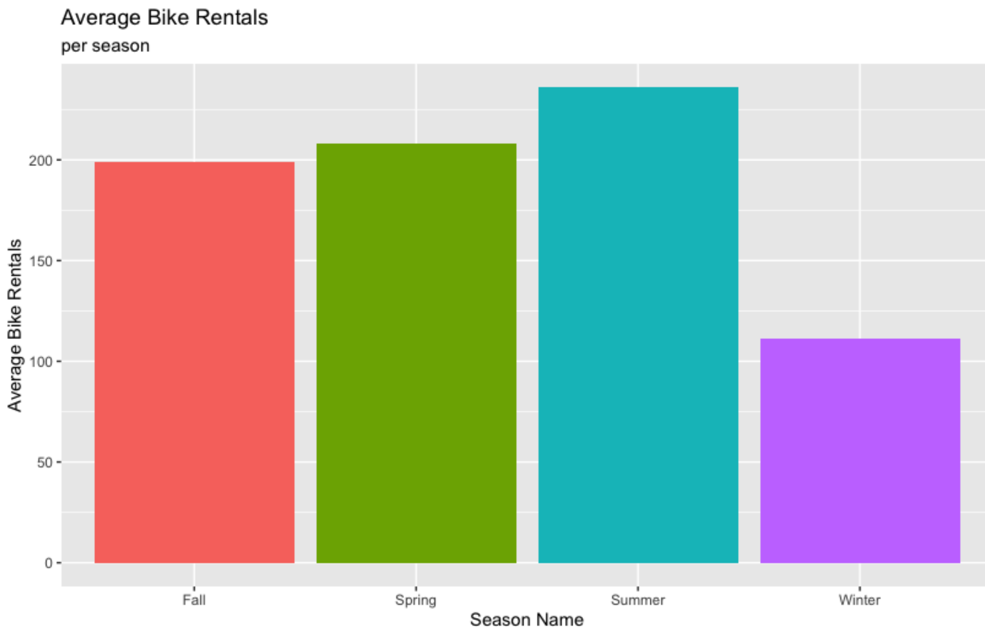
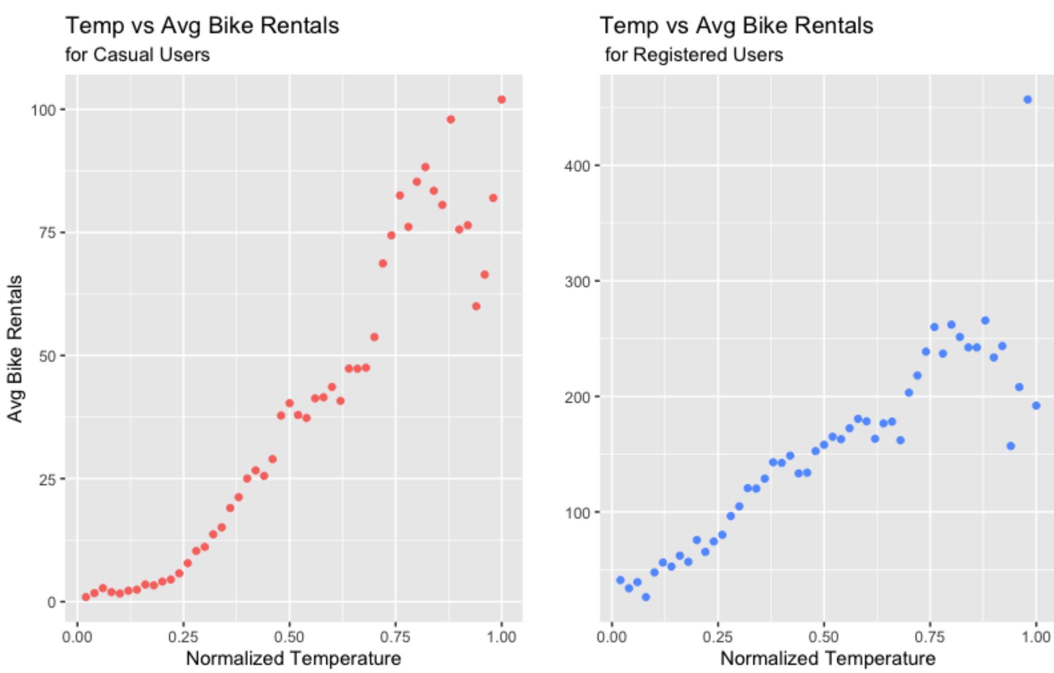


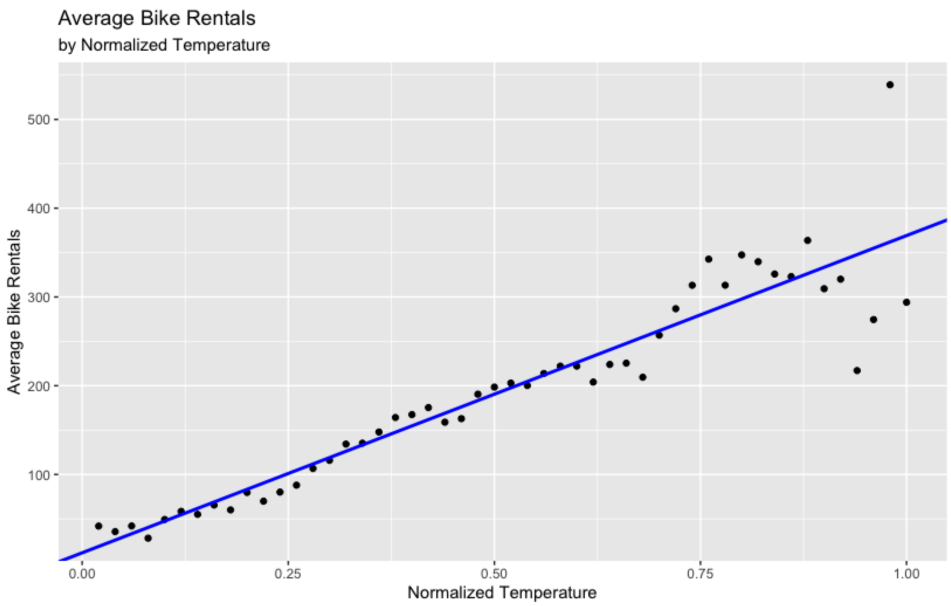
Figure 4: Average Bike Rentals per Temperature, for Casual and Registered Users



First, I ran a linear regression using actual temperature and total bike rentals. For this linear regression I found that there was a r-squared value of 0.8698 and the slope and intercept were 11.7 and 357.4 (y=11.7x+357.4). We can interrupt the slope to say that for every increase in the normalized actual temperature unit, the average number of bike rentals increases by 11.7. The intercept tells us that the predicted number of average bike rentals is 357.4 when the normalized actual temperature is 0. The r-squared value indicates a strong linear relationship between the temperature and the average number of bike rentals. I also compared the daily feeling temperature with the average number bike rentals to see if that would be a better indicator. For this regression, the r-squared value was 0.6523, the slope was 37.23 and the intercept was 278.65 (y=37.23+278.65). Again, we can interrupt the slope to say that for every increase in the normalized feeling temperature unit, the average number of bike rentals increases by 37.3. The intercept tells us that the predicted number of average bike rentals is 357.4 when the normalized feeling temperature is 0.

A larger positive r-squared value indicates a stronger linear relationship, so in this case the temperature was a better indicator of bike rentals that the feeling temperature. Figure 5 shows the plot of both temperature and feeling temperature with total bike rentals, along with the linear regression line. The results of the linear regression makes sense with what we saw in Figure 4, as the average number of bike rentals increases with the temperature.

Figure 5: Average Bike Rentals per Temperature, for Actual Temperature and Feeling Temperature



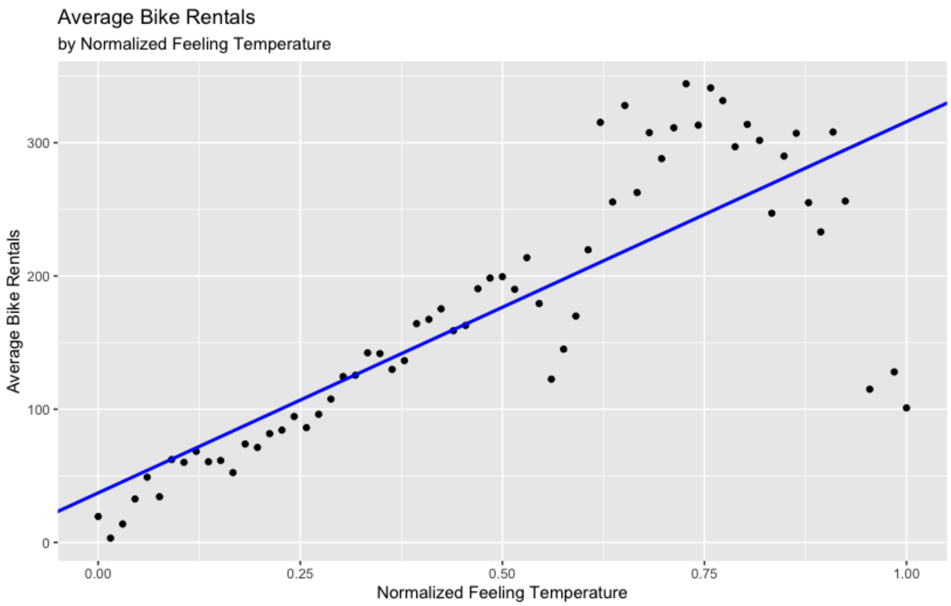


Figure 5: Results of Linear Regression, Actual Temp and Feeling Temp vs Avg Bike Rentals

	Normalized Temp vs Avg	Feeling Temp vs Avg
	357.397***	
		278.648***
	11.766	37.234**
R2	0.870	0.652
Adjusted R2	0.867	0.647
Residual Std. Error	40.728	59.156
Note:		*p<0.1; **p<0.05; ***p<0.01

Overall, from this data we can see that bike rentals really soar in the warmer months, with most bike riders tending to use bikes to commute to and from work over leisure activity. Knowing the trends of their users, Capital Bikeshare can choose to do many things with this information. For example, they can increase the number of bikes available in peak timings. Beyond that, they can also use this information to focus their marketing campaigns and promotions. They could choose to offer incentives for people to ride bikes more often during the non-peak timings, encouraging higher usage. Knowing that a lot of users choose to use bikes for commuting to and from work, Capital Bikeshare can offer incentives to people who do that, and encourage more people to do it. Not only would that result in more business for the company, it would also be a great environmental initiative.

Appendix

```
library(tidyverse)
library(ggplot2)
library(dplyr)
library(lubridate)
library(patchwork)
library(stargazer)

setwd("~/Documents/Documents/MSBA/Stats")
bikeshareData = read.csv("Capital Bike Sharing data by hour.csv")

#Question 1 What is the trend in overall bike demand over the months of the year?
bikeshareData=bikeshareData %>%
  mutate(ShareDate = as.Date(dteday, "%Y-%m-%d"))

str(bikeshareData)
anyNA(bikeshareData)

bikeshareQ1 = bikeshareData %>%
  mutate(month = month(ShareDate), year = year(ShareDate)) %>%
  group_by(month, year) %>%
  summarise(total = sum(cnt)) %>%
  arrange(year, month)

bikeshareQ1  %>%
  ggplot(aes(x = month,
             y = total, color = factor(year))) +
  geom_point() +
  scale_x_continuous(n.breaks = 12) +
  labs(x="Month",
       y="Total Bike Rentals",
       fill="",
       title = "Total Bike Rentals per Month",
       subtitle = "for 2011 and 2012") +
  scale_y_continuous(labels = scales::comma) +
  scale_color_discrete(name="Year",
                       breaks=c(2011, 2012),
                       labels=c("2011", "2012"))



#Question 2 The data science group at Capital bike share hypothesize that a.
#There must be high demand during office timings. Early morning and late
#evening can have different trends (cyclist) and low demand from 10:00pm to
#4:00am. Do you agree? b. Registered users demand more bike on weekdays compared
#to the weekend or holiday. Do you agree?

#a
bikeshareQ2a = bikeshareData %>%
  group_by(hr, workingday) %>%
  summarise(avg = mean(cnt))

bikeshareQ2a$WorkingDayText = ifelse(bikeshareQ2a$workingday==1, "Working Day", "Weekend")

bikeshareQ2a  %>%
  ggplot(aes(x = hr,
             y = avg, color=WorkingDayText)) +
  geom_point() +
  facet_wrap(~WorkingDayText,
             scales="free_x") +
  labs(x="Hour of Day",
```

```r
       y="Average Bike Rentals",
       fill="",
       title = "Average Bike Rentals per Hour",
       subtitle = "for Weekends and Work Days") +
  scale_y_continuous(labels = scales::comma) +
  scale_color_discrete(name="Type of Day")

#b
bikeshareQ2b = bikeshareData %>%
  group_by(workingday) %>%
  summarise(avgReg = mean(registered))

bikeshareQ2b$WorkingDayText = ifelse(bikeshareQ2b$workingday==1, "Working Day", "Weekend")

bikeshareQ2b  %>%
  ggplot(aes(x = as.factor(WorkingDayText),
             y = avgReg)) +
  geom_bar(stat='identity') +
  labs(x="Working Day",
       y="Total Bike Rentals",
       fill="",
       title = "Total Bike Rentals",
       subtitle = "Working Days vs Weekends (where 1=working day and 0=weekend)")+
  scale_y_continuous(labels = scales::comma)

#Question 3Is there any relationship between season and bike rental?
#Create a visualization displaying the relationship.

bikeshareQ3 = bikeshareData %>%
  group_by(season) %>%
  summarise(avg = mean(cnt))

bikeshareQ3$SeasonName[1]="Winter"
bikeshareQ3$SeasonName[2]="Spring"
bikeshareQ3$SeasonName[3]="Summer"
bikeshareQ3$SeasonName[4]="Fall"

bikeshareQ3  %>%
  ggplot(aes(x = SeasonName,
             y = avg, fill=SeasonName)) +
  geom_bar(stat='identity') +
  labs(x="Season Name",
       y="Average Bike Rentals",
       fill="",
       title = "Average Bike Rentals",
       subtitle = "per season") +
  scale_y_continuous(labels = scales::comma)+
  theme(legend.position="none")


#Question 4 What type of relationship do you see between weather and bike rental?
#Is the relationship the same for registered vs. casual users?

#Temp vs Total
bikeshareData %>%
  group_by(temp) %>%
  summarise(avg = mean(cnt)) %>%
  ggplot(aes(x = temp,
             y = avg)) +
  geom_point() +
  labs(x="Normalized Temperature",
       y="Average Bike Rentals",
       fill="",
       title = "Average Bike Rentals",
       subtitle = "by Normalized Temperature") +
```

```
  scale_y_continuous(labels = scales::comma) +
  geom_abline(slope=bikeShareTempModel$coefficients[2],
              intercept=bikeShareTempModel$coefficients[1],
              color="blue",
              size=1)

#Feeling Temp vs Total
bikeshareData %>%
  group_by(atemp) %>%
  summarise(avg = mean(cnt)) %>%
  ggplot(aes(x = atemp,
             y = avg)) +
  geom_point() +
  labs(x="Feeling Temperature",
       y="Average Bike Rentals",
       fill="",
       title = "Average Bike Rentals",
       subtitle = "by Feeling Temperature") +
  #theme_clean() +
  scale_y_continuous(labels = scales::comma) +
  geom_abline(slope=bikeShareAtempModel$coefficients[2],
              intercept=bikeShareAtempModel$coefficients[1],
              color="blue",
              size=1)

#casual users
p1 = bikeshareData %>%
  group_by(temp) %>%
  summarise(avg  = mean(casual)) %>%
    ggplot(aes(x = temp,
               y = avg)) +
      geom_point(color="#F8766D") +
  labs(x="Normalized Temperature",
       y="Avg Bike Rentals",
       fill="",
       title = "Temp vs Avg Bike Rentals",
       subtitle = "for Casual Users") +
  scale_y_continuous(labels = scales::comma)

#registered users
p2 = bikeshareData %>%
  group_by(temp) %>%
  summarise(avg = mean(registered)) %>%
  ggplot(aes(x = temp,
             y = avg)) +
  geom_point(color="#619CFF") +
  labs(x="Normalized Temperature",
       y="",
       fill="",
       title = "Temp vs Avg Bike Rentals",
       subtitle = " for Registered Users") +
  scale_y_continuous(labels = scales::comma)

p1 + p2

#Question 5 Fit a linear model predicting the total bike rental demand from daily
#temperature. What kind of insights can you generate? (make sure to write the
#linear model and interpret it in the context of the data)

bikeshareModel = bikeshareData %>%
  group_by(temp) %>%
  summarise(avg = mean(cnt))

bikeShareTempModel = lm(avg~temp, data = bikeshareModel)
bikeShareTempModel
```

```
summary(bikeShareTempModel)
str(bikeShareTempModel)

plot(bikeShareTempModel)

#Question 6 Fit another linear model predicting total daily bike rentals from
#daily feeling temperature. Write the linear model, interpret the slope, etc.
#Is the temperature or feeling temperature a better predictor of bike rentals?

bikeshareModel2 = bikeshareData %>%
  group_by(atemp) %>%
  summarise(avg = mean(cnt))

bikeShareAtempModel = lm(avg~atemp, data = bikeshareModel2)
bikeShareAtempModel
summary(bikeShareAtempModel)
str(bikeShareAtempModel)

stargazer(bikeShareTempModel, bikeShareAtempModel, type="text",
          dep.var.caption = "", dep.var.labels.include = F,
          report = "c*", df = F, model.numbers = F,
          keep.stat = c("ser","rsq","adj.rsq"),
          column.labels = c("Normalized Temp vs Avg","Feeling Temp vs Avg"))

```