PhillyCycle Rental Analysis

Michael Vitale
Wisconsin School of Business
May 2023

Executive Summary

Introduction

General Analysis

A Breakdown of Rentals Based on User Type

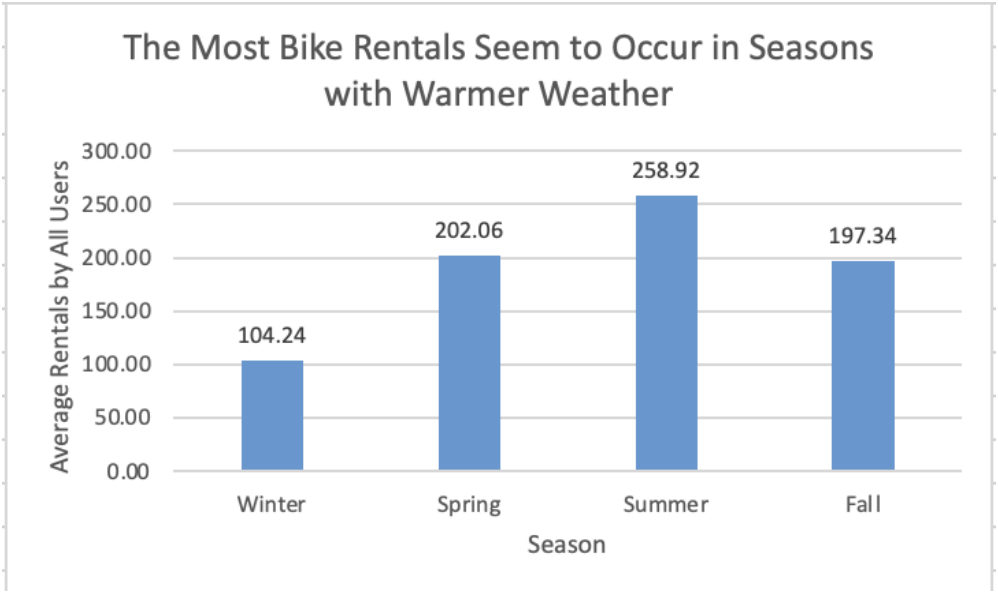
	Casual Users (Number of Rentals in an Hour)	Registered Users (Number of Rentals in an Hour)
Average	36.33	154.59
Median	16	116
Standard Deviation	50.36	153.61

I think it is valuable to include both measures to display the disconnect between them and more accurately assess the center of the data. The standard deviation is useful when considering how much the data generally varies from the average.

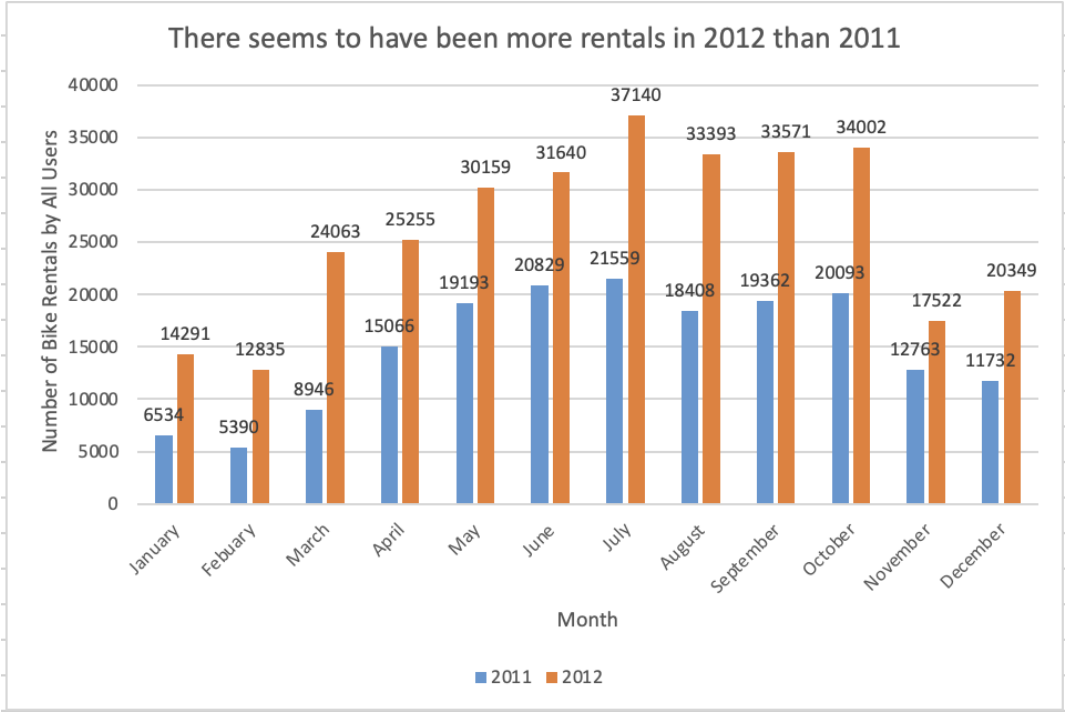
Proportion of User Type

	Casual Users	Registered Users	All Users
Sum	94012	400083	494095
Proportion	0.19	0.81	1.0

The vast majority of bike rentals are from registered users– about 80%. The remaining rentals are all from casual users– about 20%.

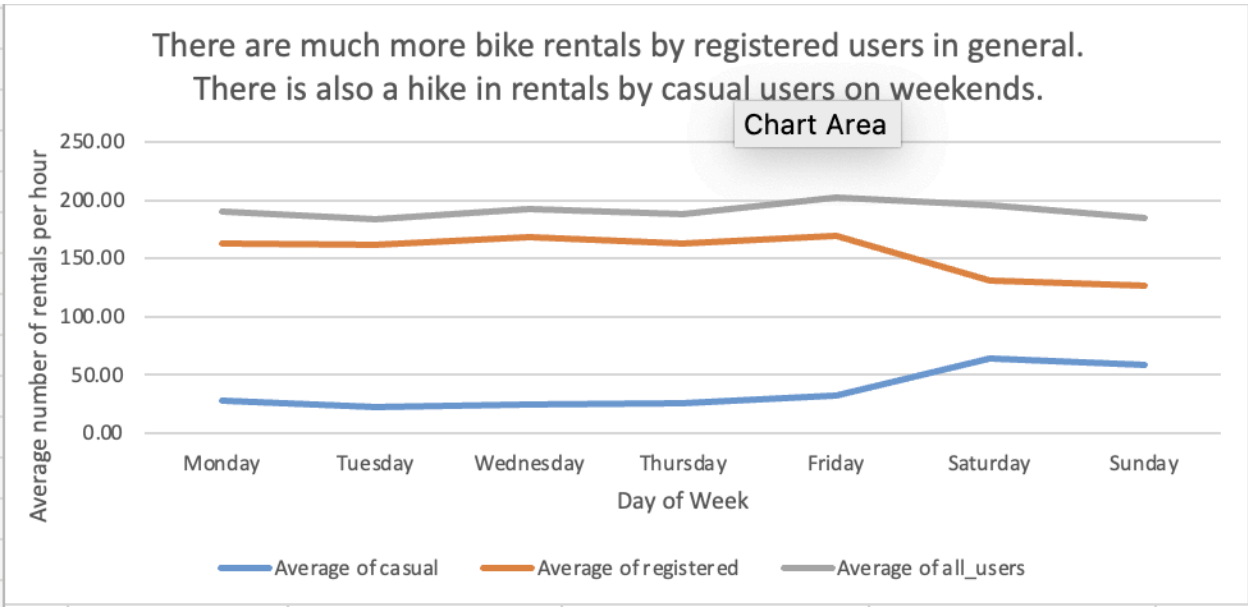


It seems bike rentals generally mirror the temperature in a given season. Summer has the most rentals, spring has the second most, fall has the third most, and winter has the least. It's also significant to mention that the difference in bike rentals between spring, summer, and fall is at most 60, but the difference in bike rentals between winter and the second lowest season is almost 100.

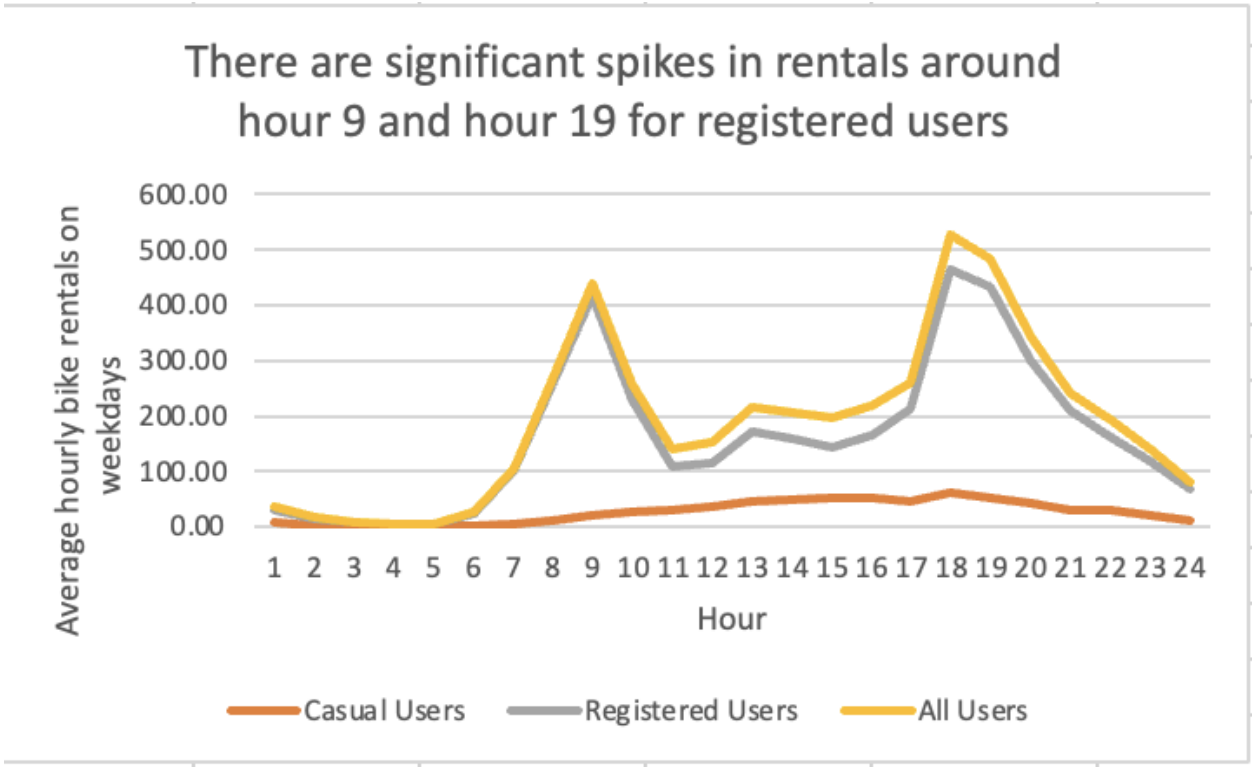


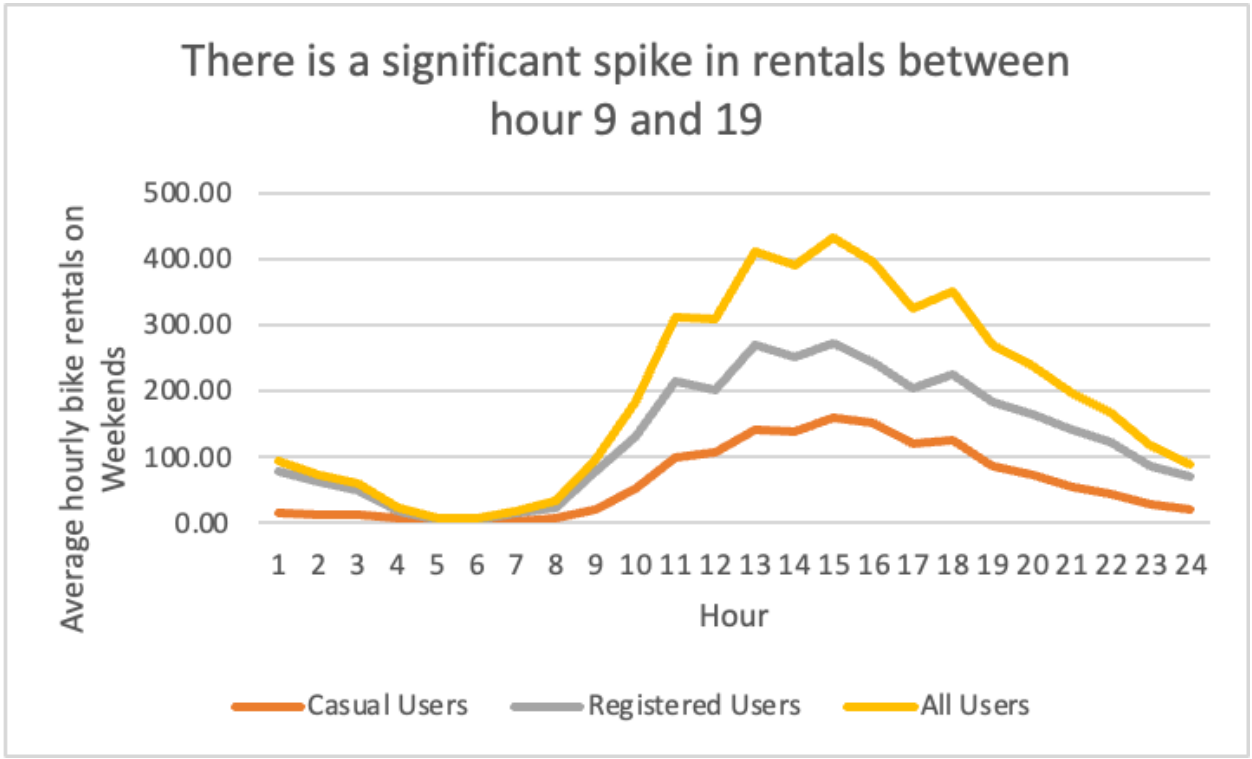
There were significantly more rentals in 2012 than 2011, potentially because of growth of the company. In general, the data from 2011 follows the same pattern as 2012. There seems to be more rentals in between March and October, and less rentals between November,

December, January, and February.



It seems the average number of rentals for casual and registered users are generally stagnant between monday and friday. Registered bike rentals are higher during this period and casual bike rentals are lower. During the weekends however, between Friday and Sunday, the quantity of bike rentals among casual users increases, and the quantity of bike rentals among registered users decreases. I would make the assumption that registered users likely use the service to get to and from work– thus they have more bike rentals during weekdays and less during the weekend. The opposite is true for casual users likely because they are using the service recreationally. All Users stay relatively stagnant the entire time because when registered user bike rentals decrease on the weekend, casual user bike rentals increase to supplement this change.

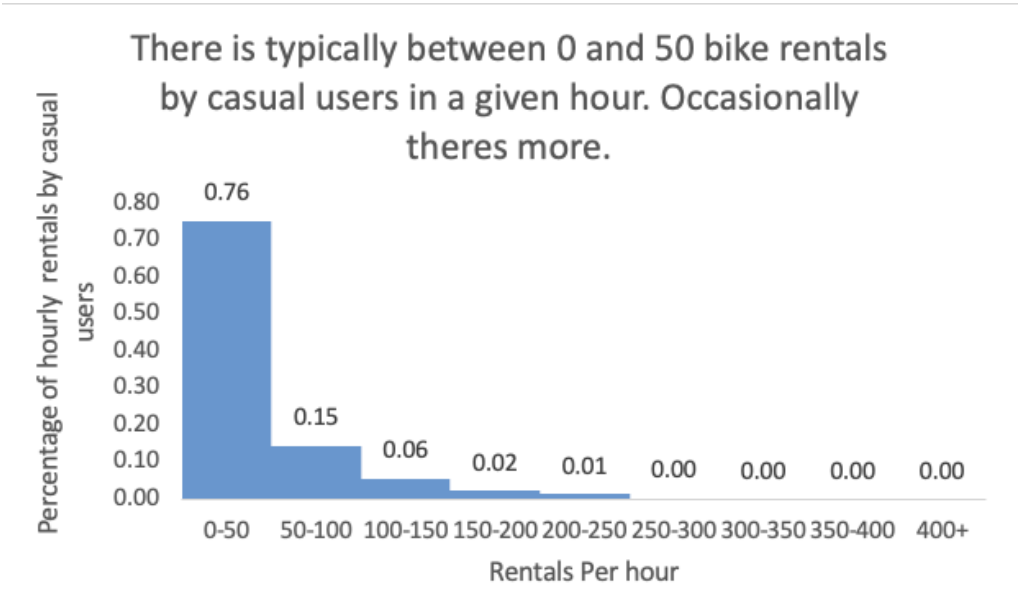


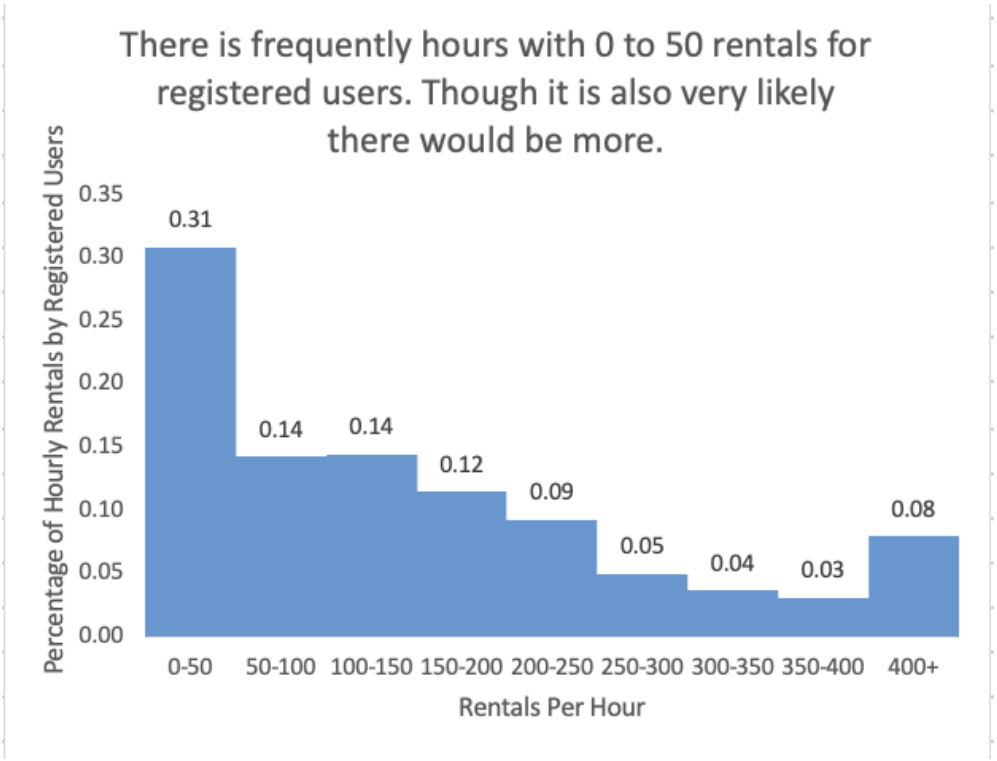


When considering the data for weekdays, it appears that there is a spike in bike rentals for registered users between hours 7 and 11, and also a spike between hours 17-21. The same

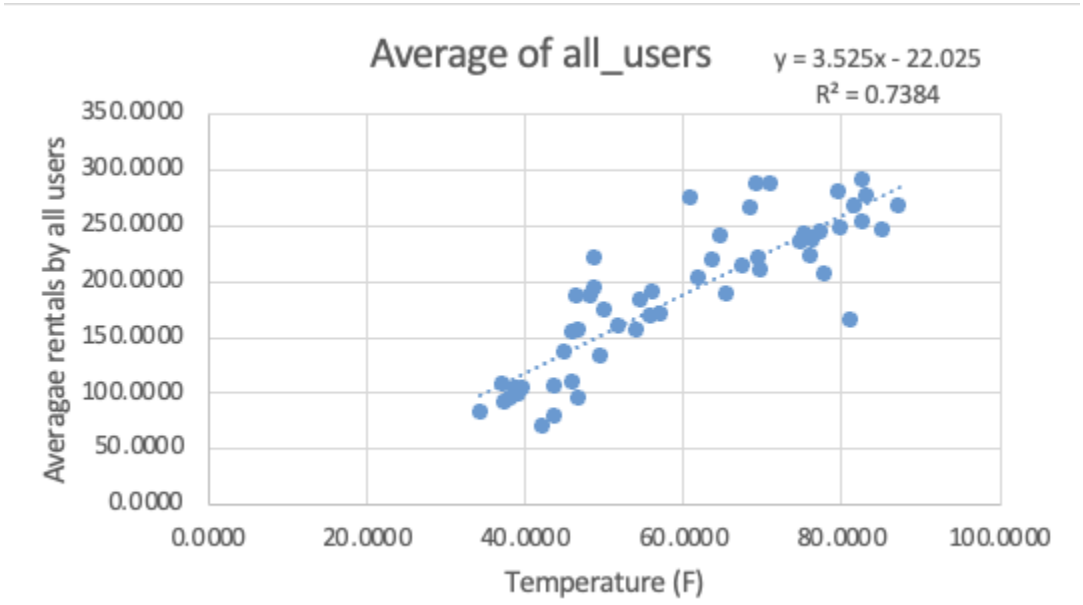
change does not occur with casual users. I would assume the spike in registered users occurs around the beginning and end of work day.

When considering the data for weekends, there is a across the board between the hours of 9 and 20. The data for registered and casual users is much closer on the weekend, though there are still more registered user bike rentals than casual user bike rentals.





It seems with casual users, the vast majority of hours have less than 50 rentals. The same is not necessarily true when considering rentals for registered users. 76% of the time, when looking at registered users, there are less than 50 rentals in a given hour. When considering the data for casual users, the same is only true 31% of the time. This is a testament to the fact that there is a much larger quantity of rentals for registered users than otherwise. Regardless, the most commonly occurring value is a rentals per hour value of less than 50.



The R^2 value in this situation is very high (.74) which points to a very high correlation between an increase in temperature and an increase in the number of average rentals. When interpreting the regression equation, the coefficient 3.525, means that for every one degree increase in temperature, a 3.53 increase in bike rentals can be expected. The y intercept of 22.025 means that the model predicts 22.025 bike rentals when the temperature is at 0 degrees fahrenheit. It is important to mention this value lies outside the data so it's hard to make this prediction without extrapolating.

Confidence Interval Analysis with respect to weekend/weekday

In this test, we tested all users, registered users, and casual users at the 95% confidence level. If the sample difference is outside the confidence interval constructed we reject the null.

	Confidence Intervals	Sample difference	Reject/Accept
Casual Users	(-5.55, 5.55)	34.62	Reject the Null
Registered Users	(-11.05, 11.05)	-35.92	Reject the Null

All Users	(-7.92, 7.92)	-1.30	Fail to Reject the Null

Being that we rejected the null for registered and casual users, we accept the alternative hypothesis– Weekend and weekday bike rentals are not equal. There is statistical evidence in support of the fact that average hourly rentals for registered and all users are not equal on weekends and weekdays. Being that we were unable to reject the null for all users, there is no statistical evidence in favor of weekend and weekday not being equal.

Regression Analysis

We made an attempt at constructing a model that factors in a variety of data to efficiently predict the number of bike rentals per hour on a given day. Our model includes factors such as weather, season, temperature, year, humidity, windspeed, and weekend/weekday. As we enter various information about the conditions on a given day into our regression equation, we can make a prediction on what the quantity of rentals in a given hour may look like.

Regression Equation:

All Rentals = 16.25 + 4.58 * Temperature + 0.28 * Wind Speed -2.62 * Humidity + 7.47 * Spring -7.52 * Summer + 61.76 * Fall + 72 .71* Year 2012 + .8 * Weekend + 17.16 * Mist + 6.55 * Precipitation

Regression Statistics

Multiple R	0.561144776
R Square	0.31488346
Adjusted R Square	0.312224878
Standard Error	153.282661
Observations	2588

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	10	27828248.81	2782824.881	118.4403863	2.7639E-203

			23495.5741
Residual	2577	60548094.66	8
Total	2587	88376343.47	

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	16.2454	20.0564	0.8100	0.4180	-23.0828	55.5736
Mist	17.1552	7.4552	2.3011	0.0215	2.5365	31.7739
Precipitation	6.5547	12.2824	0.5337	0.5936	-17.5296	30.6391
Spring	7.4730	10.7125	0.6976	0.4855	-13.5331	28.4790
Summer	-7.5184	13.8161	-0.5442	0.5864	-34.6103	19.5734
Fall	61.7596	9.4834	6.5124	0.0000	43.1637	80.3555
TempF	4.5836	0.2836	16.1645	0.0000	4.0276	5.1396
year2012	72.7105	6.0954	11.9288	0.0000	60.7581	84.6628
humidity	-2.6220	0.1885	-13.9102	0.0000	-2.9916	-2.2523
windspeed	0.2758	0.4001	0.6894	0.4907	-0.5087	1.0603
Weekend	0.8040	6.6893	0.1202	0.9043	-12.3129	13.9209

It’s important to mention how to evaluate the coefficients from the regression equation:

1. Our coefficient is based on a winter day with no mist, no precipitation, 0 degrees fahrenheit, in the year 2011, no humidity, no wind speed, and on a weekday.
2. Mist (either 0 if no or 1 if yes) =relative to clear and cloudy days, on misty days I can expect 12.37 more rentals per hour.
3. Precipitation (either 0 if no or 1 if yes) =On days with rain or snow I expect 1.36 fewer rentals per hour.
4. Spring, Summer, Fall (either 0 if no or 1 if yes) =I expect 4.425 more rentals per hour compared to variable not included (winter)
    a. Although the summer coefficient is negative, it doesn't necessarily mean there is a negative connection to rentals. → Season has a direct relationship with temperature but does not have a direct relationship with season.
5. Temp= as temperature goes up by one degree, the number of rentals per hour goes up by 4.66
6. Humidity = as humidity goes up by one point, the number of rentals per hour goes down by 2.4 rentals.
7. Year 2012 (either 0 if no or 1 if yes) = on days within the year 2012, you can expect 77.71 more rentals per hour.

8. Weekend (either 0 if no or 1 if yes) = on weekends we can expect 6.43 less rentals per hour.

There are several variables which proved statistically significant in this regression: Temperature (p-value of .00), Humidity (p-value of .00), Mist (p-value of .08), Fall (p-value of .00), and Year2012 (p-value of .00). If a variable is statistically significant it means it has a clear impact on the dependent variable and anything with a p-value less than .05 or .01 is relatively significant.

The R^2 value of this regression .31 which isn't great but shows some minor correlation between the variables. There are a lot of flaws with interpreting an R^2 variable however, so a low R^2 value, like in this case, is not necessarily indicative of a lack of connection between independent and dependent variables.

General Recommendations and Conclusion

The primary concern of Philycycle moving forward is how to go about expanding. After analyzing the data there are a variety of patterns that offer informative suggestions for the company:

1. First, the data is very clear in that temperature is a massive predictor of bike rentals. This is made evident through my My scatter plot and corresponding equation. In the equation, it is predicted that for every one degree increase in degrees fahrenheit, a 3.52 increase in hourly rentals can be expected. The R^2 value of this given graph was .74 which proves its reputability. In addition, when running my regression, the associated p-value of the temperature variable was .00 which proves there is certainly a connection between rentals and temperature. All this considered, I would strongly suggest Phillycycle expand into cities with higher temperatures on average. I would recommend cities with dense populations along the southwest or southeast coast. Simply put, people are more willing to ride a bike in warmer environments.
2. Second, in a similar way, weather proves a great predictor of bike rentals. This is evident from my regression, during which the mist and humidity p-values were both less than .05, proving their significance. Surprisingly, precipitation doesn't prove a major predictor of bike rentals. When there is misty weather, the model predicts 17.16 more rentals then if it was clear weather. For every added 1 increase in humidity, the model predicts 2.62 less bike rentals per hour. These factors should be included in the choice of the potential area of expansion as well. Phillycycle should aim for cities with low humidity, and lower precipitation.
3. Third, season is a strong predictor of bike rental as seen in the bar chart provided above. As seen in the chart, considering average rentals by all users per hour, the corresponding value of summer is 258.92, spring is 202.06, fall is 179.34, and winter is 104.24. This goes to show that considerably more rentals can be expected in the summer, spring, and fall then in the winter, and the most rentals per hour occurs in the summer. This is useful to Phyllycycle because they can adjust pricing to maximize profits in seasons with higher frequency of use, and lower pricing in seasons with minimal

rentals, like the winter. This will cause more profit in the more successful months, and hopefully more rentals in the less successful months.

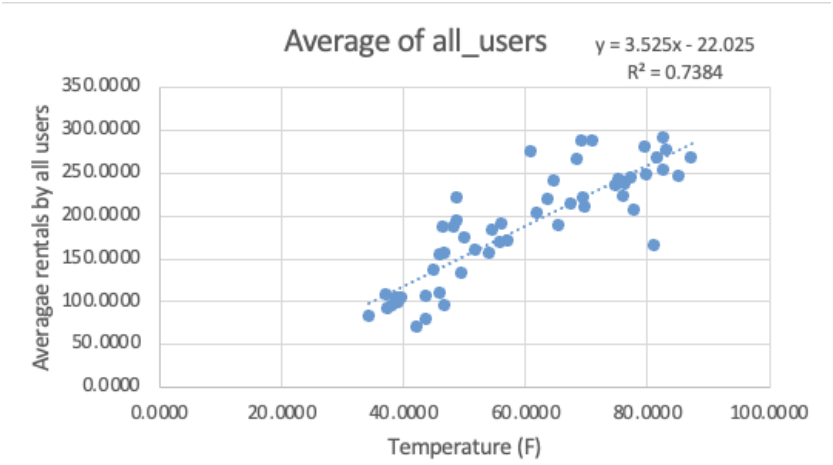
4. Fourth, weekend vs weekday is a strong predictor of bike rentals. For registered users, there are more bike rentals during the weekdays, with clear spikes around hour 9 and hour 19. There is then a slight dip in rentals on weekends for registered users, and there seems to be a spike between hours 9 and 19. For Casual users, there are few rentals throughout the weekdays, and a large spike on weekends, with the spike occurring between hours 9 and 19. I believe this is useful information in their customer segment as they can raise prices in periods of more frequent rentals to raise profits. I also think this is telling that registered users are using the service to get to and from work, based on the periods of spikes, and casual users use the service for fun on the weekends.
5. Finally, I would argue the company is already experiencing considerable growth considering the massive increase in rentals in 2012 relative to 2011. When considering average rentals per hour, every month of 2012 outperforms every corresponding month of 2011. Hopefully this can encourage the executives of Philicycle to grow their business.

I hope these suggestions prove valuable and at a minimum, Phillycycle becomes more acquainted with their individual customer segments and their intended use of the service.

Notes on Data Preparation

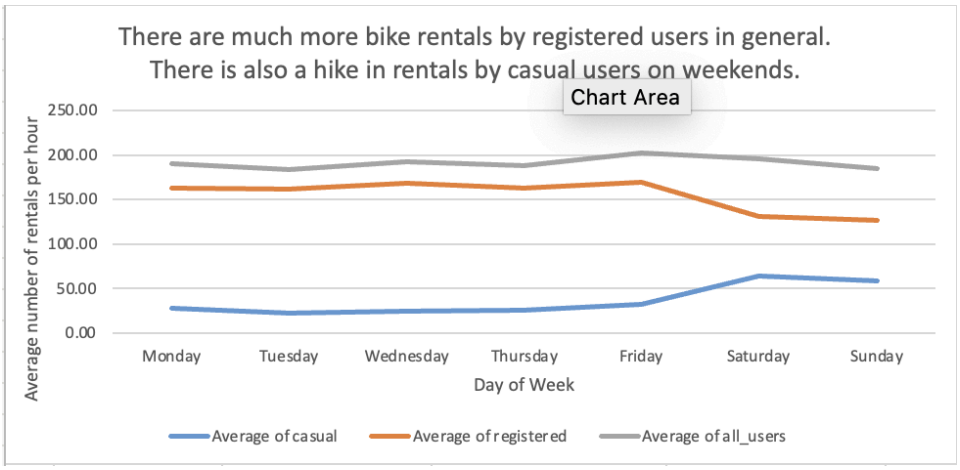
The data used in this analysis was provided by Phillycycle. The sample consisted of 2588 individual hours of operation after removing irrelevant and erroneous data. Several types of data cleansing took place prior to my analysis. To start, all duplicate observations were removed. Next, some of the data proved erroneous in regard to unrealistically high reported temperatures, so those entries were removed. I also added a variety of additional sorting categories which can be derived from the original data set– which proved valuable. Overall, I have no concerns with the data I used in this analysis and was efficiently able to organize and analyze it. Nothing ever seemed concerning or out of the ordinary through my continued investigation.

Appendix



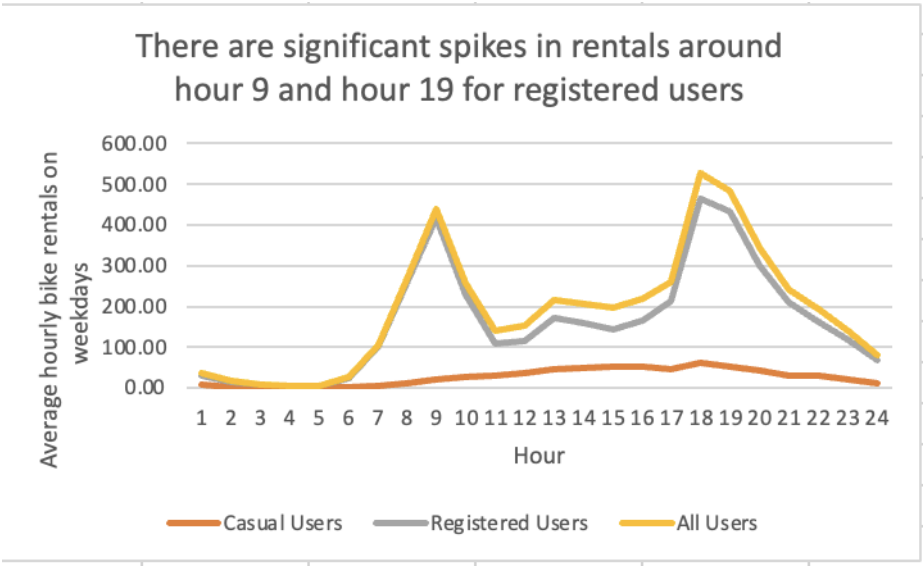
1)

This chart provides the general increase in hourly rentals when compared to temperature. In my analysis, I believe the strongest predictor of bike rentals to be temperature. This chart, proves the significance of temperature as a variable– largely through the high R^2 value which points to a strong fit of the data to the inserted trendline– and provides the nature of this relationship through the coefficient– for every one degree increase in temperature, a 3.53 increase in hourly rentals can be expected.



2)

This chart displays the pattern of daily rentals for registered users, casual users, and all users throughout the days of the week. This chart is especially valuable because it displays the rental pattern for the specific user groups. As seen from the chart, registered users have much higher hourly rentals during the weekdays, and a dip on the weekends. The opposite is true for casual users– there are much higher rates of hourly rentals on the weekends relative to the weekdays. This chart also shows that there are more rentals among registered users than casual users in general. This is telling of the patterns of the user base and Phillycycle can use this data to change pricing to match demand.



3)

The final chart I decided to show depicts the average number of users based on time of day for casual and registered users. As seen by the chart, among registered users, there are spikes in the data at hour 9 and 19. We can infer that registered users are primarily using the service to transport to and from work.