“Weather” to Rent a Bike

Teal Lucas
WISCONSIN SCHOOL OF BUSINESS
December 2024

Executive Summary

This case report will analyze the extensive data of PhillyCycle, a Philadelphia base bike sharing rental company. Starting with two years of data entries, we cleaned and sorted this data, and then analyzed it to come to conclusions about trends in the data, and how PhillyCycle can continue their expansion. Our analysis provides key insights into the factors that have driven demand and highlights potential growth opportunities.

The unsorted data began with 2,927 entries, and after removing errors, 2,916 were left. With these entries, the first thing analyzed was the user segments. Comparing the casual users to the registered users, average, median, and standard deviation were all comprised in order to show the different ways metrics represent data. Additionally, the proportion of different user segments was found. 81% of PhillyCycle's users are registered, while just 19% are casual. This gross difference represents a major opportunity for growth in the casual user segment. Whether it is marketing better and more effectively to tourists, or enhancing the user interface of the kiosks, there is room for a casual user market increase.

Furthermore, the seasonality of bike rentals was analyzed, allowing for the investigation into weather trends. To no surprise, the Summer months are the busiest time of the year for PhillyCycle. Spring and Fall follow shortly behind, and Winter is the least popular time of the year. This information provides us with the information that is needed to begin allocating marketing efforts at the peak season. Marketing in Winter will be less effective as far less people are renting from both user segments. The monthly analysis broke the seasonal analysis down even further, concluding that July and August are the most popular months. Additionally, the monthly analysis compared year one to year two, which showed massive growth. To continue this growth, PhillyCycle must be prepared when the new year comes around, and an increased number of bikes and rental stations may be needed to provide enough service for the target market.

Beyond this, time of day was analyzed, and showed insights into the uses for the user segments. Registered users most commonly rent bikes during commute times of weekdays (7:00 - 9:00 AM and 4:00 - 6:00 PM), and casual users rent bikes during leisure times of the weekend (10:00 AM - 5:00 PM). Also, a corollary test was run to analyze the effect temperature had on our rentals. There was a strong positive correlation here, as an increase in rentals happened due to the temperature increase.

In addition to this, hypothesis testing was conducted at 5% significance levels, and regression analysis was also conducted to further analyze the effect of different variables on rentals. Hourly volume and user segment were analyzed here. Casual and registered users ended up outside our established confidence intervals, but all users were contained within the interval.

Based on our testing, PhillyCycle has great room for improvement, and possible expansion in the coming years seems in the cards. Weather, season, time of day, weekday, and user segment all had effects on rental volume, and expansion to a city with a temperate climate would be beneficial.

Introduction

PhillyCycle is a Philadelphia based bike rental sharing company. Launched in 2011, tremendous growth has happened among multiple user segments, including registered users, who pay a monthly membership and receive a discount on the hourly rate, and casual users, who pay an hourly rate without having to register with the company. There is plenty of opportunity to continue the substantial growth in the coming years.

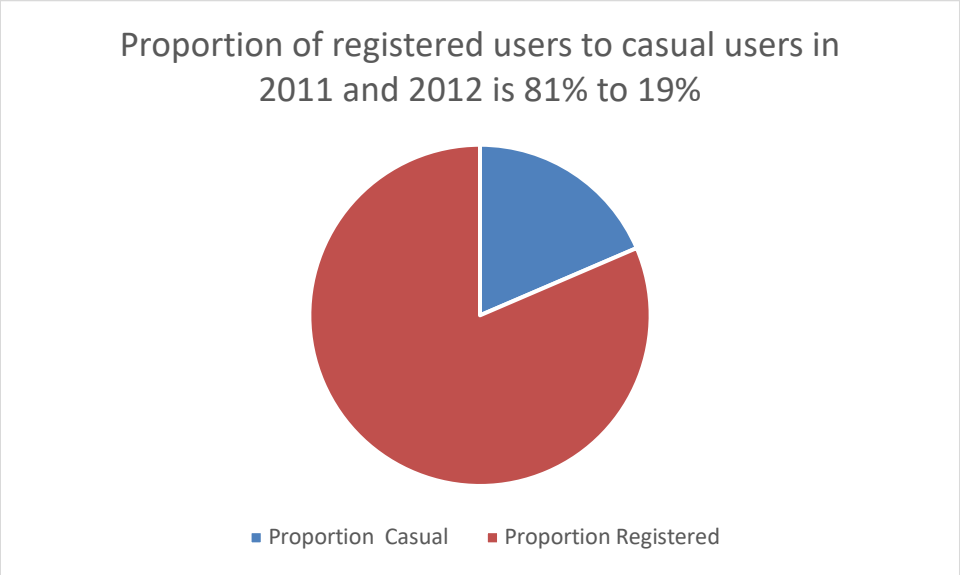
Analysis

Bike Rental Patterns

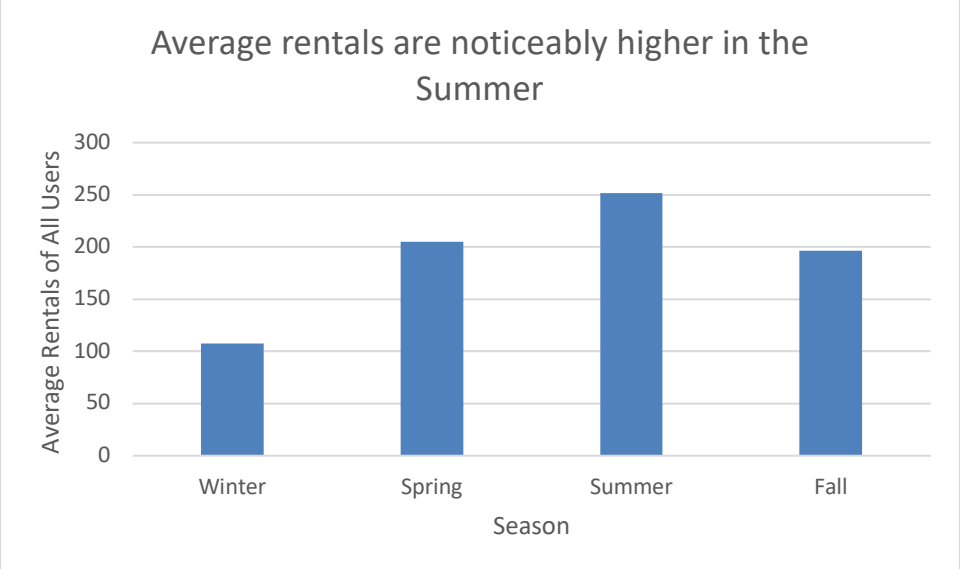
Summary Statistics by User Segment

	Casual Users	Registered Users
Average	35.32	155.15
Median	17.00	116.50
Standard Deviation	48.18	154.51

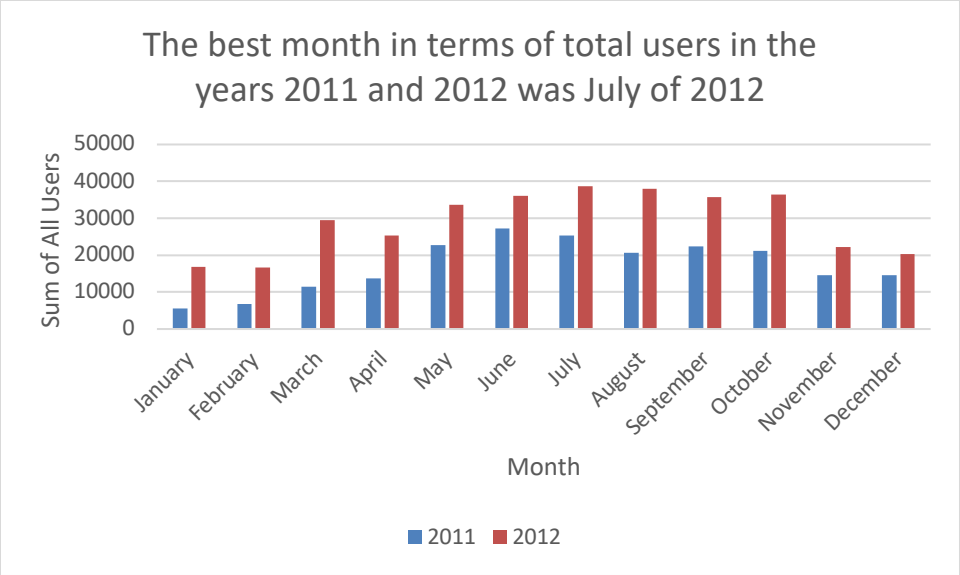
The average number of rentals by casual and registered users is 35.32 and 155.15, the median number of rentals for casual and registered users is 17.00 and 116.50, and the standard deviation of rentals of casual and registered users is 48.18 and 154.51. The average provides a clear measure of the central tendency of the data, the median helps account for potential outliers that may skew the data, and the standard deviation measures the variability in the bike rentals, which indicates how consistent the number of rentals is within the different segments. I think these three measures offer more insights into the significance of our data than would looking at a maximum or minimum value.



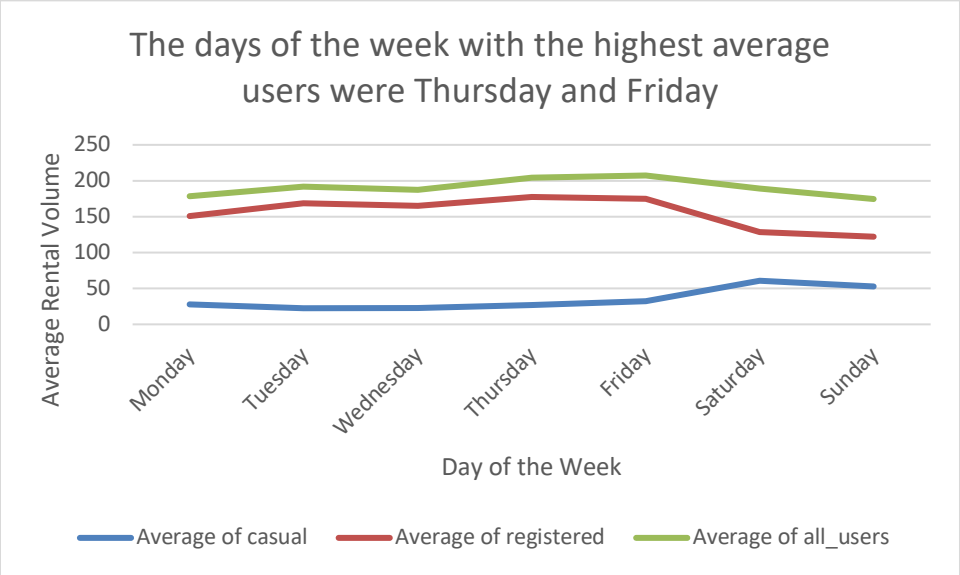
The number of bike rentals from registered users was far higher than the number from casual users. Registered users comprised 81% of rentals, while casual users only comprised 19%.



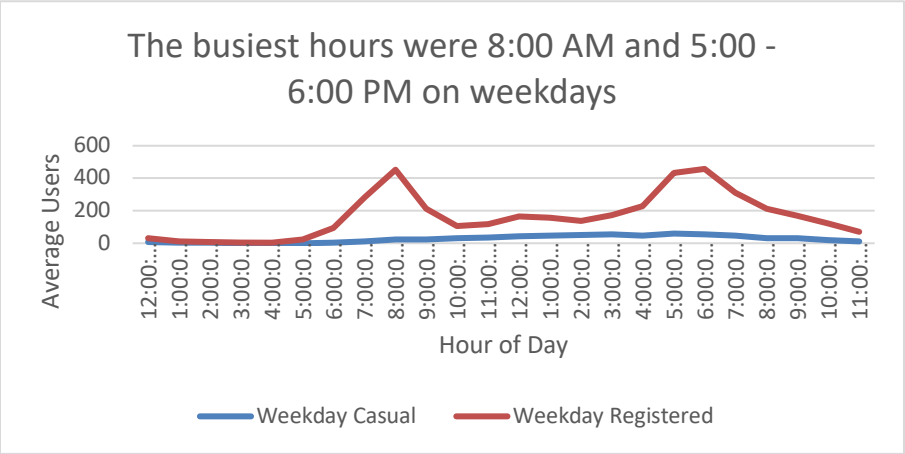
The summer season is noticeably the busiest season, followed by Spring, Fall, and then not surprisingly, Winter. Spring and Fall have very similar amounts of average rentals. In the Summer, average rentals are over twice what they are in the Winter.

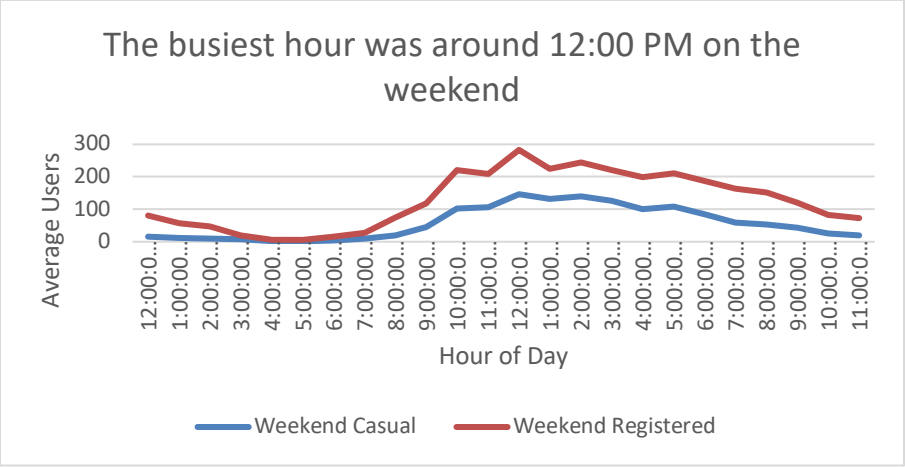


The number of users in 2012 far exceeded 2011, with January through April being twice as popular as the previous year. The most users came between May and October, with July being the most popular month, just above August.

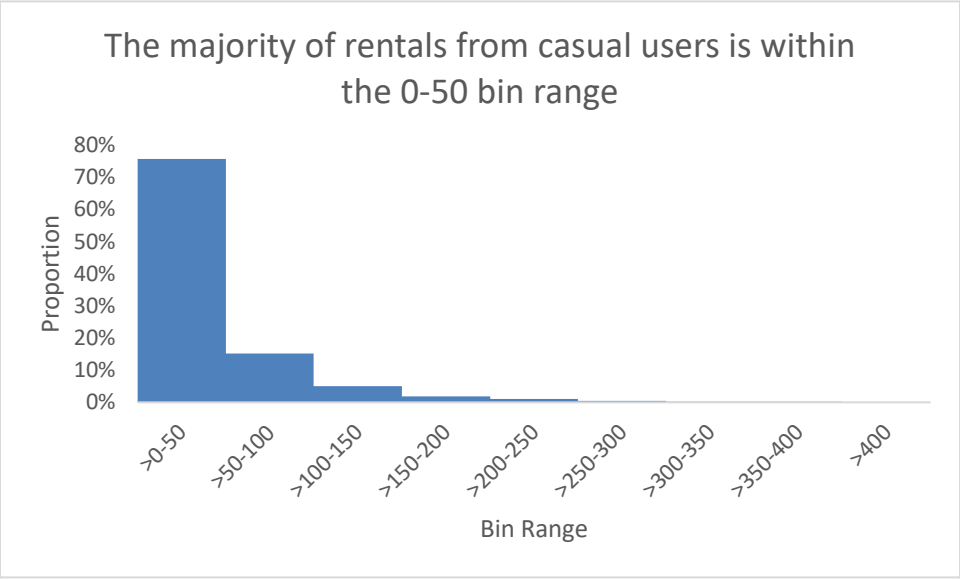


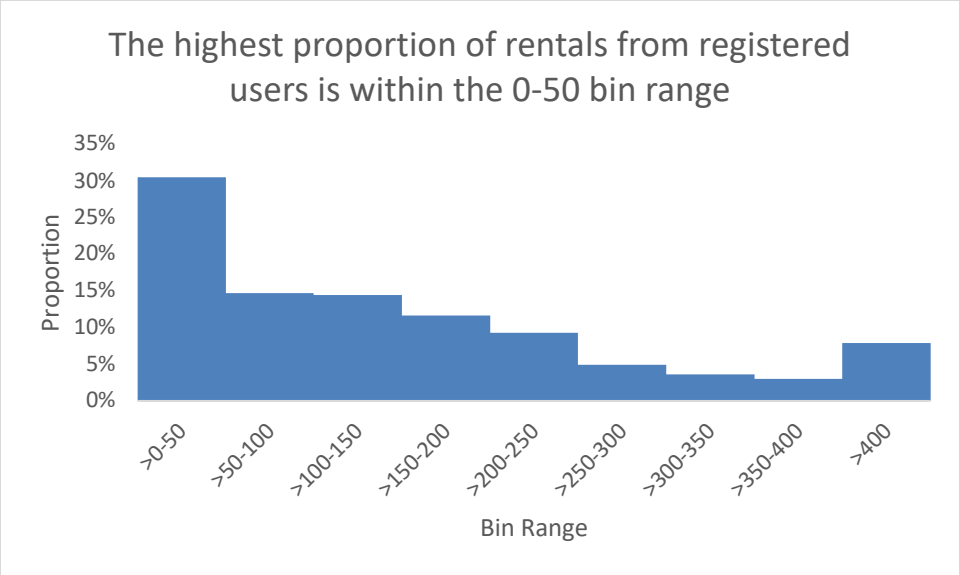
The average rental volume of registered users is substantially higher than that of casual users. Registered users have a dip in rental volume on the weekend, while casual user volume nearly doubles. When you combine casual and registered, the volume of all users is fairly consistent, with the busiest day of the week being Friday.



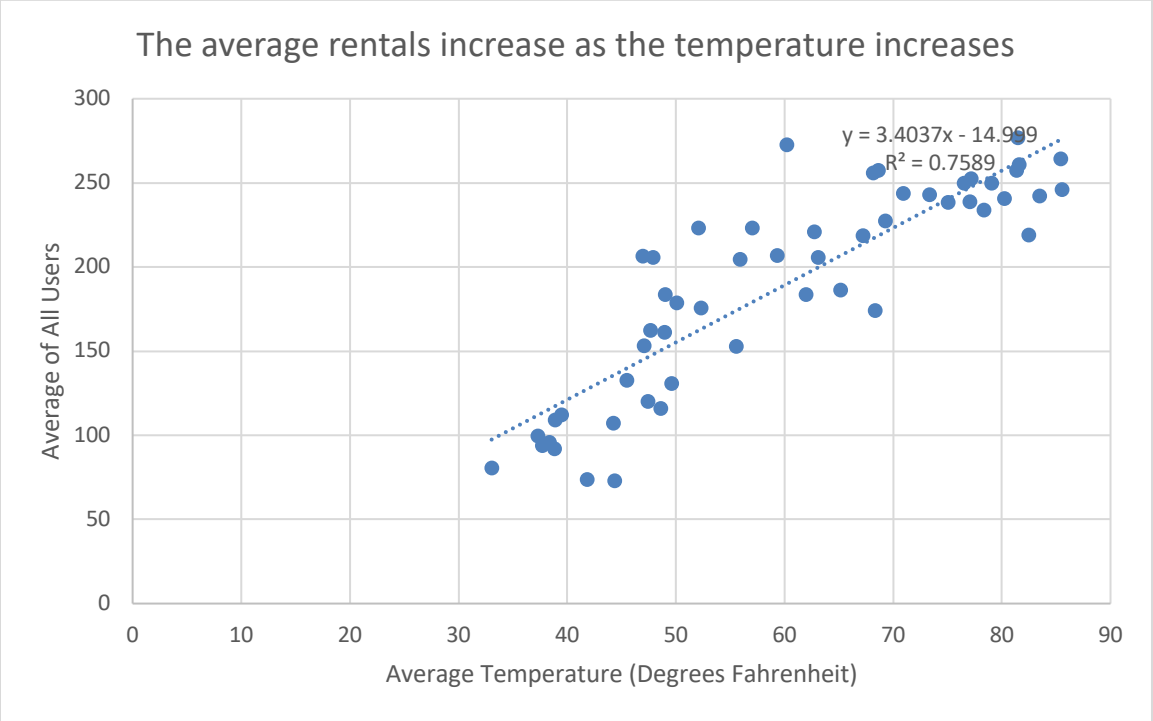


The number of rentals by hour is vastly different on weekends compared to weekdays. On weekdays, 7:00 – 8:00 AM and 5:00 – 6:00 PM are the busiest for registered users, while there are very few rentals at all for casuals. On weekends, the busiest time of day is around 12:00 PM for both casual and registered users.





With casual users, the proportions are more contained within the lower bins, with 76% being within the >0-50 bin range. With registered users, the proportion is much more spread out, and while 30% are within the >0-50 bin range, it is not the majority. Additionally, there is a small bump in the >400 range with registered users, at 8%.



An R squared value of 0.7589 indicates a strongly correlated relationship between the average rentals and the temperature. The equation y = 3.4037x – 14.999 represents a predicted value of y (average rentals) for a given value of x (temperature in degrees Fahrenheit).

Significance of Difference Between Weekend and Weekday Hourly Average Users

	All Users	Casual Users	Registered Users
Weekday Hourly Average	193.89	26.52	167.37
Weekend Hourly Average	182.16	56.68	125.48
Confidence Interval	(-14.08, 14.08)	(-4.89, 4.89)	(-10.23, 10.23)
Significant Difference	No	Yes	Yes

Hypothesis testing at a 5% significance level was conducted to determine if there was evidence of a different hourly volume on weekends compared to weekdays, for casual, registered, and all users. Using the confidence interval method, it was found that with all users, there was no significant difference because the difference between weekend and weekday hourly average was within the confidence interval. However, with casual users and registered users, there was a significant difference, as the difference in weekend and weekday hourly average was not within the confidence intervals.

Regression Analysis

A multiple regression analysis was run to detect the impact different variables had on the rental volume. The different variables analyzed were temperature (F), humidity, windspeed, season (Spring, Summer, Fall, Winter), year (2011, 2012), weekday or weekend, mist, and precipitation.

All Rentals = 3.33 + (4.74*Temperature) + (-2.49*Humidity) + (0.46*Windspeed) + (4.61*Spring) + (-16.95*Summer) + (57.96*Fall) + (74.23*Year2012) + (-7.48*Weekend) + (12.34*Mist) + (1.09*Precipitation)

Regression Statistics	
Multiple R	0.553
R Square	0.306
Adjusted R Square	0.303
Standard Error	153.531
Observations	2916.000

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.33	19.33	0.17	0.86	-34.57	41.23	-34.57	41.23
Temp (F)	4.74	0.27	17.62	0.00	4.21	5.27	4.21	5.27
humidity	-2.49	0.18	-13.96	0.00	-2.84	-2.14	-2.84	-2.14
windspeed	0.46	0.38	1.22	0.22	-0.28	1.20	-0.28	1.20
Spring	4.61	10.08	0.46	0.65	-15.15	24.36	-15.15	24.36
Summer	-16.95	12.94	-1.31	0.19	-42.32	8.43	-42.32	8.43
Fall	57.96	8.86	6.54	0.00	40.59	75.33	40.59	75.33
Year 2012	74.23	5.73	12.95	0.00	62.98	85.47	62.98	85.47
Weekend	-7.48	6.26	-1.19	0.23	-19.76	4.80	-19.76	4.80
Mist	12.34	6.99	1.77	0.08	-1.37	26.04	-1.37	26.04
Precipitation	1.09	11.25	0.10	0.92	-20.97	23.16	-20.97	23.16

In this regression, the variables that were tested at a P-Value less than 0.1 were Temp (F), humidity, Fall, Year 2012, and Mist. The variables that were tested at a P-Value greater than 0.1 were Intercept, windspeed, Spring, Summer, Weekend, and Precipitation.

The R-squared measure represents how well our data fits our regression. The R-squared value of .306 tells us that the variables in our regression explain just 30.6% of the variation in the volumes of our bike rentals.

Temp (F), humidity, Fall, and Year 2012 all show the most significant relationship to the average rentals observed, because they have a P-Values of zero. Specifically, with a one degree increase in temperature, rides increased by a factor of 4.74. With an increase in humidity by a factor of 1, rides decreased by 2.49. With our seasons, when a certain date occurs, making the unit of measure equal to one, the number of rentals increase or decrease whatever amount the coefficient is. For example, Fall would increase by 57.96. Similarly, if the year was 2012, the rentals would increase by 74.23. Our dummy variables are the variables not listed in the regression. For example, Winter is our seasonal dummy variable, and Weekday is the dummy variable for what part of the week it is. Looking at this more specifically, because the Weekend coefficient is -7.48, that means that based on the conditions, the Weekend would have 7.48 less rentals than that of a Weekday. Additionally, the Summer variable is not significant, even though it would be expected to be significant. This is because the temperature variable is already significant, and temperature is one of the most significant predictors of rentals already.

Conclusion

Based on my analysis of PhillyCycle's data, I have multiple conclusions and recommendations for the company moving forward. Some of the variables, like season, temperature, weekend, and time of day have major effects on the number of average rentals. Variability in several factors, including weather and temperature, and seasonality, explains some of the fluctuation with our rental patterns. Additionally, registered PhillyCycle users rent bikes substantially more often, however they do it at different times than our casual users. The company grew substantially from 2011 to 2012, and similar growth should be expected looking forward. This means that preparations will need to be made in order to harness this growth.

In order to capitalize on the opportunity in front of us, marketing needs to be a major point of emphasis. Allocating marketing efforts that targets the peak season (Summer) is a potential recommendation that will capitalize on the busiest time of the year. When it comes to the potential expansion into other cities, the most rapid and consistent success will come in Southern regions with more temperate climates. Additionally, enhancing casual customer experience, whether that is with top notch customer service, or user-friendliness of the rental kiosk, is another area that with the right effort, has potential for major growth. Also, marketing to tourists who don't need a full membership, or convincing non-tourist casual users to become registered is a domain that could turn more profit. Due to the expected growth, additional rental stations and kiosks should be invested in.

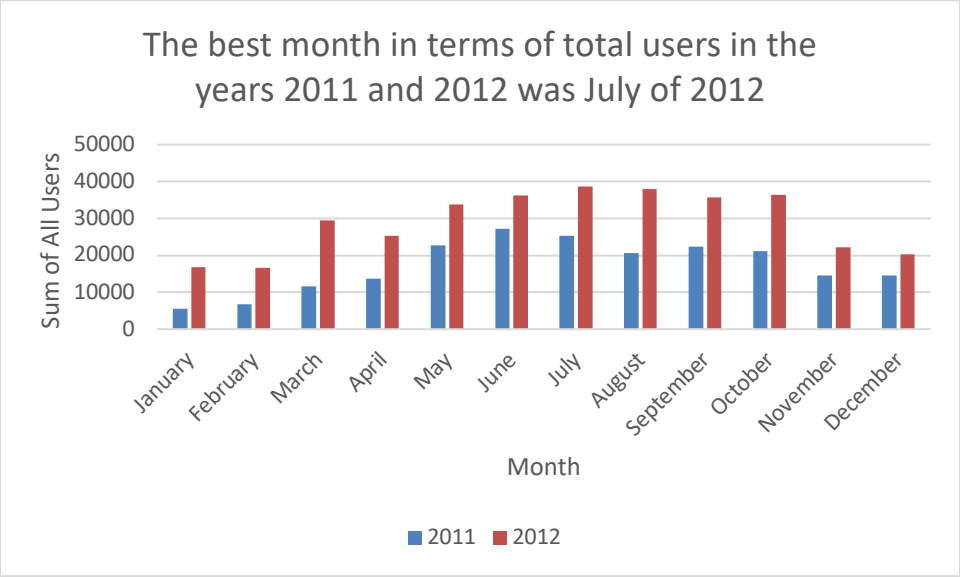
Losing sales due to a shortage of bikes may very well turn customers away for good, and lifetime customer value is extremely important.

Appendix

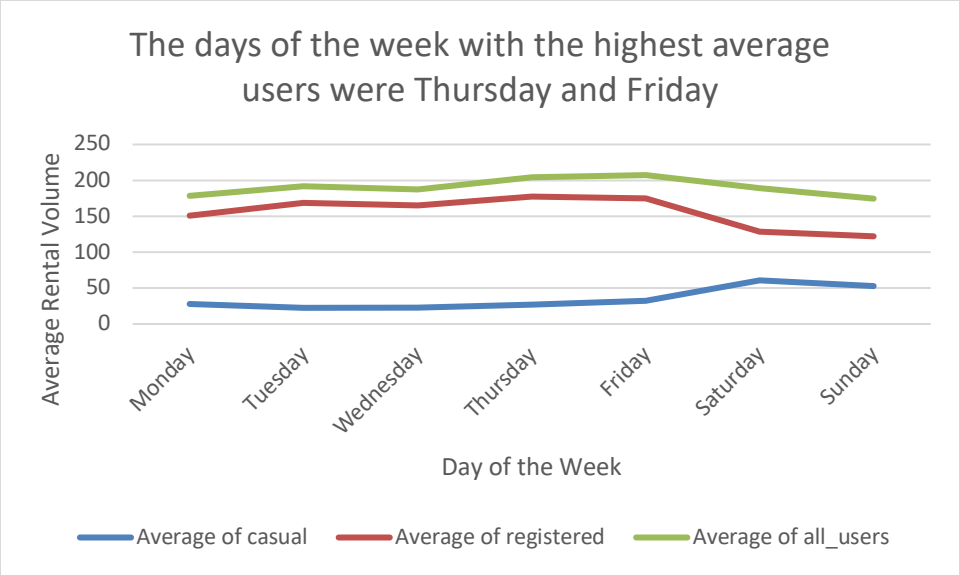
Notes on Data Preparation

The data preparation process was an extensive, yet efficient process where thousands of inputs of data were analyzed. Starting with a record number, each data entry contained a specific date and time, whether it was a holiday or not, weather, temperature (C), humidity, windspeed, and then the number rentals for each type of user (casual and registered) at that time. Beginning with 2,927 entries, erroneous data was removed, as were duplicate data entries, leaving 2,916 entries of relevant data. Columns for year, month, day of week, hour, weekend, week number, mist, precipitation, season, and year were all added. Additionally, temperature was converted to degrees Fahrenheit. The data cleaning and the addition of new variables to analyze the data allowed for a thorough and meaningful analysis of PhillyCycle's company rentals. I have no concerns with the analysis of the data, as it was a tedious and particular process to which only resulted in the removal of 11 entries. Because of the successful data preparation, several strong conclusions were made that I believe will enhance the company's future.

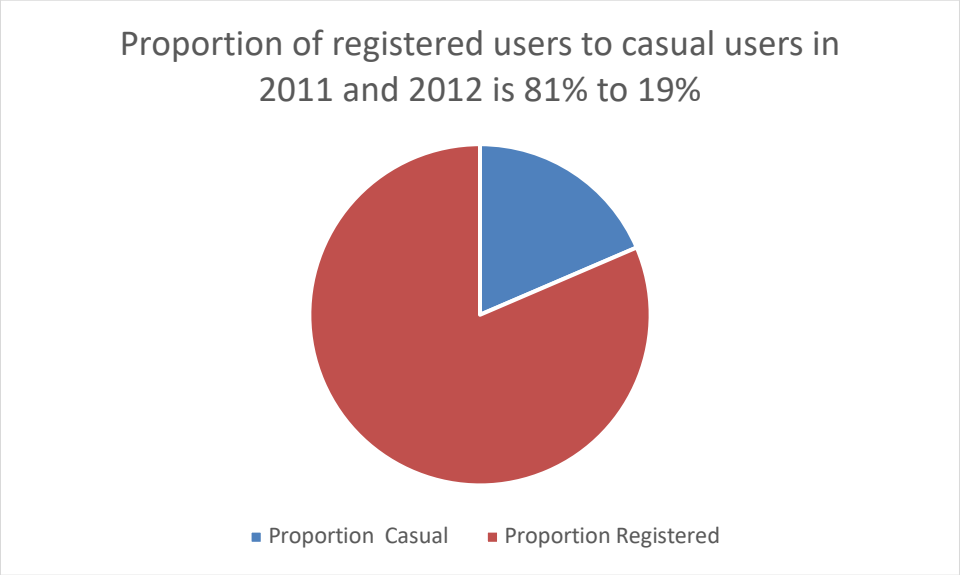
Elevator Charts



This chart displays the sum of all users throughout the months of the year in the two years of operation (2011 and 2012. I find this chart the most useful, as it shows the major growth from year one to year two, and it also displays when the busy season of the year is. This information is important to know to have a successful plan to be prepared and continue to harness the company's growth in the future.



This graph provides information pertaining to business based on day of the week for different types of users. I find this graph important, as it shows the increase in usership of casual users on weekends, but the decrease in usership of registered users in weekends. This pattern shows a clear trend that casual users use PhillyCycle more for pleasure, and registered users use it more for work. I find this graph slightly less compelling than the above one, due to its lack of year-to-year comparison.



This pie chart displays the proportion of casual and registered users in both years 2011 and 2012. It is extremely simple to read and is a quick insight into what types of users PhilllyCycle currently has, and where the market has room to grow. This chart is ranked third due to its lack of other metrics, however it still gives relevant information that can help PhillyCycle.