OzarkRide

Business Scenario:

OzarkRide is a local bike rental founded in 2011 from Fayetteville which focuses on a sustainable and affordable way to reach from point A to point B using environment-friendly technology. OzarkRide is planning to open a new branch in a city similar to Fayetteville within a few months. They have hired us at Fay Consulting to dive deep into their data and provide insights on the findings.

One of the main and most important questions they have is which month should they launch their service? Is there anything other than the month they should consider? Support with analytics.

Response

There are more than months that OzarkRide should consider when planning to open a new branch in a different city. At Fay Consulting, we believe that besides months, seasonality, weather, and population should be considered. As seasons are tied with the weather, by common sense we know that when it is cold outside, people do not do many outdoor activities and rather have indoor activities. The same when it rains; people try to stay dry and will not be riding a bike.

During our research, we found that Fall and Summer are the best time to target expansion. Based on the operation data, Fall has the highest number of users followed by Summer. We believe that beginning the summer the month of - May through October when the fall season ends - is the peak time to expand the operation in a new city.

Knowing the population density also plays a very important role in the success of the expansion. Fayetteville is a college town. Therefore, there is a high density in population during school time resulting in more people using the bike rental service. Depending on where OzarkRide plans to expand, a study of the population should be performed before making a decision. Unfortunately, the dataset provided does not have the information necessary to analyze the population.

Recommendation

Our recommendation at Fay Consulting for OzarkRide is the following:

- Have in mind to expand either during Summertime or during Fall. The earlier OzarkRide expand, the more exposure to the population.
- The best months to expand is in the range between May and October; we believe that a month in the early middle of the range is the best-case scenario. In this case, the months would be May, June, July, and August based on the analysis performed. The later OzarkRide goes, the less exposure the company has to new customers due to competition.
- Perform further analysis to study the population of the city where OzarkRide wants to open the new branch. This is key if OzarkRide wants to have higher traffic in the bike renting service.
- Study possibly competitions that provide similar services as OzarkRide. The more service varieties the targeted city has, the fewer customers the company gets in the launch unless a strong marketing campaign with promotions is included.

Appendix

Summary Statistics

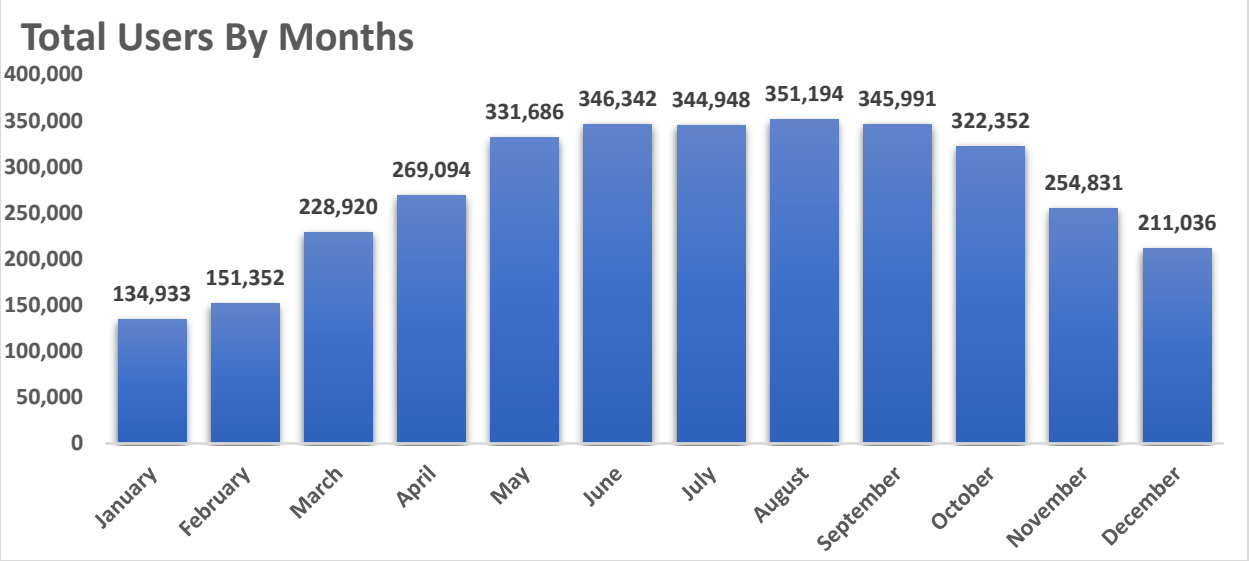
Summary Statistics

Results

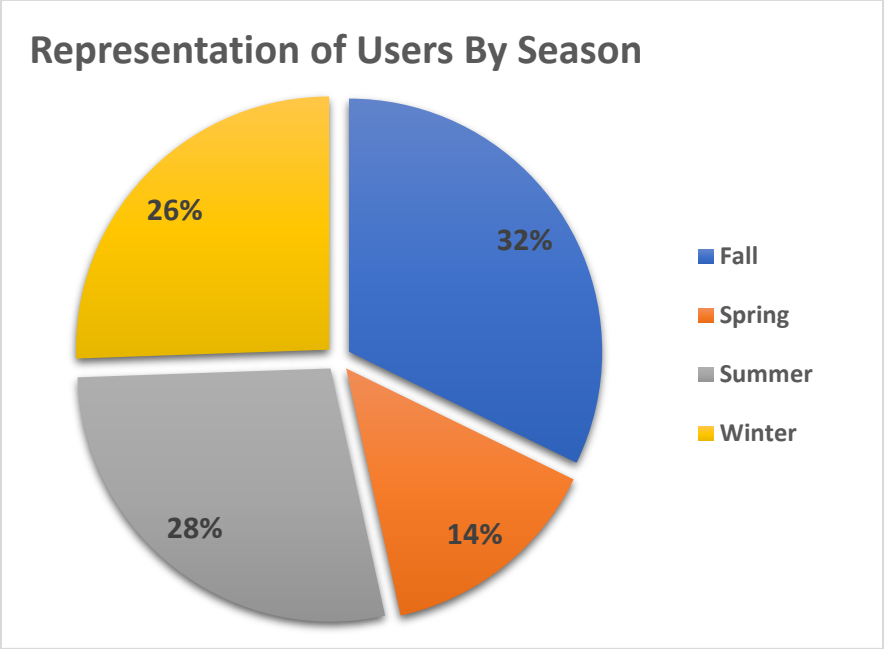
The MEANS Procedure

Analysis Variable : total_users							
season_name	N Obs	Mean	Std Dev	Std Error	Minimum	Maximum	N
Fall	4496	236.0162367	197.7116302	2.9486218	1.0000000	977.0000000	4496
Spring	4242	111.1145686	119.2240102	1.8305365	1.0000000	801.0000000	4242
Summer	4409	208.3440689	188.3624731	2.8367714	1.0000000	957.0000000	4409
Winter	4232	198.8688563	182.9679718	2.8125629	1.0000000	967.0000000	4232

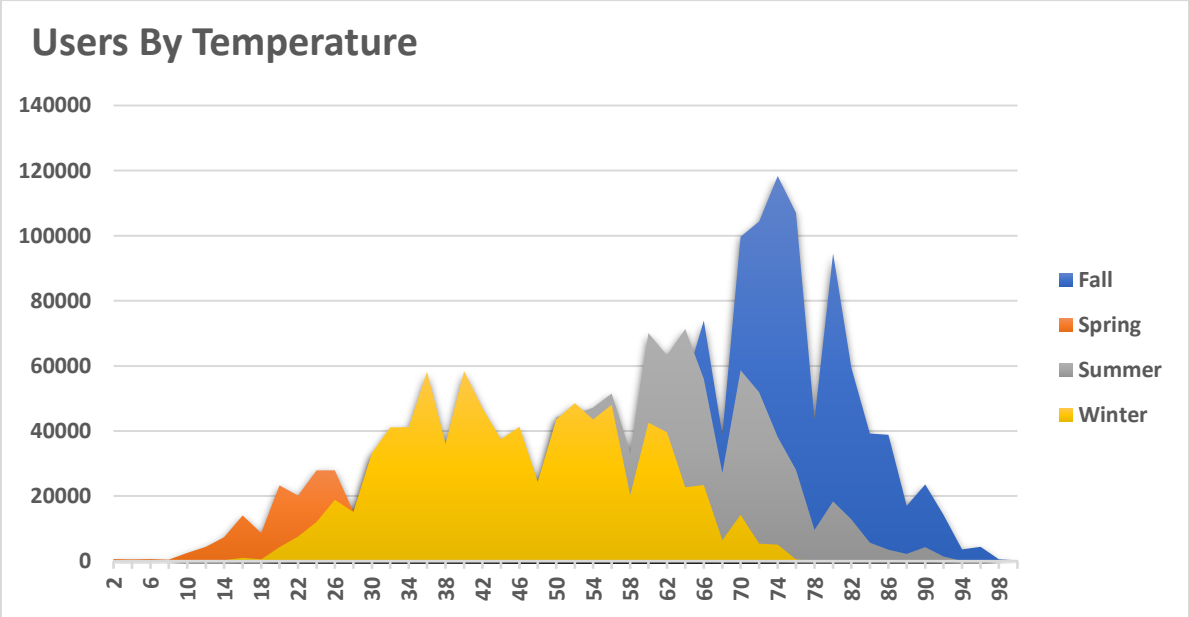
- The season with the highest average mean is Fall followed by Summer.
- The season with the lowest average mean is Spring. (*This is skewed because the data captured is wrong…January and February are considered winter, but the data has it as spring*)



- May through October have very similar number in total users with August being the highest by just 5,000.



- We clearly see that Fall has the highest representation of users followed by Summer.



- Low temperature means less people using OzarkRide service. On the other hand, as the temperature increase, the amount of rental increase for both casual and registered users.

Visualization and Descriptive Take Away

- Having the highest number of users during Fall and Summer is very normal. The weather during those two seasons is warm and people usually do outdoor activities in this case, biking.
- The reason why January and February have the lowest number of users is that it is the beginning of cold weather and people prefer to stay home or do indoor activities. Unfortunately, the data capture has spring starting in January and February, thus, skewing the spring data in favor of winter.
- OzarkRide has two types of clients: casual users and registered users. Casual users make up about 18.8% of the revenue while registered users make up about 82% of the company's revenue. Although the difference between casual users and registered users is alarming having 82% registered users is a very good number.

OzarkRide Case
By: Ruben Chung

Hypothesis Testing

ANOVA Testing

- **OzarkRide claim that the number of users is similar throughout the four seasons. Is the average number of users different throughout the four seasons?**

1. Levine's Testing

Ho: all variances are equal

Ha: at least one variance is different

2. α = 0.05

3. F = 170.65

4. p-value = 0.001 < 0.05 = α

5. Reject the null hypothesis (proceed with caution)

6. Conclusion: At least one variance is different.

One-Way Analysis of Variance

Results

The ANOVA Procedure

Levene's Test for Homogeneity of total_users Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
season_name	3	1.602E12	5.34E11	170.65	<.0001
Error	17375	5.437E13	3.1292E9		

2. One-Way ANOVA Testing

Ho: all means are equal

Ha: at least one mean is different

2. α = 0.05

3. F = 409.18

4. p-value = 0.001 < 0.05 = alpha

5. Reject the null hypothesis

6. There is a significant difference in the mean between seasons and the number of total users.

One-Way Analysis of Variance

Results

The ANOVA Procedure

Dependent Variable: total_users

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	37729357.7	12576452.6	409.18	<.0001
Error	17375	534032233.4	30735.7		
Corrected Total	17378	571761591.1			

R-Square	Coeff Var	Root MSE	total_users Mean
0.065988	92.53302	175.3159	189.4631

Source	DF	Anova SS	Mean Square	F Value	Pr > F
season_name	3	37729357.67	12576452.56	409.18	<.0001

3. Tukey's Test

The difference in the mean happens between Fall and Summer, Fall and Winter, and Fall and Spring. However, Summer and Winter means are pretty similar.

One-Way Analysis of Variance

Results

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for total_users

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	17375
Error Mean Square	30735.67
Critical Value of Studentized Range	3.63350

Comparisons significant at the 0.05 level are indicated by ***.

season_name Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Fall - Summer	27.672	18.125	37.219	***
Fall - Winter	37.147	27.500	46.795	***
Fall - Spring	124.902	115.260	134.543	***
Summer - Fall	-27.672	-37.219	-18.125	***
Summer - Winter	9.475	-0.218	19.168	
Summer - Spring	97.230	87.542	106.917	***
Winter - Fall	-37.147	-46.795	-27.500	***
Winter - Summer	-9.475	-19.168	0.218	
Winter - Spring	87.754	77.968	97.541	***
Spring - Fall	-124.902	-134.543	-115.260	***
Spring - Summer	-97.230	-106.917	-87.542	***
Spring - Winter	-87.754	-97.541	-77.968	***

One-Way ANOVA Take Away

- There is no statistical evidence to support OzarkRide's claim of having a similar number of users throughout the four seasons.
- Running ANOVA makes sense in this case as we are trying to understand how the different seasons affect the number of users.
- Tukey's test helps us understand that the average mean users of the seasons are significantly different. We can support that by looking at Tukey's graph where Fall – Summer, Fall -Winter, Fall – Spring is significantly different, and only Summer – Winter is similar. (*Having *** means they are different*)

Regression Analysis

- **From all the operation data, can temperature, humidity, and windspeed predict the total users?**

1. Overall model is significant

2.

temp is significant

hum is significant

windspeed is significant

3.

b0 = 178.81

b1 = 362.53

b2 = -273.46

b3 = 26.32

4. Adj-R = 0.2512 or 25.12% of the variability can be explained by the total users, temperature, humidity, and windspeed.

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: total_users

Number of Observations Read	17379
Number of Observations Used	17379

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	143717018	47905673	1944.57	<.0001
Error	17375	428044573	24636		
Corrected Total	17378	571761591			

Root MSE	156.95751	R-Square	0.2514
Dependent Mean	189.46309	Adj R-Sq	0.2512
Coeff Var	82.84332		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	175.81000	6.18696	28.42	<.0001
temp	1	362.53442	6.20493	58.43	<.0001
hum	1	-273.46511	6.46948	-42.27	<.0001
windspeed	1	26.31983	10.18011	2.59	0.0097

5. Based on point #3:

- If we hold all other independent variables constant, having one user increases the temperature by 362.53.
- If we hold all other independent variables constant, having one user decreases the humidity by -273.46.
- If we hold all other independent variables constant, having one user decreases the windspeed by 26.32.

- **From all the operation data, can temperature predict the number of casual users, and registered users?**

1. Overall model is significant

2.

casual is significant

registered is significant

3.

b0 = 0.42

b1 = 0.002

b2 = 0.0002

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: temp

Number of Observations Read	17379
Number of Observations Used	17379

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	145.22343	72.61172	2527.87	<.0001
Error	17376	499.11562	0.02872		
Corrected Total	17378	644.33905			

Root MSE	0.16948	R-Square	0.2254
Dependent Mean	0.49699	Adj R-Sq	0.2253
Coeff Var	34.10205		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.41570	0.00186	223.61	<.0001
casual	1	0.00152	0.00003024	50.33	<.0001
registered	1	0.00017544	0.00000985	17.81	<.0001

4. Adj-R = 0.2253 or 22.53% of the variability can be explained by the temperature, casual users, and registered users.

5. Based on point #3:

- If we hold all other independent variables constant, increasing our temperature by 1 will increase casual users by 0.002.
- If we hold all other independent variables constant, increasing our temperature by 1 will increase registered users by 0.0002.

Multiple Linear Regression Take Away

- Two regression analyses were run just to make sure we were not missing anything.
- On the first regression analysis, we can't just add one user and have our temperature and windspeed increase while humidity decrease. It is not possible as we cannot control the weather.
- In the second regression analysis, the result makes more sense because if the temperature increases there will be more people riding a bike. However, that is going to go until a certain point in the temperature. If the temperature is too high, it will be too hot outside, so people will not ride a bike. Therefore, as it gets warmer, more rental service will be used.