



Impact of Weather & Season on Yulu Sales: Using ANOVA

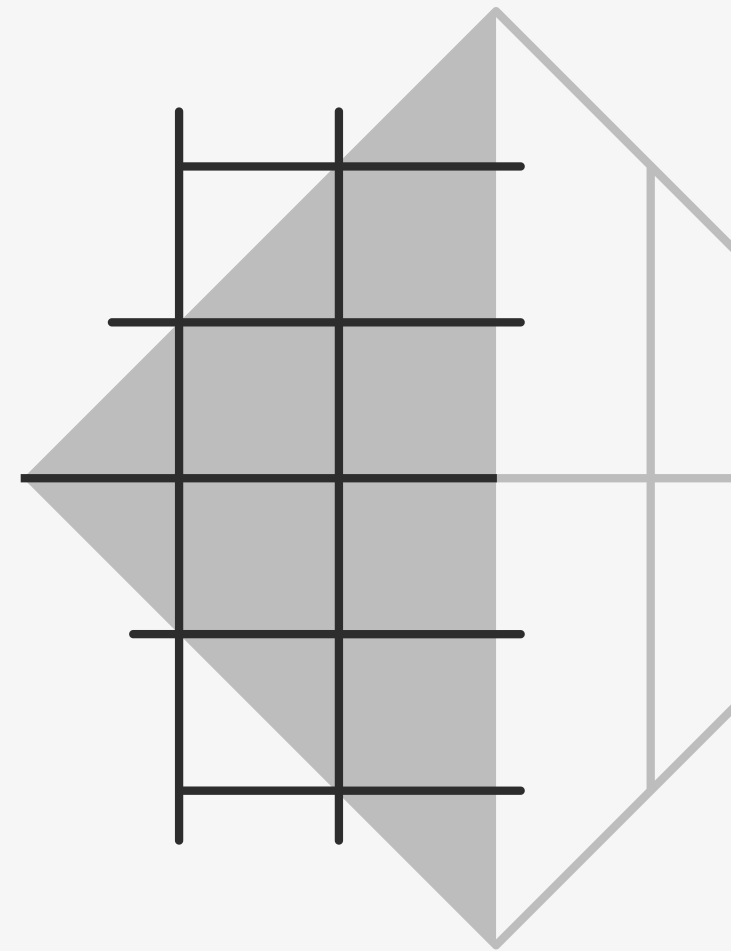
Soham Joshi

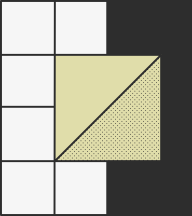
Understanding ANOVA (Analysis of Variance)

ANOVA is a statistical method used to analyze the differences among group means in a sample.

Types of ANOVA:

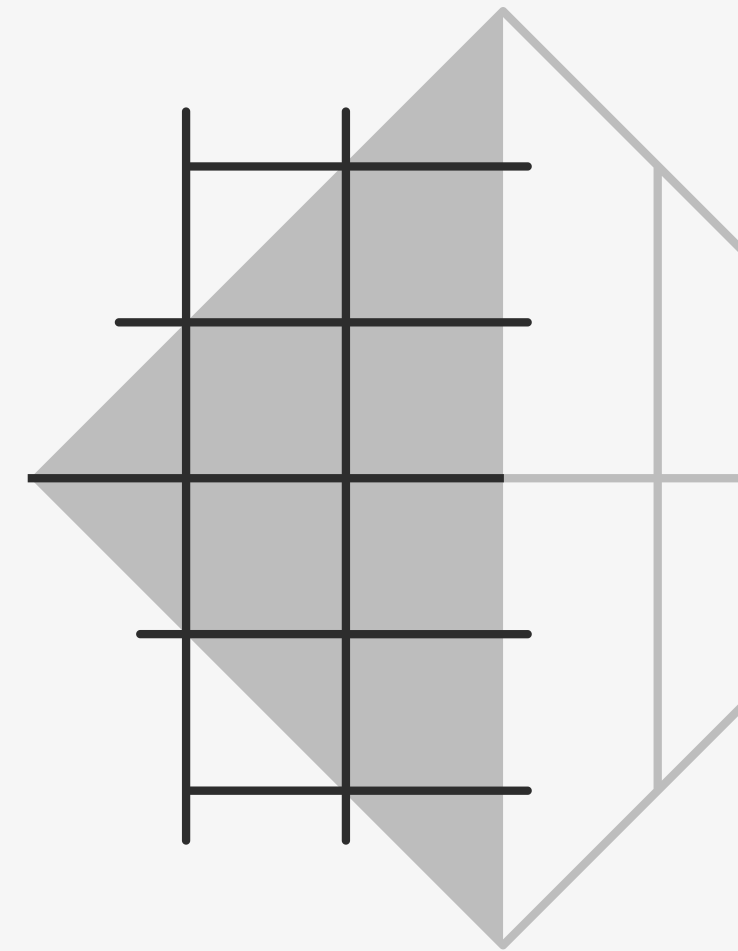
- One-Way ANOVA: Compares means across three or more independent groups.
- Two-Way ANOVA: Analyzes the effects of two categorical independent variables on a continuous dependent variable.






Understanding Factors Influencing Yulu Scooter Demand.

The dataset we have contains information about various things that might affect how many people use Yulu scooters in India. It includes details like the date and time, as well as factors like the weather, temperature, humidity, and wind speed, the season, and whether it's a holiday or not. By looking at all of this information together, we can figure out how different factors influence the number of Yulu scooter rides people take.

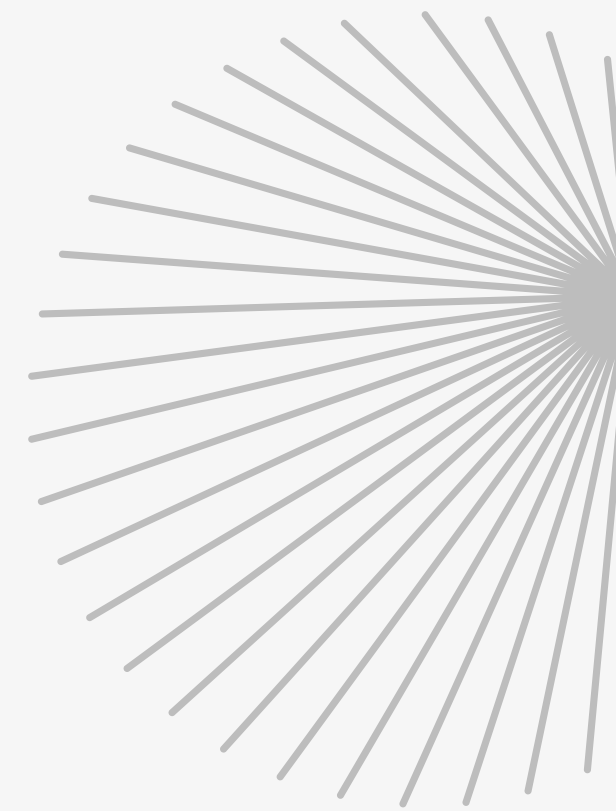




Which variables are significant in predicting the demand for shared electric cycles in the Indian market?



Determining the significant variables in predicting demand for shared electric cycles in India is important for market success, resource optimization, environmental sustainability, economic growth, policy formulation, and consumer welfare.

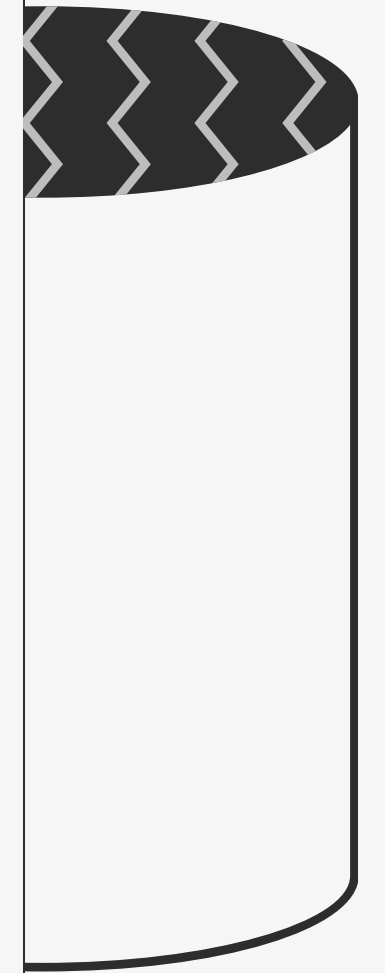


Data set

Index	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0	3	13	16
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0	8	32	40
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0	5	27	32
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0	3	10	13
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0	0	1	1
5	2011-01-01 05:00:00	1	0	0	2	9.84	12.88	75	6.0032	0	1	1
6	2011-01-01 06:00:00	1	0	0	1	9.02	13.635	80	0	2	0	2
7	2011-01-01 07:00:00	1	0	0	1	8.2	12.88	86	0	1	2	3
8	2011-01-01 08:00:00	1	0	0	1	9.84	14.395	75	0	1	7	8
9	2011-01-01 09:00:00	1	0	0	1	13.12	17.425	76	0	8	6	14
10	2011-01-01 10:00:00	1	0	0	1	15.58	19.695	76	16.9979	12	24	36



Holiday & Working day

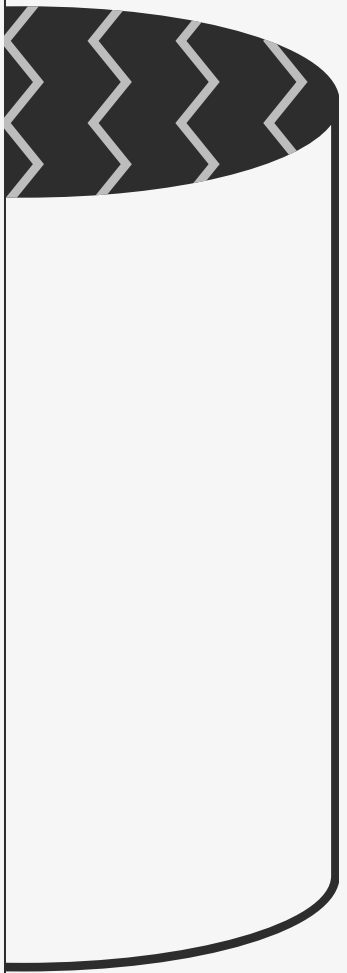


Holiday and Working day columns have binary values, which means only 2 classes (population) therefore to analyze the difference in variability it is most appropriate to use t-test.

The t-test is a statistical test used to determine if there is a significant difference between the means of two groups.



Temperature, Humidity & Windspeed



Temperature, Humidity & Windspeed are continuous random variables which need to be converted to categorical before applying ANNOVA.

Weather

```
58
59 #For Weather
60 #Defining Null and Alternative Hypothesis
61 #Null Hypothesis(H0): There is no significant difference due to Weather in Count of Vehicles Rented
62 #Alternative Hypothesis(H1): There is significant difference due to Weather(At least one pair)
63 uniqueweather = yuludf['weather'].unique() #Getting unique values aka Distinct classes
64 countbyweather = []
65 for weather in uniqueweather:
66     counts = yuludf[yuludf['weather'] == weather]['count']
67     countbyweather.append(counts)
68 #Creating list of every datapoint(Count) under each class(for each season)
69
70 #Printing ANOVA Table
71 model = ols('count ~ weather', data=yuludf).fit()
72 anova_table = sm.stats.anova_lm(model, typ=2)
73 print(anova_table)
74
75 #F tab value
76 f_tabulated = f.ppf(0.95,1,10884)
77 print("F-tabulated value:", f_tabulated)
78
79 #F cal
80 f_statistic, p_value = f_oneway(*countbyweather)
81 print("F-Statistic:", f_statistic)
82 print("P-Value:", p_value)
83
```


Weather

Output:

```
84 # =====
85 # #Output:
86 #           sum_sq      df          F      PR(>F)
87 # weather  5.911983e+06    1.0  183.185839  2.111106e-41
88 # Residual  3.512609e+08 10884.0      NaN      NaN
89 # F-tabulated value: 3.842313347173134
90 # F-Statistic: 65.53024112793271
91 # P-Value: 5.482069475935669e-42
92 # =====
93 #Decision Criterion: Reject H0 in favour of H1 at Alpha LOS iff Fcal>=Ftab
94 #Conclusion: We reject H0 in favour of H1 at LOS 0.05 Since Fcal> Ftab.
95 #Therefore we can say that there is some significant difference due to weather in count of rented vehicles.
96 #To further analyse which pairs of weather cause this variation, we must apply t test.
97
```

Season

```
18 #For Season
19 #Defining Null and Alternative Hypothesis
20 #Null Hypothesis(H0): There is no significant difference due to Season in Count of Vehicles Rented
21 #Alternative Hypothesis(H1): There is significant difference due to Season(At least one pair)
22 unique_seasons = yuludf['season'].unique() #Getting unique values aka Distinct classes
23 count_data_by_season = []
24 for season in unique_seasons:
25     counts = yuludf[yuludf['season'] == season]['count']
26     count_data_by_season.append(counts)
27 #Creating list of every datapoint(Count) under each class(for each season)
28
29 #Printing ANOVA Table
30 model = ols('count ~ season', data=yuludf).fit()
31 anova_table = sm.stats.anova_lm(model, typ=2)
32 print(anova_table)
33
34 #F tab value
35 f_tabulated = f.ppf(0.95,3,10884)
36 print("F-tabulated value:", f_tabulated)
37
38 #F cal
39 f_statistic, p_value = f_oneway(*count_data_by_season)
40 print("F-Statistic:", f_statistic)
41 print("P-Value:", p_value)
42
```

Season

Output:

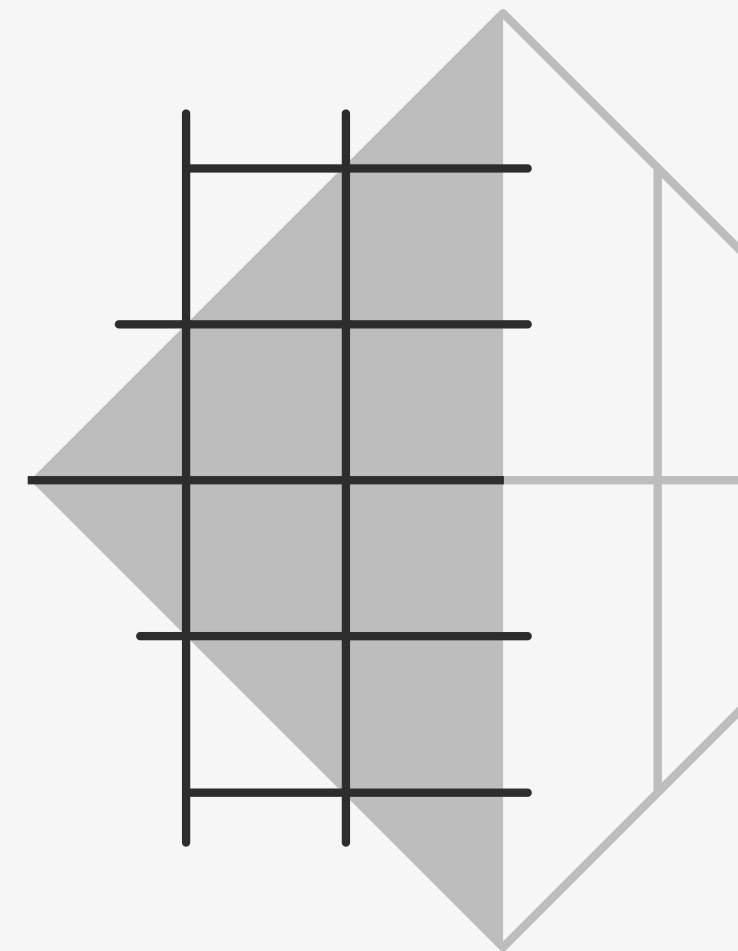
```
42
43 # =====
44 # #Output:
45 #           sum_sq      df          F      PR(>F)
46 # season    9.540914e+06      1.0  298.716206  4.758934e-66
47 # Residual  3.476320e+08  10884.0      NaN      NaN
48 # F-tabulated value: 2.6057248787213583
49 # F-Statistic: 236.94671081032106
50 # P-Value: 6.164843386499654e-149
51 # =====
52 #Decision Criterion: Reject H0 in favour of H1 at Alpha LOS iff Fcal>=Ftab
53 #Conclusion: We reject H0 in favour of H1 at LOS 0.05 Since Fcal> Ftab.
54 #Therefore we can say that there is some significant difference due to season in count of rented vehicles.
55 #To further analyse which pairs of seasons cause this variation, we must apply t test.
56
```

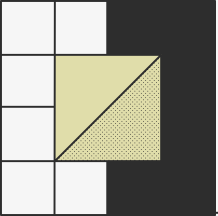


Conclusion

From the above analysis through ANNOVA, I have concluded that neither weather nor season have equal class effects. i.e. there is some significant difference in number of Yulu rented due to changes in weather and season.

To extract which season or weather class causes this significant difference we must perform t-test for pairs. After performing t-test and obtaining pairs that cost significant difference the company can implement measures to improve their sales accordingly.





Thank you

