Yulu Case study

Importing the required libraries for the dataset analysis

Out[188…		datetime	season	holiday	workingday	weather	temp	atemp	humid
	0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	
	1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	
	2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	
	3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	
	4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	
	...	...	...	...	...	...	...	...	
	10881	2012-12-19 19:00:00	4	0	1	1	15.58	19.695	
	10882	2012-12-19 20:00:00	4	0	1	1	14.76	17.425	
	10883	2012-12-19 21:00:00	4	0	1	1	13.94	15.910	
	10884	2012-12-19 22:00:00	4	0	1	1	13.94	17.425	
	10885	2012-12-19 23:00:00	4	0	1	1	13.12	16.665	

10886 rows × 12 columns

Out[189… (10886, 12)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10886 entries, 0 to 10885
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   datetime    10886 non-null  object
 1   season      10886 non-null  object
 2   holiday     10886 non-null  int64
 3   workingday  10886 non-null  int64
 4   weather     10886 non-null  int64
 5   temp        10886 non-null  float64
 6   atemp       10886 non-null  float64
 7   humidity    10886 non-null  int64
 8   windspeed   10886 non-null  float64
 9   casual      10886 non-null  int64
 10  registered  10886 non-null  int64
 11  count       10886 non-null  int64
dtypes: float64(3), int64(7), object(2)
memory usage: 1020.7+ KB
```

Out[191...
```
datetime      0
season        0
holiday       0
workingday    0
weather       0
temp          0
atemp         0
humidity      0
windspeed     0
casual        0
registered    0
count         0
dtype: int64
```
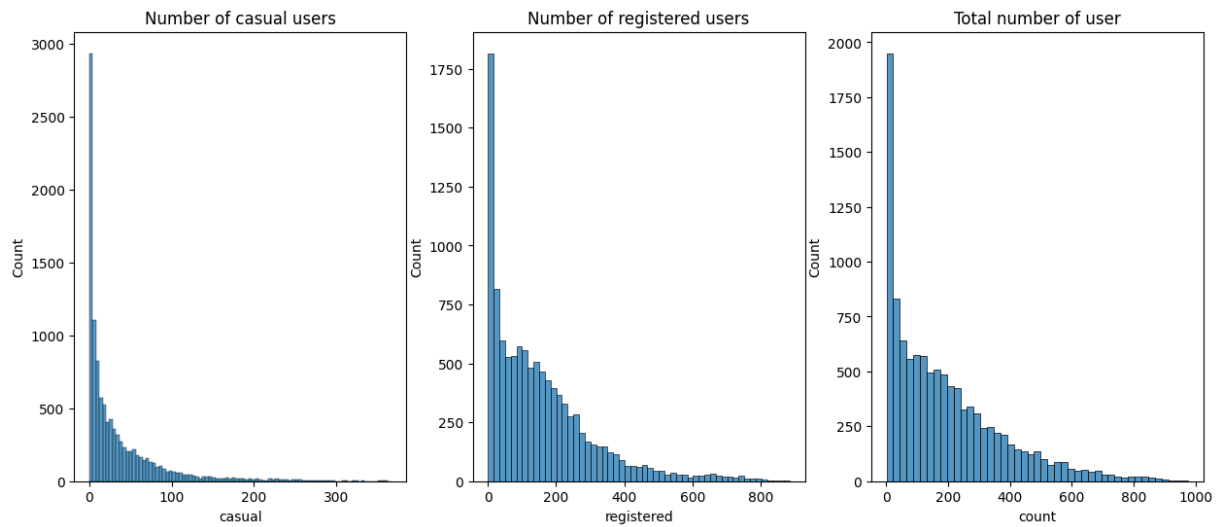
Out[192...

	holiday	workingday	weather	temp	atemp	
count	10886.000000	10886.000000	10886.000000	10886.00000	10886.000000	1
mean	0.028569	0.680875	1.418427	20.23086	23.655084	
std	0.166599	0.466159	0.633839	7.79159	8.474601	
min	0.000000	0.000000	1.000000	0.82000	0.760000	
25%	0.000000	0.000000	1.000000	13.94000	16.665000	
50%	0.000000	1.000000	1.000000	20.50000	24.240000	
75%	0.000000	1.000000	2.000000	26.24000	31.060000	
max	1.000000	1.000000	4.000000	41.00000	45.455000	

Out[193...   0

	datetime	season	holiday	workingday	weather	temp	atemp	humidity
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75

	datetime	season	holiday	workingday	weather	temp	atemp	humid
5	2011-01-01 05:00:00	1	0	0	2	9.84	12.880	
13	2011-01-01 13:00:00	1	0	0	2	18.86	22.725	
14	2011-01-01 14:00:00	1	0	0	2	18.86	22.725	
15	2011-01-01 15:00:00	1	0	0	2	18.04	21.970	
16	2011-01-01 16:00:00	1	0	0	2	17.22	21.210	
...	...	...	...	...	...	...	...	
10837	2012-12-17 23:00:00	4	0	1	3	17.22	21.210	
10838	2012-12-18 00:00:00	4	0	1	2	18.04	21.970	
10839	2012-12-18 01:00:00	4	0	1	2	18.04	21.970	
10840	2012-12-18 02:00:00	4	0	1	2	18.04	21.970	
10850	2012-12-18 12:00:00	4	0	1	3	19.68	23.485	

3694 rows × 14 columns

	datetime	temp	atemp	humidity	windspeed	casual	registered	cou
0	2011-01-01 00:00:00	9.84	14.395	81	0.0000	3	13	
1	2011-01-01 01:00:00	9.02	13.635	80	0.0000	8	32	
2	2011-01-01 02:00:00	9.02	13.635	80	0.0000	5	27	
3	2011-01-01 03:00:00	9.84	14.395	75	0.0000	3	10	
4	2011-01-01 04:00:00	9.84	14.395	75	0.0000	0	1	
...	...	...	...	...	...	...	...	
10881	2012-12-19 19:00:00	15.58	19.695	50	26.0027	7	329	3
10882	2012-12-19 20:00:00	14.76	17.425	57	15.0013	10	231	2
10883	2012-12-19 21:00:00	13.94	15.910	61	15.0013	4	164	1
10884	2012-12-19 22:00:00	13.94	17.425	61	6.0032	12	117	1
10885	2012-12-19 23:00:00	13.12	16.665	66	8.9981	4	84	

10886 rows × 11 columns

<matplotlib.legend.Legend at 0x7d406ec665f0>

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

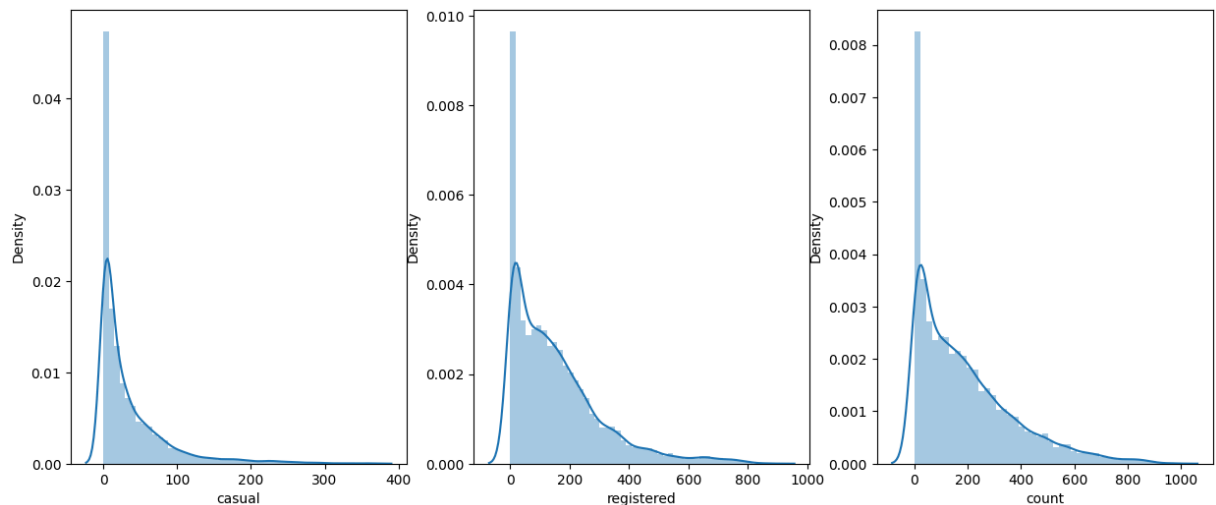|	datetime	temp	atemp	humidity	windspeed	casual	registered	count	
0	2011-01-01 00:00:00	9.84	14.395	81	0.0	3	13	16	h
1	2011-01-01 01:00:00	9.02	13.635	80	0.0	8	32	40	h
2	2011-01-01 02:00:00	9.02	13.635	80	0.0	5	27	32	h
3	2011-01-01 03:00:00	9.84	14.395	75	0.0	3	10	13	h
4	2011-01-01 04:00:00	9.84	14.395	75	0.0	0	1	1	h

```
<ipython-input-206-72467a08eb09>:5: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['casual'])
<ipython-input-206-72467a08eb09>:7: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['registered'])
<ipython-input-206-72467a08eb09>:9: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['count'])
```
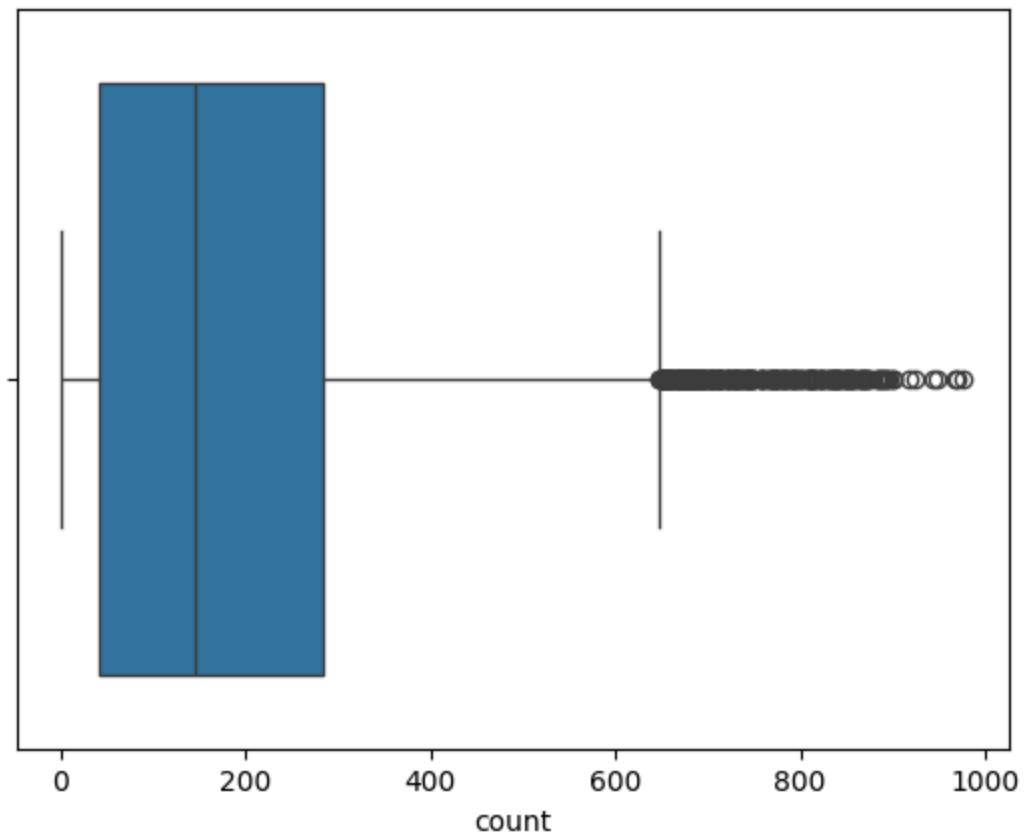


```
Out[207… <Axes: xlabel='count'>
```

Out[208...   42.0

Out[209...   284.0

Out[211...   (647.0, -321.0)

	datetime	temp	atemp	humidity	windspeed	casual	registered	cou
6611	2012-03-12 18:00:00	24.60	31.060	43	12.9980	89	623	7
6634	2012-03-13 17:00:00	28.70	31.820	37	7.0015	62	614	6
6635	2012-03-13 18:00:00	28.70	31.820	34	19.9995	96	638	7
6649	2012-03-14 08:00:00	18.04	21.970	82	0.0000	34	628	6
6658	2012-03-14 17:00:00	28.70	31.820	28	6.0032	140	642	7
...	...	...	...	...	...	...	...	
10678	2012-12-11 08:00:00	13.94	15.150	61	19.9995	16	708	7
10702	2012-12-12 08:00:00	10.66	12.880	65	11.0014	18	670	6
10726	2012-12-13 08:00:00	9.84	11.365	60	12.9980	24	655	6
10846	2012-12-18 08:00:00	15.58	19.695	94	0.0000	10	652	6
10870	2012-12-19 08:00:00	9.84	12.880	87	7.0015	13	665	6

300 rows × 11 columns

	count	count_z	count_minmax
0	16	-0.969294	0.015369
1	40	-0.836797	0.039959
2	32	-0.880962	0.031762
3	13	-0.985856	0.012295
4	1	-1.052104	0.000000
...	...	...	...
10881	336	0.797333	0.343238
10882	241	0.272866	0.245902
10883	168	-0.130146	0.171107
10884	129	-0.345454	0.131148
10885	88	-0.571803	0.089139

10886 rows × 3 columns

	count	count_z	count_minmax
6658	782	3.259570	0.800205
6659	749	3.077386	0.766393
6683	746	3.060824	0.763320
6779	801	3.364463	0.819672
6849	757	3.121552	0.774590
...	...	...	...
9935	834	3.546647	0.853484
9944	890	3.855806	0.910861
9945	788	3.292694	0.806352
10519	743	3.044262	0.760246
10534	759	3.132593	0.776639

147 rows × 3 columns

Plot a Correlation Heatmap and draw insights.

```
<ipython-input-216-adc3370a68be>:3: FutureWarning: The default value of nume
ric_only in DataFrame.corr is deprecated. In a future version, it will defau
lt to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
```
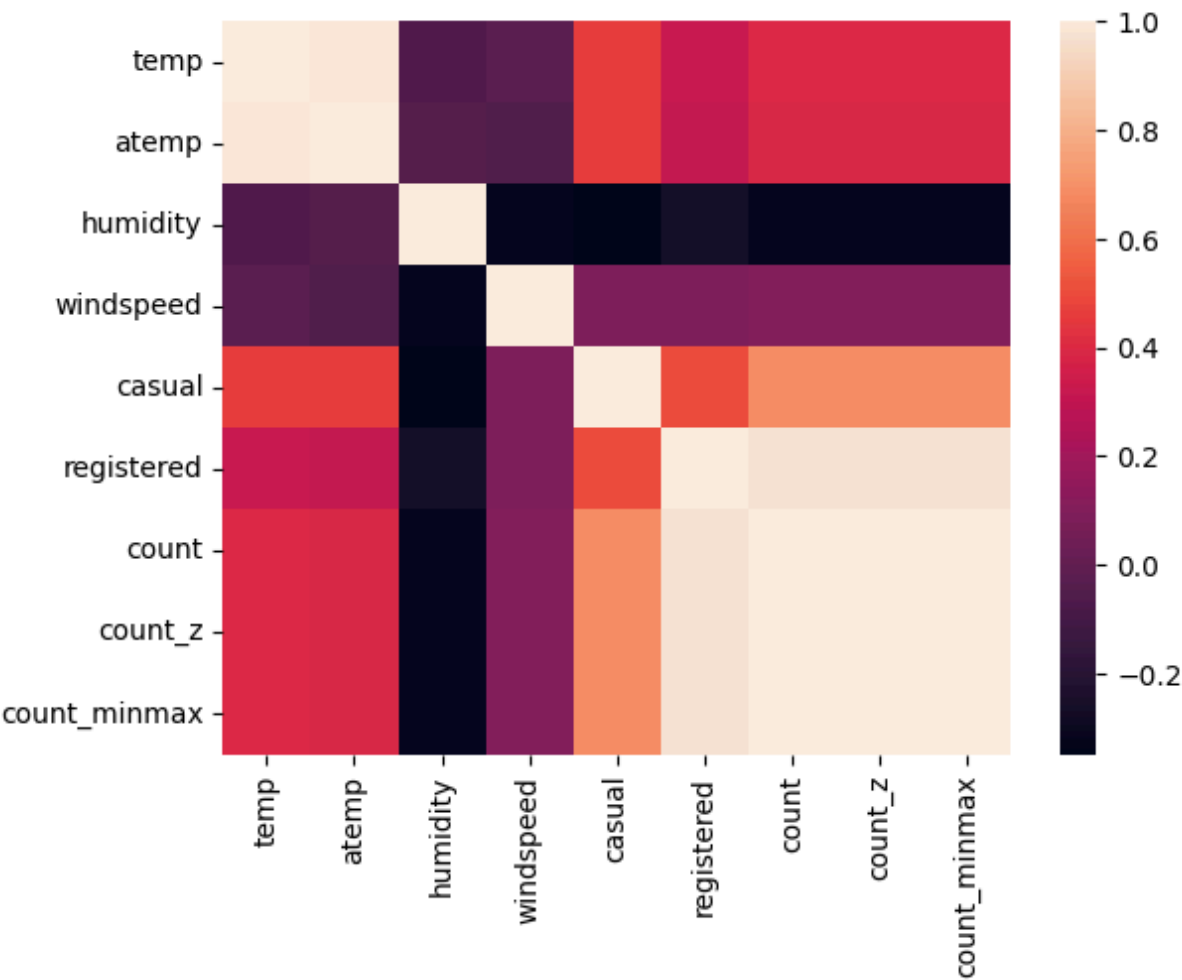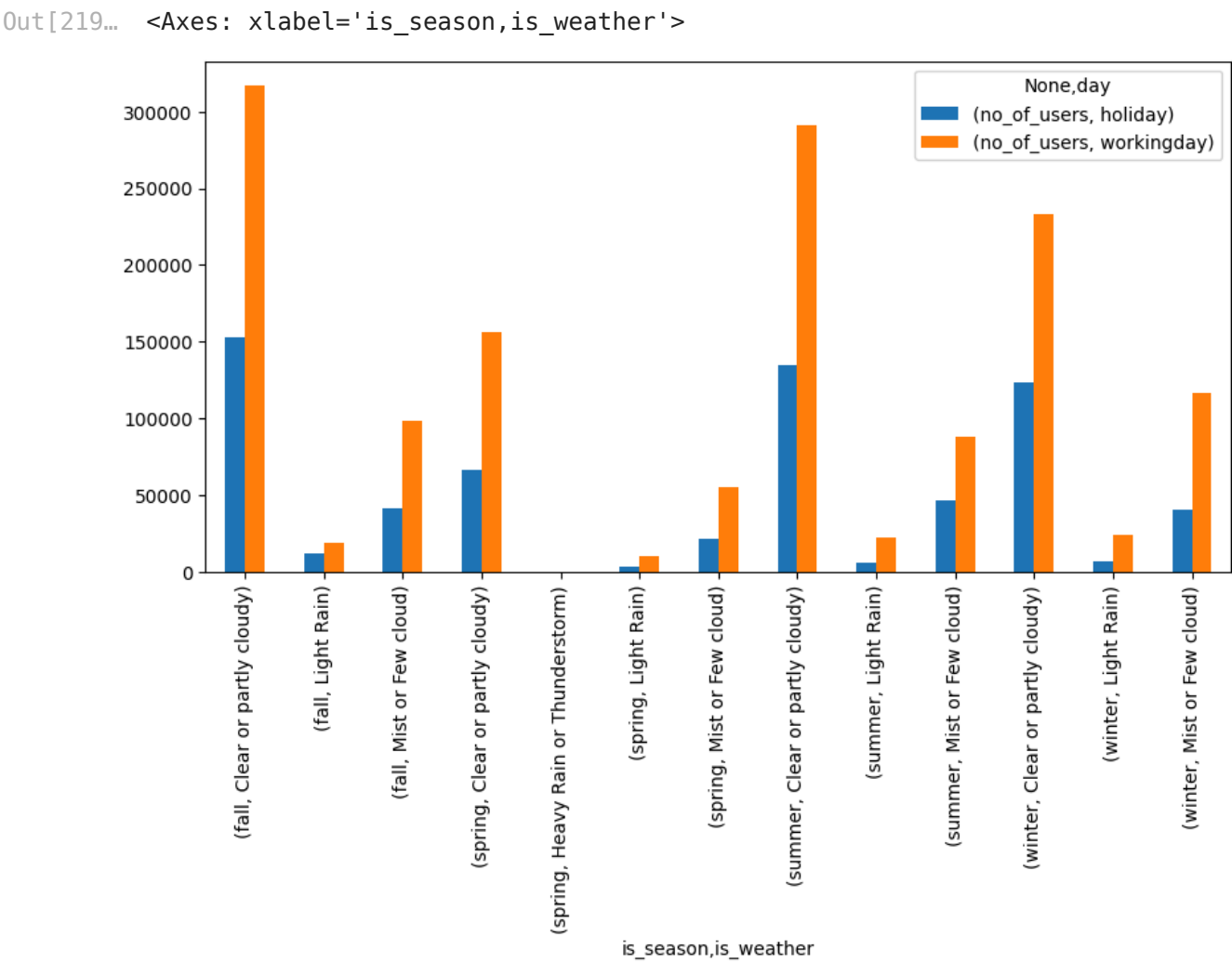
	temp	atemp	humidity	windspeed	casual	registere
temp	1.000000	0.984948	-0.064949	-0.017852	0.467097	0.3185
atemp	0.984948	1.000000	-0.043536	-0.057473	0.462067	0.3146
humidity	-0.064949	-0.043536	1.000000	-0.318607	-0.348187	-0.2654
windspeed	-0.017852	-0.057473	-0.318607	1.000000	0.092276	0.0910
casual	0.467097	0.462067	-0.348187	0.092276	1.000000	0.4972
registered	0.318571	0.314635	-0.265458	0.091052	0.497250	1.0000
count	0.394454	0.389784	-0.317371	0.101369	0.690414	0.9709
count_z	0.394454	0.389784	-0.317371	0.101369	0.690414	0.9709
count_minmax	0.394454	0.389784	-0.317371	0.101369	0.690414	0.9709

	datetime	temp	atemp	humidity	windspeed	casual	registered	count	
0	2011-01-01 00:00:00	9.84	14.395	81	0.0	3	13	16	h
1	2011-01-01 01:00:00	9.02	13.635	80	0.0	8	32	40	h
2	2011-01-01 02:00:00	9.02	13.635	80	0.0	5	27	32	h
3	2011-01-01 03:00:00	9.84	14.395	75	0.0	3	10	13	h
4	2011-01-01 04:00:00	9.84	14.395	75	0.0	0	1	1	h

Working Day has effect on number of electric cycles rented

<Axes: xlabel='is_season,is_weather'>

	datetime	temp	atemp	humidity	windspeed	casual	registered	cou
47	2011-01-03 00:00:00	9.02	9.850	44	23.9994	0	5	
48	2011-01-03 01:00:00	8.20	8.335	44	27.9993	0	2	
49	2011-01-03 04:00:00	6.56	6.820	47	26.0027	0	1	
50	2011-01-03 05:00:00	6.56	6.820	47	19.0012	0	3	
51	2011-01-03 06:00:00	5.74	5.305	50	26.0027	0	30	
...	...	...	...	...	...	...	...	
10881	2012-12-19 19:00:00	15.58	19.695	50	26.0027	7	329	3
10882	2012-12-19 20:00:00	14.76	17.425	57	15.0013	10	231	2
10883	2012-12-19 21:00:00	13.94	15.910	61	15.0013	4	164	1
10884	2012-12-19 22:00:00	13.94	17.425	61	6.0032	12	117	1
10885	2012-12-19 23:00:00	13.12	16.665	66	8.9981	4	84	

7412 rows × 13 columns

	datetime	temp	atemp	humidity	windspeed	casual	registered	cou
0	2011-01-01 00:00:00	9.84	14.395	81	0.0000	3	13	
1	2011-01-01 01:00:00	9.02	13.635	80	0.0000	8	32	
2	2011-01-01 02:00:00	9.02	13.635	80	0.0000	5	27	
3	2011-01-01 03:00:00	9.84	14.395	75	0.0000	3	10	
4	2011-01-01 04:00:00	9.84	14.395	75	0.0000	0	1	
...	...	...	...	...	...	...	...	
10809	2012-12-16 19:00:00	14.76	17.425	93	8.9981	10	99	1
10810	2012-12-16 20:00:00	15.58	19.695	82	0.0000	14	108	1
10811	2012-12-16 21:00:00	14.76	18.940	93	0.0000	14	92	1
10812	2012-12-16 22:00:00	16.40	20.455	82	12.9980	6	83	
10813	2012-12-16 23:00:00	14.76	17.425	93	8.9981	4	29	

3474 rows × 13 columns

1. Formulating the null and Alternative Hypothesis for the data of bike rentals and the type of day.

2. `Marking the significance level of 0.05 (confidence of 95%)

3. h0 : day has no effect on bike rentals.

4. ha : day has effect on bike rentals.

5. alpha = 0.05 -- 95% confidence.

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

`TtestResult(statistic=1.1826836792249074, pvalue=0.24428177337195203, df=3 8.0)`

1. The p_val we observed in the T test individual is 0.58(varies) which is greater than the alpha(significance level) of 0.05.
2. Hence we cannot reject the null and the workingday and holiday has no significant affect on the bike rentals.

is_weather	temp	atemp	humidity	windspeed	casual	registere
Clear or partly cloudy	147846.82	172565.755	407907	92723.1626	289900	118616
Heavy Rain or Thunderstorm	8.20	11.365	86	6.0032	6	15
Light Rain	16790.32	19544.905	69872	12087.2020	14983	8710
Mist or Few cloud	55587.80	65387.220	195831	34517.8506	87246	41991

Check if the demand of bicycles on rent is the same for different Weather conditions?

1. Formulating the null and alternqative hypothesis for data of bike rentals and weather conditions.
2. Marking the significance level of 0.05(95 % confidence)
3. H0 : Weather conditions has effect on bike rentals.
4. Ha : Weather conditions has no effect on tbike rentals.
5. alpha : 0.05 -- 95% confidence

`<Axes: xlabel='count', ylabel='Count'>`

<Figure size 1500x600 with 0 Axes>
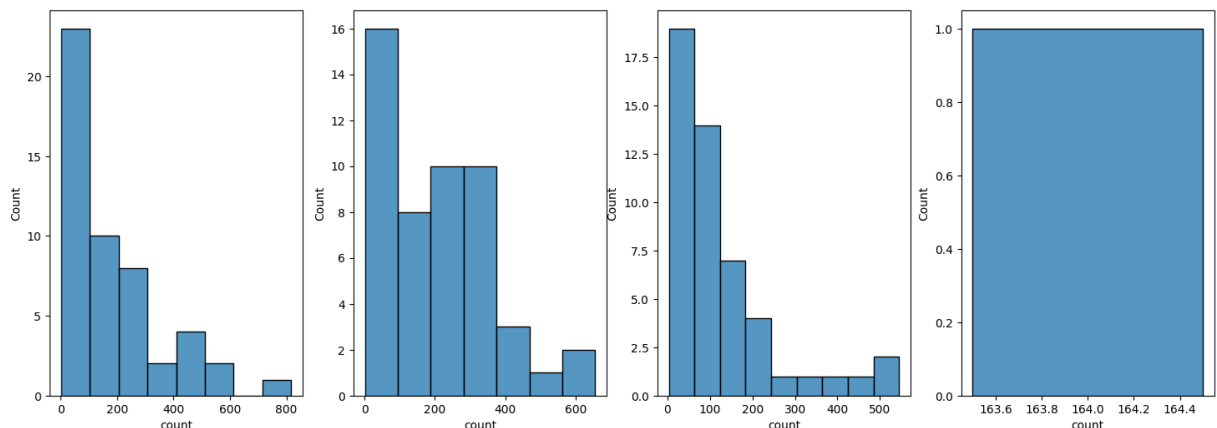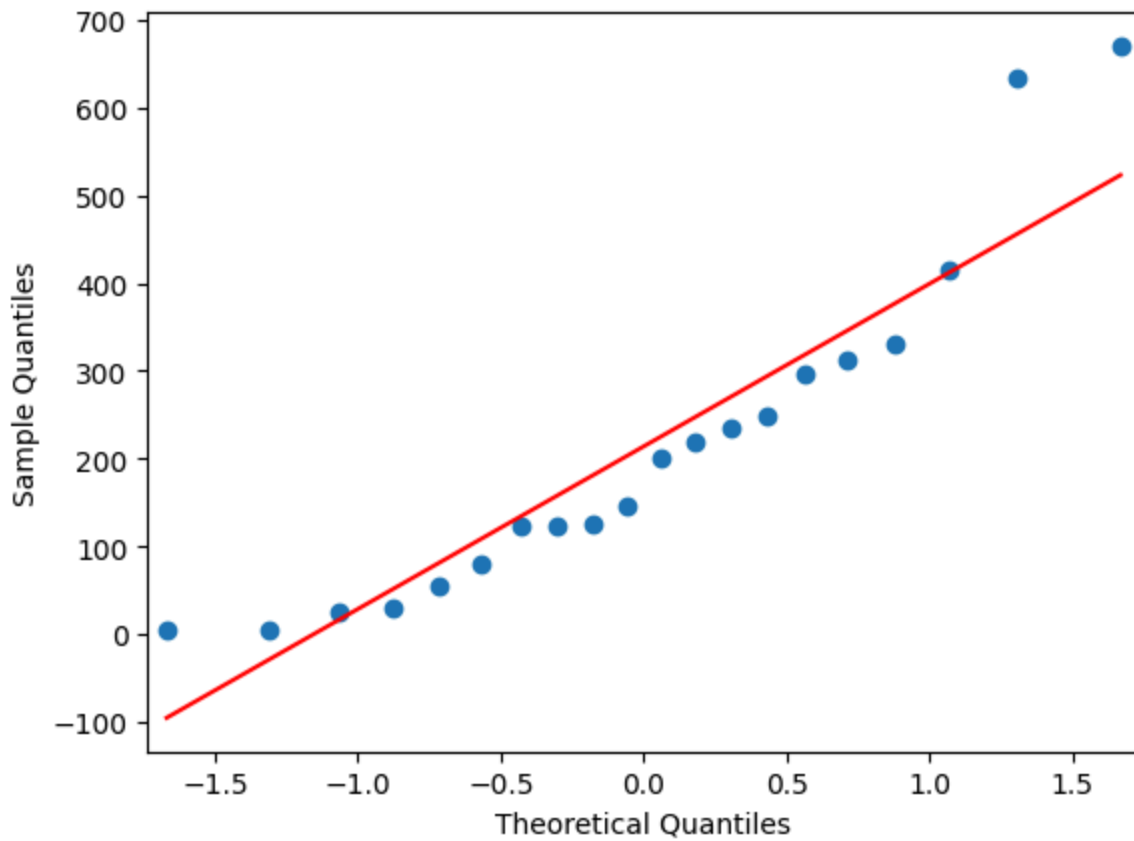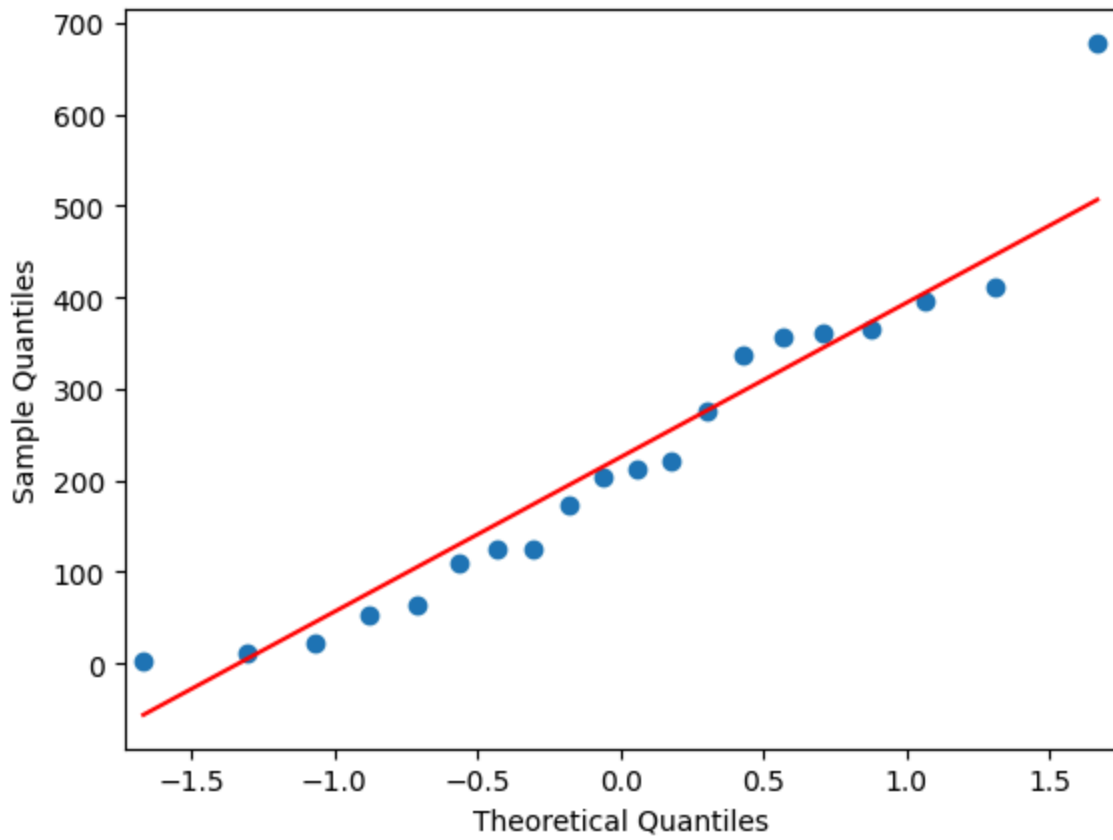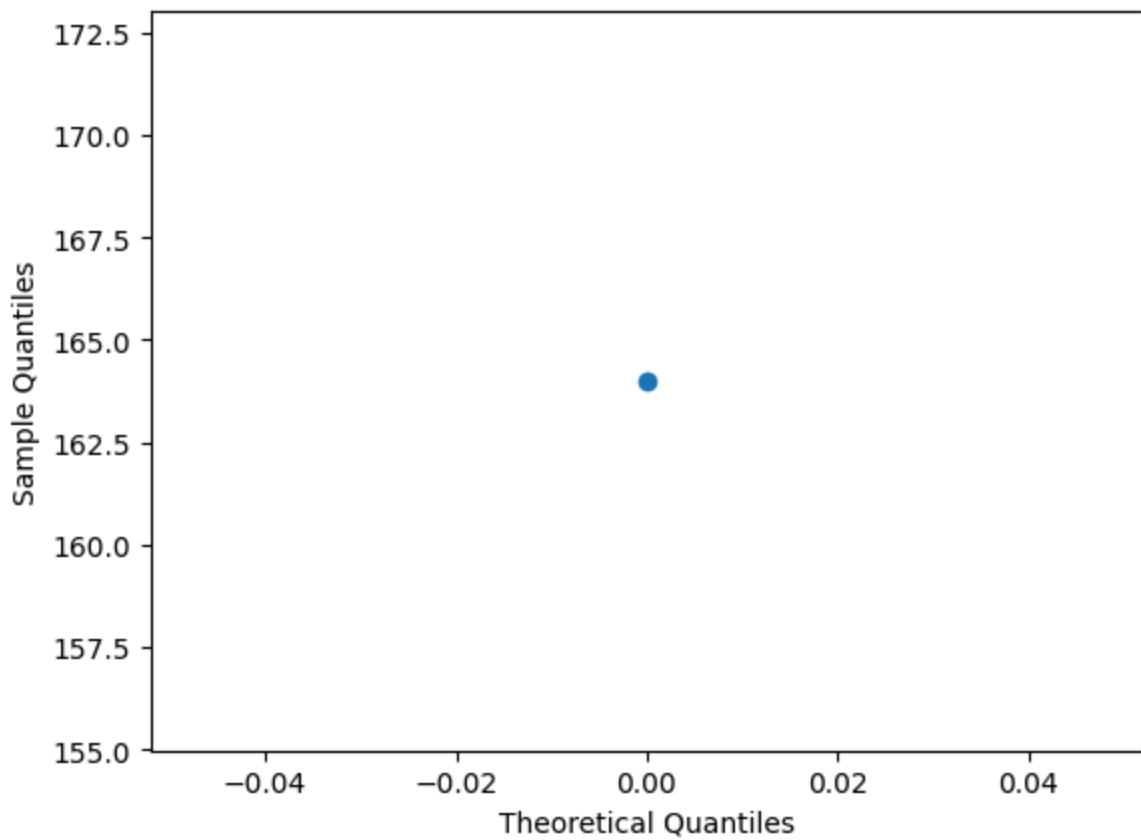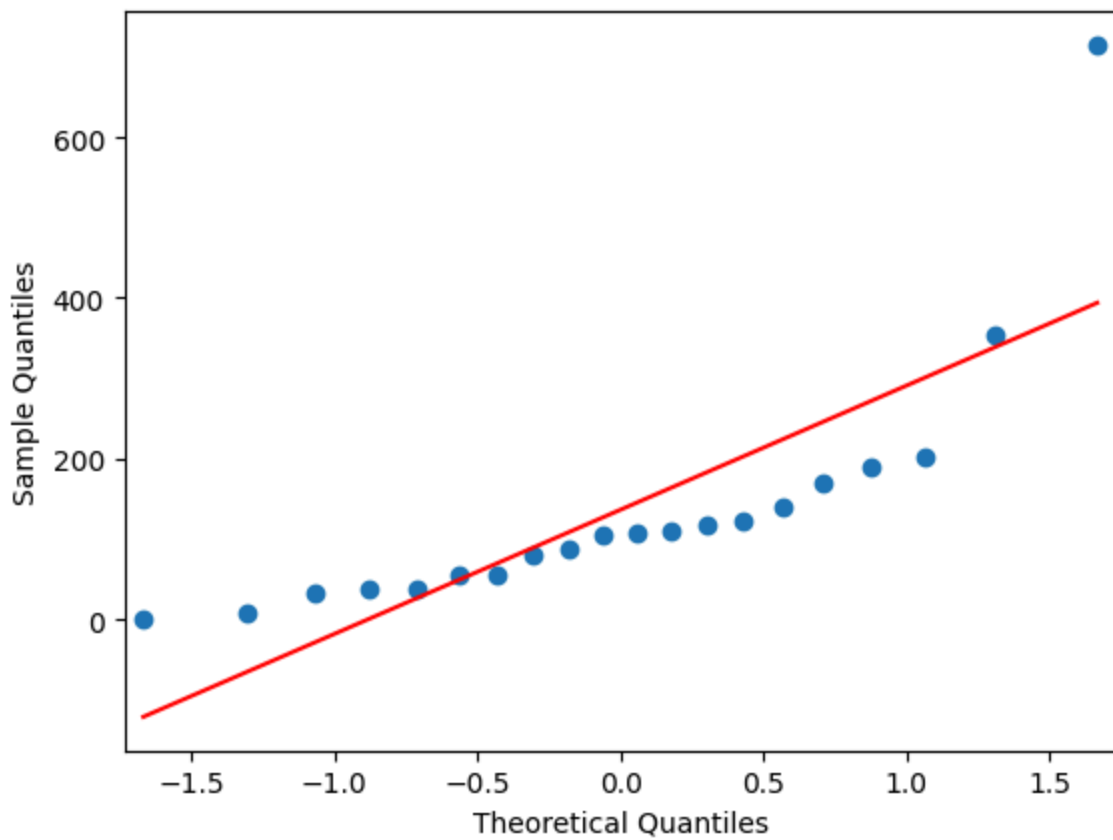
Out[231… ' We can ignore this data as there is only one data point for this filter o
f heavyrain in the df '

Out[232… `LeveneResult(statistic=1.8405200404644044, pvalue=0.14235047295888245)`

Out[233… `KruskalResult(statistic=8.862939039863937, pvalue=0.031169797575285513)`

Out[234… `F_onewayResult(statistic=2.548582152275388, pvalue=0.05810771832902403)`

1. The p _val is less than alpha value(0.05).
2. We can reject null hypothesis and coclude bike rentals depends on weather conditions.

Out[235… `0.14603275060653687`

As the p_value is very smaller and away from 1, we can conclude that the distribution is not normal

Out[236… `0.018626874312758446`

As the p_value is very smaller and away from 1, we can conclude that the distribution is not normal

Out[237… `2.068323919957038e-05`

As the p_value is very smaller and away from 1, we can conclude that the distribution is not normal

.

Check if the demand of bicycles on rent is the same for different Seasons?

	datetime	temp	atemp	humidity	windspeed	casual	registered	cou
0	2011-01-01 00:00:00	9.84	14.395	81	0.0000	3	13	
1	2011-01-01 01:00:00	9.02	13.635	80	0.0000	8	32	
2	2011-01-01 02:00:00	9.02	13.635	80	0.0000	5	27	
3	2011-01-01 03:00:00	9.84	14.395	75	0.0000	3	10	
4	2011-01-01 04:00:00	9.84	14.395	75	0.0000	0	1	
...	...	...	...	...	...	...	...	
10881	2012-12-19 19:00:00	15.58	19.695	50	26.0027	7	329	3
10882	2012-12-19 20:00:00	14.76	17.425	57	15.0013	10	231	2
10883	2012-12-19 21:00:00	13.94	15.910	61	15.0013	4	164	1
10884	2012-12-19 22:00:00	13.94	17.425	61	6.0032	12	117	1
10885	2012-12-19 23:00:00	13.12	16.665	66	8.9981	4	84	

10886 rows × 13 columns

`<Axes: xlabel='count', ylabel='Count'>`
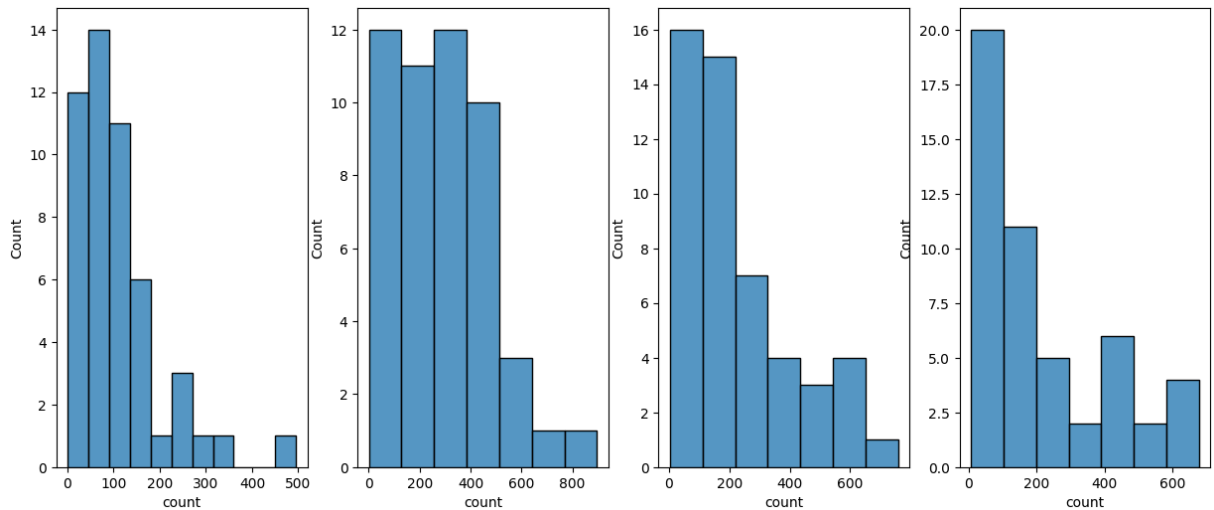
1. Formulating the NUll and Alternative Hypothesis for the Seasos data.
2. H0 : bike rentals are equal in all seasons.
3. Ha : bike rentals are not equal in all seasons.

Out[241... `LeveneResult(statistic=6.97282490618, pvalue=0.00017584915575216578)`

Out[242... `KruskalResult(statistic=20.675205459827282, pvalue=0.00012295873453147527)`

Out[243... `F_onewayResult(statistic=7.945570705995613, pvalue=5.007287903641841e-05)`

as p_val is lesser than alpha(0.05) we can reject null and conclude that bike rentals are effected by seasons.
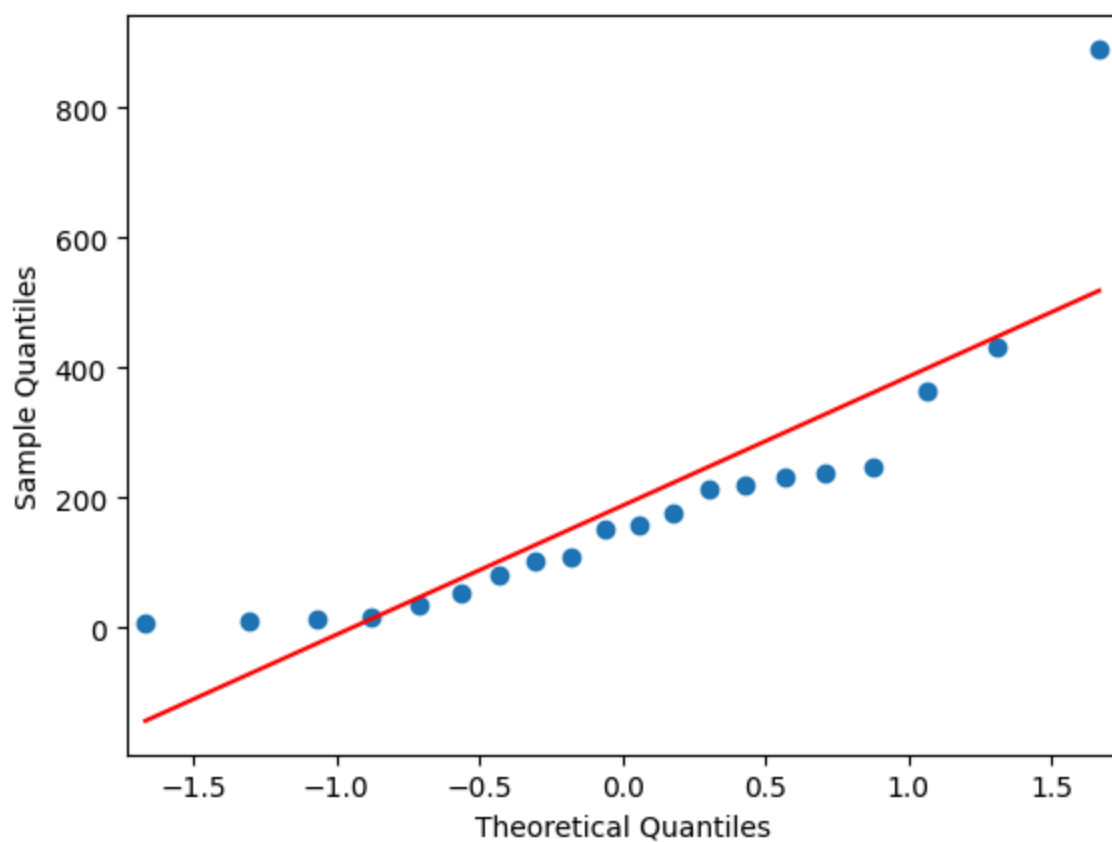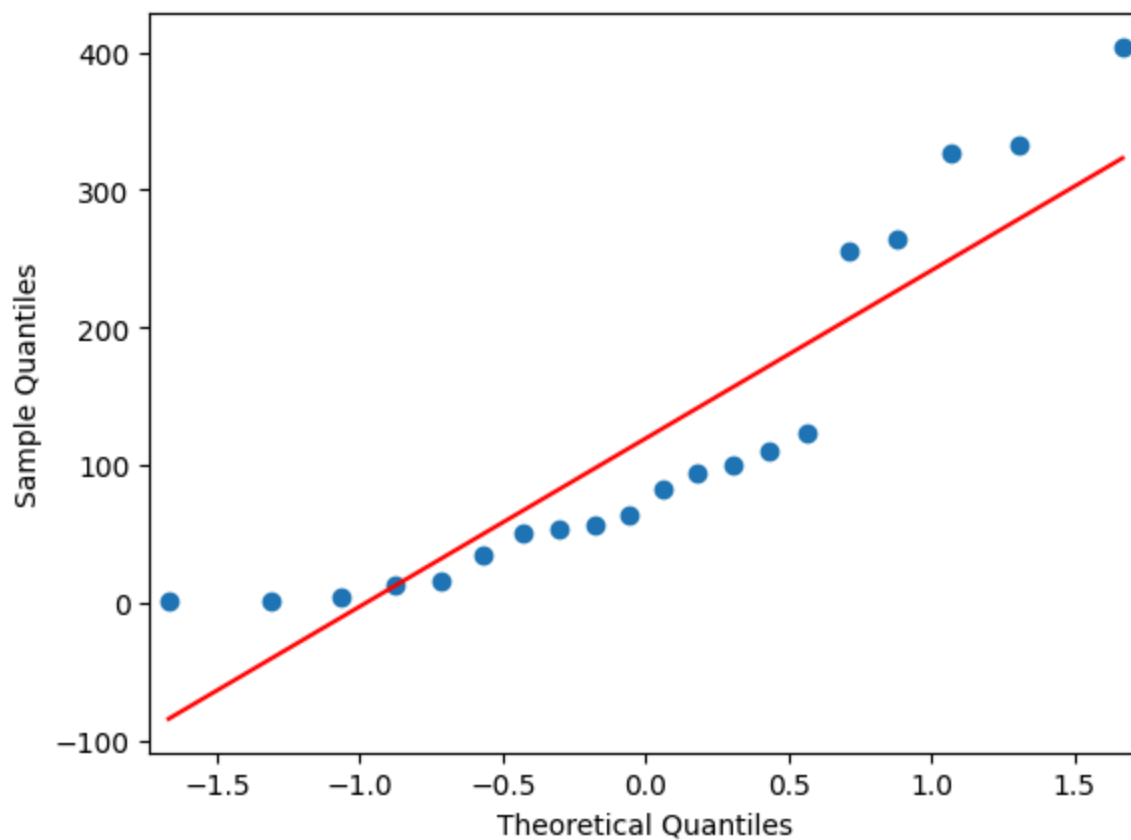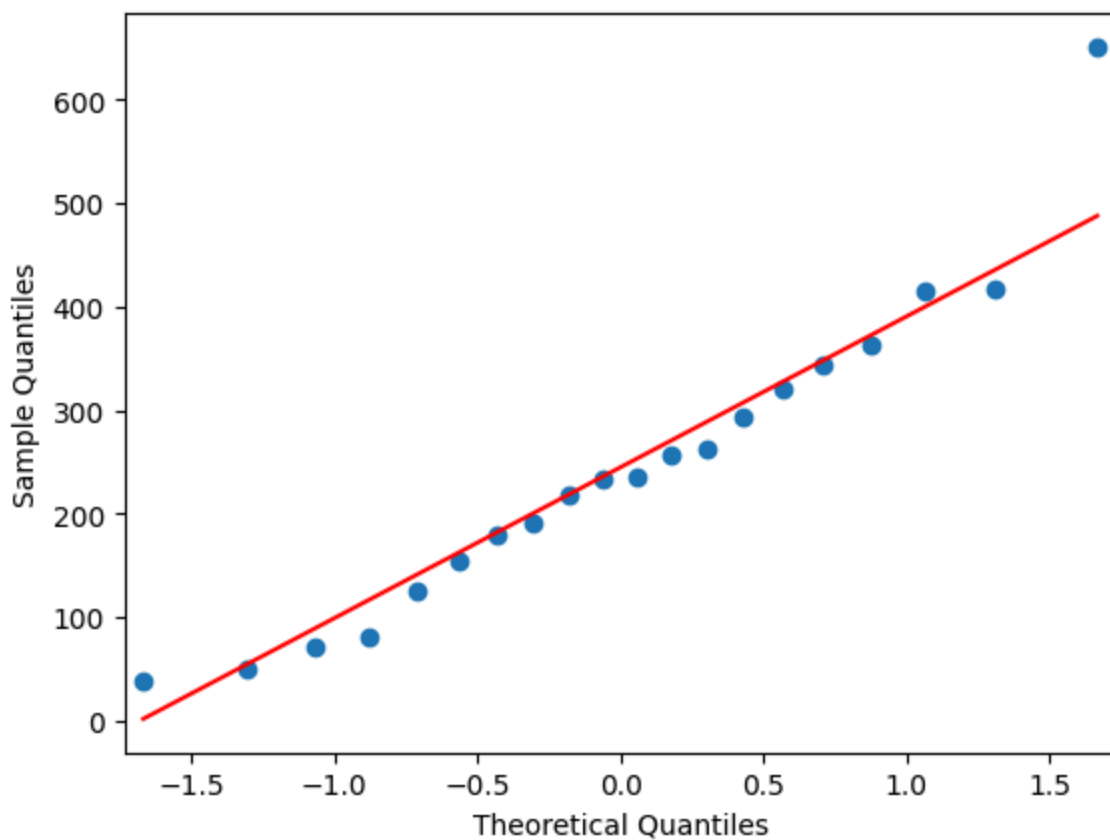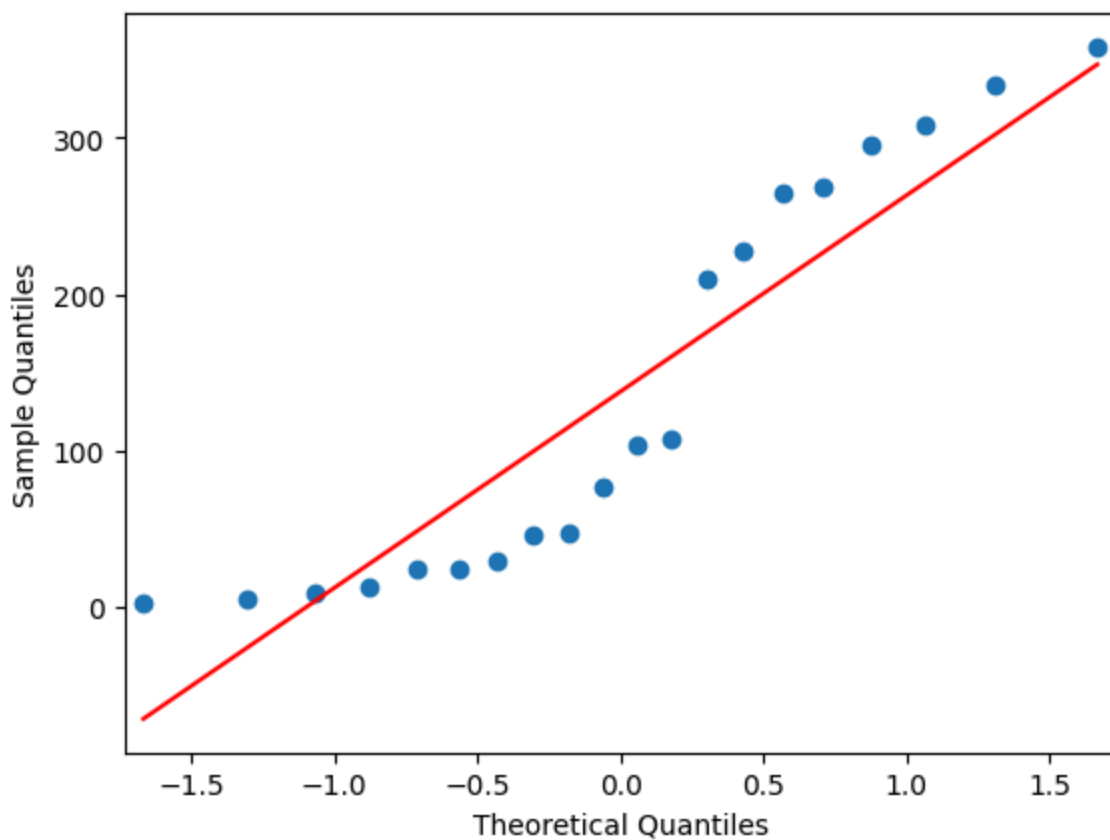
Out[250… 0.00024950012448243797

Out[251… 0.005241308361291885

Out[252… 0.25809207558631897

As the p_ values for each seasons for saamples of 20 each is very small i.e far from 1 that states the distribution is not normal.

.

Check if the demand of bicycles on rent is the same for different Seasons?

Out[253…

is_weather	Clear or partly cloudy	Heavy Rain or Thunderstorm	Light Rain	Mist or Few cloud	All
is_season					
fall	1930	0	199	604	2733
spring	1759	1	211	715	2686
summer	1801	0	224	708	2733
winter	1702	0	225	807	2734
All	7192	1	859	2834	10886

1. Formulating the Null and Alternative Hypothesis for the data of weather and season categorical columns.
2. H0 : Weather conditions are not significantly different over each season.
3. Ha : Weather conditions are significantly different over each season.
4. sigificance level -- 0.05 (95 % confidence)

Out[254… Chi2ContingencyResult(statistic=46.10145731073249, pvalue=2.8260014509929343e-08, dof=6, expected_freq=array([[1805.76352779, 215.67726229, 711.55920992],
       [1774.04869086, 211.8892972 , 699.06201194],
       [1805.76352779, 215.67726229, 711.55920992],
       [1806.42425356, 215.75617823, 711.81956821]]))

1. The p_val obtained is less then the alpha(0.05) value.
2. We can reject the null hypothesis and suport the alternative hypothesis. -- There is a siginificant change in weather conditions over each season.

Out[255… <matplotlib.legend.Legend at 0x7d406e743310>

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js
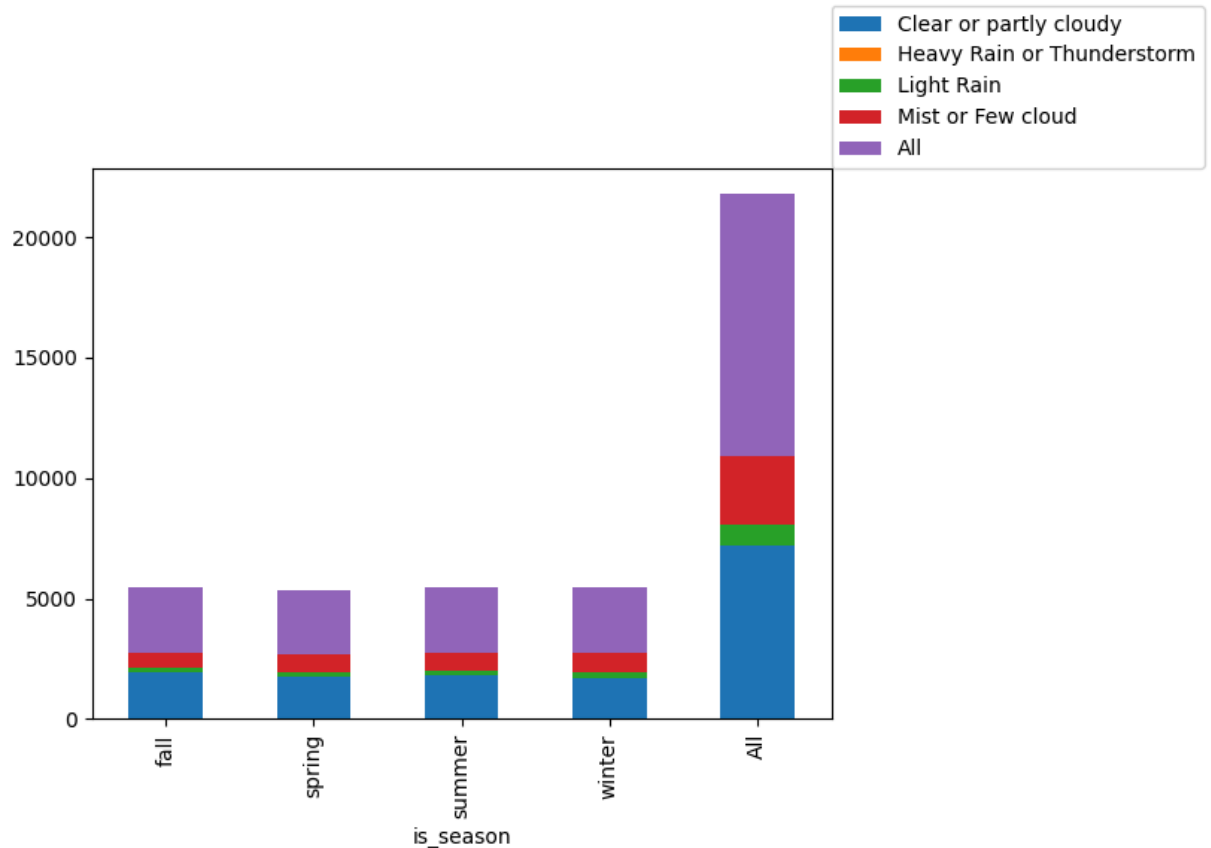
Recommendations

1. The season plays a vital role in the bike rentals and number of users.
2. There should be siginificant bike avalability on the holidays as far as at the pickup and drop points.
3. The weather conditons also has major affect on the users of bike rentals.
4. The registered users are also not much greater then the csual users, there should be necessary steps taken to attarct the users for registering into the app by offering discount rides, weather forecasting and other customer gaining techniques.
5. Most of the people won't get the bike for rentqals at peal hours of morning,afternoon and evening of working days and holidays, there should be effort to resolve this issues,
6. The seasonal weather conditions are very important for rental of the bike.
7. The temperature and humidity also effects the bike rentals, there shuld be enough forecasting data over a period for specific region of every peak point of the region inorder to help the customers for the weather conditions and other predictions of peak hour,free hours.
8. Making the customer satisfaction is key for the this bussiness as it runs on the service offered to the customers.

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

9. Having a right and timed deartments for every aspect of the company lead to the success fro long years of the company.
10. These are few recomendations from my side. Thank you.

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js