Identify the effects of weather condition on the bike sharing system

Siwei Xiao

Aug. 11th , 2022

EXECUTIVE SUMMARY

1. Painpoint

The growth of sharing bicycle markets is high with **there has been over 600 cities building up Bike Sharing Systems (BSSs),** but **'rebalancing problem'** is obvious in which the demands may not be met when needed due to geographical or climate reasons (Earth Policy Institute, 2015).

2. Context

According to Government UK, the **private transportation accounts for over 90 percent of private travels**, within which **bike travelling takes up about 20 over 694 travels,** showing an increasing trend (GOV. UK, 2021). Besides, due to the economic environment influence, **the energy price has been soring**, implying a negative effect on private car travels (Bloomberg, 2022).

3. Analysis

This report uses data from the weather in London from 2015 to 2016 including data columns **of weather type, date source, seasonal code, weekend day, windspeed, humidity, and temperature**. The statistical analysis and regression analysis will be applied to figure out the relationship between the usage of sharing bikes and weather condition.

PUBLIC TRANSPORTATION USAGE TREND

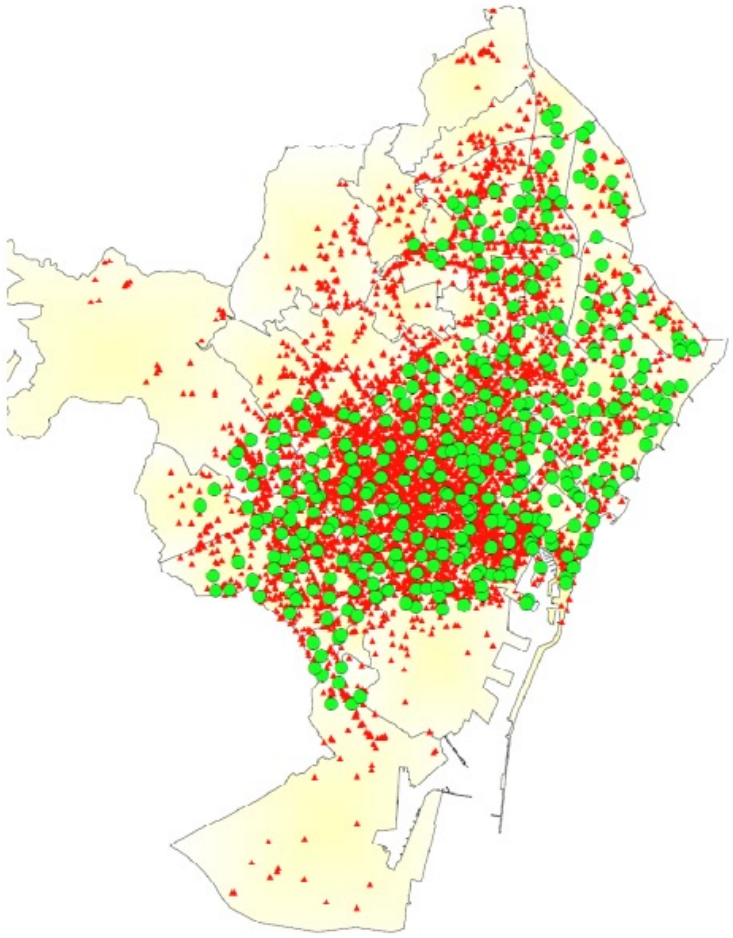

Figure 1. Sharing bike stations in Barcelona (Hampshire and Marla, 2011)

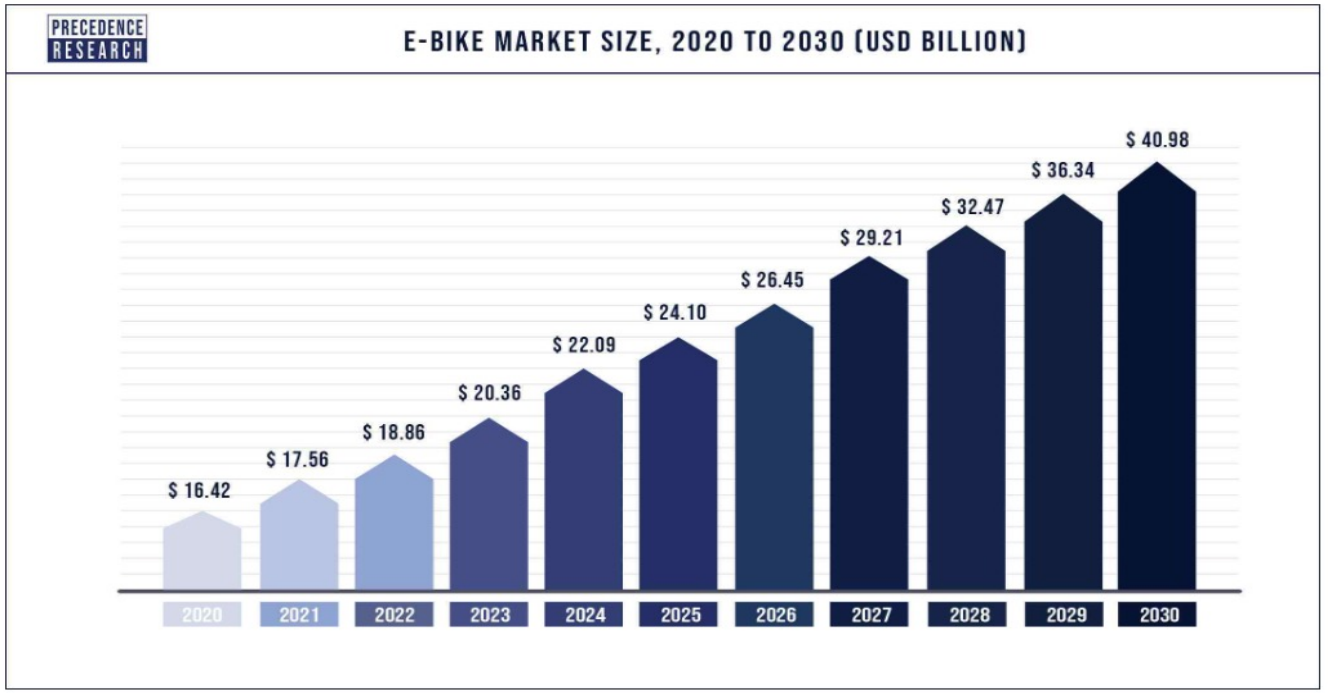

Figure 2. Sharing bike market size growth

Figure 1 provides the overview of the bike usage trend in Barcelona, it can be found that the intensity of the bike station points distribution is gathering towards the center point. In this figure, those red points indicate the usage frequency and shows a diffusion trend. Therefore, it is necessary to decide how bike station points to be set, as the growing market size proven by Figure 2.

DATA DESCRIPTION

Open source London public transportation data [1] will be used in this project. This dataset is colleceted by Tfl, a free transportation data serivce, from three sources - Tfl data collection, weather data website and government regulations on public holidays, which can be found in the note at the bottom.

Detailed data description is given below:

- *Timestamp* - Date data.
- *Number* - Counting the new bike shares behaviour.
- *Temperature* - Real temperature denoted as “t1”.
- *Temperature-feel* - Temperature individual feels denoted as 't2'.
- *hunidity* - humidity in percentage.
- *windspeed* - wind speed in km/h.
- *weathercode* - Category code of the weather.
- *isholiday* - category denoted whether a day is holidy. (1 - holiday ; 0 - non holiday)
- *isweekend* - categody denoted whether a day is weekend (1 - yes; 0 -not)
- *season* - category of seasons (0-spring; 1-summer; 2-autumn; 3- winter

[1] https://www.kaggle.com/datasets/hmavrodiev/london-bike-sharing-dataset?resource=download

IDENTIFYING WHAT DATA ARE KEY TO ANALYSIS

Based on data from Kaggle database, we totally have 9 columns of data in the period between 2015 and 2016. To identify the valuable information to our project, here based on the importance and how easy they are to perform analytics, use 0-5 scale for ranking.

Data	importance	Availability (How easy for analysis on a scale)
timestamp	5	4
count	5	2
temperature	5	5
temeprature-feel	4	2
humidity	4	5
windspeed	4	5
weathercode	5	5
isholiday	4	5
isweekend	4	5
season	4	5

STATISTICAL INFORMATION OF DATA

	count	mean	std	min	25%	50%	75%	max
cnt	17414.0	1143.101642	1085.108068	0.0	257.0	844.0	1671.75	7860.0
t1	17414.0	12.468091	5.571818	-1.5	8.0	12.5	16.00	34.0
t2	17414.0	11.520836	6.615145	-6.0	6.0	12.5	16.00	34.0
hum	17414.0	72.324954	14.313186	20.5	63.0	74.5	83.00	100.0
wind_speed	17414.0	15.913063	7.894570	0.0	10.0	15.0	20.50	56.5
weather_code	17414.0	2.722752	2.341163	1.0	1.0	2.0	3.00	26.0
is_holiday	17414.0	0.022051	0.146854	0.0	0.0	0.0	0.00	1.0
is_weekend	17414.0	0.285403	0.451619	0.0	0.0	0.0	1.00	1.0
season	17414.0	1.492075	1.118911	0.0	0.0	1.0	2.00	3.0

According to this statistics result, we have totally **17414 observations** for data analysis. Four of them are continuous which can be used in specific quantitative specular process and there are four categorical variables for further exploration.

Three topics are to be explored in addressing the intended problems of relationship exploration and effects clarification.

Hypothesis 1a
- Time has an effect on number of cycles.

Hypothesis 2a
- Season influences usage of sharing bikes.

Hypothesis 3a
- Weather condition has an effect on usage of sharing bikes

HEATMAP ANALYSIS BASED ON PEARSON MARTIX

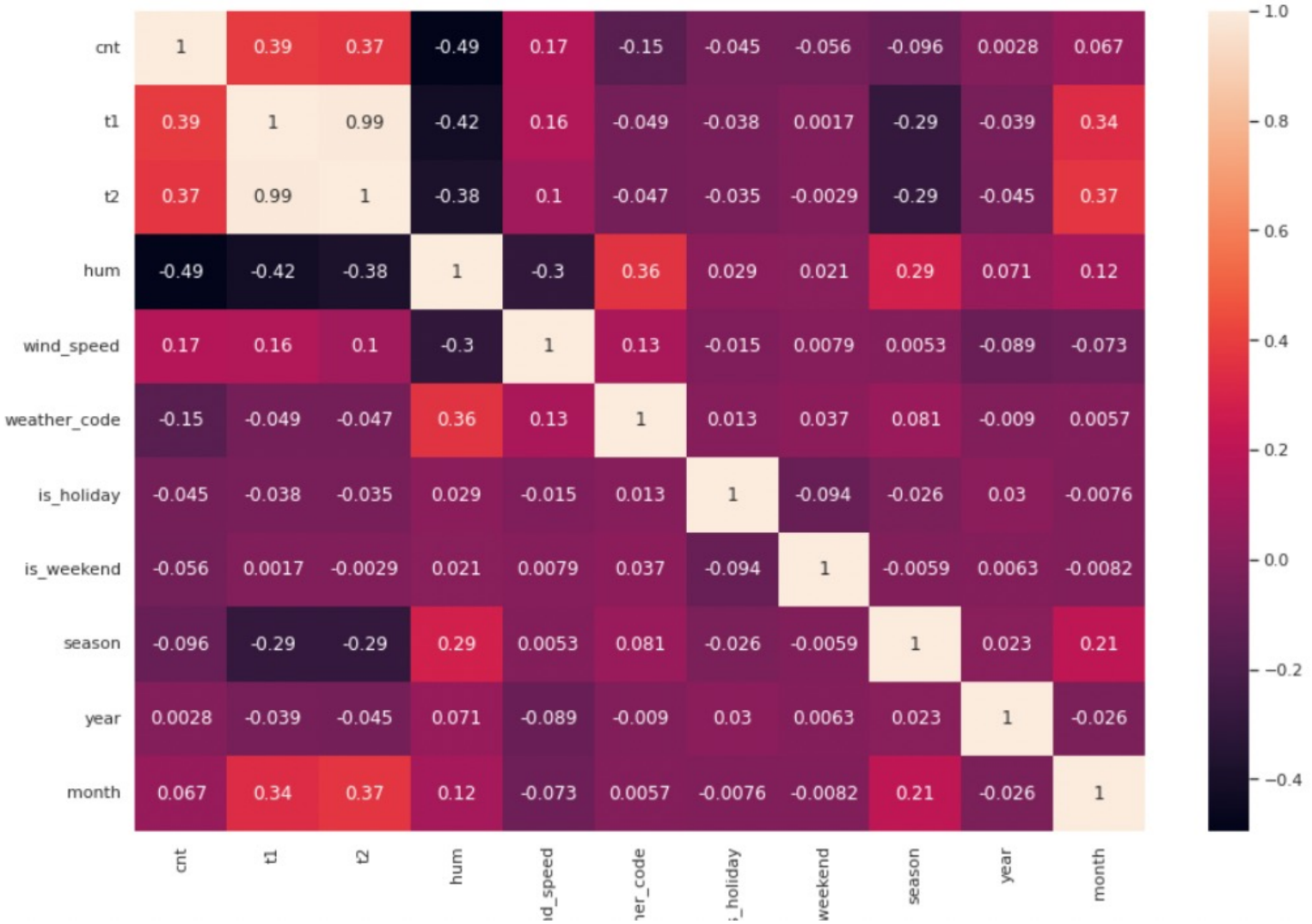

Figure 3. Pearson Matrix of Data Used

Before applying a regression analysis, first the **cross-relation between independent variables** are to be identified. Peason matrix is a powerful in realizing this goal and the result has been shown in Figure **3.**

This matrix table considers both independent and dependent variables. The lighter color indicates a stronger positive relationship while a darker indicates a strong negative relationship. Usually, the relation ratio above **0.4** indicates a strong relationship, and value above **0.7** indicates a very strong one. Thus, we have:

- Bike usage is positively related with temperature with a relatively strong relationship.
- Independent variables have interactive effects which should be considered in regression model, preventing bias.

EXPLORATION DATA ANALYSIS

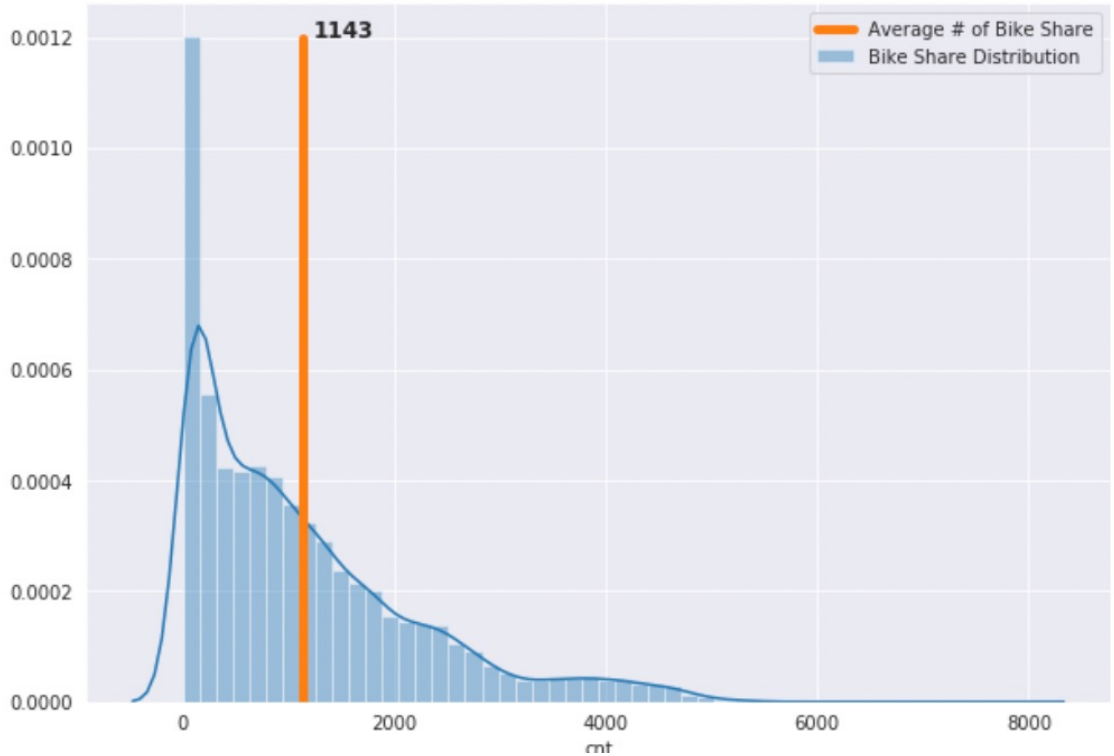

Figure 4. Bike usage distribution and average value

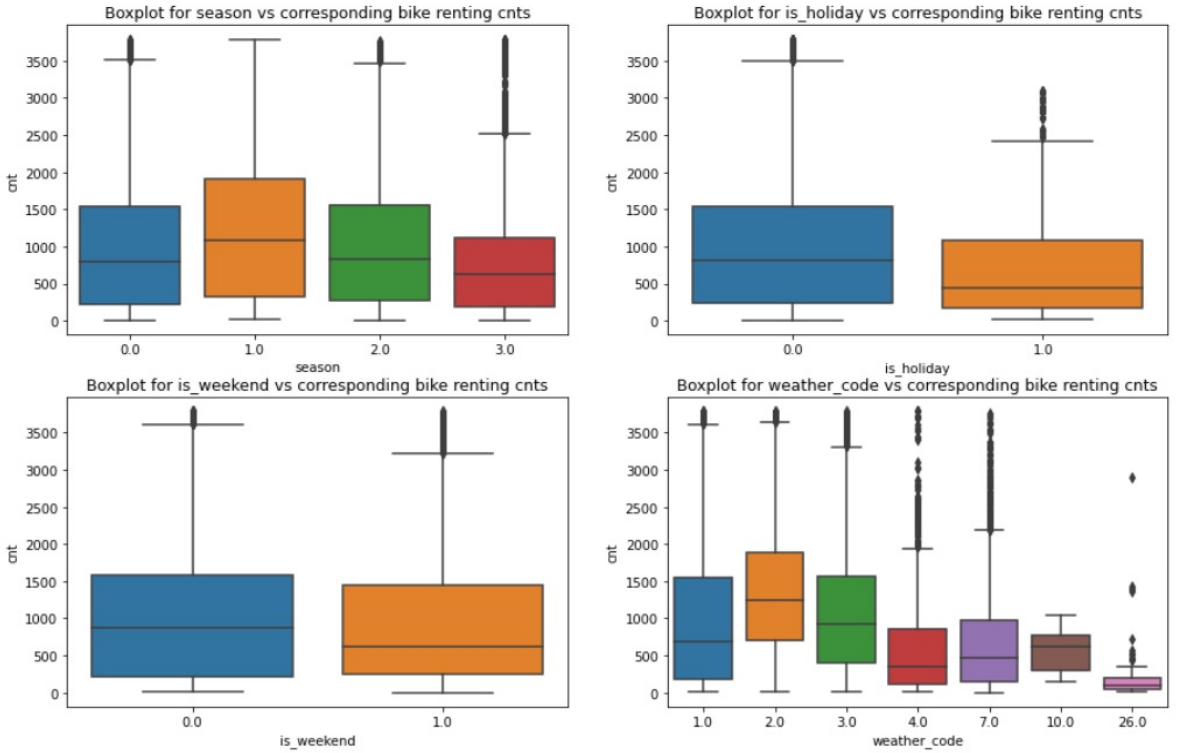

Figure 5. Box plot for independent and dependent variables

From above figures, we can find that the bike usage distribution is not a norm one. But **follows an exponential or log function**. For ease exploration, later we may need to normalize data in building the mathematical model. Figure 2 tells that there are many outliers in variables, so we visualize it in next slide.

EXPLORATION DATA ANALYSIS

To remove outliers, we delete 'count' less than Q1 - 1.5 IQR and greater than Q3 + 1.5 IQR. Therefore, our new datasets covers the following range with the number of observations is **16739**:

[Q1 – 1.5 IQR, Q3 + 1.5 IQR]

Figure 6 shows the cleaned dataset result for regression analysis.

Figure 6. Scatter plotting of cleaned dataset

REGRESSION ANALYSIS

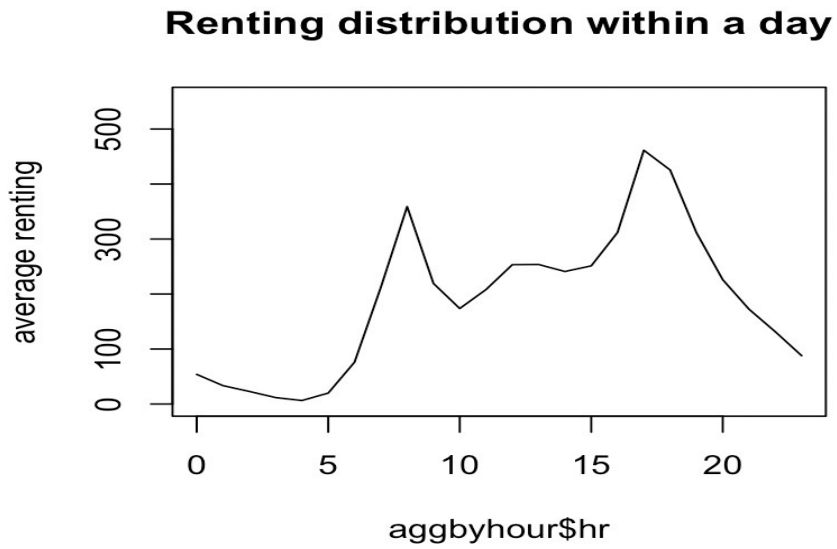

Figure 7. Sharing bicycle average amount by hour

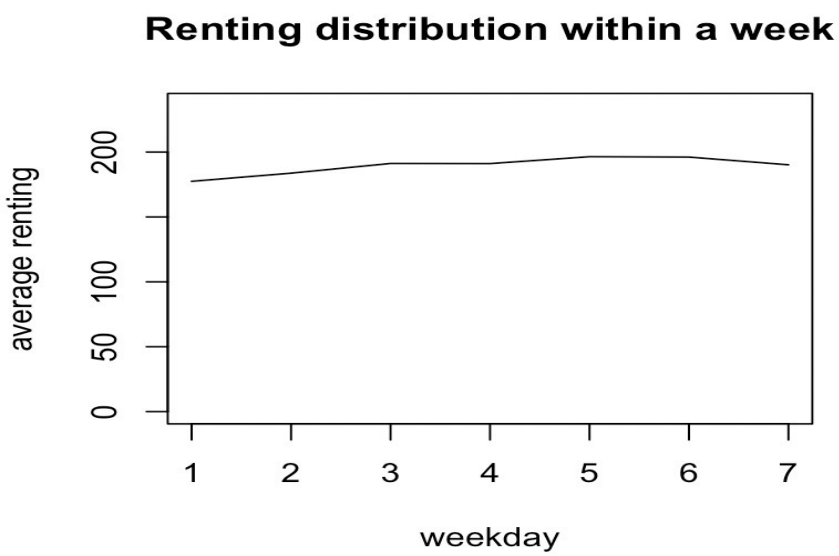

Figure 8. Sharing bicycle average amount by day

Sharing bicycling features high flexibility and traffics as well as uncertain **time and environmental climates**. To figure out their relationships, this report picks up a relevant dataset from UCI.

Research 1 – Descriptive depiction of renting rate distribution within a day and a week
(See figures above - "Renting distribution within a day" & "Renting distribution within a week")
Based on them and the regression result, we have the following fundamental conclusions:

- Average renting count varies during a day, demands more lie on 8-9 am and 18-19pm. (see figure 7)
- Average renting count is almost constant within one week (as you can see a flat straight line in (figure 8)
- From correlation analysis, the renting rate is weakenly related to daily hours, but closely connected to working days.

Therefore, further exploration needs to be made on seasonal and environmental influence on the sharing bicycles usage. **First** is a bubble diagram using *aggregated renting count on seasons*. **Secondly**, a new regression model on weather factors and significance level explanation to it.

REGRESSION ANALYSIS

From right bubble chart,, the average renting count is influenced by seasonal factors (1=winter, 2=spring, 3=summer, 4=fall)

Considering weather differentiation, here picking climate factors of weather situation, temperature, feeling temperature, humidity and windspeed to build a regression model.

The result tells humidity has the biggest and negative influence on sharing bicycle usage with the correlation of -304.714, $p<0.05$.
Next, feeling temperature has the second biggest and positive effects on dependent variable with a correlation of 292.556, $p<0.05$.

To summarize, weather factors pose significant effects on using sharing bicycles, and the level of issues users paying attention to is: hum (-) >atemp(+) > temp(+) >windspeed(+)>weather situation (+).

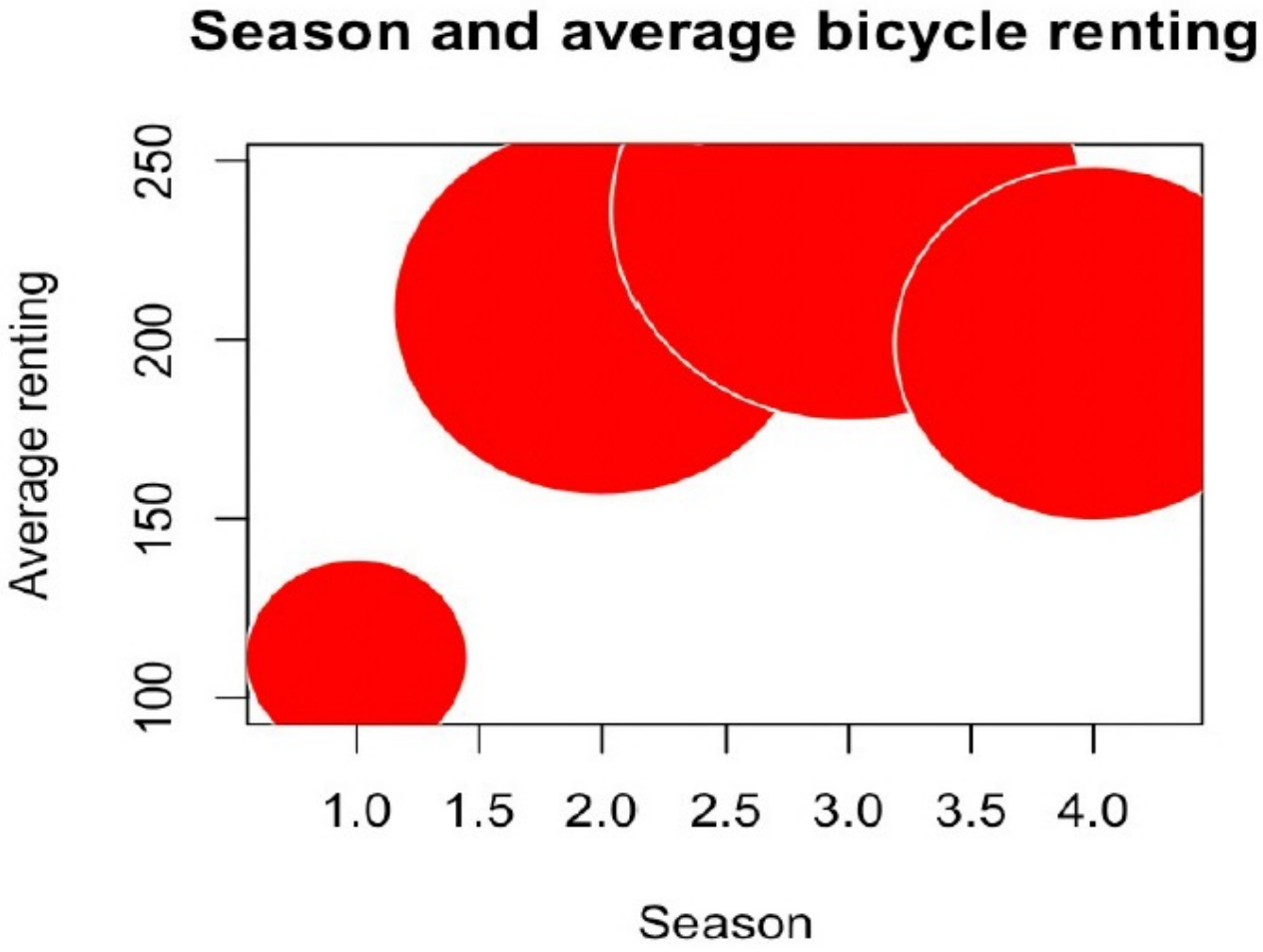

Figure 9. Sharing bicycle amount group in season

VISUALIZING WEATHER CONDITIONS ON BIKE SUAGE

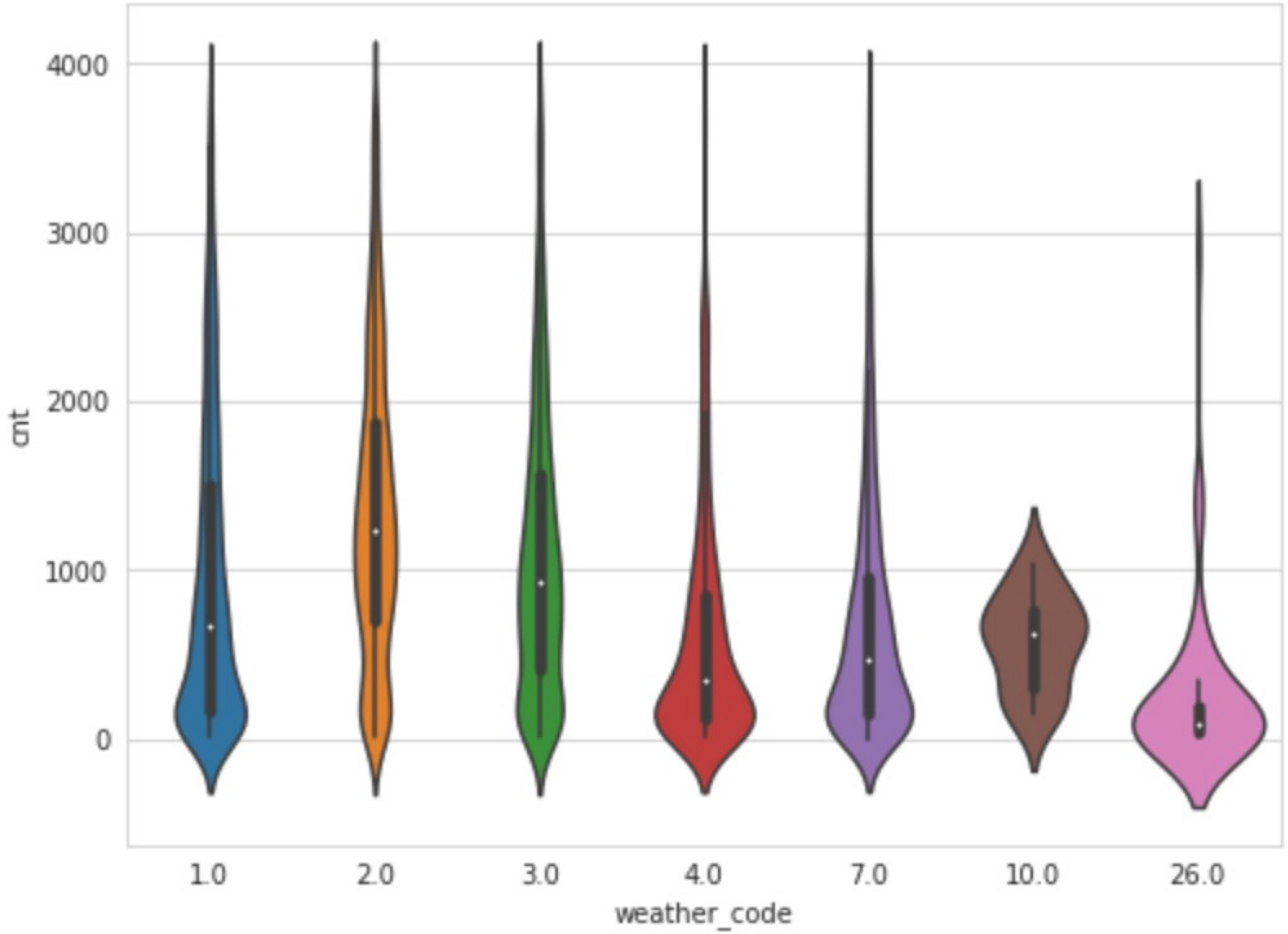

Figure 10. Violin chart of weather condition and bike usage

Observations

- Weather code 1 has more number of cycles are shared overall.
- So, does Weather 2 but little less than Weather 1.
- Weather 3 is around, 0 - 50K
- However, Weather 10.0 and Weather 26.0 is highly populated around 0-16K. I think they are regular cyclist on any condition, they ride a bike.

Reference

Earth Policy Institute. (2015, februari 15). Datasets on population, health and society. Retrieved from http://www.earth-policy.org/data_center/C26

Bloomberg (2022). UK Road Fuel Prices Surge to Record High as Fill-Up Nears £103. *Bloomberg.com*. [online] 15 Jun. Available at: https://www.bloomberg.com/news/articles/2022-06-15/uk-road-fuel-prices-surge-to-record-high-as-fill-up-nears-103.

GOV.UK. (2021). *Transport Statistics Great Britain: 2021*. [online] Available at: https://www.gov.uk/government/statistics/transport-statistics-great-britain-2021/transport-statistics-great-britain-2021.

Hampshire, R. and Marla, L. (2011). *An Analysis of Bike Sharing Usage: Explaining Trip Generation and Attraction from Observed Demand 2 3 School of Public Policy and Management School of Information Systems and Management School of Public Policy and Management School of Information Systems and Management*. [online] Available at: https://nacto.org/wp-content/uploads/2012/02/An-Analysis-of-Bike-Sharing-Usage-Explaining-Trip-Generation-and-Attraction-from-Observed-Demand-Hampshire-et-al-12-2099.pdf [Accessed 10 Aug. 2022].

UK Road Fuel Prices Surge to Record High as Fill-Up Nears £103. (2022). *Bloomberg.com*. [online] 15 Jun Available at: https://www.bloomberg.com/news/articles/2022-06-15/uk-road-fuel-prices-surge-to-record-high-as-fill-up-nears-103 [Accessed 10 Aug. 2022].