

**INFO151: Databases**

Course Coordinator & Lecturer: Tiong Goh

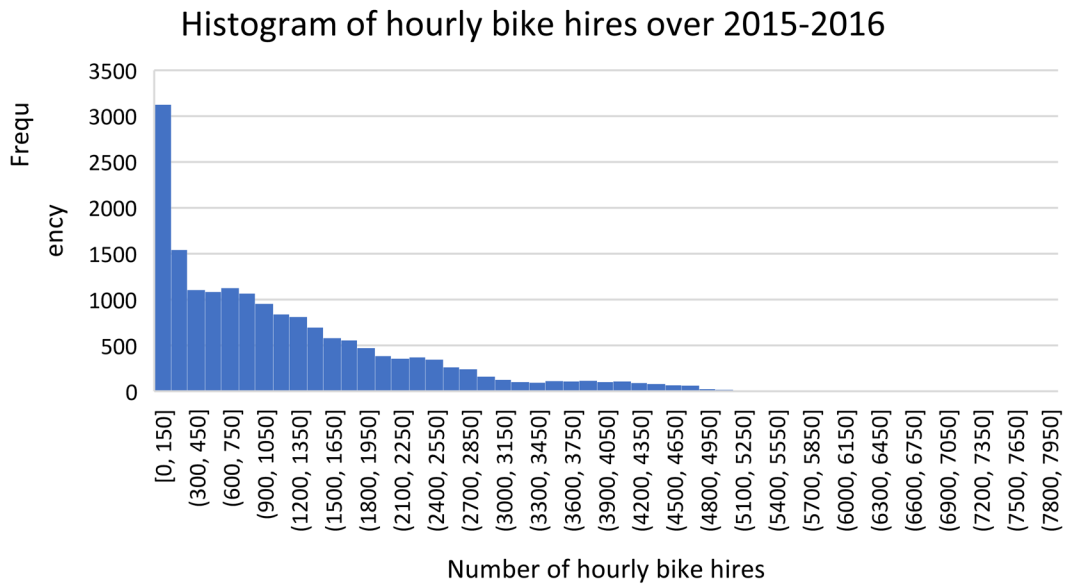
# Lab Report #1

**Student Name:** PHAM NGOC KHANH DUNG (Junn Pham)

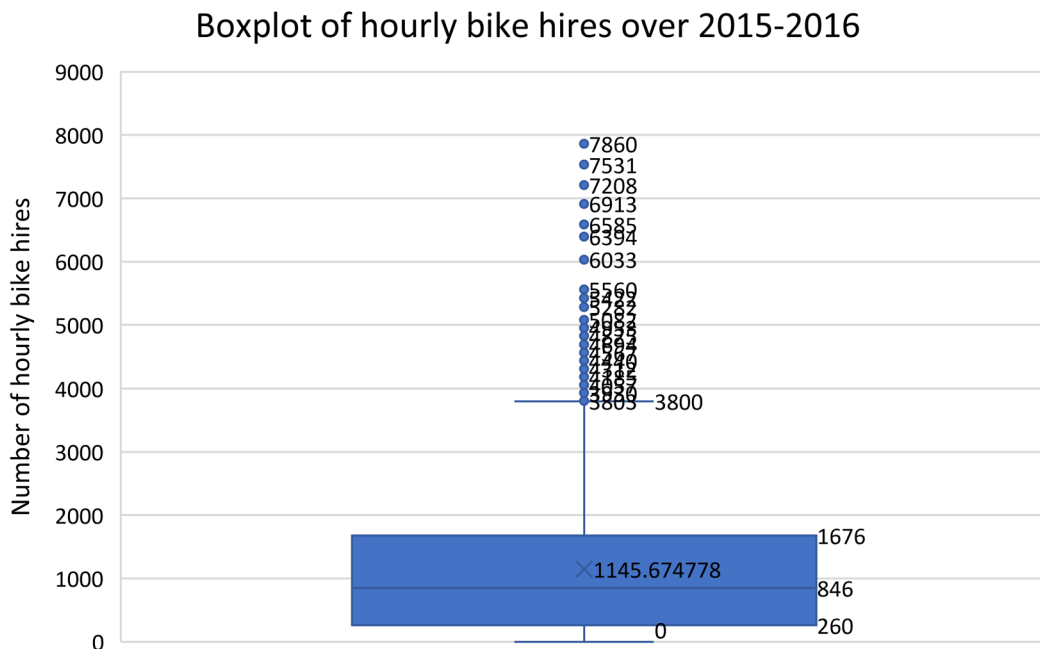
**Student ID:** 300538618

### Question 1:

Measure	Symmetric Distribution		Asymmetric/Skewed Distribution	
Centrality	Mean	1145.6748	Median	846
Dispersion	SD	1085.9234	Interquartile	1416

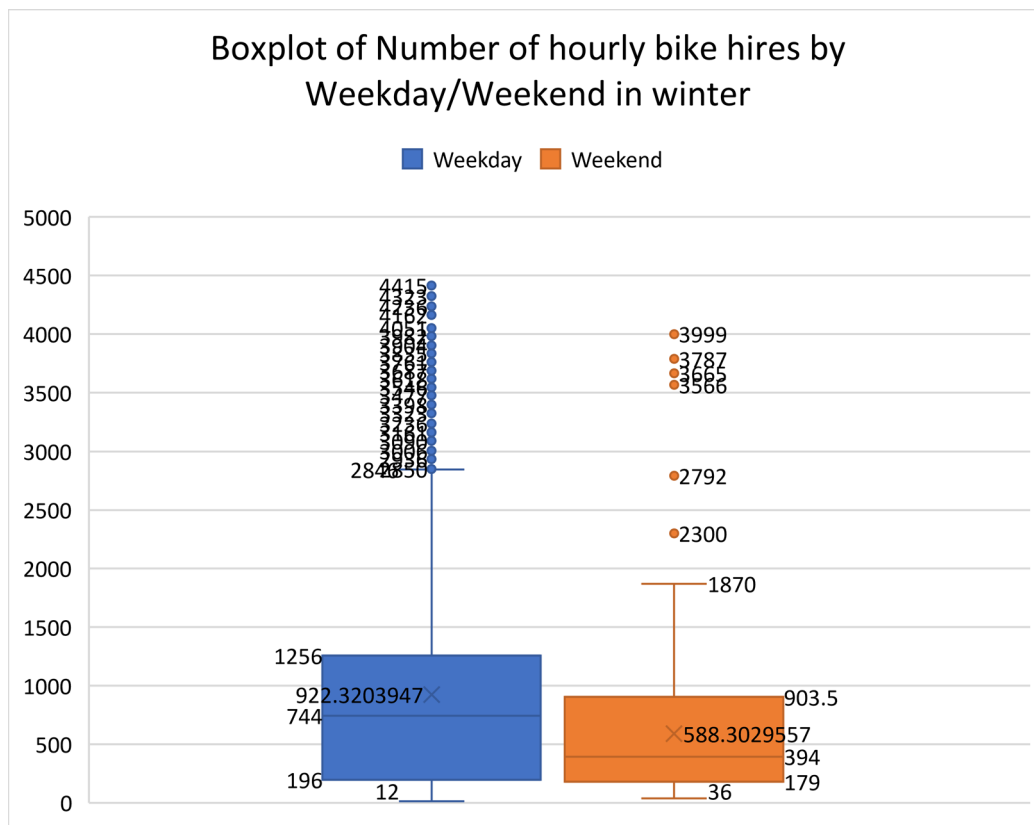


The histogram illustrates that the number of bike hires per hour which is in the range of 0 to 150 bikes accounts for the most distribution over 17,342 observations. Furthermore, the hourly bike hires distribution is unimodal and left skewed with a long right tail. Therefore, the most appropriate measures of centrality and dispersion to use is median and interquartile range.



The boxplot shows that a half of the number of hourly bike hires during 2015-2016 distributes from 260 to 1676 bikes. There are period when no bike is hired, and the minimum number of hourly bike hires equals to 0. The middle value of the data set is 846. There are outliers and the maximum observation is 7860.

## Question 2:



	Weekday	Weekend
Median	922.320395	588.302956
First quartile	196	179
Third quartile	1256	903.5
Interquartile	1060	724.5
Minimum	12	36
Maximum	4415	3999

According to the boxplot, the number of bikes hired per hour on working days in winter is higher than that figure on the weekend. Specifically, the median number of hourly bikes hired on working days is 922.32, which is higher than the figure of weekend which is 588.30. The dispersion is considerably greater for the weekday observations as well. The interquartile of the weekday data set is 1060, while that number of the weekend is much lower, which equals 724.5.

There are outliers on the weekday boxplot and the maximum value of them is up to 4415, while the maximum number of hourly bike hires on weekend in winter is only 3999.

### Question 3:

Regression Statistics	
Multiple R	0.503172752
R Square	0.253182819
Adjusted R Square	0.253053597
Standard Error	938.5208175
Observations	17342

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	5177340747	1725780249	1959.285272	0
Residual	17338	15271680131	880821.3249		
Total	17341	20449020878			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	2831.786203	54.06205181	52.38029465	0	2725.819131	2937.753275	2725.819131	2937.753275
Temperature_2	35.78310069	1.180222501	30.31894467	8.5609E-197	33.4697456	38.09645578	33.4697456	38.09645578
Humidity	-28.68567119	0.565476993	-50.72827286	0	-29.79406311	-27.57727927	-29.79406311	-27.57727927
Wind_Speed	-1.669720108	0.941836145	-1.772835027	0.076273602	-3.515813908	0.176373691	-3.515813908	0.176373691

### Interpreting the output:

The adjusted R Square equal to 25.3% for this regression suggests that the independent variables explain 25.3% of the variation in the dependent variable.

The significance F or p-value corresponding to the ANOVA table is 0, which is extremely small, Therefore, we can say that the regression model is statistically significant, and the independent variables are useful in forecasting the number of hourly Bikes hires.

The coefficient values suggest that with a one-degree increase in “Feel like” temperature, the number of Bike hires is predicted to increase by approximately 36 bikes. On the other hand, a one percent rise in humidity and a one km/h increase in wind speed lead to a drop of nearly 29 bikes and 2 bikes, respectively.

Lastly, the p-values of the independent variables show that Humidity and “Feel like” temperature is important in predicting the number of Bike hires. While the p-value of wind speed is not as small as the numbers of the other variables, it is worth using in forecasting the number of hourly bike hires.

**The estimated linear equation is:**

$$\begin{aligned}\text{Predicted number of hourly Bike hires} &= 2831.79 + 35.73 * \text{Temp} + (-28.69) * \text{Humidity} \\ &+ (-1.67) * \text{Wind Speed}\end{aligned}$$

A prediction of bike hires when the "feels like" temperature is 22°C, the humidity is 50%, and the wind speed is 5 km/h:

$$\begin{aligned}\text{Predicted number of hourly Bike hires} &= 2831.79 + 35.73 * 22 + (-28.69) * 50 + (-1.67) * 5 \\ &= 2176 \text{ (bikes)}\end{aligned}$$