# Springboard Capstone Project 1 Storytelling

Tzu-Ying Chen

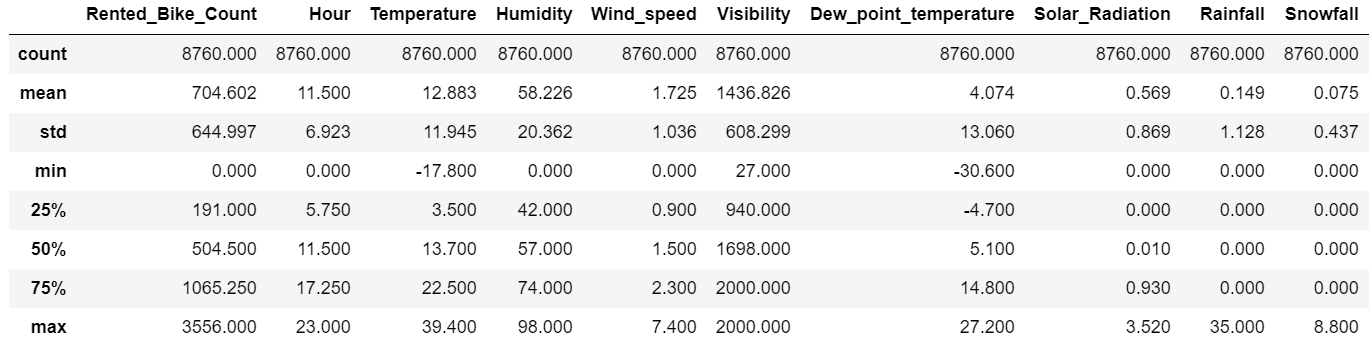
## Data Type

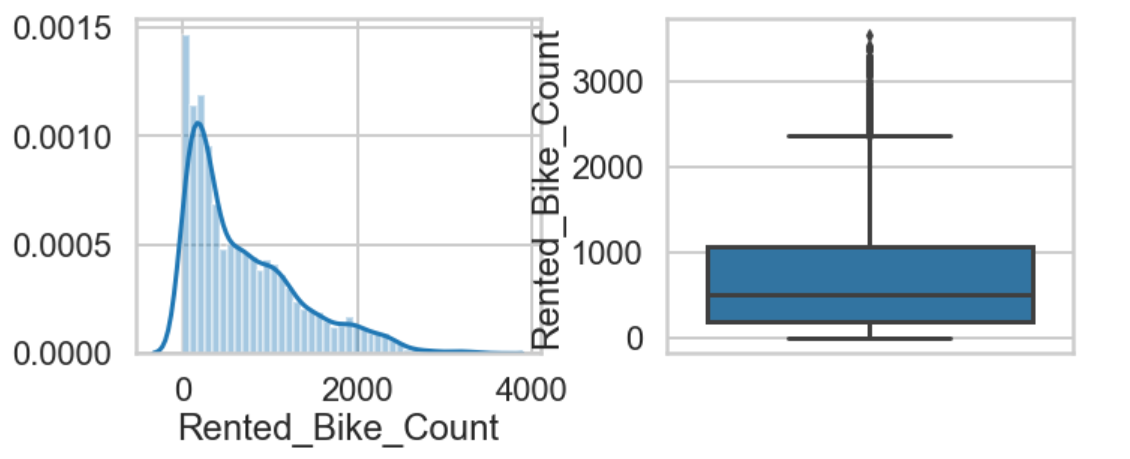
The goal of the project is to predict rental bike count for each hour, hence the target variable for this project is rented\_bike\_count. The predictor variables include temperature, humidity, wind speed, dew point temperature, solar radiation, rainfall, snowfall, holiday, functioning day, seasons, date, and hour. The following chart show the overview of the data set in terms of the data type and category.

## Exploratory Data Analysis

### Univariate Analysis

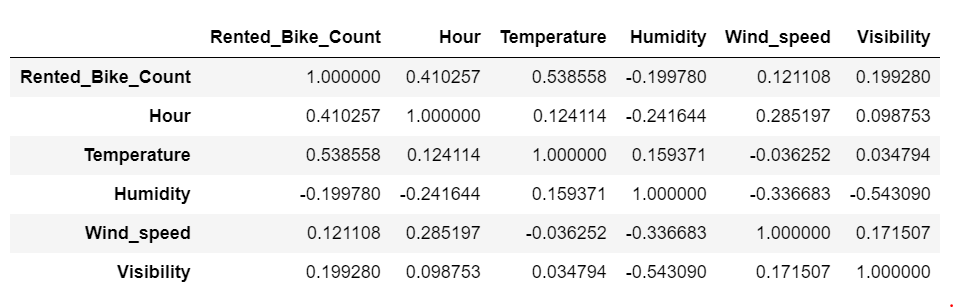
When conducting univariate analysis, variables are analyzed one at a time. For the continuous variable, we can observe its distribution and identify the outliers. In the project, there are 10 continuous variables in the data set.



In the above data overview, we can summarize that there are average of 705 bikes are rented on an hourly basis, but the data has a fairly high standard deviation, indicating that there is a high variation in bike rentals. In addition, the distribution of bike rental amount shows left skewed, and the box chart providing the idea of there are some extreme cases.

### Bivariate Analysis

When conducting bivariate analysis, we are examining the relationship of two variables. When looking at the correlation between two continuous variables, there is no strong relationship between any of the two.



However, when plotting the time series data on temperature, we can see a trend. One noticeable thing is, I have to exclude 2017 data due as the data set doesn’t include full year of data. In addition, the chart below is aggregated into date level instead of the hourly level.

