

NICD Bike Rental Report

Author: Louise Braithwaite

Code and detailed report can be accessed online: github.com/NCL-LouB/Bike_rental_project

1. Background

Bike rental data was collected over a two-year period. The data set contains:

- the number of daily rentals (rental count)
- daily weather conditions (temperature, humidity, wind speed and precipitation type)
- day type (season, year, month, holiday, weekday, working day)

1.1. Project aim

Analyse the data to derive business insights into how the change in weather effects the number of bikes rented per day.

1.2 Data preparation

Data preparation code and explanatory text can be found in the project's github repository.

1.3 Data assumptions

The following data assumptions were made.

- Temperature is measured in degrees celsius (°C)
- Humidity measure is relative humidity (%)
- Wind speed is measured in miles per hour (mph)
- Weathersit variables are an indication of precipitation, "GOOD" is equivalent to "NO.RAIN"
- Temperature, humidity and wind speed can be rounded to integers to ease interpretation

2. Methods

1. Initial exploratory data analysis (EDA) of all variables to identify key variables (see github repository)
2. A pairs plot and correlation matrix were generated for the key variables. This included a predictor variable, 'rental.count', and four explanatory variables: 'temperature', 'humidity', 'wind.speed' and 'precipitation'
3. Further EDA was undertaken on the variable that had the strongest linear correlation with number of daily rentals, temperature

3. Data analysis

3.1 Correlation coefficients

A correlation coefficient measures the strength and direction of the linear relationship between two variables. A value of exactly 1.0 means there is a perfect positive relationship between the two variables. For a positive increase in one variable, there is also a positive increase in the second variable. A value of -1.0 means there is a perfect negative relationship between the two variables.

variable.1	variable.2	coefficient
rental.count	temperature	0.63
rental.count	precipitation	-0.30
rental.count	wind.speed	-0.24
rental.count	humidity	-0.10

Table 1: Summary of correlation coefficients

The correlation coefficients in table 1 show:

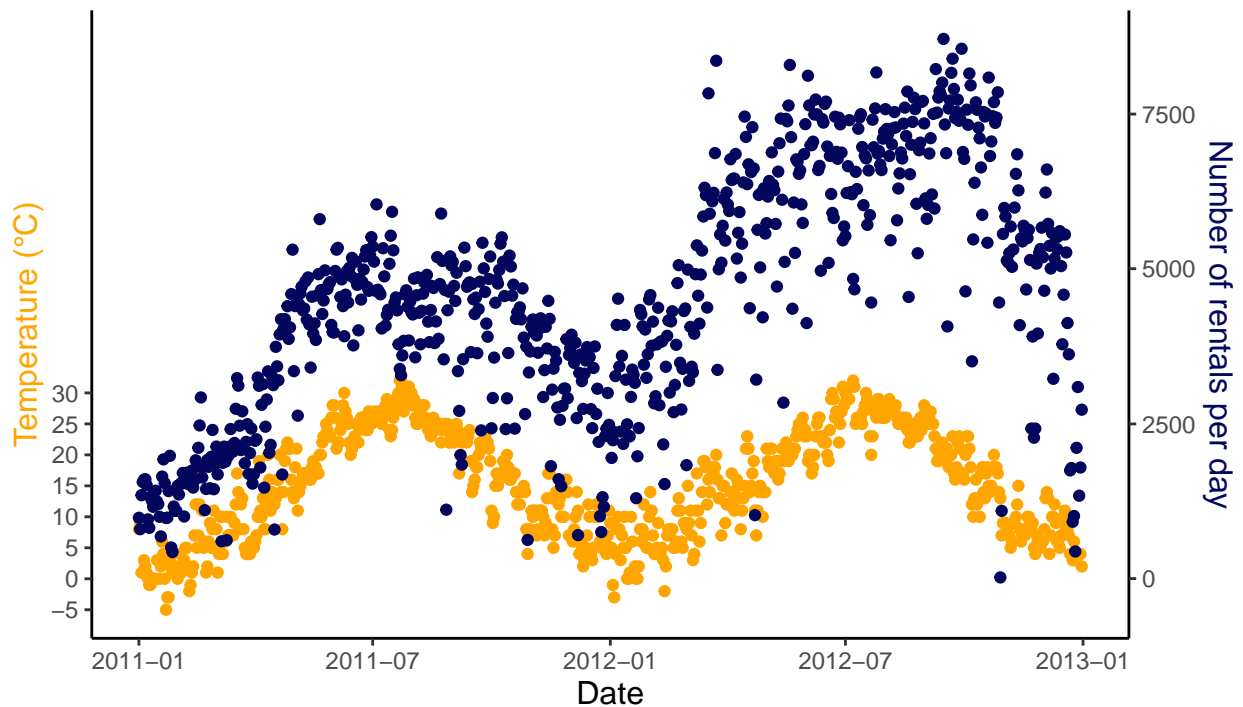
- The correlation between rental.count and temperature, 0.63, indicates a moderately strong positive linear relationship
- The correlation between rental.count and weather.category, -0.3, suggests there is some indication of a negative linear relationship between the two variables. This supports the assumption that as the weather goes from no rain (“GOOD”), to “MISTY”, to “RAIN/SNOW/STORM” the rental.count will decrease.
- The correlation between rental.count and wind.speed, -0.24, and rental.count and humidity, -0.1, are negative but are too low to indicate a linear relationship.

Key finding: temperature has the most significant linear relationship with rental count. Figures 1 and 2 explore this relationship in more detail.

3.2 Daily temperature and rental count trend over two years

Figure one plots the daily temperature and daily rental count across the two year period. The plot shows that the annual trend of daily rentals follows the annual temperature trend.

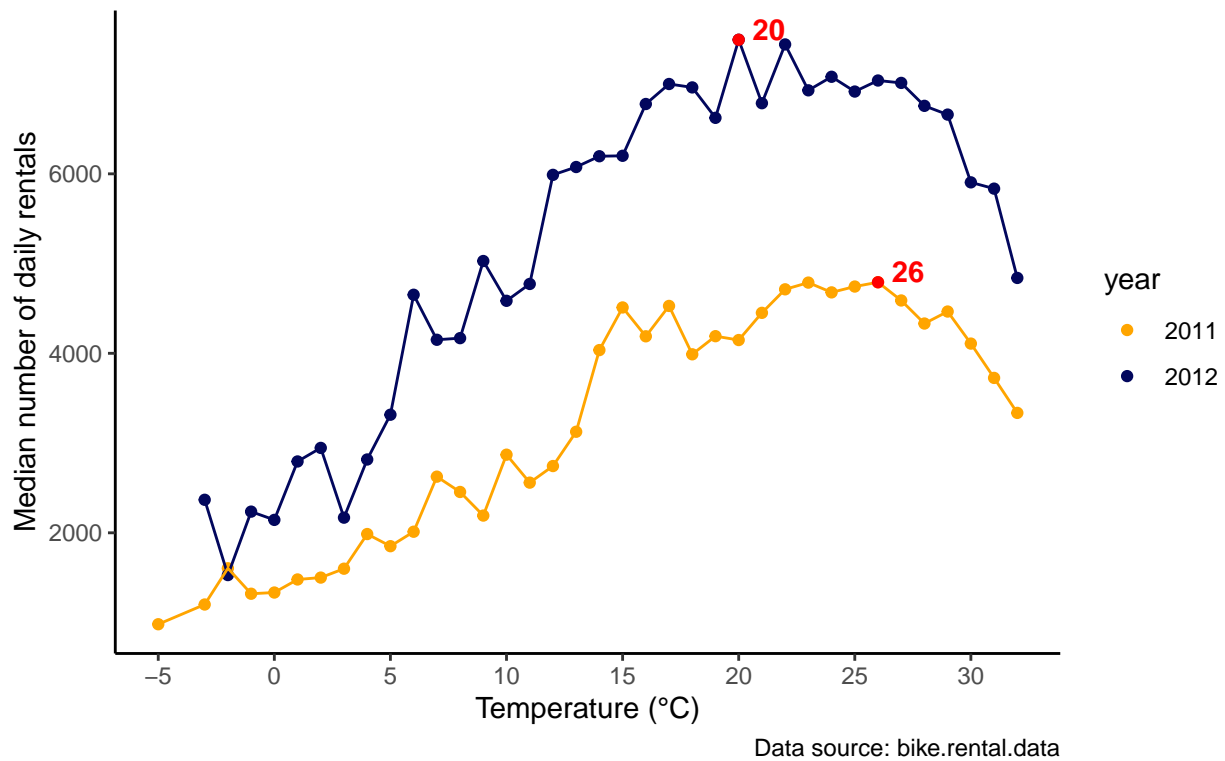
Figure 1
Rental count and temperature per day



Data source: bike.rental.data

Figure 2: Median number of rentals per day by temperature and year

Maximum median value



3.3 Median rental count by temperature

Figure 2 shows that the median number of daily rentals increases with the temperature up to the mid-20s, when the median daily rentals starts to decrease.

4. Future work

To build upon this work

- Use data to build a model to predict rental