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. A least squares linear regression model was fitted to the key variables but only 47.21% of the variation in rental.count could be explained by regression on the four standardised predictors. This leaves a large proportion of the variation unexplained and so the model was rejected.
4. 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 (0-1) and direction (+ or -) of the linear relationship between two variables. A value of exactly 1.0 means there is a perfect positive relationship between the two variables, so a positive increase in one variable relates to a positive increase in the second variable. .

.
.
.
.
.

Table 1: Summary of correlation coefficients

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

.

The correlation coefficients in table 1 show:

- The correlation between rental.count and temperature is 0.63, which indicates a moderately strong positive linear relationship
- The correlation between rental.count and weather.catergory is -0.3, which 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" rental.count will decrease
- The correlation between rental.count and wind.speed is -0.24 and rental.count and humidity id -0.1. Both are negative but the coefficients are too low to indicate a linear relationship

---

Key finding: temperature has the most significant linear relationship with rental count.
The relationship between temperature and rental count is explored in more detail in figures 1 and 2.

3.2 Daily temperature and rental count trend over two years

Figure 1 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. Temperatures and number of daily rentals peak in the summer and early fall months and are at their lowest in January and February.

.

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. After this point the median daily rentals starts to decrease.

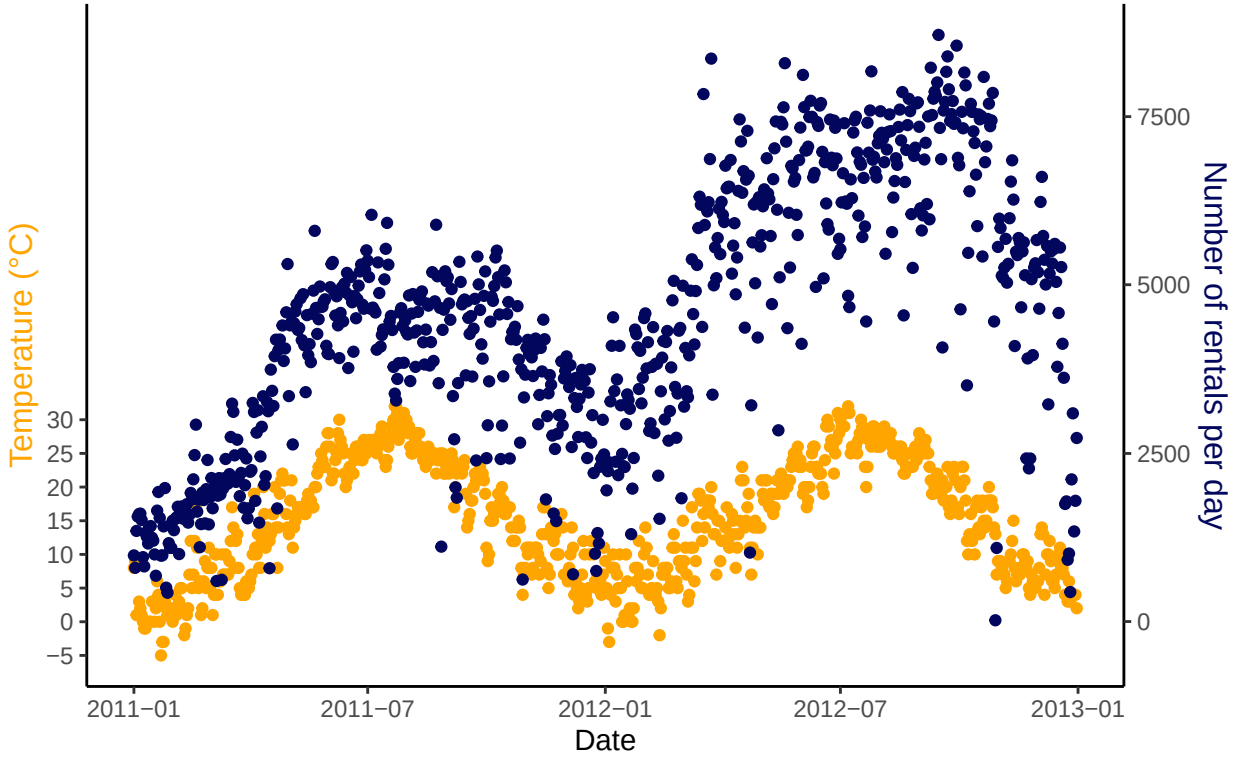
This figure provides an indication of what the bike rental company could expect their rental numbers to be on a average day given a certain temperature.

.

4. Future work

The analysis may be improved by identifying a continuous measure of daily rainfall, in addition to the categorical value used to identify, no rain, mist, rain, snow and storm. To build upon this initial exploratory data analysis work, further statistical modeling could be conducted to identify a model which incorporates all four key weather indicators.

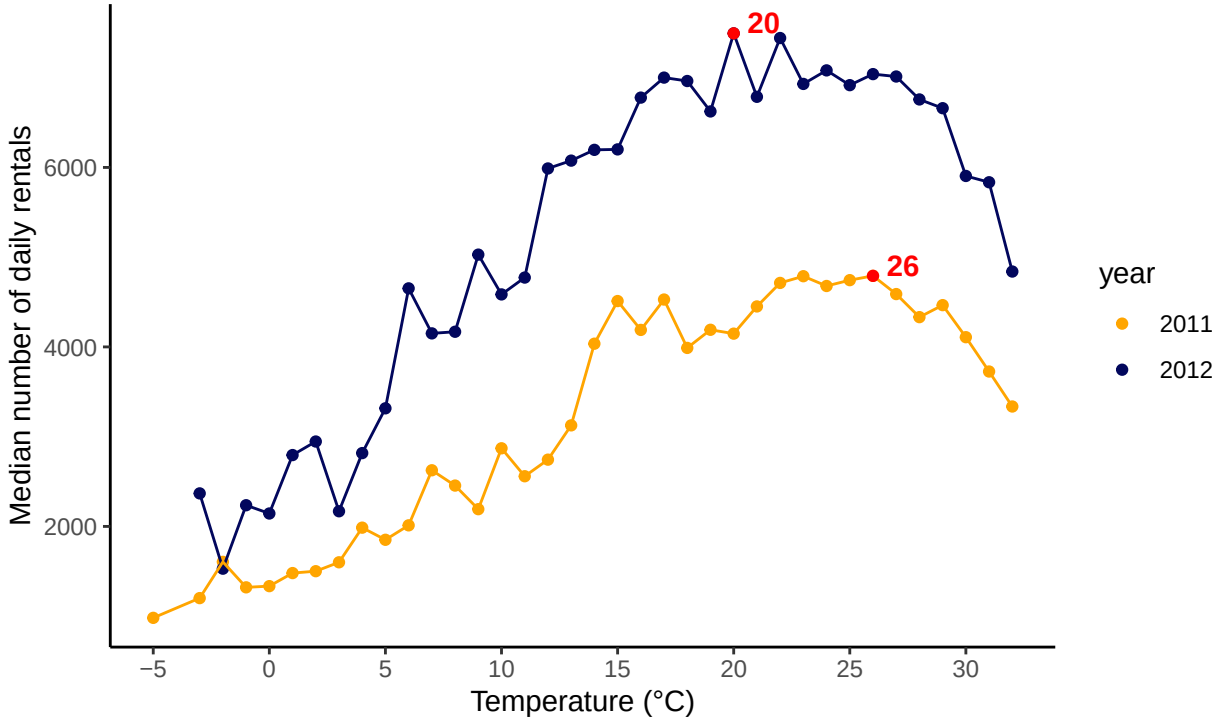
Figure 1: Daily rental count and temperature over 2 years



Data source: bike.rental.data

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

Maximum median value



Data source: bike.rental.data