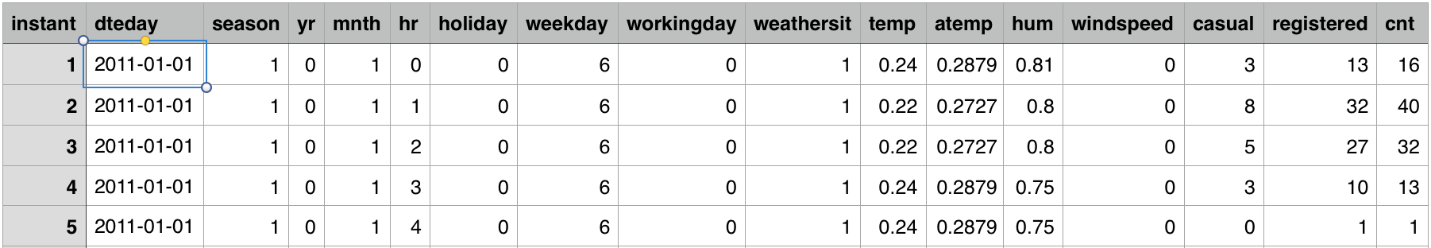
Many U.S. cities have communal bike sharing stations where you can rent bicycles by the hour or day. Washington, D.C. is one of these cities. The District collects detailed data on the number of bicycles people rent by the hour and day.

[Hadi Fanaee-T](http://www.liaad.up.pt/area/fanaee) at the [University of Porto](http://www.up.pt/) compiled this data into a CSV file, which you'll work with in this project. The file contains 17380 rows, with each row representing the number of bike rentals for a single hour of a single day. You can download the data from the [University of California, Irvine's website](http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset). If you need help at any point, you can consult the solution notebook [in our GitHub repository](https://github.com/dataquestio/solutions/blob/master/Mission213Solution.ipynb).

Here's what the first five rows look like:



Here are the descriptions for the relevant columns:

* instant - A unique sequential ID number for each row
* dteday - The date of the rentals
* season - The season in which the rentals occurred
* yr - The year the rentals occurred
* mnth - The month the rentals occurred
* hr - The hour the rentals occurred
* holiday - Whether or not the day was a holiday
* weekday - The day of the week (as a number, 0 to 7)
* workingday - Whether or not the day was a working day
* weathersit - The weather (as a categorical variable)
* temp - The temperature, on a 0-1 scale
* atemp - The adjusted temperature
* hum - The humidity, on a 0-1 scale
* windspeed - The wind speed, on a 0-1 scale
* casual - The number of casual riders (people who hadn't previously signed up with the bike sharing program)
* registered - The number of registered riders (people who had already signed up)
* cnt - The total number of bike rentals (casual + registered)

In this project, you'll try to predict the total number of bikes people rented in a given hour. You'll predict the cnt column using all of the other columns, except for casual and registered. To accomplish this, you'll create a few different machine learning models and evaluate their performance.

**Instructions**

* Use the [pandas](http://pandas.pydata.org/) library to read bike\_rental\_hour.csv into the dataframe bike\_rentals.
* Print out the first few rows of bike\_rentals and take a look at the data.
* Make a histogram of the cnt column of bike\_rentals, and take a look at the distribution of total rentals.
* Use the [corr](http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.corr.html" \t "_blank) method on the bike\_rentals dataframe to explore how each column is correlated with cnt.

### Learn

It can often be helpful to calculate features before applying machine learning models. Features can enhance the accuracy of models by introducing new information or distilling existing information.

For example, the hr column in bike\_rentals contains the hours during which bikes are rented, from 1 to 24. A machine treats each hour differently, without understanding that certain hours are related. We can introduce some order into the process by creating a new column with labels for morning, afternoon, evening, and night. This bundles similar times together, enabling the model to make better decisions.

### Instructions

* Write a function called assign\_label that takes in a numeric value for an hour, and returns:
  + 1 if the hour is from 6 to 12
  + 2 if the hour is from 12 to 18
  + 3 if the hour is from 18 to 24
  + 4 if the hour is from 0 to 6
* Use the [apply](http://pandas.pydata.org/pandas-docs/stable/generated/pandas.Series.apply.html) method on series objects to apply the function to each item in the hr column.
* Assign the result to the time\_label column of bike\_rentals.

Github solution - https://github.com/dataquestio/solutions/blob/master/Mission213Solution.ipynb