Projects (I use the data of city bike as an example: 2-3 weeks)

1. Given the data, please read the descriptions based on the link to understand the data
2. Data processing (excel, python), e.g., handling missing values (based on the slide Ambient sensors and sensor data on 2/9)
3. Machine Learning methods (explore the weka software application)

* Supervised learning (classification; regression)
  + We want to predict the gender of a new user
  + labels: male, female
  + data: features / attributes + label
    - …….. label
  + Split the data into training (70%) and testing (30%)
  + Use the training data to build up a model (decision tree)
    - Python (e.g., sklearn package); weka
  + Use the learned model to predict the labels of the test data (weka: csv file or txt)
  + If we use weka, it will show the prediction measures:
    - Accuracy, precision, confusion matrix….
* Unsupervised learning (clustering)
  + Weka