

Name: _____

Date: _____

IDS Unit 4

Design Project and Oral Presentation

Water Usage

At the beginning of this unit, you explored a 2010 dataset from the Los Angeles Department of Water and Power (DWP). You also created a participatory sensing campaign to investigate water usage around your community.

For this assignment, you will use both datasets to apply what you have learned in unit 4 and to answer the research question from the beginning of the unit:

How can we help city officials use participatory sensing to find out how water is being used around your neighborhood?

Your assignment is as follows:

1. You and a partner will predict water usage for the month of June using a subset of the `dwp_2010` dataset, which is called `dwp_student`.
 - Load the `dwp_student` dataset.
 - Using this data, create two datasets: training and testing. Name these datasets `student_train` and `student_test`.
 - Create the best prediction model that you can based on your training data. Remember to `set.seed(123)` when creating your own training and testing data.
 - You're building this model with data from July 2010 to May 2011. You will use your model to predict water usage for June 2011.
 - After you settle on a specific model, submit your model (code) to your teacher. If you created any new variables, submit the code you used to create them as well.
 - ***What do the variables included in your prediction model say about how Angelenos use water?***
 - You will evaluate the prediction accuracy based on a separate set of data. Your teacher will give you another dataset. Use this dataset to evaluate your prediction. The pair with the smallest prediction error based on mean squared error (MSE), is the winner.
2. Using your participatory sensing data explain how water is being used in your neighborhood. Make sure you use evidence from your PS data analysis. Be sure to answer the research question and your statistical questions.

Create a four-to-five-slide, 5-minute presentation that explains your model, the predicted value for June 2011 water consumption, and the findings using your campaign data. Be sure to include detailed explanation of how you and your partner decided to create your prediction model and how it performed on the test dataset your teacher provided in your presentation. Each person must participate in the presentation. In addition to the presentation, submit a two to four page, double-spaced summary of your analysis including plots/graphs.