

TRADER PROGRAMMING CHALLENGE INSTRUCTIONS

Please complete the following programming and analysis problem in whichever language you think the problem is best suited for.

The Question:

The Local Air Quality is a function of the amount of Sulfur Dioxide (SO₂) in the air in a given area. Attached you will find data which measures the SO₂ at State Street and Madison Ave in Downtown Chicago along with a number of covariates. The file "TRAINING_DATA.txt" includes the data to train your model on. The file "TESTING_DATA.txt" is the file for which you will produce projected values

Covariates Included:

- Car.count = count of cars to pass through intersection in the interval observed
- Wind.velocity
- Wind.direction = measured in degrees. 0 = N
- Temperature1 = Degrees Celsius at 1 meter altitude
- Temperature30 = Degrees Celsius at 30 meters altitude

Additional Variables:

- Time.of.day: The hour of the day of the observation of SO₂ and all covariates
- Day.index: The day of the observation starting at Feb, 1st 2010. I.e. 1 = Feb 1st, 2010

From the covariates included in the file, build the best linear model to explain the SO₂ level.

What you need to submit:

- A file with predicted values for each record in the "TESTING_DATA.txt" file.
- Your model and summary statistics (including t-stats of variables and R-squared)
- A plot of the residuals of your model (Y) plotted against the fitted values of your model (X).
- An explanation of your approach
 - If you choose to transform any of the variables please explain why
 - If you choose to include any interaction terms please explain why
 - If you choose to remove any covariates please explain why
 - Please include any additional plots that you think would support the decisions above.
- Your code
- An estimate of how long it took you to complete this problem

You will be graded on the following criterion:

- Whether your model accurately predicts SO₂ out of sample.
 - The best R-Squared out of sample to date is 0.62, with a reasonable number of variables.
 - Whether your model breaks any of the assumptions of linear regression
 - Your approach