Practical Data Science
(Predictive Analysis-1st Attempt)

Solve the following problems.

Problem1: Working with Data Frames

The data file rainfall.dat(available at algorithmica github repository under datasets branch) records hourly rainfall at a certain location in Canada, every day from 1960 to 1980. Answer the following questions:

a. Load the data set and make it a dataframe called rain.df. What command did you use?

b. How many rows and columns does rain.df have? How do you know? (If there are not 5070 rows and 27 columns, you did something wrong in the first part of the problem.)

c. What command would you use to get the names of the columns of rain.df? What are those names?

d. What command would you use to get the value at row 2, column 4? What is the value?

e. What command would you use to display the whole second row? What is the content of that row?

f. Create a new column called daily which is the sum of the 24 hourly columns

g. Make a histogram of the daily rainfall amounts.

Problem2: A First attempt at Predictive Analytics Problems

Go through the following problem of kaggle:
https://www.kaggle.com/c/sf-crime/data

Do the following tasks:

a. Apply random predictions to each observation and find out how much accurate your predictions are by submission to kaggle?

b. Find out the pattern in the data manually and then hard-code the logic. Resubmit your predictions and compare with random predictions?