Boston University
Department of Mathematics and Statistics

MA575 – Fall 2018
Linear Models
Lab Session 2: Projects

Project Report Due: Week of the 24th of September 2018 at beginning of lab

The purpose of this lab session is to get you thinking about the topic of your project and analyze questions of interest that your regression model will explain. As part of this course, you will be working, along with your group, on a data analysis project.

Pre-lab activities

1. Please read project2018description.pdf file.

Lab activities

1. Get familiar with your group members. Each group should have 5—6 people. Your first task as a group is to come up with a creative name for your group. What is the name of your group? You should also decide if somebody (and who) in your group would be responsible for bringing a computer and for printing and handing in the assignments.

2. As part of your work in this course, you are asked to complete a course project. For this project, you should work on one data set from the following 3 choices:

   • *Bike Sharing*. For this project, you will use data on the usage of bike sharing resources (e.g., like the Hubway system in Boston).

   • *Forest Fires*. In this project you will use data from Portugal to predict how susceptible an area is to forest fires.

   • *Facebook Social Media Metrics* The data is related to posts published during the year of 2014 on the Facebook's page of a renowned cosmetics brand. This dataset contains 500 of the 790 rows and part of the features analyzed by Moro et al. (2016).

   These data sets are located in the bulearn website under

   "Content -> Project and Lab Materials -> Project Data"

   Please read the project2018description.pdf file for a general description of the 3 data sets. For more details, please read the cited papers that are also contained.

3. Explore the 3 data sets. Read the DataSetDescription.txt file for each of the data sets. Now start discussing the potential data set project in your group. At this point, try to identify an interesting question that can be approached by inference and prediction.

   For example, from the bike sharing data, suppose that we have data for the daily number of bike rentals for 2012 and other observations such as day, month, year, holiday, weather, etc. An interesting prediction question to ask is *What are the daily bike rentals for 2013 ?* As a business you are interested in having the least inventory possible to reduce your costs.

Post-lab activities

Please read the project2018description.pdf file.

You have a project deliverable, “Description of Potential Topic Areas” due at the beginning of lab in two weeks. At most one page, it should summarize your tentative topic (bike rentals, forest fires or social media metrics) and list interesting questions that can be approached by regression analysis, prediction and inference. Please put at least 3 questions of interest.