

# MAST 6251: HW1

## Bike Share Regression

Due 11:59pm the night before Week 2 Class:

Sections 1 and 2: Wednesday 1/18 at 11:59pm

Section 3: Thursday 1/19 at 11:59pm

## 1 Background

Bike sharing systems are new generation of traditional bike rentals where whole process from membership, rental and return back has become automatic. Through these systems, user is able to easily rent a bike from a particular position and return back at another position. Currently, there are about over 500 bike-sharing programs around the world which is composed of over 500 thousands bicycles. Today, there exists great interest in these systems due to their important role in traffic, environmental and health issues.

Apart from interesting real world applications of bike sharing systems, the characteristics of data being generated by these systems make them attractive for the research. Opposed to other transport services such as bus or subway, the duration of travel, departure and arrival position is explicitly recorded in these systems. This feature turns bike sharing system into a virtual sensor network that can be used for sensing mobility in the city. Hence, it is expected that most of important events in the city could be detected via monitoring these data.

## 2 Data

Bike-sharing rental process is highly correlated to the environmental and seasonal settings. For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors. The core data set is related to the two-year historical log corresponding to years 2011 and 2012.

- The data is available in *HW1 bikeshare.csv*
- A readme text file outlines more information on the variables *HW1 readme.txt*

### 3 Objectives and Deliverables

- Work in small groups (less than or equal to five) and complete a regression analysis on the data, the goal is to predict the number of daily bike users. You decide whether “casual”, “registered” or “total” is most valuable as a dependent variable.
- **Pretend that I am a manager that wants new insights from the data.** I am interested in making a decision based on this analysis that I would not otherwise make. As in the real world, your job is to determine an interesting insight that would catch my attention.
- **Tell a story with the data**, which variables are related, are there outliers, how are variables distributed? But be succinct - your manager does not have an MSBA degree.
- Complete the analysis in R Markdown. Submit 1) the code and 2) the output file as a PDF. There is no need to show the code in the output file, if I am interested I will look at the raw .rmd file. This also goes for, say, model selection: show your complete work in the .rmd file but suppress irrelevant output for the PDF summary.
- Summarize the analysis in a succinct memo: **3 pages at most of PDF writeup, which includes graphs and tables.**
- **Make sure to use the model for decision making.** What new information or insights can be gained from the model? Are there any potential actions that a manager could take in order to increase revenues? What ideas do you have? Get creative! And don’t be afraid to introduce *reasonable* assumptions (e.g., promotional costs, etc.) for your ideas. What is the impact of your suggestion? Remember, the model is just a means to an end - tell me what we should change based on the model findings.
- Upload your files onto Canvas. **Please make your groups in Canvas and upload one set of files (.rmd and .pdf) per group.**
- Please include all team member names and SMU IDs on the .pdf output