**Assignment 1.2 – Course Project Step 1**

Filipp Krasovsky

University of San Diego

# Background

The data I will be working with for the remainder of this course is bike sharing information generated by the Laboratory of AI and Decision Support (LIADD) from the University of Porto from 2011-2012.

Bike sharing data from the Capital Bikeshare System (Capital Bikeshare, 2020) was

aggregated on an hourly basis along with weather and temperature conditions for the DC Metropolitan area from freemeto.com (Freemeto, 2020). In particular, the data includes an index time series, available at both a daily (n=732) and hourly (n=17380) granularity, and includes dummy variables for holidays, seasons, temperature conditions, and working days, as well as data about the month and day of the week. Further weather data includes normalized weather in Celsius, which includes feeling temperature, humidity, and windspeed. Finally, bike usage data consists of ridership by casual riders as well as those who have a membership in the bikeshare system. According to the readme file from the UCI Machine Learning Repository, the intended purpose of aggregating this data is twofold; the first being the more mundane purpose of forecasting ridership in the DC Metropolitan area, and the second being the detection of anomalous events which are associated with large fluctuations in rentals like natural disasters (Fanaee-T, 2013).

Table of Contents

[Background 2](#_Toc50938167)

[Intent 4](#_Toc50938168)

[Similar Applications 5](#_Toc50938169)

[References 6](#_Toc50938170)

# Intent

In my exploration of this dataset, I intend to create a multiple-regression forecasting model that would explore whether relative changes in temperature are a key driver of bike sharing activity or absolute temperature serves as a better indicator. Given the climate diversity of the United States, this data could provide insight into the likelihood of adoption of similar programs in areas like Texas & Nevada, which tend to have high temperatures throughout the year. The primary cause for distinguishing between relative variation and absolute temperature differences involves an inquiry into whether or not states with chronically high temperatures are less likely to have significant ridership compared to the DC area, which has a regular seasonal variation in weather conditions.

# Similar Applications

Analogous usage of large-scale bike sharing data has been applied in two notable cases.

The first of these is an analysis of ridership by young commuters in the Nanjing metropolitan area in China. The study surveyed smart-card data from bikeshare commuters aged 18-35 over a three-week period (n=1,154) with the purpose of determining ridership factors including socio-economic demographics, travel-related attributes, and more appropriately, weather patterns. This information was intended to be applied in shaping policy on bikeshare program integration in different parts of the country based on the aforementioned variables. The study found, with a p-value < 0.001, that a model consisting of built environmental factors, socio-economic makeup, and travel duration and time of departure provides explanatory power that exceeds the null model, although the findings were somewhat ambiguous about the effect of weather. (Liu, Ji, Yang, & Timmermans, 2020).

The second use of this data was by Noland & Gebhart in 2014, who analysed the effect of weather on bikeshare activity in Washington D.C from 2011 to 2012, much like the study in question, and found that increasing humidity, rain, and low temperatures tend to decrease ridership overall, with a compounding effect if the bikeshare station is not near a metro stop. (Gebhart, Noland, 2014)

# References

Capital Bikeshare. (n.d.). *System Data*. Retrieved September 13, 2020, from https://www.capitalbikeshare.com/system-data

Fanaee-T, Hadi, and Gama, Joao, "Event labeling combining ensemble detectors and background knowledge", Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg, doi:10.1007/s13748-013-0040-3.

Freemeto. (n.d.). *Freemeto*. Https://I-Weather.Com/. Retrieved September 13, 2020, from https://freemeteo.nl

Fanaee-T, H. (2013). *Bike Sharing Dataset*. UCI Machine Learning Repository. <https://archive.ics.uci.edu/ml/machine-learning-databases/00275/>

Liu, Y., Ji, Y., Feng, T., & Timmermans, H. (2020). Understanding the determinants of young commuters’ metro-bikeshare usage frequency using big data. Travel Behaviour and Society, 21, 121–130. https://doi-org.sandiego.idm.oclc.org/10.1016/j.tbs.2020.06.007

Gebhart, K., & Noland, R. B. (2014). The Impact of Weather Conditions on Bikeshare Trips in Washington, DC. Transportation, 41(6), 1205–1225