Name (Last, First):

Kunjangada Sushma Arun

NEU ID:

002473132

NEU COE INFO6105 Final Project Proposal

(due 11/09/2024)

UCI Bike Sharing Dataset

• License: This dataset is made available under the Creative Commons Attribution 4.0 International License (CC BY 4.0)

• Modified: This dataset is unmodified and is presented as-is. (12/19/2013)

• Publisher: UCI Machine Learning Repository.

• Dataset Link: https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset

Dataset Information: Contains 17,379 hourly records and 731 daily records from Capital Bikeshare system spanning 2011-2012, with both temporal and weather-related features. This real-world dataset provides comprehensive information about bike rental patterns influenced by environmental and seasonal factors, making it valuable for analyzing urban mobility patterns and optimizing bike-sharing operations.

Question 1: "How do weather conditions (temperature, humidity, and wind speed) correlate with daily bike rental patterns during different seasons?"

Methodology: The study will employ correlation analysis to quantify relationships between weather variables and rental frequencies. Scatterplots with regression lines will provide visual representation of these relationships. Multiple regression analysis will identify significant weather predictors of rental behavior. Confidence intervals for regression coefficients will establish the reliability of our findings. This comprehensive approach will reveal how weather conditions influence user behavior across different seasons.

Question 2: "Is there a significant difference in mean bike rental numbers between weekdays versus weekends across different seasons?"

Methodology: A two-way Analysis of Variance (ANOVA) will examine the main effects and interactions between day type and seasonal factors. Descriptive statistics will summarize rental patterns, while box plots will visualize distributions. Confidence intervals for mean differences will quantify the significance of observed patterns. This analysis will reveal temporal usage patterns crucial for operational planning.

NEU COE INFO6105 Final Project Guidelines

(updated 11/02/2024)

One of the requirements for INFO6105 is the final project (***20% of the grade***). **The project must be done individually.** This is an opportunity for you to be creative in solving a problem that interests you and demonstrate your understanding of the concepts and principles and your proficiency with R or Python. The project should be challenging enough so that you could discuss it at future interviews with potential employers. Another possible benefit of this project is that it can give your professor a good topic for discussion should you ever desire a reference.

Before beginning your project of cited literature, you MUST submit via Gradescope a few sentences (250 words max, describing your project proposal (due November 09, 2023, at noon ET, worth 5% of the final project). This is so your professor can assess the appropriateness of your idea.

The project is due by December 07, 2023, at noon ET and includes a 4-minute presentation of your work. Your verbal and written communication and presentation skills will be judged for 4 Extra Credit points.

In addition to the presentation, you are required to submit the following through email attachment in one zip file called **<username>_final_project.zip**:

- **One pdf file of your final project report (no more than eight pages)**
- **One .R or .py file (including instructions on how to run your code written as comments).**

Here are some key criteria to consider when determining your final project and its elements.

1. It must be original work and not something that might be proprietary to your company or generated by ChatGPT or similar AI tools.
2. The presentation and well-documented code should be at the level that other students and lay people can understand your project. Do not use advanced math or industry terms that require a lot of explanation.
3. You can choose any topic that interests you and conforms to the above criteria. The most important thing to remember is that this final project is meant to demonstrate your ability to apply what was learned in the course. You will not be judged on the originality of your topic or the difficulty of implementation. That said, if your project is overly simplistic, that will be held against you. We want you to show us the concepts and principles taught in this course that are well-understood and conform to best practices. Again, think of this as something you can show an employer as an example of why they should hire you as a data engineer, a data analyst, or a data scientist.

4. Your project must include at least the minimum number of each of the following:
    a. One data set from reputed sites such as (and cite the data source):
        i. UC Irvine Machine Learning Repository (https://archive.ics.uci.edu/)
        ii. CMU StatLib (https://lib.stat.cmu.edu/datasets/)
    b. Two meaningful questions about the data set are asked, to be answered with two or more methods taught in this course.
    c. Use R or Python for all the data analysis and visualization procedures.
    d. Write a final project report (**no more than eight pages**) with the following sections: Introduction, Methods, Results, Discussion and Conclusion, References (cite all the sources).
    e. Include a URL link to a video you made that recorded the process of running your R or Python program to complete your data analysis and visualization. (You can use YouTube, Vimeo, or other cloud video providers as long as you can provide your professor with a link. YouTube allows you to create an unlisted video that's not publicly available, but anyone with the link can view the video.)
    f. **It is important to remember that longer is not better, and you will lose credit for not being succinct in your presentation.**

When submitting your project, include a .R or .py file with your code plus instructions (written as comments) for running the code and installing any 3rd party modules or libraries. We cannot grade your project if it uses proprietary modules or if we cannot run it.

Finally, we want to emphasize that **your project MUST NOT contain proprietary, nonpublic, or confidential algorithms and data from your employer or other sources and MUST NOT be generated by ChatGPT or similar AI tools.**

Good luck, and have fun!