Proposal for the Final Project STAT 5320

Anti Li(T00751339), Feng Gu(T00751197), Yuzhuo Ye(T00751492)

March 12, 2025

1 Team members

Our group has three members:
Anti Li(T00751339), Feng Gu(T00751197), Yuzhuo Ye(T00751492).

2 Meeting and discussion

We discussed the project with one meeting and one discussion. The meeting was on Mar 12 afternoon, in person. The discussion was held online.
Discussion(online):

- Time: Mar 9, around 18:00

- Content: Discussed the major task type in our project, and talked about the skills might be used in the project. Decided the formal meeting time and the possible data set we might choose.

Meeting(in person):

- Time: Mar 12, from 14:00 to 15:40

- Content: Discussed the general idea, steps and the exercises that can be finished in the project. Decided the data set we are going to use. Divided the work into different parts and roughly assigned the tasks to each member.

3 Data set to be used

The data set we are going to use is the **Energy Appliances** data set. Its shape is **(19735, 29)** and the data is about the energy consumption of appliances in a low energy building. The data set is collected in a house with a total of 10 electrical appliances. It is noticeable that the data was collected every 10 minutes for about 4.5 months, which means there are some time series features in it. However, we would not focus on the time series analysis that is beyond the scope of this project and thereby we would make some assumptions to simplify our analysis.

Table 1 shows the data variables in the Energy Appliances data set.

Table 1: Data Variables in the Energy Appliances Data Set

Data Variables	Units	Number of Features
Appliances energy consumption	Wh	1
Light energy consumption	Wh	2
T_i: Temperature in room i	˚C	3-21
RH_i: Humidity in room i	%	4-20
Temperature outside	˚C	21
Pressure	mm Hg	22
Humidity outside	%	23
Windspeed	m/s	24
Visibility	km	25
Dew point	˚C	26
Random Variable 1 (RV_1)	-	27
Random Variable 2 (RV_2)	-	28
Number of seconds from midnight	s	29
Week status (weekend or weekday)	-	30

4 Major tasks and exercises will be done

4.1 Major tasks:

- Data cleaning and preprocessing, including missing value imputation, detecting and handling outliers, data transformation, etc.

- Exploratory data analysis, including data visualization, correlation analysis.

- Model building and comparison, including using different improvement techniques to improve the model performance, like feature selection, hyper-parameterλ tuning, etc.

- Model evaluation, including using different metrics to evaluate the model performance. Test the selected model on the test set and assess its performance.

4.2 Corresponding exercises:

- Residual analysis.

- Multiple linear regression.

- Creating and using dummy variables.

- Variables selection.

- Perform statistics inference on a coefficient.

- *We might try to change the assumption about the residual in MLE for estimating the co-efficients in regression, which we had tried a little bit in the assignment 2 last week. But it depends on the time and the progress of the project.

  Click here to see the previous code of trying it.

5 Attestation

We attest that **no team member** has analyzed the data in any relevant way previously. We will also not plagiarize ideas or solutions from others to complete the project, even though other people might have done similar projects on top of the same data set.

6 Temporary Schedule for the Project

Tasks	Content	Name	Due Date
Draft abstract	Summary of our report, covering the general work and conclusions	Yuzhuo Ye	mat 20th
Introduction	Why doing the research and how	Yuzhuo Ye	Mar 24th
Data preparation	Preparing, cleaning, transforming, and description	Anti Li	Mar 24th
Fitting model and doing tests	Searching a good model and making sure models pass statistical tests	Anti Li, Feng Gu	Mar 25th
Results	Explaining results for practical meanings, thinking drawbacks and possible improvements	Yuzhuo Ye, Feng Gu	Mar 27th
Discussion/Conclusion	Potential future work derived from our work and the following work that could be done to improve it. The meaning of our report in the real world.	Anti Li, Yuzhuo Ye	Mar 28th
Checking and refining	Checking core contents in the report, making possible improvements and organizing them in a clearer way.	Feng Gu	Mar 31th

Table 2: Overview of Tasks and Responsibilities