

Introduction to Data Science
IST687

Project Update 2

Submitted by

Bhavika Karale

Net ID: bmkarale

Abhishekh Shinde

Net ID: ashinde

Goutham Sri Vishwesh Bikkumalla

Net ID: gbikkuma

Minnie McMillian

Net ID: msmcmill

Prasanna Lakshmi Chelliboyina

Net ID: pchellib

Under the guidance of

Prof. Anderson, Prof. Saltz

Professor

Fall 2023

Contents

1	Completed Tasks:	1
2	In Progress Tasks:	1
3	Challenges faced:	1

1 Completed Tasks:

1. Obtained datasets via AWS.
2. Downloaded all the weather and home energy datasets, matched weather data sets based on county ID with house energy data sets, merged five datasets related to weather and housing energy in the time column.
3. Investigated every dataset to determine which variables to incorporate into the combined dataset.
4. Worked on utilizing R loop to combine all the datasets. Combined energy usage by datetime; weather using datetime and county code.
5. Sought to see if there was a more efficient way to import and merge all the datasets.
6. Perform data cleaning, transformation and visualization.
7. Analyse exploratory data.
8. Perform data cleaning, transformation and visualization.

2 In Progress Tasks:

1. Investigate the operation of time series and the data formatting. Merge data to develop a time series model.
2. Work on Linear Regression Model.
3. Work on association models with some of our dataset and practice on some simple data visuals.
4. Create report and presentation.

3 Challenges faced:

1. Merging all the datasets together is difficult.
2. Understanding parquet format and data import techniques.
3. The dataset is too large. Hence, analyzing the datasets is time consuming.
4. Choosing a Machine Learning model to obtain better prediction