



**RMIT**  
**UNIVERSITY**

## **MELBOURNE DATATHONE**

### **Insights category**

**SUBMISSION DATE** : 18 October 2020  
**STUDENT NAME** : Rashbir Singh Kohli  
**EMAIL** : rashbirkohli@gmail.com  
**PHONE** : +61 468787652  
**LINKEDIN** : <https://www.linkedin.com/in/rashbir-singh-kohli-a617b4112/?originalSubdomain=au>  
**GITHUB** : <https://github.com/RashbirSingh>

### Link To Code

[https://github.com/RashbirSingh/Melbourne\\_datathone.git](https://github.com/RashbirSingh/Melbourne_datathone.git)

## 1. The Challenge

We want you to analyse some electricity consumption data and determine if a change can be detected that may be attributable to covid-19.

## 2. Category

Insights category

## 3. Data

This data set for those who are seeking mastery in timeseries data analysis. The data set contains hourly power usage of 2storied house located in Houston, Texas, USA. The data set contains hourly power usage in kwh starting from 01-06-2016 to August 2020. The dataset has marked notes for weekdays, weekends, COVID lockdown & vacation days in notes category column.

Power usage during day time is different from night time. The electrical devices that are inside the house are security DVR and POI cameras, 2 x refrigerators, 2 x 50gallon water heater that are on during day time. At night several electrical bulbs, TV's, washing machine, dryer and AC run from evening 6pm to morning 8am.

Another data contains historical weather report of Houston, Texas starting from 01-06-2016 to August 2020. Thanks to wonderful weather at Houston, Texas we are blessed with almost 9 months of summer. Starting from Feb month to Nov month and winter is only for two months December and January for most of the years.

The power usage data, extracted from "TRIEAGLE ENERGY LP, The Woodlands, Texas 77393"

The historical weather data for Houston, Texas[1]

## 4. Data Columns

- Vacation setting:- "'AC and electric bulbs turned off"
- COVID- Lockdown:- 'AC is turned on during day time, laptops, monitors etc., are on"
- Weekday:- 'Morning 7am to 5pm "AC is 84 F temperature during summer and heating set at 60 F during winter".
- Weekends:- 'Room Temperature is set at 78 F during summer and 68 F for heating during winter" [2]

## 5. Exploratory Data Analysis

Following results came from EDA

- Energy consumption is somewhat directly proportional temperature (figure 1).
- Energy consumption is somewhat directly proportional dew point (figure 1).
- Temperature point is inversely proportional to atmospheric pressure.
- Dew point is inversely proportional to atmospheric pressure.
- So, for feature selection pressure won't be considered as they correlate.
- Energy consumption was low in the year 2020 as compared to the year 2019 and 2018 (Figure 2, 3 and 4).
- In actual people consumed less than the previous year for the year 2020 as compared to the year 2019 and 2018 (Figure 5).

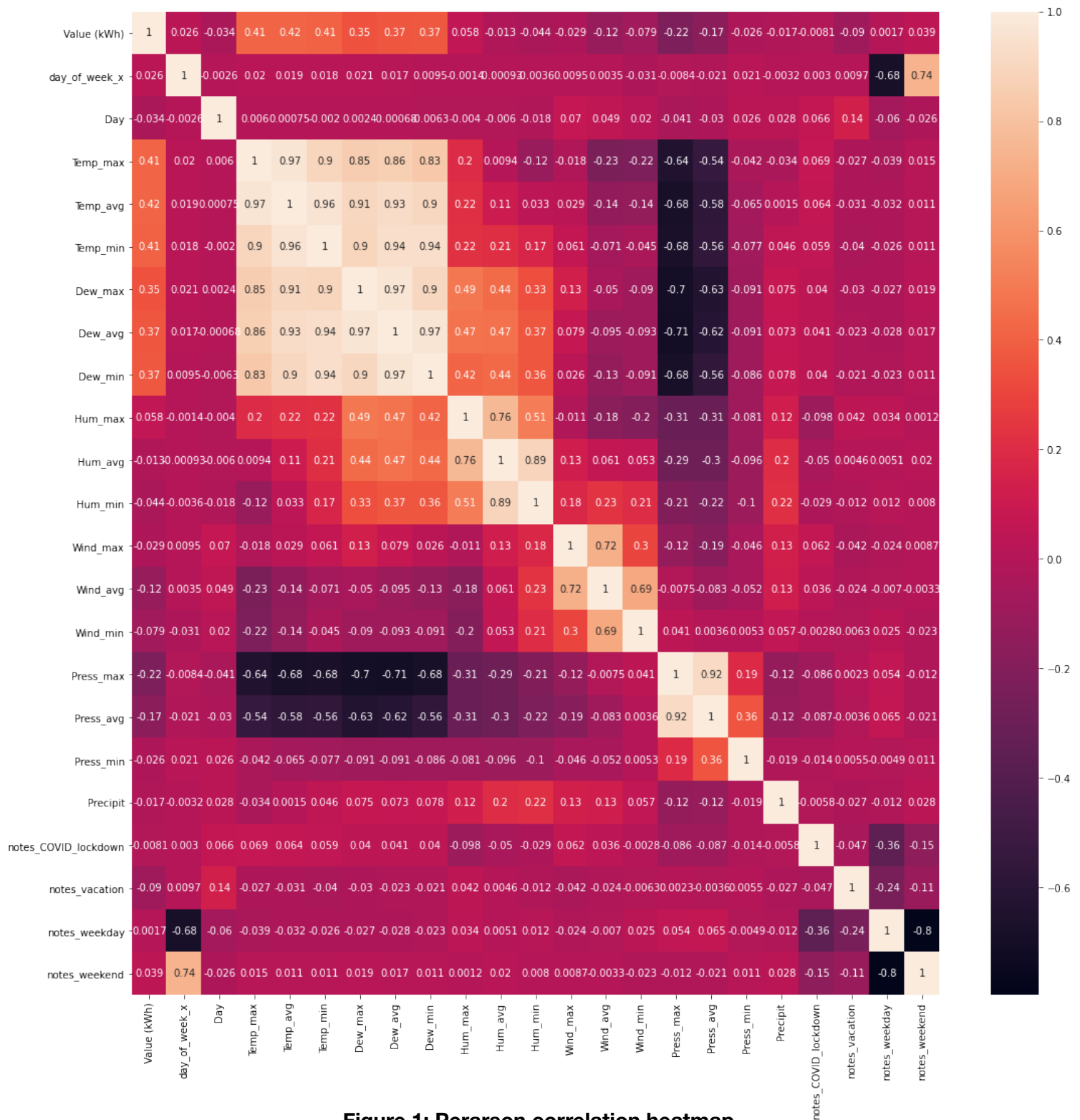
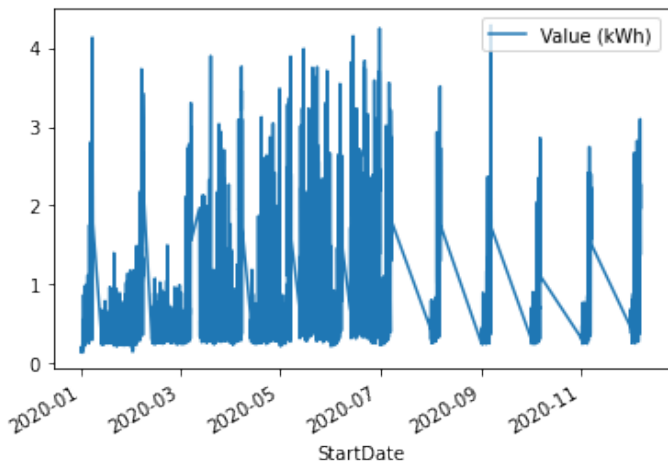


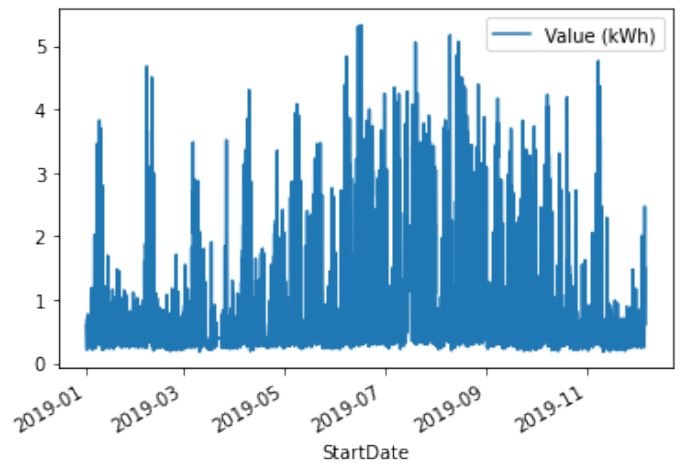
Figure 1: Pearson correlation heatmap

## 6. Data pre-processing

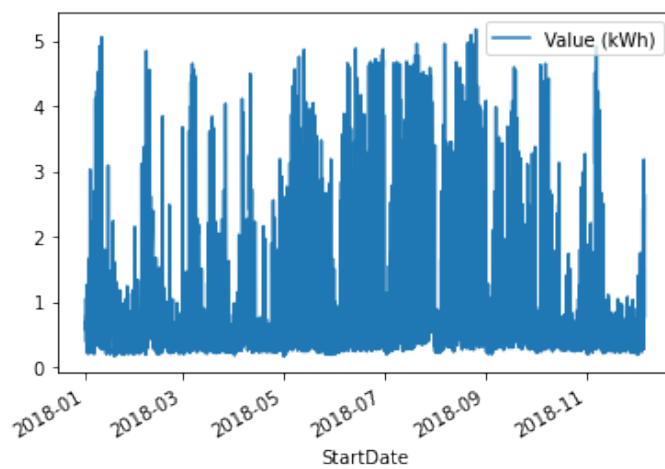
- Forward filling of data was performed after merging the weather data with energy consumption data.
- One hot encoding was performed using dummies function of pandas to convert object-based categorical data to encoded format.
- Data was scaled using scaler in range of 0 and 1 to reduce high variance for further processing.



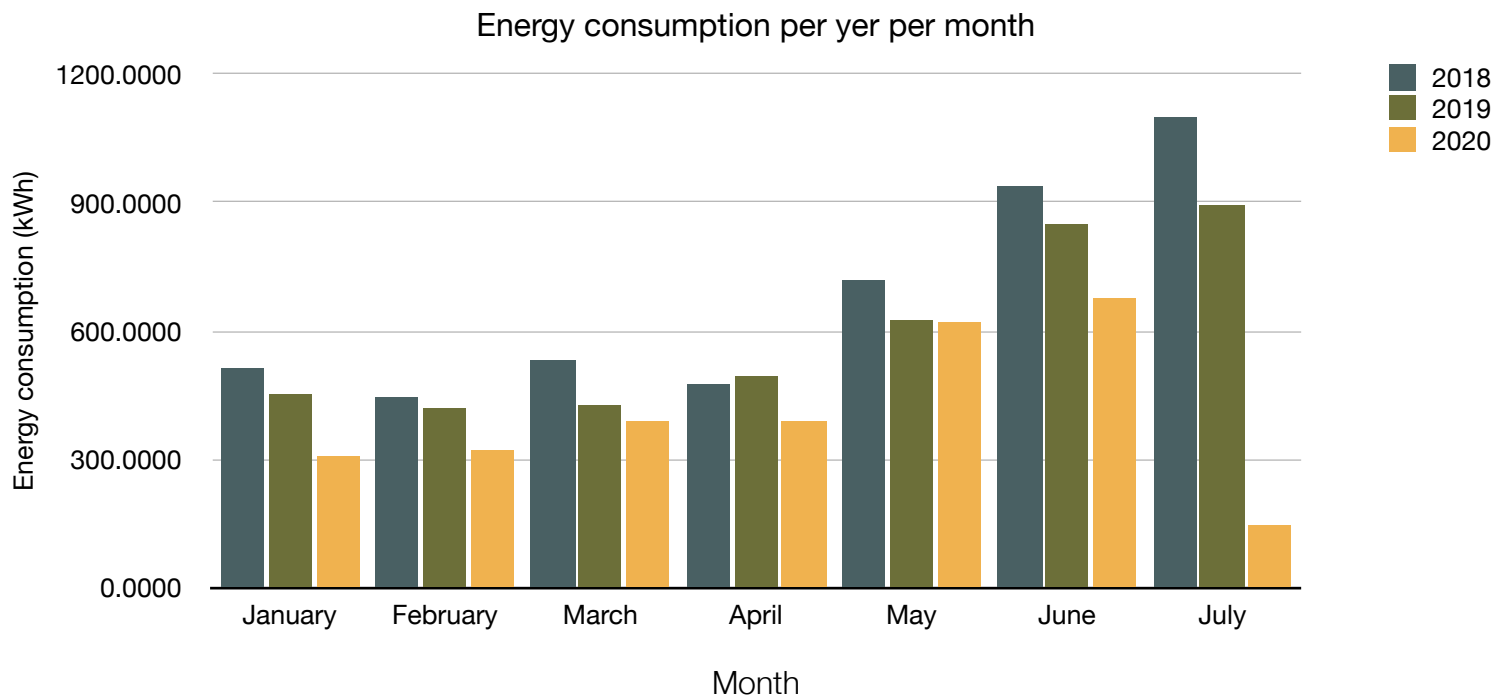
**Figure 2. Energy consumption in year 2020**



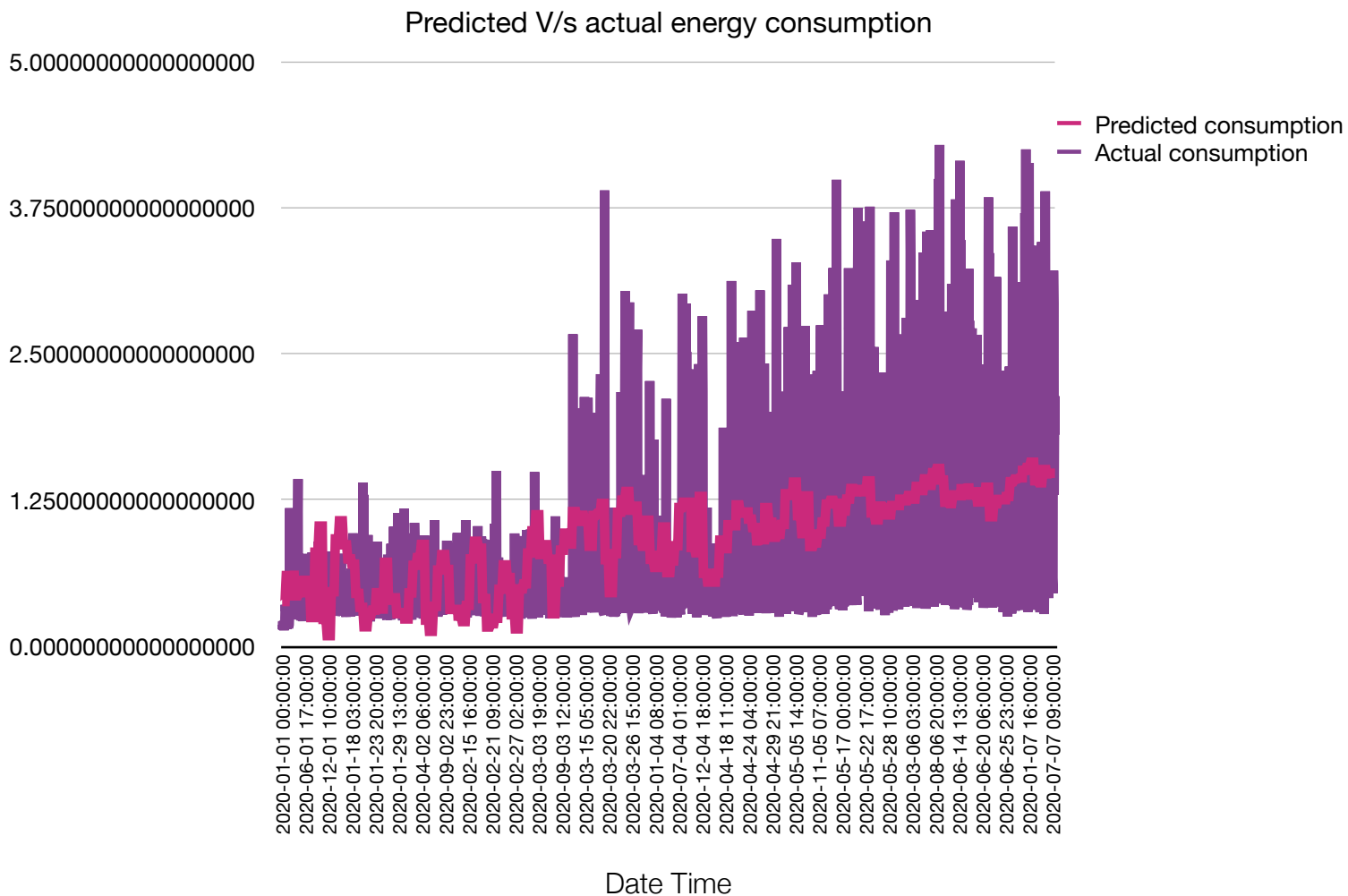
**Figure 3. Energy consumption in year 2019**



**Figure 4. Energy consumption in year 2018**



**Figure 5. Energy consumption in year 2018**



**Figure 6. Predicted V/s actual result**

## 7. Model selection and tuning

For the analysis, three models were considered and accuracy was compared after hyperparameter tuning:

- Simple Linear Regression
- Lasso Regression
- Ridge Regression

The models were trained from the year 2016 to 2018 and trained in the year 2019 using the hyperparameter tuning results the best-selected model is Lasso with alpha as 0 and with 72.14% accuracy.

Based on the analysis a new model was trained from the year 2018 to 2019 and prediction was made for the year 2020 based on the selected features using selectKBest and selected features are:

- Temp\_max
- Temp\_avg
- Temp\_min
- Dew\_min

The results are shown in figure 6, on average people consumed 0.1756 kWh more during the covid lockdown.

## 8. References

[1] [www.wunderground.com](http://www.wunderground.com) fetched on 18 October 2020

[2] <https://www.kaggle.com/srinuti/residential-power-usage-3years-data-timeseries> fetched on 18 October 2020