**Data Analytics and Visualization of Smart Home Appliances**

**19BCE1460, KISHORE DS, Vellore Institute of Technology, Chennai, India**

**19BCE1726, DHAYA SAGAR K, Vellore Institute of Technology, Chennai, India**

**ABSTRACT**

When it comes to an automated system there is always a need for continuous monitoring and data visualization in order to check and ensure if the system runs normally and smoothly, and also a proper visualization may even let a lay man understand the pattern of the system functioning. So, in our project we considered a dataset containing records of energy generation and consumption by various appliances at every minute along with other oriented environmental data like temperature, humidity, etc. We use the dataset contents to visualize and predict various parameters oriented with the smart home appliances’ functioning and their energy usage, environmental conditions, energy generation, etc.

**KEYWORD**

Smart home appliances, data visualization and analytics, prediction

1. **INTRODUCTION**

The world is changing at a very faster pace where everything is getting automated and the for the management, analysis and visualization of the data utilized and generated by these devices had become very crucial.

Our project would mainly concentrate on visualizing and predicting various parameters and data that are oriented with a dataset containing information regarding smart home appliances, their functioning, their energy consumption, overall energy production, environmental conditions and other lively monitored environmental parameters.

We have visualized the data using various density plots, time series plots, simple exponential smoothening and predictive models like holt’s linear trend , exponential trend and additive damp trend. Through these visualizations we were able to make many clear inferences on the dataset’s nature and other patterns followed by the parameters of the dataset.

1. **Literature Survey**

The researcher in this paper [1] talks about the four domains' clusters namely service, finance, technology and organization. There were 15 core clusters and 52 sub-clusters found in all. In order to make the text more readable. The goal of this paper is to provide a comprehensive representation of Smart Living literature's current structure.

In this research [2], The functioning concept of smart home systems has been clearly stated after an intensive study on various sorts of smart home systems. To assess the benefits and drawbacks of various smart home systems, a comparative analysis of smart home systems was conducted. Because there are so many smart home technologies on the market, it's critical for customers to have a strong understanding of the various smart home systems so that they can choose the best system for their budget.

Paper [3] discuss about the application of heuristics based on linear programming to solve the cost reduction problem of air conditioning systems for residential and commercial loads was examined.

This [4] paper talks about using mixed integer linear programming, an optimal home load scheduling model was suggested.

The paper [5] defines the decision variable used to build a mathematical MILP and line programming (LP) or heuristic optimization problem for smart home appliance management. The number of appliances in a home area network can be raised to accommodate more types of appliances (HAN). Other dynamic pricing models can be used to investigate optimized appliance scheduling. In a smart grid setting, smart home energy management is becoming increasingly vital. It has the ability to handle demand response and management strategies. The basic goal of the home energy management challenge is to reduce overall energy usage by optimizing the schedule of various home equipment. As a result, the authors provide a complete analysis of solution strategies and methods for optimizing smart home energy management systems in this study.

1. **Feasibility Study**

There are plenty of data, parameters and patterns that can be extracted from the dataset that we have considered. Here, we specifically extract few of the parameters and patterns from the dataset. Few motivations of our project are:

* To analyze the pattern of total and energy production and the total usage by smart appliances.
* Make time series and periodic plots to analyze the pattern of usage, generation, temperature, etc. with respect time, days, month,etc.
* Apply various predictive models and check their accuracies with respect to few of the parameters in the dataset.
* To make appropriate analysis and obtain inferences on the visualizations in order to the pattern that the parameter follows, its impact and comparison with real world.

#### The most feasible solution to plot few density plots, make few times series analysis, make few prediction using certain existing models and inferring the outputs and data as per our requirements, and make analysis on it.

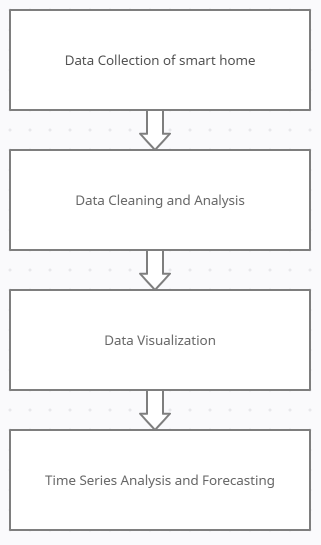
1. **About The Dataset**

We use a dataset that provides readings of house appliances every minute in kW from a smart meter, as well as meteorological data for that specific house. House appliances connected to and retrieved from smart meter are below:

* Amount of house energy consumption
* Dishwasher
* Furnace
* Home office
* Fridge
* Wine cellar
* Garage door
* Kitchen
* Barn
* Well
* Microwave
* Living room
* Solar power generation

By using this dataset, we can understand the relationship between energy consumption by appliances and time period, or detect anomalous usage of appliances, or clarify the relationship between weather information and energy generated by solar power.

1. **Design and flow of models**



**Fig: Design and flow of model**

for the analysis we have used the following modules and analysis parameters:

**Module 1: Data Collection of Smart Home**

For this module we have collect a dataset based on Smart Home Appliances from Kaggle and we imported and integrated it into google colab for further processing.

**Module 2: Data Cleaning and Analysis**

We used various data cleaning methods in python and removed all the unwanted values, null values and we have modified few columns as per our requirement inorder to make the further analysis, processing and plotting easier. This module includes removing the unwanted data from the dataset, adding the missing values, and arranging the data properly for further processing.

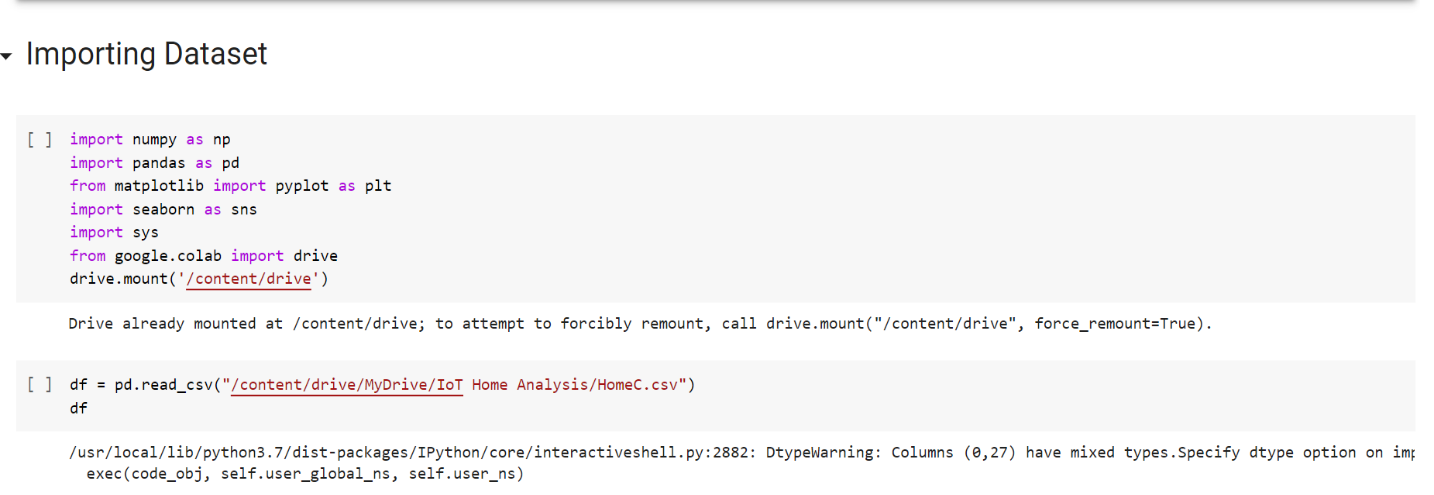
**Module 3: Data Visualization**

This is one of the most important and core modules of our project, this module includes various visualizations of different parameters of the dataset like energy usage, energy consumption, and temperature visualizations include correlation plots, density distribution plots, time series plots, and a few prediction plots.

**Module 4: Time Series and Forecasting**

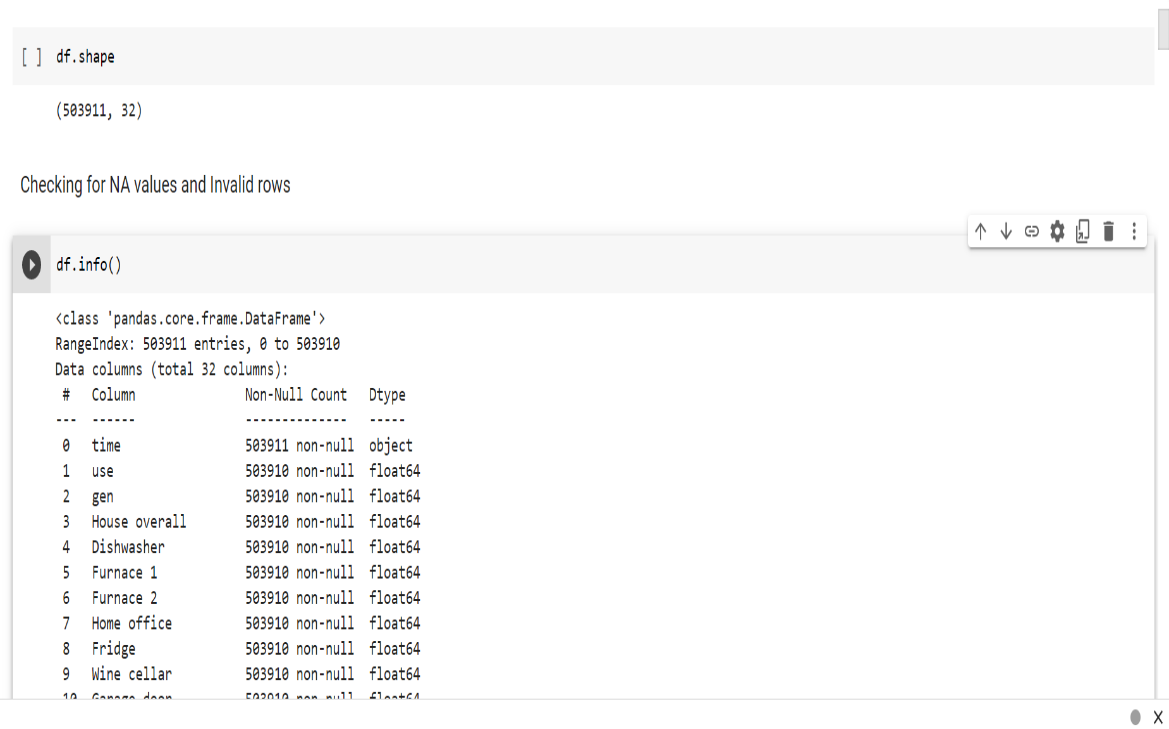
Time series and forecasting are also one of the crucial components of our project, this module includes various time series and forecasting plots like seasonal decomposition, exponential smoothing, holt’s linear trend, exponential trend, additive damped trend, etc. These plots help us get a deeper understanding on the data present in dataset making further analysis easier.

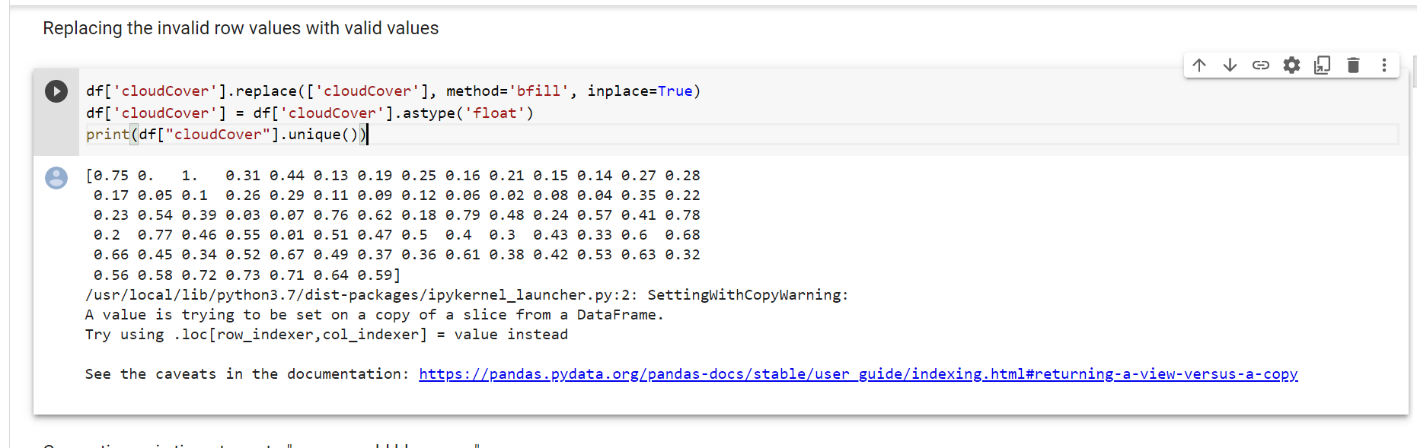
**6. IMPLEMENTATION**

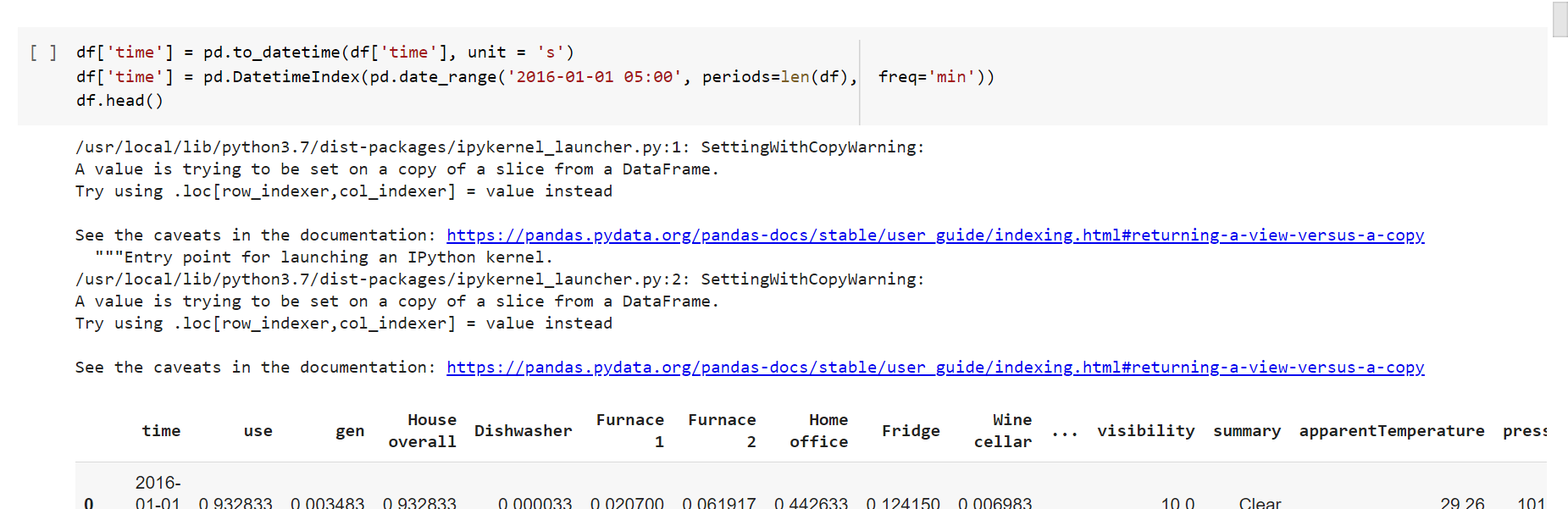
1. First we import modules that can help us collect, integrate and plot the data.

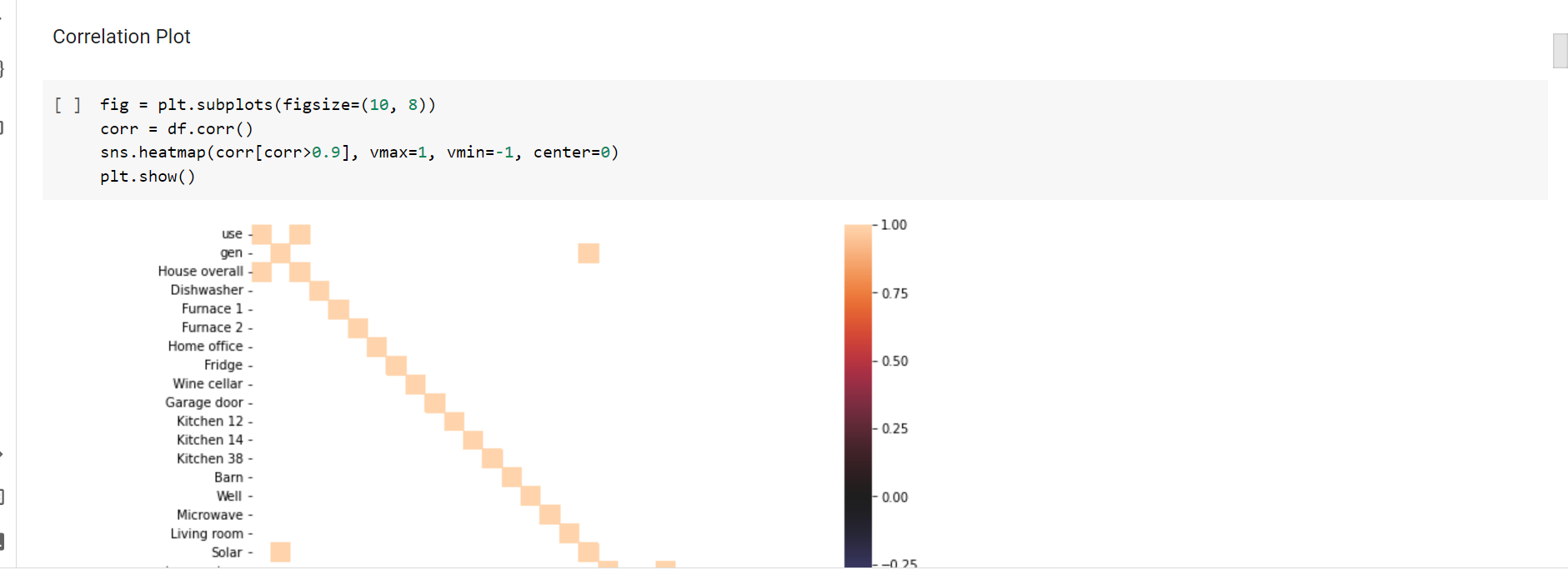
**Importing all the libraries and modules**

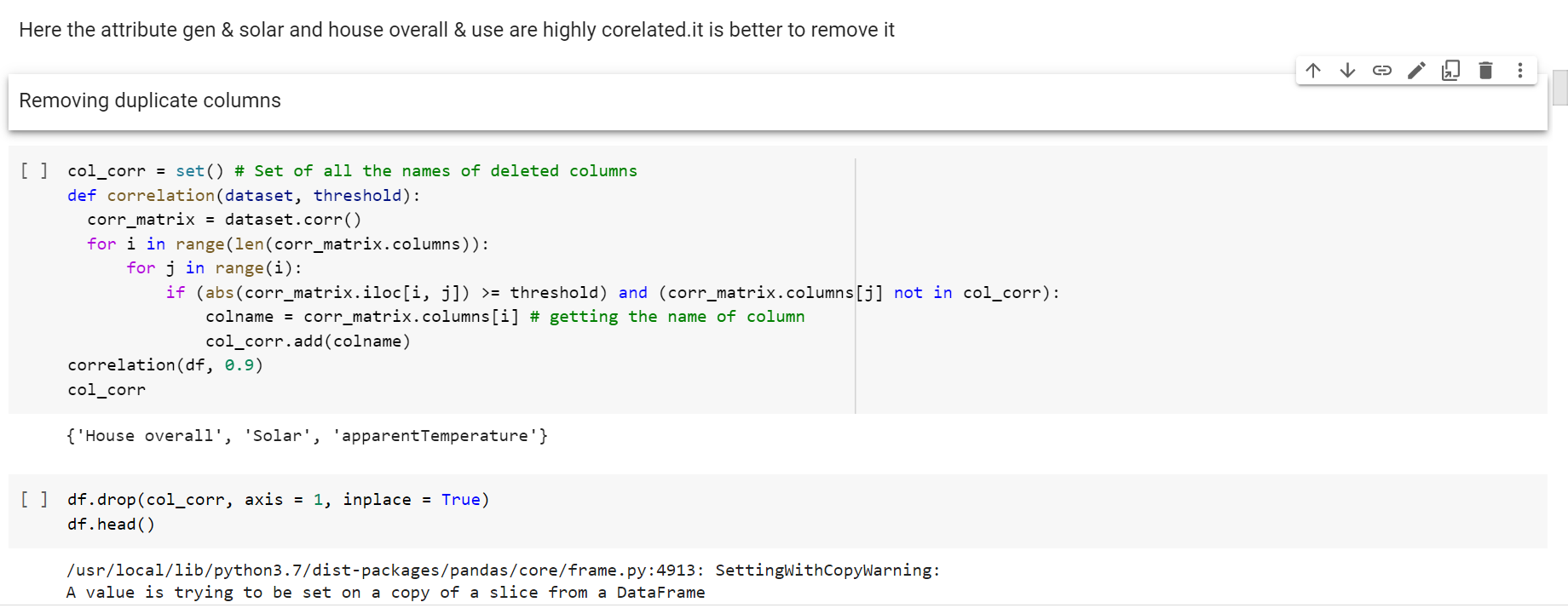
**2.Dataset analysis**

**Before & during preprocessing**









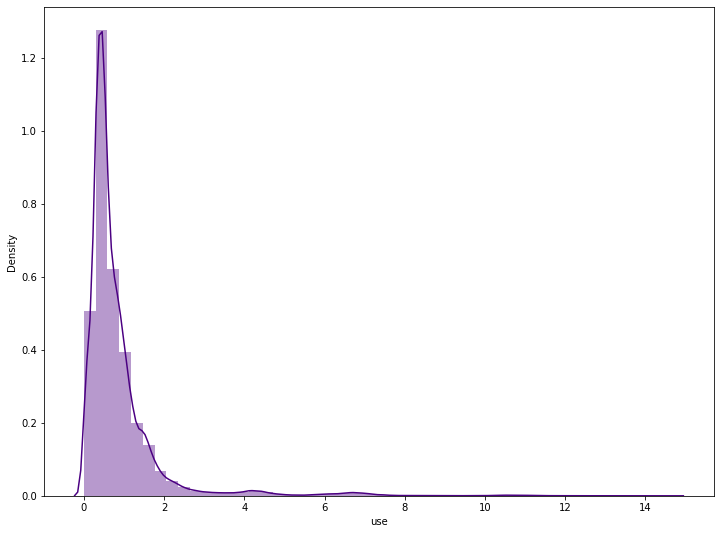
We have removed all the null and unwanted values, and we have modified them in order to make the data more compatible in order to facilitate the further processing for much better analysis and to obtain accurate results.

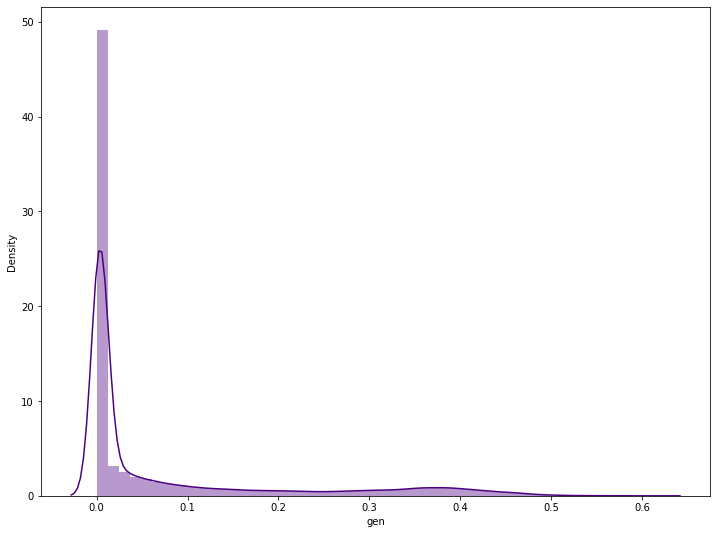
**After preprocessing**

After the preprocessing is over, we have the dataset ready to analyze, deduce and plot further required results and analysis. The following are the outputs and plots obtained using which the required analysis and inferences are made.

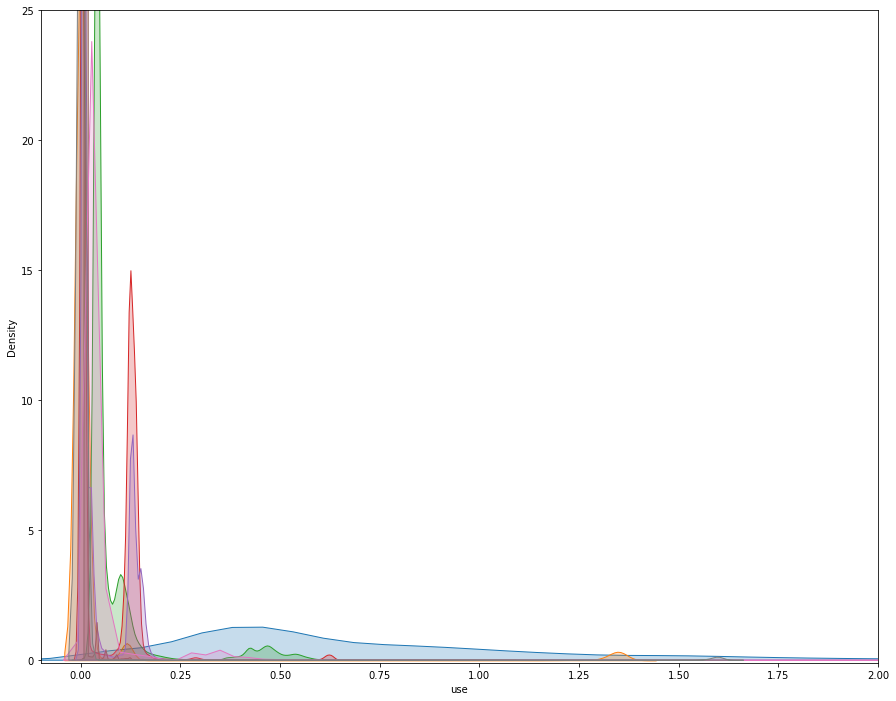
**Exploratory Data Analysis [EDA]**

**Density plots for Total usage of energy and Generation of energy**

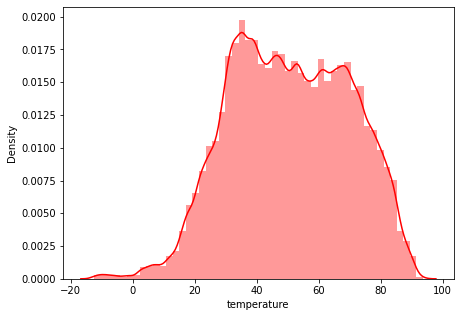


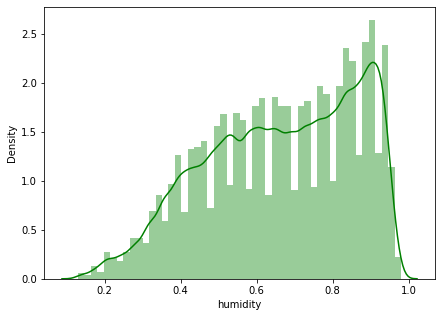
****

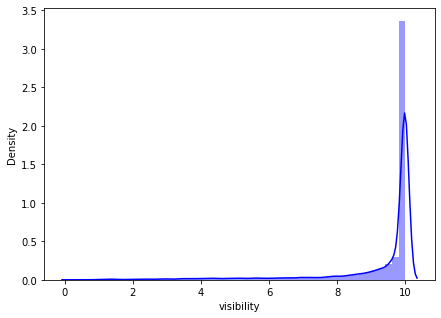
**Stacked density plot for appliances using energy**

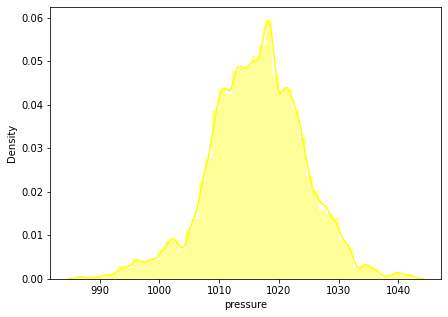


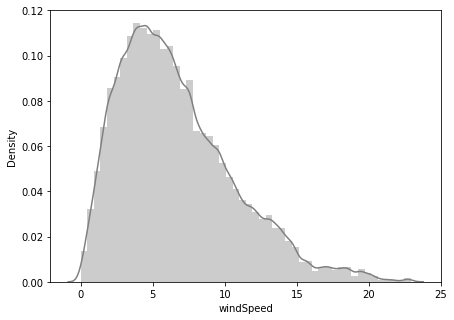
**Density plot for weather-based attributes**

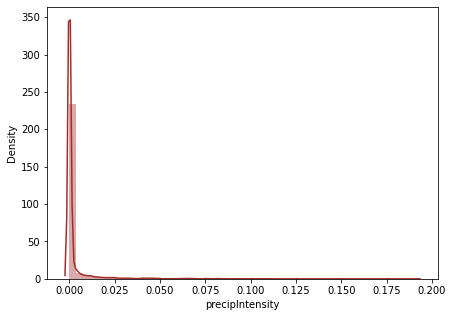
****

****

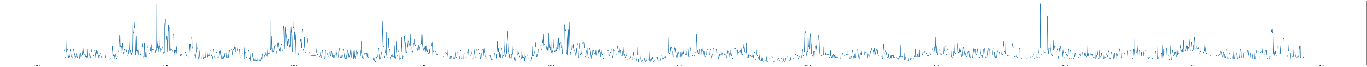
****

****

****

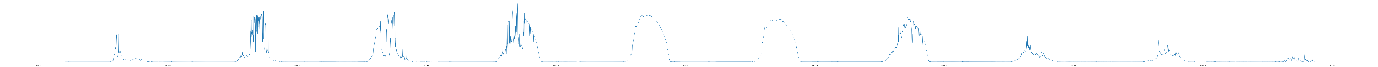
****

**Plot of Energy Usage Over 10 Days**



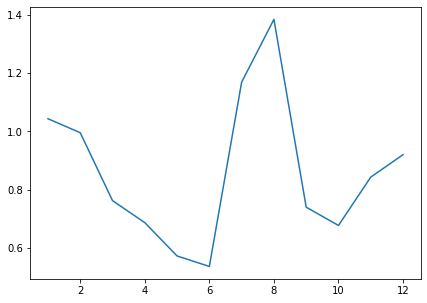
This has a low random up and down variation but has high up and down variation during the start of the day

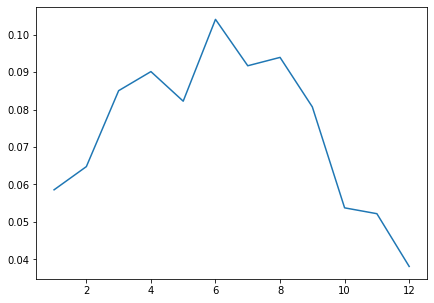
**Plot of Energy Generation Over 10 Days**



This has constant energy generation but has high generation on every 2nd half of a day

**Grouping Energy Generation and Energy Usage by month**

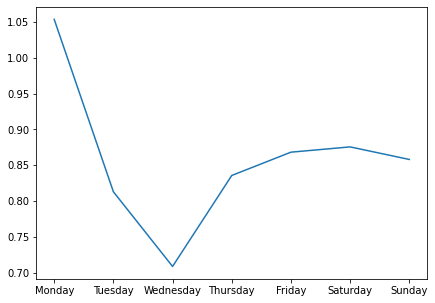


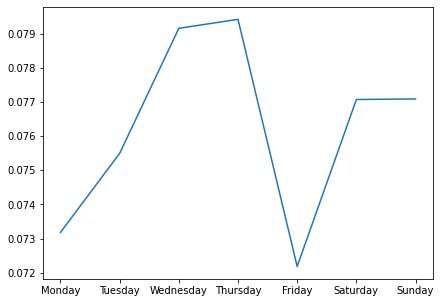


The Energy Usage is low during the May & July.

Has a spontaneous raise during July and spontaneous fall during September. Energy usage is highest at August. Energy Generation is lowest at Winter Season.

**Grouping Energy Generation and Energy Usage by weekday**

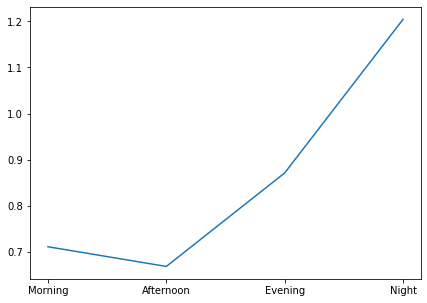
****

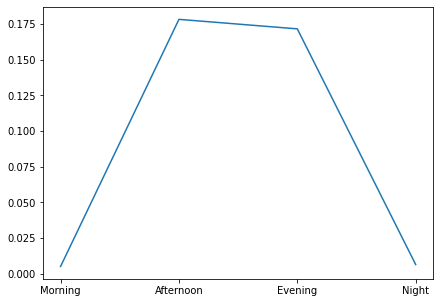
****

Energy usage is lowest on Wednesday

Energy generation is lowest on Friday

**Grouping Energy Generation and Energy Usage by Timing**

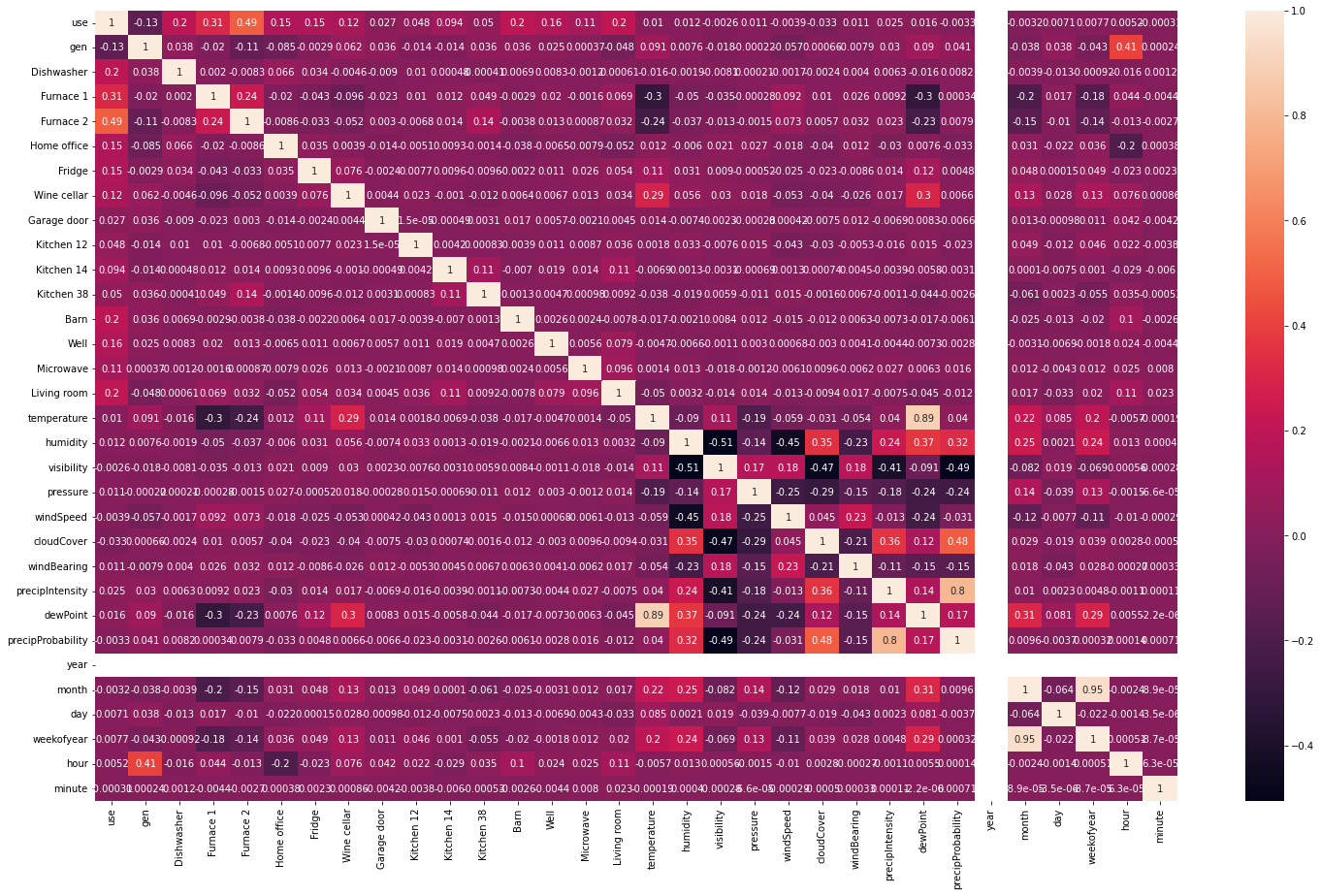
****

****

Energy usage raises after Afternoon

Energy generation gradually raises during Morning and gradually falls during Night

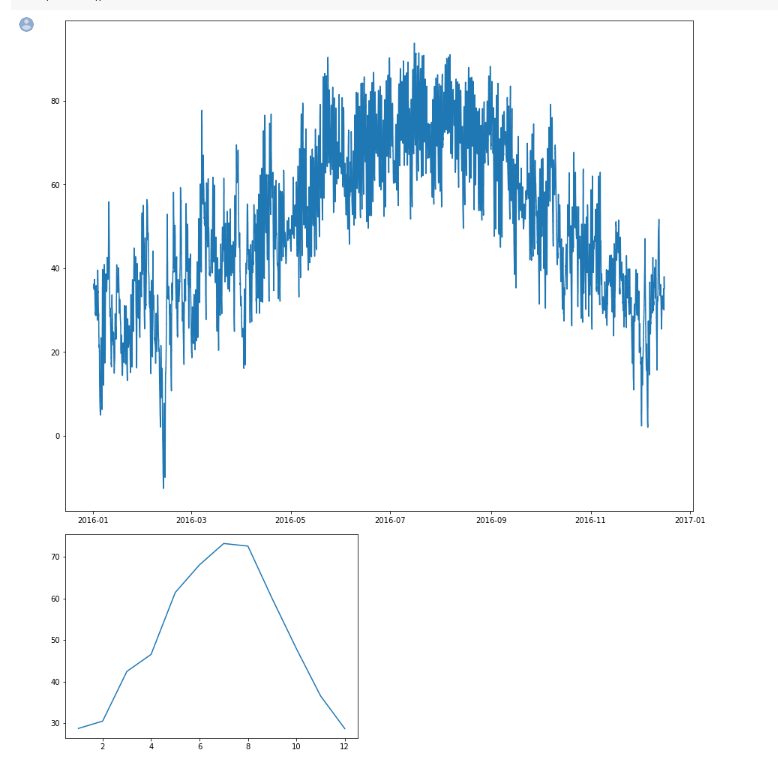
**Correlations Plot**



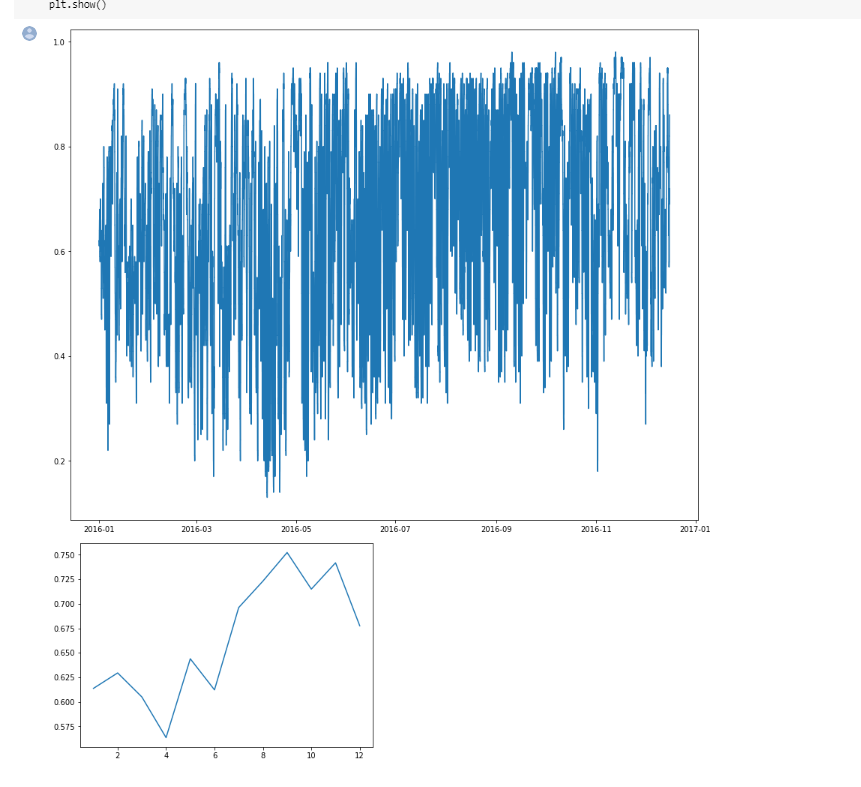
Total Energy Usage highly depends on appliances like Dishwasher, Furnace, Home Office, Fridge, Wine Cellar, Barn, Well, Microwave and Living Room.

Usage of Furnace, Fridge and Wine Cellar increases the Temperature

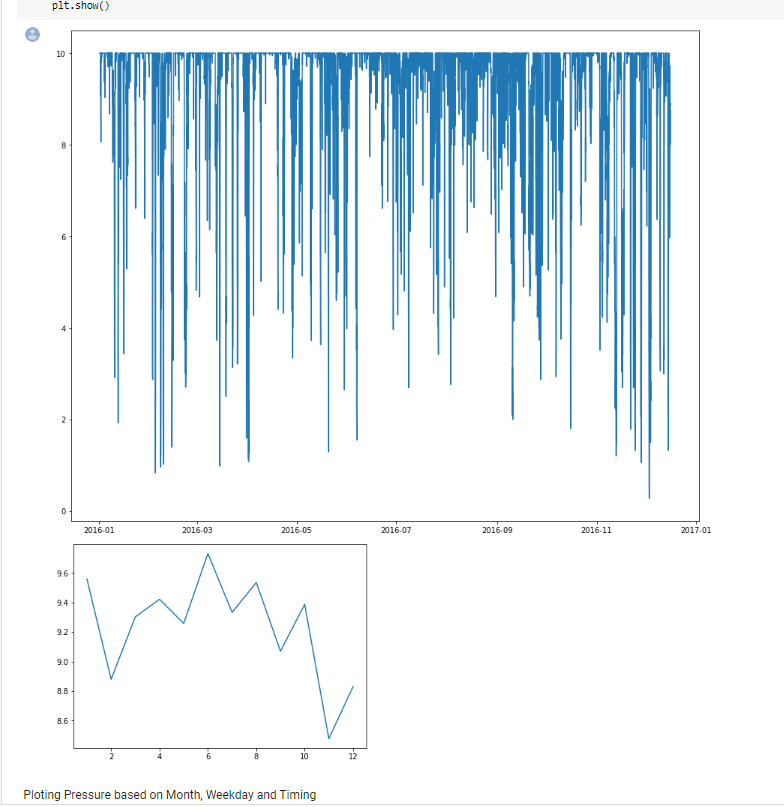
**Ploting Temperature based on Month, Weekday and Timing**



**Ploting Humidity based on Month, Weekday and Timing**



**Ploting Visibility based on Month, Weekday and Timing**

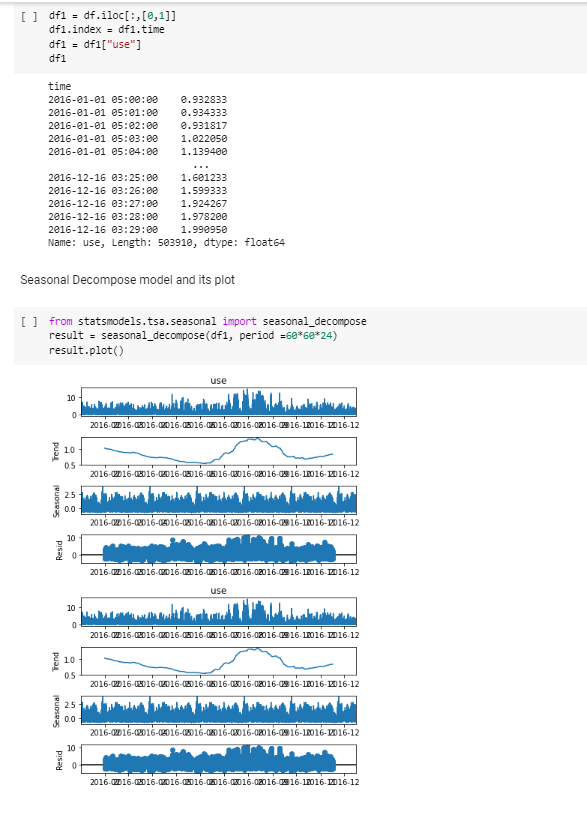


All the Weather Attributes follows the natural pattern like visibility is low in winter & nights, visibility is low when humidity is high etc

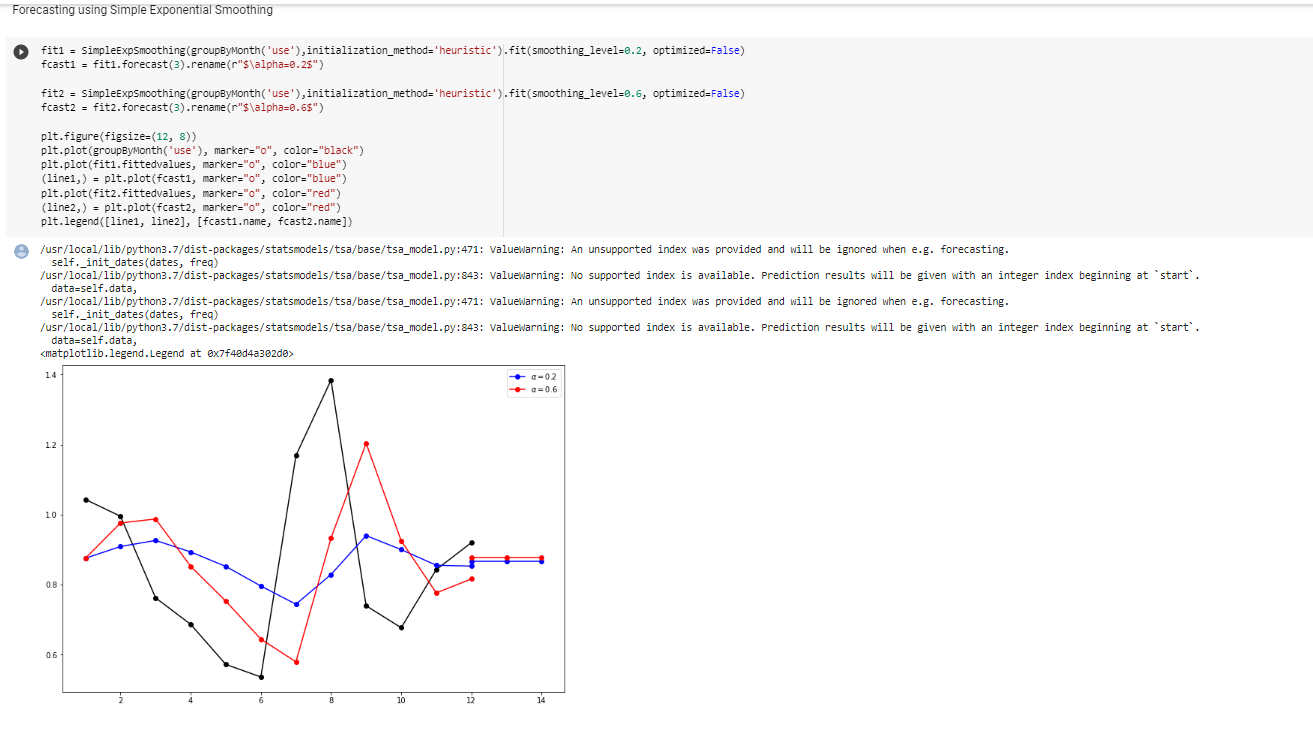
The only exception is the temperature which doesn't seem to obey the natural pattern

**Time Series**

**Separating Energy usage from the main dataframe with the time for time series analysis and Seasonal Decompose model and its plot**

****

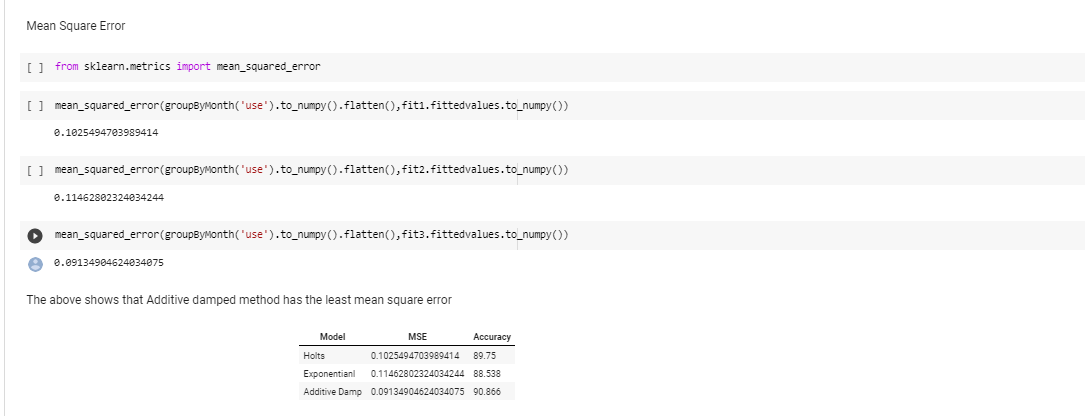
**Forecasting using Simple Exponential Smoothing**

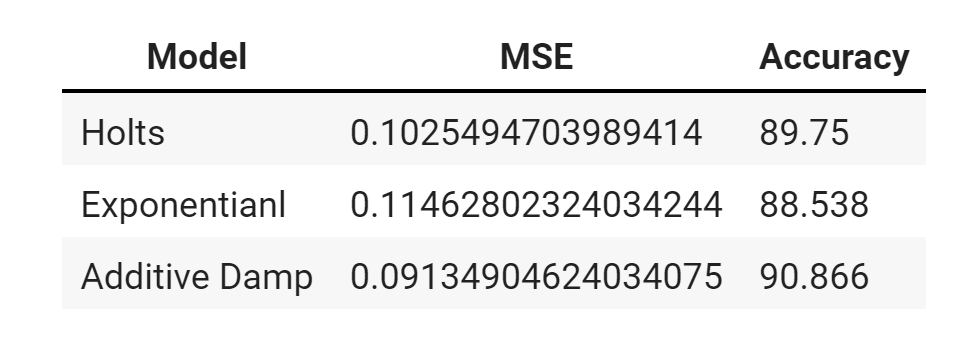
****

**Forecasting using Holts ,Exponential Smoothing and Additive damped model**

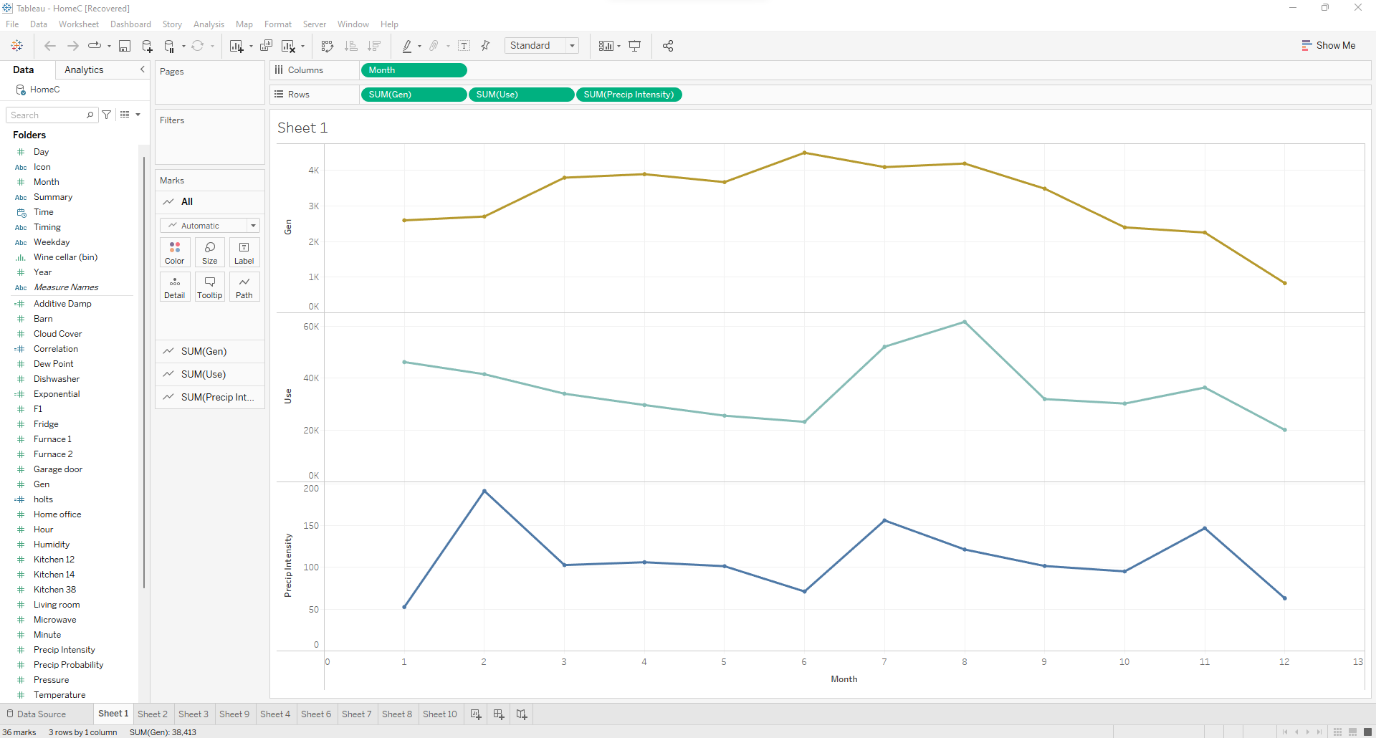
****

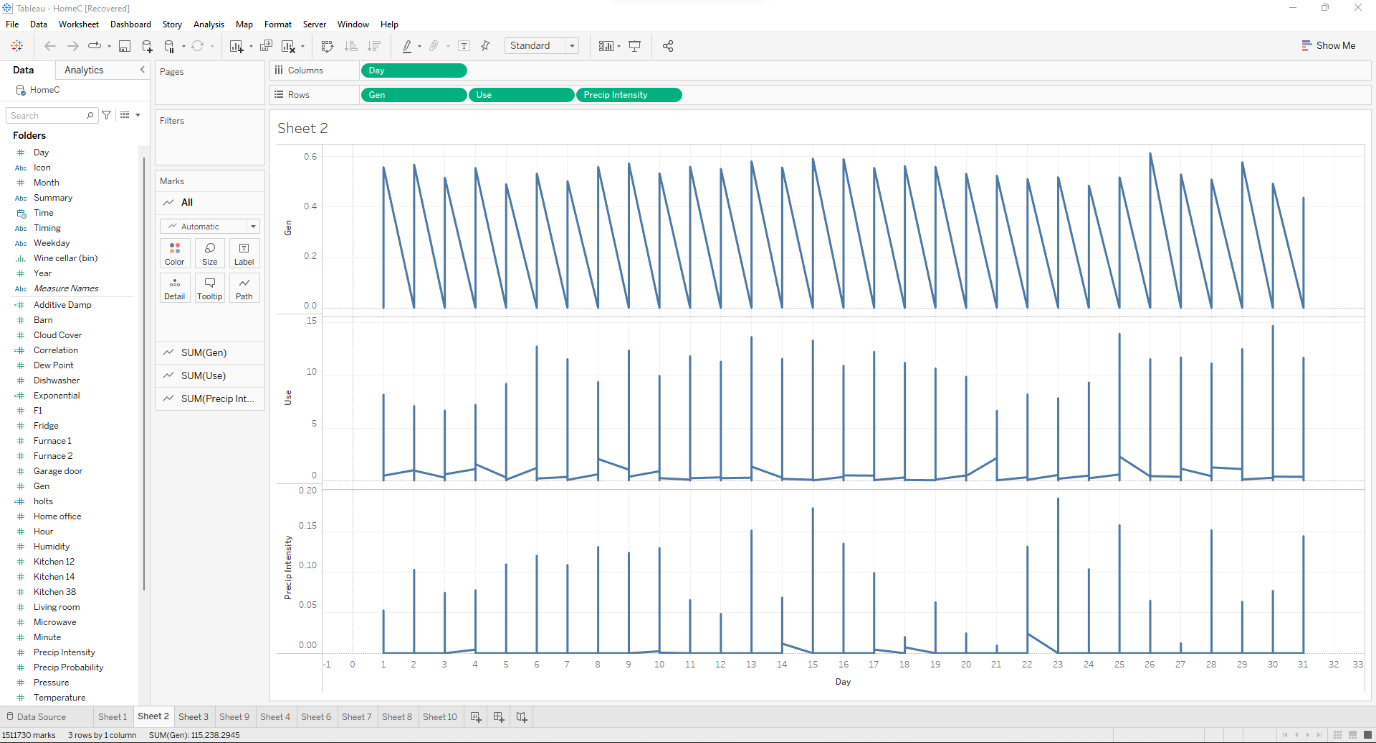
**Mean Square Error**

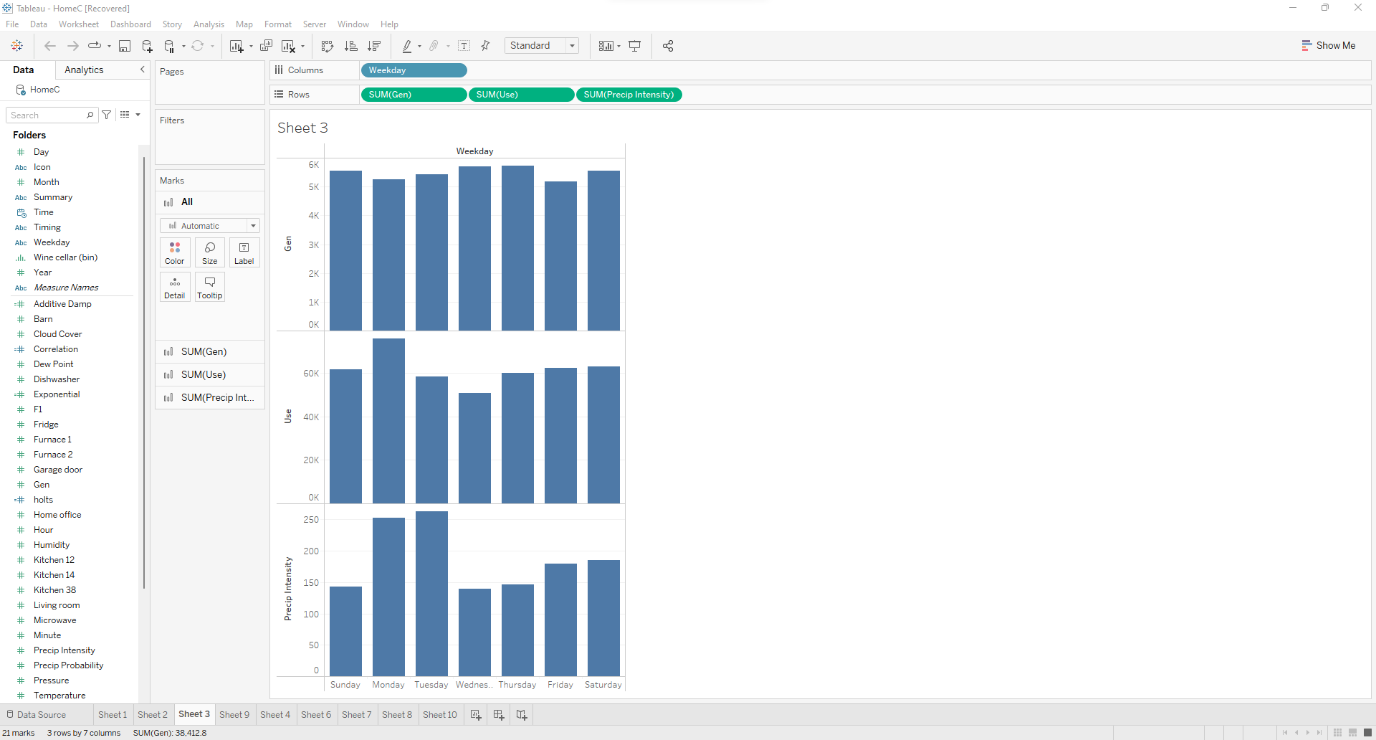
****

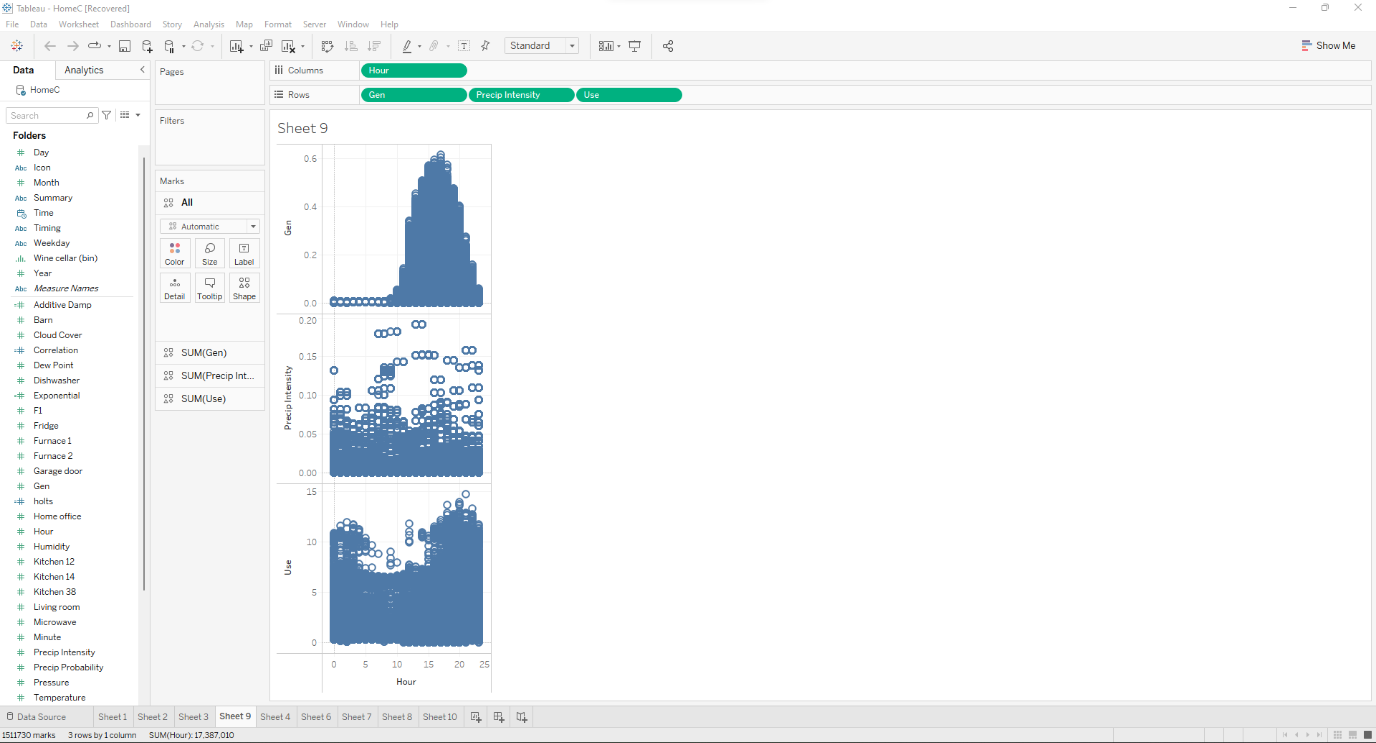
****

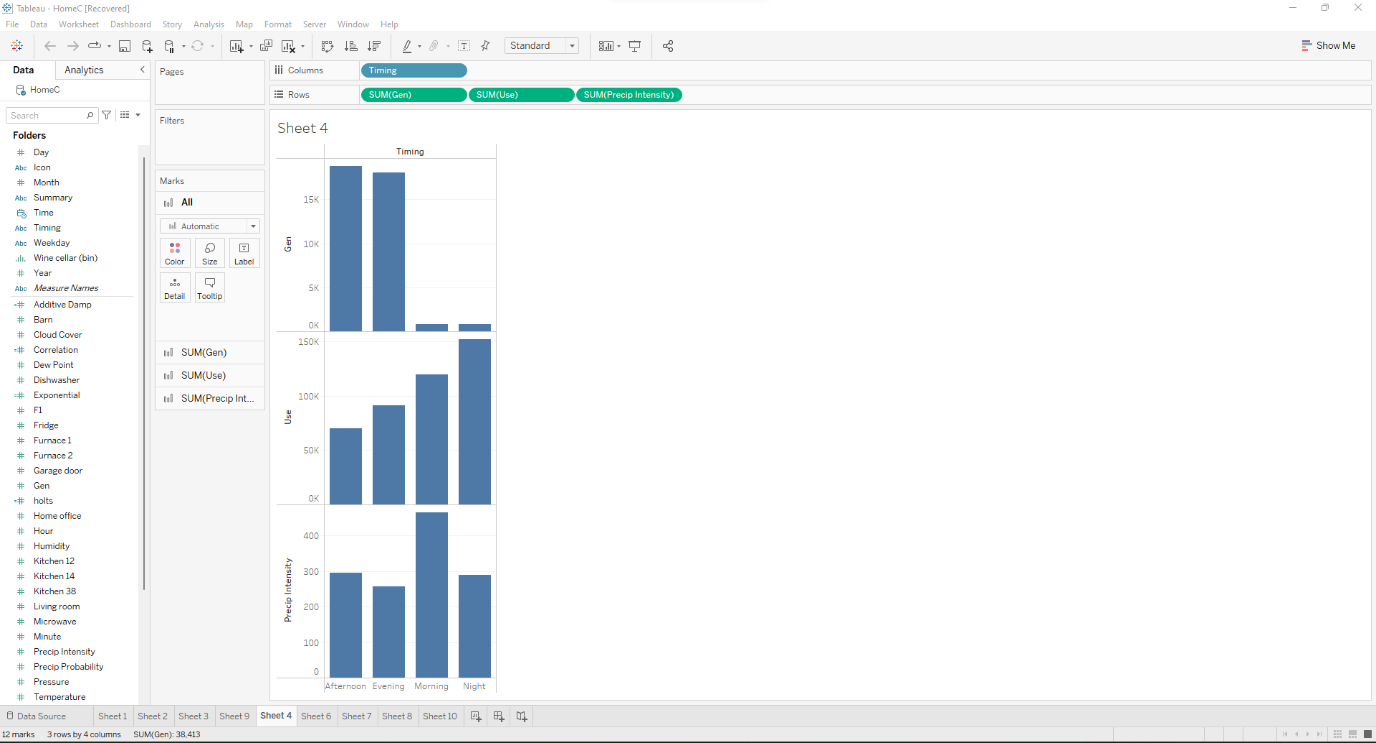
**Tableau Visualizations**

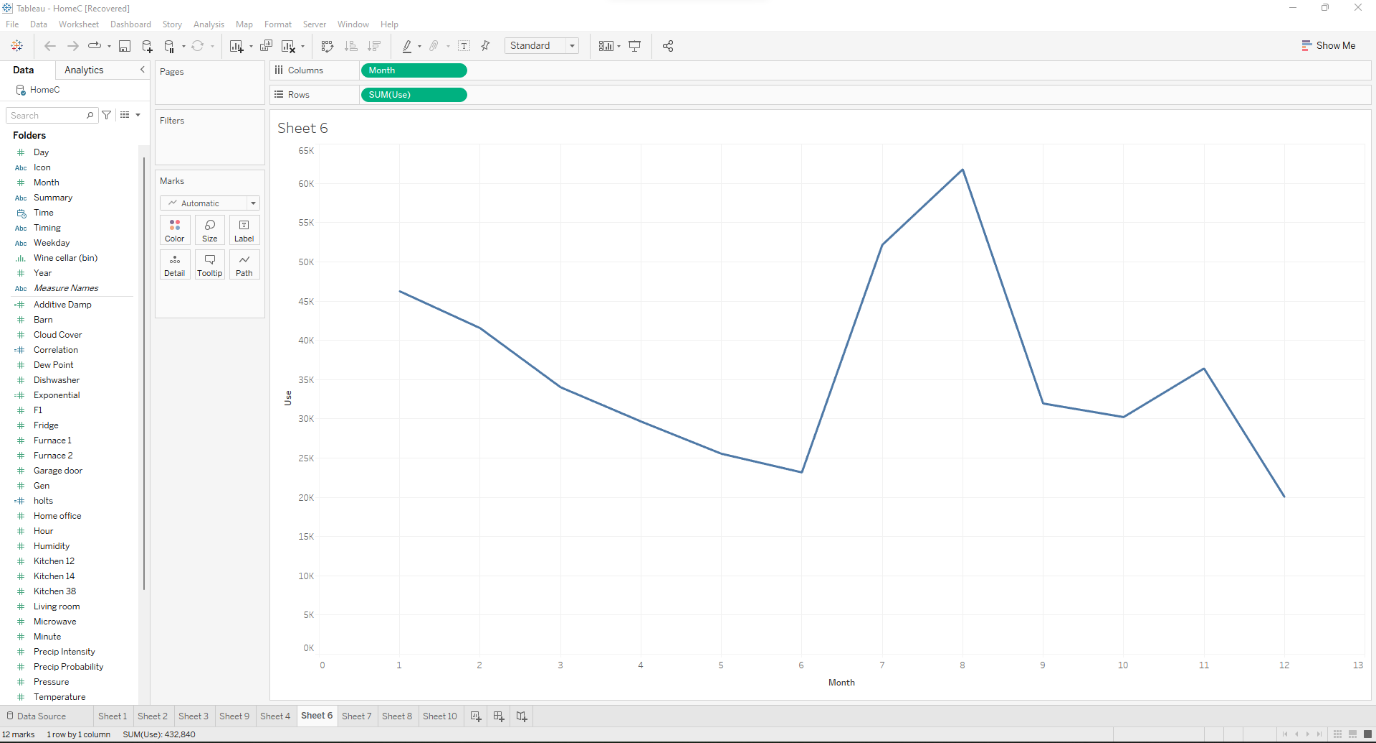


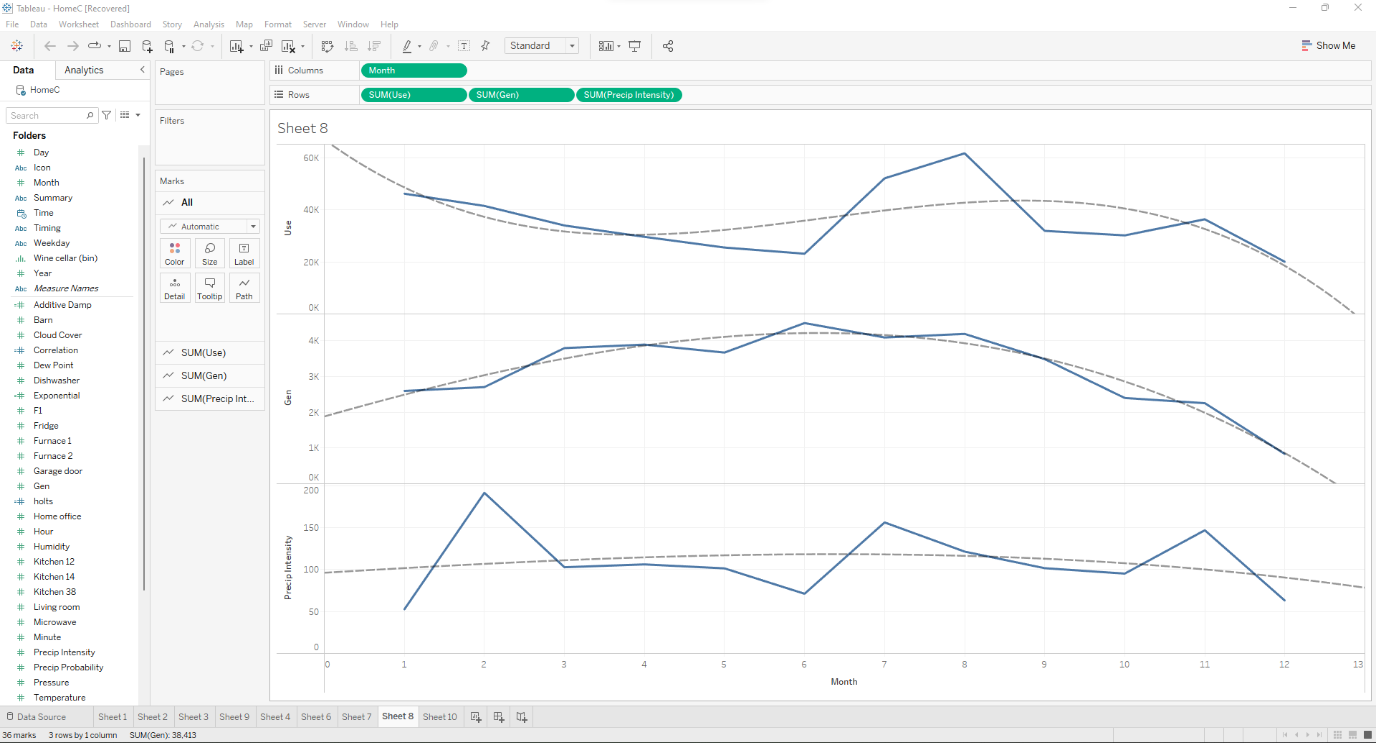












**CONCLUSION**

In the start we have insisted the demand and need of visualization and also the advantages of using a visualization to represent a data, that too while dealing with the smart home appliances data it has become furher more important and crucial as almost every minute a data is generated and we also encounter a lot of variations in the data and this makes data visualization highly important as reading or analysing data generated every minute is highly difficult as every minute a data is generated.

We have inferred a lot of patterns and flows from the visualizations made using the smart home appliance dataset. Considering the same inferences on a large scale we found that the energy consumption is so high which may lead to loss of many non-renewable resources, also the temperature data that we have analysed, when taken on a large scale is definitely a sign of global warming and the pattern of temperature fluctuation seems to be really alarming. Hence, we conclude saying that we should switch to an eco friendly pattern of life and monitor all our consumptions, using smart technologies and visualizations and use the resources in a controlled manner in order to reduce the chances of energy deficits and environmental threats.

**REFERENCES:**

[1] Solaimani, S., W. Keijzer-Broers, and H. Bouwman. “What We Do - and Don't - Know about the Smart Home: an Analysis of the Smart Home Literature.” Indoor and Built Environment 24.3 (2013): 370–383. Web.

[2] Hasan, Mehedi, et al. "Smart home systems: Overview and comparative analysis." 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). IEEE, 2018.

[3] M. Arikiez, F. Grasso, M. Zito. Heuristics for the cost-effective management of a temperature controlled environment. 2015 IEEE Innovative Smart Grid Technologies— Asia (ISGT ASIA), Bangkok, 2015: 1–6.

[4] F. Y. Melhem, O. Grunder, Z. Hammoudan, et al. Optimal residential load scheduling model in smart grid environment, 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Milan, 2017: 1–6.

[5] Molla, Tesfahun, Baseem Khan, and Pawan Singh. "A comprehensive analysis of smart home energy management system optimization techniques." Journal of Autonomous Intelligence 1.1 (2018): 15-21.

[6] L. Geng, Z. Lu. A Novel Model for Home Energy Management System Based on Internet of Things, 2016 IEEE International Conference on Power and Renewable Energy.

[7] F. Y. Melhem, O. Grunder, Z. Hammoudan, et al. Optimal residential load scheduling model in smart grid environment, 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Milan, 2017: 1–6.

[8] Yantai Huang, Lei Wang, Qidi Wu. A Hybrid PSO-DE Algorithm for Smart Home Energy Management. ICSI 2014, Part II, LNCS 8795, 2014: 292–300.