Capstone Project-2
Appliances Energy Prediction

Prepared by:

Suhail Shaikh

Contents

- Defining problem statement
- Reading the data set
- Data cleaning
- Dealing with outliers
- Exploratory data analysis
- Identifying correlation
- Model implementation
- Model evaluation
- Conclusion
- References

Problem Statement

- To predict Appliance energy consumption for a house based on factors like temperature, humidity & pressure .

- In order to achieve this, we need to develop a supervised learning model using regression algorithms.

Reading the Data Set

Data set has 19735 rows and 28 columns. The columns in data set have information as mentioned below :

- **date** : date in (yy-mm-dd) and time in (hr:min:sec)
- **Appliances** : energy use in Wh
- **lights** : energy use of light fixtures in the house, in Wh
- **T1** : Temperature in kitchen area, in Celsius
- **RH1** : Humidity in kitchen area, in %
- **T2** : Temperature in living room area, in Celsius
- **RH2** : Humidity in living room area, in
- **T3** : Temperature in laundry room area

%

Cont…

- **RH3** : Humidity in laundry room area, in %
- **T4** : Temperature in office room, in Celsius
- **RH4** : Humidity in office room, in %
- **T5** : Temperature in bathroom, in Celsius
- **RH5** : Humidity in bathroom, in %
- **T6** : Temperature outside the building (north side), in Celsius
- **RH6** : Humidity outside the building (north side), in %
- **T7** : Temperature in ironing room , in Celsius
- **RH7** : Humidity in ironing room, in %
- **T8** : Temperature in teenager room 2, in Celsius
- **RH8** : Humidity in teenager room 2, in %

Cont…

- **T9** : Temperature in parents room, in Celsius
- **RH9** : Humidity in parents room, in %
- **To** : Temperature outside (from Chievres weather station), in Celsius
- **Press_mm_hg** : (from Chievres weather station), in mm Hg
- **Rh_out** : Humidity outside (from Chievres weather station), in %
- **Windspeed** : (from Chievres weather station), in m/s
- **Visibility** : (from Chievres weather station), in km
- **Tdewpoint** : (from Chievres weather station), Â°C
- **Rv1** : Random variable 1, nondimensional
- **Rv2** : Random variable 2, nondimensional

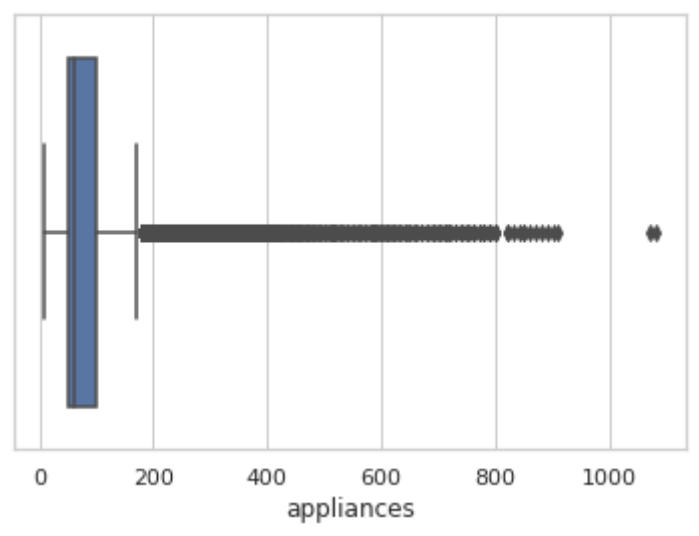
Data Cleaning

- Data set has no null values

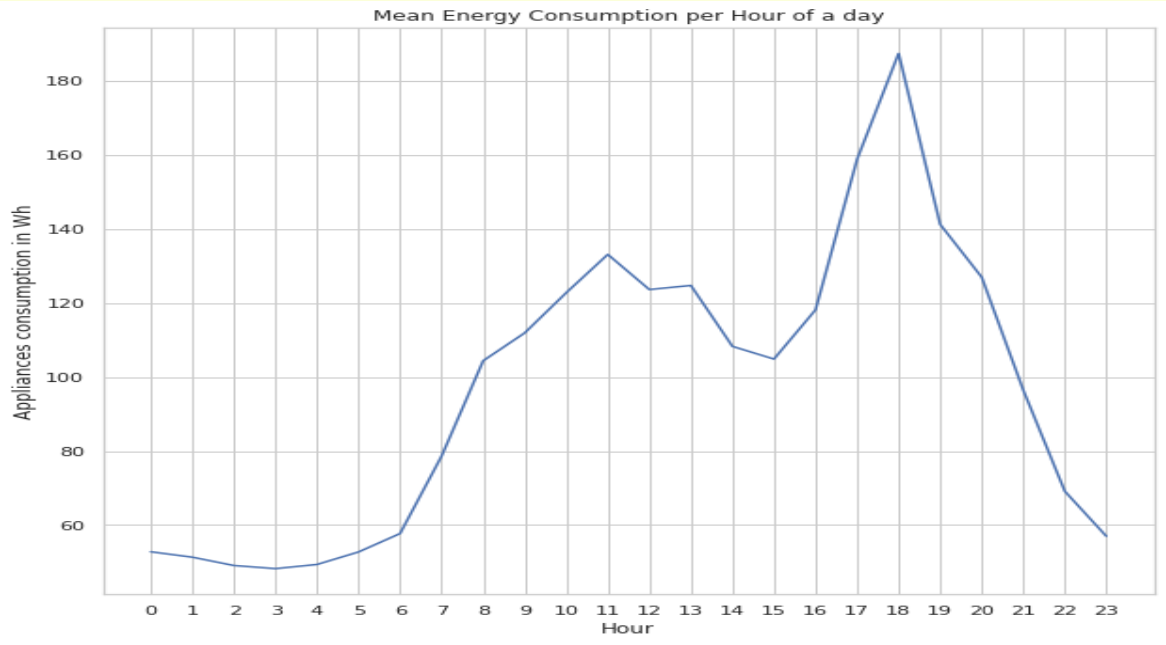
Column Name	Null Values	Column Name	Null Values
appliances	0	t7	0
lights	0	rh_7	0
t1	0	t8	0
rh_1	0	rh_8	0
t2	0	t9	0
rh_2	0	rh_9	0
t3	0	t_out	0
rh_3	0	press_mm_hg	0
t4	0	rh_out	0
rh_4	0	windspeed	0
t5	0	visibility	0
rh_5	0	tdewpoint	0
t6	0	rv1	0
rh_6	0	rv2	0

Dealing with Outliers

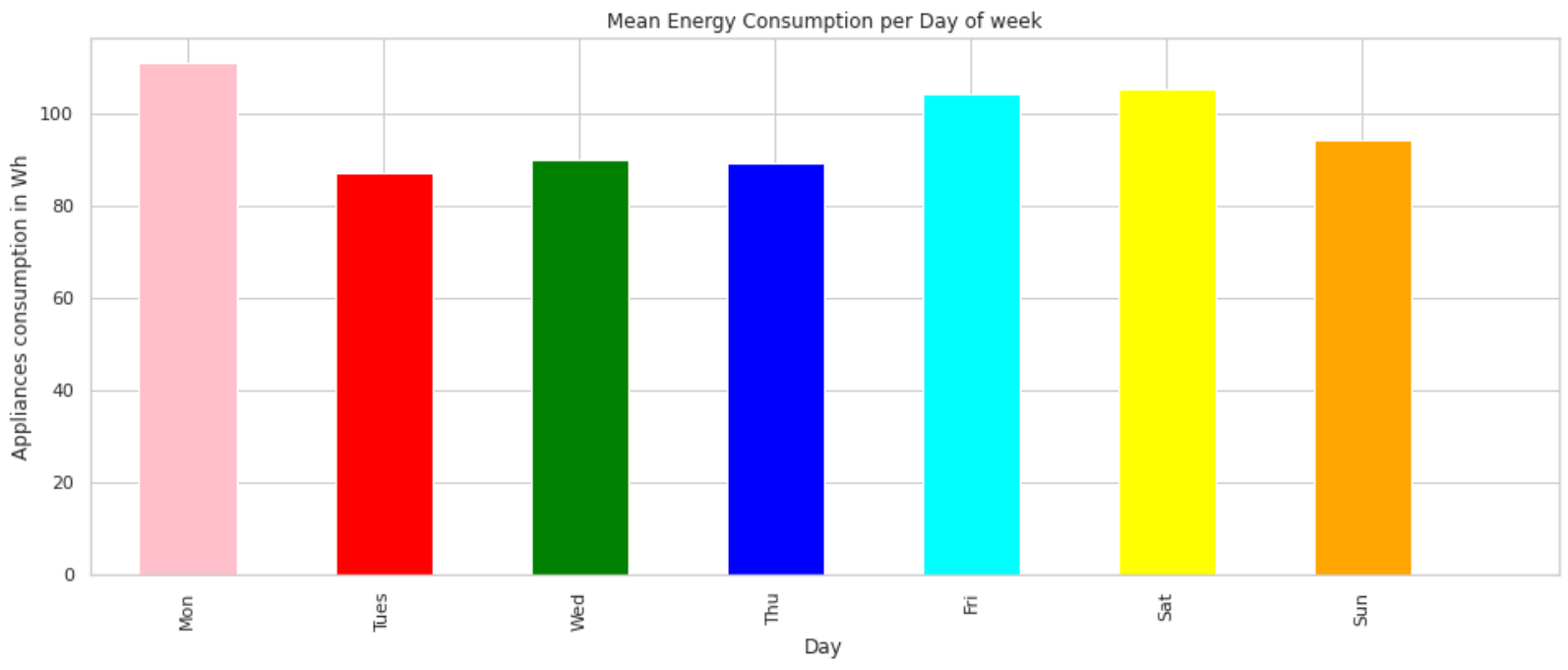
- Energy consumption of appliances cannot be negative and appliances having high energy consumption (greater than 790 Wh) are removed from data set.



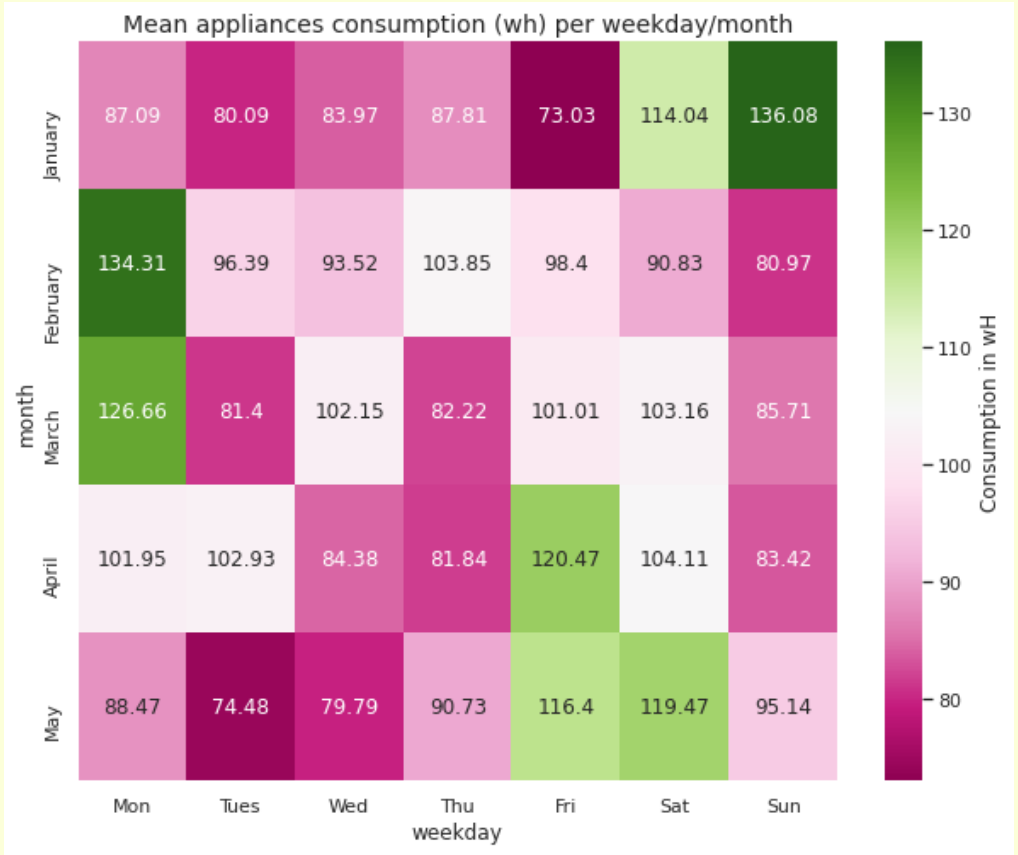
Exploratory Data Analysis



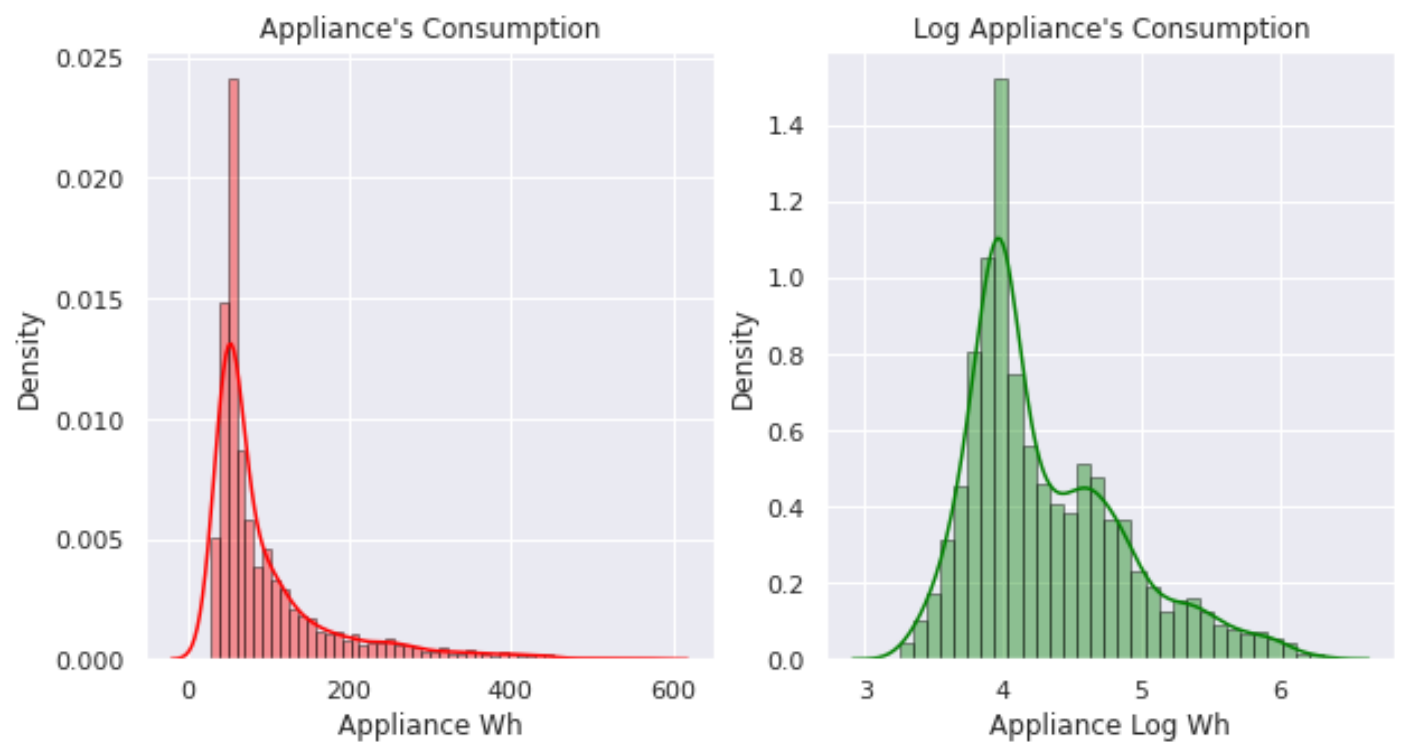
- **Energy consumption is higher in evening from 5pm to 7pm.**



- **Energy consumption is higher on start of week i.e. Monday**

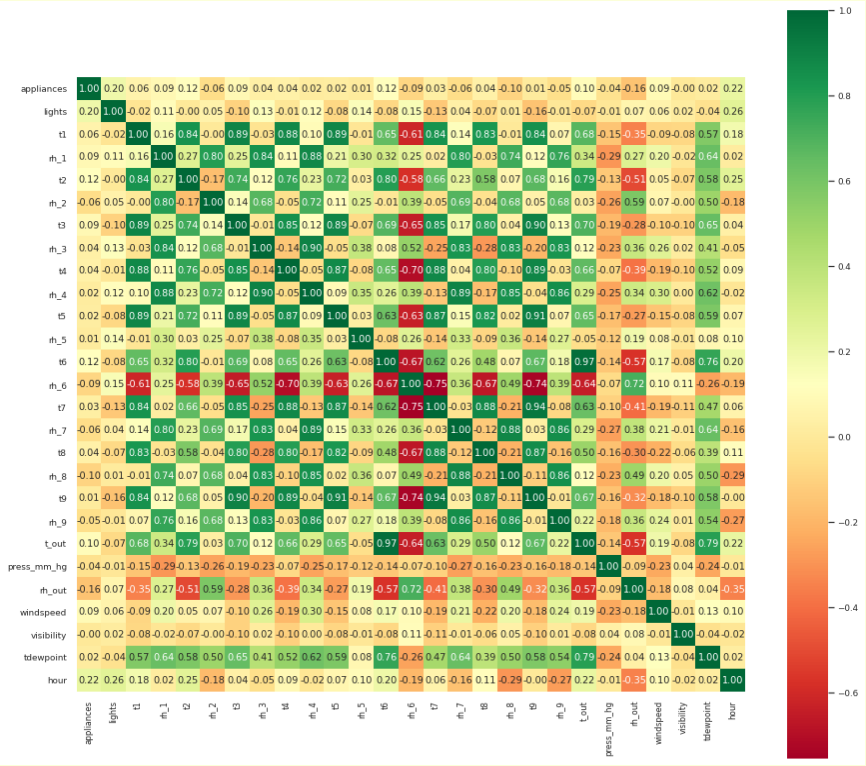


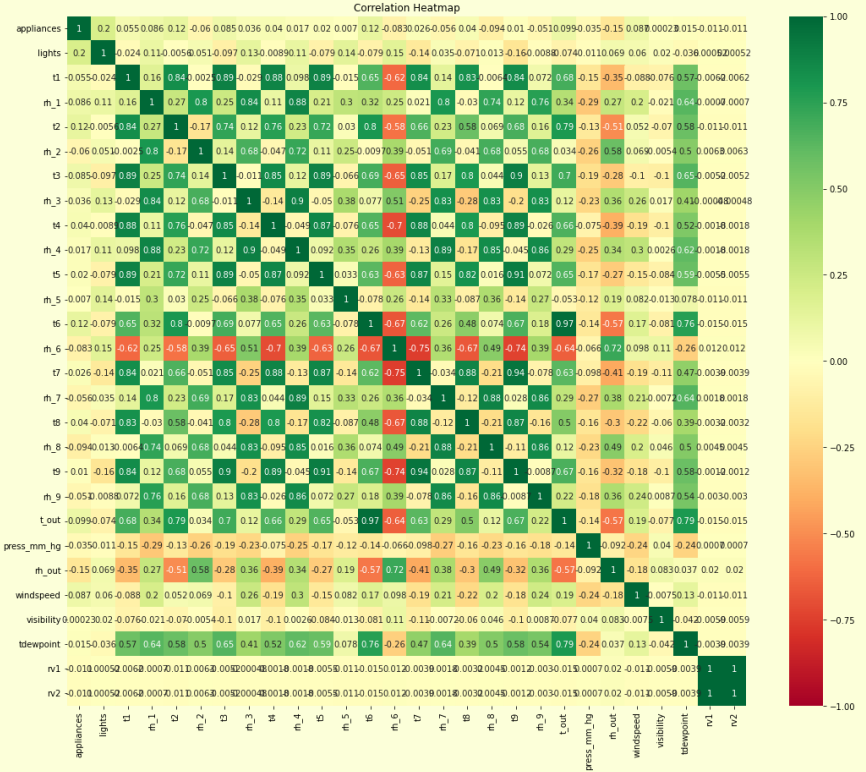
- **Energy consumption was higher on Sunday in month of January. In other months it was high on Monday.**



- **In order to increase accuracy of model and make data more balanced we have applied log to Appliance column.**

Identifying Correlation





Model Implementation

- Model Implementation

```
from sklearn import linear_model
lin_model = linear_model.LinearRegression()
lin_model.fit(X_train,y_train)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

- **We have Implemented Linear Regression Model to data.**

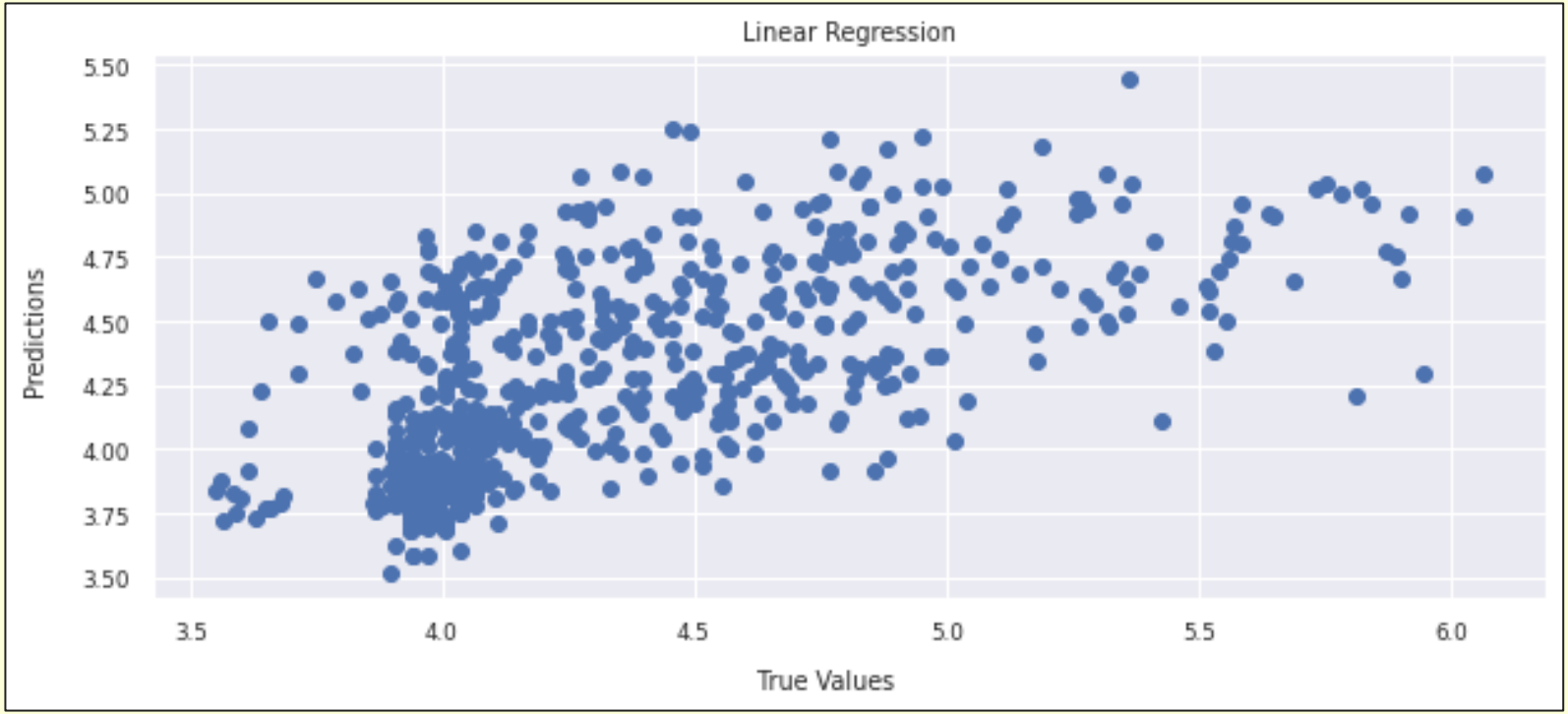
Model Evaluation

```
evaluate(lin_model,X_test,y_test)
```

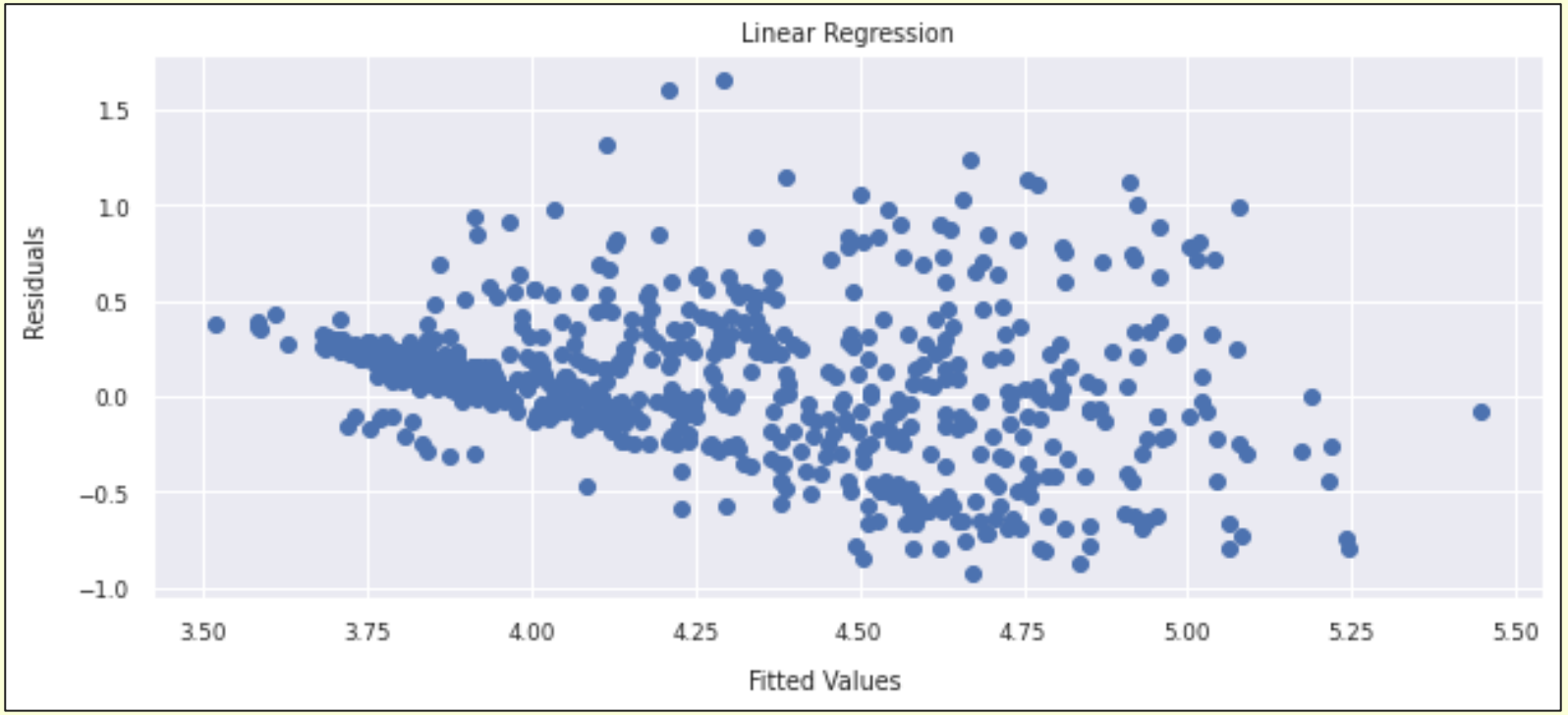
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

Average Error          :0.2998 degrees
Variance Score         :32.30%
Accuracy               :93.31%
```

- Linear Regression model has Accuracy of 93.31%

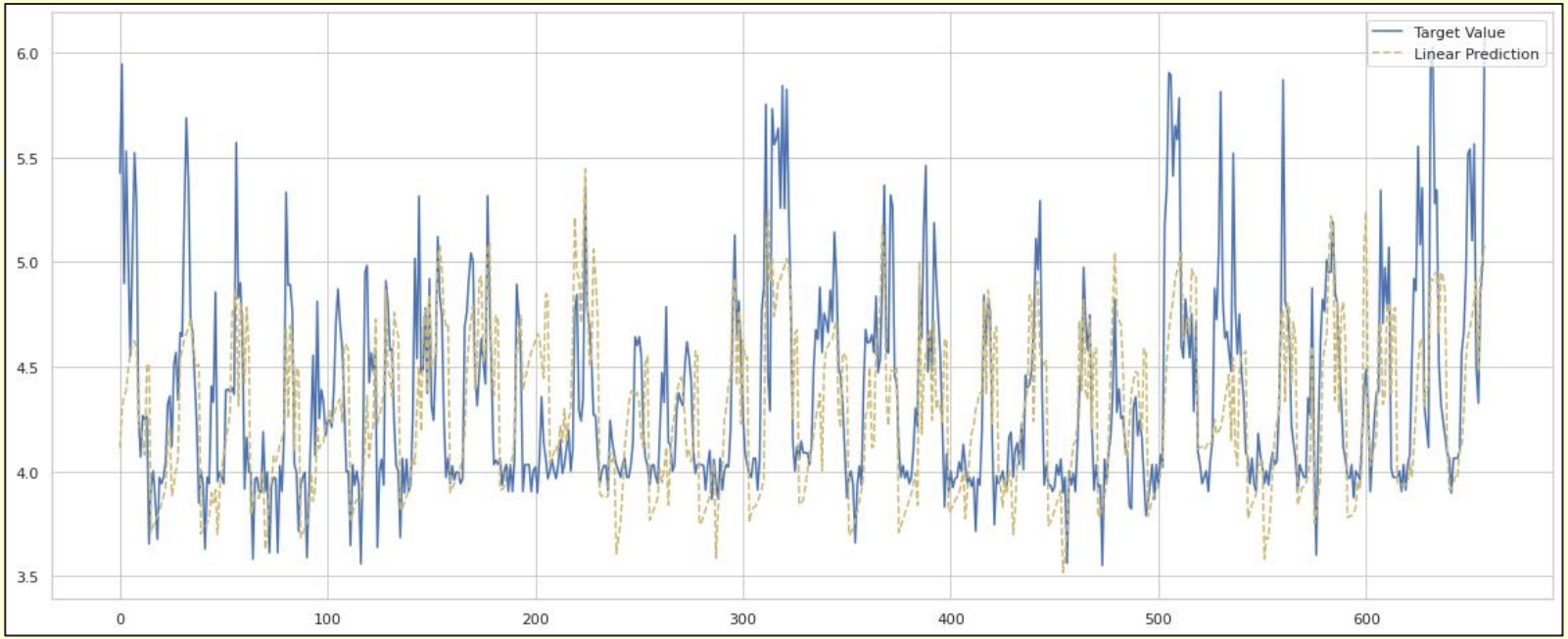


❑ **Model Predictions Vs True Values.**



❑ **Residuals Vs Fitted Values.**

Conclusion



- Model has Accuracy of 93.31 % for appliances energy prediction.

Conclusion

- We have implemented Linear Regression Model with Accuracy of 93.31% .
- From EDA we can conclude that Energy Consumption is higher in evening from 5 pm to 7 pm.
- Energy consumption is higher on Monday in every week.
- In month of January energy consumption is higher on Sunday and in other month it is higher on Monday.

References

- Kaggle
- Youtube
- Github
- Towards data science
- Analytics Vidya
- Code basics
- Stack over flow

Thank You