**ADVANCES IN DATA SCIENCE**

**Energy Data Set**

**Team 8**

**REPORT BY:**

**Komal Ambekar**

**Rishi Rajani**

**Sayali Borse**

OBJECTIVE

The report summarizes the design and implementation of the data wrangling performed on the Appliances Energy Consumption data set. This report is divided into 5 sections.

Section 1: Review of papers

Section 2: Exploratory Data Analysis

Section 3: Feature engineering and Feature Selection

Section 4: Prediction algorithms

Section 5: Model Validation and Selection

Section 6: Final pipeline

SECTION 1: REVIEW OF PAPERS

We were given three papers for review and analysis. Each document had to offer different features. The three papers were:

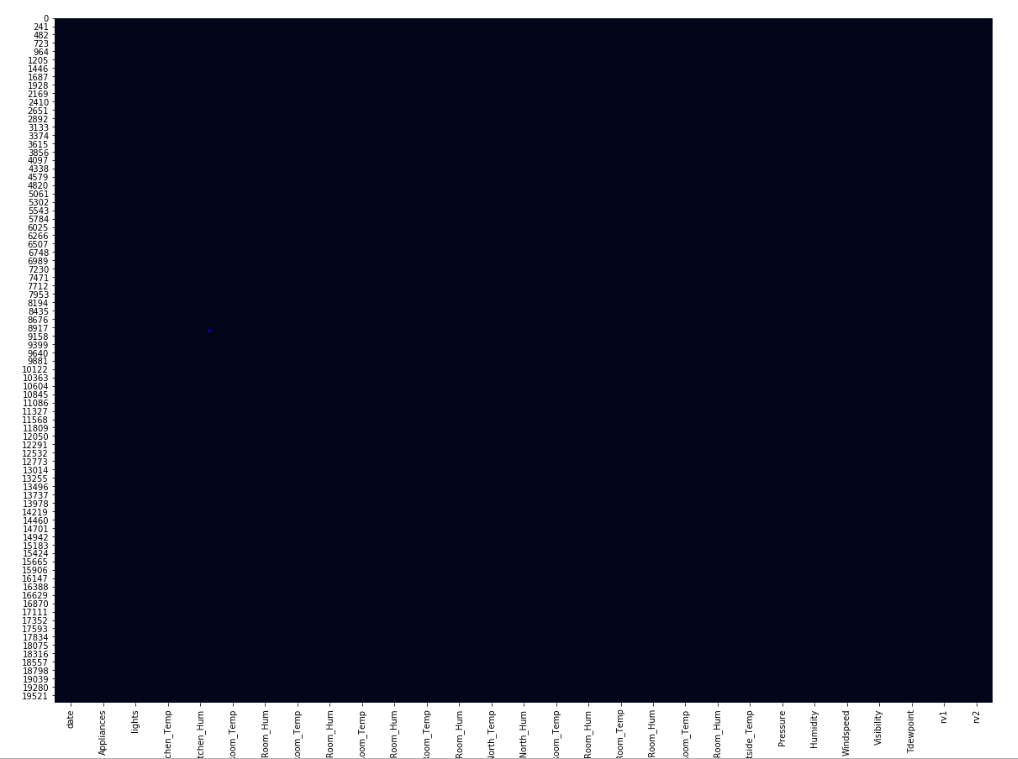
1. Data driven prediction models of energy use of appliances in a low-energy house
2. A review of artificial intelligence-based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models
3. Prediction of appliances energy use in smart homes

SECTION 2: EXPLORATORY DATA ANALYSIS

As interest in IOT and sensors pick up steam, companies are trying to build algorithms and systems to understand consumer behavior to help them make better decisions. One such application is energy modeling. Though, most consumers are aware of their aggregate consumption of energy, few are aware of how and where energy is consumed. With increasing sensors in equipment, it is becoming easier to find out which equipment/instruments consume the most power. The energy (Wh) data logged every 10 min for the appliances is the focus of this analysis. The 10 min reporting interval was chosen to be able to capture quick changes in energy consumption. Data used include measurements of temperature and humidity sensors from a wireless network, weather from a nearby airport station and recorded energy use of lighting fixtures. The wire-less sensor network’s temperature and humidity recordings were averaged for the corresponding 10 min periods and merged with the energy data set by date and time. The time span of the data set is 137 days (4.5 months). Fig. 1 shows the energy consumption profile for the period. The energy consumption profile shows a high variability. Fig. 2 shows a histogram of the data.

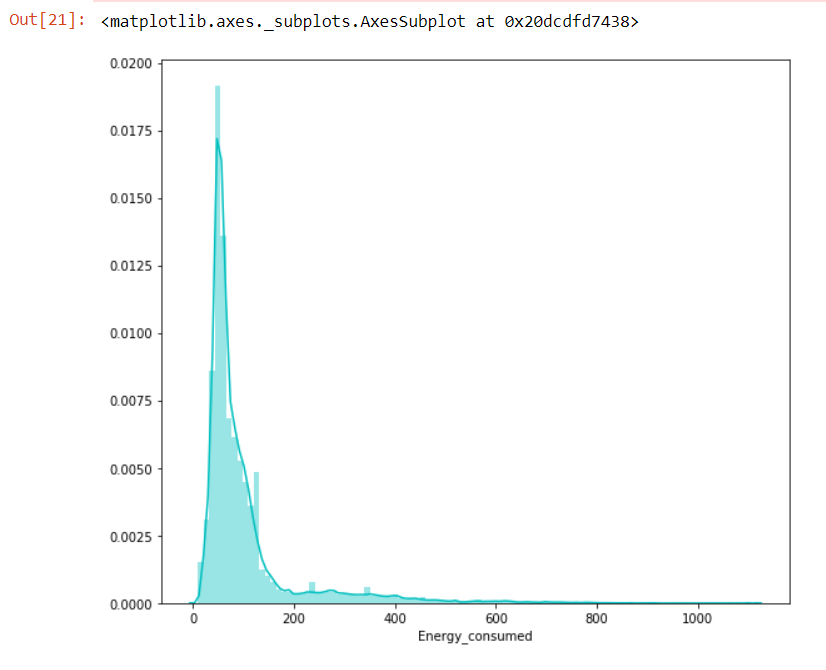
* To check all null values in the dataset:



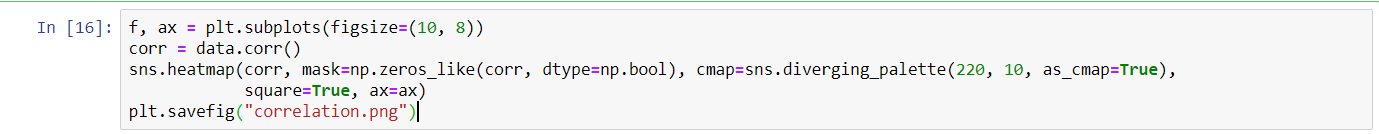


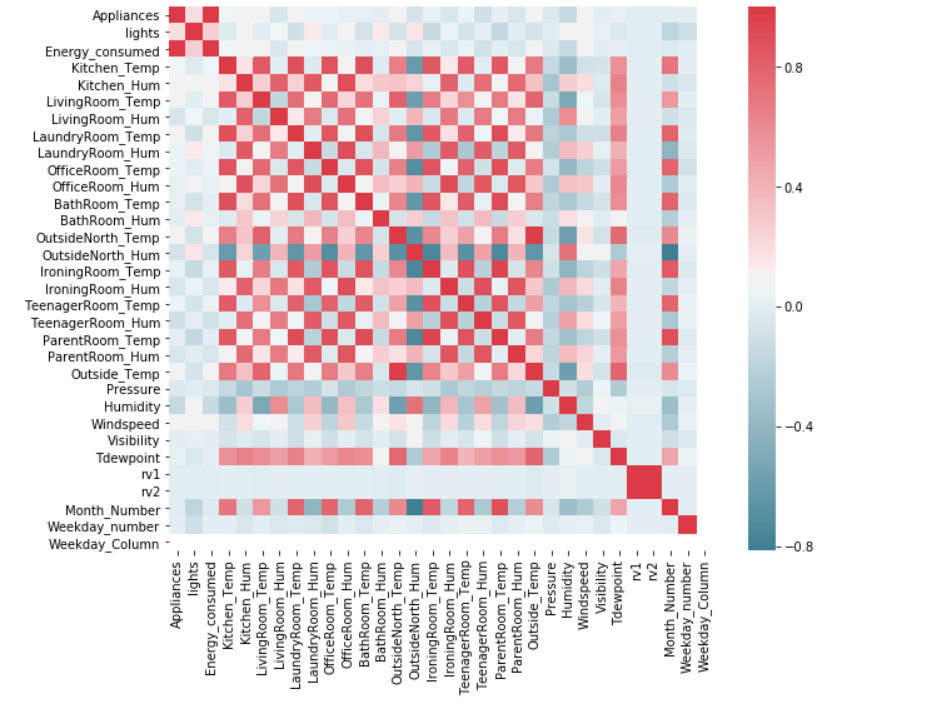
* Energy Consumed:



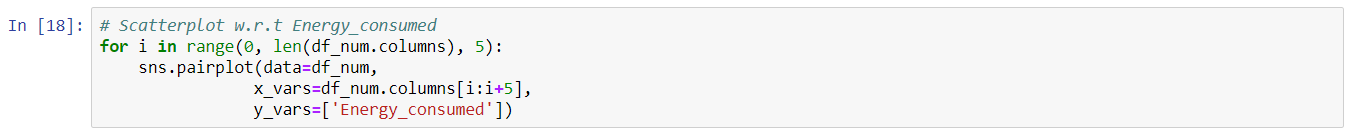


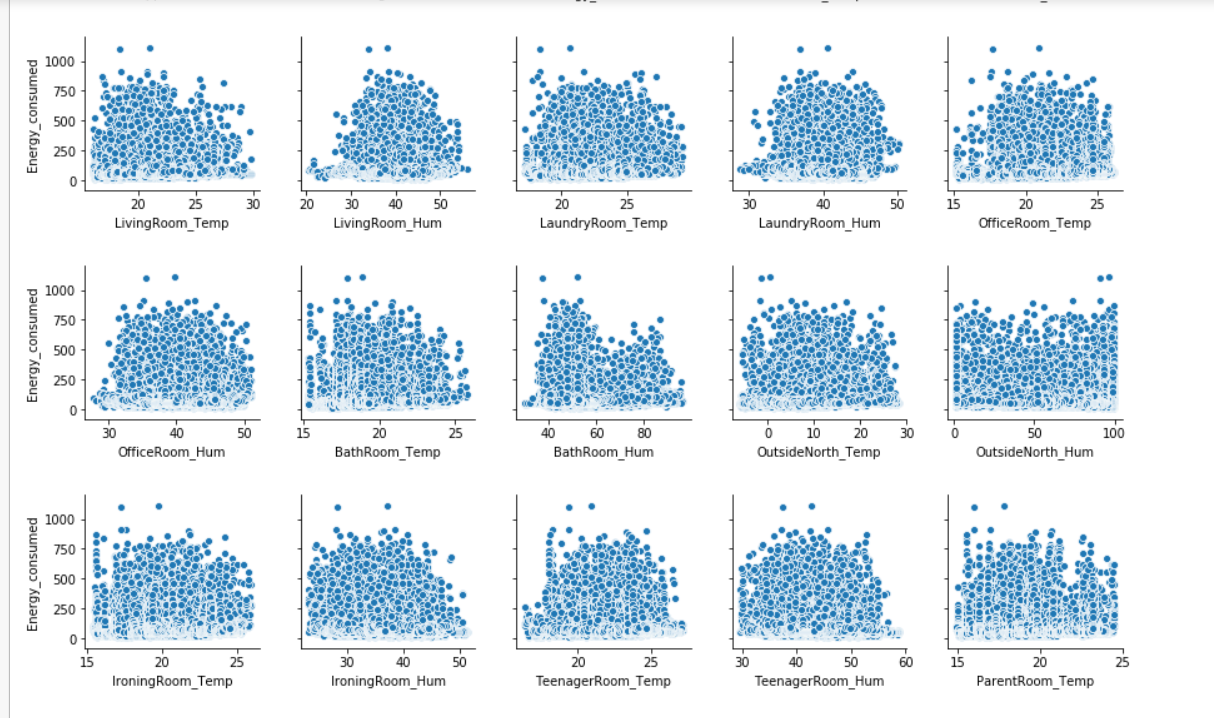
* Correlation Matrix



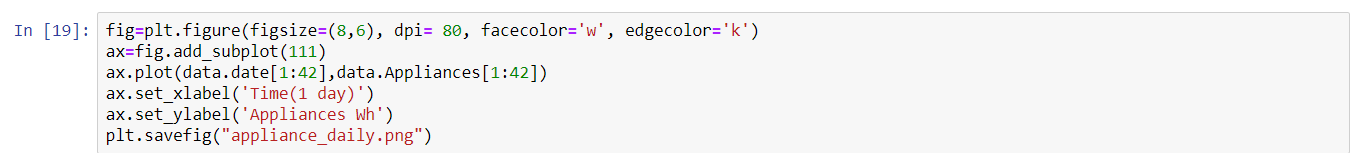


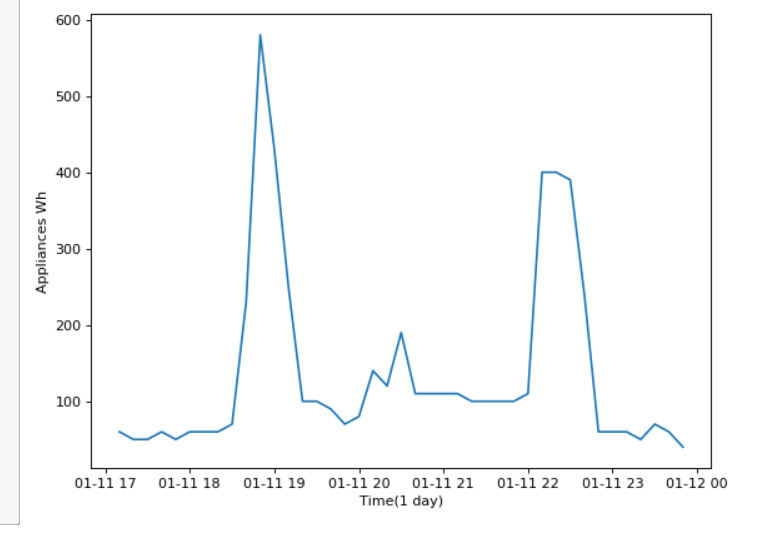
* Scatter plot wrt Energy Consumed



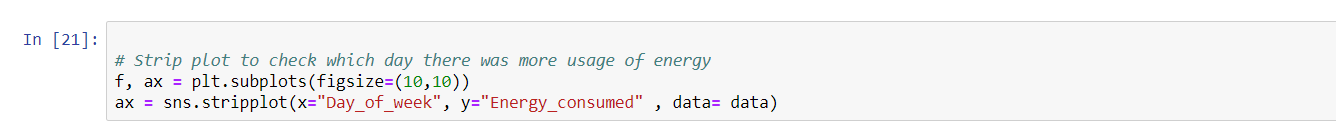


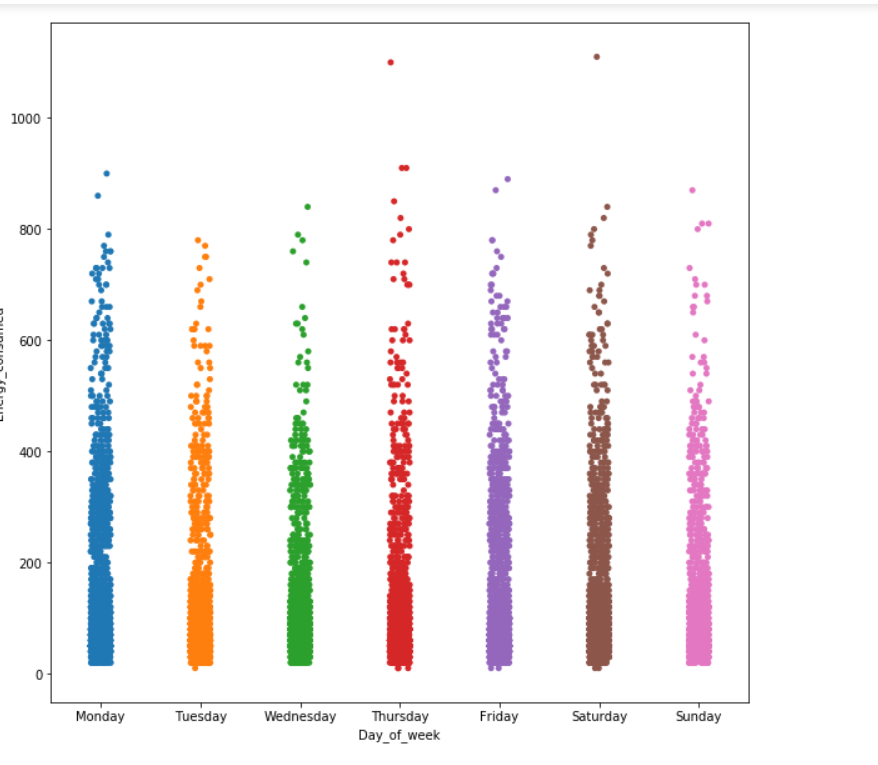
* Daily Energy Consumption





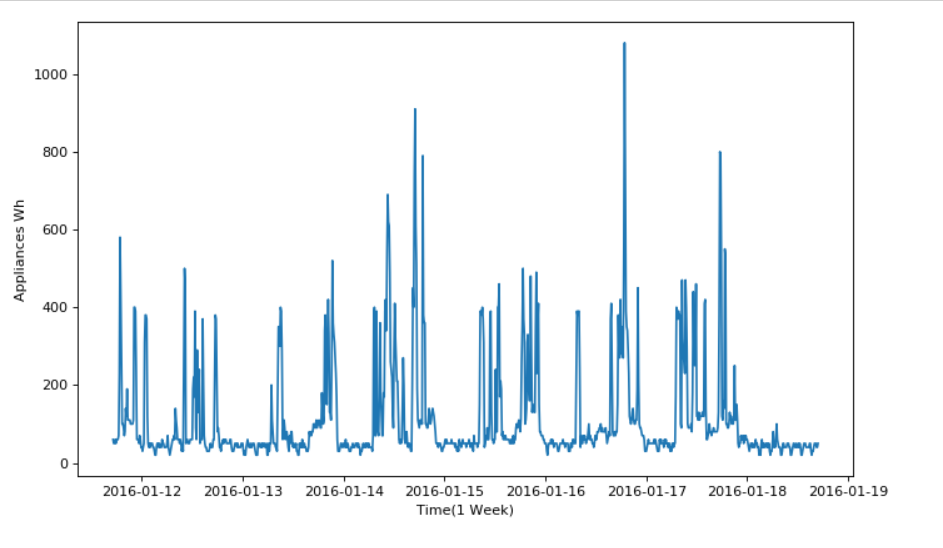
* Weekday wise energy consumption





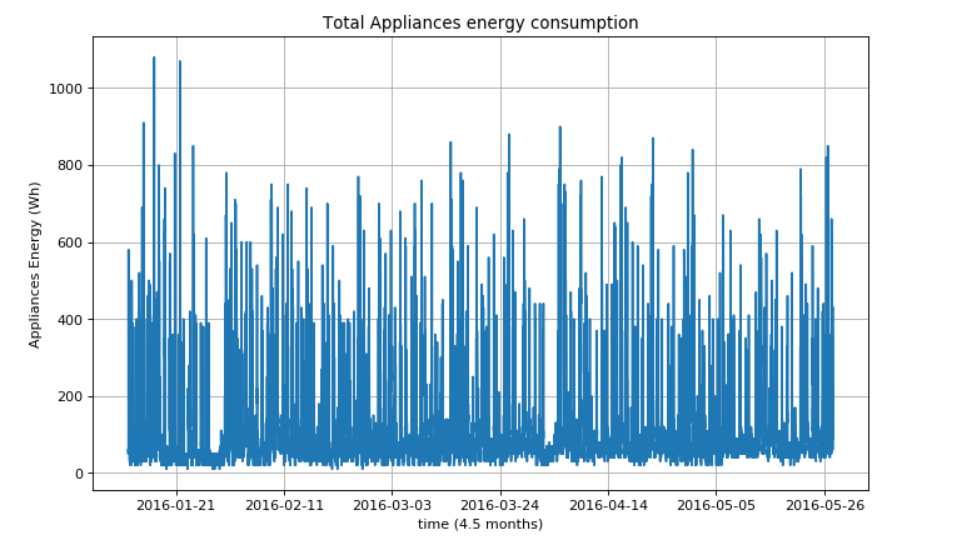
* Weekwise Energy Consumption



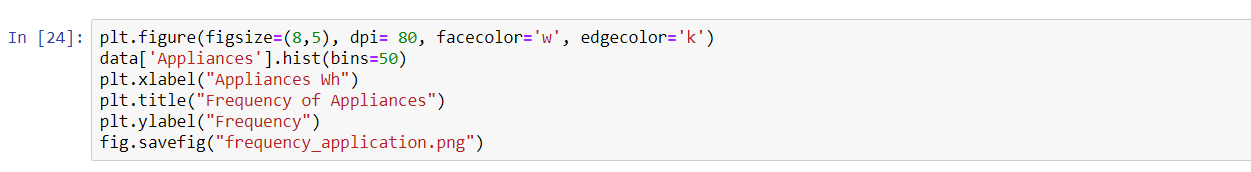


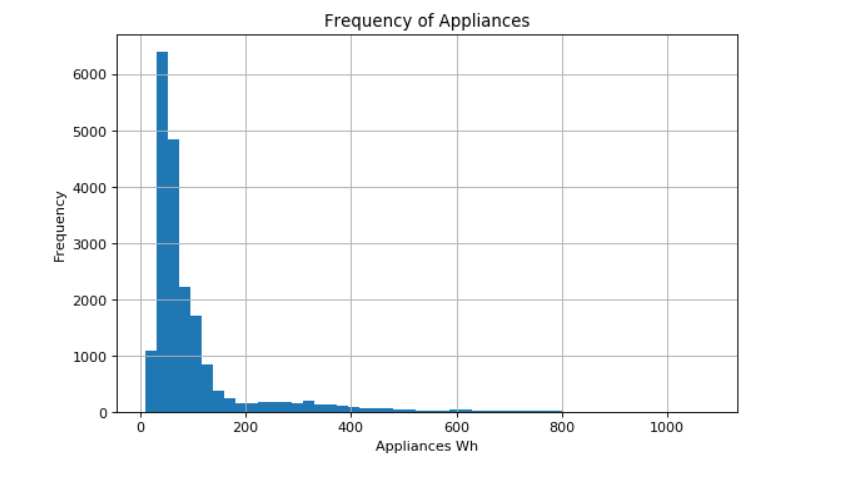
* Total Appliances energy consumption





* Frequency of Appliances





* Plotting Histogram of Numerical Features

