Data Visualization of DCMS live data on energy consumption

Author: Minjie Xu

Content

Abstract

The data set is Department for Culture, Media and Sport live data on energy consumption. The main body includes three parts: data preprocessing, basic charts based on descriptive statistics and tables on inferential statistics. The tool of data visualization used here is python with libraries of seaborn, plotly, matplotlib.

Introduction

The dataset is the energy consumption data for the DCMS headquarters building at 100 Parliament Street. It's generated in real-time from data taken every 5 seconds from the on-site meters. Getting these energy data out of some buildings is harder than others, but in general, the buildings contain a small low-power computer that takes very frequent readings from the electricity meters and stores the data.

The dataset consists of two different CSV files: electricity and heat. Energy use is measured here in kilowatt-hours (kWh), which are the standard units of a home energy bill (1kWh is the amount of electricity used by ten 100W light bulbs in one hour). For electricity, this number represents the amount of energy that flows into a building through the meter and excludes distribution losses. For district heating, it reflects a flow of temperature into the building over time (after the heat produced by burning the fuel has been transported to the meter, which involves other losses). Each of these numbers, while all being measured in kWh, mean very different things.

Data Preprocessing

At the beginning of the CSV files does give the note: empty cells representing half-hourly consumption mean that there is no data available for that time and numbers followed by an 'E' are unreliable numbers. These kinds of data are not helpful for data analysis.

Firstly, rows with empty cells and invalid cells [Fig. 1] should be removed.

ility	Unit	Date	00:00	00:30	01:00	01:30	02:00	...	19:30	20:00	20:30	21:00	21:30	22:00	22:30	23:00	23:30	Total
city	kWh	2011-02-02	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0E
city	kWh	2011-02-03	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0E
city	kWh	2011-02-04	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0E
city	kWh	2011-02-05	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0E
city	kWh	2011-02-06	0.0	0.0	0.0	0.0	0.0	...	75.0	80.0	75.0	65.0	75.0	65.0	70.0	70.0	70.0	25710E

Fig.1

Secondly, some outliers which affect data visualization should be removed. For example, after removing the maximum outliers in the district heating dataset, the violin-plot gives a clearer visualization with the average and other important descriptive statistic information[fig.2].

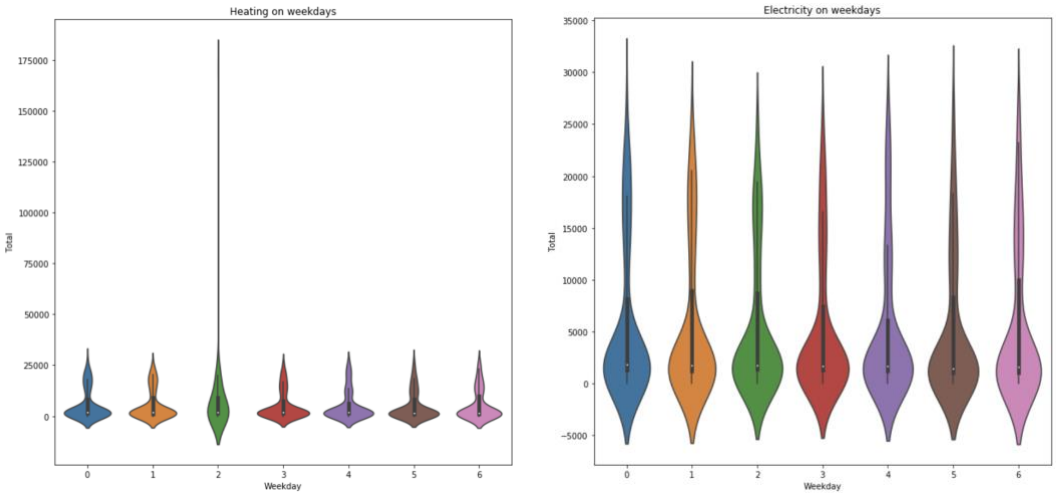


Fig.2

Basic charts

Overall trend: Line charts with animate buttons

I drew three charts controlled by three animate buttons which depict different trends of electricity and heating consumption over the period from 2010 to 2015. Three animate buttons not only create interaction with users but also help to display different data sets separately. (“Effectiveness of Animation in Trend Visualization,” ten years later, 2020)
The first chart compares electricity consumption with heating consumption using red and green lines. Interestingly, although they have fully different trends, they get similar averages. [Fig.3]

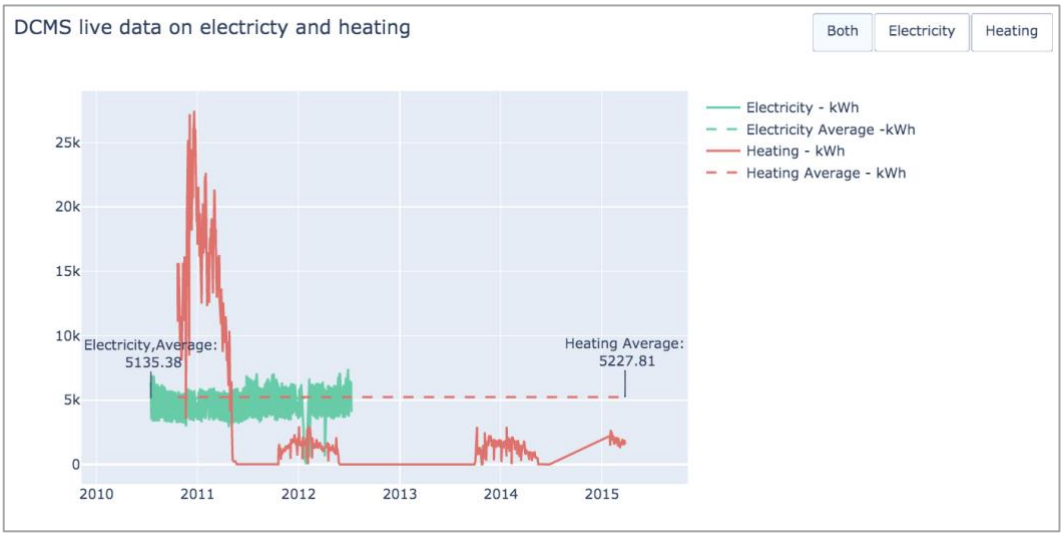


Fig.3

The line chart of electricity consumption gives the information of the average (5135.38 kWh) using the breakpoint line, the maximum value (7415 kWh),

and the minimum value (10 kWh). We can also find that it fluctuates steadily over the period so that we can try to find reasons for fluctuation later.[Fig.4]

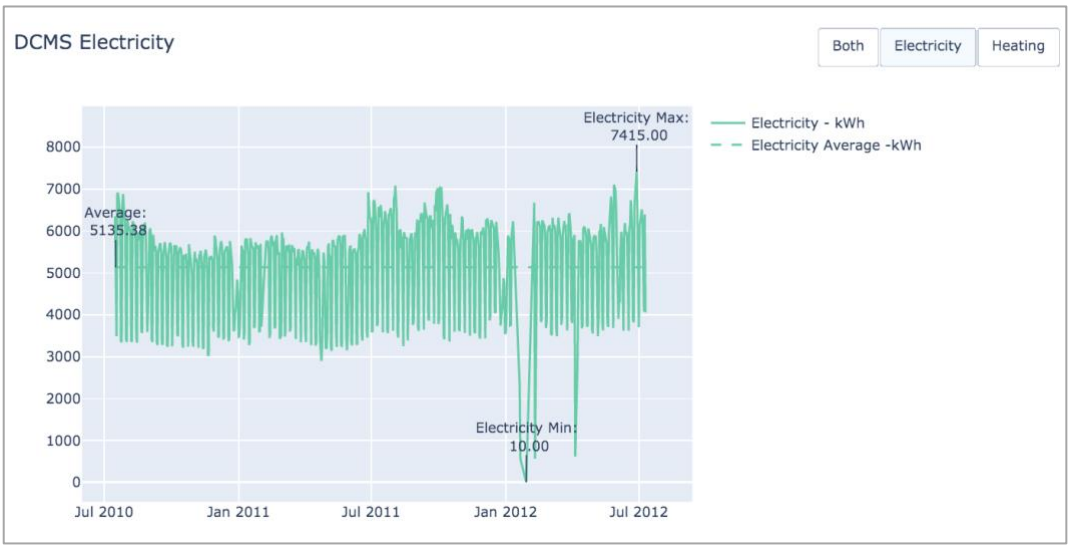


Fig.4

The third chart shows the trend of heating consumption using the red line. The average line split the line into two parts. Before mid-2011, the consumption of heating fluctuates dramatically over 10k kWh. However, it declined sharply after mid-2011. [Fig.5] The efforts of HMRC's finding no-cost and low-cost ways to reduce our carbon emissions paid off.

These ways include voltage optimization; temperature adjustments similar to 100 Parliament Street and replacing halogen lights with low energy LED alternatives. Besides, the building receives heat via the Whitehall district heating system, which serves other Government HQ Sites locally. The site has a mixture of naturally ventilated and mechanically cooled offices and rooms. (100 Parliament Street, London - CarbonCulture, 2020)

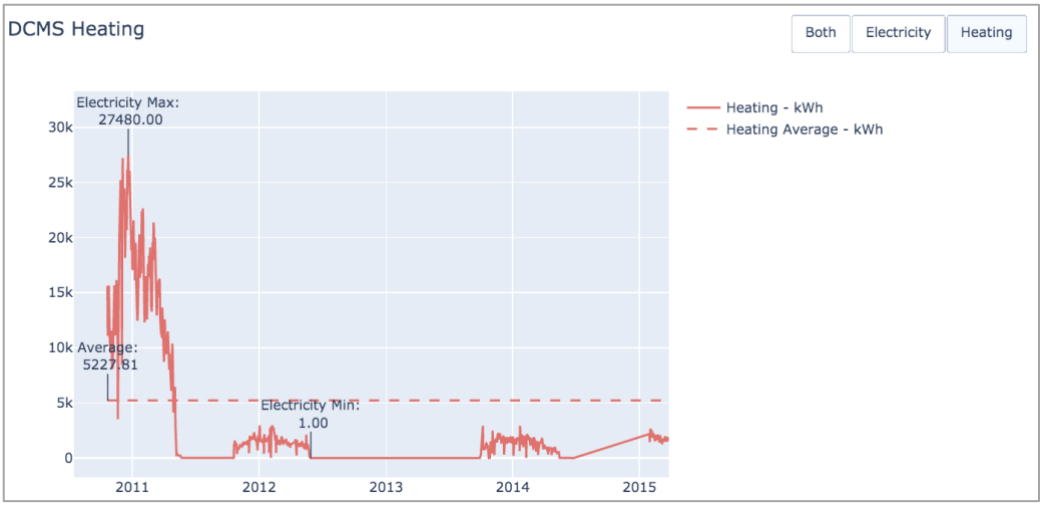
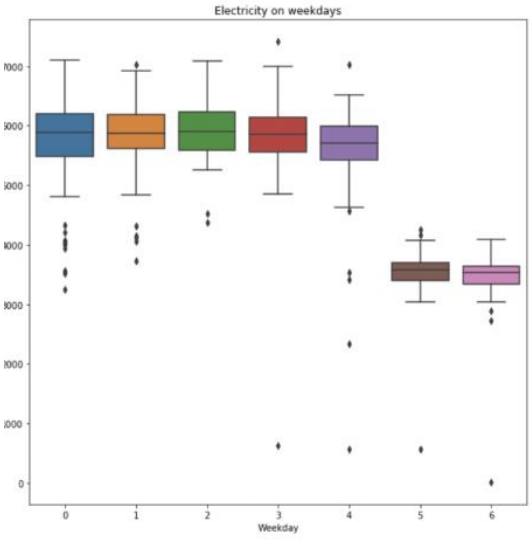


Fig.5

Descriptive Statistics based on weekdays: Box plots

To find the reason for fluctuation in electricity consumption, I drew the box plot of electricity consumption dataset grouped by weekdays. As the box plot shows, the electricity consumption of working days gets a similar average (about 6000 kWh) while that of weekends get a smaller average (about 3500 kWh).[Fig.6]



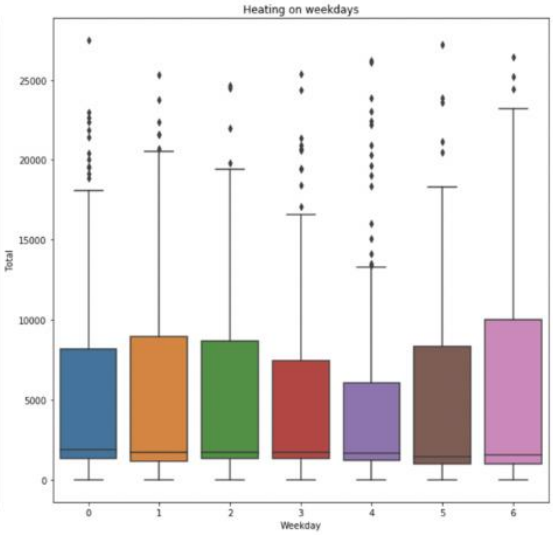


Fig.6 Fig.7

However, the box plot of heating consumption does not have the same conclusion. The trend is not related to weekdays. [Fig.7]

Descriptive Statistics based on months and years: Mixed Plots

Fig.8

The electricity consumption dataset is grouped by months whose trend is shown in the left subplot. [Fig.8] The line chart fluctuates dramatically with 5711 kWh

as the maximum value and 4247 kWh as the minimum value. The right subset shows the box plot of electricity consumption every year. Obviously, the average electricity consumption increased continuously. The largest average of electricity consumption over three years is nearly 6000 kWh which is in the year 2012.

According to the mixed plot of heating consumption [Fig.9], we still get the conclusion that the heating consumption decreased sharply after mid-2012. The largest consumption appeared in December 2010 with 22.45k kWh.

Fig.9

Inferential statistics

Based on the results of descriptive statistics, I proposed the hypotheses:

H0: μ1=μ2=μ3; (Year does not affect electricity consumption)
∂1=∂2=….=∂10=∂11=∂12; (Month does not affect electricity consumption)
ß1=ß2=ß3=…=ß6=ß7; (Weekday does not affect electricity consumption)
H1: μ(s) are not equal; ß(s)are not equal; ∂(s) are not equal;

Firstly, according to the central limit theorem, they all follow a normal distribution. The result of the homogeneity test of variances(p-Value:0.33) shows that groups based on “Year” have the same variance. [Fig.10] Then we test groups base on “Month” and “Weekday” respectively. The p-Value of the

former is 0.9(≥0.05) while the p-Value of the latter is ≤0.05.

```
leveneResult=scipy.stats.levene(y1,y2,y3)
p=leveneResult[1]
if p<0.05:
    print("variances of groups are not equal")
else:
    print("variances of groups are equal, p-Value:",p)
```

```
variances of groups are equal, p-Value: 0.3330070158076228
```

Fig.10

Finally, we test the variance of two factors: Year and Month. The table result of ANOVA [Fig.11] shows that they reject H0. Thus, "Year" and "Month" are the influenced factors of electricity consumption.

```
df = pd.DataFrame(elec_new, columns=['Total', 'Year','Month'])
formula = 'Total ~ C(Year)  +C(Month)+C(Year):C(Month)'
model = ols(formula, df).fit()
aov_table = anova_lm(model, typ=2)
print(aov_table)
```

	sum_sq	df	F	PR(>F)
C(Year)	1.657434e+08	2.0	59.076826	5.481429e-14
C(Month)	6.826239e+08	11.0	44.238431	1.533135e-45
C(Year):C(Month)	4.384348e+07	22.0	1.420670	1.197391e-01
Residual	9.300421e+08	663.0	NaN	NaN

Fig.11

Conclusion and evaluation

The visualization shows clearly different trends of heating and electricity consumption. Also, it considers factors of different trends and the interaction with users. However, more aesthetic factors should be presented to attract the attention of the audience instead of simple lines and graphics, such as line charts and box plots. In addition, the visualization ignores analyzing and visualizing the outliers of datasets, especially the maximum value of heating consumption which is removed in the step of data preprocessing.

Reference

Platform.carbonculture.net. 2020. *100 Parliament Street, London* – Carbon culture. [online] Available at: <https://platform.carbonculture.net/places/100-parliament-street/8928/>.

Medium. 2020. "Effectiveness of Animation In Trend Visualization," Ten Years Later. [online] Available at: <https://medium.com/@FisherDanyel/effectiveness-of-animation-in-trend-visualization-ten-years-later-e2f52b433526>.