

ОТЧЕТ

по командной игре "Умная аналитика для умного дома"

Команда: "К8" (г. Нижний Новгород)

Участники: Жирков А.Р.

Список и количество файлов данных, с которыми работали.

В качестве исходного набора данных были использованы файлы из публичного датасета "Smart meters in London" (<https://www.kaggle.com/jeanmidev/smart-meters-in-london>):

informations_households.csv - информация о датчиках энергопотребления и краткие сведения о домах, в которых эти датчики установлены;

block_0.csv .. block_111.csv - 112 файлов, содержащих обобщенные данные об энергопотреблении домохозяйств за каждые сутки;

weather_daily_darksky.csv - обобщенные данные о погоде за сутки для всего интервала дат в файлах **block_x.csv**.

С помощью специально разработанного скрипта (файл **ddprep.ipynb**), данные из файлов **block_0.csv .. block_111.csv** были предварительно очищены и объединены в один файл **daily_energy.csv**. Также с помощью этого скрипта из файла **weather_daily_darksky.csv** был сформирован файл **daily_weather.csv**. Отвратительное качество данных в файле **weather_daily_darksky.csv** вызвало необходимость разработки специального алгоритма идентификации даты, к которому относится та или иная запись в файле **weather_daily_darksky.csv**. Дополнительно с помощью указанного скрипта был сформирован файл **weekdays.csv**, содержащий номер дня недели в стандарте ISO для каждой даты анализируемого периода.

Описание логики анализа данных.

Нами было проведено исследование изменений среднего суточного энергопотребления домохозяйств (ССЭД) по годам и месяцам, а также по различным дням недели. Проанализировано влияние сезонных факторов, таких как температура воздуха и продолжительность светового дня на ССЭД. Изучено распределение ССЭД по совокупности 112 домов и по совокупности домохозяйств в домах с самыми высокими и с самыми низкими показателями ССЭД. Исследовано распределение максимального суточного энергопотребления домохозяйств по совокупности 112 домов. Получено распределение ССЭД по формам оплаты за электроэнергию. Исследовано распределение ССЭД по геодемографическим группам системы ACORN.

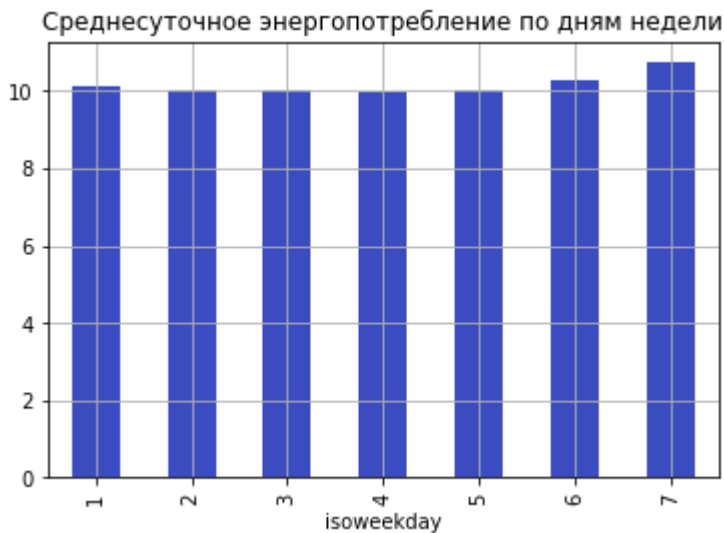
Результаты анализа факторов, влияющих на энергопотребление.

1. Распределение среднего суточного энергопотребления домохозяйств по годам и месяцам.



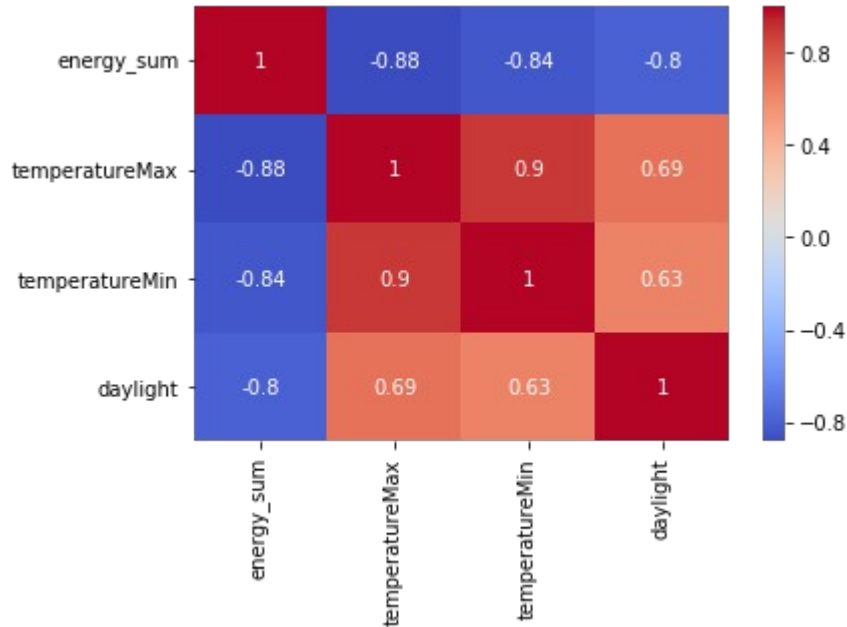
Видна ярко выраженная сезонность энергопотребления.

2. Распределение среднего суточного энергопотребления домохозяйств по дням недели.



Наблюдается незначительный разброс энергопотребления по дням недели с минимумом в четверг и максимумом в воскресенье.

3. Корреляция среднего суточного энергопотребления домохозяйств с основными сезонными факторами: температурой окружающего воздуха и продолжительностью светового дня.



Четко видна сильная обратная корреляция (с коэффициентами корреляции более -0.8) ССЭД с температурой окружающего воздуха (чуть в большей степени с максимальной суточной температурой temperatureMax, чуть в меньшей – с минимальной температурой temperatureMin). Обратная корреляция с продолжительностью светового дня чуть меньше, но тоже весьма существенна (коэффициент корреляции -0.8).

В рассматриваемом географическом местоположении (Лондон) наблюдается существенная прямая корреляция продолжительности светового дня и температуры окружающего воздуха (с коэффициентом корреляции более 0.6). Очевидно, что температурный фактор будет иметь максимальное влияние на энергопотребление только в домах с электрическим отоплением. В то же время, фактор продолжительности светового дня влияет на энергопотребление в любых домах.

4. Распределение среднего суточного энергопотребления домохозяйств по домам.



Наблюдается очень большое различие (более, чем в три раза) в среднем суточном энергопотреблении домохозяйств, расположенных в различных домах.

5. Распределение максимального суточного энергопотребления домохозяйств по домам.



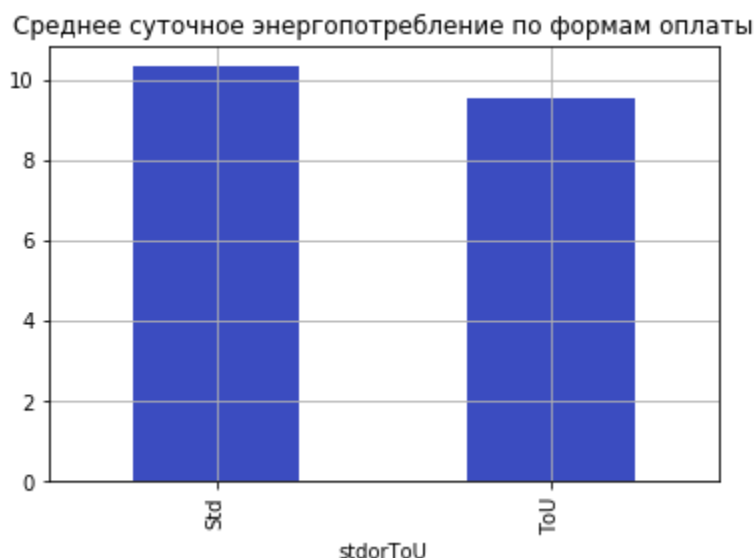
Наблюдаем очень большие величины суточного энергопотребления, превышающие 300 Квт, в некоторых домах. При этом различия в максимальном суточном энергопотреблении между домами превышают 8 раз.

6. Распределение среднего суточного энергопотребления домохозяйств одного дома.



Наблюдаются очень большие (в 25 раз) различия в среднесуточном энергопотреблении домохозяйств одного дома. Все домохозяйства в этом доме относятся к группе ACORN-A ('Lavish Lifestyles' - "Щедрый образ жизни"), т.е. имеют высокий уровень благосостояния. Это наводит на мысль, что помимо уровня благосостояния, на энергопотребление существенно влияет образ жизни членов домохозяйств.

7. Распределение среднего суточного энергопотребления домохозяйств по формам оплаты за электроэнергию.



Наблюдаем незначительную зависимость энергопотребления от формы оплаты. Оплату, зависящую от времени суток, используют более экономные домохозяйства.

7. Распределение среднего суточного энергопотребления домохозяйств по группам ACORN.



Наблюдаются очень существенные различия в энергопотреблении различных геодемографических групп населения. Максимальное среднее суточное энергопотребление (19.14 кВт) наблюдается в группе ACORN-A ('Lavish Lifestyles' - "Щедрый образ жизни"). Минимальное энергопотребление (6.64 кВт) - в группе ACORN-P ('Struggling Estates' - "Борющиеся имения").

Рейтинг факторов, влияющих на показатели энергопотребления.

Основываясь на проведенном анализе, может быть предложен следующий рейтинг факторов, влияющих на показатели энергопотребления домохозяйств:

1. Образ жизни членов домохозяйства (“Lifestyle”).

Обуславливает различия в уровне энергопотребления от 25 до 8000 раз внутри групп с одинаковым уровнем благосостояния.

2. Уровень благосостояния домохозяйства, детерминируемый во многом принадлежностью к определенной группе классификации ACORN.

Обуславливает различия в уровне энергопотребления более 3-х раз между различными группами ACORN.

3. Сезонные факторы: температура окружающего воздуха и продолжительностью светового дня.

Обуславливает различия в уровне энергопотребления между зимними и летними месяцами до 75%.

4. Форма оплаты за электроэнергию.

Домохозяйства, использующие тариф, зависящий от времени суток, имеют уровень энергопотребления на 8.3% ниже, чем использующие стандартный тариф.

5. День недели.

Наблюдаются различия в уровне энергопотребления по дням недели до 7.5%.