

## Командная игра "Умная аналитика для умного дома"

Уважаемые участники! Поздравляем вас с выходом на новый этап изучения Python для анализа данных. На этом этапе вы, параллельно с изучением оставшихся материалов по Pandas, вам предлагается поработать над объемным проектом, связанным с анализом данных об использовании датчиков энергопотребления в 112 домах Великобритании.

Полученный отчет **вы можете использовать как часть своего будущего портфолио аналитика.**

### Описание задачи:

Представьте, что вы – команда аналитиков, участвующая в реализации проекта по разработке оборудования для умного дома. Компания заключила договор с производителями датчиков энергопотребления, которые только что закончили тестирование своего оборудования в 112 домах Великобритании.

**Ваша задача** – подготовить отчёт, который наглядно покажет какие факторы повлияли на энергопотребление. Отчет должен быть визуально понятным, чтобы эксперты компании на основе него могли дать рекомендации по энергосбережению.

Заказчик предоставил вам все необходимые показатели. **(Ссылка на данные и инструкция по ним в разделе "Приложения")**.

Вам нужно будет выбрать из предложенного для работы датасета те файлы для анализа, извлечь из них нужную информацию и представить её в виде отчёта о факторах, влияющих на уровень потребления электрической энергии.

### Отчет должен включать в себя:

1. Название команды, список участников, которые принимали участие в работе над задачей;
2. Список и количество таблиц, с которыми работали;
3. Описание логики анализа данных;
4. Графическую визуализацию отчетов:

- О наиболее высоких и низких данных электропотребления в разные периоды времени для одного дома;

- О наиболее высоких и низких данных электропотребления в один период для разных домов;

- О взаимосвязи дополнительных показателей(погодных условий, тарифного плана, уровня благосостояния домовладельцев и т.п.) и уровня энергопотребления;

#### 5. Рейтинг факторов, влияющих на показатели энергопотребления.

Желательно использовать самые различные вариации визуализации отчетов: графики, различные виды диаграмм и др. Важно, чтобы отчет был наглядный, понятный и на основе него руководители могли принять то или иное решение.

**Дедлайн - 4 июня 2019 г.**

#### **Рекомендуемый план работ:**

1. Изучите предоставленные данные и выберите из них те, на основе которых будет проводиться анализ (описание файлов на русском языке предоставлено в Приложении к данной инструкции).
2. Опишите логику по которой вы анализируете данные. Сколько файлов анализировали, какие именно файлы. Каким показателям было уделено особое внимание?

***Например:** проанализировали среднесуточные показатели энергопотребления в доме пенсионеров, оплачивающих электричество по стандартному тарифу и в доме молодого холостяка с доходом выше среднего. Уровень энергопотребления был сопоставлен с показателями максимальной и минимальной температуры воздуха, скоростью ветра и данными об осадках.*

Описание логики можно сделать в конце работы, но в презентации разместить вторым пунктом.

3. Изучите представленные в датасете показатели энергопотребления:
  - 3.1. Выберите наиболее высокие и наиболее низкие данные, зафиксированные для одного дома в разные периоды времени;
  - 3.2. Для разных домов в один и тот же период.
4. Сравните данные о погодных условиях, измеренные в период высокого или низкого энергопотребления. Сопоставьте полученные сведения с показателями энергопотребления.

5. Изучите сведения о домовладельцах: способ оплаты за электричество, к какой категории потребителей по системе ACORN (см. Приложение) они относятся. Сравните уровень энергопотребления в домах, принадлежащих разным категориям владельцев.
6. Разработайте иллюстрации, демонстрирующие изменения энергопотребления в разных условиях:
  - 6.1. Суточная динамика энергопотребления в домах, принадлежащих разным категориям потребителей.
  - 6.2. Сезонная динамика энергопотребления и её связь с погодными условиями (можно попробовать отобразить два показателя на одном графике: например, среднесуточная динамика энергопотребления и изменение среднесуточной температуры воздуха в течение того же периода).
  - 6.3. Сравнение средних показателей энергопотребления у разных групп населения при одинаковых погодных условиях.
7. Составьте рейтинг факторов, влияющих на показатели энергопотребления. Обоснуйте каждую позицию рейтинга.
8. Подготовьте отчет о проведенном анализе. Отчет должен включать текстовую, графическую и табличную информацию. Рекомендованный формат отчёта – pdf. При желании можно также сделать дашборд в любом формате, для просмотра которого не требуется установка коммерческого ПО.
9. Загрузите результаты работы (текстовый отчет и ноутбук с кодом и графиками) на GitHub (<https://github.com>) и предоставьте ссылку на репозиторий организаторам.

## Критерии оценки работы

Критерий	Пояснения	Максимальный балл
Содержание отчёта	Присутствие в отчёте всех пунктов плана	5
	Объём данных, использованных для анализа	10
	Полнота набора показателей, использованных при составлении рейтинга	10
	Обоснованность разработанного рейтинга.	10
Качество кода	Оформление, наличие комментариев	5
	Корректное использование стандартных функций	10
	Эффективность кода	10
	Универсальность кода и возможность его использования для решения аналогичных задач с другими данными	10
	Оформление репозитория на GitHub, наличие файла с поясняющей информацией	5
Оформление отчёта	Единство стиля оформления, продуманный дизайн	10
	Соответствие типов наглядной информации (графики, таблицы, диаграммы) поставленным задачам	5
	Наличие на графиках элементов, облегчающих восприятие (подписи осей, легенды и т.п.)	10
Итого		100

## Приложение

### Общая характеристика датасета

При выполнении данного задания вы будете работать с набором данных, содержащим информацию об уровне потребления электроэнергии жителями 112 домов Великобритании. В каждом из этих домов были установлены около 50 датчиков, фиксирующих уровень энергопотребления 1 раз в 30 минут. Также датасет содержит файлы с информацией о погодных условиях, об экономическом статусе владельцев домов, о способе оплаты за электроэнергию.

Общий объем датасета составляет 1 гигабайт, но для выполнения задания нет необходимости работать со всеми этими файлами. Файлы можно скачать по адресу: <https://www.kaggle.com/jeanmidev/smart-meters-in-london>.

Одной из особенностей датасета является то, что в нем большое внимание уделяется особенностям экономического статуса и потребительской активности жителей домов, в

которых были установлены счетчики. Для ознакомления с методикой классификации и основными характеристиками групп потребителей Вы можете использовать руководство пользователя, доступное по адресу: [https:// acorn.caci.co.uk/downloads/Acorn-User-guide.pdf](https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf). Язык руководства – английский.

Характеристика файлов, входящих в состав датасета:

informations\_households.csv

Информация о датчиках энергопотребления и краткие сведения о домах, в которых эти датчики установлены. Столбцы датафрейма:

- ✓ LCLid – id датчика;
- ✓ stdorToU – форма оплаты за электроэнергию в доме, в котором установлен датчик (Std – стандартная, ToU – оплата зависит от времени суток);
- ✓ Acorn, Acorn\_grouped – информация о том, к какой категории потребителей по системе ACORN относится семья, проживающая в данном доме;
- ✓ file – имя файла, содержащего показатели счетчиков.

halfhourly\_dataset.zip

Архив содержит 112 файлов, содержащих данные об энергопотреблении, получаемые с каждого счетчика 1 раз в 30 минут. Каждый файл содержит информацию с датчиков, установленных в одном доме. Столбцы датафрейма:

- ✓ LCLid – id датчика;
- ✓ tstp – дата и время фиксации показателей;
- ✓ energy(kWh/hh) – уровень энергопотребления.

daily\_dataset.zip

Архив содержит 112 файлов, содержащих обобщенные данные об энергопотреблении за каждые сутки. Каждый файл содержит информацию с датчиков, установленных в одном доме. Столбцы датафрейма:

- ✓ LCLid – id датчика;
- ✓ day – дата;

- ✓ energy\_median – медиана суточных показателей;
- ✓ energy\_mean – среднее арифметическое суточных показателей;
- ✓ energy\_max – максимальное значение энергопотребления, зафиксированное датчиком в течение дня;
- ✓ energy\_count – количество показателей, зафиксированных датчиком в течение дня;
- ✓ energy\_std – стандартное отклонение;
- ✓ energy\_sum – сумма значений всех показателей;
- ✓ energy\_min – минимальное значение энергопотребления, зафиксированное датчиком в течение дня.

acorn\_details.csv

Данные о группах потребителей согласно классификации ACORN. При характеристике групп в данном датафрейме используется сравнение каждой группы с общенациональными показателями. Так, если значение ячейки по какому-либо показателю составляет 150 это означает, что в рассматриваемой группе этот показатель встречается в 1,5 раза чаще, чем в целом по стране. Столбцы датафрейма:

- ✓ MAIN CATEGORIES, CATEGORIES, REFERENCE – показатели, по которым осуществляется сравнение групп;
- ✓ ACORN-A, ACORN-B, ACORN-C, ACORN-D, ACORN-E, ACORN-F, ACORN-G, ACORN-H, ACORN-I, ACORN-J, ACORN-K, ACORN-L, ACORN-M, ACORN-N, ACORN-O, ACORN-P, ACORN-Q – частота встречаемости каждого из рассмотренных признаков в каждой группе.

P.S. Использование данного файла при проведении анализа не является обязательным. Если Вы владеете английским языком в достаточной степени, то возможно, Вам будет удобнее использовать для ознакомления с характеристиками групп руководство пользователя, доступное на официальном сайте.

weather\_daily\_darksky.csv

Обобщённые данные о погоде за день. Столбцы датафрейма:

- ✓ temperatureMax – максимальное значение температуры воздуха;

- ✓ temperatureMaxTime – время, когда была зафиксирована максимальная температура воздуха;
- ✓ windBearing – направление ветра (по азимуту);
- ✓ icon – стандартизованное словесное описание погодных условий;
- ✓ dewPoint – точка росы;
- ✓ temperatureMinTime – время, когда была зафиксирована минимальная температура воздуха;
- ✓ cloudCover – облачность;
- ✓ windSpeed – скорость ветра;
- ✓ pressure – атмосферное давление;
- ✓ apparentTemperatureMinTime – время, когда была зафиксирована минимальная температура комфорта;
- ✓ apparentTemperatureHigh – дневная температура комфорта;
- ✓ precipType – тип осадков;
- ✓ visibility – видимость (в милях);
- ✓ humidity – относительная влажность;
- ✓ apparentTemperatureHighTime – время, когда была зафиксирована дневная температура комфорта;
- ✓ apparentTemperatureLow – ночная температура комфорта;
- ✓ apparentTemperatureMax – максимальная температура комфорта;
- ✓ uvIndex – UV-индекс;
- ✓ time – время начала сбора данных;
- ✓ sunsetTime – время заката;
- ✓ temperatureLow – минимальная ночная температура;
- ✓ temperatureMin – минимальная температура за сутки;

- ✓ temperatureHigh – максимальная дневная температура;
- ✓ sunriseTime – время восхода;
- ✓ temperatureHighTime – время, когда была зафиксирована максимальная дневная температура;
- ✓ uvIndexTime – время, когда был зафиксирован максимальный UV-индекс;
- ✓ summary – словесное описание погоды в течение дня (не рекомендуется использовать для автоматизированного анализа!);
- ✓ temperatureLowTime – время, когда была зафиксирована минимальная ночная температура;
- ✓ apparentTemperatureMin – минимальная температура комфорта;
- ✓ apparentTemperatureMaxTime – время, когда была зафиксирована максимальная температура комфорта за сутки;
- ✓ apparentTemperatureLowTime – время, когда была зафиксирована минимальная ночная температура комфорта;
- ✓ moonPhase – фаза луны. weather\_hourly\_darksky.csv

Почасовые сведения о погоде. Столбцы датафрейма:

- ✓ visibility – видимость в милях;
- ✓ windBearing – направление ветра (по азимуту);
- ✓ temperature – температура воздуха;
- ✓ time – время записи показателей;
- ✓ dewPoint – точка росы;
- ✓ pressure – атмосферное давление;
- ✓ apparentTemperature – температура комфорта;
- ✓ windSpeed – скорость ветра;



- ✓ `precipType` – тип осадков;
- ✓ `icon` – стандартизованное словесное описание погодных условий;
- ✓ `humidity` – относительная влажность;
- ✓ `summary` – нестандартизованное словесное описание погодных условий.