



Анализ данных интеллектуальных счетчиков Лондона

Чтобы лучше следить за потреблением энергии, правительство хочет, чтобы поставщики энергии устанавливали интеллектуальные счетчики в каждом доме в Англии, Уэльсе и Шотландии. Есть более 26 миллионов домов для поставщиков энергии, чтобы к 2020 году каждый дом имел умный счетчик.

Команда 09TeamPy

Участники:

1. Черных Сергей Викторович
2. Волков Дмитрий Александрович
3. Деев Вадим Александрович
4. Сидорова Юлия Викторовна
5. Сайфутдинов Фарин Ринатович
6. Вершинин Алексей Николаевич
7. Чернышев Иван Владиславович
8. Ветошкин Дмитрий Александрович
9. Тодорук Дмитрий Павлович
10. Шуравин Евгений Петрович
11. Мусорин Сергей Викторович

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?

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

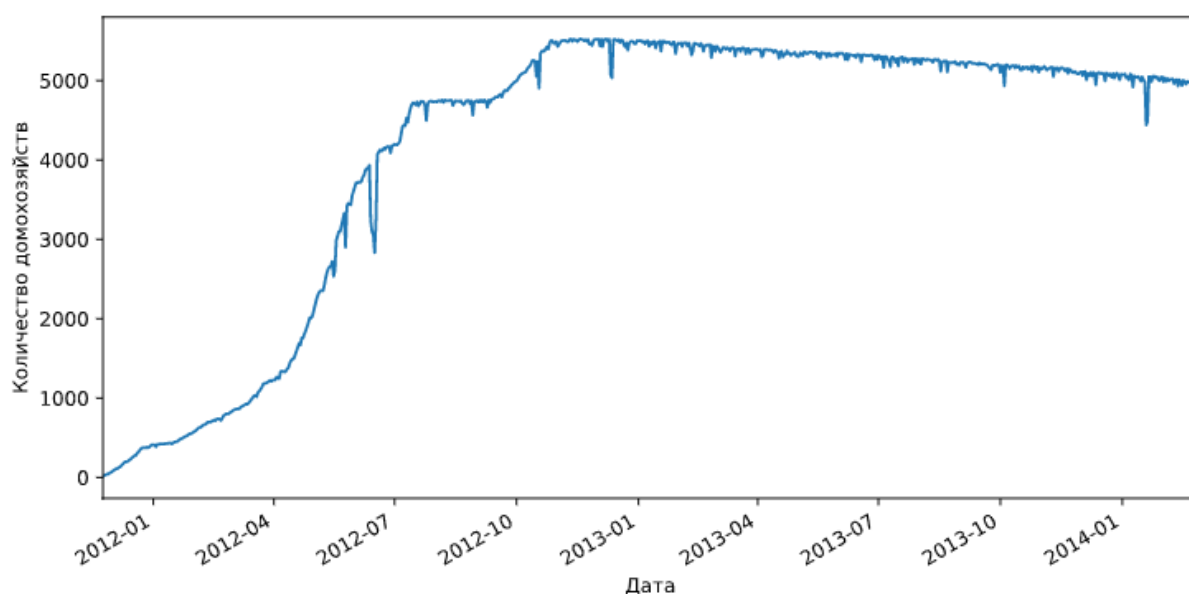
3. Изучите представленные в датасете показатели энергопотребления:
 - 3.1. Выберите наиболее высокие и наиболее низкие данные, зафиксированные для одного дома в разные периоды времени;
 - 3.2. Для разных домов в один и тот же период.
4. Сравните данные о погодных условиях, измеренные в период высокого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

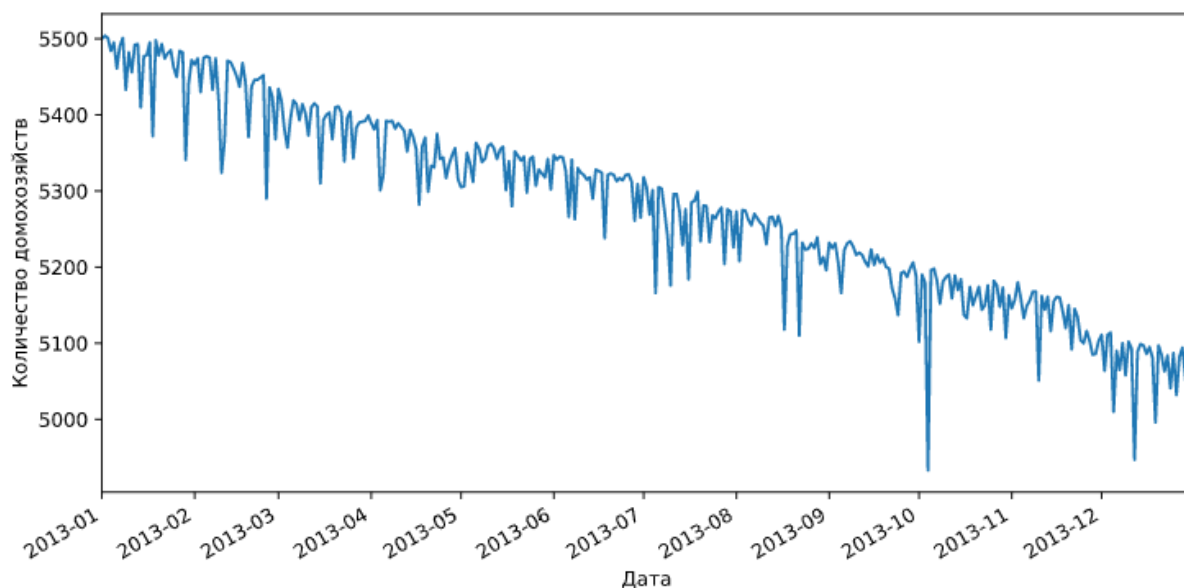
Файлы данных, которые были использованы в ходе проведения анализа:

informations_households.csv	Информация о датчиках энергопотребления и краткие сведения о домах, в которых эти датчики установлены.
weather_daily_darksky.csv	Обобщённые данные о погоде за день.
weather_hourly_darksky.csv	Почасовые сведения о погоде.
acorn_details.csv	Данные о группах потребителей согласно классификации ACORN.
uk_bank_holidays.csv	Информация о национальных праздниках в Великобритании
файлы block_*.csv из архива daily_dataset.zip	Архив содержит 112 файлов, содержащих обобщенные данные об энергопотреблении за каждые сутки.

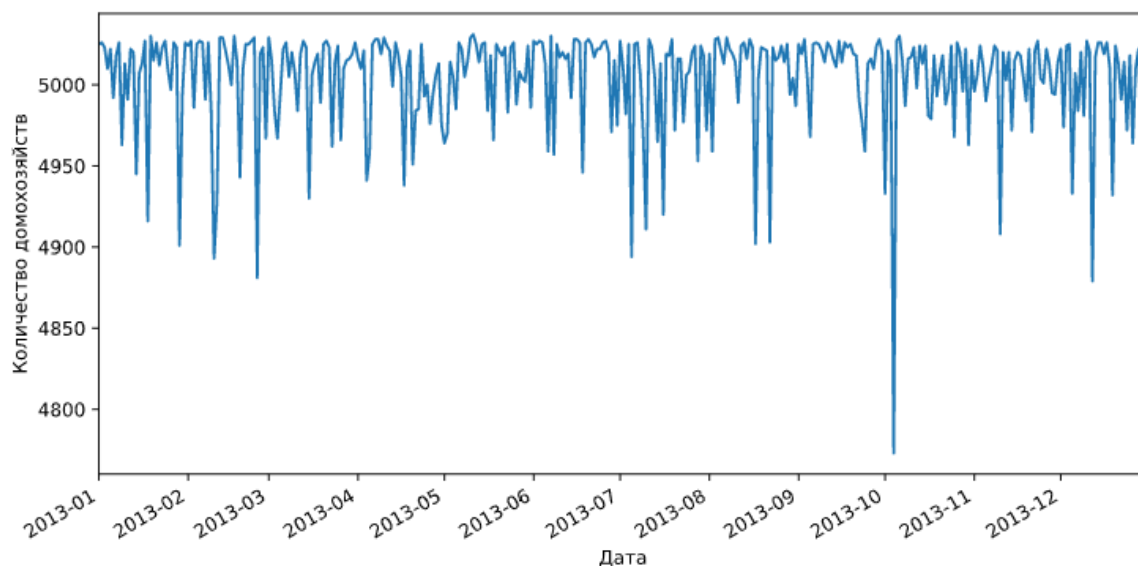
После проведения предварительного анализа на графике видно, что с конца 2011 года наблюдается резкое увеличение числа доступных датчиков, пик которого достигнут в 2013 году. т.е .конец 2011 года и полностью 2012 год - это видимо период, когда датчики устанавливались. Далее видно плавное снижение количество доступных датчиков - период нормальной эксплуатации - некоторые датчики выходят из строя. Поскольку в дальнейшем мы будем анализировать потребление в зависимости от разных факторов, в том числе от сезона, то для анализа нужно взять период кратный году. Наиболее подходящий - это 2013 год, когда количество работающих датчиков менялось незначительно.



На графике видно, что с конца 2011 года наблюдается резкое увеличение числа доступных датчиков, пик которого достигнут в 2013 году. т.е.конец 2011 года и полностью 2012 год - это видимо период, когда датчики устанавливались. Далее видно плавное снижение количество доступных датчиков - период нормальной эксплуатации - некоторые датчики выходят из строя. Поскольку в дальнейшем мы будем анализировать потребление в зависимости от разных факторов, в том числе от сезона, то для анализа нужно взять период кратный году. Наиболее подходящий - это 2013 год, когда количество работающих датчиков менялось незначительно.



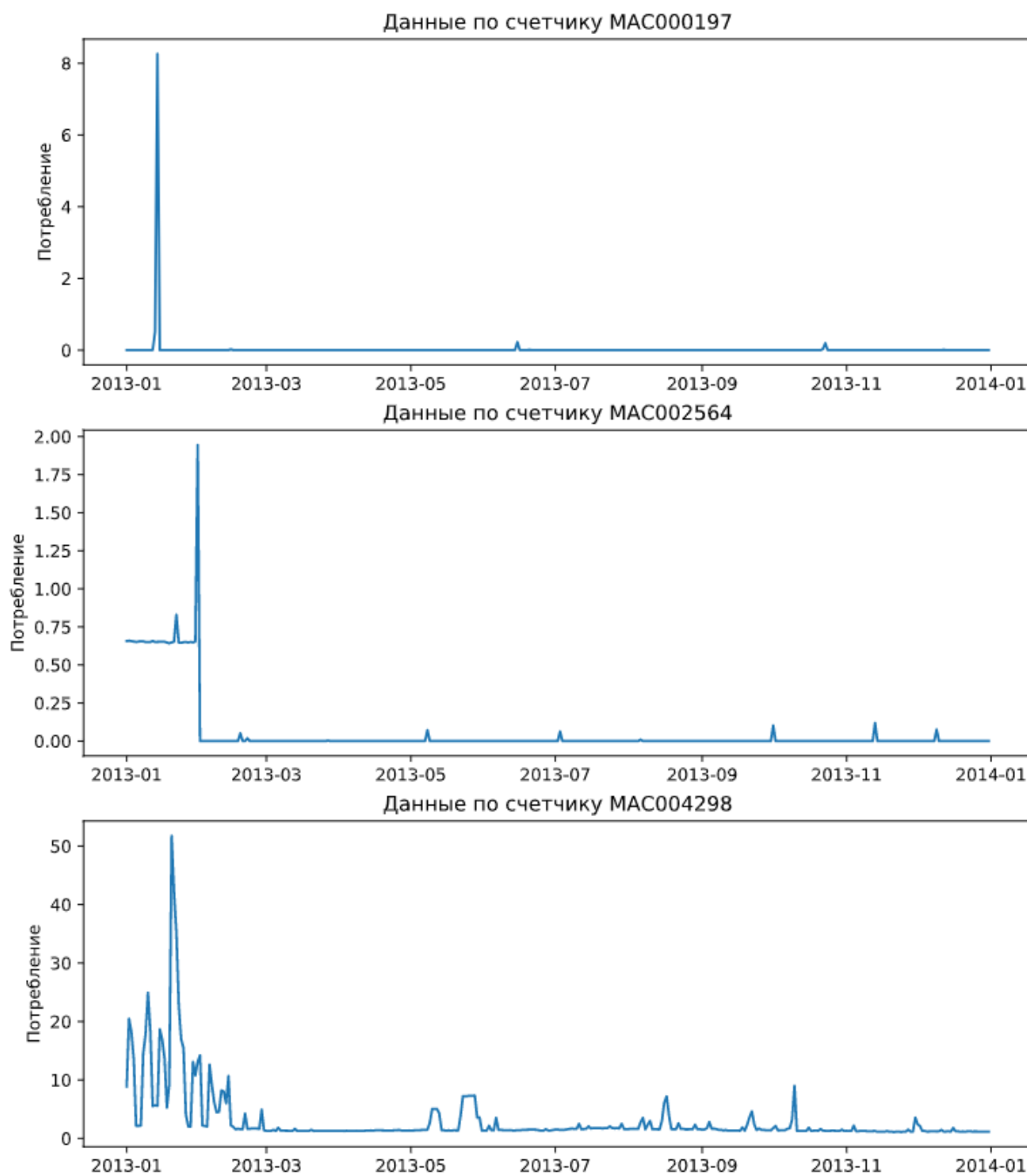
Новый график выглядит более гладким, тем не менее, из него понятно, что за 2013 год вышло из строя около 400 датчиков. Желательно их исключить, так как они могут исказить картину, поскольку работали не весь год, а это может исказить распределение потребления по сезонам. К тому же, вышедший из строя датчик - потенциально не надежен - с большой вероятностью он мог передавать неверные показания. Оставим для анализа только датчики, которые проработали не менее 350 дней в году. Может показаться, что отброшено слишком много данных, но тем не менее, для анализа мы используем более половины всего датасета:



Сгруппируем датчики по максимальному и среднему потреблению. Потом отсортируем по соотношению максимальное/среднее потребление (слишком большее соотношение будет подозрительно):

	LCLid	max	mean	over
167	MAC000197	8.259	0.025699	321.375685
2305	MAC002594	0.854	0.003492	244.575924
1894	MAC002136	0.042	0.000188	222.970588
4223	MAC004672	0.302	0.001989	151.836565
2277	MAC002564	1.945	0.060756	32.013212
1152	MAC001309	1.487	0.050438	29.481532
1206	MAC001372	22.988	1.234202	18.625797
1748	MAC001976	8.341	0.471846	17.677372
3883	MAC004298	51.756	3.040348	17.023052
3988	MAC004410	120.603	7.100767	16.984503

Обнаружили несколько датчиков с аномально высоким соотношением. Построим графики потребления по дням для некоторых из них:



Видно, что MAC000197 передал ненулевые значения только 4 дня. MAC002564 передавал данные в течении января, а после тоже только почти нулевые. Последний MAC004298 с соотношением 17, показывает более равномерный график в течении года, но тоже похоже, что нормальная работа этого датчика была только первые два месяца.

Отсортируем датафрейм по среденедневному потреблению. Учитывая, что потребление только холодильника около 0,5 кВт/час в день - более низкие значения скорее всего некорректны:

	LCLid	max	mean	over
3258	MAC003627	0.000	0.000000	NaN
3673	MAC004067	0.000	0.000000	NaN
1894	MAC002136	0.042	0.000188	222.970588
4223	MAC004672	0.302	0.001989	151.836565
2305	MAC002594	0.854	0.003492	244.575924
167	MAC000197	8.259	0.025699	321.375685
1152	MAC001309	1.487	0.050438	29.481532
2277	MAC002564	1.945	0.060756	32.013212
32	MAC000037	0.862	0.081527	10.573123
2119	MAC002388	0.491	0.332074	1.478586
4581	MAC005069	4.195	0.458177	9.155854
1748	MAC001976	8.341	0.471846	17.677372
357	MAC000408	1.484	0.541570	2.740180
4800	MAC005313	5.160	0.676602	7.626342
2559	MAC002873	11.849	0.785458	15.085460

Так и есть. Это все те же самые датчики с высоким соотношением максимума к среднему. Исключим эти датчики из анализа как неисправные:

MAC003627, MAC004067, MAC002136, MAC004672 ,MAC002594, MAC000197 ,MAC001309, MAC002564, MAC000037, MAC002388,MAC005069, MAC001976 , MAC001372

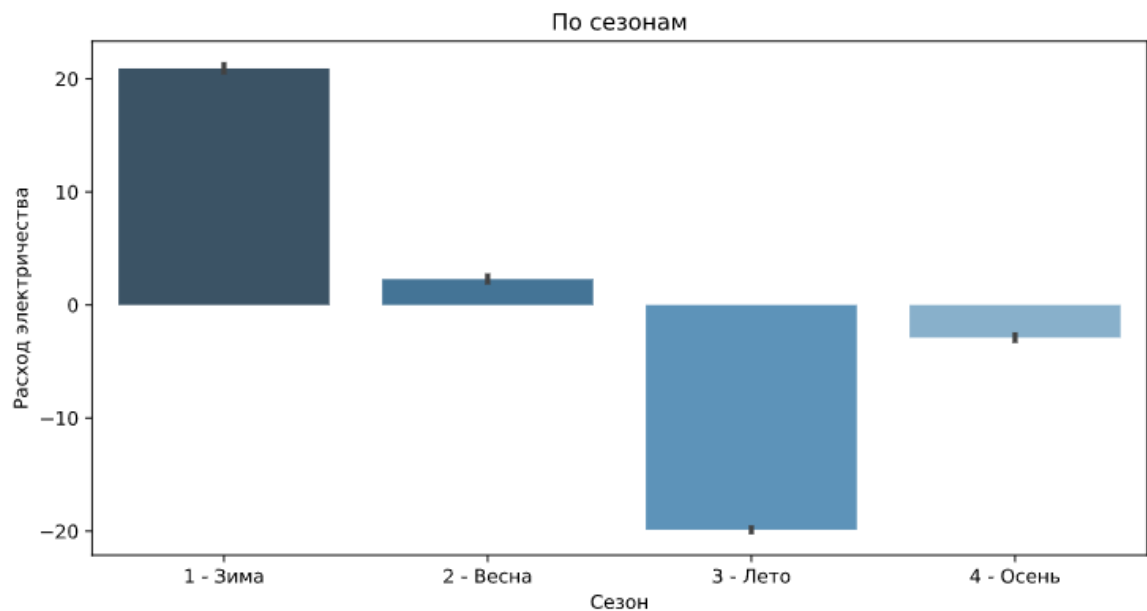
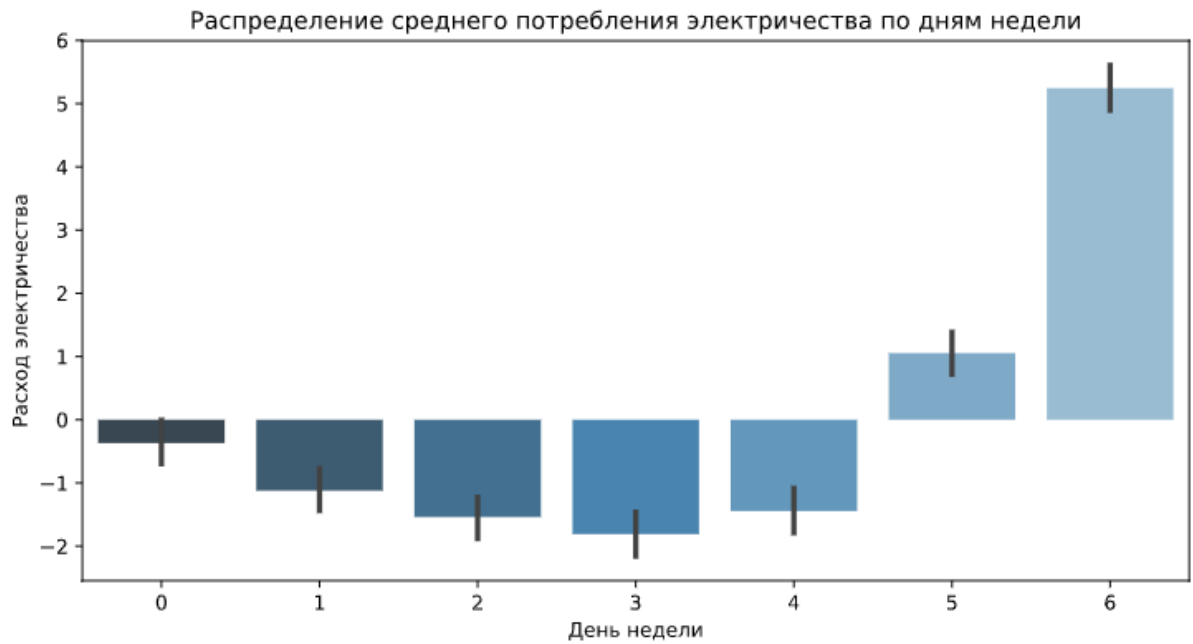
Рассмотрим **hhblock_dataset**

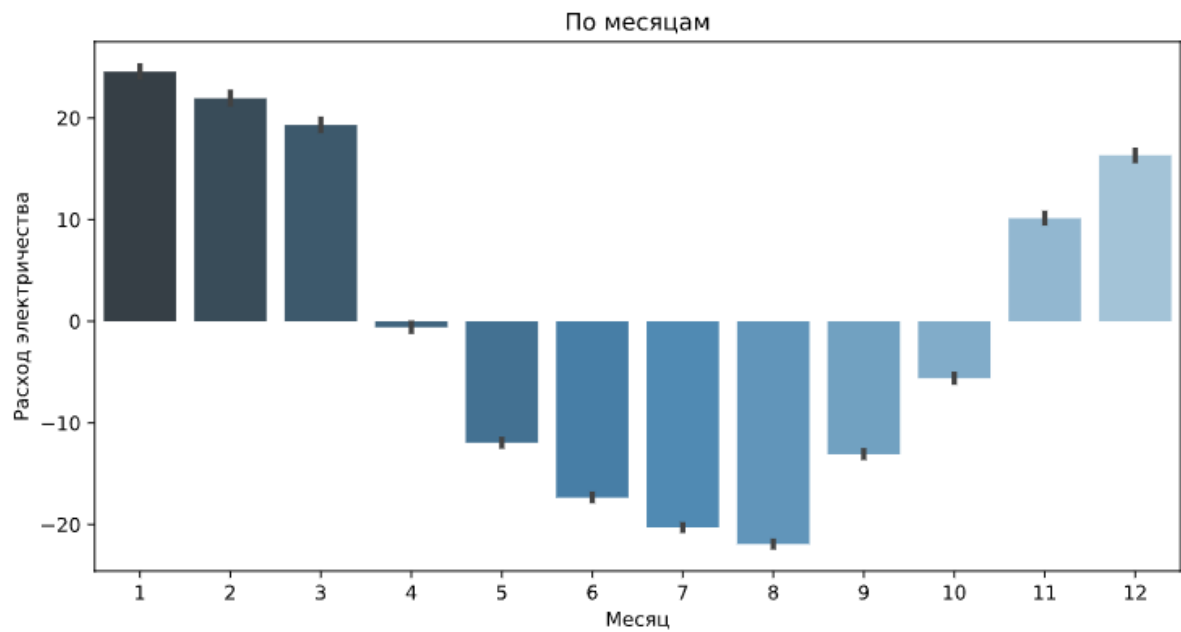
каждый из 112 файлов в папке содержит получасовые данные об энергопотреблении за каждые сутки. Одна строка файла содержит информацию по получасовым значениям энергопотребления одного домовладения за одни сутки.

Рассмотрим **halfhourly_dataset**

Добавим 3 колонки - день недели, месяц и время года

Рассмотрим распределение среднего потребления электричества по дням недели:

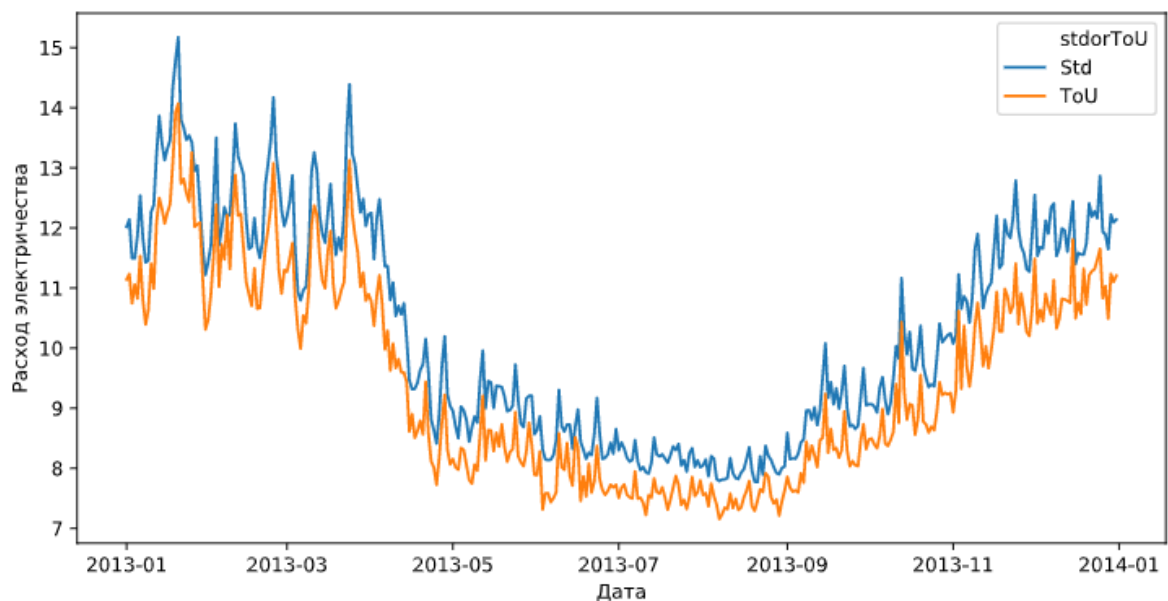




Вывод:

Пик потребления приходится на воскресенье, более +5% от среднего, в течении других дней недели потреблению более равномерное. Сильно зависит среднее потребление от сезона, до -20% летом и +20% зимой. Максимальное потребление достигается в декабре, а минимум в августе. Таким образом сезон - более важный фактор с потреблением эл.энергии, разброс до 40%, в то время как от дня недели зависимость более слабая.

Построим график потребления электричества в зависимости от типа счетчика и даты

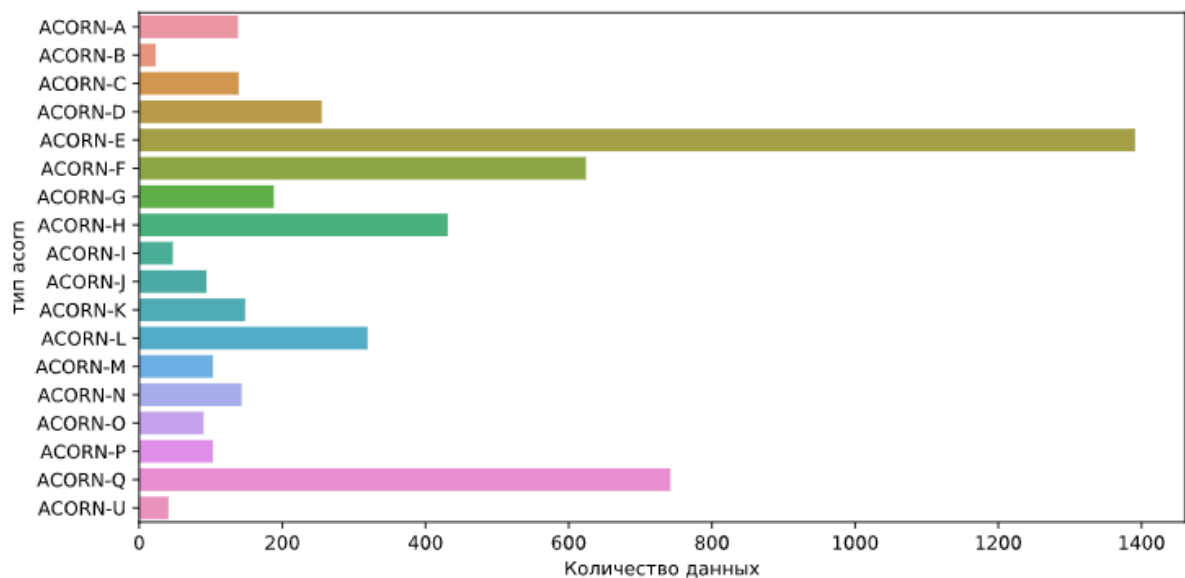


По графику видно, что люди со стандартным тарифом тратят электричество больше, чем люди с динамическим тарифом.

Среднесуточное потребление при установленном стандартном счетчике: 10.275 кВт/час, при динамическим: 9.450 кВт/час

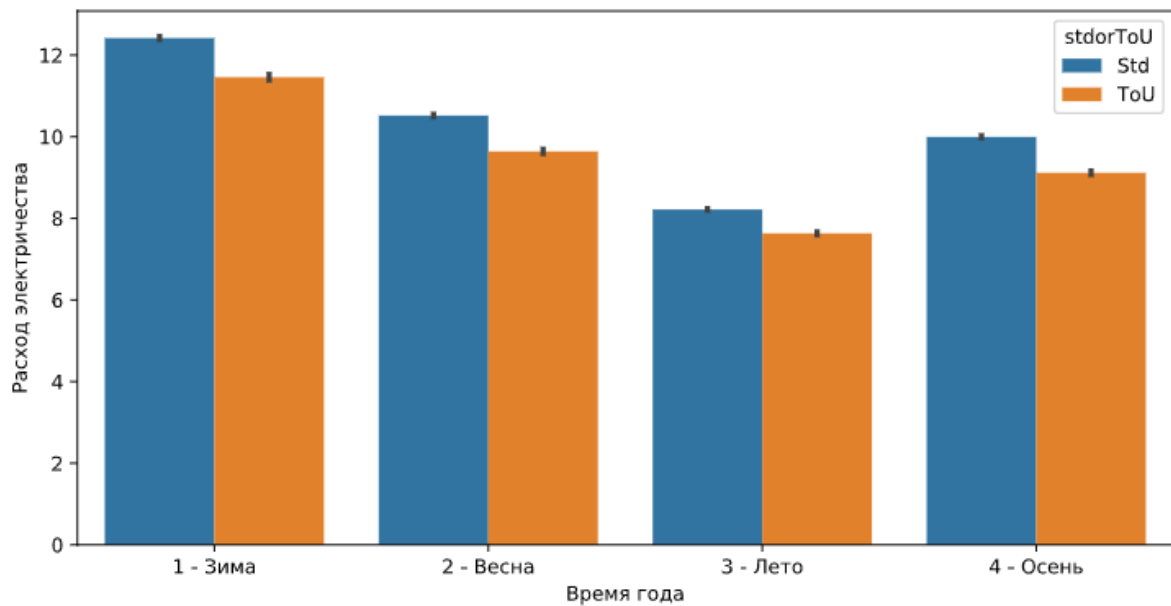
Потребление при стандартном счетчике больше на 8.74 % чем при динамическим.

Проанализируем количество установленных датчиков по группам потребителей Acorn:



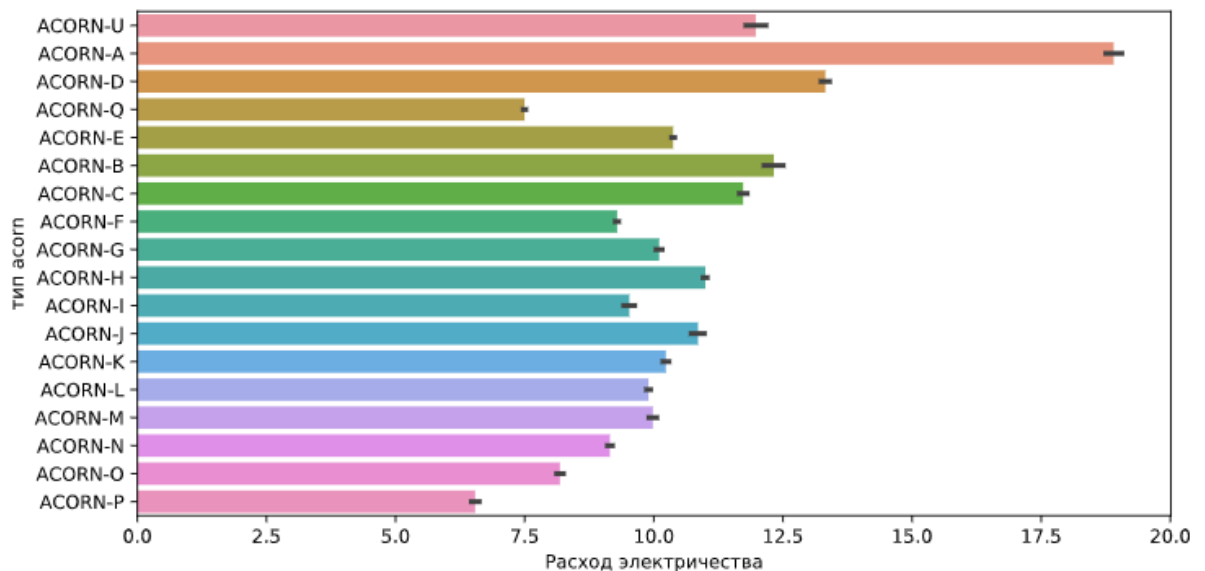
Вывод: имеется значительная неравномерность в распределении датчиков по группам, так Acorn-E представлены около 1400 датчиков, а количество датчиков в группах Acorn-B,I,U не дотягиват и до сотни. Таким образом, наибольший вклад в результаты исследования вносят группы Acorn -E,F,Q ввиду многочисленности датчиков. Целесообразно в дальнейшем рассмотреть статистику потребления по "малочисленным" Acorn отдельно.

График потребления электричества в зависимости от типа счетчика и времени года



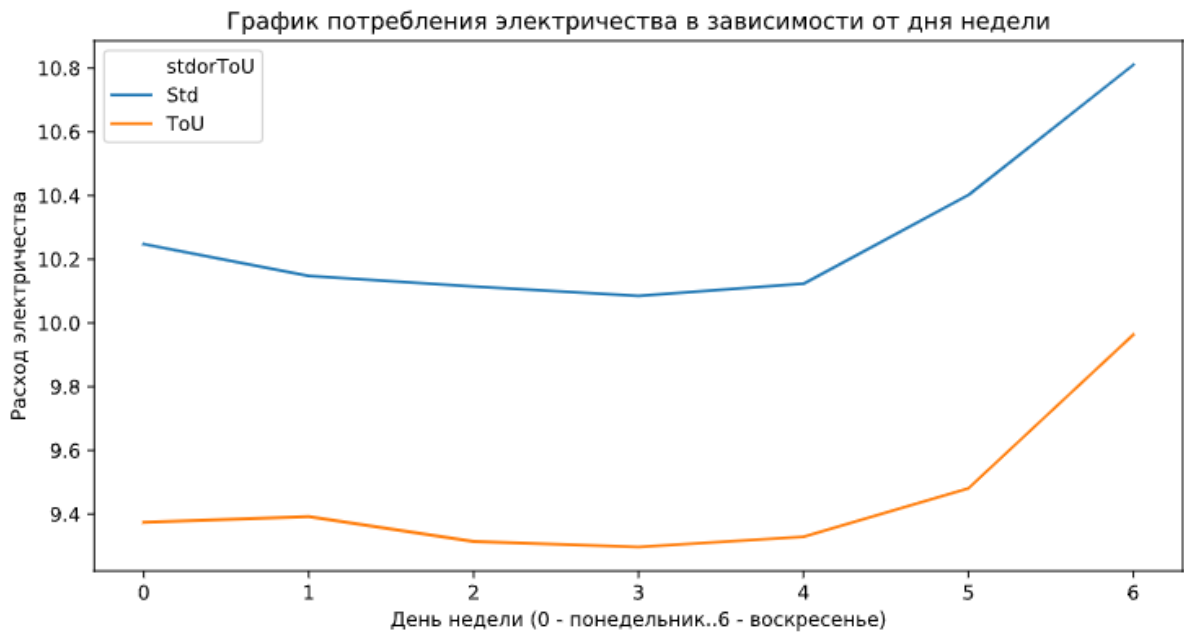
По графику видно, что существует связь между потреблением электроэнергии и временем года. Сезонный эффект очень очевиден, т.к. могут использовать электричество в качестве источника тепла. В теплое время (лето) электроэнергии тратится меньше, чем в холодное время.

График потребления электричества в зависимости от типа Acorn



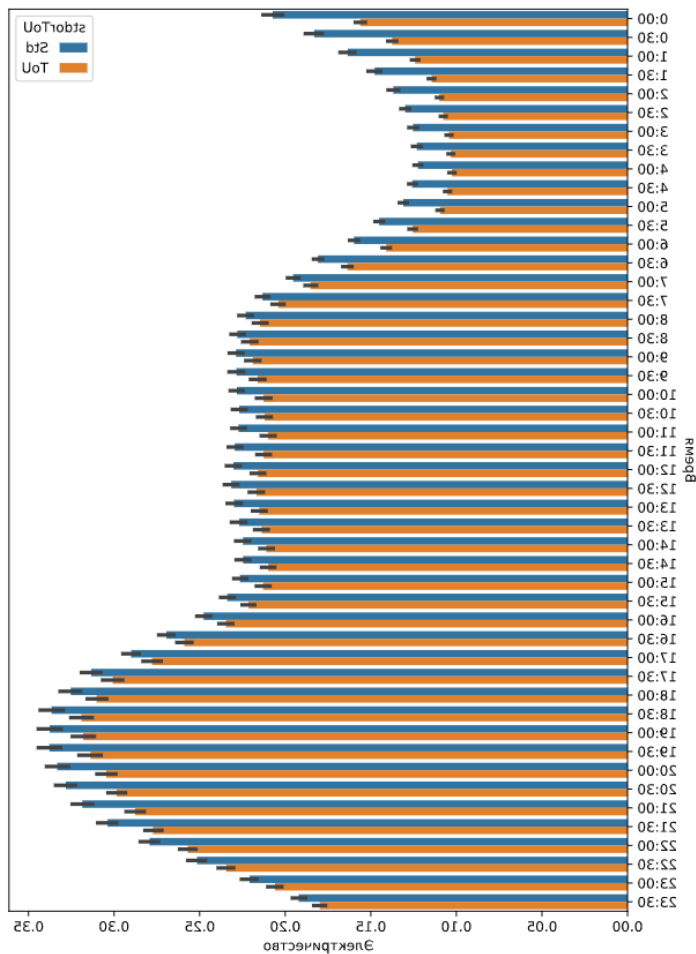
В целом наблюдается тенденция снижения потребления от группы Acorn-A к группе Acorn-Q, т.е. от богатых к бедным. Люди из ACORN-A тратят примерно в три раза больше электроэнергии, чем люди из группы ACORN-P, т.е. это разница в 300% и пока это наиболее значимый фактор.

График потребления электричества в зависимости от дня недели



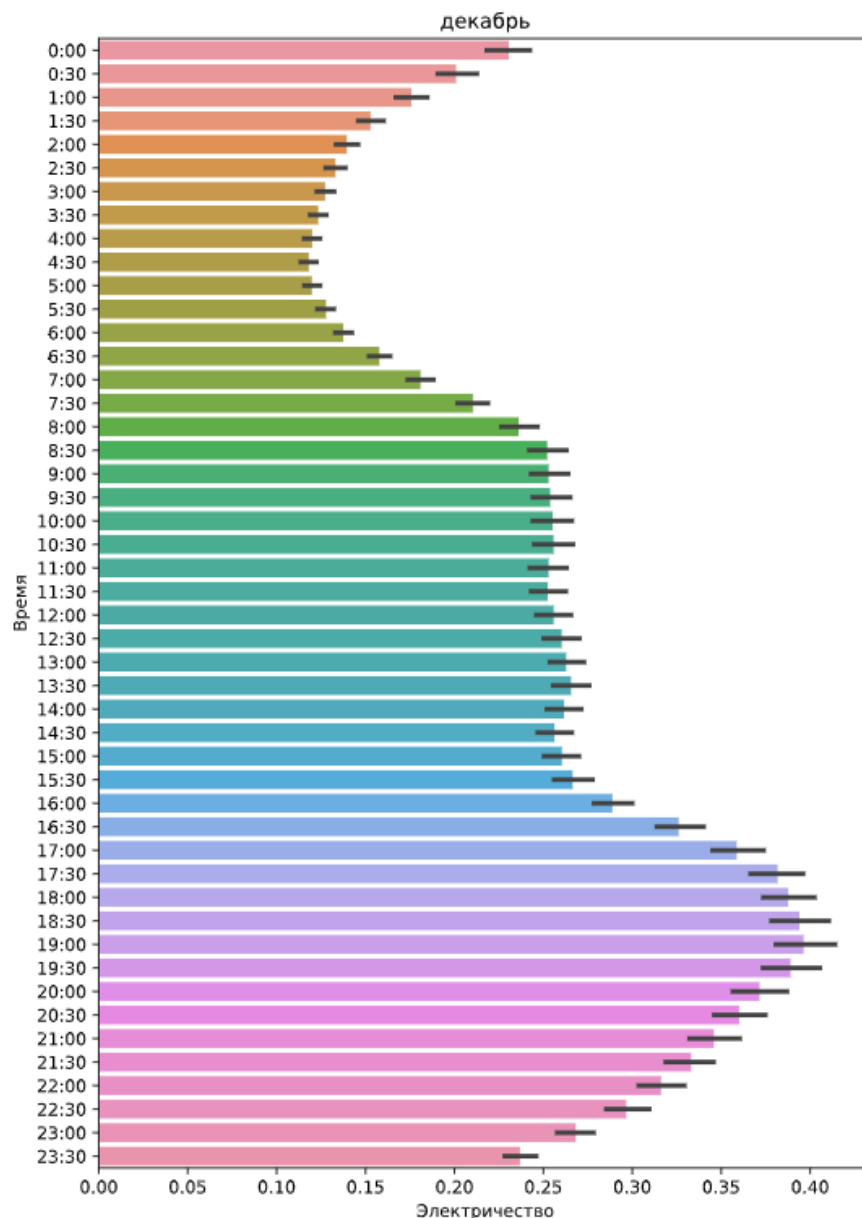
По графику видно, что в выходные дни потребление электричества увеличивается, по сравнению с рабочими днями.

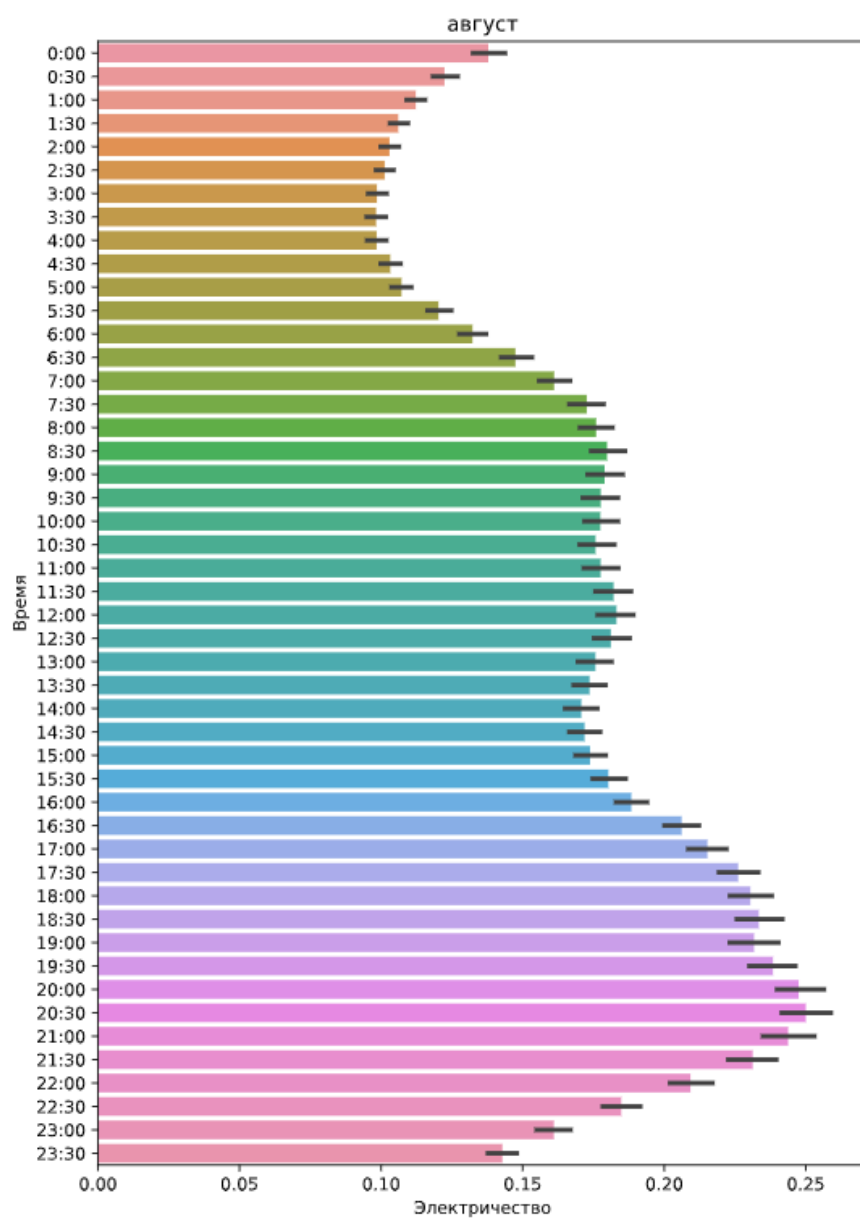
Построим графики почасового потребления электроэнергии



Из графика видно, что пик потребления электричества приходится на вечернее время с 18:30 до 20:00 часов, наименьшее потребление электроэнергии с 02:00 до 04:00 часов. Разница в потреблении примерно трехкратная. Причем, это верно как для стандартного счетчика, так и для многотарифного. Многотарифная оплата за электроэнергию призвана в первую очередь стимулировать потребителей потреблять энергию в часы наименьшей нагрузки и снижать потребление в часы пик, т.е. выравнивать потребление в течении суток. Как видно из графика - ожидания многотарифных счетчиков не оправдываются. Хотя в целом, люди, кто установил многотарифный счетчик электроэнергию используют более экономно.

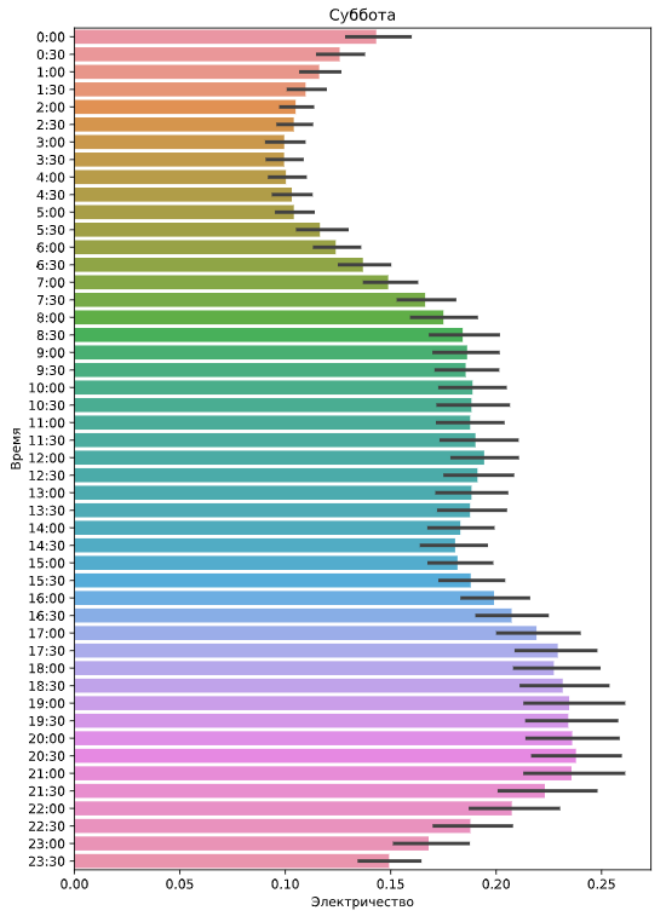
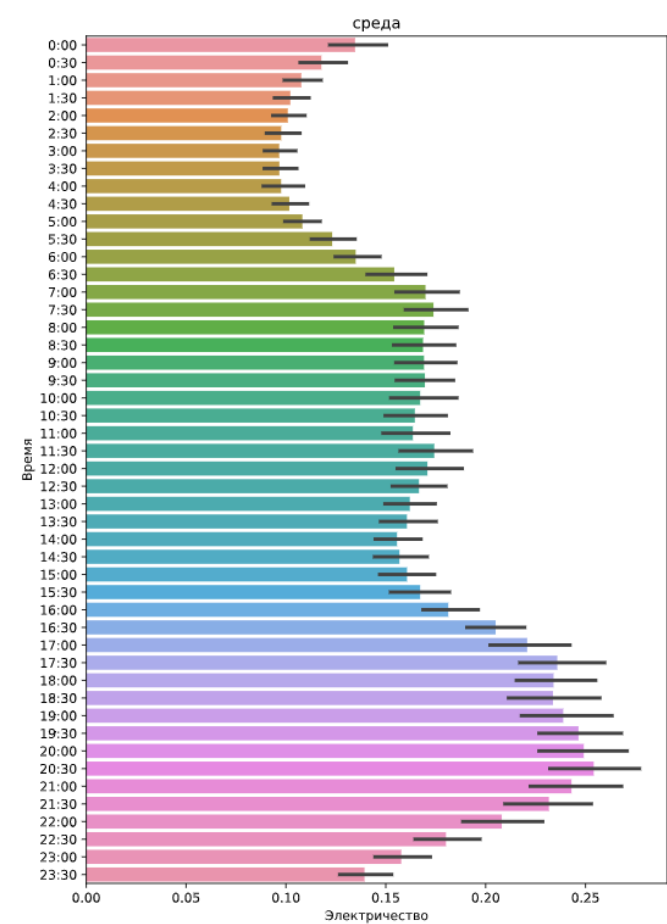
Поскольку есть сильная зависимость потребления от сезона, то построим графики часового потребления отдельно для декабря и для августа:

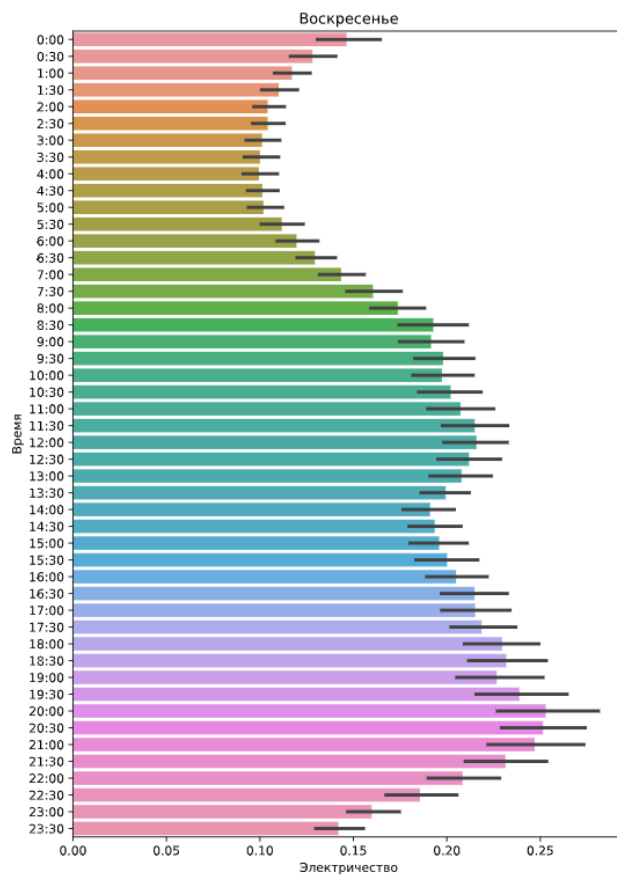




Графики похожи, пики и минимумы приблизительно в одно и то же время, видимо это связано с рабочими часами, а не с солнечным ритмом. Минимум потребления и в декабре и в августе - ночью, около 0,1 кВтт/час, а вот днем заметна разница. По видимому сезонная разница больше связанна не с отоплением, а с коротким солнечным днем и недостатком света зимой, когда приходится чаще использовать искусственное освещение.

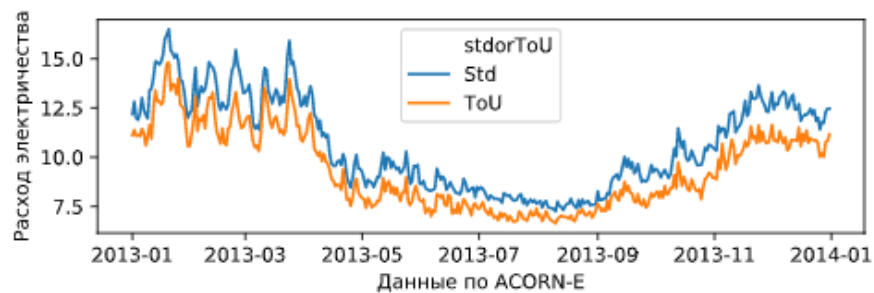
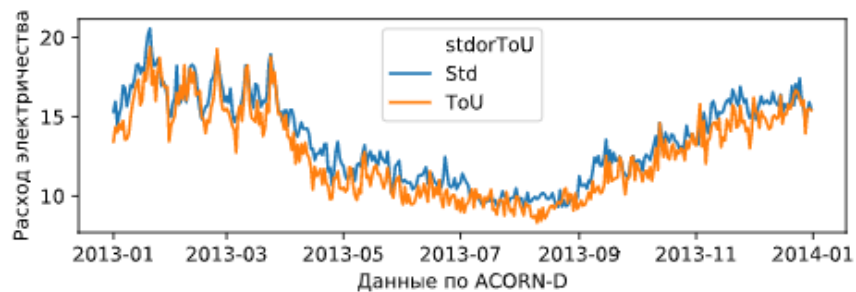
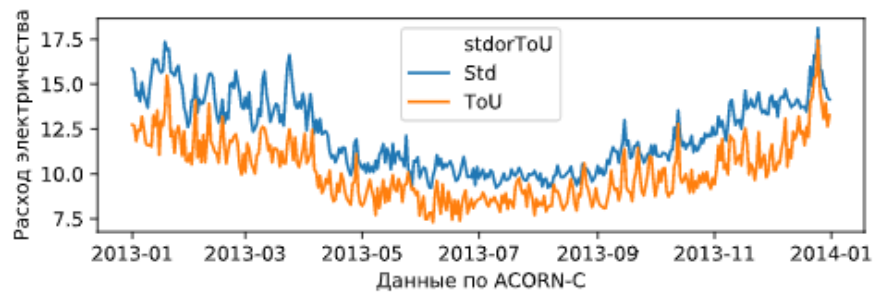
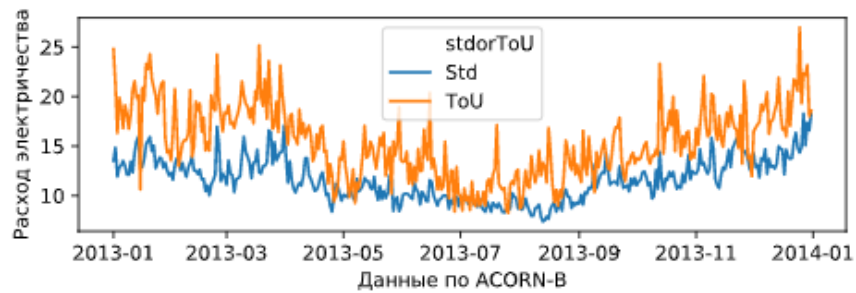
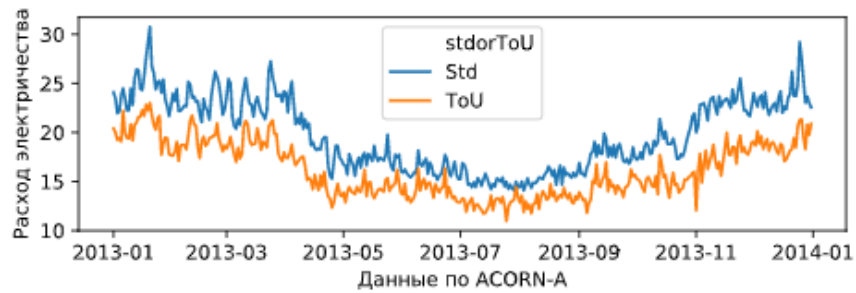
Посмотрим как распределяется потребление в рабочие и выходные дни:

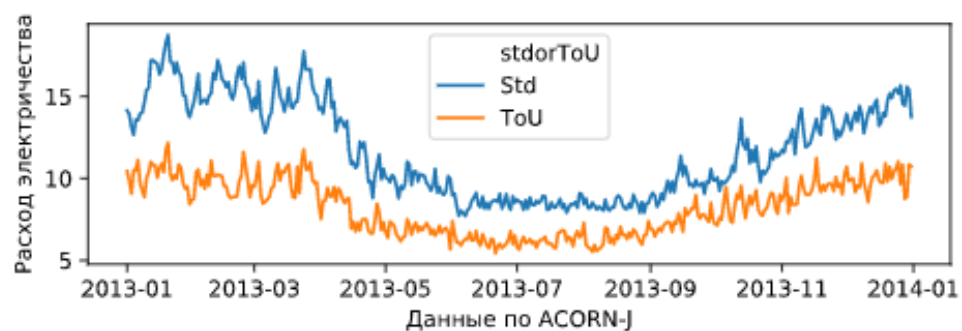
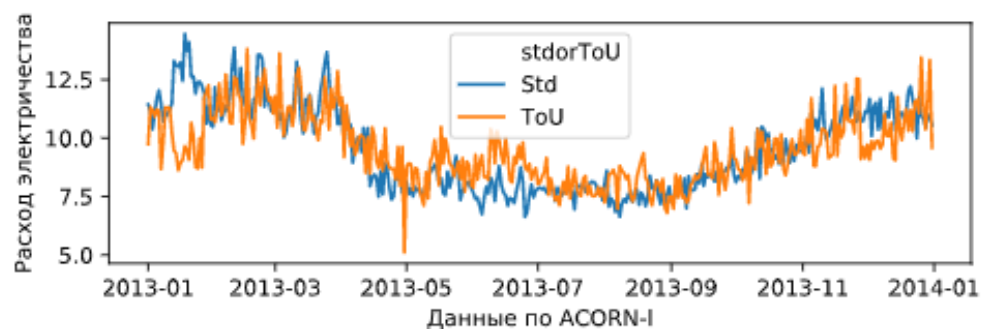
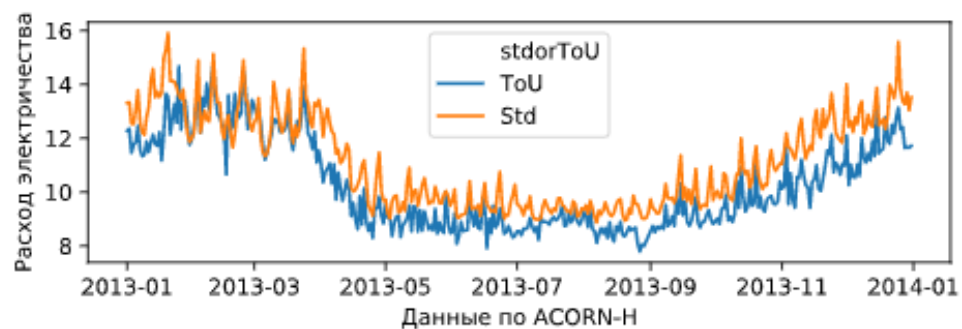
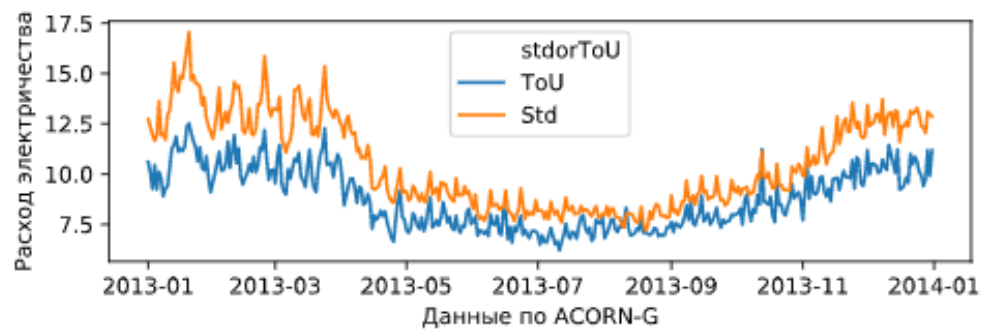
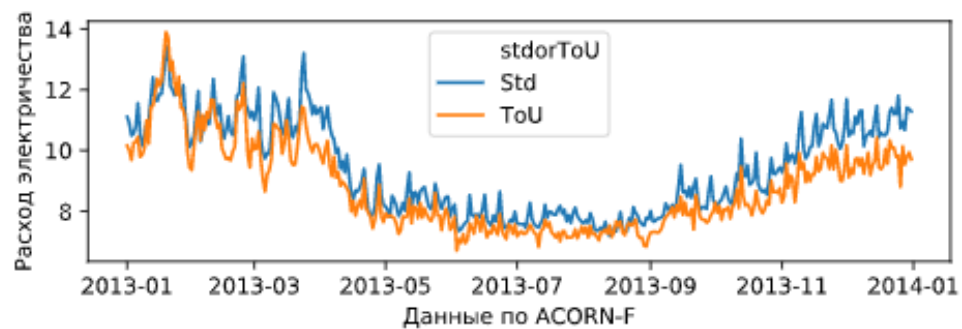


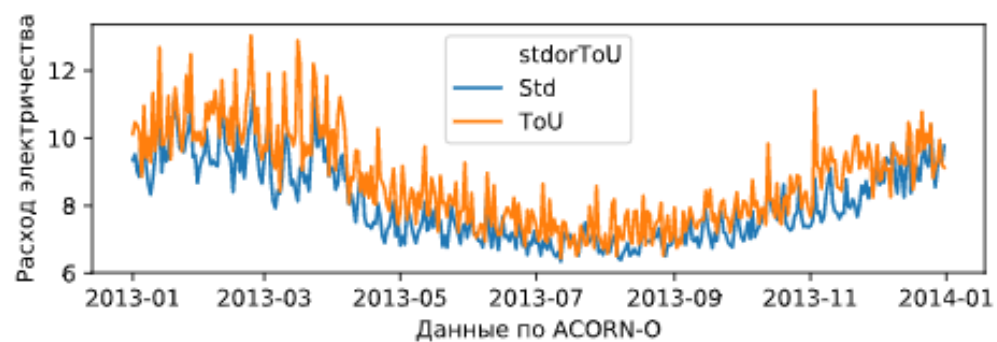
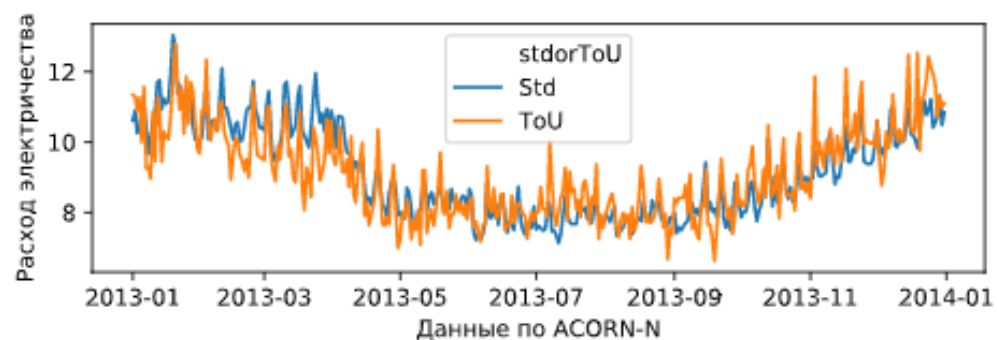
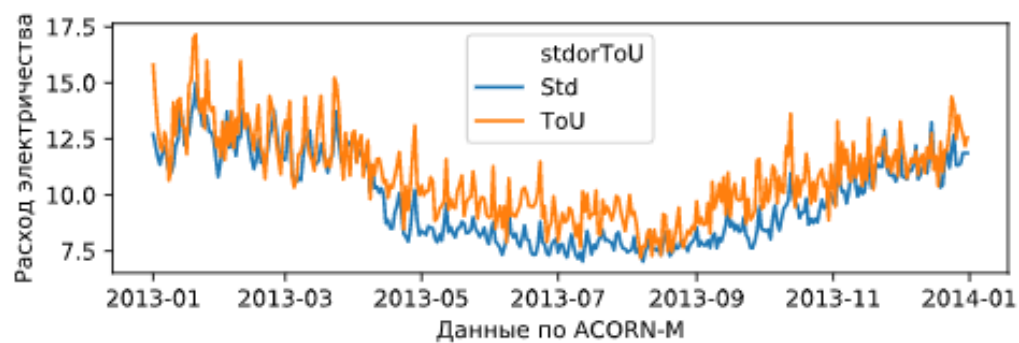
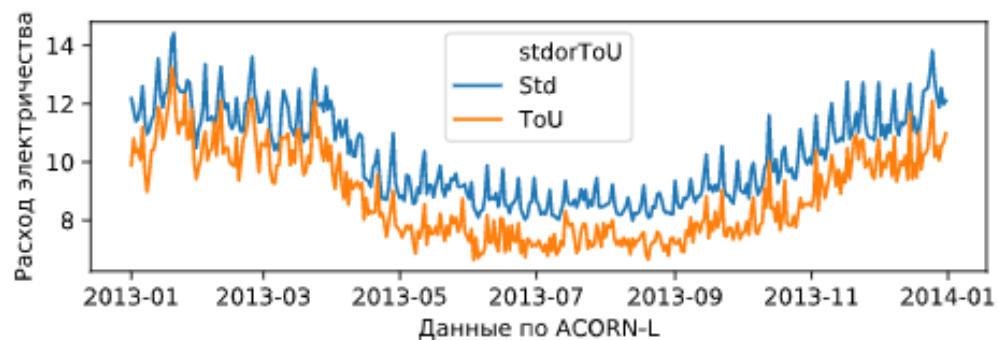
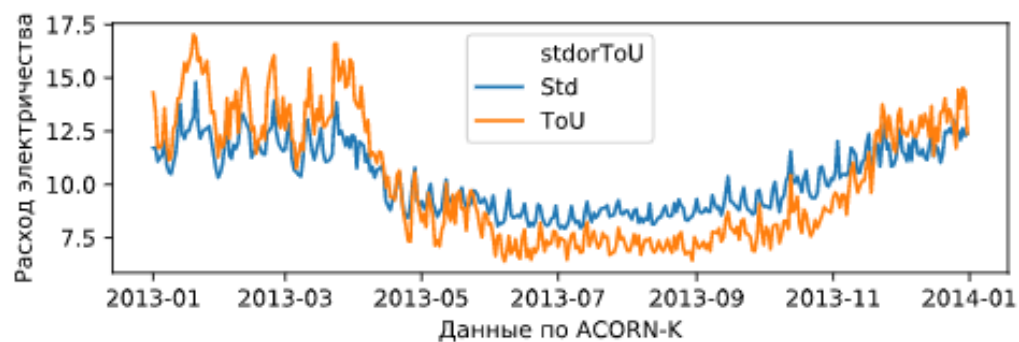


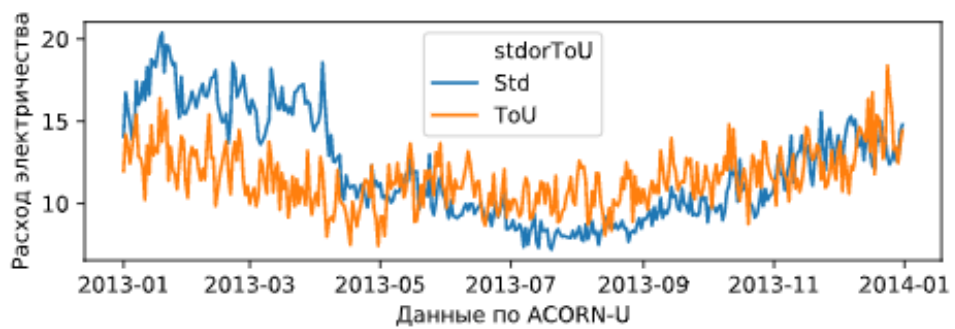
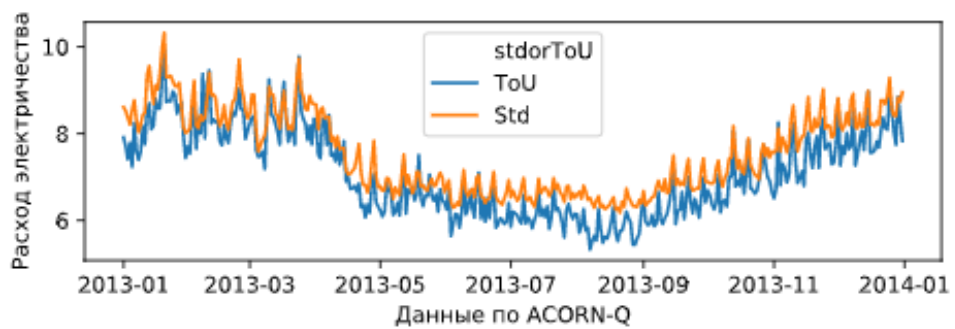
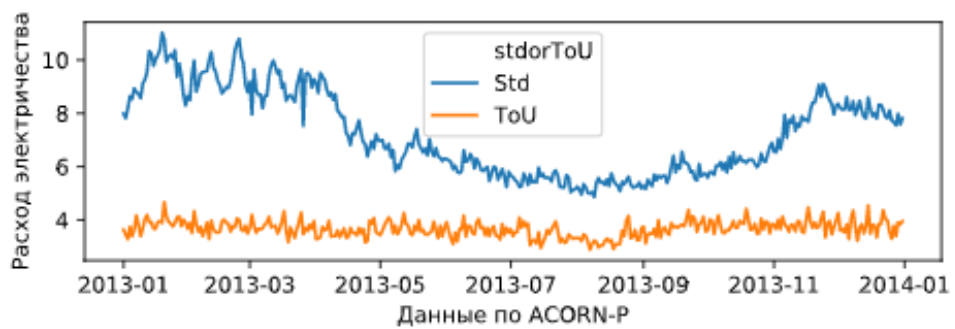
В целом в выходные дни потребление днем более равномерно, чем в будни, особенно характерно воскресенье - потребление в течении дня с 8:00 до 20:00 равномерное, пик 20:00 не так выражен.

Проанализируем потребление ЭЭ отдельно по каждому из Asorn.

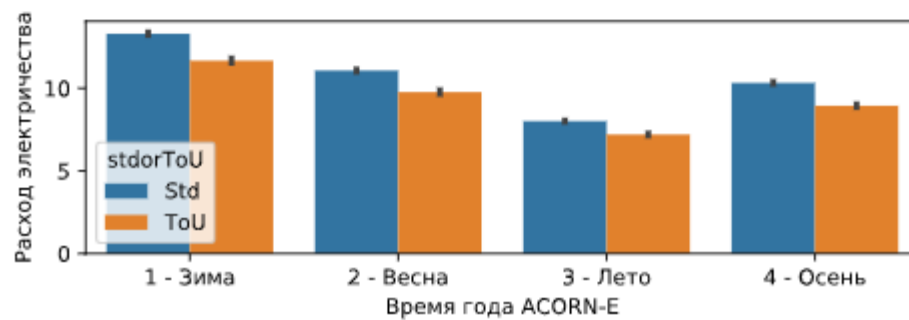
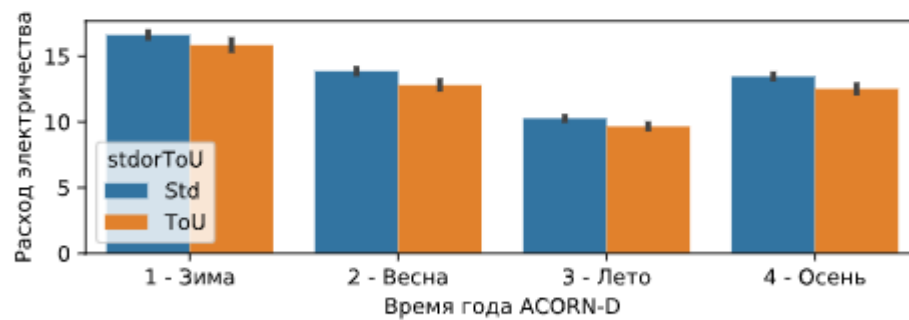
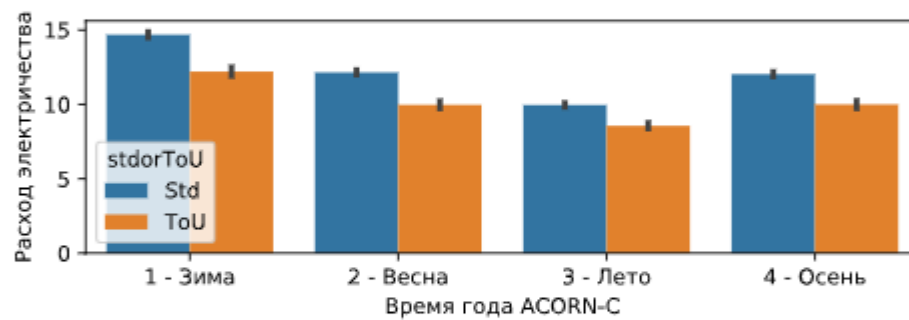
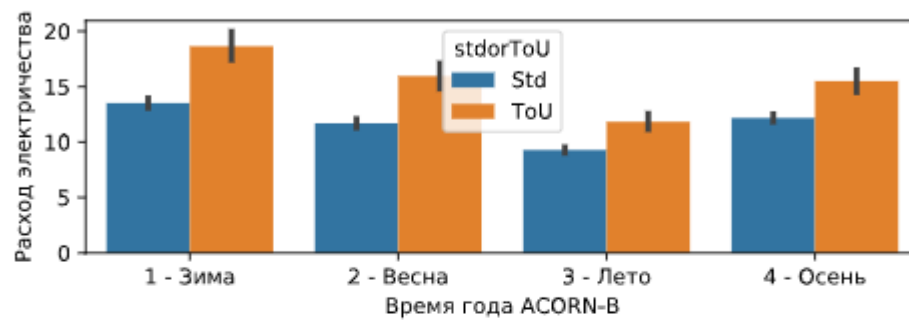
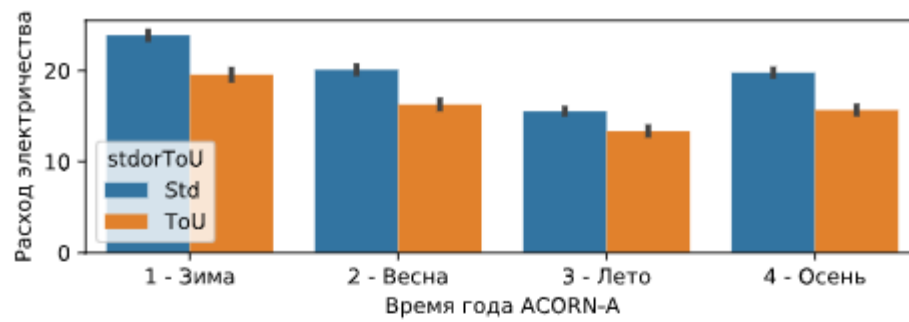


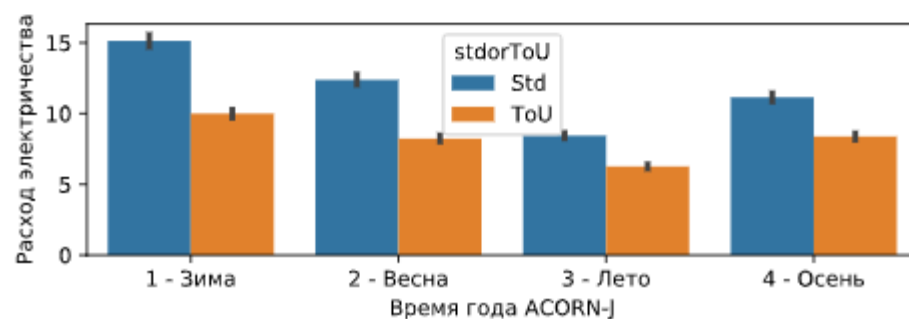
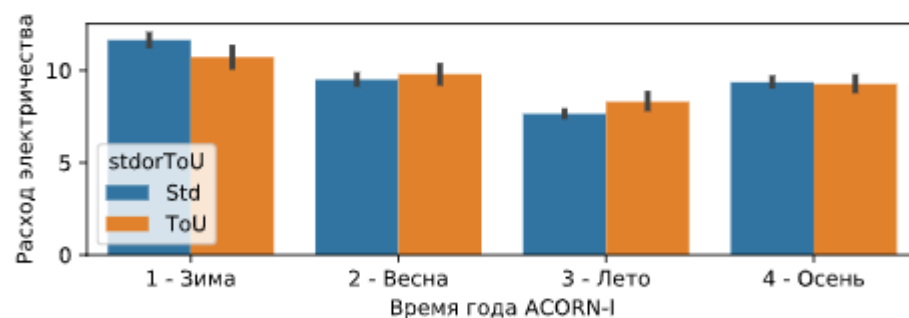
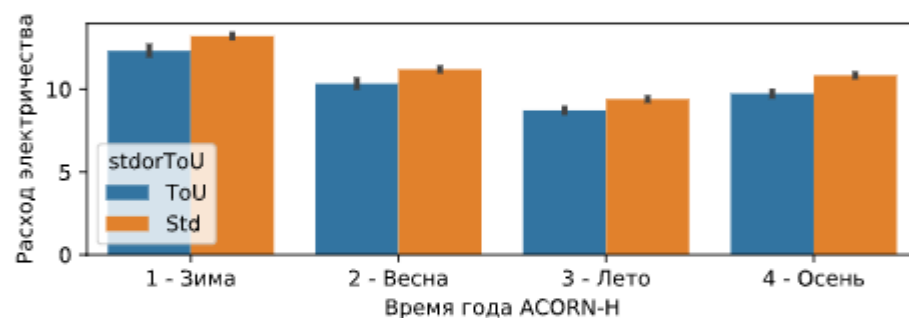
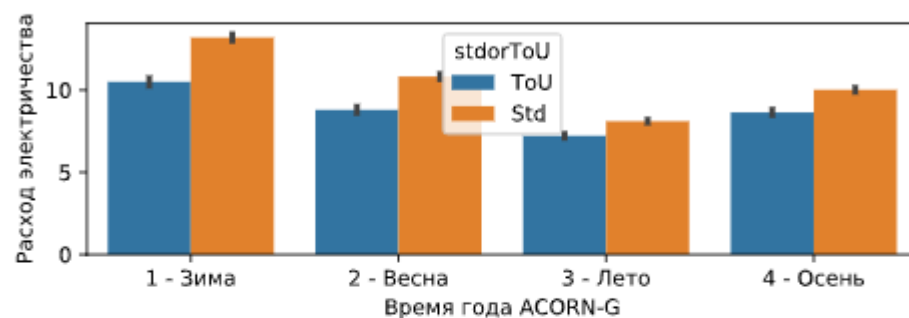
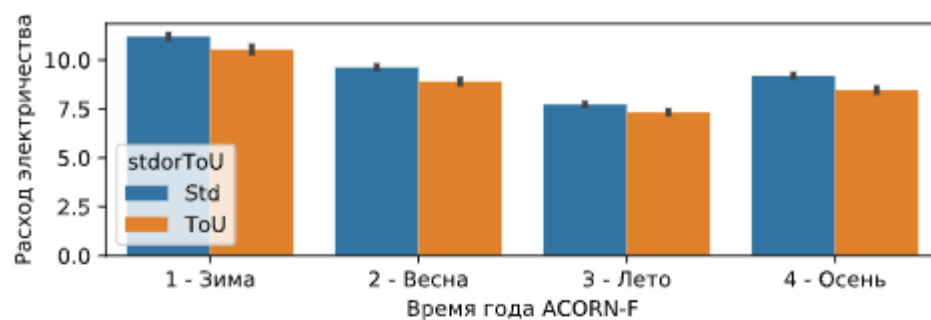


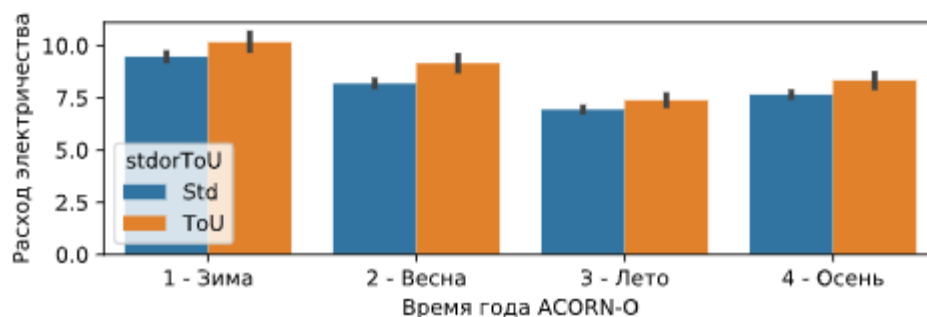
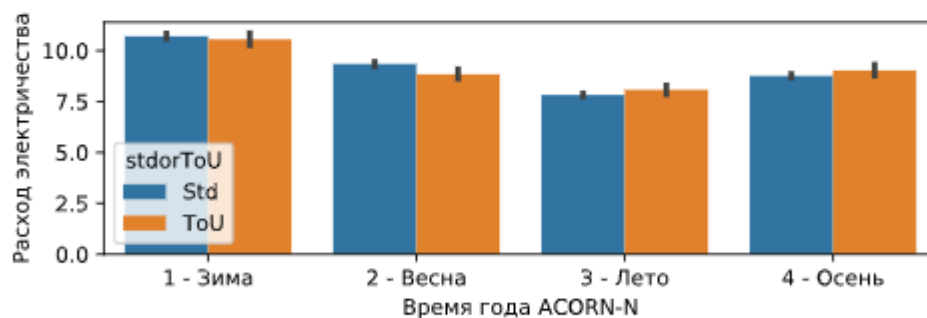
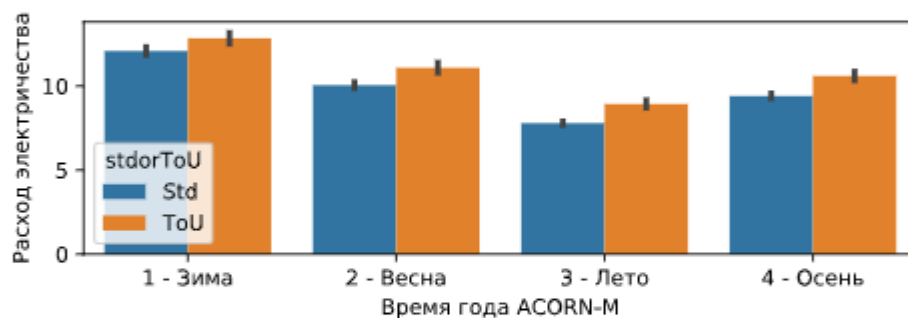
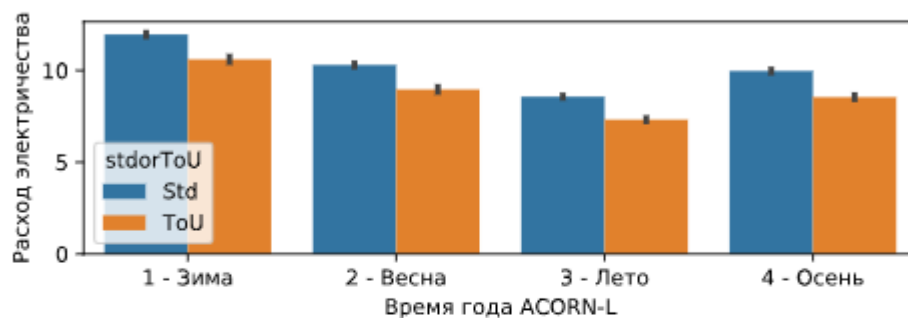
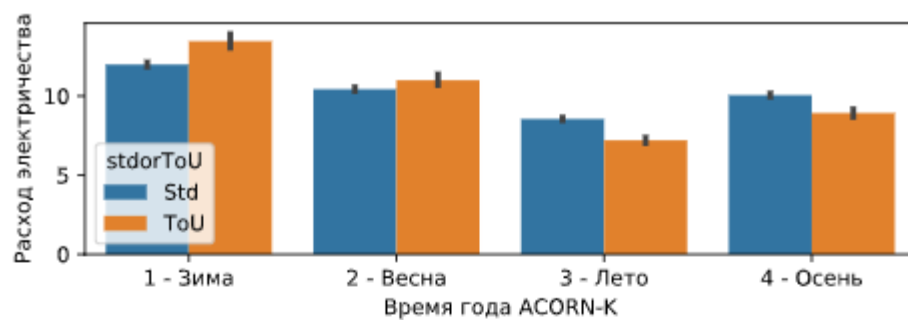


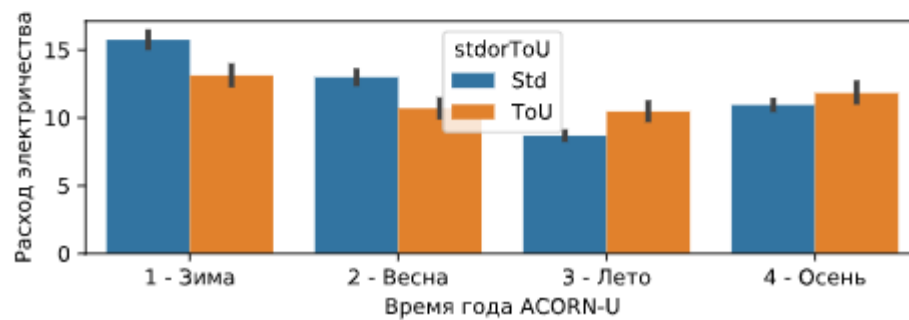
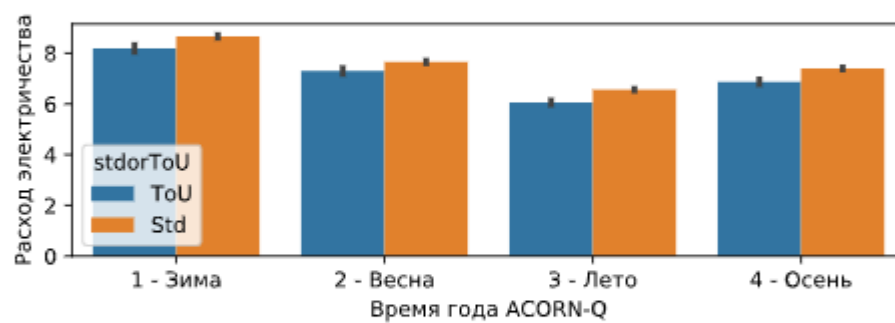
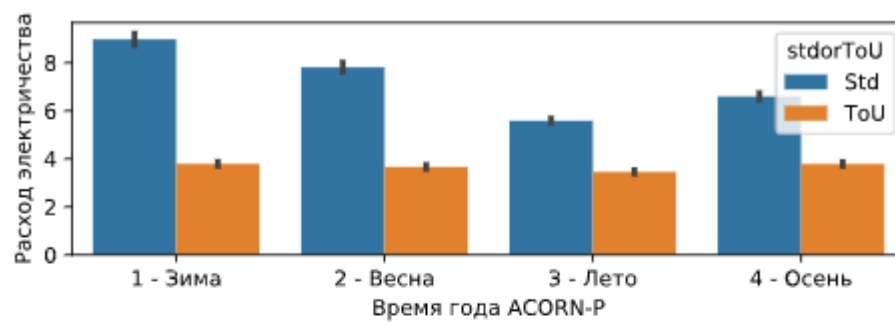


Потребление электроэнергии в зависимости от сезона для различных групп Acorn

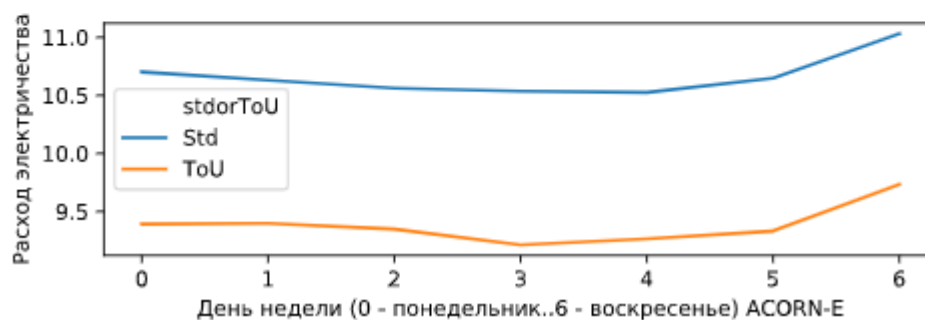
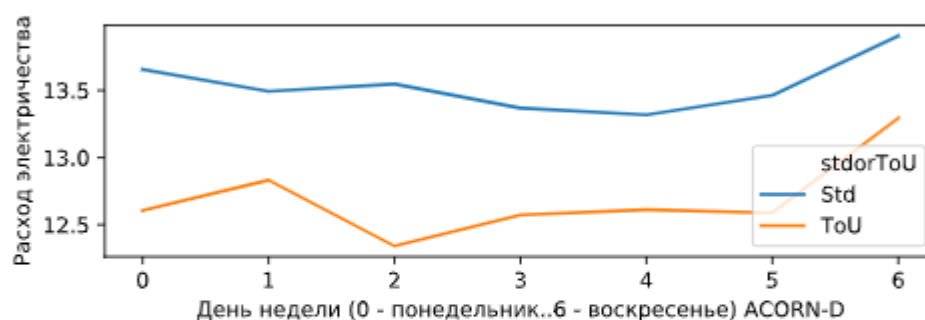
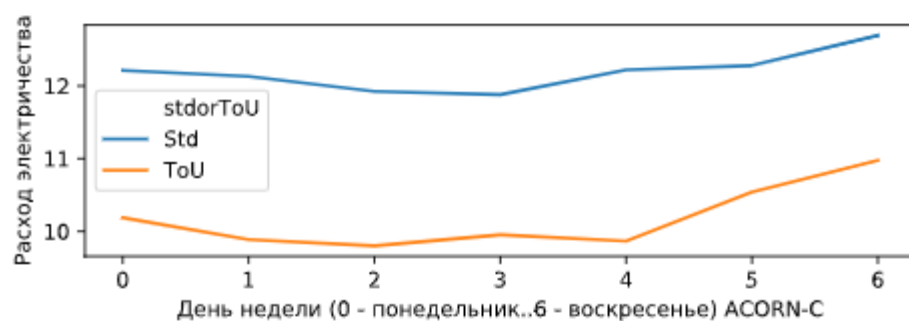
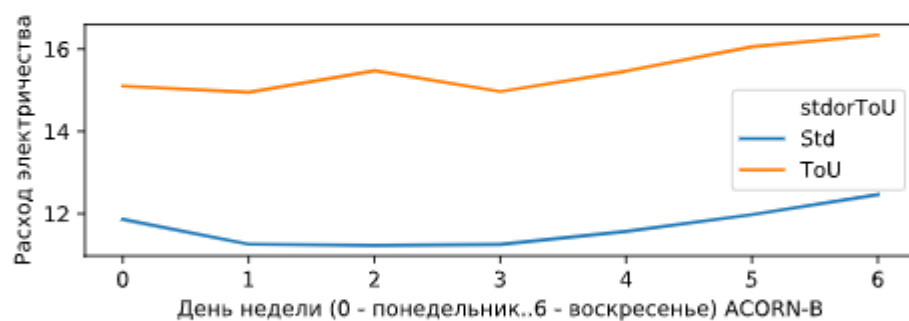
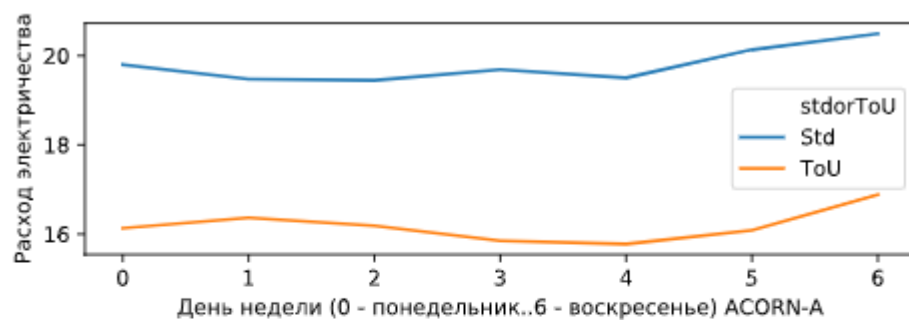


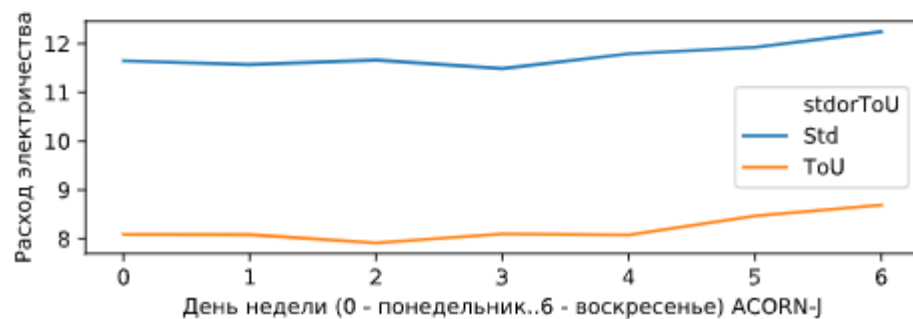
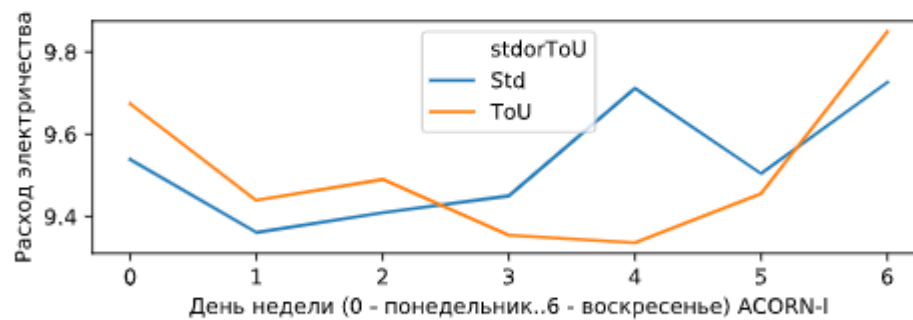
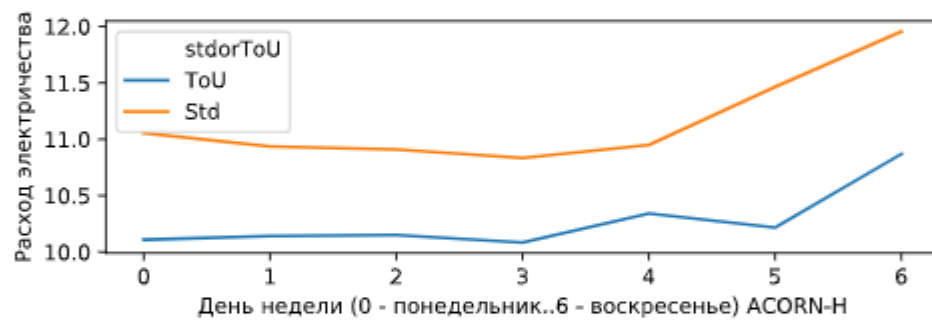
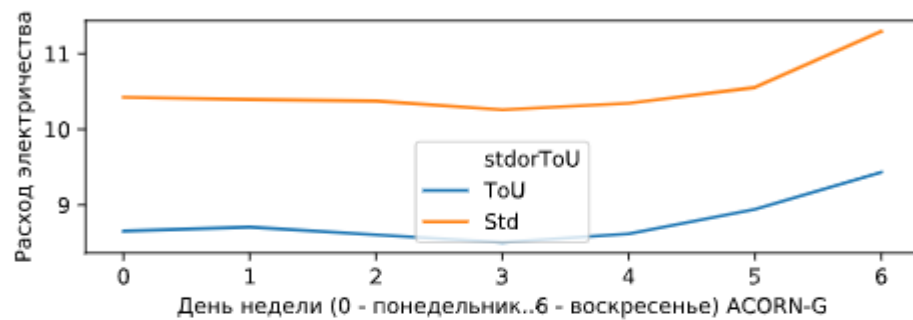
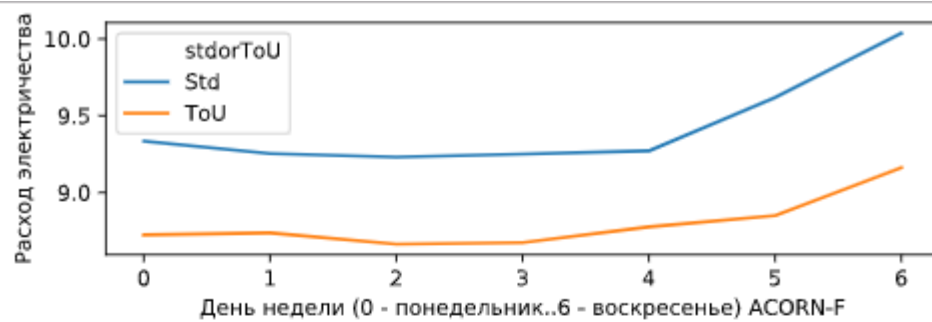


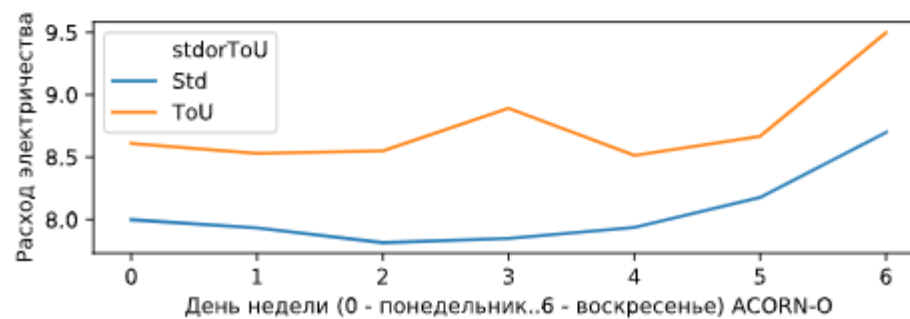
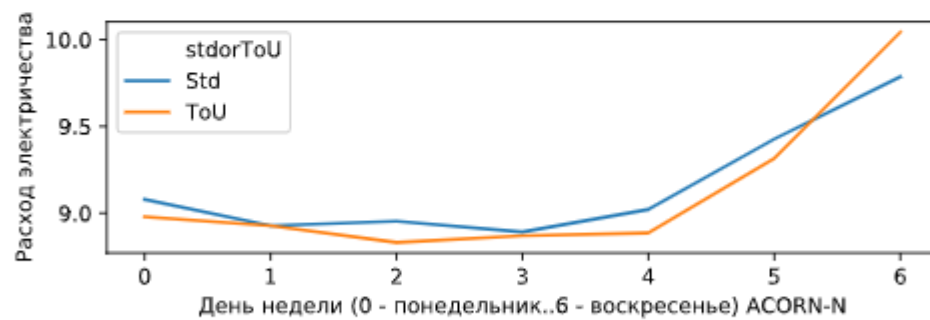
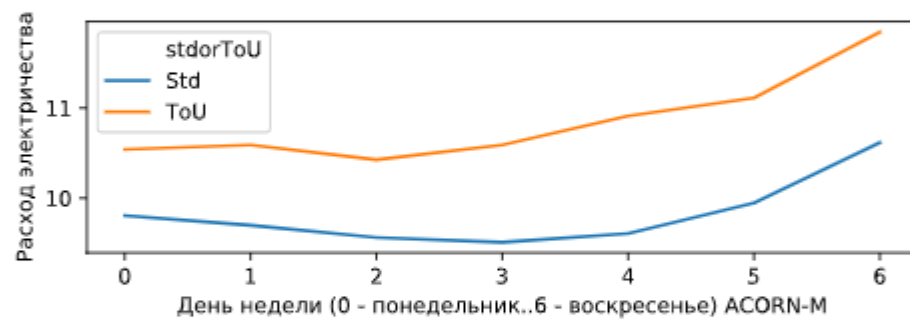
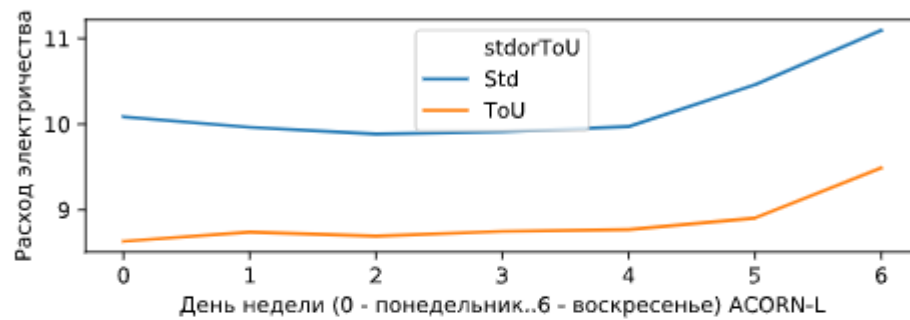
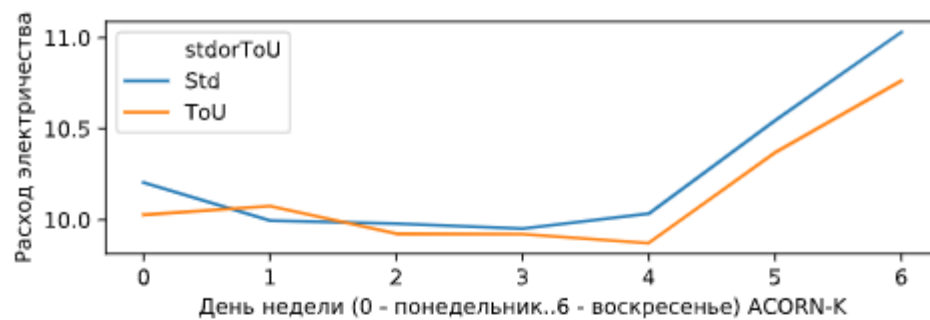


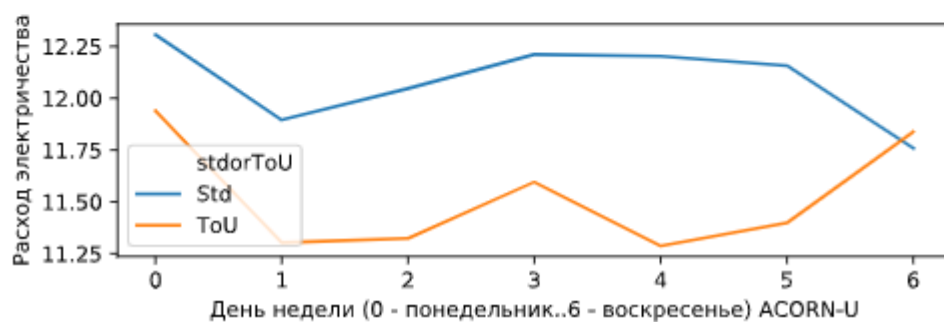
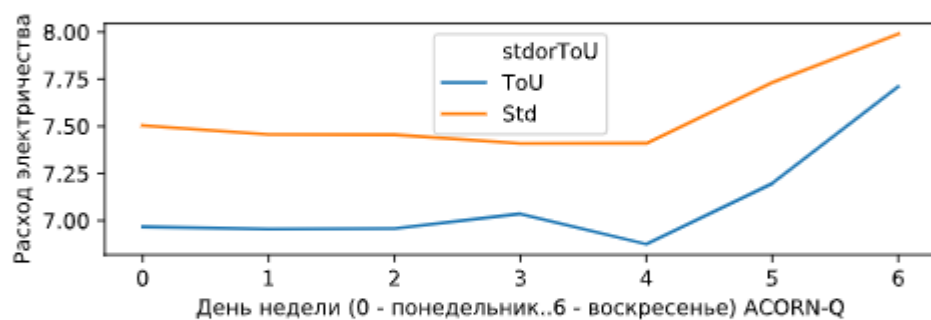
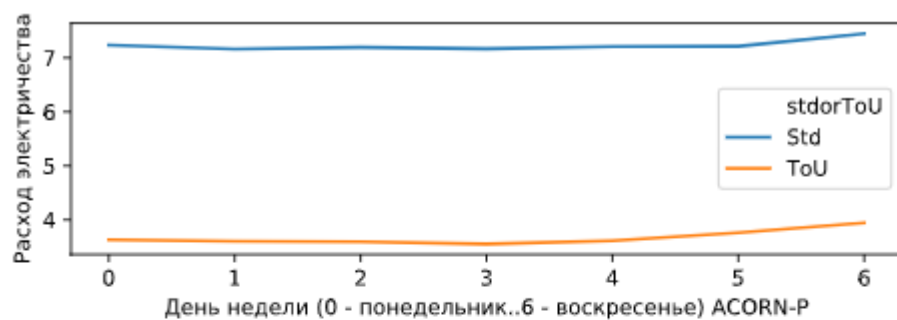


Распределение потребления по дням недели для различных групп Acorn

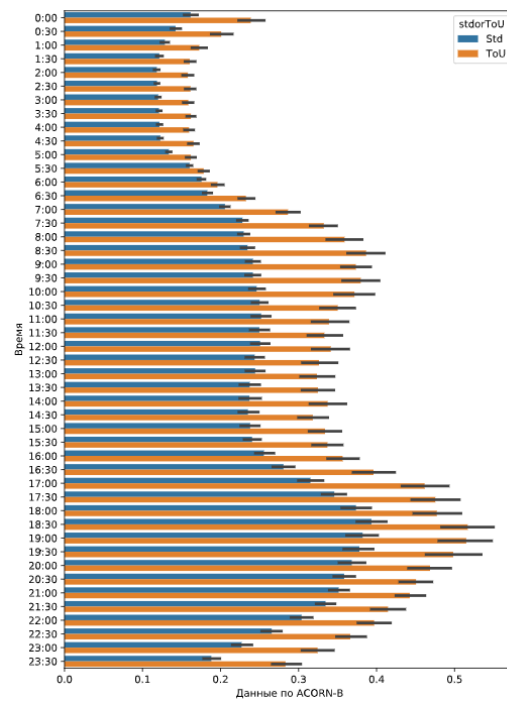
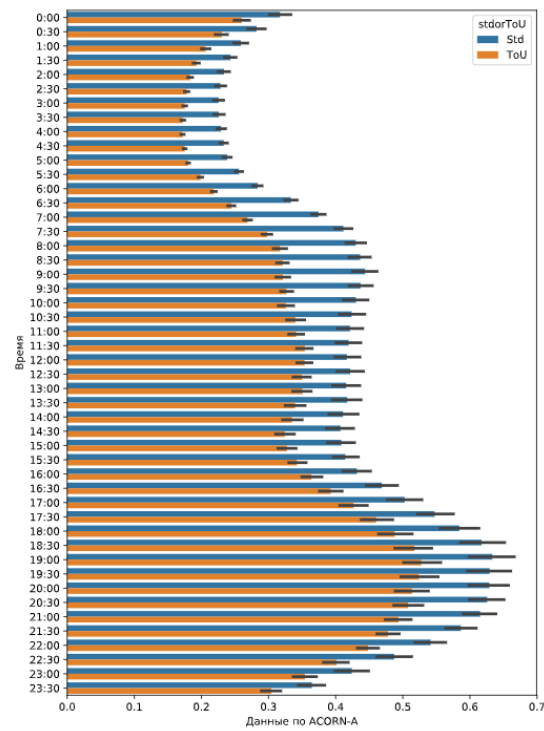


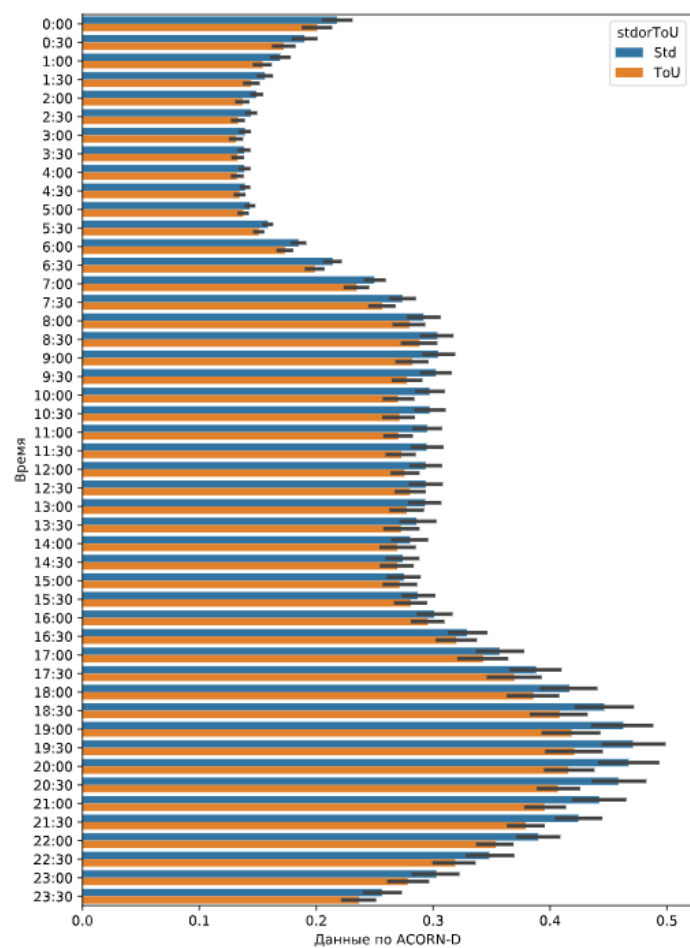
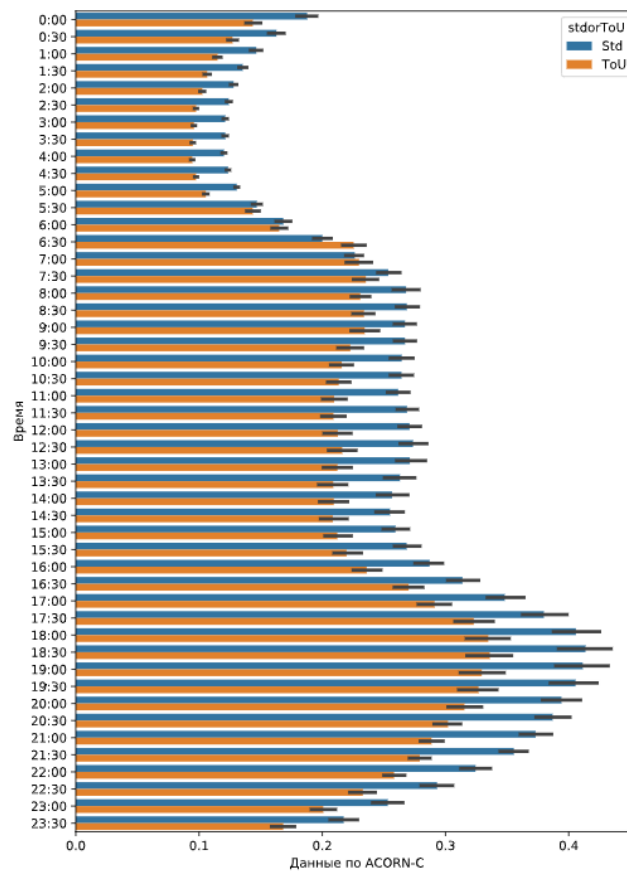


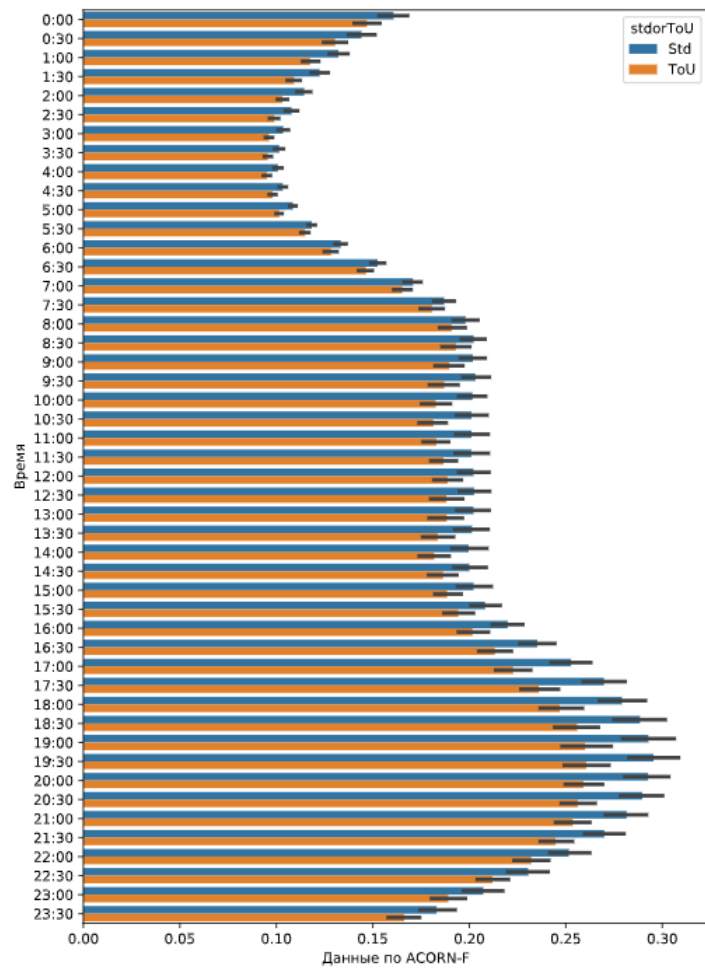
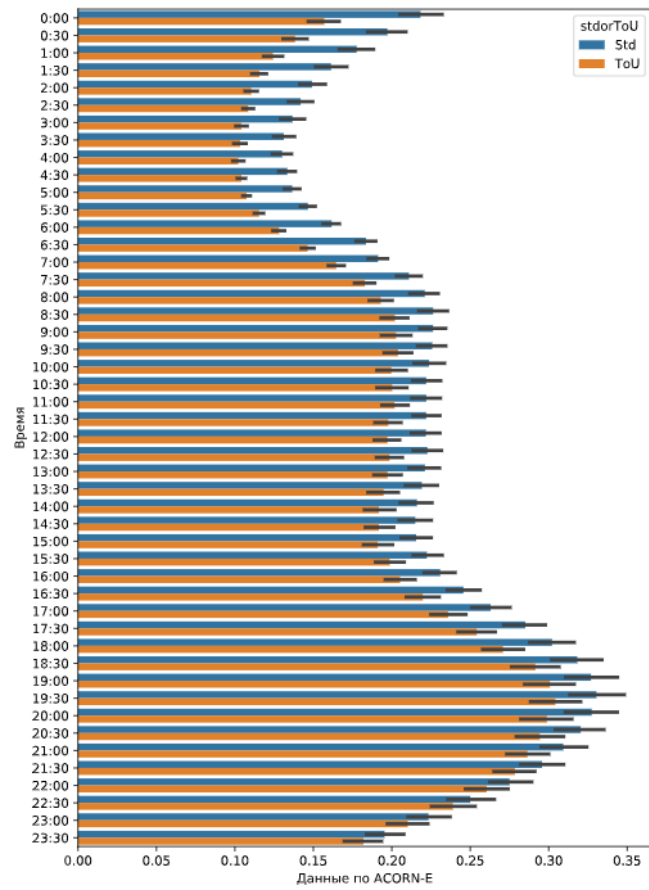


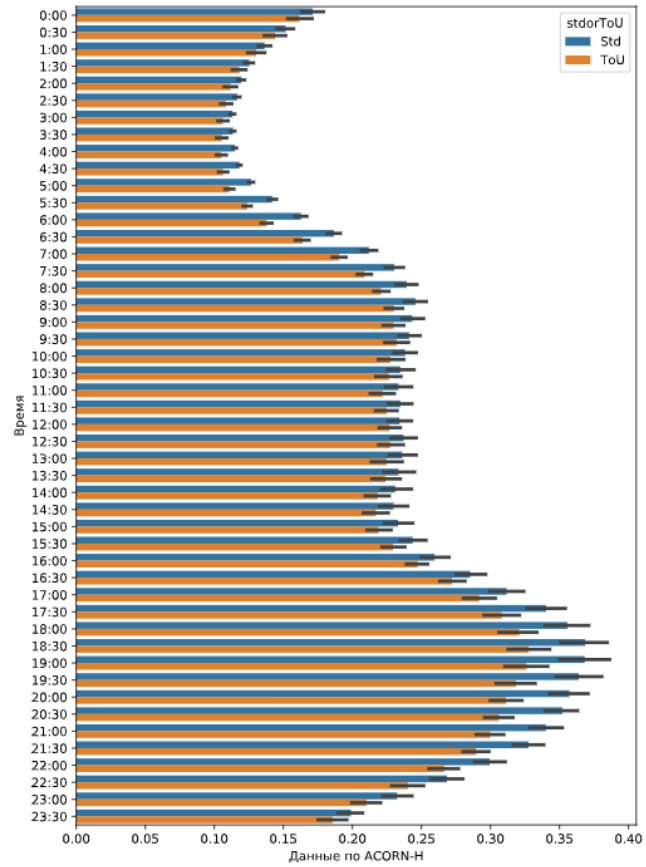
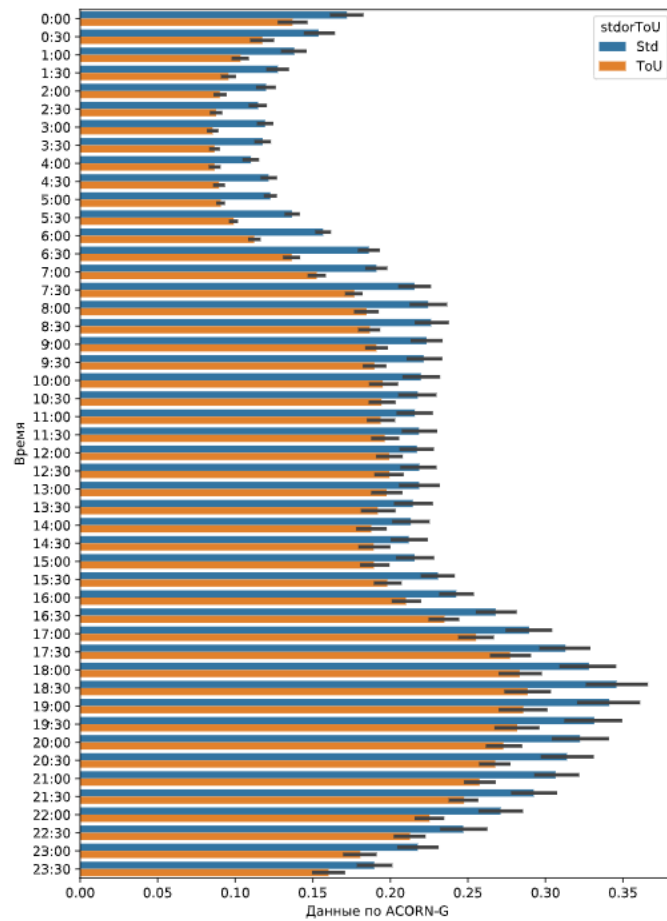


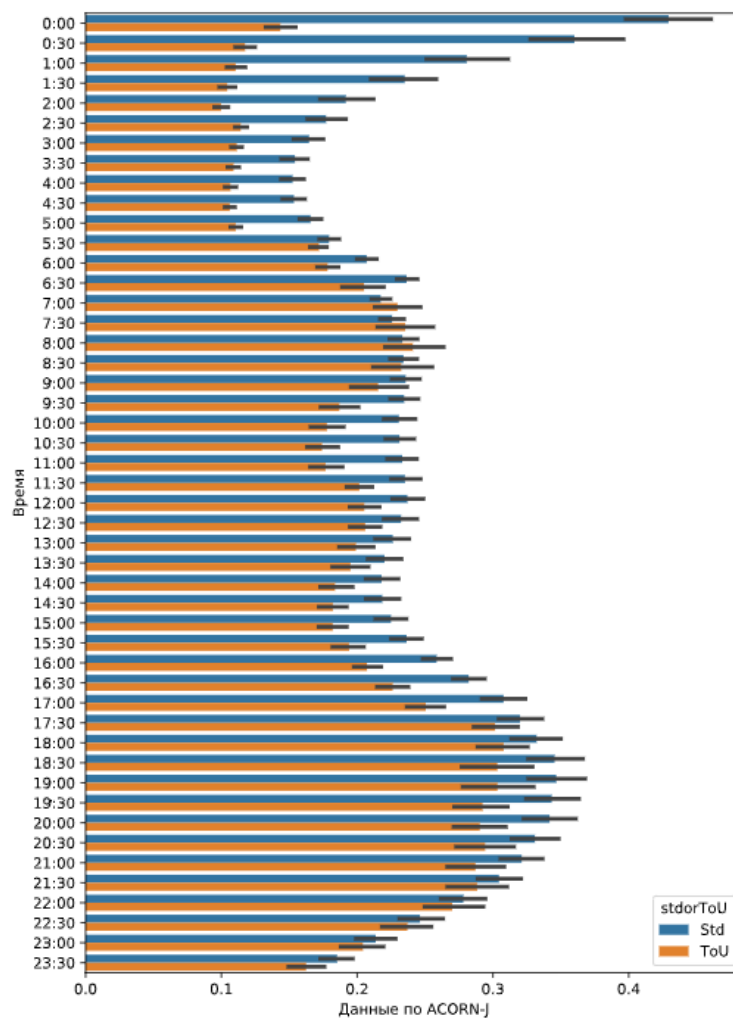
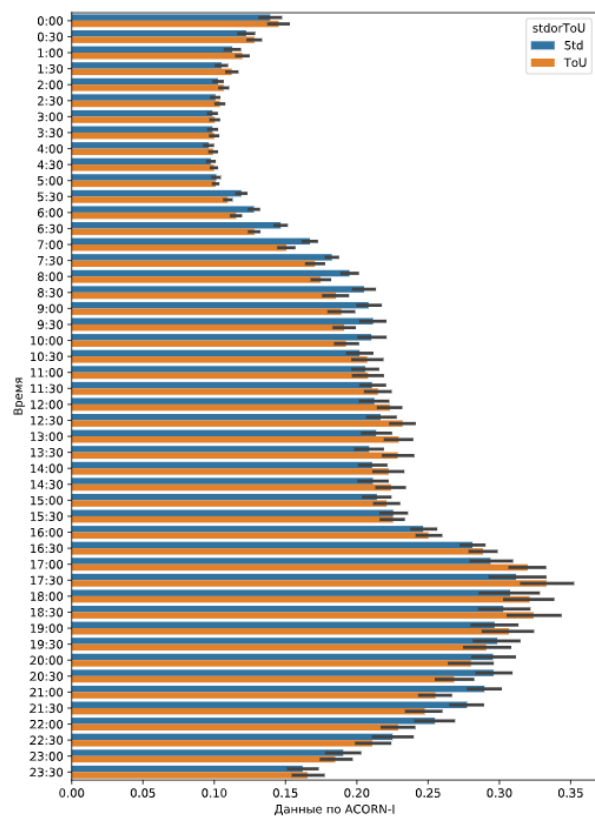
Почасовое распределение потребления электроэнергии по различным группам Acorn

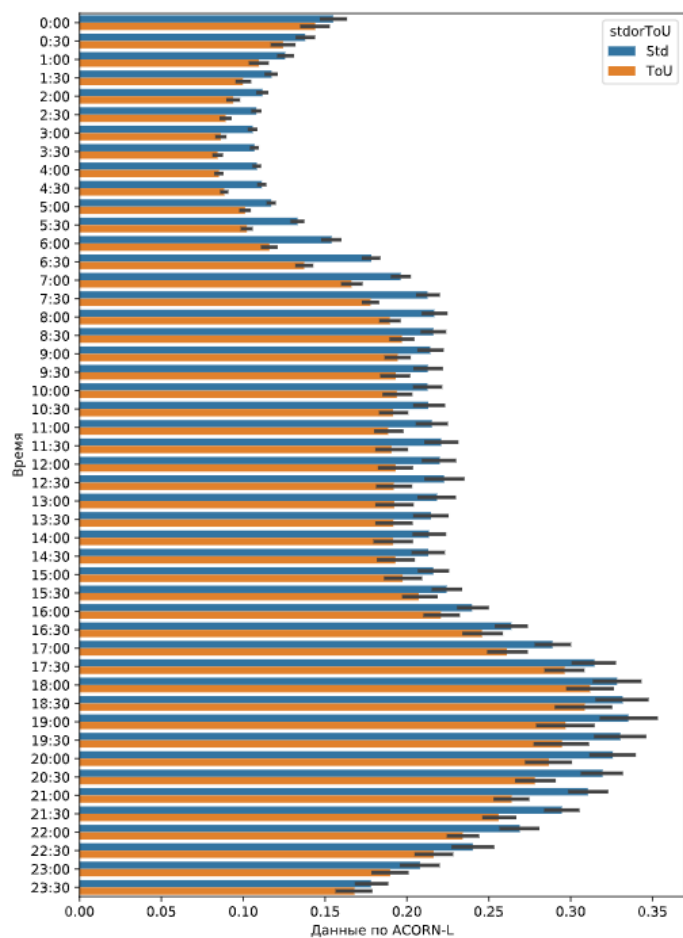
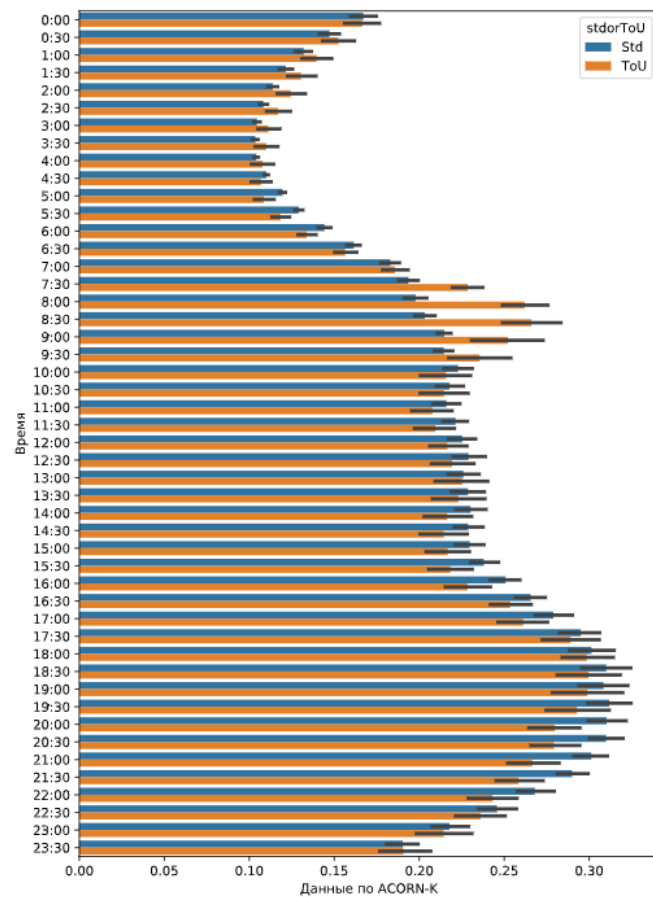


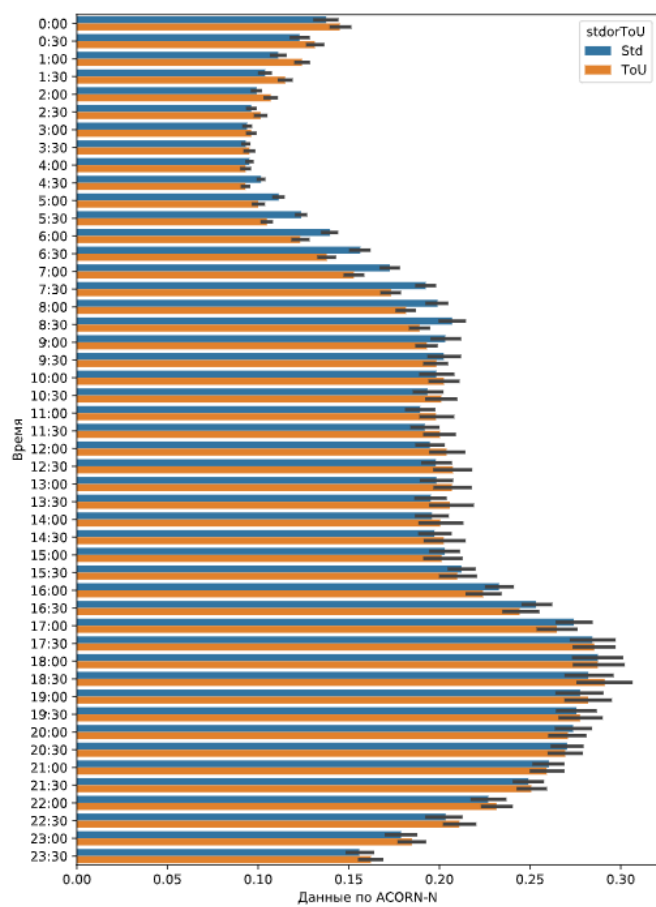
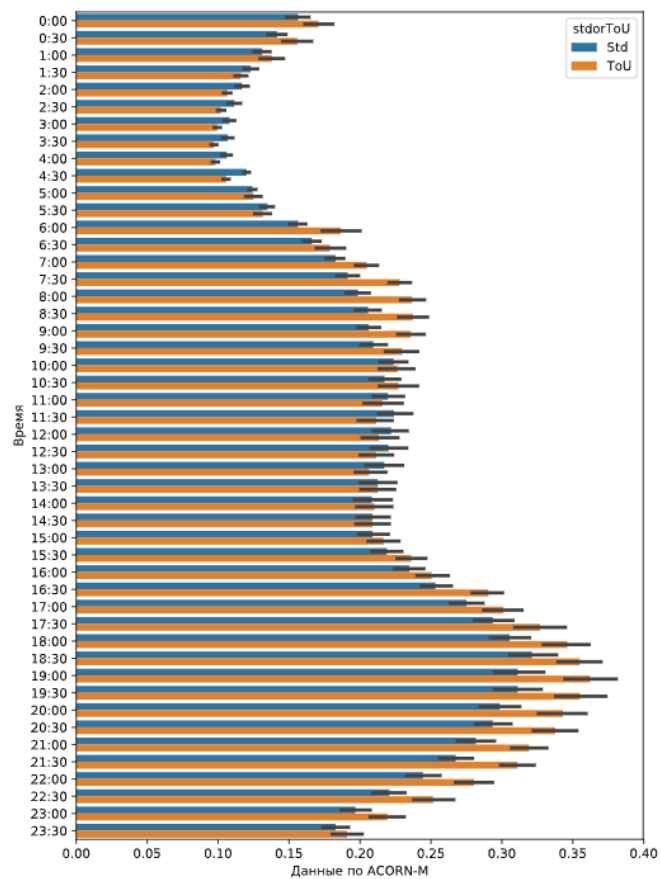


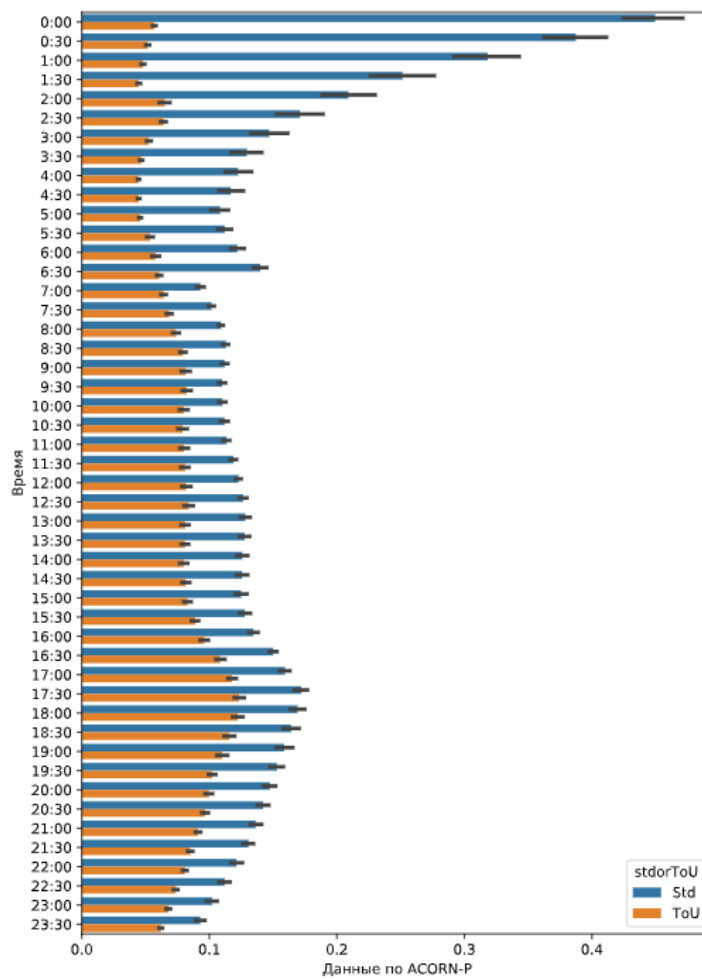
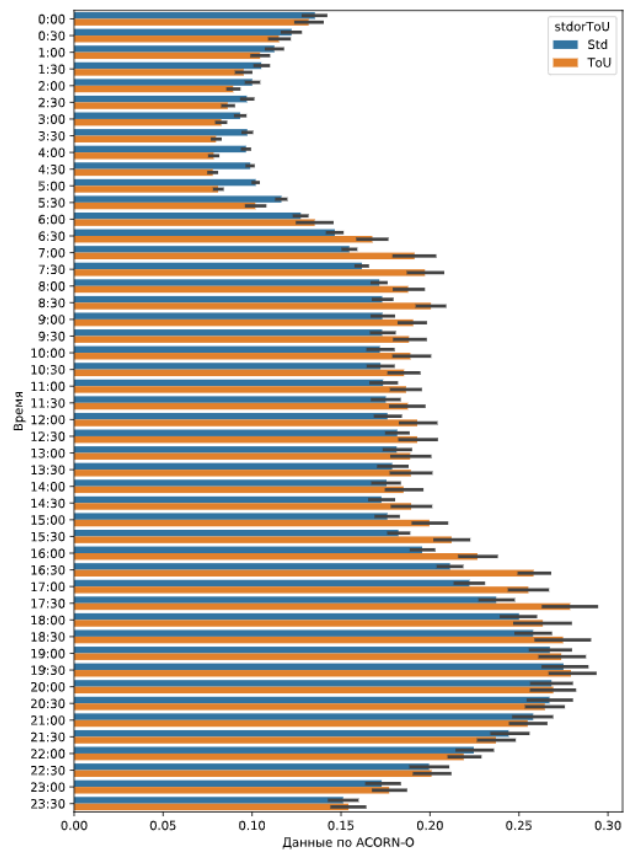


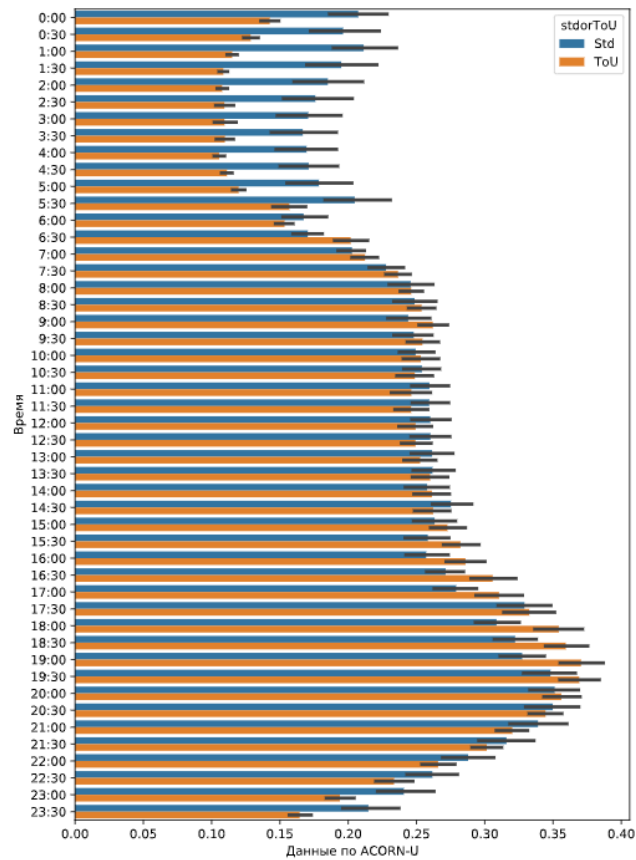
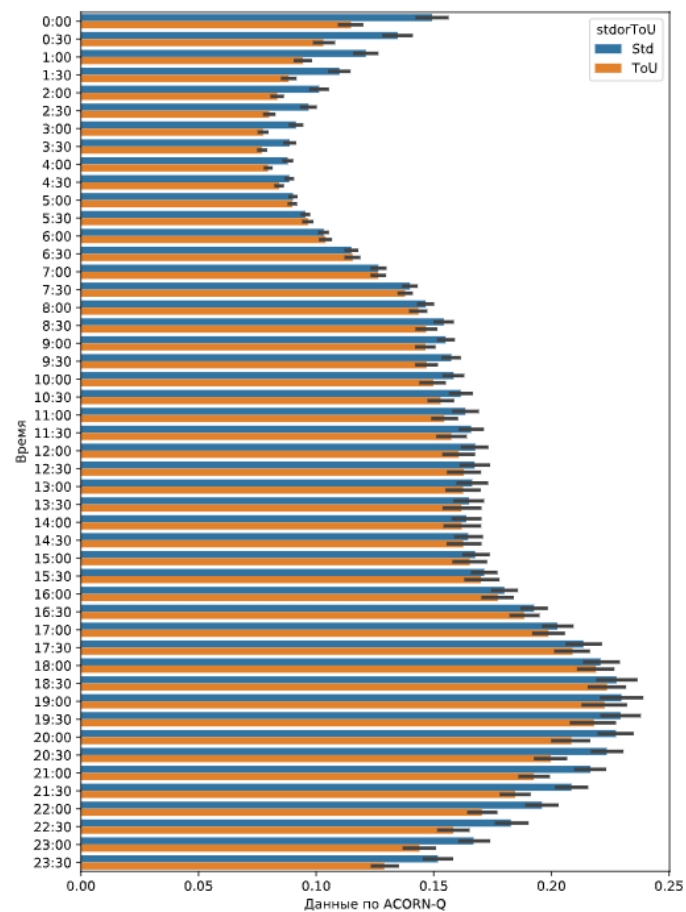












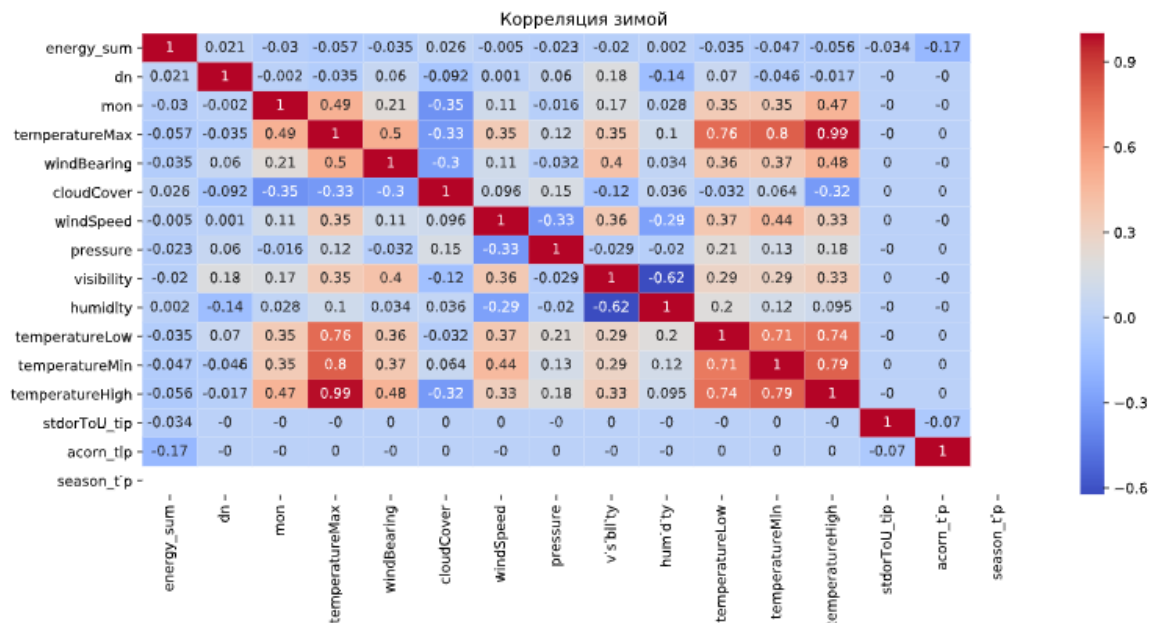
Анализ зависимости количества потребленного электричества от погодных факторов

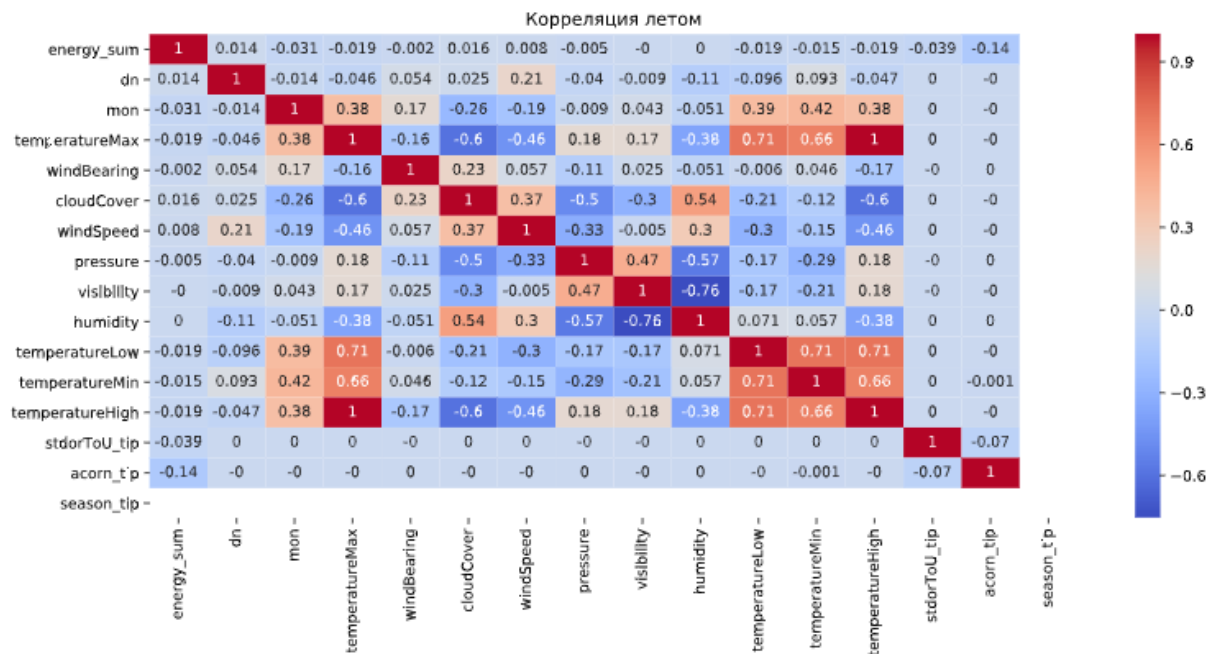


По графику видно, что существует связь между потреблением электроэнергии и временем года.



Посмотрим на корреляцию влияющих признаков и потребления электроэнергии:





Вывод: существует связь между потреблением электроэнергии и временем года. Однако нельзя сказать, что электричество идет на обогрев, т.к. слишком малая в среднем мощность чтобы что-то обогревать. Потребление ночью и зимой и летом одинаково: 0,1, а вот дневное различается. Сезонная разница зависит от того что зимой день короткий, и нужно больше электричества для освещения. Конечно есть корреляция с температурой и потреблением: 0,18, но это именно потому что это сезонная корреляция, зимой холодно, но еще и темно. Если посмотреть корреляция температура-потребление только зимой, то коэффициент всего около 0,04 -0,05 т.е. намного слабее. А летом вообще 0,02.

Выводы:

В данной работе мы исследовали влияние различных факторов (погодных, социальных, экономических) на уровень потребления электрической энергии. Для этого проанализировали исходные данные:

1. выбрали оптимальный период для исследования - 2013 год
2. не учитывали данные, где было малое число показаний счетчиков в течении суток
3. не учитывали данные, где было малое количество дней в году, когда проводились измерения потребления энергии
4. не учитывали неполные данные

Далее, мы проанализировали влияние на уровень энергопотребления следующих факторов:

1. сезон
2. среднесуточная температура
3. дни недели
4. выбранный тарифный план
5. время суток
6. принадлежность к определенной группе (Acorn)

Наибольшее влияние на потребления электрической энергии оказывает среднесуточная температура (сезонность), далее - принадлежность к определенной группе (Acorn).

Так же выяснили, что пик потребления приходится на воскресенье, более +5% от среднего, в течении других дней недели потребление более равномерное. Сильно зависит среднее потребление от сезона, до -20% летом и +20% зимой. Максимальное потребление достигается в декабре, а минимум в августе.

Так же выяснили, что пик потребления электричества приходится на вечернее время с 18:30 до 20:00 часов, наименьшее потребление электроэнергии с 02:00 до 04:00 часов. Разница в потреблении примерно трехкратная. Причем, это верно, как для стандартного счетчика, так и для многотарифного. Многотарифная оплата за электроэнергию призвана в первую очередь стимулировать потребителей потреблять энергию в часы наименьшей нагрузки и снижать потребление в часы пик, т.е. выравнивать потребление в течении суток. В целом, люди, которые установили многотарифный счетчик, электроэнергию используют более экономно.