Постановка задачи

1. Задача регрессии с числовыми данными (объем потребления коммунальных услуг, объем инвестиций), с категориальными данными (внешние факторы, такие как климат и т.д.) с прогнозом коммунальных расходов (числовые данные) в горизонте 3 лет.
2. Создание «удобного» интерфейса с возможностью управления параметрами.
3. Качество и наглядность визуализации данных в пользовательском интерфейсе.
4. Возможность классификации объектов анализа по степени энергоэффективности.

Уточнения по задачам

1. Язык программирования python.
2. Исключить влияние дохода организации на прогноз расходов коммунальных платежей.
3. Климатические показатели, данные о тарифах на коммунальные услуги, налогов и инфляции берём из открытых источников.

Решение

Отсутствует соответствия для разных годов, придется дополнять данные. Берем 2019 год для начала. Сбербанк индекс для регионов.

Данные Extensions

Код статьи, код вида расходов, наименование вида расходов объединяем в один с помощью binary encoder.

Строки повторяются только с разными улицами. Нужно узнать сколько улиц и разделить данные на строки.

Год сметы можно выкинуть.

Добавить первую дату для января.

Убрать итоговые строки.

Направление расходов не обязательно.

Инфляция экономических расходов не совпадает с инфляцией. Разные тарифы для коммерческих и не коммерческих видов (гаражи и т.д.).

Целостность здания влияет на вбухивание денег на ремонт.

Данные KC\_TR\_KR

KR. В том числе проектирование и капремонт строчки выкидываем.

Итого Дальневосточное ГУ Банка России г. Владивосток строчки по итогу для улиц выкидываем, так как их можно получить из вышестоящих, содержащих улицы.

KS. Пояснения по состоянию закупочных процедур и работ на объектах выкидываем.

Пользовательский интерфейс

Пользовательский интерфейс с управлением параметров и выводом визуализации статистики.

Сделать основную задачу для крупных регионов и только потом для городов дополнительно.

Поискать данные для погоды и тд, коммуналку готовые датасеты.

Скрипт, пандос. Синтетика+.

Если используем сторонние данные упоминать об этом на презентации.

Streamlit. Презентацию после стоп кода.

На презент 5 минут вопросы и тех часть до 10 слайдов,рекоменды после второго чека.