**2. Доступное моделирование и связи в данных**

Возможные подходы к моделированию:

* *Классификация.* Данные содержат столбец «Current Status», который позволяет разделить разрешения, например, на те, что завершены («complete»), и на прочие. Это позволяет построить задачу бинарной или мультиклассовой классификации. В качестве признаков можно использовать тип разрешения, оценочную стоимость, количество предлагаемых этажей, район (Supervisor District), почтовый индекс и другие.
* *Регрессия.* При наличии количественных переменных, таких как «Estimated Cost» и «Revised Cost», можно попытаться предсказывать окончательную стоимость проекта или продолжительность его выполнения.

Наличие связей:

В датасете наблюдаются взаимосвязи между:

* Типом разрешения и оценочной стоимостью.
* Количеством этажей (существующих или предлагаемых) и стоимостью/типом проекта.
* Географическим положением (Supervisor District, Zipcode) и типами или сложностью проектов.

Такие взаимосвязи позволяют применять модели, способные работать как с числовыми, так и с категориальными признаками (например, Random Forest, Gradient Boosting).

**3. Возможные верхнеуровневые задачи**

На основе данного датасета можно решать следующие задачи:

* *Классификация статуса разрешения*: Прогнозирование, будет ли разрешение завершено успешно или возникнут задержки/проблемы.
* *Регрессия по стоимости или срокам*: Предсказание окончательной стоимости проекта или времени до его завершения на основе оценочных данных и характеристик объекта.
* *Анализ временных трендов*: Выявление сезонных или временных зависимостей в подаче и выдаче разрешений.
* *Кластеризация*: Группировка объектов по схожим характеристикам для выявления паттернов или сегментации регионов/типов проектов.
* *Оценка рисков*: Выявление факторов, способствующих задержкам или перерасходу бюджета, что может помочь в оптимизации процессов выдачи разрешений.

**7. Применение модели в реальной бизнес-ситуации и выбор бизнес-метрик**

Применение модели:

В реальном проекте модель можно использовать для прогнозирования успешности выполнения строительных проектов или своевременности выдачи разрешений. Например, если модель предсказывает, что определённое разрешение с высокой вероятностью не будет выполнено вовремя, управленцы могут заблаговременно вмешаться — перераспределить ресурсы, провести дополнительный аудит или оптимизировать процессы.

Бизнес-метрики для контроля модели:

* *Точность (Accuracy, Precision, Recall, F1-score)*: Эти метрики позволяют оценить, насколько правильно модель прогнозирует статус разрешения.
* *Время отклика (Time-to-Decision)*: Насколько быстро модель может обработать заявку и предоставить прогноз — важный показатель для оперативного принятия решений.
* *ROI (Return on Investment)*: Оценка экономической выгоды от применения модели (например, снижение затрат за счёт предупреждения задержек и перерасхода бюджета).
* *Показатели качества обслуживания*: Например, удовлетворённость клиентов и подрядчиков, что может быть косвенным индикатором эффективности работы модели.
* *Финансовые показатели (NPV, ROI, EBITDA)*: Если модель позволяет предотвратить перерасход средств или оптимизировать распределение ресурсов, эти метрики помогут оценить её влияние на бизнес-процессы.