

Бриф учебного кейса «Модель прогнозирования стоимости жилья для агентства недвижимости»

Ваша задача — разработать модель, которая позволила бы агентству недвижимости обойти конкурентов по скорости и качеству совершения сделок.

[Датасет для работы](#)

Вам необходимо:

1. Провести разведывательный анализ и очистку исходных данных. Обратите внимание, что данные в таблице реальные: в результате во многих признаках присутствуют дублирующиеся категории, ошибки ввода, жаргонные сокращения и т.д. Вам предстоит отыскать закономерности, самостоятельно расшифровать все сокращения, найти синонимы в данных, обработать пропуски и удалить выбросы.

Постарайтесь уделить этапу очистки максимум внимания, так как это напрямую повлияет на результаты моделирования.

2. Выделить наиболее значимые факторы, влияющие на стоимость недвижимости.
3. Построить модель для прогнозирования стоимости недвижимости.
4. Разработать небольшой веб-сервис, на вход которому поступают данные о некоторой выставленной на продажу недвижимости, а сервис прогнозирует его стоимость.

Описание данных:

- `'status'` — статус продажи;
- `'private pool'` и `'PrivatePool'` — наличие собственного бассейна;

- `'propertyType'` — тип объекта недвижимости;
- `'street'` — адрес объекта;
- `'baths'` — количество ванных комнат;
- `'homeFacts'` — сведения о строительстве объекта (содержит несколько типов сведений, влияющих на оценку объекта);
- `'fireplace'` — наличие камина;
- `'city'` — город;
- `'schools'` — сведения о школах в районе;
- `'sqft'` — площадь в футах;
- `'zipcode'` — почтовый индекс;
- `'beds'` — количество спален;
- `'state'` — штат;
- `'stories'` — количество этажей;
- `'mls-id'` и `'MlsId'` — идентификатор MLS (Multiple Listing Service, система мультилистинга);
- `'target'` — цена объекта недвижимости (**целевой признак, который необходимо спрогнозировать**).

Важные сведения о данных:

1. Признаки `'homeFacts'` и `'schools'` представляют собой словари и содержат несколько типов сведений. Необходимо десериализовать содержимое этих признаков. Если в нём есть полезные данные, необходимо создать из них новые признаки.
2. [Типы жилья и их основные характеристики](#) (может пригодиться при обработке признака `propertyType`):
 - `apartment` — съёмная квартира (нельзя купить);
 - `condo` — кондоминиум (можно купить);
 - `co-op` — квартира в жилищном кооперативе;
 - `single-family (detached, tiny home)` — односемейный (отдельный, маленький) дом;
 - `townhome, townhouse` — таунхаус;

- cape cod — черепичная крыша, облицовка деревянным сайдингом, дверь в центре здания, окна по бокам, один-два этажа;
- colonial home — минимум два этажа, симметрия, лестница в центре здания, строгий внешний вид;
- contemporary — чистые, простые линии, нейтральные цвета, натуральные текстуры;
- cottage — небольшая уютная веранда, небольшие жилые помещения;
- craftsman — деревянные элементы ручной работы, выступающие балки, большие колонны;
- Greek revival — большие белые колонны, украшения в греческом стиле, нарядный вход;
- farmhouse — прямоугольная планировка, большое крыльцо, амбарная крыша;
- French country — влияние прованса, облицовка камнем, состаренный вид;
- Mediterranean — белая штукатурка, дерево и камень в тёплых тонах, черепичные крыши, элементы испанских и итальянских вилл;
- midcentury modern — сочетание плавных природных и строгих геометрических линий, гладкость, лаконичность, большие окна;
- ranch — один этаж, низкая крыша, открытые жилые помещения, задний двор;
- split-level — жилые помещения разделены короткими лестничными пролётами;
- Tudor — деревянные балки, каменная кладка, асимметричная крутая двускатная крыша;
- Victorian (Queen Anne Houses) — два-три и более этажей, богатый декор, крутая двускатная крыша, небольшие башенки, яркий фасад;
- European Houses style — кирпич или камень, высокая крутая крыша, высокие окна, часто со ставнями, традиционные декоративные детали (фронтоны, замковые камни);
- log home — деревянный дом.

3. Некоторые жаргонные сокращения — для исправления дублирования (может пригодиться при обработке признака `propertyType`):

- `'manufactured home (house)' = 'mobile' = 'prefab' = 'modular';`
- `'mobile' = 'mo2 le' = 'mo2le';`
- `'cabin' = 'ca2 n' = 'ca2n';`
- `midcentury = mid century = mid-century;`
- `single family = single-family home.`

4. Некоторые термины, используемые в сфере купли-продажи недвижимости:

- `status` — статус, состояние.
- `estate` — объект недвижимости.
- `'foreclosed' ('foreclosure')` — «лишён права выкупа».

Это процесс, когда недвижимость переходит к кредитору из-за неуплаты ипотеки заёмщиком. Продажа foreclosure-недвижимости отличается от традиционных продаж.

- `'pre foreclosure' ('pre foreclosure auction')` — «до выкупа», «аукцион до выкупа».
- `'under contract showing' ('under contract show', 'under contract', 'active under contract')` — предложение о покупке уже сделано, но не принято продавцом окончательно, и недвижимость продолжают показывать другим потенциальным покупателям.
- `'under contract backups', 'active backup', 'backup contract'` — предложение о покупке уже сделано и принято продавцом, однако он открыт для резервных предложений.
- `'contingency'` — дополнительные обстоятельства (критерии, которым должен соответствовать продавец/покупатель до завершения сделки).
- `'contingency contract'` — контракт на случай возникновения дополнительных обстоятельств.

- **'active contingency'** — сделка уже практически состоялась, однако продавцу/покупателю для её завершения необходимо выполнить какие-то условия.
- **'insp inspection contingency'** — ситуация, когда покупатель может официально проинспектировать объект недвижимости в определённый период времени.
- **'pending escape clause'** — договор, предполагающий свободный выход сторон из него.
- **'pending backup wanted'** — договор уже заключен, однако продавец хочет продолжать показывать недвижимость другим потенциальным покупателям и принимать резервные предложения.
- **'pending take backups'** — условие, предполагающее, что продавец может принимать резервные предложения, если сделка с текущим покупателем не состоится.
- **'pending continue show'** — договор уже заключен, однако продавец хочет продолжать показывать недвижимость другим потенциальным покупателям на случай, если текущая сделка не состоится.
- **'pending inspection'** — «в ожидании инспекции».
- **'due diligence period'** — период, в течение которого покупатель имеет право проинспектировать недвижимость и изучить всю важную для заключения сделки информацию.
- **'activated'** — «активен», открыт для покупки.
- **'active with contract'** — продавец уже принял предложение о покупке, однако сделка ещё не закрыта.
- **'active with offer'** — продавец принял предложение о покупке, однако ищет резервные предложения на случай, если сделка с текущим покупателем не состоится.
- **'active auction', 'auction active', 'auction'** — продажа недвижимости с помощью аукциона.
- **'cooling-off period'** — период, в течение которого покупатель может расторгнуть договор после завершения сделки.