**Вводные слова**

Каждое наблюдение в наборе данных – это дом, который был продан в городе Эймс, штат Айова. Известны цена продажи, а также различные характеристики самого жилья. Задача – провести разведочный анализ, обработать данные, а затем разбить проданное жилье на кластеры – сначала на основе отобранных признаков, а затем – после уменьшения размерности.

**Список признаков**

* SalePrice – цена продажи жилья в $
* MSZoning – тип локации, где находится дом
* LotArea – общая площадь
* Utilities – наличие коммунальных благ
* Condition1 – расположенный рядом объект
* Condition2 – еще один расположенный рядом объект (если их больше 1)
* OverallQual – общая оценка дома
* OverallCond – общая оценка состояния дома
* YearBuilt – год постройки дома
* YearRemodAdd – год капитального ремонта
* RoofStyle – стиль крыши
* RoofMatl – материал крыши
* Exterior1st – материал отделки
* Exterior2nd – еще один материал отделки (если использовано больше 1)
* ExterQual – качество отделки
* ExterCond – состояние отделки
* Foundation – тип фундамента
* BsmtQual – качество подвала
* BsmtCond – состояние подвала
* TotalBsmtSF – общая площадь подвала
* Heating – тип отопления
* HeatingQC – качество отопления
* GrLivArea – жилая площадь
* BedroomAbvGr – спальни выше подвала
* KitchenAbvGr – кухни выше подвала
* KitchenQual – качество кухни
* TotRmsAbvGrd – общее количество комнат выше подвала
* Fireplaces – количество каминов
* FireplaceQu – качество каминов
* GarageArea – площадь гаража
* GarageQual – качество гаража
* GarageCond – состояние гаража
* PoolArea – площадь бассейна
* PoolQC – качество бассейна
* Fence – тип забора
* MiscFeature – дополнительные характеристики
* MiscVal – значения этих дополнительных характеристик

**Важно:** на любом этапе вы вольны удалять признаки, которые вам кажутся бесполезными, если при этом будет дано обоснование, почему вы так решили. Кроме того, вы можете создавать новые признаки.

|  |  |
| --- | --- |
| **Критерий** | **Балл** |
| 1. Дан общий обзор набора данных | 1 |
| 1. Проведен одномерный анализ данных, включая визуализацию и применение статистических методов | 2 |
| 1. Проведен многомерный анализ данных, включая визуализацию и применение статистических методов | 2 |
| 1. Выявлено наличие или отсутствие выбросов; в случае наличия – выбросы обработаны | 1 |
| 1. Выявлено наличие или отсутствие пропусков; в случае наличия – пропуски обработаны | 1 |
| 1. Проведена обработка признаков в зависимости от их типа | 2 |
| 1. Проведена кластеризация наблюдений, дана интерпретация кластеров и оценка качества кластеризации | 2 |
| 1. Проведено уменьшение размерности данных, обоснован выбор нового числа признаков | 2 |
| 1. Проведена повторная кластеризация на данных в пространстве уменьшенной размерности, дано сравнение результатов с п. 7 | 2 |
| 1. Общее качество работы: файл структурирован, код соответствует нормам программирования в Python, рисунки имеют все необходимые подписи, даны аналитические комментарии и т.п. | 5 |