

Номер темы	Набор данных	Основная тема	Примерная формулировка задачи
1.	https://www.kaggle.com/gustavomodelli/forest-fires-in-brazil	Закономерности в частотах и силе пожаров	выделить статистические характеристики пожаров различных регионах Бразилии; найти зависимости между частотой или силой пожаров с другими параметрами в наборе данных; выделить аномальное уменьшение или увеличение количества или силы пожаров в некотором регионе
2.	https://www.kaggle.com/elikplim/forest-fires-data-set	Закономерности в частотах и силе пожаров	выделить статистические характеристики пожаров; найти зависимости между частотой или силой пожаров с другими параметрами в наборе данных; выделить аномальное увеличение или ослабление силы или частоты пожаров в определенное время
3.	https://www.kaggle.com/sumitm004/forest-fire-area	Предсказание силы пожара и его последствий	найти зависимости между силой пожаров с другими параметрами в наборе данных; выделить аномальное увеличение или ослабление силы или частоты пожаров в определенное время
4.	https://www.kaggle.com/new-york-city/ny-2015-street-tree-census-tree-data	Популярные деревья в районах Нью-Йорка	выделить районы, в которых состояние деревьев аномально хорошее или плохое; выделить наилучший район по состоянию деревьев; выделить виды деревьев (или более общие элементы классификации), обладающие наилучшими показателями «здоровья»
5.	https://www.kaggle.com/mrisdal/minneapolis-incidents-crime	Закономерности в преступлениях	выделить районы с наибольшим или наименьшим количеством преступлений, выделить наиболее популярные преступления в каждом районе, объединить районы в кластеры по набору наиболее распространенных преступлений; построить предсказание типа преступления по информации о времени и месте
6.	https://www.kaggle.com/paultimothy/mooney/denver-crime-data	Закономерности в преступлениях	выделить районы с наибольшим или наименьшим количеством преступлений, выделить наиболее популярные преступления в каждом районе, объединить районы в кластеры по набору наиболее распространенных преступлений; построить предсказание типа преступления по информации о времени и месте
7.	https://www.kaggle.com/roxiroxi/diamond-price-prediction	Предсказание цен на предметы роскоши или недвижимость	построить математическую модель определения стоимости бриллианта на основе его характеристик, выделить кластеры в наборе данных, построить оценку стоимости драгоценных камней в каждом кластере и сравнить точность с моделью, построенной по всем данным, выделить аномально дорогие или дешевые драгоценные камни

8.	https://www.kaggle.com/shree1992/housedata	Предсказание цен на предметы роскоши или недвижимость	построить математическую модель определения стоимости недвижимости по её характеристикам, выделить аномально дорогие или дешевые объекты недвижимости; построить несколько классификаторов на основе данных, кластеризованных по району, времени; ответить на вопрос, что происходит с ценами недвижимости в определенных сегментах
9.	https://www.kaggle.com/quantbruce/real-estate-price-prediction	Предсказание цен на предметы роскоши или недвижимость	построить математическую модель определения стоимости недвижимости по её характеристикам, выделить аномально дорогие или дешевые объекты недвижимости; построить несколько классификаторов на основе данных, кластеризованных по району, времени; ответить на вопрос, что происходит с ценами недвижимости в определенных сегментах
10.	https://www.kaggle.com/mrisdal/open-exoplanet-catalogue	Кластеризация или классификация экзопланет	построить классификаторы планет по размеру («юпитеры», «нептуны», «земли» и промежуточные классы), температуре и расстоянию до звезды; выделить планеты, обладающие аномально высокими или низкими показателями по каждой классификации; подобрать планеты, похожие по своим характеристикам на Землю или планеты солнечной системы; подобрать планеты, на которых может существовать жидкая вода.
11.	https://www.kaggle.com/muhakabay/sloan-digital-sky-survey-drl6	Классификация астрономических объектов	определить важные для классификации характеристики, выделить аномально высокие и низкие значения параметров, построить физически интерпретируемые границы классов (звезда, квазар, галактика), сравнить работу простых методов классификации на представленном наборе данных
12.	https://www.kaggle.com/danielkyrka/bmw-pricing-challenge	Предсказание цен на автомобиль	определить важные для определения цены характеристики, выделить аномальные значения характеристик в наборе данных, построить прогноз стоимости автомобиля по его характеристикам
13.	https://www.kaggle.com/hugoncosta/price-of-flats-in-moscow	Предсказание цен на автомобиль	построить математическую модель определения стоимости недвижимости по её характеристикам, выделить аномально дорогие или дешевые объекты недвижимости; построить несколько классификаторов на основе данных, кластеризованных по расстоянию от центра, и сравнить модель с построенной по всем объектам недвижимости
14.	https://www.kaggle.com/imdevskp/corona-virus-report	Кластеризация заболевших коронавирусом	выделить аномально быстрое или медленное распространение вируса; выделить группы населения, наиболее сильно подверженные вирусу; выделить подгруппы среди заболевших – определить, какие разновидности вируса наиболее опасны.
15.	https://www.kaggle.com/ronitf/heart-	Задача классификации	Построить несколько классификаторов. Что общего в каждом классе.

	disease-uci	сердечно-сосудистых заболеваний по 4 классам	Применить алгоритмы кластеризации. Что общего между объектами в каждом кластере? Выделить аномальные объекты, которые далеко отстоят от кластеров.
16.	https://www.kaggle.com/mlg-ulb/creditcardfraud	Определение мошеннических операций с картами	Построить несколько бинарных классификаторов. Какие параметры вносят наибольший вклад при определении мошеннических операций? Являются ли мошеннические операции выбросами?
17.	https://www.kaggle.com/arjunbhasin/2013/ccdata	Кластеризация держателей кредитных карт.	Использовать несколько алгоритмов кластеризации, сравнить результаты их работы между собой. Что общего между объектами в каждом кластере? Какое оптимальное количество кластеров для алгоритма k-средних? Много ли данных, которые трудно отнести к какому-либо кластеру?
18.	https://www.kaggle.com/annavictoria/speed-dating-experiment	Данные быстрых свиданий.	Каковы наименее желательные качества в мужчине? Отличается ли это для женщин? Являются ли общие интересы более важными, чем общее расовое происхождение? Произвести кластеризацию, выявить аномальные случаи. Что в них особенного?
19.	https://www.kaggle.com/jessemostipak/hotel-booking-demand	Данные о бронировании отелей.	Применить алгоритмы кластеризации. Что общего между объектами в каждом кластере? Выделить аномальные объекты, которые далеко отстоят от кластеров. Построить модель прогнозирования отмены брони.
20.	https://www.kaggle.com/mirichoi0218/insurance	Предсказание стоимости медицинской страховки.	Построить модель предсказания, оценить вклад каждого компонента. Есть ли выбросы среди данных, какие? Применить алгоритмы кластеризации. Что общего между объектами в каждом кластере?
21.	https://www.kaggle.com/shivam2503/diamonds	Предсказание цены бриллианта.	Построить модели предсказания, оценить вклад каждого компонента. Построить статистические оценки параметров и гистограммы распределений. Есть ли выбросы среди данных, какие? Применить алгоритмы кластеризации. Что общего между объектами в каждом кластере?
22.	https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009	Определение качества вина.	Построить модели предсказания качества, оценить вклад каждого компонента. Построить статистические оценки параметров и гистограммы распределений. Построить бинарный классификатор: хорошее/плохое вино. Выделить аномальные объекты
23.	https://www.kaggle.com/uciml/student-alcohol-consumption	Употребление алкоголя учащимися.	Построить прогноз итогового бала учащегося. Насколько сильно влияет употребление алкоголя? Какие параметры вносят наибольший вклад в предсказание? Построить графики статистической оценки параметров. Применить алгоритмы кластеризации. Что общего между объектами в

			каждом кластере?
24.	https://www.kaggle.com/spscientist/students-performance-in-exams	Успеваемость учащихся.	Построить модели для предсказания оценок по математике, чтению и письму. Какие параметры вносят наибольший вклад в предсказание? Определить зависимость между: полом и оценкой за чтение, этической группой и оценкой по математике? У каких признаков наибольшая корреляция? Построить графики статистической оценки параметров.
25.	https://www.kaggle.com/ananthu017/california-wildfire-incidents-20132020	Закономерности в частотах и силе пожаров	выделить статистические характеристики пожаров; найти зависимости между частотой или силой пожаров с другими параметрами в наборе данных; выделить аномальное увеличение или ослабление силы или частоты пожаров в определенное время