

Условие задачи:

В компании имеются пять моделей автомобилей: 1, 2, 3, 4 и 5. Имеется 20 автомобилей, распределенных между моделями следующим образом: 1-5, 2-6, 3-4, 4-5.

Для каждого автомобиля записаны данные о маршрутах, включающие объем потраченного бензина и его октановое число, а также время в пути (туда и обратно).

Необходимо выполнить следующие задачи:

Определить общее количество маршрутов для всех автомобилей (если у каждой конкретной машины одинаковое количество уникальных путей, но поездок разное число).

Выявить подозрительные маршруты, на которых было потрачено больше бензина, чем требуется для данного маршрута.

Для решения задач необходимо использовать данные о моделях автомобилей и маршрутах, включающих объем потраченного бензина, его октановое число и время в пути.

Для определения общего количества маршрутов можно просуммировать количество маршрутов для каждого автомобиля.

Для выявления подозрительных маршрутов можно использовать сравнение объема потраченного бензина с расходом бензина, необходимого для данного маршрута, и среднего расхода бензина для соответствующей модели автомобиля. Если объем потраченного бензина превышает требуемый расход бензина для данного маршрута, то этот маршрут можно считать подозрительным.

Для определения требуемого расхода бензина для каждого маршрута можно использовать данные об октановом числе и расходе бензина для каждой модели автомобиля. Для каждой поездки на маршруте можно рассчитать расход бензина, используя формулу: объем потраченного бензина = расход бензина \* расстояние, где расход бензина зависит от октанового числа и может быть рассчитан для каждой модели автомобиля.

Таким образом, для решения поставленных задач необходимо использовать данные об автомобилях, маршрутах и расходе бензина для каждой модели автомобиля. Можно применить различные методы анализа данных, такие как сравнение, агрегация и расчеты. Результатом работы должно быть количество маршрутов для всех автомобилей и список подозрительных маршрутов.

Обозначения столбцов в датасете: car\_model - модель автомобиля, route\_id – идентификатор маршрута, distance -дистанция, fuel\_octane – октановое число бензина, fuel\_consumed - количество израсходованного топлива, time\_spent – время в пути.

Решение задачи:

1. Для решения этой задачи с помощью библиотеки scikit-learn необходимо определить, какие модели автомобилей потребляют больше топлива, а также какие поездки являются подозрительными по объему потраченного топлива.
2. Если у каждой конкретной машины одинаковое количество уникальных путей, но поездок разное число, то можно использовать метод кластеризации для определения уникальных путей.

Для этого необходимо провести кластеризацию на основе данных о затратах на топливо и времени в пути для каждой машины, с учетом марки автомобиля.

Например, можно использовать метод кластеризации K-means из библиотеки `scikit-learn`. Сначала необходимо разделить данные на тренировочный и тестовый наборы. Затем провести кластеризацию на тренировочном наборе, определить количество кластеров и применить модель к тестовому набору, чтобы определить уникальные пути для каждой машины.