Описание:

Для понимания структуры платящих пользователей очень часто используются алгоритмы кластеризации, изначально мы не знаем какие классы удастся выделить и сколько их будет. В общем случае нам интересно выделить пользователей, которые приносят в наш продукт большую часть денег, пользователей которые представляют некий портрет среднего пользователя нашего продукта и пользователей, которые по каким-то причинам не очень охотно платят.

В качестве “подопытного кролика” предлагается рассмотреть набор данных PaymentsInfoDataset.csv со следующей структурой:

CustomerId - уникальный идентификатор пользователя

Monetary - объем средств, который пользователь принес в продукт за анализируемый период времени

Recency - время последнего платежа пользователя за анализируемый период времени

Frequency - количество платежей, которое пользователь совершил за анализируемый период времени

factor\_\* - замаскированные показатели

Есть описательные статистики разных пользователей. Найдите среди всех пользователей похожие группы, проведите кластерный анализ и подумайте есть ли здесь вообще зависимости или нет?

Задачи:

1 - Необходимо разработать модель, которая позволит разделить пользователей на классы по их платежному поведению, должна получится ясная картина того, в какой класс попал каждый пользовать и почему, количество классов определяете Вы, руководствуясь аналитикой и здравым смыслом. Возможна ситуация, когда ряд пользователь не попал ни в один из выделенных классов, таких пользователей можно отнести к классу Other, но он не должен быть слишком большим, не забывайте, что цель разработки данной модели - понять структуру платящих пользователей продукта, а не запутаться еще больше. В качестве основы модели могут использоваться алгоритмы машинного обучения или методы классической аналитики.

2 - показать средние значения факторов в разных кластерах

3 - ответить на вопрос, какие из факторов не несут информации. Ответом на Задание 3 являются названия факторов без какой-либо информации через запятую

4 - в коде напишите, какие кластеры Вы выделили и почему

5 - визуализируйте кластеры