Курсовой проект состоит из 3х частей

1. Студенты весь курс работают с retail-датасетом: данные о покупках 2.5k клиентов 80k товаров, признаковое описание пользователей и товаров. 3 недели покупок недоступны студентам. По мере развития курса мы строим все более сложные и продвинутые RS, которые улучшают качество рекомендаций. В конце курса студентам предлагается построить алгоритм на исторических данных и предсказать какие товары пользователи купят в будущем.

Для обучения доступны 95 недель + информация о пользователях и товарах

96 - 98 недели недоступны. Нужно предсказать какие товары пользователи купят за эти 3 недели

Качество оценивается по метрике precision@5, усредненной по всем пользователям. Критерий сдачи: значение метрики > 0.27

Во время разработки модели выделите 3 известные недели (93 - 95) в качестве валидационной выборки. Продемонстрируйте на валидационной выборке требуемое значение precision@5

Ноутбук должен содержать описание pipeline’а Вашей модели. (Какие модели использовали, какие признаки генерировали и тд)

Бизнес-ограничения:  
1. Нельзя рекомендовать top 3 самых популярных товаров

2.Нельзя рекомендовать товары, которые не продавались последние 12 месяцев

3.Нельзя рекомендовать товары, с общим числом продаж < 50

4.Нельзя рекомендовать товары, которые стоят < 1$

5.Каждому пользователю нужно порекомендовать 5 товаров, один из них должен быть обязательно товаром, который данный пользователь никогда не покупал

1. В рамках решения основной задачи студенты должны произвести предварительный и визуальный анализ данных. В рамках его нужно построить диаграммы распределения признаков, проанализировать некоторые категории товаров, их цену, среднее число покупок и тд. Это творческое задание, единственное конкретное требование: каждый студент должен найти в данных что-то уникальное (то чего не нашли одногруппники).

1. \* Провести некоторое исследование по любой теме, разобранной в курсе, тему исследования необходимо согласовать с преподавателем.