Домашнее задание

Дисциплина	Python для инженерии данных
Тема	Тема 9. Знакомство с Apache Spark
Форма проверки	**Самопроверка. Студент выполняет задание и самостоятельно проверяет его.**
Имя преподавателя	Дмитрий Клабуков
Время выполнения	1 час
Цель задания	Научиться читать и обрабатывать данные с помощью Spark
Инструменты для выполнения ДЗ	jupyter notebook или google colab
Правила приема работы	Прикрепите ссылку в LMS на выполненное задание в Google colab или GitHub (если вы использовали Jupyter Notebook) **Важно:** убедитесь в том, что по ссылке есть доступ в Google colab (иногда в колабе нет доступа для другого логина).
Критерии оценки	**Задание считается выполненным, если:** - прикреплена ссылка на файл с выполненным заданием - доступ к файлу открыт - код дает правильный ответ к задаче **Задание не выполнено, если:** - файл с заданием не прикреплен или отсутствует доступ по ссылке - код выдаёт ошибку или дает неправильный ответ
Дедлайн	**7 дней с даты проведения соответствующего вебинара**

Перед выполнение задания установите jupyter notebook либо используйте google colab

В файле movies.csv лежит база фильмов. Название фильма записано во втором столбце title.

Задача:

С помощью Spark разбейте названия фильмов на отдельные слова и посчитайте, какое слово встречается чаще всего.

Чек-лист самопроверки

Критерии выполнения задания	Отметка о выполнении
Установлен jupyter notebook либо используется google colab	
Создан профиль на https://github.com (при использовании jupyter notebook)	
Для вычислений использован Spark	
Посчитано, какое слово встречается чаще всего в названиях фильмов	
Прикреплена на учебной платформе ссылка на выполненное задание в Google colab или Github (если вы использовали jupyter notebook)	
Если используется Google colab, то по ссылке есть доступ (иногда в колабе нет доступа для другого логина)	