**ЗАДАНИЕ НА ПОЗИЦИЮ**

**ML Engineer**

**Модели классификации на основе машинного обучения**

В тестовом задании представлены данные из нескольких розничных сетей по товарам кондитерской категории из различных продуктовых категорий. Вам представлено для работы 2 файла:

1) уже правильно классифицированные SKU (file\_1.csv),

2) Список SKU для классификации(file\_2.csv).

*Ключевая задача – распределить неклассифицированные товары (из file\_2) по продуктовым подкатегориям, которые вы видите в исходном массиве (file\_1), а именно - Product\_category, Product\_Type.*

**Чтобы прочитать файлы:**

Изображение выглядит как текст, Шрифт, линия, снимок экрана

Автоматически созданное описание

**Шаги вашего исследования:**

* 1. Проведите исследование данных в Python и выявите аномалии/проблемы (обозначьте их в вашем ноутбуке в отдельной ячейке)
  2. Реализуйте классификацию (дерево решений, опорные вектора и/или др.), которая поможет распределить по имеющимся группам (*Product\_category, Product\_Type) -* новые товары из file\_2
  3. Подготовьте краткий обзор выбранного вами метода с аргументацией почему он наилучшим образом подходит под данную задачу

Отправьте результат домашнего задания в виде ноутбука с описанием логики его решения и кратких выводов (на 2-3 предложения) в удобном формате.

Не забывайте, что вы можете обращаться с вопросами, уточнениями и дополнительными деталями.