

Выполнил Сироткин-Шретер Лев (@Schroter)

Модуль 1 Задание

Ваша задача - проанализировать следующие наборы данных, предоставленные компанией, и выявить все проблемы с качеством данных. После того как вы просмотрите эти наборы данных, составьте перечень всех проблем с качеством данных (если они есть).

Проблемы:

Сравнив элементы выяснил, что ошибка в данных есть в таблице emotions - странный и короткий cid vl20see20.

В названии столбца cid в таблице emotions имеется ненужный пробел, из-за чего могут быть ошибки при выводе данных.

В таблице emotions уникальный номер (cid) 6ad9452a-4d4b-47e4-8c4e-447fd874c095 повторяется 5 раз и ответы в surveys отличаются.

Также имеются ошибки во времени, так как в 2021/08/10 12:18:54 PM GMT+3
2021/08/10 12:43:55 PM GMT+3 ошибка в времени.

Модуль 2 Задание

Создайте презентацию, в которой будет изложен способ, с помощью которого мы определим: какие эмоции испытывали зрители, какие эмоции испытывали разные когорты ЦА из респондентов проекта, участвующих в исследовании и любые другие взаимосвязи. Объясните алгоритмы выполнения задания по всем трем фазам: Исследование данных, разработка модели и интерпретация.

Алгоритм выполнения задания:

1. Визуальный осмотр датасетов;
Осмотрел датасет, на явные недостатки в нем.
2. Извлечение данных;
Передал данные в среду разработки, для последующего анализа и работы с ними.
3. Изучение данных и визуализация;
Тут начал предпринимать последующие пути для группировки и визуализации данных.
4. Обобщение данных;
Поиск по каким вариантам данные можно обобщить и классифицировать.
5. Группировка данных;
Сгруппировал данные, например самое явное по трем классам заработка.
6. Определение моделей и тенденций;
Определил взаимосвязи. Например, что у более обеспеченных респондентов меньше ярких эмоций.
7. Идентификация отношений;
Визуализировал все группы данных, для поиска проблемных мест.
8. Поиск аномальных данных;

Находил ошибки в имеющихся данных во время анализа.

Дополнительно есть презентация: Neuro Project Тестовое задание (Модуль 2)

Модуль 3 Задание

Пожалуйста, создайте дэшборд, который мы сможем представить заказчику на нашей следующей встрече. Представьте сводку данных и результаты анализа в виде дэшборда. В частности, ваша презентация должна содержать результат анализа данных и найденные инсайты в датасете.

Neuro Project Тестовое задание (Модуль 3)