Задача 2.

Всего 2626 дата поинтов (количество строк информации). Отсутствующих значений так же не наблюдается (Хорошо).

Самые количественно- значимые регионы с инцидентами за период с 28/03/2022 по 31/03/2022:

Самарская область – 265 инцидентов

Вологодская область – 155 инцидентов

Республика Башкортостан – 150 инцидентов

Калужская область – 90 инцидентов

Ярославская область – 89 инцидентов

Белгородская область – 79 инцидентов

Основной группой тем является Экономика и бизнес – рост цен (2583 инцидента) за весь период.

Следом идет Торговля (26 инцидентов)

Коронавирус (17инцидентов)

Из анализа так же следует что наивысшие точки пика инцидентов пришлись на 29/03/2022 и 31/03/2022. Возможно, здесь преследуется какая-то причина, но у нас не так много данных что бы делать заключения.

Из возможных метрик, я бы добавил категории продуктов и соотносил каждую жалобу к категории что бы можно было проследить тенденцию роста цен на конкретную продукцию.

Возвращаясь к первой задаче, здесь я бы добавил NLP Машинку что бы разбить предложения на составляющие слова/ словосочетания и попробовал бы сгруппировать инциденты в категории и дальше посмотреть к какой категории относятся инциденты (Продовольствия, услуги и тд.)