В данном коде содержится фильтрация данных и визуализация. Этот набор данных содержит обзорную информацию из Google Maps (ID рецензента, имя, время просмотра (unix time), рейтинг, текст отзыва, фото отзыв, бизнес-ответ на отзыв, идентификатор компании) до сентября. 2021 год в США. Брала штат Северная Дакота для анализа.

В начале загружаю данные в Google collab, импортирую библиотеку pandas и вывожу таблицу. Далее смотрю информацию по таблице. Вижу, что есть нулевая переменная (None), после избавляюсь от нее и еще перевожу время из unix time в формат гггг.мм.дд и вывожу датасет.

На этом этапе я столкнулась с проблемой, после перевода формата даты, везде дата была одна это 1970.01.01., что в свою очередь не корректно.

Далее я приступила к графику «соотношения количества отзыва от рейтинга». Проблема на этом этапе заключалась в том, что по заданию надо было учесть фактор даты, но у меня это не получилось.

Первое, формат даты не корректный, а второе при попытке построения гистограммы с датой, критерий «отзыв», не получалось перевести в числовой формат (как на гистограмме соотношения кол-во отзыва от рейтинга).

Еще была проблема при импорте данных. На сайте было несколько видов данных, это сокращенные и полные. У меня получилось загрузить только сокращенные, так как для полных нужно было заполнять гугл форму и ждать ответа, ответ пришел не сразу. Поэтому я работала только с критериями «отзыв», «рейтинг» и «дата».

Но даже после того как прислали полные данные, когда я на них посмотрела, то возник вопрос. Как выделить только книжные магазины из столбца критерий? Если в одном объекте указано несколько вариантов, например вместо hotel (гостиница) написано Bed & breakfast, Indoor lodging (Ночлег с завтраком, Жилье в помещении), что усложняет поиск конкретной категории. Я пыталась отсортировать категорию Book store, но мне просто ничего не выводит.