Практическая работа № 6

На собственном наборе данных (не менее 250МБ) выполнить следующий алгоритм:

1. Загрузить набор данных из файла
2. Провести анализ набора данных по следующим параметрам:
   a. Объем памяти, который занимает файл на диске
   b. Объем памяти, который занимает набор данных при загрузке в память
   c. Вычислить для каждой колонки занимаемый объем памяти, долю от общего объема, а также выяснить тип данных
3. Полученный набор данных отсортировать по занимаемому объему памяти. Вывести в файл (json) данные по колонкам с пометкой, что это статистика по набору данных без применения оптимизаций.
4. Преобразовать все колонки с типом данных «object» в категориальные, если количество уникальных значений колонки составляет менее 50%.
5. Провести понижающее преобразование типов «int» колонок
6. Провести понижающее преобразование типов «float» колонок
7. Повторно провести анализ набора данных, как в п. 2, сравнив показатели занимаемой памяти
8. Выбрать произвольно 10 колонок для дальнейшем работы, прописав преобразование типов и загрузку только нужных данных на этапе чтения файла. При этом стоит использовать чанки. Сохраните полученный поднабор в отдельном файле.
9. Используя оптимизированный набор данных, построить пять-семь графиков (включая разные типы: линейный, столбчатый, круговая диаграмма, корреляция и т.д.)

Набор данных подгружать в репозиторий не нужно, достаточно указать ссылку, откуда его можно скачать.