Для отладки аналитического модуля необходимо иметь достаточно большой набор данных. Набор данных, подходящий для данной задачи был найден на сайте <https://www.kaggle.com/marcinex1423/salesdb-grocery>. Данный набор содержит в себе информацию о продажах сети продуктовых магазинов Market Grocery Company за 2018 год и включает более 6 миллионов транзакций. Данный набор представляет собой csv-файл объемом 518 мегабайт. Содержимое файла представлено на рисунке @номер

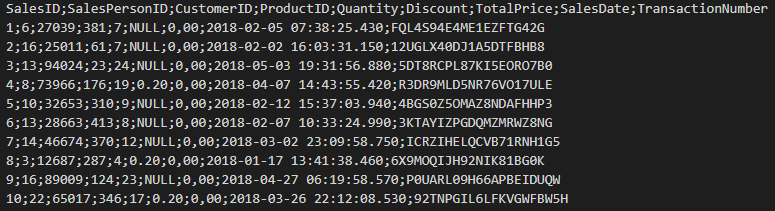


Рисунок @номер – Содержимое файла с продажами

Информация о покупателях, продавцах и самих товарах находится в отдельных csv-файлах.

Для обработки csv файлов использовалась библиотека ChoETL.NETStandard. Данная библиотека позволяет считывать данные из csv-файлов, десериализуя их в пользовательские типы, а также формировать LINQ-запросы к этим данным не загружая в память файл целиком.

Обработка 100000 загруженных транзакций заняла порядка 30 секунд.

