Тестовое задание на позицию “Аналитик данных”

Имеется набор данных. Необходимо проанализировать его и представить ответы на вопросы с визуализацией каждого пункта.  
 Ссылка: https://docs.google.com/spreadsheets/d/1DkdhWMrVjtflQfmCIKl8hCjSlIxph1L42\_fgyFo0QPs/edit?usp=sharing

Задание 1  
 а) Какие подгруппы товаров наиболее часто покупают за все время продаж (минимум 4 группы)?  
 б) Какие подгруппы товаров наиболее часто покупают за последние два года (минимум 4 группы)?  
 в) Какие подгруппы товаров наиболее часто покупают за последний год(минимум 4 группы)?  
  
 Подгруппа – Sub-Category  
 Дата – Order\_date  
 Сделать вывод на основе полученных результатов.  
  
 Задание 2  
 Построить boxplot («Ящик с усами») на основе продаж (Sales). Найти мажоритарную черту (т.е. избавиться от аномалий и представить четкую картину распределения величин).  
 Можно использовать правило трех сигм. Однако любые другие решения приветствуются.  
  
 Задание 3  
 Для этого задания необходимо разбить все покупки на энное количество групп “Sale\_group” (Допустим маленькие продажи, средние и высокие) на основе Sales

Сгруппировать данные на основе региона и группы продаж (Region, Sale\_group). Определить основные тенденции и паттерны. Выделить наиболее «прибыльную» группу.  
  
 Замечания: Работу лучше выполнить в jupyter и там же написать выводы по каждому заданию. После выполнения загрузите код на github.