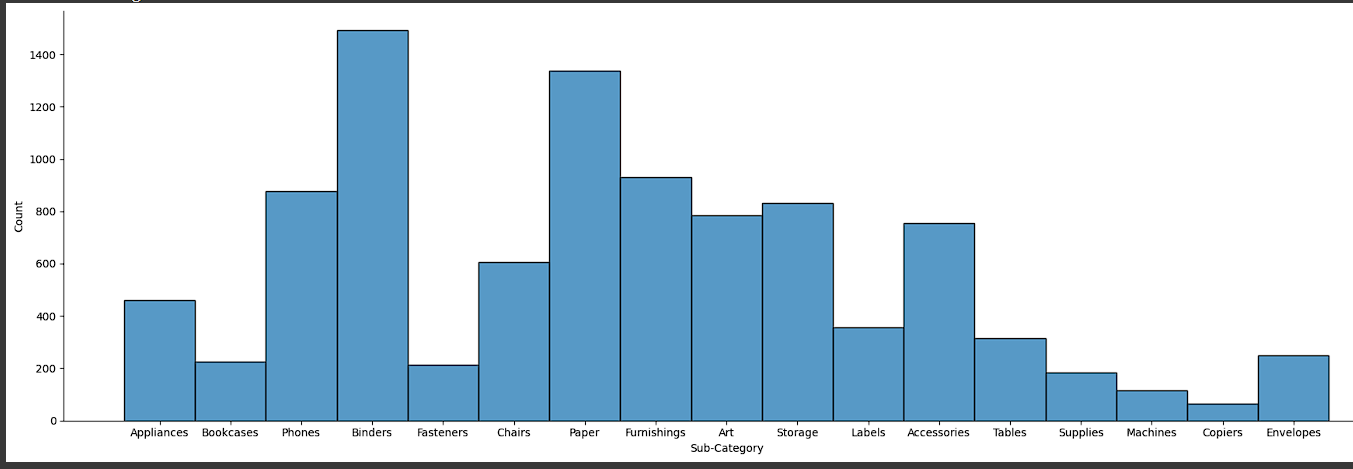
***Тестовое задание на позицию “Аналитик данных”***

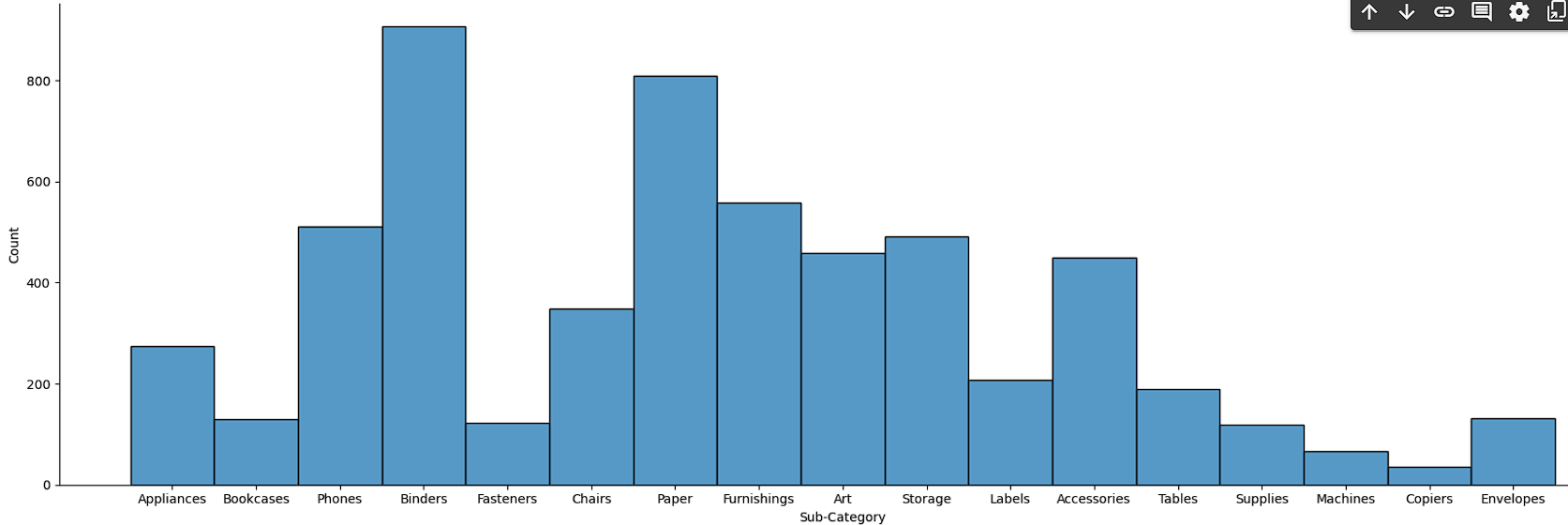
Начальные данные:  
https://docs.google.com/spreadsheets/d/1DkdhWMrVjtflQfmCIKl8hCjSlIxph1L42\_fgyFo0QPs/edit?usp=sharing

* Задание 1

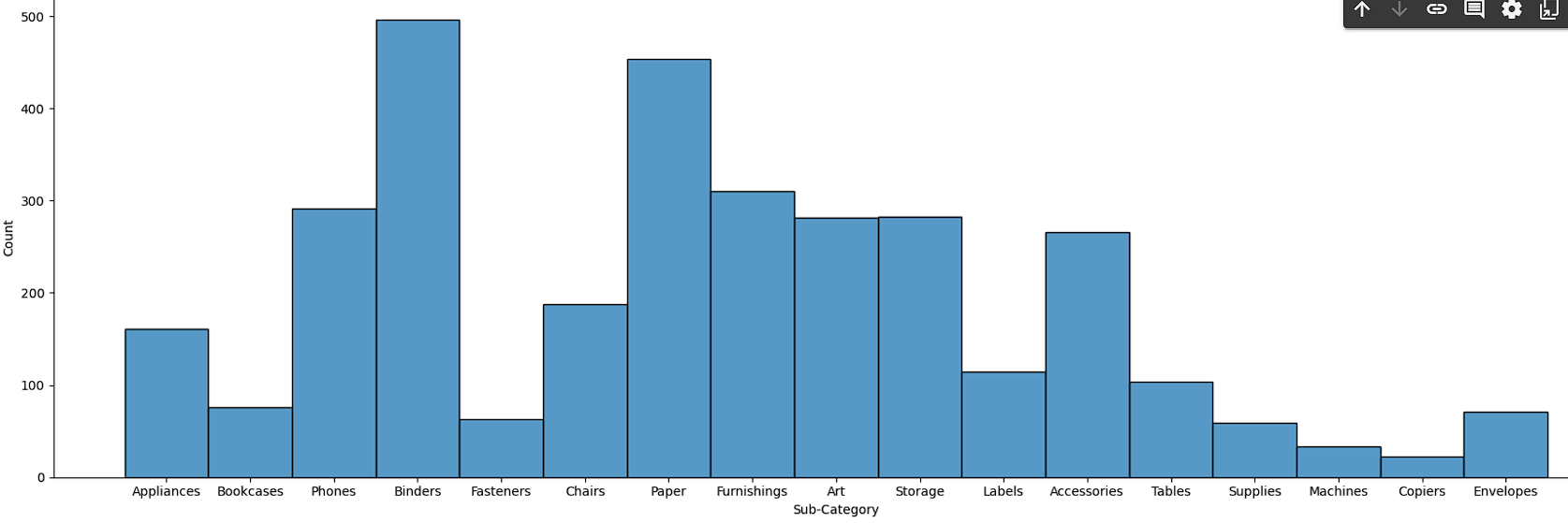
а) Какие подгруппы товаров наиболее часто покупают за все время продаж (минимум 4 группы)?



Ответ: Binders, Paper, Furnishings, Phones.  
  
б) Какие подгруппы товаров наиболее часто покупают за последние два года (минимум 4 группы)?



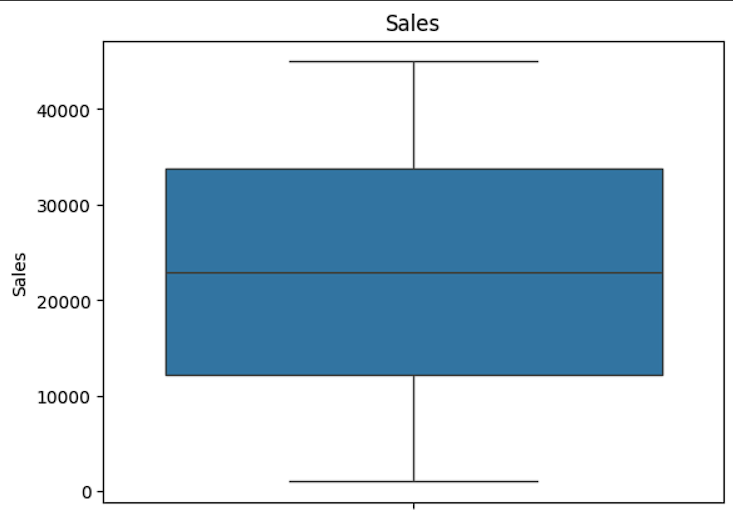
Ответ: Binders, Paper, Furnishings, Phones.  
в) Какие подгруппы товаров наиболее часто покупают за последний год(минимум 4 группы)?

  
Ответ: Binders, Paper, Furnishings, Phones.

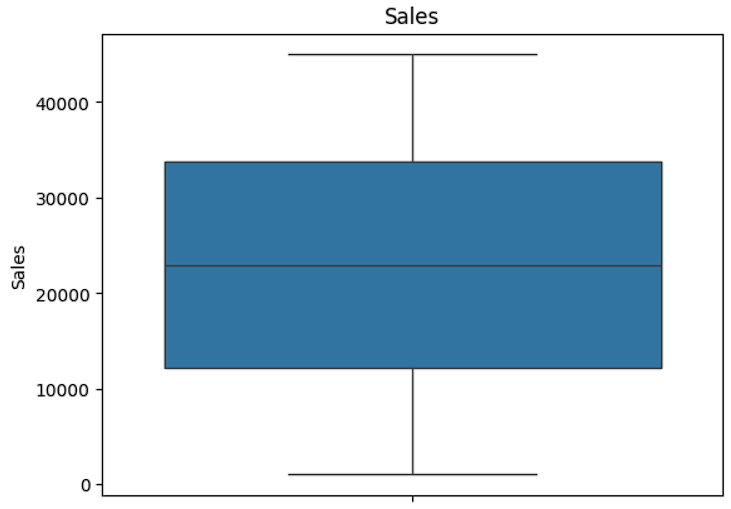
**Вывод:** “Наиболее часто покупаемые товары” неизменны на протяжении долгого времени.

* Задание 2

а) Построить boxplot («Ящик с усами») на основе продаж (Sales).



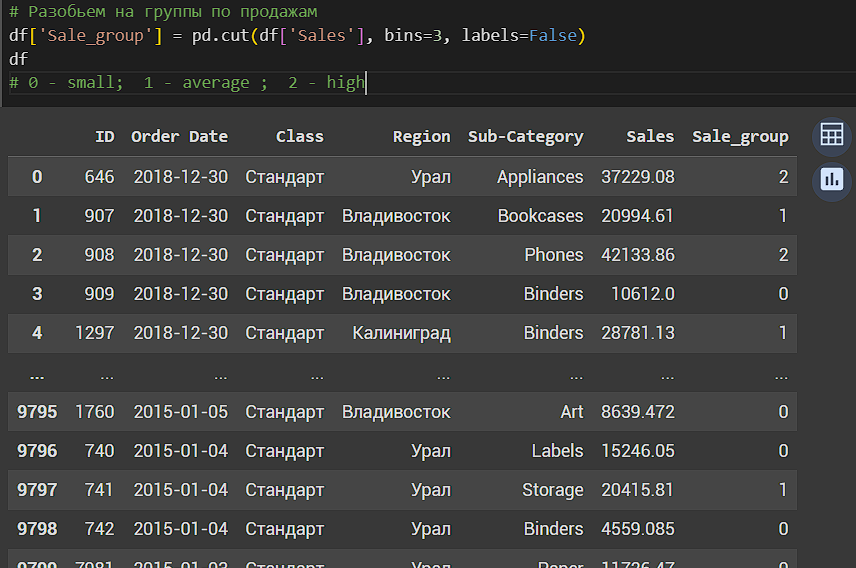
б) Найти мажоритарную черту (т.е. избавиться от аномалий и представить четкую картину распределения величин).



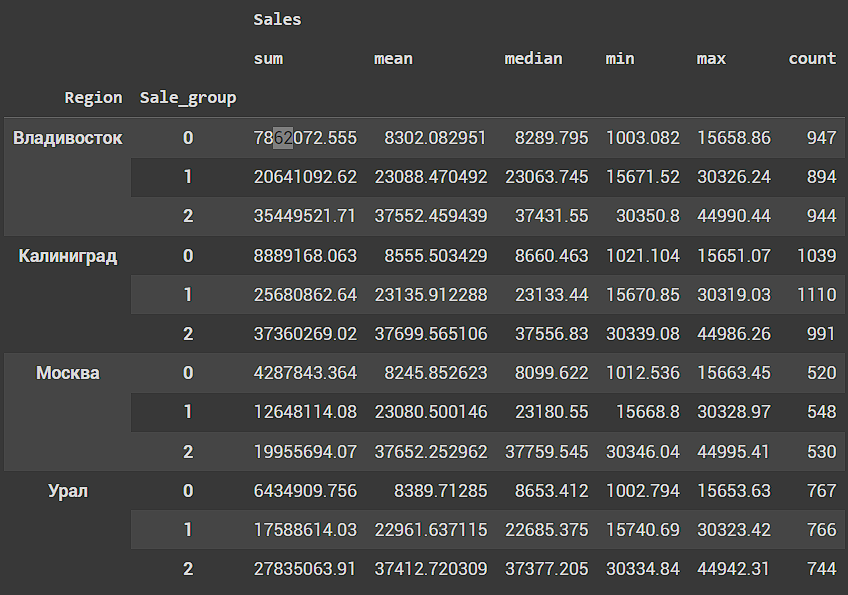
**Вывод:** Выбросы в +- 3\*σ от МО не были обнаружены. После проведения аналитического обзора я пришел к такому же выводу.

* Задание 3

а) Разбить все покупки на n-ое количество групп “Sale\_group” (допустим маленькие продажи, средние и высокие) на основе Sales.



б) Сгруппировать данные на основе региона и группы продаж (Region, Sale\_group).



в) Определить основные тенденции и паттерны. Выделить наиболее «прибыльную» группу.

*\*Скриншоты не стал прикреплять, все в коде есть\**

**Вывод:** Очевидно, что наибольшая выручка с продаж в регионах с “высокой” группой. Лидирует Калининград, Москва, Владивосток - крупные города, очень важные торговые узлы. Наибольшая сумма продажи в Москве - самый развитый центр с большими средними ЗП. По количеству продаж лидирует средняя группа - самая обширная прослойка общества со средними доходами.