

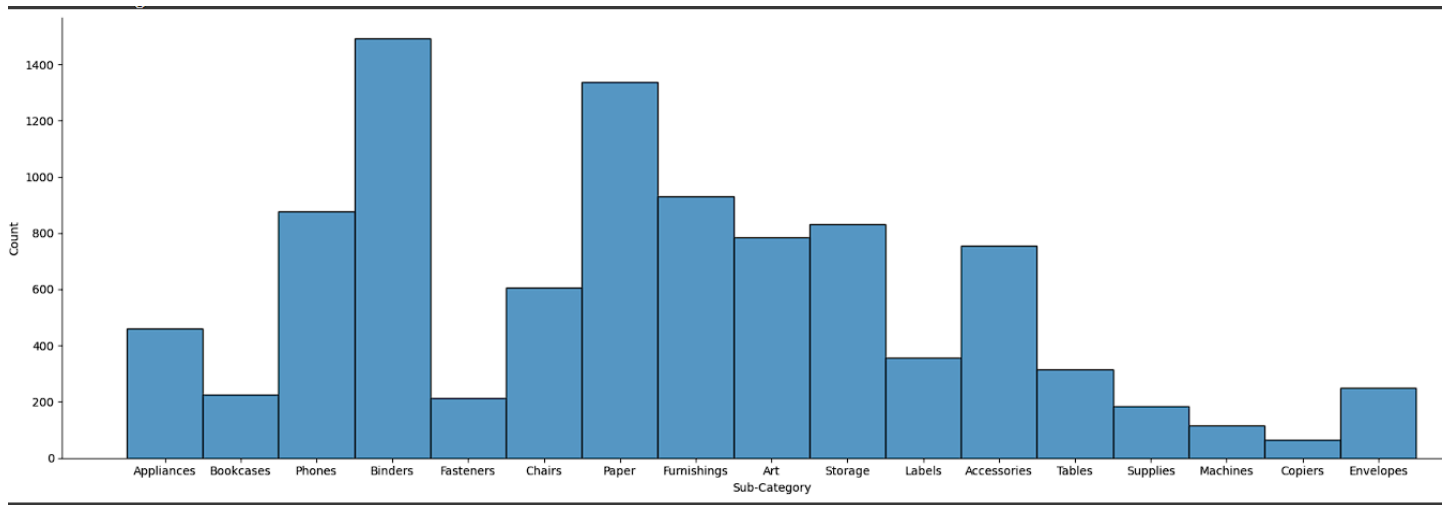
Тестовое задание на позицию “Аналитик данных”

Начальные данные:

https://docs.google.com/spreadsheets/d/1DkdhWMrVjtflQfmCIKl8hCjSIxph1L42_fgyFo0QPs/edit?usp=sharing

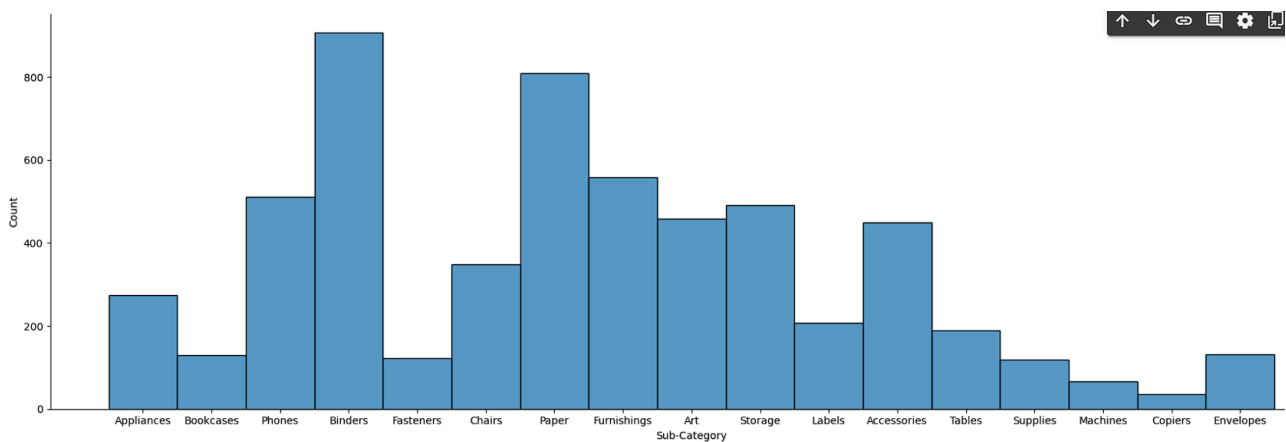
● Задание 1

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



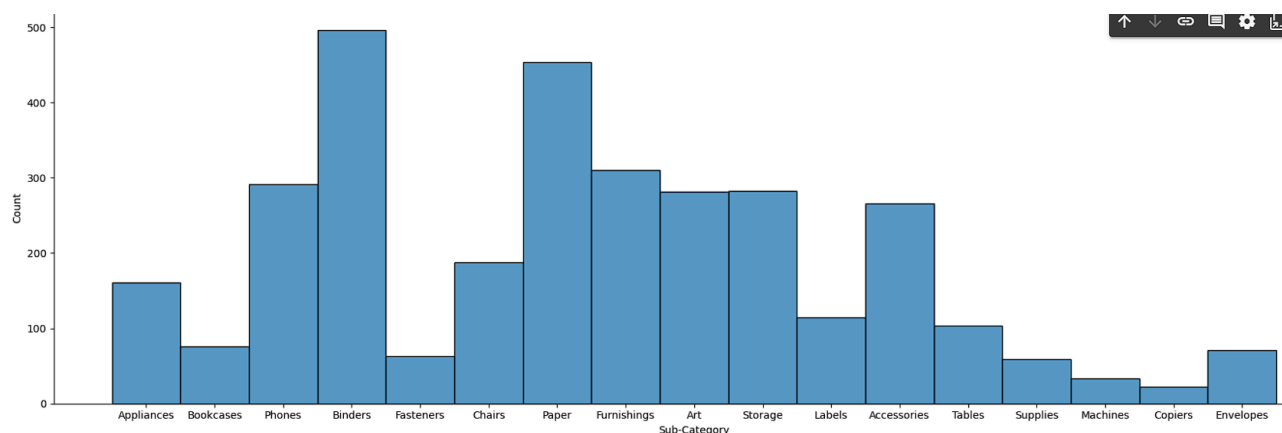
Ответ: Binders, Paper, Furnishings, Phones.

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)?

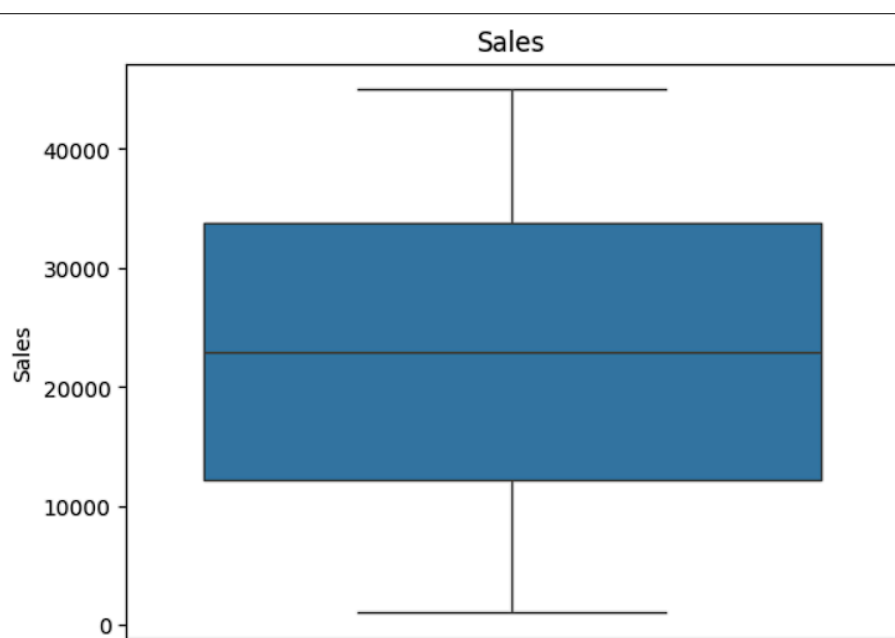


Ответ: Binders, Paper, Furnishings, Phones.

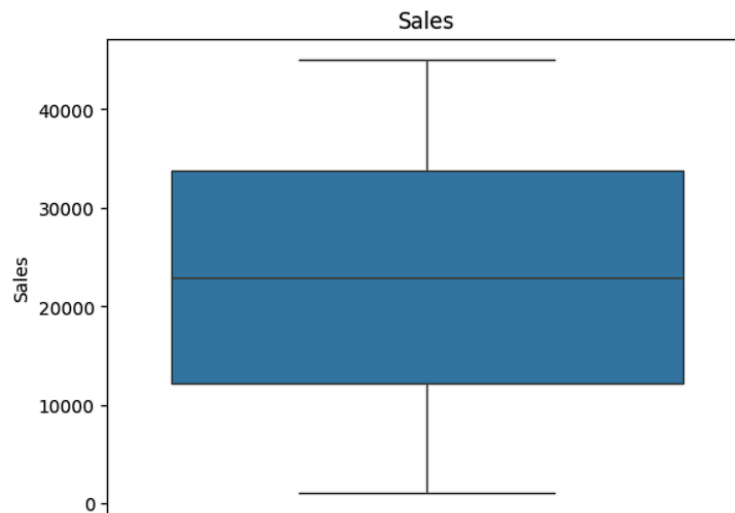
Вывод: “Наиболее часто покупаемые товары” неизменны на протяжении долгого времени.

- 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 $\pm 3\sigma$ 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.

```
# Разобьем на группы по продажам
df['Sale_group'] = pd.cut(df['Sales'], bins=3, labels=False)
df
# 0 - small; 1 - average ; 2 - high
```

	ID	Order Date	Class	Region	Sub-Category	Sales	Sale_group
0	646	2018-12-30	Стандарт	Урал	Appliances	37229.08	2
1	907	2018-12-30	Стандарт	Владивосток	Bookcases	20994.61	1
2	908	2018-12-30	Стандарт	Владивосток	Phones	42133.86	2
3	909	2018-12-30	Стандарт	Владивосток	Binders	10612.0	0
4	1297	2018-12-30	Стандарт	Калининград	Binders	28781.13	1
...
9795	1760	2015-01-05	Стандарт	Владивосток	Art	8639.472	0
9796	740	2015-01-04	Стандарт	Урал	Labels	15246.05	0
9797	741	2015-01-04	Стандарт	Урал	Storage	20415.81	1
9798	742	2015-01-04	Стандарт	Урал	Binders	4559.085	0
9799	7081	2015-01-02	Стандарт	Урал	Books	11726.47	0

б) Сгруппировать данные на основе региона и группы продаж (Region, Sale_group).

		Sales					
		sum	mean	median	min	max	count
Region	Sale_group						
Владивосток	0	7862072.555	8302.082951	8289.795	1003.082	15658.86	947
	1	20641092.62	23088.470492	23063.745	15671.52	30326.24	894
	2	35449521.71	37552.459439	37431.55	30350.8	44990.44	944
Калининград	0	8889168.063	8555.503429	8660.463	1021.104	15651.07	1039
	1	25680862.64	23135.912288	23133.44	15670.85	30319.03	1110
	2	37360269.02	37699.565106	37556.83	30339.08	44986.26	991
Москва	0	4287843.364	8245.852623	8099.622	1012.536	15663.45	520
	1	12648114.08	23080.500146	23180.55	15668.8	30328.97	548
	2	19955694.07	37652.252962	37759.545	30346.04	44995.41	530
Урал	0	6434909.756	8389.71285	8653.412	1002.794	15653.63	767
	1	17588614.03	22961.637115	22685.375	15740.69	30323.42	766
	2	27835063.91	37412.720309	37377.205	30334.84	44942.31	744

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.