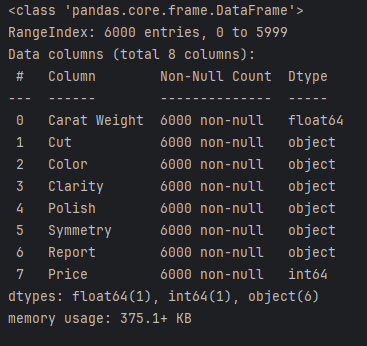
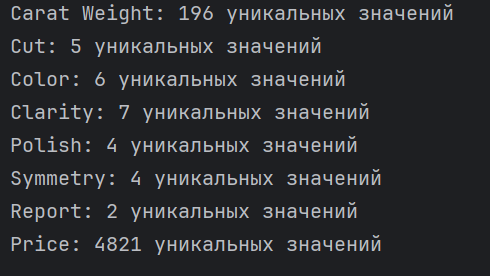
Анализ БД “Diamond.csv”

1. df.info()





Carat - **Дискретные переменные.**

Cut - **Дискретные переменные.**

Color **Дискретные переменные.**

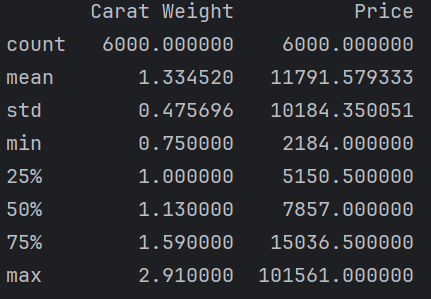
Clarity - **Дискретные переменные.**

Polish - **Дискретные переменные.**

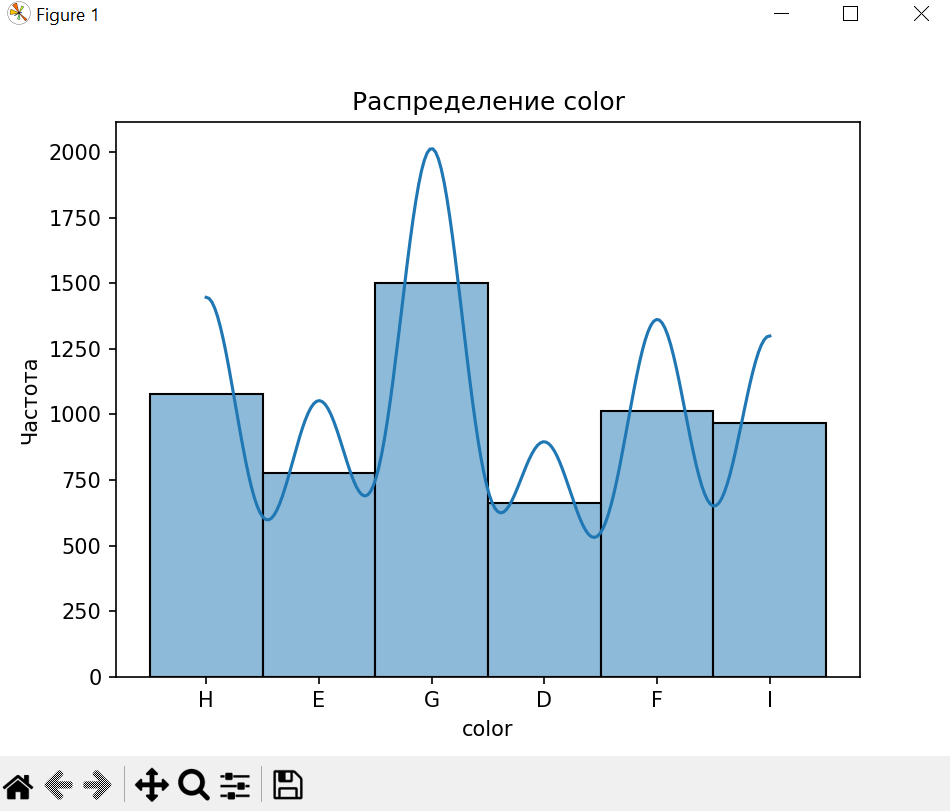
Symmetry - **Дискретные переменные.**

Report - **Дискретные переменные.**

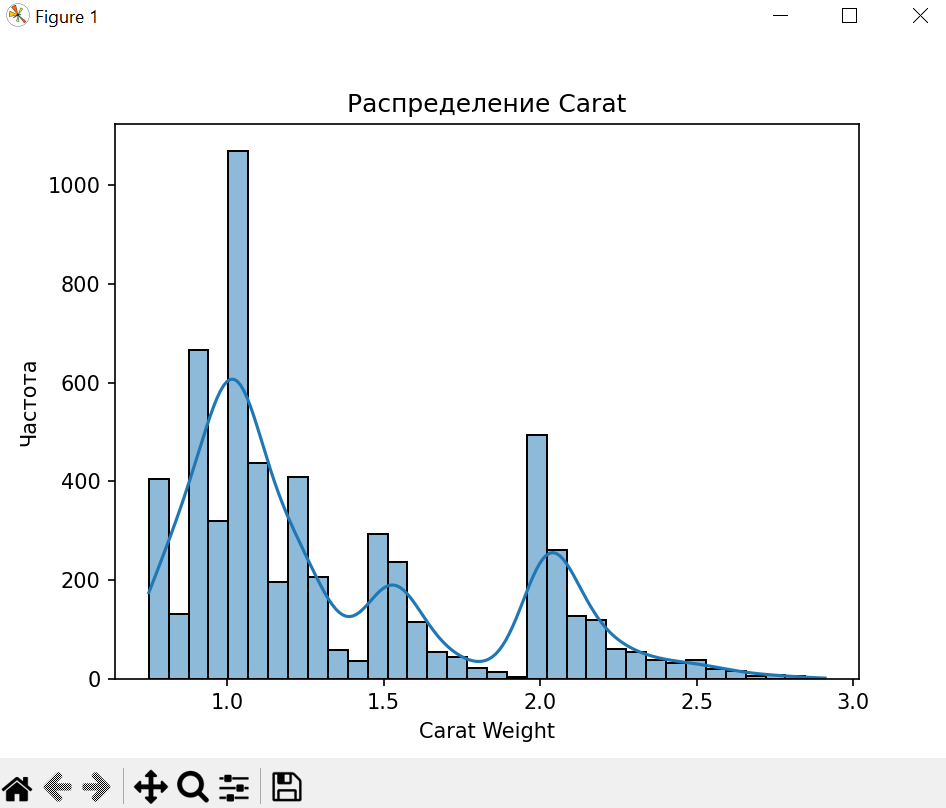
Price - **Непрерывные переменные.**



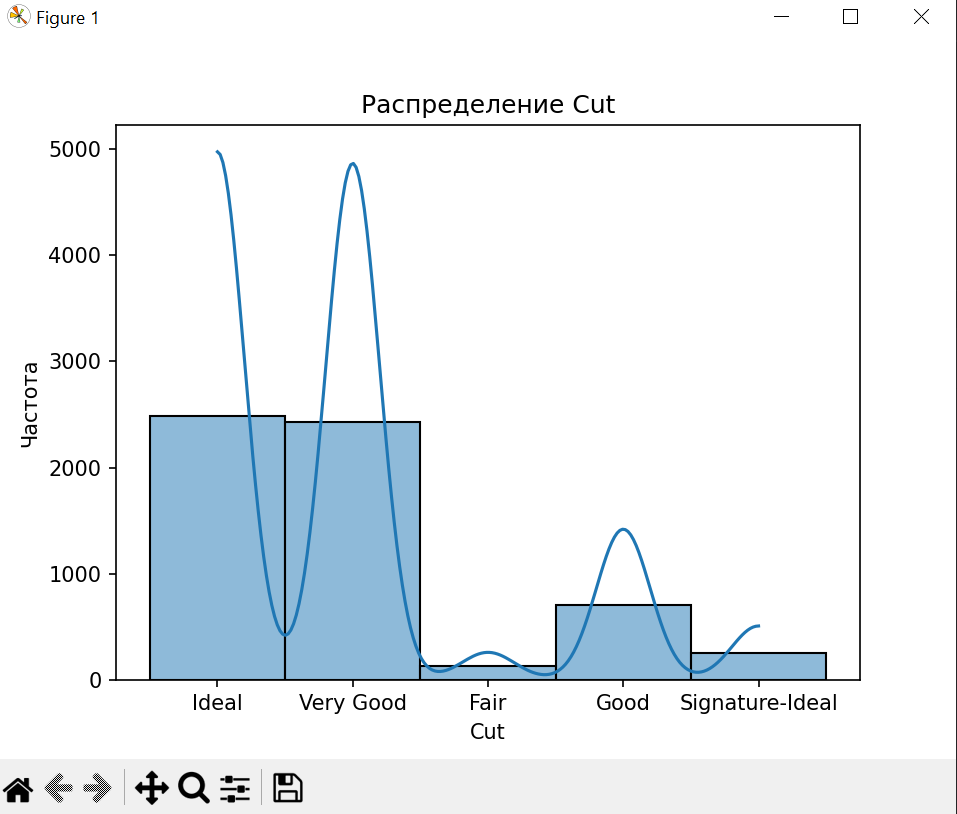
Медиана делит статистику на отметке 1.130000 7857.000000



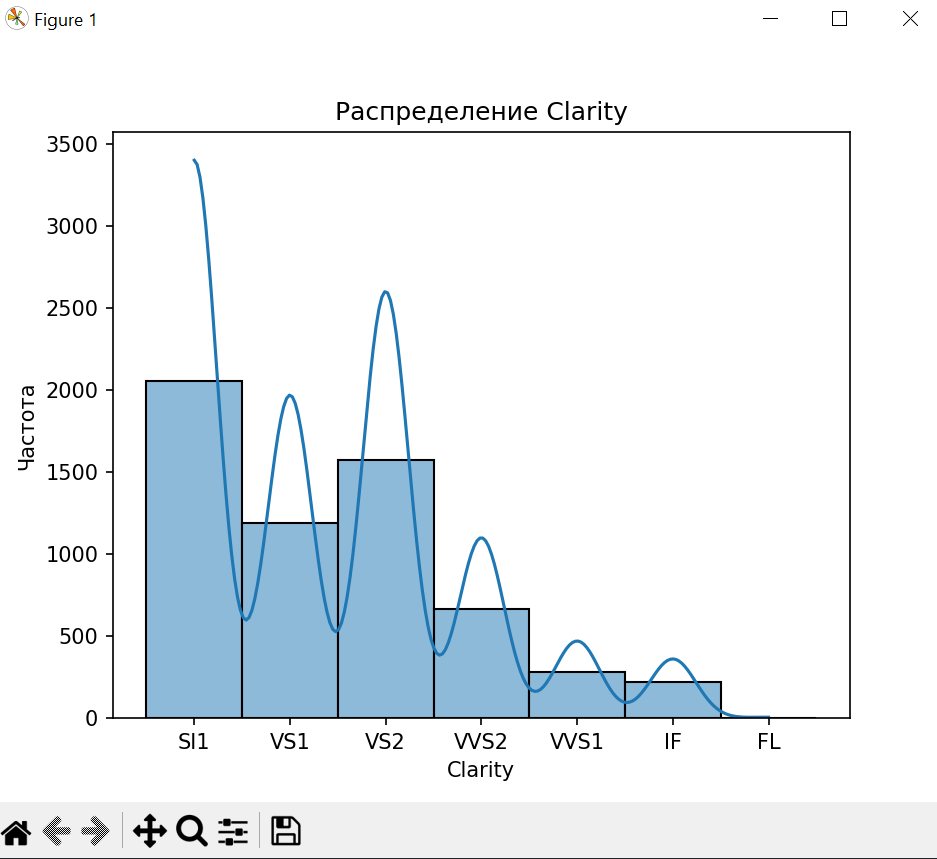
В цвете мы не наблюдаем нормального распределения. Но цвет «G» - преобладает.



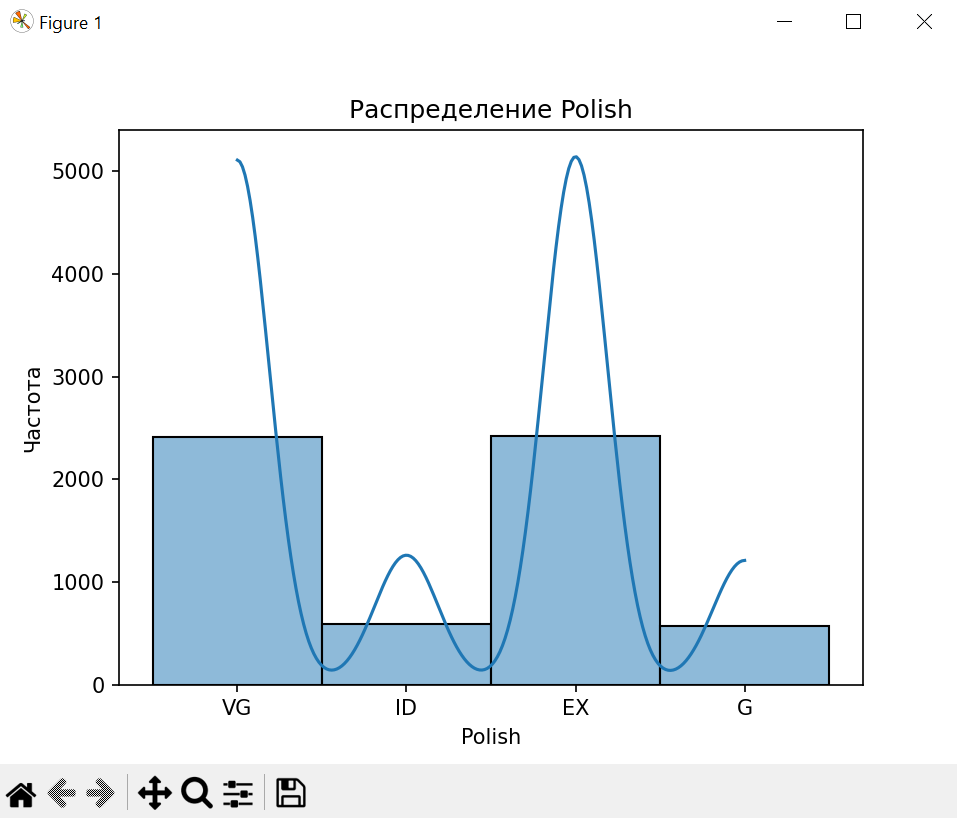
На графике мы можем наблюдать бимодальное распределение. Моду захватывает распределение в окрестности значения «1» и «2». Наиболее часто повторяемый признак на значении «1».



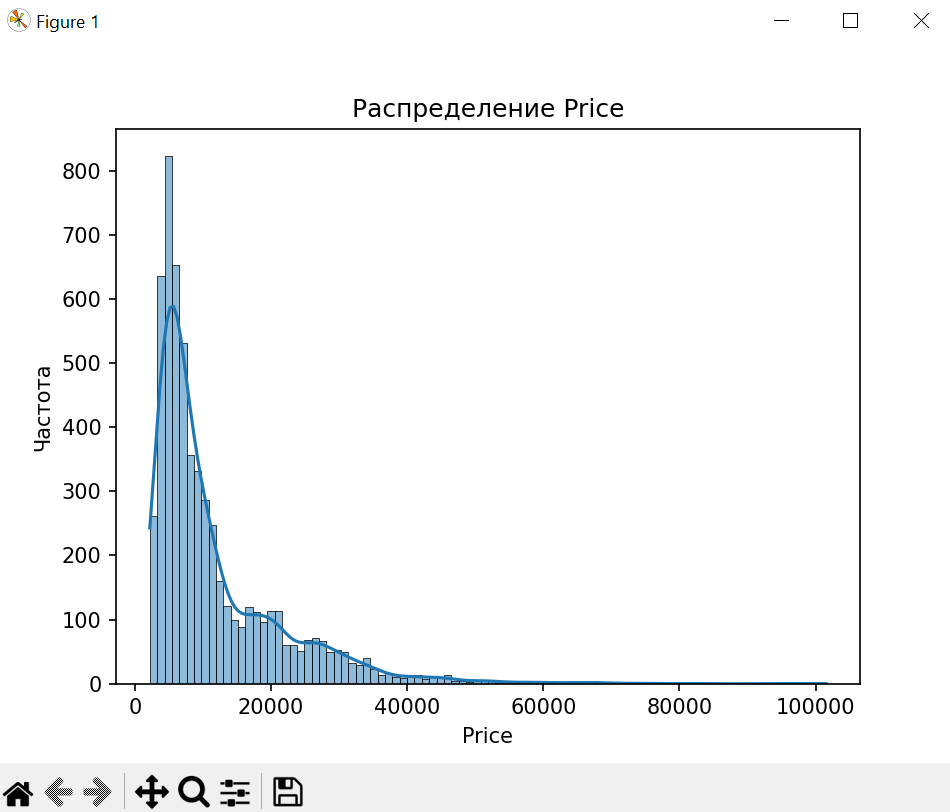
Вновь мы встречаем бимодальное распределение между «ideal» и «Very good»



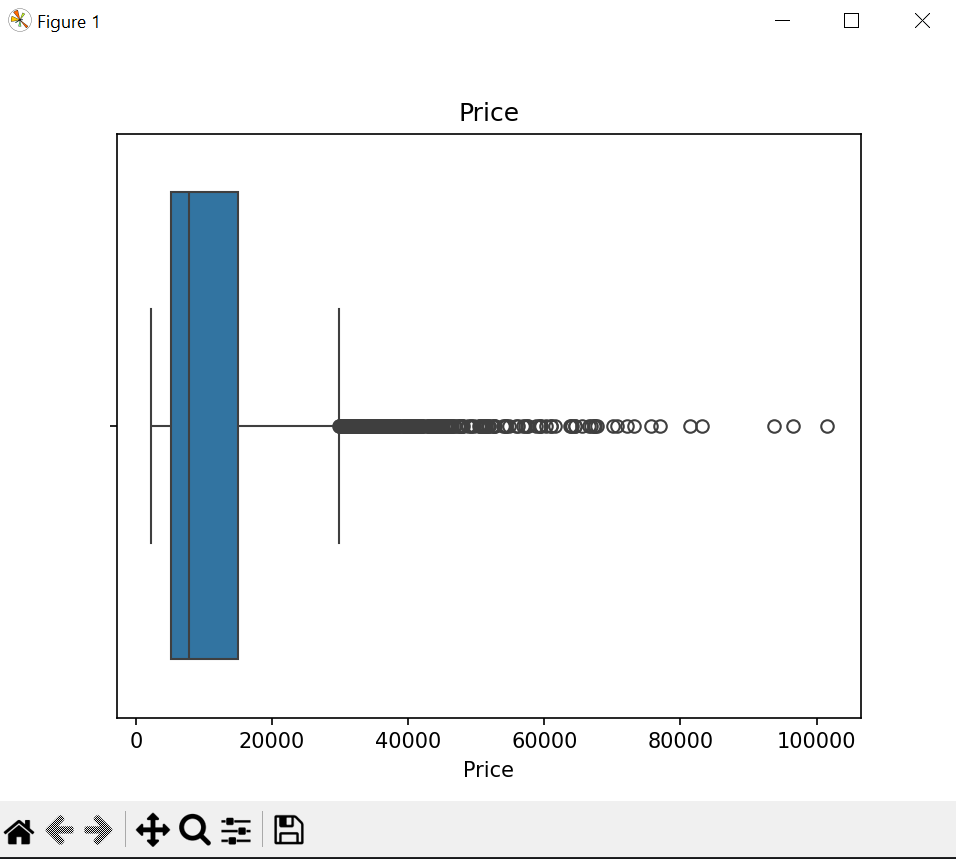
Большинство значений в Clarity уходит в SI1, VS2, VS1. FL – Можно не учитывать.



Polish: Бимодальное распределение между «VG» и «EX».

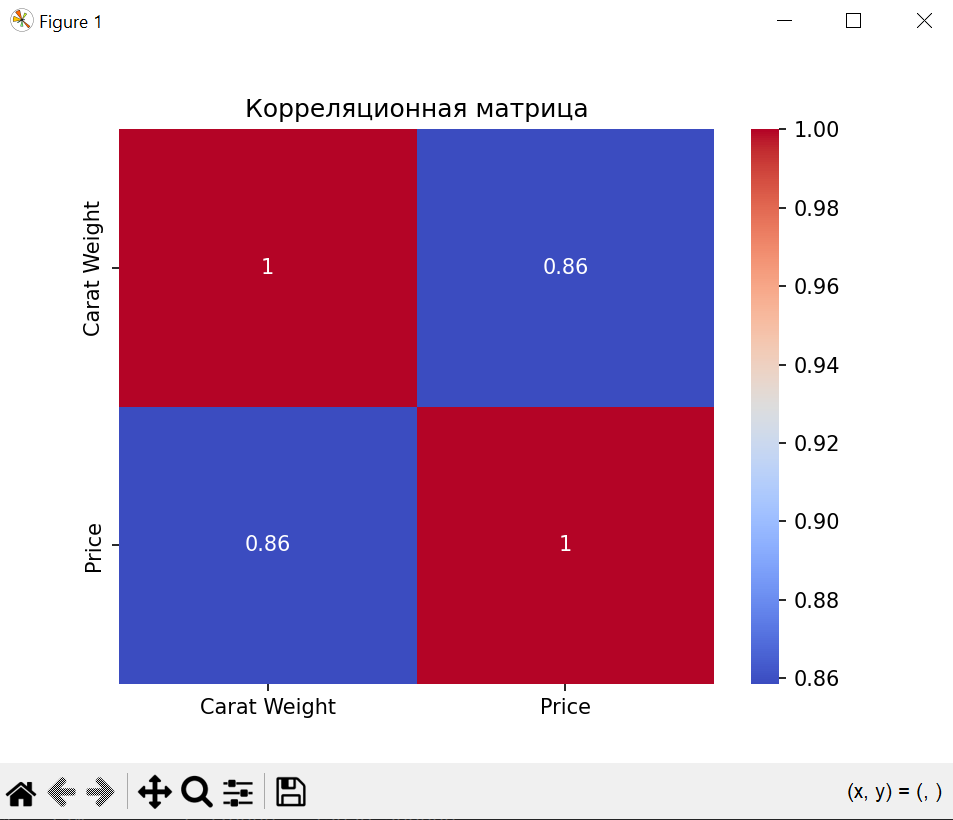


Price: Нормальное распределение до 20000. После 40000 – уникальные экземпляры, которые можно не учитывать.



На данном графике мы отчётливо видим медиану и наиболее частые значения близ отметки 9000. Выбросами считаются все значения после 30000.

**Корреляционная модель.**

****

**В данной модели мы можен наблюдать, что price напрямую зависит от Carrat weight.**