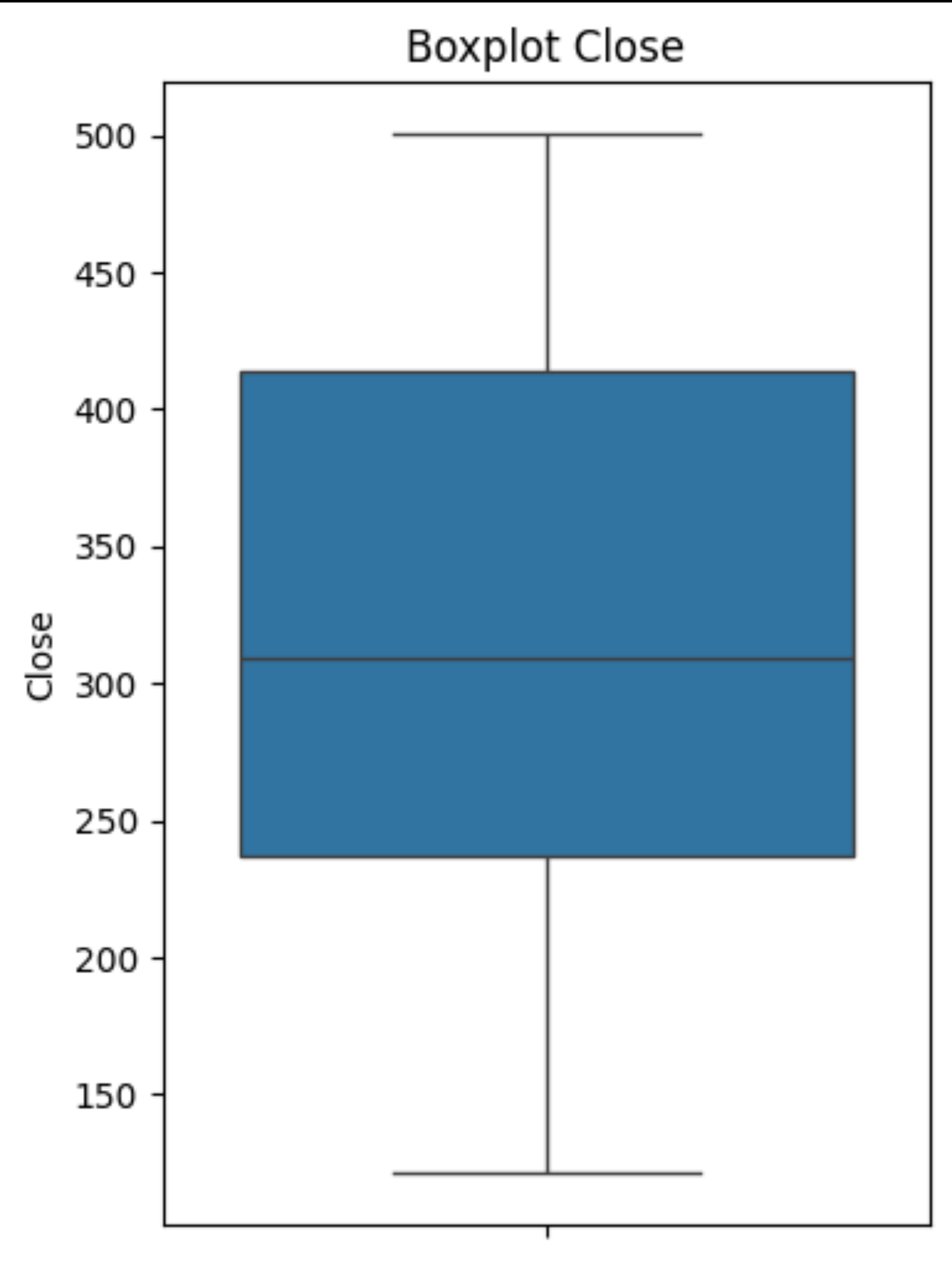
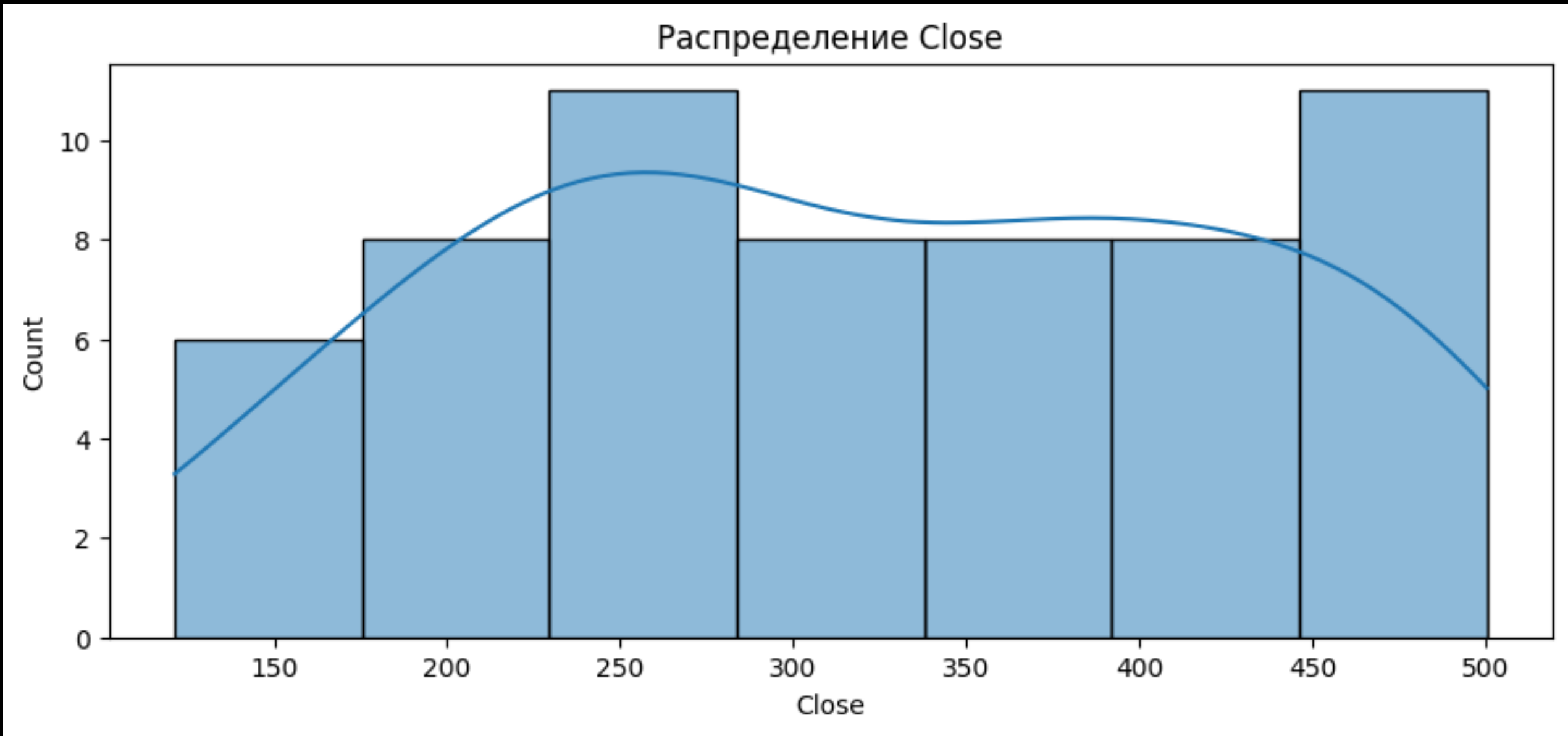


Stock Market Dataset

Динамика цены закрытия (Close) за период



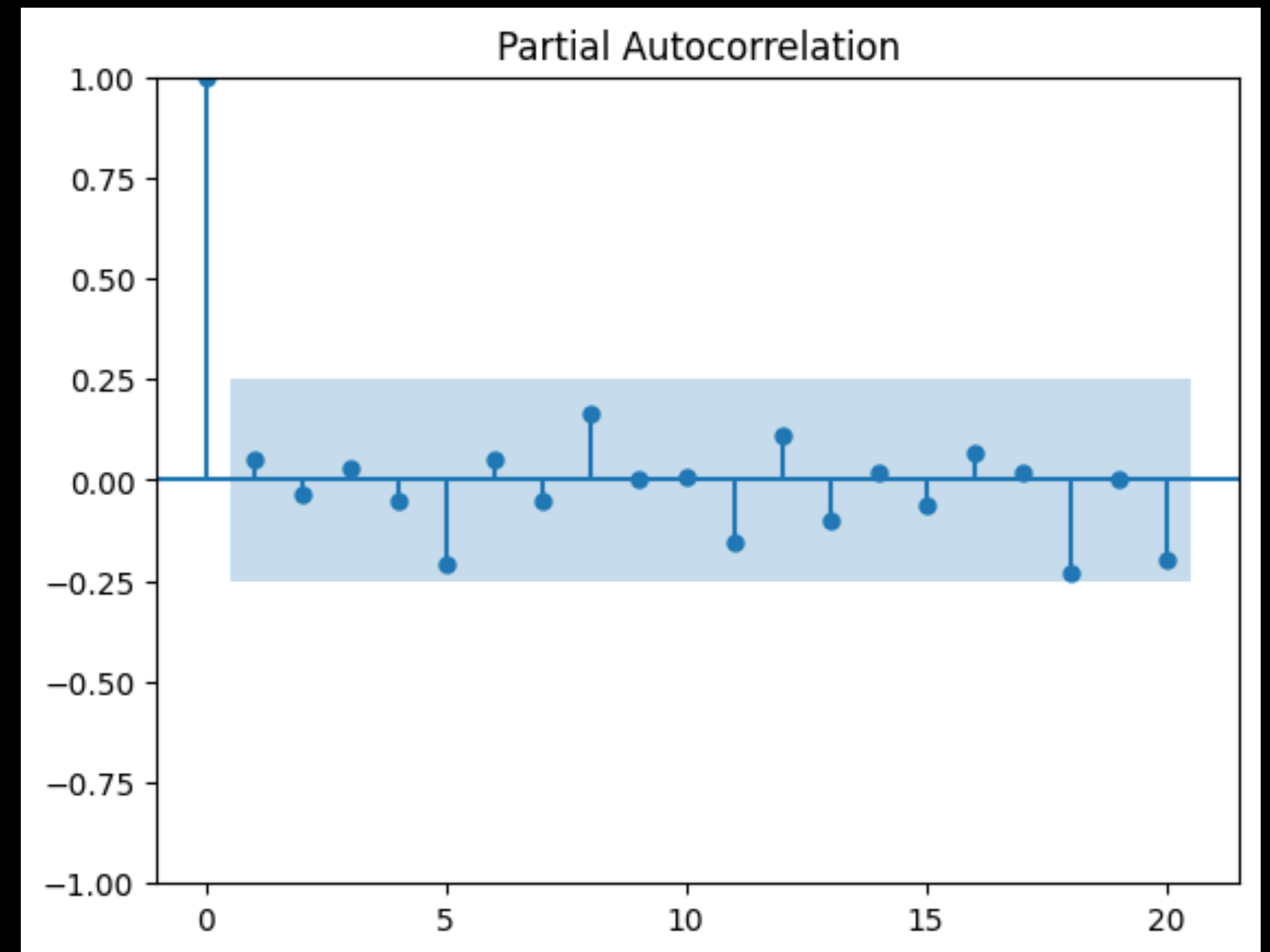
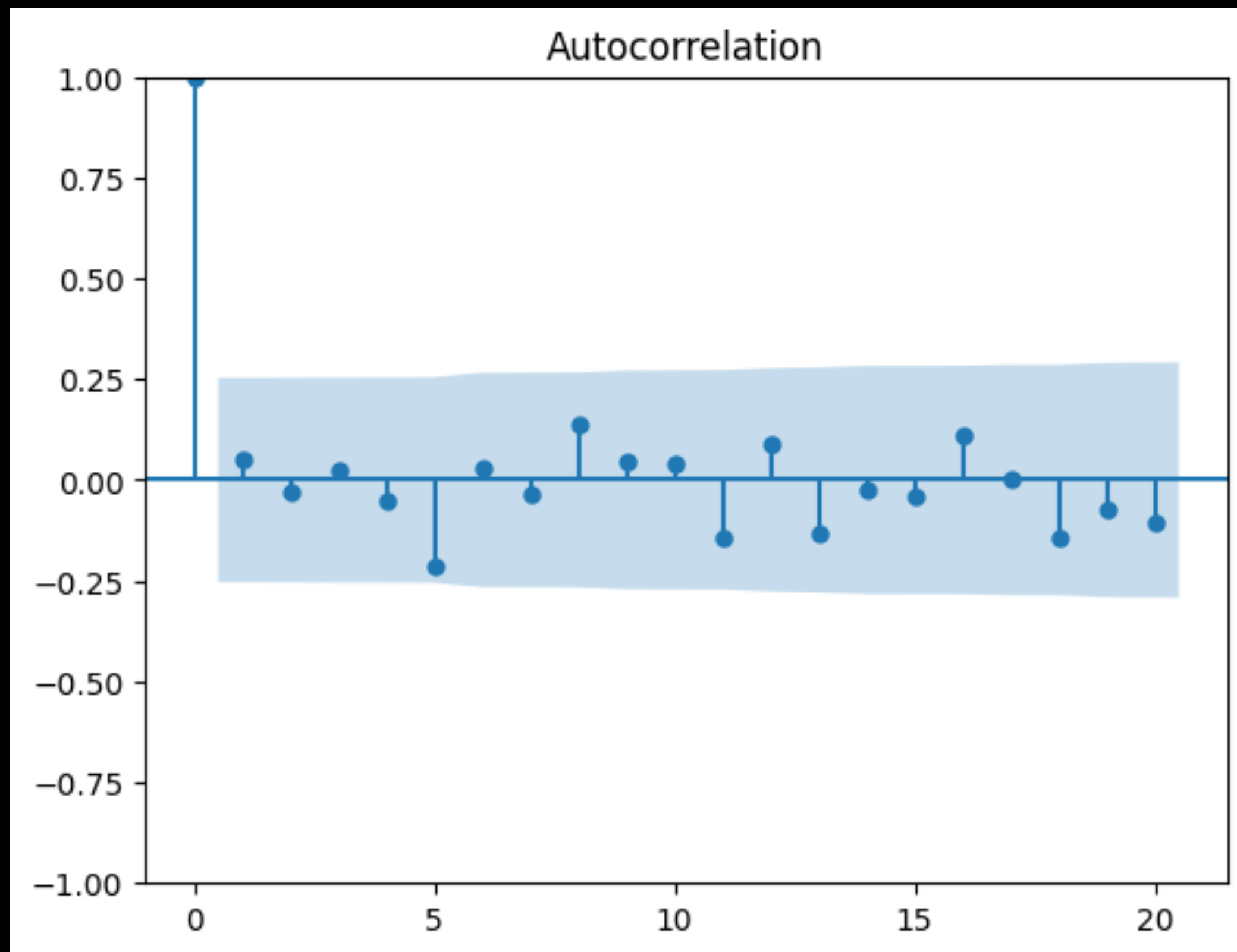
Распределение Close



Корреляционная матрица исходных числовых признаков



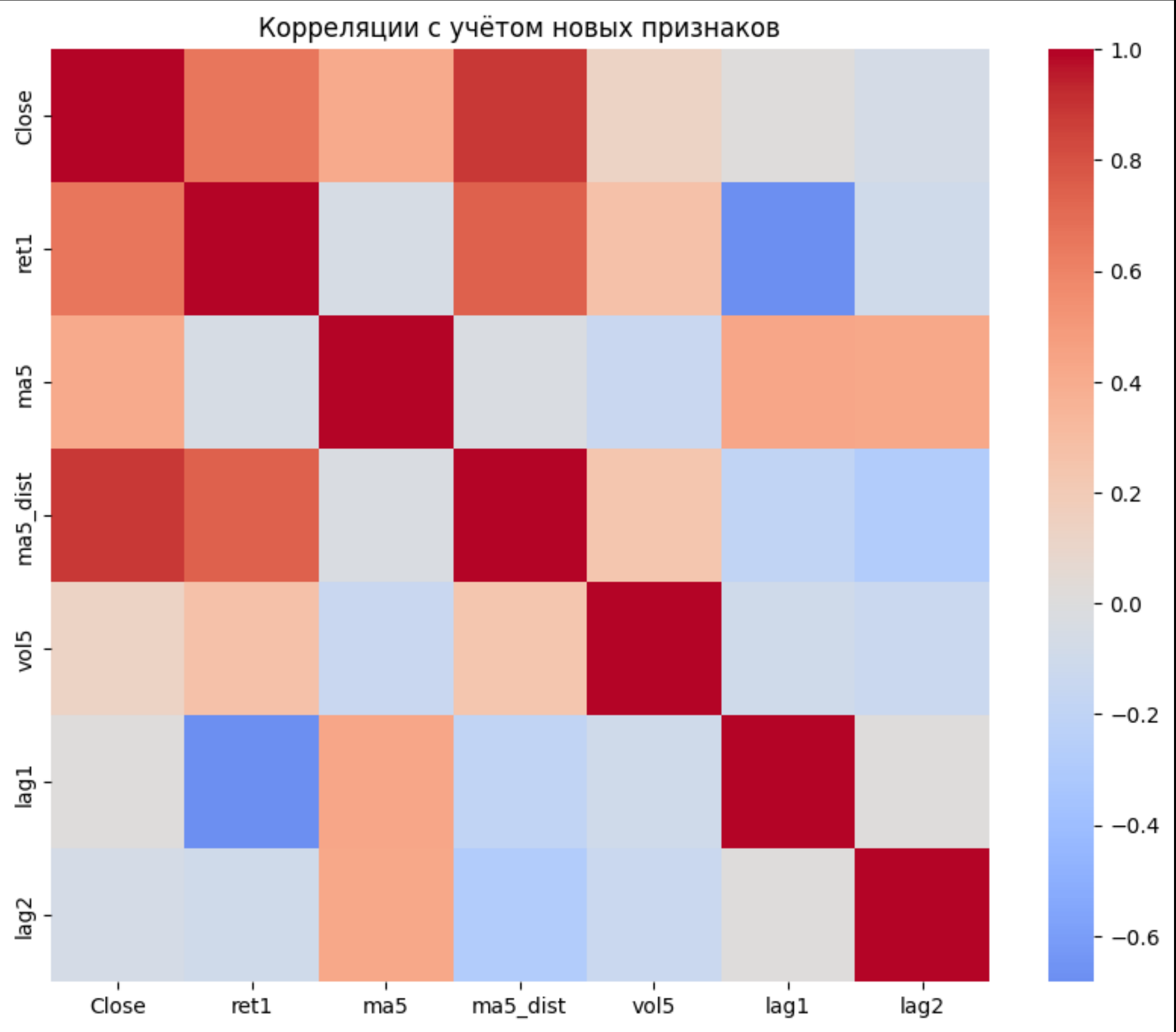
ACF и PACF



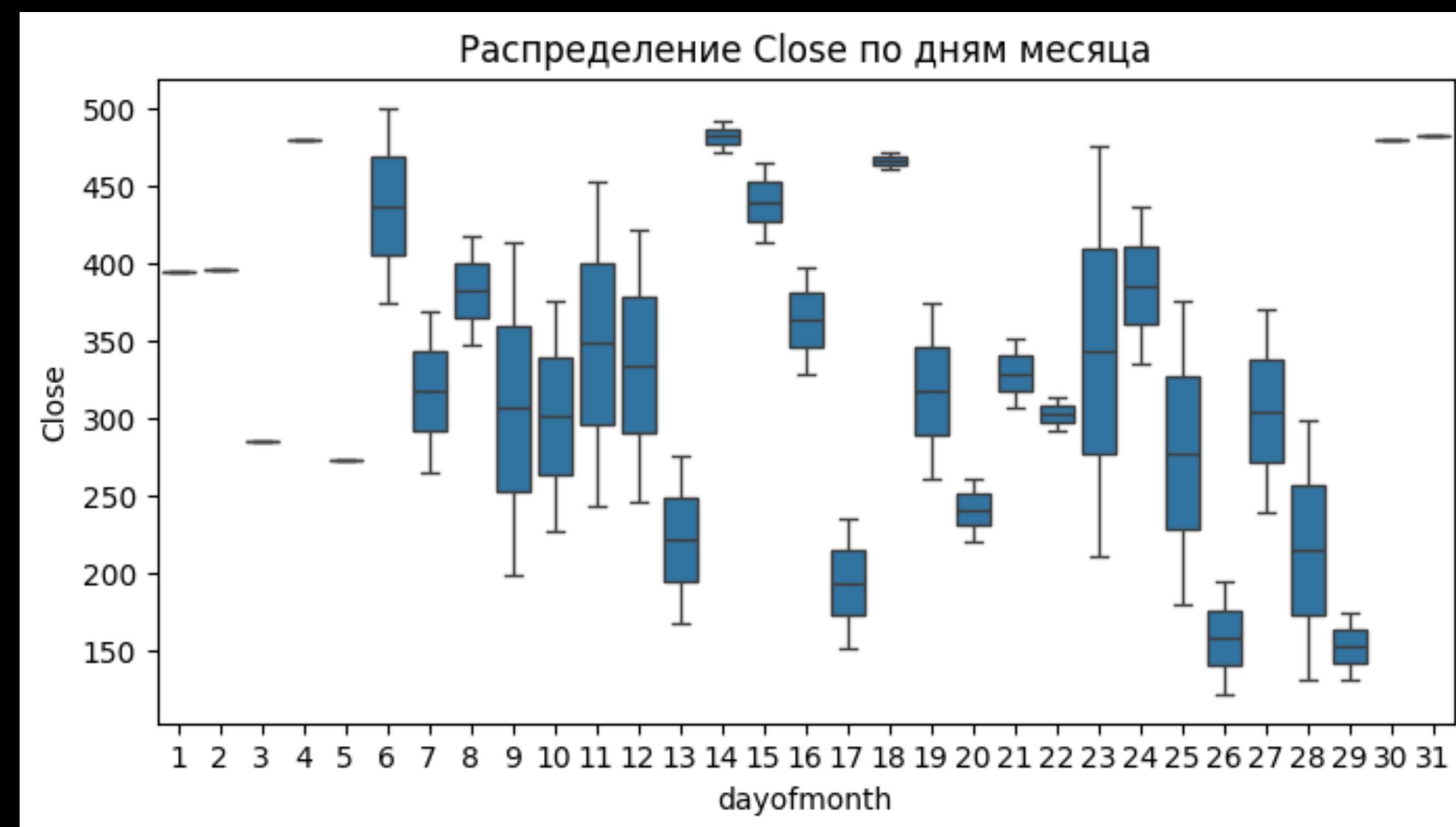
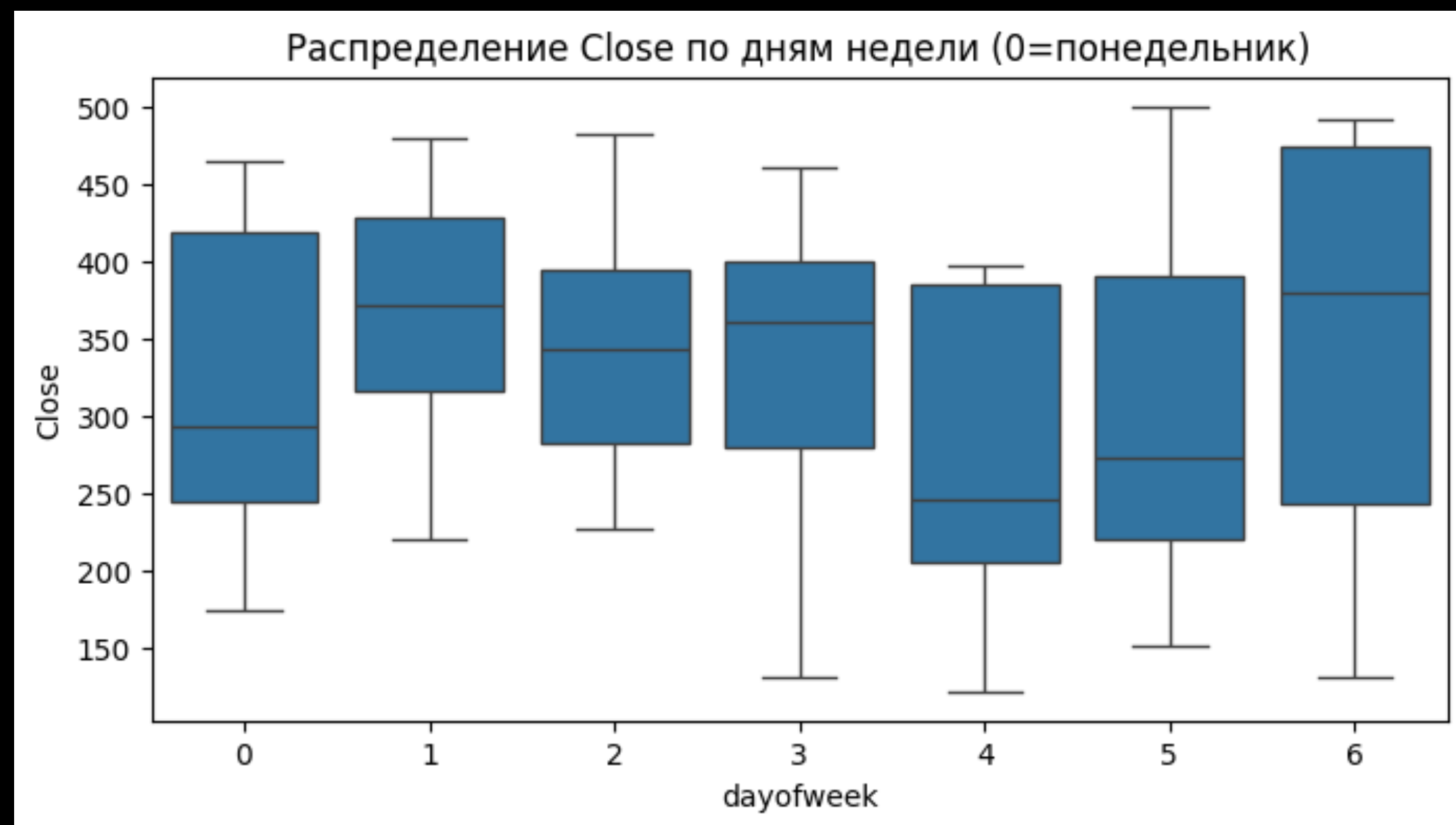
Feature engineering

- `dayofweek` — номер дня недели. Помогает уловить возможные календарные эффекты.
- `dayofmonth` — число месяца. Может отражать слабую сезонность внутри месяца.
- `ret1` — однодневная доходность. Показывает направление и силу последнего движения (рост/падение)
- `ma5` — 5-дневная скользящая средняя цены закрытия. Гладкий «локальный тренд» за последние несколько дней, сглаживает шум.
- `ma5_dist` — отклонение текущей цены от `ma5`. Показывает, насколько цена сейчас выше или ниже своего локального тренда.
- `vol5` — оценка волатильности за 5 дней (стандартное отклонение однодневной доходности). Характеризует, насколько сильно колебалась цена в недавнем прошлом.
- `lag 1, 2` — запаздывающее значение ряда, то есть значение признака в прошлом.

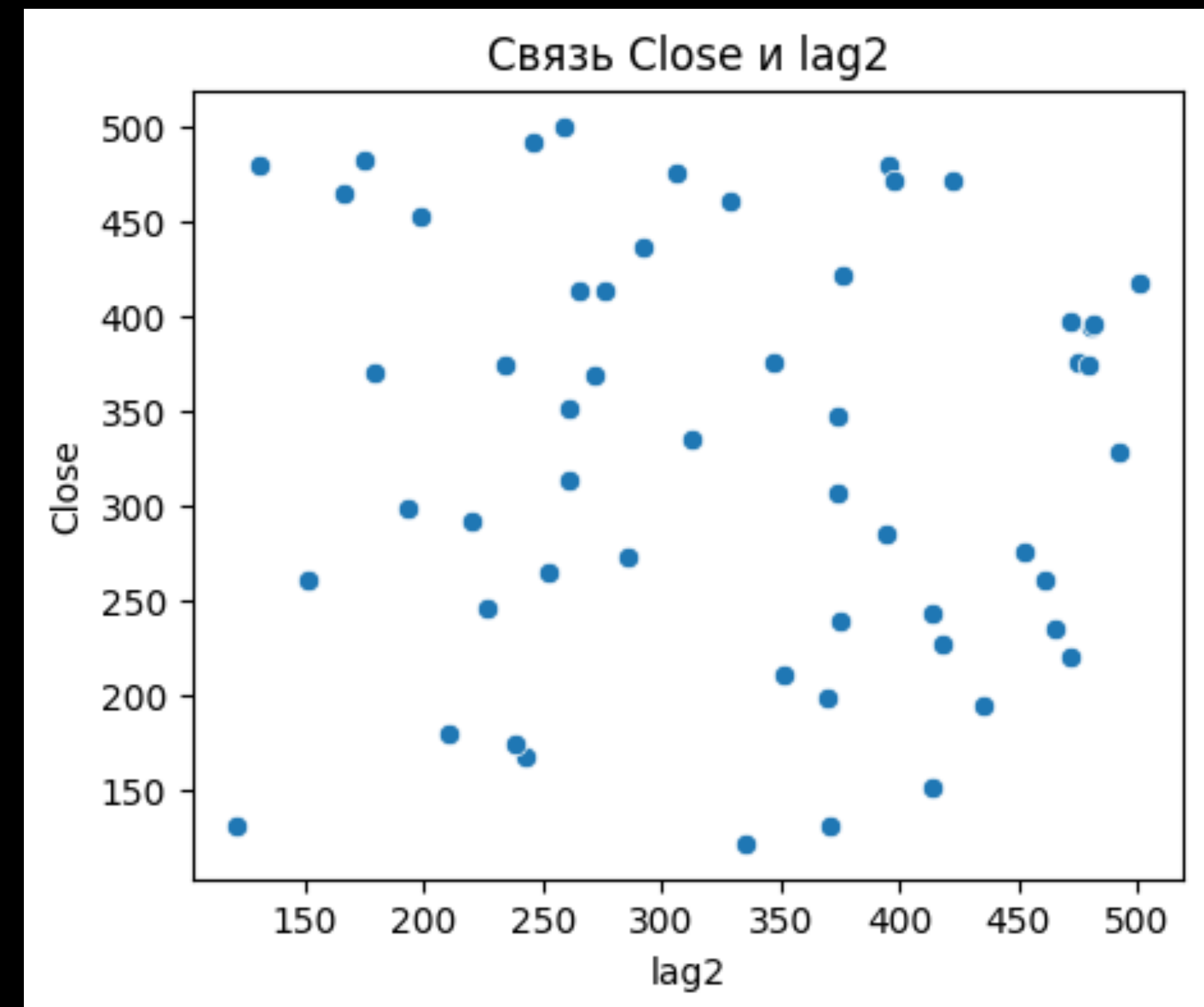
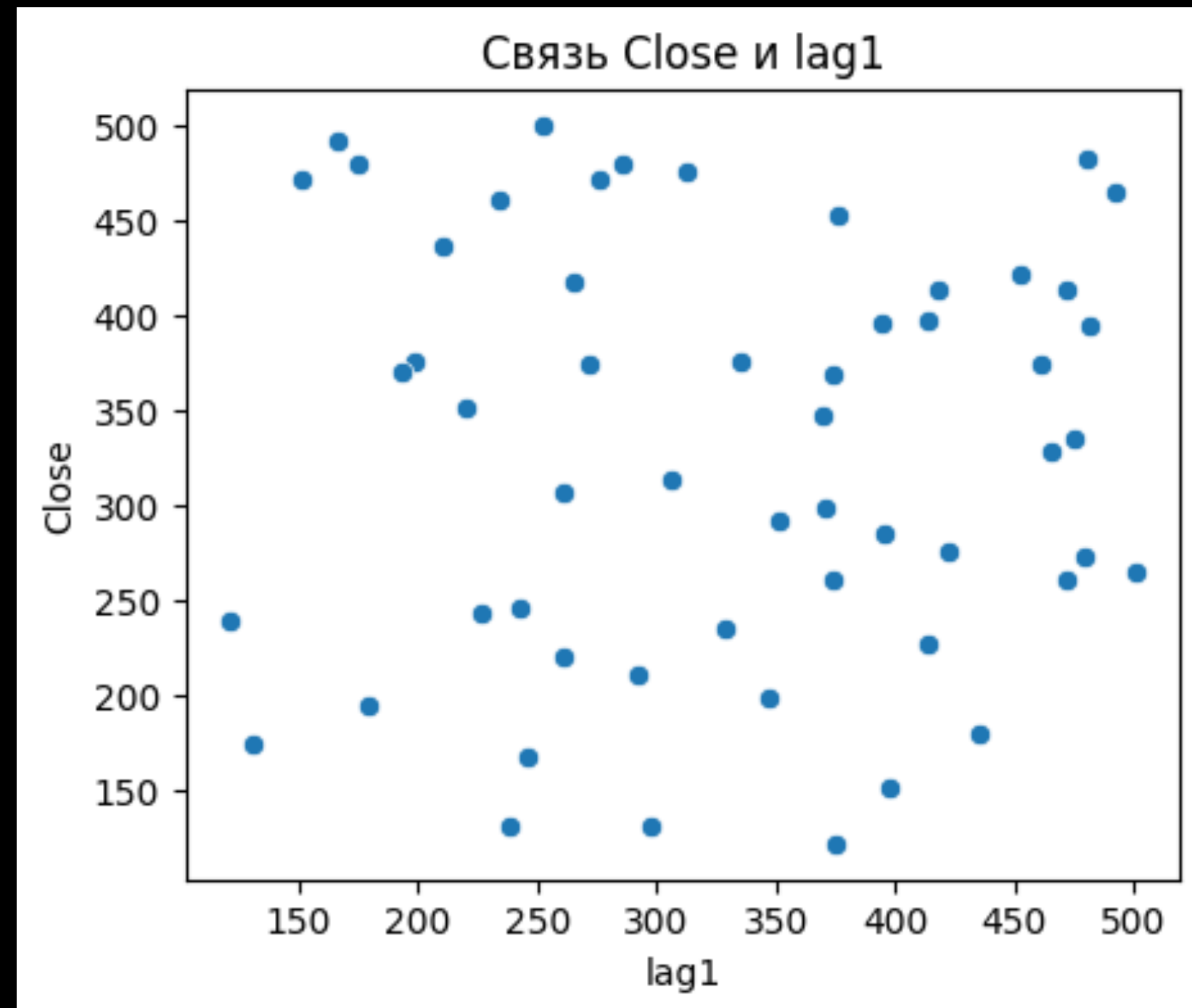
Корреляционная матрица с новыми признаками



Распределение Close по дням



Связь Close и lag



Вывод по EDA

В рамках feature engineering были добавлены лаговые признаки lag1–lag5 (цена закрытия за 1–2 предыдущих дней) и календарный признак dayofweek.

Анализ корреляций показал, что линейная автокорреляция ряда слабая, а лаги не дублируют друг друга и целевую переменную. Boxplot по дням недели указывает на возможные отличия в уровне цен в разные дни недели, однако из-за малого объёма данных эти эффекты статистически ненадёжны.

Таким образом, созданные признаки не обладают сильной линейной предсказательной силой сами по себе, но могут быть полезны для нелинейной модели (XGBoost), которая способна извлечь более сложные зависимости между лагами, календарными эффектами и целевой переменной.

Оценка подбора гиперпараметров

method	RMSE	MAE	MAPE
GridSearch	140.217923	130.225697	59.594750
RandomSearch	134.987226	124.907141	58.577967
Optuna	120.579570	108.785819	52.420229

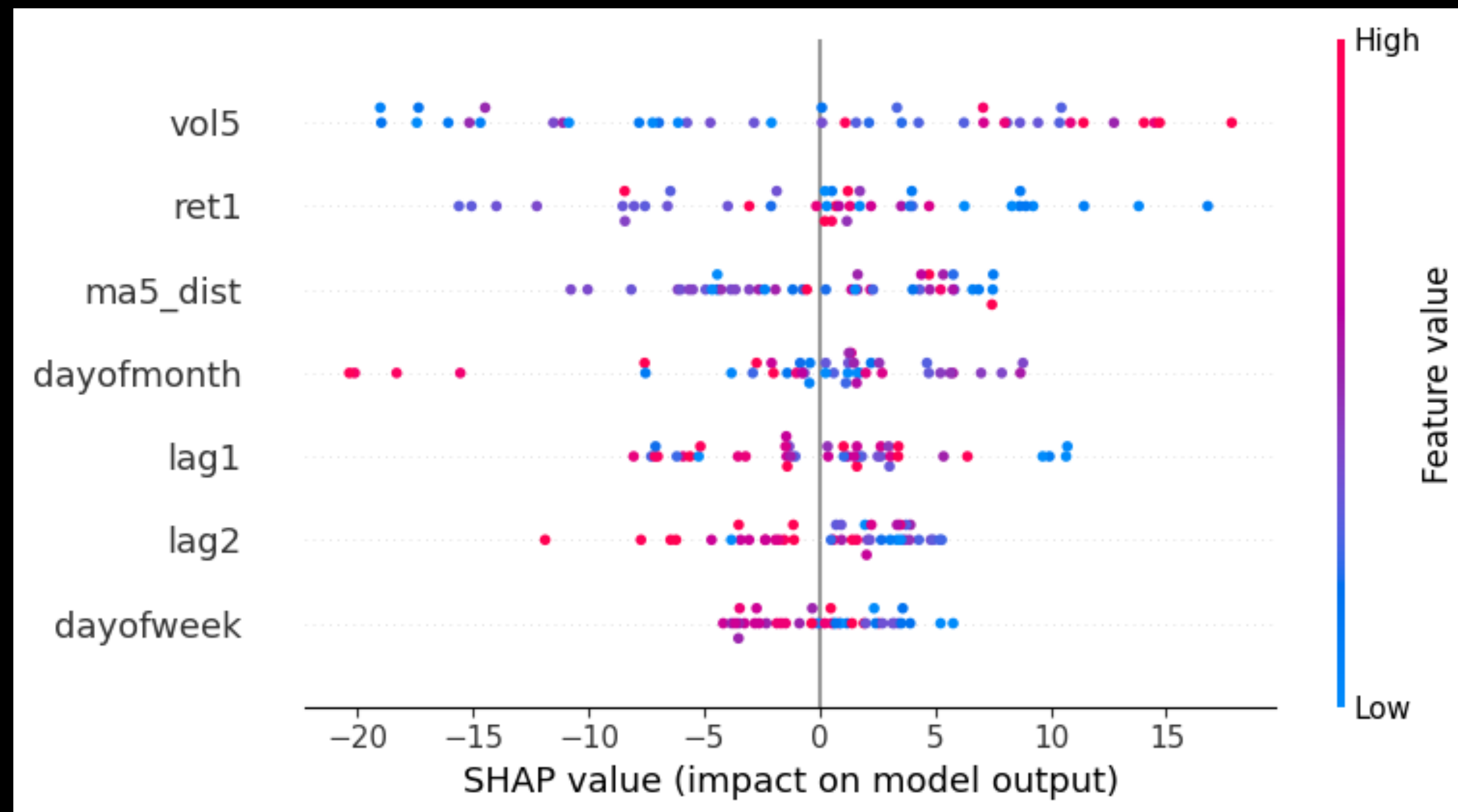
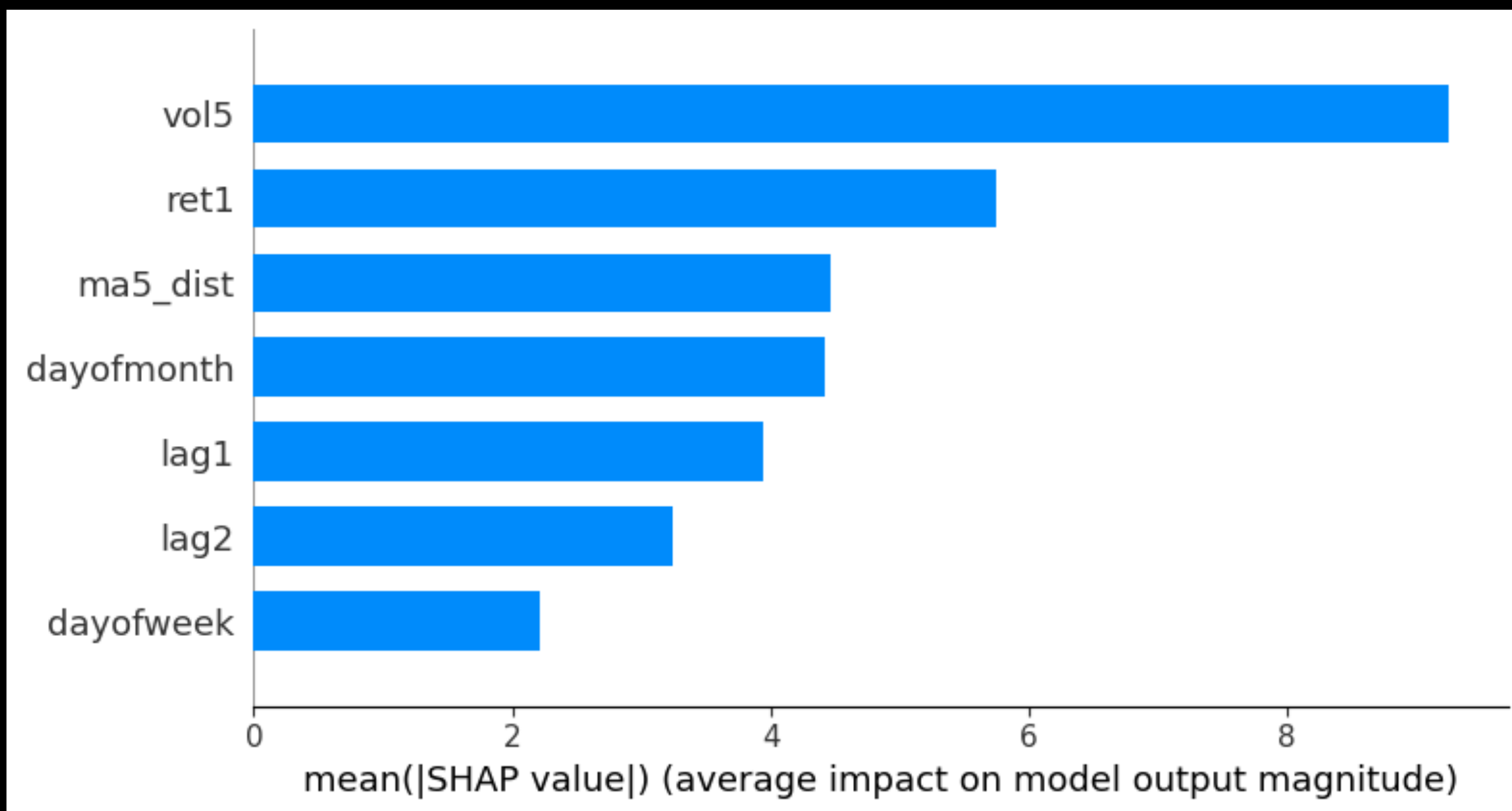
Анализ методов подбора гиперпараметров

Для подбора гиперпараметров XGBoost были протестированы три подхода: GridSearchCV, RandomizedSearchCV и Optuna.

Наилучшее качество показала модель с гиперпараметрами, найденными Optuna: RMSE снизился примерно на 14% по сравнению с GridSearch и на 11% по сравнению с RandomSearch, а MAPE сократился с ~59–60% до ~52%.

Таким образом, адаптивная байесовская оптимизация в Optuna оказалась существенно эффективнее классического перебора и случайного поиска, позволяя при том же числе запусков получить заметно более точную модель прогноза цены закрытия акций.

Shap значения



Вывод

SHAP-анализ показал, что для построенной модели XGBoost наибольшее влияние на прогноз цены закрытия оказывают недавняя волатильность (vol5) и однодневная доходность (ret1), а также отклонение цены от локального тренда (ma5_dist). Календарные признаки (dayofmonth, в меньшей степени dayofweek) и лаговые значения (lag1, lag2) дополняют информацию, но играют вспомогательную роль.