

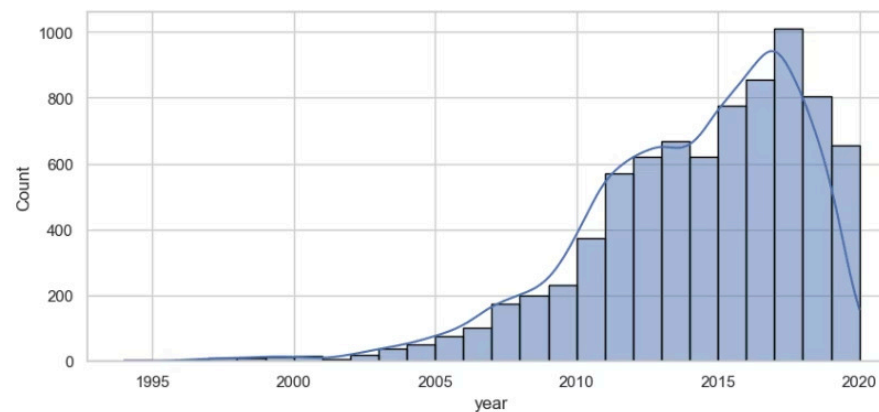
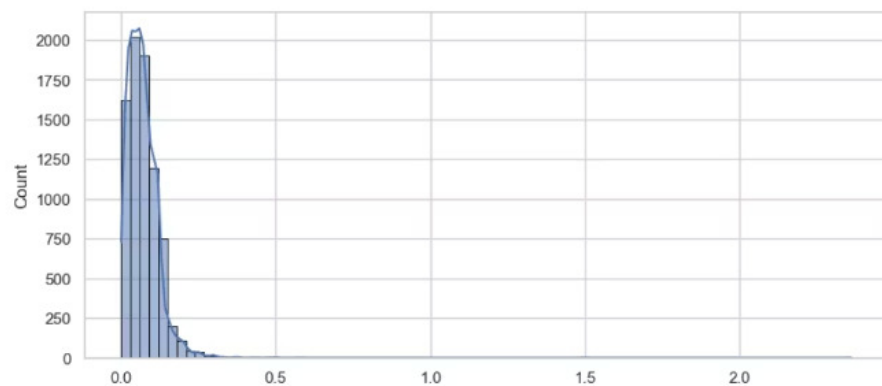
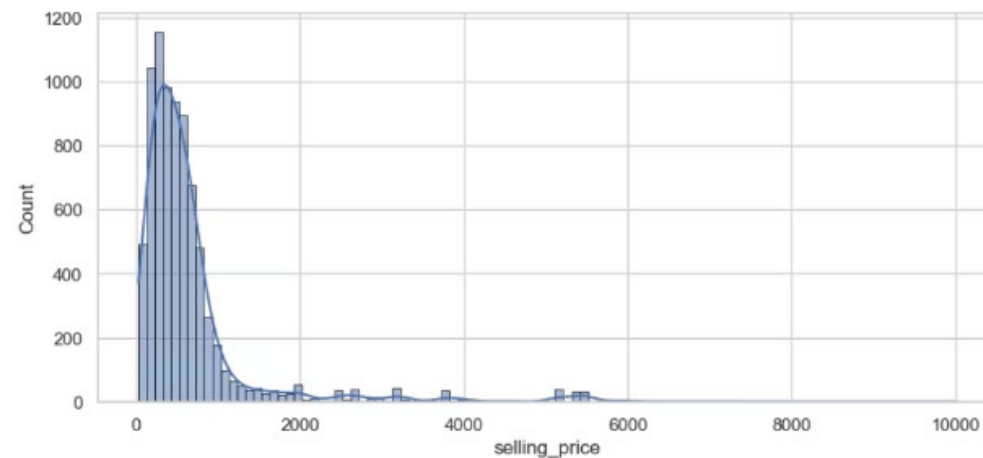
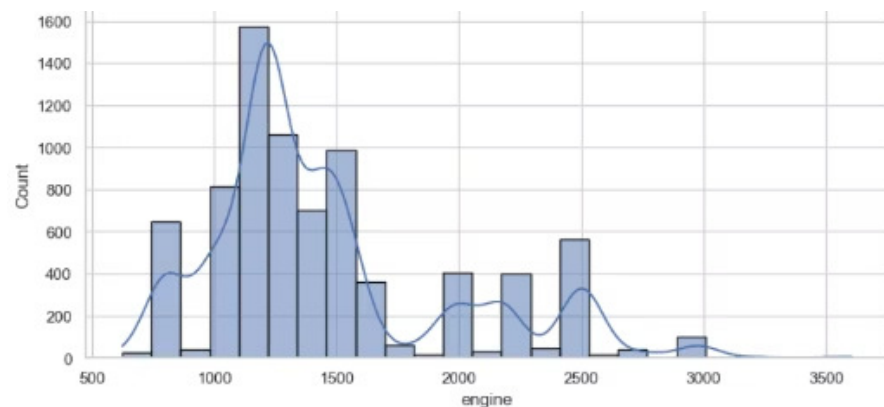
Прогнозирование цены подержанных автомобилей

Цель: разработка модели для предсказания стоимости подержанных автомобилей

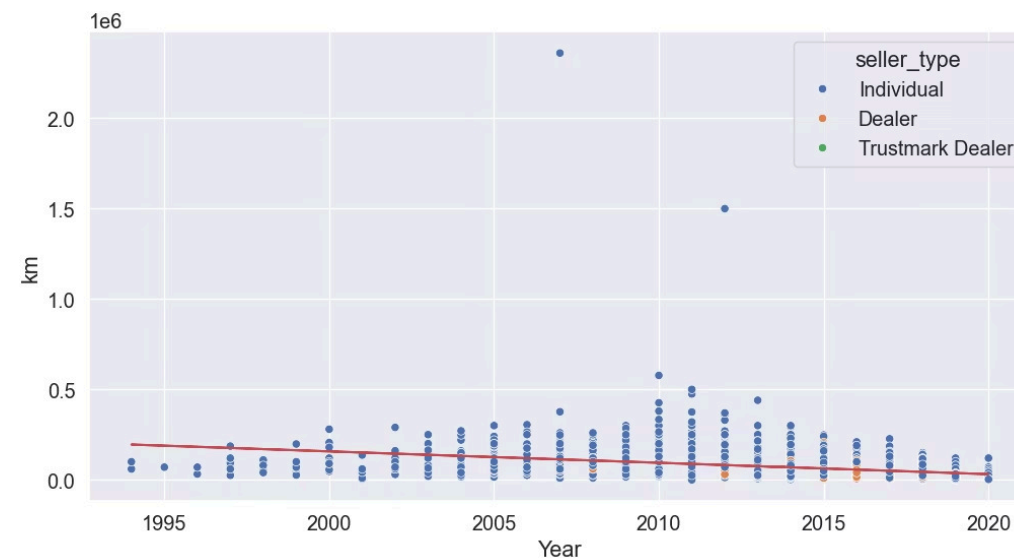
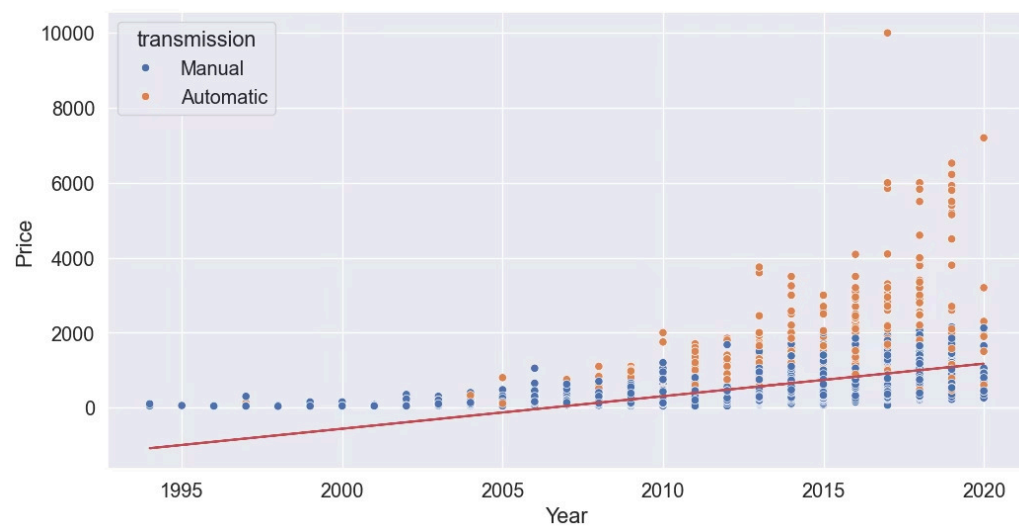
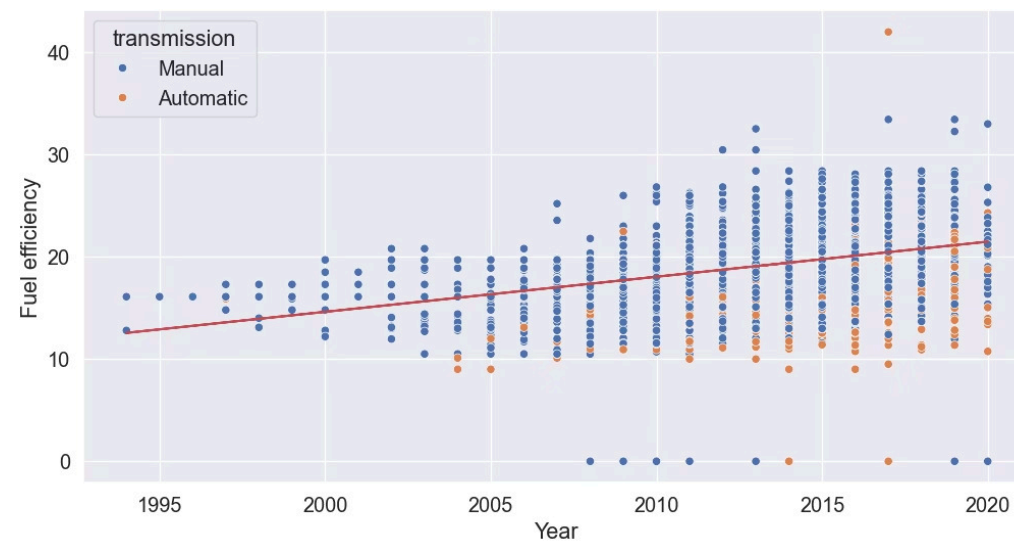
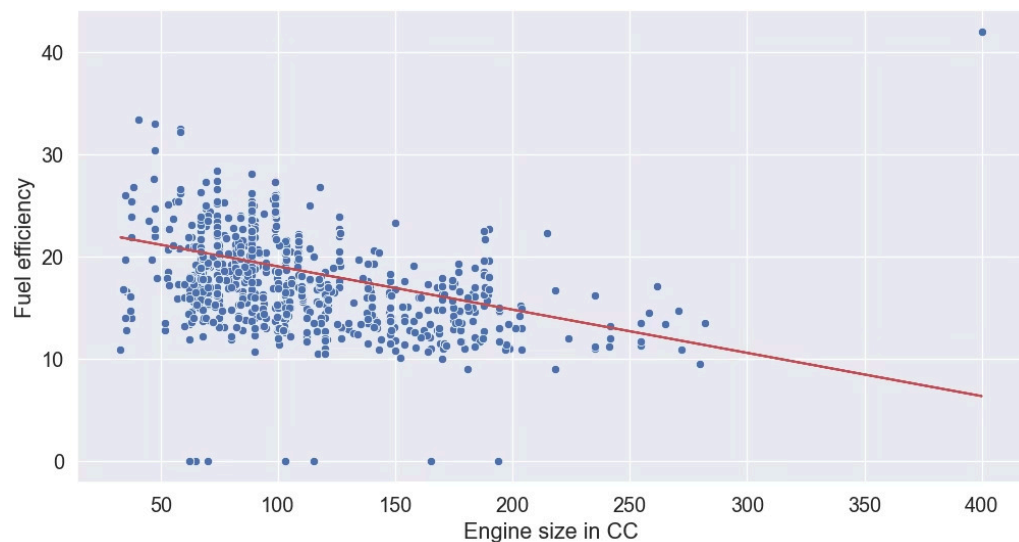
Датасет: 7906 автомобилей после очистки



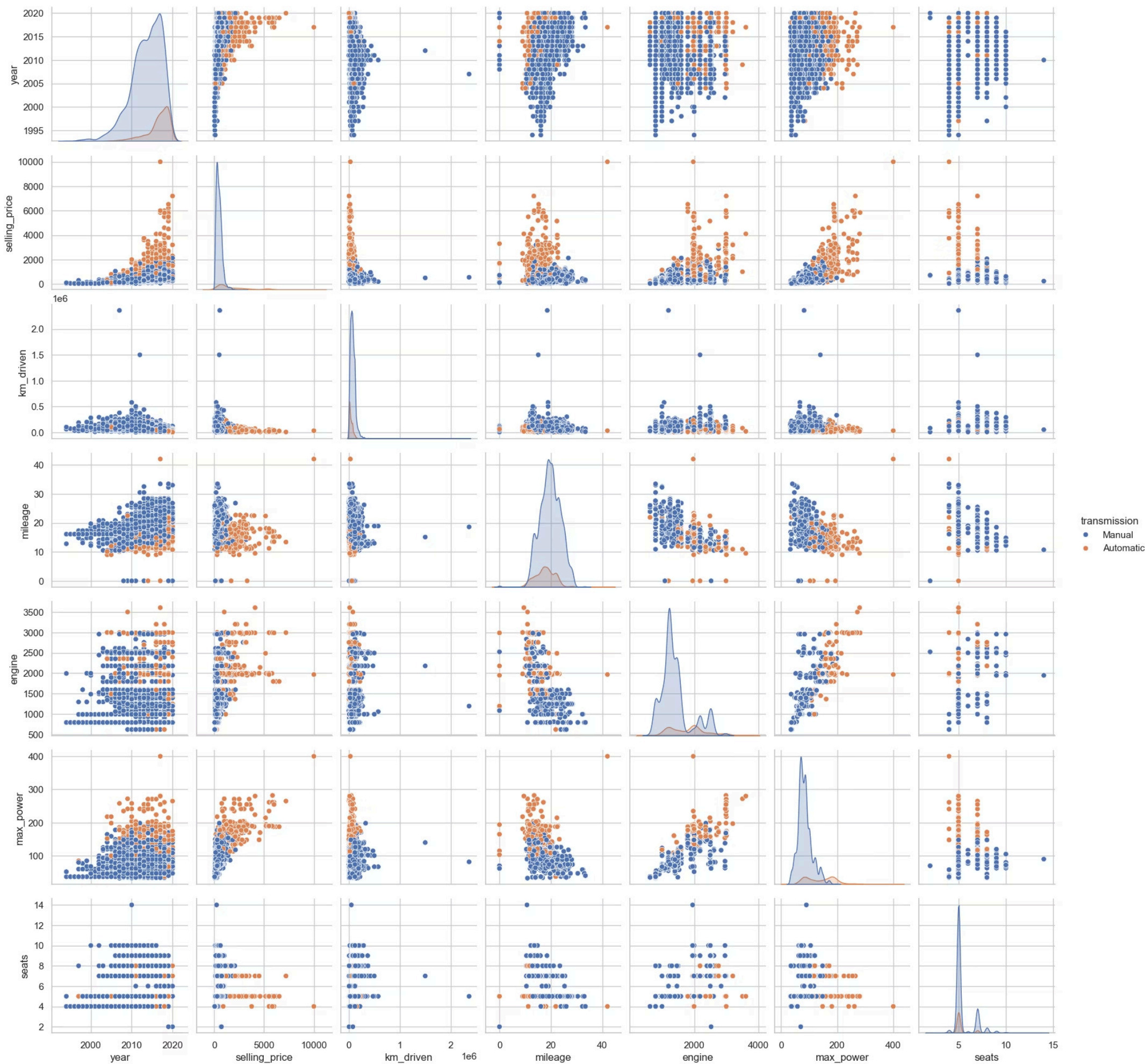
Распределение некоторых параметров



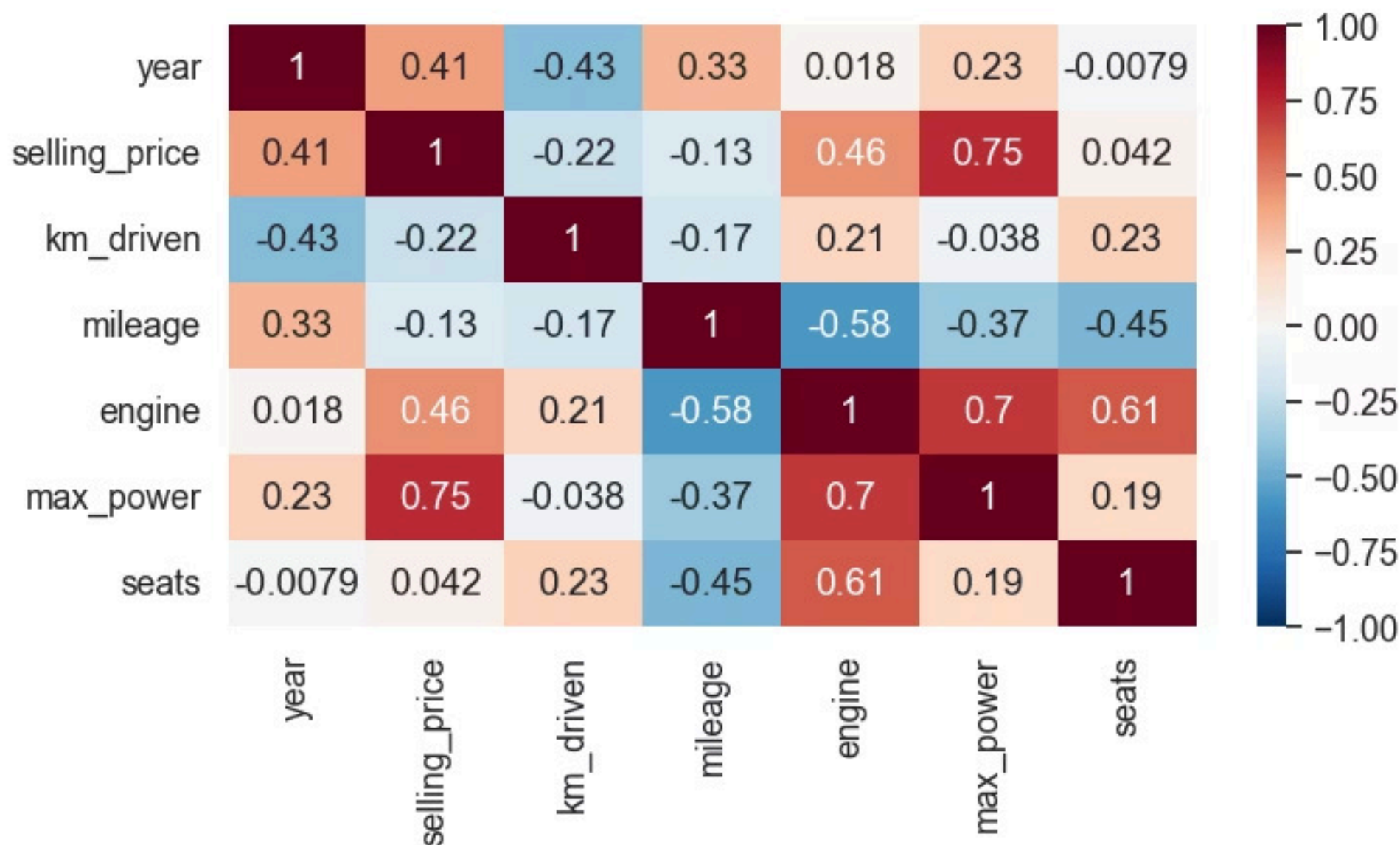
Зависимости между параметрами



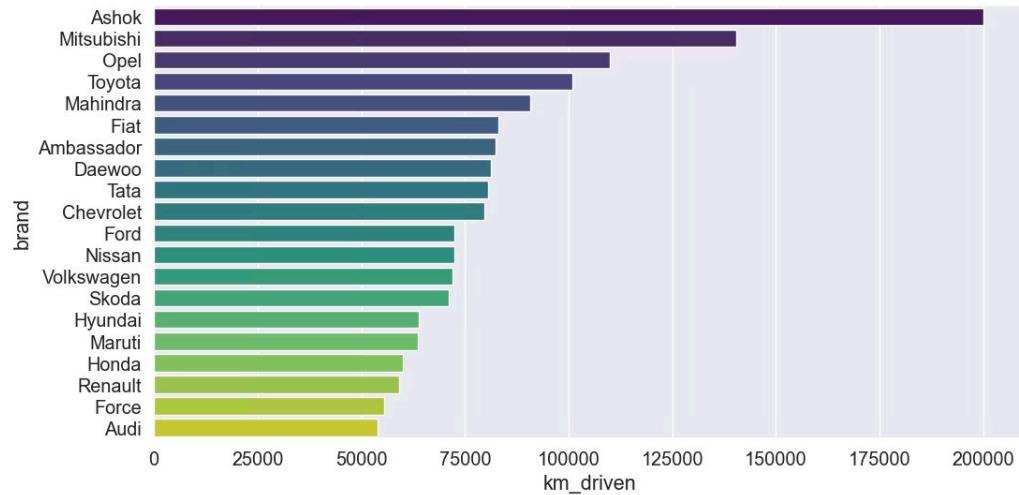
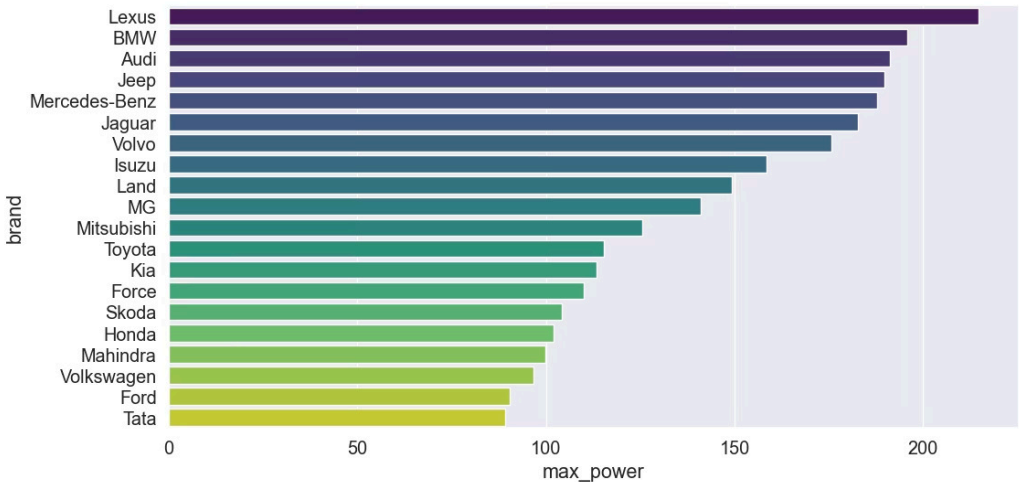
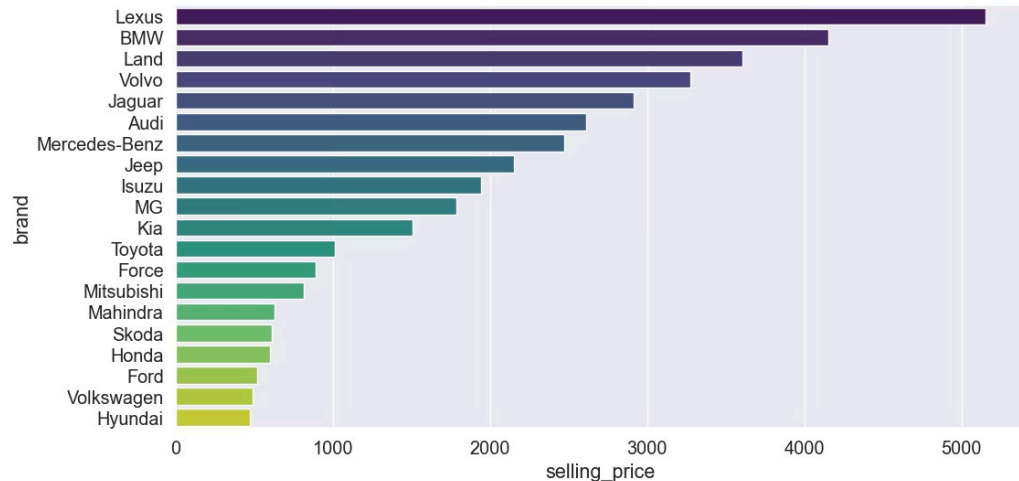
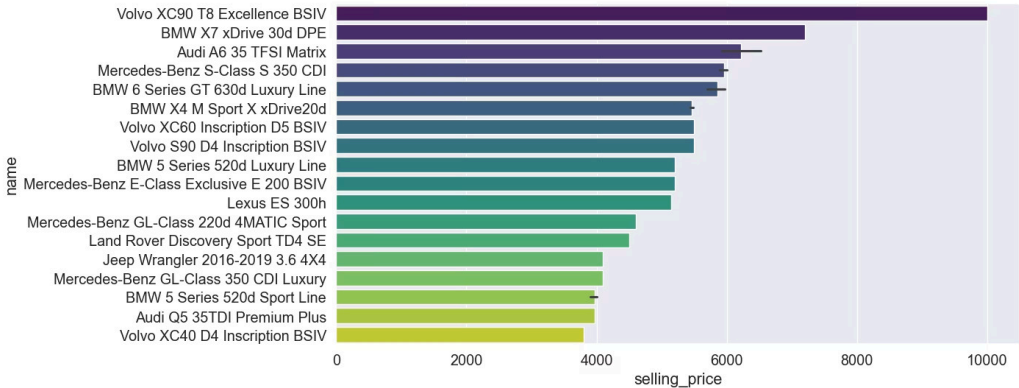
Зависимости между параметрами с разделением на тип КПП



Корреляция между параметрами



Параметры в зависимости от марки автомобиля



Car Price Correlations

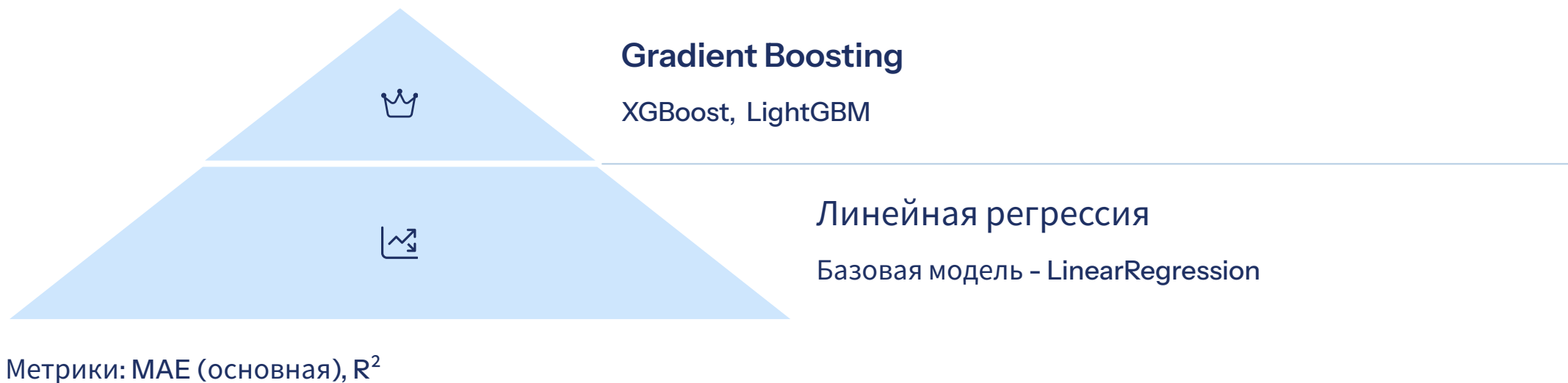
Car Price Plots Word Plots



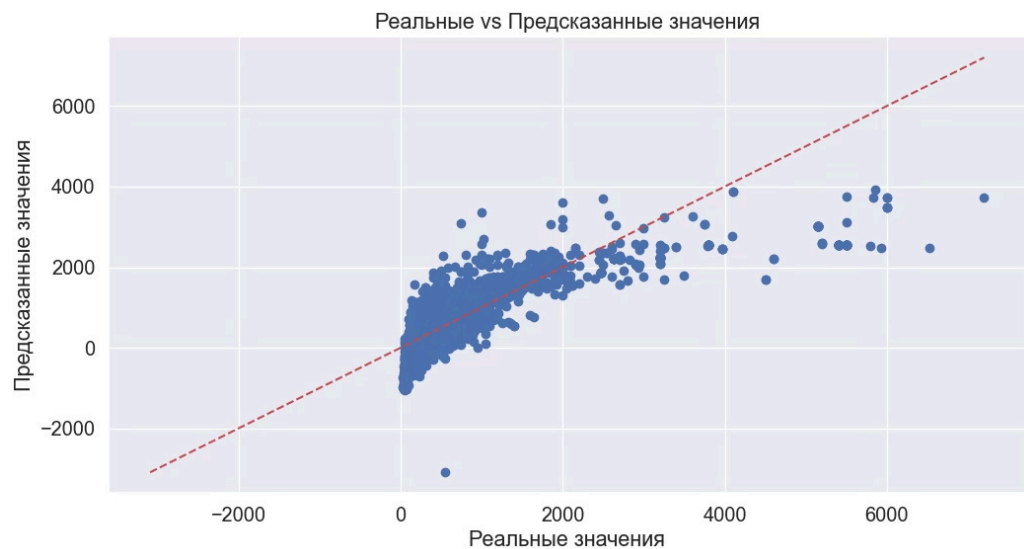
Разведочный анализ (EDA)

- Распределение цен
Средняя цена ~650, разброс от 30 до 10,000
- Год выпуска
Новее автомобиль — выше цена
- Пробег
Больше пробег — ниже стоимость
- Тип топлива
Дизельные автомобили дороже бензиновых
- Тип трансмиссии
Автомобили с АКПП стоят дороже, чем автомобили с МКПП
- Марка автомобиля
Цена меняется в зависимости от модели

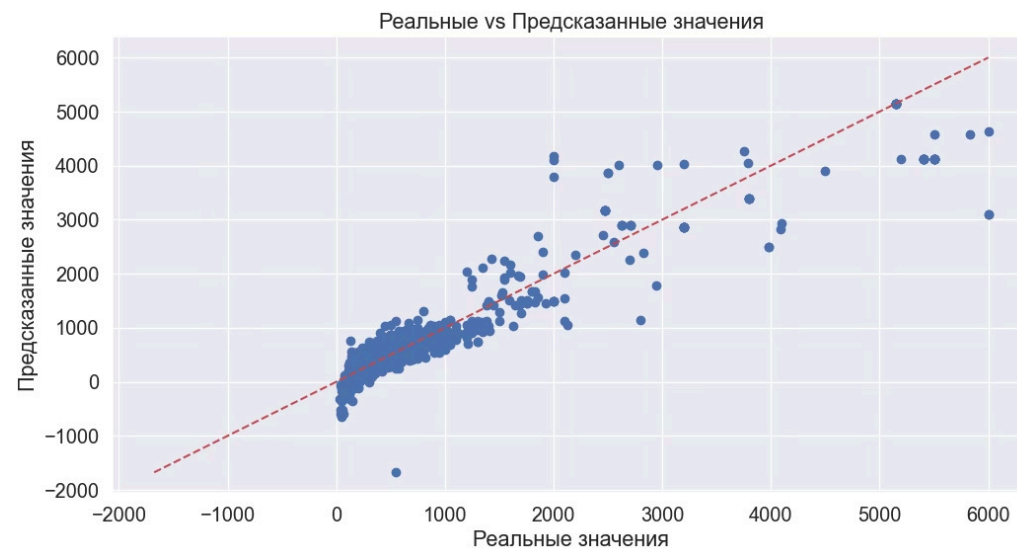
Выбор модели и метрики



Результаты линейной регрессии

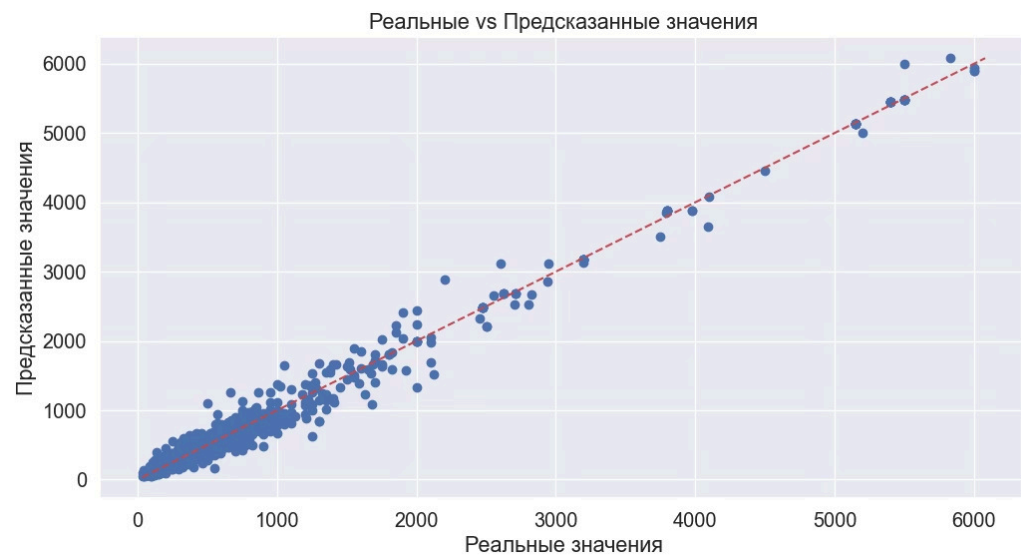


Без учета категориальных признаков

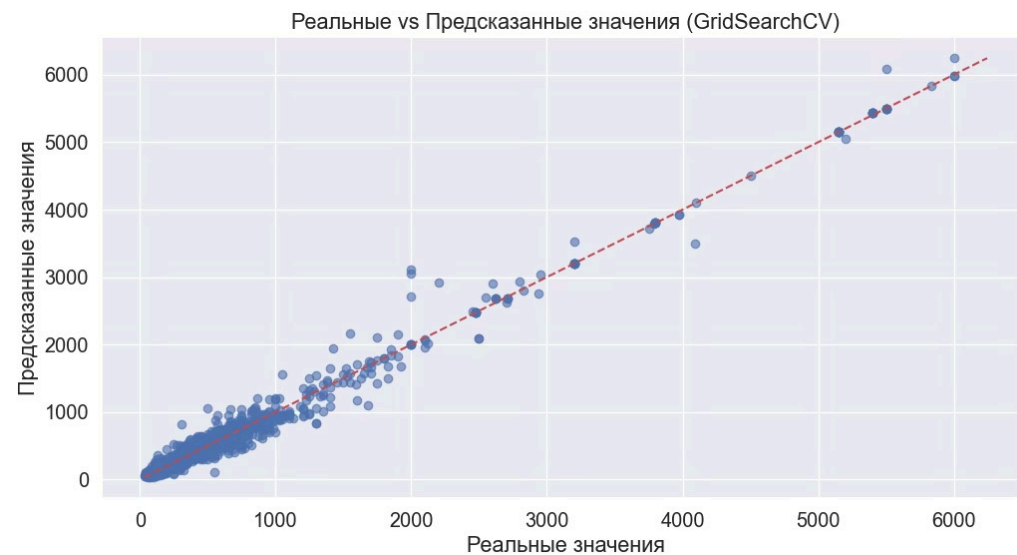


С учетом

Результаты XGBoost



Без GridSearch

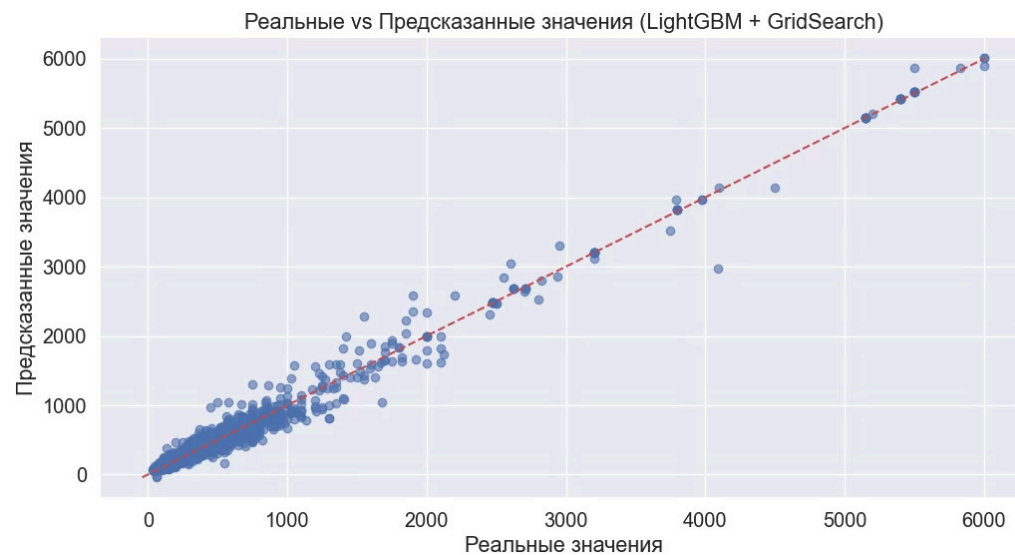


С GridSearch

Результаты LightGBM



Без GridSearch

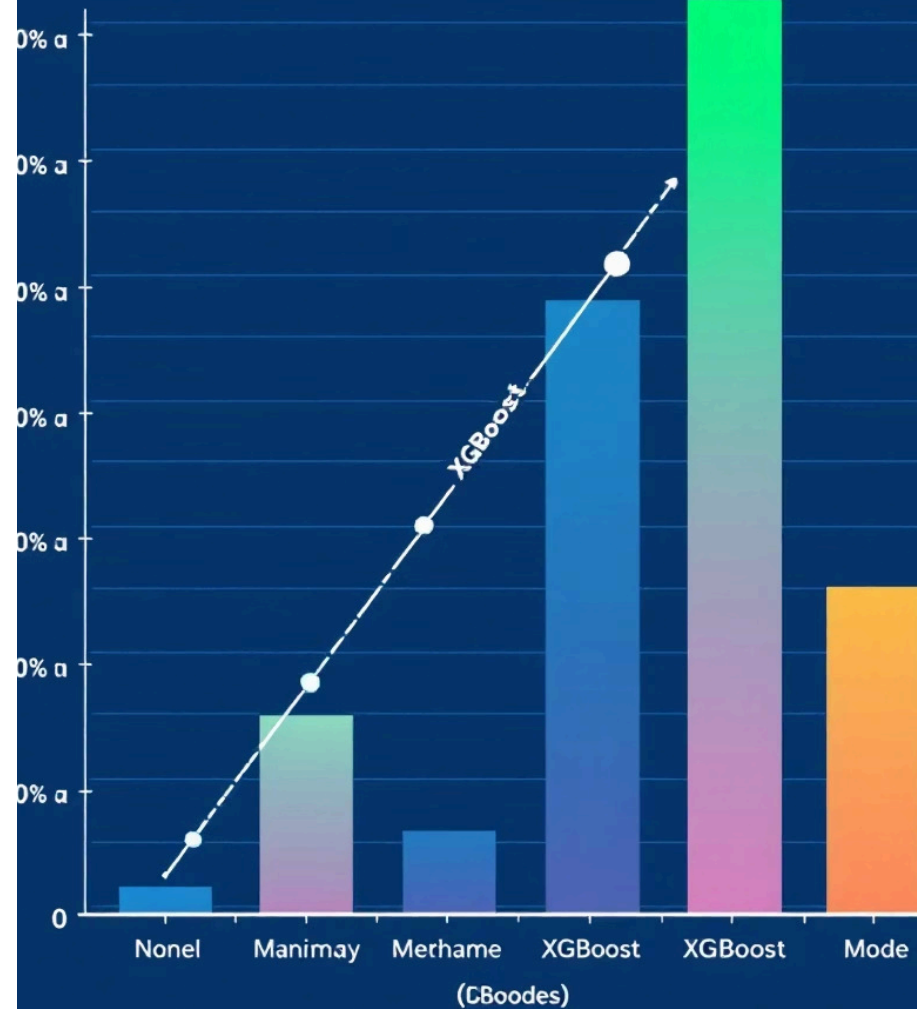


С GridSearch

Результаты

Модель	MAE (на тесте)	R ²
Линейная регрессия	~292\163	0.63\0.88
XGBoost	~70\63	0.98\0.98
LightGBM	~74\66	0.98\0.98

ХGBoost показал лучший результат



Made with **Gamma**