

Проект по МАПЗ:

Сбор и анализ данных о ноутбуках

Вяткин Артём Андреевич

19.Б07-мм

st076419@student.spbu.ru

24.03.23

Данные о ноутбуках

- Сайт [DNS](#)
- Использовались *Selenium* и *BeautifulSoup*
- Сбор по разным городам
- Всего собрано 8004 объекта
- Количество признаков — 86, представлены все типы:
 - Бинарные (*Игровой ноутбук, Имеет сенсорный экран* и т.д.)
 - Количественные (*Цена, Плотность пикселей (ppi)* и т.д.)
 - Категориальные (*Модель встроенной видеокарты, Материал корпуса* и т.д.)

Данные о ноутбуках

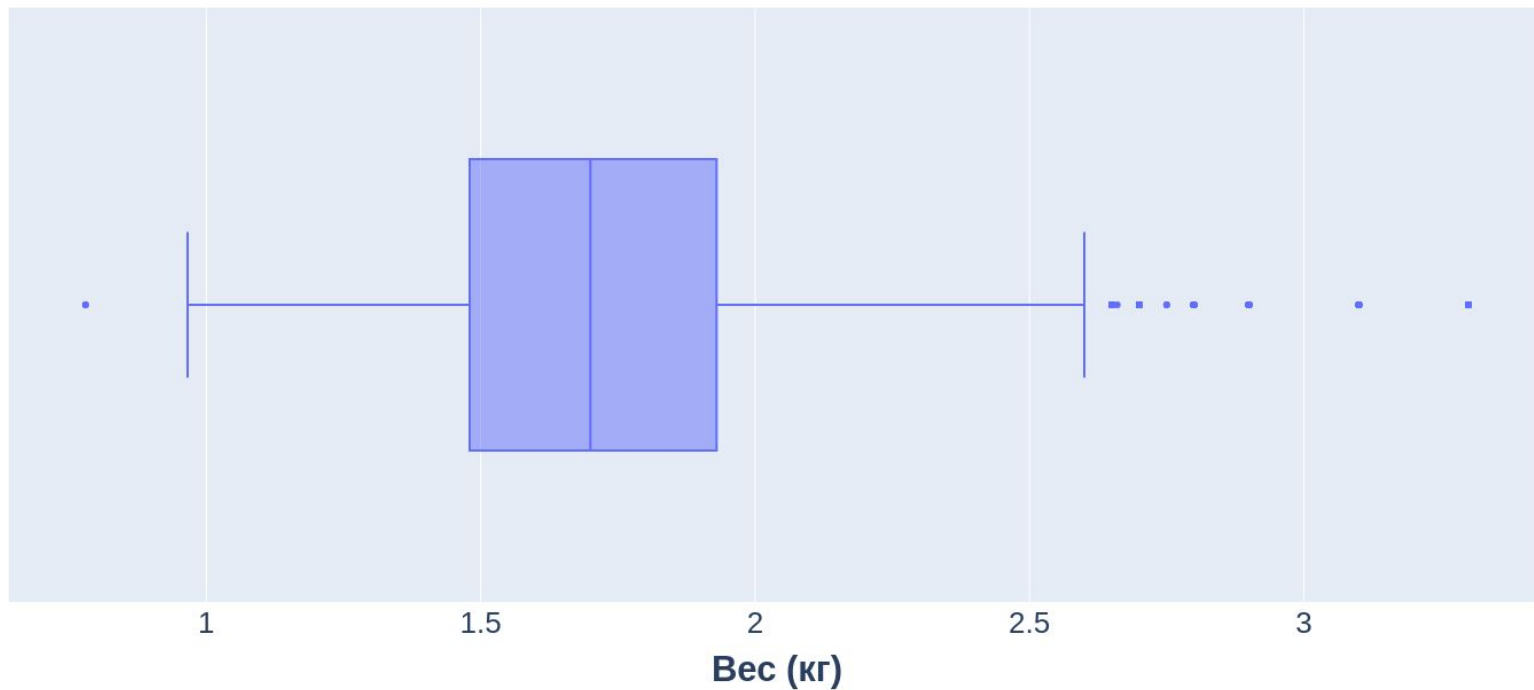
	Гарантия от производителя (лет)	Тип	Модель	Код производителя	Год релиза	Операционная система	Игровой ноутбук	Конструктивное исполнение	Цвет верхней крышки	Материал крышки	...	Цветовой охват	Гарантия от продавца (лет)	Скорость вращения шпинделя (об/мин)	Объем видеопамяти (Гб)	Производитель видеочипа	Тактовая частота с ускорением (Мгц)	Тип видеопамяти	Расширенные возможности
0	1.0	ноутбук	Irbis NB77	[NB77]	2021.0	Windows 10 Home Single Language	False	классический	черный	пластик	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1.0	ноутбук	Irbis NB257	[NB257]	2021.0	Windows 10 Home Single Language	False	классический	серый	пластик	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	1.0	ноутбук	Irbis NB283	[NB283]	2021.0	Windows 10 Pro	False	классический	серый	пластик	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	1.0	ноутбук	DEXP Aquilon	[C14-ICW201]	2022.0	Windows 11 Pro	False	классический	серебристый	пластик	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	1.0	ноутбук	DEXP Aquilon	[C14-ICW202]	2022.0	Windows 11 Home Single Language	False	классический	серый	пластик	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
7999	1.0	ноутбук	Apple MacBook Pro	[MNWE3LL/A]	2023.0	macOS	False	классический	серебристый	алюминий	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8000	1.0	ноутбук	Apple MacBook Pro	[MNWA3LL/A]	2023.0	macOS	False	классический	серый	алюминий	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8001	1.0	ноутбук	MSI Raider GE78 HX 13V1-095RU	[9S7-17S111-095]	2023.0	Windows 11 Home Single Language	True	классический	черный	пластик	...	NaN	NaN	NaN	16.0	NVIDIA	2040.0	GDDR6	технологии Resizable BAR, технологии DLSS, ноу.
8002	2.0	ноутбук	MSI CreatorPro X17 A12UMS-205RU	[9S7-17Q121-205]	2022.0	Windows 11 Pro	False	классический	черный	алюминий	...	NaN	NaN	NaN	16.0	NVIDIA	1640.0	GDDR6	NaN
8003	2.0	ноутбук	MSI Titan GT77 HX 13V1-096RU	[9S7-17Q211-096]	2023.0	Windows 11 Home Single Language	True	классический	черный	пластик	...	NaN	NaN	NaN	16.0	NVIDIA	2040.0	GDDR6	технологии Resizable BAR, технологии DLSS, ноу.

8004 rows × 86 columns

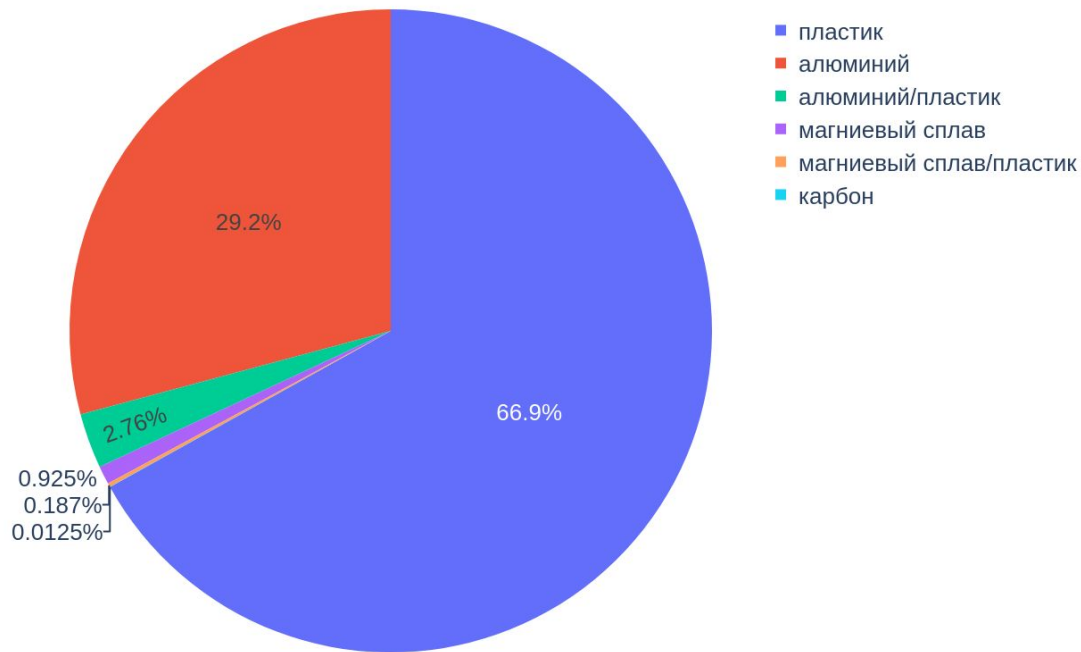
Подсчет первичных описательных статистик

- Медиана
 - Цена — 57999 руб.
 - Общее количество ядер — 4
 - ...
- Среднее
 - Цена — 75641 руб.
 - Вес — 1.78 кг
 - ...
- Мода
 - Материал корпуса — пластик
 - Модель процессора — Intel Core i3-1115G4
 - ...

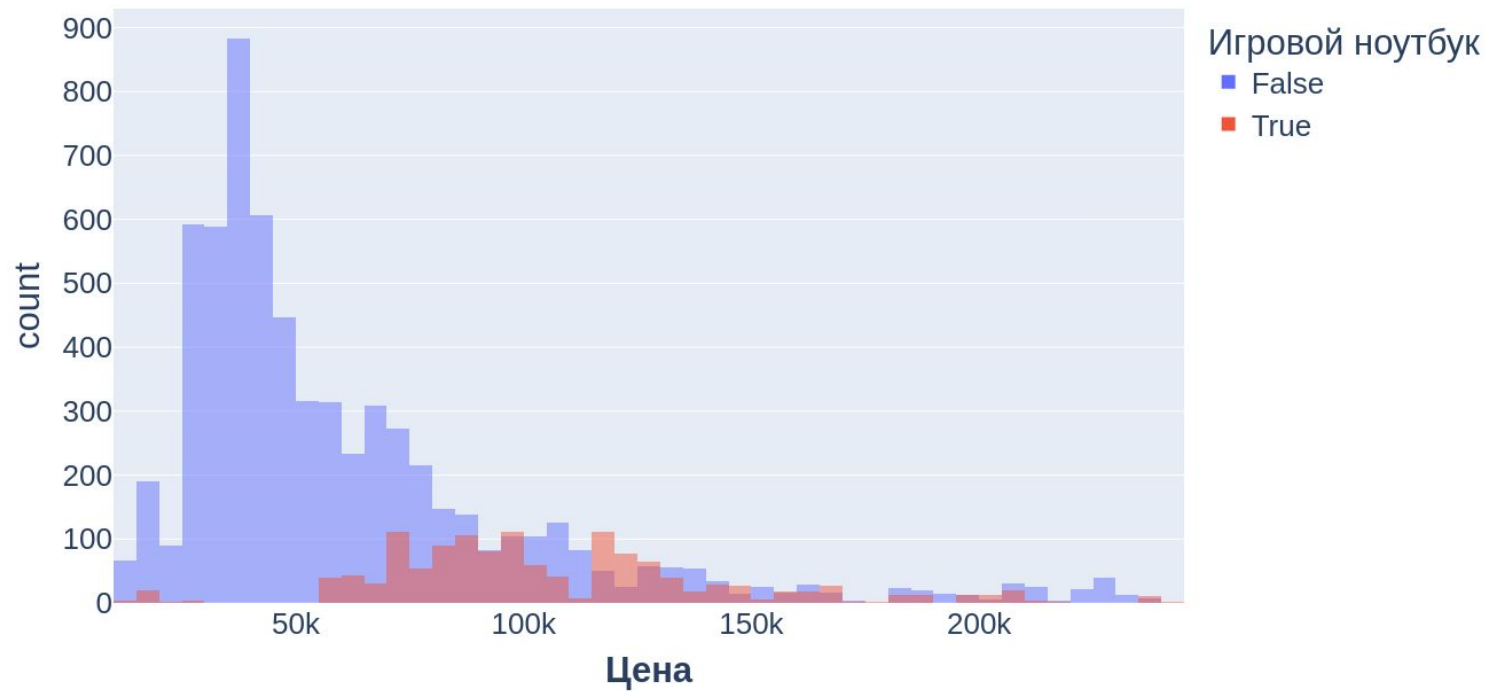
Распределение веса ноутбуков



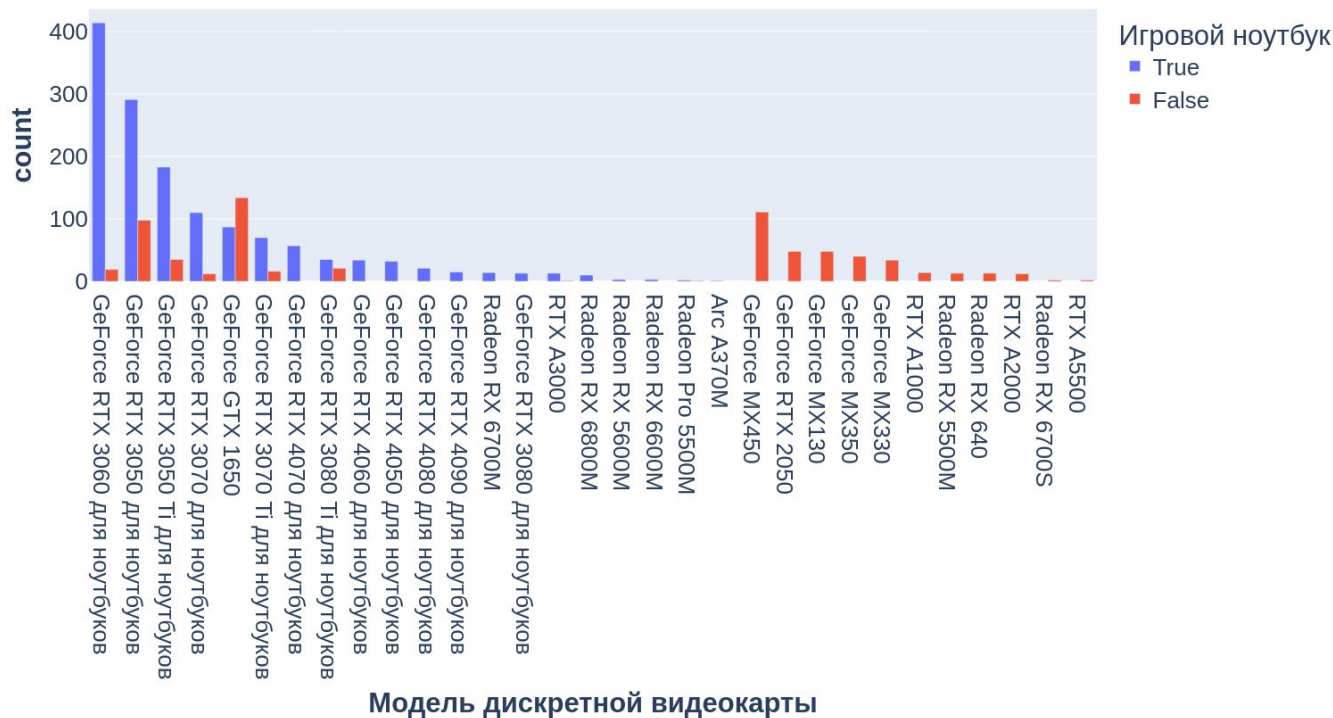
Распределение материалов корпусов



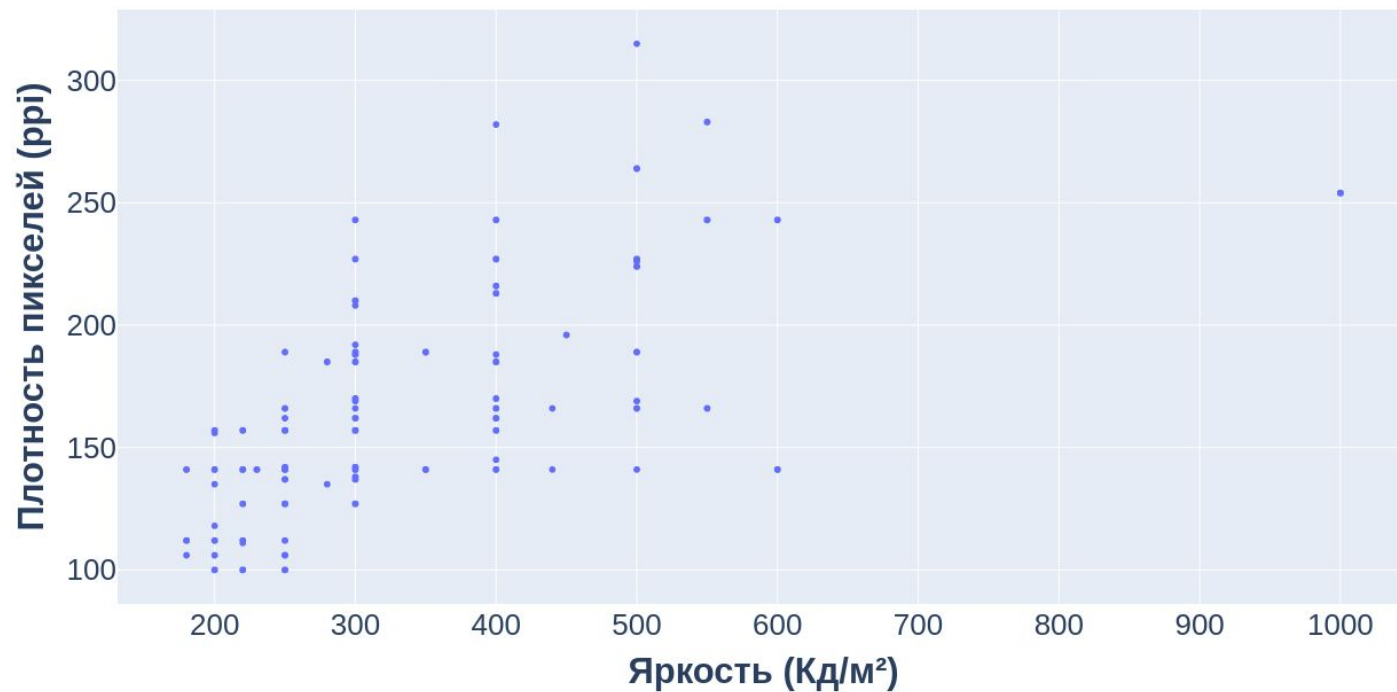
Стоимость игровых и неигровых ноутбуков



Дискретные видеокарты у игровых и неигровых ноутбуков

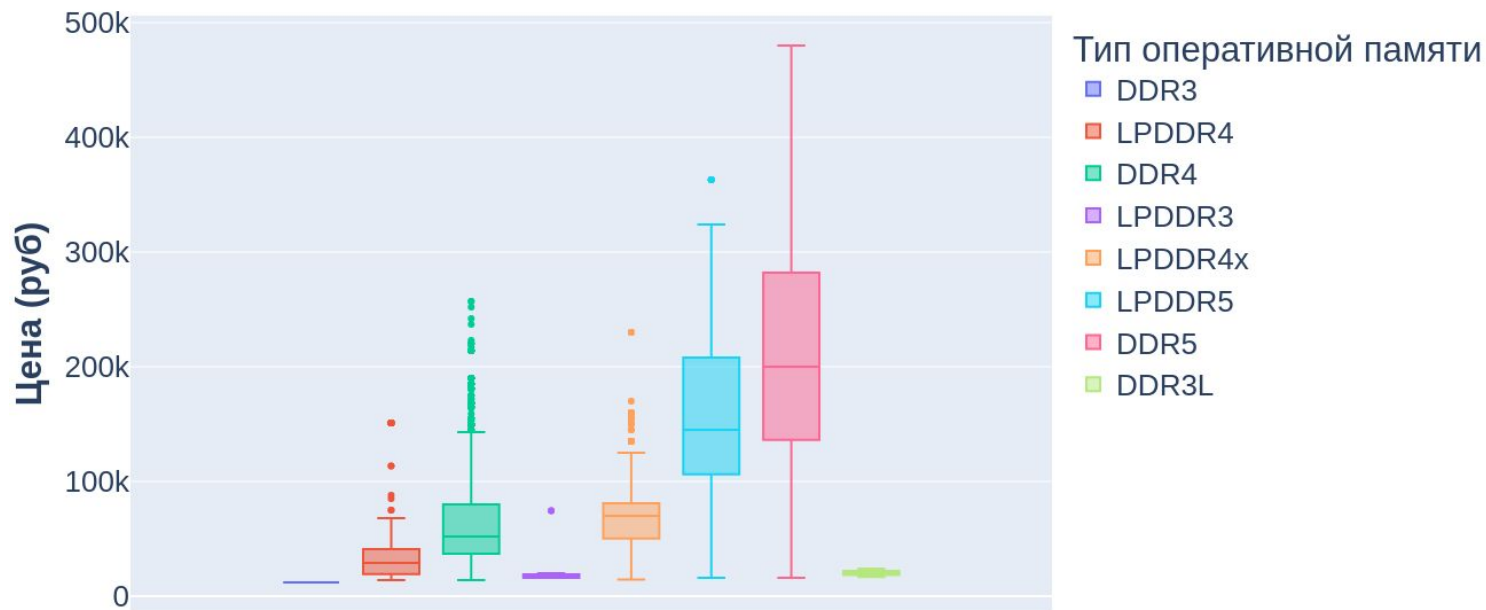


Яркость и плотность пикселей



Цена и тип оперативной памяти

Зависимость цены от типа оперативной памяти



Задача регрессии

- Целевая переменная — *Цена*
 - Соответственно задача — регрессия
- Признаки для предсказания *Цены*:
 - *Год релиза*
 - *Диагональ экрана (дюйм)*
 - *Максимальная частота обновления экрана (Гц)*
 - *Частота процессора (ГГц)*
 - *Объем оперативной памяти (Гб)*
 - *Общий объем SSD (Гб)*
 - *Максимальный объем памяти (Гб)*
 - *Автоматическое увеличение частоты (ГГц)*
 - *Максимальное число потоков*
- Дубликаты удалены

Обучение трех моделей

- Применена кросс-валидация (4 блока)

	GradientBoostingRegressor				LinearRegression				RandomForestRegressor			
R ²	0.37	0.14	0.45	0.45	0.25	0.05	0.48	0.55	0.4	0.3	0.38	0.35
RMSE	8794	38610	26078	47040	12358	37113	25374	48640	13081	41248	27617	45531
MAE	6707	26912	17556	37058	9401	26229	16960	38674	9680	30144	18684	35141

- Лучшая — **GradientBoostingRegressor**

Поиск гиперпараметров

- Модель — GradientBoostingRegressor
- Использован GridSearchCV
- Лучшие значения параметров:
 - *max_depth* — 3
 - *min_samples_leaf* — 15
 - *min_samples_split* — 2
 - *n_estimators* — 100
- Сравнение

	Best	Standart
R ²	0.42	0.4
RMSE	27389	27845
MAE	18217	18395



Лучше?

Проект по МАПЗ:
Сбор и анализ данных о
ноутбуках

Вяткин Артём Андреевич

19.Б07-мм

st076419@student.spbu.ru

24.03.23