

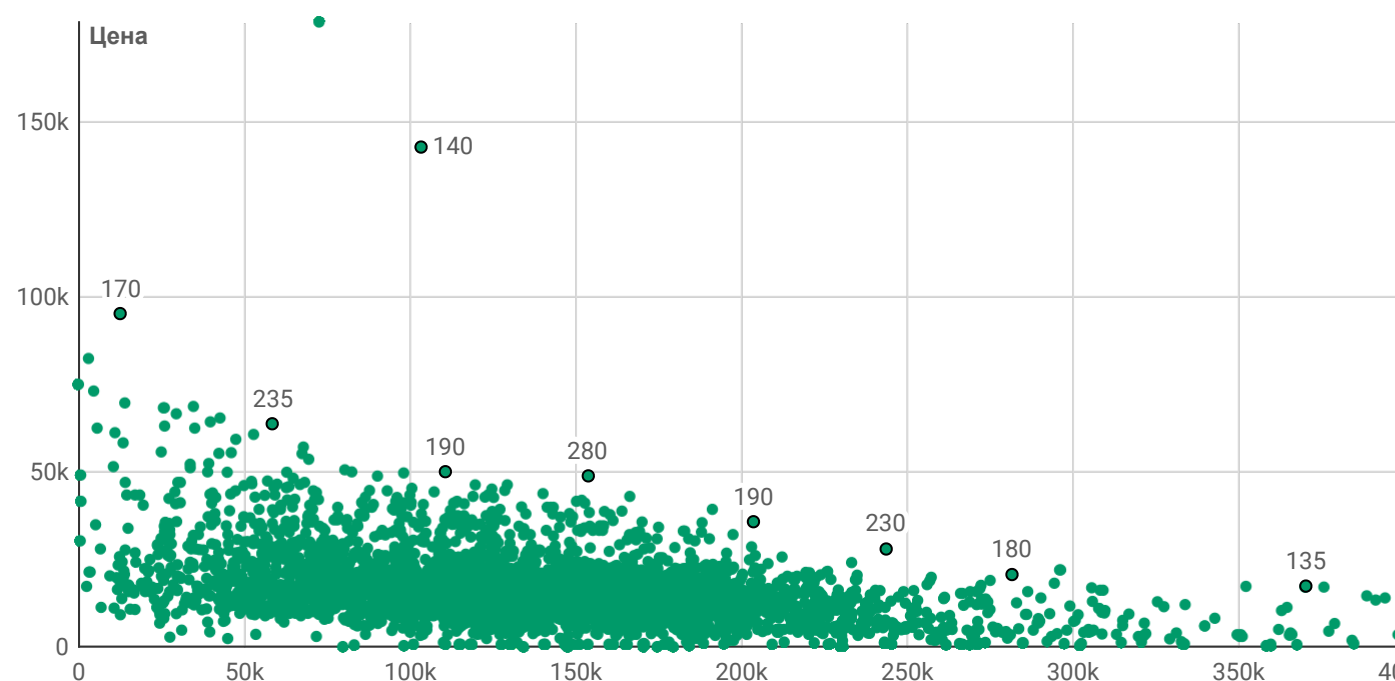
Дата сторителлинг по датасету с аукциона продажи автомобилей BMW

Источник данных: [BMW pricing challenge](#)

Предоставленные данные состоят из почти 5000 реальных автомобилей BMW, которые были проданы на аукционе b2b в 2018 году. Цена, указанная в таблице, является самой высокой ставкой, которая была достигнута в ходе аукциона. Оценка стоимости подержанного автомобиля является одной из основных повседневных задач в автомобильном бизнесе, поэтому визуализация этих данных приходится для этой задачи как нельзя кстати. Она наглядно позволит оценить стоимость автомобилей в данном датасете, а сторителлинг упростит этот процесс.

Конечно, перед оценкой стоимости автомобилей, проданных на аукционе хочется подумать над вопросом: а от каких же факторов может зависеть стоимость автомобиля? Очевидно, что этими факторами будут являться пробег, мощность и год выпуска автомобиля. Кроме того, в таблице предоставляются данные по дополнительным опциям к автомобилям, зависимость которых на стоимость мы также попробуем проследить. Итак, рассмотрим график зависимости цены автомобиля от пройденного пробега.

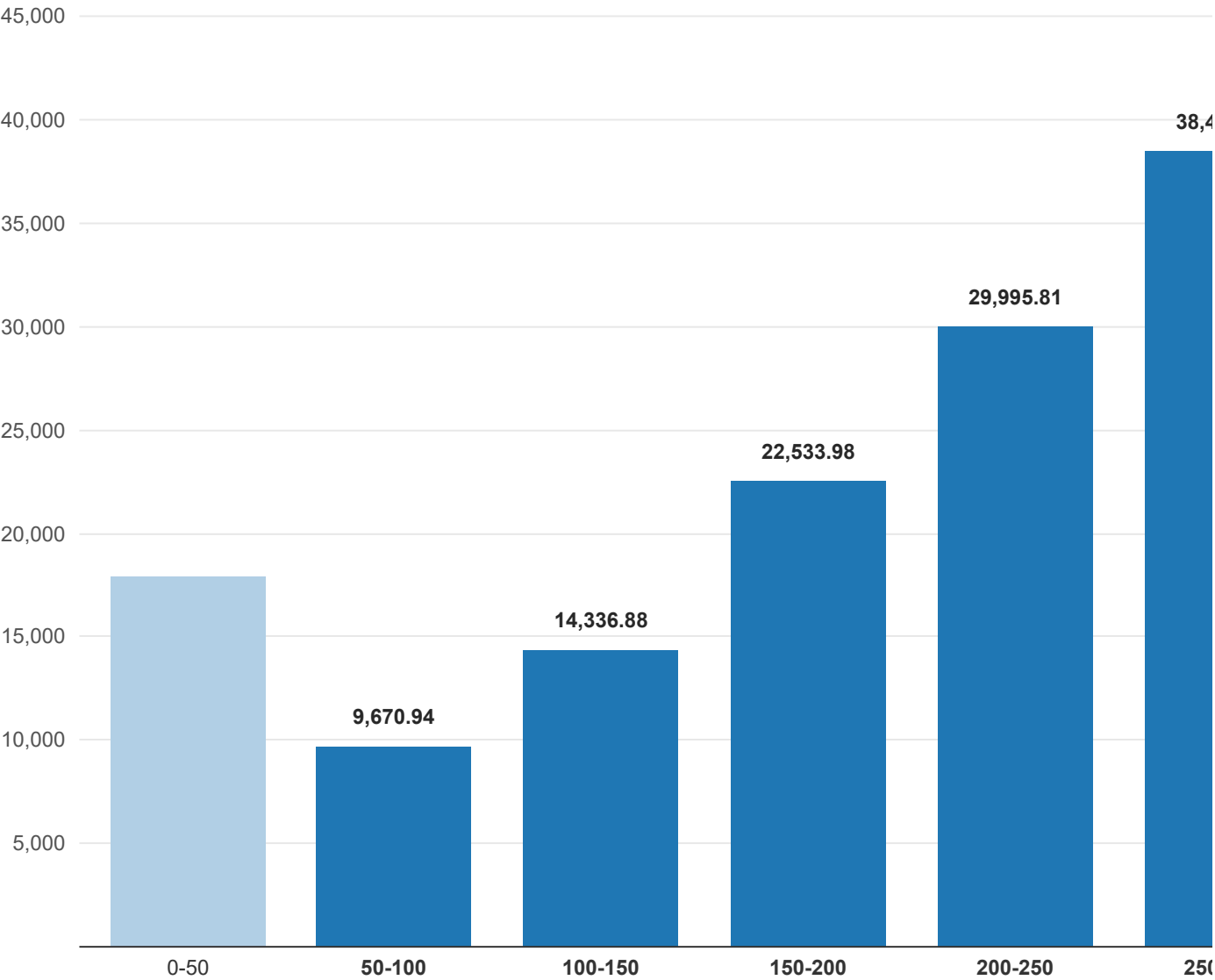
Точечный график зависимости цены автомобиля от его пробега



Из данного графика по скоплениям точек мы видим, что цена имеет тенденцию к снижению с увеличением пробега автомобиля, что является весьма логичным фактом. Значит, соответственно, чем меньше пробег автомобиля, тем дороже его стоимость. Кроме того, на данном графике, как можно заметить, верхнюю сторону "треугольника концентрации точек" занимают автомобили с довольно высокими показателями мощности.

Рассмотрим второй фактор, от которого, как мы предполагаем, должна зависеть цена автомобиля. Обратимся к гистограмме зависимости цены от мощности автомобиля, ведь, как говорится, за лошадиные силы надо платить.

Гистограмма зависимости средней цены от мощности

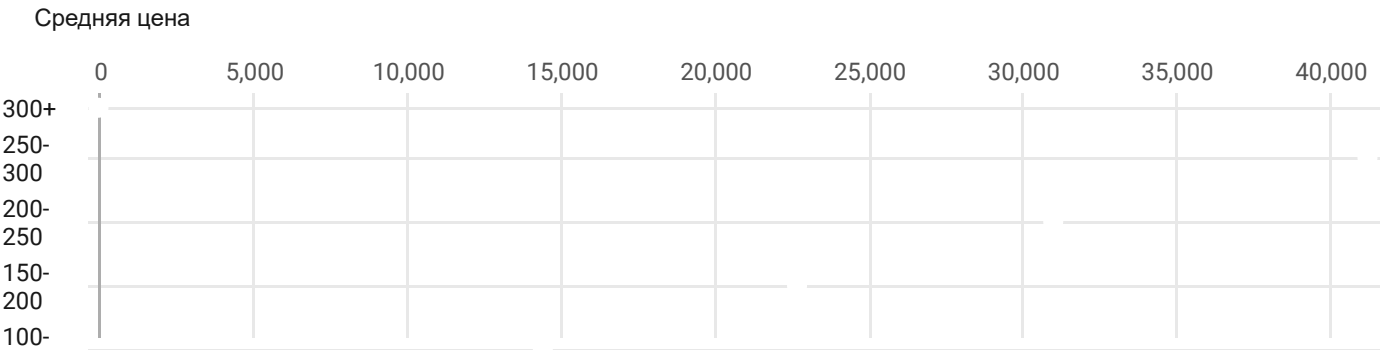


Source: [BMW pricing challenge](#)

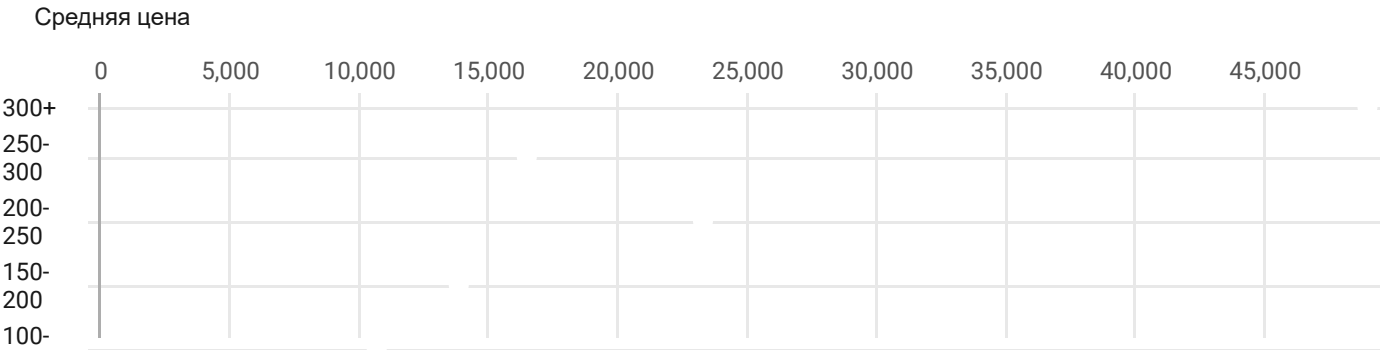
На данной гистограмме мощность была разделена на некоторые диапазоны, для каждого из которых была посчитана средняя стоимость автомобилей из данного диапазона мощности. Как мы видим, общий принцип зависимости стоимости от мощности - с увеличением лошадиных сил увеличивается и цена. Обратим внимание на столбец со значением мощности от 0 до 50 лошадиных сил, потому что, как видно из гистограммы, данный столбец выше столбца со средней стоимостью более мощных автомобилей.

Для более детального исследования зависимости цены от мощности, обратимся также к анализу типов топлива, которые используются автомобилями.

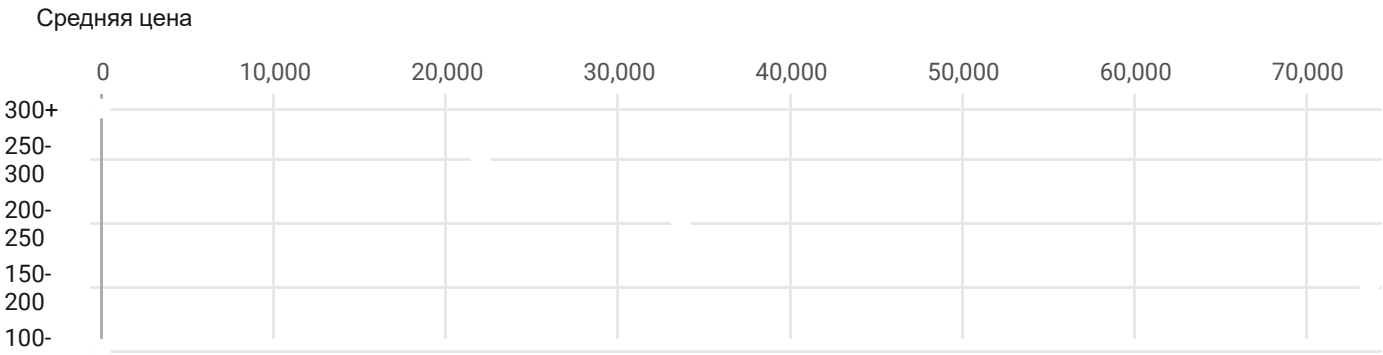
Зависимость средней цены от мощности для дизельных автомобилей



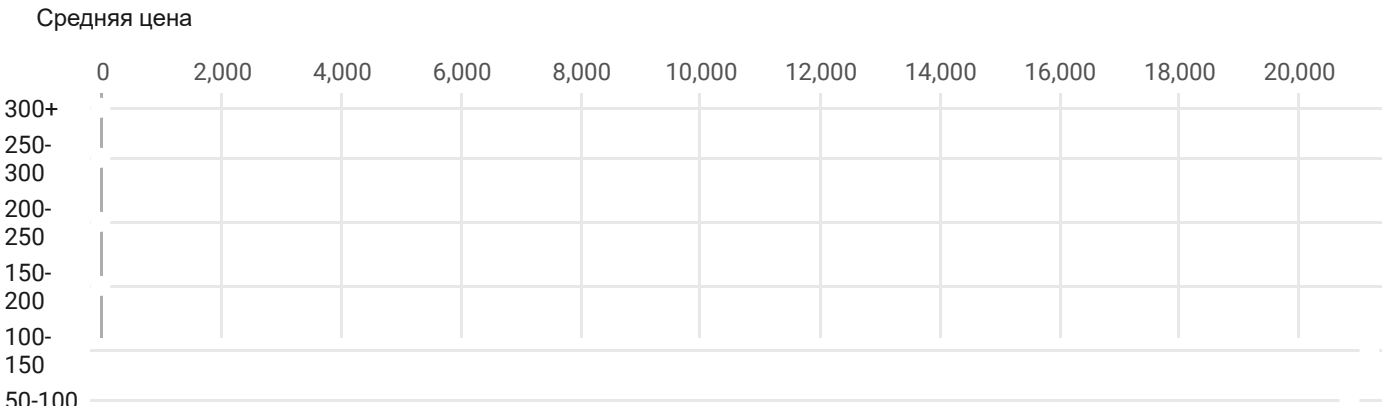
Зависимость средней цены от мощности для бензиновых автомобилей



Зависимость средней цены от мощности для гибридных автомобилей



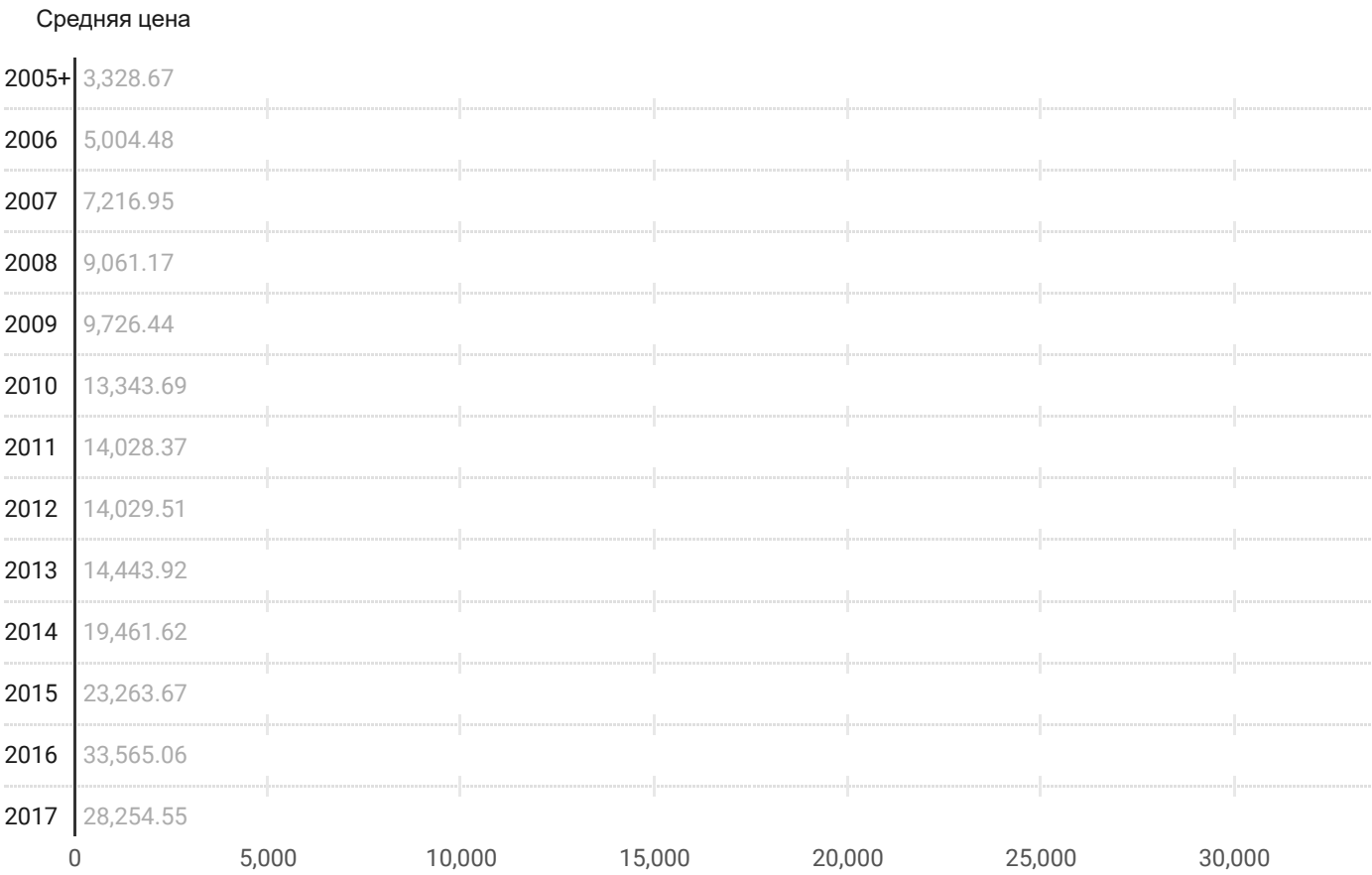
Зависимость средней цены от мощности для электромобилей



Данные точечные графики являются аналогом гистограммы, но для обозначения средней цены для определенного диапазона значений мощности используется не высота столбца, а отдаленность точки от нулевого значения. Итак, нас интересует диапазон значений мощности от 0 до 50, который не имеет нулевого значения средней цены автомобилей из него. Из всех четырех точечных графиков можно найти лишь одну такую не нулевую точку - на графике зависимости средней цены от мощности для гибридных автомобилей. Из это более детального исследования следует, что не смотря на свою небольшую мощность двигателя, автомобили с гибридным типом двигателя стоят дороже остальных (видимо, из-за более сложного устройства двигателя таких автомобилей). Кроме того, дополнительно из данных точечных графиков по типам топлива можно отметить, что самыми дорогими автомобилями, проданными на аукционе, были гибридные автомобили, мощность которых находилась в диапазоне от 150 до 200 лошадиных сил.

Проверим следующий фактор, от которого зависит цена на автомобиль - год выпуска машины. Для этого исследуем следующую гистограмму.

Зависимость цены от года выпуска автомобиля



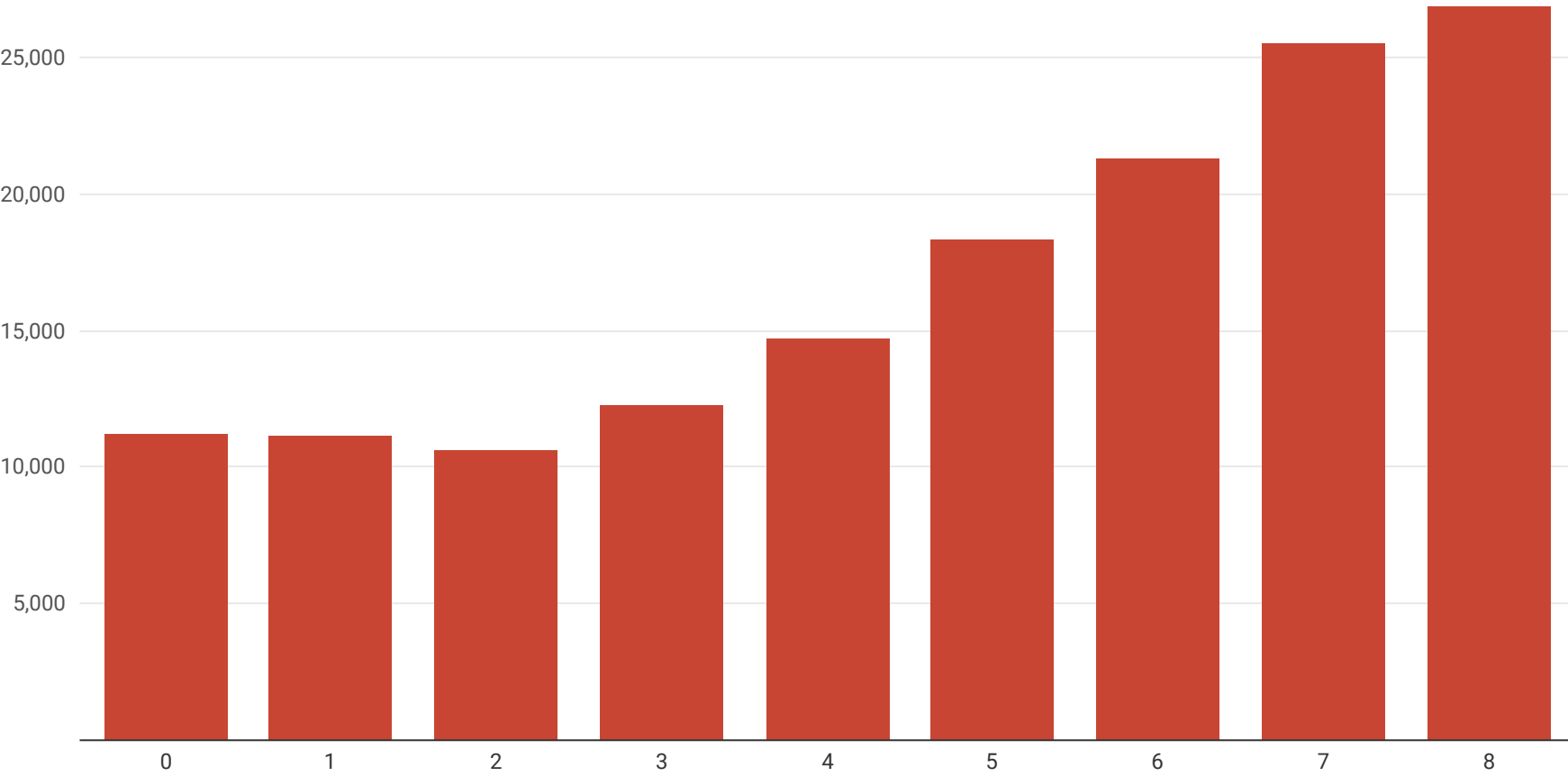
Source: [BMW pricing challenge](#) • [Get the data](#) • Created with [Datawrapper](#)

Отсюда мы видим, что чем старше автомобиль, тем меньше его стоимость. Конечно, эта зависимость прослеживается не для всех годов выпуска(например, средняя цена на автомобили 2016 года выше, чем на машины 2017 года). Мы можем лишь предположить, что автомобили 2016 года выпуска, которые были

проданы на аукционе, обладали другими факторами, которые могли положительно повлиять на среднюю цену для данного года выпуска.

Итак, последний фактор, который был взят для рассмотрения зависимости средней цены автомобиля - количество установленных дополнительных опций. Рассмотрим последнюю гистограмму нашего исследования.

Зависимость цены от количества дополнительных опций, которыми оснащены автомобили



Source: [BMW pricing challenge](#) • [Get the data](#) • Created with [Datawrapper](#)

В данном случае получаются снова логические выводы: общая тенденция - чем больше опций, тем выше цена проданного автомобиля. Конечно, есть столбец, который говорит, что с двумя опциями машины в среднем дешевле, чем с одной, но разница между данными столбцами минимальна, а также это можно списать на то, что неизвестны сами опции, на которые могут отличаться автомобили, ведь логично предположить, что цена на разные опции может отличаться.

Подводя итоги данного исследования, можно с уверенностью сказать, что цена проданных на аукционе автомобилей увеличивалась при: уменьшении пробега, увеличении мощности, увеличении года выпуска, увеличении количества установленных опций. Кроме того, мог влиять на стоимость и тип топлива автомобиля.

>