

Отчёт по исследованию автомобильного рынка Казахстана для автоцентра холдинга «Меркур Авто»

Введение

В рамках проекта был проведён анализ рынка новых автомобилей Казахстана за 2019 год. Основной задачей было оценить текущую позицию автоцентра «Меркур Авто» на рынке и выявить потенциальные точки роста.

Для этого был проанализирован датасет, содержащий более 39 тысяч записей о продажах. Был выполнен разведочный анализ данных (EDA) с акцентом на структуру рынка, ценовые сегменты, географию продаж и характеристики автомобилей.

Актуальность

Автомобильный рынок Казахстана демонстрирует устойчивый спрос. В условиях высокой конкуренции и изменяющихся потребительских предпочтений автодилерам критически важно понимать свою долю рынка, эффективность портфеля брендов и географическое присутствие. Анализ позволяет принимать обоснованные стратегические решения — от расширения модельного ряда до выхода в новые регионы.

Цель работы

1. Подготовить качественный, пригодный для последующего анализа датасет.
2. Провести разведочный анализ рынка Казахстана.
3. Оценить основные показатели автоцентра «Меркур Авто» на авторынке.
4. Предложить рекомендации по развитию бизнеса.

Описание процесса очистки

Подробный процесс очистки с промежуточными выводами можно посмотреть в файле: https://github.com/darthKwisatzHaderach/mephi-data-preparation-and-analysis/blob/main/Kazakhstan_AutoMarket_Research.ipynb

Исходные данные содержали множество ошибок: полные дубликаты (более 28 тыс. строк), пропуски в полях (тип топлива, объём двигателя, тип привода), дублирование категорий (например, «бензин» и «Бензин»), а также технические артефакты (неразрывные пробелы, некорректные значения).

Удаление полностью дублирующихся записей

Сначала были выявлены **28 775 полностью совпадающих строк** (около 72% от исходного объёма).

Были проверены сделки, где кол-во автомобилей больше одного. Было найдено 62 такие сделки в каждом месяце с порядковыми номерами в датасете идущими один за другим. Такие дубли — результат технической ошибки при выгрузке данных. Реальные повторные продажи нескольких автомобилей в один месяц одним дилером маловероятны, особенно в таком масштабе.

Т.к. дубликаты были найдены в каждом месяце было принято решение удалить все полные дубликаты с помощью `drop_duplicates(keep='first')`.

В результате объём данных сократился до 16 495 строк, что позволило провести корректный анализ без искажений.

Унификация названия компаний, стран, типов топлива, привода и трансмиссии

Для обеспечения согласованности данных и корректного агрегирования была выполнена унификация категориальных признаков.

- Были исправлены названия компаний. Учитывались опечатки, пробелы и дублирующие записи (например, «Mercur Autos» и «Mercur Auto», «Astana Motors» и «Astana Motors»).
- Значения из колонки страны-производителя были закодированы в международные трёхбуквенные коды ISO ALPHA-3 (например, «Российская Федерация» в «RUS»).
- Выделены 4 типа топлива: бензин (F), дизель (D), гибрид (HYB) и электро (E), с объединением вариантов написания («бензин»/«Бензин»).
- Тип привода сведён к трём категориям — передний (FWD), задний (RWD) и полный (4WD), с учётом множества синонимов («quattro», «AWD», «4x4» в «4WD»).
- Значения поля "Трансмиссия" сведены в два основных типов — «Автомат» и «Механика». Туда были сгруппированы все разновидности (AT, CVT, DSG, Tiptronic и др. в «Автомат»; MT, МКПП, 5MT и т.п. в «Механика»). Ошибочные и неинтерпретируемые значения заменялись на NaN.

Приведение числовых полей к корректному формату

Были обработаны столбцы `quantity`, `price_usd` и `total_sale_usd`, которые изначально имели тип `float64`, но могли содержать артефакты парсинга или неявные текстовые значения.

Для надёжности **все три поля были явно преобразованы** с помощью `pd.to_numeric(..., errors='coerce')`, что позволило заменить некорректные значения (например, строки вместо чисел) на NaN.

После этого был проведён анализ аномалий: **найден единственный возврат** (`quantity = -1`), удалён, т.к. в контексте других данных это незначительная доля.

Проверены экстремальные значения (например, сделки на 115 автомобилей или

цены свыше \$300 тыс.). Это реальные сделки - крупные заказы или продажи автобусов.

Также была установлена обратная зависимость между количеством автомобилей в сделке и ценой за единицу — **при увеличении объёма цена снижалась, что соответствует практике оптовых скидок и подтверждает корректность данных.**

Восстановление пропущенных значений

Пропущенные значения в технических полях — «объём двигателя» (`engine_volume`) и «тип топлива» (`fuel_type`) — восстанавливались в два этапа. На первом этапе применялась группировка по комбинации `brand + model`: внутри каждой группы пропуски заполнялись медианным значением объёма двигателя и модой типа топлива, если такие значения существовали. Это позволило корректно восстановить характеристики для большинства массовых моделей, где **хотя бы одна запись содержала полные данные.**

Однако, для ряда популярных моделей (например, Renault Logan, Nissan Kaptur, Mazda CX-5) в датасете отсутствовали **все** записи с заполненными техническими характеристиками. Это похоже на какую-то техническую ошибку сбора данных.

В таких случаях использовалась внешняя справочная информация. На основе общедоступных данных о комплектациях автомобилей, актуальных для рынка СНГ в 2019 году, задавались значения объёма двигателя и типа топлива. Например, для Logan и Sandero был указан объём 1.6 л и бензин, для Kaptur — 2.0 л бензин, для электромобиля Jaguar I-Pace — тип топлива «электро» и объём двигателя 0.0. **Этот подход обеспечил высокое качество восстановления даже в условиях неполных исходных данных.**

Создание временного признака «дата продажи»

Также был создан единый признак «дата продажи» (`sale_date`). Исходные данные содержали год и месяц продажи в отдельных столбцах, причём месяц указывался текстом на русском языке (например, «Июль», «Сентябрь»).

Сначала названия месяцев были заменены на соответствующие числовые значения с помощью словаря (`month_mapping`), после чего для каждой строки **был определён последний день месяца** с использованием функции `calendar.monthrange()`, **так как точная дата сделки не указывалась.** На основе трёх компонент — года, месяца и последнего дня месяца — сформирован полноценный объект типа `datetime64[ns]`.

Это позволило проводить агрегацию по времени, строить временные ряды и сравнивать сезонные паттерны продаж. В завершение исходные столбцы `year`, `month` и вспомогательный `day` **были удалены из датасета.**

Результат очистки данных

В результате очистки исходного датасета (39 966 записей) был получен качественный и пригодный для последующего анализа набор данных, содержащий **16 258 уникальных записей** о продажах новых автомобилей в Казахстане за 2019 год.

Очищенный датасет включает **18 колонок**: от **базовых характеристик автомобиля** (бренд, модель, год выпуска, объём двигателя, тип топлива, привод, трансмиссия) до **бизнес-показателей** (цена, количество, общая сумма сделки), а также **географических** (регион, область) и **временных данных** (дата продажи в формате `datetime`).

Все категориальные признаки унифицированы и приведены к типу `category` , числовые поля обработаны, пропуски в технических характеристиках восстановлены на основе группировки по бренду/модели и справочной информации, а дубликаты и аномалии удалены.

Датасет охватывает только автомобили 2017–2019 годов выпуска и **полностью готов для проведения разведочного анализа**, сегментации рынка и оценки эффективности автоцентра «Меркур Авто».

Анализ рынка

Наблюдается повышенный спрос на внедорожники и кроссоверы: на этот сегмент пришлось **52% общей выручки** (около \$340 млн из \$657 млн), несмотря на то, что по количеству проданных единиц он уступает легковым автомобилям. Легковые модели составили **29% выручки**, но при этом охватили значительную долю массового сегмента — почти 12,7 тыс. автомобилей. Средняя цена одного автомобиля составила **\$27 814**, что отражает доминирование бюджетных и средних ценовых категорий.

Подавляющее большинство покупателей выбирали **бензиновые двигатели** — их доля достигла **88%**, тогда как дизельные выбирали лишь 10% клиентов, а гибриды и электромобили находились на уровне менее 2%.

Что касается трансмиссии, то несмотря на растущую популярность автоматических коробок передач в мире, в Казахстане в 2019 году **механика** всё ещё оставалась более распространённой — **59%** всех проданных автомобилей были оснащены «механикой». Это говорит о высокой чувствительности рынка к цене и стоимости владения.

Позиции автоцентра «Меркур Авто»

Занимает относительно небольшую нишу: его доля составляет всего **1,28% по количеству проданных автомобилей** (376 единиц) и **2,24% по выручке** (\$14,7 млн). Но есть важная особенность — компания демонстрирует **чёткое премиальное позиционирование**. Средний чек «Меркур Авто» составляет **\$39 273**, что на **41% выше рыночного среднего**. Это свидетельствует о фокусе на дорогих моделях и состоятельных клиентах, преимущественно в крупных городах.

Компания занимается автомобилями **Volkswagen Group**: основной вклад в объём продаж вносит **Volkswagen** (74% сделок), оставшиеся 26% составляют **Porsche** и **Audi**.

При этом Porsche и Audi генерируют значительную часть выручки из-за высокой стоимости, особенно в сегменте премиальных внедорожников и спортивных автомобилей.

Деятельность «Меркур Авто» в основном представлена в городах **Алматы** и **НурСултан**. Притом **62% продаж** приходится на **Алматы**, **14%** — на **Нур-Султан**. Остальные 24% распределены между Атырау, Карагандой, Уральском и Костаное. Это указывает на **ограниченное присутствие** за пределами двух крупнейших городов Казахстана и упущенные возможности в других регионах с растущим спросом.

Рекомендации

Для устойчивого роста и снижения рисков зависимости от одного бренда рекомендуется **расширить модельный ряд**: **увеличить долю Porsche и Audi**, а также **рассмотреть возможность добавления других премиальных марок** (например, BMW или Mercedes-Benz), **особенно в сегменте внедорожников**, где наблюдается наибольшая выручка.

Также имеет смысл **усилить продажи премиальных внедорожников**, т.к. они имеют высокую маржу и устойчивый спрос. Важным направлением станет **расширение городов присутствия** — выход в такие города, как Шымкент, Караганда и Павлодар, где формируется новый класс покупателей, готовых рассматривать премиальные предложения.

Наконец, стоит активнее развивать **корпоративные B2B-продажи**: заключение договоров с крупными компаниями, госструктурами и лизинговыми организациями позволит увеличить средний размер сделки, сделать поток заказов более стабильным и снизить зависимость от колебаний розничного спроса.