

AutoBridge Analytics: Data Methodology

Методология данных (Русский)

1. Философия и Прозрачность

Точность и абсолютная прозрачность лежат в основе AutoBridge Analytics. Мы убеждены, что грузинский автомобильный рынок, являясь крупнейшим реэкспортным хабом в регионе, нуждается в надежной цифровой инфраструктуре. Наша статистика строится не на опросах продавцов или ручных оценках экспертов, а представляет собой прямой, непрерывный срез реального цифрового автомобильного рынка.

2. Сбор данных (Data Collection)

Наши ETL-алгоритмы (Extract, Transform, Load) ежедневно агрегируют данные с крупнейших автомобильных маркетплейсов и платформ Грузии.

- **Объем:** Наша база данных ежемесячно пополняется десятками тысяч новых уникальных предложений.
- **Частота:** Данные отслеживаются в реальном времени, что позволяет нам фиксировать снижение или повышение цен, изменение пробега и финальный статус публикации (снято с продажи).

3. Очистка и Верификация (ETL Pipeline)

Сырые данные всегда полны "шума", дублирования и спама. Мы применяем строгие инженерные фильтры:

1. **Дедупликация:** Мега-дилеры часто публикуют один и тот же автомобиль на разных сайтах или создают дублирующие объявления для обновления даты. Наши скрипты склеивают такие публикации: мы считаем **уникальные автомобили**, а не количество их рекламных размещений.
2. **Аномалии и Спам:** Объявления с ценой \$1, \$999,999 или указанным пробегом 1 км для старых авто автоматически исключаются из аналитических выборок.
3. **Изоляция сегментов:** В метриках легковых автомобилей мы алгоритмически исключаем грузовики, спецтехнику, битые автомобили на запчасти (разборку) и посуточную аренду.

4. Ключевые метрики

- **Медианная цена (Median Price):** Для оценки стоимости модели мы **никогда** не используем среднее арифметическое (Average). Один ошибочно дорогой или битый автомобиль может сильно исказить среднюю цену. Медиана показывает реальную цену, по которой продается "середина рынка".
- **Индекс ликвидности (Liquidity Index):** Этот уникальный показатель отражает среднее время (в днях) жизни объявления на платформе. Чем ниже индекс (меньше дней), тем быстрее автомобиль перепродается, что сигнализирует о горячем спросе.

- **Доли рынка (Market Share):** Определяются пропорционально объему валидных, очищенных от дублей публикаций определенного бренда за выбранный таймфрейм.

5. Ограничения данных (Limitations)

Как и любая аналитическая платформа, мы имеем метрические ограничения. Мы анализируем "цифровой след" рынка. Сделки, которые осуществляются дилерами "в офлайне" (напрямую на авторынке без публикации в интернете), не попадают в нашу статистику до момента их оцифровки. Однако масштабы нашей цифровой выборки обеспечивают точность трендов выше 95%.
