

Исследование данных (Data exploration)

Разделы: [Бизнес-задачи](#)

Исследование данных представляет собой процедуру, которая обычно выполняется перед [анализом данных](#) и включает их визуальное изучение. Оно позволяет понять, что содержится в данных, а также оценить объем и полноту информации, корректность и связи между элементами.

Исследование обычно проводится с использованием комбинации автоматизированных и ручных действий. Первые включают [профайлинг](#), [визуализацию](#), формирование отчетов, которые позволят аналитику получить первоначальное представление о данных и понимание их ключевых свойств, важных с точки зрения анализа.

Все эти действия направлены на формирование общего представления и понимания данных аналитиком, а также определение основных [метаданных](#), которые будут использоваться при дальнейшем анализе.

По результатам исследования могут быть исключены из рассмотрения не пригодные для анализа данные, произведена первоначальная оценка их [качества](#), сформулированы цели и разработаны сценарии дальнейшей аналитической обработки.

Не следует путать исследование данных и [разведочный анализ данных](#) Тьюки. Первый механизм направлен в основном на получение общего представления о данных, формулирование первичных целей и задач их анализа, и опирается главным образом на визуализацию и несложные манипуляции с ними типа профайлинга. В то же время разведочный анализ ориентирован больше на формулирование статистических [гипотез](#) и их проверку, а также применение статистических моделей (например, [регрессии](#)), [понижения размерности](#) данных и отбор переменных в аналитические модели.

В аналитической платформе Loginom содержится широкий набор инструментов для исследования данных: профайлинг и оценка качества данных с формированием соответствующих [отчетов](#), а также разнообразные [визуализаторы](#).