



Декомпозиция аналитической задачи

Любая задача прилетает аналитику «сырой», в самой общей форме. Пытаясь решить её с наскока, вы рискуете совершить логические ошибки, неверно обработать данные и даже упустить из виду важные части задачи. Нужно делить её на блоки, а те на подзадачи. Это называется декомпозиция, и ей тоже стоит научиться.

В проектах учебной программы мы декомпоновали задачи за вас. Сейчас вам предстоит сделать это самостоятельно:

1. Вместе с заказчиком выработайте образ конечного результата

Например, «на основе персональных данных сформировать список пользователей, которым надо отказать в кредите».

Конкретизируйте форму представления результата: таблица, график, презентация или что-то другое. Всё, что касается заказчиков, обычно происходит и может потребовать нескольких итераций. Будьте терпеливы и настроены на сотрудничество, не отчаивайтесь после первых безрезультатных переговоров.

2. Выделите этапы решения

Какие шаги приведут вас к желаемому результату? Каждый этап должен иметь **один** конечный результат, например:

1. выгрузка данных;
2. график;
3. алгоритм.

3. Вы можете предварительно визуализировать декомпозицию в виде:
 1. блок-схемы;
 2. майнд-карты.
4. Проверьте себя. Если вы успешно завершите каждый из описанных шагов, получите ли вы заранее определённый конечный результат?

В трекере можете группировать похожие задачи. Например, «Знакомство с данными», «Предобработка», «Исследовательский анализ данных».

Не создавайте по задаче на каждый мелкий шаг: "Открыть Jupyter Notebook", "вызвать функцию print()". Изучите, как были сформулированы шаги в учебных проектах. Объединяйте мелкие действия наподобие «Избавиться от пропусков» и «Нормализовать данные» в одну задачу «Предобработка».

Вот неудачный пример декомпозиции задачи:



Теперь приведём пару удачных примеров. Это первый:

EDA-1: Выявление профилей потребления

🔖 1. Формирование техзадания

EDA-1: Выявление профилей потребления

🔖 2. Загрузка, первичный осмотр и предобработка данных

EDA-1: Выявление профилей потребления

🔖 3. Проведение исследовательского анализа данных (EDA)

EDA-1: Выявление профилей потребления

🔖 4. Сегментация покупателей на основе истории их покупок

EDA-1: Выявление профилей потребления

🔖 5. Проверка статистических гипотез

EDA-1: Выявление профилей потребления

🔖 6. Презентация

Второй, более детальный:

EDA-75: Декомпозиция задачи

🔖 Шаг 1. Изучение общей информации. Подготовка данных.

EDA-75: Декомпозиция задачи

🔖 Шаг 2. Добавление нужных столбцов. Исследовательский анализ.

EDA-75: Декомпозиция задачи

🔖 Шаг 3. Провести категоризацию столбца 'product' и кластеризацию по столбцам 'price' и 'quantity'. Найти заказы в которых было куплено только одно наименование товара

EDA-75: Декомпозиция задачи

🔖 Шаг 4. Оценить качество разных способов разбивки по категориям.

EDA-75: Декомпозиция задачи

🔖 Шаг 5. Провести окончательную границу между категориями. Построить графики, сформулировать и проверить статистические гипотезы.

EDA-75: Декомпозиция задачи

🔖 Шаг 6. Выводы, рекомендации.

