

Для наглядного представления специфик работы аналитика данных, представляем вам своеобразное "интервью с аналитиком".

Ева - аналитик данных. Она работает в компании, которая разрабатывает мобильных игр (англ. *game development*, «разработка игр»). Очень популярные мобильные игры. В прошлом году выручка компании превысила половину ВВП одной небольшой страны... Как же проходит день специалиста по анализу данных?

Ева

Мы, специалисты в сфере данных, всегда в тени. Хотя без нас мир был бы совсем другим...

Впрочем, мы отвлеклись. Прежде чем передать руководству полученные выводы, я собираю информацию о проблеме, затем продумываю направление и конкретный алгоритм действий — что-то подобное ты уже делал, когда анализировал поломки робокотов. Потом я программирую — открываю и исследую данные, строю графики и преобразую значения, чтобы проверить гипотезы. Например, гипотезу о том, что новая фишка сильно влияет на баланс игры.



Постановка задачи

Получив задачу, специалист всегда её уточняет. Проблема обозначена: пользователи игр массово жалуются на новую функцию — уплотнитель. На онлайн-соревнованиях она фактически определяет победителя, потому что превзойти применяющих уплотнитель игроков невозможно. Осталось провести исследование и выяснить, так ли это.

Продуктовый менеджер



А ты не можешь побыстрее это посчитать без всяких уточнений? Там же вроде ничего сложного. Просто посмотри, сколько игроков применяют уплотнитель. Если таких будет слишком много, возьмём да и уберём его.

Ева



Предлагаю ещё раз проговорить задачу. *Какую бизнес-проблему ты хочешь решить?*

Продуктовый менеджер



Ладно. Только давай недолго, у меня следующая встреча через 15 минут.

Проблема такая: игроки заваливают поддержку жалобами. Нужно не допустить разрастание негатива и потерю пользователей.

Ева



А какую долю такие жалобы ежедневно составляют от общего числа обращений?

Продуктовый менеджер



Этого мы не считали. Но в целом количество негативных обращений выросло. Думаю, что причина как раз в уплотнителе.

Ева



Понятно. За какой период мне изучить данные?

Продуктовый менеджер



Хм, мы внедрили уплотнитель полгода назад... Давай за последние шесть месяцев.

Ева



Хорошо, тогда вот предварительный вариант задачи: нужно посчитать количество применявших уплотнитель игроков за последние шесть месяцев. Верно?

Продуктовый менеджер



Да, похоже, что так.

Уточнение задачи не всегда лёгкий, но крайне важный процесс. От правильной постановки задачи зависят результаты исследования, от результатов исследования — принятые руководством решения, а от принятых решений — успешность всего бизнеса. Не стесняйтесь задавать вопросы. Даже если вашим коллегам этот этап кажется необязательным.

Продуктовый менеджер



Ну что, всё? Можно расходиться?

Ева



Почти, ещё несколько вопросов. *Каковы обязательные требования к результату?*

Продуктовый менеджер



Что ты имеешь в виду? Надо быстро, понятно и data-driven. Чтобы я мог принять решение.

Ева



Например, требование может быть таким: из результата видны различия в какой-нибудь метрике между игроками с уплотнителем и без.

Продуктовый менеджер



Да нет, просто посчитай, сколько пользователей применяют уплотнитель.

Ева



А как ты будешь использовать эту информацию?

Продуктовый менеджер



Я пойму, насколько всё плохо. Смогу решить, стоит ли менять игру и убирать уплотнитель.

Ева



А как ты поймёшь, что «всё плохо»?

Продуктовый менеджер



Если игроки с уплотнителем и правда постоянно выигрывают соревнования, то это беда.

Ева



Ага, вот оно! То есть нужно смотреть на метрику «количество побед».

Если сформулирована гипотеза, которые содержат цель и действие: «Выяснить причину поломок роботов, чтобы снизить количество возвратов». Это продуктовые гипотезы. Ещё можно выдвигать статистические гипотезы, которые составляют только часть продуктовой. Например: «Все сломанные роботы принадлежат новой партии».

Это исследовательская часть, и над ней можно работать отдельно.

Ева

Значит, гипотеза звучит так: «Игроки с уплотнителем чаще побеждают на соревнованиях, чем игроки без уплотнителя». Проверять будем на данных за последние шесть месяцев.



Кстати, а какие категории пользователей включать в анализ?

Продуктовый менеджер

Давай все. Сейчас что-то совсем ничего не понятно.



Тебе нужна ещё какая-то информация, чтобы начать работу? Данные у тебя вроде есть?

Ева



Последнее — *в каком формате ты хочешь получить результат?*

Продуктовый менеджер



Как обычно: презентация, графики какие-нибудь, выводы. Твои рекомендации тоже будут кстати, ты в этой теме разбираешься.

Ева

Хорошо. Давай финализируем задачу.

Есть гипотеза о пользователях, которые применяют уплотнитель. Нужно выяснить, есть ли разница в количестве побед между игроками с уплотнителем и без. По результатам исследования я подготовлю отчёт с выводами и рекомендациями.



Продуктовый менеджер



Отлично, всё так! Удачи. А я побежал на следующую встречу.

Мы можем видеть, что при решении текущих задач аналитик должен следовать следующим правилам.

Уточнять смысл задачи

Акцентируйте внимание на том, что сделала Ева в разговоре с продуктовым менеджером:

1. **Уточнила первоначальный запрос.** Ева узнала, как менеджер планирует применять запрашиваемую информацию, и это помогло ей выдвинуть гипотезу.

Примеры вопросов: Какую бизнес-проблему надо решить? Какую долю жалобы составляют от общего числа обращений ежедневно? За какой период изучить данные?

2. **Сформулировала промежуточный результат.** Это позволило сфокусировать беседу.

Пример вопроса: Нужно посчитать количество применявших уплотнитель игроков за последние шесть месяцев. Верно?

3. **Уточнила требования.** То есть определила пожелания менеджера к результату.

Примеры вопросов: Каковы обязательные требования к результату? Как ты будешь использовать эту информацию?

Задача про уплотнитель приобрела законченный вид и даже немного изменилась — теперь нужно проанализировать не просто распределение уплотнителя по игрокам, но и связь этого показателя с числом побед на соревнованиях.

Какие ещё бывают задачи

Глава отдела финансов

СФО

Нужна выгрузка по всем платящим пользователям за вчерашний день! Очень желательно в ближайшее время.

Ведущий разработчик

СТО

Надо проанализировать, как пользователи работают с нашим приложением на Android.

Директор по маркетингу

СМО

В пять часов выступаю с презентацией перед генеральным директором! Нужно срочно добавить в презентацию количество пользователей, которые воспользовались нашим специальным предложением.

На работе вы получите очень много подобных запросов с пометкой «срочно». Обижаться на коллег не стоит, они не хотят вам зла. Просто именно так чаще всего происходит первичная постановка бизнес-задачи. Специалист направляет её в правильное русло: задаёт вопросы, выдвигает гипотезы и утверждает итоговую формулировку.

Глава отдела финансов

СФО

Нам нужна выгрузка по всем платящим пользователям за вчерашний день! Очень желательно в ближайшее время.

Ivan

Зачем?



Глава отдела финансов

СФО

Хотим посмотреть, какие подписки они оформили.

Даже один простой вопрос существенно меняет дело: оказывается, узнать нужно не количество платящих пользователей и даже не суммы их покупок, а категорию платежей.

Сценарий уточнения задачи всегда зависит от контекста и содержания самого запроса — универсальной формулы нет.

Ведущий разработчик

СТО

Надо проанализировать, как пользователи работают с нашим приложением на Android.

Ivan

Какой элемент в работе с приложением вы хотите проанализировать?



Ведущий разработчик

СТО

Как пользователи начинают играть: из своего профиля или через главное меню.

Часто важно спросить, данные по каким пользователям необходимо проанализировать. Например, это могут быть новые пользователи или совершившие уже несколько покупок.

И ещё — всегда полезно уточнить время. Человек может легко потеряться в днях и часах: отправить вам email после полуночи или вообще из другого часового пояса. В таких случаях есть риск разминуться в определении таких понятий, как «завтра».

Декомпозиция

Ева уточнила задачу и выдвинула гипотезу. Далее следует продумать этапы решения.

Как ни странно, всё начинается с результата, то есть *образа конечного результата*. Если сразу не решить, куда вы движетесь и зачем, можно столкнуться с одной из самых частых проблем в анализе данных. Эта проблема на профессиональном сленге называется «закопаться в задаче».

Большинство аналитических задач имеет множество ответвлений, которые способны увести далеко от первоначальной идеи.

Поэтому важно сразу понять, какой именно результат решит вашу задачу. Например, конечным результатом, в рассматриваемом примере Евы, будет подтверждённая или опровергнутая гипотеза, сформулированные выводы и отчёт с визуализацией.

Сформировать образ конечного результата поможет информация, полученная при уточнении задачи. Иногда могут быть расхождения: например, заказчик хочет получить только один график, но вам для этого нужно подготовить целое исследование. Тогда исследование тоже будет частью конечного результата — вы просто сохраните его на будущее.

Когда образ конечного результата сформирован, можно приступать к декомпозиции задачи. То есть разбивке её на небольшие шаги. Правильно декомпозировать задачу поможет знание классического процесса анализа данных. Вот его этапы:

Постановка задачи

Уточнение запроса и формирование образа конечного результата.

Получение данных

Данные всегда где-нибудь хранятся: в базе, файле, на сайте, в онлайн-таблице. Если вы хорошо знаете, где взять данные для проекта, на этом этапе важно получить их в удобном для анализа формате. Если же вам предстоит работа с новыми источниками данных, обратитесь к разработчикам или менеджерам — они подскажут, где лежат данные и как их получить.

Предобработка данных

На практике данные часто содержат пропуски, дубликаты, ошибки. Например, при заполнении формы регистрации на сайте кто-нибудь обязательно забудет написать фамилию, а цена товара вместо ожидаемых рублей может быть выражена копейками. Ваша задача во время предобработки — найти и устранить такие проблемы. Обнаружить абсолютно все оплошности вам вряд ли удастся, но результатом этого этапа будут готовые к дальнейшему анализу данные.

Исследовательский анализ данных

Исследовательский анализ данных, или EDA, предполагает построение графиков по каждому столбцу в наборе данных. Такая визуализация позволяет увидеть некоторые особенности, закономерности и аномалии. Например, график может показать, что определённый курьер в компании доставляет заказы неправдоподобно быстро. Сделанные наблюдения помогут вам разобраться в результатах дальнейшего анализа.

Подготовка прототипа решения

На этом этапе создается первый рабочий вариант решения, или прототип. Работа над аналитической задачей — циклический процесс. Иногда прототип может стать её финальным решением, но обычно хороший результат требует неоднократных доработок.

Проверка результатов

Решение подготовлено — далее следует изучить результаты. Это очень ответственный момент: вы определяете, решена ли задача или нужно ещё что-нибудь доделать. Как это определить? Во-первых, соотнесите результаты с поставленной задачей и образом конечного результата. Во-вторых, оцените правдоподобность результатов, руководствуясь здравым смыслом. Могут ли быть такие значения в этом контексте? Почему значения именно такие? Результат этого этапа — завершение проекта или выявленные недоработки.

Доработка решения

Обнаружив несовершенства в прототипе решения во время проверки результатов, вы переходите к его доработке. Этот этап замыкает цикл из трёх основных шагов решения аналитической задачи: подготовка прототипа решения —> проверка результатов —> внесение изменений. Цикл повторяется до тех пор, пока результат вас не удовлетворит.

Оформление результатов

Задача решена, но заказчик об этом пока не знает. Вы можете сделать выгрузку данных или подготовить отчёт с визуализацией результатов, выводами и рекомендациями. Как узнать, в каком формате заказчик ожидает результат? Лучше всего просто спросить. Дополнительно уточните, как заказчик планирует применить полученное решение: возможно, вы предложите более подходящий вариант представления результата.

Проверка на практике (проверка реальностью)

Наиболее интересный, но и самый сложный этап. Результаты работы аналитика обычно трудно сразу же оценить. Например, вы изучили информацию о пользователях приложения и представили свои рекомендации по повышению продаж. После этого заказчик исследования, обычно менеджер из команды развития бизнеса, решает — принять ваши предложения или нет. Если ответ — да, через некоторое время можно будет оценить их эффективность. Так вы узнаете, действительно ли ваше решение было корректным.

Каждый этап хорошо декомпозированной задачи имеет какой-нибудь весомый результат: подготовленные к анализу данные, график, программа. Выполнение всех шагов приводит к заранее определённому конечному результату.

Вернемся к примеру с Евой.

Она отлично поработала. В ходе исследования выяснила, что игроки определённого стиля и уровня игры действительно имеют преимущество, применяя уплотнитель. Однако остальных игроков это никак не затрагивает и таких большинство.

Что делать дальше?

- Показать заказчикам исследование с кодом и графиками;
- Отобрать графики и выводы для презентации — по промежуточным и финальной гипотезам;
- Подготовить презентацию.

Заказчик не всегда бывает согласен с исполнителем. Если так происходит, постарайтесь выяснить аргументы заказчика. Иногда на несогласие есть веские причины: например, внедрить предложенные решения невозможно из-за недостатка работников или бюджета.

В иных случаях важно уметь защитить свою позицию. Специалист владеет в таких спорах особым оружием: любые аргументы будут весомее, если подкрепить их конкретными цифрами и выводами из исследования. Всегда полезно показать заказчику, как текущее решение влияет на бизнес и что будет, если его изменить.

Так сделала и Ева.

На встрече с менеджером она предложила не отключать уплотнитель для всех игроков: лояльные пользователи, которые играют давно и не входят в «несправедливый» сегмент, могут отказаться от игры. Ева представила презентацию на основе своего исследования и показала, что для решения проблемы достаточно лишь изменить уплотнитель в том сегменте, где его применение делает игру нечестной и вызывает недовольство игроков. Менеджер и остальные коллеги после некоторых препирательств приняли эту рекомендацию. Вскоре жалобы игроков на уплотнитель прекратились.

Проблемы и возможности

Аналитик активно взаимодействует с бизнесом по всем вопросам: от рутинных задач до крупных исследований. Это логично, ведь аналитики влияют на принятие важнейших решений в компании, а для этого важно быть глубоко погружённым в бизнес-процесс.

Задачи аналитика чаще всего состоят из проблем и гипотез, которые формулирует бизнес. При этом потребность в аналитике бизнес испытывает обычно по двум направлениям: исследование текущих проблем или поиск точек роста.

Исследование проблем

Допустим, компания запустила новую форму регистрации на сайте. Сразу после обновления графики суточной посещаемости и количества регистраций обвалились, и никто не понимает почему. Показатели не нулевые: кто-то всё же регистрируется, но таких очень мало.

Паниковать не стоит. Как аналитик, вы можете изучить данные и выяснить, что все зарегистрировавшиеся пользователи перешли на сайт по ссылкам из разных рекламных источников. Однако один такой источник — самый крупный рекламный баннер — не принёс вам ни одной регистрации. Проверив баннер, команда выясняет, что в него просто забыли добавить ссылку на ваш сайт. Расследование таких случаев — часть работы аналитика.

Поиск точек роста

Когда проблем нет, у аналитика также много работы: нужно помогать бизнесу находить «узкие» места и новые точки роста. Например, пользователи на сайте компании проходят очень длинный путь от стартовой страницы до начала работы с вашим сервисом. Вы анализируете метрики и видите, что из-за этого компания теряет множество потенциальных клиентов. Дайте свои рекомендации: добавить на главную страницу возможность попробовать сервис. Важно только убедить команду сделать это сначала в порядке эксперимента на небольшом числе пользователей: если он будет успешным, то вы обнаружили новую точку роста — можно запускать нововведение; если нет, вы просто откажетесь от этой идеи и будете исследовать другие, не нанеся компании никакого вреда.