



01.

Принятие решений на данных

Курс аналитики Доставки, 2023



Функции аналитики



Что такое аналитика?



В чем функции аналитики? **и какие задачи решают аналитики?**

Что важно понимать?

*Аналитик (как и продакт, как и дизайнер, как и UX-исследователь и менеджер) не решает задачи, он решает **ПРОБЛЕМЫ***

Исполняемость заказов в одной из опций ниже в 4 раза, чем средняя

- можно было спросить у людей, что случилось?
(рисерчер/продакт)
- можно было самому заказать 10 раз
(тестировщик/дизайнер)
- а можно посмотреть воронку и исследовать аномалию (аналитик/продакт/бизнес)

Задачи у всех одинаковые

Отличаются базовые подходы к решению:

- Инструментами аналитика являются данные
- Инструментами дизайнерами являются интерфейсы (?)
- Инструментами ресерчера являются люди
- Инструментами продакта являются продукты...

Главный вопрос аналитика



Зачем?

Главный вопрос аналитика



Зачем?

The diagram consists of two identical sets of overlapping rounded rectangles. Each set has a dark gray rectangle in the background and a light gray rectangle in the foreground. The light gray rectangles contain the text 'Зачем?' and 'Какую проблему решаем?' respectively. The rectangles are slightly offset to the right and down from the background ones.

Какую проблему
решаем?

Главный вопрос аналитика



Зачем?

Какую
проблему
решаем?

Как понять что
мы сделали
хорошо?
(критерии
успеха)

Стратегия и
принятие
решений

Аналитика в
продукте и
бизнесе

Функции аналитики

Отчетность и
мониторинг

Поддержка
продуктовых и
операционных
процессов

Research и
генерация
гипотез

Стратегия и принятие решений:

- Сегментация
- Разработка стратегии
- Оптимизация затрат
- ...

Функции
аналитики

Отчетность и мониторинг целевых показателей бизнеса и продукта для принятия data-driven decisions:

- Отчетность и мониторинг ключевых метрик;
- Воронка метрик/Пирамида метрик/Дерево метрик;
- **KPI / OKR**
- Единая методология расчета целевых метрик, единая методология проведения экспериментов (на сколько это возможно).

Функции
аналитики

Функции аналитики

Research и генерация гипотез:

- Решение задач анализа бизнес-показателей, позволяющие находить точки роста в продукте;
- Сегментация пользователей на основе данных, поиск и выделение фокусных сегментов, поиск инсайтов на данных;
- Генерация гипотез на основе данных для качественных и UX-исследований;
- Сложные кросс-функциональные задачи на стыке бизнес-аналитики, operations, аналитики роста и продуктовой аналитики, анализ когорт, retention;
- Теоретические исследования поведения; построения моделей, поиск зависимостей, изучение маркетинг-плейса и т.п.

Поддержка процессов:

- Оценка задач в показателях, в которых оценивается эффективность бизнеса;
- Принятие решений на данных про делать/не делать; приоритезация, планирование спринтов;
- Участие в discovery и grooming процессах;
- Поддержка экспериментов, методология проведения тестов, логирование результатов; задачи на dwh и разработку, связанные с данными;
- Ретроспективная оценка результата продуктовых и бизнесовых изменений; рекомендация по раскатке (в т.ч. катим/не катим); оценка долгосрочного эффекта от изменений;
- Участие в презентации результатов, доведения до бизнеса эффектов и инсайтов

Функции аналитики

Компетенции аналитика

Базовые навыки аналитика данных



Для того, чтобы эффективно справляться с описанными задачами, аналитику нужно:

- иметь отличную математическую базу – за аналитиком не хочется перепроверять выкладки и формулы
- понимать базовую теорию вероятности и математическую статистику – нужно уметь проверять гипотезы, понимать ошибки разного рода, зависимость / независимость испытаний и т. д.
- иметь математическую культуру – если аналитик использует метод или алгоритм, он должен знать область его применимости

Базовые навыки аналитика данных



А также..

- обладать критическим мышлением – реже, чем другие люди попадаться в ловушки когнитивных искажений
- обладать продуктовым мышлением – уметь оцифровывать пользовательский опыт в метриках, а также видеть за метриками пользователей, пытающихся решить определенную задачу
- обладать бизнесовым мышлением – уметь оцифровывать бизнес-процессы компании и изменения рынка, связать это воедино с продуктом и пользователями

**Что мы проверяем на
собеседованиях ?**

Критическое мышление

Критическое мышление

Журналисты спросили у 20 детей в одном детском саду, живут ли их родители вместе.

Двое детей ответили, что не живут.

На следующий день была опубликована статья под заголовком "Каждая десятая семья с маленькими детьми в нашем городе разведена!"

Перечислите основные причины, из-за которых такая оценка может быть сильно смещена.

(с) Николай Мацук

Критическое мышление и понимание статистики

Журналисты спросили у 20 детей в одном детском саду, живут ли их родители вместе. Двое детей ответили, что не живут. На следующий день была опубликована статья под заголовком "Каждая десятая семья с маленькими детьми в нашем городе разведена!"

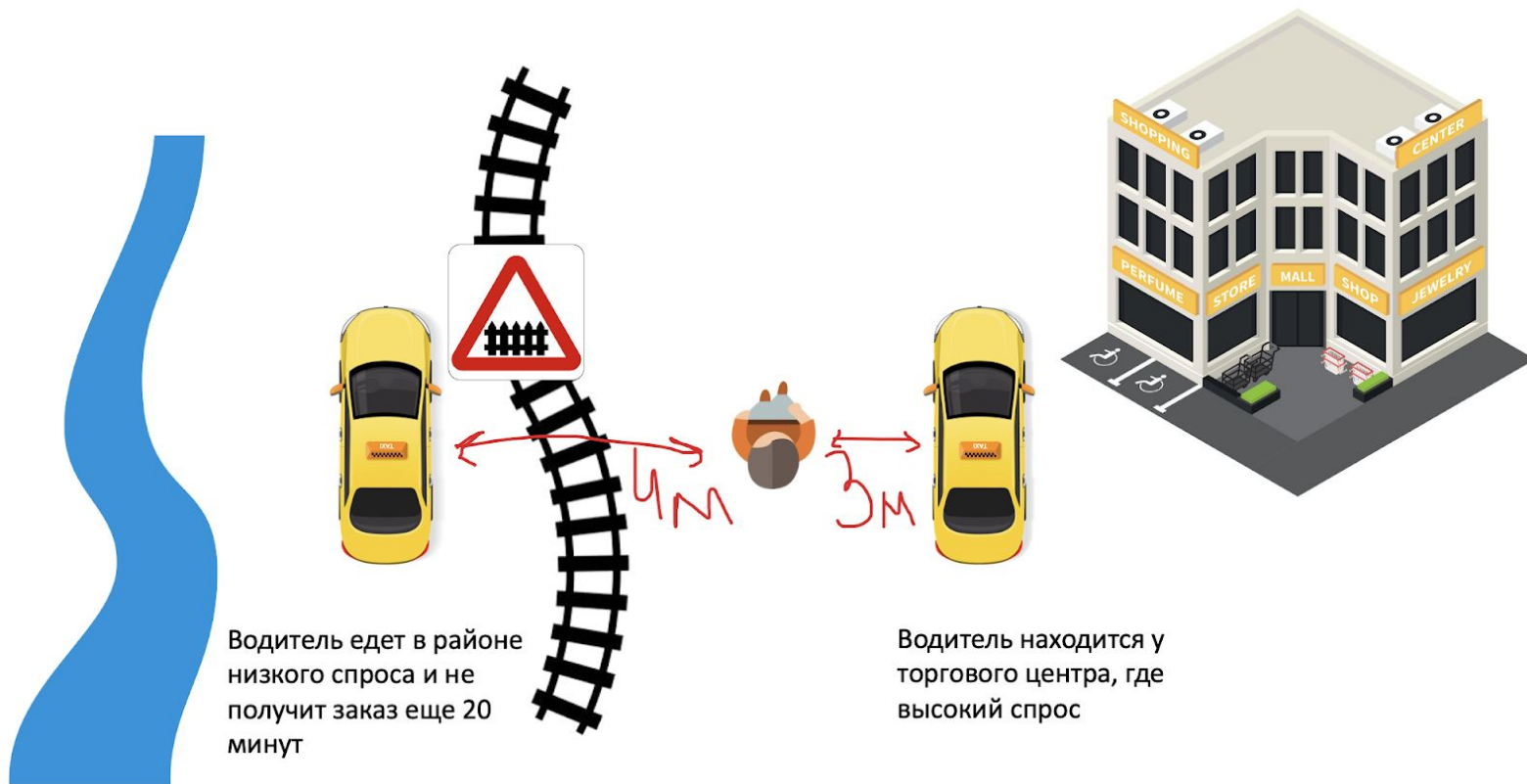
Перечислите основные причины, из-за которых такая оценка может быть сильно смещена.

(с) Николай Мацук

1. Малый размер выборки
2. На одну семью с детьми в среднем приходится больше одного ребенка
3. Вдовы/вдовцы != разведенная семья
4. Детский сад может располагаться в наиболее или наименее благополучном районе
5. Не во всех семьях детей водят в детские сады
6. Дети могут скрывать/обманывать

Системное мышление

Кому отдать заказ?



Продуктовое мышление





2 лифта

Представьте два идентичных (“статистически”) офисных здания.

В них работает одинаковое количество людей, одинаково распределенное по этажам и одинаково распределенное по времени приходы и ухода на работу и с работы.

В каждом из зданий для простоты есть по одному лифту. Лифт - классический, то есть с кнопкой вызова снаружи (на каждом этаже) и кнопками этажей внутри.

Оба лифта с технической точки зрения тоже одинаковы (скорость передвижения, грузоподъемность и другие характеристики идентичны).



2 лифта

К вам как к аналитику приходит продакт менеджер, его команда написала новый алгоритм движения лифта, который должен быть “лучше” старого потому, что решает ряд выявленных в ходе качественных интервью проблем.

Алгоритм движения лифта - это то, как он реагирует на нажатия кнопок внутри и снаружи и как принимает решение о движении и остановках.

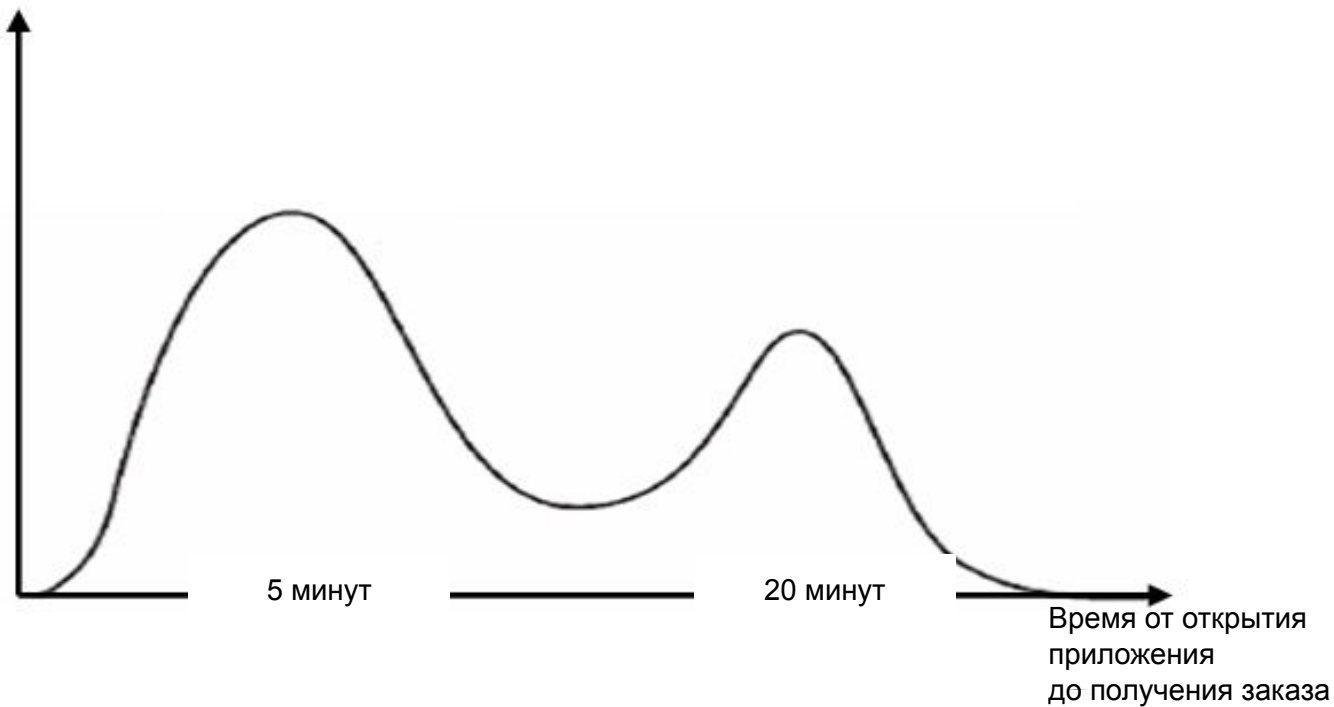
У вас есть данные о положении лифта в каждый момент времени, логи нажатия на кнопки внутри и снаружи и датчик веса в лифте.

Расскажите, как вы будете принимать решение о новом алгоритме движения, лучше ли он старого?

Генерация гипотез

Генерация продуктовых гипотез

Доля
водителей
(сглаженная
гистограмма
по бакетам)



Исследования и генерация гипотез в широком смысле

- Анализ существующих проблем и поиск точек роста; анализ корневых причин, или причинно-следственная диаграмма
- Поиск корреляций и взаимосвязей в данных и метриках
- Сравнение двух процессов, в которых что-то отличается (пример: регионы, сегменты, и тп); бенчмаркинг
- Поиск аномалий: все что не укладывается в ожидаемое поведение продукта, сервиса, пользователей и тп
- Временной анализ, тренды и прогнозирование, аплифт моделирование
- Сегментация и кластеризация, кумулятивные кривые распределений, CDF
- Когортный анализ

Основные проблемы в интерпретации метрик и манипуляции данными

Не запрещённый контент: радикальный феминизм

За шесть лет «радфем» стал почти
в 100 раз популярнее

2014 год

292

сообщений

2020 год

33 000

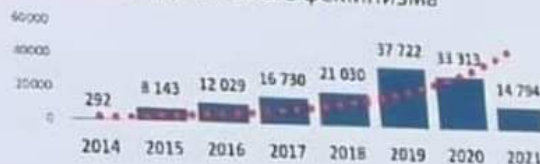
сообщений

Радикализация:

- Мужененавистничество
- Ненависть к семье
- Атаки на «неправильных» женщин
- Политизация, протесты

**Радфем ведёт к мужененавистничеству, к разрушению
семьи и фрагментации общества**

Динамика сообщений с признаками
радикального феминизма



Я стала феминисткой, потому что
4 июля в 6:26

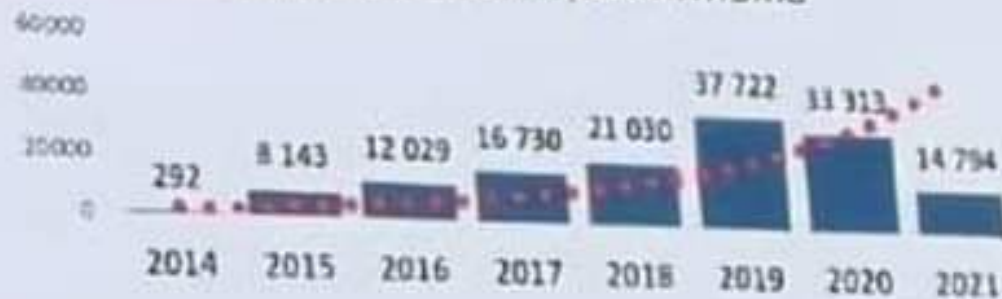
в большинстве своем мужчины - кретины, выращенные такой культурой патриархата, агрессивные под воздействием тестостерона и не умеющие себя контролировать, от которых можно ожидать чего угодно, и только: держась вместе, мы сможем сохранить все то доброе и светлое, что еще осталось в этом мире

А. Далиса Питвинова

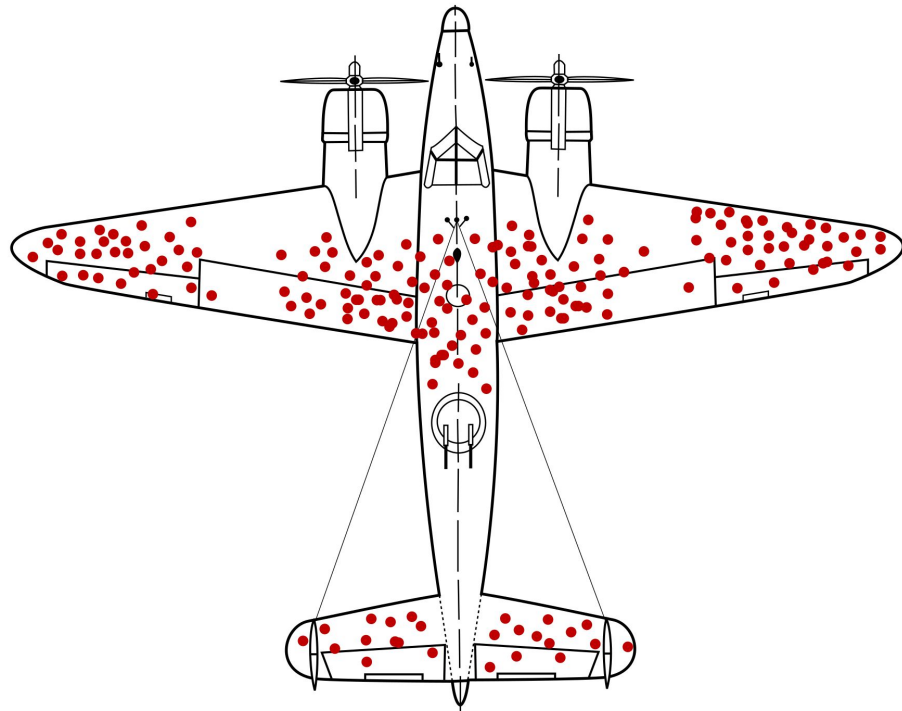
173 11

С. Т. 74

Динамика сообщений с признаками радикального феминизма



Ошибка
выжившего
(survivorship bias)



Неслучайное - случайно

Исследование о том, вредны ли линии электропередач для здоровья.

В течении 25 лет исследовали корреляции между близостью ЛВП и заболеваниями (более 800 видов).

Исследование показало, что уровень заболеваемости детской лейкемией был в 4 раза выше среди тех, кто жил ближе к линии электропередач.

Ошибка меткого стрелка

Гипотеза о том, морепродукты улучшают работу мозга и повышают результаты на экзаменах.

Собрана группа студентов, которым предложили добавлять суши к их рациону.

Через несколько месяцев они сдали экзамены – но оказалось, что результаты не лучше, чем у остальных.

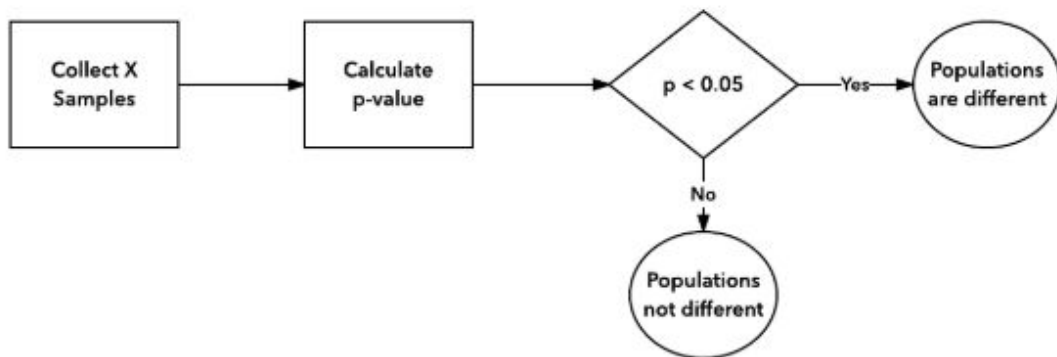
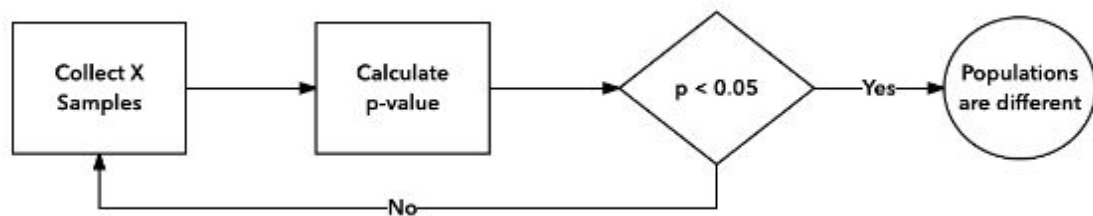
Однако, если выделить в группе первокурсников появляется корреляция. А если ограничить группу до белых мужчин, результат выглядит еще лучше.

Нельзя
подглядывать!

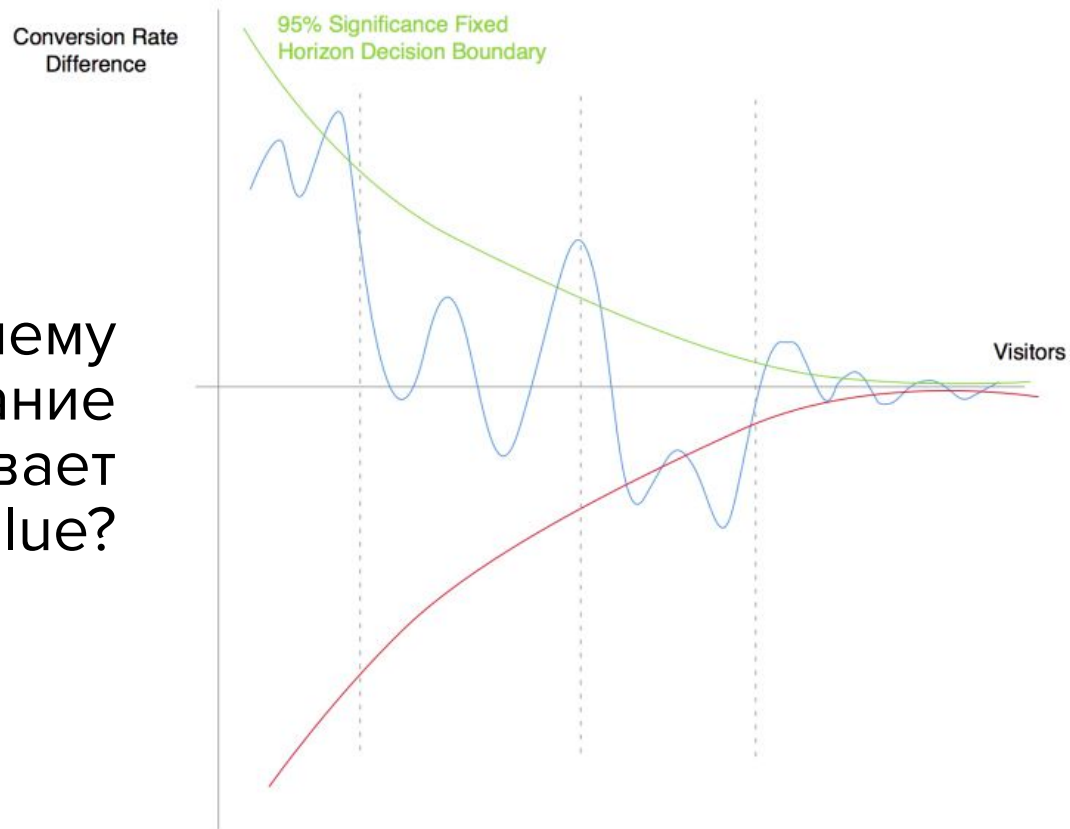


Почему
подглядывание
увеличивает
p-value?





Почему
подглядывание
увеличивает
p-value?



Почему подглядывание увеличивает p-value?

Даже если две группы идентичны, то разница конверсий может периодически выходить за границы зоны неразличимости по мере накопления наблюдений. Это совершенно нормально, так как границы сформированы так, чтобы при тестировании одинаковых версий лишь в 95% случаев разница оказывалась в их пределах.

Сетевой эффект

Сетевой эффект — это ситуация в экспериментах, когда поведение одного потребителя услуги или товара может влиять на характеристики или даже возможность предоставления услуги/товара для другого потребителя. Этот эффект особенно актуален для двустороннего маркетплейса, потому что предположение об отсутствии влияния (SUTVA) здесь не может быть принято.

Пример

Курьерам поставили персональные цели

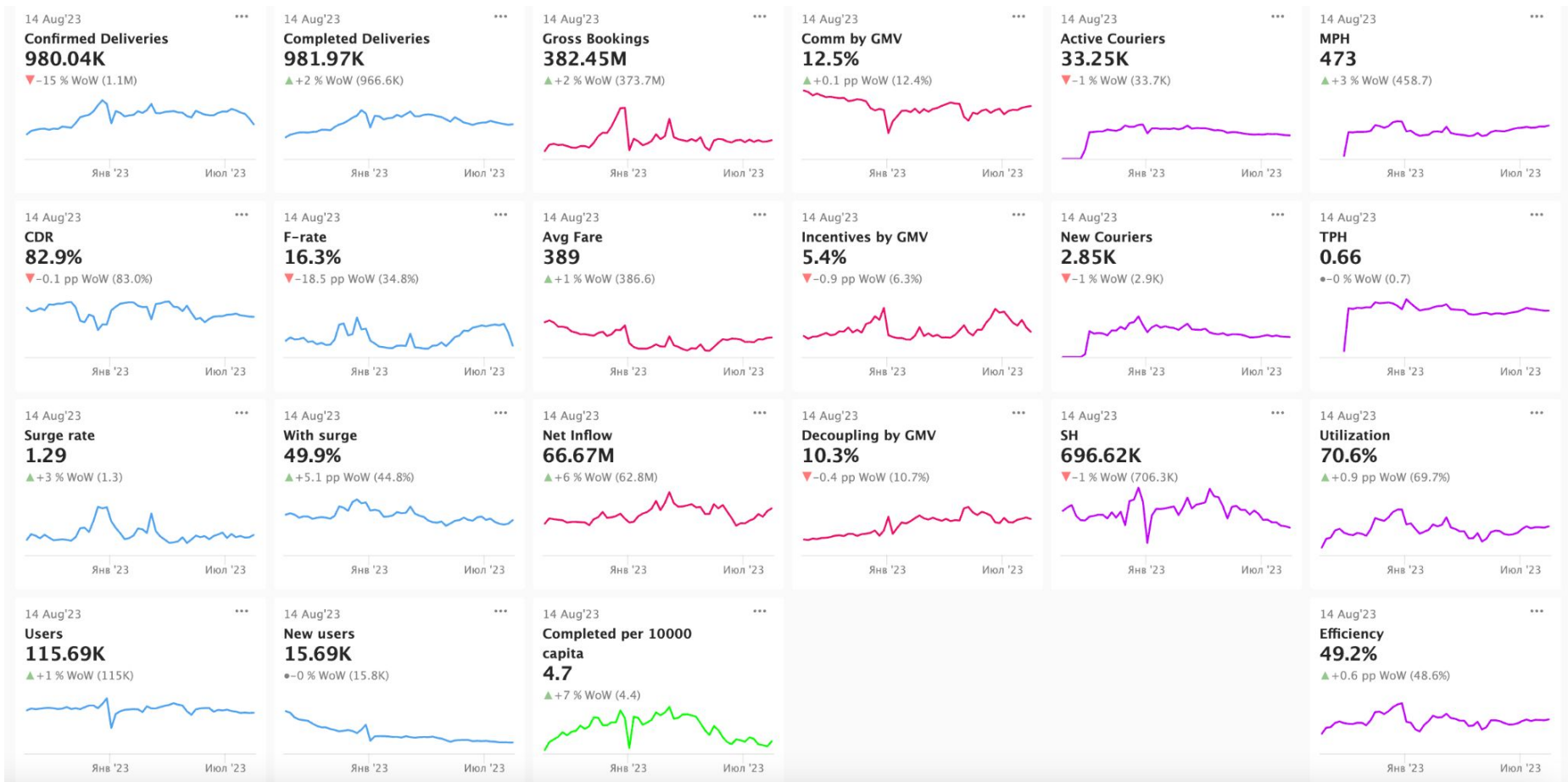
В А/Б измеряем количество заказов, нормированное на размер группы

Имеем каннибализацию экспериментом контрольной группы

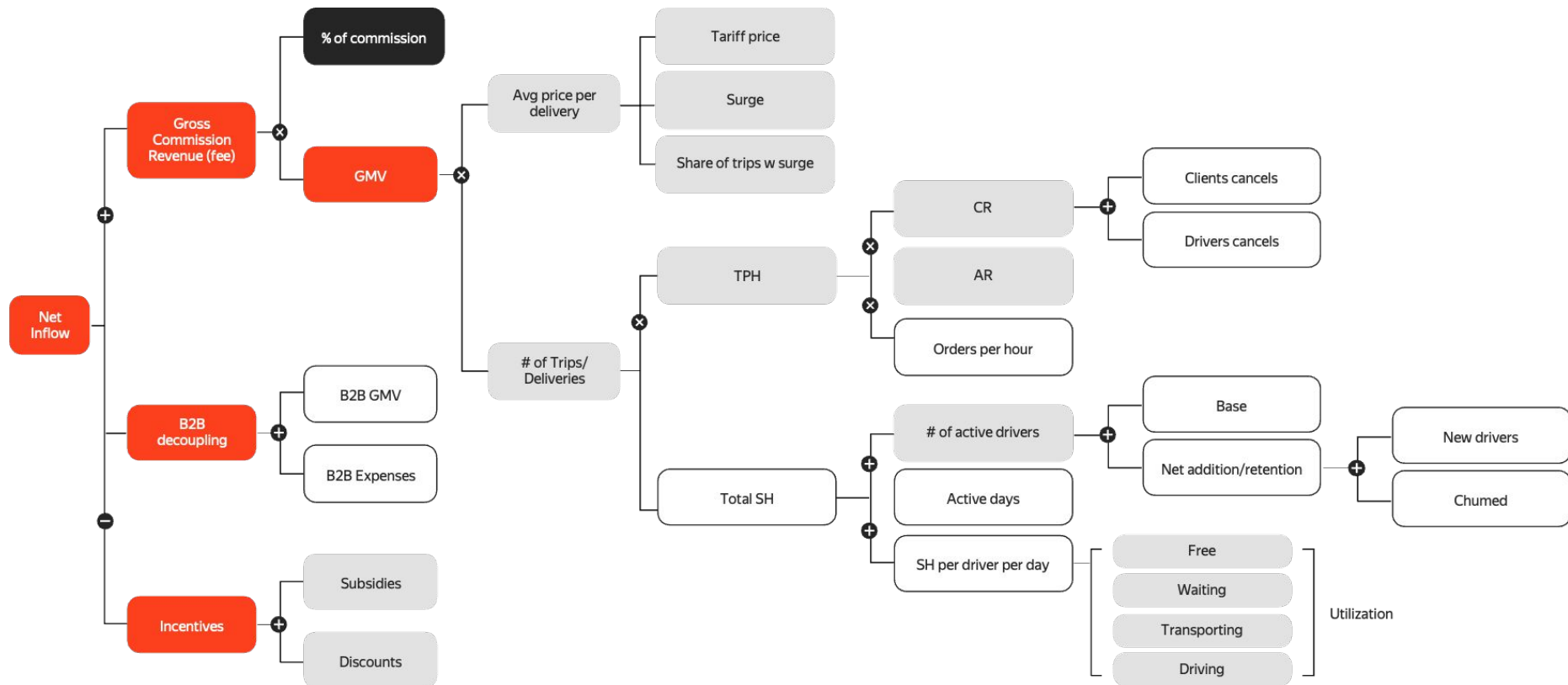
Систематические ошибки анализа

- Ошибки восприятия и интерпретации;
- Манипуляции с данными (выбрать сегмент, на котором есть результат уже после проведения эксперимента);
- Ошибка «меткого стрелка»;
- Ошибки выжившего и парадокс доступности информации;
- Сетевые эффекты.

Метрики и иерархия метрик



Пример декомпозиции метрики в виде дерева метрик



Алгоритм построения иерархий и деревьев метрик

В целом алгоритм построения иерархий можно описать следующим образом:

- ❑ 1) Какие **цели** у нашего бизнеса/продукта? Где мы хотим быть через 5 лет? через 2 года? через год?
- ❑ 2) Как мы будем **измерять** наш успех? какие у нас могут быть OKR (Objectives and Key Results «цели и ключевые результаты»)?
- ❑ 3) Далее мы можем сформулировать одну или несколько **ключевых метрик** (обычно финансовых), по которым мы будем оценивать успех компании (например, капитализация, прибыль, оборот и тп). Эту метрику нужно перевести в бизнес-метрики или продуктовые метрики
- ❑ 4) Как быть, если ключевых метрик получилось несколько? Есть два основных подхода:
 - они созависимы: например, прибыль и выручка, и одну можно поставить выше, выразив через другую (прибыль = выручка - koszty)
 - они находятся в сложной нетривиальной зависимости (например, покупатели на Авито и выручка компании с продавцов), тогда можно либо повторить первый метод, либо сделать два отдельных дерева метрик, где одни и те же показатели будут встречаться на разных уровнях
- ❑ 5) Основной алгоритм, который применяется для движения от «корней» дерева к «листам», похож на современные алгоритмы построения деревьев решений. Смотрим, от чего зависит данная вершина дерева больше всего, а затем пытаемся сократить энтропию в возникающей модели.

Принятие решений на данных

Как принимать решение на данных?

- Формулировка проблемы, которую мы решаем;
- Обоснование проблемы в виде качественных и количественных сигналов:
 - CX, researches, аналитика
- Четкая формулировка цели, исходя из сформулированной проблематики:
 - **Что мы хотим получить в качестве результата?** (не путать результат решения задачи и «результат» для продукта / бизнеса)
- Формулируем важные допущения (assumptions), которые мы сделали при оценке проекта и необходимости его реализации, а также в оценке потенциала;
- Формулируем гипотезы, которые помогут принять решение по проекту. Как можно эти гипотезы подтвердить / опровергнуть?
- Формулируем задачу и то, что хотим сделать в качестве решения. Важна четкая логика — что есть *проблема*, что есть *средство ее решения*, что есть *результат*?
- Существует и приведено исчерпывающее описание возможных решений проблемы и способов достижения цели, а также результаты предварительного анализа каждого из вариантов; ака **Почему мы для решения этой проблемы делаем именно это, а не что-то еще?**
- От продуктовых метрик в задачах (конверсий, usage, и тп) переходим к метрикам роста/качества (увеличение частотности, LTV, дополнительные поездки, дополнительный GMV, новые пользователи) и метрикам эффективности (сокращение затрат на заказ/пользователя).



Спасибо!

Курс аналитики Доставки, 2023

