



Визуализация данных

Tinkoff.ru

Кто я?



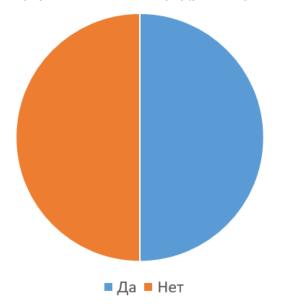
- Ольга Костина
- Руководитель отдела аналитики по кредитованию наличными
- → > 5 лет в Тинькофф
- → 2015 Мехмат МГУ, кафедра математической статистики

Зачем вам эта лекция?



Пример отчета из рабочего письма.

Согласие на оформление нового продукта в прошлом месяце



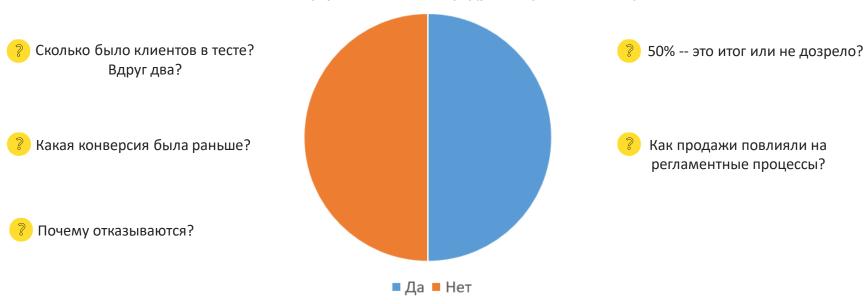
Хорошо это или плохо?

Зачем вам эта лекция?



Пример отчета из рабочего письма.

Согласие на оформление нового продукта в прошлом месяце



О чем лекция?





Визуализация: что это такое и зачем она нужна?



Как визуализировать данные + разбор неудачных примеров



Инструменты визуализации в Тинькофф



Что такое метрики? Простейшие метрики продукта

Зачем нужна визуализация?



Центр сбора	Доход	Семейно е положен ие	Образова ние	Кол-во членов семьи	Количест во детей	пол	Первый платеж	Должнос ть	Номер договора	Заемщик _ID	Дата выдачи договора	Адрес рабочий	Адрес регистра ции	Адрес залога		Номер телефона основной	Номер телефона домашни й	Номер телефона рабочий	Номер телефона дополнит ельный	
Москва	30000	Женат	Среднее	2	0 N	И		Электром	624/0000-:	1-DO25K-7	17.11.2015	1	1	0	0	1	0	1	0	RUR
Москва	270000	Женат	Высшее	4	2 N	И	5187,92	коммерче	625/2200-0	1-63D5G-1	03.03.2014	1	1	0	0	1	0	0	0	RUR
Москва	270000	Женат	Высшее	4	2 N	И		коммерче	633/2703-0	1-63D5G-1	09.02.2012	1	1	0	0	1	0	0	0	RUR
Москва	40000	Женат	Среднее	3	1 X	K	0		625/0000-0	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	40000	Женат	Среднее	3	1 X	K			633/0203-0	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	30000	Не женат	Высшее	1	К 0	K	4062,7	Юрист	625/2000-0	1-PG-42	14.11.2005	1	1	0	0	1	1	0	0	RUR
Москва	87290	Не женат	Высшее	1	0 N	4		консульта	624/0000-0	1-9JVII-14	22.12.2009	1	1	0	1	0	1	1	1	RUR
Москва	400000	Женат	Дополните	4	0 N	4	15281,08	первый за	625/3703-0	1-64VSL-6	10.11.2014	0	1	0	0	1	1	0	0	RUR
Москва	120169	Женат	Высшее	4	2 N	Л		командир	624/0000-0	1-9SV6I-31	25.12.2007	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	Л	1229,41	командир	625/1203-0	1-9SV6I-31	28.12.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	Л		командир	633/0000-0	1-9SV6I-31	18.04.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	Л	3889,8	командир	625/1203-0	1-9SV6I-31	29.07.2014	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	Л	3486,11	командир	625/3500-0	1-9SV6I-31	22.12.2011	0	1	0	0	0	0	1	1	RUR
Москва	30450	Не женат	Высшее	1	0 Ж	K	3675,78	оператор	625/2200-0	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	30450	Не женат	Высшее	1	0 Ж	K	1018,35	оператор	625/2200-0	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	50000	Разведен			N	Л	0	Руководит	625/0000-0	1-DTTQH-	10.06.2015	0	1	0	0	1	1	1	1	RUR
Москва	48074	Не женат	Высшее	1	0 Ж	K		Менедже	633/2810-0	1-BKUP4-6	17.07.2012	1	1	0	0	1	1	0	0	RUR
Москва	48074	Не женат	Высшее	1	0 Ж	K		Менедже	624/0000-:	1-BKUP4-6	03.12.2012	1	1	0	0	1	1	0	0	RUR
Москва	87000	Не женат	Среднее с	1	0 N	И		менедже	633/2610-0	1-BYW4Y-€	12.05.2014	1	1	0	1	1	0	0	0	RUR
Калуга (Вс	24641,23	Вдовец(в,	Среднее с	1	0 Ж	K	0	Санитарка	625/0000-0	1-7FG0L-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Калуга (Вс	24641,23	Вдовец(в,	Среднее	1	0 Ж	K		Санитарка	633/1103-0	1-7FG0L-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Тамбов (В	850000		Высшее	2	0 N		0		625/0000-0		27.03.2012	0	0	0	0	1	1	1		RUR

Зачем нужна визуализация?



Центр сбора	Доход	Семейно е положен ие	Образова ние	Кол-во членов семьи	Количест во детей	пол	Первый платеж	Должнос ть	Номер договора	Заемщик _ID	Дата выдачи договора	Адрес рабочий	Адрес регистра ции	Адрес залога	Адрес фактичес кий	Номер телефона основной	Номер телефона домашни й	Номер телефона рабочий	Номер телефона дополнит ельный	
Москва	30000	Женат	Среднее с	2	0 N	1		Электром	624/0000-:	1-DO25K-7	17.11.2015	1	1	C	0	1	0	1	0	RUR
Москва	270000	Женат	Высшее	4	2 N	1	5187,92	коммерче	625/2200-0	1-63D5G-1	03.03.2014	1	1	0	0	1	0	0	0	RUR
Москва	270000	Женат	Высшее	4	2 N	1		коммерче	633/2703-0	1-63D5G-1	09.02.2012	1	1	0	0	1	0	0	0	RUR
Москва	40000	Женат	Среднее	3	1 X	<	0		625/0000-0	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	40000	Женат	Среднее	3	1 X	<			633/0203-0	1-AQMJ2-1	02.11.2012	1	1	0	1	1	1	0	0	RUR
Москва	30000	Не женат	Высшее	1	K 0	<	4062,7	Юрист	625/2000-0	1-PG-42	14.11.2005	1	1	0	0	1	1	0	0	RUR
Москва	87290	Не женат	Высшее	1	0 N	1		консульта	624/0000-0	1-9JVII-14	22.12.2009	1	1	0	1	0	1	1	1	RUR
Москва	400000	Женат	Дополните	4	0 N	1	15281,08	первый за	625/3703-0	1-64VSL-6	10.11.2014	0	1	0	0	1	1	0	0	RUR
Москва	120169	Женат	Высшее	4	2 N	1		командир	624/0000-0	1-9SV6I-31	25.12.2007	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	1	1229,41	командир	625/1203-0	1-9SV6I-31	28.12.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	1		командир	633/0000-0	1-9SV6I-31	18.04.2013	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	1	3889,8	командир	625/1203-0	1-9SV6I-31	29.07.2014	0	1	0	0	0	0	1	1	RUR
Москва	120169	Женат	Высшее	4	2 N	1	3486,11	командир	625/3500-0	1-9SV6I-31	22.12.2011	0	1	0	0	0	0	1	1	RUR
Москва	30450	Не женат	Высшее	1	K 0	<	3675,78	оператор	625/2200-0	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	30450	Не женат	Высшее	1	0 Ж	<	1018,35	оператор	625/2200-0	1-8F9RM-1	19.09.2012	1	1	0	1	1	0	1	0	RUR
Москва	50000	Разведен			N	1	0	Руководит	625/0000-0	1-DTTQH-	10.06.2015	0	1	0	0	1	1	1	1	RUR
Москва	48074	Не женат	Высшее	1	K 0	<		Менедже	633/2810-0	1-BKUP4-6	17.07.2012	1	1	0	0	1	1	0	0	RUR
Москва	48074	Не женат	Высшее	1	K 0	<		Менедже	624/0000-:	1-BKUP4-6	03.12.2012	1	1	0	0	1	1	0	0	RUR
Москва	87000	Не женат	Среднее с	1	0 N	1		МЕНЕДЖЕ	633/2610-0	1-BYW4Y-6	12.05.2014	1	1	0	1	1	0	0	0	RUR
Калуга (Вс	24641,23	Вдовец(в,	Среднее с	1	И 0	<	0	Санитарка	625/0000-0	1-7FG0L-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Калуга (Вс	24641,23	Вдовец(в,	Среднее с	1	0 Ж	<		Санитарка	633/1103-0	1-7FG0L-3	01.07.2010	1	1	0	1	1	1	1	0	RUR
Тамбов (В	850000		Высшее	2	0 N				625/0000-0		27.03.2012	0	0	C	0	1	1	1		RUR
		-	_	-									_	-			-	-		



И что?

Зачем нужна визуализация?





Анализ данных



Регулярная отчетность, мониторинг



Презентации





?



→ Покомпонетное сравнение (например, сравниваем доли клиентов из разных городов)



- → Покомпонетное сравнение (например, сравниваем доли клиентов из разных городов)
- Э Позиционное (показать, какой канал привлекает больше заявок)



- → Покомпонетное сравнение (например, сравниваем доли клиентов из разных городов)
- Э Позиционное (показать, какой канал привлекает больше заявок)
- Временное



- → Покомпонетное сравнение (например, сравниваем доли клиентов из разных городов)
- Э Позиционное (показать, какой канал привлекает больше заявок)
- → Временное
- Частотное (показать концентрацию, распределение)
- Корреляционное

Типы диаграмм



- → Круговая
- Э Линейчатая
- → Гистограмма
- → График
- → Точечная
- Другие (лепестковая, свечная, пузырьковая...)

Покомпонентное сравнение

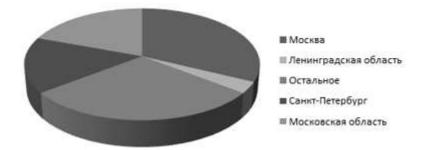


Общая задача: понять распределение по типам

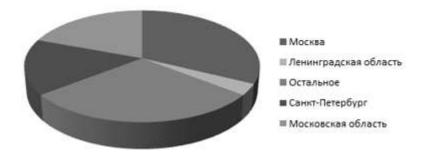
Примеры конкретных задач, которые можно решать с помощью круговых диаграмм

- ? Распределение клиентов по регионам
- ? Распределение пользователей приложения по марке телефона
- ? Структура доходов и расходов продукта
- 🥐 Доля клиентов, пользующихся конкретным видом товара



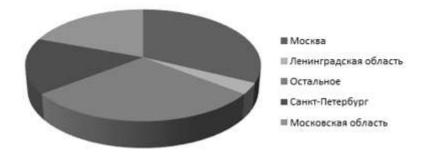






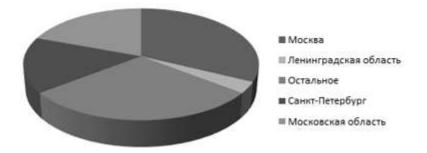
Рет заголовка





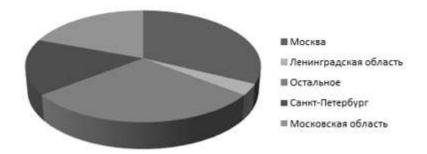
- Нет заголовка
- 🥐 Цвета плохо различимы





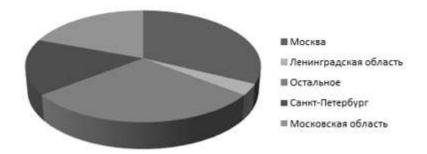
- Нет заголовка
- 🥐 Цвета плохо различимы
- Не отсортировано





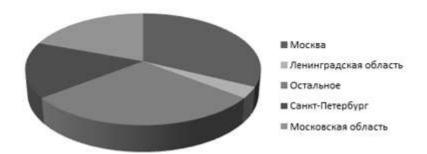
- Нет заголовка
- 🥐 Цвета плохо различимы
- Не отсортировано
- 🥐 Не указаны ни числа, ни доли





- 🥐 Нет заголовка
- 🥐 Цвета плохо различимы
- Не отсортировано
- 🥐 Не указаны ни числа, ни доли
- 🥐 Может, лучше плоскую?

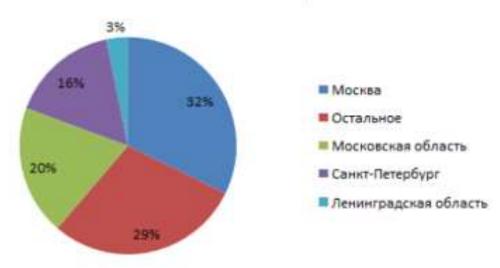




Нет заголовка

- 🥐 Цвета плохо различимы
- Не отсортировано
- Не указаны ни числа, ни доли
- 🥐 Может, лучше плоскую?

Разбивка клиентов по регионам



Позиционное сравнение



Общая задача: сравнить определенный показатель в разных группах

Позиционное сравнение



Общая задача: сравнить определенный показатель в разных группах

Примеры конкретных задач, которые можно решать с помощью линейчатых диаграмм

- ? Актуальность предложения в зависимости от региона
- ? Конверсия по каналам поступления заявки
- ? Проникновение доп. услуг в зависимости от типа продукта
- ? Средний доход заявителей по региону

Что используем: упорядоченные линейчатые диаграммы, гистограммы

Линейчатая диаграмма





 $0.00\%\ 10.00\%\ 20.00\%\ 30.00\%\ 40.00\%\ 50.00\%\ 60.00\%\ 70.00\%\ 80.00\%\ 90.00\%100.00\%$

Линейчатая диаграмма



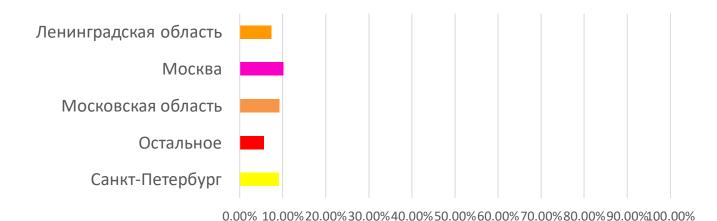


0.00% 10.00% 20.00% 30.00% 40.00% 50.00% 60.00% 70.00% 80.00% 90.00%100.00%

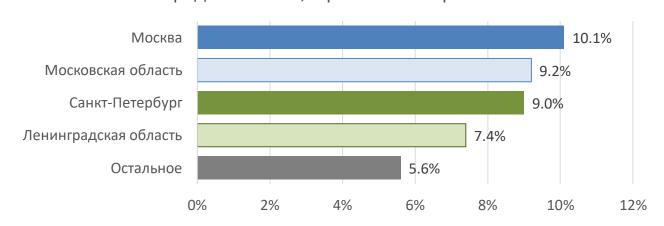
- Нет заголовка
- От выбора цветов вытекают глаза (отвлекается внимание)
- Не отсортировано
- 🥐 Нет подписей данных
- Масштаб не позволяет наглядно сравнить группы

Линейчатая диаграмма





Доля клиентов, воспользовавшихся предложением, в разбивке по регионам



Частотное сравнение



Общая задача: нарисовать распределение по некоторому числовому параметру

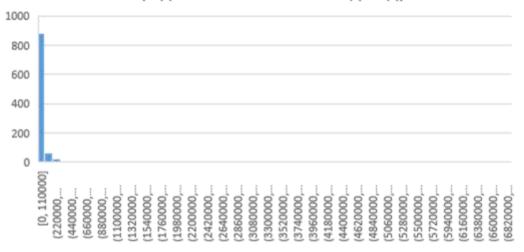
Примеры конкретных задач, которые можно решать с помощью частотных сравнений

- ? Распределение заявителей по доходу, возрасту
- ? Распределение клиентов по тратам за месяц
- ? Распределение клиентов по количеству продуктов в банке
- ? Распределение по количеству входов в мобильное приложение
- ? Распределение слайдов по числу упоминаний слова «распределение»

Гистограмма



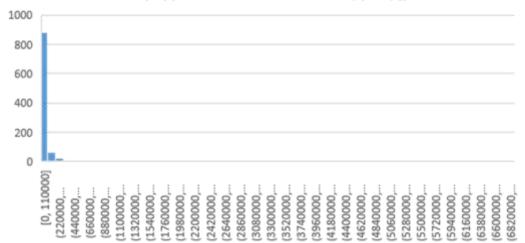
Распределение заявителей по доходу



Гистограмма



Распределение заявителей по доходу

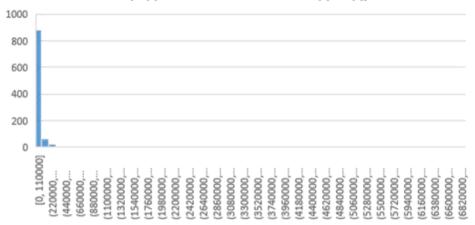


- Не ясна структура
- Большое число столбиков не информативны
- 🥐 Плохой выбор интервалов
- 🥐 Многоточия? Загадочно..

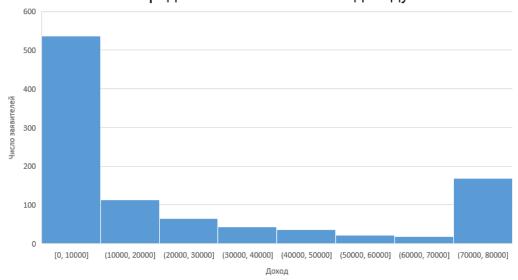
Гистограмма



Распределение заявителей по доходу



Распределение заявителей по доходу



Временное сравнение



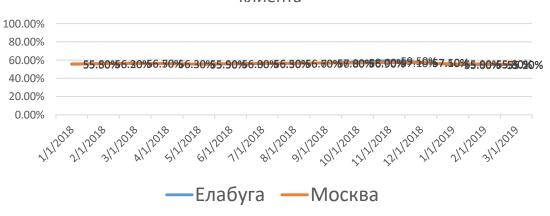
Общая задача: посмотреть динамику показателей

Примеры конкретных задач, которые можно решать с помощью временного сравнения

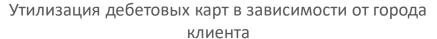
- 🥐 Конверсии в каналах
- ? Число клиентских обращений
- ? Доля клиентов, утилизирующих продукт
- 🥐 Доля клиентов, использующих бонусную программу
- ? График числа участников зум-конференции по времени



Утилизация дебетовых карт в зависимости от города клиента



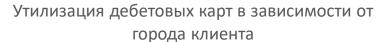






Мои глаза! Подписи явно лишние.



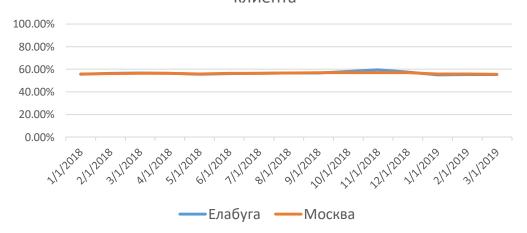




- 🥐 Мои глаза! Подписи явно лишние.
- Не видно различий между линиями из-за масштаба
- 🥐 Не видно временных эффектов



Утилизация кредитных карт зависимости от города клиента



Утилизация кредитных карт зависимости от города клиента



Корреляционное сравнение



Общая задача: найти зависимость между параметрами

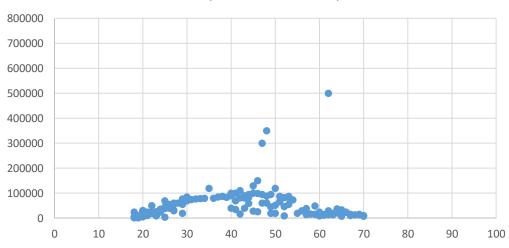
Примеры конкретных задач, которые можно решать с помощью корреляционного сравнения

- ? Связь между числом жителей города и конверсией
- ? Связь между уровнем дохода и утилизацией
- ? Связь между возрастом и уровнем дохода
- Пояльность к бренду: число продуктов и «стаж» клиента

Точечная диаграмма



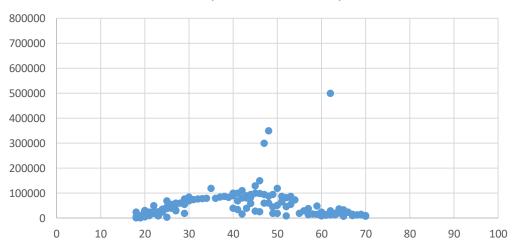
зависимость между доходом и возрастом клиента



Точечная диаграмма





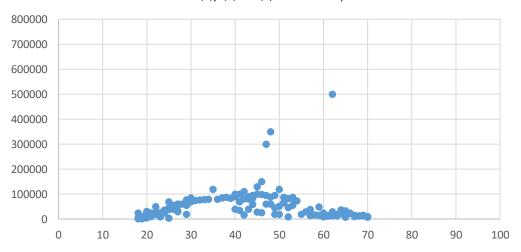


- Не видно тренда
- 🥐 Мешают выбросы
- 🥐 Неудачный масштаб
- 🥐 Не подписаны оси

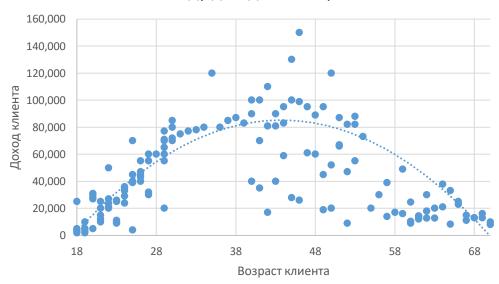
Точечная диаграмма



зависимость между доходом и возрастом клиента



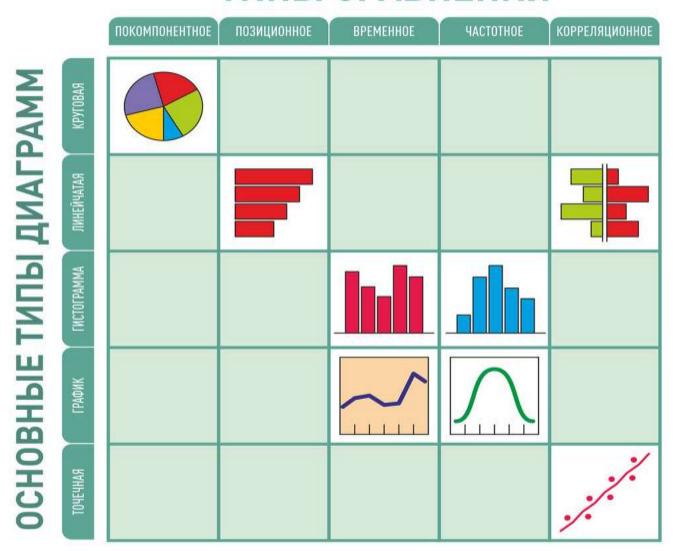
зависимость между доходом и возрастом клиента



Типы диаграмм и типы сравнений



ТИПЫ СРАВНЕНИЯ



Инструменты визуализации



- → Excel
- → Python
- → SAS EG
- → SAP BO aka BI
- → Zeppelin
- → Tableau

Excel: плюсы и минусы



🖒 Быстро и легко строятся графики

- Тяжело работает с большими таблицами
- <u></u> Широкий простор для форматирования
- Нет автоматизации

- 🔥 Вычисление по формулам без кода
- Нет подключения к базам данных

ტ Дружит с Outlook и PowerPoint

Нет наглядности: по каким правилам собраны данные

Python: плюсы и минусы



- **А**втоматизация
- Большое количество библиотек для визуализации
- 🖒 Есть подключение к базе
- Удобно обмениваться кодом
- Высокая скорость обработки

- Графики сложнее настроить
- Нет простой интеграции с продуктами Office
- Нельзя дать менеджеру данные «покрутить»

SAS EG



Автоматизация

- Корпоративный платный инструмент
- Подходит для автоматических рассылок отчетов
- 📗 Субъективно: не очень удобно

🖒 Есть подключение к базе

Нельзя дать менеджеру данные «покрутить»

👍 Удобно обмениваться кодом

Tableau: плюсы и минусы



- 🖒 Отчеты интерактивные
- Подходит для управленческой отчетности
- 🖒 Есть подключение к базе
- <u></u> Интуитивно понятный интерфейс

- Инструмент платный
- Если данных много, возможны зависания

Пример, который будем рассматривать



Карта для сельхоз рабочих



В чем продукт?

- Дебетовая карта
- → Вклад от 50 000 рублей



- → На покупки в магазине «Крестьянин» повышенный кешбек
- 🔶 Онлайн-карта миграции колорадских жуков в приложении

Пример, который будем рассматривать



Процесс



Клиент узнаёт о нас из интернета, объявлений или от партнера – магазина «Крестьянин»



Заполняет заявку или в интернете, или по телефону, или через партнера



Банк связывается с клиентом для назначения встречи



На встрече представитель банка и клиент подписывают договор, клиент получает карту. Клиенту становятся доступны мобильное приложение и интернет-банк.

Пример, который будем рассматривать



В чем выгода для банка?



Годовое обслуживание платно



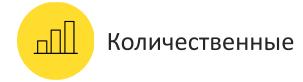
Вклады



Магазин «Крестьянин» платит банку процент с каждой покупки

Простые метрики







Количественные метрики



Абсолютные величины: количество, сумма, среднее, медиана

Примеры из реальной жизни



Количество

- заявок на продукт
- открытых счетов
- клиентов
- 3BOHKOB
- обращений



Среднее значение, медиана

- средняя сумма одной покупки
- сумма транзакций клиента в месяц
- время разговора
- время встречи



Сумма в деньгах

- портфель продукта
- доходность
- затраты











заявки на продукт









заявки на продукт



Заметили проблему, раскрыли дело, если нужно – приняли меры.



Возвращаемся к примеру с сельхоз картой на широком горизонте





Возвращаемся к примеру с сельхоз картой на широком горизонте





Возвращаемся к примеру с сельхоз картой на широком горизонте



Осознали тенденцию, учли это в стратегии развития.



Возвращаемся к примеру с сельхоз картой на широком горизонте



Находим перспективы развития продукта.

Относительные метрики



Всевозможные доли

Примеры из реальной жизни



- доля мужчин среди заявителей
- доля автовладельцев среди клиентов
- конверсия из обращения в заявку
- конверсия из заявки в утилизицацию



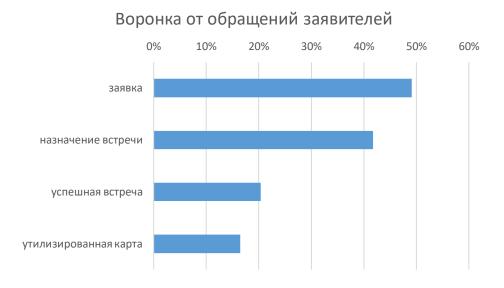
Воронка продаж – путь клиента до целевого действия. Состоит из нескольких стадий.

Конверсия – доля клиентов, успешно прошедших какойто этап воронки.

Зачем нужны относительные метрики



На нашем примере карты для конкретного села Кукуево



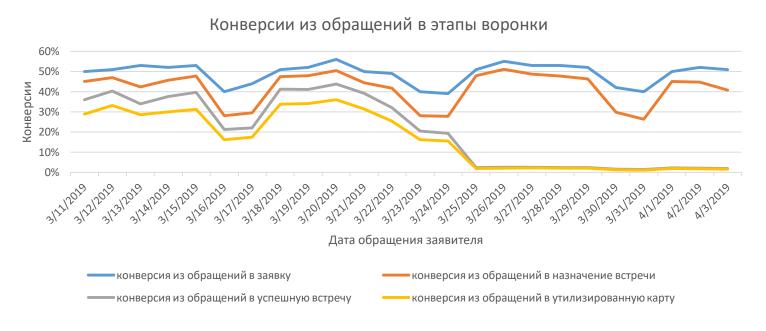
- 🖒 Понятен в общих чертах процесс
- 🖒 Легко сравнить, наглядно

- Не видна динамика: меняется ли поведение заявителей?
- Нельзя мониторить, не сломалось ли что в процессе.

Зачем нужны относительные метрики



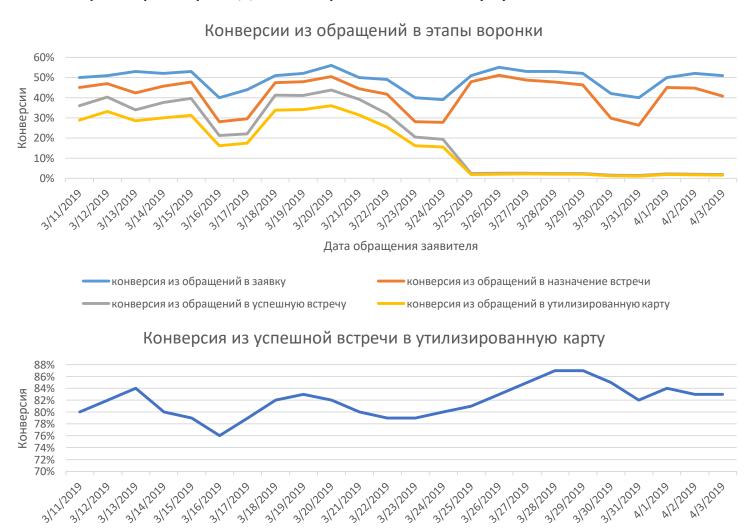
На нашем примере карты для конкретного села Кукуево



Зачем нужны относительные метрики



На нашем примере карты для конкретного села Кукуево



Операционные метрики

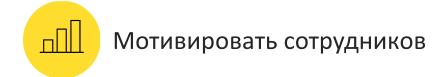


Операционные метрики – метрики, связанные с работой сотрудников

Операционные метрики



Цели операционных метрик

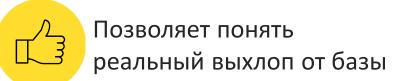








Базовая конверсия – доля успехов от базы





Из-за качества базы можем выставить неверную мотивацию для сотрудников



Базовая конверсия – доля успехов от базы





Из-за качества базы можем выставить неверную мотивацию для сотрудников

Пример – дозаполнение заявок на нашу сельхоз карту, поступивших от магазина «Крестьянин»



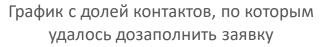
Базовая конверсия – доля успехов от базы





Из-за качества базы можем выставить неверную мотивацию для сотрудников

Пример – дозаполнение заявок на нашу сельхоз карту, поступивших от магазина «Крестьянин»







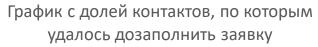
Базовая конверсия – доля успехов от базы





Из-за качества базы можем выставить неверную мотивацию для сотрудников

Пример – дозаполнение заявок на нашу сельхоз карту, поступивших от магазина «Крестьянин»





Операторы стали работать хуже?



Пример – дозаполнение заявок на нашу сельхоз карту, поступивших от магазина «Крестьянин»





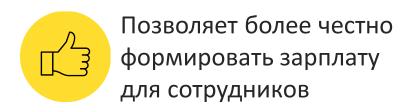
Пример – дозаполнение заявок на нашу сельхоз карту, поступивших от магазина «Крестьянин»



По базовой конверсии не можем понять, стали ли операторы хуже разговаривать. Но точно видим, что их работа стала сложнее, и успешных звонков у них меньше.











Продажная конверсия – успех / (успех + отказ)



Позволяет более честно формировать зарплату для сотрудников



Уязвима к фроду

Пример – дозаполнение заявок на нашу сельхоз карту, поступивших от магазина «Крестьянин»







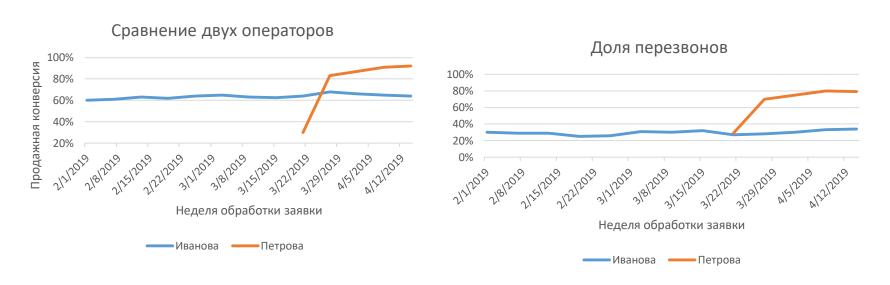






Операционные метрики. Антифродовые





Операционные метрики. Антифродовые









Операционные метрики. Антифродовые





Доля перезвонов, недозвонов, успехов



Доля претензий после перезвона, недозвона, успеха



Доля одобренных банком

Операционные метрики. Конверсия Максимова





Операционные метрики. Конверсия Максимова





 \rightarrow

Идея: конверсия относительно среднего



Позволяет сравнивать операторов между собой независимо от базы

Операционные метрики. Конверсия Максимова





У Идея: конверсия относительно среднего

Позволяет сравнивать операторов между собой независимо от базы

Зачем это?

- Мотивация операторов
- Приоритизация обработки выделенных клиентов

Операционные метрики



Основная идея – смотреть на совокупность метрик!

Резюме



На что стоит обратить внимание при визуализации данных

- Выбрать подходящий тип графика
- → Выбрать подходящий инструмент
- → Убрать все лишнее (ненужные данные, выбросы)
- Выбрать масштаб
- → Смотреть на совокупность метрик (разных по сути, разных по типу)
- Смотреть на динамику