



РОРО эффект:

Измеряем эффект рекламы для офлайн продаж

Серия вебинаров МТТ TALKS  
V КОНФЕРЕНЦИЯ - "АВТОМАТИЗАЦИЯ МАРКЕТИНГА"  
4 сентября 2020, Москва

<https://event.marpeople.com/kam>

# Предисловие

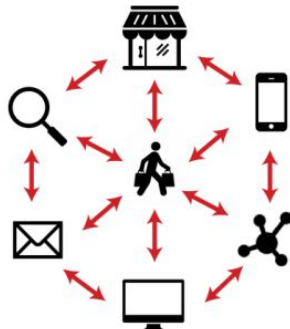
Современная аналитика продаж формирует запрос на решения:

**Универсальные** - применимы в любых категориях



88%\* потребителей используют веб перед походом в мебельный центр или гипермаркет

**Адекватные задаче** - используют коннекторы, связи и модели ML



Потребители давно привыкли жить кроссплатформенно и омниканально

**Понятная ценность** - экономит значительные ресурсы компании

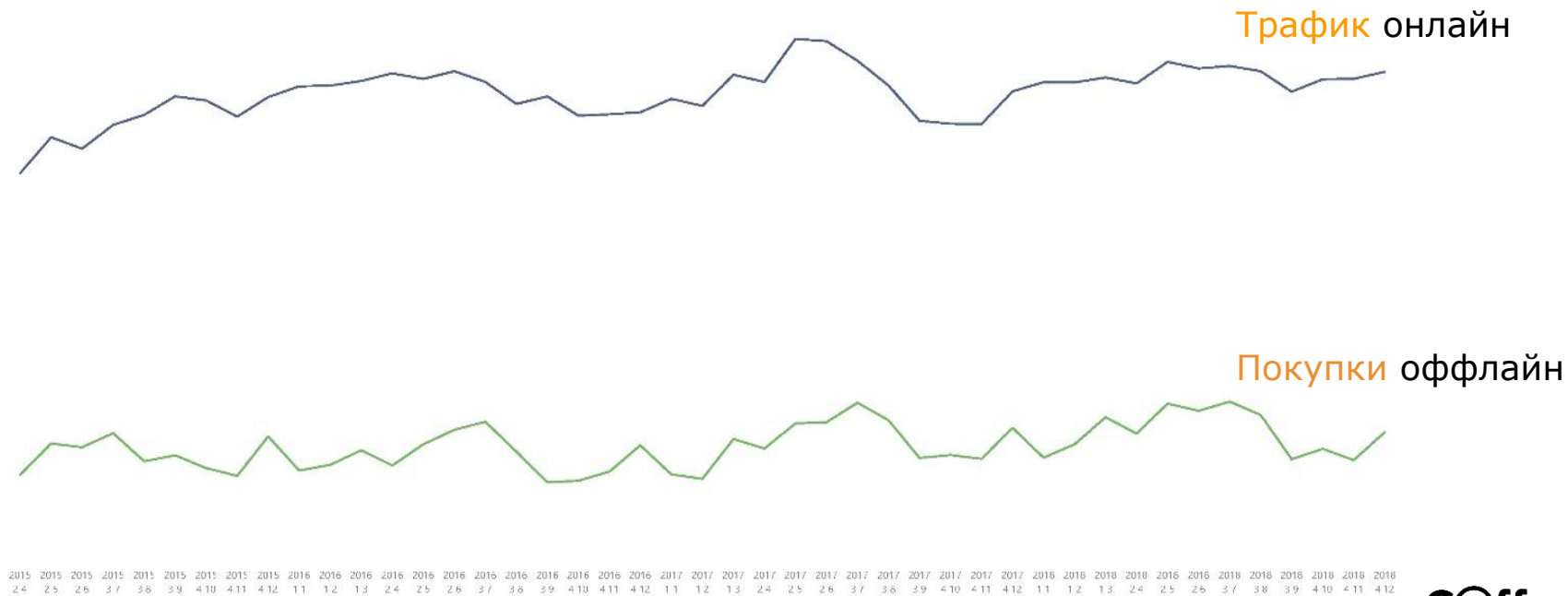


Широкое кол-во факторов в продвижении усложняют атрибуцию и снижают точность

# Задача: посчитать доход от рекламы

Как определить, есть ли связь рекламы и офф-лайн продаж?

● Покупки ● Трафик



# Задача: посчитать доход от рекламы

Как определить, есть ли связь рекламы и офф-лайн продаж?

● Покупки ● Трафик

Трафик онлайн

Как связаны  
временные  
ряды?

Корреляция XXX

Покупки оффлайн

Почему корреляция ещё не обозначает причинно-следственную связь

2016 2018 2016 2018 2016 2018 2016  
26 37 38 39 410 411 412

# ЧТО ТАКОЕ ROPO

Сегодня подход актуален как никогда

**Метод аналитики ROPO** (Research Online, Purchase Offline — поиск в онлайн, приобретение в оффлайне) определяет, как связана реклама в Интернете с ростом продаж в традиционных оффлайн магазинах и рознице.

## Универсальный



88% потребителей используют веб перед походом в мебельный центр или гипермаркет

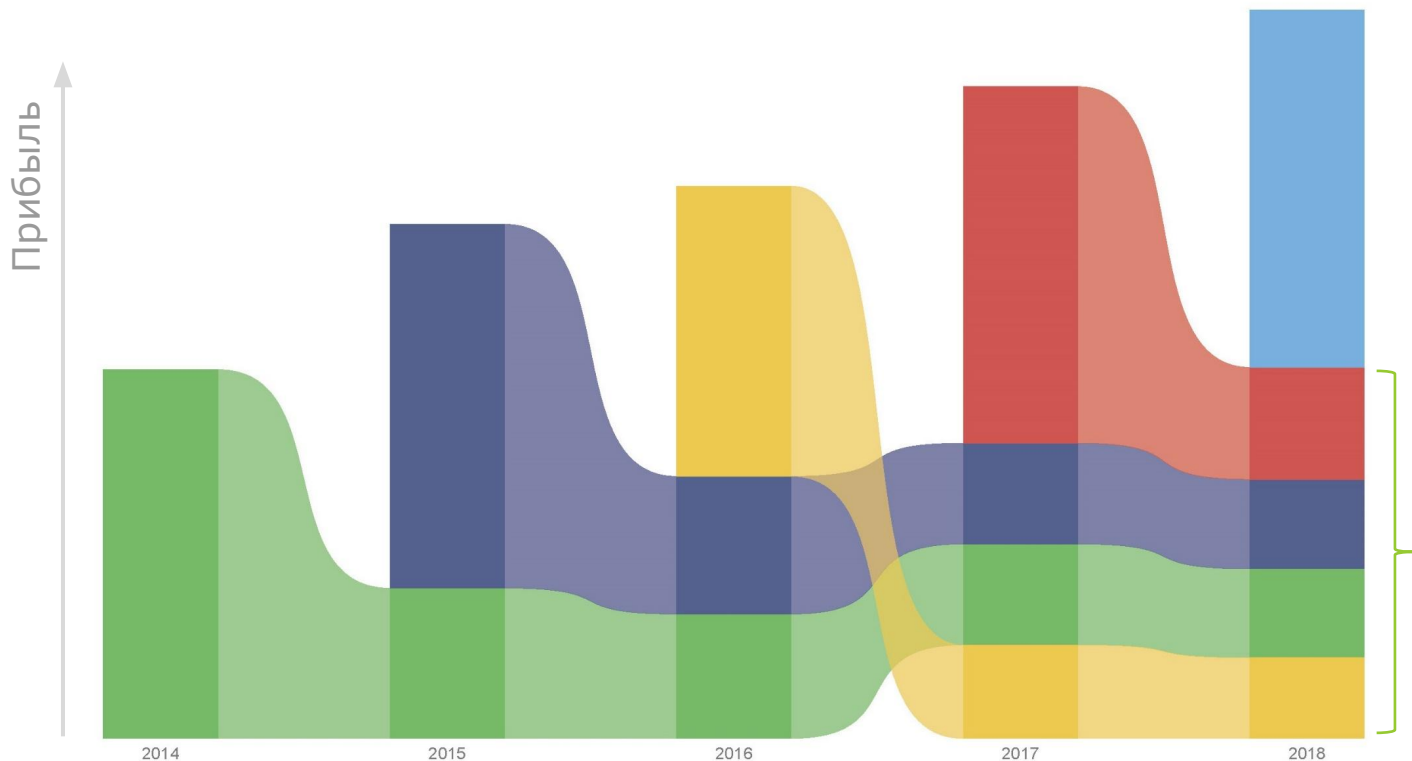
**Карантинный режим высвободил свободное время.** Часть этого времени люди тратят на поиск товаров в Интернете, но покупок при этом не совершают.

**В силу экономии, большее число покупателей** ищут информацию о товарах длительного пользования в Интернете перед тем, как купить его в традиционном магазине.

# ROPO для магазина нижнего белья

Инсайт аналитики продаж

Когорта ● 2014 ● 2015 ● 2016 ● 2017 ● 2018

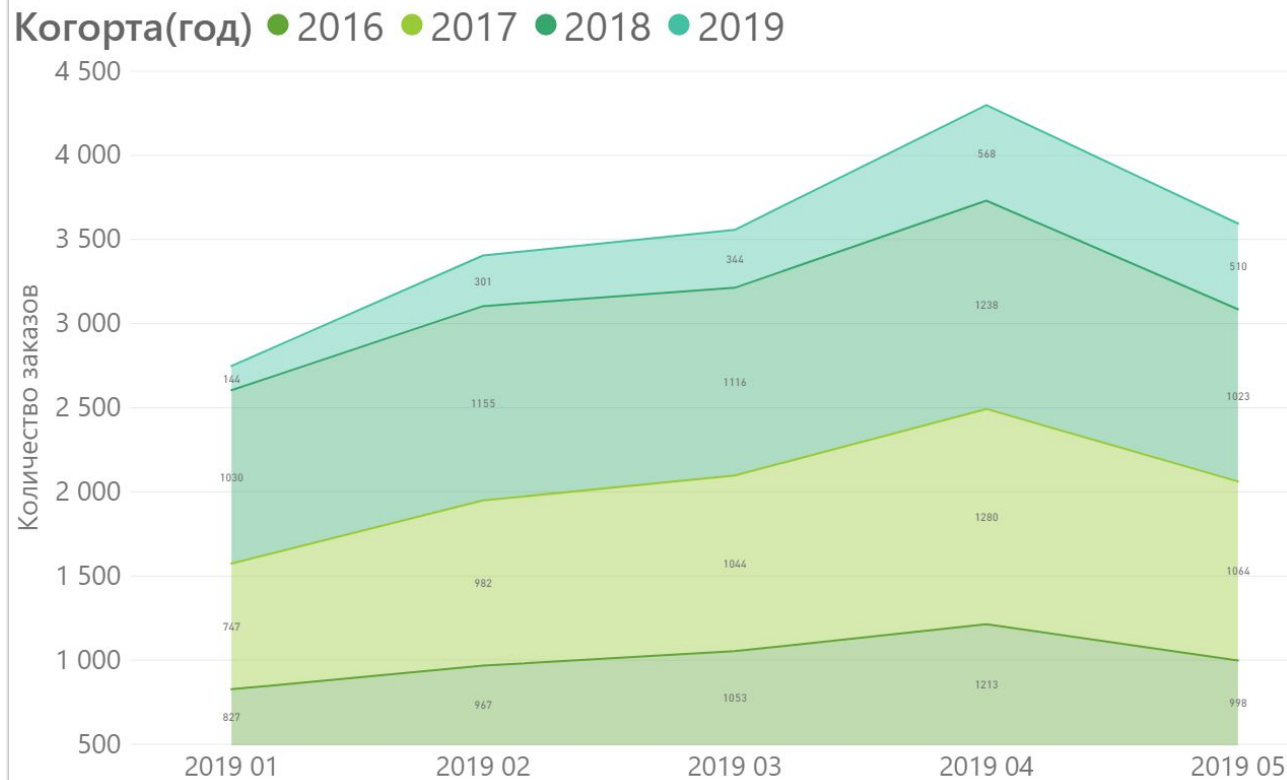


50  
%

Прибыли  
приносят  
когорты  
прошлых  
лет

# ROPO анализ для сети кафе

Доминируют заказы клиентов прошлых лет



Лишь незначительная часть клиентов пришла в этом году

# 3 способа отследить онлайн в офлайн

Которые можно комбинировать



**Коннектор**

Промокод

Телефон/Email

Call-tracking



**Спросить**

Анкетирование

Опросы(после покупки)



**Посчитать**

Сравнение(до/после)

A/B тест

Оценка влияния (корреляция)

- **Коннектор** – это любое (уникальное) значение, которое существует в привязке к клиенту он-лайн (в GA или ЯМ) и одновременно в привязке к его покупке офф-лайн (в 1С, на платформе коллтрекинга, в записях продавца)
- **Если коннектора нет**, проводить опросы «где вы о нас узнали» до или после покупки
- **Если пункты выше недоступны** или имеют большую погрешность, можно рассчитать несколькими способами, как изменение в рекламе повлияло на изменение продаж

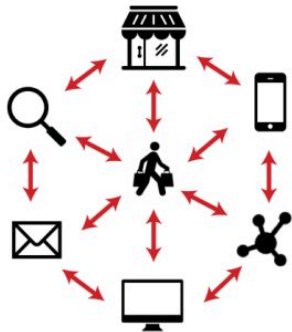


# Как связать данные

Что влияет на величину ROPO-коэффициента

**Полная картинка с доступными данными** позволит построить точную картину и выявить факторы влияния на оффлайн продажи

## Адекватный



Потребители давно привыкли жить кроссплатформенно и омниканально

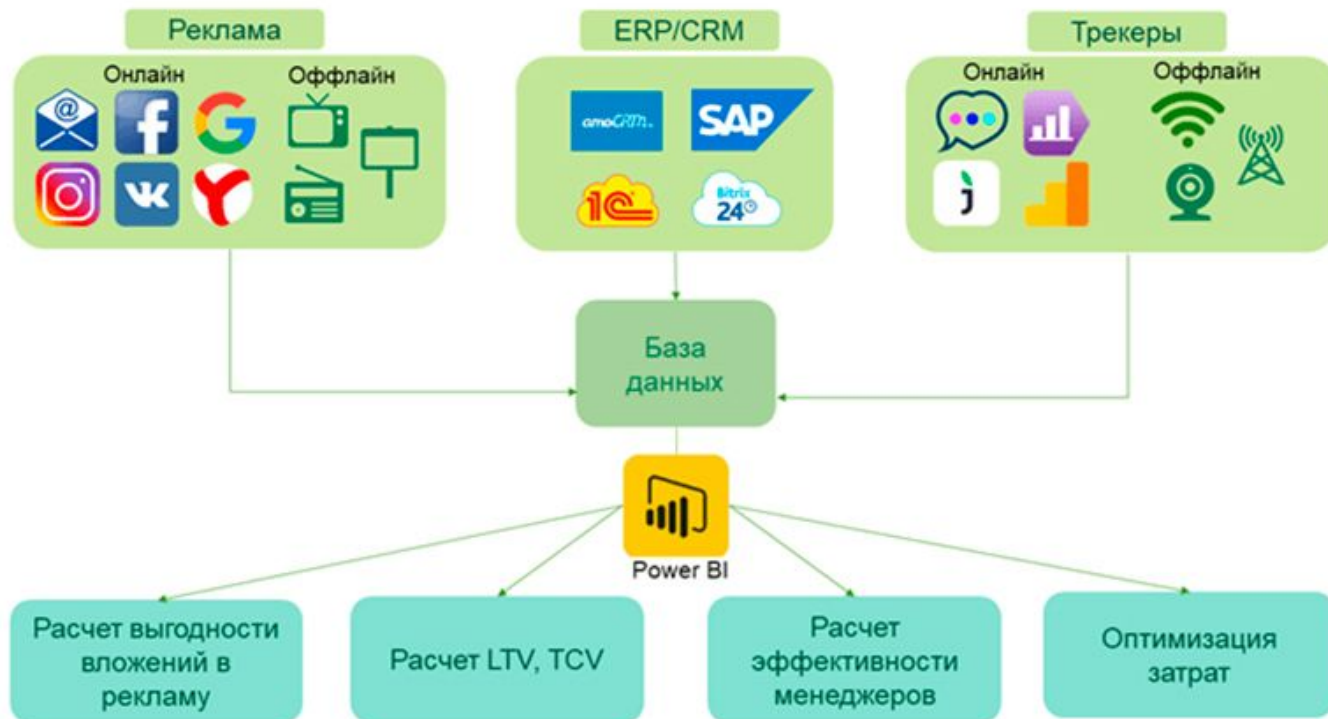
**Присутствие компании на картах.** Большинство пользователей ищут товары и услуги максимально близко к месту своего нахождения.

**Отзывы других покупателей** — текстовые или видео — помогают узнать больше о конкретном продавце, особенностях взаимодействия компании с покупателями.

**Каталог товаров и общий контент сайта:** максимально подробную, исчерпывающую информацию о товаре нужно «выдавать» из первых рук.

# Интеграция данных

Собрать и интегрировать все доступные данные



Данные из системы аналитики сайта:

userId	источник	Код скидки
1448653	Yandex/Direct	Gv2j4b3k4jyb
5462783	Google/Adwords	agsvja7739

Данные из системы продаж:

Код скидки	Номер транзакции	Сумма продажи
Gv2j4b3k4jyb	173671498	14 200
agsvja7739	153746891	31 500



Уникальный анонимный идентификатор пользователя на сайте

- Настроить корректный сбор userId
- Добавить на сайт уникальные коды скидки или форму заполнения для клиента
- Передавать код скидки или контактные данные клиента в систему web-аналитики сайта
- Сбирать коды скидки или контактные данные клиента при продаже в 1С или в таблицах

# Провести анкетирование и опросы

Как настроить:

Данные из таблички опроса клиентов

Дата и время	источник	Номер транзакции	Клиент (телефон)
09.07.2020 12:00	Яндекс	173671498	+7 919 367 32 98
09.07.2020 13:30	От знакомых	153746891	+7 920 322 32 12

Данные из системы продаж

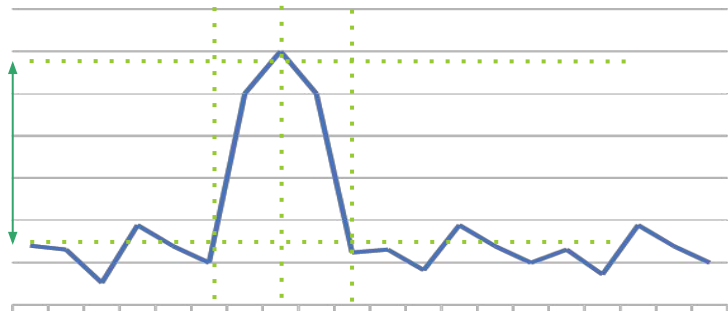
Дата и время	Номер транзакции	Сумма продажи
09.07.2020 12:00	173671498	14 200
09.07.2020 13:30	153746891	31 500



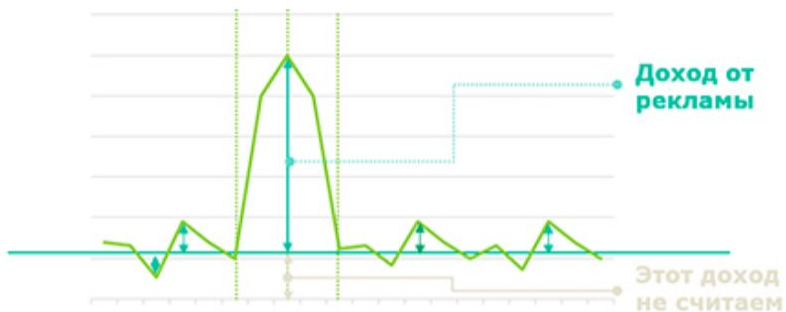
- Вести базу клиентов (номера телефонов/почта)
- Опрашивать клиента до покупки в магазине или после покупки по телефону
- Записывать время покупки, «откуда вы о нас узнали», номер транзакции
- Выгрузить инфо из 1С по номеру транзакции

# Посчитать корректно

Как настроить тест:



Определить величину всплеска



Прогноз: определение базовой линии



- Выделить сегменты для анализа (гео, товарная категория, новые покупатели, временной период...)
- Отследить изменение трафика или создать его искусственно (можно в одном сегменте)
- Выявить все факторы, которые могли повлиять в это время на продажи (реклама, конкуренты, кризис..)
- Определить величину влияния всплеска трафика на продажи с учетом других факторов

**А теперь подробно**

**Как будем считать, если:**



**ЕСТЬ СВЯЗЬ**



**Нет связи**



# Есть связь

Решение в Excel:

Есть связь

Formula bar: `=ВПР(C2;Продажи[#Все];3;0)`

А	В	С	Д	Е	Ф	Г	Н
userId	источник	Код скидки	продажи		Код скидки	Номер транзакции	Сумма продажи
1448653	Yandex/Direct	Gv2j4b3k4jyb			Gv2j4b3k4jyb	173671498	14 200
5462783	Google/Adwords	agsvja7739			agsvja7739	153746891	31 500

Слева табличка с источниками рекламы, Id пользователей и кодом скидки. Справа табличка данных продаж, куда также вносили код скидки

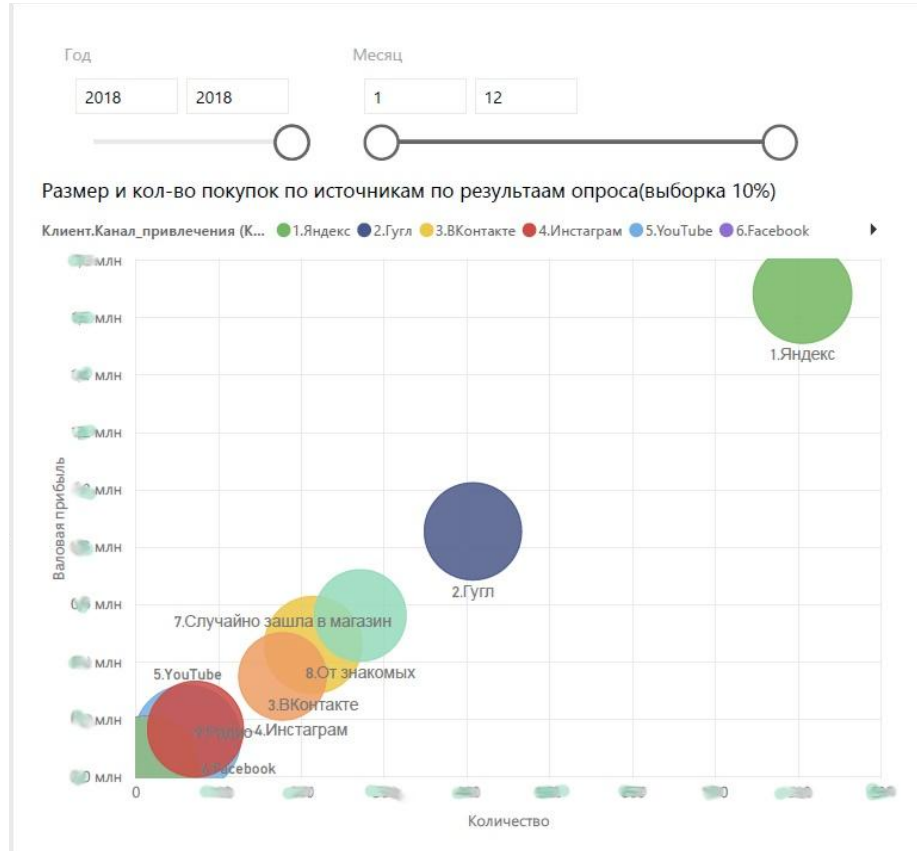
Соединяем данные продаж и рекламы по нашему ключу, считаем, сколько продаж принесли пользователи, пришедшие по рекламе



# Есть связь

Анкетирование: пример результата расчетов

Есть связь



**Опрос.** Для создания опроса стоит убедиться, что получится опросить достаточно большую выборку. Для этого можно использовать [он-лайн калькулятор](#). Или [pwr.t.test\(\)](#) в R для более точного определения достоверности каждого полученного результата.

**Результат.** В правом верхнем углу этого графика мы видим канал рекламы, принесший наибольшее кол-во покупок и дохода. Размер кружка означает доход на 1 клиента



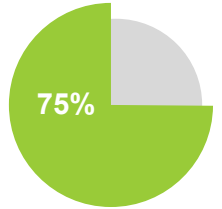


# Есть связь

Что необходимо учесть:

**Для какой части от всех продаж** за анализируемый период собраны данные об источниках:

- Можем ли мы утверждать, что оставшиеся 25% продаж точно не имеют отношения к рекламе? Ответ: нет, не можем\*
- Можем ли мы экстраполировать результаты, полученные на 75% данных на остальные 25%? Ответ: нет, не можем\*



С помощью коннекторов мы можем рассчитать, сколько мы получили дохода от пользователей, пришедших по рекламе для 75% продаж.

Для оставшихся 25% продаж мы считаем влияние рекламы так же, как и для продаж без связи

\* Без дополнительных исследований. Доп. Исследования см на слайде №25



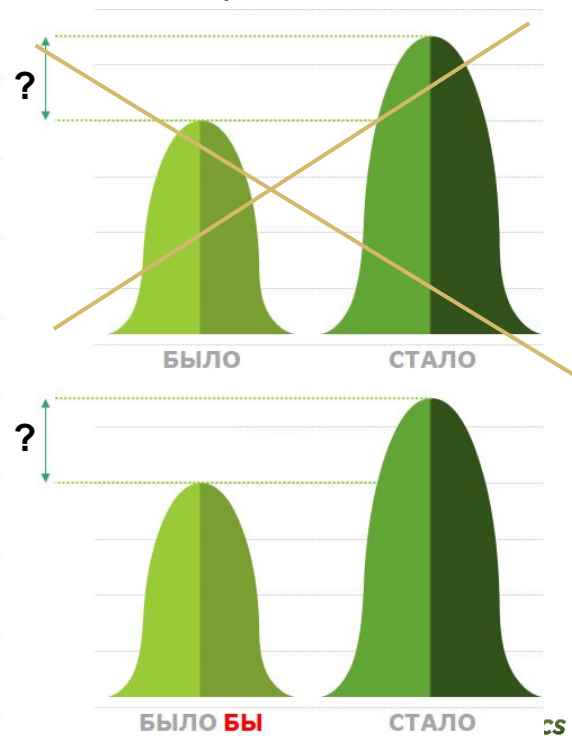
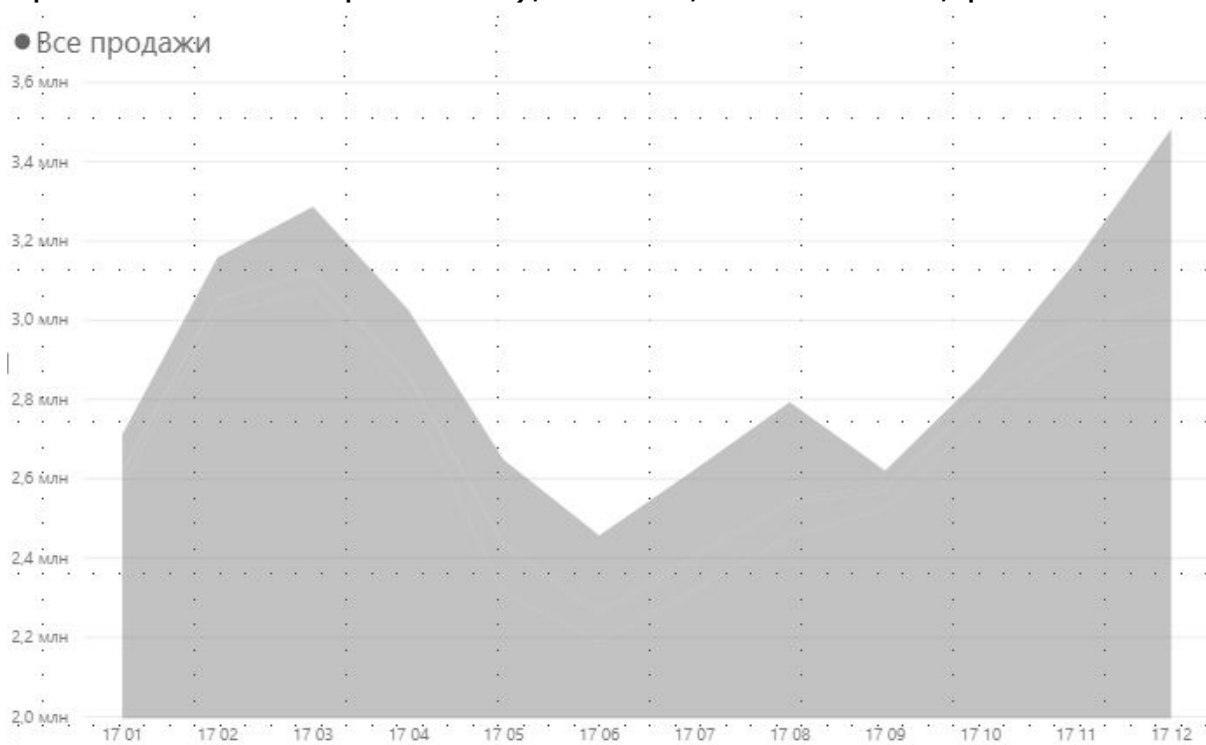
# Нет связи

Начинаем с аналитики продаж:

**Базовый уровень - прогноз продаж без рекламы.**

Если реклама не принесла дополнительного дохода (или отключение рекламы не принесло его сокращение), значит, мы считаем, реклама не влияет на продажи.

Нет связи





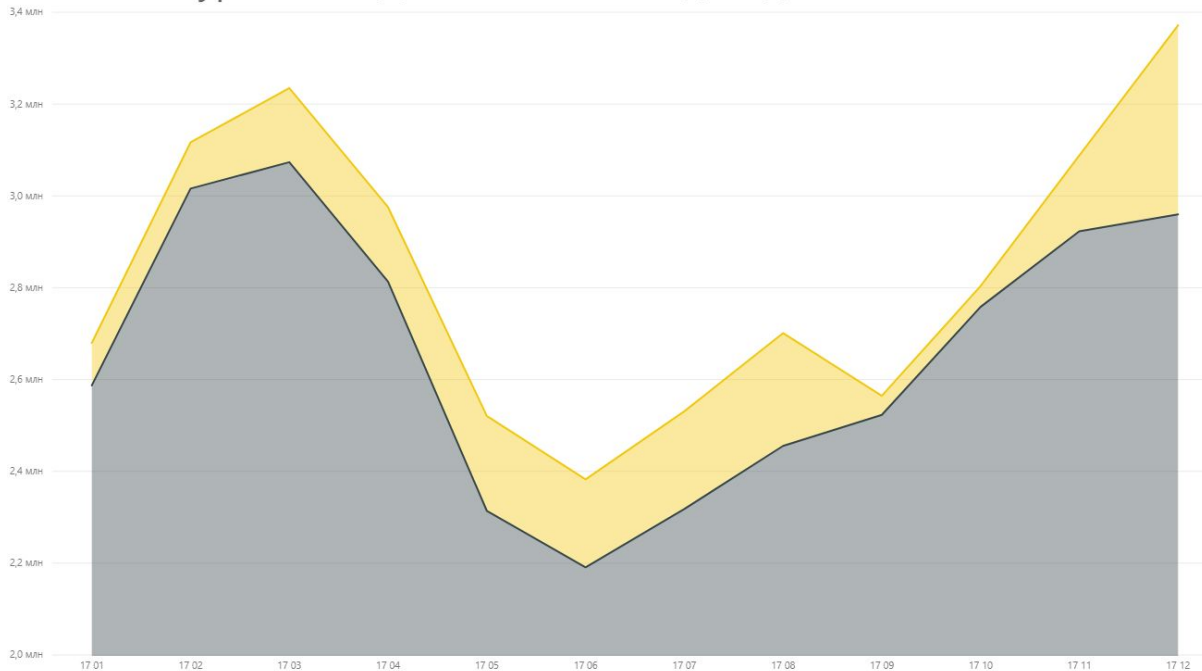
# Нет связи

Начинаем с аналитики продаж:

## Базовый уровень - прогноз продаж без рекламы.

Если реклама не принесла дополнительного дохода (или отключение рекламы не принесло его сокращение), значит, мы считаем, реклама не влияет на продажи.

● Базовый уровень ● Дополнительный доход



**Базовый уровень** на графике выделен серым. Его значение мы и рассчитываем.

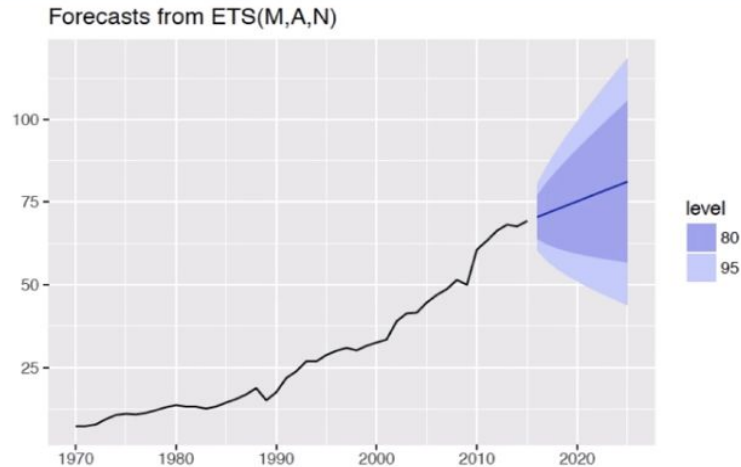
Базовый уровень считается как:

- Прогноз по историческим данным
- Прогноз по похожему сегменту

# Метод прогнозирования на основе прошлого

Как строить прогноз:

```
> ausair %>% ets() %>% forecast() %>% autoplot()
```



**Вариант 1.** Делаем прогноз на основе исторических данных.

Для более точного прогноза используют методы моделирования ARIMA или ETS.

При прогнозе учитывается среднее значение, тренд, изменения к аналогичному периоду в прошлом со свежими данными, учитывается сезонность, вероятность случайных изменений.

В случае знакомства с основами R и прогнозирования можно попробовать подбирать эти модели самостоятельно или с помощью автоматических функций `auto.arima()` и `ets()`.

# ARIMA или ETS

## Выбор модели



**ARIMA** (Auto Regressive Integrated Moving Average) хорошо использовать когда:

- данные стационарны(стабильны)
- подходит авторегрессия(прошлое хорошо объясняет будущее)

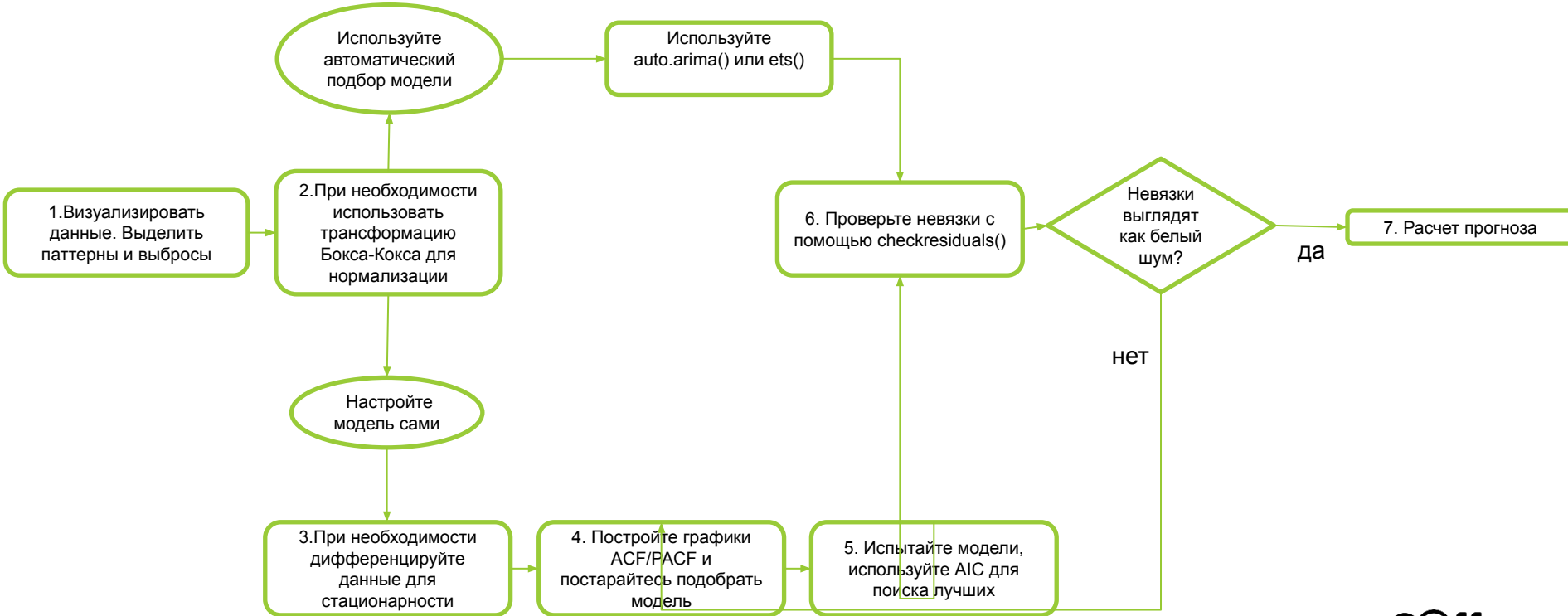
**ETS** (Error Trend and Seasonality, or exponential smoothing)

- подходит экспоненциальное сглаживание (последние события более важны, чем предыдущие)
- выраженный тренд и сезонность

Обязательно попробуйте обе методологии и проверьте, какая лучше подходит в вашем случае

# Подбор и проверка модели

## Алгоритм



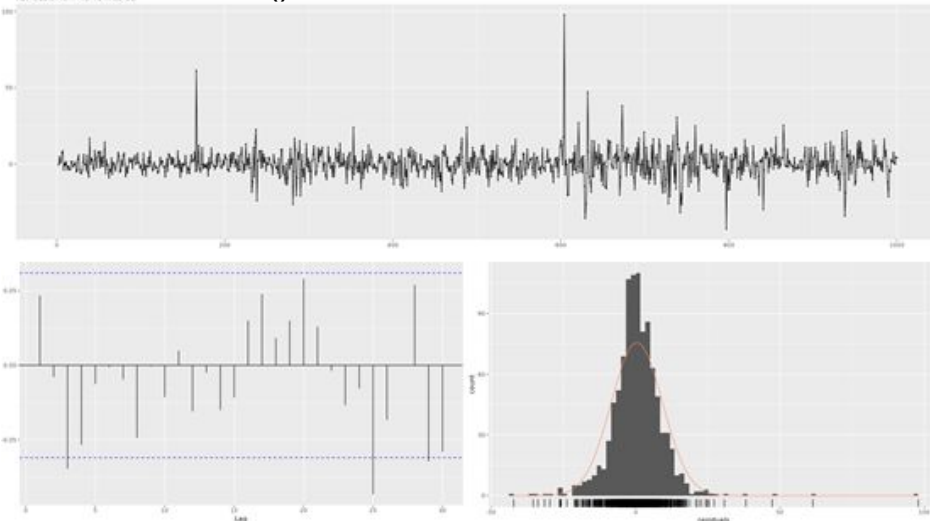
# Проверка модели

Замеряем невязки (Residuals)

Тренировочные данные

Прогнозные данные

```
library(forecast)
checkresiduals()
```



**Невязка = Прогноз - Факт**

Невязка, в отличие от ошибки определяется для тренировочного датасета. Ошибка же вычисляется только тестовых данных, которые не участвовали в создании модели

**Невязка должна выглядеть, как Белый шум:**

- Не должны иметь корреляции
- Среднее значение = 0
- Должна иметь нормальное распределение
- Должны иметь постоянную дисперсию случайной величины (разброс значений)

# Проверка модели

Замеряем ошибки

Тренировочные данные

Тестовые данные

```
> accuracy(fc, test)
      ME  RMSE  MAE  MPE  MAPE  MASE  ACF1  Theil's U
Training set  9.874 52.56 39.43 2.507 12.571 1.0000 0.1802      NA
Test set     21.602 35.10 29.98 3.964  5.778 0.7603 0.4030  1.185
```

## Ошибка = Тест - Прогноз

MAE - средняя ошибка (модуль)

MSE - средняя квадратичная ошибка

MAPE - средняя абсолютная ошибка в процентах

MASE - средняя абсолютная масштабированная ошибка

Не подходят, когда сравниваются слишком разные по размеру данные

Лучше, но если все данные позитивны и не имеют нулей

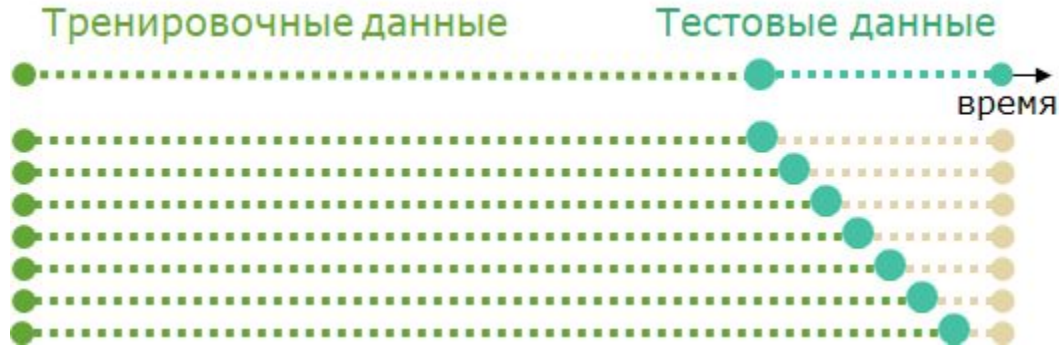
**Выбор**

$MASE = MSE/Q$  где  $Q$  - константа масштабирования



# Проверка модели

Кроссвалидация ошибки



## Проверка на несколько шагов вперед

С течением времени ошибка может увеличиваться. Для того, чтобы этого не происходило используют кросс валидацию

**tsCV** - считает ошибку прогноза для каждого периода на шаг вперед

```
> e <- tsCV (oil, forecastfunction = naive, h = 1)
> mean(e^2 , na.rm = TRUE)
[1] 2355.753
```

# Метод прогнозирования по похожему сегменту

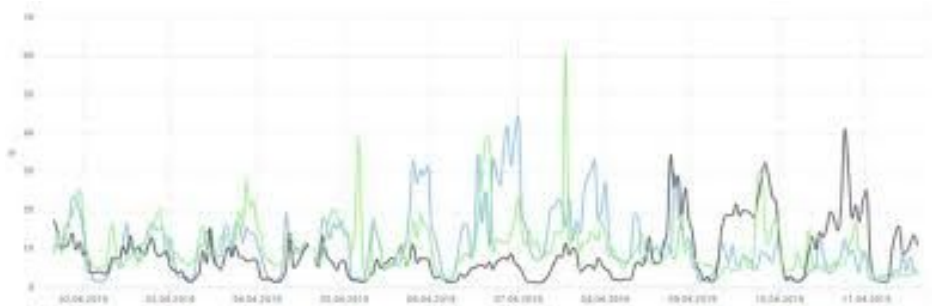
Как строить прогноз:

**Вариант 2.** Сравнить похожие сегменты при включенной и выключенной рекламе. Например: регионы. Так мы узнаем точнее, как поведут себя продажи, чем при прогнозе только на основе исторических данных.

**На «включенный» сегмент** будут действовать все факторы, которые есть в реальном времени (пример: режим самоизоляции).

Данные продаж по регионам:

```
> str(region_data[-1])
'data.frame':  12 obs. of  21 variables:
 $ month      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ адыгея     : int  1 0 1 0 2 2 2 5 1 2 ...
 $ алтайский  : int  18 7 7 10 6 12 28 20 13 15 ...
 $ астраханский : int  28 5 8 7 10 14 18 14 7 5 ...
 $ забайкальский : int  1 0 0 1 3 0 1 3 1 2 ...
 $ ингушетия  : int  3 0 2 0 1 1 0 0 0 0 ...
 $ иркутский  : int  47 10 6 20 15 22 18 40 24 25 ...
 $ кабардино  : int  3 4 4 3 0 13 9 13 3 2 ...
 $ .....
```



# Метод прогнозирования по похожему сегменту

Как строить прогноз:

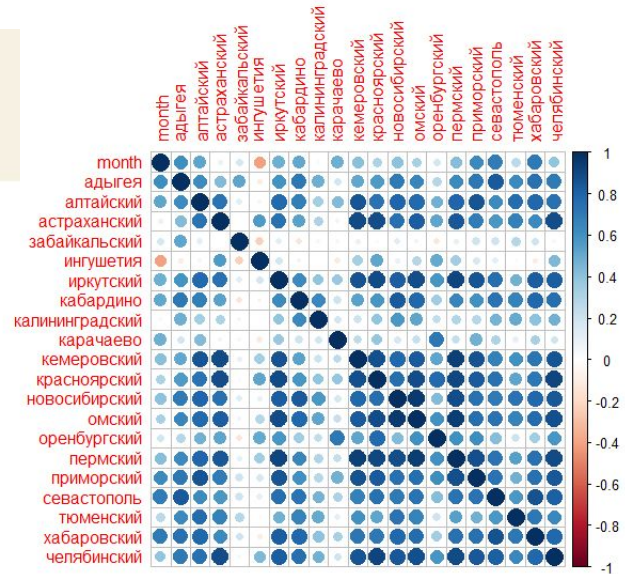
## Порядок действий:

1. Выбираем регионы, в которых графики продаж похожи. Для этого считаем их корреляцию.
2. Чем ближе к 1, тем больше похожи. Анализ в R группирует похожие регионы в кластеры (скопление **тёмных кругов**)

```
library(corrplot)
library(dplyr)

region_data %>% cor() %>% corrplot()
```

	тюменский	хабаровский	челябинский
адыгея	0.63387913	0.74379828	0.70387937
алтайский	0.76051504	0.78333047	0.74414220
астраханский	0.67212762	0.63706547	0.88871379
забайкальский	0.28269356	0.20245681	0.04631558
ингушетия	-0.04305662	-0.09950822	0.44317386
иркутский	0.47101459	0.82816766	0.82203534
кабардино	0.66295366	0.78216058	0.75396711
калининградский	0.50591184	0.40215507	0.47062558
карачаево	0.11185584	0.32313781	0.19681641
кемеровский	0.60611579	0.72757646	0.85721893
красноярский	0.52461437	0.71106966	0.91871515



# Метод прогнозирования по похожему сегменту

Как строить прогноз:

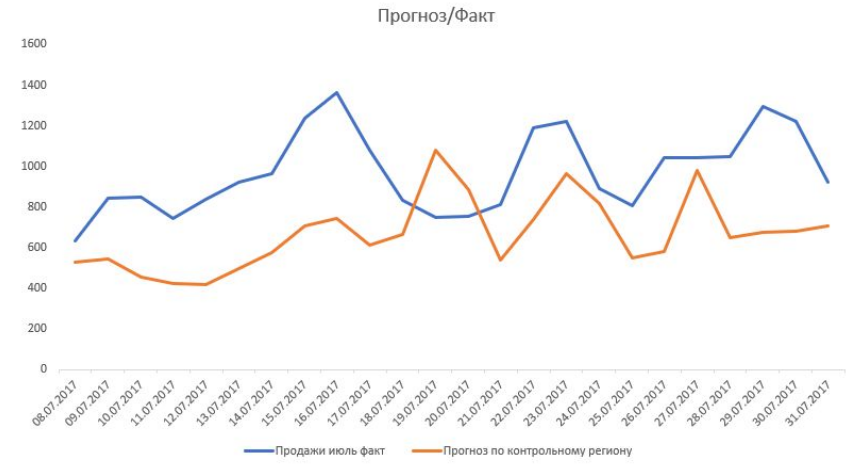
**Вариант 2 (продолжение):** После проведения A/B теста сравнить число продаж в регионе с “включенной” рекламой и в похожем регионе без рекламы.

Для точного прогнозирования используются модели ARIMA и ETS, которые подошли лучше + регрессия по контрольному региону. Auto.arima в R так же работает с регрессией.

```
fit <- Arima(y, xreg=x, order=c(1,1,0))
```

xreg – линия продаж контрольного региона

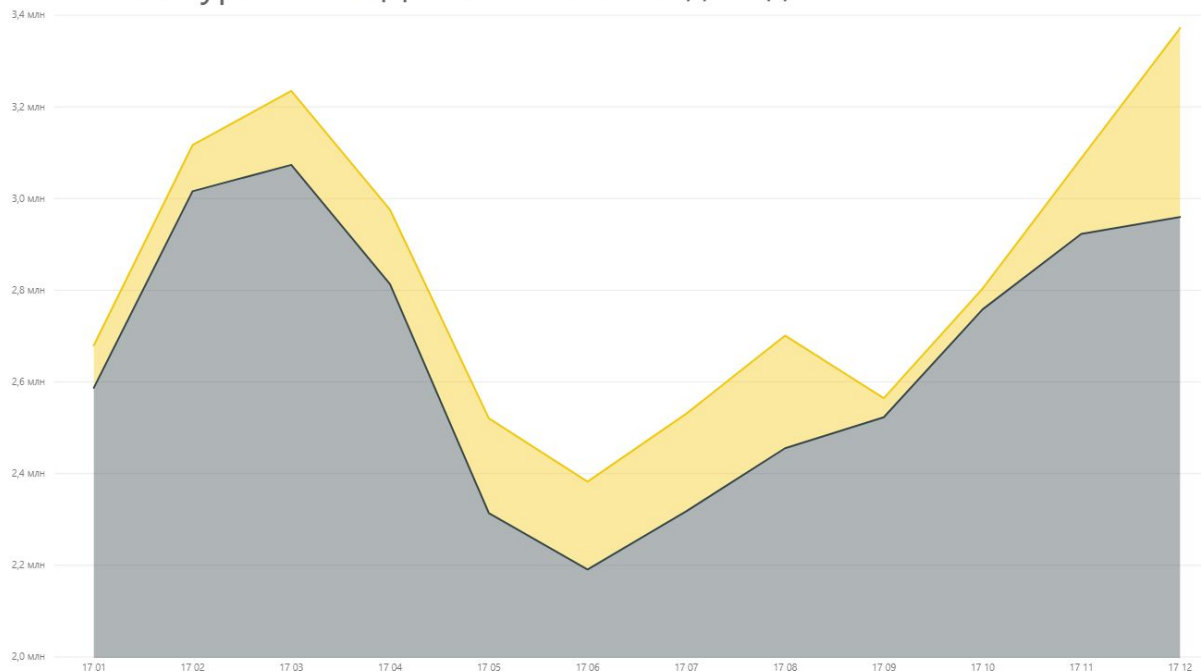
В остальном подбор модели происходит по тому же алгоритму, что и в первом случае



# Считаем дополнительный доход

Шкура убитого медведя (выделена жёлтым)

● Базовый уровень ● Дополнительный доход

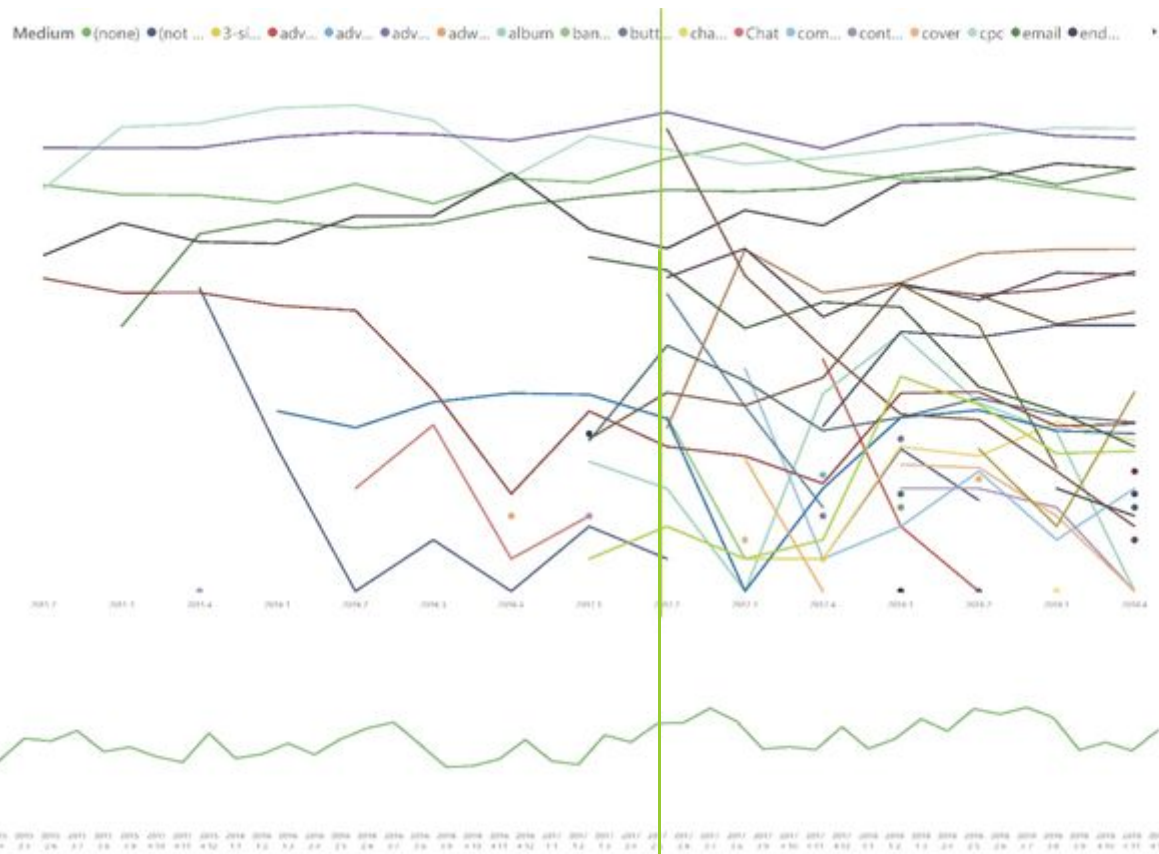


Теперь вычитаем из того, что получили, наш прогноз (+- ошибка прогноза) и получаем чистый добавочный доход - то, что заработали сверх того, что должны были заработать.

Либо, если отключали рекламу, мы считаем, сколько мы НЕ заработали

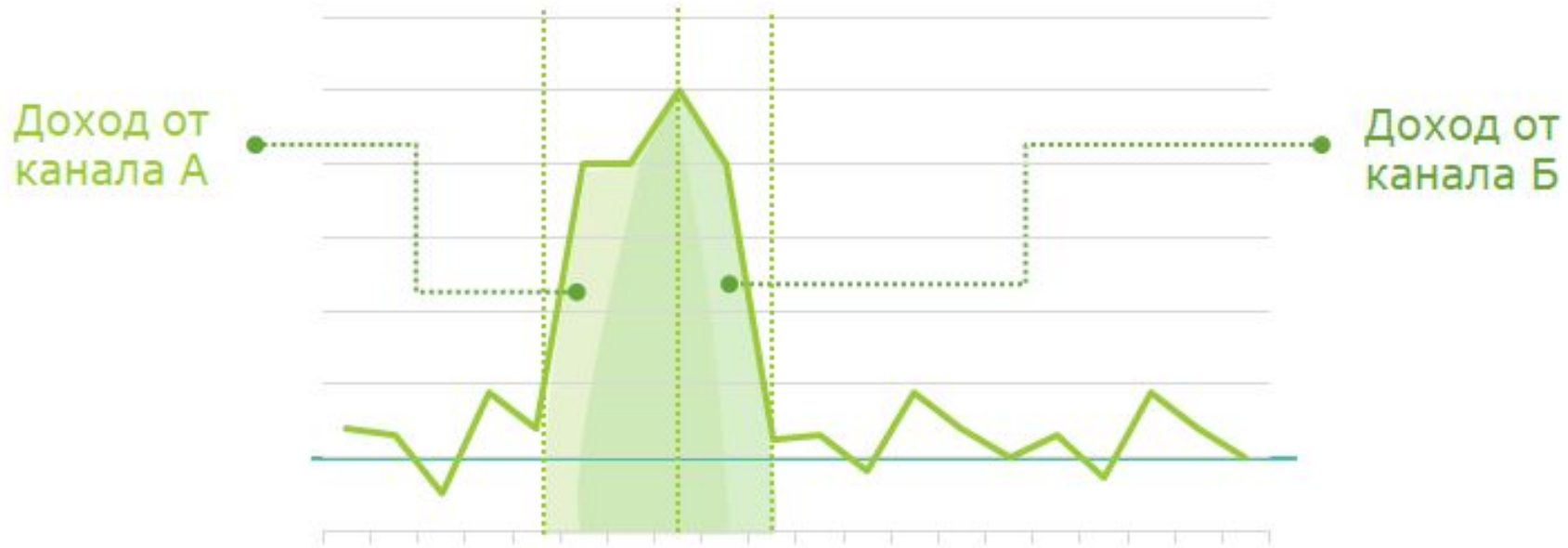
# Считаем влияние рекламы

Что из этого повлияло на продажи?



# Считаем влияние рекламы

А вот теперь корреляция



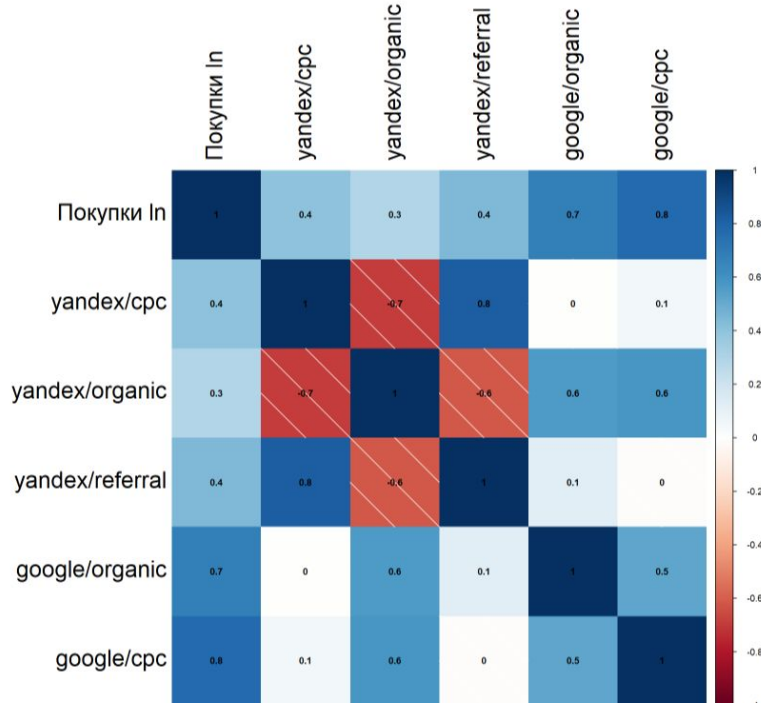
Корреляция это не что иное, как сопоставление “всплесков” трафика(рекламы) и “всплесков” продаж(или контактов). Иногда такие всплески могут идти “с отставанием”. Кроме того, каналы влияют друг на друга, и всплески в одном канале могут быть вызваны всплеском в другом.



# Считаем влияние рекламы

Как посчитать корреляцию каналов:

```
_data %>% cor() %>% corrplot()
```



1. Считаем корреляцию - влияние трафика на продажи. Все, что от 0.6 – повлияло.

*Рекомендуется считать все типы контактов, а не только последний перед продажей. В том числе можно считать всё, что угодно. Например баннер над входом, 1- висел, 0 – не висел.*

1. Нормализовать данные.

*Для этого можно использовать натуральный логарифм (Ln), или в R – `boxcox(x, ...)` - трансформацию Бокса-Кокса (автоматически подбирает лучший вариант нормализации)*

3. Интерпретация. Чем более тёмный синий цвет в первом столбике – тем больше изменение трафика по этому каналу повлияло на продажи. В данном случае на продажи более всего повлияли каналы Google Organic и Google CPC



# Как считать вклад каналов

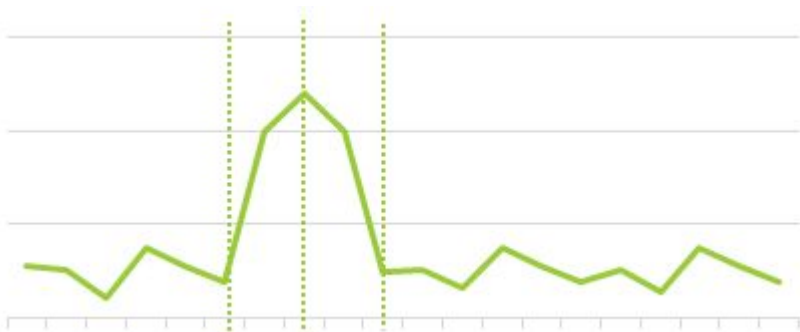
Как атрибутировать добавочный доход каналам рекламы:

Атрибуция дохода каналам рекламы выходит за рамки данной темы, так как подбор модели атрибуции может сильно отличаться для каждого случая и трудно описать его кратко.

Для начала можем предложить простую модель атрибуции по степени влияния на результат:

1. Выделяем сегменты. Например: время выхода рекламы, география. В рамках этого сегмента считаем для продаж: факт - прогноз\* = дополнительный доход.

Период: июль, город: Санкт-Петербург



Доход за месяц	500,000.00 Р
Прогноз дохода без рекламы	400,000.00 Р
Добавочный доход	100,000.00 Р

\*Как считать прогноз без рекламы мы подробно [рассмотрели](#) на слайдах ранее

# Как считать вклад каналов

Как атрибутировать добавочный доход каналам рекламы:

2. Выделяем каналы рекламы с охватом в сегменте: Санкт-Петербург, Июль. Считаем корреляцию с продажами, как на слайдах [ранее](#). Высчитываем долю влияние каждого канала:

Канал рекламы	Корреляция с продажами	Доля в прибыли
yandex/cpc	0.4	$=\frac{@[Корреляция с продажами]}{СУММ([Корреляция с продажами])}$
yandex/organic	0.3	12%
yandex/refferal	0.4	15%
google/organic	0.7	27%
google/cpc	0.8	31%

Дополнительно аналогично можно взвесить доли по силе охвата (кол-во показов, переходов, целевых действий) или другим значимым параметрам:

Канал рекламы	Корреляция с	Доля К	Охват	Доля O	Доля в прибыли
yandex/cpc	0.4	15%	150	15%	$=\frac{([Доля К]+@[Доля O])}{200\%}$
yandex/organic	0.3	12%	40	4%	8%
yandex/refferal	0.4	15%	300	30%	23%
google/organic	0.7	27%	500	50%	39%
google/cpc	0.8	31%	1	0%	15%

# Как считать вклад каналов

Как атрибутировать добавочный доход каналам рекламы:

Получившуюся долю влияния каждого канала умножаем на добавочный доход и получаем долю прибыли, которую уже можно сопоставить с затратами на каждый канал в данном сегменте (Петербург, Июль) и рассчитать ROI

Доход за месяц	500,000.00 ₽						
Прогноз дохода без рекламы	400,000.00 ₽						
Добавочный доход	100,000.00 ₽						
Канал рекламы	Корреляция с	Доля К	Охва	Доля С	Доля в	Доля в прибыли, р.	
yandex/cpc	0.4	15%	150	15%	15%	=[@[Доля в прибыли]] * \$E\$3	
yandex/organic	0.3	12%	40	4%	8%		7,787.39 ₽
yandex/refferal	0.4	15%	300	30%	23%		22,828.53 ₽
google/organic	0.7	27%	500	50%	39%		38,688.58 ₽
google/cpc	0.8	31%	1	0%	15%		15,435.07 ₽



**Благодарю за внимание.**  
Вопросы?

Руководитель отдела аналитики  
Coffee Analytics

[anna@coffee-analytics.ru](mailto:anna@coffee-analytics.ru)

8 919 017-74-34

[www.coffee-analytics.ru](http://www.coffee-analytics.ru)

# Дополнение: Power analysis in R: T-Test

Шпаргалка для подбора размера выборки для опроса

## Power analysis in R: T-Test

```
library(pwr)

pwr.t.test(power = 0.8,
           sig.level = 0.05,
           d = 0.2)
```

Two-sample t test power calculation

```
      n = 393.4057
      d = 0.2
sig.level = 0.05
power = 0.8
alternative = two.sided
```

NOTE: n is number in \*each\* group

**Параметры, которые вставляем (выделены зеленым):**

**Power** - надежность

**Sig.level** - (доверительный интервал (" +/- лапоть"))

**d** - наблюдаемая разница (кол-во покупок, на которое один канал принес больше другого)

**N** - сколько нужно покупок для каждой группы (каждого канала рекламы), чтобы говорить о достоверности

# Дополнение: Основные модели ARIMA

Шпаргалка для подбора моделей

ARIMA (0, 0, 0)	—	WHITE NOISE
ARIMA (0, 1, 0)	—	RANDOM WALK
ARIMA (1, 0, 0)	—	AUTOREGRESSIVE MODEL (order 1)
ARIMA (0, 0, 1)	—	MOVING AVERAGE MODEL (order 1)
ARIMA (1, 0, 1)	—	SIMPLE MIXED MODEL

# Дополнение: Проверка модели (AIC)

Шпаргалка для выбора оптимальной модели

После подбора своих вариантов моделей, нужен критерий - как выбрать лучшую.

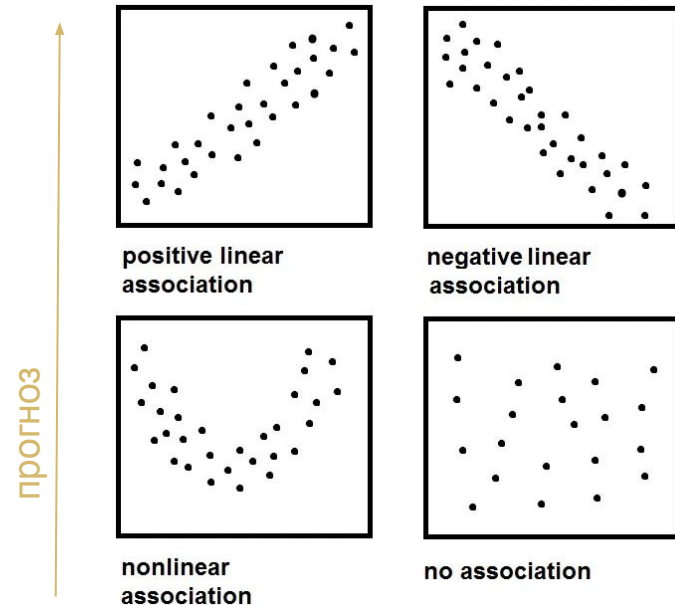
Существуют несколько подходов, один из распространенных - графический. Построить диаграмму рассеяния (см. Рис. справа) и посмотреть, есть ли связь. Недостаток этого метода в том, что связь не всегда видна.

Предлагается использовать метрики точности прогнозов AIC\* (команда CV() в R):

```
CV(fit.consMR)
#>      CV      AIC      AICc      BIC      AdjR2
#> 0.1163 -409.2980 -408.8314 -389.9114 0.7486
```

Мы хотим найти модель с наименьшим значением этих величин

\*кроме правдоподобия ещё учитывает кол-во используемых данных при построении модели



факт

# Дополнение: Ссылки на самоизучение

1. Курс Forecasting in R Data Camp  
<https://learn.datacamp.com/courses/forecasting-in-r>
  2. Курс Time Series Analysis in R Data Camp
  3. <https://learn.datacamp.com/courses/time-series-analysis-in-r>
  4. Учебник "Forecasting: Principles and Practice" Rob J Hyndman and George Athanasopoulos. Monash University, Australia  
<https://otexts.com/fpp3/>
1. Курс A/B Testing in R <https://learn.datacamp.com/courses/ab-testing-in-r>
  2. Курс Machine Learning for Marketing Analytics in R  
<https://learn.datacamp.com/courses/machine-learning-for-marketing-analytics-in-r>

Наше видео с облегченной версией ROPO Аналитики: Мебельный e-commerce: аналитика от бюджета до продажи

<https://www.youtube.com/watch?v=Fnj3WYaDpIE>