



ТИНЬКОФФ

Лекция 10. Модели

«Урок, который я извлек, касается всего статистического анализа и заключается в том, что даже самые точные измерения или вычисления не должны противоречить здравому смыслу» - Чарльз Уиллан (Голая статистика)

Начнем с определения



Модель - это инструмент анализа данных, который используется для создания прогнозов и оптимизации процессов в различных сферах, включая экономику, финансы, инженерию и медицину.



Модель - это математическое, статистическое или компьютерное представление реальной системы или явления, которое используется для анализа данных и прогнозирования поведения системы в различных условиях.



Модель - это упрощенное представление сложной системы, которое помогает аналитикам лучше понять ее особенности и принимать решения на основе этого понимания.



ТИНЬКОФФ

Разминка

Какие модели вы знаете?

Можете поднять руку и назвать свой вариант устно, а можете просто написать свои мысли в чат



Простой пример

Вы - **владелец**ы кофейни

Продажи за W1

2 000



Капучино x 100



Апельсиновый сок x 50

Продажи за W2

3 000



Апельсиновый сок x 150

Продажи за W3

?



Капучино x 130



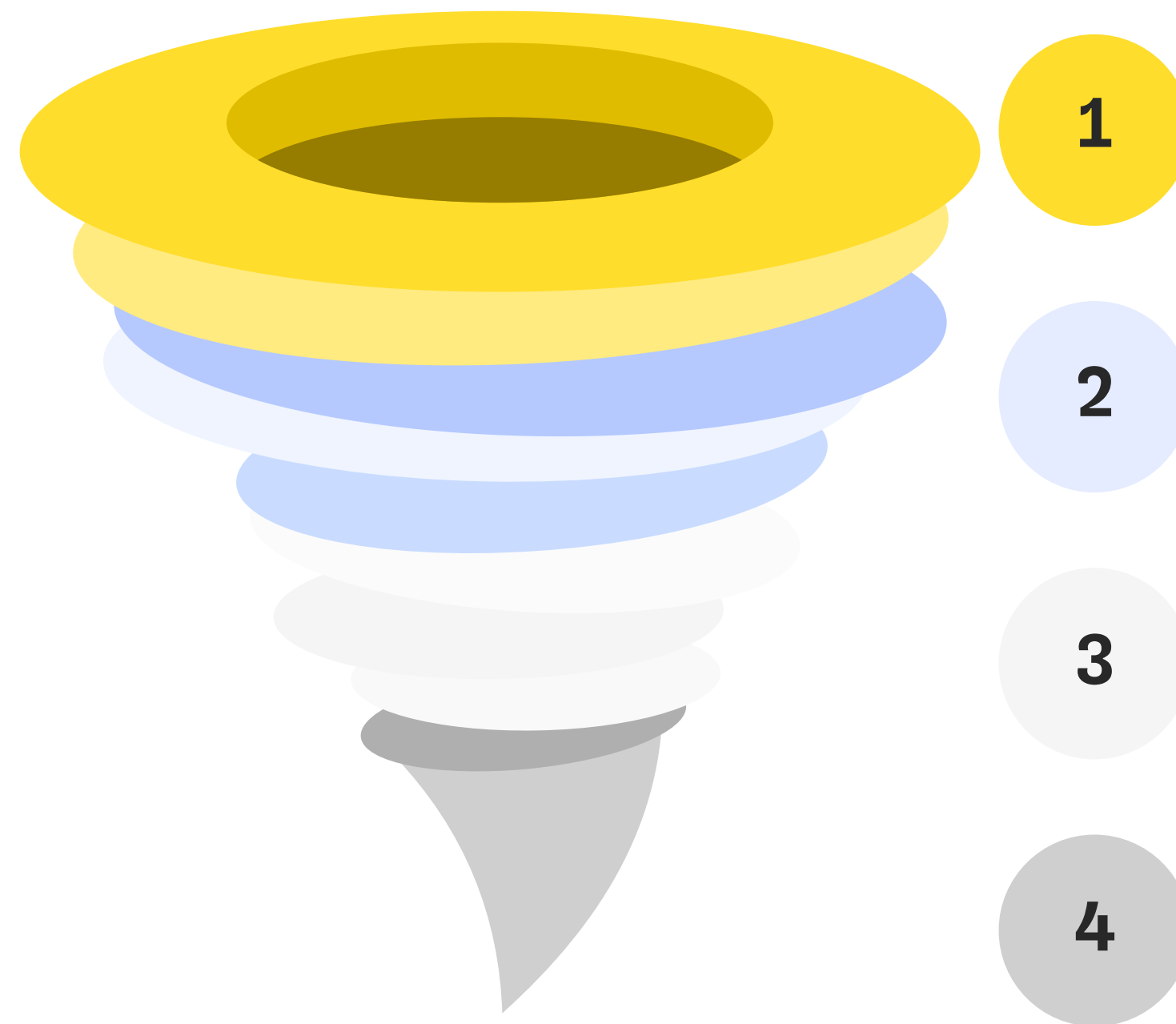
Апельсиновый сок x 60



Яблочный пирог x 20

Пример чуть сложнее

Знакомимся с воронкой вашей кофейни



1

Трафик людей рядом

Сколько людей проходит рядом с
вашей кофейней

2

Заглянули посмотреть меню

Потенциальные гости вашей кофейни

3

Спросили что-то у бариста

Еще чуть чуть и гость сделает заказ

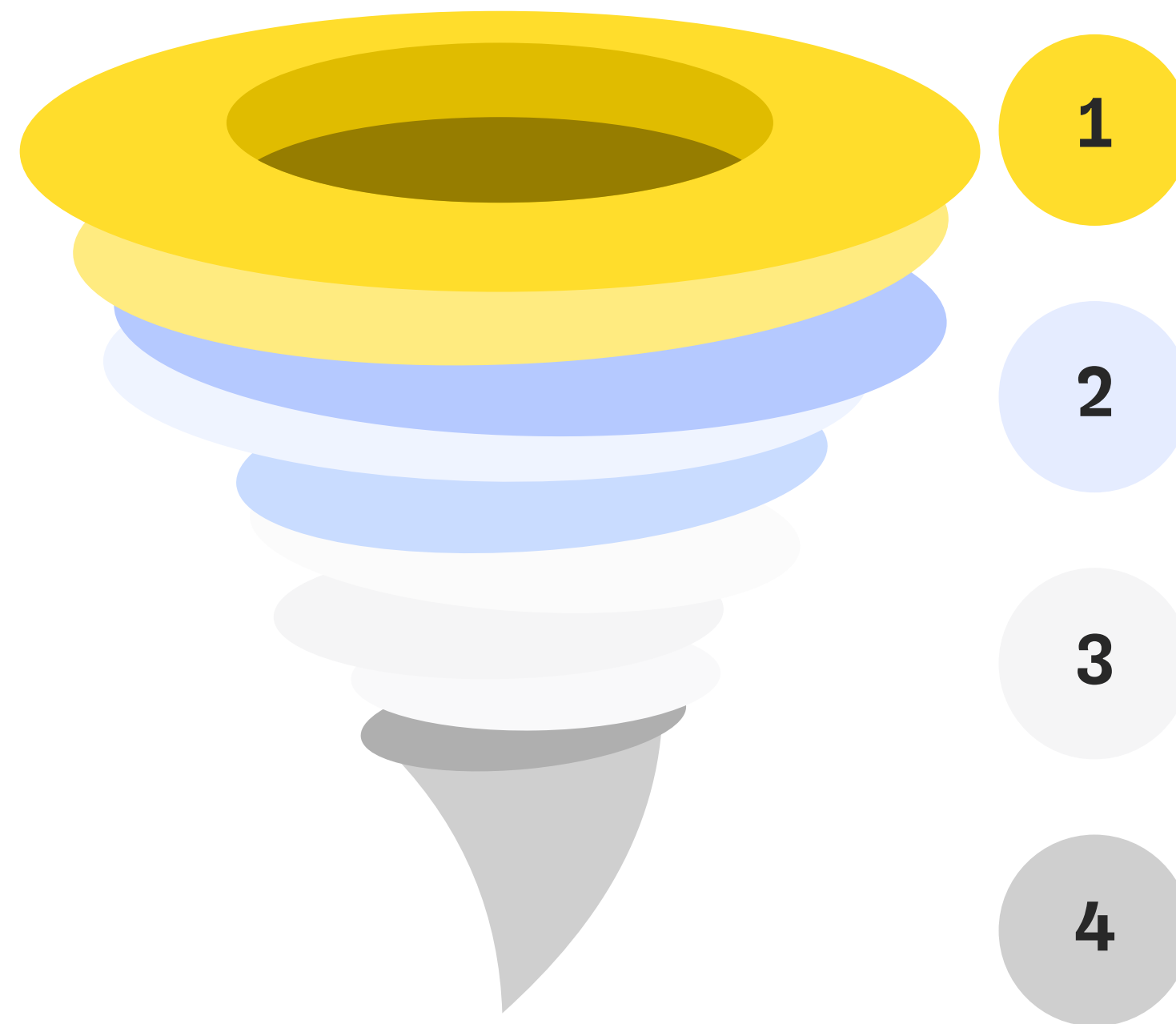
4

Сделали заказ и оплатили его

Ура! Мы что-то заработаем

Без духоты

Дадим еще немного ВВОДНЫХ данных



Трафик людей рядом

1 000 человек

Заглянули посмотреть меню

60% человек зашли

Спросили что-то у бариста

Новички что-то купят с вероятностью 50%,
Старички с вероятностью 80% [пусть 50 на 50](#)

Сделали заказ и оплатили его

Сколько будет заказов? **390**

Пусть:

- 1. Мы открываем по одной новой кофейне каждый месяц, кофейни открываем в одинаковых районах и имеем один и тот же трафик по объему.
 - 2. В марте, мае и декабре случается сезонные пики и рядом с кофейнями ходит в полтора раза больше людей;
 - 3. Доля старичков будет расти линейно от 0% в 1ый месяц работы до 35% на десятый месяц.
- Пользователи уникальные и делают один заказ в месяц.

Задание на дом №1

**Посчитайте
количество
заказов по
итогу года
экспансии**



На примерах мы затронули процесс аппроксимации

Аппроксимация - это процесс построения простой модели, которая приближает нас к пониманию сложной системы или явления.

Аппроксимация может использоваться для того, чтобы упростить модель, сократить время вычислений и уменьшить количество данных, необходимых для построения модели. Часто проявляется в виде допущений при вычислениях.

Когда вам нужно построить модель какого-то сложного процесса или оценить потенциальный эффект от внесения какого-то изменения в продукт, вам на помощь приходит аппроксимация.

Однако стоит быть очень осторожным с тем что мы допускаем и проверять любые вводные.

Для чего нужно строить модели?

Обычно вы сначала найдете вопрос или проблему, а затем решите моделировать процесс, чтобы получить конкретный ответ или проверить чей-то готовый вывод альтернативным путем.

Важно понимать, что бывает наоборот: имея расплывчатую задачу, либо нелокализованную проблему, вы сначала построите модель текущего процесса, чтобы получить представление «изнутри», как улучшить процесс или где теория не соответствует практике.

А какие виды моделей бывают?



Математические модели

модели, основанные на математических уравнениях. Они могут использоваться для прогнозирования поведения системы в различных условиях. Например, модель экономического роста может использоваться для прогнозирования изменений в экономике на основе различных факторов, таких как инфляция, безработица и т.д.



Статистические модели

модели, основанные на статистических данных. Они могут использоваться для анализа данных и прогнозирования поведения системы. Например, модель регрессии может использоваться для прогнозирования цены на недвижимость на основе различных факторов, таких как количество комнат, площадь и т.д.



Компьютерные модели

модели, которые используют компьютеры для моделирования поведения системы. Они могут использоваться для анализа больших объемов данных и прогнозирования поведения системы в различных условиях. Например, модель машинного обучения распространения эпидемии при учете множества факторов.



Физические модели

модели, которые основаны на физических уравнениях и принципах. Они могут использоваться для моделирования поведения физических систем, таких как течение жидкости или электрическое поле. Например, модель течения жидкости может использоваться для прогнозирования изменений в течении реки на основе различных факторов, таких как скорость ветра, температура и т.д.

Важные нюансы при работе с моделями



Целевая переменная, набор метрик и область их определения

Объем продаж, кол-во клиентов, выручка, ...



Параметры, их область значений

Кол-во дней в периоде, Меню, Акции,
Конкуренты, и многое-многое другое...



Решение об экстраполяции

А если бы мы делали прогноз на квартал / год?



«Устойчивость» полученных выводов

Что нам лучше продавать?

Финансовые модели



финансовую модель можно использовать для прогнозирования доходности инвестиций на основе различных факторов, таких как ставки процента, инфляция, рыночные условия и т.д.



Вы хотите открыть свою кофейню. Размер инвестиций 300к рублей.

Как оценить доходность ваших вложений?

На чем мы зарабатываем?

- Маржа с продажи товаров
- Размещение рекламы
- Чаевые

Что мы тратим?

- Себестоимость продуктов и товаров;
- Налоги;
- Аренда;
- ФОТ;
- Амортизация оборудования;
- Прочие издержки.

финансовые модели

Месяц	0	1	2	3	4	5	6	7	8	9	10	11	12
Инвестиции	-300												
Выручка	0	50	55	61	67	73	81	89	97	107	118	130	143
Маржа		40%	40%	40%	40%	40%	40%	41%	41%	41%	41%	41%	41%

Прибыль	0	20	22	24	27	30	33	36	40	44	48	53	58
CF	-300	-280	-258	-234	-207	-177	-145	-109	-69	-26	22	75	134

Мы инвестировали 300 тыс. рублей, а за год заработали 434 тыс. рублей. Мы в плюсе?

$$\text{ROI} = \frac{(\text{доход} - \text{затраты})}{\text{затраты}} \times 100\% \quad \text{ROI} = 45\%$$

$$NPV = \sum_{k=1}^n \frac{Pk}{(1+i)^k} - IC$$

n – период расчета;

P_k – денежные потоки за выбранный период времени;

i – ставка дисконтирования;

IC – размер первоначальных вложений

Обычно ставка дисконтирования берется как 15% годовых.

NPV	-	-280	-259	-235	-210	-182	-152	-119	-83	-44	-2	44	95
-----	---	------	------	------	------	------	------	------	-----	-----	----	----	----

Рекламная кампания

20 тыс. руб.

Так как все было онлайн, вы
смогли посчитать, что кампания
привела вам 100 человек.

CAC vs LTV

200 vs 1 200

За 6 месяцев работы по
получаем с пользователей в 6
раз больше, чем потратили на их
привлечение. Хорошее
соотношение 1:4.

CAC

200 руб.

Customer acquisition cost –
расходы организации на
привлечение одного пользователя

LTV (Life Time Value)

валовая прибыль, которую средний пользователь
принесет за все время использования продукта
! Валовая прибыль — это разница между выручкой и
всеми переменными затратами, которые напрямую
ассоциированы с реализованной продукцией или
услугой (COGS или Cost of Goods Sold).

Как посчитать LTV?

LTV -> Когортный анализ, а можно и аппроксимацией.

$$LTV = ARPU * LT$$

Пусть гость берет в кофейне 2 чашки кофе за 250 рублей в месяц, а наша
маржа при этом 40%. $ARPU = \sum GP * Margin$ И пусть LT = 6 месяцев

Хороший пример

Пример модели роста

Month	LG_number	Volume_rub
фев.21	218	39525927
апр.21	561	60780987
ноя.20	221	44873845
янв.21	222	47372236
дек.20	223	58804445
мар.21	326	61965168

В какие месяца наблюдался рост/спад оборотов?
Как это можно объяснить?

Month	Volume_rub	Points
ноя.20	44,87	221
дек.20	58,80	223
янв.21	47,37	222
фев.21	39,53	218
мар.21	61,97	326
апр.21	95,97	561

- Сортировка месяцев и формат таблицы ✓
- Рубли перевели в млн. рублей ✓

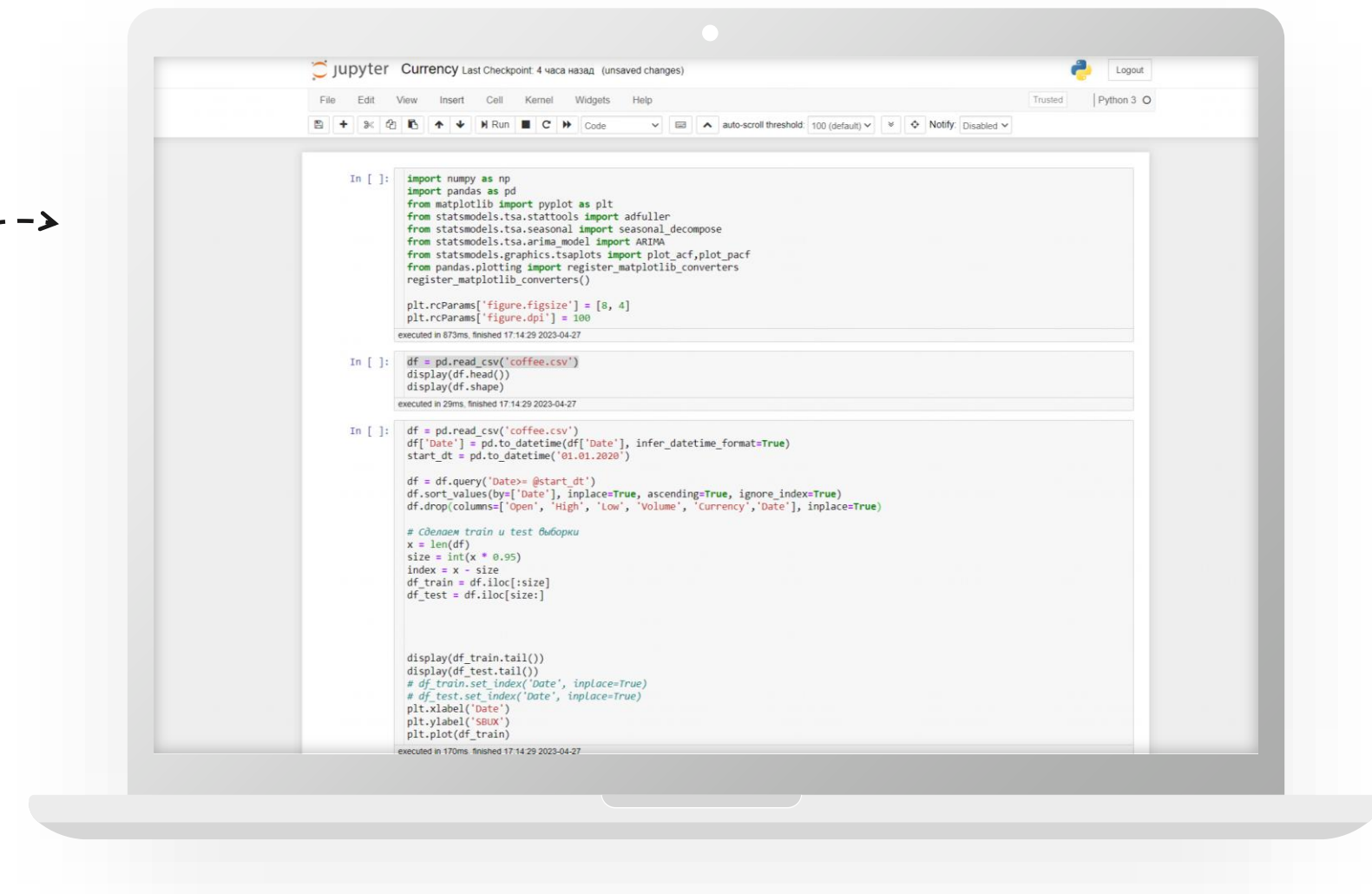
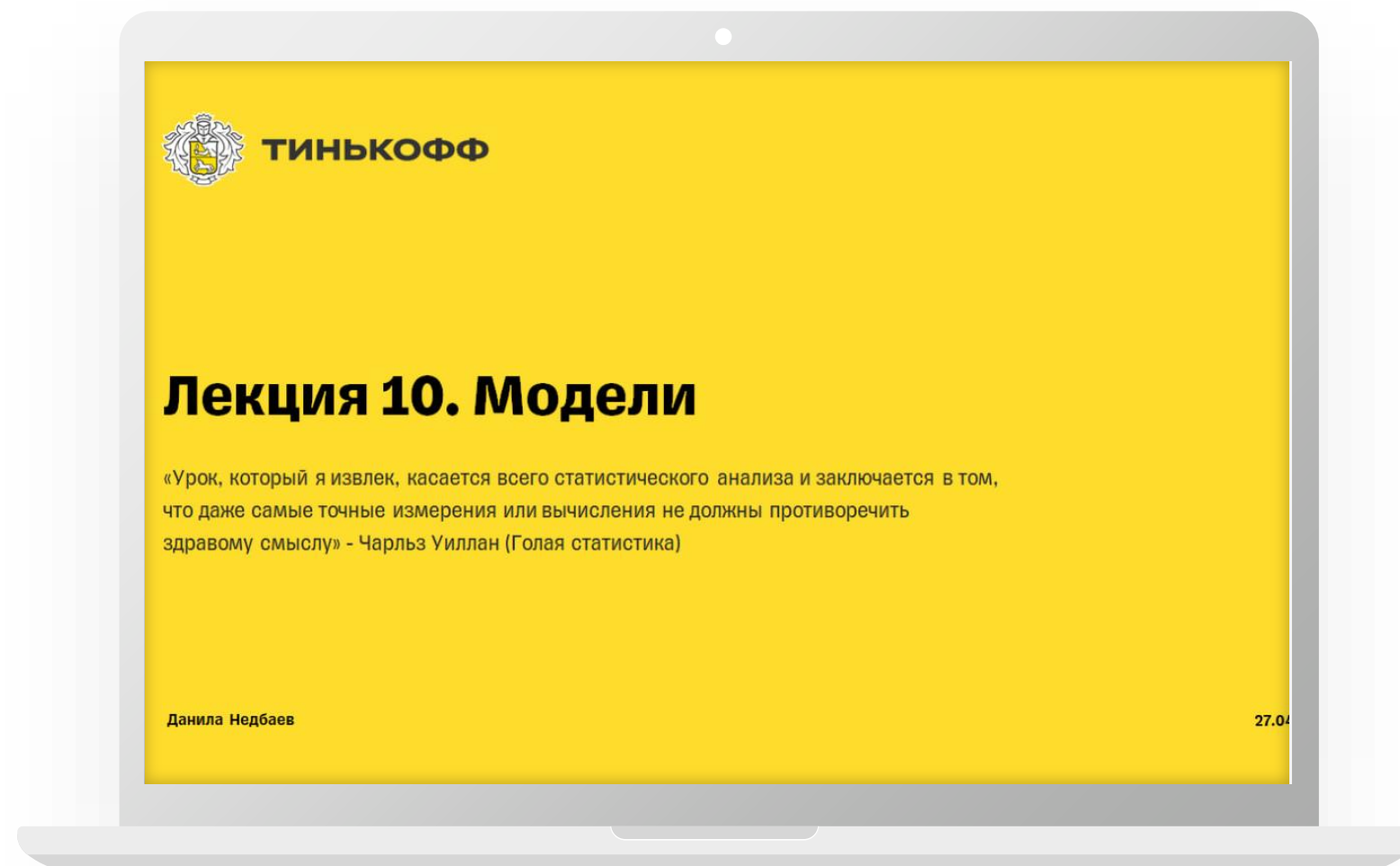
Пример модели **роста**

В какие месяца наблюдался рост/спад оборотов?
Как это можно объяснить?

Month	Volume_rub	Points	Vol/Point	Seasonality
ноя.20	44,87	221	203	99.7%
дек.20	58,80	223	264	129.6%
янв.21	47,37	222	213	104.6%
фев.21	39,53	218	181	88.9%
мар.21	61,97	326	190	93.3%
апр.21	95,97	561	171	84.0%
			Av vol/point	203,67

Дальше будет немного Питона

Перейдем к прогнозированию временных рядов



Используя датасет о стоимости кофе из Kaggle нам
нужно построить модель прогнозирования
стоимости кофе в будущем



ТИНЬКОФФ



Вам было полезно?