

# Прикладной анализ данных в ритейле

## Вводная лекция

Элен Теванян

свободный художник



# Вводное занятие

## 1. Организационное

# План занятий

- Занятие 1. Вводная лекция
- Занятие 2. Прогноз спроса
- Занятие 3. Uplift- и look-alike-моделирование
- Занятие 4. Рекомендации
- Занятие 5 и занятие 6. Ценообразование
- Занятие 7. Эксперименты

# Что будет на курсе

## Как оценим курс

- **Формула:**  $0.25\text{Проект\_1} + 0.25\text{Проект\_2} + 0.5*\text{Проект\_3}$
- **Проект 1: Прогнозирование спроса или пользовательской метрики**
- Суть: Небольшая лабораторная работа с заданиями от меня на не очень простом датасете.
- **Проект 2: Оценка нестандартных экспериментов**
- Суть: В этой проектно-лабораторной работе практикуемся в анализе эксперимента, задизайненного не как АБ-тест.
- **Проект 3: ML System Design**
- Суть: я предлагаю вам формулировку задачи, а вам нужно сделать ее решение

# Вводное занятие

1. Организационное
- 2. Про ритейл, цели и данные**

# Ритейл

## Элемент цивилизации





# Логистика Тоже изменяется



# Ритейл и сервисы доставки продуктов

## Где серьезно работают с данными

 Пятёрочка

 лента

 OZON

 ВкусВилл

 Перекрёсток

 Delivery Club

 MAGNIT

 самокат

 Яндекс Еда

 Перекрёсток  
быстро

Яндекс  Лавка

 УТКОНОС:  
онлайн-гипермаркет

 СБЕР MARKET

Яндекс  Маркет

 Vprok.ru  
Перекрёсток



Логистика  
Тоже изменяется

# Логистика

## Тоже изменяется

### РТО

---

Розничный  
товарооборот

Выручка

# Логистика

## Тоже изменяется

### РТО

---

Розничный  
товaroоборот

Выручка

### Прибыль

---

Маржа/EBITDA

Что скажут финансы

# Цель бизнеса для DS-решений

РТО

Прибыль



# Цель бизнеса для DS-решений

РТО

Затраты

Прибыль

# Цель бизнеса для DS-решений

РТО ↑

Затраты ↓

Прибыль ↑

# Классический ритейл: Две сущности



Магазин



Покупатели

# Доставка: Три сущности



Магазин



Покупатели



Курьеры



# Источники Данных

# Источники Данных

- Чеки
- Промо
- Цены
- Магазины
- Товары
- Лояльность
- Внешние отчеты

# Источники Данных

- Чеки
- Промо
- Цены
- Магазины
- Товары
- Лояльность
- Внешние отчеты

>20 ГБ в день

# Вводное занятие

1. Организационное
2. Про ритейл, цели и данные
- 3. Задачи**





# Основные задачи

- Спрос
- Ценообразование
- Промо
- Ассортимент
- Маркетинг
- Риски
- Монетизация

# Спрос

## Core задача

Прогноз количества  
товара в магазине на  
X дней

- Снижение списаний и заполненности складов
- Снижение упущенных продаж
- Оптимизация ценообразования
- Оптимизация наполнения промо
- Оптимизация ассортимента

# Ценообразование

## Управление ценами

Разработка системы  
для автоматического  
управления ценами

- Решается как оптимизационная задача с ограничениями с максимизацией выручки или маржи
- Оценка эластичности цен

# Ценообразование

## Управление ценами

Разработка системы  
для автоматического  
управления ценами

Стоимость доставки  
или стоимость  
минимальной корзины  
для доставки

- Решается как оптимизационная задача с ограничениями с максимизацией выручки или маржи
  - Оценка эластичности цен
- 
- Максимизируется качество сервиса
  - Базируется на прогнозах нагрузки



Промо

Еще один core продукт

Выбор продуктов для  
наполнения промо

# Эффективность промо сейчас:

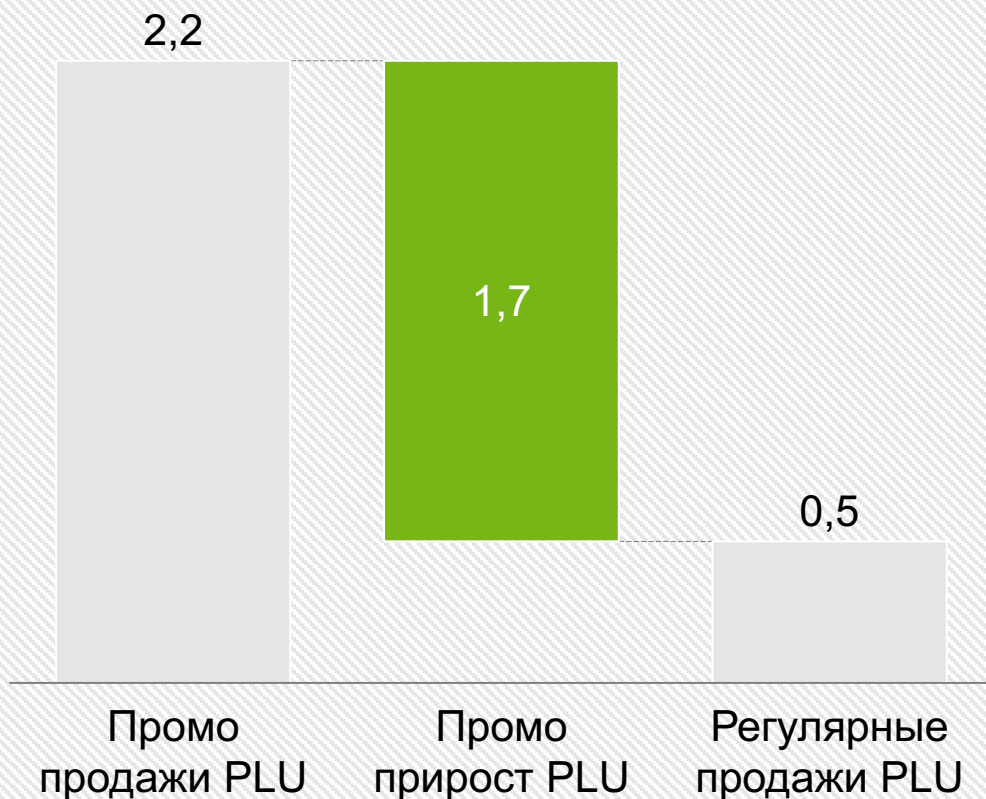
Выбор товара на основе увеличения продаж в промо



## Какой из товаров вы бы скорее поставили в промо?

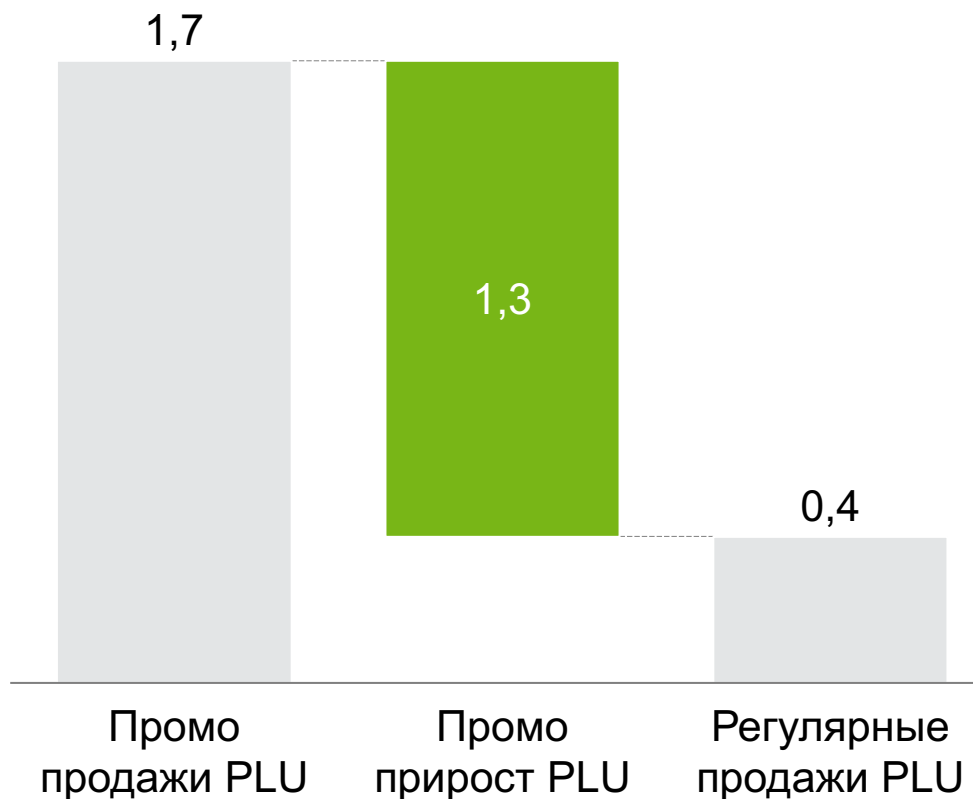
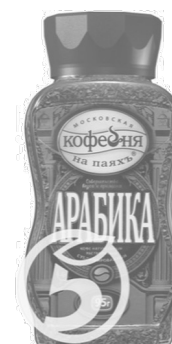
### Изменение продаж товара в промо

Тыс. руб./неделя/магазин



### Изменение продаж товара в промо

Тыс. руб./неделя/магазин



# Промо

## Еще один core продукт

Выбор продуктов для  
наполнения промо

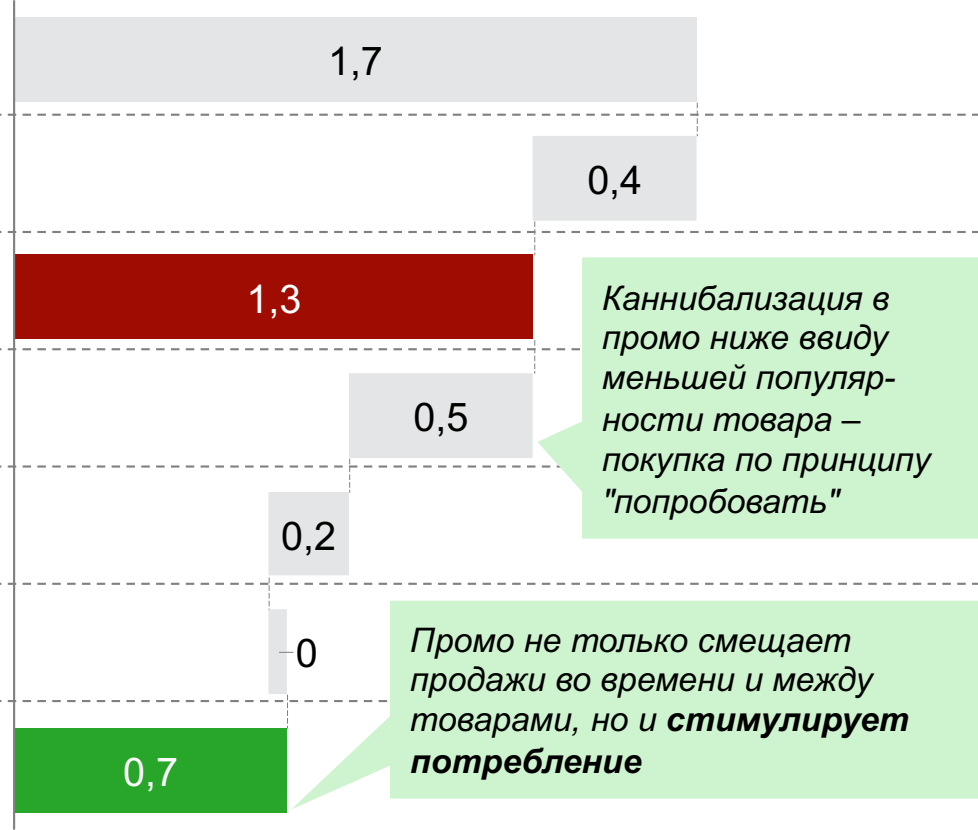
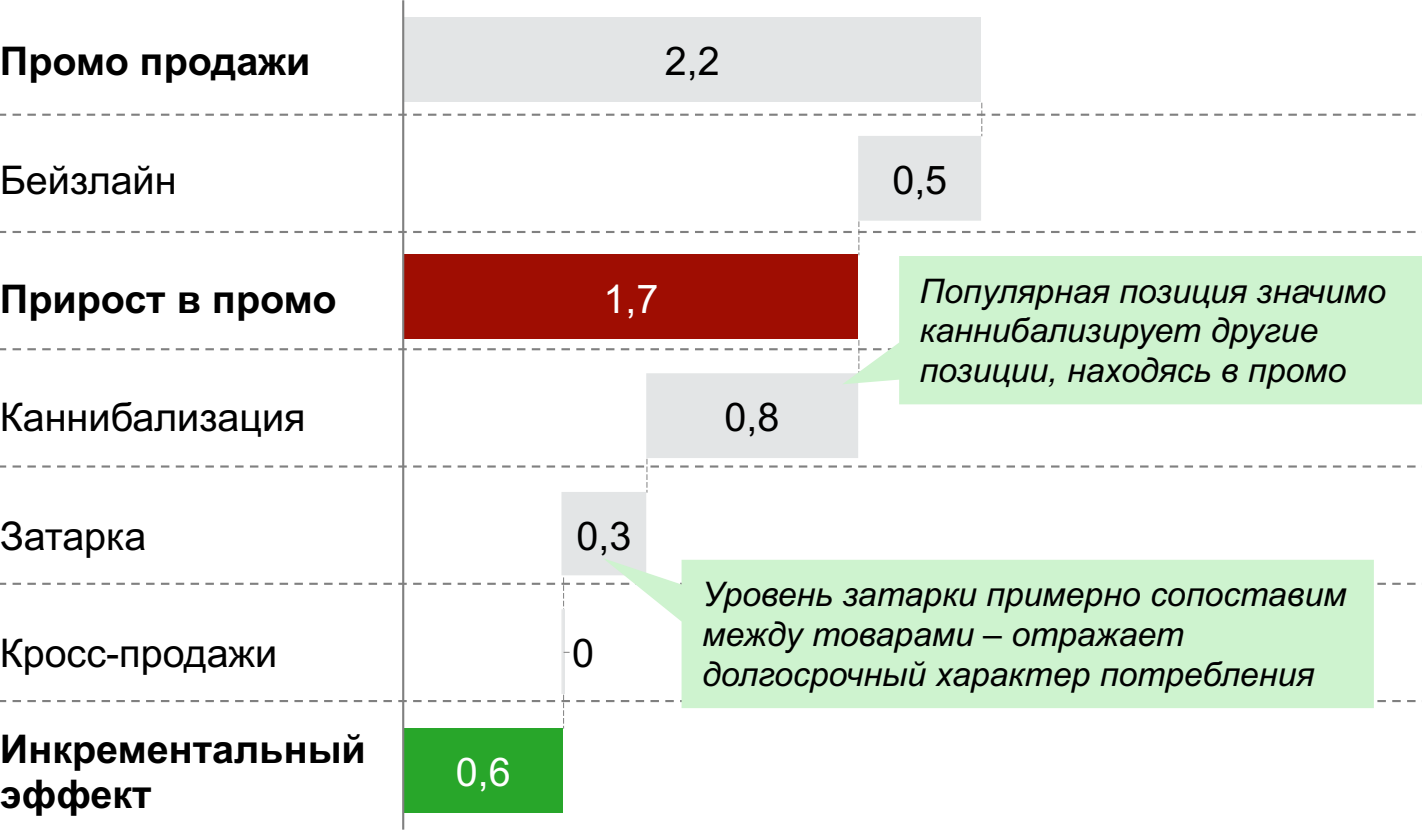
- Классическая метрика эффективности – прирост продаж в период промо
- Снижение продаж после промо, каннибализация продаж товаров-заменителей
- Переход к системе из 4х показателей:
  - Прямой эффект
  - Каннибализация
  - Затарка
  - Кросс-продажи
- Основная часть магии происходит за счет построения CDT



**Эффект промо на РТО на магазин**  
JACOBS Кофе GOLD нат. раст. субл.95 г.



**Эффект промо на РТО на магазин**  
Кофе АРАБИКА раст. субл. стекло 95 г.



# Маркетинг

## Задачи в маркетинге

Упор на  
персонализацию  
маркетинга

- Прогноз клиентских метрик:
  - РТО, кол-во чеков, средних чек, доля РТО
- Прогноз оттока
- Рекомендации товаров
- Рекомендации категорий
- Uplift для выбора сегмента под оффер
- Uplift для выборка канала
- Аналитика кампаний по моделям
- Методология оценки кампаний

# Монетизация данных

## Решения для рынка

Продукты,  
помогающие  
участникам рынка  
принимать лучшие  
решения

- Делают все, у кого есть данные

# Риски

## Проверка гипотез

Каждая инициатива  
несет в себе риск.

- Царство АБ-тестов
- Есть особенности

# Deep Learning

## Компьютерное зрение

- Детекция пустот
- Определение количества очередей

- Всегда требуется дополнительное железо – дорогое и теперь сложно доступное
- Требуется точных разметок исходных данных
- Требуется мэтчинг



# Deep Learning

## Компьютерное зрение



# Deep Learning Speech Recognition

Разработка голосовых  
и чат-ботов для  
автоматизации  
процесса и удобства  
пользователя

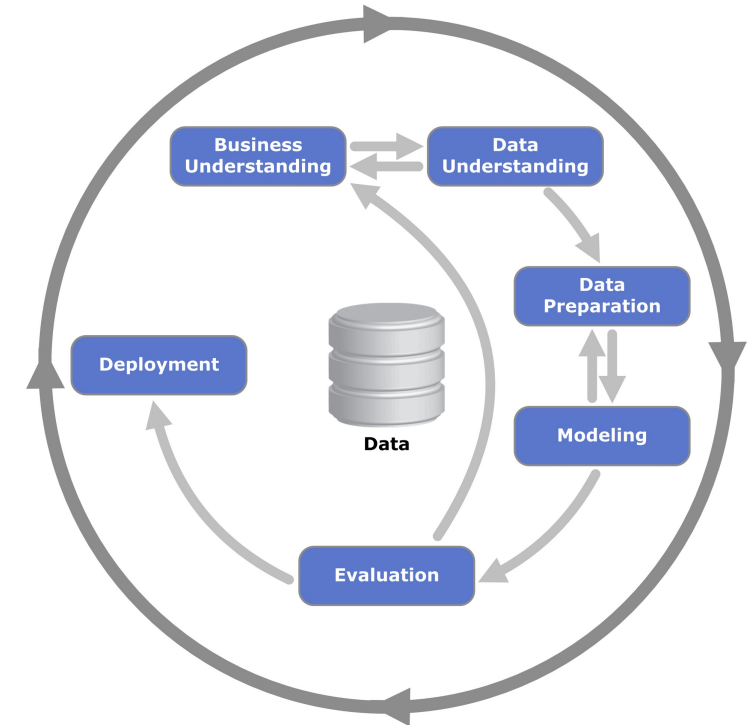
- Очень инженерный продукт
- Используются в том числе открытые решения по распознаванию речи

# Вводное занятие

1. Организационное
2. Про ритейл, цели и данные
3. Задачи
- 4. CRISP-DM**

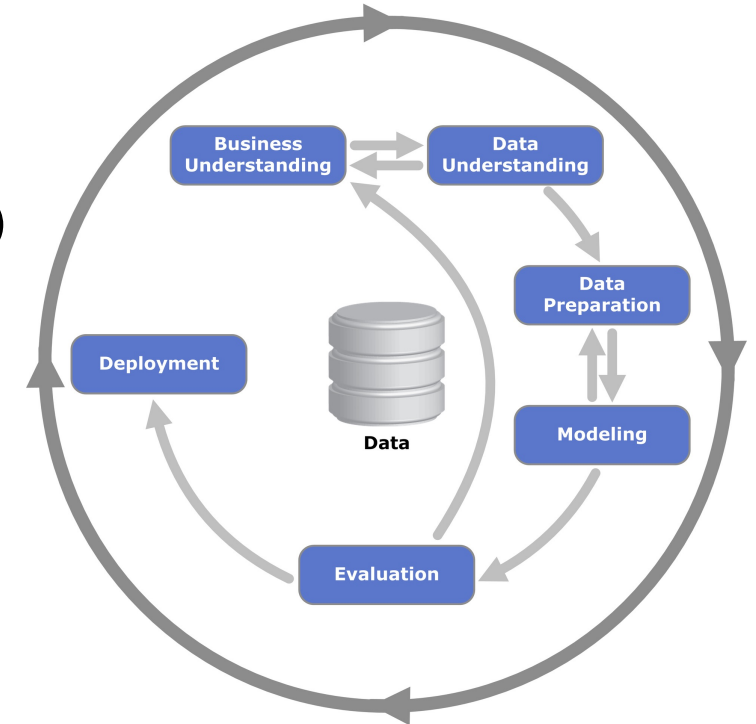
# CRISP-DM

- **Cross-Industry Standard Process for Data Mining** –  
распространенная методология по исследованию данных
- лидер по опросам последних 10 лет
- 1996: DaimlerChrysler, SPSS и Teradata  
+ 200 компаний



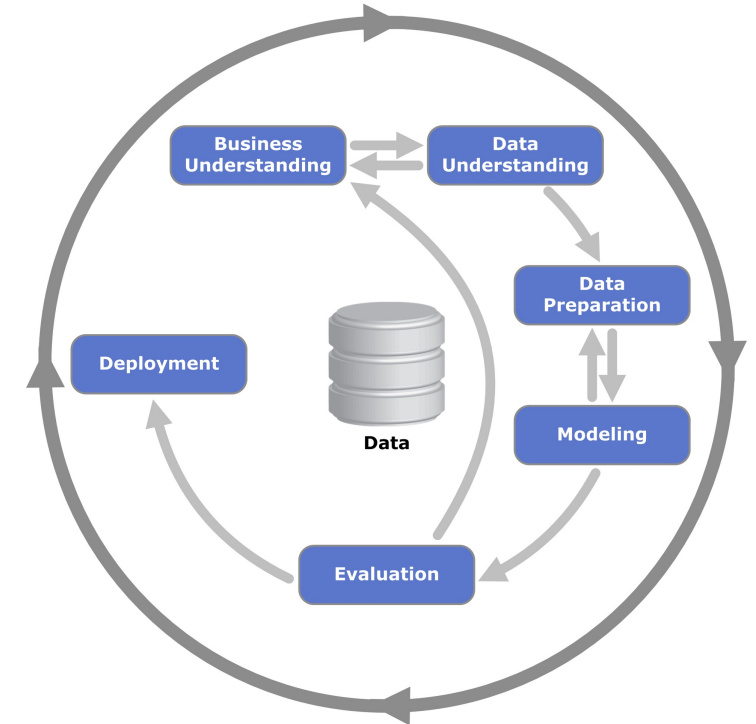
# CRISP-DM

- **Cross-Industry Standard Process for Data Mining** – распространенная методология по исследованию данных
- В ней выделяется 6 фаз:
  - Понимание бизнес-целей (Business Understanding)
  - Начальное изучение данных (Data Understanding)
  - Подготовка данных (Data Preparation)
  - Моделирование (Modeling)
  - Оценка (Evaluation)
  - Внедрение (Deployment)



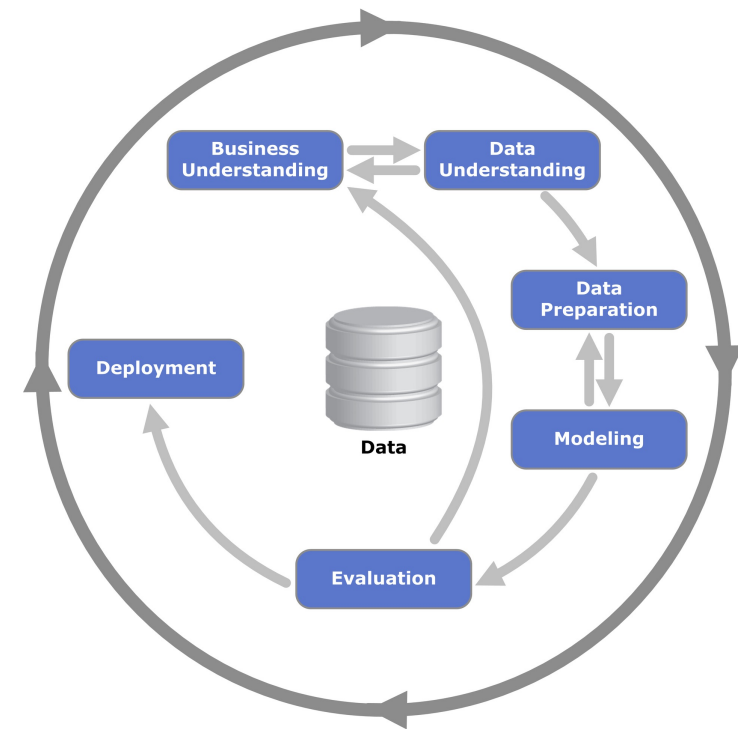
# CRISP-DM

- **Cross-Industry Standard Process for Data Mining** – распространенная методология по исследованию данных
- Это концепция, как стоит подойти к вопросу
- Не дает инструкций, как организовать работу на практике



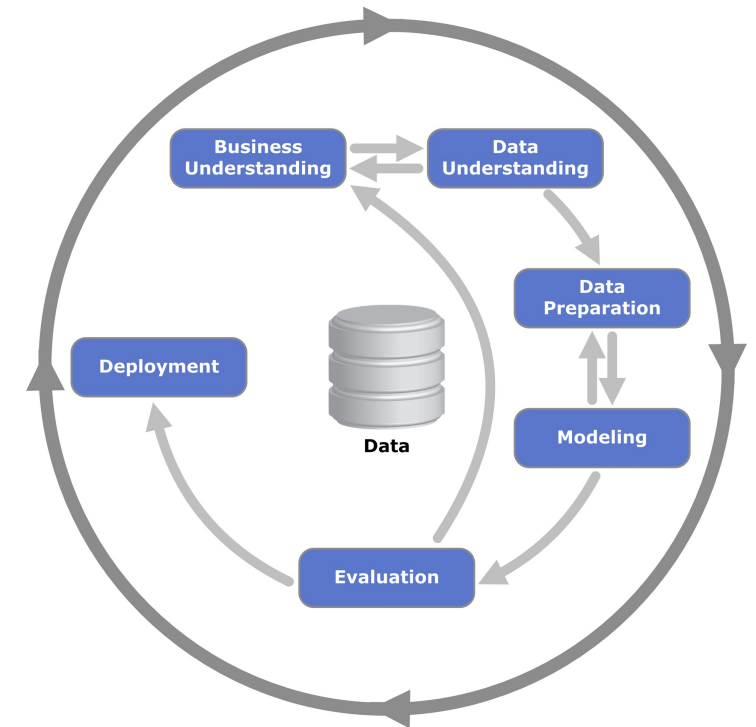
# CRISP-DM: понимание бизнес-целей

- Определение бизнес-целей:
  - Какую проблему хотим решить?
  - Какая команда занимается проектом?
  - Существуют ли решения, которые уже решают задачу?
- Оценка текущей ситуации
  - Хватит ли ресурсов для проекта: данные, железо, люди?
  - Риски проекта



# CRISP-DM: понимание бизнес-целей

- Определение целей аналитики
  - Определение метрик качества модели
  - Определение минимума / среднего приемлемого решения
- Подготовка плана проекта
  - Диаграммы Ганта и т.п.





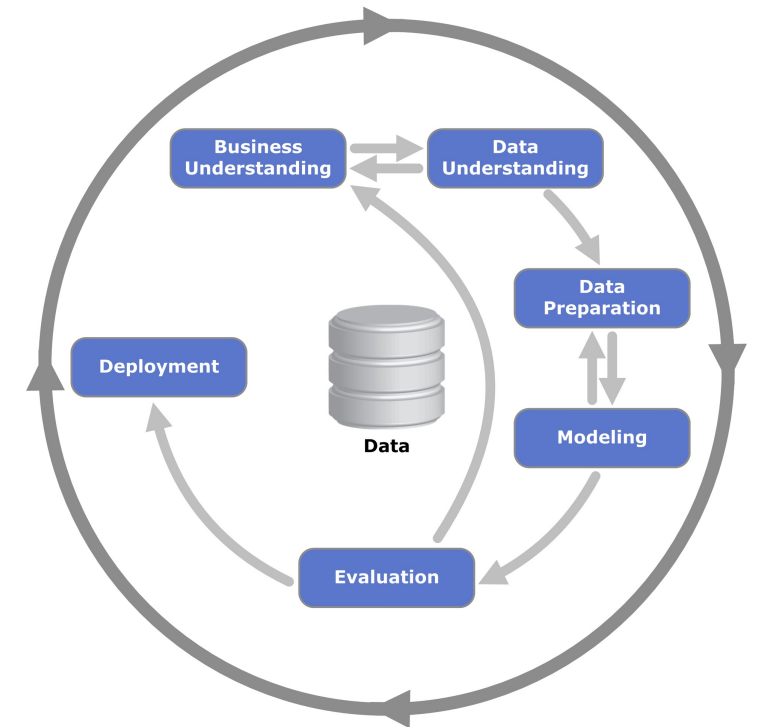
# CRISP-DM: начальное изучение данных

- **Цель:** понять сильные и слабые стороны представленных данных
- **Инструменты:** описательные статистики, графики
- Сбор данных
- Описание данных
- Исследование данных
- Качество данных



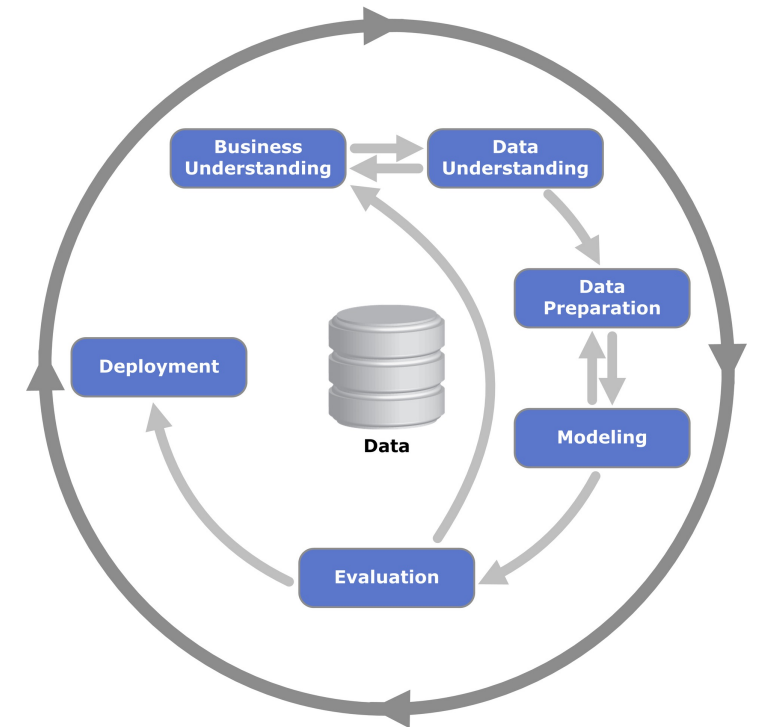
# CRISP-DM: подготовка данных

- Отбор данных
- Очистка данных
- Генерация данных
- Интеграция данных
- Форматирование данных



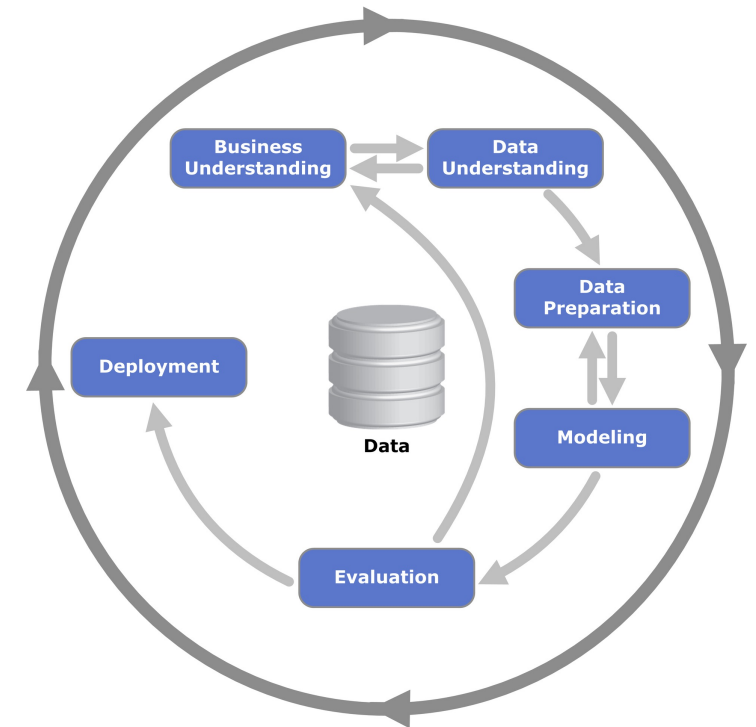
# CRISP-DM: моделирование

- Выбор моделей
- Выбор стратегии валидации
- Обучение
- Оценивание результатов



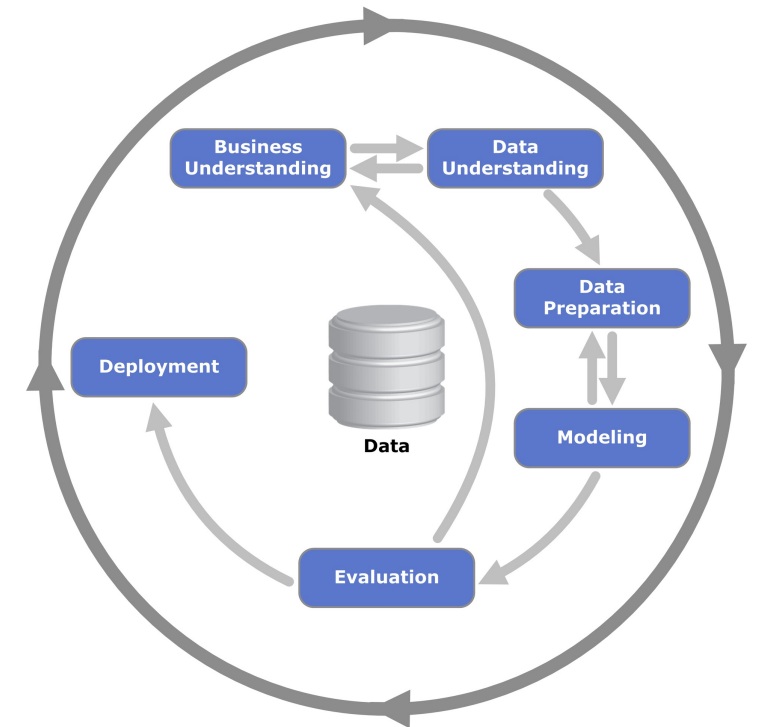
# CRISP-DM: оценка результатов моделирования

- Оценка результата моделирования
- Оценка процесса работы
- Определение следующих шагов



# CRISP-DM: внедрение

- Внедрение
- Планирование мониторинга и поддержки
- Подготовка отчетов
- Ревью проекта



# Вводное занятие

1. Организационное
2. Про ритейл, цели и данные
3. Задачи
4. CRISP-DM
- 5. ML System Design**

# ML System Design

- комплексный многоэтапный процесс проектирования, внедрения и поддержки систем на основе машинного обучения; включает в себя комбинацию подходов и навыков из разных областей и ролей, включая машинное обучение, глубинное обучение, разработка программного обеспечения, управления проектами, управления продуктом лидерства

*Machine Learning System Design*, Arseny Kravchenko, Valerii Babushkin

# Содержание документа MLSD

## Версия AWS

1. Обзор документа:  
3-5 предложений о сути документа
2. Мотивация:  
зачем делаем проект
3. Метрики успеха
4. Требования и ограничения
5. Методология:
  1. Постановка проблемы
  2. Данные
  3. Технологии
  4. Эксперименты и валидация
6. Имплементация:
  1. архитектура решения – общая схема
  2. Инфраструктура
  3. Нагрузка на систему
  4. Безопасность
  5. Приватность данных
  6. Мониторинги и системы предупреждения
  7. Стоимость системы
  8. Точки интеграции
  9. Риски и неопределенность
7. Приложения
  - Риски, альтернативы, дедлайны и тп



# Содержание документа MLSD

## Версия Reliable ML

### 1. Цели и предпосылки

1. Зачем идти в разработку продукта
2. Бизнес-требования и ограничения
3. Что входит в скоуп проекта/итерации
4. Предпосылки решения

### 2. Методология

1. Постановка задачи
2. Блок-схема решения
3. Этапы решения задачи

### 3. Подготовка пилота

1. Способы оценки пилота
2. Что считаем успешным пилотом
3. Подготовка пилота

### 4. Внедрение

1. архитектура решения – общая схема
2. Инфраструктура
3. Нагрузка на систему
4. Безопасность
5. Приватность данных
6. Мониторинги и системы предупреждения
7. Стоимость системы
8. Точки интеграции
9. Риски и неопределенность

# Содержание документа MLSD

## Версия Reliable ML

- Почитаем один из них:
- [Прогноз спроса в FMCG](#)