цифра

Организация работы с данными в бизнесе: от А до Я

Александр
Белугин

Директор AI проектов

Цифра – это

- Мониторинг оборудования и сбор данных
- Виртуальные сенсоры
- Материальный баланс
- Управление снабжением
- Предсказание спроса
- Автоматическое управление
- Беспилотное оборудование
- …

Переход от цифрового управления отдельной установкой к управлению всем производственным процессом

цифра

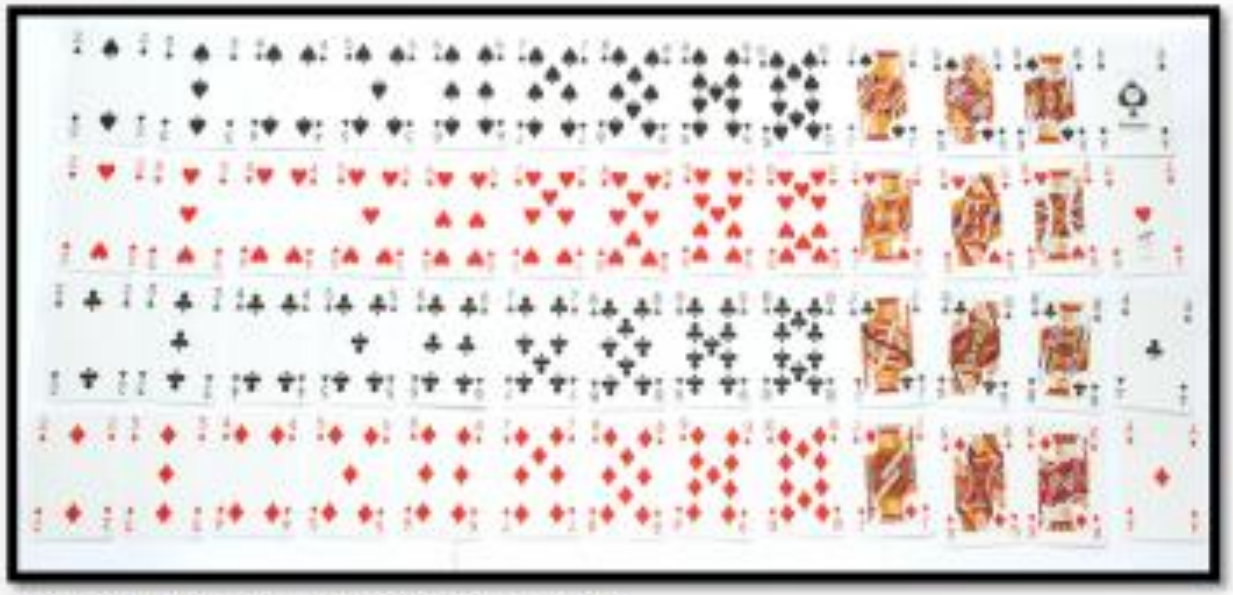
Junior DS

- Доказать, что данные нужны
- Да, все данные
- Ждать доступа
- Получить 18000 Excel файлов

Relational algebra refresher – Cartesian Product

$$R \times S$$

Ranks = {A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, 2}
Suits = {♠, ♥, ♦, ♣}
Ranks × Suits = {
(A, ♠), (A, ♥), (A, ♦), (A, ♣),
(K, ♠), ...,
(2, ♠), (2, ♥), (2, ♦), (2, ♣)
}



By Trainler - Own work, CC BY 3.0,
https://commons.wikimedia.org/w/index.php?curid=7104281

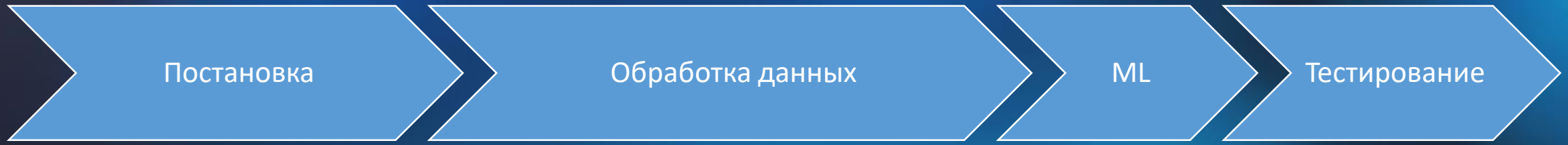
Middle DS

- Предсказывать отток клиентов!
- Предсказывать спрос на товары!
- Выдавать рекомендации на товары!

Senior DS

Пошёл к заказчику и узнал…

- Заказчик не знает, что такое MAPE
- Ему можно объяснить, но ему всё равно
- Нужно решение, а не ML



цифра

Chief DS

Сбор идей

- Сверху – от KPI
- Снизу – от возможностей

Технически реализуемые и бизнес-обоснованные

цифра

Сбор идей и требований

- Мозговой штурм
- Бенчмаркинг
- Функциональная декомпозиция
- Анализ документов
- Интервью
- Наблюдение

цифра

Про каждую идею…

- Владелец
- Спонсор
- Организационный объем
- Бизнес-метрика
  - Текущее значение
  - Целевое значение

цифра

Про каждую идею…

- Экономический эффект
- Длительность и стоимость внедрения
- Простота внедрения
- Наличие и качество данных
- Релевантный опыт
- Масштабируемость

Эффект, млн. руб.	Срок, мес.	Затраты на разработку	Затраты на данные	Уверенность в успехе	Простота внедрения	Наличие и качество данных	Релевантный опыт	Наследственность	Мета оценка
161	9	53	6	2	2	2	0	2	38%
382	8	110	22	3	4	4	0	0	57%

Регулярный процесс

- Идеи собираются
- Готовые идеи ранжируются
- Топ ранжированных идей бюджетируется
- Накапливаются извлеченные уроки

цифра

CDO

Обоснование инвестиций

- Поддерживающие инициативы не несут бизнес-ценности
- Но влияют на скорость и стоимость бизнес-инициатив
- Их нужно обосновать и делать наперёд

эффект

Реестр бизнес-инициатив

Реестр наборов данных

Реестр информационных систем

Реестр поддерживающих инициатив

цифра

связанные затраты

Процессы управления данными

- Data Governance – planning, supervision and control over data management and use
- Data Architecture Management – as an integral part of the enterprise architecture
- Data Development – analysis, design, building, testing, deployment and maintenance
- Database Operations Management – support for structured physical data assets

цифра

Процессы управления данными

- Data Security Management – ensuring privacy, confidentiality and appropriate access
- Reference & Master Data Management – managing golden versions and replicas
- Data Warehousing & Business Intelligence Management – enabling access to decision support data for reporting and analysis

цифра

Процессы управления данными

- Document & Content Management – storing, protecting, indexing and enabling access to data found in unstructured sources (electronic files and physical records)
- Meta Data Management – integrating, controlling and delivering meta data
- Data Quality Management – defining, monitoring and improving data quality

цифра

Исходные данные

Исходные данные

Исходные данные

Исходные данные

- Генерируются датчиками
- Вносятся вручную
- Дорогое, устаревшее и разнородное хранение

Все или почти все исходные данные

Предобработанные наборы данных

- Единое дешевое хранилище
- Богатый набор интеграционных возможностей
- Архив данных и запросы к нему
- Предобработанные наборы готовы для доступа аналитиков и подрядчиков

Пользовательская песочница

Пользовательская песочница

Пользовательская песочница

- Индивидуальный доступ
- Пространство для собственных версий данных

Качество данных – 1

- Полнота
- Уникальность
- Дискретность
- Прослеживаемость
  - по времени
  - по логистике

цифра

Качество данных – 2

- Оригинальность
- Чистота
- Наличие пропусков
- Использование справочников
- Документация

цифра

Выводы

- Офис управления данными
- Своя компетенция
- Метод отбора инициатив
- Умение быстро пилотировать
  - Доступность данных
  - Четкая метрика успеха

цифра

hello@zyfra.ai