

Data Understanding. Excel

Пётр Емельянов

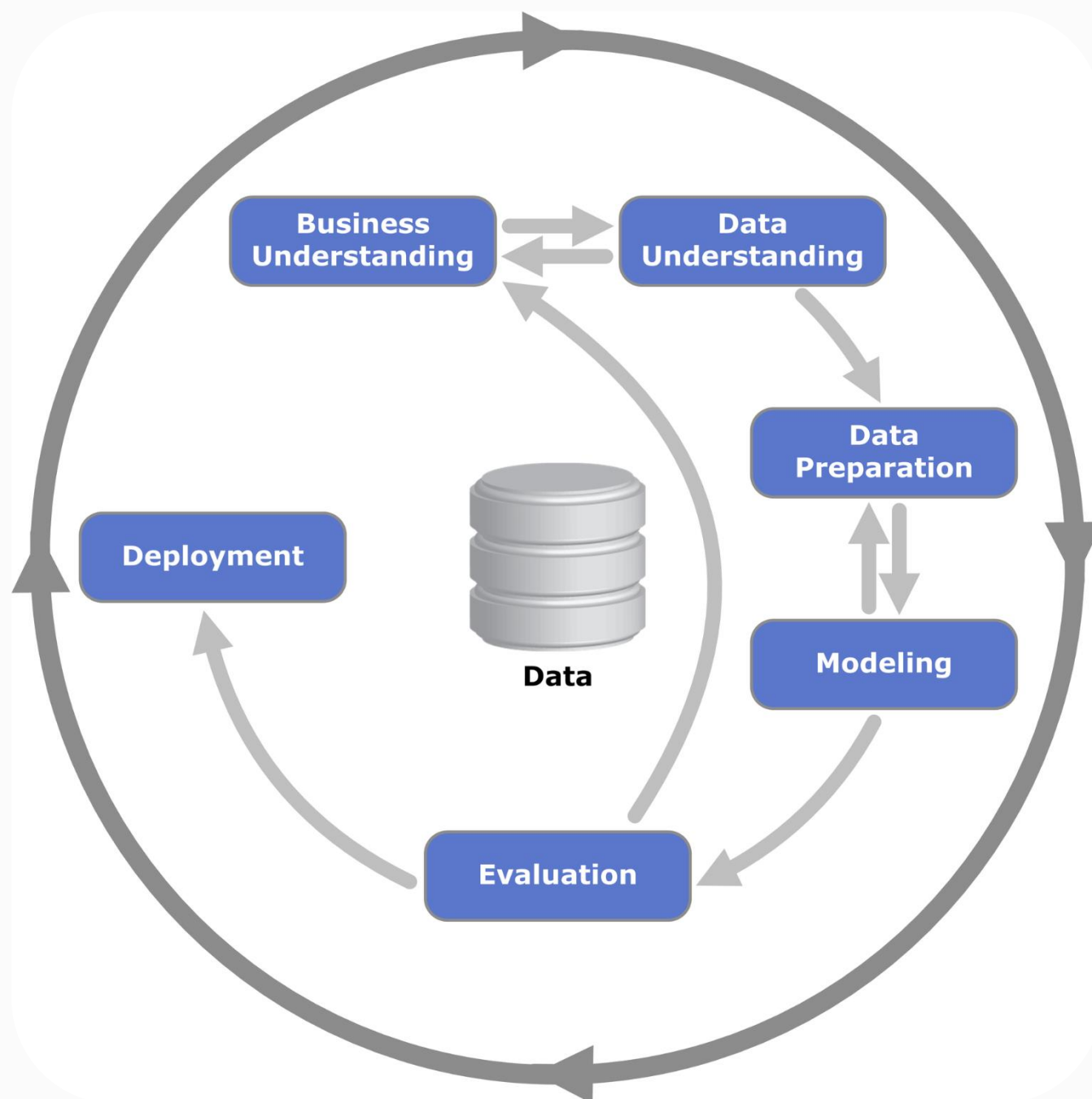
R&D Директор в Ubic Technologies

Skillbox

Что такое Data Understanding?

Что такое Data Understanding?

Data Understanding



Что такое Data Understanding?

Артефакты Business Understanding

- 1 Бизнес-процесс
- 2 Бизнес-цель и критерии успеха
- 3 Задача по анализу данных

Что такое Data Understanding?

“

Дайте мне начальные
данные частиц всего
мира, и я предскажу
вам будущее мира.

Пьер-Симон Лаплас

Французский математик, механик,
физик и астроном

Что такое Data Understanding?

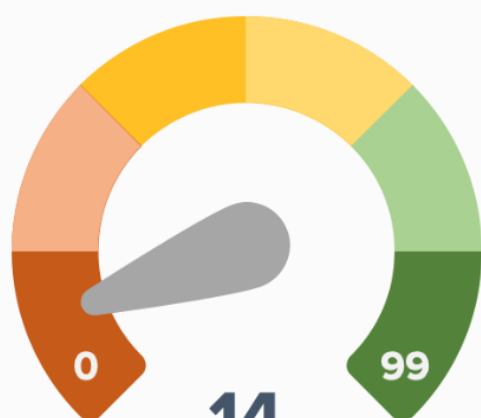
Data Understanding

- 1 Собираем данные
- 2 Описываем данные
- 3 Исследуем данные
- 4 Проверяем качество данных

Что такое Data Understanding?

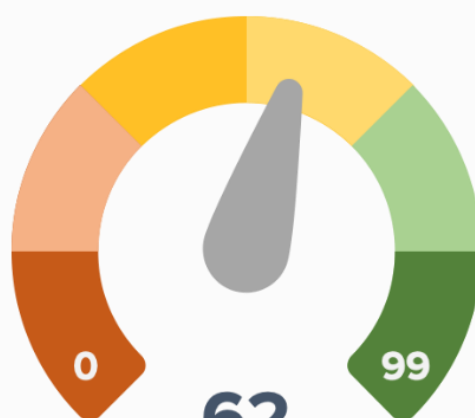
Оценка кредитного риска

Отказать в кредите



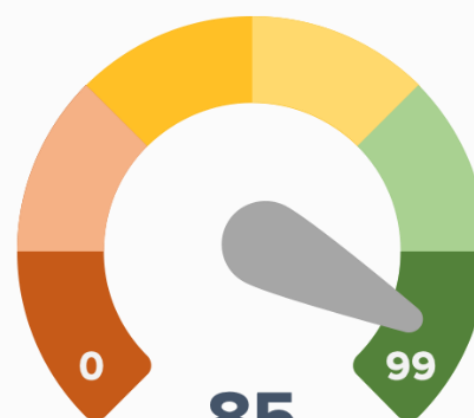
БЕЗОТВЕТСТВЕННЫЙ

Подумать еще



СРЕДНИЙ

Выдать кредит



ОТВЕТСТВЕННЫЙ

Собрать данные означает
изучить **все источники**
данных, которыми мы
располагаем в рамках
текущей задачи.

Что такое Data Understanding?

Источники данных банка

АБС



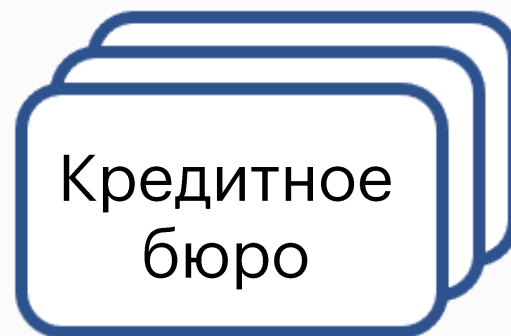
Данные
партнёров



Открытые
источники



Кредитное
бюро



Источники данных банка

Источник данных	Тип источника
Автоматизированная банковская система	Собственные (1rd) данные
Данные партнёров	Сторонние (3rd) данные
Данные кредитного бюро	Сторонние (3rd) данные
Статистика с портала открытых данных	Открытые данные

Данные в виде таблицы

Возраст	Профессия	Семейное положение	Образование	Была ли просрочка
58	Менеджер	Женат	Высшее	Нет
44	Слесарь	Холост	Среднее полное	Нет
33	Предприниматель	Женат	Среднее полное	Нет
47	Разнорабочий	Женат	Неизвестно	Нет
33	Неизвестно	Холост	Неизвестно	Нет
35	Менеджер	Женат	Высшее	Нет
28	Менеджер	Холост	Высшее	Нет
42	Предприниматель	Разведён	Высшее	Да
58	Пенсионер	Женат	Среднее общее	Нет
...

Что такое Data Understanding?

ТАБЛИЦА УМНОЖЕНИЯ

2 X 2 = 4
2 X 3 = 6
2 X 4 = 8
2 X 5 = 10
2 X 6 = 12
2 X 7 = 14
2 X 8 = 16
2 X 9 = 18
2 X 10 = 20

3 X 2 = 6
3 X 3 = 9
3 X 4 = 12
3 X 5 = 15
3 X 6 = 18
3 X 7 = 21
3 X 8 = 24
3 X 9 = 27
3 X 10 = 30

4 X 2 = 8
4 X 3 = 12
4 X 4 = 16
4 X 5 = 20
4 X 6 = 24
4 X 7 = 28
4 X 8 = 32
4 X 9 = 36
4 X 10 = 40

5 X 2 = 10
5 X 3 = 15
5 X 4 = 20
5 X 5 = 25
5 X 6 = 30
5 X 7 = 35
5 X 8 = 40
5 X 9 = 45
5 X 10 = 50

6 X 2 = 12
6 X 3 = 18
6 X 4 = 24
6 X 5 = 30
6 X 6 = 36
6 X 7 = 42
6 X 8 = 48
6 X 9 = 54
6 X 10 = 60

7 X 2 = 14
7 X 3 = 21
7 X 4 = 28
7 X 5 = 35
7 X 6 = 42
7 X 7 = 49
7 X 8 = 56
7 X 9 = 63
7 X 10 = 70

8 X 2 = 16
8 X 3 = 24
8 X 4 = 32
8 X 5 = 40
8 X 6 = 48
8 X 7 = 56
8 X 8 = 64
8 X 9 = 72
8 X 10 = 80

9 X 2 = 18
9 X 3 = 27
9 X 4 = 36
9 X 5 = 45
9 X 6 = 54
9 X 7 = 63
9 X 8 = 72
9 X 9 = 81
9 X 10 = 90

Что такое Data Understanding?

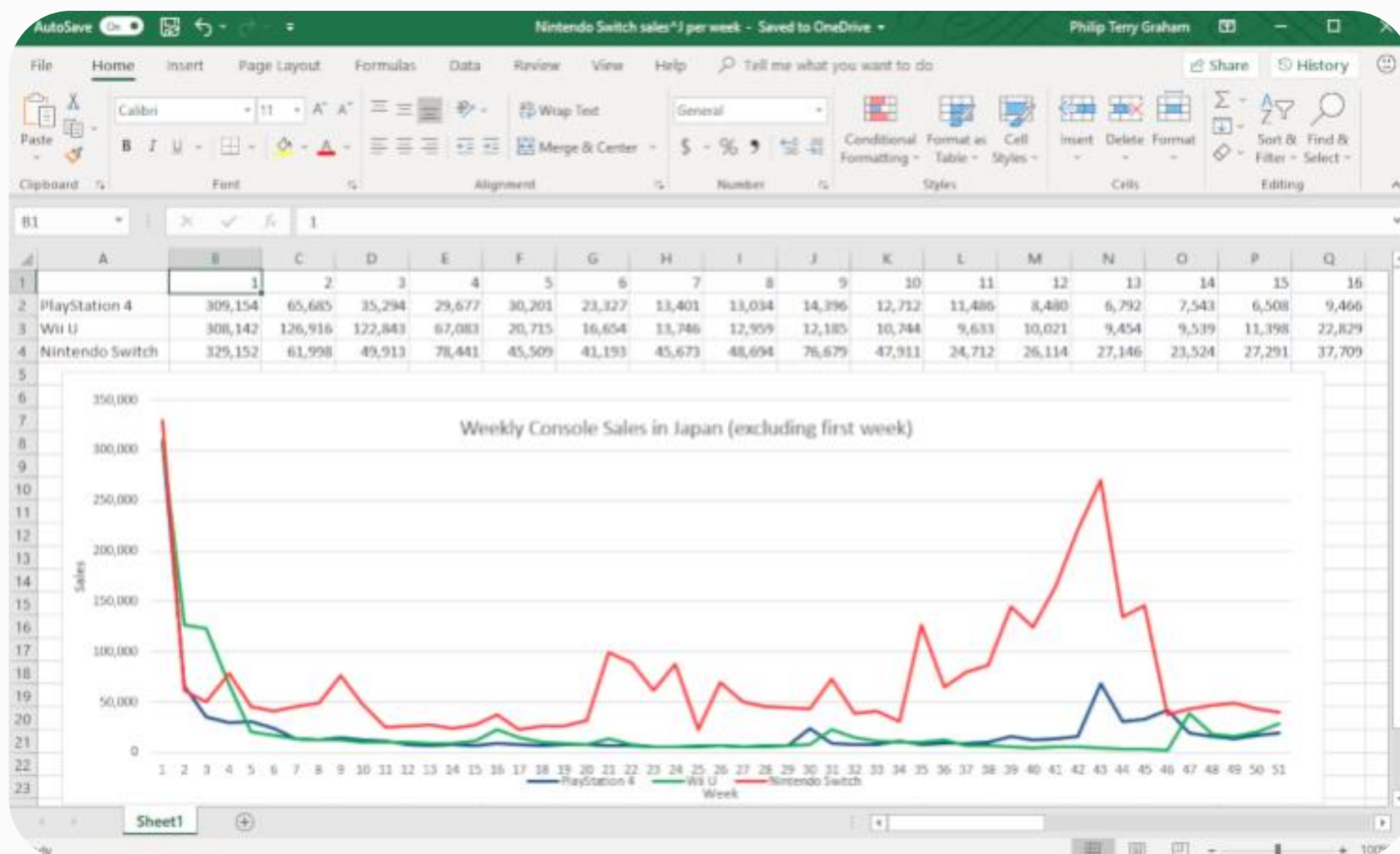
Свои собственные
данные банк хранит
в базе данных.

Реляционная база
данных — это способ
организации таблиц
и SQL — язык запросов
к данным.

Файл —
последовательный набор
данных. У файла есть имя
и формат.

Что такое Data Understanding?

XLSX-файл в родной среде



API — это набор правил,
следуя которым одна
программа может
взаимодействовать
с другой.

Источники данных банка

Источник данных	Тип источника	Метод доступа
Автоматизированная банковская система	Собственные (1rd) данные	База данных, SQL
Данные партнёров	Сторонние (3rd) данные	Выгрузки в виде csv-файлов
Данные кредитного бюро	Сторонние (3rd) данные	Доступ по API
Статистика с портала открытых данных	Открытые данные	Выгрузки в виде xlsx-файлов

Описательная
статистика — раздел
статистики, который
изучает методы
представления основных
свойств данных.