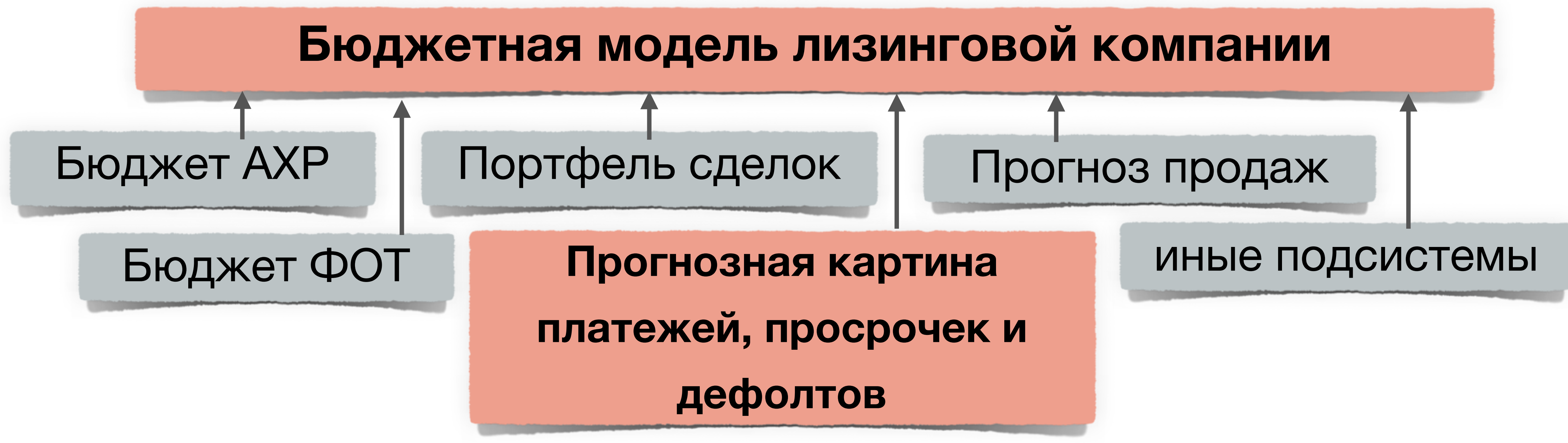


РЕАЛИЗАЦИЯ АЛГОРИТМА  
ПОСТРОЕНИЯ КАРТИНЫ  
ДЕФОЛТОВ С ИСПОЛЬЗОВАНИЕМ  
ЯЗЫКА PYTHON

Серкин Александр

# Описание предметной области



UID	01.2018	02.2018	03.2018	04.2018	05.2018	01.2018	02.2018	03.2018	04.2018	05.2018
UID1	1200	1200	1200	1200	1200	... без просрочки	без просрочки	1-30	31-60	дефолт
UID2	1100	1100	1100	1100	1100	...	1-30	без просрочки	1-30	без просрочки
...	...	...	...	...	...	...	...	...	...	...
UID75670	700	700	700	700	700	...	1-30	31-60	дефолт	дефолт
UID75671	600	600	600	600	600	...	без просрочки	без просрочки	без просрочки	без просрочки

# Формулирование проблемы

---

Имеющийся в компании алгоритм построения картины дефолтов реализован на базе решения «1-С Предприятие».

Время исполнения алгоритма - 4-5 часов.

---

**Цель:** разработать инструмент для оперативной генерации картины дефолтов

---

## Задачи:

- реализовать модель средствами VBA Excel
- адаптировать модель с использованием векторных вычислений Python
- сравнить полученные результаты, выбрать модель по сочетанию «скорость обработки» - «соответствие заданным параметрам»
- по возможности доработать выбранную модель, повысив скорость и/или точность обработки

Подготовительный

Основной

Завершающий

# Итоговая абстрактная последовательность реализации:

## Подготовительный этап

1. задать параметры логгирования;
2. прочитать параметры прогнозирования, в случае отсутствия каких-либо параметров, задать их;
3. прочитать данные для обработки, проверить валидность;

**set logging params**

**set picture params**

**get data**

**proceeding**

# Итоговая абстрактная последовательность реализации:

## Основной этап

4. в цикле от первого месяца горизонта планирования до последнего:
  - 4.1. для каждого вида рейтинга повторить назначение рейтинга клиенту до момента удовлетворения условию, заполняя рейтинг в столбец «рейтинг»
  - 4.2. добавить столбец для отражения бакета
  - 4.3. для каждого статуса платежа выполнить подбор бакета, анализируя значения предыдущего и заполняя значения в столбце текущего
  - 4.4. подвести итоговые доли распределения рейтингов в общей сумме платежей
  - 4.5. подвести итоговые доли распределения платежей по бакетам в общей сумме платежей

Подготовительный

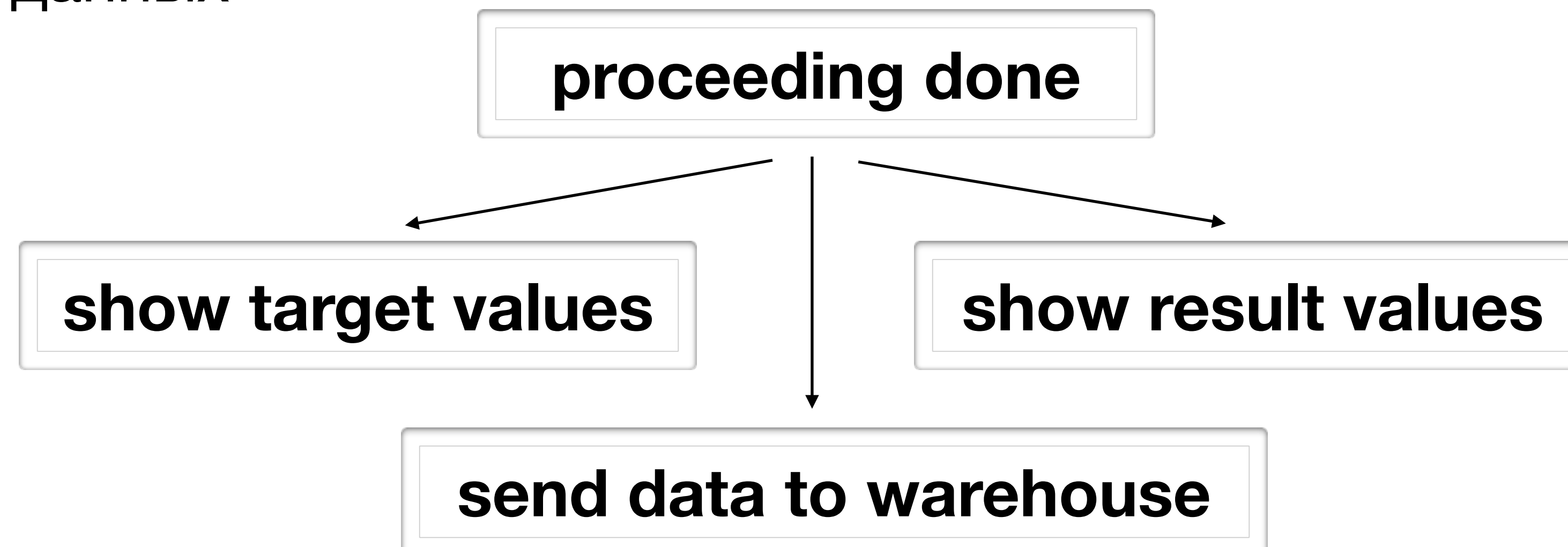
Основной

Завершающий

# Итоговая абстрактная последовательность реализации:

## Завершающий этап

5. вывести целевые параметры
6. вывести результирующую статистику картины дефолтов
7. передать обработанный результат для загрузки в базу данных



# Реализация модели средствами VBA Excel

Доли рейтингов по годам в горизонте 5 лет

Рейтинг клиента	2018	2018	2019	2019	2020	2020	2021	2021	2022	2022
	VBA	цель	VBA	цель	VBA	цель	VBA	цель	VBA	цель
высокий	<b>53%</b>	52%	50%	50%	46%	46%	<b>41%</b>	40%	<b>35%</b>	34%
средний	<b>27%</b>	26%	<b>28%</b>	27%	<b>31%</b>	30%	34%	34%	<b>36%</b>	35%
низкий	<b>21%</b>	20%	21%	21%	22%	22%	24%	24%	25%	25%
без ФА	<b>0%</b>	2%	<b>1%</b>	2%	<b>1%</b>	2%	<b>1%</b>	2%	<b>4%</b>	6%

**Среднее время  
исполнения  
алгоритма:  
от 1 до 1,5  
часов**

Доли бакетов по годам в горизонте 5 лет

Бакет	2018	2018	2019	2019	2020	2020	2021	2021	2022	2022
	VBA	цель	VBA	цель	VBA	цель	VBA	цель	VBA	цель
без про-срочки	89%	93%	87%	90%	86%	89%	83%	88%	82%	85%
1-30	7%	5%	8%	6%	8%	6%	9%	6%	9%	7%
31-60	2%	1%	3%	2%	4%	2%	4%	2%	4%	2%
дефолт	2%	1%	3%	2%	4%	3%	4%	4%	5%	6%

**MS Excel на  
рабочей  
станции  
аналитика  
недоступен на  
все время  
исполнения**

# Адаптация с использованием средств векторных вычислений Pandas

Доли рейтингов по годам в горизонте 5 лет

Рейтинг клиента	2018	2018	2019	2019	2020	2020	2021	2021	2022	2022
	pd	цель	pd	цель	pd	цель	pd	цель	pd	цель
высокий	52%	52%	<b>49%</b>	50%	<b>45%</b>	46%	40%	40%	<b>35%</b>	34%
средний	<b>25%</b>	26%	<b>26%</b>	27%	<b>28%</b>	30%	<b>31%</b>	34%	<b>33%</b>	35%
низкий	20%	20%	21%	21%	<b>21%</b>	22%	<b>23%</b>	24%	<b>24%</b>	25%
без ФА	<b>3%</b>	2%	<b>4%</b>	2%	<b>5%</b>	2%	<b>7%</b>	2%	<b>8%</b>	6%

Доли бакетов по годам в горизонте 5 лет

Бакет	2018	2018	2019	2019	2020	2020	2021	2021	2022	2022
	pd	цель	pd	цель	pd	цель	pd	цель	pd	цель
без просрочки	91%	93%	90%	90%	88%	89%	87%	88%	83%	85%
1-30	6%	5%	5%	6%	6%	6%	6%	6%	7%	7%
31-60	2%	1%	2%	2%	2%	2%	2%	2%	2%	2%
дефолт	1%	1%	3%	2%	4%	3%	4%	4%	7%	6%

Среднее время исполнения алгоритма:  
от 15 до 20 минут

Аналитик может продолжать работу в MS Excel параллельно с вычислениями

Требуется установка дополнительного ПО и обучение аналитика

# Выбор модели по сочетанию «время обработки» - «соответствие заданным параметрам»

## Модель VBA:

- время обработки 1-1,5 часов
- результаты отличаются на 1-2 п.п.

Выявленные недостатки:

- MS Excel на рабочей станции аналитика недоступен на все время исполнения

## Модель Pandas:

- время обработки 15-20 минут
- результаты отличаются на 1-3 п.п.

Выявленные недостатки:

- требуется установка отдельного программного обеспечения Python, обучение аналитика работе



**Выбор - модель Pandas:**  
скорость исполнения в 3 раза выше, чем модель VBA  
высокая точность генерируемых результатов

# Доработка выбранной модели с использованием средств Pandas

## Целевые параметры улучшения:

- увеличить скорость подбора
- сохранить точность подбора (соответствия заданным ожиданиям)

## Реализовано:

- 1) время исполнения сокращено до диапазона 2-3 минуты за счет:
  - вынести цикл подбора параметров в отдельную процедуру
  - не пересчитывать на каждом шаге все статистики, а производить вычисления суммирующих статистик по клиенту, добавлять к ранее вычисленным значениям
- 2) время исполнения сокращено до диапазона 1:02 - 1:06 минуты за счет:
  - присваивать в первую очередь рейтинги с наименьшей долей

**Время генерации картины дефолтов сокращено до 1 минуты с исходных 5 часов**

# Выводы

- 1) имеющийся в компании алгоритм генерации картины дефолтов реализован в модели на VBA Excel
- 2) разработанная модель адаптирована с использованием векторных вычислений в языке Python с использованием библиотеки Pandas
- 3) из двух моделей выбрана реализация на Python
- 4) реализация на Python доработана с учетом особенностей векторных вычислений и не загружает рабочую станцию аналитика

## Цель достигнута:

разработан инструмент для оперативной генерации картины дефолтов

### Дополнительный эффект:

отдельные методы, разработанные в модели Pandas, будут использоваться в дальнейшей работе и уже легли в основу , например, сегментации клиентской базы в прогнозируемом периоде в заданных параметрах

**Спасибо за внимание!**