



# Text2SQL



**Миша**

МИПТ + Skoltech  
Работаю в R77  
Живу в Испании

# Программа

01 Что за задача и для чего это надо?

02 Какая выглядит архитектура?

03 Что по оценке качества?

04 Переходим от Text2SQL к агенту?

05 Some tips and tricks?

06 Как защититься от ошибок?

07 Как работает?

08 Куда дальше?

# Задача



**Text2SQL** — это задача преобразования текста на естественном языке в SQL-запрос, соответствующий намерению пользователя. Цель — автоматически сгенерировать корректный SQL-запрос по описанию, не требующему знаний программирования.

# Зачем надо

## Заменить Аналитиков Данных

- Автоматизация рутинной работы.
- Снижение нагрузки на команду.
- Ускорение создания отчетов и дашбордов.
- Помощь с выводами.

# Зачем надо

## Заменить Аналитиков Данных

- Автоматизация рутинной работы.
- Снижение нагрузки на команду.
- Ускорение создания отчетов и дашбордов.
- Помощь с выводами.

## Предоставление данных пользователям

- Аналитика.
- Поддержка.
- Отчеты.

# Зачем надо

## Заменить Аналитиков Данных

- Автоматизация рутинной работы.
- Снижение нагрузки на команду.
- Ускорение создания отчетов и дашбордов.
- Помощь с выводами.

## Предоставление данных пользователям

- Аналитика.
- Поддержка.
- Отчеты.

## Доступ к данным для нетех. сотрудников

- менеджеры, маркетологи (~~и прочая нечисть~~) не тратят ваше время

# Зачем надо

## Заменить Аналитиков Данных

- Автоматизация рутинной работы.
- Снижение нагрузки на команду.
- Ускорение создания отчетов и дашбордов.
- Помощь с выводами.

## Предоставление данных пользователям

- Аналитика.
- Поддержка.
- Отчеты.

## Доступ к данным для нетех. сотрудников

- менеджеры, маркетологи (~~и прочая нечисть~~) не тратят ваше время



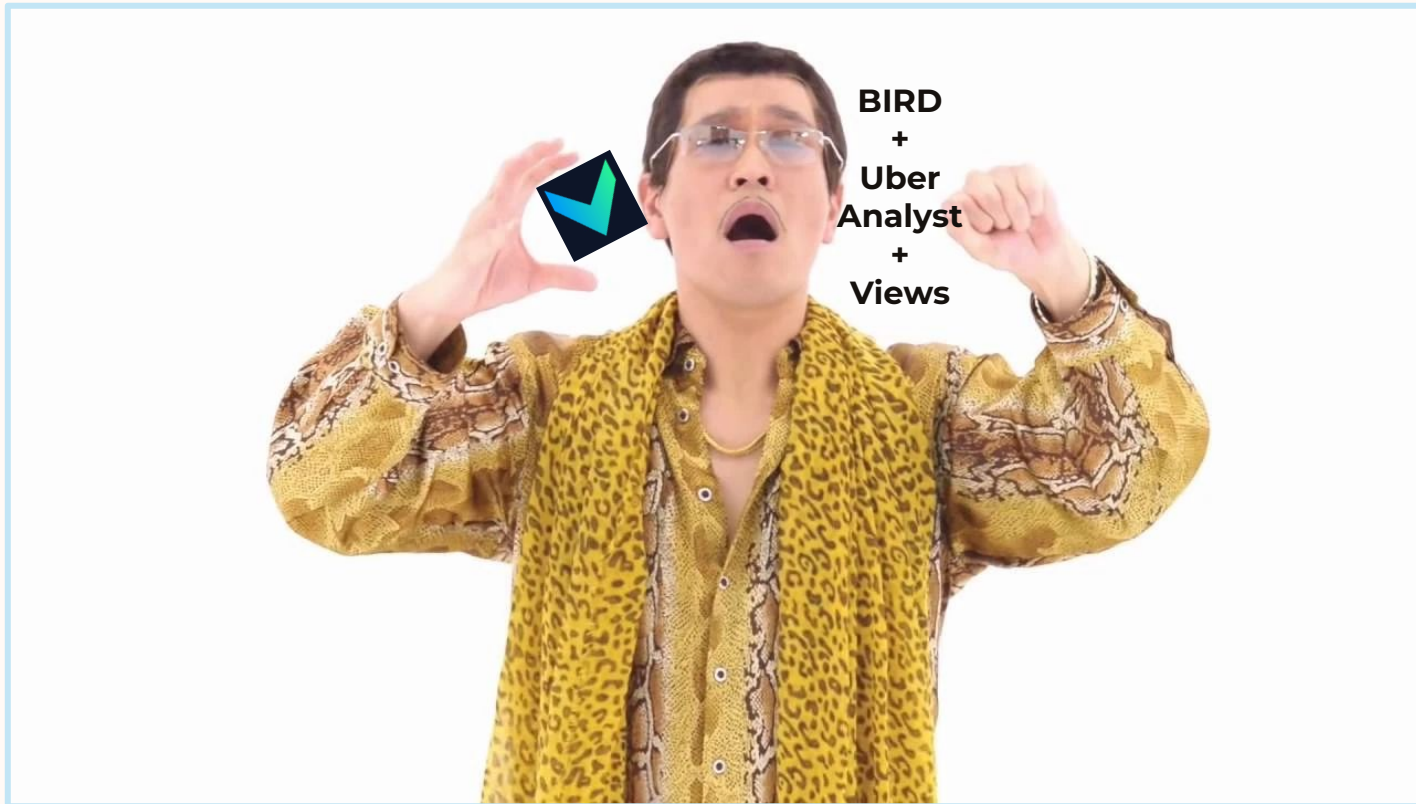
A woman with short dark hair, wearing a red textured sweater and blue jeans, is smiling while working on a laptop at a dark wooden standing desk. The desk is in a bright, modern office space with large windows on the right. On the desk, there is a laptop, a green mouse, a white mug, and some papers. In the background, there is a wooden ladder, a wire grid with papers, a potted plant, and a wooden shelf with books and a plant. A white text box is overlaid at the bottom of the image.

Аналитик, которому не надо писать SQL



Сборка решения

# Архитектура



**DDL** ~ CREATE TABLE

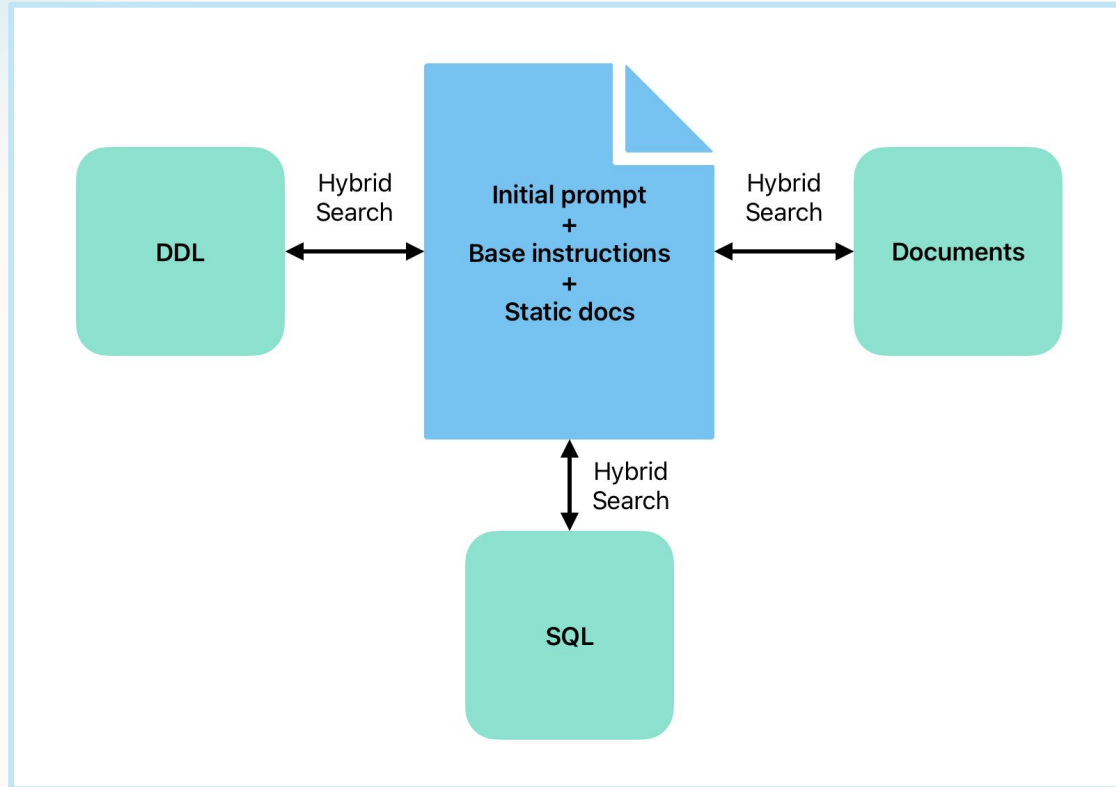
**Documents** - описание колонок таблиц, определения, правила выдачи, методы расчета

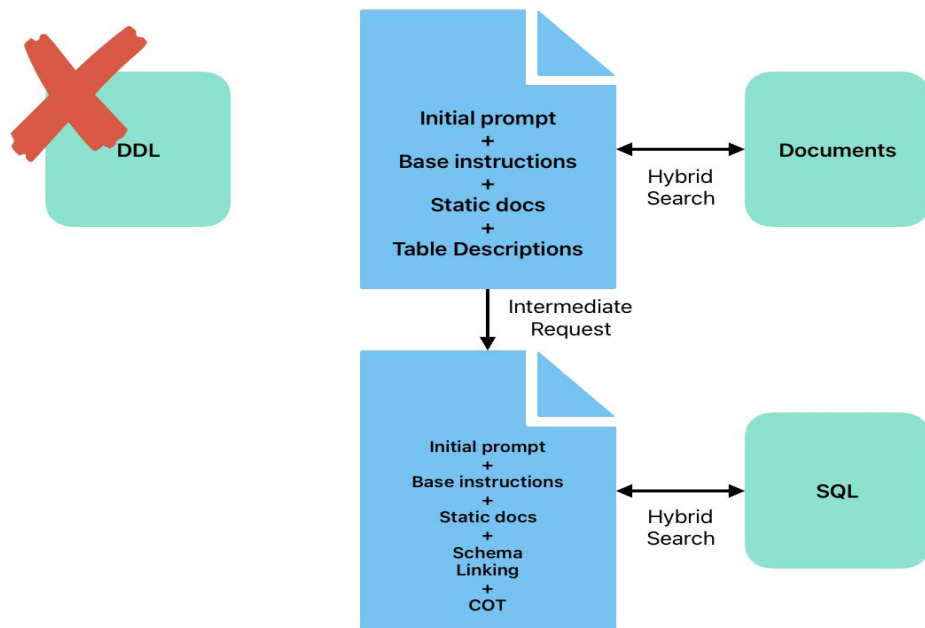
**SQL** - пары вопрос-sql

**Initial instructions** - инструкции, специфичные для проекта, описания таблиц.

**Base instructions** - общие инструкции по генерации SQL

**Static docs** - самые важные документы, которые идут в промпт всегда





**DDL** - имеет смысл, когда много таблиц

**Table Descriptions** ~  
DESCRIBE database.table

**Schema linking** - связь запроса пользователя и названий таблиц/колонок

**COT** - Chain of Thoughts

01. Как облегчить  
задачу модели?

- Сырые таблицы - фуу 👎  
Агрегаты под разные задачи - воо 👍

01. Как облегчить задачу модели?

- Сырые таблицы - фуу 👎  
Агрегаты под разные задачи - воо 👍

02. Сколько у вас таблиц?

- 8

01. Как облегчить задачу модели?

- Сырые таблицы - фуу 👎  
Агрегаты под разные задачи - воо 👍

02. Сколько у вас таблиц?

- 8

03. Откуда взять первые примеры вопрос-sql?

- Синтетика + проверка экспертом



01. Как облегчить задачу модели?

- Сырые таблицы - фуу 👎  
Агрегаты под разные задачи - воо 👍

02. Сколько у вас таблиц?

- 8

03. Откуда взять первые примеры вопрос-sql?

- Синтетика + проверка экспертом

04. Что самое важное в подготовке данных?

- Консистентность (таблицы + примеры)



Оценка качества

# Метрики

1. **EX** - точное совпадение результатов выполнения скриптов.

1. **EX** - точное совпадение результатов выполнения скриптов.
2. **R-VES - EX** с наградой, которая учитывает время выполнения скриптов.

1. **EX** - точное совпадение результатов выполнения скриптов.
2. **R-VES - EX** с наградой, которая учитывает время выполнения скриптов.
3. **Soft F1-Score** - считаем кол-во соответствий по строкам -> считаем **FP, FN, TP**

1. **EX** - точное совпадение результатов выполнения скриптов.
2. **R-VES - EX** с наградой, которая учитывает время выполнения скриптов.
3. **Soft F1-Score** - считаем кол-во соответствий по строкам -> считаем **FP, FN, TP**
4. **Soft-Accuracy** - кастомная метрика. Метрика не штрафует за порядок колонок, не штрафует за лишние колонки, но штрафует за лишние строки и неправильную сортировку.

$$acc_{soft} = \frac{tp}{\max(n_{rows}^{pred}, n_{rows}^{gt}) \cdot \min(n_{cols}^{pred}, n_{cols}^{gt})}$$

1. **Размеченный экспертами датасет.** Датасет содержит пары вопрос-sql на разные темы, интересующие заказчика.

1. **Размеченный экспертами датасет.** Датасет содержит пары вопрос-sql на разные темы, интересующие заказчика.
2. **BIRD mini-dev.**

### Difficulty Distribution

- Simple: 30%
- Moderate: 50%
- Challenging: 20%

### Database Distribution

- Debit Card Specializing: 30 instances
- Student Club: 48 instances
- Thrombosis Prediction: 50 instances
- European Football 2: 51 instances
- Formula 1: 66 instances
- Superhero: 52 instances
- Codebase Community: 49 instances
- Card Games: 52 instances
- Toxicology: 40 instances
- California Schools: 30 instances
- Financial: 32 instances

Описание датасета BIRD



1. **Размеченный экспертами датасет.** Датасет содержит пары вопрос-sql на разные темы, интересующие заказчика.
2. **BIRD mini-dev.**

Model	EX	R-VES	Soft F1	Soft Acc
gpt-4.1	25.0	23.66	74.96	84.3

Результаты на экспертном датасете

## Difficulty Distribution

- Simple: 30%
- Moderate: 50%
- Challenging: 20%

## Database Distribution

- Debit Card Specializing: 30 instances
- Student Club: 48 instances
- Thrombosis Prediction: 50 instances
- European Football 2: 51 instances
- Formula 1: 66 instances
- Superhero: 52 instances
- Codebase Community: 49 instances
- Card Games: 52 instances
- Toxicology: 40 instances
- California Schools: 30 instances
- Financial: 32 instances

Описание датасета BIRD



Собираем агента

Агент 

Зачем нужен агент?

# Зачем нужен агент?

The screenshot shows a chat interface with a light blue border. At the top, a user message (red icon) asks "как дела?". Below it, a system message (S icon) displays a SQL query: `1 SELECT 'Все хорошо!' AS response;`. This is followed by a table with two columns: "response" and a value "0". The next system message (S icon) says "Все хорошо!". At the bottom, there is a text input field with the placeholder "Задайте мне вопрос про ваши данные" and a right-pointing arrow button.

response
0

# Зачем нужен агент?

Не всегда нужно ходить в базу данных для ответа на вопрос пользователя.

как дела?

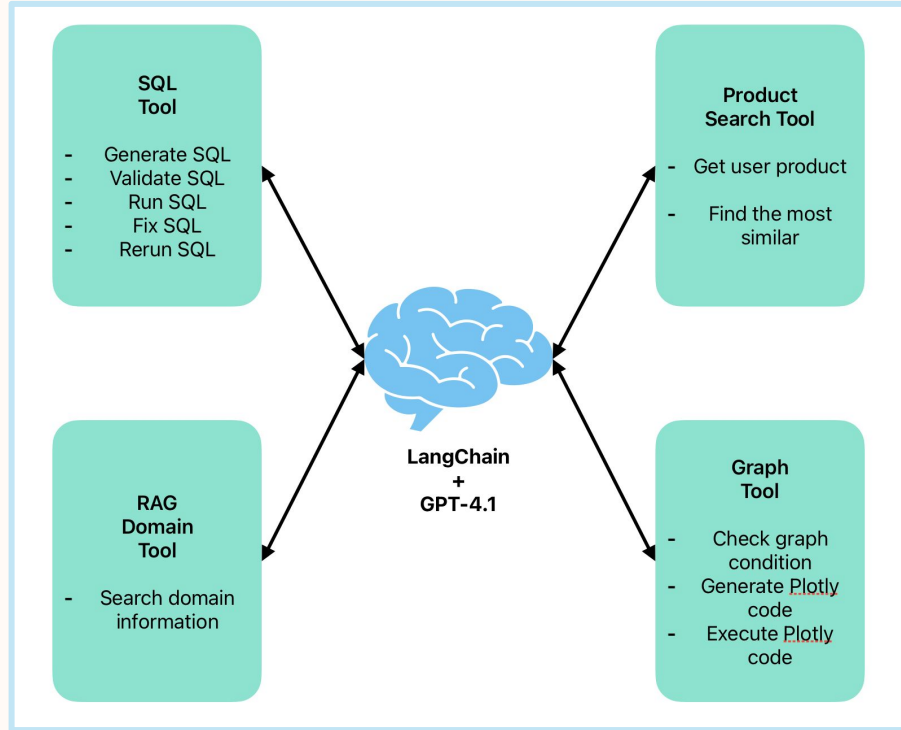
**S** `1 SELECT 'Все хорошо!' AS response;`

	response
0	Все хорошо!

**S** Все хорошо!

Задайте мне вопрос про ваши данные

## Собираем агента - Архитектура



- Если SQL не был выполнен, prompt на его исправление
- Строим график, когда есть численные значения и строчек в таблице больше 3
- Поиск продуктов с помощью BM25
- Валидация SQL - проверка, есть ли SELECT + ?

→ Ака как решали проблемы

# Tips and Tricks

-Что больше всего  
влияет на ответ  
модели?



-Что больше всего  
влияет на ответ  
модели?

- **Пары вопрос-ответ (few-shot)**
  - Можно подавать в промпт
  - Можно подавать в историю диалога

-Как увеличить  
влияние инструкции  
при длинном  
контексте?

-Как увеличить  
влияние инструкции  
при длинном  
контексте?

- Помещать инструкции **И в начало, И в конец**. Но если вы очень экономите токены, то – **в начало**

“Wb”, “вб”, “вайлдбериз” - **запрос**

“Wb” - **в бд**

“Wb”, “вб”, “вайлдбериз” - **запрос**

“Wb” - **в бд**

- Определения с синонимы в векторную таблицу с документами

-Что делать, если  
агент не понимает,  
что такое прошлый  
месяц?

-Что делать, если агент не понимает, что такое прошлый месяц?

- Дата и время в промпт
- Доступный диапазон дат в промпт
- Определение в статические документы
- Проверить консистентность синтетики

-Что делать, если  
ответы не  
стабильны?



-Что делать, если  
ответы не  
стабильны?

- Температура 0
- Закрепить seed

-Что делать, если  
ответы не  
стабильны?

- Температура 0
- Закрепить seed

Метрика	Улучшение
Процент совпадения	49%
Soft Accuracy	11%



Как не рассказать лишнего

~~Ошибки~~

01. Как не рассказать лишнего?

- Защита через промпт + **проверка account\_id** + добавление account\_id (промпт)

01. Как не рассказать лишнего?

- Защита через промпт + **проверка account\_id** + добавление account\_id (промпт)

02. Как быть, если SQL сгенерировался не корректно?

- Починка скрипта - ошибка, вопрос и ответ в промпт с просьбой починить

01. Как не рассказать лишнего?

- Защита через промпт + **проверка account\_id** + добавление account\_id (промпт)

02. Как быть, если SQL сгенерировался не корректно?

- Починка скрипта - ошибка, вопрос и ответ в промпт с просьбой починить

03. Как быть, если починить не удалось?

- Ответ пользователю, что модель глупенькая и ничего не поняла и нужно уточнить запрос + похожие вопросы



Секретная информация

# Примеры

### **Товарная аналитика**

- Какие товары набирают популярность?
- Какую скидку я могу дать при условии поддержания маржинальности в пределах 5%?
- Какие категории товаров стоит расширить?

### **Финансы**

- Какой канал продаж моего бизнеса приносит максимальную прибыль?
- Как изменилась выручка за последние 3 месяца?

### **Оборачиваемость**

- Какие товары нужно закупить снова?
- Почему некоторые товары не продаются?



### **Реклама**

- Какие рекламные кампании самые эффективные?
- Какая у меня окупаемость рекламы?

### **Внешняя аналитика**

- Какие ниши сейчас набирают популярность?
- Что мне стоит дропнуть к своим товарам?
- В каких категориях сейчас мало продавцов?

### **Общие вопросы**

- Как мне поднять средний чек?
- Почему у меня низкая маржинальность?
- Как увеличить выручку в 3 раза?



Какие товары имеют высокий потенциал, но плохо оборачиваются?



```
1 WITH last_month AS (
2   SELECT date_trunc('month', (SELECT MAX(month) FROM monthly_product_me
3 )
4 SELECT product_name, abc_profit_segment, xyz_segment
5 FROM monthly_product_metrics
6 WHERE insales_account_id = 140259
7   AND month = (SELECT month FROM last_month)
8   AND abc_profit_segment = 'A'
9   AND xyz_segment = 'Z'
10 LIMIT 5;
```



	product_name	abc_profit_segment	xyz_seg
0	BMW Scanner 1.4 - автомобильный сканер	A	Z
1	Galletto 1260 RUS - диагностический программатор чип тюнинга	A	Z
2	Автосканер Delphi DS150e CDP Pro (Одноплатный) Bluetooth	A	Z
3	Набор переходников AutoCom для грузовых автомобилей	A	Z
4	Программа софт Delphi Autocom 2021.11 dvd диск + активация	A	Z



Вот список товаров с высоким потенциалом (категория A по ABC-анализу), но с низкой оборачиваемостью (категория Z по XYZ-анализу):

1. BMW Scanner 1.4 - автомобильный сканер
2. Galletto 1260 RUS - диагностический программатор чип тюнинга
3. Автосканер Delphi DS150e CDP Pro (Одноплатный) Bluetooth
4. Набор переходников AutoCom для грузовых автомобилей
5. Программа софт Delphi Autocom 2021.11 dvd диск + активация

Сделай столбчатую диаграмму: сравнение выручки по каналам за последний месяц.

```
1 WITH last_month AS (  
2   SELECT date_trunc('month', (SELECT MAX(month) FROM monthly_marketplace_metrics)) AS month  
3 )  
4 SELECT marketplace,  
5        ROUND(SUM(total_revenue), 2) AS total_revenue  
6 FROM monthly_marketplace_metrics  
7 WHERE month = (SELECT month FROM last_month)  
8    AND insales_account_id = 5630772  
9 GROUP BY marketplace  
10 ORDER BY total_revenue DESC;
```

	marketplace	total_revenue
0	wb	25812456.63
1	ozon	19667106

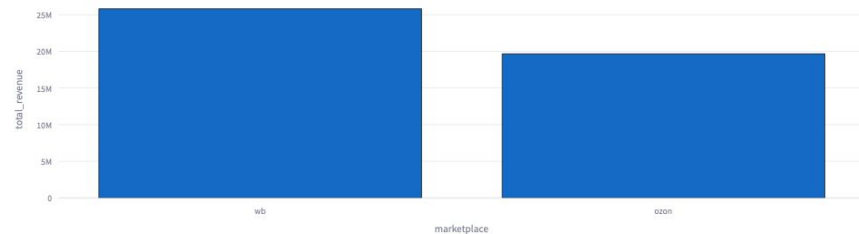
Столбчатая диаграмма сравнения выручки по каналам за последний месяц (февраль 2025) показывает, что канал wb принес 25,8 млн руб., а ozon — 19,7 млн руб.

Данные взяты с помощью SQL-запроса к агрегированным значениям выручки по каналам за последний доступный месяц. Это дает наглядное сравнение эффективности разных маркетплейсов.

Если нужно сравнить результаты по более ранним периодам или включить другие каналы — дайте знать!

```
1 import plotly.express as px  
2  
3 fig = px.bar(df, x='marketplace', y='total_revenue', title='Сравнение выручки по каналам за последний месяц')
```

Сравнение выручки по каналам за последний месяц





Куда двигаться дальше

# Развитие

- 01. Поиск по другим категориям (как по товарам)
- 02. Локальная LLM + дообучение для русского языка
- 03. Улучшение RAG
- 04. Структурированный вывод
- 05. Ризонинг модели
- 06. Куча продуктовых фичей



...что пришли!