

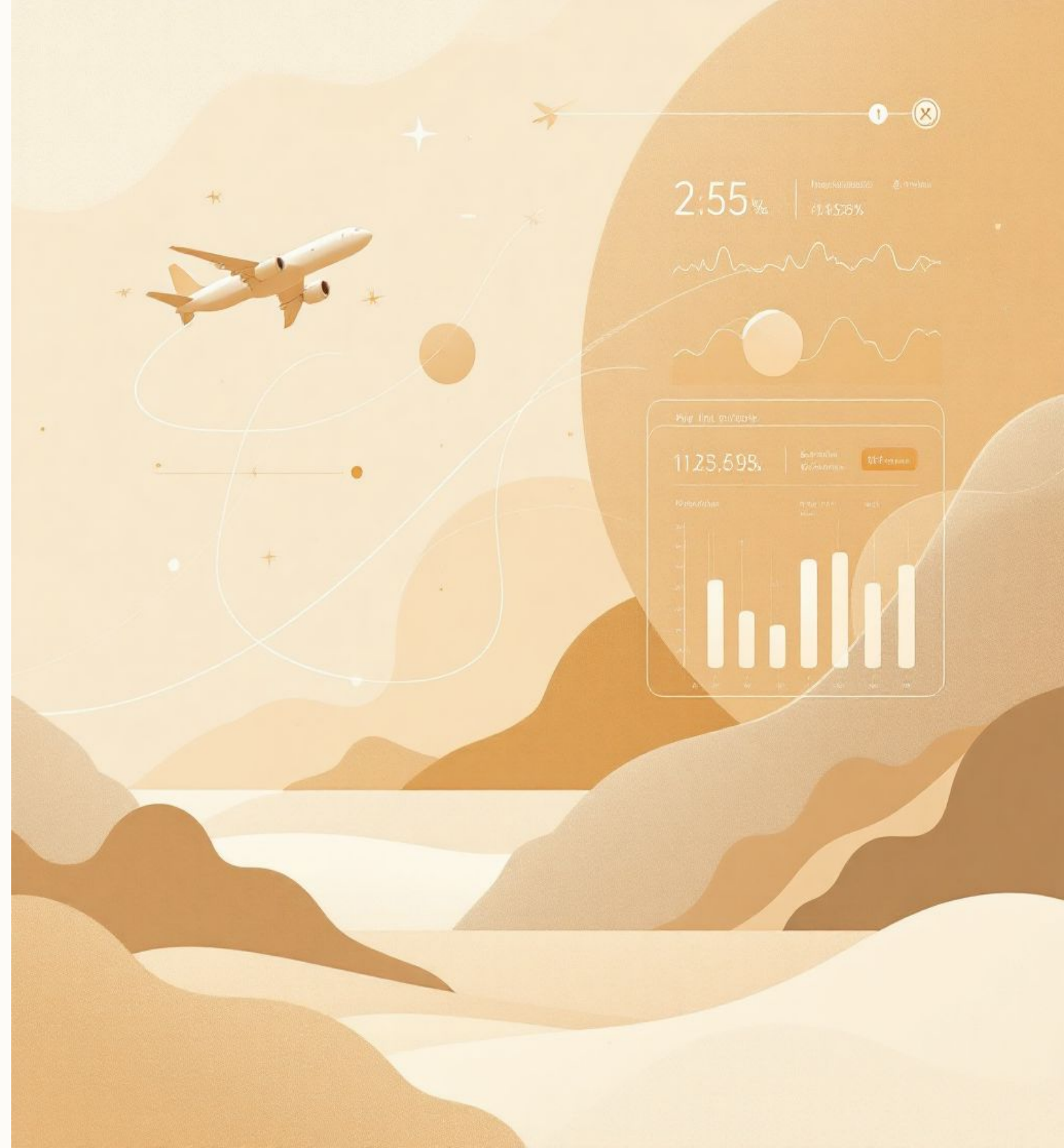
SkyPrice AI: Мониторинг и аналитика цен на авиабилеты

SkyPrice AI — это интеллектуальная система мониторинга цен на авиабилеты, которая собирает исторические данные и предоставляет оптимальные рекомендации по покупке. Система автоматизирует весь процесс от сбора данных до получения полезных инсайтов.

Задорожный Александр Сергеевич

ITMO

NAPOLEON IT



Предполагаемые даты по запросу: ...

Февраль

Вс	Пн	Вт	Ср	Чт	Пт	Сб
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28

Март

Вс	Пн	Вт	Ср	Чт	Пт	Сб
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

Июнь

Вс	Пн	Вт	Ср	Чт	Пт	Сб
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

Июль

Вс	Пн	Вт	Ср	Чт	Пт	Сб
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Октябрь

Вс	Пн	Вт	Ср	Чт	Пт	Сб
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Ноябрь

Вс	Пн	Вт	Ср	Чт	Пт	Сб
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

■ Вариант 2 ■ Вариант 3 ■ Вариант 4 ■ Вариант 5

Обзор: основные возможности



Умная генерация маршрутов

Персонализированная настройка поездок и динамическое создание маршрутов.



Сбор и обработка данных

Агрегация из нескольких источников и непрерывное архивирование цен.



Анализ и визуализация

Интерактивные графики трендов и инструменты сравнения маршрутов.

Архитектура системы: определение проекта

Домен

Оптимизация путешествий и транспорта.

Бизнес-Проблема

Снижение затрат на привлечение клиентов и улучшение времени покупки билетов.

Цели

Определение оптимальных окон покупки, прогнозирование колебаний цен и предоставление рекомендаций.

Select Dashboard

- ☐ Best Flights
- ☐ Route Combinations
- ☒ Price Trends
- ☐ Aggregator Analysis
- ☐ Data Quality

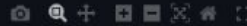
Data Sources:

- gold_request_best_flights_mv
- gold_request_best_route_combo_mv
- gold_request_price_timeseries_mv
- data quality from sessions

Flight Price Trends Over Time

Select Request ID (or leave empty for all):

2



Price Time Series for Request 2



Price Heatmap for Request 2

Price Heatmap by Date and Route Type



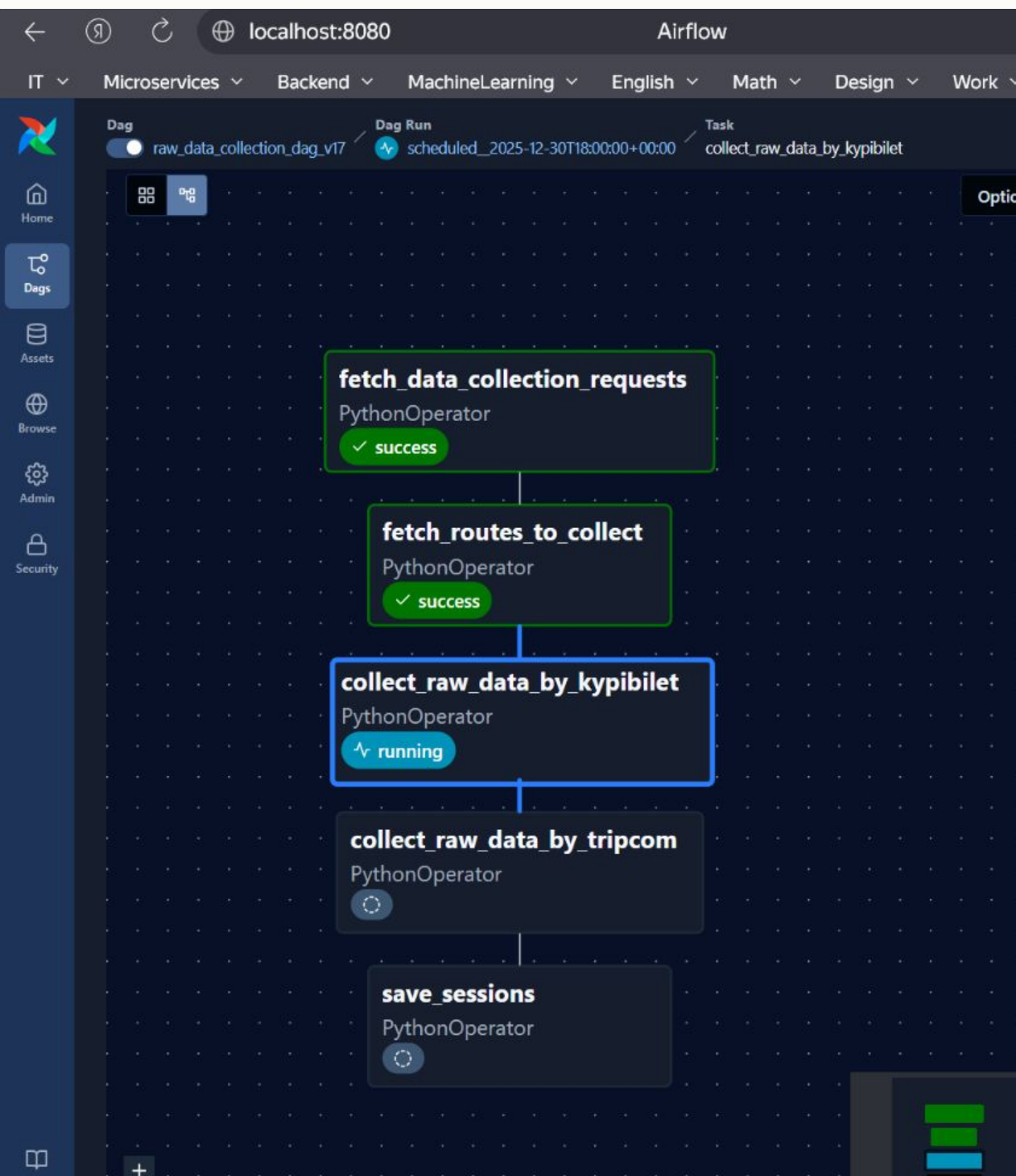
Умная генерация маршрутов

Персонализированная конфигурация

- Гибкие даты и окна поездок
- Предпочтения по дням вылета/прилета
- Односторонние или туда-обратно
- Предпочтения по времени суток

Система генерирует все возможные комбинации маршрутов на основе пользовательских предпочтений и начинает сбор данных.





Конвейер сбора и обработки данных

01

Агрегация из нескольких источников

Поддержка Trip.com и Kypibilet с расширяемой архитектурой.

02

Скрапинг в реальном времени

Использование Selenium для динамического извлечения контента.

03

Создание исторических данных

Непрерывный сбор и архивирование данных о ценах для анализа трендов.

04

Автоматическое планирование

Apache Airflow управляет всеми рабочими процессами сбора данных.

Технологический стек

Оркестрация	Apache Airflow	Планирование и автоматизация рабочих процессов
Хранение данных	PostgreSQL + SQLAlchemy	Хранение обработанных/анализируемых данных
Сырые данные	S3/MinIO	Хранение сырых данных и архивов
Визуализация	Streamlit	Интерактивные дашборды и аналитика
Сбор данных	Selenium (Python)	Веб-скрапинг для цен в реальном времени
Обработка данных	Pandas, NumPy	Очистка и преобразование данных
Оркестрация контейнеров	Docker Compose	Простое развертывание и управление сервисами

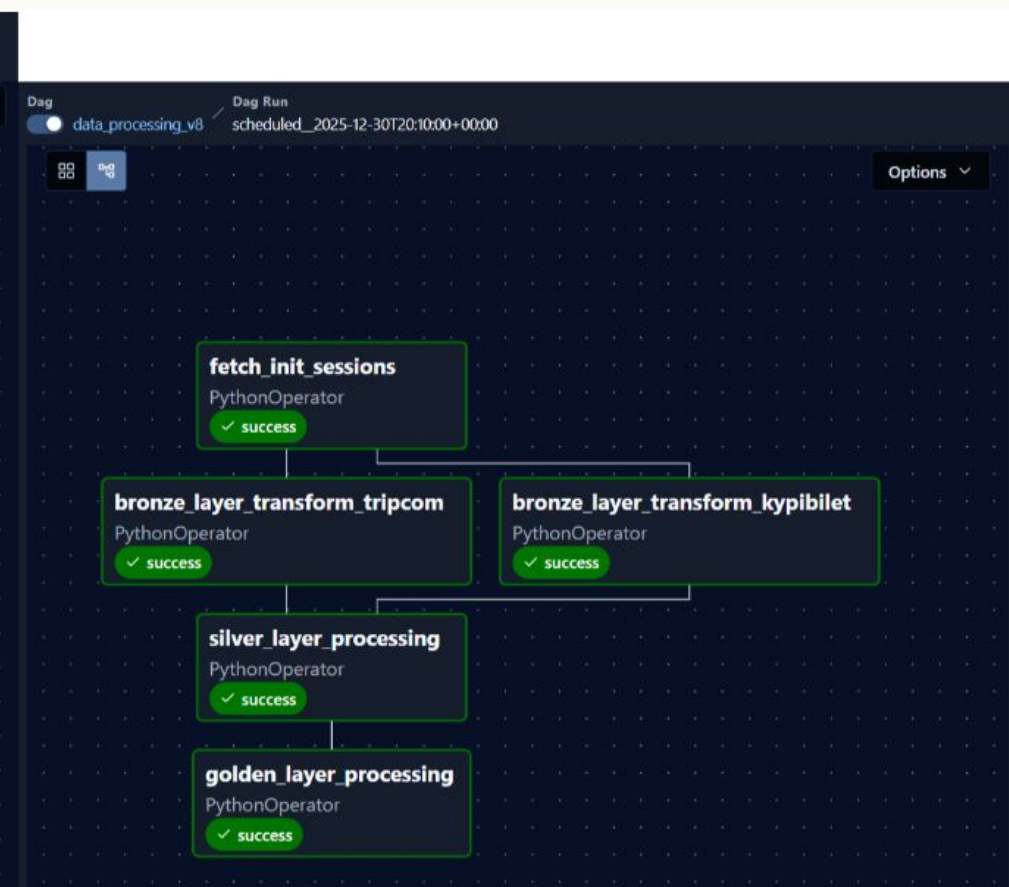
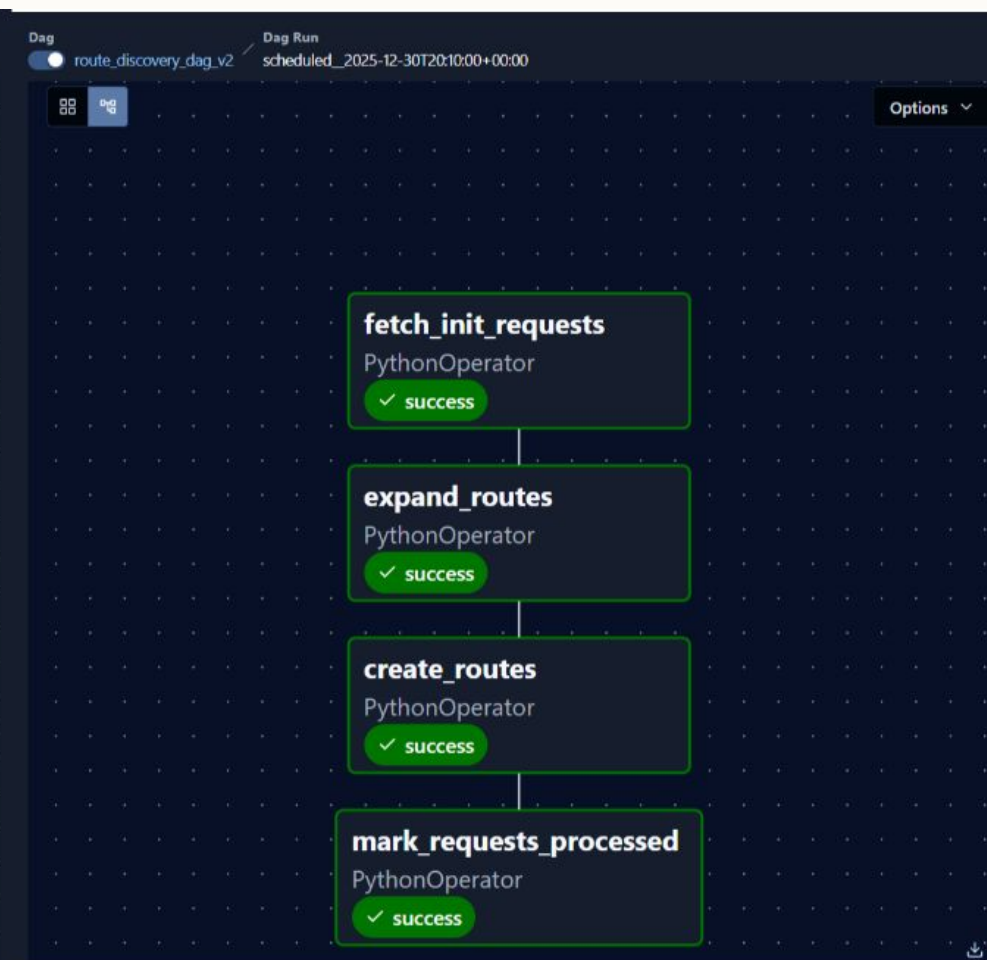
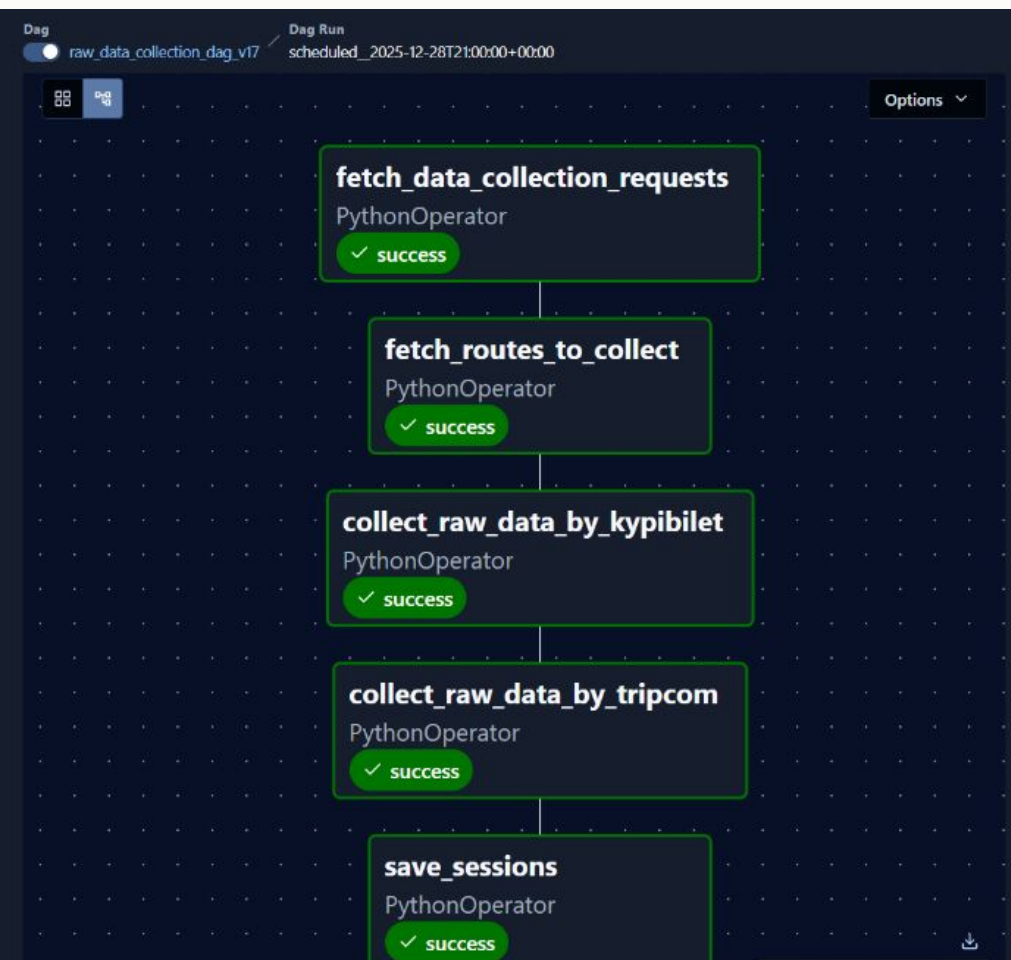
Очистка и предварительная обработка данных

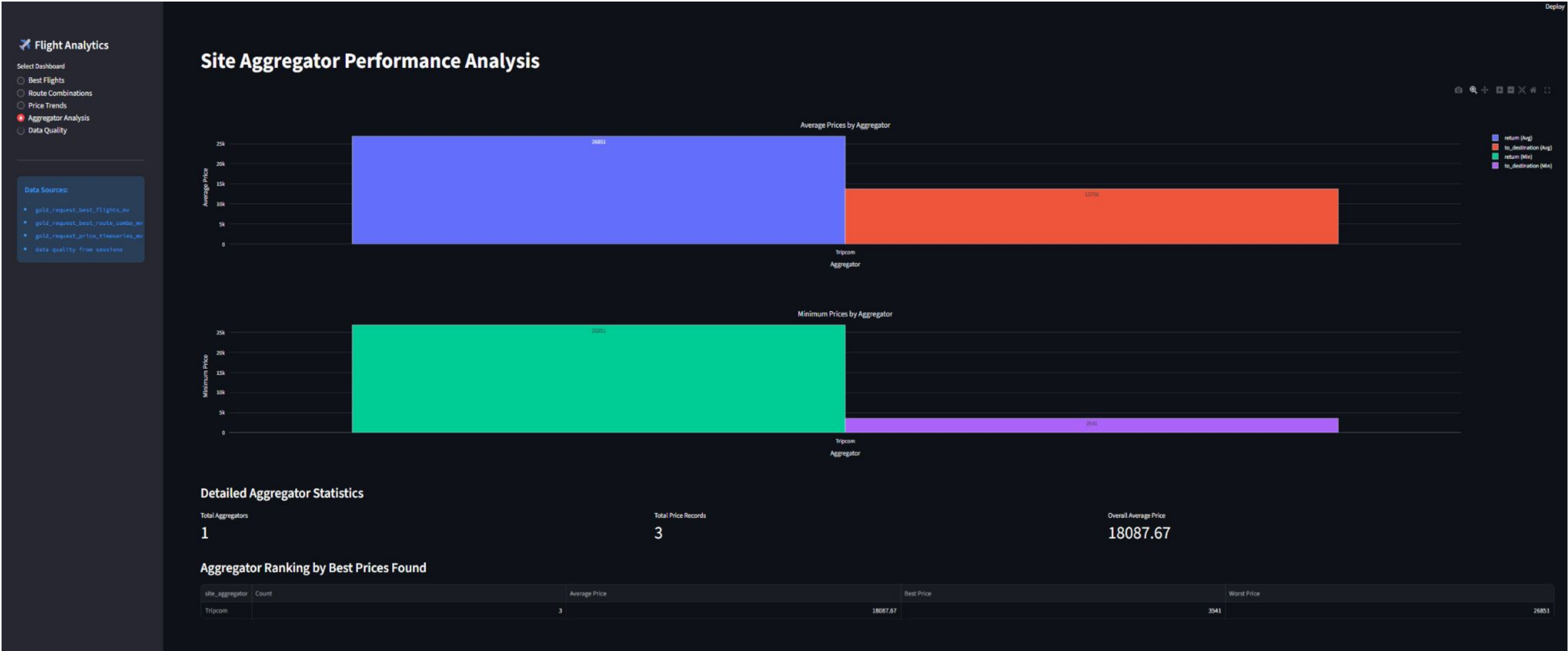
Очистка данных (Бронзовый слой)

- Автоматическая проверка на пропущенные цены и некорректные форматы.
- Валидация цен, стандартизация форматов дат, удаление дубликатов.
- Нормализация валют и сохранение метрик качества данных.

Предварительная обработка данных

- Нормализация цен, создание временных признаков.
- Обработка пропущенных значений (интерполяция, заполнение).
- Кодирование категориальных данных (авиакомпания, аэропорты).





Хранение данных и визуализация

Стратегия хранения

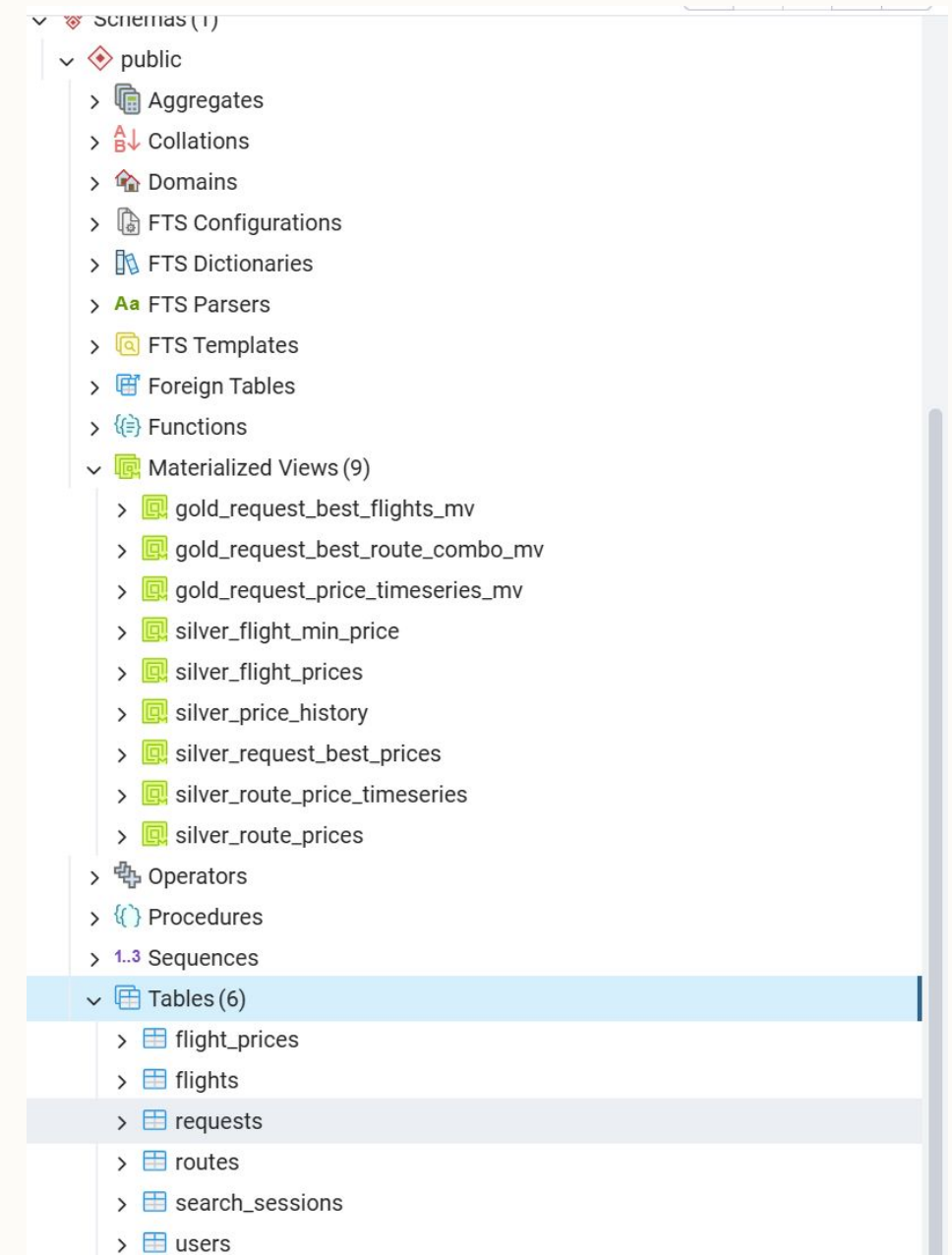
Сырые данные: S3/MinIO для необработанных данных.

Обработанные данные: PostgreSQL для аналитических данных.

Аналитические данные: Оптимизированные таблицы для быстрых запросов.

Анализ и визуализация

- Интерактивные графики трендов цен.
- Индикаторы оптимального времени покупки.
- Инструменты сравнения маршрутов и историческая аналитика.



←

↺

↻

🌐 127.0.0.1:9001

MinIO Console

⋮

🔖

🗨 80 %

🌐 Перевести

🔴

🔵

✍

📺

🔄

▶

↔

🤖

🔋

🗨

⬇

IT ▾ Microservices ▾ Backend ▾ MachineLearning ▾ English ▾ Math ▾ Design ▾ Work ▾ vpn ▾ CoolStuff ▾

Другие закладки ▾

MINIO

OBJECT STORE

Community Edition

🔍 Filter Buckets

Buckets

📁 kypibilet-raw-data

📁 tripcom-raw-data

📖 Documentation

Object Browser

🔍 Start typing to filter objects in the bucket

🔍 ☀

📁

kypibilet-raw-data

Created on: Sun, Dec 28 2025 18:37:48 (GMT+3)

Access: PRIVATE

238.9 KiB - 24 Objects

Rewind ⏮

Refresh ↺

Upload 📤

⏪

kypibilet-raw-data

📁

Create new path 📁

<input type="checkbox"/>	▲ Name	Last Modified	Size
<input type="checkbox"/>	📄 route_1_2025-12-28T15:37:48.json	Sun, Dec 28 2025 18:38 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_1_2025-12-28T16:01:14.json	Sun, Dec 28 2025 19:01 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_1_2025-12-28T16:47:46.json	Sun, Dec 28 2025 19:48 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_1_2025-12-28T18:00:04.json	Sun, Dec 28 2025 21:01 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_1_2025-12-28T19:31:06.json	Sun, Dec 28 2025 22:31 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_1_2025-12-28T19:55:34.json	Sun, Dec 28 2025 22:56 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_1_2025-12-28T21:00:03.json	Mon, Dec 29 2025 00:00 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_1_2025-12-29T19:55:34.json	Sun, Dec 28 2025 23:36 (GMT+3)	8.5 KiB
<input type="checkbox"/>	📄 route_2_2025-12-28T15:37:48.json	Sun, Dec 28 2025 18:38 (GMT+3)	7.0 KiB
<input type="checkbox"/>	📄 route_2_2025-12-28T16:01:14.json	Sun, Dec 28 2025 19:02 (GMT+3)	7.0 KiB
<input type="checkbox"/>	📄 route_2_2025-12-28T16:47:46.json	Sun, Dec 28 2025 19:48 (GMT+3)	7.0 KiB
<input type="checkbox"/>	📄 route_2_2025-12-28T18:00:04.json	Sun, Dec 28 2025 21:02 (GMT+3)	7.0 KiB
<input type="checkbox"/>	📄 route_2_2025-12-28T19:31:06.json	Sun, Dec 28 2025 22:32 (GMT+3)	7.0 KiB
<input type="checkbox"/>	📄 route_2_2025-12-28T19:55:34.json	Sun, Dec 28 2025 22:56 (GMT+3)	7.0 KiB
<input type="checkbox"/>	📄 route_3_2025-12-28T19:31:06.json	Sun, Dec 28 2025 22:32 (GMT+3)	12.8 KiB
<input type="checkbox"/>	📄 route_3_2025-12-28T19:55:34.json	Sun, Dec 28 2025 22:56 (GMT+3)	12.8 KiB

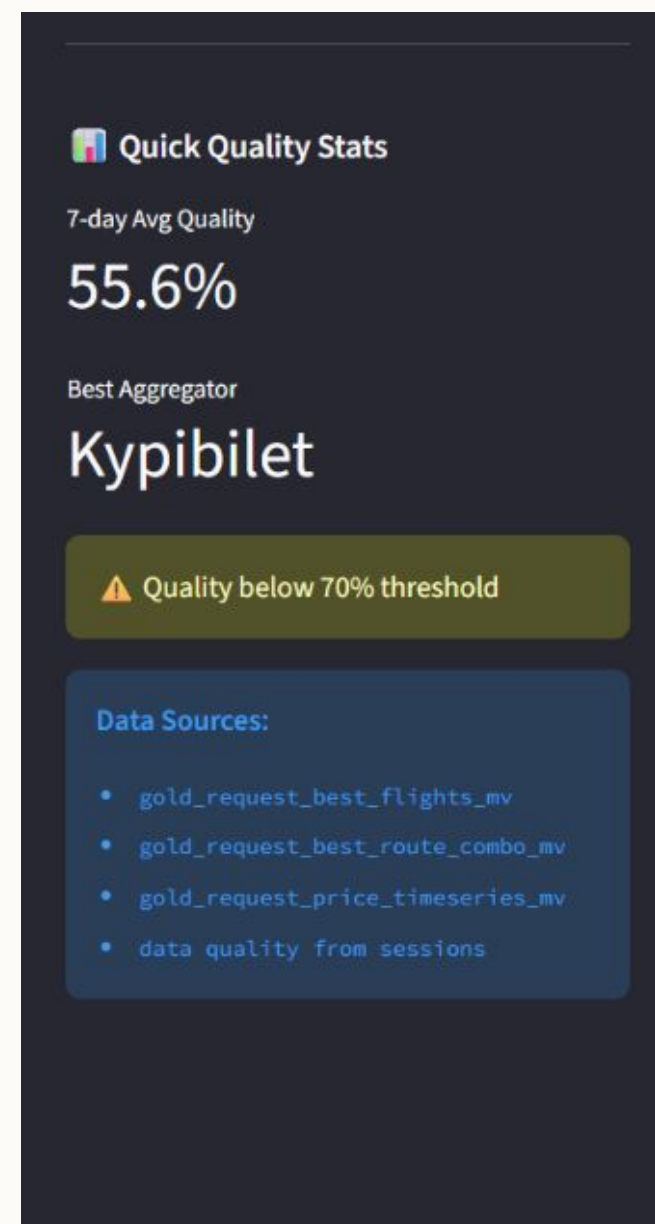
Обеспечение качества данных и расширяемость

Метрики Качества

- Полнота: Процент успешного сбора данных.
- Точность: Метрики парсинга и обработки данных.
- Согласованность: Межисточниковая валидация.

Расширяемость Платформы

- Легкое добавление новых агрегаторов.
- Интеграция новых моделей машинного обучения.
- Расширение возможностей экспорта данных.



Search Session Data Quality Analysis

Total Sessions

16

Sessions with Quality Data

16

Avg Completeness

55.58%

Avg Valid Ratio

60.63%

Date Range

2025/12/28 – 2025/12/30

Site Aggregator

Tripcom

Kypibilet

Quality Overview

Data Quality Scorecard

Total Sessions

16

Sessions with Quality Data

16

Avg Completeness

0.56

Avg Valid Ratio

0.61

Total Valid Flights

1362

Total Invalid Flights

425

Overall Success Rate

0.76

Trends

Distribution

Aggregator Comparison

Warnings Analysis

Quality Distribution

Overall Quality Distribution

poor

22.2%

excellent

77.8%

Completeness Score Distribution

Count

7

6

5

4

3

2

1

0

0

0.2

0.4

0.6

0.8

Completeness Score

Metrics Correlation

Quality Metrics Correlation Matrix

overall_completeness	1	1	1	1	0.3380617	0.431644
valid_ratio	1	1	1	1	0.3380617	0.431644
completeness_ratio	1	1	1	1	0.3380617	0.431644
avg_completeness	1	1	1	1	0.3380617	0.431644
total_warnings	0.3380617	0.3380617	0.3380617	0.3380617	1	0.9256947
valid_flights	0.431644	0.431644	0.431644	0.431644	0.9256947	1
invalid_flights						

Aggregator Quality Comparison

Aggregator Quality Comparison

Mean Completeness by Aggregator

Score

1

0.8

0.6

0.4

0.2

0

0.7

0.57

Kypibilet

Tripcom

Valid Flight Ratio

Ratio

1

0.8

0.6

0.4

0.2

0

0.78

0.39

Kypibilet

Tripcom

Average Warnings per Session

Warnings

60

40

20

0

1

61

Kypibilet

Tripcom

Session Count

Count

8

6

4

2

0

9

7

Kypibilet

Tripcom

Ценность проекта и будущие планы

Экономия затрат

Определение оптимальных окон покупки для максимальной экономии.

Принятие решений на

основе данных

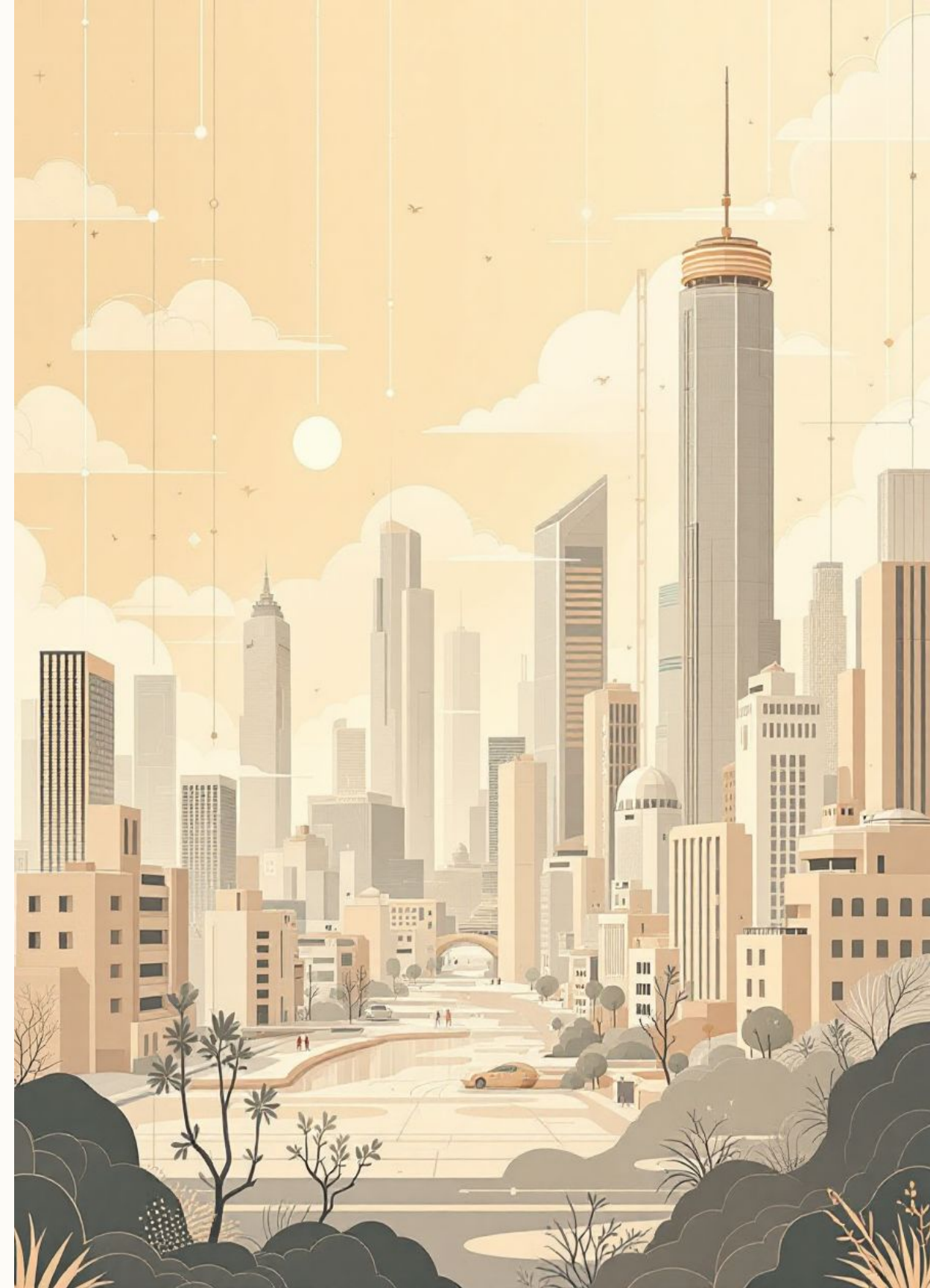
Эмпирический анализ заменяет догадки.

Масштабируемая архитектура

Обработка растущих объемов данных и пользовательских нагрузок.

Будущие планы

- Добавление Telegram-бота и новых агрегаторов.
- Интеграция ML-моделей для прогнозирования цен и оптимальных периодов покупки.
- Обработка пользовательских запросов с помощью LLM.



Контакты



Александр Задорожный

Руководитель отдела ML ООО “Мобил-груп”

8 909 611 96 73

alexander.zadorozhnyy@yandex.ru



ООО “Мобил-груп”