

ПРОЕКТНА РОБОТА

- Спеціалізації Analytics Engineering

- Project Author: Fedotova Hanna



Використані технології:

- PySpark – обробка та трансформація даних
- MongoDB – зберігання сирих даних (raw)
- Supabase (PostgreSQL) – централізоване сховище даних (DWH)
- dbt – моделювання даних: bronze → silver → gold
- Apache Airflow – автоматизація пайплайну
- Power BI – візуалізація та бізнес-аналітика
- Nominatim API – геокодування українських міст

Джерела даних:

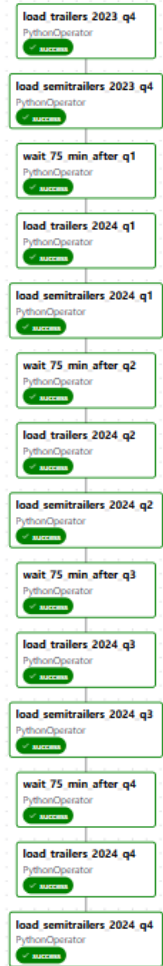
Джерело	Тип	Формат	Обсяг	Частота	Опис
RIA API	REST API	JSON	~10,000	Разове	Реальні оголошення, бренди, області та міста
MongoDB	NoSQL (raw layer)	JSON	~10,500	Разове	Початкове збереження даних з API та геоданих
Nominatim API	Геокодинг API	JSON	~400+	Разове	Координати міст України (широта/довгота)
Синтетичні дані	Генерація вручну	SQL	100	Разове	Фейкові оголошення для країн ЄС у Supabase
Supabase (PostgreSQL)	SQL DB	Таблиці	~10,500	Разове	Об'єднані дані для моделювання та аналітики
dbt models	SQL/dbt	—	—	Автоматично	Staging → clean → gold (денормалізовані аналітичні моделі)

Архітектура та технологічний стек

Компонент	Технологія / Інструмент	Роль у проєкті
Джерела даних	ukr.auto.ria API, Autoline	Основні API/JSON джерела по ринку тягачів та причепів
	OpenCage Geocoding API	Геокодування міст/областей України
	Python (synthetic data)	Генерація даних для країн ЄС
Bronze	MongoDB Atlas (free tier)	Зберігання необроблених JSON-даних з API
ETL	PySpark	Очистка, трансформація, валідація, завантаження в staging
		Завантаження з MongoDB → PostgreSQL
DWH / Staging	PostgreSQL (Supabase)	Зберігання staging-даних, marts, зіркова схема (Kimball)
Моделювання	dbt	Створення моделей, тестів, документації
Оркестрація	Apache Airflow (Docker)	DAG для початкового імпорту з API → MongoDB (тривалий процес)
Візуалізація	Power BI (через ODBC)	Аналітичні дашборди напряду з PostgreSQL



Apache
Airflow



Dag
ria_trailers_5_quarters_load
Dag Run
2025-07-05, 11:27:11



Options

load_trailers_2023_q4
load_semitrailers_2023_q4
wait_75_min_after_q1
load_trailers_2024_q1
load_semitrailers_2024_q1
wait_75_min_after_q2
load_trailers_2024_q2
load_semitrailers_2024_q2
wait_75_min_after_q3
load_trailers_2024_q3
load_semitrailers_2024_q3
wait_75_min_after_q4
load_trailers_2024_q4
load_semitrailers_2024_q4

2025-07-05, 11:27:11 success

Logical Date
2025-07-05, 11:27:11

Run Type
manual

Start
2025-07-05, 11:27:14

End
2025-07-05, 16:27:58

Duration
05:00:44

Add a note

Clear Run

Task Instances Events Code Details

Search Tasks

All States

Task ID	Map Index	State	Start Date	End Date	Try Number	Operator	Duration
load_semitrailers_2024_q4		success	2025-07-05, 16:27:55	2025-07-05, 16:27:58	1	PythonOperator	3.21s
load_trailers_2024_q4		success	2025-07-05, 16:27:50	2025-07-05, 16:27:54	1	PythonOperator	3.81s
wait_75_min_after_q4		success	2025-07-05, 15:12:49	2025-07-05, 16:27:50	1	PythonOperator	01:15:00
load_semitrailers_2024_q3		success	2025-07-05, 15:12:47	2025-07-05, 15:12:49	1	PythonOperator	2.17s
load_trailers_2024_q3		success	2025-07-05, 15:12:42	2025-07-05, 15:12:46	1	PythonOperator	3.64s



<https://fedotovaanna.github.io/TrailerMarket/#!/overview>

Overview



Project



Database



Group

Sources



public



factvehicles_comparison



factvehicles_new



raw



DimCountries



DimCurrency



dim_body_styles



dim_categories



dim_cities



dim_marks



dim_models



dim_states

Projects



TrailerMarket



models



gold



silver



staging

Projects



TrailerMarket



models



gold



fact_vehicles_gold



silver



fact_vehicles_silver



staging



stg_fact_vehicles



stg_factvehicles_comparison



name

- Выбрать все
- (Пусто)
- Автовоз
- Бетономішалка (Міксе...
- Бітумовоз
- Борт
- Бортовий напівпричіп
- Газовоз
- Для перевезення твар...
- Для перевезення твар...
- Зерновоз - напівпричіп
- Зерновоз - причіп
- Ізотермічна будка
- Інші причеи
- Контейнеровоз
- Контейнеровоз напівп...
- Лафет
- Легковий причіп
- Лісовоз / Сортиментов...
- Лісовоз / Сортиментов...

загальна кількіс
оголошень

102 ТЫС.

кількість марок на рінках

302

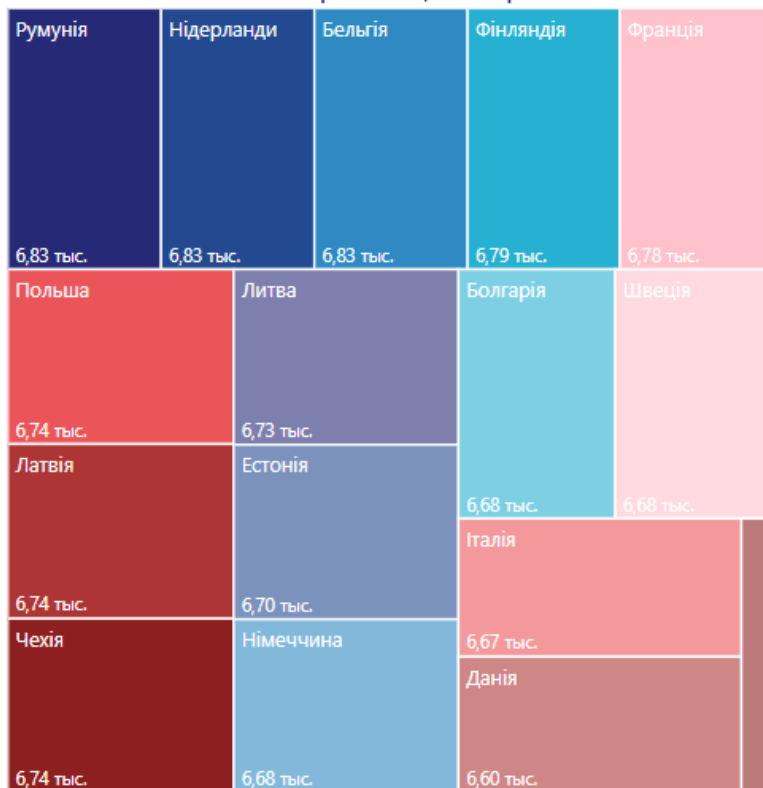
кількість країн

57

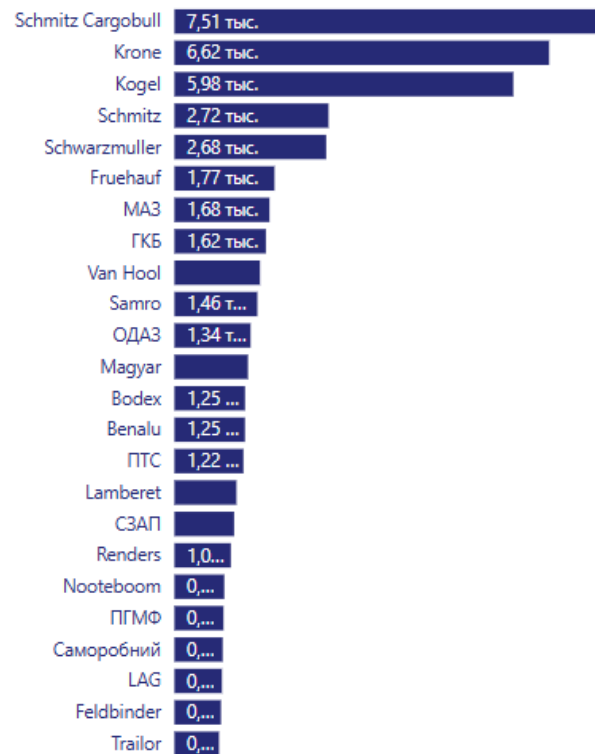
середня ціна продажу

9,78 ТЫС.

Кількість пропозицій по країнам



марки



Автор проекту щиро вдячний:

- викладачу Igor Bethley за глибокі, змістовні та справді цікаві лекції й заняття, які не просто передавали знання, а мотивували досліджувати самостійно. Завдяки вашій подачі складні технічні теми стали зрозумілими навіть тим, хто починав із нуля — як я.
- ментору Oleg Koval за перевірку домашок та зворотній зв'язок
- моїм колегам в групі команда Gosling за цікаві питання, допомогу та підтримку



мої постійні соавтори. Їх принципова відмова оцінювати ідеї як надто дивні сприяли збереженню авторської креативності. Відсутність перебивань у порівнянні з традиційною резиновою качечкою підтверджує ефективність альтернативних методів комунікації у процесі розробки

