AutoCommodity Data Collection & Analysis
Analyzing the Relationship Between Used Car & Commodity Prices

Group 206 — Dongyuan Gao, Ramiro Diez-Liebana, Cyriel Van Helleputte

November 2025

Contents

Group 206

Dongyuan Gao · Ramiro Diez-Liebana · Cyriel Van Helleputte

1 Project Overview

1.1 The Storyline (Potential Business Problem)

The Swiss used car market is highly competitive. Our **fictional client** AutoHelvetia AG, a leading national **used car dealer**, faces the challenge of optimizing their pricing & purchasing strategy. In recent years, commodity prices are volatile and affects pricing of cars. So AutoHelvetia AG delegated the task to us: to understand the relationship between used car prices and commodity prices.

1.2 Our Solution

This project delivers an **advanced data collection** and **anylsis** framework. Our goal is to collect valuable **market data** and uncover relationships between used car prices and key commodity markets. We develop a tool box using **web scraping of AutoScout24.ch** and integrating with **Yahoo Finance commodity data**, to provide AutoHelvetia AG data-driven insights for:

- **Optimize Pricing Strategies**
- **Gain Competitive Advantage**

1.3 Project Structure

```
project_scraping_CIP_analysis_car_commodity_price/
   Analysis/                  # Analysis notebooks and scripts
       RQ1/                   # Research Question 1 script & analysis
       RQ2/                   # Research Question 2 script & analysis
       RQ3/                   # Research Question 3 script & analysis
   Data/                      # Data storage
       API_data_pull/         # API-fetched commodity data & script
       clean_data/            # Processed and cleaned datasets & script
       Scraping/              # Web scraped data and scripts & scraper script
   Documentation.md           # This documentation file
   README.md                  # Project overview
```

1.4 Initiall Findings

2 Feasibility Research

2.1 Ethical Feasibility of Web Scraping AutoScout24.ch

This web scraping project was evaluated for both technical and legal feasibility. We focused on the academic research context and our analysis of AutoScout24.ch's robots.txt file and terms of service indicates that the project operates within acceptable boundaries for academic research purposes. - **robots.txt Analysis**: - Allowed: General listing pages without filters - Restricted: User account pages (`/de/account/`, `/de/member/`) - Restricted: Filtered search results with specific URL parameters (e.g., `sort=`, `pricefrom=`) - Restricted: Administrative functions

2.1.1 Technical Feasibility

- **Data Extraction**: Ethically extracts vehicle specifications, pricing, and listing details with Scraper and Yahoo Finance API, involving selenium and beautifulsoup.
- **Data Availability**: We found consistent and abundant data, which is appropriate for analysis for both used car listings and Commodity Data

2.1.2 Analytical Feasibility

- **Statistical Methods**: Appropriate statistical methods can be appllied for analysis, including correlation analysis, regression analysis, and time series analysis, etc.
- **Potential Conclusions**: The project can provide potential valuable insights into the relationship between used car prices and commodity prices, helping stakeholders make informed decisions

3 Data Collection

3.1 Web Scraping Implementation

3.1.1 Target Website

- **Primary Source**: AutoScout24.ch (https://www.autoscout24.ch)
- **Target Path**: /de/autos/alle-marken (All car listings)
- **Scope**: Used car listings across all makes and models available on the platform

3.1.2 Technical Implementation

3.1.2.1 Core Technologies

- **Selenium WebDriver**: For browser automation and dynamic content loading
- **BeautifulSoup4**: For HTML parsing and data extraction
- **Custom Parser**: Combines multiple extraction methods(json, html, css, regex) for robustness

3.1.2.2 Scraping Methodology

1. **Pagination Handling**:
   - Iterates through listing pages systematically and click on next page
   - Implements smart navigation with randomized delays (5-15s between pages)
2. **Data Extraction Strategy**:
   - **Primary Method**: Structure-aware parsing using SVG icon titles and sibling elements
   - **Combination of Methods**:
     - JSON structured data extraction
     - CSS class-based element targeting
     - Regular expression fallbacks for critical fields

3.1.3 Data Points Collected

Data Field	Description	Example
car_model	Full vehicle make and model	"Volkswagen Golf 2.0 TDI"
price_chf	Listing price in CHF	25,900
mileage	Vehicle mileage in km	85,200
engine_power_hp	Engine power in HP	150
power_mode	Fuel/power type	Diesel, Petrol, Electric, Hybrid
transmission	Transmission type	Automat, Manuell, Halbautomatik
production_date	Production date	2018
listing_url	Direct URL to the listing	[Link]

3.2 API Integration

[Document the API integration, including: - APIs used (Yahoo Finance, etc.) - Authentication process - Data retrieval methods - Rate limits and handling]

4 Data Processing

[Document the data processing pipeline, including: - Data cleaning steps - Data transformation - Handling missing values - Data validation]

5 Analysis and Methodology

5.1 Research Questions

1. [Research Question 1]
2. [Research Question 2]
3. [Research Question 3]

5.2 Methodology

[Describe the analytical methods used, including: - Statistical methods - Machine learning models (if any) - Validation techniques]

6 Results and Findings

[Present key findings, including: - Summary statistics - Visualizations - Key insights - Limitations]

7 Conclusion and Future Work

[Provide conclusions and potential future improvements]

8 References

[List all references and data sources]

9 Appendices

9.1 A. Data Dictionary

[Document the structure and meaning of all data fields]

9.2 B. Troubleshooting

[Common issues and solutions]

9.3 C. Contributing

[Guidelines for contributing to the project]