Analyzing the Relationship Between Commodity Prices and Used Car Depreciation in Switzerland

Feasibility Study

Project Team: Student AB (AutoScout24 Scraping), Student C (Yahoo Finance API & Commodity Data Pipeline), Student ABC (Data Cleaning, Analysis, Visualization)

Dongyuan Gao
Ramiro Diez-Liebana
Cyriel Van Helleputte

Table of Contents

1. Introduction and Project Scope 1
2. Research Questions 1
3. Data Sources and Acquisition Plan 1
   I. Primary Source: AutoScout24.ch (Used Car Listings) 1
   II. Secondary Source: Yahoo Finance (Commodity Price Data) 1
4. Risks, Constraints, and Mitigation Strategies 2
   I. Web Scraping Risks 2
   II. API and Data Quality Risks 2
   III. Timeline and Resource Constraints 2
5. Backup Plans and Alternative Strategies 2

1. Introduction and Project Scope

The automotive industry represents one of the most commodity-intensive manufacturing sectors, with materials such as steel, aluminum, copper, and increasingly lithium comprising a significant portion of vehicle production costs. This project investigates whether fluctuations in global commodity prices correlate with used car pricing patterns in the Swiss market. Our central hypothesis is that changes in raw material costs influence not only new car prices but also cascade into the used car market through multiple mechanisms, including altered production economics, shifting consumer preferences, and changing replacement costs.

The scope of this project resolves around two data streams that will be merged for analysis. First, we will scrape secondhand car listings from AutoScout24.ch, Switzerland's largest online vehicle marketplace, capturing price, mileage, model specifications, and power characteristics for popular car brands. Second, we will construct a time-series commodity price database using Yahoo Finance data, tracking spot and futures prices for key automotive materials including crude oil, copper, aluminum and steel from 2011 onward. By aligning these datasets and analyzing correlations, we seek to discover whether material price levels unfold in observable patterns within the secondhand car market.

Through this project, we aim to develop practical proficiency in Python-based data acquisition. We will work with web scraping technologies including Selenium and BeautifulSoup, and API integration through the yfinance library. We will deepen our skills on data cleaning methodologies by addressing real-world challenges such as missing data, outlier detection, etc. Finally, we will strengthen our visualization skills by creating graphical representations that represent relationships between commodity markets and automotive asset values.

2. Research Questions

RQ1: How do used car prices in Switzerland correlate with historical commodity price indices for key automotive materials (steel, aluminum, copper and crude oil)? We will construct a weighted composite commodity index reflecting material costs and examine Pearson and Spearman correlations with median used car prices over time, controlling for vehicle age and mileage.

RQ2: Do different vehicle power modes (petrol, diesel, electric, hybrid) exhibit different sensitivity to specific commodity price movements? For instance, we hypothesize that electric vehicles may show stronger correlation with rare material prices, while conventional combustion engine vehicles may be more sensitive to crude oil and steel price fluctuations.

RQ3: How does the relationship between commodity prices and used car values vary across popular car brands in Switzerland? We will segment our analysis by brand to determine whether luxury manufacturers, volume producers, or specific market segments exhibit different sensitivities to raw material cost pressures.

3. Data Sources and Acquisition Plan

I. Primary Source: AutoScout24.ch (Used Car Listings)

AutoScout24.ch serves as our primary data source for used car market information. This platform is Switzerland's dominant online marketplace for vehicle sales, providing extensive listings with detailed specifications. Our acquisition methodology employs Selenium with undetected-chromedriver to automate browser interactions and handle dynamic page elements, combined with BeautifulSoup for HTML parsing. A critical technical insight from our preliminary research is that AutoScout24 embeds structured JSON-LD data within a hidden script tag specifically designed for search engine indexing. This approach provides superior stability compared to scraping visual HTML elements, as structured data formats are less susceptible to frontend design changes.

For each listing, we will collect car model and brand information, price in Swiss Francs, mileage in kilometers, engine power in horsepower, and power mode classification including petrol, diesel, electric, and hybrid variants. Our target is to collect a minimum of two thousand listings spanning multiple popular brands in Switzerland, including Volkswagen, BMW, Mercedes-Benz, Audi, and Toyota, which collectively represent a substantial portion of the Swiss used car market. The scraping will be conducted with randomized delays between requests to mimic human browsing patterns and minimize detection risk.

II. Secondary Source: Yahoo Finance (Commodity Price Data)

Our commodity price pipeline will utilize the yfinance Python library to access Yahoo Finance's public endpoints without requiring API authentication. This approach enables retrieval of both spot prices, reflecting immediate delivery and settlement values, and futures prices for selected contract maturities extending to November 2027 where available. The commodities tracked will include West Texas Intermediate crude oil as a proxy for energy costs affecting transportation and production, copper and aluminum representing primary structural materials, steel indices where available, and lithium as the critical battery material for electric vehicles. The yfinance library's multi-ticker download capability will be leveraged to minimize request counts and reduce the risk of rate limiting.

4. Risks, Constraints, and Mitigation Strategies

I. Web Scraping Risks

The primary technical risk involves IP blocking or CAPTCHA challenges from AutoScout24. Anti-bot measures could halt data collection if our scraper is detected as automated activity. To mitigate this risk, we have implemented several defensive strategies. The script incorporates time limiting with randomized five to ten second pauses between page loads to simulate human browsing behavior. We employ undetected-chromedriver specifically designed to present a standard browser user agent and evade common detection heuristics. Additionally, the script includes CAPTCHA detection logic that pauses execution and allows for manual intervention to solve challenges before resuming automated collection.

Website structure changes represent another concern, as AutoScout24 could modify its HTML or CSS architecture, potentially breaking our data extraction logic. However, our reliance on structured JSON-LD data provides significant resilience against this risk. Because this data format is designed for machine readability and search engine optimization, it experiences far less frequent changes than visual presentation layers. Nevertheless, we will maintain version-controlled scraper code and document all selector logic to enable rapid updates if structural changes do occur.

II. API and Data Quality Risks

Yahoo Finance endpoint reliability presents a secondary risk, as yfinance relies on unofficial public endpoints that could change without notice or introduce rate limiting. To address this, we implement conservative request pacing at approximately one request per second and utilize multi-ticker downloads to reduce total request volume. We also maintain awareness that yfinance is a community-maintained library and could experience breaking changes with Yahoo's infrastructure updates.

Coverage gaps in commodity data represent a known limitation, particularly for lithium, which has incomplete historical data on Yahoo Finance and lacks individual futures contracts. Our mitigation strategy involves flagging lithium data as potentially incomplete and preparing alternative data pipelines using sources such as Trading Economics or World Bank Pink Sheet data. These alternative sources will be documented separately to prevent mixing different data provenance within our primary dataset.

Data quality concerns include potential missing values, outliers, and inconsistent formatting across scraped listings. Our cleaning pipeline will enforce strict data type validation, establish plausible value ranges for fields like mileage and price, and implement outlier detection using interquartile range methods. Listings missing critical fields such as price or model information will be flagged and excluded from analysis to maintain dataset integrity.

III. Timeline and Resource Constraints

The compressed timeline to October ninth creates execution risk. Our mitigation approach prioritizes collecting sufficient data for statistical validity over exhaustive coverage. We will focus on three to five high-volume car models that ensure adequate sample sizes for correlation analysis. If scraping encounters technical obstacles consuming excessive time, we will reduce the scope to a single popular brand with multiple model variants rather than attempting comprehensive brand coverage. For commodity data, the automated nature of API calls reduces timeline risk, but we have allocated buffer time for debugging rate limiting or authentication issues.

5. Backup Plans and Alternative Strategies

If AutoScout24 proves resistant to scraping despite our mitigation strategies, we have already prepared autohero.nl as our primary backup data source. We have conducted preliminary technical reconnaissance on this platform and verified that its structure is amenable to our scraping approach, meaning we could pivot to this alternative with minimal delays.

For commodity data, if direct commodity price tracking through Yahoo Finance becomes unreliable, we have explored using proxy instruments that closely track underlying material prices. Specifically, we can leverage commodity-focused exchange-traded funds such as those tracking industrial metals, energy sectors, or battery materials as reliable proxies for the commodities themselves. These ETFs aggregate exposure to the underlying physical commodities and often exhibit strong correlations with spot prices while being readily accessible through the same yfinance infrastructure we are already implementing. Additionally, we have identified key publicly traded companies whose stock performance serves as a proxy for specific commodity price movements.

If the direct commodity-to-car-price correlation proves weak or inconclusive, we can pivot our analytical focus toward depreciation curve modeling independent of commodity effects, which would still provide valuable insights into Swiss used car market dynamics. Alternatively, we could narrow our scope to electric vehicles specifically and their relationship with lithium and battery-related commodity prices, where the theoretical connection may be more direct and observable.