

From Machine to Masterpiece : Artification and Price Formation of Classic Cars at Auctions (2025)

Noam Chabot & Léna Rebours - Université PSL

Repository purpose

This repository provides the datasets used in the research article. It does not include analysis scripts or regression code.

Overview

This repository accompanies the research article *From Machine to Masterpiece: Artification and Price Formation of Classic Cars at Auctions (2025)*.

The project investigates how industrial automobiles become cultural luxury assets and how auction prices are formed through institutional mechanisms, narrative qualification, and market contexts.

The study combines historical archival research on early automobile auctions, institutional and qualitative analysis of the contemporary auction market, combined with hedonic price regressions

Key findings include positive price associations with pedigree, rarity, and narrative qualification; statistically significant brand fixed effects (e.g., premium brands); contextual event effects associated with specific sale environments; and evidence that auction prices reflect institutional and narrative framing rather than intrinsic value alone.

Data

This study relies on two datasets.

Early Public Automobile Auction Sales in France (1898–1930)

First, a dataset containing early public auctions of automobiles and automobile-related goods in France during the late nineteenth and early twentieth centuries. It contains 41 auction lots drawn from three public sales held in 1898, 1922, and 1930. The data were collected from official auction records (*procès-verbaux*) consulted at the Paris Archives. The sales were identified through a systematic review of *La Gazette Drouot* beginning with its first publication in 1891, which allowed the identification of one of the earliest documented public automobile auctions in France (December 1898). Two additional sales from 1922 and 1930 were selected to complement this early

observation. For each lot, the dataset records the sale date, a textual description of the item, the sale price when available, and the identities of the auctioneer, buyer, and seller as reported in the archival documents. The lots include automobiles, trucks, and various types of automotive equipment, such as spare parts and accessories.

Classic Car Sales, Primarily in France (2025)

Secondly, a dataset containing information on collectible automobiles sold at public auctions in 2025, primarily in France. It includes more than 600 auction lots and was constructed as part of an academic research project in Art Economics. The dataset focuses on vehicles manufactured or first registered before 1995, in order to capture a coherent segment of the collectible car market. The data were collected manually from auction catalogues and published results between September and December 2025. Auctions were identified through La Gazette Drouot and through targeted consultation of the websites and catalogues of major auction houses active in the collectible automobile market, including Christie's, Sotheby's, Bonhams, Phillips, and Artcurial. The dataset is not intended to be exhaustive; auctions with insufficient information, particularly missing hammer prices or estimates, were excluded to ensure data quality. For each vehicle, the dataset records core auction and vehicle characteristics, including the year of manufacture, brand, textual lot description, low and high pre-sale estimates, and, when available, the hammer price. Additional metadata include the auction house, sale name, auction dates, lot number, and information on buyer's fees. When possible, final prices including buyer's premiums were computed, and all prices are reported in both euros and U.S. dollars. In addition to manual annotation, a large language model was used to extract structured variables from the textual lot descriptions using a standardized, rule-based prompt. This process was limited to information explicitly stated in the catalogues and allowed the extraction of supplementary variables related to vehicle characteristics (e.g., color, engine configuration), administrative information (e.g., registration documents, chassis numbers), and, when mentioned, elements of provenance or condition.

Prompt for LLM

A large language model (Qwen2.5-7B) was used to extract structured variables from auction lot descriptions. Some extracted features were subsequently excluded when the underlying textual information was insufficiently precise. For transparency, the original standardized prompt including all initially defined variables is reproduced below.

You are an expert automotive auction analyst. Your task is to extract structured, ordered information from the following auction car description.

CAR DESCRIPTION:
\"\\\"{car_description}\\\"\\\"

CORE PRINCIPLES:

- Extract ONLY information explicitly stated in the text.
- Do NOT infer, assume, or guess missing information.
- If a piece of information is not present, write exactly: None
- For Yes/No fields: output Yes only if it is explicitly stated; output No only if it is explicitly stated as absent; otherwise output None.
- You MUST analyze carefully before producing the final structured output.

FIELD DEFINITIONS (STRICT):

A. Identity & Legal

0) Brand (Make)

- Extract the brand/make exactly as stated (e.g., “Porsche”, “Citroën”, “Ferrari”).
- If multiple brands are mentioned, select the brand of the vehicle being sold.
- If not mentioned: None.

1) Registration Document (Yes / No / None)

- Yes: explicitly mentions a registration document (e.g., “carte grise”, “registration document”, “title”, “logbook”).
- No: explicitly states there is no registration document.
- None: not mentioned.

2) Chassis Number Mentioned (Yes / No / None)

- Yes: a chassis/VIN number is explicitly written or clearly referenced as present in the description.
- No: explicitly states chassis/VIN is unknown/absent/not provided.
- None: not mentioned.

3) Matching Numbers (Yes / No / None)

- Meaning: engine/chassis (and possibly gearbox) numbers match the original factory specification.
- Yes: explicitly states “matching numbers” or an equivalent statement.
- No: explicitly states it is not matching numbers.
- None: not mentioned.

B. Physical Characteristics

4) Color

- Extract the color exactly as stated (e.g., “red”, “black”, “silver”, “two-tone blue/white”).
- If multiple colors are stated, keep them as written in a short phrase.
- If not mentioned: None.

5) Engine Configuration

- Output ONLY one of the following values (exact spelling):

Inline-4

Inline-6

V6

V8

V10

V12

Boxer/Flat

Rotary

Other

None

- Choose the closest category ONLY if explicitly stated (e.g., “V8 engine” → V8; “flat-six” → Boxer/Flat).
- If the engine layout is mentioned but does not fit the list (e.g., W12, V16, straight-8, electric): output Other.
- If not mentioned: None.

6) Right-hand Drive (Yes / No / None)

- Yes: explicitly states right-hand drive (RHD).
- No: explicitly states left-hand drive (LHD) or explicitly says it is not RHD.
- None: not mentioned.

7) Three-wheeler (Yes / No / None)

- Yes: explicitly states it has three wheels / is a three-wheeler.

- No: explicitly states it is not a three-wheeler / has four wheels (rare, but if explicit).
 - None: not mentioned.

C. Production & Authenticity

8) Replica (Yes / No / None)

- Yes: explicitly states replica / tribute / recreation.

- No: explicitly states it is not a replica.

- None: not mentioned.

9) Prototype / Concept (Yes / No / None)

- Yes: explicitly states prototype / concept car / pre-production / show car concept.

- No: explicitly states it is not a prototype/concept.

- None: not mentioned.

10) Production Unique (Yes / No / None)

- Yes: explicitly states one-off / unique build / single example produced.

- No: explicitly states it is not unique / produced in multiple examples (explicitly).

- None: not mentioned.

D. History & Condition

11) Mileage (km)

- Extract the mileage value if explicitly stated.

- If the description provides miles only, keep the number as stated and add “(miles)” (do NOT convert).

- If not mentioned: None.

12) Pedigree

- Extract any explicit notable ownership/collection/provenance statements (e.g., “ex-collection X”, “owned by Y”).

- If not mentioned: None.

13) Repaired / Restored for Resale (Yes / No / None)

- Yes: explicitly states the car was repaired/restored/refurbished/serviced (e.g., “restored”, “mechanical overhaul”, “repainted”, “rebuilt”) in connection with sale or preparation for sale.

- No: explicitly states it was not repaired/restored for resale.

- None: not mentioned or repairs/restoration are mentioned without any link to resale preparation.

IMPORTANT RULES:

- 1) Think step-by-step BEFORE writing the standardized output.
- 2) The justification must briefly cite the phrases or facts from the description that led to each extracted value.
- 3) Do not add fields. Do not change field names. Do not change ordering.
- 4) The structured output must be EXACTLY parseable line-by-line.
- 5) If uncertain, output None.

RESPONSE FORMAT (STRICT — EXACTLY THIS STRUCTURE):

Line 1: Justification: [2-6 sentences. Reference evidence for multiple fields; do not exceed what is needed.]

Line 2: ### Extracted Information

A. Identity & Legal

Brand: [value or None]

Registration Document: [Yes / No / None]

Chassis Number Mentioned: [Yes / No / None]

Matching Numbers: [Yes / No / None]

B. Physical Characteristics

Color: [value or None]

Engine Configuration: [Inline-4 / Inline-6 / V6 / V8 / V10 / V12 / Boxer/Flat / Rotary / Other / None]

Right-hand Drive: [Yes / No / None]

Three-wheeler: [Yes / No / None]

C. Production & Authenticity

Replica: [Yes / No / None]
Prototype / Concept: [Yes / No / None]
Production Unique: [Yes / No / None]

D. History & Condition
Mileage (km): [value or None]
Pedigree: [Yes or None]
Repaired / Restored for Resale: [Yes / No / None]

Data Source Notice

Auction catalogues and sale results remain the intellectual property of the respective auction houses and publishers. This repository provides structured research data compiled for academic purposes only.

Authors

Noam Chabot - Université PSL - noam.chabot@psl.eu

Léna Rebours - Université PSL - lena.rebours@psl.eu

License

This dataset is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). See the LICENSE file for details.