

ALT/FNDATA ACADEMY · MODULE 15

Art market fundamentals

This session asks a single plain question and answers it with real transaction data: what forms the price of a work of art.

75 minutes · sandbox.altfndata.com

What today looks like

- Length: 75 minutes.
- Level: general art audience, no coding or finance background assumed.
- Access: the no-code sandbox at sandbox.altfndata.com only. No API key is needed. Students self-register with a work or school email and are approved automatically.
- Goal: students leave able to name what forms an artwork's price, build a comparables set for one artist by hand, and read price dispersion and estimate versus realized price for that artist.

OBJECTIVES

By the end of class students can

- Name the factors that form an artwork's price, including artist, medium and format, size, period, rarity, provenance, condition, and market timing, and explain how each one moves a price up or down.
- Distinguish the primary market, where a work first sells through a gallery or dealer, from the secondary market, where a work resells at auction, and state which market this dataset captures.
- Define pre-sale estimate as the auction house's formalized judgment of comparable value and realized price as what a buyer actually paid, and explain how the two relate.
- Build a comparables set for a single artist from sold lots, and read it for its typical price, its range, and its outliers.
- Describe price dispersion within one artist's market, using the minimum, median, and maximum realized price, and explain why two works by the same artist can sell for very different amounts.
- Identify the limits of auction data for art market fundamentals, including that it captures the secondary market only, that documented fields do not include medium, size, provenance, or condition directly, and that the newest periods are still being ingested and should not be read as a trend.

How the session runs

- 0 to 10 min: What forms a price. Introduce the factors, artist, medium and format, size, period, rarity, provenance, condition, and timing, and explain that this session tests these ideas against real sold lots rather than treating price as unexplainable.
- 10 to 20 min: Sandbox orientation. Confirm students can open the fine art data table and locate designer, vendor, `usd_price_decimal`, `sale_estimates_high_usd_price`, and `status` in the data dictionary, with no query written yet.
- 20 to 35 min: Guided demo, build a comparables set for one artist, ordering sold lots by realized price.
- 35 to 50 min: Guided demo, read price dispersion for the same artist using minimum, median, and maximum realized price, and compare estimate to realized price.
- 50 to 60 min: Guided demo, a rough medium and format proxy using text search on `item_title`, framed explicitly as an imperfect keyword match rather than a clean field.
- 60 to 70 min: Small-group exercise, students pick a different artist and build their own comparables set and dispersion figures.
- 70 to 75 min: Class discussion and homework assignment, groups share what surprised them about their artist's range.

The in-class walkthrough

- Open sandbox.altfndata.com, sign in, and select the fine art data table from the SQL editor dropdown.
- Open the coverage browser tab and confirm how a well-known artist is written in the designer field. Spelling and partial-name matching both matter for a clean comparables set.
- Return to the SQL editor and run the first guided query, a comparables set for one artist, which lists that artist's sold lots ordered from highest to lowest realized price.
- Read the list aloud with the class. Point out that the works clustered near the top and bottom are the outliers, and ask what might explain a lot near the top, a larger size, a rarer subject, a better provenance, versus a lot near the bottom.
- Run the second guided query, price dispersion for the same artist, returning the minimum, median, and maximum realized price alongside the sold-lot count. Explain that the gap between minimum and maximum is the range a comparables set has to explain, and that the median, not the average, is the safer typical figure when a handful of very high prices are present.
- Run the third guided query, estimate versus realized price for the same artist's sold lots, and explain that the estimate is the house's formalized judgment of comparable value before the sale, while realized price is what a real buyer decided the work was worth on the day. Point out where the two agree and where they diverge.

The core query

- Run this in the sandbox SQL editor:

```
- SELECT designer AS artist,  
-           item_title,  
-           vendor,  
-           sale_date,  
-           usd_price_decimal AS realized_price_usd  
- FROM all_fine_art_data  
- WHERE status = 'sold'  
- AND designer LIKE '%Warhol%'
```

DISCUSS

Questions for the room

- A comparables set for one artist shows lots ranging from a modest sum to many times that amount. What factors, artist reputation aside, might explain a work near the top of that range compared with one near the bottom?
- Why is the median a safer measure of an artist's typical price than the average, when a comparables set includes a few very high sales?
- The pre-sale estimate is the auction house's formalized judgment of comparable value. What does it mean when realized price consistently lands above that judgment, and what does it mean when it consistently lands below?
- This dataset captures the secondary market only, resales at auction, not the primary market of gallery and dealer sales. What kind of artist's true market would this data most understate, and why?
- A text search on `item_title` for a phrase like oil on canvas is a rough proxy for medium, not a clean field. What kinds of works might this search miss or wrongly include, and what would you want to check before trusting the result?

This week's assignment

- Each student selects one artist represented in the fine art data and writes a two-page comparables memo, as if preparing a client for an upcoming sale.
- The memo reports the artist's comparables set of recent sold lots, states the minimum, median, and maximum realized price, and identifies at least two outlier lots with a plausible explanation for why each sits where it does in the range, drawing on medium, size, period, rarity, provenance, or condition as concepts even where the data cannot confirm them directly.
- The memo compares estimate to realized price for at least three lots and states what the pattern suggests about how the house's judgment held up against real buyers.
- The memo must include the SQL queries used as an appendix, at least one exported table, and an explicit limitations section addressing the secondary-market-only nature of the data, the imperfect nature of any item_title text search used, and the recency caveat.
- Grading criteria:

NEXT

Wrapping up

- Restate the one-sentence takeaway before students leave.
- Point to the facilitator guide for discussion guidance and teaching notes.
- Questions or a class API key: info@altfndata.com