

Facilitator guide: Module 16, data science for art market research

A teacher-facing companion to Module 16. The module is the full lesson; this guide adds the facilitation layer: a preparation checklist, timing cues, guidance for the discussion questions, and notes on what to emphasize and where students get stuck. Slides and a printable version are below.

At a glance

- **Length:** 90 minutes.
- **Level:** graduate art business or analysts, comfort with numbers and tables assumed, no heavy math and no statistics prerequisite beyond the idea of a median.
- **Access:** the no-code sandbox at sandbox.altfndata.com is the primary path for the entire session. The optional Python and API extension needs a class API key from your instructor account, requested from info@altfndata.com.
- **Goal of the session:** students leave able to build a defensible comparable set, construct and honestly read an artist-level market index, pull demand signals that do not depend on price, and run a cohort comparison across artists without mistaking noise for a finding.

Before class

- Register your own sandbox account and run all four guided queries end to end, so you have known-good results on screen, including the cohort comparison with the minimum-lot threshold.
- Open the coverage browser on your chosen demo artist beforehand and confirm how the name appears in the designer field and roughly how many sold lots that artist has. The whole session depends on a comparable set large enough to be read with confidence.
- Confirm the exact documented field names in the schema browser, or via GET `/v1/tables` at docs.altfndata.com, before the demo: `designer`, `item_title`, `sale_date`, `usd_price_decimal`, `sale_estimates_high_usd_price`, `status`, and `vendor`.
- Load the slides and confirm the projector or screen share shows the sandbox clearly, especially the count column next to the median in the index query.
- Optional: for the API extension, request a class key from info@altfndata.com ahead of time and confirm the downloadable Python client or tutorials notebook opens on your machine.

Timed agenda with cues

Time	Segment	What to do	Watch for
0 to 12	What makes art market research rigorous	Introduce the comparable set as the atomic unit of the analysis and preview the four methods the class will build: an index, a hedonic-style split, demand signals, and a cohort comparison.	Students treating "rigorous" as a synonym for "more data" rather than a defensible comparable set and honest caveats.

Time	Segment	What to do	Watch for
12 to 22	Sandbox and schema orientation	Confirm students can open the fine art data table and locate designer, item_title, sale_date, usd_price_decimal, sale_estimates_high_usd_price, status, and vendor in the schema browser, and confirm one artist's lot count is large enough to support an index.	Students guessing field names instead of confirming them; an artist chosen with too few lots to support a reliable index.
22 to 42	Guided demo, artist-level market index	Run Query 1, median realized price by year with lot count shown alongside, and read it aloud pointing to the count column first. State the recency caveat before anyone reads the newest one or two periods as a signal.	Students reading a rising median as proof of appreciation without checking whether the count shows the mix of works changed.
42 to 55	Guided demo, demand signals and hedonic split	Run Query 2, sell-through and bought-in rate for the same artist, then Query 3, the rough hedonic-style split on an item_title keyword. Be explicit that the keyword is an imperfect proxy, not a documented characteristic field.	Students treating the item_title keyword match as a true, reliable attribute rather than a rough proxy that will misclassify some lots.
55 to 70	Guided demo, cohort comparison	Run Query 4, comparing several artists at once with a minimum-lot threshold (HAVING COUNT(*) >= a chosen number). State the secondary-market-only caveat: none of the four queries say anything about primary or gallery sales.	Students wanting to include a low-volume artist in the table "for completeness" despite the threshold existing to protect against noise.
70 to 82	Small-group exercise	Students pick a different artist or a small cohort, reproduce the index and the demand signals, and flag which of their own recent periods should be treated with caution.	Groups that pick an artist with too few lots, or that skip flagging the recent, provisional periods.
82 to 90	Class discussion and wrap-up	Groups report what their index and demand signals show. Restate the central caveats and assign the homework.	Leave real time for the recency and secondary-market-only caveats to land, not just the mechanics of the queries.

Guidance for the discussion questions

Use these as the "what to listen for" behind each question in the module. They are talking points, not a graded key.

- 1. What a loosely defined comparable set breaks.** Listen for: a loose artist filter mixes in a different maker or a misattributed lot, a loose date range blends periods that should be read separately, and a loose status filter can quietly include unsold lots in what should be a sold-only price figure. You would know it had gone wrong because the count of matching rows would look larger or stranger than expected, and a spot check of a few rows would surface names, dates, or statuses that do not belong.
- 2. The two reasons the index moves, and why the count column matters.** Listen for: the index can move because the artist's market genuinely shifted, or because the mix of works that happened to sell that period changed, for example a run of major paintings selling one year and mostly minor prints the next. Showing the lot count next to the median lets a reader see when a big swing in the median lines up with a small or unusual count, which is the tell that composition, not the market, is driving the number.
- 3. Why the index measures method and composition, not pure appreciation.** Listen for: the median summarizes whatever happened to sell in a period, and what sold is not a random or fixed sample of the

artist's output, so a change in the index reflects both any real market movement and the changing set of works offered. Separating the two would require controlling for the works' characteristics period to period, which the hedonic-style thinking in Query 3 only gestures toward with a single rough keyword.

4. What sell-through and bought-in rate add beyond a price index. Listen for: these two signals describe whether buyers are actually showing up to bid regardless of the price level, so they can confirm or contradict a price-based story, for example a rising median alongside a falling sell-through would flag that fewer, pricier lots are selling while more are being turned away. They cannot say anything about how much a work sold for or whether the price level itself is high or low.

5. What the hedonic-style keyword split misclassifies. Listen for: item_title text can omit the medium entirely, use a term the keyword list does not anticipate, or use loose or promotional language that does not map cleanly to print, photograph, or painting, so some lots land in "other or unspecified" and some are misbucketed. You would check the error rate by pulling a sample of rows in each bucket and reading the titles directly against the designer's known body of work.

6. Why the cohort comparison needs a minimum-lot threshold. Listen for: without a threshold, an artist with only a handful of lots can show an extreme median or sell-through purely from small-sample noise, and placing that figure next to a well-supported artist's number implies a false equivalence. To a client who wants every artist shown regardless of count, explain that the table would then mix reliable and unreliable figures without any way for the reader to tell which is which, which is worse than omitting the unreliable ones.

7. Why the newest periods are provisional, not a current trend. Listen for: the newest one or two periods are still being ingested, so their lot counts are understated and their medians are computed on an incomplete sample, making them not comparable to a median computed on a complete period. To a client who wants "this year's number," explain that the honest answer is a provisional figure clearly labeled as such, not a number presented with the same confidence as a completed year.

8. What auction data alone cannot answer. Listen for: any question about primary-market or gallery activity, such as a work's original sale price at a gallery, a living artist's current gallery representation and pricing, or private treaty sales that never come to auction. Answering those would require a different source entirely, for example gallery price lists, artist estate or foundation records, or private sale disclosures.

Teaching notes

- **Most common misconception:** students read a rising median index as proof an artist's market is appreciating. Fix: always pull up the count column next to the median in the same view and ask the class whether the count changed enough to explain the move on its own before allowing any appreciation claim.
- **Second misconception:** students treat the newest one or two periods in the index as the current trend, because the number is right there on screen. Fix: before showing the class the full index, tell them in advance which periods are still being ingested, then let them find the same instability themselves when they look at the count column for those periods.
- **60 minute variant:** compress the opening discussion to 8 minutes, run only Query 1 and Query 2 live as the guided demo, and fold Query 3 and Query 4 into the homework as a take-home extension rather than a live demo.
- **90 minute variant (as scripted above):** run all four guided queries live, keep the full small-group exercise, and use the extra time to have one group present the API extension results to the class as a preview of Module 8's visualization work.

- **If the sandbox is slow or blocked on the room network:** fall back to the pre-run screenshots you captured in preparation for all four queries, walk the class through the count column and the caveats from the screenshots, and have students run the live queries themselves as part of the homework instead.

For the full lesson content, queries, and homework, see Module 16. Questions or a class API key: info@altfndata.com.