

Study sheet: Module 5, library data literacy workshop

This workshop is a general-audience introduction to reading and questioning a real dataset, designed for a public or academic library setting rather than a credit-bearing course. Participants need no background in finance, statistics, or computers beyond comfort using a web browser. Working entirely in a no-code sandbox, they explore a dataset of auction and resale transactions for watches, jewelry, handbags, and other luxury goods, drawn from over 100 auction houses with history to the late 1990s, and practice the basic moves of data literacy: asking a question, finding the right table and fields, running a filter, and reading the result honestly, including its limits.

What you should be able to do

1. Navigate a real, professionally maintained dataset using a point-and-click browser tool, with no coding required.
2. Explain what a data dictionary is and use one to understand what a field in a dataset actually measures.
3. Filter a dataset to answer a simple, self-generated question (for example, what has a particular brand sold for recently).
4. Read a chart built from filtered data and describe, in plain language, what it does and does not show.
5. Identify at least one reason to be cautious about a data-driven conclusion, such as an incomplete category or an unstable recent period.

Key terms

- **Field:** a single column of information in the dataset, such as a sale date or a price, defined plainly in the data dictionary.
- **Filter:** a rule that narrows the dataset to only the rows meeting a condition, such as one brand or one status.
- **Data dictionary:** a plain-language guide to what each field actually records, used before trusting any number pulled from it.
- **Coverage browser:** the sandbox tool for checking how many records and vendors exist for a brand or category before filtering to it.
- **Status:** the field marking whether a listed item actually sold or went unsold, the difference between an asking price and a real one.
- **Realized price:** what a buyer actually paid for an item, as opposed to what a seller asked or an auction house expected.
- **Vendor:** the auction house or platform where an item was sold.
- **Recency gap:** the tendency for the newest few months in the dataset to look thin simply because collection has not caught up yet, not because the market slowed.

Core sandbox query

The query the module is built around. Run it, then change the brand or category and compare.

```
SELECT designer, item_title, sale_date, usd_price_decimal
FROM all_handbags_data
WHERE designer LIKE '%Hermès%'
      AND status = 'sold'
ORDER BY sale_date DESC
LIMIT 25;
```

Common mistakes to avoid

1. Treating an asking price as if it were a sale price, rather than checking status for "sold" first.
2. Reading a dip in the most recent months as a market trend, when it is usually a recency gap in data collection.
3. Trying a brand name once in the coverage browser, finding few results, and concluding there is no data, instead of trying a shorter or differently spelled version of the name.
4. Reading a chart of one filtered brand or category as if it described the whole market, rather than just that slice.
5. Forgetting what filter produced an exported CSV, so the saved file becomes hard to interpret later.
6. Assuming a higher record count for one brand means it is more popular, rather than simply better covered by the auction houses in the dataset.

Full lesson, more queries, and the homework are in Module 5 at academy.altfndata.com. Questions: info@altfndata.com.