

Problem set 1: Alternative data in finance

Module 1. Total: 100 points. Format: no-code, entirely in the sandbox.

Before you start

Register at sandbox.altfndata.com with your work or school email address. Registration is auto-approved, so you can begin the moment you submit the form. You will not need an API key for this problem set. Keep the coverage browser, the schema tab, and the SQL editor open in separate tabs as you work.

Problem 1 (10 points): Access and surfaces

ALT/FNDATA offers three access surfaces: the sandbox, the documentation site, and the production API. In two to three sentences, describe what each surface is for and which one you, as a student, will use to complete this problem set. Name the specific site for each surface.

Problem 2 (15 points): Reading the coverage browser

Open the coverage browser in the sandbox and answer each of the following in one sentence:

- a. Roughly how many transaction records does the platform hold, and how far back does the history go?
- b. Name four categories of physical luxury assets covered by the platform.
- c. Roughly how many distinct brands are represented, and roughly how many auction houses and marketplaces feed the dataset?

Problem 3 (25 points): The data dictionary

Open the schema tab for a category of your choice (for example watches) and locate the following documented fields: `designer`, `model`, `item_title`, `sale_date`, `usd_price_decimal`, `sale_estimates_high_usd_price`, `status`, `vendor`, `stock_ticker`.

- a. For each of the nine fields above, write a one line plain English definition in your own words (do not copy the schema description verbatim).
- b. Which one field lets you connect a single auction record to a publicly traded company? What is the trading identifier called in this dataset (name the field, not an example value)?

Problem 4 (20 points): Your first sandbox query

In the sandbox SQL editor, write and run a query against the table for your chosen category that returns the `item_title`, `designer`, `sale_date`, `status`, and `usd_price_decimal` for the 10 most recent sold records. Paste your SQL below and describe, in one or two sentences, what the result set shows you.

```
-- your query here
```

Problem 5 (15 points): Why this counts as alternative data

In 150 to 250 words, explain why auction and resale pricing data on physical luxury assets qualifies as alternative data rather than traditional financial data. Your answer should address at least two of the following: the source of the data (auction houses and resale marketplaces rather than filings or exchanges), the kind of signal it can produce that a balance sheet cannot, and one limitation an analyst should keep in mind when using it.

Problem 6 (15 points): Point-in-time correctness

Every record carries a `sale_date` and a status of sold or unsold. Explain, in three to four sentences, why an analyst building a time series from this data must always filter and group by `sale_date` rather than by the date the record was pulled or ingested, and why the status field matters when computing a price based metric.

Then write a single SQL query that counts how many records in your chosen category have status equal to unsold.

-- your query here

Submission. Turn in this file with your written answers and SQL filled in, or paste your work into your course's submission portal if your instructor has set one up.