Table 9: Data source (DBnomics) example.

Field	Entry
id	meteofrance_TEMPERATURE_celsius.07005.D
source	dbnomics
combination_of	N/A
question	What is the probability that the daily average temperature at the French weather station at Abbeville will be higher on resolution_date than on forecast_due_date?
background	The history of Average temperature by day and by station for France - Degree Celsius - ABBEVILLE - Daily from Météo-France is available at https://db.nomics.world/meteofrance_TEMPERATURE_celsius.07005.D.
market_info_ resolution_ criteria	N/A
market_info_ open_datetime	N/A
market_info_ close_datetime	N/A
url	https://db.nomics.world/meteofrance_TEMPERATURE_celsius.07005.D
resolution_ criteria	Resolves to the value found at https://db.nomics.world/meteofrance_TEMPERATURE_celsius.07005.D once the data is published.
freeze_datetime_ value	17.95
freeze_datetime_ value_explanation	The daily average temperature at the French weather station at Abbeville.
freeze_datetime	2024-07-12T00:00:00+00:00
source_intro	DBnomics collects data on topics such as population and living conditions, environment and energy, agriculture, finance, trade and others from publicly available resources, for example national and international statistical institutions, researchers and private companies. You're going to predict how questions based on this data will resolve.
resolution_dates	["2024-07-28", "2024-08-20", "2024-10-19", "2025-01-17", "2025-07-21", "2027-07-21", "2029-07-20", "2034-07-19"]

Data collection 39 superforecasters provided forecasts, rationales, and additional information about they way they forecast in our latest survey round. We will further manually check all of their responses to ensure the anonymity of the dataset.

B.2.3 LLM FORECAST SET

This dataset provides the same data (aside from user_id) as outlined in Table 11 and Table 12, only provided by language models. Each individual .json file was created by a model for the given question set.

Data collection Beyond informing teams their forecasts will be made public, we will not check the rationales.

Table 10: Combination (DBnomics) example.

Field	Entry
id	["meteofrance_TEMPERATURE_celsius.07117.D", "meteofrance_TEMPERATURE_celsius.07240.D"]
source	dbnomics
combination_of	[An array containing dictionary entries of both questions.]
question	We are presenting you with two probability questions. Please predict the probability that both will happen, that one will happen but not the other, and that neither will happen. In other words, for each resolution date please provide 4 predictions.
background	N/A
market_info_resolution_criteria	N/A
market_info_open_datetime	N/A
market_info_close_datetime	N/A
url	N/A
resolution_criteria	N/A
freeze_datetime value	N/A
freeze_datetime value_explanation	N/A
freeze_datetime	2024-07-12T00:00:00+00:00
human_prompt	We are presenting you with two probability questions. Please predict the probability that both will happen, that one will happen but not the other, and that neither will happen. In other words, for each resolution date please provide 4 predictions.
resolution_dates	["2024-07-28", "2024-08-20", "2024-10-19", "2025-01-17", "2025-07-21", "2027-07-21", "2029-07-20", "2034-07-19"]

Table 11: Data dictionary of headers for forecast set.

Field	Description	Required	Data Type
organization	The organization name as it should be displayed on the leaderboard.	✓	string
model	The model name as it should be displayed on the leader board.	✓	string
question_set	The name of the question set file these forecasts are associated with.	✓	string
forecast_due_date	The date the forecasts were due in ISO 8601 format YYYY-MM-DD.	✓	string
forecasts	All forecasts for this question set.	✓	array<object>

Table 12: Public forecast set data dictionary of entries in `forecasts` array from Table 11.

Field	Description	Required	Data Type
id	A unique identifier string given `source`. If instead of a string it's an array of strings, then this is a combination question.	✓	string | array<string>
source	Where the data comes from.	✓	string
forecast	The forecast $\in [0, 1]$.	✓	number
resolution_date	The resolution date this forecast corresponds to. `null` for market questions.	✓	string |null
reasoning	The rationale underlying the forecast. During data anonymization, we insert `[redacted to maintain anonymity]` wherever text has been redacted.	✓	string
direction	If `id` has an array value, this is an array of the same length. Each entry is an integer $\in \{-1, 1\}$. If the value is 1, the question was asked in the normal direction. If the value is -1, the question was negated in the combination question e.g., for a question asking for $P(\neg Q1 \cap Q2)$, the value would be $[-1, 1]$. All possible values are: $[1, 1]$, $[-1, 1]$, $[1, -1]$, $[-1, -1]$, and `null`.	✓	array<number> |null
user_id	A randomly generated string associated with the human respondent who submitted the forecast. This value contains no information which could identify said participant and was assigned to the dataset after personal identifiers had been removed. Only required for participants from the general public and superforecaster surveys.		string

Table 13: Superforecaster forecast set data dictionary of entries in `forecasts` array from Table 11 (additional fields to those in Table 12).

Field	Description	Required	Data Type
searches	An array of search terms used in researching the topic.	✓	array<string> |null
consulted_urls	A list of useful URLs	✓	array<string> |null