

Will New York’s high temperature land at 79–80°F?

Expected Parrot

Open-source research tools

Vorhersage | Forecast research report

2026-09-16T13:29:10.549914Z

Assessment

Our estimate was 28.6%—about three chances in ten. We completed the forecast at 8:26 a.m. EDT on September 16, 2026, before seeing the market price and before the day’s high was known. It was close to Kalshi’s earlier 28% quote. That agreement is encouraging for this case, but the forecast still rests on an imperfect match between our weather data and the contract’s settlement source.

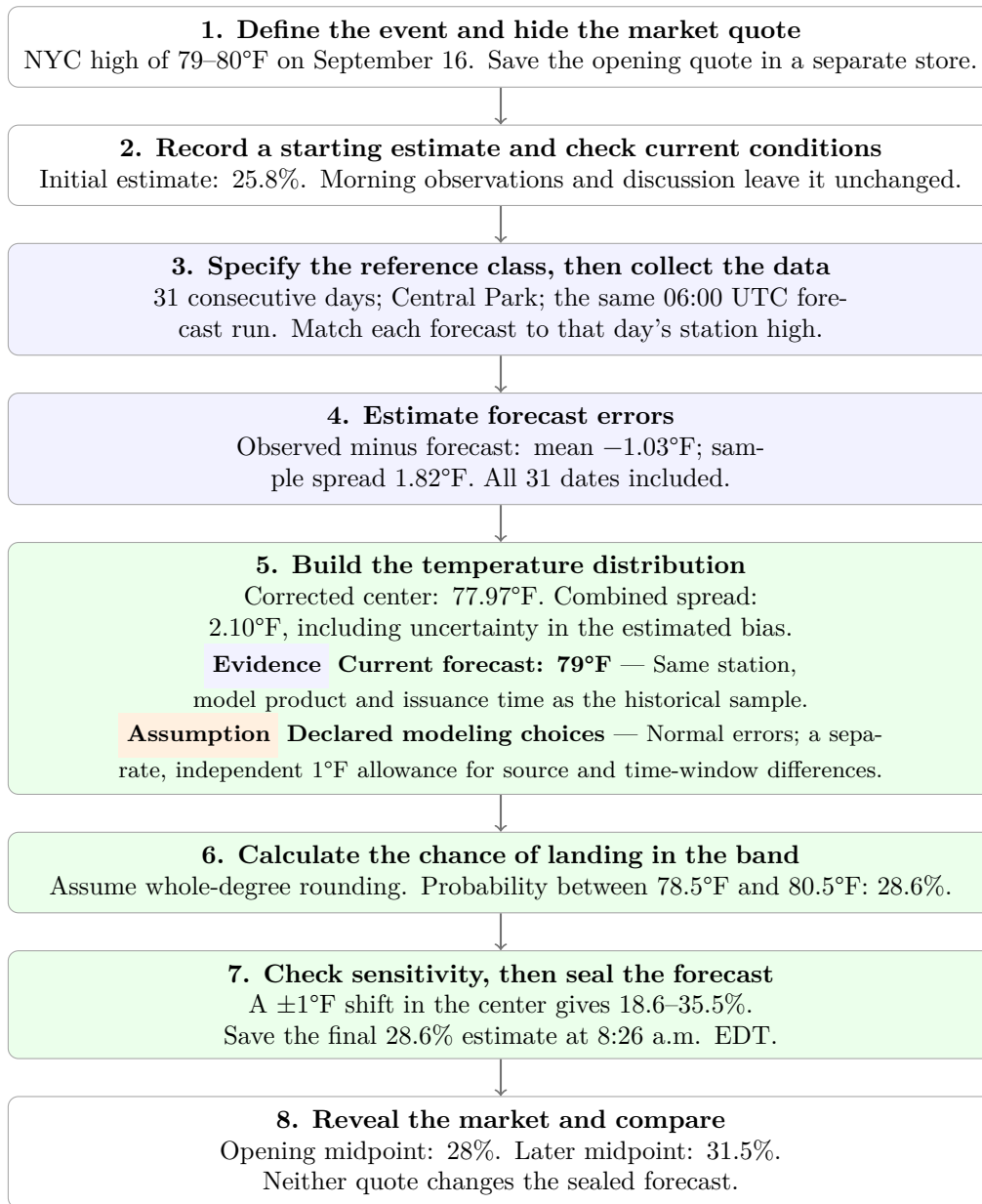
What would count as a Yes?

The question concerns the maximum temperature reported for New York City (CLINYC) on September 16, 2026. The contract pays Yes if that maximum is 79°F or 80°F; an ordinary final value outside that range pays No.

The Weather Company’s final report determines settlement. A National Weather Service forecast helps us predict the temperature, but does not determine the result. This distinction matters when a one-degree difference can move the outcome into or out of the winning band.

Methodology at a glance

Read from top to bottom. The market quote is held outside the forecasting process until the estimate is sealed. Blue boxes identify evidence; amber identifies assumptions; green marks calculated results or a completed decision.



This chart reconstructs the completed workflow. The sample window was specified before collecting historical outcomes; the distribution and extra uncertainty were chosen after inspecting the sample, but before revealing market prices.

How the estimate changed

We started with a weather forecast, then investigated two specific uncertainties: whether the morning conditions changed the outlook, and how much error to expect in a same-day high-temperature forecast. The market price stayed hidden until the final estimate was saved.

Stage	Estimate	What changed
Starting model	25.8%	An 80°F forecast with an assumed 3°F error spread.
Morning weather check	25.8%	No quantified reason to change the model.
Historical error analysis	28.6%	A narrower error spread, offset by a correction for forecasts running warm.

Method: selection, timing, and price blinding

This was a prospective, one-question development exercise. The target was the market’s opening probability, not the eventual temperature outcome. We selected an unresolved contract, saved its quote separately, recorded an initial estimate, declared each research step, and sealed the final estimate before revealing the quote. We did not use the day’s eventual maximum.

Selection was practical rather than random. An initial contract for a high above 82°F failed the two-sided quote check. We switched to the 79–80°F bracket, which contained the NWS point forecast. That bracket initially failed a ten-contract depth requirement; we retried with a one-contract minimum. No prices were displayed by those failed checks. This selection history limits how representative the example is of other markets.

The quote was captured at 8:18 a.m. EDT. The initial estimate at 8:19 already used a weather forecast encountered during screening, so it was not a judgment made before any research. The 31-day historical sample was specified at 8:20, before its outcomes were collected. The final estimate was sealed at 8:26; the market comparison was revealed at 8:30.

The same assistant context conducted the work. Prices were kept in a separate store that was not inspected during research; no fresh model context or enforced access barrier was used. The historical sample window and forecast issuance time were declared before extraction. The normal distribution and extra 1°F uncertainty allowance were chosen after examining the sample, although still before seeing market odds. Thus the entire statistical procedure was not specified in advance.

The morning check did not warrant a revision

The available Central Park observations showed overnight and early-morning temperatures around 61–64°F. The afternoon high was still ahead. The local forecast discussion supported fair weather and did not give us a numerical reason to change the starting estimate.

This research was useful for checking that the event remained unresolved and the forecast was plausible. It did not justify increasing confidence simply because several related weather products agreed.

Source: NWS Central Park hourly observations

Source: NWS New York area discussion issued 07:02 EDT September 16

Method: constructing the reference class

We used the Iowa Environmental Mesonet archive of the National Blend of Models' short-range station forecasts, identified as NBS, for Central Park station KNYC. The historical window was every date from August 16 through September 15, 2026. The comparison holds the station and forecast issuance time fixed; it is not a collection of forecasts from different cities or different lead times.

For each date, we selected the 06:00 UTC run—2 a.m. EDT—and its maximum-temperature value reported at 00:00 UTC the following day. The product calls this field TXN. According to its definition, that maximum covers 12:00 UTC on the forecast date through 06:00 UTC the next day. The current September 16 forecast uses the same run time and field.

We joined each forecast to the same date's NYC station daily maximum from the IEM daily-summary archive. Missing forecasts or observations would have been excluded and listed. All 31 dates had both values; there were no exclusions. Every date received equal weight. We did not select only sunny days, forecasts near 79°F, or cases with small errors.

This matches station, date, and issuance time, but not necessarily the measurement window or provider. A station daily summary can cover different hours from TXN, and the contract settles from The Weather Company. The estimated errors may therefore combine forecasting error with measurement differences. We retrieved historical archives during the exercise; we did not establish that every historical value was an unrevised, as-issued vintage.

Example date	Forecast high	Observed high	Error
August 16	81°F	80°F	−1°F
August 17	84°F	81°F	−3°F
August 18	85°F	86°F	+1°F

Source: IEM archived NBS forecasts

Source: IEM station daily maxima

Source: NOAA NBM text product definitions

Source: IEM daily-summary definition

Method: estimating bias and error spread

For each paired day, error means observed high minus forecast high. A negative error means the forecast ran warm. For example, a forecast of 84°F followed by an observed 81°F gives an error of −3°F.

The 31 errors sum to −32°F. Their average is therefore $-32 / 31 = -1.0323^\circ\text{F}$. We apply the full average correction to the current 79°F forecast: expected high = $79 - 1.0323 = 77.9677^\circ\text{F}$. We did not partially shrink this correction toward zero or fit a trend.

To measure spread, we subtract the average error from each daily error, square those differences, sum them, divide by 30 (the sample size minus one), and take the square root. This sample standard deviation is 1.8163°F. It measures variation around the average error; it is not the average absolute error or the uncertainty of the estimated average itself.

Transporting these errors to September 16 assumes the recent month is informative about today's forecast. We did not fit separate effects for wind, cloud cover, season, or forecast temperature. The current weather model also reported a 2°F spread, but the final calculation uses the historical residual estimate. We did not multiply agreement between NWS products as if it were independent evidence.

Source: IEM archived NBS forecasts

Source: IEM station daily maxima

Source: Current KNYC NBS forecast

Source: NWS current Central Park grid forecast

Method: specifying the predictive distribution

We approximate the day's high with a normal, bell-shaped distribution. Its center is 77.9677°F after the bias correction. Its spread must cover a new day's forecast error, uncertainty in our estimate of the average bias, and differences between our data and the settlement process.

Let $s = 1.8163^\circ\text{F}$ be the sample error spread and $n = 31$ the number of days. The declared predictive variance is $s^2 + s^2/n + 1^2$. These terms are about 3.2989, 0.1064, and 1.0000 square degrees, respectively. Taking the square root of their sum gives a predictive standard deviation of 2.0989°F.

The s^2/n term is the usual variance of a sample mean under independent, similarly distributed errors. Adding it allows for our estimate of the average bias being imprecise. Adding the separate 1^2 term assumes an additional, independent error with mean zero and a 1°F standard deviation. This is why we combine squared spreads rather than simply adding 1°F to 1.8163°F.

The 1°F allowance is subjective. It does not demonstrate that provider or time-window differences are actually centered at zero. The normal curve is also an approximation: we did not establish normality, account fully for uncertainty in the estimated spread, or use a Student-t predictive distribution. Correlation among adjacent days could make s^2/n too small.

Model ingredient	Value	Basis
Current forecast high	79°F	Weather-model forecast for Central Park
Bias correction	−1.03°F	Average observed minus forecast temperature over 31 days
Historical error spread	1.82°F	Standard deviation in the paired sample
Extra source/window uncertainty	1°F	Subjective allowance
Final distribution	Mean 77.97°F; spread 2.10°F	Combination of the inputs above

Source: Current KNYC NBS forecast
Source: NWS current Central Park grid forecast
Source: IEM archived NBS forecasts
Source: IEM station daily maxima
Source: NOAA NBM text product definitions
Source: IEM daily-summary definition

Method: converting temperature into a contract probability

The contract is about a two-degree band, so a point forecast alone is insufficient. We need the fraction of the predictive temperature distribution that would be reported as either 79°F or 80°F.

Assuming rounding to the nearest whole Fahrenheit degree, that band corresponds to an underlying high from 78.5°F up to 80.5°F. This half-degree adjustment is a modeling assumption about reporting, not a verified property of The Weather Company’s settlement feed.

To evaluate the interval, subtract the model’s center and divide by its spread. The lower endpoint is $(78.5 - 77.9677) / 2.0989 = 0.2536$ standard deviations above the center; the upper endpoint is $(80.5 - 77.9677) / 2.0989 = 1.2065$. A standard normal cumulative distribution function gives the probability below each endpoint. Subtracting those two probabilities gives the probability inside the band.

Using the unrounded parameters, the calculation is $0.8861827 - 0.6000939 = 0.2860888$, or 28.6%. The implementation uses Python’s NormalDist cumulative distribution function. This is a forecast-error model, not a likelihood-ratio update or a multiplication of evidence weights.

Calculation	Probability
High below 80.5°F	88.62%
High below 78.5°F	60.01%
High between the two endpoints	28.61%

Source: IEM archived NBS forecasts
Source: IEM station daily maxima
Source: Current KNYC NBS forecast
Source: NOAA NBM text product definitions
Source: IEM daily-summary definition

Method: sensitivity checks and alternative estimates

Before revealing the market quote, we calculated how the result changes when the expected temperature moves by 1°F and when the extra provider/window uncertainty ranges from 0°F to 2°F. The rows below change one assumption at a time. They are alternative calculations, not separate research checkpoints, and their range is not a confidence interval.

We also tried a direct empirical check: add each of the 31 historical errors to today's 79°F forecast and count how often the resulting high is 79°F or 80°F. That requires an error of exactly 0°F or +1°F. Seven days satisfy that condition, giving $7/31 = 22.6\%$. This calculation reuses the discrete errors directly; it does not fit a bell curve or add the extra provider uncertainty.

The difference between 22.6% and 28.6% shows that the distributional choice matters. Neither estimate was selected by comparing it with the hidden market price. However, we did not compare candidate estimators on a separate validation sample, so we cannot claim the normal approximation was the better-calibrated choice.

Thirty-one neighboring days are a small sample, potentially correlated and limited to one part of the season. We did not perform a bootstrap, estimate an effective sample size, or validate against past TWC contract settlements. Those are remaining methodological gaps, not checks that silently passed.

Alternative assumption	Probability
Final declared model	28.6%
Expected temperature 1°F lower; same spread	18.6%
Expected temperature 1°F higher; same spread	35.5%
No extra provider/window uncertainty; same center	30.2%
Extra provider/window uncertainty of 2°F; same center	24.6%
Direct historical-error frequency; no normal curve or extra allowance	22.6%

Source: IEM archived NBS forecasts

Source: IEM station daily maxima

Source: Current KNYC NBS forecast

Source: NOAA NBM text product definitions

Source: IEM daily-summary definition

What the market comparison tells us

At 8:18 a.m. EDT, the saved Kalshi quote was 27% bid and 29% ask, giving a midpoint of 28%. Our initial estimate was 2.24 percentage points away from that midpoint; the final estimate was 0.61 points away and inside the opening spread.

By the time we revealed the price at 8:30 a.m., the market had moved to a 31.5% midpoint. That 3.5-point market move was larger than our own 2.84-point revision. The later price is a separate observation, not a replacement for the original comparison target.

The package computes agreement as $(\text{forecast probability} - \text{opening midpoint})^2$. This fell from about 0.000500 to 0.000037, an improvement of 0.000463. These are squared differences between probabilities, not Brier scores against a resolved Yes/No outcome. The model research and the

later quote reflect different observation times; a sequential one-case exercise does not isolate a causal effect of research.

Liquidity was modest: roughly 24 contracts were available at each best quote when the target was saved, and only six at the best bid in the later snapshot. This is one case of closer agreement with an opening market quote, not evidence that the method is generally accurate.

The lesson for the next forecast

The most useful research targeted a number that drove the result: the size and direction of forecast errors. It replaced an unsupported assumption with a small, inspectable reference class. General weather corroboration did not change the probability.

For the next case, we should choose the error model and the treatment of settlement-source differences before inspecting the historical sample, and seek observations from the exact settlement source. We should then test that process against a new hidden market quote. Changing this forecast to match a price we have already seen would not be an independent test.

How to reproduce the calculation

The report's technical appendix contains the saved calculation file, all 31 forecast/observation pairs, exclusion list, and sensitivity results. The accompanying `analyze.py` reads the frozen NBS forecast and station-observation downloads, applies the date and run-time filters, constructs the paired errors, and recomputes the probabilities without making network requests.

The calculation file's SHA-256 hash was recorded in a finding before the forecast was sealed. The report includes that same file as a supplement. The methodology explanation here was expanded after the market reveal; it documents the original calculation and its limits rather than revising the forecast.

Source: IEM archived NBS forecasts

Source: IEM station daily maxima

Source: Current KNYC NBS forecast

Source: NOAA NBM text product definitions

About this report

This explanation was written at report generation from the recorded research. It does not change the sealed predictions. The complete evidence and calculation records are in the technical appendix.

A Technical appendix

Complete records follow. These do not change the assessment above.

Summary

Workbench workbench_306959beccd8453c97fb: latest estimate 28.6%; revealed.

Question and resolution

Created at: 2026-09-16T12:16:24.873252Z

Specification:

Domain: weather

Event deadline: 2026-09-17T05:00:00Z

Event group: KXHIGHTNY-26SEP16

Id: nyc_high_79_80_20260916

Kind: real

No: The final qualifying reported maximum lies outside the inclusive 79–80°F bracket.

Profile: general

Resolution source: The Weather Company final report for New York City CLINYC, per frozen Kalshi market rules; weather.com/kalshi. NWS forecasts are evidence, not the settlement authority.

Resolve after: 2026-09-18T05:00:00Z

Text: Will the maximum temperature reported for New York City (CLINYC) on September 16, 2026 be 79–80°F?

Void: Nonbinary last-fair-price settlement, unresolved material data error, or unavailable final source data is outside this binary comparison.

Yes: The Weather Company reports a final maximum in the inclusive 79–80°F bracket for CLINYC for September 16, 2026.

Version: 1

Workbench overview

Case id: workbench_306959beccd8453c97fb

State: revealed

Question version: 1

Method: NWS forecast plus explicit temperature-error model; sequential current-state and reference-class research

Mode: prospective

Latest estimate: 28.6%

Sealed at: 2026-09-16T12:26:32.228788Z

Market status: Revealed

Opening target captured at: 2026-09-16T12:18:15.716650Z

Frozen contract and case selection

Question:

Created at: 2026-09-16T12:16:24.873252Z

Specification:

Domain: weather

Event deadline: 2026-09-17T05:00:00Z

Event group: KXHIGHTNY-26SEP16

Id: nyc_high_79_80_20260916

Kind: real

No: The final qualifying reported maximum lies outside the inclusive 79–80°F bracket.

Profile: general

Resolution source: The Weather Company final report for New York City CLINYC, per frozen Kalshi market rules; weather.com/kalshi. NWS forecasts are evidence, not the settlement authority.

Resolve after: 2026-09-18T05:00:00Z

Text: Will the maximum temperature reported for New York City (CLINYC) on September 16, 2026 be 79–80°F?

Void: Nonbinary last-fair-price settlement, unresolved material data error, or unavailable final source data is outside this binary comparison.

Yes: The Weather Company reports a final maximum in the inclusive 79–80°F bracket for CLINYC for September 16, 2026.

Version: 1

Contract:

Event group: KXHIGHTNY-26SEP16

Market id: KXHIGHTNY-26SEP16-B79.5

Resolution source: See frozen Kalshi rules and series contract.

Rules: If the maximum temperature recorded at New York City (CLINYC) for Sep 16, 2026, is between 79–80° fahrenheit according to The Weather Company, then the market resolves to Yes.

Not all weather data is the same. While checking a source like AccuWeather or Google Weather may help guide your decision, the official and final value used to determine this market is the maximum/minimum temperature as reported by the Weather Company.

Reading weather data Preliminary Weather Company data may be subject to rounding and conversion differences from the final reported value. Use caution when interpreting preliminary Weather Company readings.

If it is determined in the sole discretion of the Exchange that there is a material error in the initial non-preliminary publication, Expiration may be held until the publication of a non-materially-erroneous data revision or until the Expiration Date, whichever comes first. If no data is available at the end of this period, all markets will resolve to the last fair price determined by Kalshi. Please note, a status of "Final" on weather.com/kalshi is provided for convenience only, and does not guarantee data to be error-free and thus exempt from the above.

Scheduled close: 2026-09-17T05:00:00Z

Title: Will the maximum temperature be 79-80° on Sep 16, 2026?

Venue: kalshi

Yes description: 79° to 80°

Selection:

Contract match rationale: Exact CLINYC September16 date and inclusive 79–80F bracket. TWC final value controls, with error-revision and nonbinary-settlement clauses preserved in frozen rules. The 05UTC following-day boundary is the exchange close; calendar-date assignment is governed by TWC.

Eligibility rationale: Selected at about 08:05 EDT before the afternoon maximum. NWS page shows 64F at 00:51 EDT and forecasts daytime high near80F. No qualifying observation or final result is known; first research step will check the elapsed-day record.

Market id: KXHIGNY-26SEP16-B79.5

Max spread: 0.05

Method: NWS forecast plus explicit temperature-error model; sequential current-state and reference-class research

Min contracts each side: 1

Mode: prospective

Question:

Question id: nyc_high_79_80_20260916

Version: 1

Venue: kalshi

Prediction history

Checkpoints are recorded judgments, not necessarily issued forecasts. All timestamps retain their recorded time zones.

Stage	Initial
Probability	25.8%
Recorded at	2026-09-16T12:19:15.466027Z
Gap to opening market (pp)	-2.24
Squared market error	0.000500
Step improvement	Not applicable

Stage	Research 1: Current state, latest forecast and settlement-station alignment
Probability	25.8%
Recorded at	2026-09-16T12:20:13.628648Z
Gap to opening market (pp)	-2.24
Squared market error	0.000500
Step improvement	+0.000000
Stage	Research 2: Empirical short-horizon high-temperature error distribution at Central Park
Probability	28.6%
Recorded at	2026-09-16T12:26:32.118256Z
Gap to opening market (pp)	+0.61
Squared market error	0.000037
Step improvement	+0.000463

Market comparison

Frozen opening midpoint: 28.0%; bid–ask 27.0%–29.0%. Captured 2026-09-16T12:18:15.716650Z.

Later midpoint: 31.5%; bid–ask 31.0%–32.0%. Captured 2026-09-16T12:30:37.889669Z.

Squared market error = (estimate – opening midpoint)². Positive step improvement means closer agreement.

Market agreement is not an outcome score. The later quote is separate from the frozen opening target.

Target:

Ask: 0.29

Ask contracts within 2c: 258.65999999999997

Ask size: 23.66

Bid: 0.27

Bid contracts within 2c: 148.12

Bid size: 24.0

Midpoint: 0.28

Spread: 0.02

Target captured at: 2026-09-16T12:18:15.716650Z

Target method: Opening order-book midpoint, frozen before initial assessment.

Net squared error improvement: 0.00046261464161725106

Followup quote:

Ask: 0.32

Ask contracts within 2c: 297.38

Ask size: 18.38

Bid: 0.31

Bid contracts within 2c: 125.45

Bid size: 6.0

Midpoint: 0.315

Spread: 0.01

Followup captured at: 2026-09-16T12:30:37.889669Z

Market movement pp: 3.4999999999999973

Followup error: Not recorded

Contract changed: No

Eligible for blind comparison: Yes

Qualification reasons:

None recorded.

Limitations:

Item 1.

Market agreement is not an outcome score or proof of forecast accuracy.

Item 2.

Step improvements are descriptive; sequential research does not identify causal effects.

Item 3.

Research may include information arriving after the opening target. Inspect market movement and timestamps.

Item 4.

No exposure reported is an agent declaration, not independently verified isolation.

Item 5.

Bid-ask spread is not a confidence interval. Missing follow-up data does not replace the opening target.

Research journal 1: Initial

Recorded 2026-09-16T12:19:15.466027Z · pre_reveal

Entry workbench_entry_b3f56559ce8048df8d75

Assumptions:

Item 1.

Forecast center80F from NWS

Item 2.

Residual standard deviation3F, subjective and deliberately broad

Item 3.

Approximately unbiased Gaussian errors and nearest-integer final reporting

Item 4.

NWS Central Park forecast is a usable proxy for TWC CLINYC measurement

Evidence refs:

Packet id	Record id
pkt_9fb3c2b218bac923948abc3c	point80

Probability: 25.8%

Rationale: NWS point forecast80F plus an explicitly assumed Gaussian residual SD3F gives $P(78.5 \leq \text{latent Tmax} < 80.5)$. This continuity correction approximates integer reported79–80F; provider rounding and residual SD are unverified. This is a starting model, not an empirical calibration.

Research status: in_progress

Uncertainties:

Item 1.

Current observed temperature and revised NWS forecast

Item 2.

Forecast residual spread at this station and short horizon

Item 3.

Provider/date/rounding alignment

Research journal 2: Plan

Recorded 2026-09-16T12:19:15.573541Z · pre_reveal

Entry workbench_entry__0853ad349b6240de8c2e

Higher if: Fresh forecasts cluster around79–80F and morning observations support that trajectory

Lower if: Latest forecast moves outside the band, or current maximum already exceeds80F

Search plan: Read current NWS Central Park hourly observations, point forecast and area discussion; inspect TWC source information without opening prediction odds.

Stopping rule: Stop after obtaining current station observations plus a fresh official point forecast and discussion, or documenting inaccessible sources.

Uncertainty: Current state, latest forecast and settlement-station alignment

Why it matters: The old80F center may be stale; an already-observed out-of-band high could resolve NO, and a wrong station or provider would invalidate the model.

Research journal 3: Checkpoint

Recorded 2026-09-16T12:20:13.628648Z · pre_reveal

Entry workbench_entry_45345efeb5184023a6c9 · Research plan workbench_entry_0853ad349b6240de8c2e

Changed assumptions:

None recorded.

Evidence refs:

Packet id	Record id
pkt_72933450ceaa54cf5a0ba4da	elapsed
pkt_72933450ceaa54cf5a0ba4da	regime

Findings: NWS observed 62.1F at 06:51 EDT; overnight readings shown are 61–64F. The 07:02 EDT discussion leaves the forecast broadly unchanged, with dry ridging and a possible southerly wind shift.

Limitations:

Item 1.

Text-only point-forecast and api.weather.gov opens failed in web retrieval. TWC page also inaccessible. Exact TWC daily rounding remains unverified. NWS hourly records do not fully substitute for the final TWC extreme.

Item 2.

Model/tool costs unmetered; no numeric usage counters claimed.

Probability: 25.8%

Rationale: Retain center 80F and SD 3F: current evidence confirms plausibility and eligibility but supplies no calibrated reason to change either number. Do not shrink uncertainty solely because two correlated NWS products agree.

Remaining uncertainties:

Item 1.

Uncalibrated 3F residual SD

Item 2.

TWC versus NWS measurement/rounding relationship

Item 3.

Exact afternoon sea-breeze effect

Sources checked:

Item 1.

<https://forecast.weather.gov/data/obhistory/KNYC.html>

Item 2.

<https://forecast.weather.gov/product.php?site=OKX&issuedby=OKX&product=A&format=CI&version=1&glossary=1>

Item 3.

<https://weather.com/kalshi>

Item 4.

<https://api.weather.gov/points/40.781,-73.966>

Research journal 4: Plan

Recorded 2026-09-16T12:20:13.745881Z · pre_reveal

Entry workbench_entry_569c6855c540473282cb

Higher if: Matched past forecasts exhibit smaller residual dispersion and little bias near the band

Lower if: Residuals are wider, biased away from the band, or station/provider alignment is poor

Search plan: Before inspecting outcomes, select all dates August16–September15 2026 with KNYC NBS/NBM06UTC same-day maximum forecasts and completed station daily maxima. Prefer an exact matched archive. If unavailable, inspect NBS current uncertainty or published verification; do not substitute model disagreement for calibrated error or count unrelated stations as matched cases.

Stopping rule: Inspect IEM MOS archive/API and NOAA NBM product definitions; attempt a matched local extraction. Stop after these sources or a clear availability failure, preserve exclusions and do not invent an empirical residual estimate.

Uncertainty: Empirical short-horizon high-temperature error distribution at Central Park

Why it matters: The assumed3F spread largely determines the chance of a two-degree bracket, even with an accurate80F center.

Research journal 5: Checkpoint

Recorded 2026-09-16T12:26:32.118256Z · pre_reveal

Entry workbench_entry_a04ef60f997746f9a889 · Research plan workbench_entry_569c6855c540473282cb

Changed assumptions:

Item 1.

Center now 77.967742 F after empirical bias correction to current NBS forecast.

Item 2.

Predictive SD now 2.098891 F, including a declared subjective 1 F provider/window error.

Item 3.

Evidence supports a narrower local residual distribution, but does not validate exact contract calibration.

Evidence refs:

Packet id	Record id
pkt_29f3824c0ae023774478baca	residuals

Packet id	Record id
pkt_29f3824c0ae023774478baca	current_high
pkt_29f3824c0ae023774478baca	mismatch
pkt_29f3824c0ae023774478baca	calculation

Findings: All 31 selected dates were available. NBS forecasts averaged 1.03 F above station daily highs, with residual SD 1.82 F. Current NBS and NWS forecasts are 79 F, with NBS XND 2 F. Narrower dispersion raises the band probability; correcting the warm bias lowers it. The net revision is +2.84 percentage points. TWC settlement and NBM/day-summary time windows remain imperfectly matched.

Limitations:

Item 1.

IEM station daily highs are a proxy, not TWC contract settlement records.

Item 2.

NBM maximum covers 12 UTC through 06 UTC; station daily summary windows can differ.

Item 3.

31 adjacent days are a small, serially correlated sample; no out-of-sample validation.

Item 4.

One-degree extra SD and normality are subjective; Gaussian smoothing differs from empirical frequency.

Item 5.

No adjustment for weather-regime or seasonal differences; no exact TWC rounding verification.

Item 6.

No exact matched TWC reference class. This is a proxy reference class with an explicit allowance, not the exact extraction originally preferred.

Item 7.

Research occurred in the existing root context, with evaluator files and outside event odds uninspected; no separate context was launched.

Item 8.

Usage and monetary cost are unmetered. The initial judgment already included the point forecast.

Probability: 28.6%

Rationale: Use the current 79 F NBS forecast plus the full predetermined 31-day mean residual (-1.0323 F), replacing the subjective 80 F center. Replace the unsupported 3 F spread with the 1.8163 F sample residual SD, a simple mean-estimation term $\hat{s}^2/n$, and an explicitly assumed independent 1 F provider/window SD. Gaussian integration over 78.5 to 80.5 F gives 28.61%. Normality, constant bias, nearest-integer rounding and the 1 F cushion are judgments, not calibrated facts. The un-smoothed residual estimate is 7/31 (22.6%); this difference warns against interpreting decimal precision as confidence.

Remaining uncertainties:

Item 1.

TWC versus IEM/NWS reporting and rounding.

Item 2.

Stationarity and serial dependence in 31 adjacent late-summer days.

Item 3.

Gaussian smoothing versus empirical distribution shape; changing mean by ± 1 F gives 18.6% to 35.5% with other final assumptions fixed.

Sources checked:

Item 1.

<https://mesonet.agron.iastate.edu/cgi-bin/request/mos.py?station=KNYC&model=NBS&sts=2026-08-16T00%3A00Z&ets=2026-09-16T00%3A00Z&format=csv>

Item 2.

https://mesonet.agron.iastate.edu/cgi-bin/request/daily.py?sts=2026-08-16&ets=2026-09-15&network=NY_ASOS&stations=NYC&var=max__temp_f&format=csv

Item 3.

<https://mesonet.agron.iastate.edu/api/1/mos.json?station=KNYC&model=NBS&runtime=2026-09-16T06%3A00Z>

Item 4.

<https://api.weather.gov/gridpoints/OKX/34,45/forecast>

Item 5.

<https://vlab.noaa.gov/web/mdl/nbm-textcard-v5.0>

Item 6.

<https://mesonet.agron.iastate.edu/cgi-bin/request/daily.py?help=>

Research journal 6: Finish

Recorded 2026-09-16T12:26:32.228788Z · pre__reveal

Entry workbench__entry__4689004d46c6405ba0b4

Exposure notes: No market price or outside probability for this event has been observed. Screening returned only terms or eligibility failures. Weather point forecasts and NBM dispersion are evidence used to build the event probability. Original root context was retained; evaluator store has not been read.

Market exposure: none

Outcome status: unresolved

Stopping reason: Completed current-state checks and the predetermined local reference-class extraction. Remaining TWC alignment and distribution-shape uncertainties are documented and sensitivity calculated; stop before revealing odds.

Research journal 7: Reflection

Recorded 2026-09-16T12:31:14.148966Z · post_reveal

Entry workbench_entry_afe82524faf8449f81d9

Next method change: On a new price-hidden case, specify the residual-distribution estimator and provider/window adjustment before reading reference-class outcomes. Prefer exact settlement-source observations and the same forecast issuance horizon. Save a deterministic extraction, exclusions, alternate-distribution sensitivity, and quote-age/depth alongside the comparison. Judge the method on fresh cases; do not retune this sealed forecast toward either revealed quote.

What did not: The final match does not validate the Gaussian choice or subjective 1 F provider allowance: empirical residual frequency was 22.6%, and a 1 F mean shift gives 18.6%-35.5%. Settlement-source mapping remains incomplete. The midpoint moved from 28% to 31.5% between snapshots, more than our 2.84-point revision. Selection was opportunistic and the initial 10-contract requirement failed before a retry with minimum1; the successful snapshot later had about24 contracts each side, while follow-up bid depth was6. Root context was retained; blinding is a declaration and file separation, not enforced isolation.

What helped: The planned 31-day extraction replaced an unsupported error spread with inspectable paired data and exposed a warm bias. The final 28.61% is closer to the frozen 28% midpoint than the initial 25.76%, and lies inside the opening 27%-29% spread. Recording an unchanged first checkpoint prevented qualitative corroboration from masquerading as numerical evidence.

Model and calculation

No structured model artifact was attached to this case. The recorded rationales and assumptions above are the available model description. Supplemental files, if supplied, are identified separately below.

Limitations and research cost

Case limitations:

Item 1.

Prices are separated from this project, not protected by an operating-system sandbox.

Item 2.

Selection checks spread and top-of-book size only; neither establishes market accuracy.

Item 3.

The agent must verify the event is still unknown; an open market alone cannot establish this.

Reported usage:

None recorded.

Usage note: Usage is agent-reported. Missing usage is unmetered, not zero; partial records are not a total.

Supplemental model or research file: calculation.json

Supplied at report generation. This is not automatically part of the sealed research record; its hash can be compared with a previously recorded commitment. No code in this file was executed.

Name: calculation.json

Sha256: 07b91b063ba56cfcebd1b8aa4d0445afed8fc21ca86ce886cd59ceb733a503dd

Provenance: report_time_attachment

Content:

Selection: All dates August 16 through September 15, 2026; KNYC NBS 06 UTC; next 00 UTC
TXN

Pairs:

Date	Forecast	Observed	Error
2026-08-16	81.0	80.0	-1.0
2026-08-17	84.0	81.0	-3.0
2026-08-18	85.0	86.0	1.0
2026-08-19	86.0	85.0	-1.0
2026-08-20	82.0	84.0	2.0
2026-08-21	78.0	79.0	1.0
2026-08-22	76.0	77.0	1.0
2026-08-23	81.0	80.0	-1.0
2026-08-24	79.0	79.0	0.0
2026-08-25	79.0	78.0	-1.0
2026-08-26	82.0	81.0	-1.0
2026-08-27	81.0	77.0	-4.0
2026-08-28	82.0	84.0	2.0
2026-08-29	80.0	76.0	-4.0
2026-08-30	82.0	79.0	-3.0
2026-08-31	79.0	78.0	-1.0
2026-09-01	83.0	82.0	-1.0
2026-09-02	74.0	71.0	-3.0
2026-09-03	84.0	84.0	0.0
2026-09-04	84.0	84.0	0.0
2026-09-05	80.0	79.0	-1.0
2026-09-06	74.0	76.0	2.0
2026-09-07	80.0	78.0	-2.0
2026-09-08	83.0	80.0	-3.0
2026-09-09	86.0	82.0	-4.0
2026-09-10	88.0	84.0	-4.0
2026-09-11	80.0	79.0	-1.0
2026-09-12	79.0	78.0	-1.0
2026-09-13	80.0	79.0	-1.0
2026-09-14	74.0	75.0	1.0
2026-09-15	72.0	71.0	-1.0

Excluded dates:

None recorded.

N: 31

Bias observed minus forecast f: -1.032258064516129

Residual sample sd f: 1.8162942303445209

Current nbs high f: 79.0

Current nbs xnd f: 2.0

Initial probability: 0.2576462938849167

Bias corrected gaussian probability: 0.3031132141751074

Empirical residual probability: 0.22580645161290322

Final mean f: 77.96774193548387

Final sd f: 2.0988905778994633

Assumed extra provider window sd f: 1.0

Final probability: 0.2860888041077261

Sensitivity:

Extra sd f	Mean shift f	Probability
0	-1	0.1753740704340384
0	0	0.30151285168854003
0	1	0.39687168270681006
1	-1	0.18648959266100196
1	0	0.2860888041077261
1	1	0.3554894481802137
2	-1	0.18955393786180796
2	0	0.24641942703785835
2	1	0.28154042981086763

Limitations:

Item 1.

IEM station daily highs are a proxy, not TWC contract settlement records.

Item 2.

NBM maximum covers 12 UTC through 06 UTC; station daily summary windows can differ.

Item 3.

31 adjacent days are a small, serially correlated sample; no out-of-sample validation.

Item 4.

One-degree extra SD and normality are subjective; Gaussian smoothing differs from empirical frequency.

Item 5.

No adjustment for weather-regime or seasonal differences; no exact TWC rounding verification.

Raw sha256:

Nbs current.json: 9ea1a97ea17ca142bb7a167a4f54c34617ed35fca7ac61e46841409ecf4e3709
Nbs history.csv: 50bba89c87866be68e277e9e845eadb25dba57be437f971d09d17e144b16033b
Nws forecast.json: 514ee022a33af97d7e5e8351a715902591d753ae4bb688842509369d44eddc03
Nws points.json: f1521160148001508aea1d4c0a2c84914d5a2310f46fd7950bd85db2f328dee6
Observed history.csv: ea9c4635a1f0dec2e4f7a6276bbe57d0cbc5d3086fa379f878fdabe5aca23fee

Finding: pkt_9fb3c2b218bac923948abc3c:point80

Sources: [1]

Claim: The pre-research screening forecast centers the September16 Central Park high near80F; overnight observation64F does not establish the daily maximum.

Claim type: official_statement

Entity ids:

None recorded.

Id: point80

Observed at: 2026-09-16T12:03:56Z

Provenance:

Adapter: vorhersage.research_bundle.v1

Value: Not recorded

Finding: pkt_72933450ceaa54cf5a0ba4da:elapsed

Sources: [2]

Claim: Observed overnight/morning temperatures do not establish the79–80F final-high event; afternoon maximum remains ahead.

Claim type: observation

Entity ids:

None recorded.

Id: elapsed

Observed at: 2026-09-16T12:20:13.440822Z

Provenance:

Adapter: vorhersage.research_bundle.v1

Value: Not recorded

Finding: pkt_72933450ceaa54cf5a0ba4da:regime

Sources: [3]

Claim: Fresh discussion supports the earlier fair-weather forecast but does not quantify station-specific temperature uncertainty.

Claim type: official_statement

Entity ids:

None recorded.

Id: regime

Observed at: 2026-09-16T12:20:13.440887Z

Provenance:

Adapter: vorhersage.research_bundle.v1

Value: Not recorded

Finding: pkt_29f3824c0ae023774478baca:residuals

Sources: [4], [5]

Claim: Predetermined 31-day proxy sample: observed minus NBS same-day forecast mean=-1.032258F, sample SD=1.816294F; no dates excluded. Forecasts averaged about 1F warmer than station daily highs.

Claim type: inference

Entity ids:

None recorded.

Id: residuals

Observed at: 2026-09-16T12:26:04.710249Z

Provenance:

Adapter: vorhersage.research_bundle.v1

Value: Not recorded

Finding: pkt_29f3824c0ae023774478baca:current_high

Sources: [6], [7]

Claim: Current NBS and NWS point forecasts both center near 79F; NBS XND is 2F. They are correlated sources, not two independent votes.

Claim type: observation

Entity ids:

None recorded.

Id: current_high

Observed at: 2026-09-16T12:26:04.710249Z

Provenance:

Adapter: vorhersage.research_bundle.v1

Value: Not recorded

Finding: pkt__29f3824c0ae023774478baca:mismatch

Sources: [8], [9]

Claim: The weather-model daytime maximum and IEM station daily maximum are only approximate matches to the TWC final daily contract. Provider/window/rounding uncertainty remains.

Claim type: inference

Entity ids:

None recorded.

Id: mismatch

Observed at: 2026-09-16T12:26:04.710249Z

Provenance:

Adapter: vorhersage.research_bundle.v1

Value: Not recorded

Finding: pkt__29f3824c0ae023774478baca:calculation

Sources: [4], [5], [6], [8]

Claim: Declared Gaussian approximation uses mean77.967742F, SD2.098891F including subjective1F provider/window error, and interval[78.5,80.5], yielding28.6088804%. Raw empirical residual frequency is7/31=22.580645%. Calculation JSON SHA256=07b91b063ba56cfcebd1b8aa4d0445afed8fc21ca86ce886cd59ceb733a503dd; raw inputs and executable analyze.py saved alongside case.

Claim type: inference

Entity ids:

None recorded.

Id: calculation

Observed at: 2026-09-16T12:26:04.710249Z

Provenance:

Adapter: vorhersage.research_bundle.v1

Value: Not recorded

Source [1]: NWS Central Park forecast, issued September15 20:23 EDT

Excerpt: September16 forecast high near80F with mostly sunny conditions and afternoon southwest wind near5mph. The page displayed KNYC64F at00:51EDT. The forecast was issued20:23EDT September15.

Excerpt kind: paraphrase

Id: nws_initial

Published at: 2026-09-16T00:23:00Z

Retrieved at: 2026-09-16T12:03:56Z

Title: NWS Central Park forecast, issued September15 20:23 EDT

Url: <https://forecast.weather.gov/MapClick.php?lat=40.781&lon=-73.966>

Source [2]: NWS Central Park hourly observations

Excerpt: September16 observations through06:51EDT range from61F to64F in the displayed hourly series;06:51 temperature62.1F. No displayed reading exceeds80F. The series is hourly, not a complete minute-by-minute daily extreme record.

Excerpt kind: paraphrase

Id: obs

Retrieved at: 2026-09-16T12:20:13.440822Z

Title: NWS Central Park hourly observations

Url: <https://forecast.weather.gov/data/obhistory/KNYC.html>

Source [3]: NWS New York area discussion issued07:02EDT September16

Excerpt: The morning update made no significant changes. A ridge supports mostly dry weather today; light southwest flow becomes more southerly in the afternoon. Timing of the metro-area wind shift may vary by one to two hours.

Excerpt kind: paraphrase

Id: afd

Published at: 2026-09-16T11:02:00Z

Retrieved at: 2026-09-16T12:20:13.440887Z

Title: NWS New York area discussion issued07:02EDT September16

Url: <https://forecast.weather.gov/product.php?site=OKX&issuedby=OKX&product=AFD&format=CI&version=1&glossary=1>

Source [4]: IEM archived NBS forecasts

Excerpt: All 31 selected dates have KNYC 06UTC forecasts, paired with next-day00UTC TXN for the same-day daytime maximum.

Excerpt kind: paraphrase

Id: history

Retrieved at: 2026-09-16T12:26:04.710249Z

Title: IEM archived NBS forecasts

Url: <https://mesonet.agron.iastate.edu/cgi-bin/request/mos.py?station=KNYC&model=NBS&sts=2026-08-16T00%3A00Z&ets=2026-09-16T00%3A00Z&format=csv>

Source [5]: IEM station daily maxima

Excerpt: 31 completed station daily maxima are available for the predetermined August16 through September15 period. These are not TWC contract settlements.

Excerpt kind: paraphrase

Id: observed

Retrieved at: 2026-09-16T12:26:04.710249Z

Title: IEM station daily maxima

Url: https://mesonet.agron.iastate.edu/cgi-bin/request/daily.py?sts=2026-08-16&ets=2026-09-15&network=NY_ASOS&stations=NYC&var=max_temp_f&format=csv

Source [6]: Current KNYC NBS forecast

Excerpt: The September16 06UTC run predicts a79F high with XND2F for the period reported at September17 00UTC.

Excerpt kind: paraphrase

Id: current

Retrieved at: 2026-09-16T12:26:04.710249Z

Title: Current KNYC NBS forecast

Url: <https://mesonet.agron.iastate.edu/api/1/mos.json?station=KNYC&model=NBS&runtime=2026-09-16T06%3A00Z>

Source [7]: NWS current Central Park grid forecast

Excerpt: API generated at10:30UTC September16 reports sunny with a high near79F today, south wind0to5mph.

Excerpt kind: paraphrase

Id: nws

Retrieved at: 2026-09-16T12:26:04.710249Z

Title: NWS current Central Park grid forecast

Url: <https://api.weather.gov/gridpoints/OKX/34,45/forecast>

Source [8]: NOAA NBM text product definitions

Excerpt: TXN maxima cover12UTC currentday through06UTC nextday and are reported at00UTC nextday. XND is the maximum/minimum temperature standard deviation.

Excerpt kind: paraphrase

Id: definition

Retrieved at: 2026-09-16T12:26:04.710249Z

Title: NOAA NBM text product definitions

Url: <https://vlab.noaa.gov/web/mdl/nbm-textcard-v5.0>

Source [9]: IEM daily-summary definition

Excerpt: Daily summaries mix computed values and explicit official totals. Airport station windows typically follow standard time and can differ from NBM daytime maximum windows.

Excerpt kind: paraphrase

Id: daily_definition

Retrieved at: 2026-09-16T12:26:04.710249Z

Title: IEM daily-summary definition

Url: <https://mesonet.agron.iastate.edu/cgi-bin/request/daily.py?help=>

Report provenance

This report renders stored records without new research, model calls, or probability revisions. Missing work remains missing. The JSON export preserves the complete report input.

Schema version: vorhersage.report.v1

Generated at: 2026-09-16T13:29:10.549914Z

Record sha256: bc1c5296a6b0f331db27d1d7cb9199cdc2ff8d456216d0d955f659549fac90ed

Format: latex

Latex engine: XeLaTeX or LuaLaTeX

Narrative sha256: b9fd54872d8cda0071bd9fe30e7da0617ad5988a3765a752e3e895169934627e