Overall Summary (Post data aggregation and analysis)

The data aggregation captured 414,833 hourly price observations across **630** prediction markets from ***Polymarket (384 markets)*** and ***Kalshi (246 markets)***. The data revealed several key characteristics of prediction market (*p.m*) dynamics that aren't entirely clear to the day-to-day *p.m* bettor.

Price Behavior: Markets are heavily skewed toward low probabilities at any point after the *p.m*'s inception. These probabilities have mean prices of 19% (Polymarket) and 26% (Kalshi). This reflects that most binary markets in the dataset involve events that are perceived as unlikely. Price trajectories show characteristic prediction market patterns; long periods of stability punctuated by sharp information-driven jumps (i.e. breaking news, quarterly reports, etc.), with gradual convergence toward 0 or 1 as market resolution approaches.

Return Characteristics: Returns are highly non-normal with extreme leptokurtosis[1] (kurtosis ~70,000 combined), indicating frequent small changes with occasional dramatic moves. This fat-tailed distribution is consistent with information arrival being, for lack of a better word, "lumpy" rather than continuous, a hallmark of event-driven markets.

Platform Differences:

Kalshi provides denser data per market (882 vs 515 observations) despite having fewer markets, suggesting its structured series-based approach (e.g. daily S&P predictions) generates more consistent time series.

Polymarket offers broader market coverage with more diverse topics but sparser individual market data (there are a ton of markets on ***Polymarket*** without much liquidity). This is likely due to the nature of ***Polymarket's*** building blocks, it's on the blockchain! yikes.

Trading Patterns:

Activity clusters around US evening hours (7pm-11pm EST), suggesting a primarily American retail user base engaging with the *p.m*'s after work hours. Weekend activity is slightly elevated, consistent with retail participation patterns.

Risk Profile:

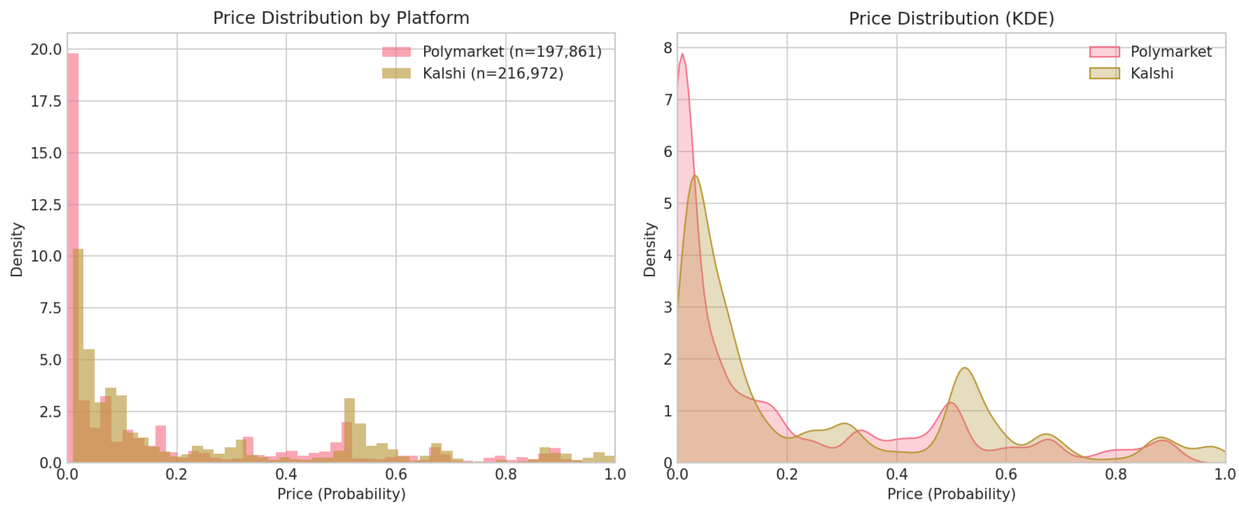
---

[1] Kurtosis is how "peaky" a distribution is. Sharp peak with a low std, we're probably looking at

Median market volatility is moderate (11.5% annualized) with typical drawdowns[2] around 6%, but tail risks are significant, some markets experience near-complete probability collapses[3], reflecting the binary nature of prediction market outcomes.

Plot Descriptions

1. **Price Distribution**



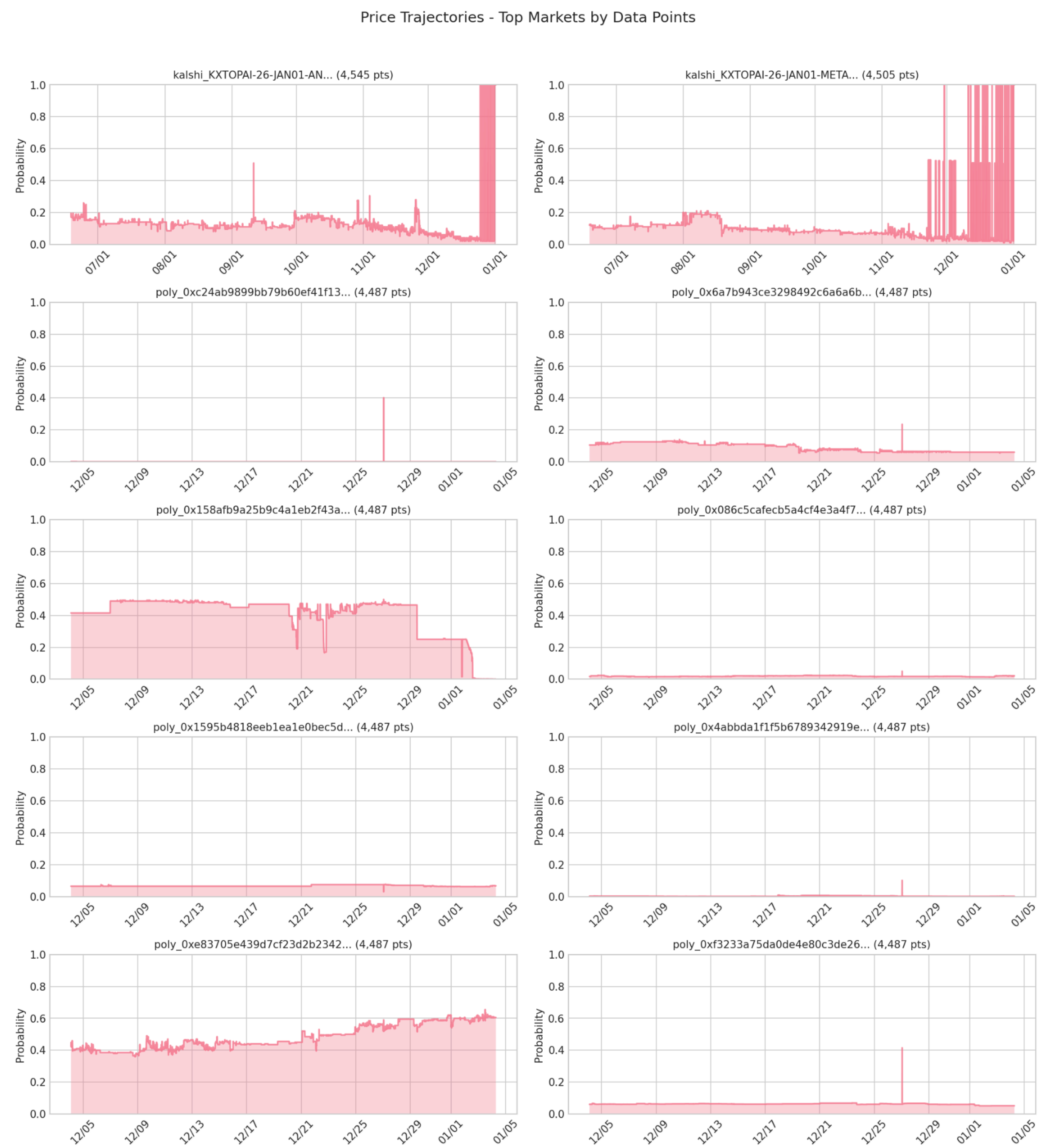
Left panel (Histogram): Shows the distribution of prices (probabilities 0-1) for both platforms. Both Polymarket (pink, n=197,861) and Kalshi (tan, n=216,972) show strong right-skew with most prices clustered near 0. Polymarket has an extremely tall spike near 0 or slightly above 0. Both ***Kalshi*** and ***Polymarket*** show a secondary mode[4] around 0.50.

Right panel (KDE): Smoothed density curves reveal Polymarket prices are heavily concentrated below 0.10 (probability <10%), while Kalshi has a more spread distribution with a notable bump around 0.45-0.55. This suggests ***Polymarket*** has more "long-shot" markets while Kalshi markets tend toward more uncertain outcomes.

It seems as if bettors are either really certain about a *p.m*'s market resolution or are uncertain about its resolution.

2. **Price Trajectories**

[2] peak-to-trough decline in value of an investment or portfolio
[3] It's often the case that near market resolution the collective is fairly confident how the market will ultimately resolve itself.
[4] A second "hump" appears in the original distribution.

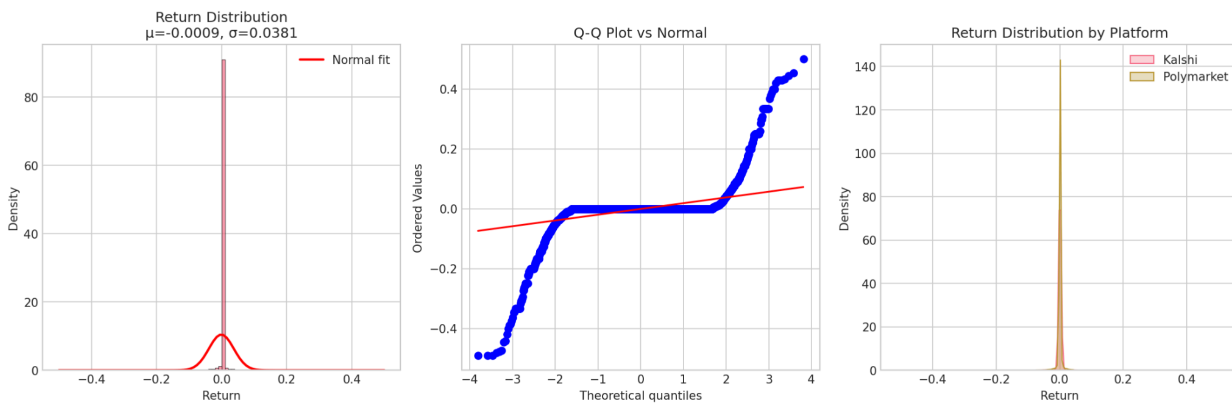


Shows price evolution over time for the top 10 markets by data density (~4,500 points each). Key observations:

- Top Row (***Kalshi***): Two S&P 500 index markets (***KXTOPM***) showing volatile intraday movements between probabilities (0.2-0.4)
- **Remaining panels (*Polymarket*)**: Various markets showing characteristic market patterns:

- Several markets show long periods of stability followed by sudden jumps (information shocks)
- Some markets trend gradually toward 0 or 1 as resolution approaches
- A few show high volatility throughout their entire lifecycle

3. Return Distribution



Quick clarification.
Here, in this context, "return" refers to the *percentage price change* between consecutive hourly observations, not profit/loss from betting. It's calculated as:

```python
return = (price_t - price_{t-1} / price_{t-1}
```

Left panel: Distribution of hourly returns centered near zero (μ=-0.0009, σ=0.0381). Extremely leptokurtic (peaked) with fat tails, most returns are near zero with occasional large moves.

Middle panel (Q-Q Plot)[5]: Compare return distribution to normal. The S-curve shows returns are far from normal, much heavier tails than Gaussian, with

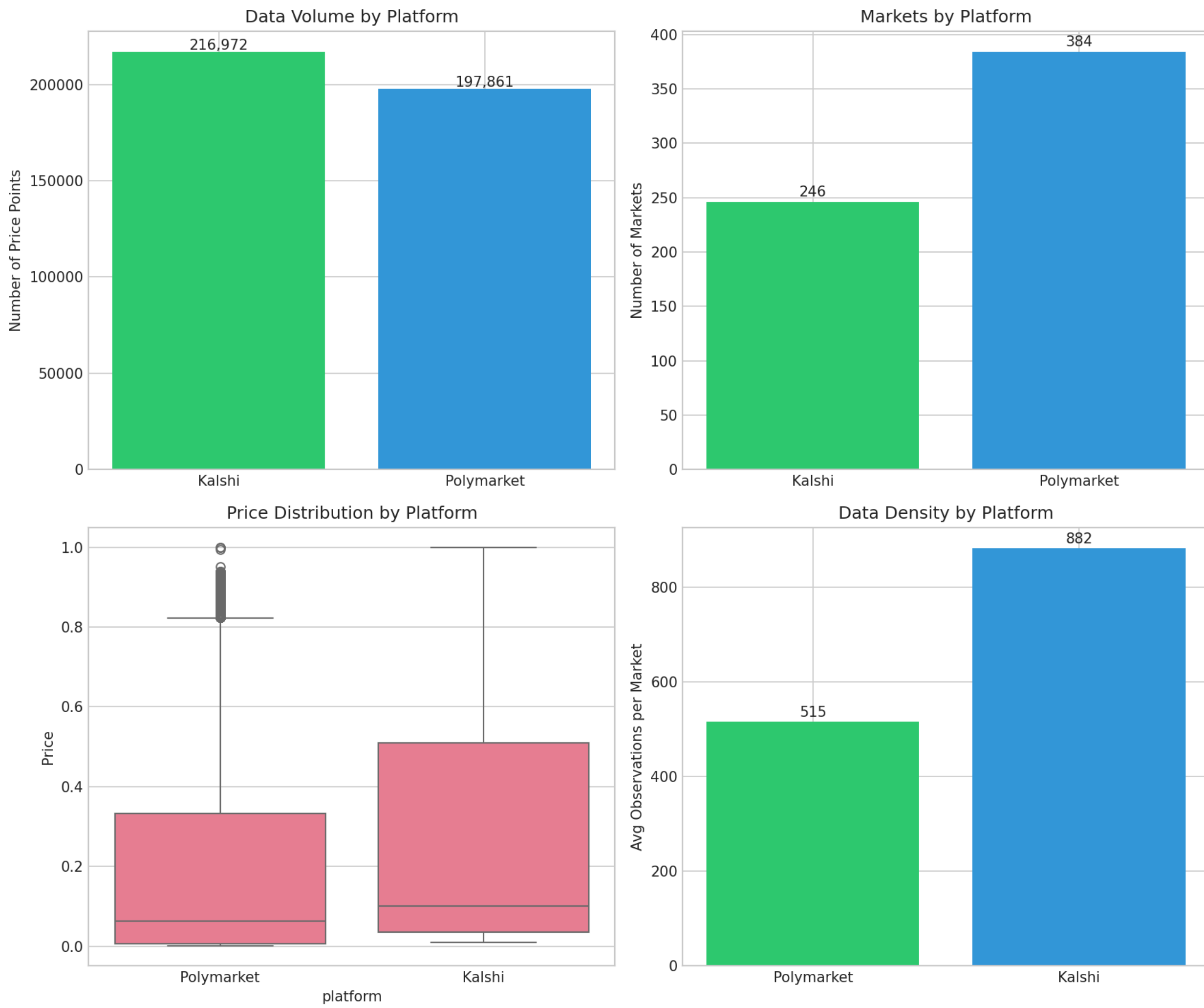
[5] A Quantile-Quantile plot compares two probability distributions by plotting their quantiles against each other. The x-axis shows theoretical quantiles from a reference distribution (here, normal/Gaussian), while the y-axis shows the actual quantiles from our data. If the data were perfectly normal, all points would fall on the diagonal red line. Deviations reveal how the distributions differ:

extreme positive and negative returns occurring more frequently than a normal distribution would predict. This is shown by all the blue points above and below the red fit line.

Right panel: Platform comparison shows both Kalshi and Polymarket have similar return distributions, though Kalshi appears slightly more dispersed.

Unlike equities which trade continuously with constant price discovery, prediction markets are event-driven. Prices only move when new information arrives about the underlying event. Between information shocks, there's no fundamental reason for prices to change, a *p.m* asking "Will X happen by December?" has no new information most hours, so the price stays flat. This creates extreme concentration at zero returns. The occasional large moves occur when actual news breaks (polls released, official announcements, unexpected developments), causing sudden probability reassessments. This "waiting for news" dynamic is fundamentally different from stock markets where earnings expectations, interest rates, and macro factors create continuous price pressure.

4. **Platform Comparison**



Top-left (Data Volume): ***Kalshi*** has slightly more price points (216,972) than Polymarket (197,861).

Top-right (Markets)*: Polymarket* has more unique markets (384) compared to ***Kalshi*** (246).

Bottom-left (Box Plot): Price[6] distributions differ, ***Polymarket's*** median price is very low (~0.05) with outliers extending to 1.0. ***Kalshi's*** median is higher (~0.10) with a wider interquartile range extending to ~0.35.

---

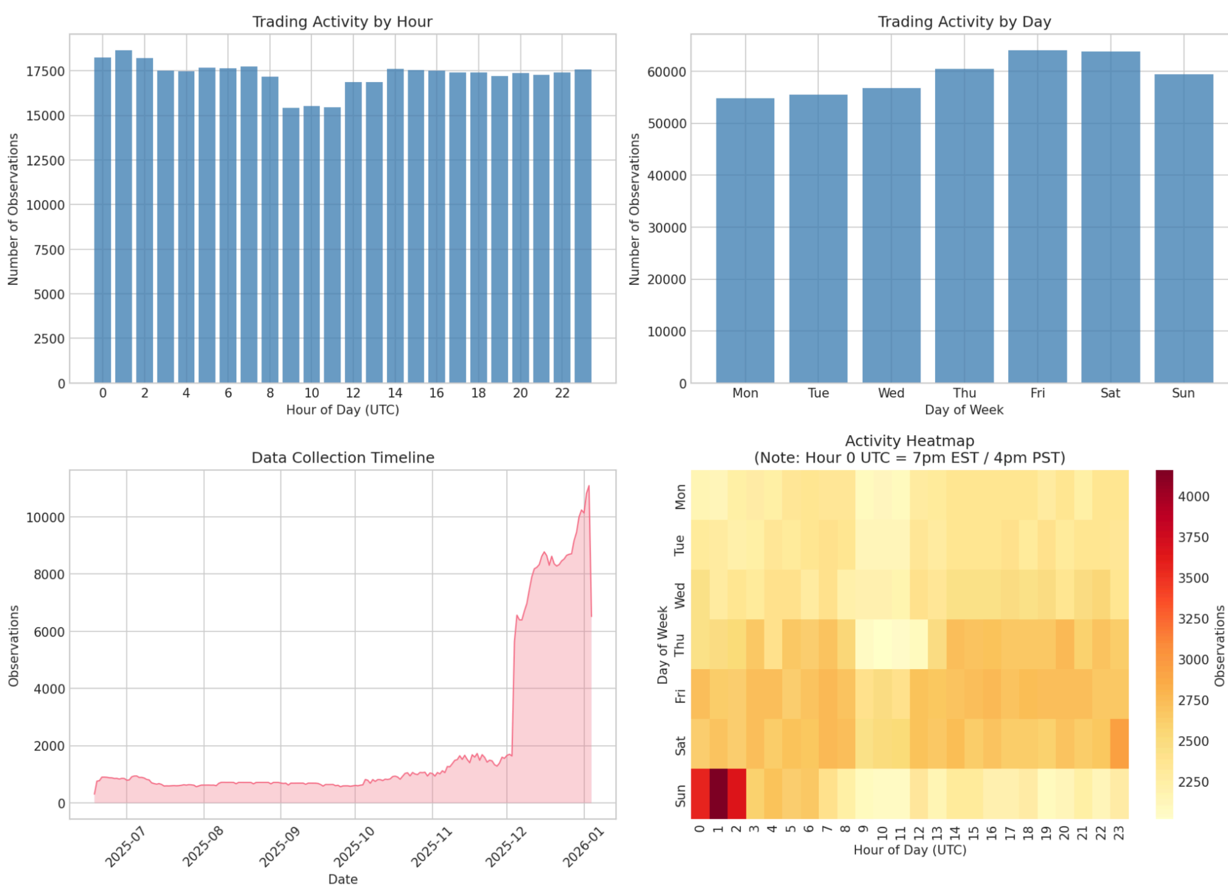
[6] In *p.m* "price" refers to the cost of one "Yes" share (or ticket), these prices range from $0.00 to $1.00. The price directly represents the market's implied probability.

Bottom-right (Data Density): ***Kalshi*** has more observations per market (882 avg) compared to ***Polymarket*** (515 avg)**,** suggesting ***Kalshi*** markets are tracked for longer periods or have more active trading.

Why are median prices so low? I can resolve this down to 3 potential reasons.

1. Resolved markets skew to "No": Many events don't happen (e.g., "Will X resign by January?" , typically not)
2. Long-shot markets: Platforms list many speculative events with low base rates (this is an especially prevalent issue on ***Polymarket*** since anyone can propose an event!
3. Selection bias: High-volume markets (which we filtered for) may include many "Will [unlikely thing] happen?" style questions

5. Time Patterns



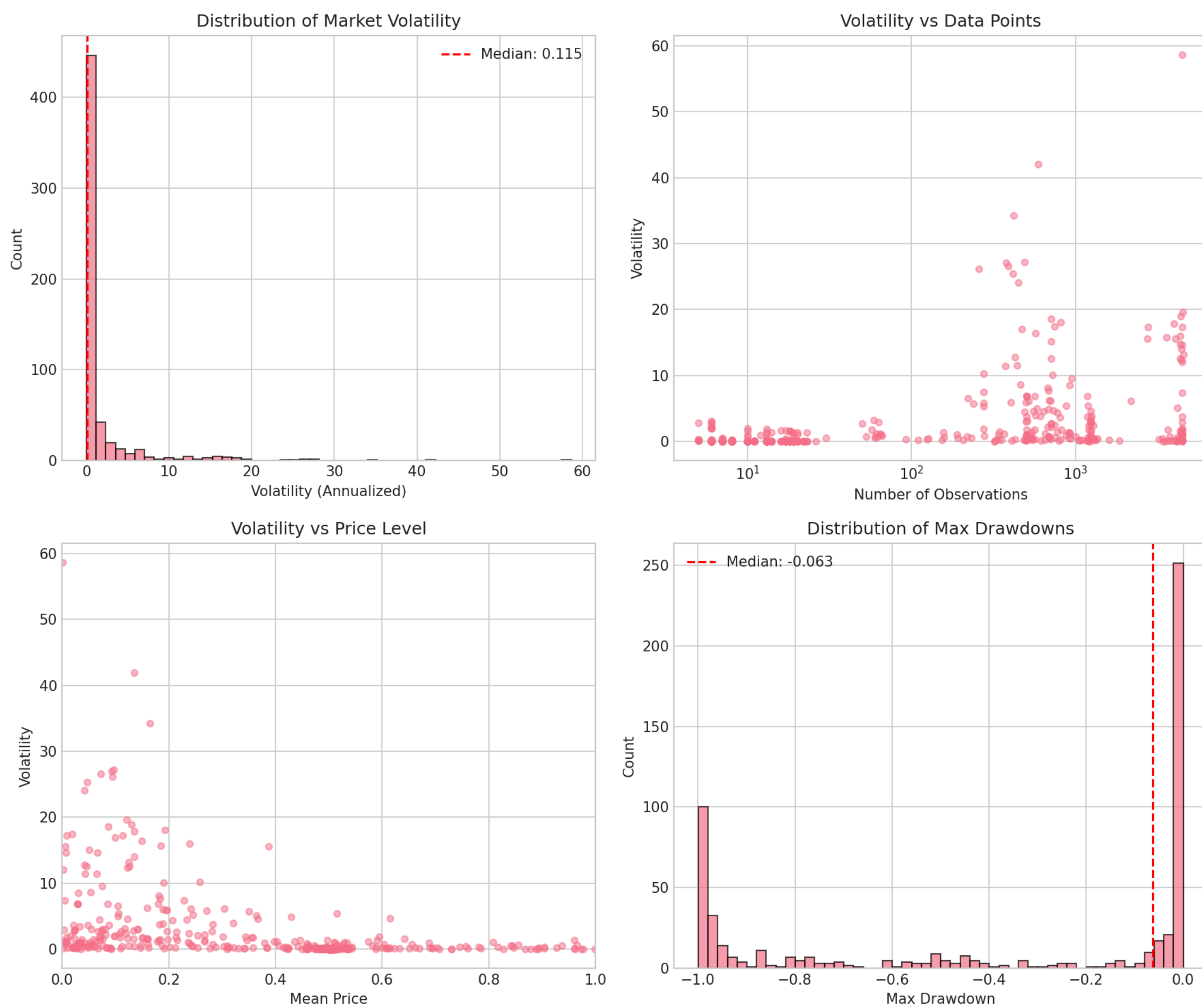
Top-left (Hourly): Activity peaks during UTC hours 0-4 and 14-18, which corresponds to US evening hours (7pm-11pm EST) and morning trading (9am-1pm EST).

Top-right (Daily): Relatively even distribution across weekdays, with slight peaks on Friday-Saturday and a dip mid-week.

Bottom-left (Timeline): Data collection ramps up dramatically around late December 2025 through January 2026, showing an exponential increase in observations.

Bottom-right (Heatmap): Confirms the hourly/daily pattern. Sunday shows the highest concentration at hour 0-2 UTC (Saturday evening US time). The note correctly identifies that Hour 0 UTC = 7pm EST.

6. Volatility Analysis



Volatility measures how much a price fluctuates over time, higher volatility means bigger, more frequent price swings. How it's calculated:

Calculate returns for each time period:

```Python
return = (price_t - price_{t-1 / price_{t-1}
```

Take the standard deviation of those returns:

```Python
standard_deviation = std(r_1, r_2, ..., r_n)
```

Annualize it by multiplying by √(periods per year)

Top-left (Volatility Distribution): Most markets have low annualized volatility (median 0.115), but a long right tail extends to 60+, indicating some highly volatile markets.

Top-right (Volatility vs Data Points): No strong correlation between number of observations and volatility. Markets with more data points tend to have moderate volatility (5-20 range).

Bottom-left (Volatility vs Price Level): Lower-priced markets (near 0) show higher volatility. Markets priced 0.3-0.7 tend to be less volatile, consistent with probability theory (variance maximized at p=0.5 for Bernoulli).

Bottom-right (Max Drawdown): Most markets experience modest drawdowns (median -6.3%), but some markets see extreme drawdowns approaching -100%, representing complete probability collapses.

What motivates me to pursue Equity Research at Fidelity is the opportunity to apply rigorous, first-principles analysis to real-world decision making at scale. My background is somewhat unconventional for asset management, but that is precisely what draws me to the role. As a deep learning researcher at the University of Chicago Geophysical Sciences Department, I study stochastic systems where noise, regime shifts, and uncertainty are unavoidable, skills that I think would translate naturally to analyzing markets, companies, and industries under imperfect information

In parallel, my work as an XLab Research Fellow mapping global AI data center infrastructure has exposed me to the intersection of technology, energy demand, capital investment, and geopolitical risk. Evaluating satellite data, infrastructure build-outs, and power constraints has sharpened how I think about long-term secular trends, capital intensity, and second-order effects, core questions in equity research. My experience as a Full-Stack Developer has also made me comfortable working directly with large datasets, building reproducible pipelines, and validating assumptions rather than relying on surface-level narratives.

Equity research appeals to me because it rewards intellectual honesty, probabilistic thinking, and the ability to turn technical detail into clear investment insight. Fidelity's long-term, fundamentals-driven approach supports how I already think about complex systems. This means understanding structure, stress-testing assumptions, and respecting uncertainty. Professionally, I see equity research as a path where I can combine deep analytical work with real economic impact, while continuing to grow my understanding of markets, industries, and capital allocation alongside experienced investors.