Anduril

Air Defense Case Study

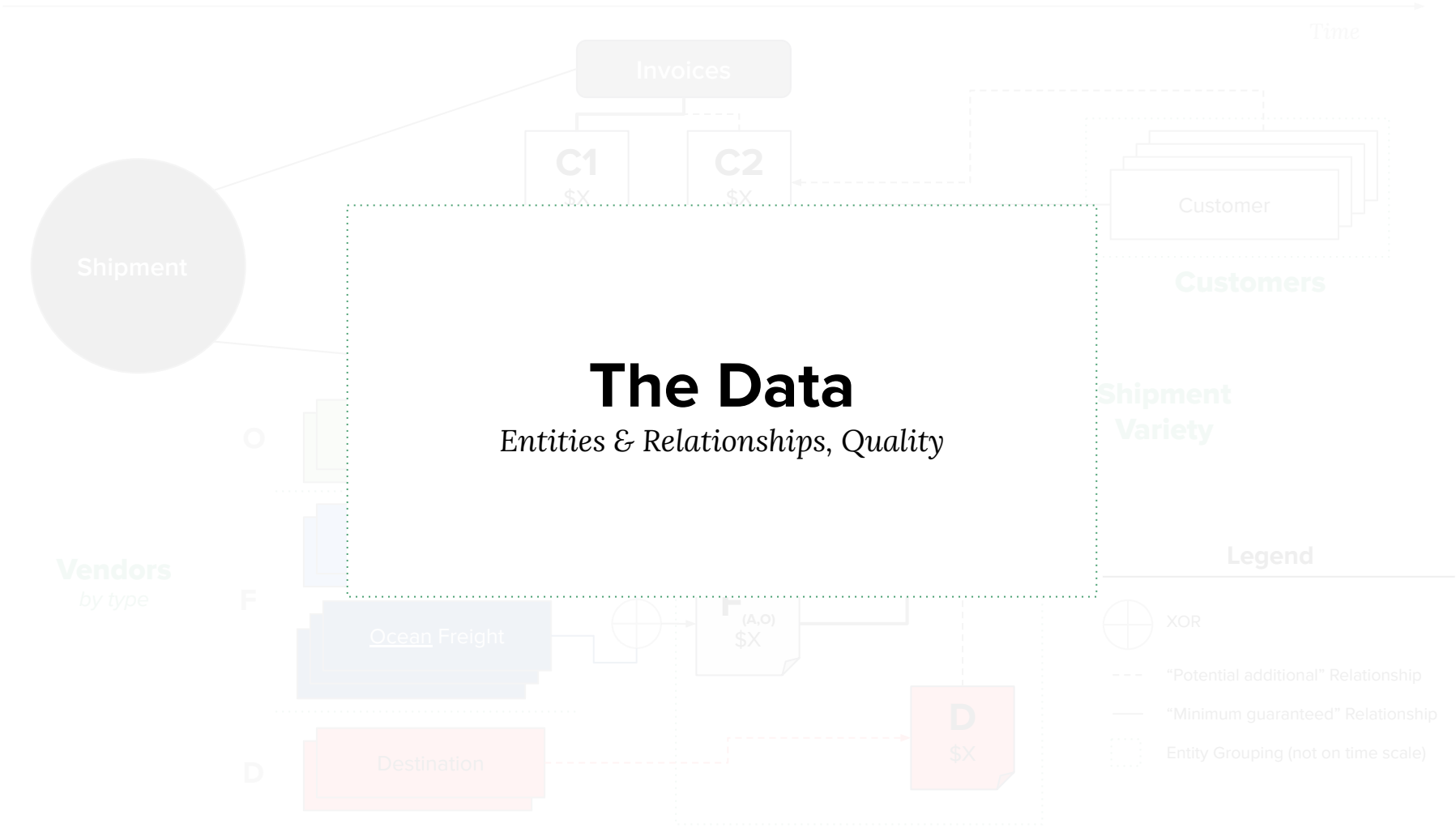
Ravi Dayabhai
April 8, 2024

Contents

Context

Context

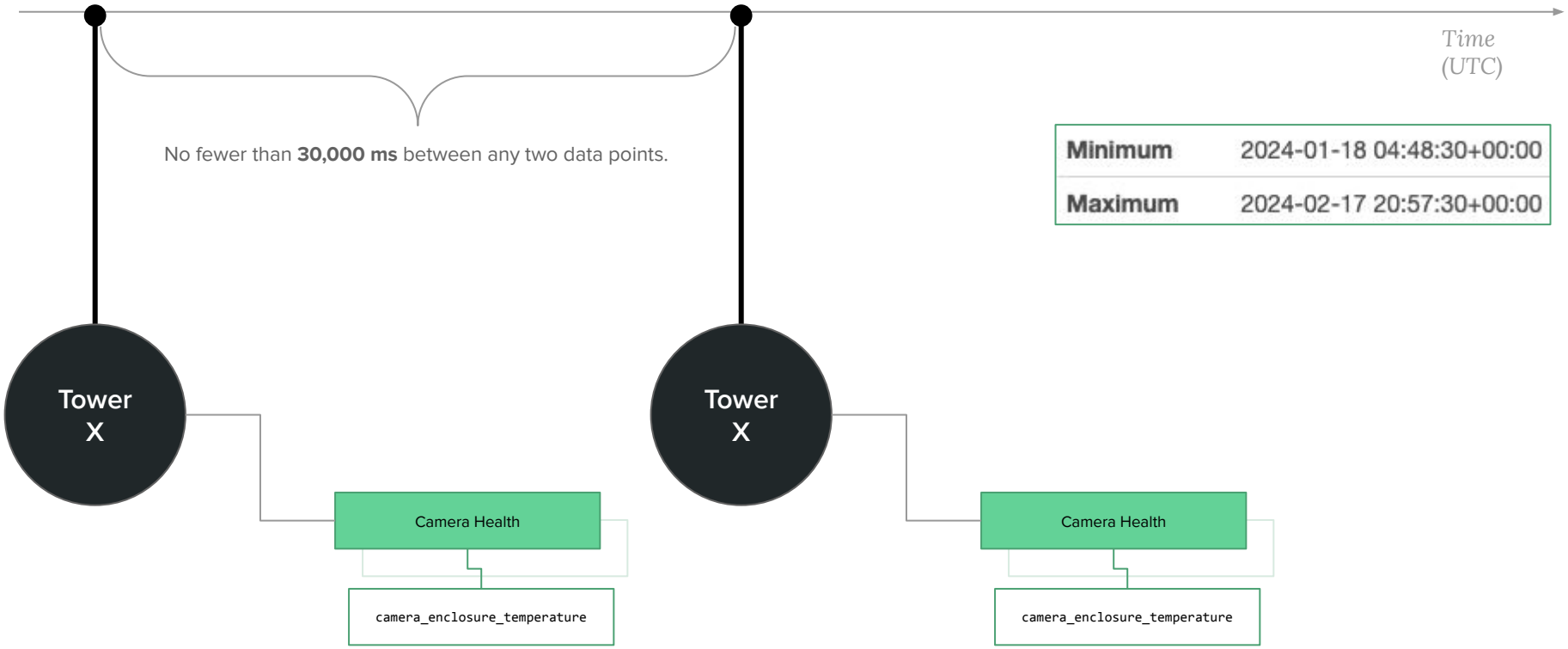
- **The Task:** **exploratory data analysis** + answer select **questions about tower performance** and reporting
- **The Data:** raw readings for **30+ metrics** from **4 towers** spanning **c. 1 month period**
  - Data is structurally sound, but the representations were non-uniform (e.g., CSV, JSONL)
  - Units were assumed from metric names or inferred via data or standards
  - No assumptions are made about missing data, though significant nullity is worth further investigation
- **The Artifacts:** various **data products** to enable subsequent analysis
  - Idempotent ETL logic (i.e., code to build data warehouse)
  - Static site of the profiled data (e.g., summary statistics, cuts by metric & by tower)
  - Chronicled investigations (i.e., notebooks with deeper dives)



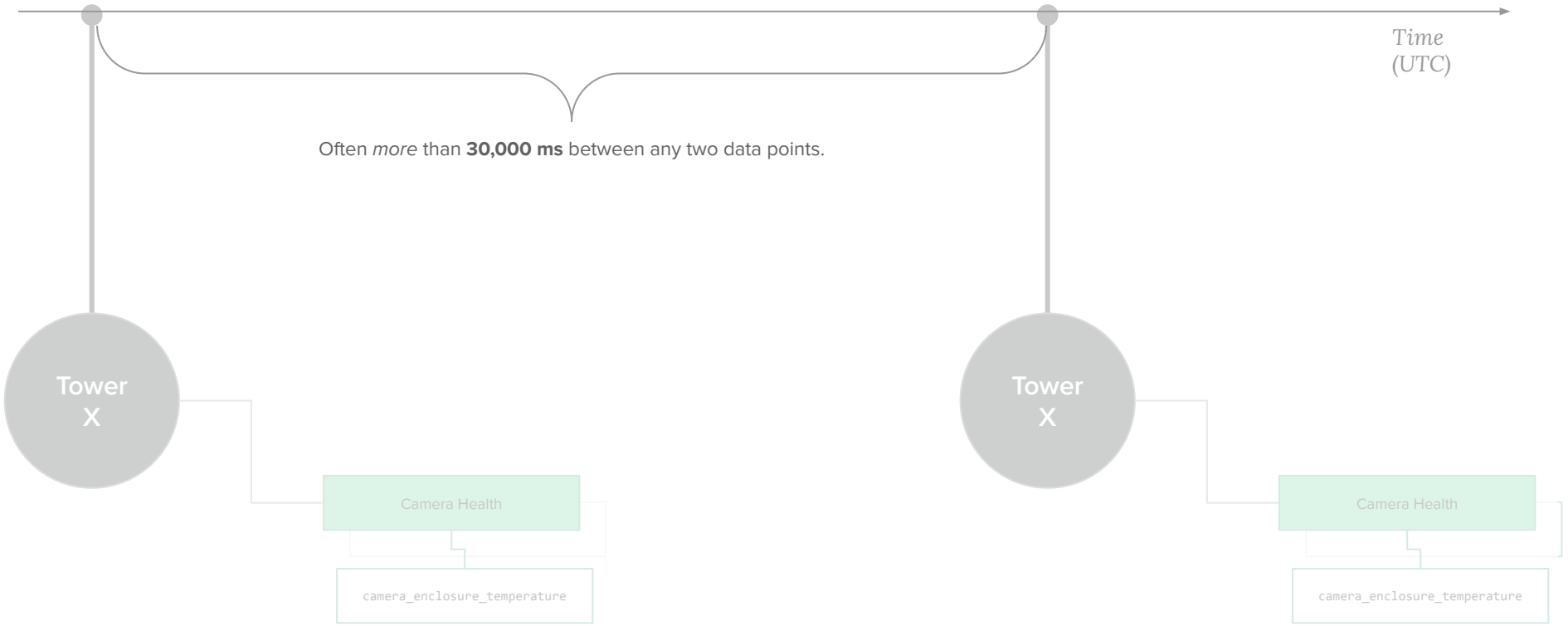
The Data

Entities & Relationships, Quality

Each **tower** reports **time-series data** on various collections of **metrics.**

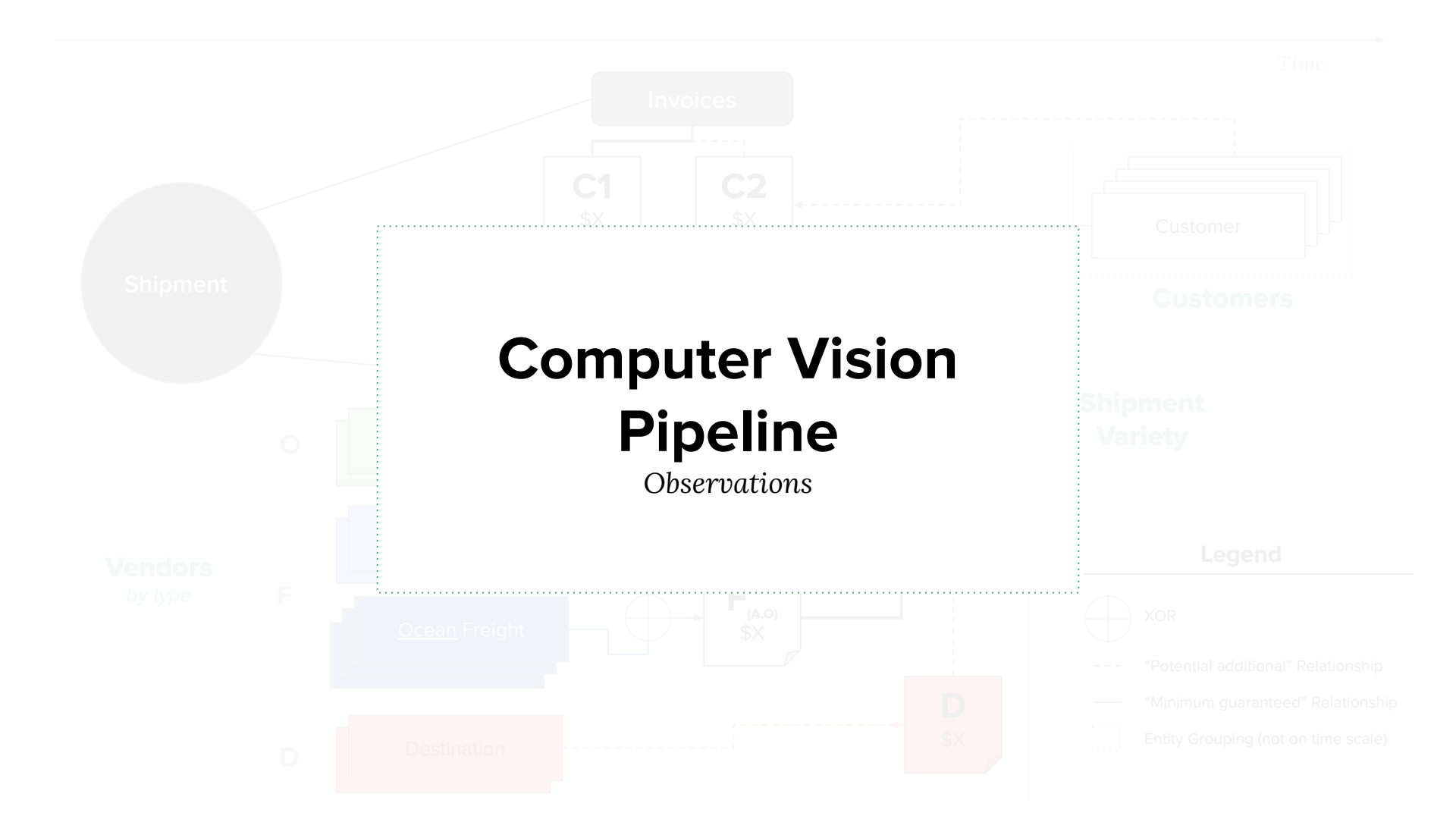


The **sampling frequency** over the period for any **tower - metric** pair is **30 seconds**.



The **data availability** over the period **show patterns**, though the **underlying causal explanation is not clear**.

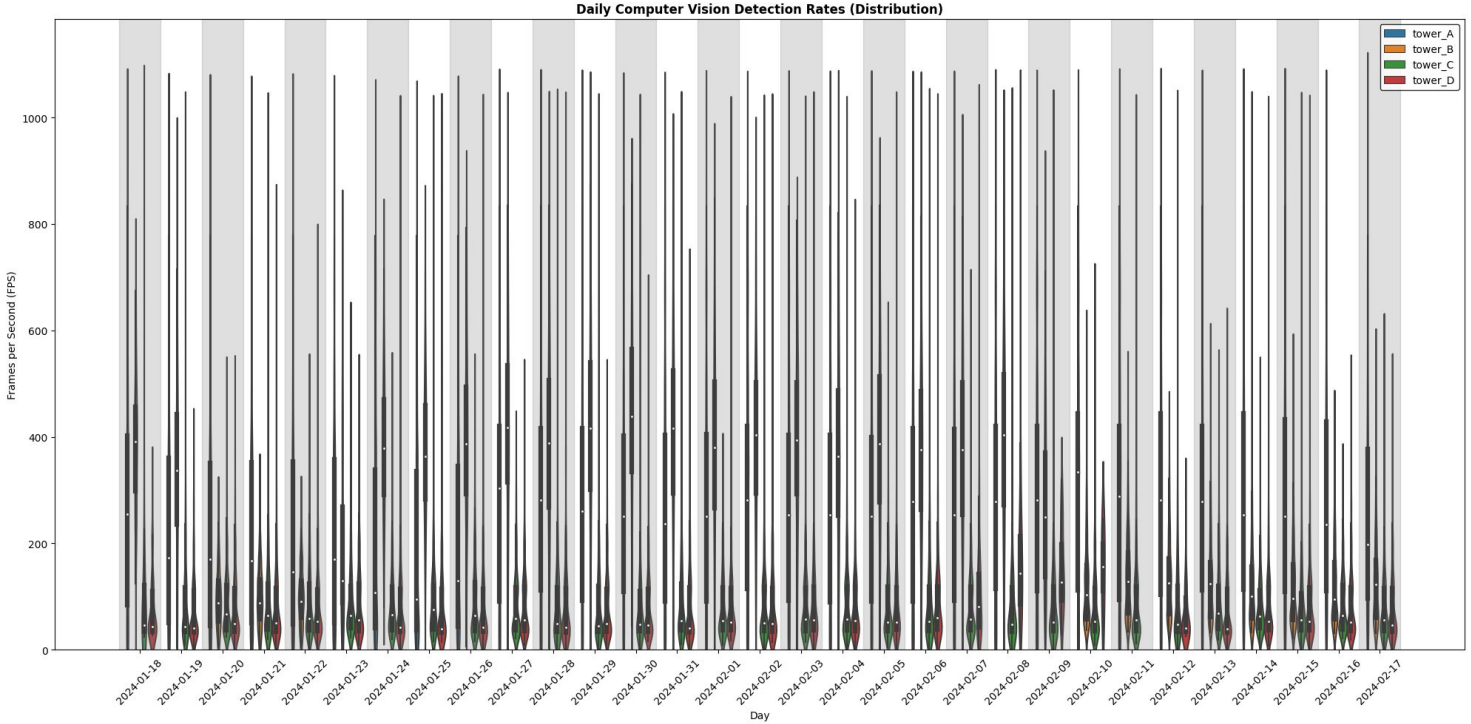
Analysis



Computer Vision Pipeline

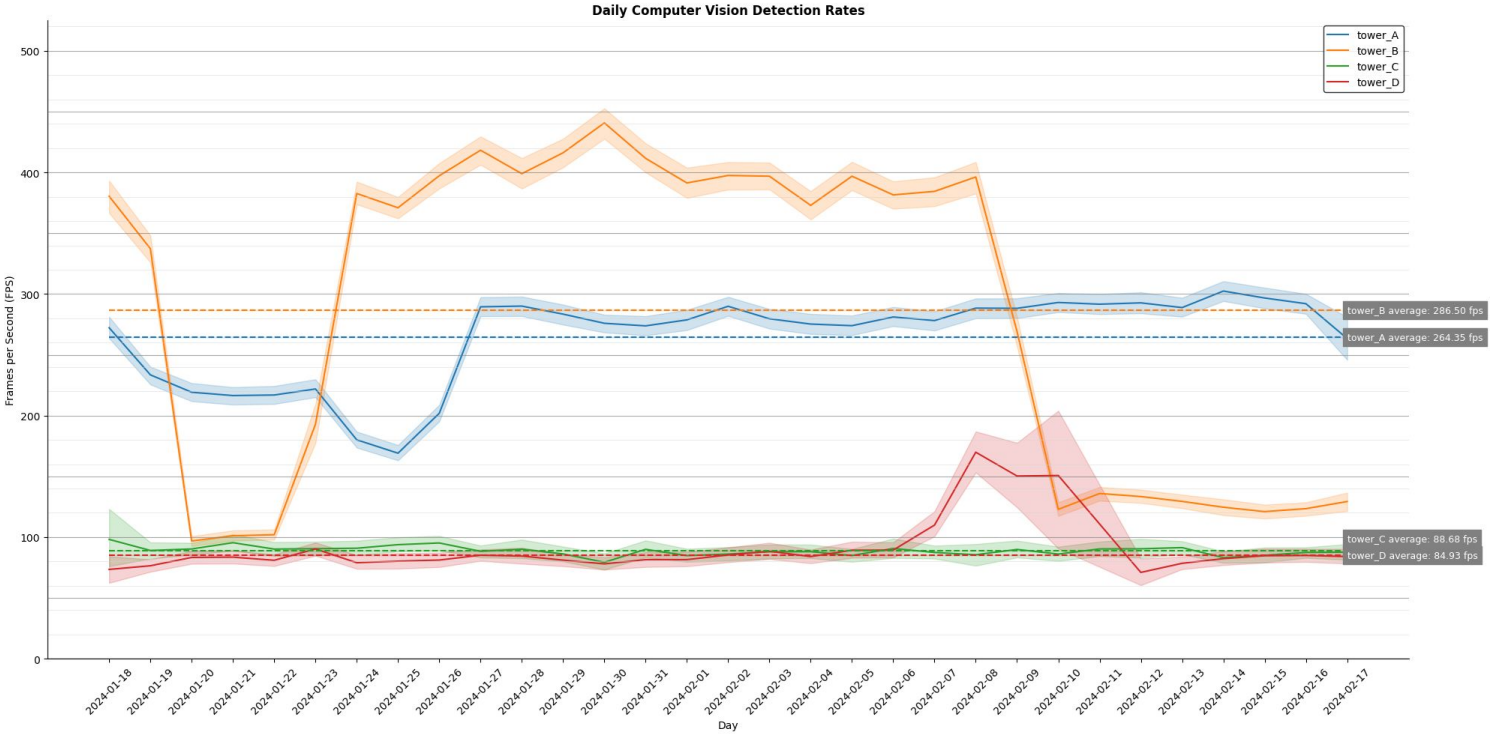
Observations

The distributions of CV detection rates are bounded but vary in shape...



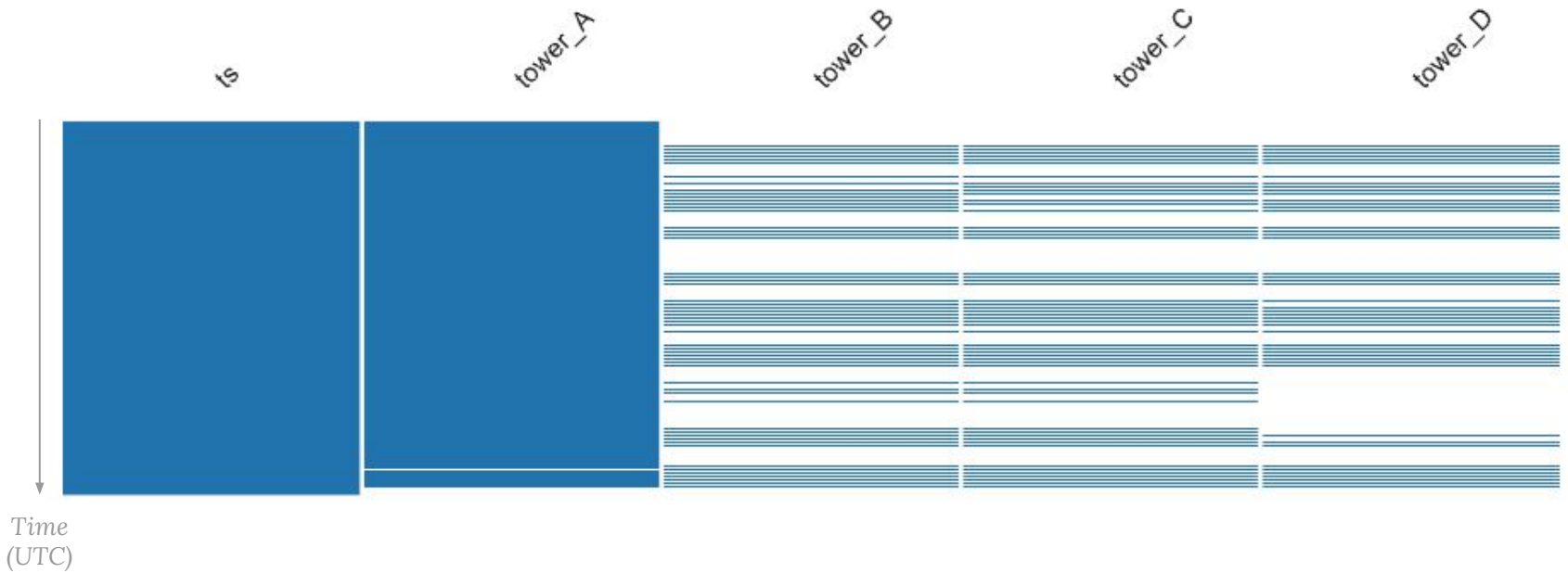
➔ Towers **A & B show greater average, and more varied, FPS** than their counterparts overall and per day.

…and there are notable fluctuations over particular sub-periods…



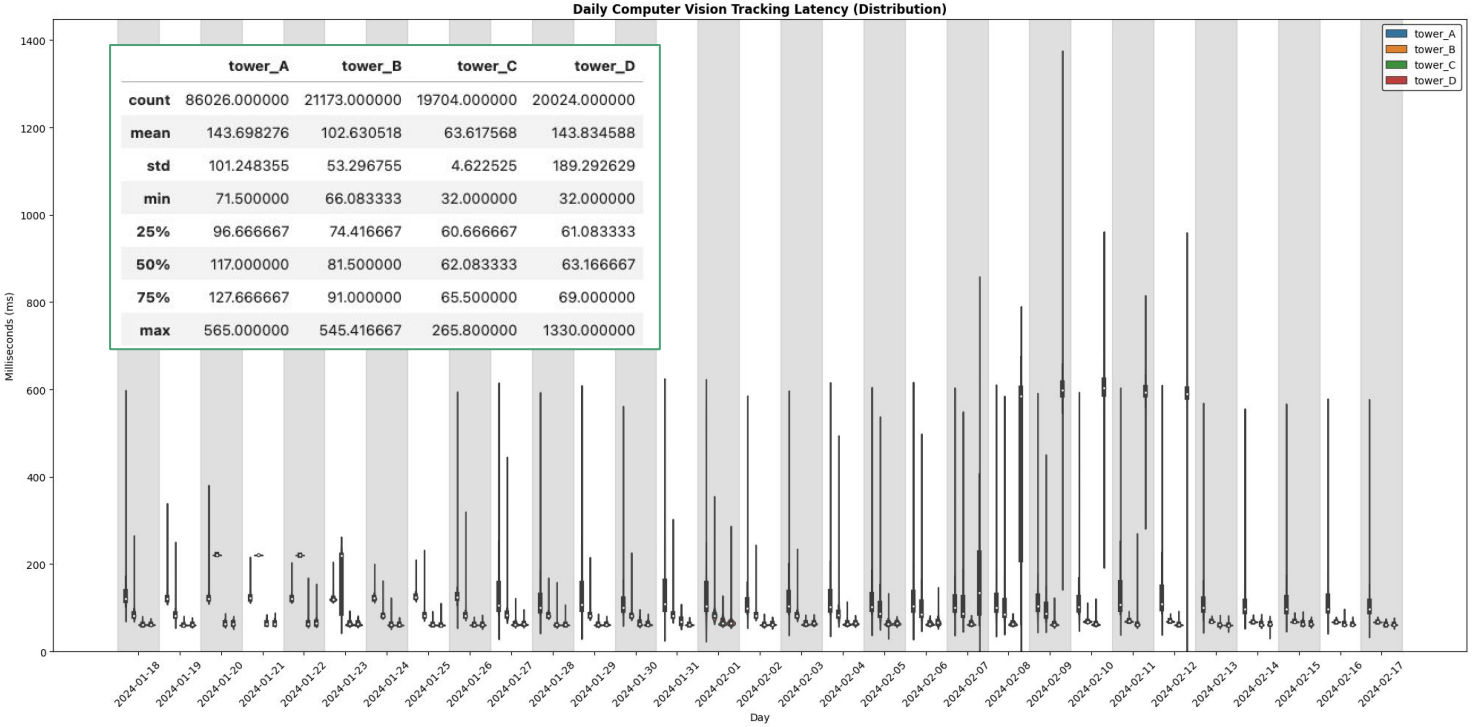
- ➔ **Tower B**'s sees a **c. 4x *increase*** in the middle of the period. A **similar phenomenon affects Tower D** later in the period (**to a lesser extent**).
- ➔ The **latter pair of towers have c. 3x *lower* FPS** over the period on average.

...and non-reporting for Towers B, C, D are awfully similar.



➔ This is first (of many) examples of **patterned nullity** in the dataset; this and **the next two metrics share similar nullity**.

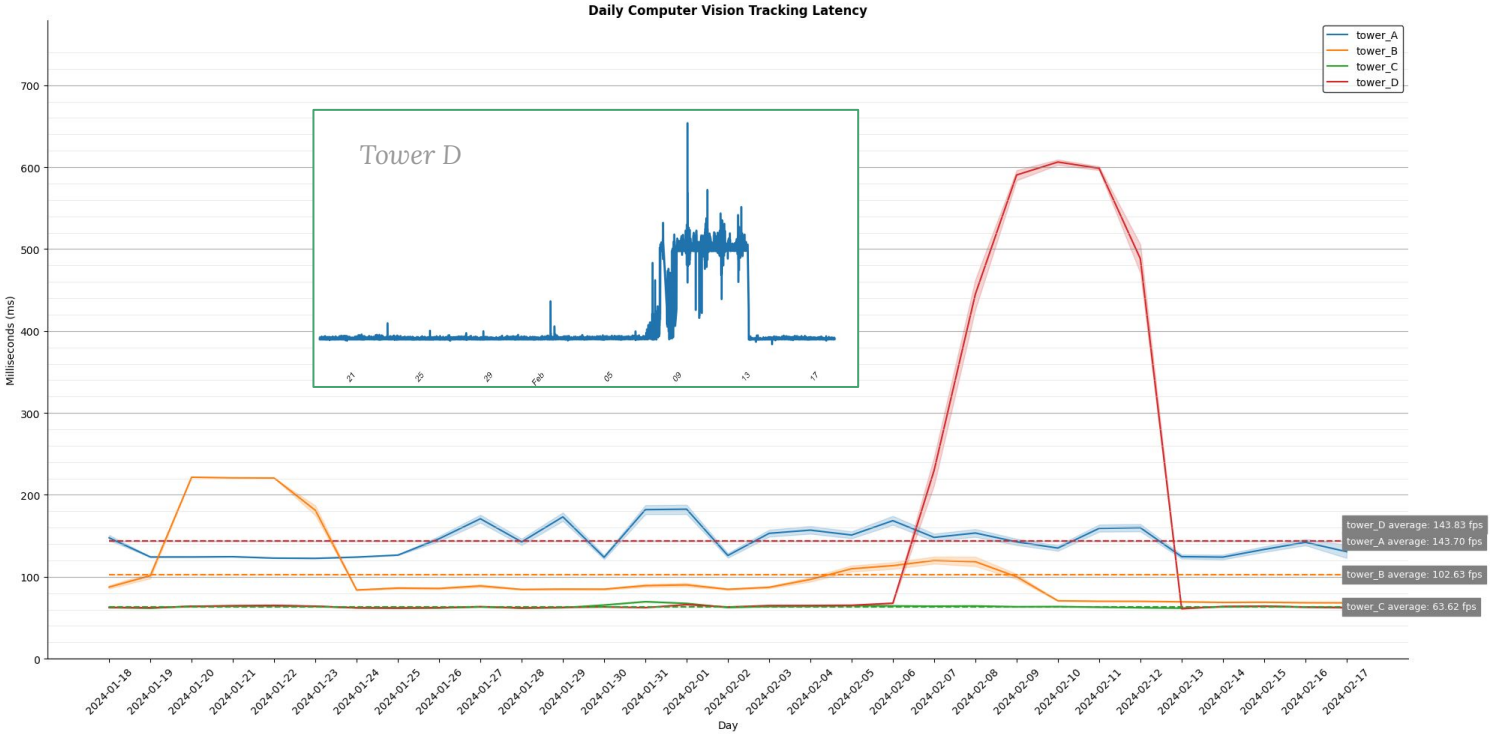
CV tracking latency is also non-uniform across towers and days...



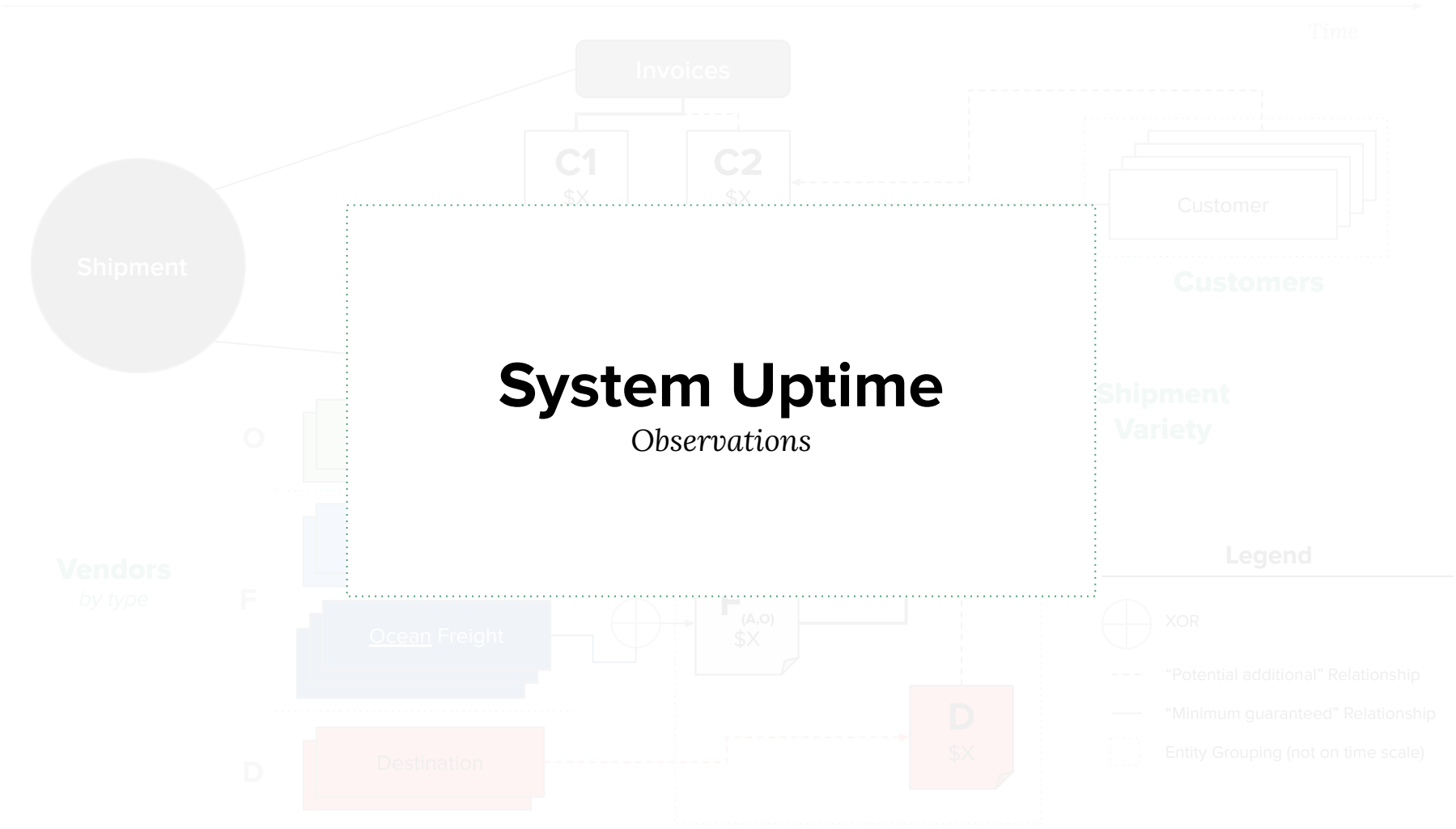
	tower_A	tower_B	tower_C	tower_D
count	86026.000000	21173.000000	19704.000000	20024.000000
mean	143.698276	102.630518	63.617568	143.834588
std	101.248355	53.296755	4.622525	189.292629
min	71.500000	66.083333	32.000000	32.000000
25%	96.666667	74.416667	60.666667	61.083333
50%	117.000000	81.500000	62.083333	63.166667
75%	127.666667	91.000000	65.500000	69.000000
max	565.000000	545.416667	265.800000	1330.000000

➔ Putting aside the anomalous readings, in general, **the first pairs of towers are more similar to each other** than to the second pair in terms of CV tracking latency.

...with the average latency for Tower D spiking in early February.



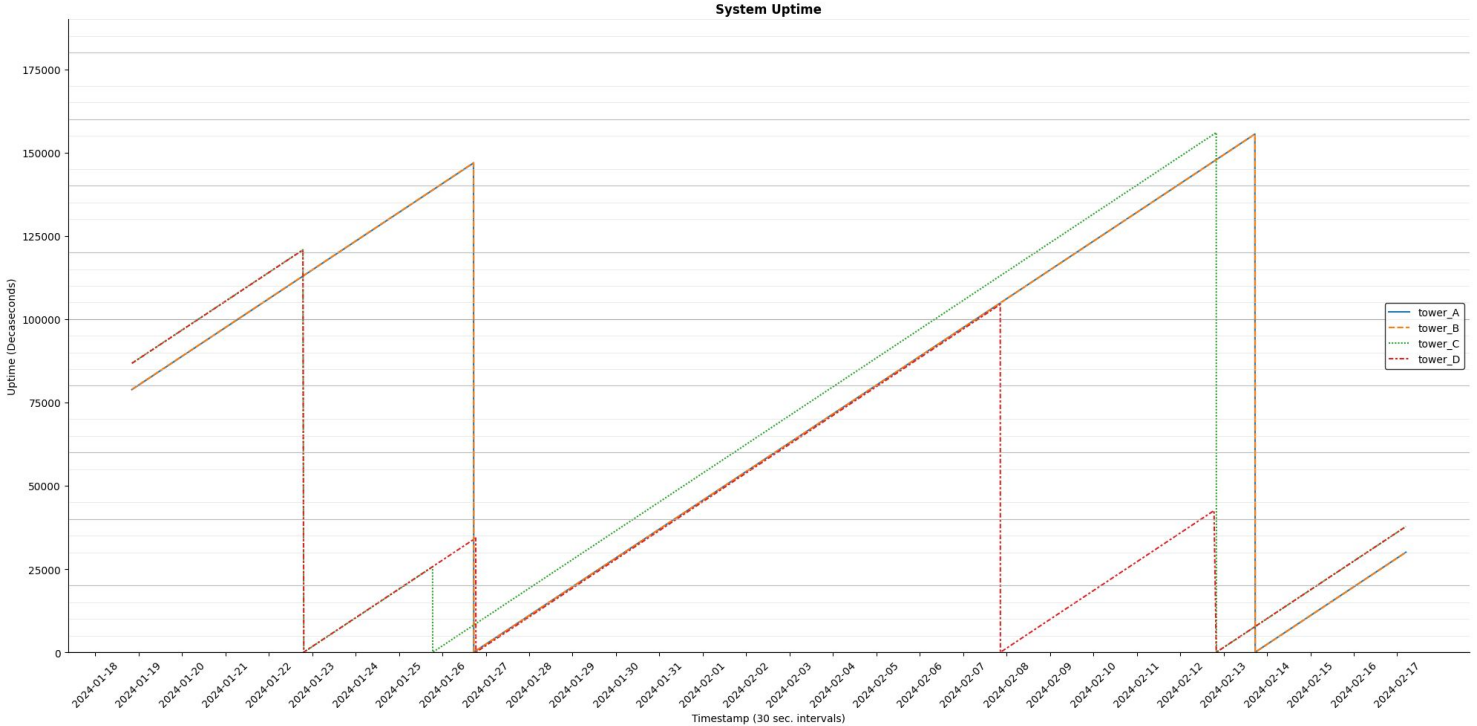
➔ Towers **C & D tend to have lower latencies**, with the notable **exception being tower D during early February**.



System Uptime

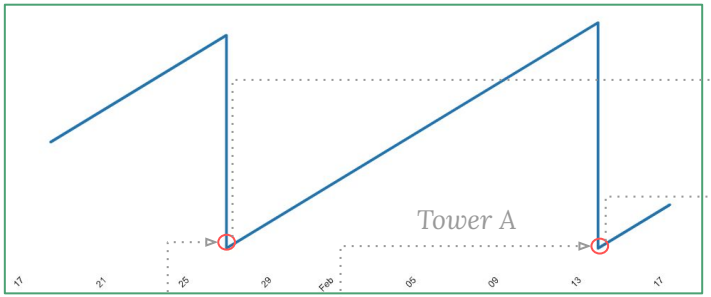
Observations

System uptime readings are *mostly* consistent...



➔ The **maximum** uptime is **c. 18 days**, but **there are aberrations** according to a few towers.
➔ The *inferred* units are *decaseconds* (1 decasecond = 10 seconds).

...but it's not clear whether monotonicity of the metric is reliable.



	tower_A	tower_B	tower_C	tower_D
1	0 days 00:05:00	18 days 00:08:00	0 days 03:54:00	3 days 23:18:00
2	18 days 00:02:00	NaT	2 days 23:42:00	12 days 02:04:00
3	0 days 00:05:00	NaT	18 days 01:22:00	4 days 22:12:00
4	NaT	NaT	2 days 22:26:00	NaT

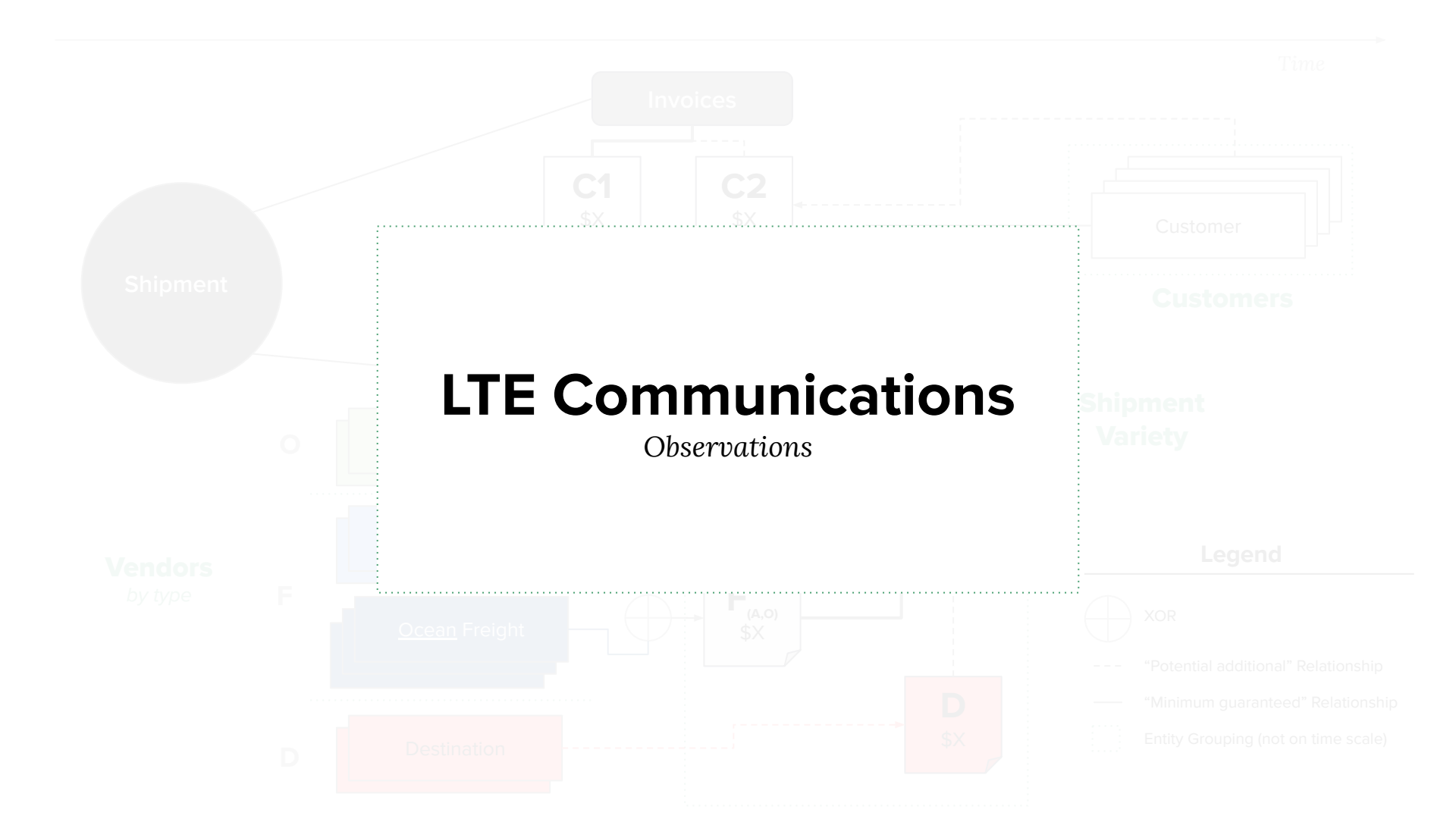
ts	tower_A	prev_value	next_value	is_peak
2024-01-26 17:12:00+00:00	146868.200	73287.700	97917.666667	True
2024-01-26 17:17:00+00:00	58762.040	58758.440	27.600000	True
2024-02-13 17:19:00+00:00	155555.600	27.600	77786.450000	True
2024-02-13 17:24:00+00:00	38907.875	38903.375	28.300000	True

Peak-to-Peak Cycle Means over Reporting Period

```
tower_A     6 days 00:04:00
tower_B    18 days 00:08:00
tower_C     6 days 00:51:00
tower_D     6 days 23:51:20
dtype: timedelta64[ns]
```

⚠️

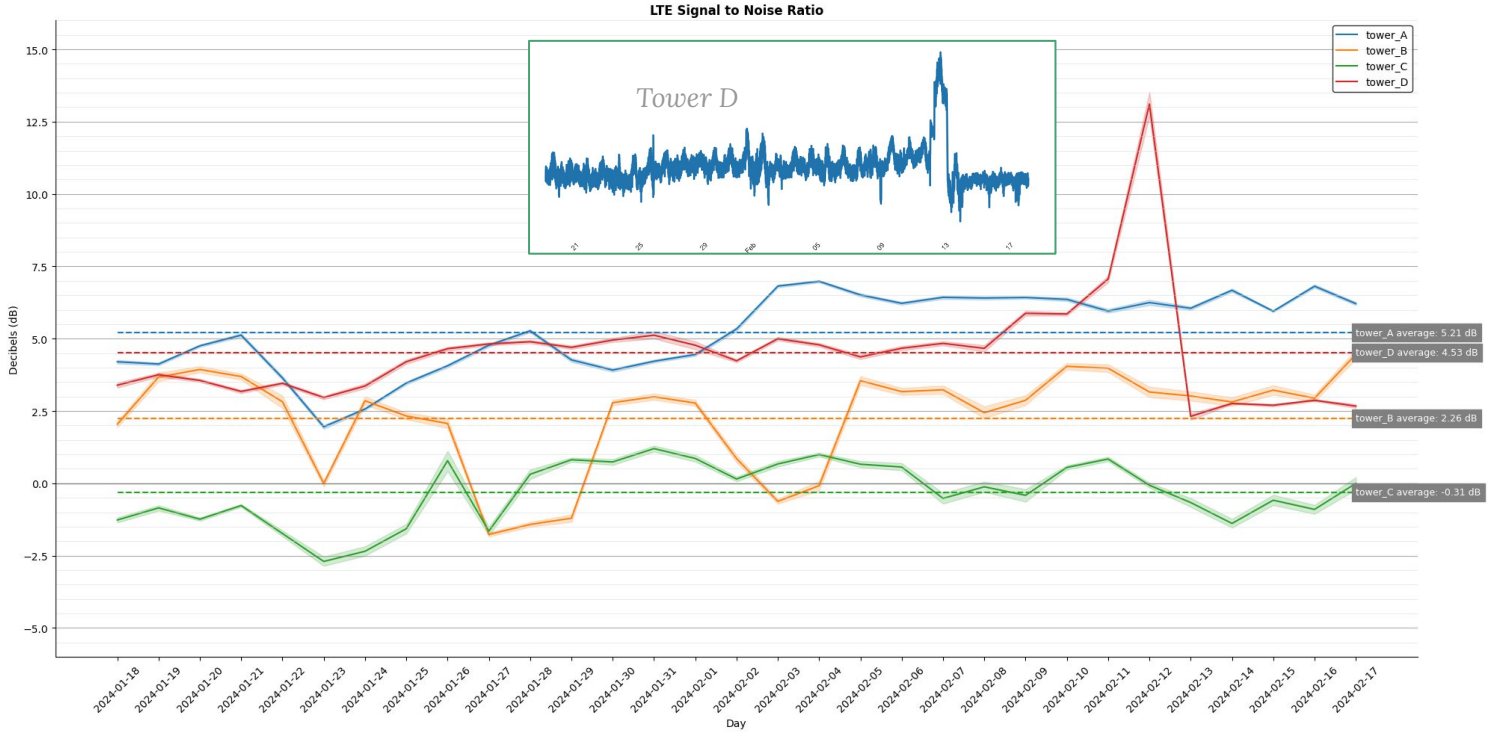
➔ Peak-to-peak calculations **reveal "micro-cycles"**, which have an **outsized effect on simple averages**.



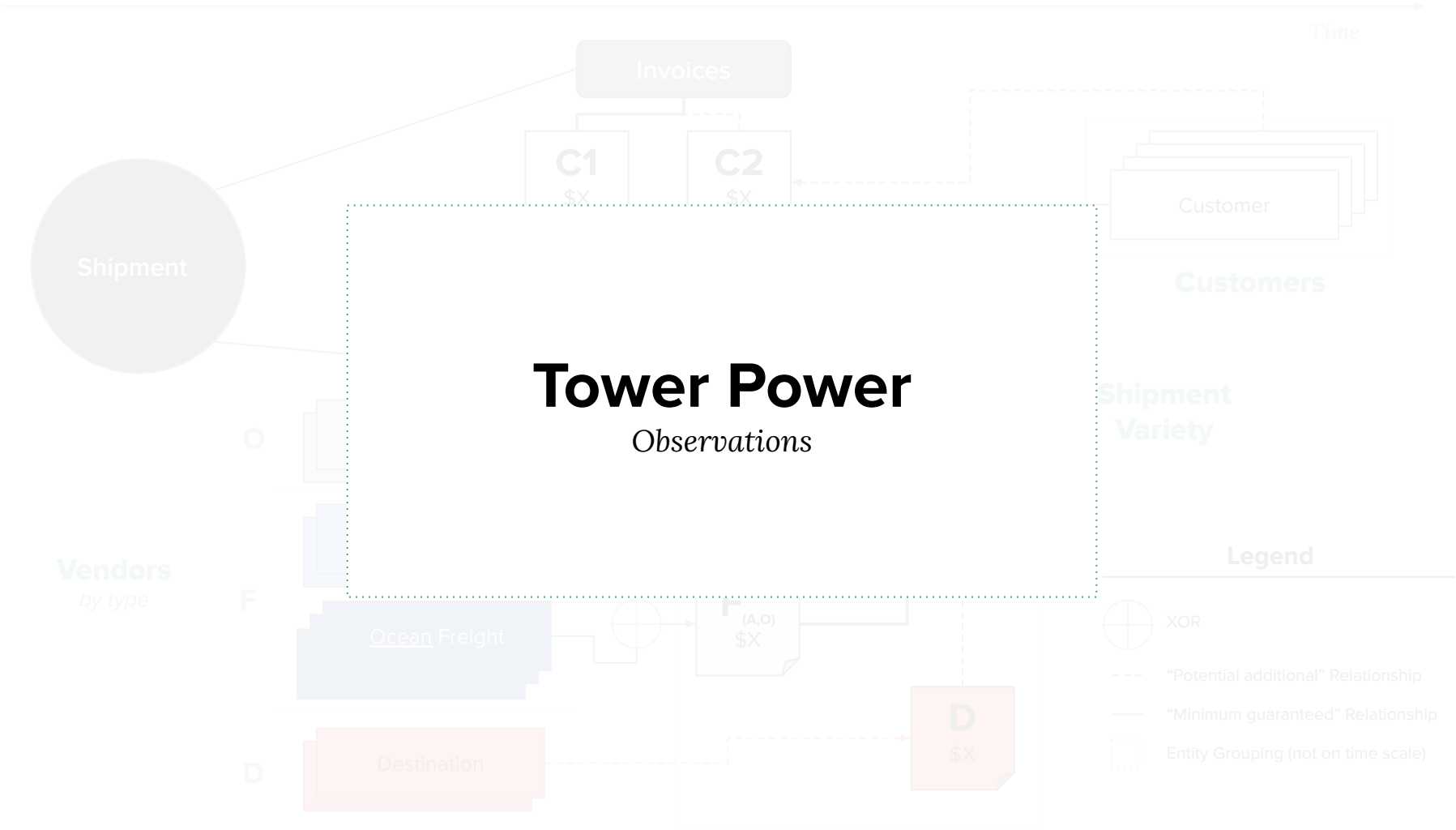
LTE Communications

Observations

LTE SNR looks relatively stable (absent baseline).



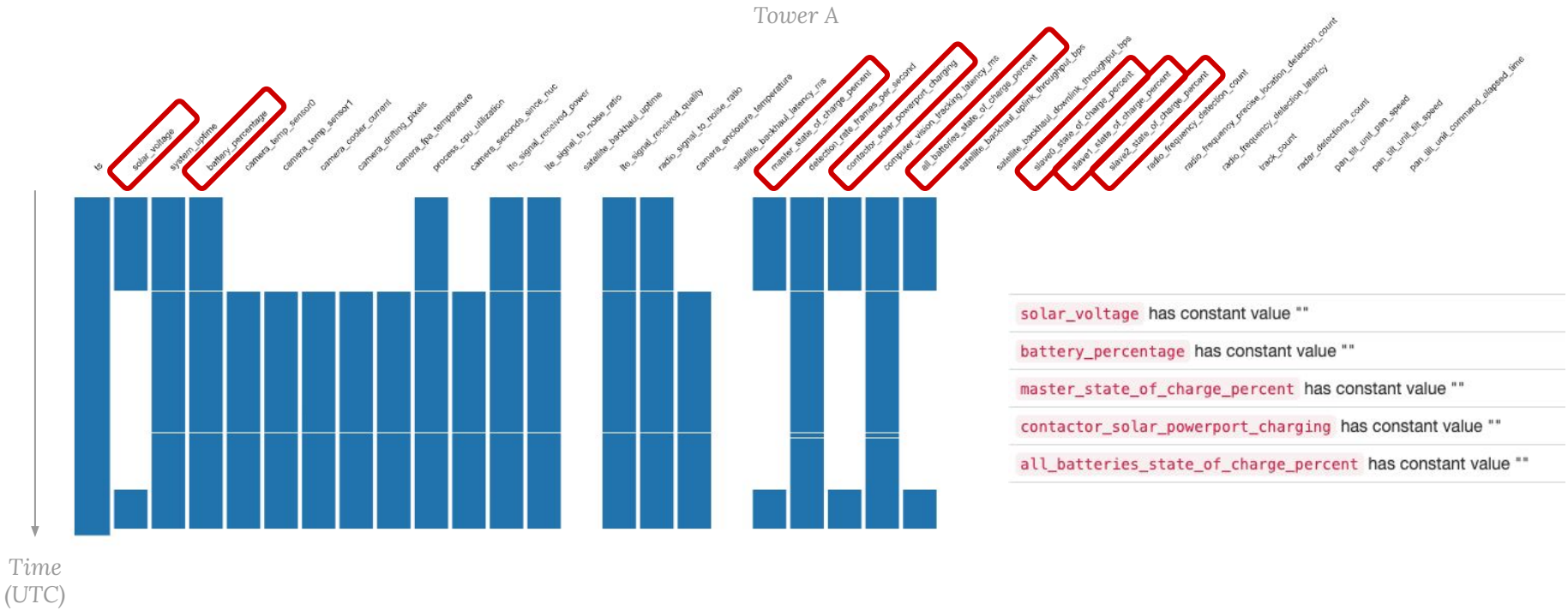
➔ **Tower C** is the **only tower with a *negative* ratio**, on average.
➔ **Tower D** is the **only tower with a *bimodal* distribution**, from the **peak on February 12th**.



Tower Power

Observations

Nullity and indiscernible power consumption patterns...



`solar_voltage` has constant value ""

`battery_percentage` has constant value ""

`master_state_of_charge_percent` has constant value ""

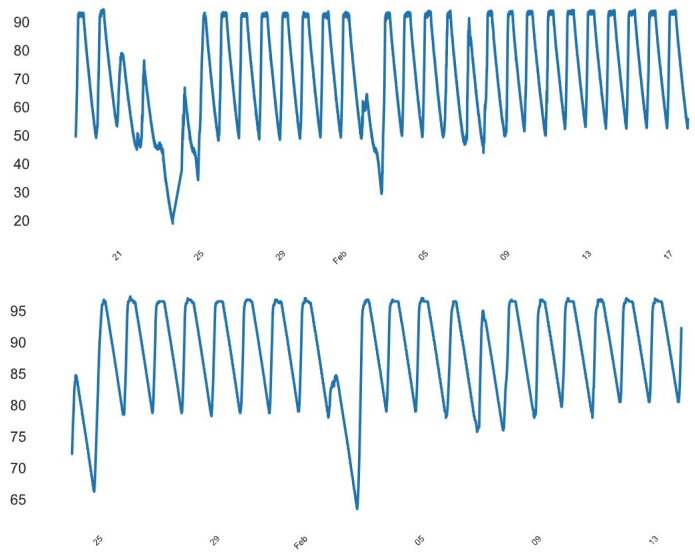
`contactor_solar_powerport_charging` has constant value ""

`all_batteries_state_of_charge_percent` has constant value ""

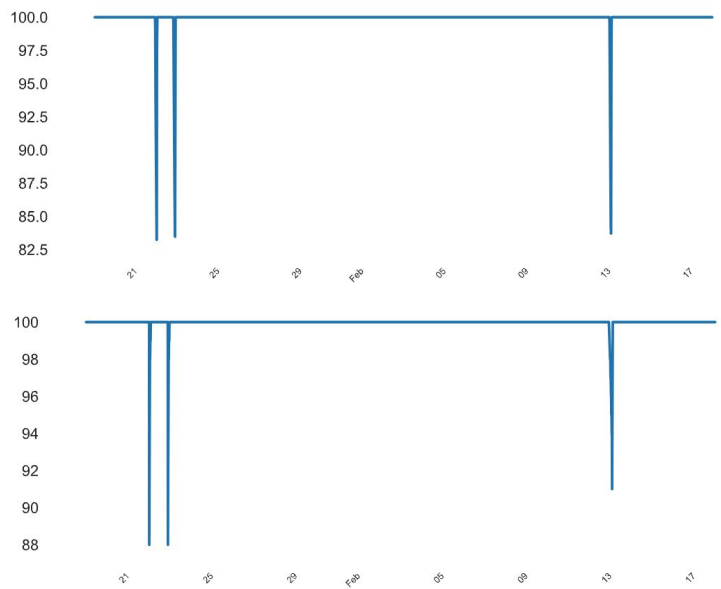
➔ Unless power doesn't vary, **Tower A** is effectively **providing no information about power consumption**, as all related fields are constant (100%) or missing entirely.

...obscure consistent monitoring of tower power.

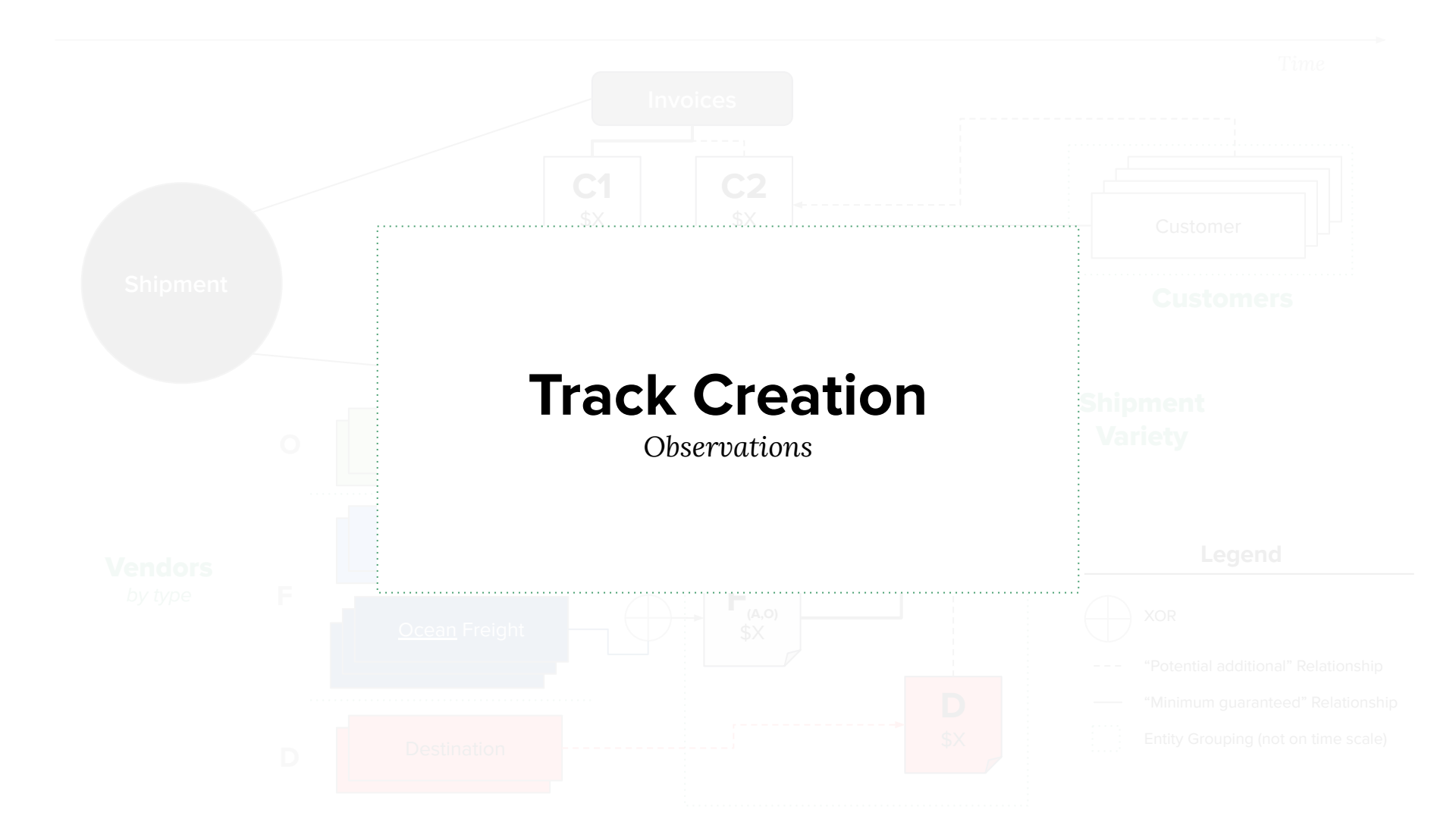
Tower A: Battery Percentage & All Batteries State of Charge Percentage



Tower D: Battery Percentage & All Batteries State of Charge Percentage



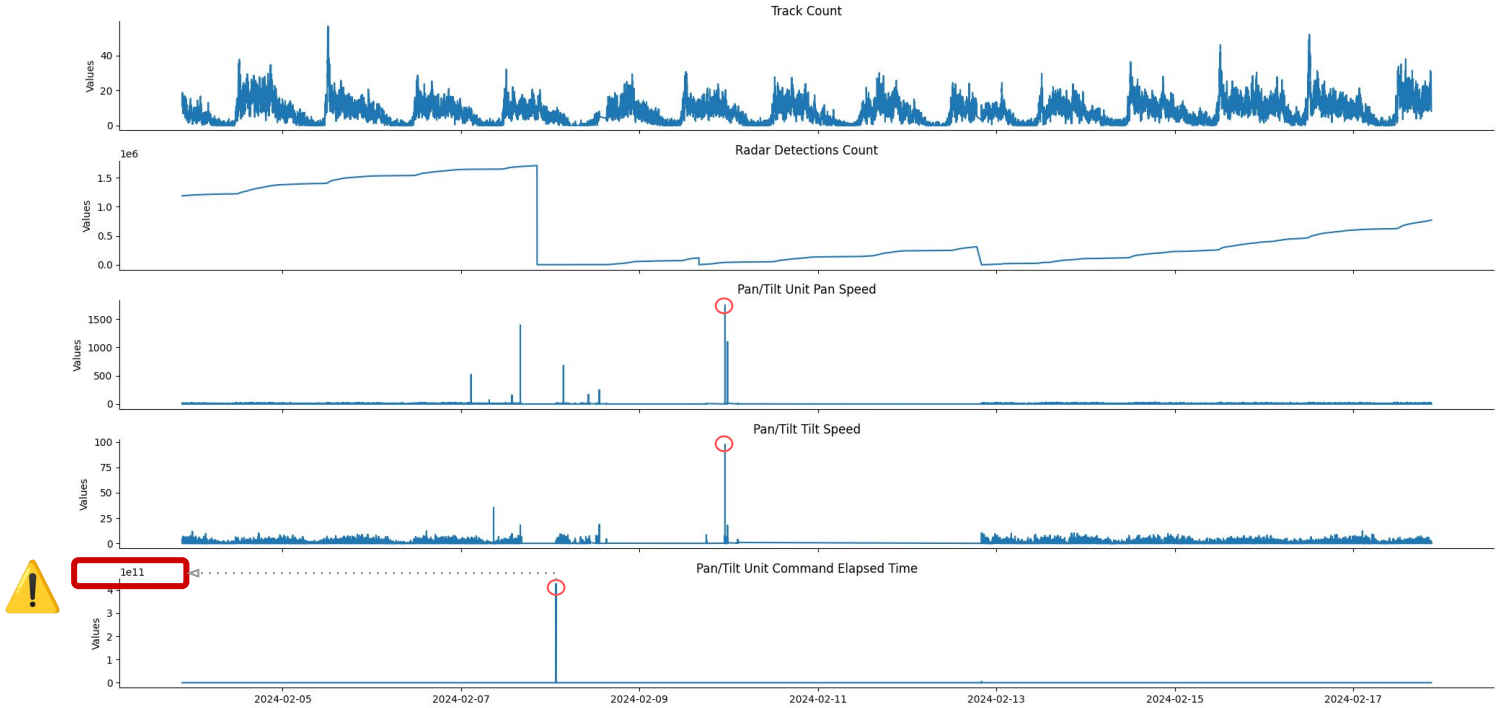
➔ It **does not seem there is a clear and consistent picture** across the four towers if consumption patterns are expected to be similar.



Track Creation

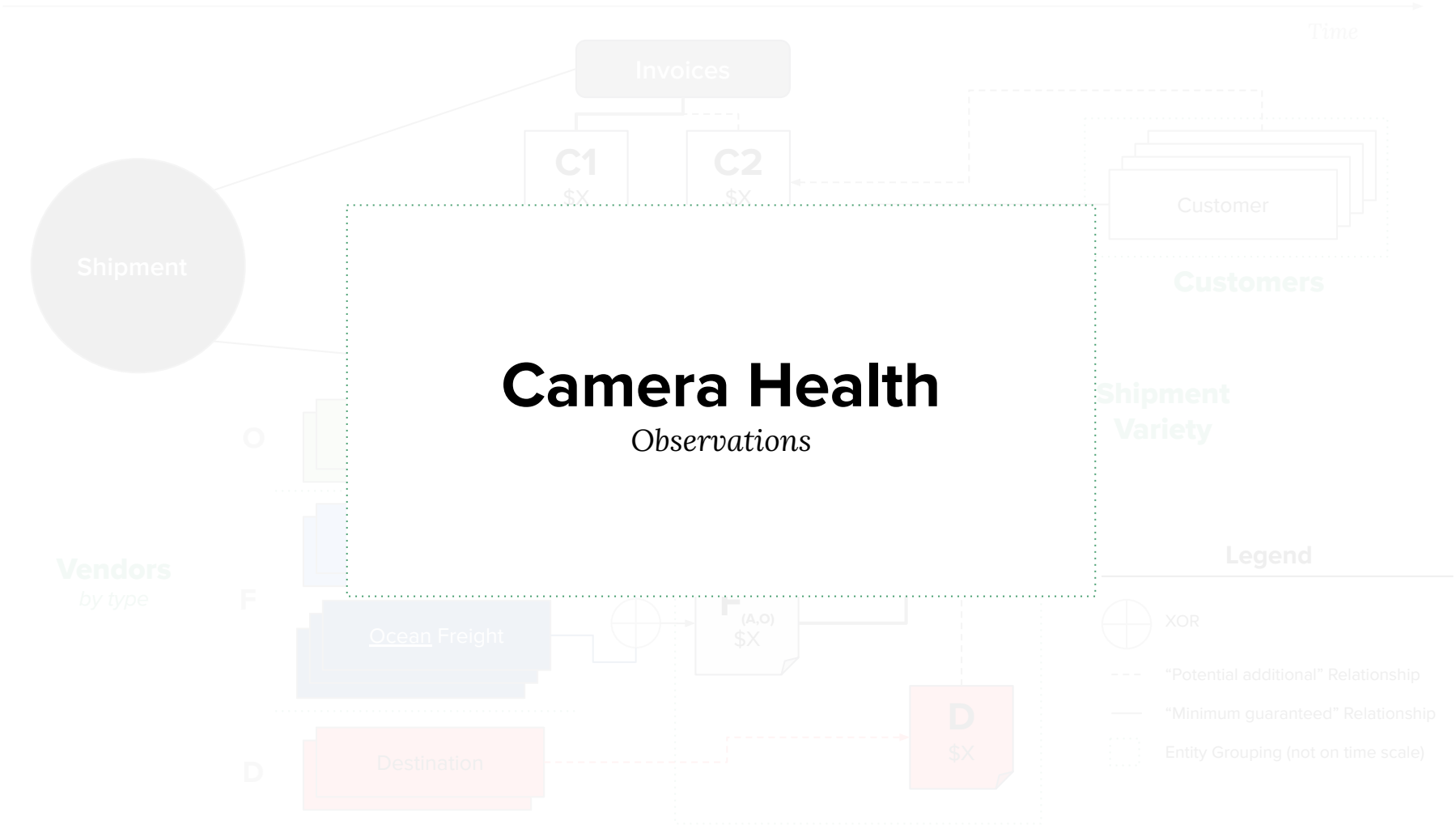
Observations

Track creation metrics contain notable outliers.



⚠️

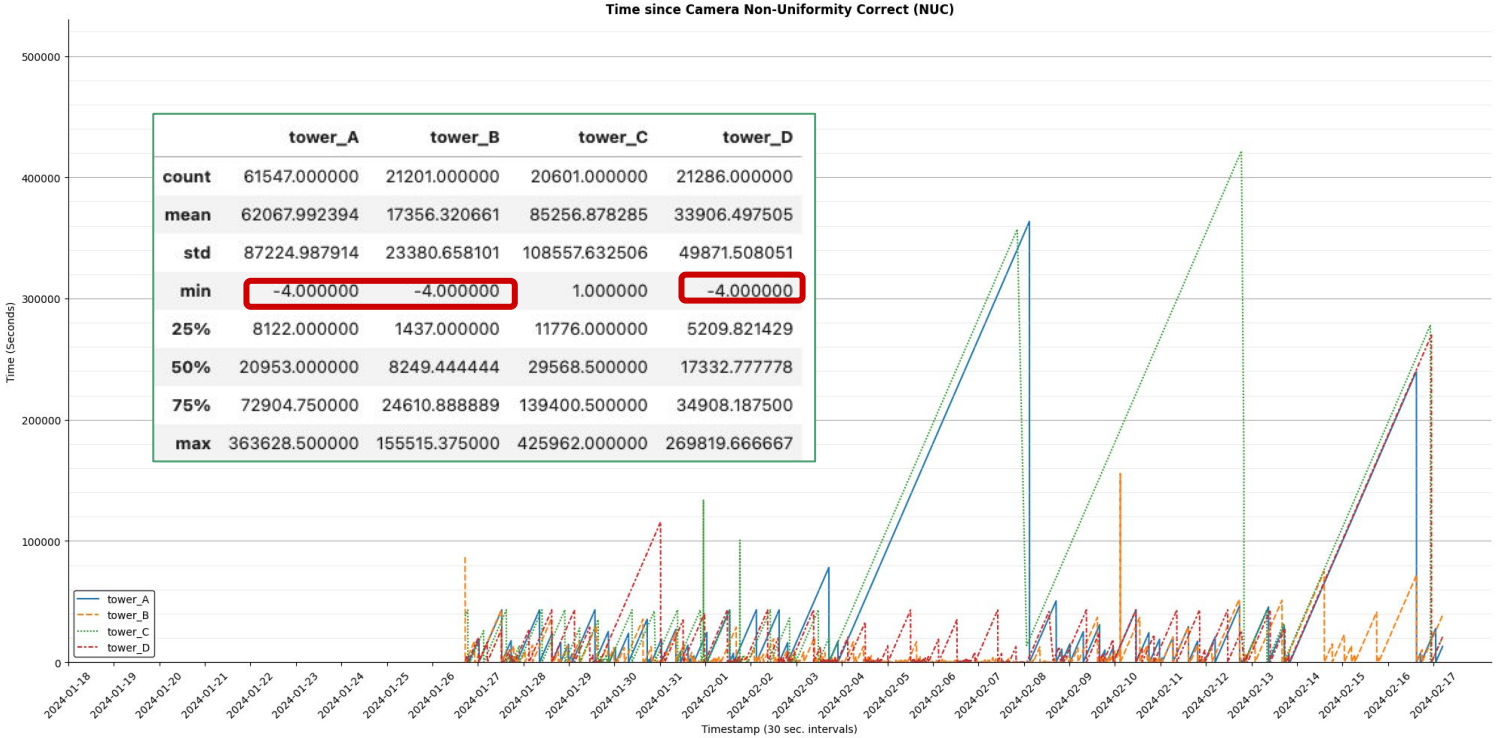
➔ The **pan/tilt unit elapsed command time** contains a record that, if measured in *seconds*, amounts to **c. 13k-14k *years*.**



Camera Health

Observations

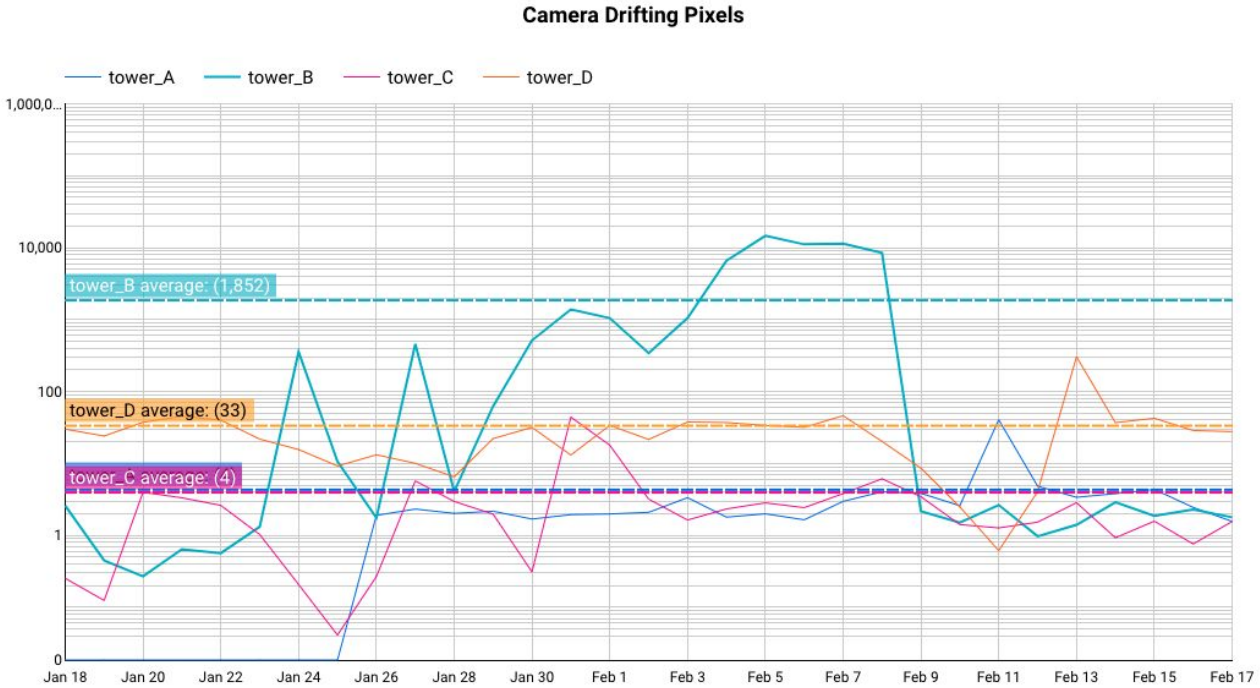
Time between Non-Uniformity Corrections (NUCs) ranges widely…



	tower_A	tower_B	tower_C	tower_D
count	61547.000000	21201.000000	20601.000000	21286.000000
mean	62067.992394	17356.320661	85256.878285	33906.497505
std	87224.987914	23380.658101	108557.632506	49871.508051
min	-4.000000	-4.000000	1.000000	-4.000000
25%	8122.000000	1437.000000	11776.000000	5209.821429
50%	20953.000000	8249.444444	29568.500000	17332.777778
75%	72904.750000	24610.888889	139400.500000	34908.187500
max	363628.500000	155515.375000	425962.000000	269819.666667

➔ It is **unclear** what the **expected behavior** or underlying cause for this variation is.

...and the variation across towers doesn't stop there.



- ➔ **Tower B** experiences an **exponential increase** in camera drift from Jan. 26th to Feb. 9th.
- ➔ Roughly, coincides with **Computer Vision detection rate uptick** mentioned earlier.

Q&A