**Software stack:**

* Initially started working on a google collab ($50 monthly) for gpu access
* Once RAM became an issue turned to Saturn cloud computing, which became expensive (but hopefully worth it)

**Competition Basics:**

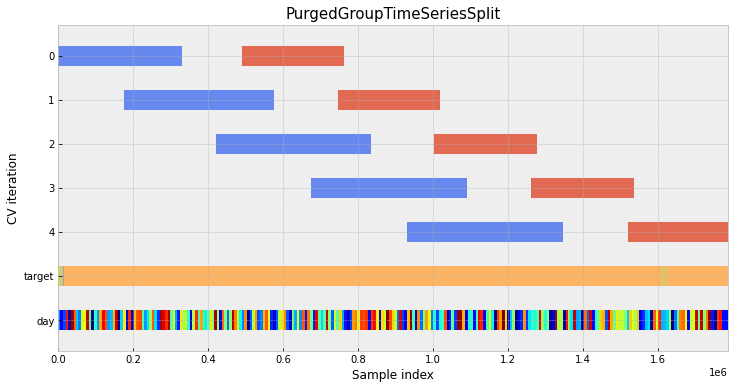
* Crypto forecasting competition for 14 different crypto currencies
* Train data: 1 Jan 2018 - 9 Sept 2021
  + Caveat, data given through training period at two different dumps, before final submission data was available to 24 Jan 2022
* Forecasting period is from Feb 1st 2022 to May 3rd 2022
* Scoring metric is weighted correlation with target value
  + Target value is log(t+1) / log(t+16) adjusted with a multiplier for overall market trend (of the 14 assets) and weighted value of the asset given
  + Asset weights range from 1.03 to 6.4
  + Understood this metric to be used over mse in order to reduce the influence of outliers of returns, but enhance the impact of identifying outliers amongst the market trend

**Data Basics:**

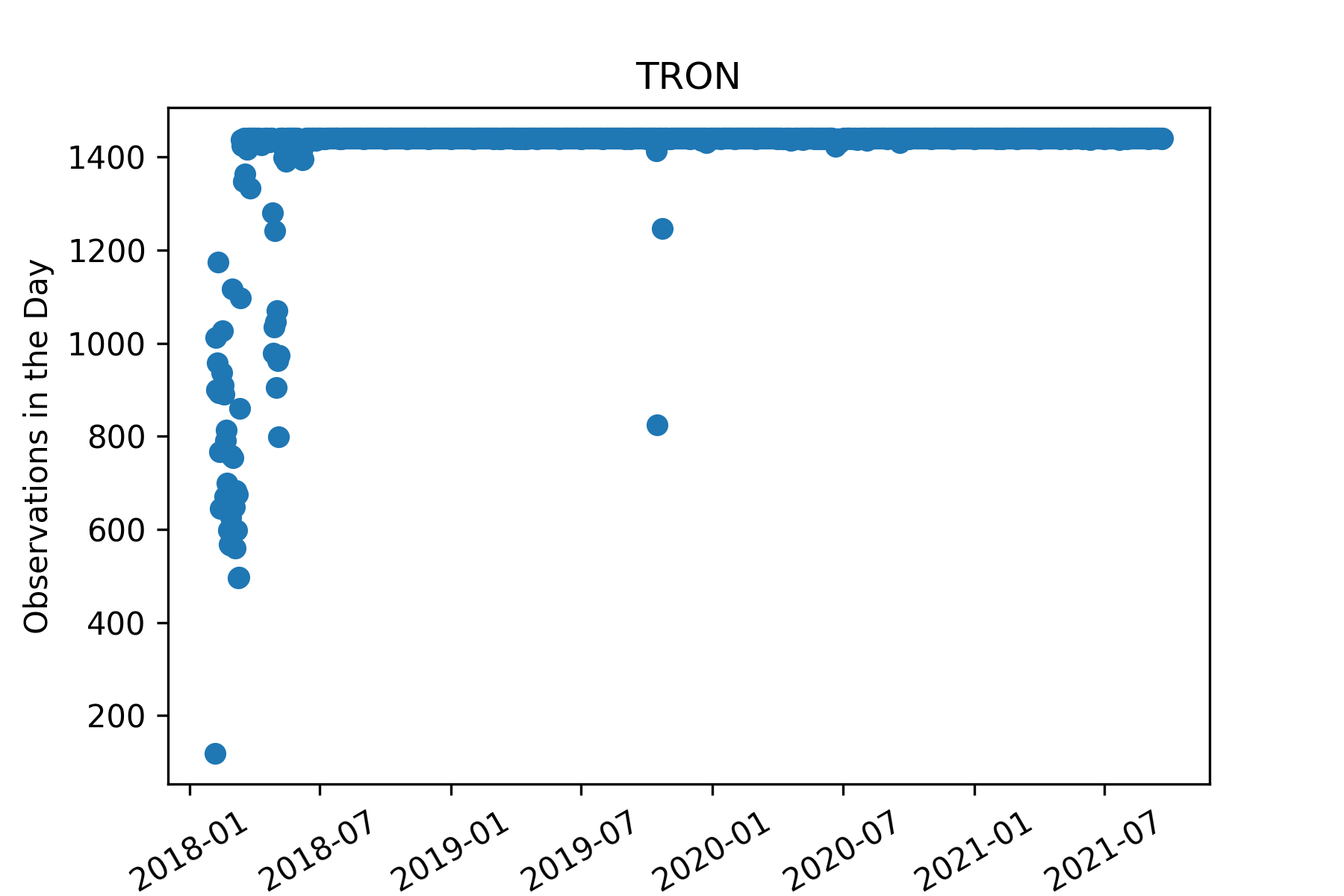
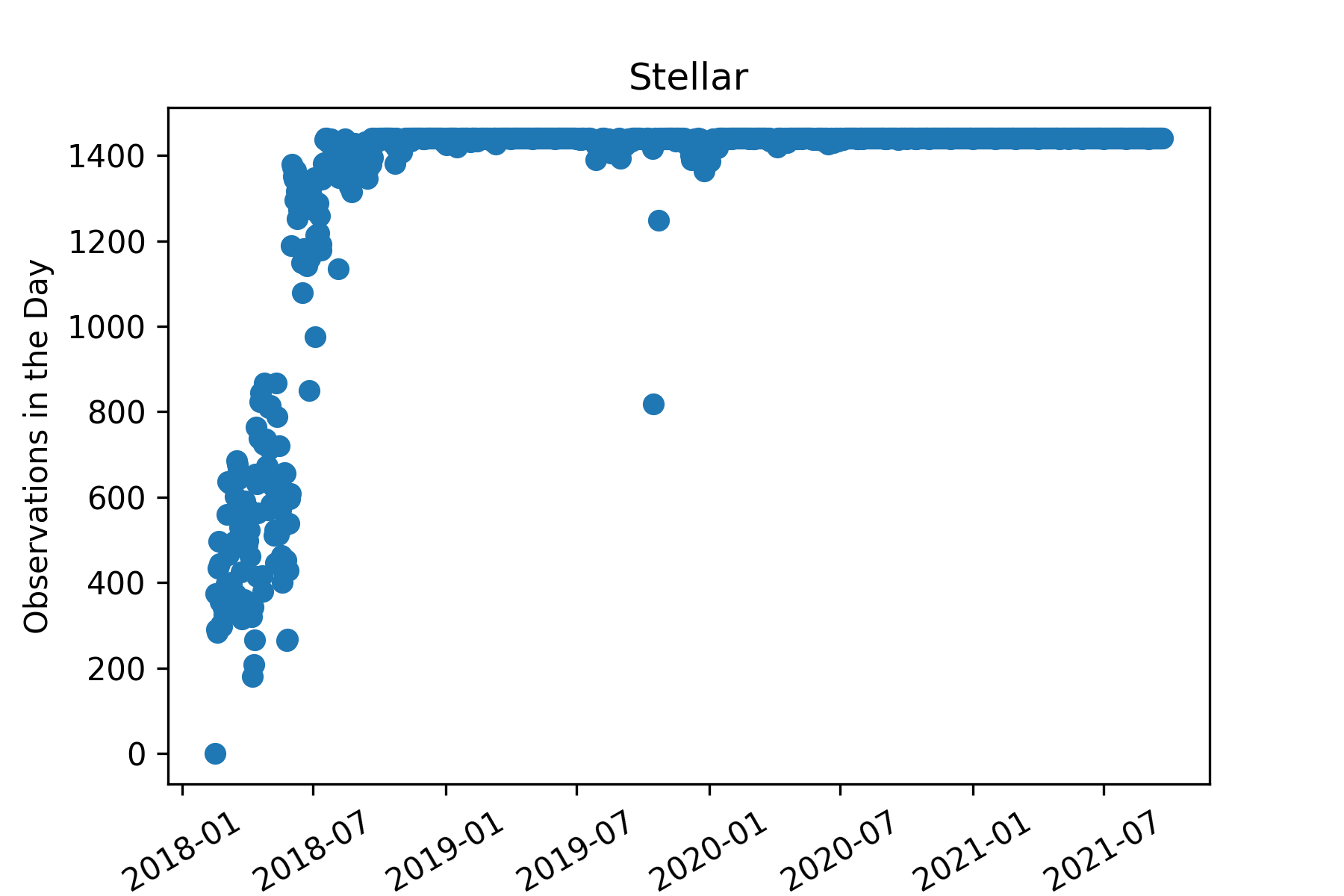
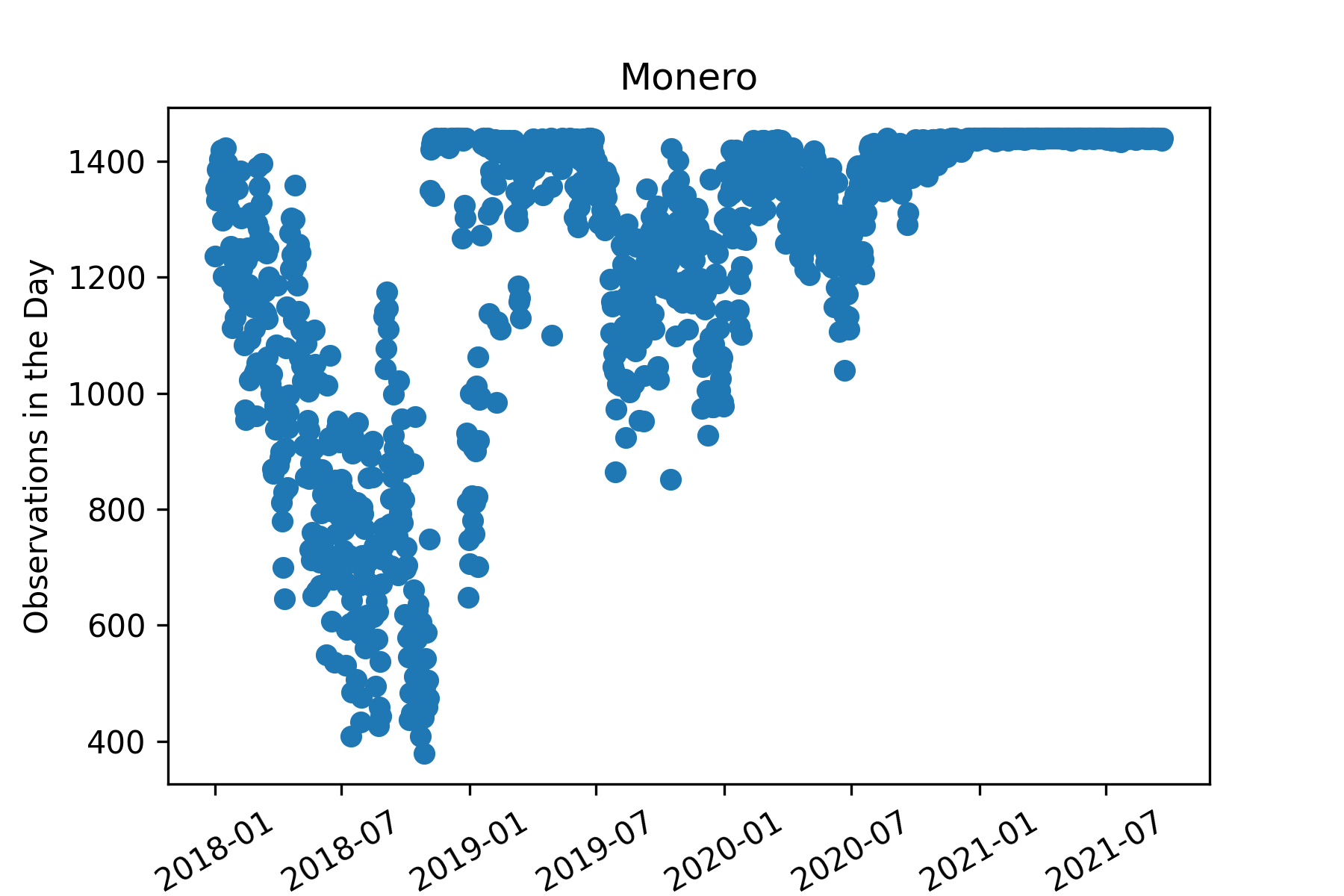
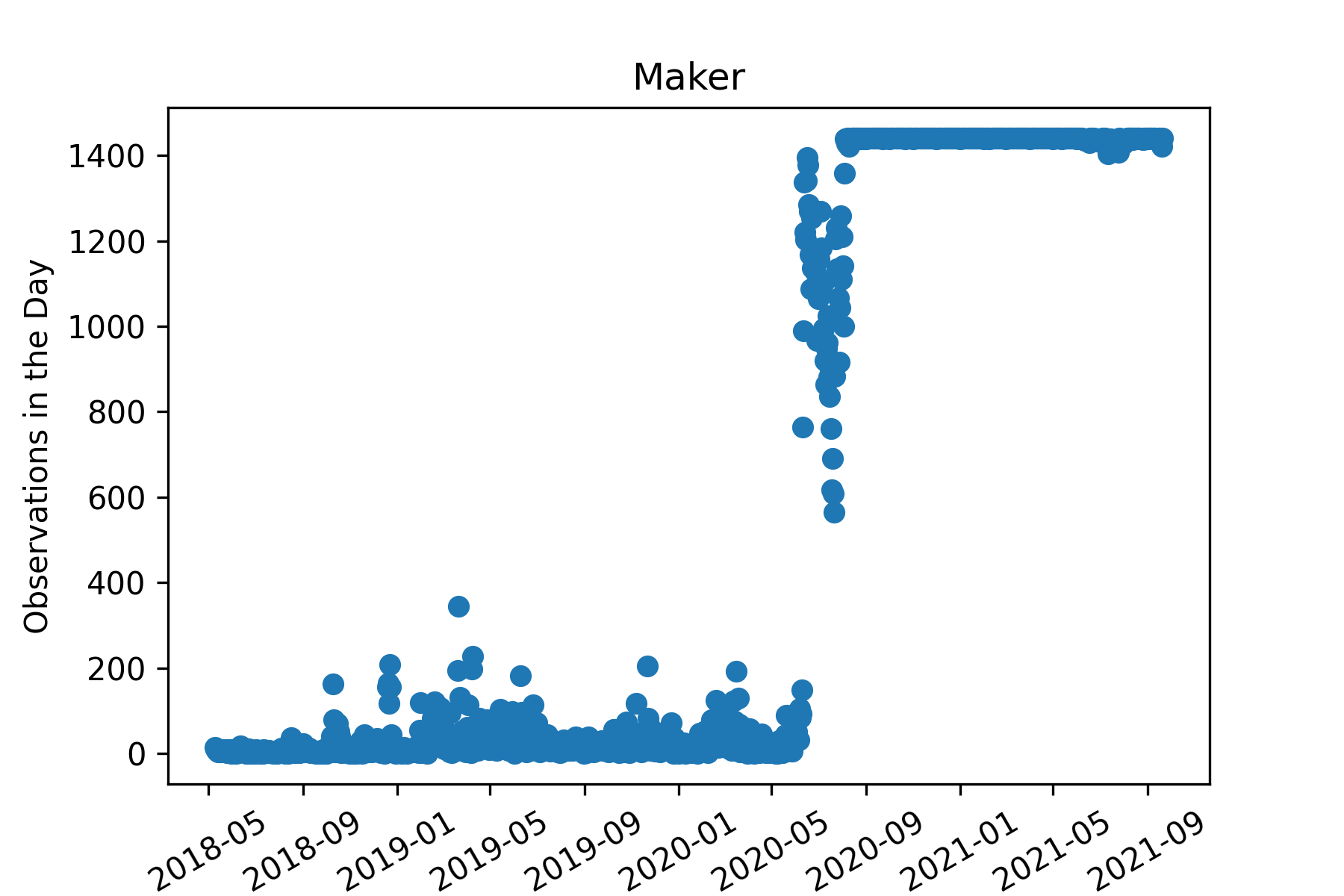
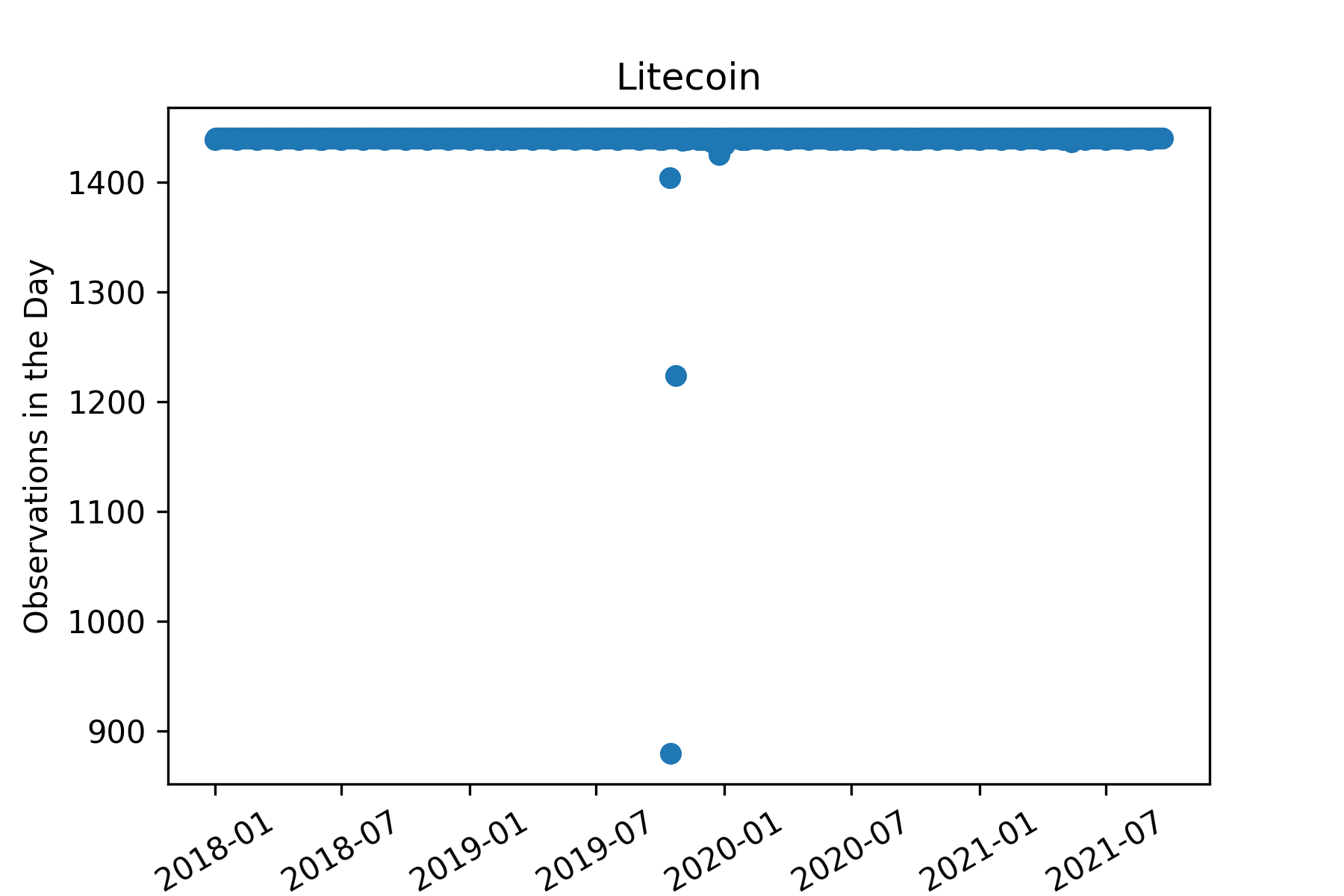
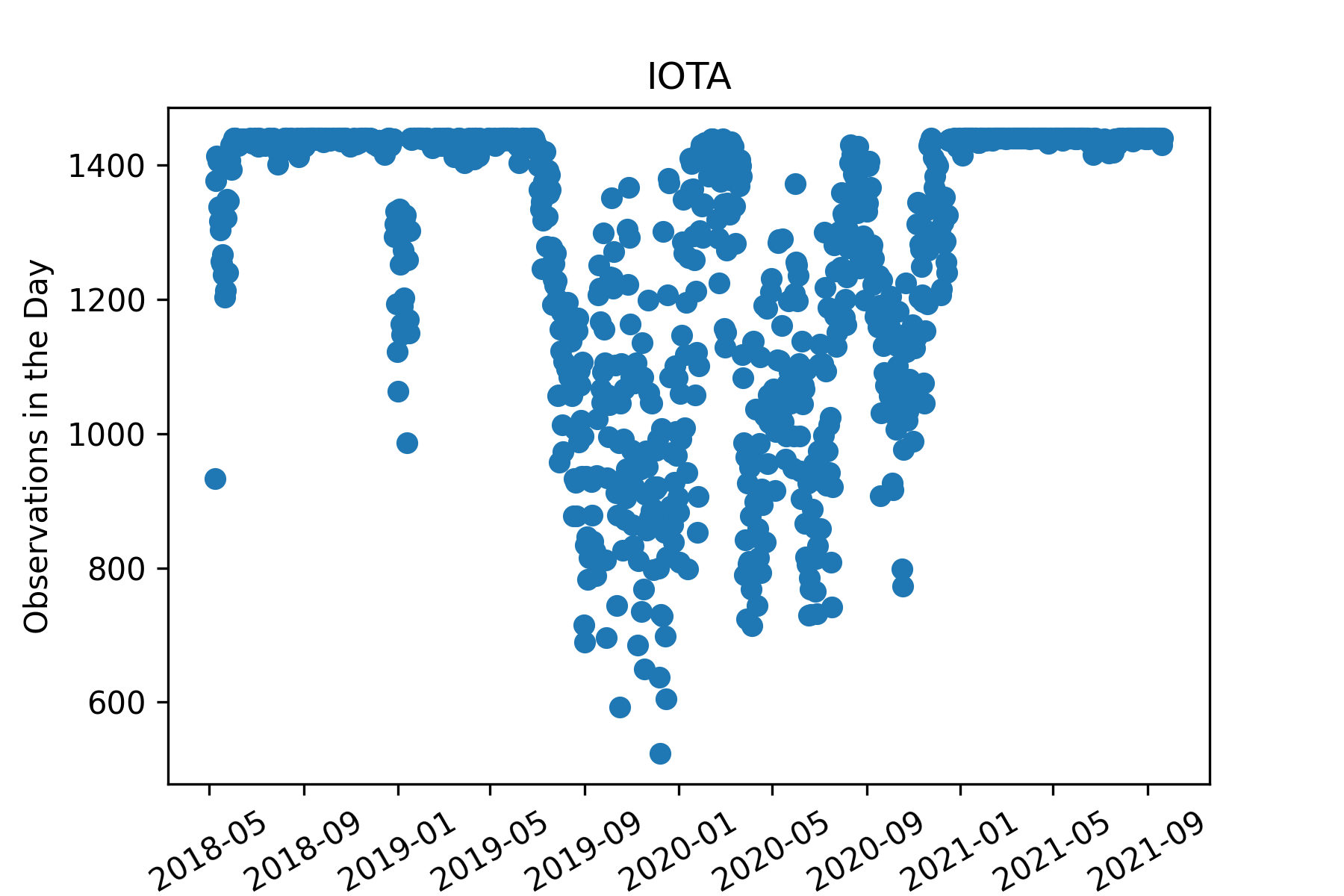
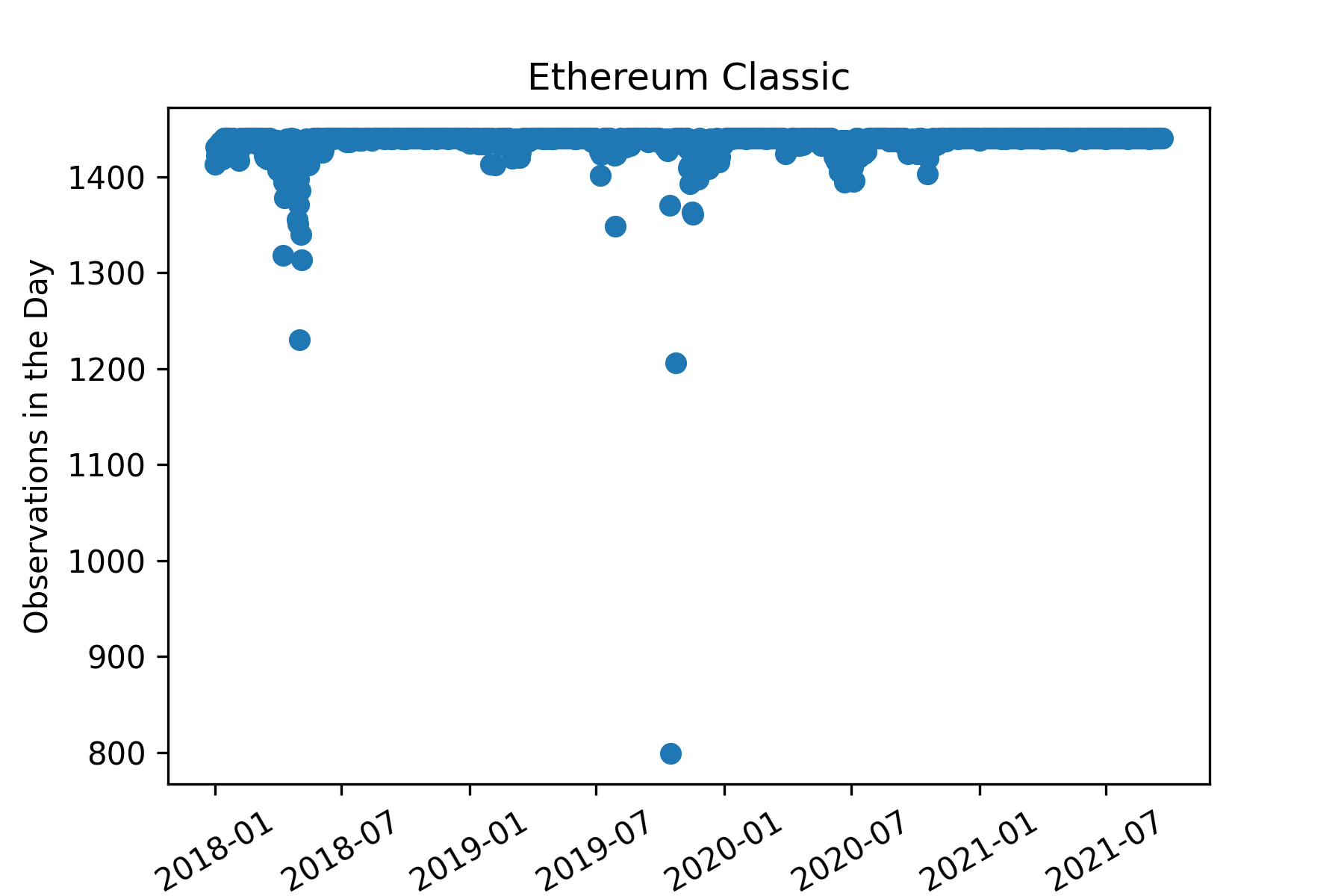
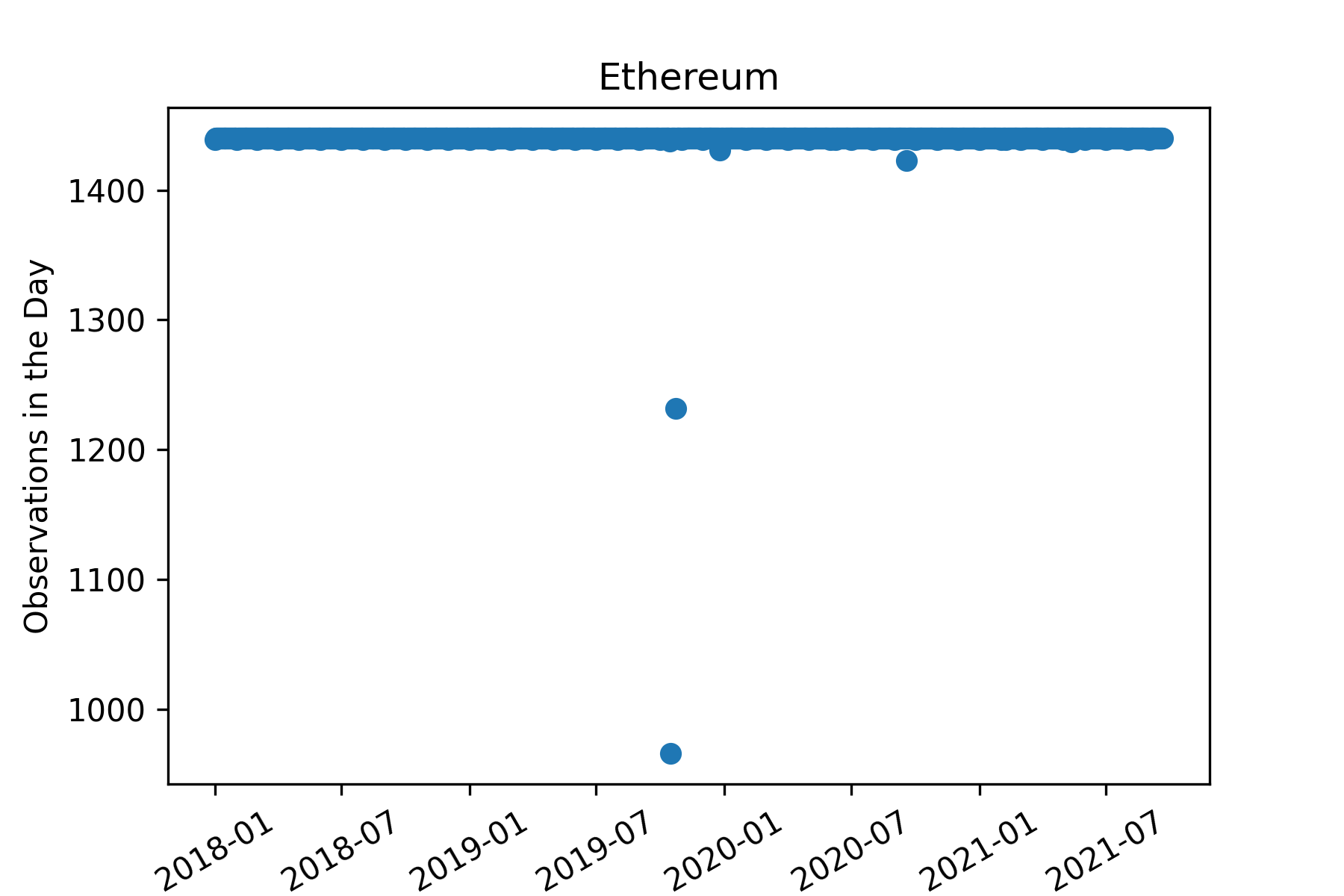
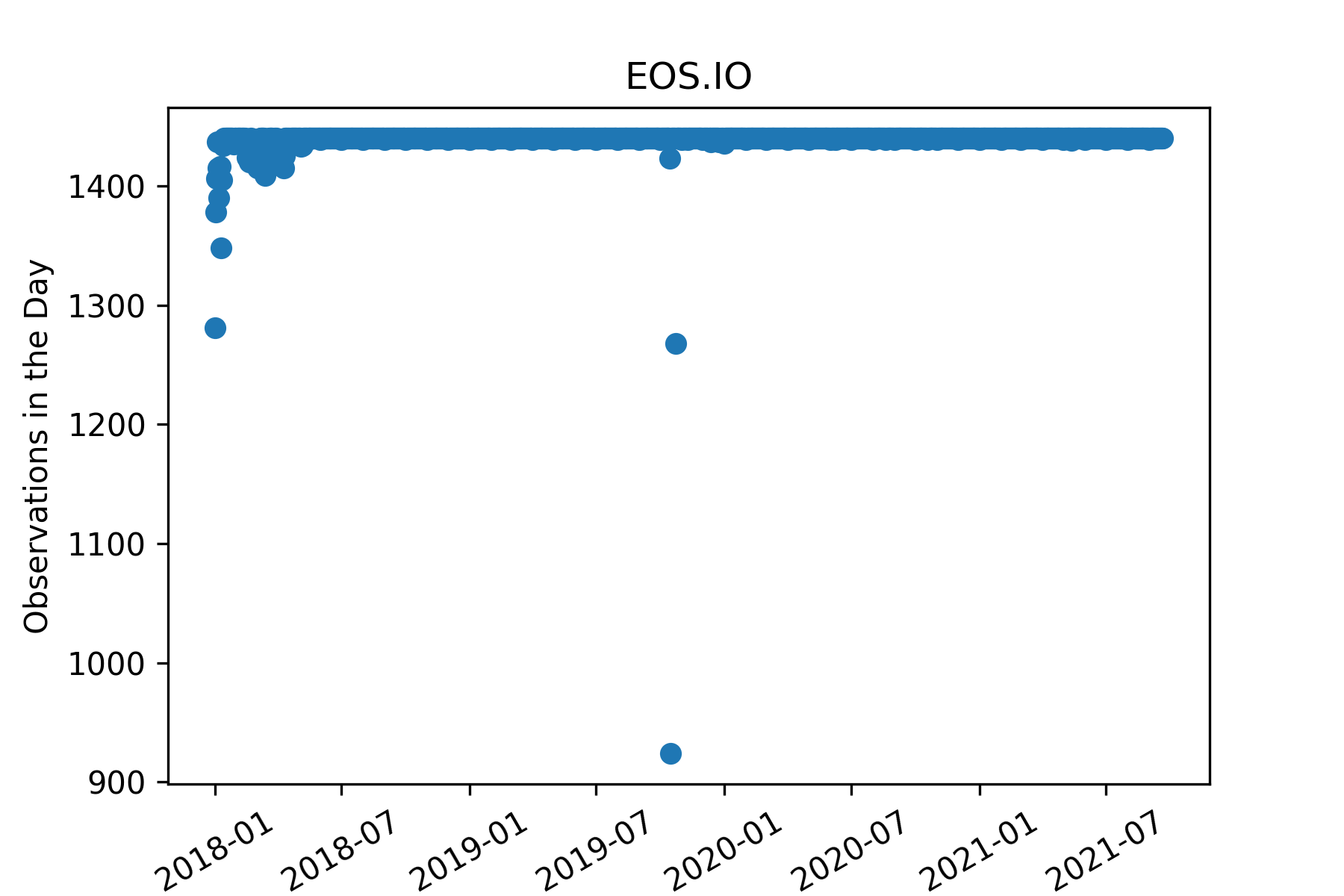
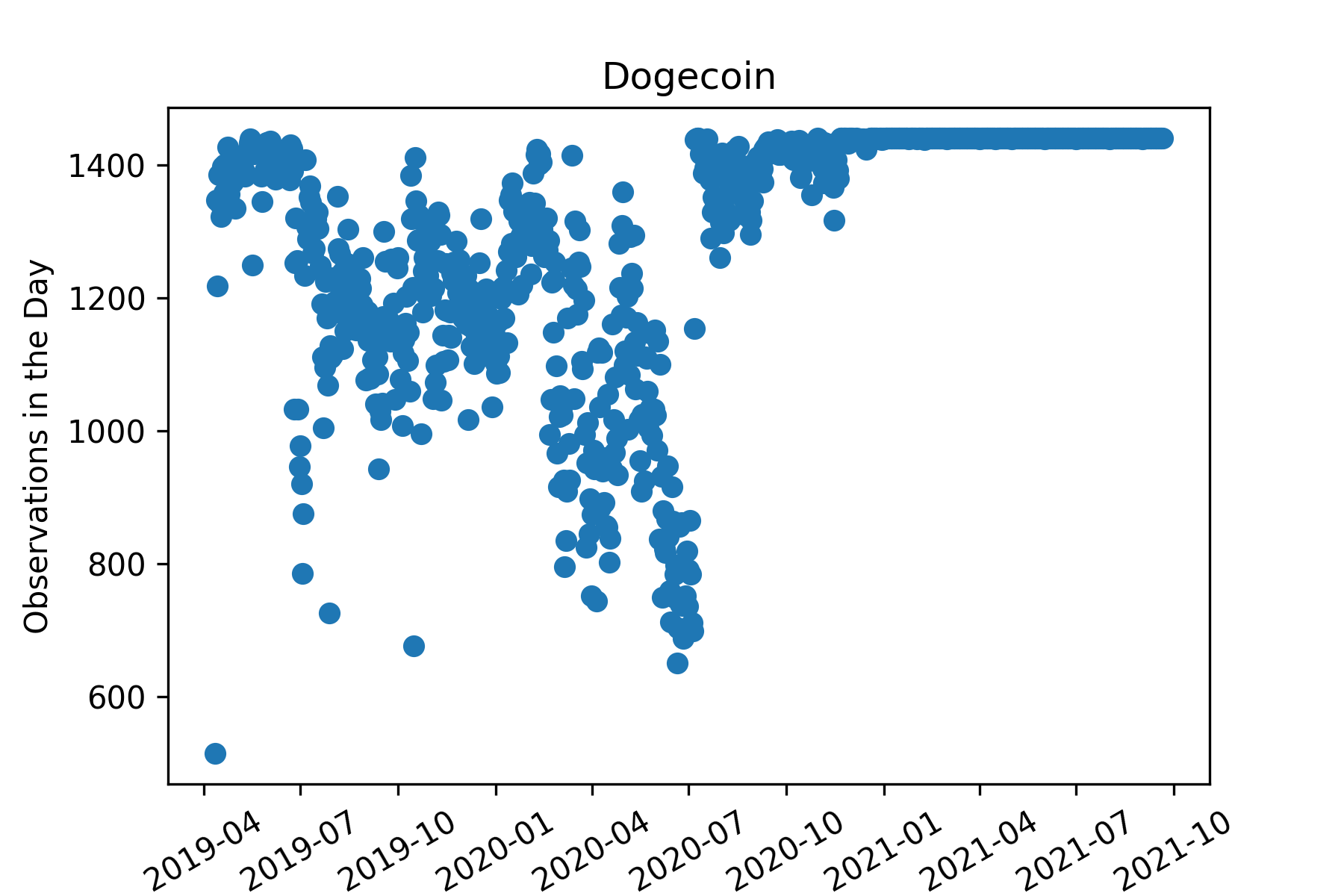
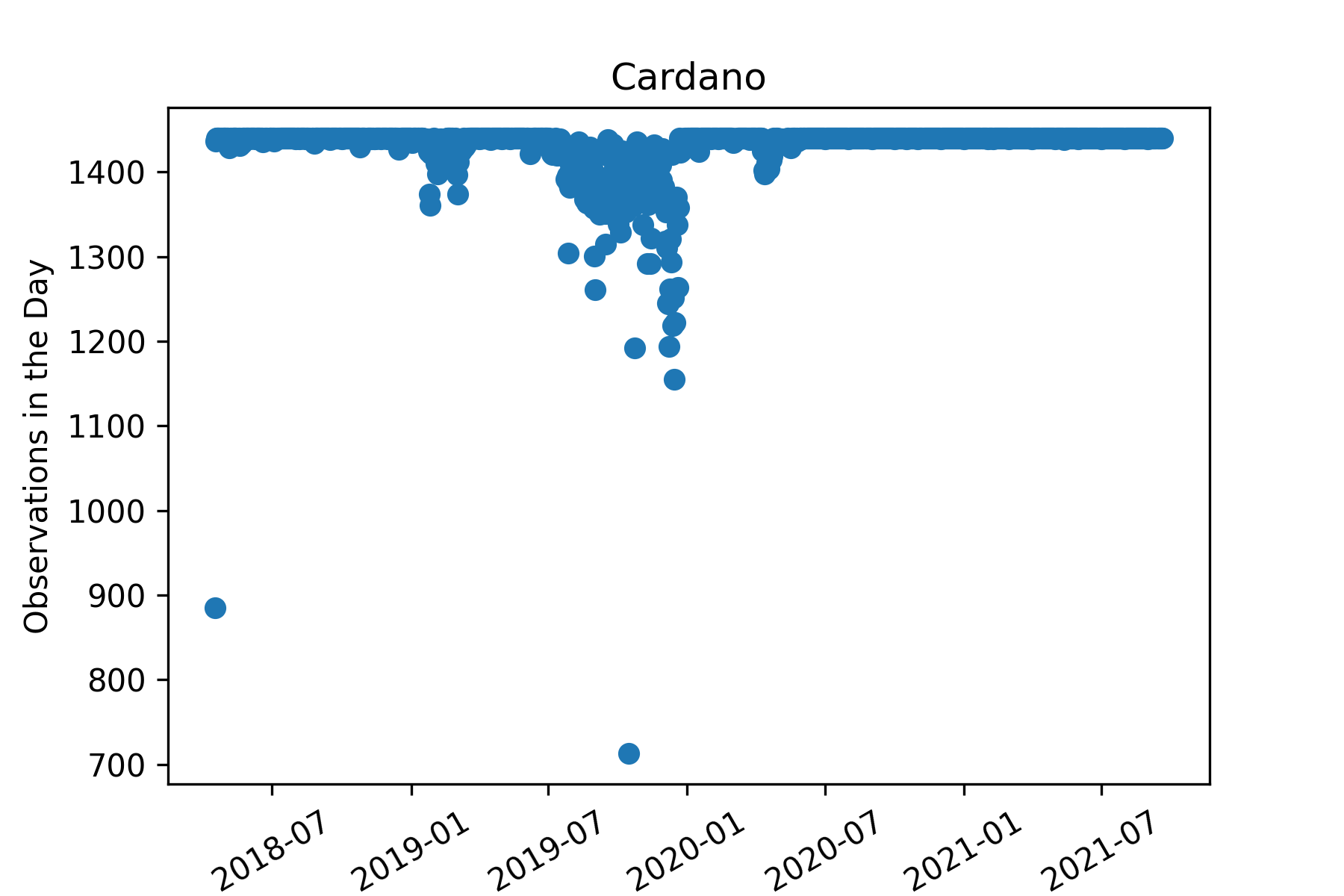
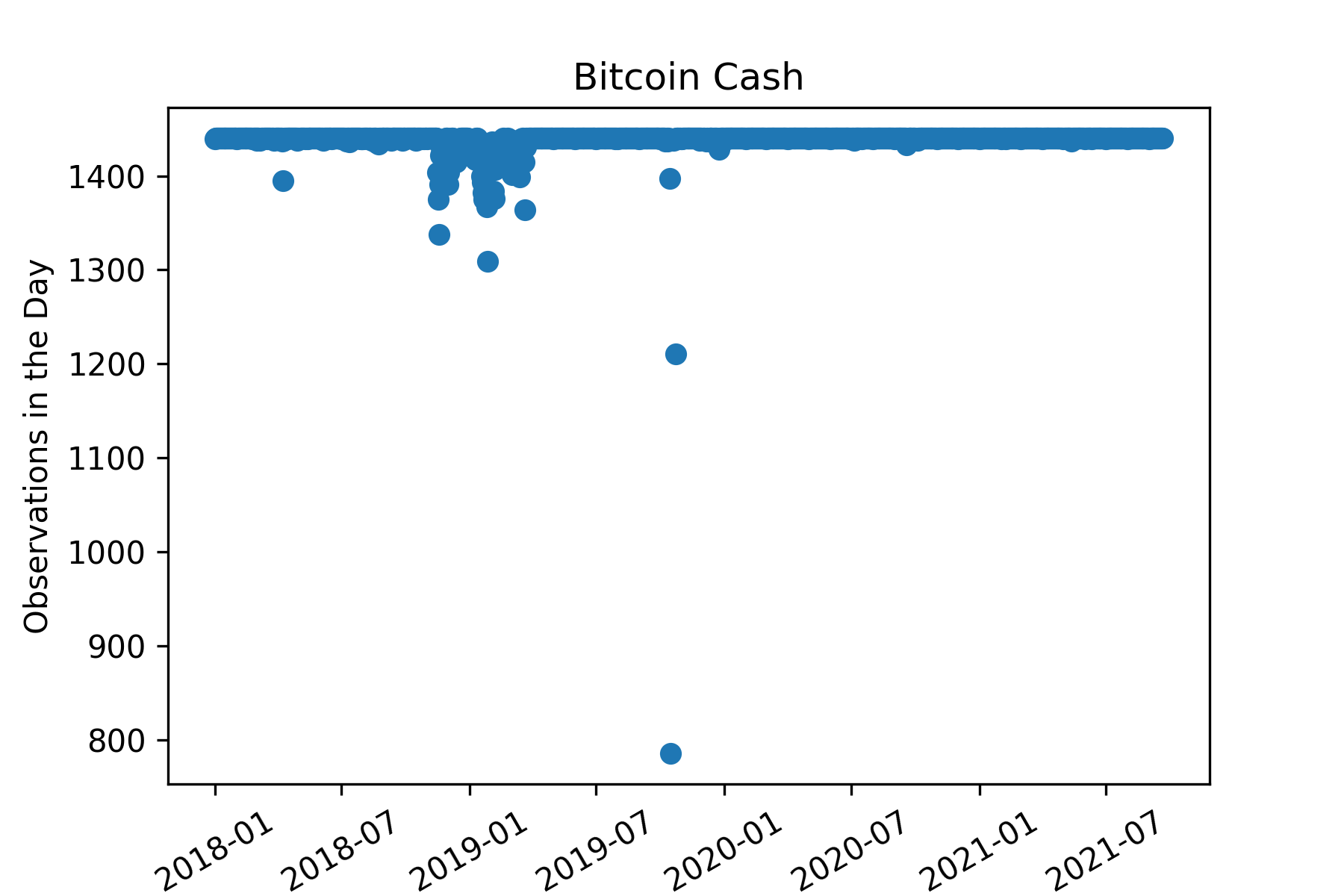
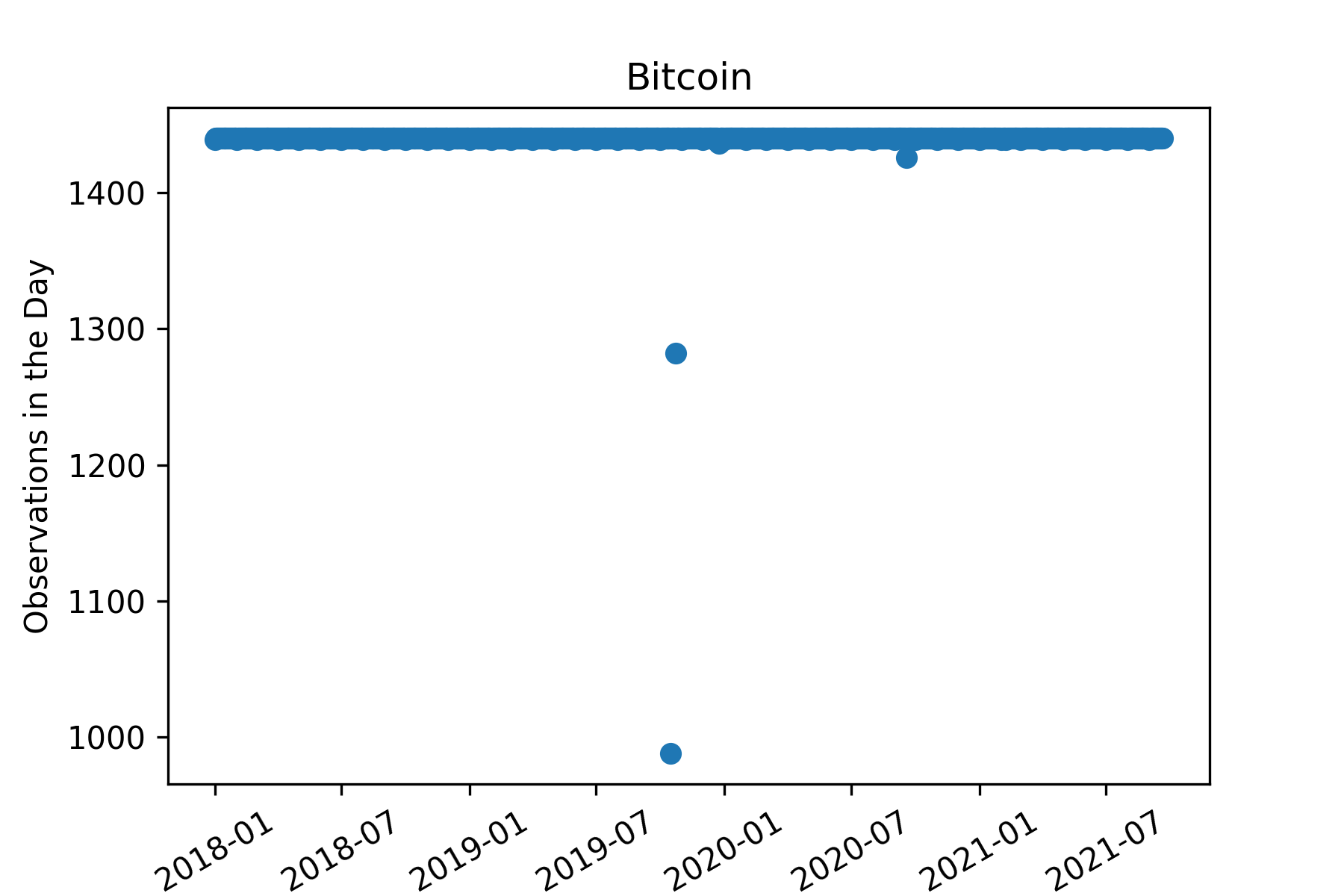
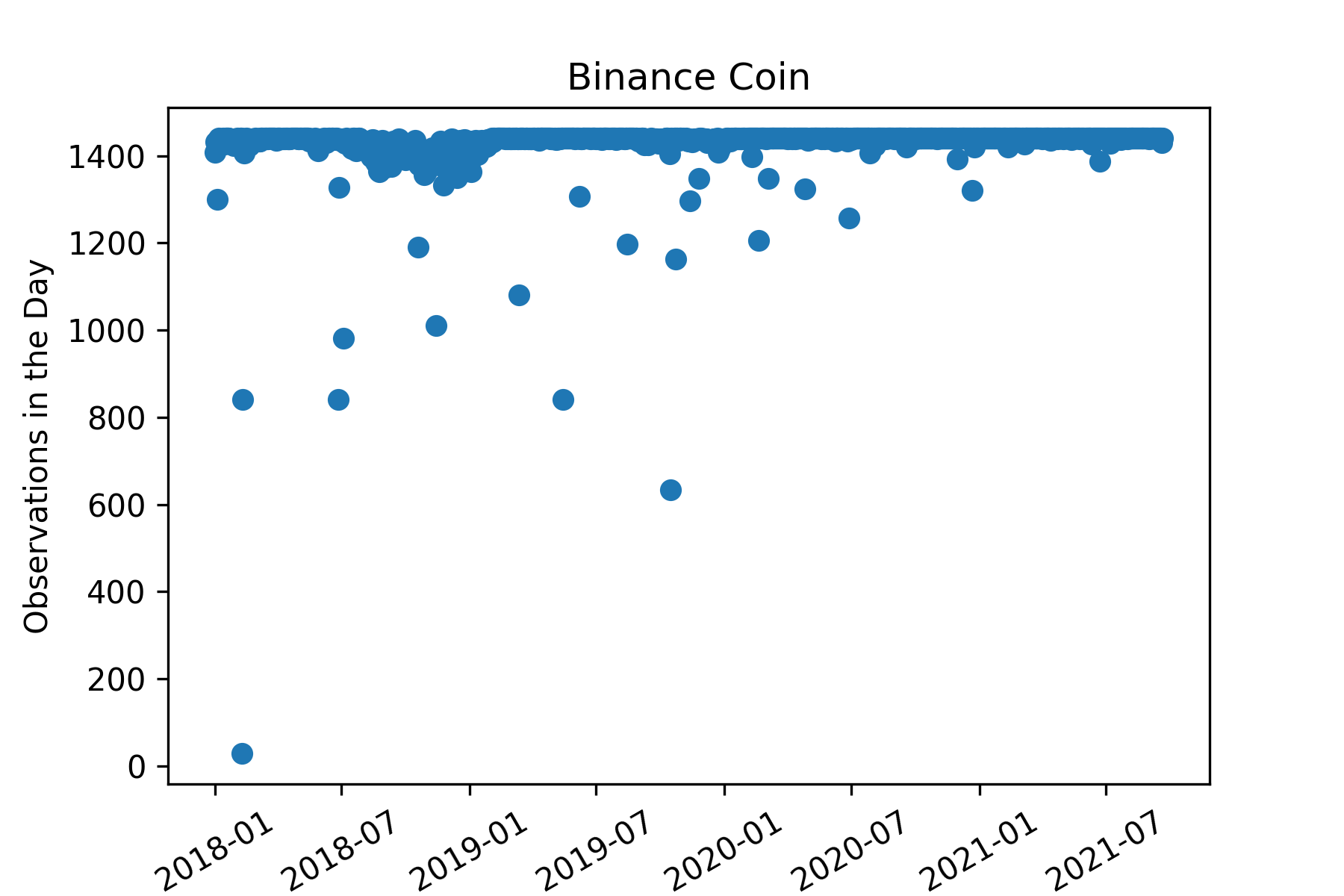
* Although every coin is described as going back to 1 Jan 2018, a good chunk of coins (sometimes referenced as shit coins) did not start or have clean data.
  + I.e. Dogecoin, monero, maker, tron, stellar, and EOS.io
* Total rows: 24,236,806; total columns: 10
  + Each row represents a coin’s values at a given minute
  + The columns are: timestamp, asset\_id, count, volume, vwap, open, high, low, close, target
  + If all the data is complete that means there would be 14 rows per minute (this is often not the case)

**Data Deep dive:**

* Many contestants used a purged group time series split (<https://www.kaggle.com/yamqwe/bottleneck-encoder-mlp-keras-tuner>)
  + This CV excels in time series data because it splits the data into equal groups based on how many days are available, then it provides a gap between the train data and the test data in order to prevent data leakage.



* Upon a close look at the data, many of the days across various coins do not have an equal amount of observations per period, thus this causes a major problem for this cross validation strategy to apply to our data:



* Coins like Maker, Monero, IOTA and doge would really struggle with this underlying cross validation strategy. Thus we wrote our own CV that did not split on days but instead collected the desired number of test windows, test window size, and gap, starting from the final data observations, and then used the remaining number of observations within the data to split into equal groups. This strategy seemed fair to us and we used a 3 fold validation, thus giving each fold around a year’s worth of data. A key reason this strategy seemed fair is because the identification of an individual day is not important to use in this category for training, for testing we tested on a 90 day window (5 days of data was used as an embargo to avoid any leakage, and then tested on the following 85 days for assessment)

**Feature Engineering:**

* The basics:
  + Time features augmented for cos and sin: Minute of day, month, total days into month (adjusted for month), day of week
  + Minute difference ((close/high) - (open/high)) & Max minute difference ((high/close) - (low / close))
  + After creation of the above features, log10 (Count, open, high, low, close, volume, VWAP, max min difference) in order to make the distribution more normal
    - Replacing any infinites with nan’s
* Aggregation across current period:
  + Groupby timestamp, mean, median, min, max, std across Count, Volume, Minute difference, mas min difference
  + For each row difference from market mean and median
  + Indicator column on if it the highest or smallest value for for the count, minute difference, or max min difference
* Lag features (60 minutes & 1 day (1440 minutes); rare cases different versions of lags) via bottleneck:
  + Mean and standard deviation across an assets values, and market trend mean and standard deviation
  + Lag peak detection (<https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data> )
    - Z-score threshold of -3 and 3
  + Comparison across lags

**Modeling:**

* Through initial Tuning of NN’s and boosted models it appeared that boosted models would outperform NN’s
* Final settle was on 1 set of models for each coin, each set contained 3 seeds across the top 3 architectures from hyperparameter tuning
  + Ensemble just simple average

**Reflection:**

**Things that didn’t work:**

* Neural nets
* Percentage of market (time component)

**Things that ran out of time to implement:**

* Target recreation
* Ensemble creativity
* Adding extra lags
* Bottleneck rank
* Rolling correlation
* Combining (shit) coins in order to give more data
* Model for market trend
* Model for weekdays vs weekend
  + Weekday avg of NA’s: 104,607.2
  + Weekend avg of NA’s: 113,652.5

**Learning Lessons:**

* Datatable and Jay files
* Pydata
  + Bottleneck
* Numpy speed

**Appendix:**

**2 lags trained on train data only, test on supplemental:**

Ensemble for Asset #0 ------> 0.05489978685598183

Model #0 ------> 0.055605301931092706

Model #1 ------> 0.03150891425452708

Model #2 ------> 0.04552724923447004

Model #3 ------> 0.05549736020953802

Model #4 ------> 0.04250921143316239

Model #5 ------> 0.02449433087441122

Model #6 ------> 0.04395408493347775

Model #7 ------> 0.030494428277514776

Model #8 ------> 0.059882000400826176

Ensemble for Asset #1 ------> 0.03482448584907717

Model #0 ------> 0.025517437806547444

Model #1 ------> 0.025520723239452647

Model #2 ------> 0.04161836511135604

Model #3 ------> 0.03596672522921222

Model #4 ------> 0.02612990413181417

Model #5 ------> 0.040653984739159155

Model #6 ------> 0.0316304524694895

Model #7 ------> 0.041131337674182346

Model #8 ------> 0.029084481586897724

Ensemble for Asset #2 ------> 0.022956567368125416

Model #0 ------> -0.010505979387490646

Model #1 ------> 0.004696940674313141

Model #2 ------> 0.03814045861704887

Model #3 ------> -0.011286084581107797

Model #4 ------> 0.06558633124780618

Model #5 ------> -0.036505228918525714

Model #6 ------> 0.018680757944038725

Model #7 ------> 0.013955549900876416

Model #8 ------> 0.0005017729970249732

Ensemble for Asset #3 ------> 0.11857614005980194

Model #0 ------> 0.09753269724548648

Model #1 ------> 0.08998812859483174

Model #2 ------> 0.11254710177273987

Model #3 ------> 0.11389111909269041

Model #4 ------> 0.12107327168224638

Model #5 ------> 0.11036218663297272

Model #6 ------> 0.11129838195752764

Model #7 ------> 0.1042534573261027

Model #8 ------> 0.10079517334625857

Ensemble for Asset #4 ------> -0.007339482411860706

Model #0 ------> 0.005355812784814241

Model #1 ------> -0.036476706926822644

Model #2 ------> -0.0012228818333863238

Model #3 ------> 0.010166138836122489

Model #4 ------> -0.0032863497038309415

Model #5 ------> 0.00915614740131027

Model #6 ------> -0.006081137934105178

Model #7 ------> 0.003231886854223952

Model #8 ------> -0.00577166430183598

Ensemble for Asset #5 ------> 0.009483944659693468

Model #0 ------> -0.021738609615861108

Model #1 ------> 0.0403009616771241

Model #2 ------> 0.019950424557962348

Model #3 ------> 0.004756793663758685

Model #4 ------> 0.017023766258562778

Model #5 ------> 0.011073548098456103

Model #6 ------> -0.005790441236430271

Model #7 ------> -0.0018119194941972803

Model #8 ------> 0.003122513021802349

Ensemble for Asset #6 ------> -0.00123692666713881

Model #0 ------> 0.0020250021183445257

Model #1 ------> 0.008857796877412285

Model #2 ------> 0.0027813822073275234

Model #3 ------> 0.00455365311557369

Model #4 ------> -0.007637699383901387

Model #5 ------> 0.012460032257835613

Model #6 ------> -0.0014285500990616358

Model #7 ------> -0.013347560332302803

Model #8 ------> -0.00913725661423536

Ensemble for Asset #7 ------> 0.013534762198076254

Model #0 ------> 0.02205048333915212

Model #1 ------> 0.002223070621136548

Model #2 ------> 0.014445892254155308

Model #3 ------> 0.017450630966822575

Model #4 ------> -0.005208419871568901

Model #5 ------> 0.0226131711097532

Model #6 ------> 0.012158883902096667

Model #7 ------> 0.01915223935959196

Model #8 ------> -0.003915934184682412

Ensemble for Asset #8 ------> 0.03234950117406302

Model #0 ------> 0.017583692037804178

Model #1 ------> 0.01021967993702651

Model #2 ------> 0.0009908040351366571

Model #3 ------> 0.008017778885414247

Model #4 ------> 0.014875203551105947

Model #5 ------> 0.026198656364133946

Model #6 ------> 0.03218483999257563

Model #7 ------> 0.02886584470779041

Model #8 ------> 0.017093228842346243

Ensemble for Asset #9 ------> 0.027884121585575725

Model #0 ------> 0.02775836325828246

Model #1 ------> 0.029914257492724607

Model #2 ------> 0.029480174073590267

Model #3 ------> 0.02737267379578931

Model #4 ------> 0.02255817434814509

Model #5 ------> 0.026908737223271806

Model #6 ------> 0.023104561538854356

Model #7 ------> 0.02548479313890824

Model #8 ------> 0.025529030964724385

Ensemble for Asset #10 ------> 0.03289864612842949

Model #0 ------> 0.034260976242433114

Model #1 ------> 0.04150628433723162

Model #2 ------> 0.045972902625914566

Model #3 ------> 0.01731163785675921

Model #4 ------> 0.0019539793172354894

Model #5 ------> 0.03260967326915544

Model #6 ------> -0.02420503471420639

Model #7 ------> 0.011012508816987378

Model #8 ------> 0.035962187945150347

Ensemble for Asset #11 ------> 0.008490996738880962

Model #0 ------> -0.0017716582030696227

Model #1 ------> 0.0020044877139272395

Model #2 ------> 0.005135752517889964

Model #3 ------> 0.009476167555816766

Model #4 ------> -0.004400432748436358

Model #5 ------> 0.022989346556524976

Model #6 ------> 0.020400227577608888

Model #7 ------> 0.004188535877164469

Model #8 ------> 0.012783736290447892

Ensemble for Asset #12 ------> -0.015060256843022498

Model #0 ------> -0.0182894827666569

Model #1 ------> -0.018055178648273988

Model #2 ------> -0.020625831340325588

Model #3 ------> -0.01960410186256491

Model #4 ------> -0.005462617923753512

Model #5 ------> 0.0017478925728764236

Model #6 ------> 0.0004592821286504606

Model #7 ------> -0.006977011345340167

Model #8 ------> -0.022424444303078998

Ensemble for Asset #13 ------> 0.03747178972821625

Model #0 ------> 0.03516504826096025

Model #1 ------> 0.023046798767399592

Model #2 ------> 0.047964492139656784

Model #3 ------> 0.021450229603581276

Model #4 ------> 0.0323565688906301

Model #5 ------> 0.02999853760570405

Model #6 ------> 0.03086322423940253

Model #7 ------> 0.0007740029401947963

Model #8 ------> 0.018085759815238956

**4 lags with rank on 2 trained on train data only, test on supplemental:**

Ensemble for Asset #0 ------> 0.043050296567046176

Model #0 ------> 0.038874619792531186

Model #1 ------> 0.0547932975104669

Model #2 ------> 0.052335856837285276

Model #3 ------> 0.0546335664885307

Model #4 ------> 0.02971314466975484

Model #5 ------> 0.06068992312042598

Model #6 ------> 0.0075368427200273435

Model #7 ------> 0.031173757847298124

Model #8 ------> 0.01401232887551888

Ensemble for Asset #1 ------> 0.04382086624322857

Model #0 ------> 0.012094766253698873

Model #1 ------> 0.02549184814795384

Model #2 ------> 0.03528475582669088

Model #3 ------> 0.038479688897242605

Model #4 ------> 0.04153994414989542

Model #5 ------> 0.03466770290285093

Model #6 ------> 0.054451912229656674

Model #7 ------> 0.04632668713229702

Model #8 ------> 0.04263995804136259

Ensemble for Asset #2 ------> 0.02310266870829585

Model #0 ------> 0.004932455869215748

Model #1 ------> 0.04053188737942581

Model #2 ------> 0.04372386469024854

Model #3 ------> 0.00694976068301534

Model #4 ------> -0.00978567260990626

Model #5 ------> -0.0012768501834834768

Model #6 ------> 0.02459316081436302

Model #7 ------> 0.058906753476727414

Model #8 ------> -0.002559684602212865

Ensemble for Asset #3 ------> 0.11783010437848213

Model #0 ------> 0.12231376526913892

Model #1 ------> 0.11172675473257293

Model #2 ------> 0.10782330226048274

Model #3 ------> 0.11095782131156595

Model #4 ------> 0.11316761074649122

Model #5 ------> 0.11244167878635451

Model #6 ------> 0.11322544948549684

Model #7 ------> 0.11026561640731267

Model #8 ------> 0.11317359784873894

Ensemble for Asset #4 ------> -0.013328881146636771

Model #0 ------> -0.011403488970050826

Model #1 ------> -0.00899155482368153

Model #2 ------> -0.03890374347372186

Model #3 ------> -0.02295866716175948

Model #4 ------> -0.005481913081313332

Model #5 ------> 0.03062022047219496

Model #6 ------> -0.003400849862169105

Model #7 ------> -0.025360931784711764

Model #8 ------> -0.00940198789671617

Ensemble for Asset #5 ------> 0.011705333437464535

Model #0 ------> 0.010418296930486749

Model #1 ------> 0.01638393929105574

Model #2 ------> 0.0033514500972035466

Model #3 ------> 0.010559442829378192

Model #4 ------> 0.010437396105568864

Model #5 ------> 0.001444730657311613

Model #6 ------> 0.011942746822966937

Model #7 ------> 0.02187462132831324

Model #8 ------> 0.010332579701227675

Ensemble for Asset #6 ------> -0.0021592823244161777

Model #0 ------> 0.0002445076463184744

Model #1 ------> -0.007359170609918752

Model #2 ------> -0.0047178820289291105

Model #3 ------> -0.002014466934622423

Model #4 ------> 0.0012592970804924397

Model #5 ------> 0.0022834997134015614

Model #6 ------> -0.002001483680546416

Model #7 ------> -0.0015011204683148052

Model #8 ------> -0.0029483062274193515

Ensemble for Asset #7 ------> 0.023058469616629756

Model #0 ------> 0.03124731472730907

Model #1 ------> 0.024954556797408087

Model #2 ------> 0.00291430612169319

Model #3 ------> 0.02221903817414169

Model #4 ------> 0.021347717660899254

Model #5 ------> 0.029797589277667697

Model #6 ------> 0.021711230509011043

Model #7 ------> 0.018695670239299318

Model #8 ------> 0.007292440428198338

Ensemble for Asset #8 ------> 0.03660434492619522

Model #0 ------> 0.02731987717821833

Model #1 ------> 0.029603258902347735

Model #2 ------> 0.0288393971641042

Model #3 ------> 0.02787238288969766

Model #4 ------> 0.03377930845154292

Model #5 ------> 0.036467621670176385

Model #6 ------> 0.02792857312212685

Model #7 ------> 0.028563901906603694

Model #8 ------> 0.0348851343279684

Ensemble for Asset #9 ------> 0.0005494854274450786

Model #0 ------> 0.0008818695340270425

Model #1 ------> -0.007128841440498494

Model #2 ------> 2.6367732936954865e-05

Model #3 ------> 0.0009788805958819485

Model #4 ------> -0.008146543932006584

Model #5 ------> -0.0016869341790532032

Model #6 ------> 0.010466036432021613

Model #7 ------> 0.0004785469914241015

Model #8 ------> 0.005525172683532352

Ensemble for Asset #10 ------> 0.005257245108997063

Model #0 ------> 0.0003124201837292257

Model #1 ------> 0.007194201583935411

Model #2 ------> 0.005186142688370716

Model #3 ------> 0.00545504315428577

Model #4 ------> -0.0028863186203982767

Model #5 ------> 0.008048973758632023

Model #6 ------> 0.006496469651264343

Model #7 ------> 0.004673389525248803

Model #8 ------> -0.0025813317768673735

Ensemble for Asset #11 ------> 0.014112852906942534

Model #0 ------> 0.02339024883368815

Model #1 ------> 0.005981769889280012

Model #2 ------> 0.006808607695221492

Model #3 ------> 0.004430026618696107

Model #4 ------> 0.0027097701710667816

Model #5 ------> 0.01179170730682084

Model #6 ------> 0.00842179793686802

Model #7 ------> 0.01599002724822653

Model #8 ------> 0.028264659860623386

Ensemble for Asset #12 ------> -0.005777887574824087

Model #0 ------> 0.005514593138915706

Model #1 ------> 0.010312377936574746

Model #2 ------> 0.01639539033246443

Model #3 ------> -0.019821477136744887

Model #4 ------> -0.005401062760997682

Model #5 ------> -0.0014407617867482614

Model #6 ------> -0.030644467359564786

Model #7 ------> -0.003179025655502778

Model #8 ------> -0.02879207247658946

Ensemble for Asset #13 ------> 0.04994905270526877

Model #0 ------> 0.04914231745035585

Model #1 ------> 0.04473760034892098

Model #2 ------> 0.04420992108672259

Model #3 ------> 0.03604077182848184

Model #4 ------> 0.05106913458337379

Model #5 ------> 0.04569558612276077

Model #6 ------> 0.038722031199410346

Model #7 ------> 0.05357800514496348

Model #8 ------> 0.043560242984038176

