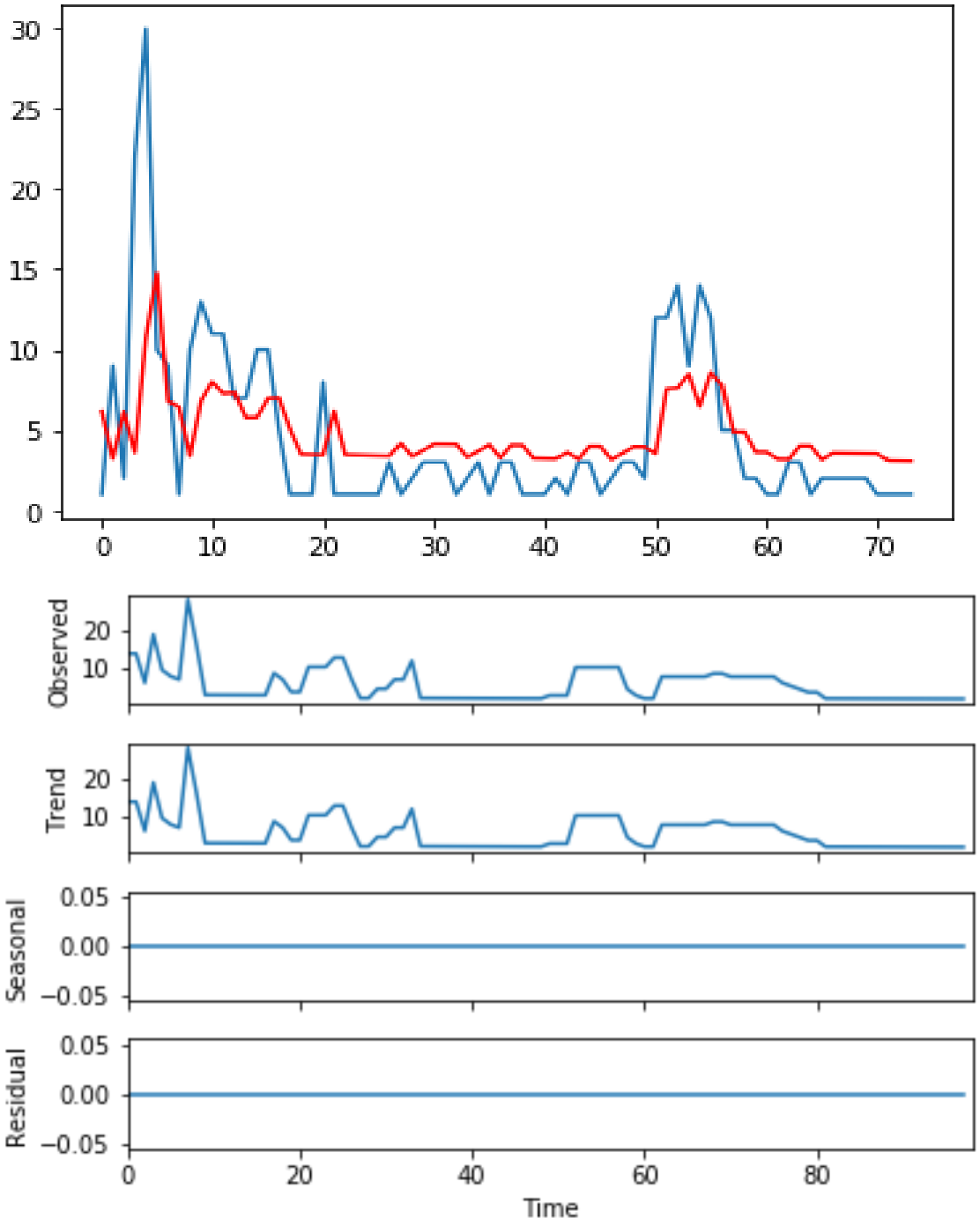
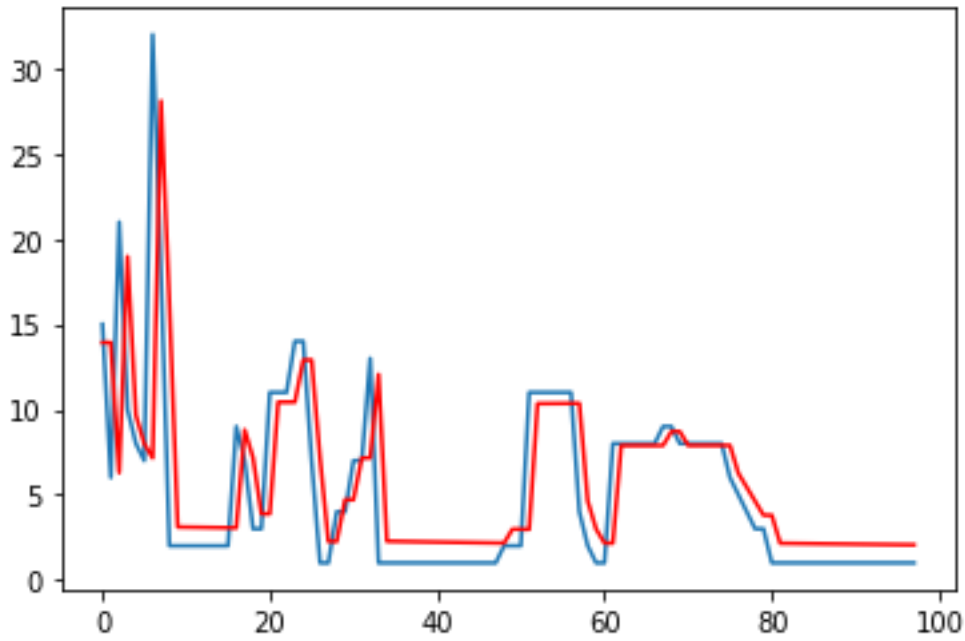
Assignment_5

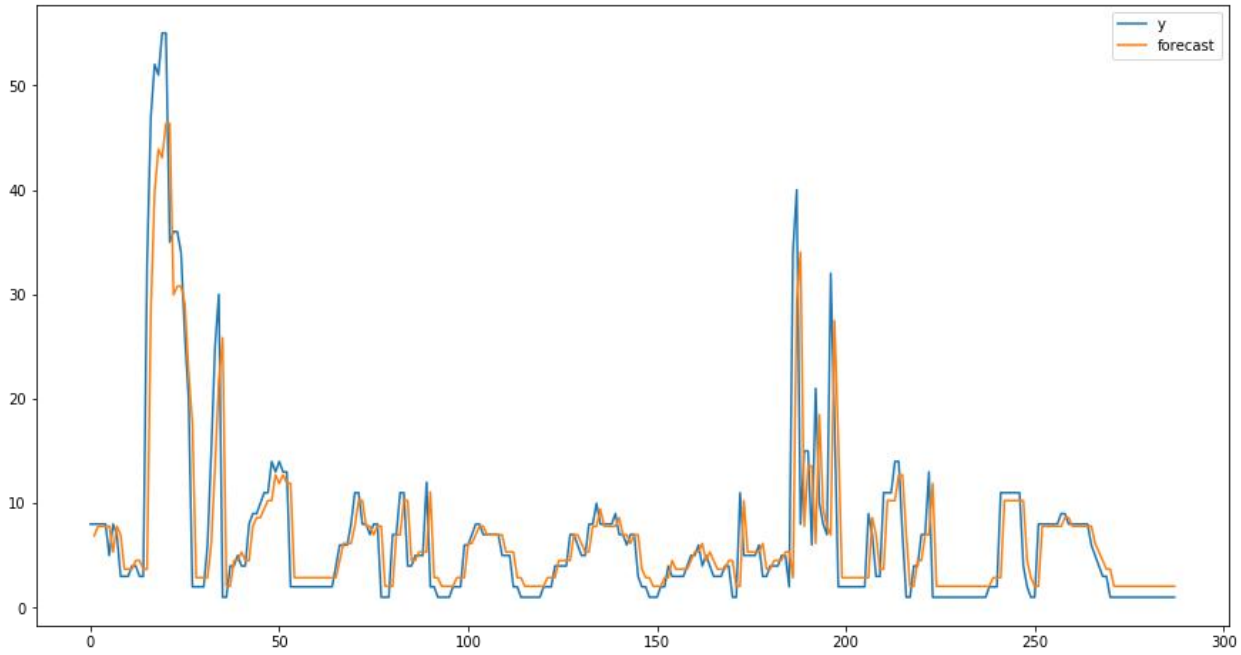
Rupa Rajendran

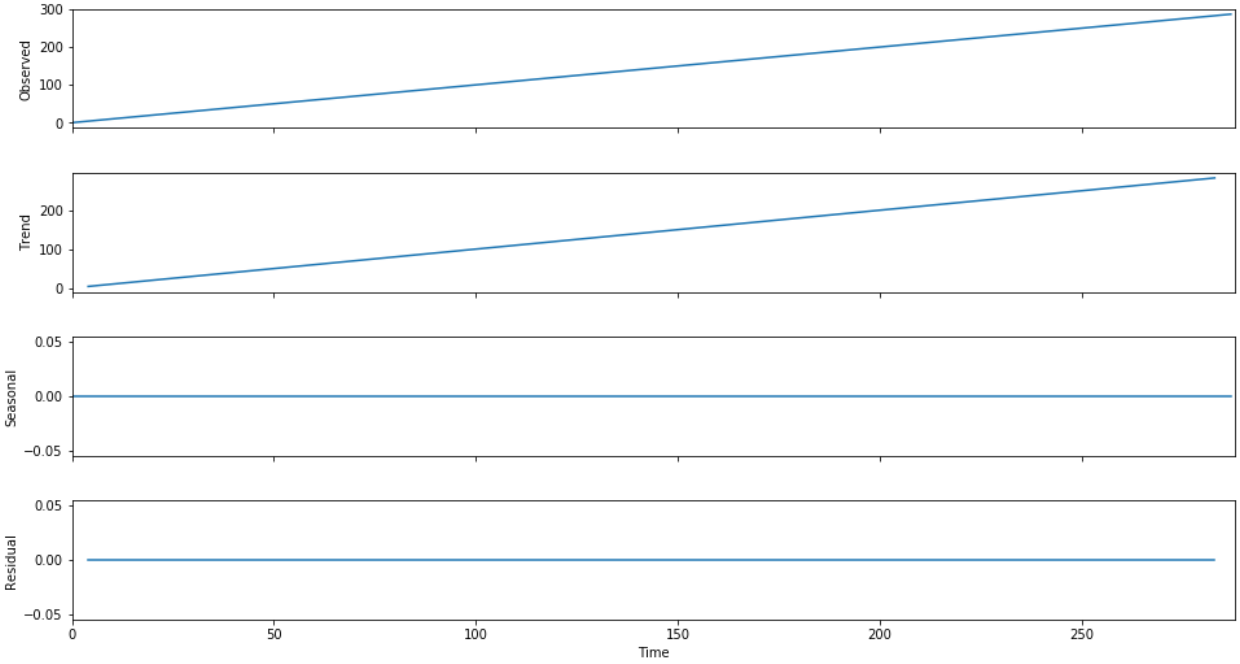
- Are there any useful patterns of outliers in the dataset?
- What are the UCL (Upper Control Limit) and LCL (Lower Control Limit) for certain issue metrics?
- Can the outliers detect hidden problems in the given dataset?
- What is the correlation between outliers and hidden problems in the given dataset?
- How to detect if there is problem hidden in the given dataset?
- How to detect if certain engineer deliberately creates issues with Priority Critical?
- How to detect if certain origination phase causes majority of the in progressCritical-Bug issues?
- Can you chart the patterns of outliers in the dataset?
- Can you create the right pivot stackedbar chart? How to group multi-levels?
- Group by Origination phase or Category for example. How many Levels of indexing?
- Should the Priority be displayed in a pivotchart of DetectionPhase for example
- What is the avg number of issues opened per DetectionPhase?
- What is the avg turn around time per issue (time from the day the issue created till it got closed)?
- What is the avg number of rejected issues opened per eningeer?
- What is the avg number of critical issues opened per eningeer?
- What is the avg number of rejected issues per OriginationPhase?
- What is the avg number of critical issues per OriginationPhase?
- What is the avg number of created issues per OriginationPhase?
- What is the avg number of rejected critical issues per OriginationPhase?
- What is the ratio of total number of critical to medium issues per OriginationPhase?
- Which month got the maximum number of Critical issues created? Which week got the minimum number of issues created?

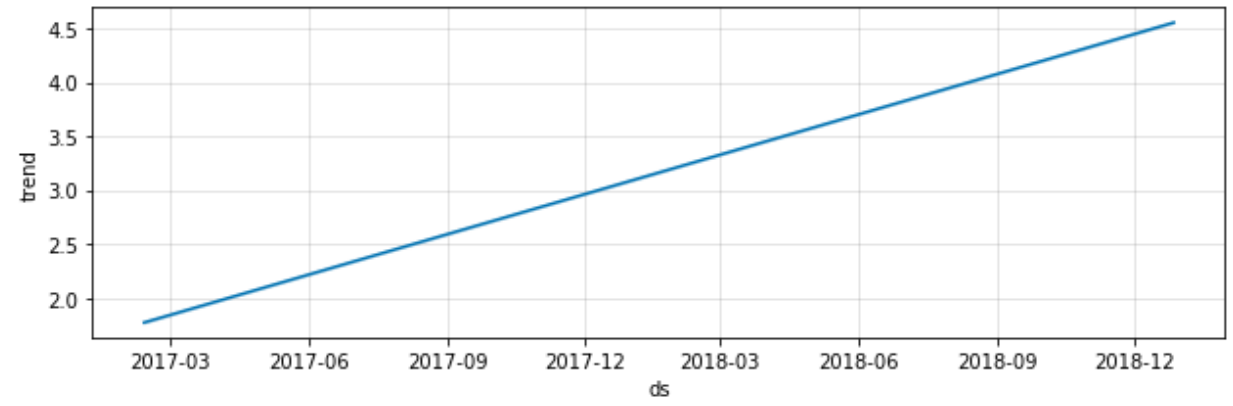
  For the given data the following graphs gives us an insight on how the date set works and we can try analysing from those graphs the key points.

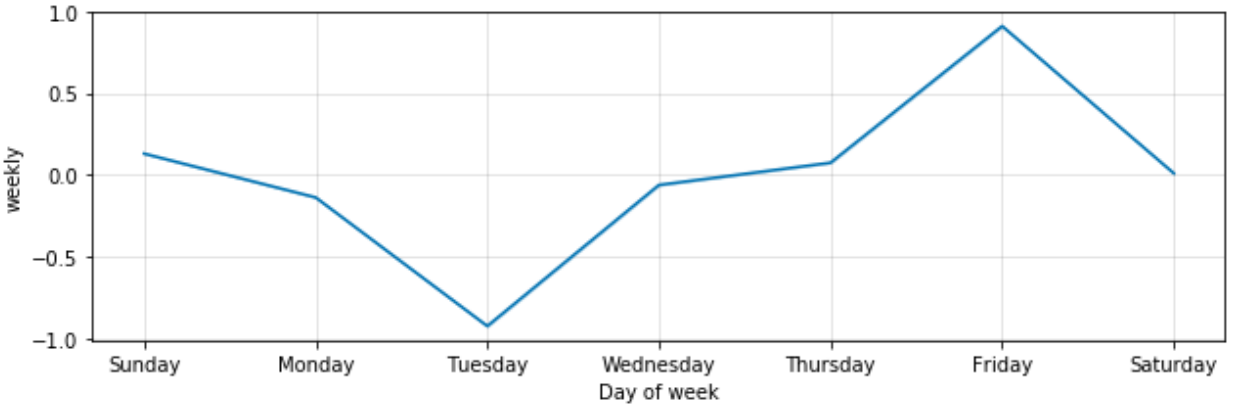

30
25
20
15
10
5
0
0
10
20
30
40
50
60
70
Observed
20
10
Trend
20
10
Seasonal
0.05
0.00
−0.05
Residual
0.05
0.00
−0.05
0
20
40
60
80
Time

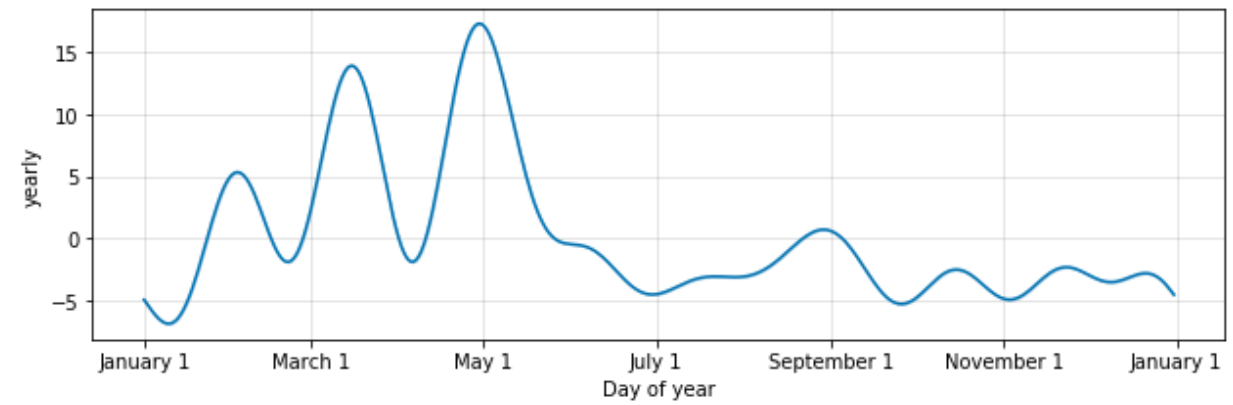

30
25
20
15
10
5
0
0
20
40
60
80
100

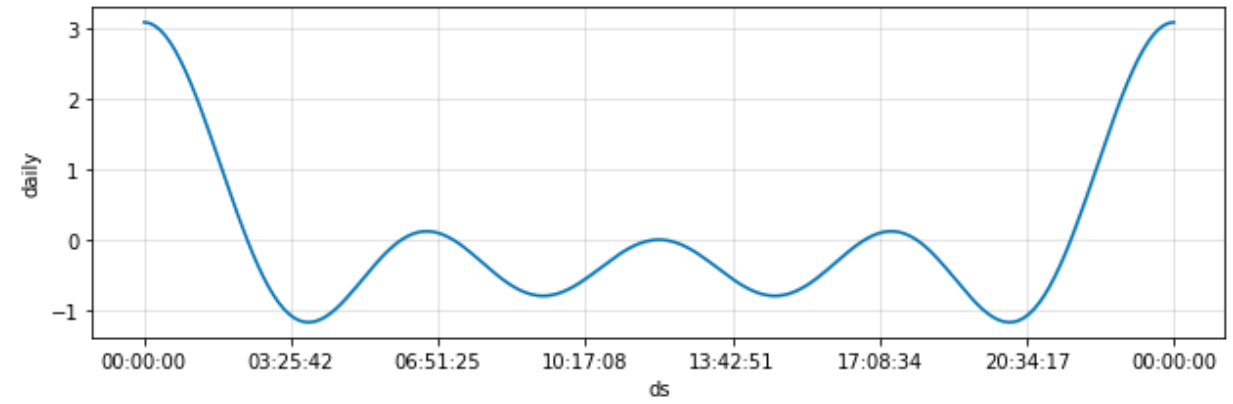

y
forecast
50
40
30
20
10
0
0
50
100
150
200
250
300

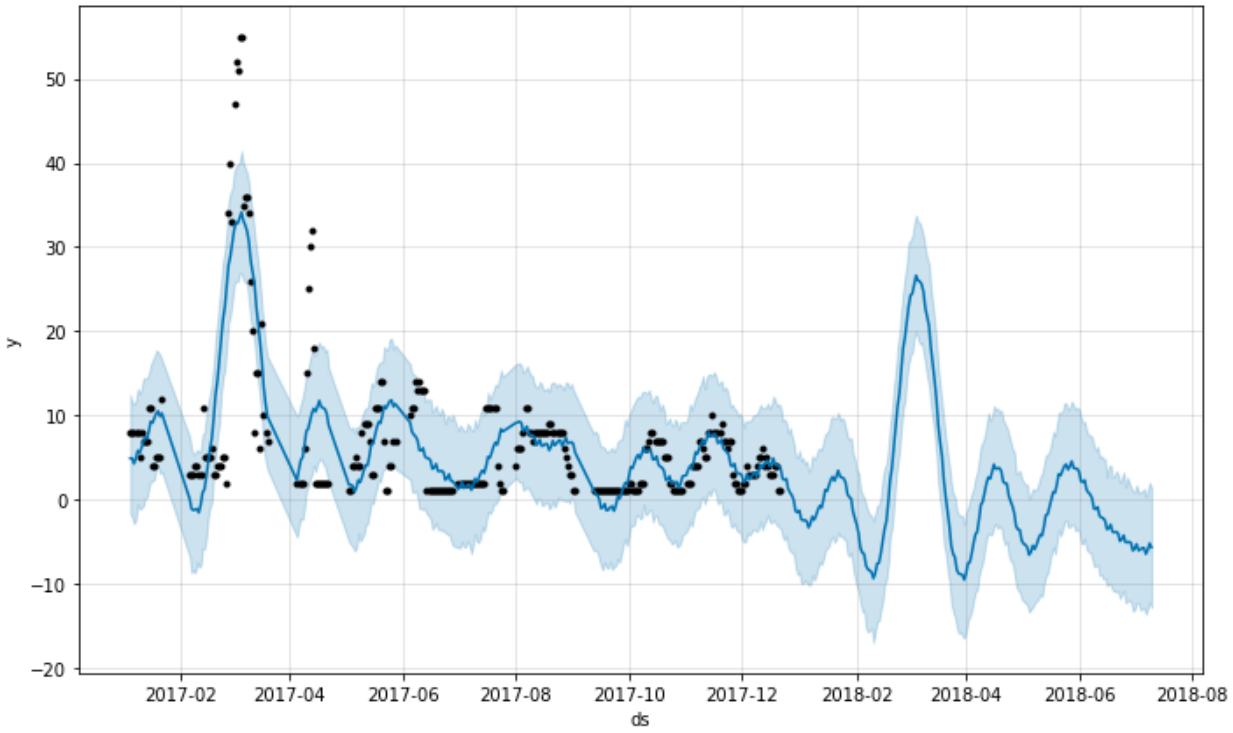

Observed
300
200
100
0
Trend
200
100
0
Seasonal
0.05
0.00
−0.05
Residual
0.05
0.00
−0.05
0
50
100
150
200
250
Time

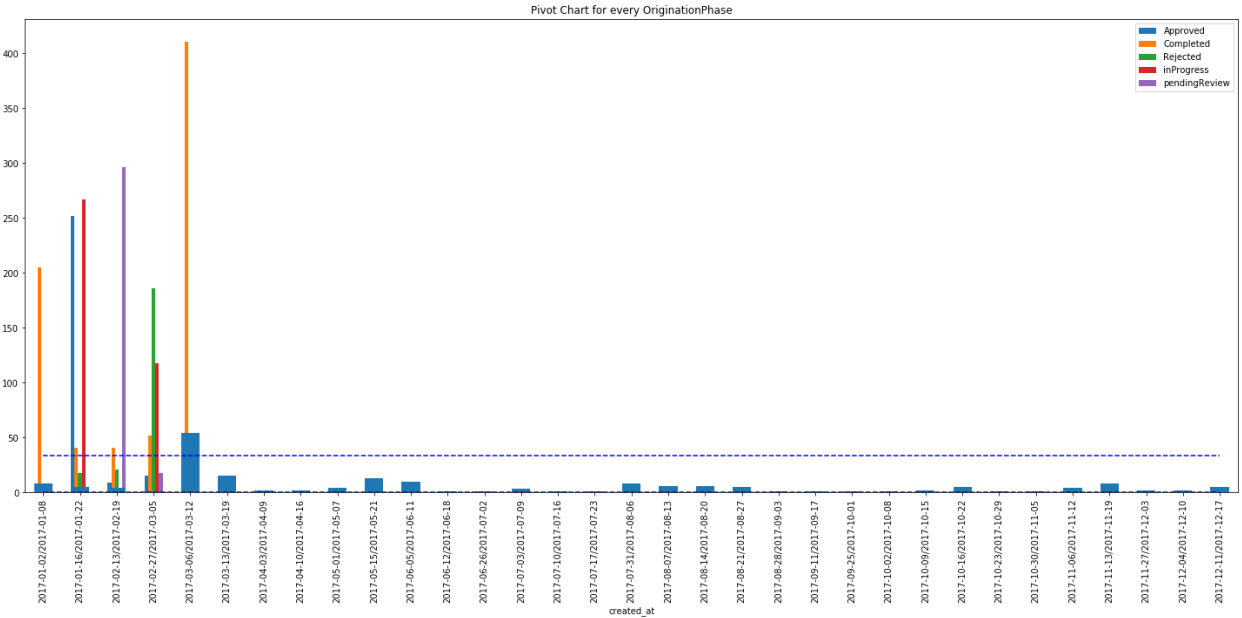

trend
4.5
4.0
3.5
3.0
2.5
2.0
2017-03
2017-06
2017-09
2017-12
2018-03
2018-06
2018-09
2018-12
ds


weekly
1.0
0.5
0.0
−0.5
−1.0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week

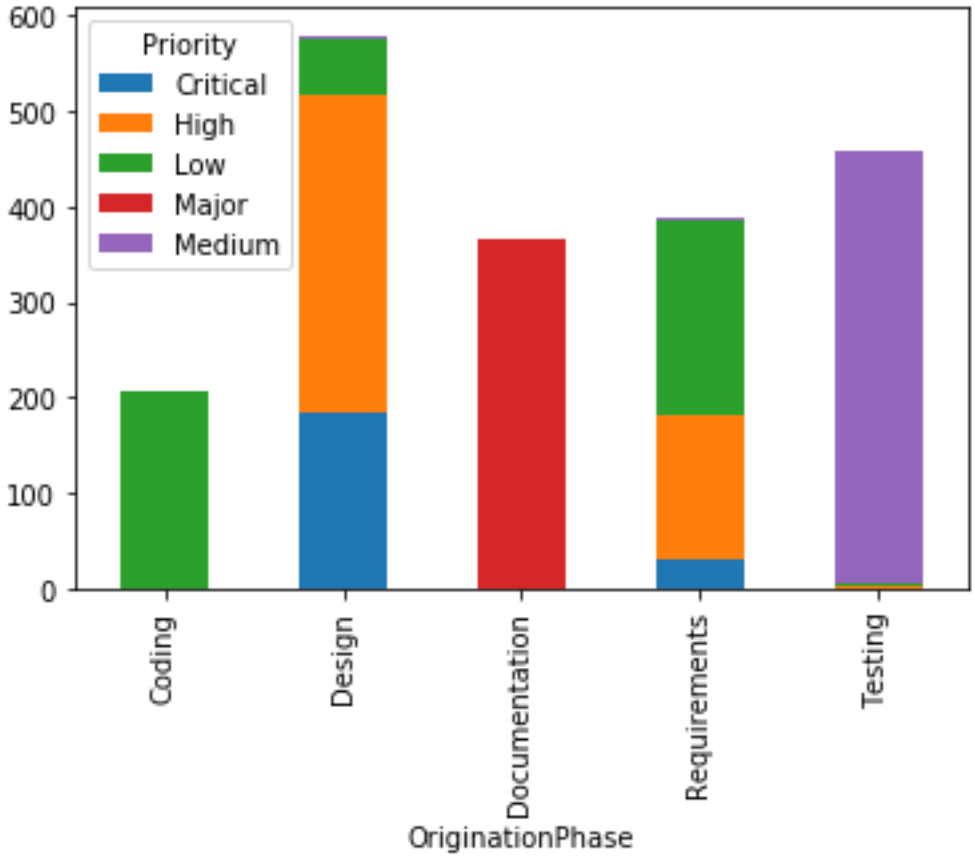

yearly
15
10
5
0
−5
January 1
March 1
May 1
July 1
September 1
November 1
January 1
Day of year

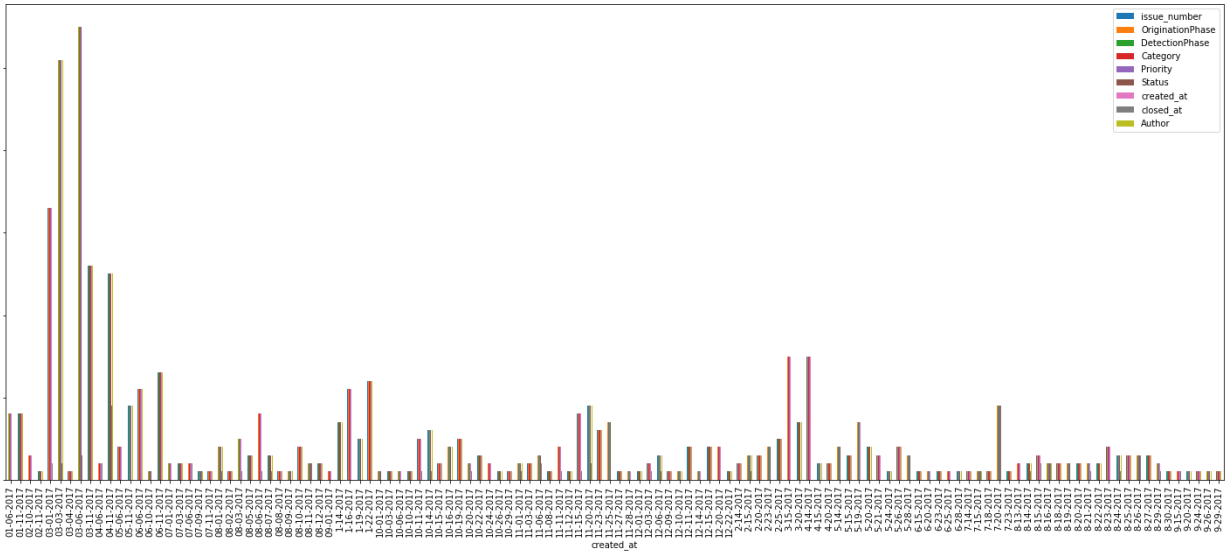

daily
3
2
1
0
−1
00:00:00
03:25:42
06:51:25
10:17:08
13:42:51
17:08:34
20:34:17
00:00:00
ds

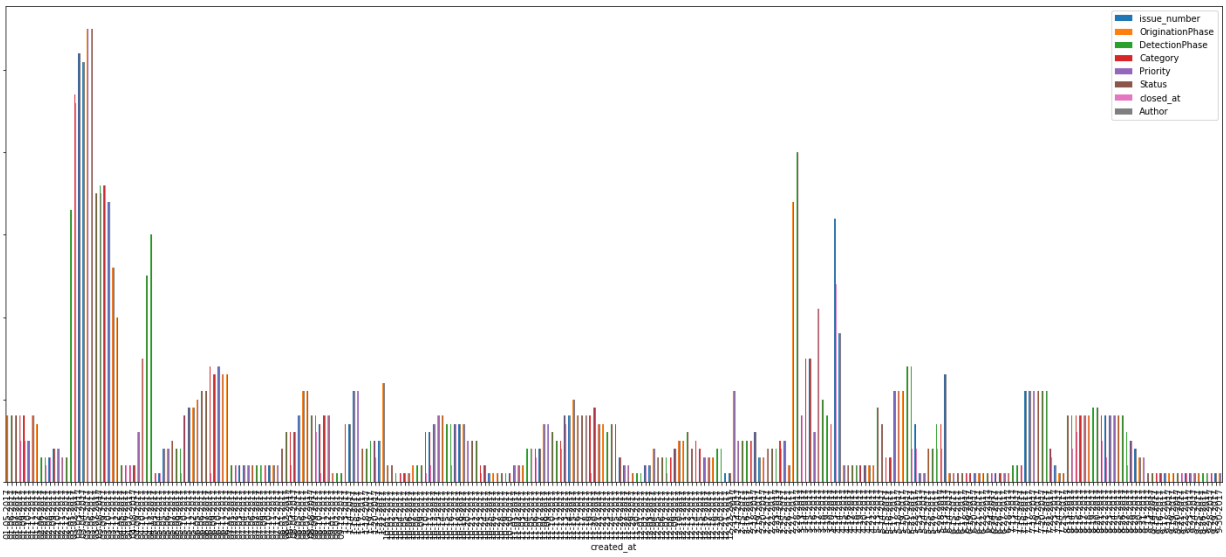

y
ds
50
40
30
20
10
0
−10
−20
2017-02
2017-04
2017-06
2017-08
2017-10
2017-12
2018-02
2018-04
2018-06
2018-08

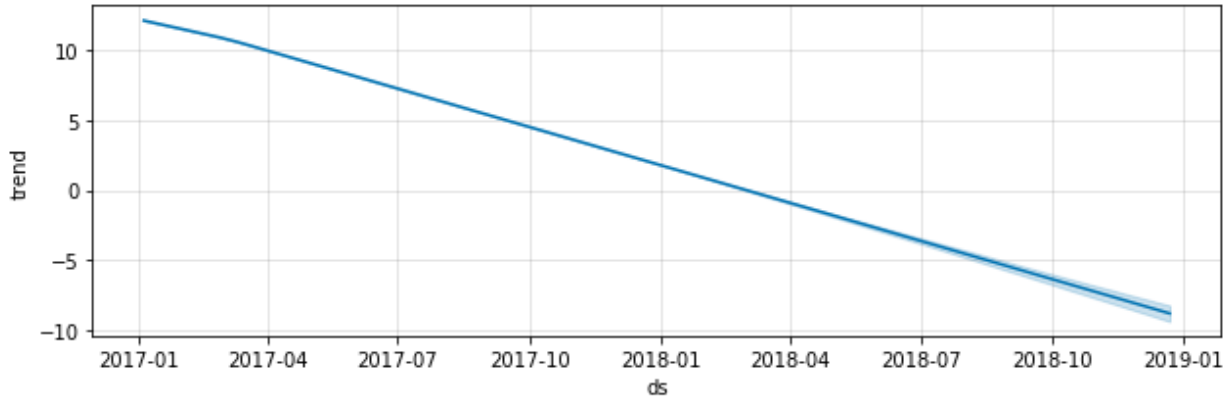

Pivot Chart for every OriginationPhase
Approved
Completed
Rejected
inProgress
pendingReview
created_at

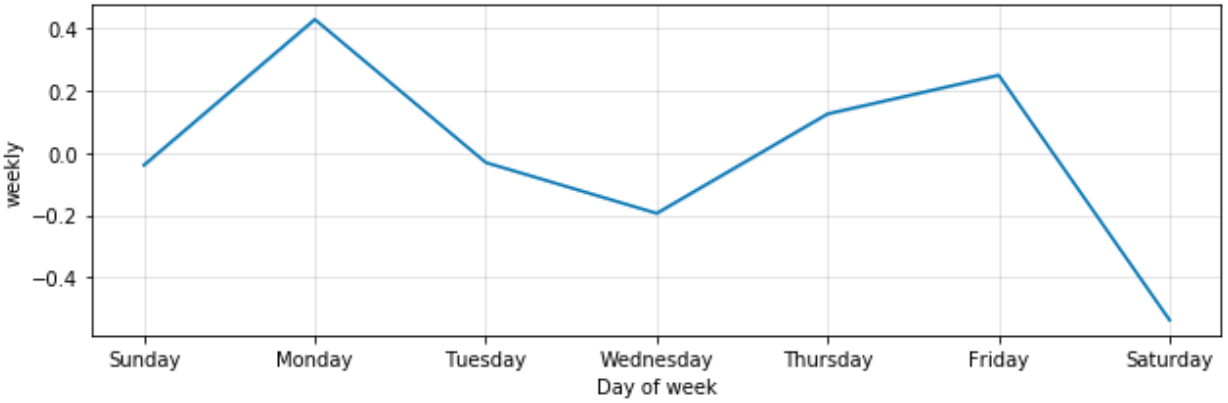

Bar chart for every detection phase
closed_at
06-08-2017
09-09-2017
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Testing
DetectionPhase

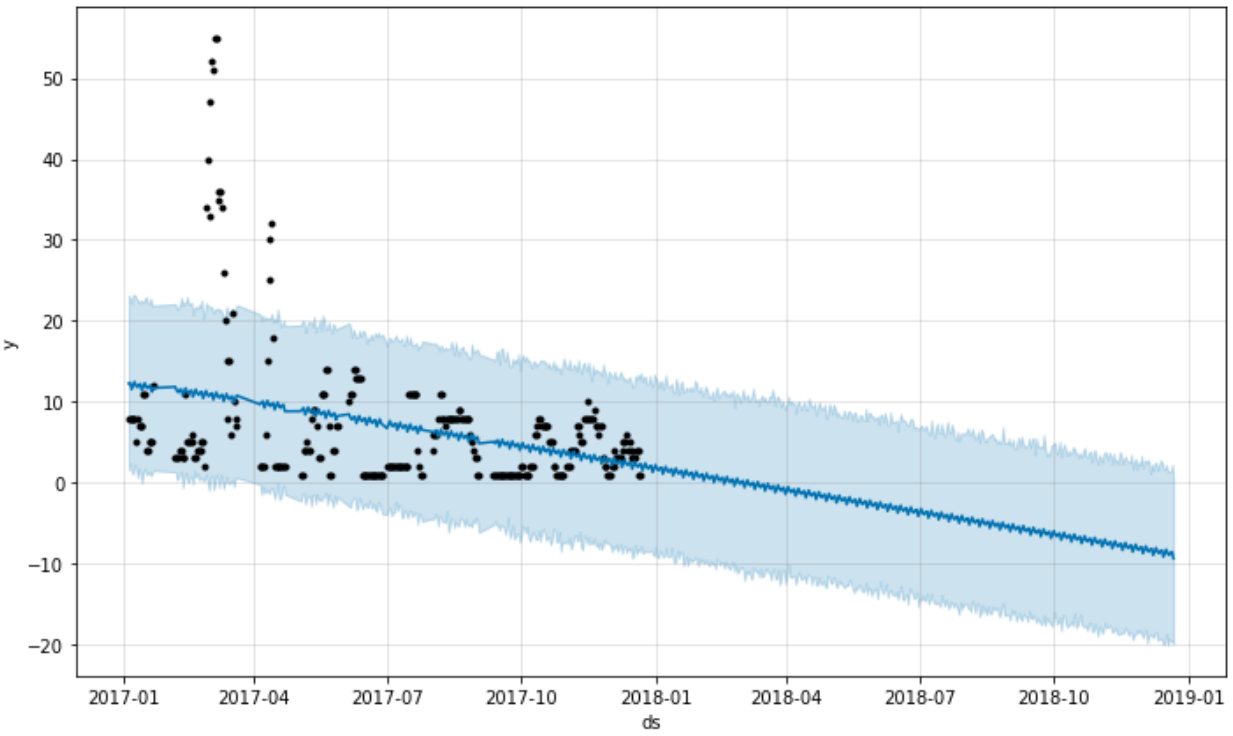

Priority
Critical
High
Low
Major
Medium
600
500
400
300
200
100
0
Coding
Design
Documentation
Requirements
Testing
OriginationPhase

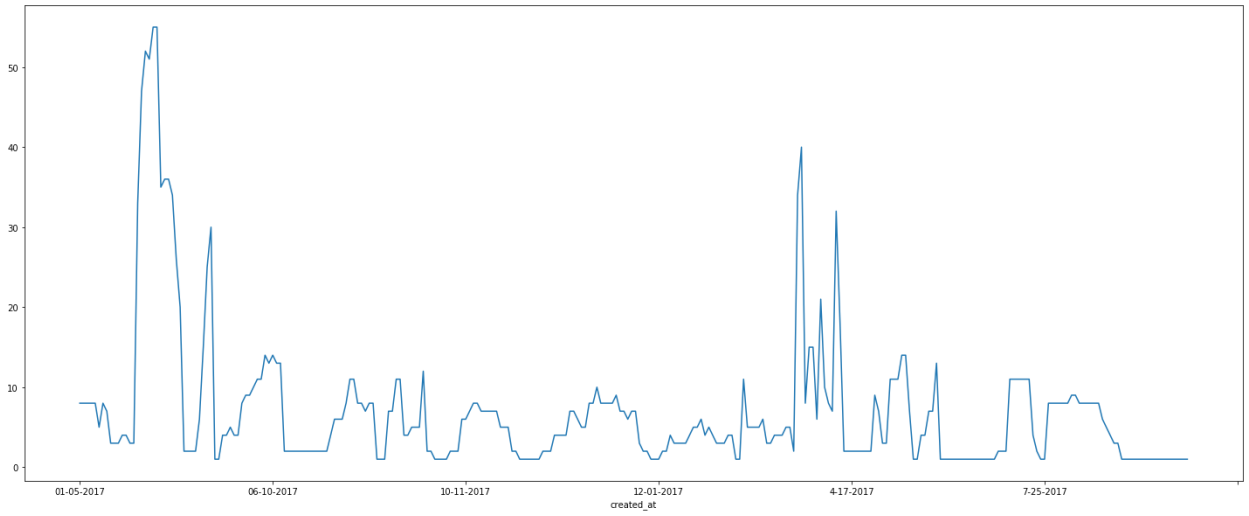

issue_number
OriginationPhase
DetectionPhase
Category
Priority
Status
created_at
closed_at
Author
created_at

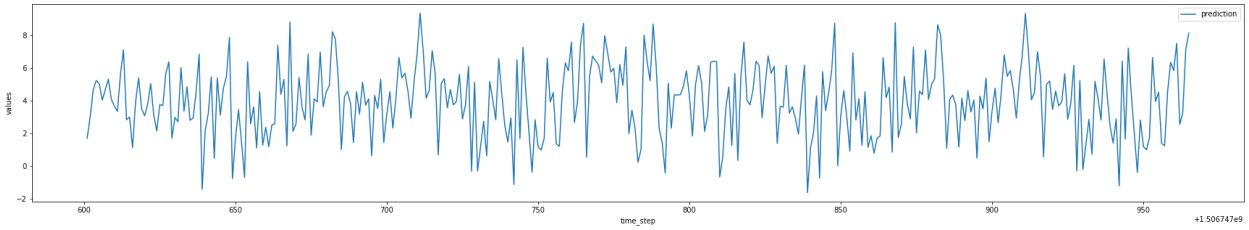

issue_number
OriginationPhase
DetectionPhase
Category
Priority
Status
closed_at
Author
created_at

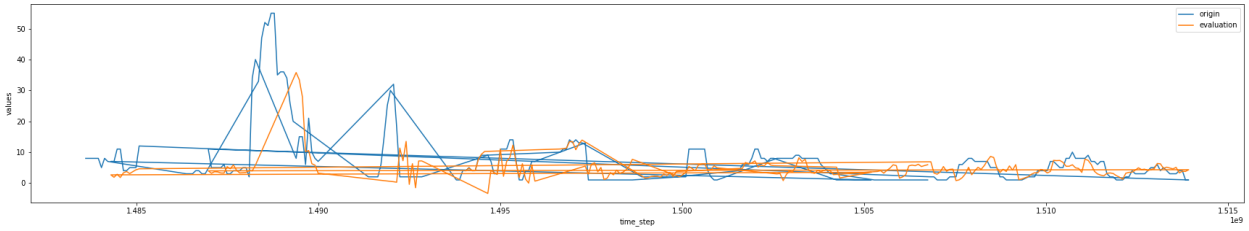

trend
10
5
0
−5
−10
2017-01
2017-04
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10
2019-01
ds

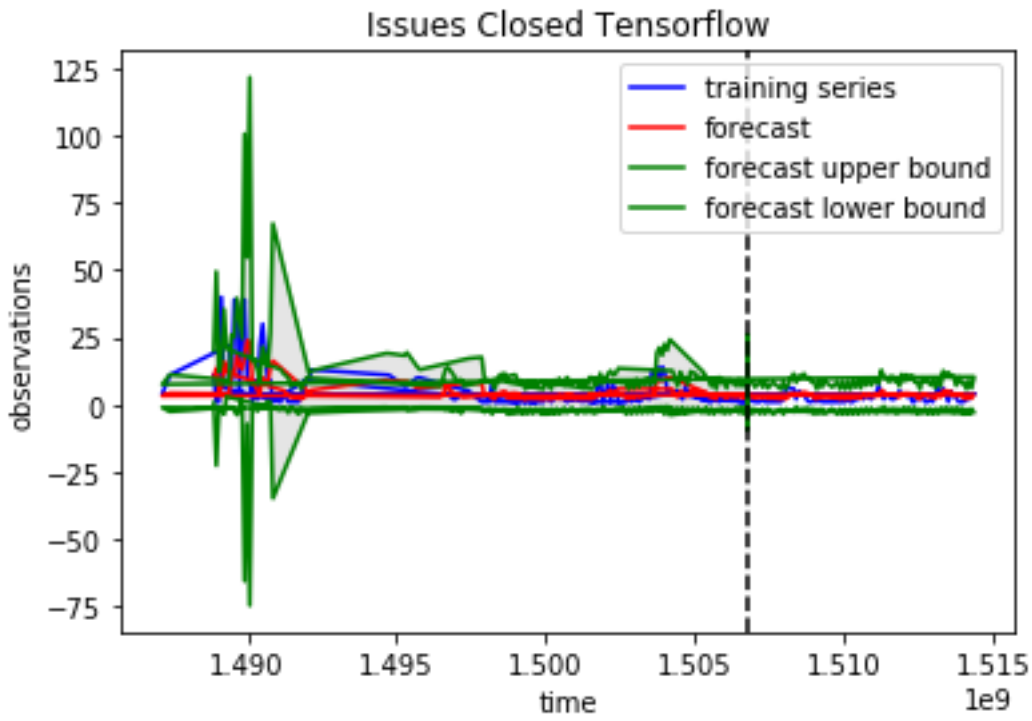

weekly
0.4
0.2
0.0
−0.2
−0.4
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week

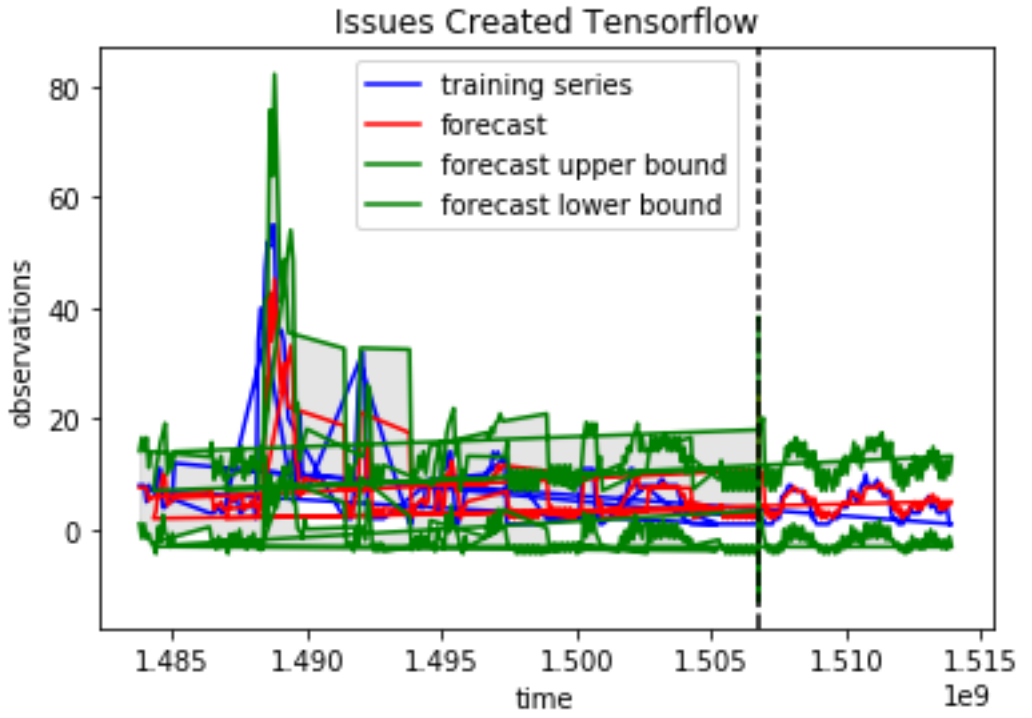

y
50
40
30
20
10
0
−10
−20
2017-01
2017-04
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10
2019-01
ds


50
40
30
20
10
0
01-05-2017
06-10-2017
10-11-2017
12-01-2017
4-17-2017
7-25-2017
created_at


prediction
values
8
6
4
2
0
−2
600
650
700
750
800
850
900
950
time_step
+1.506747e9


origin
evaluation
values
50
40
30
20
10
0
1.485
1.490
1.495
1.500
1.505
1.510
1.515
time_step
1e9


Issues Closed Tensorflow
training series
forecast
forecast upper bound
forecast lower bound
observations
125
100
75
50
25
0
−25
−50
−75
1.490
1.495
1.500
1.505
1.510
1.515
time
1e9


Issues Created Tensorflow
training series
forecast
forecast upper bound
forecast lower bound
observations
80
60
40
20
0
1.485
1.490
1.495
1.500
1.505
1.510
1.515
time
1e9