I/O Performance Analysis Framework on Measurement Data from Scientific Clusters

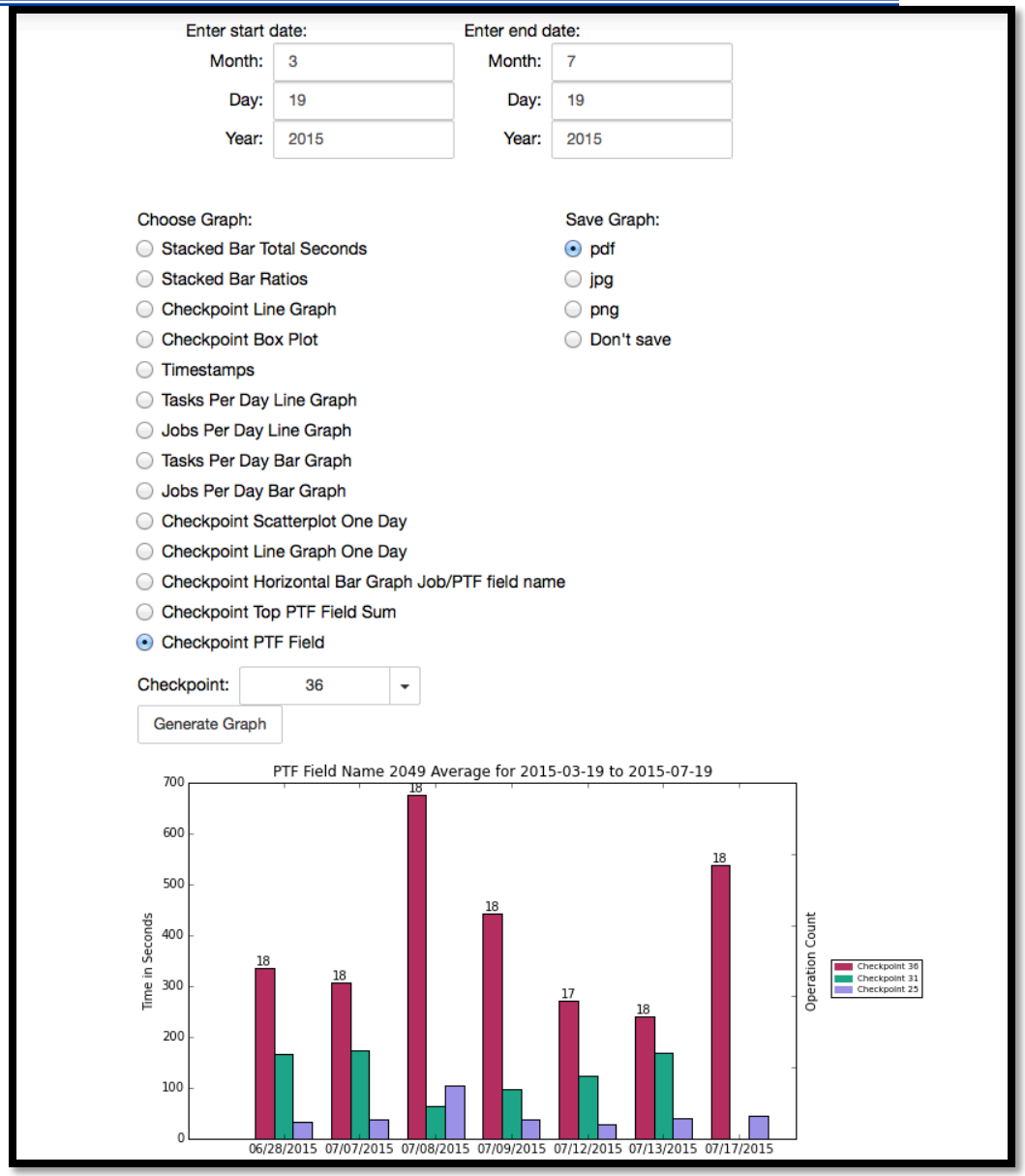
Michelle Koo

University of California, Berkeley
&
Scientific Data Management Research Group
Computational Research Division
Lawrence Berkeley National Laboratory

SC'15 ACM SRC

I/O Analysis Framework Tool

- Helps analyze I/O performance characteristics
- Can be applied to many different scientific applications
- Allows users to interactively visualize and understand data

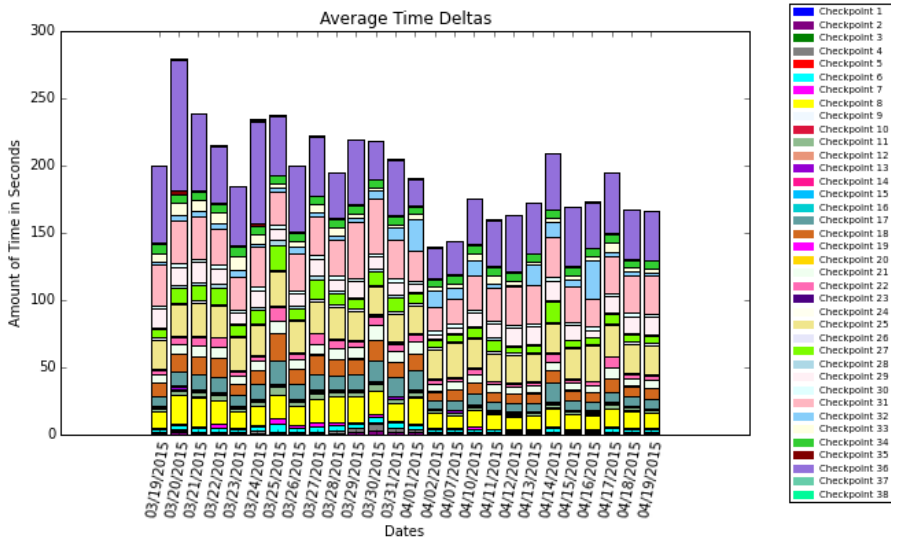


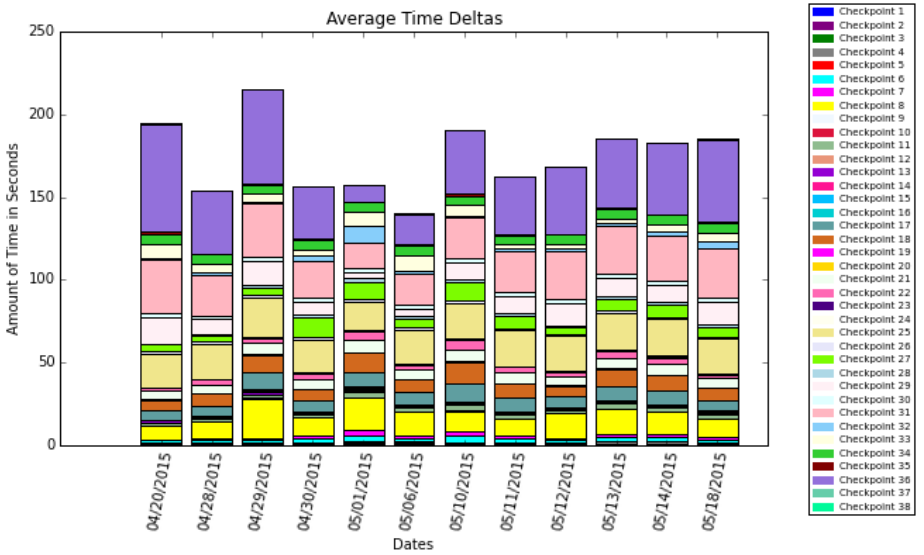
APPLICATION:
PALOMAR TRANSIENT FACTORY

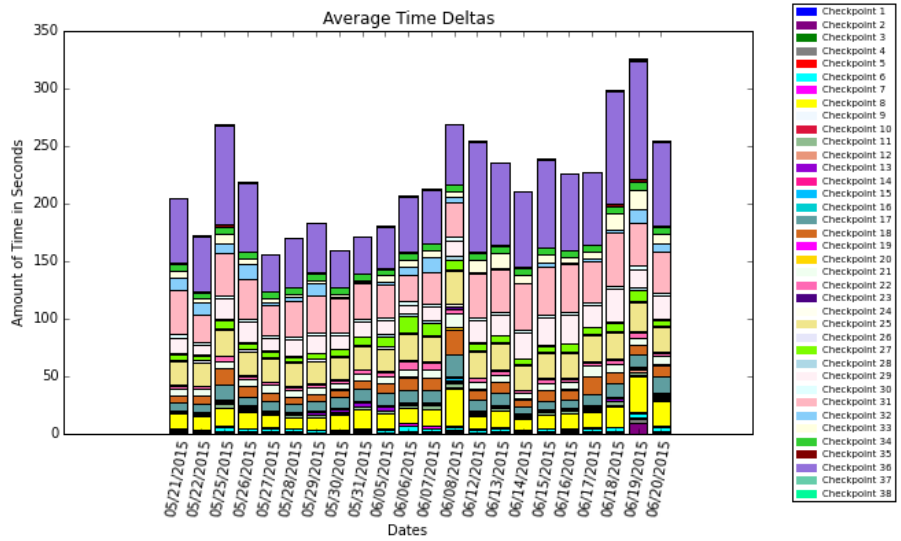
Palomar Transient Factory (PTF) Background

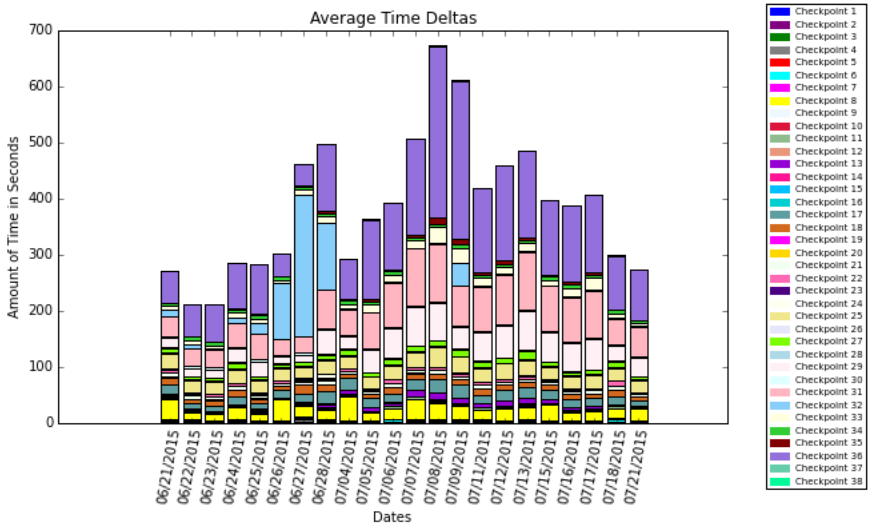
- Data from cameras analyzing the sky are processed at NERSC and LBNL through PTF real-time image subtraction pipeline
- PTF data analysis pipeline consists of 38 checkpoints
- 1.6 TB logs processed from database of timestamps recording how long execution of PTF pipeline and each checkpoint takes
- Apache Spark was used in the backend to distribute computations to multiple nodes and was executed in parallel
- Subsequent graphs show results on PTF data analysis pipeline discovered from framework tool

Each Operation Total Daily Average Time Execution

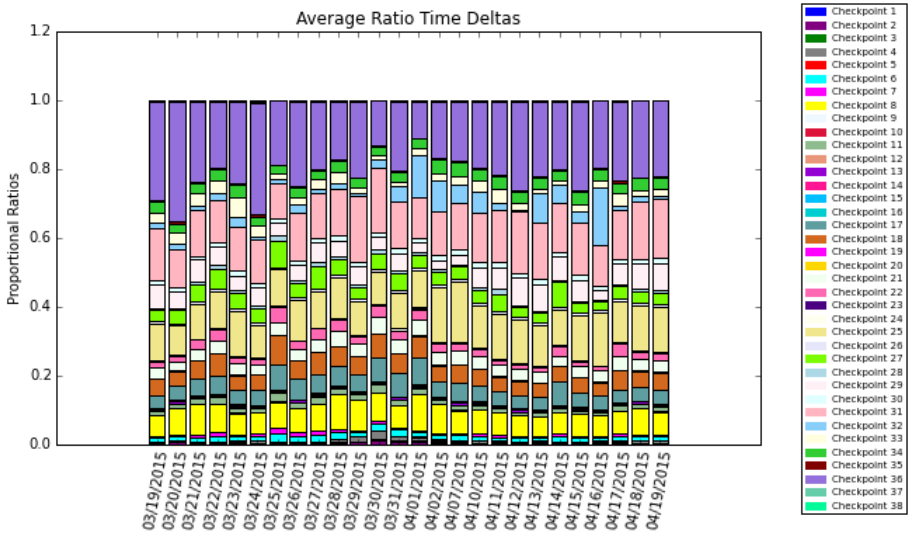


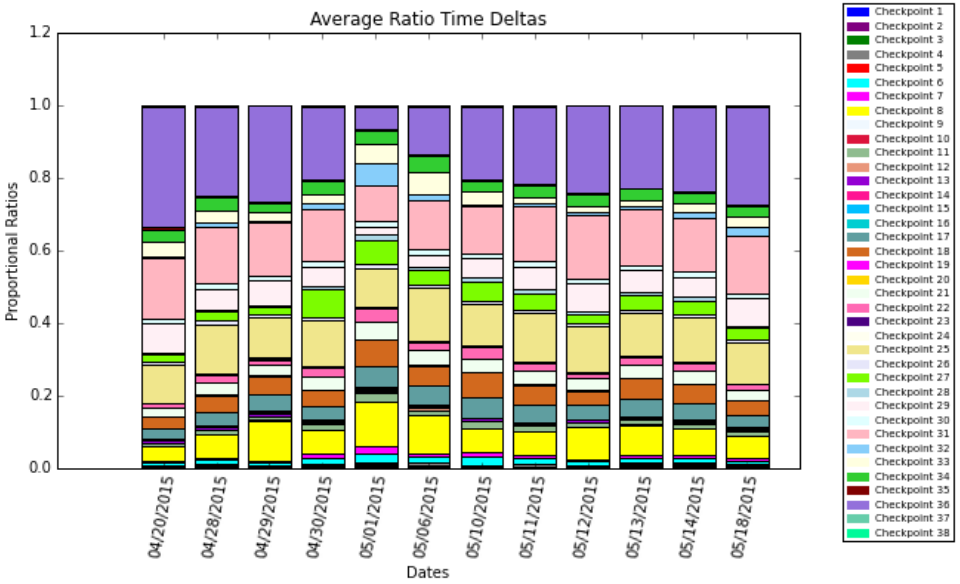


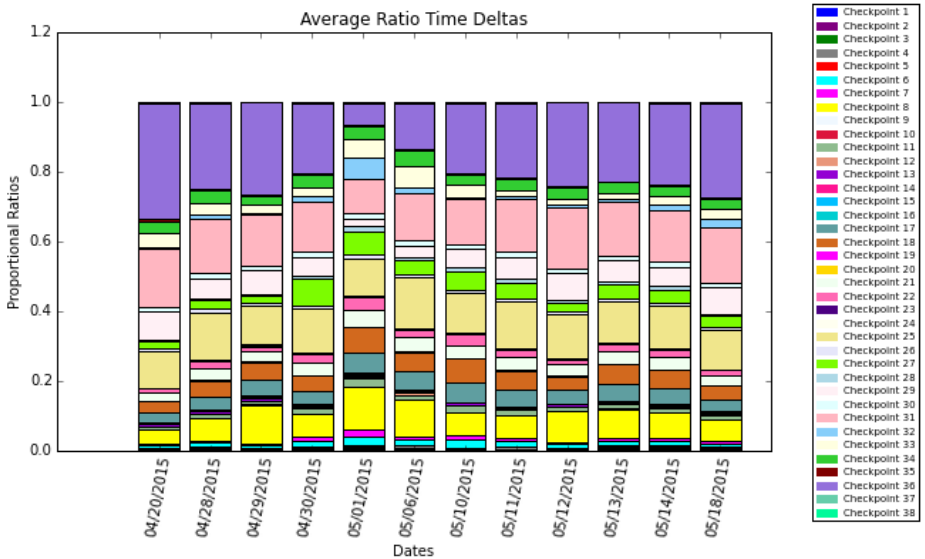


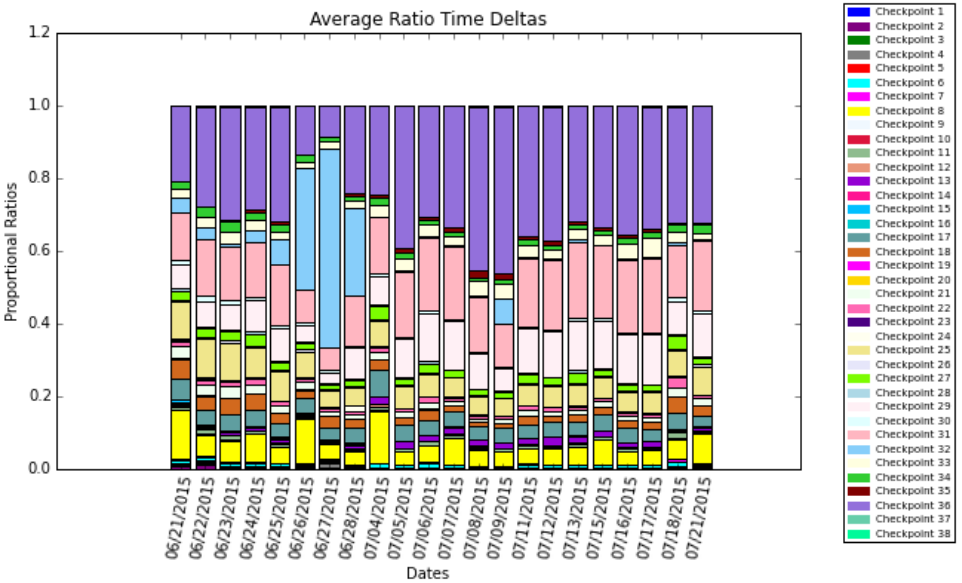


Each Operation
Daily Proportional Ratios









Top 5 Checkpoints

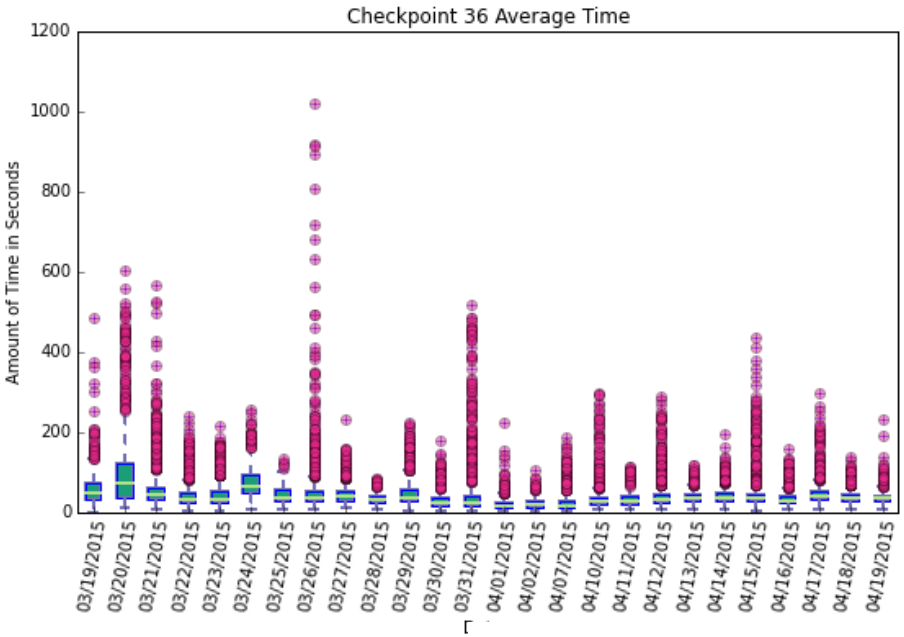
- Top checkpoints with the longest execution times
- Data taken over a span of 64 days

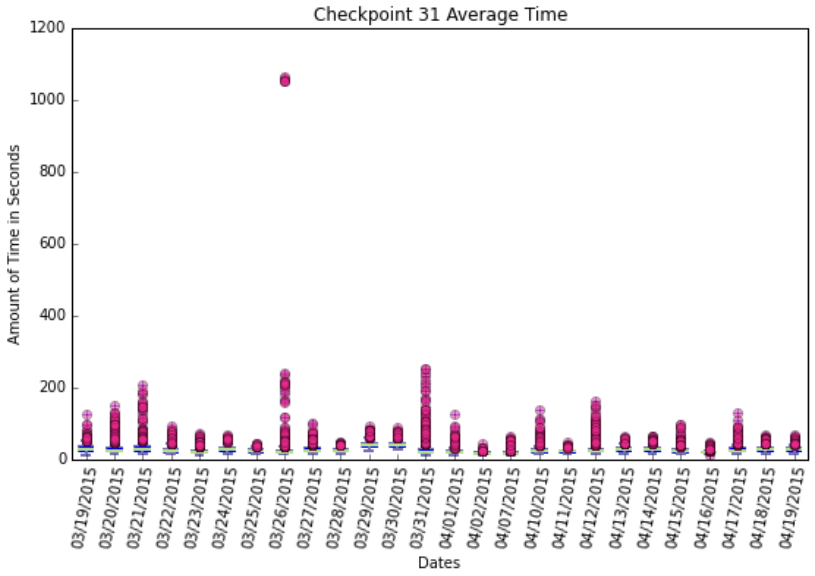
Checkpoint	Average Percentage of Total PTF Pipeline Time Execution
Checkpoint 36	23.72%
Checkpoint 8	7.29%
Checkpoint 25	11.16%
Checkpoint 29	6.22%
Checkpoint 31	14.79%

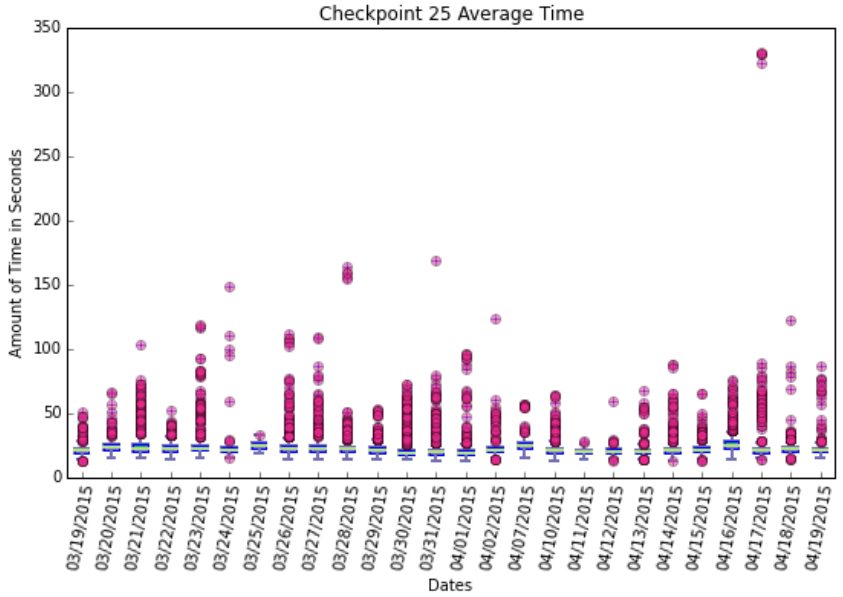
TOP 3 CHECKPOINTS WITH LONGEST TIME EXECUTION: 36, 31 AND 25

Checkpoints 36, 31, and 25 Daily Timestamp Data

DATES: 3/19/2015 – 4/19/2015

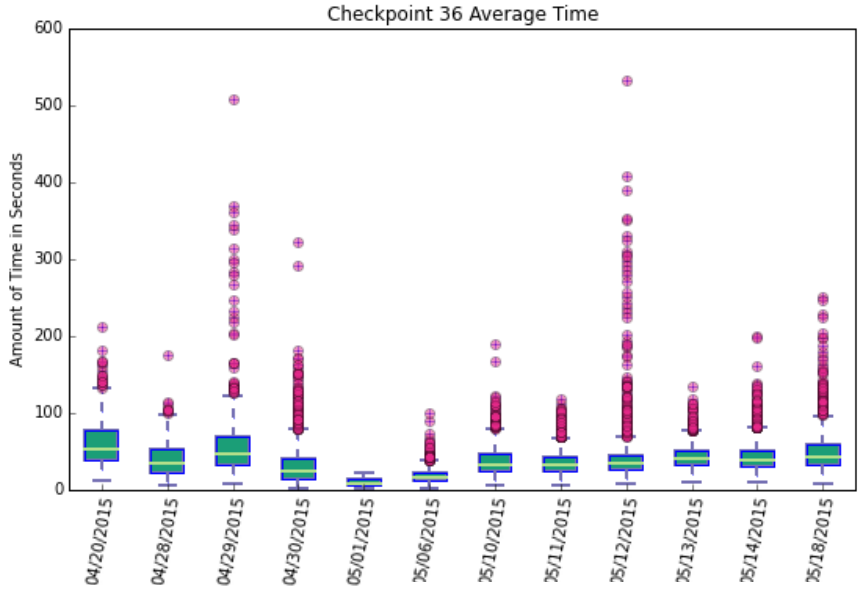


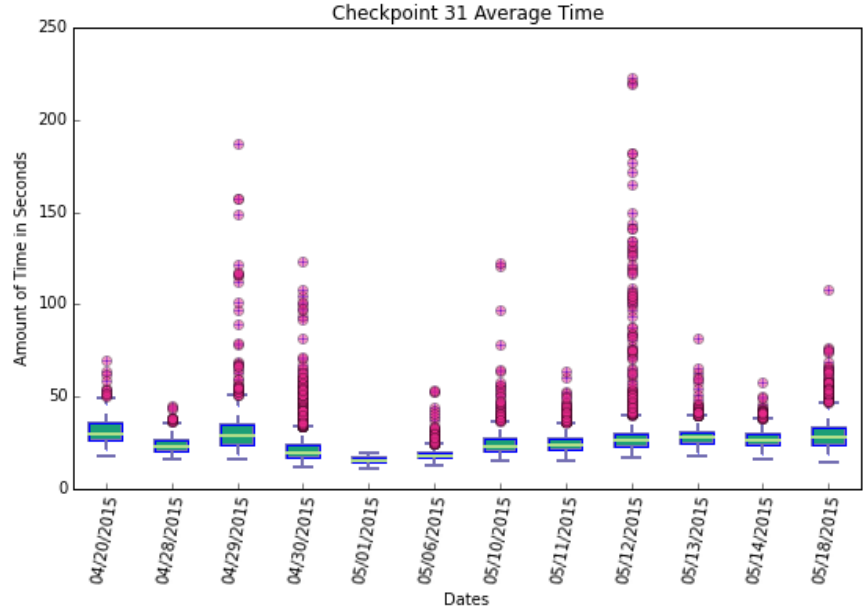


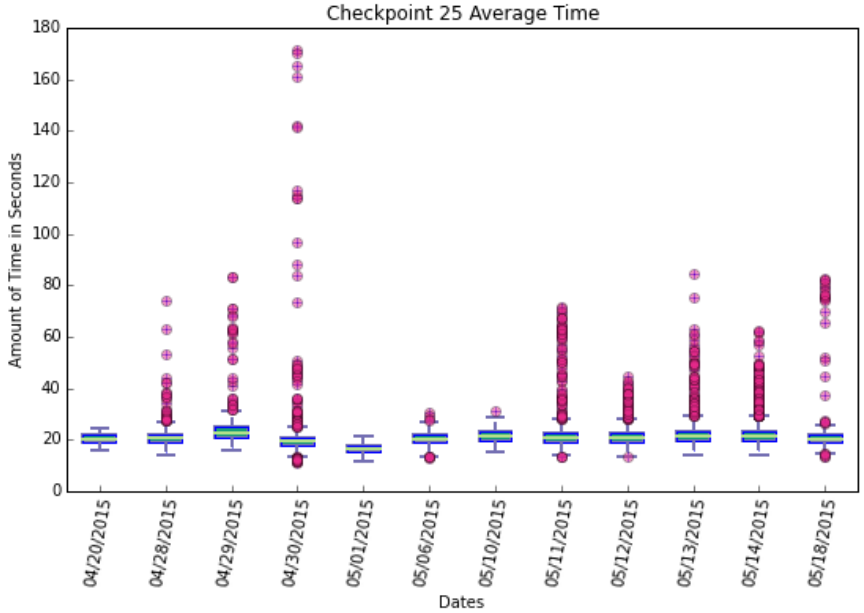


Checkpoints 36, 31, and 25 Daily Timestamp Data

DATES: 4/20/2015 – 5/20/2015

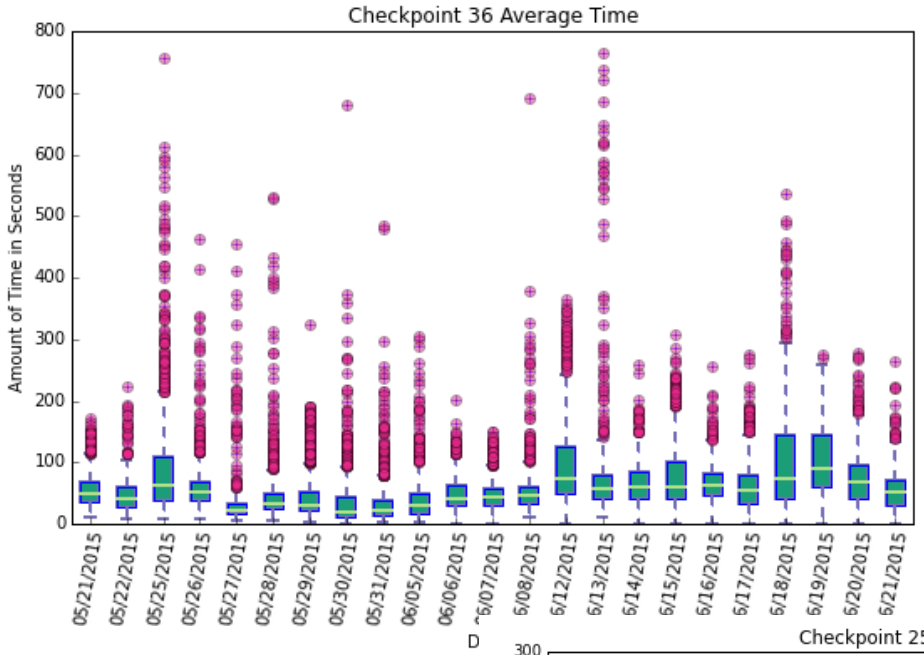


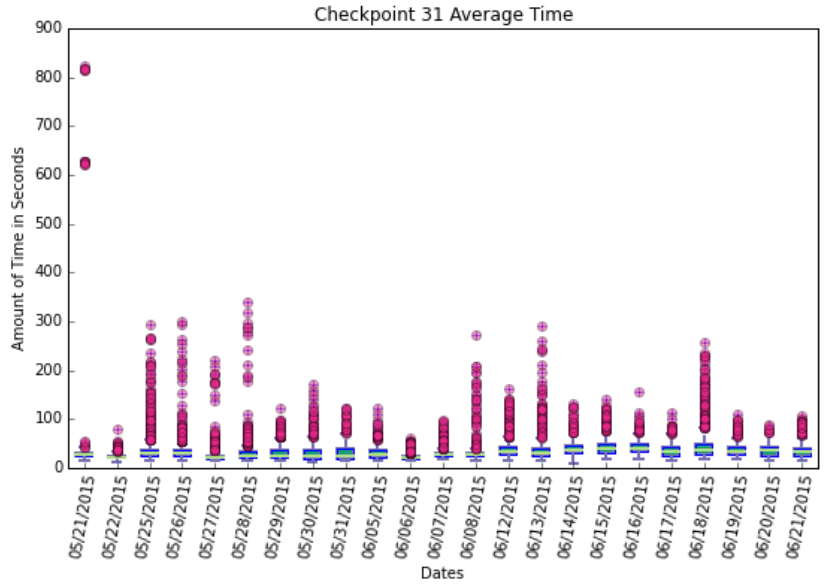


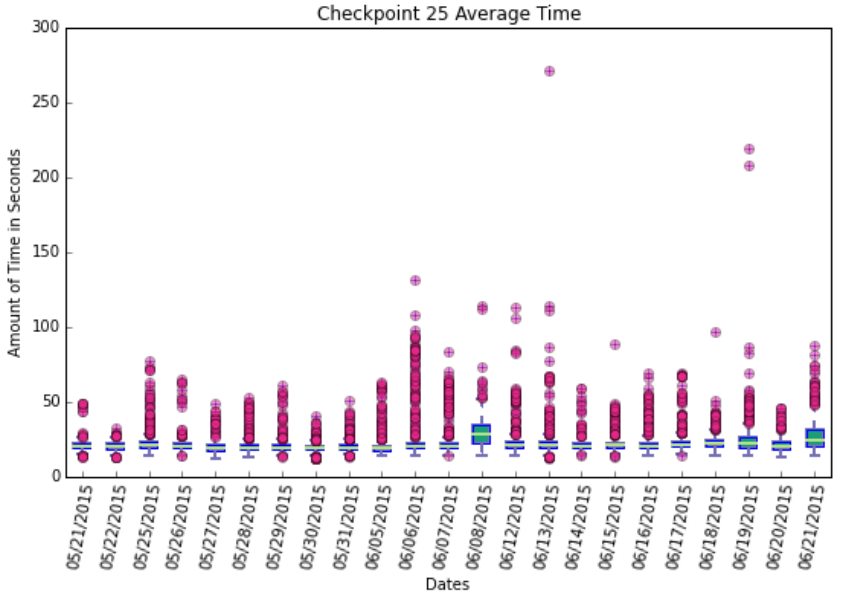


Checkpoints 36, 31, and 25 Daily Timestamp Data

DATES: 5/21/2015 – 6/21/2015

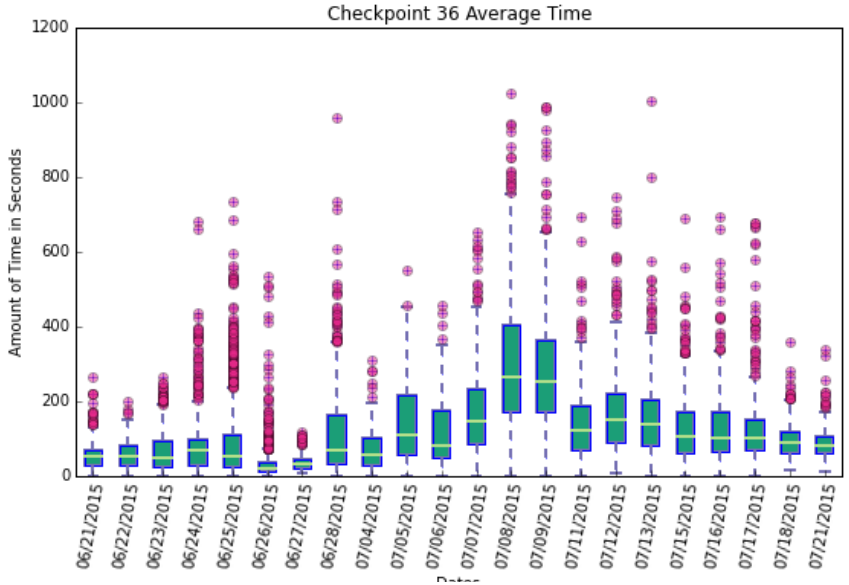


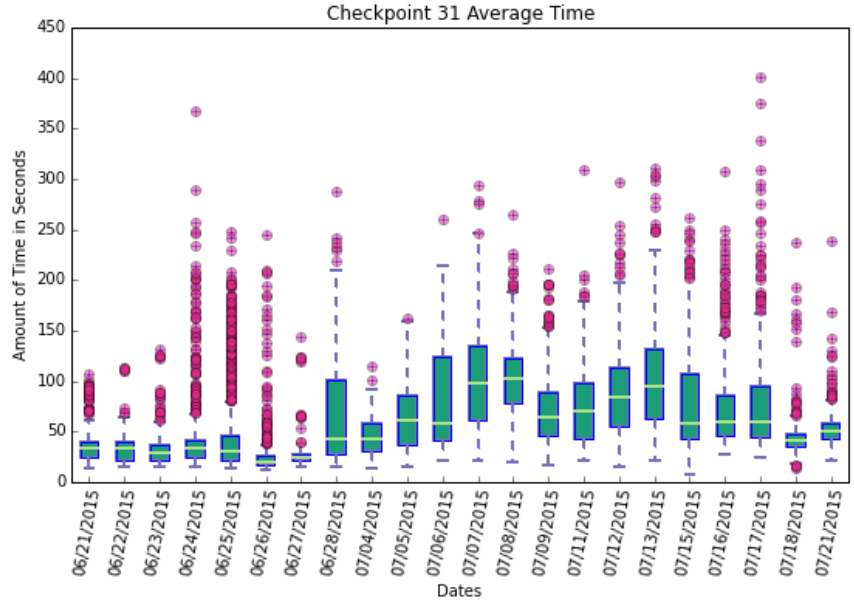


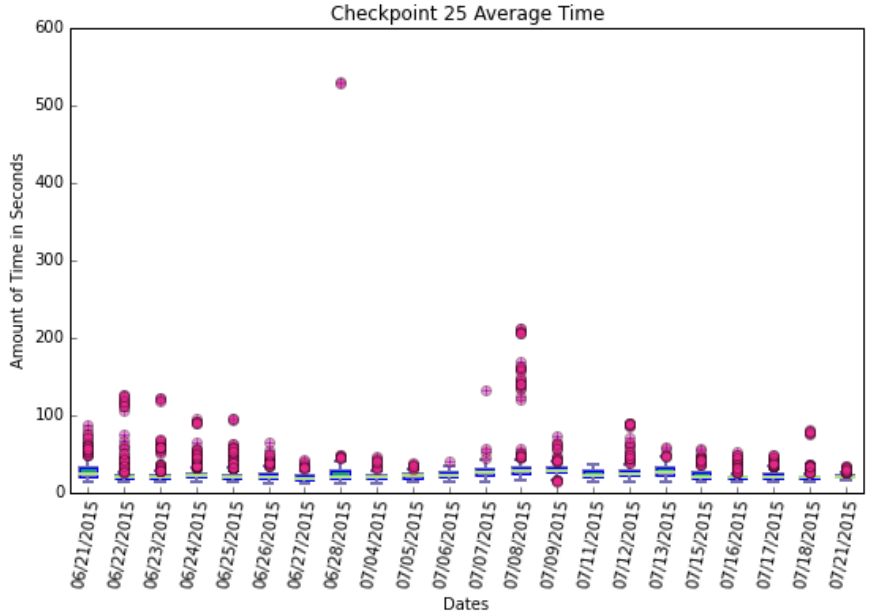


Checkpoints 36, 31, and 25 Daily Timestamp Data

DATES: 6/21/2015 – 7/21/2015



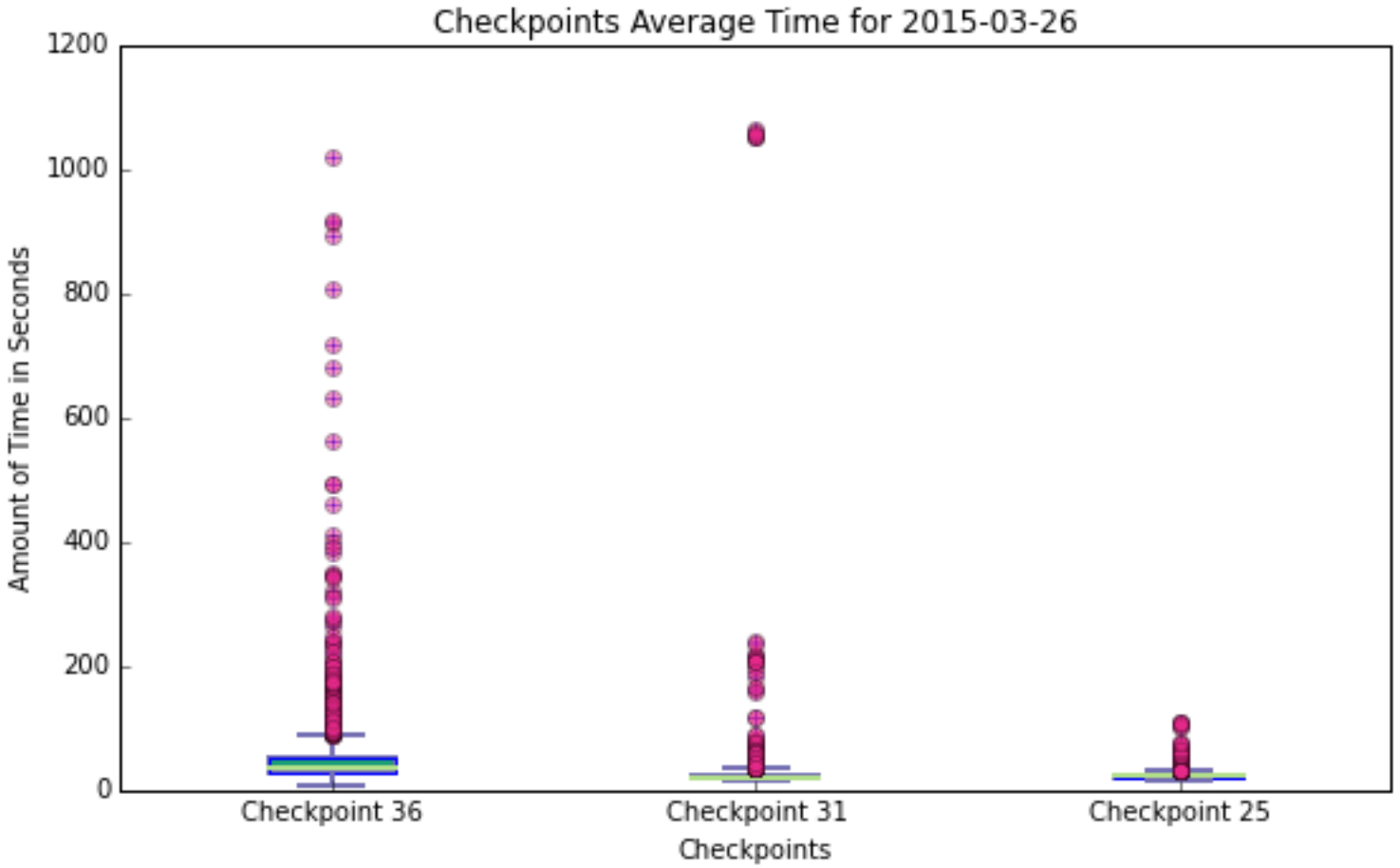




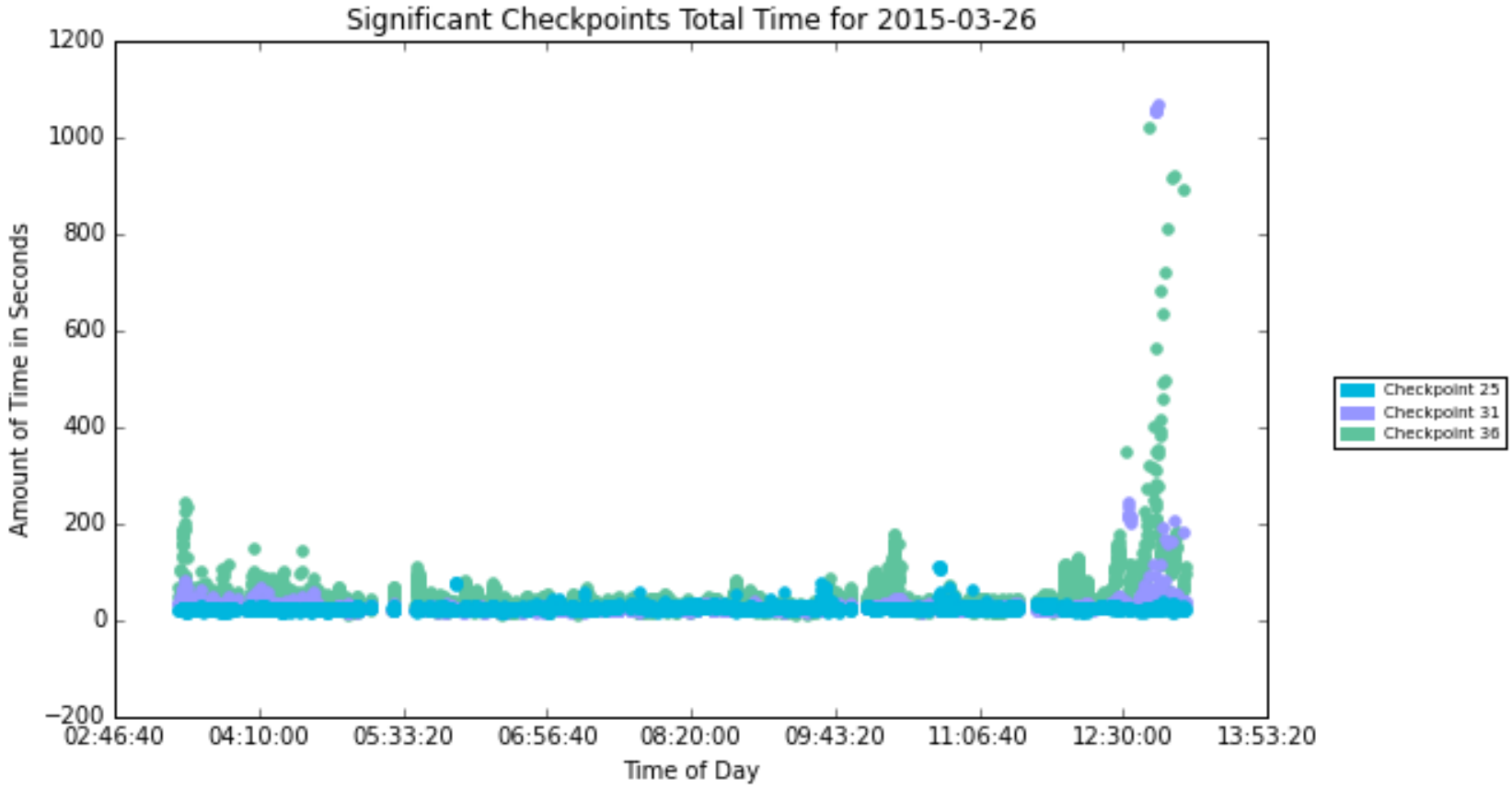
DATES ANALYZED:
3/26/2015
5/25/2015
6/13/2015
6/25/2015

These dates contained a vast variety of outliers in previous box plots

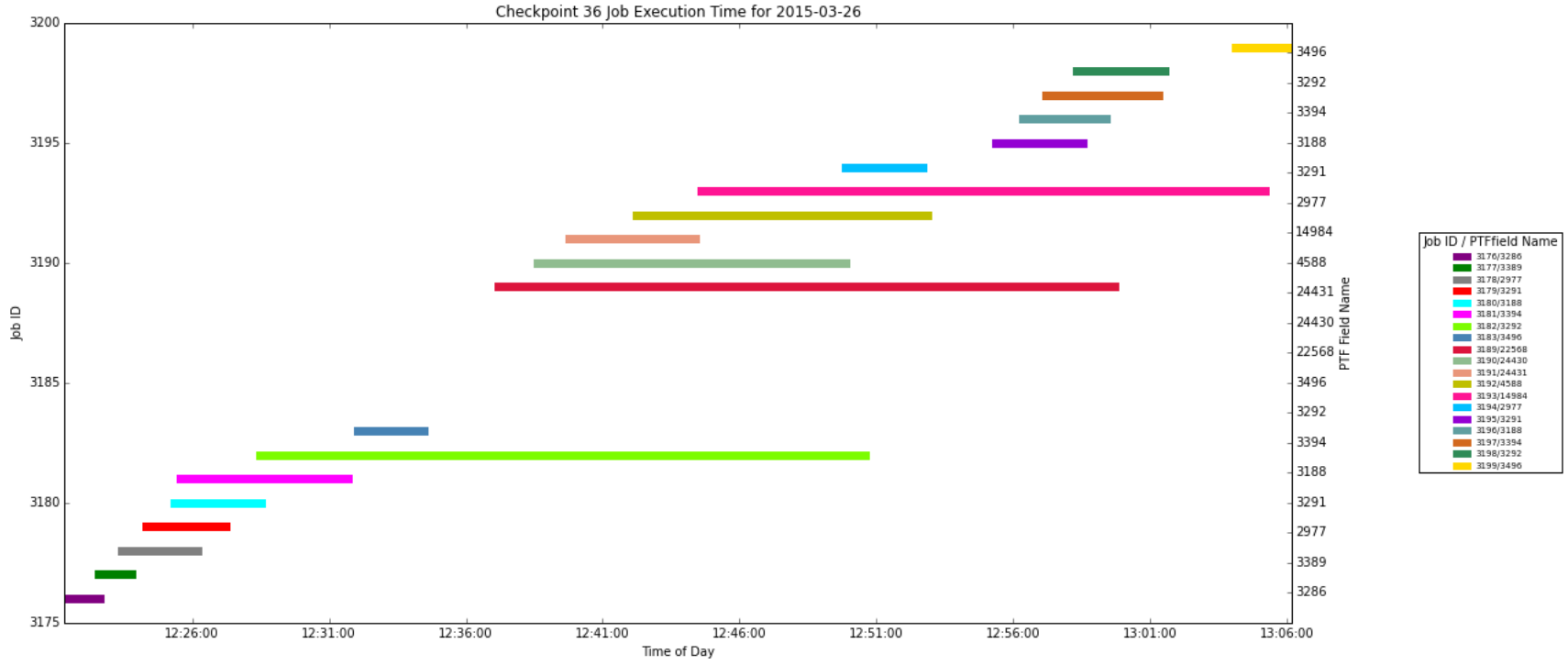
Checkpoints 36, 31, and 25 Average Time on 3/26/2015



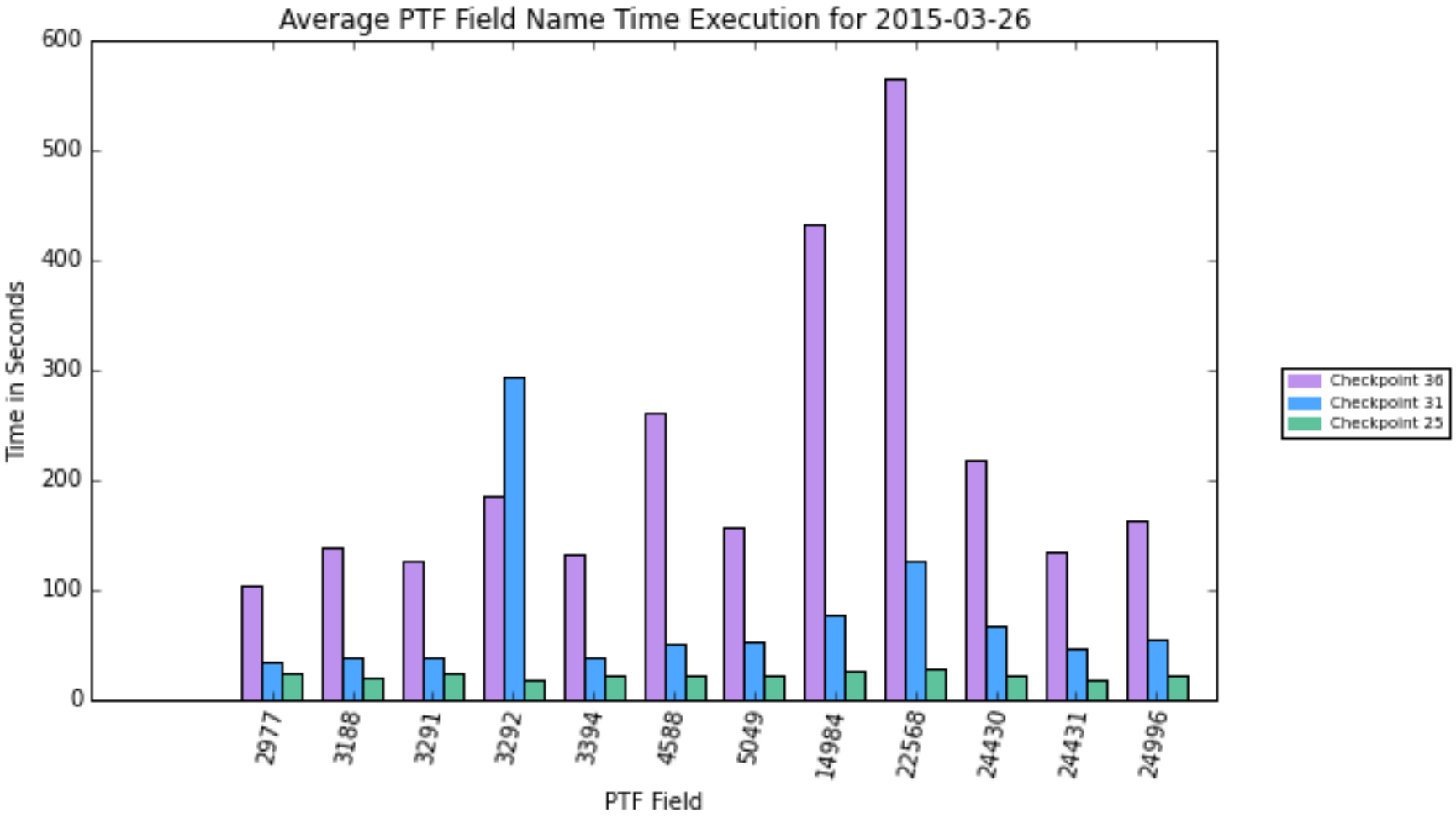
Checkpoints 36, 31, and 25 Time Execution on 3/26/2015



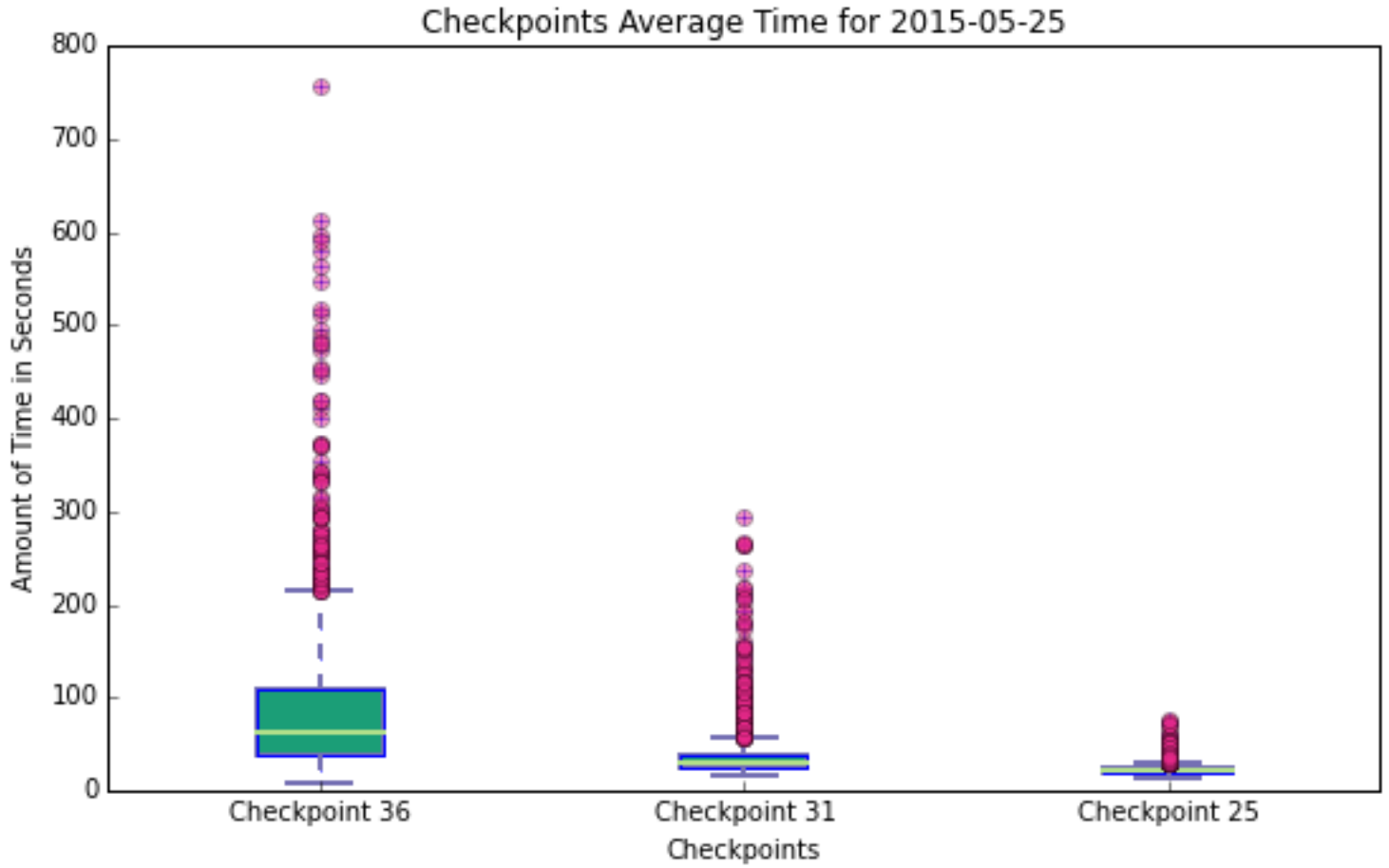
Checkpoint 36 Job ID & PTF Field Names on 3/26/2015



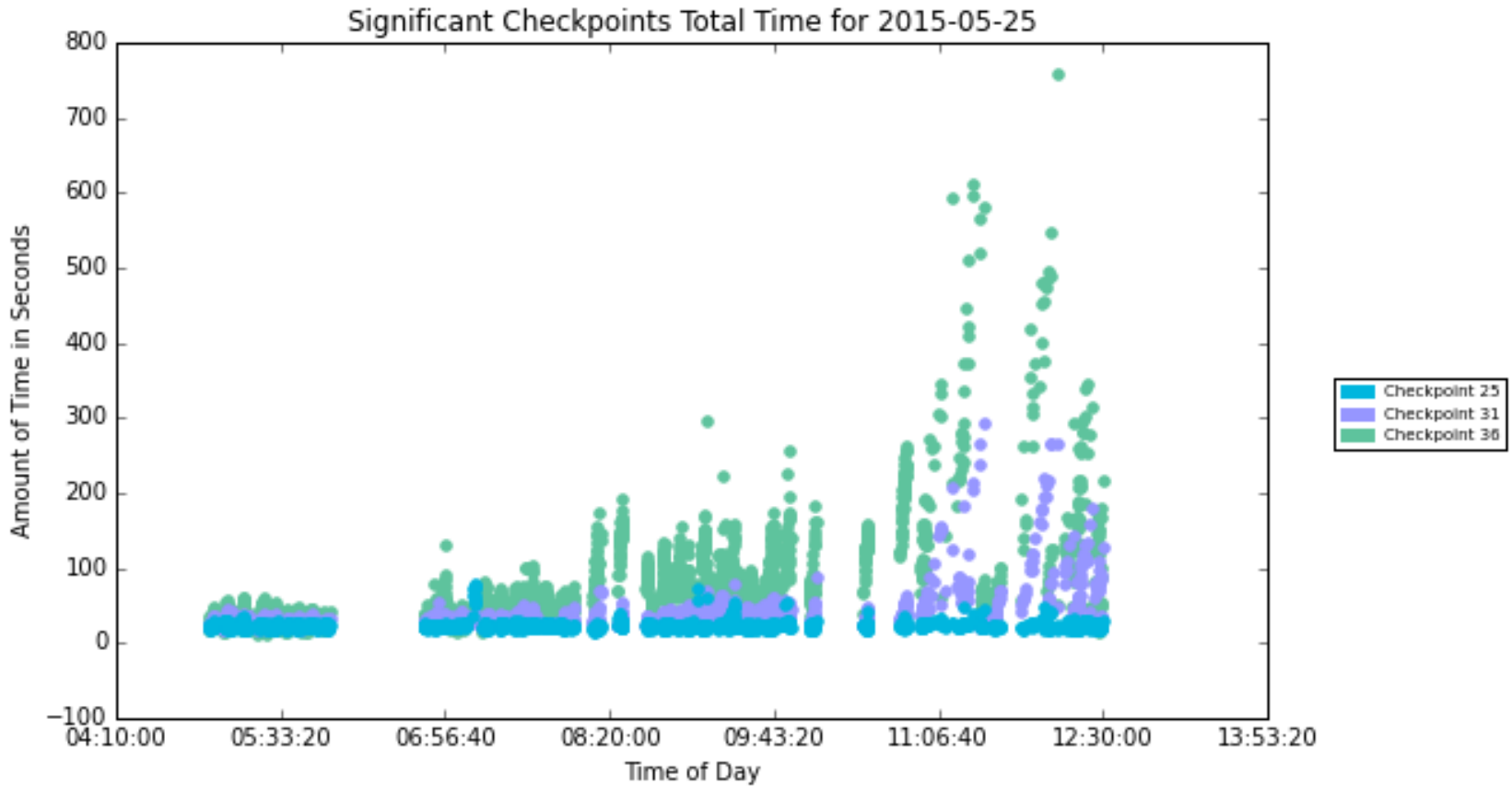
Checkpoints 36, 31, and 25 PTF Fields Average Time on 3/26/2015



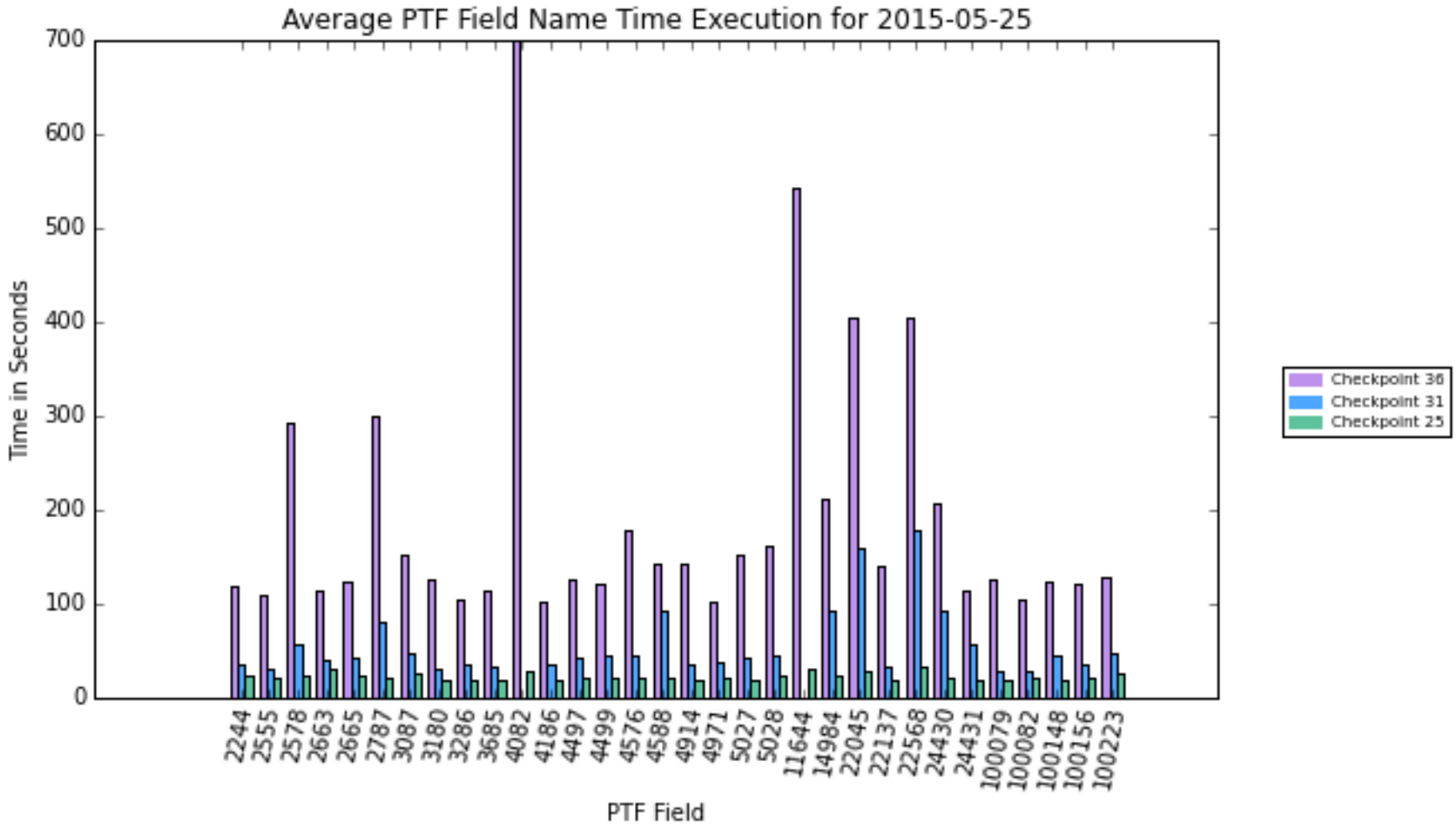
Checkpoints 36, 31, and 25 Average Time on 5/25/2015



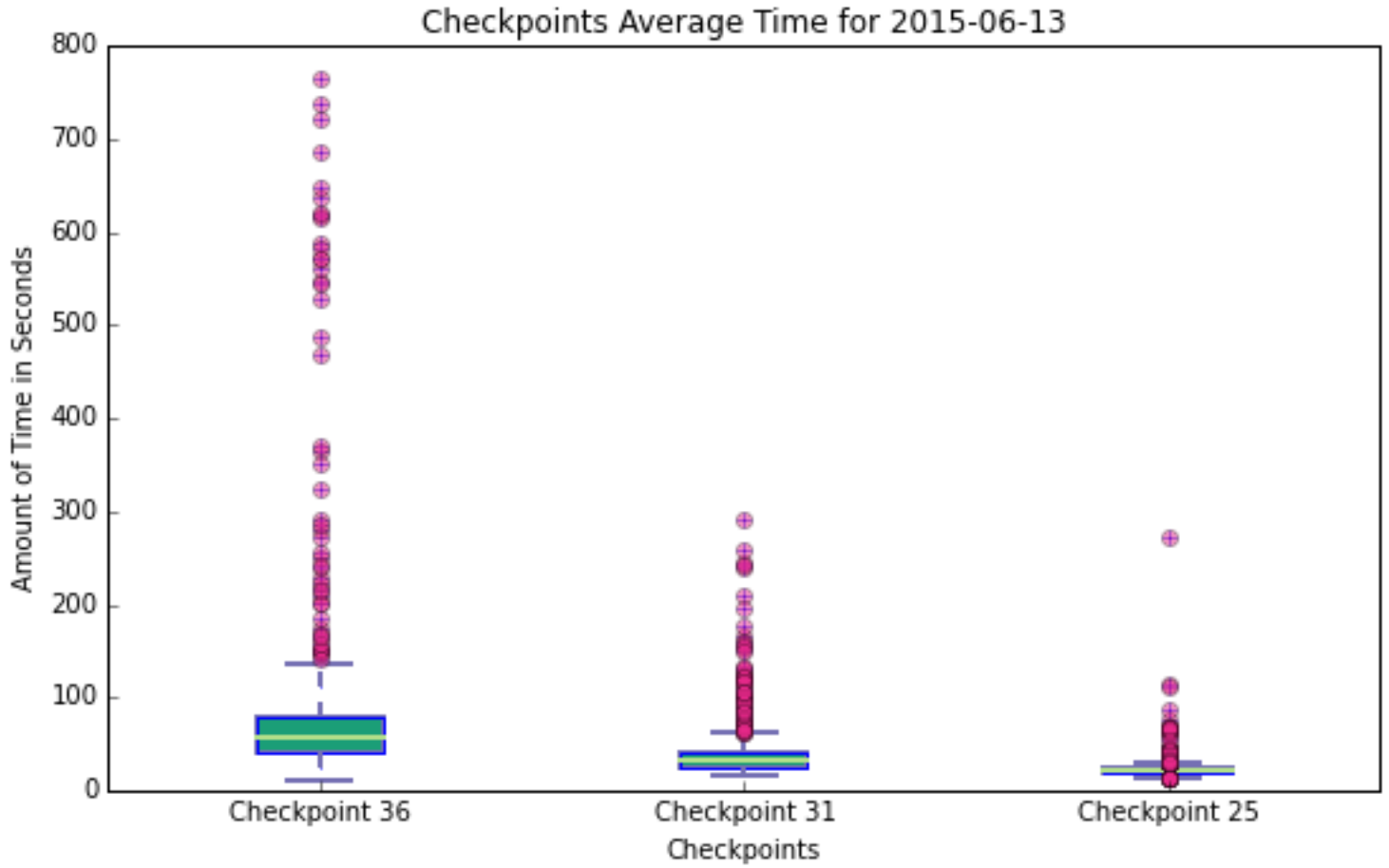
Checkpoints 36, 31, and 25 Time Execution on 5/25/2015



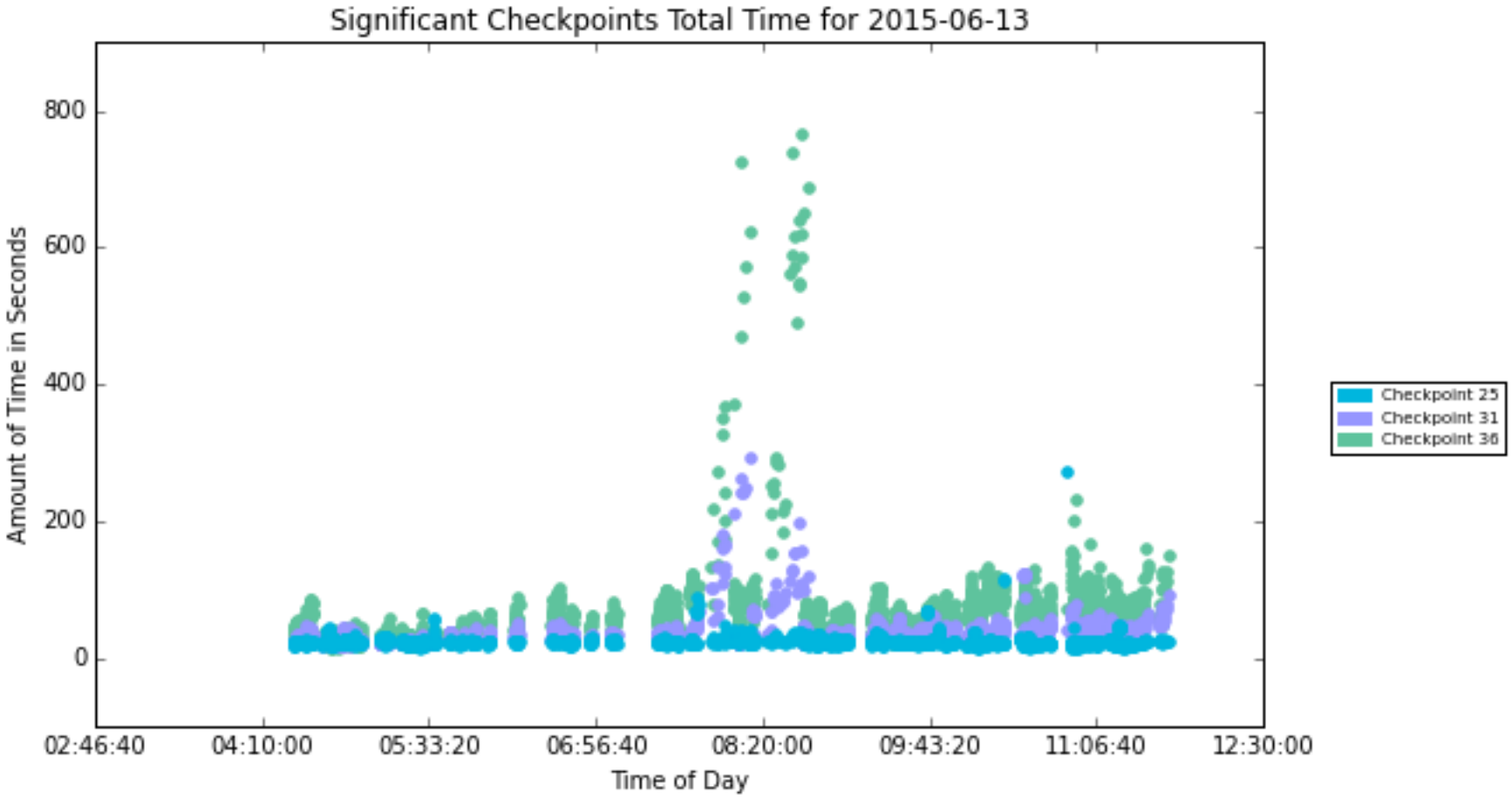
Checkpoints 36, 31, and 25 PTF Fields Average Time on 5/25/2015



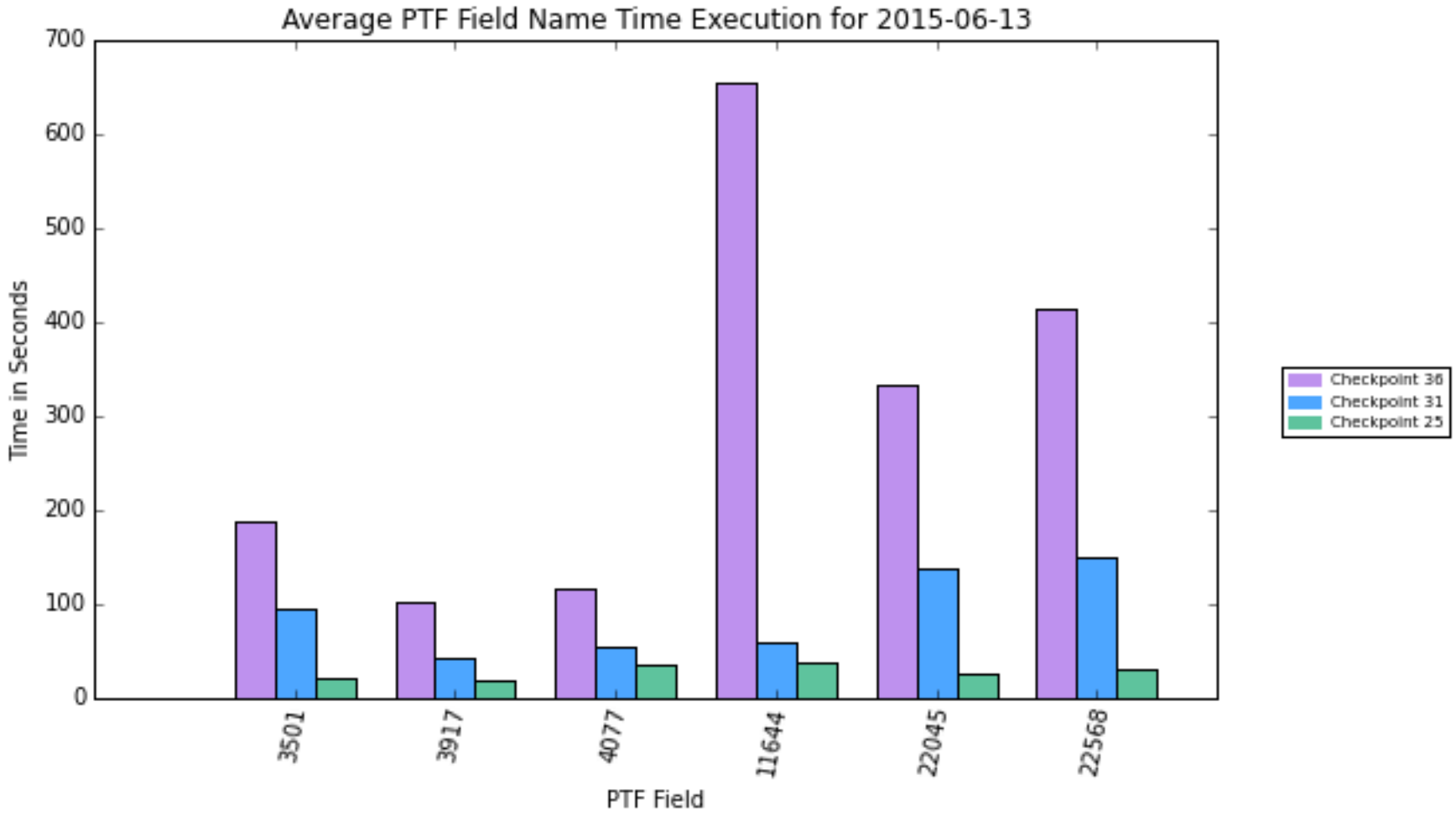
Checkpoints 36, 31, and 25 Average Time on 6/13/2015



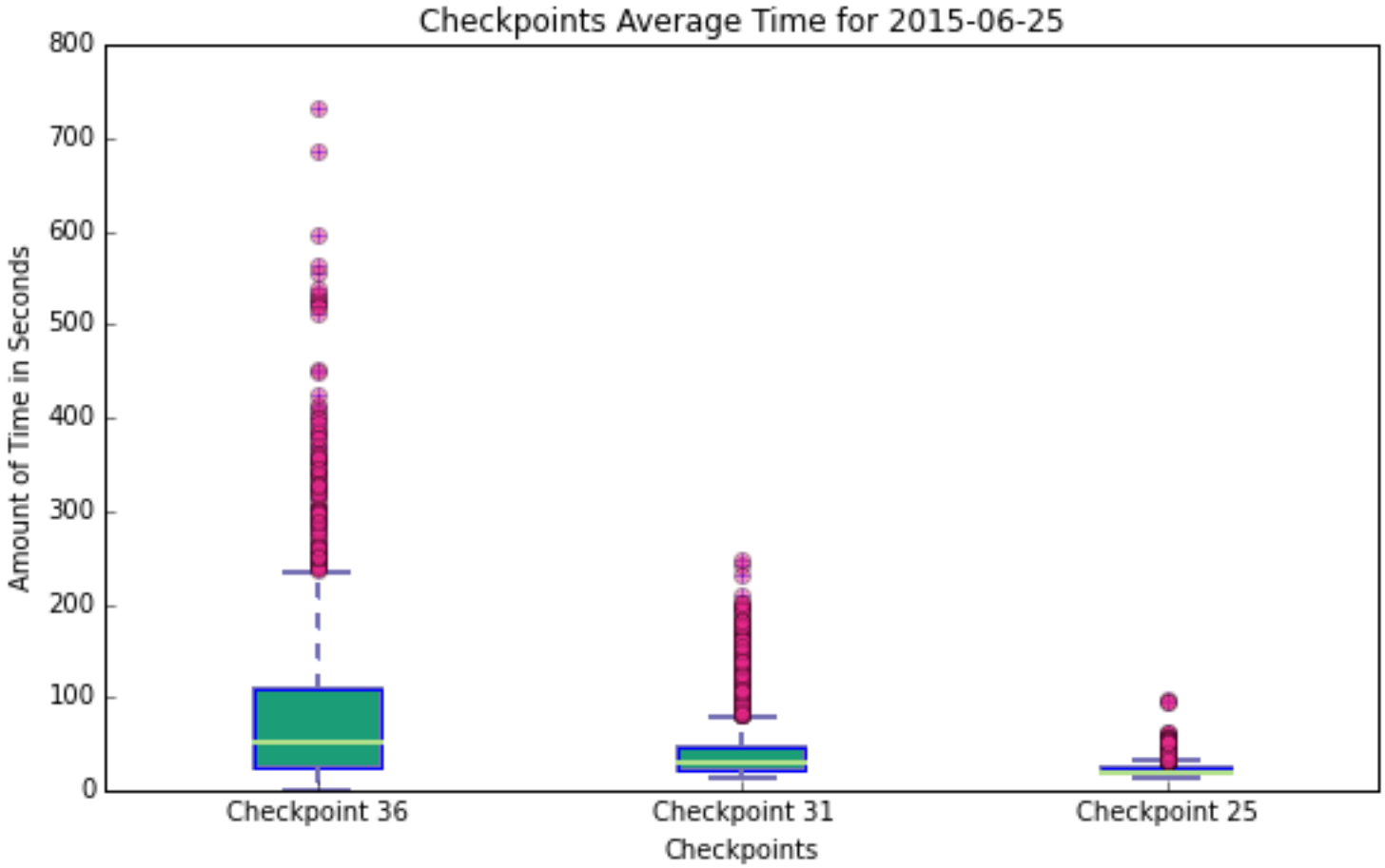
Checkpoints 36, 31, and 25 Time Execution on 6/13/2015



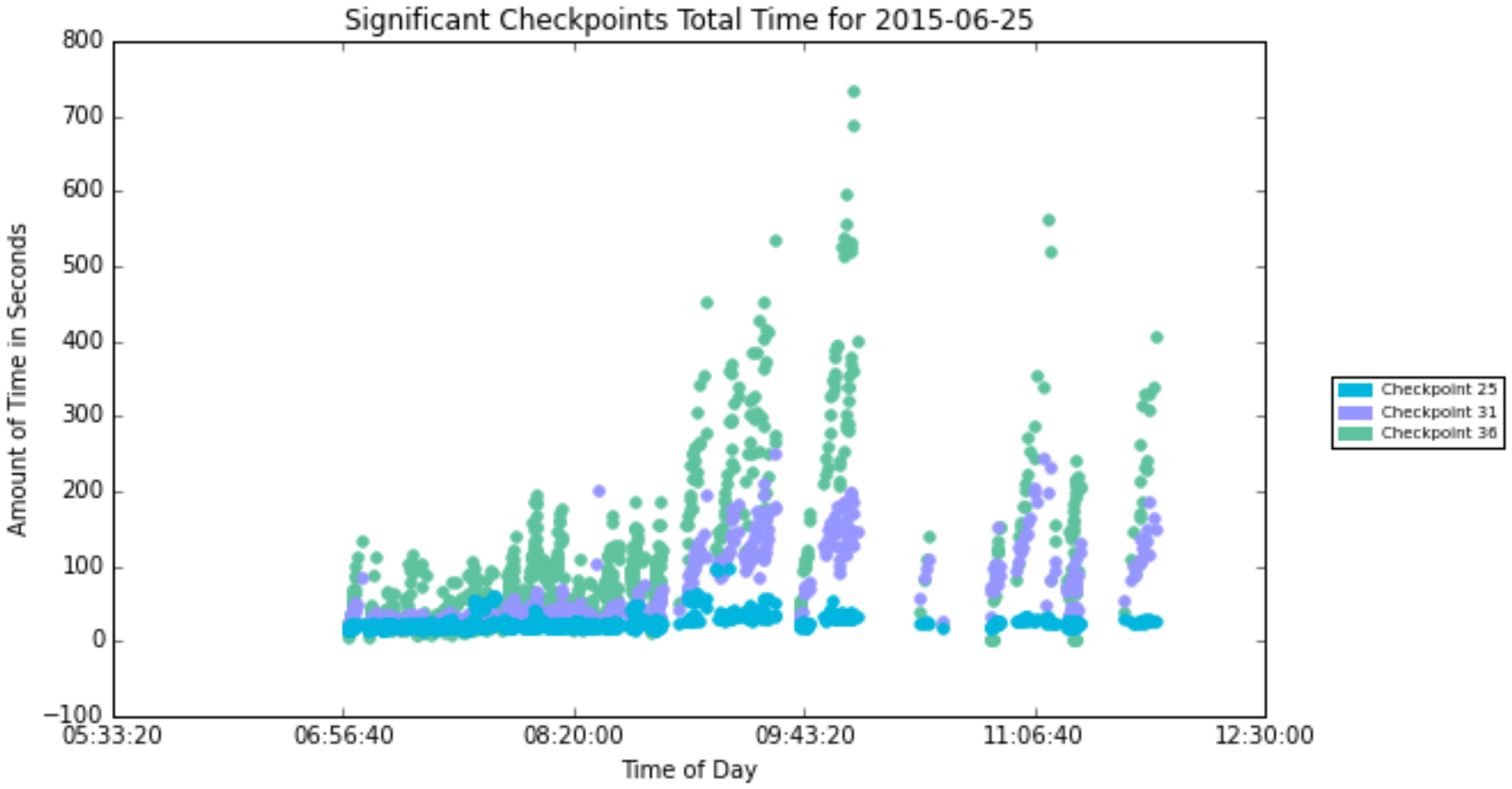
Checkpoints 36, 31, and 25 PTF Fields Average Time on 6/13/2015



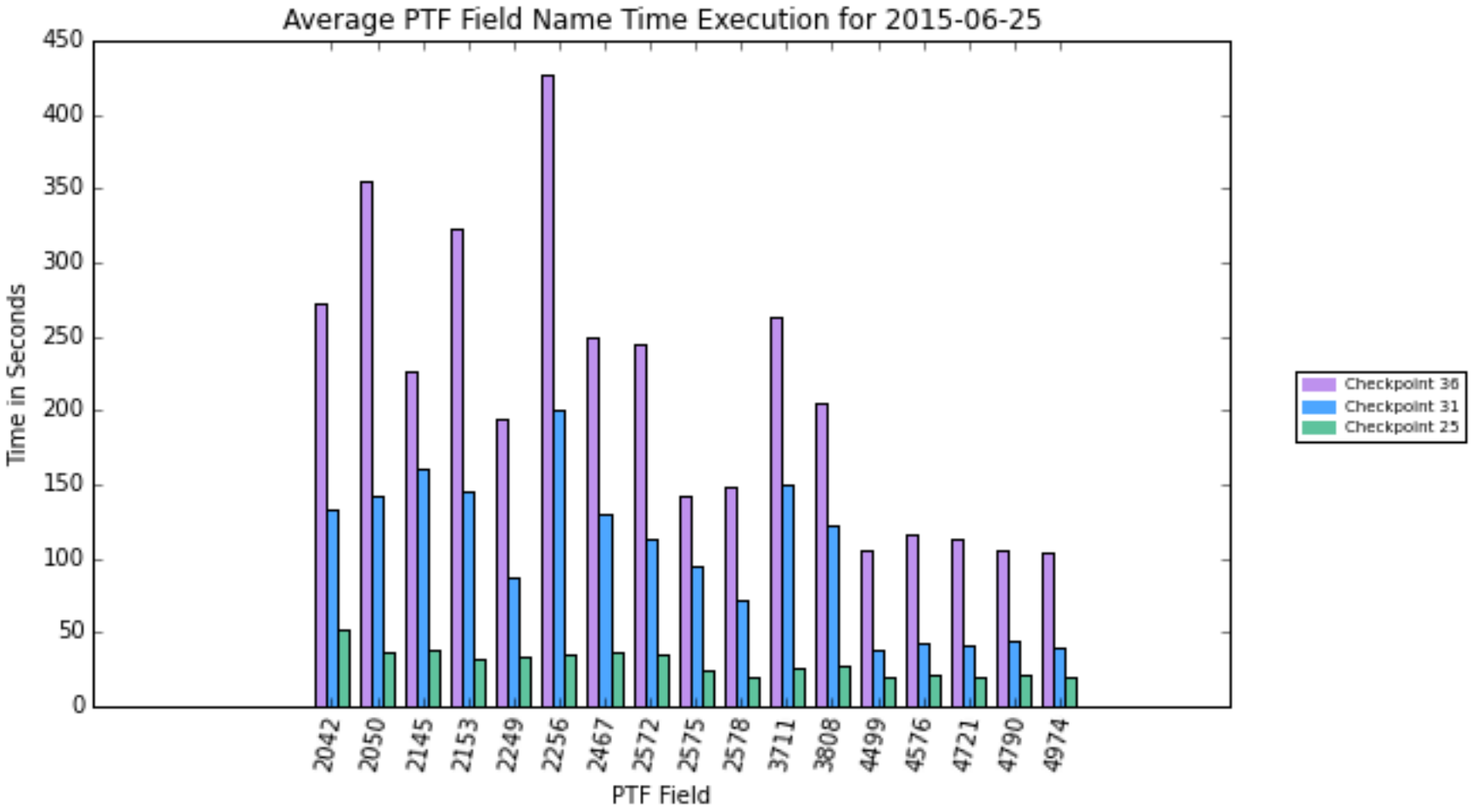
Checkpoints 36, 31, and 25 Average Time on 6/25/2015



Checkpoints 36, 31, and 25 Time Execution on 6/25/2015

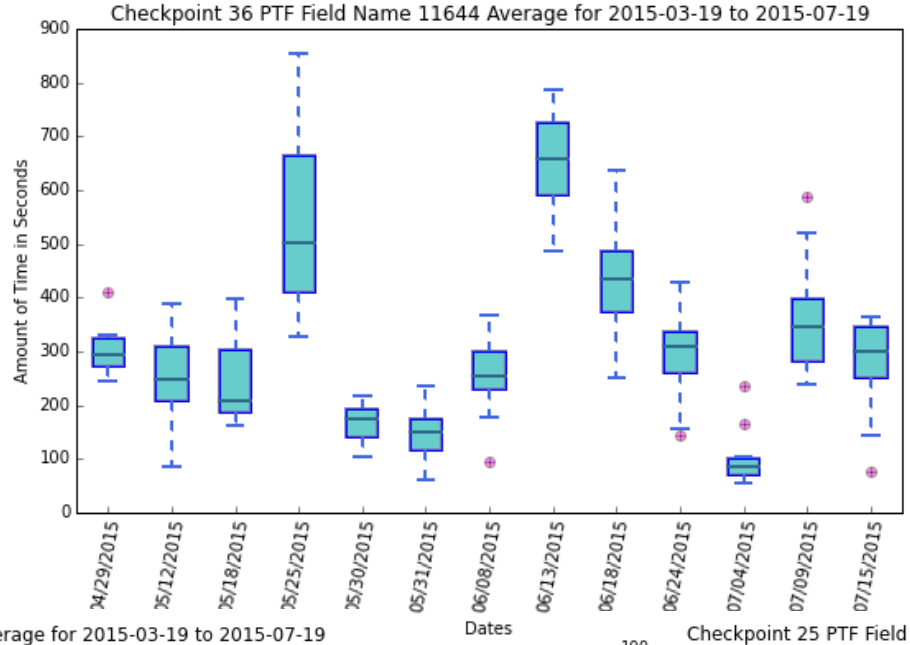


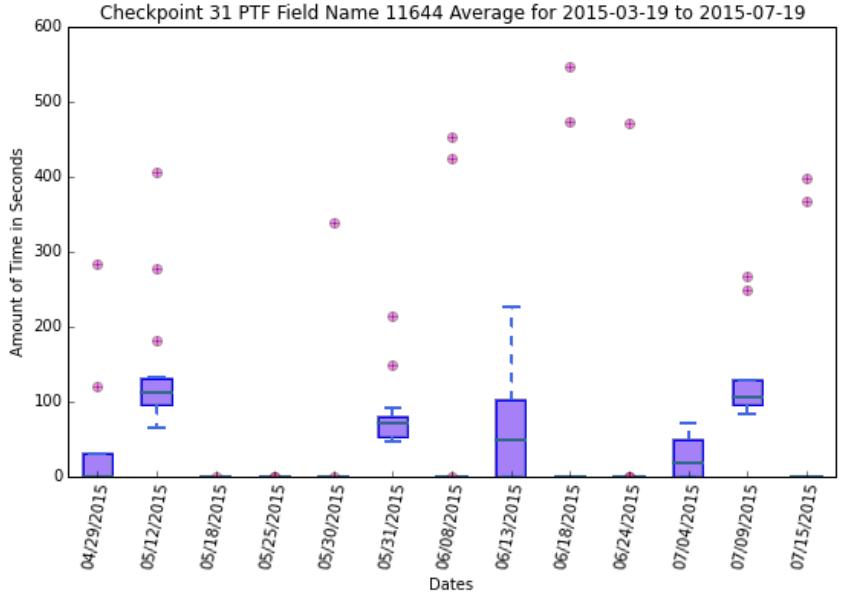
Checkpoints 36, 31, and 25 PTF Fields Average Time on 6/25/2015

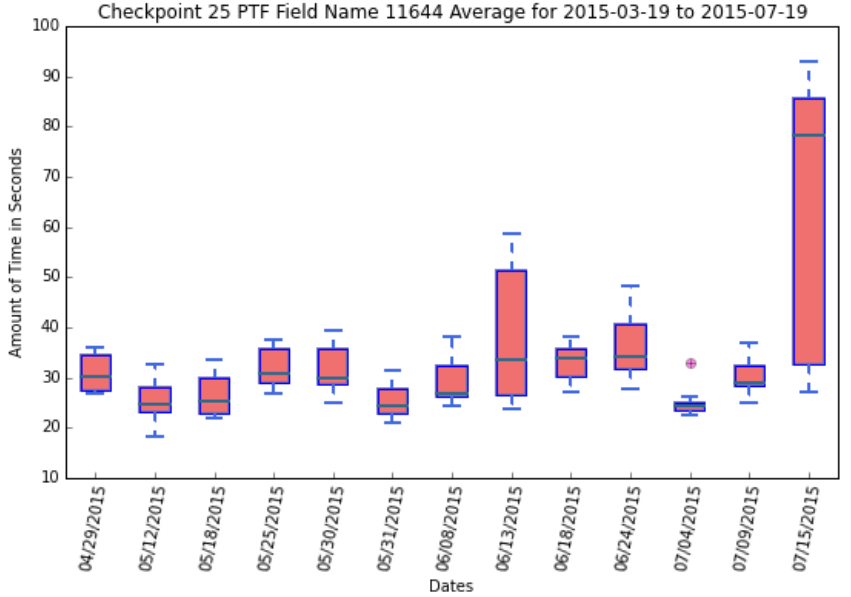


ANALYSIS WITH COMMON PTF FIELD NAMES BETWEEN CHECKPOINTS 36, 31, AND 25

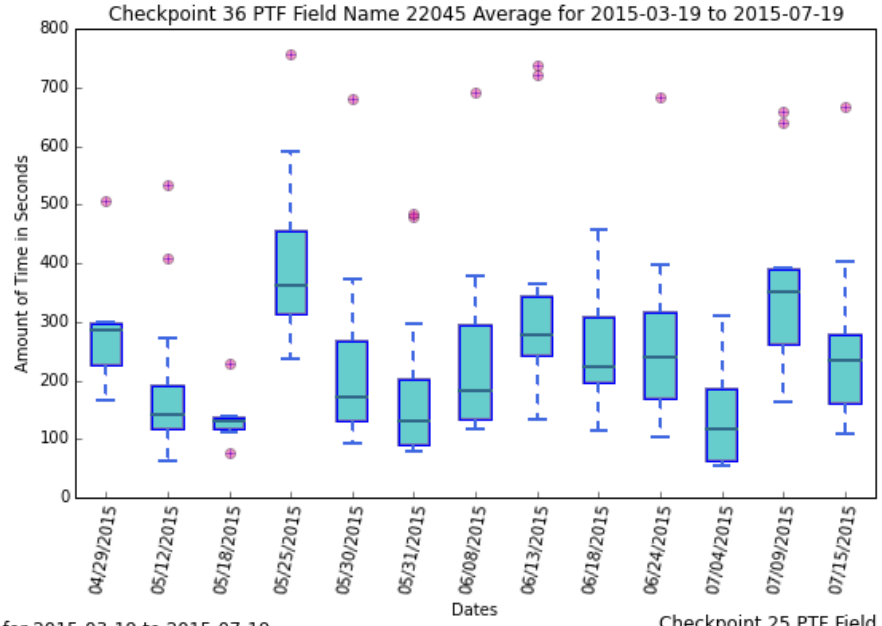
Checkpoints 36, 31, and 25 PTF Field 11644 Average Time

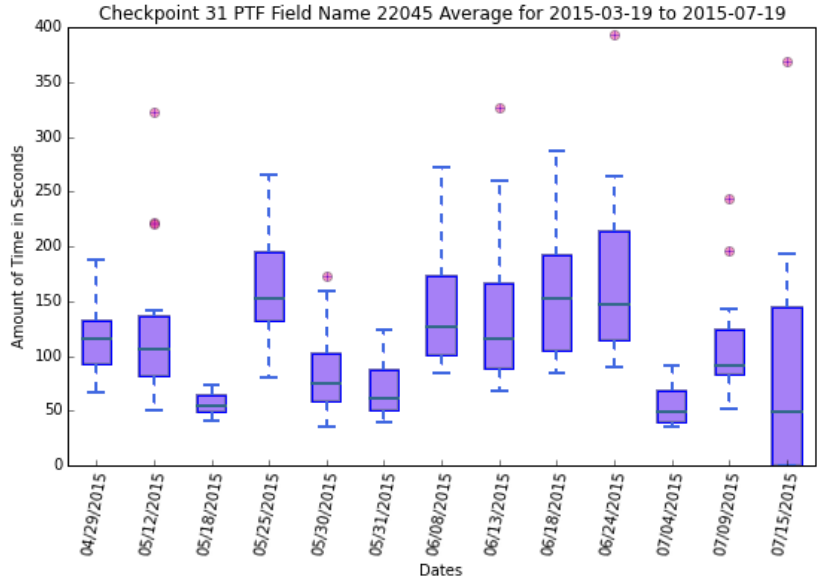


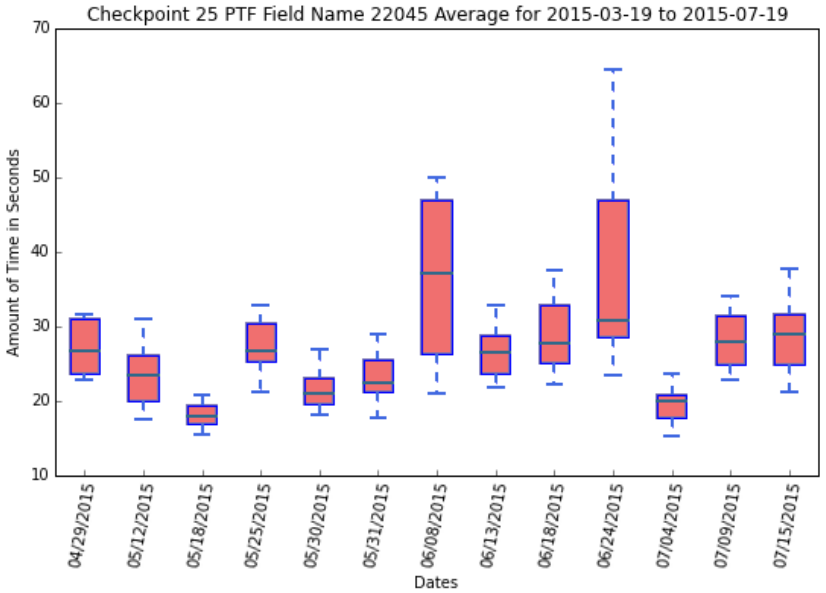




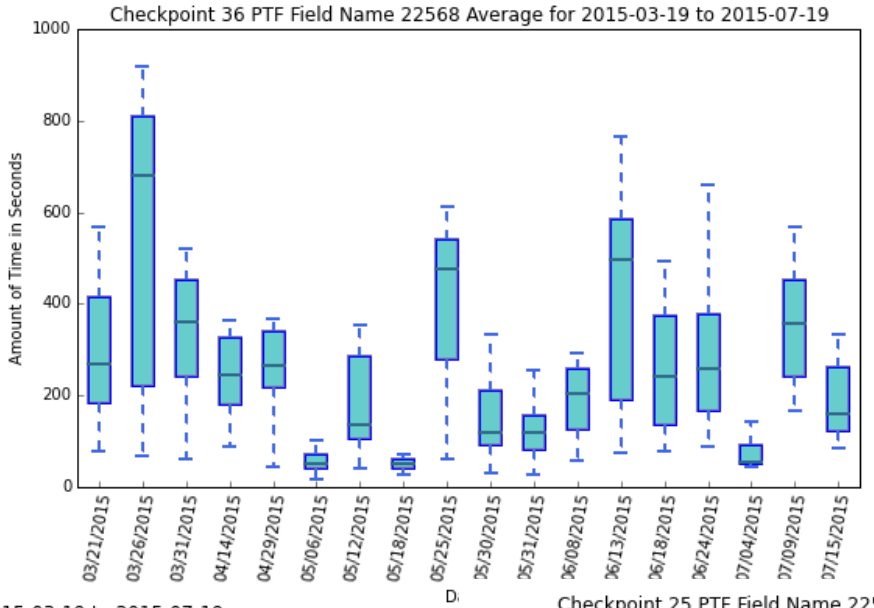
Checkpoints 36, 31, and 25 PTF Field 22045 Average Time

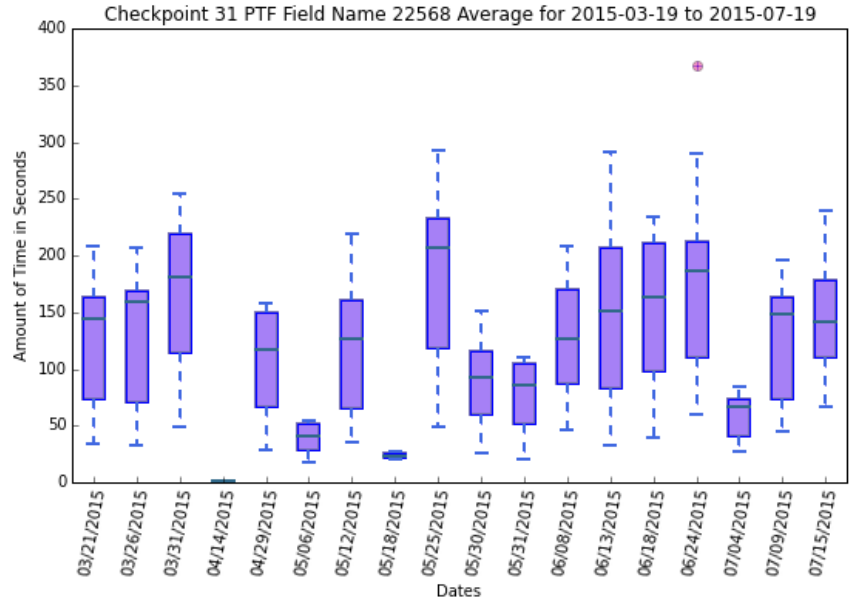


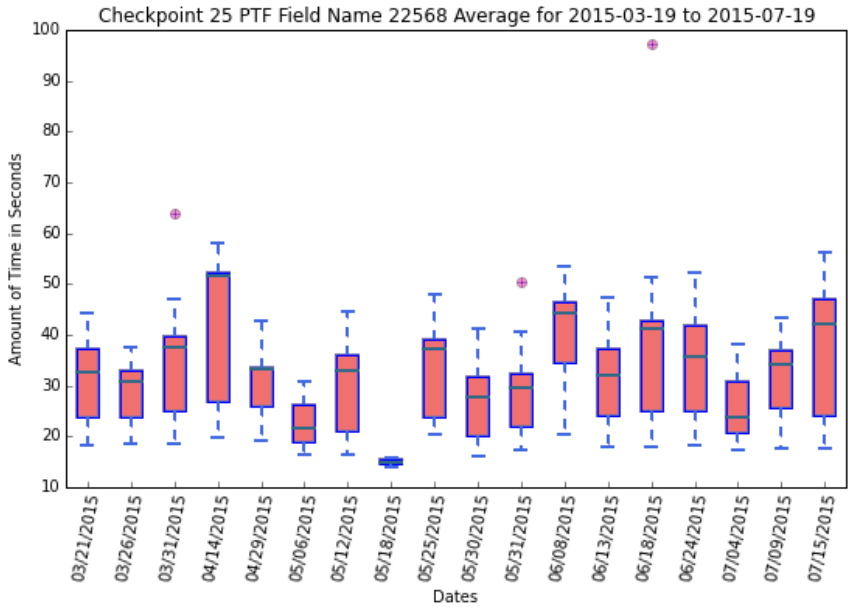




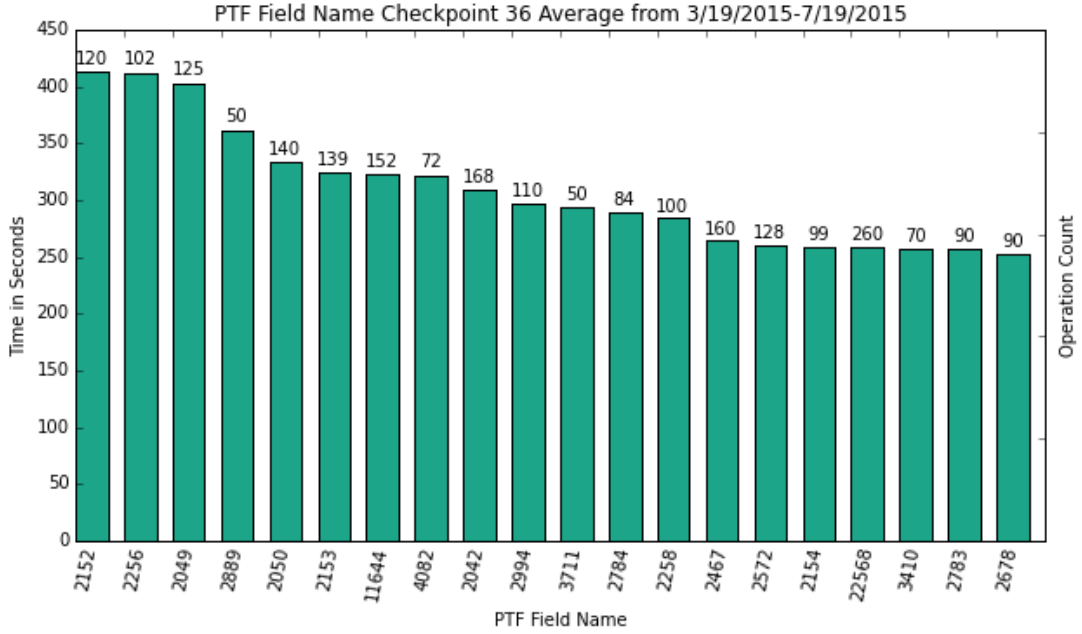
Checkpoints 36, 31, and 25 PTF Field 22568 Average Time

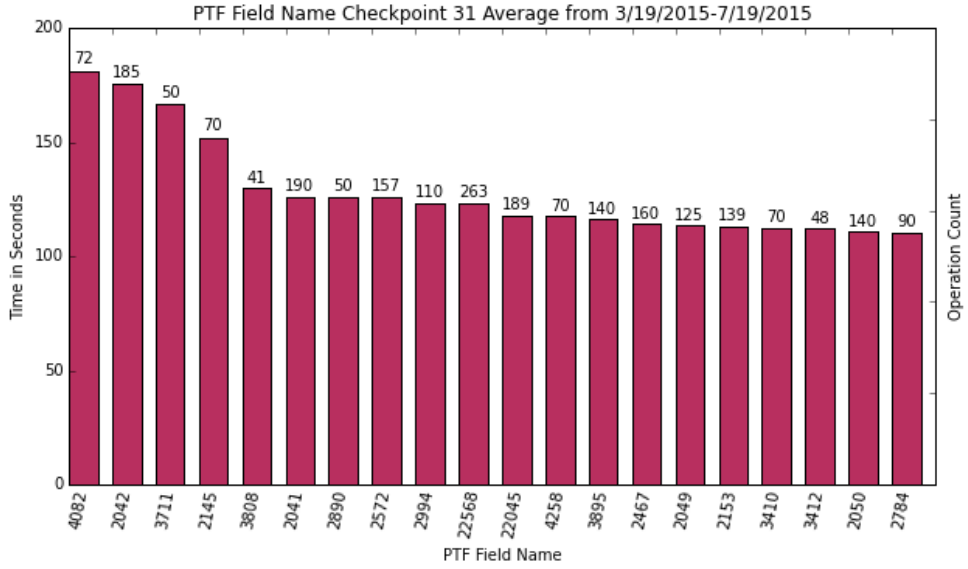


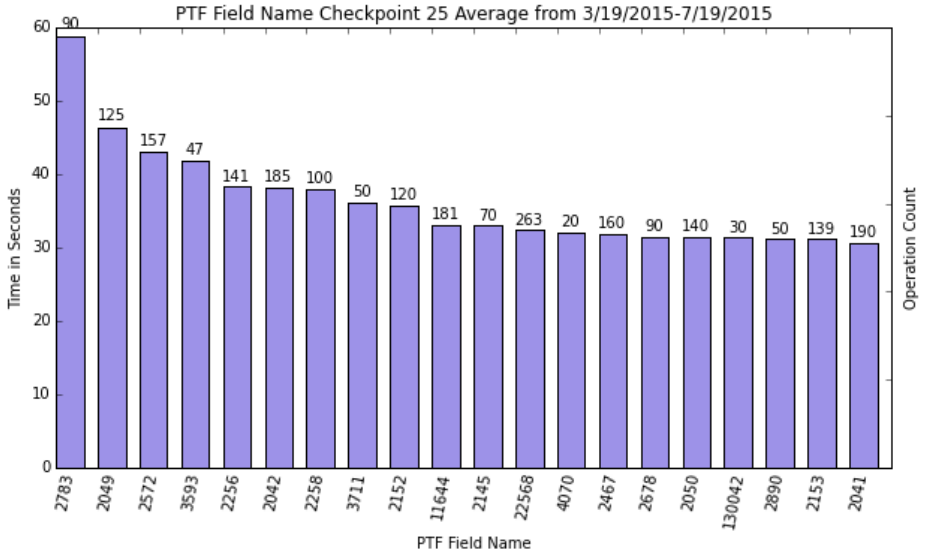




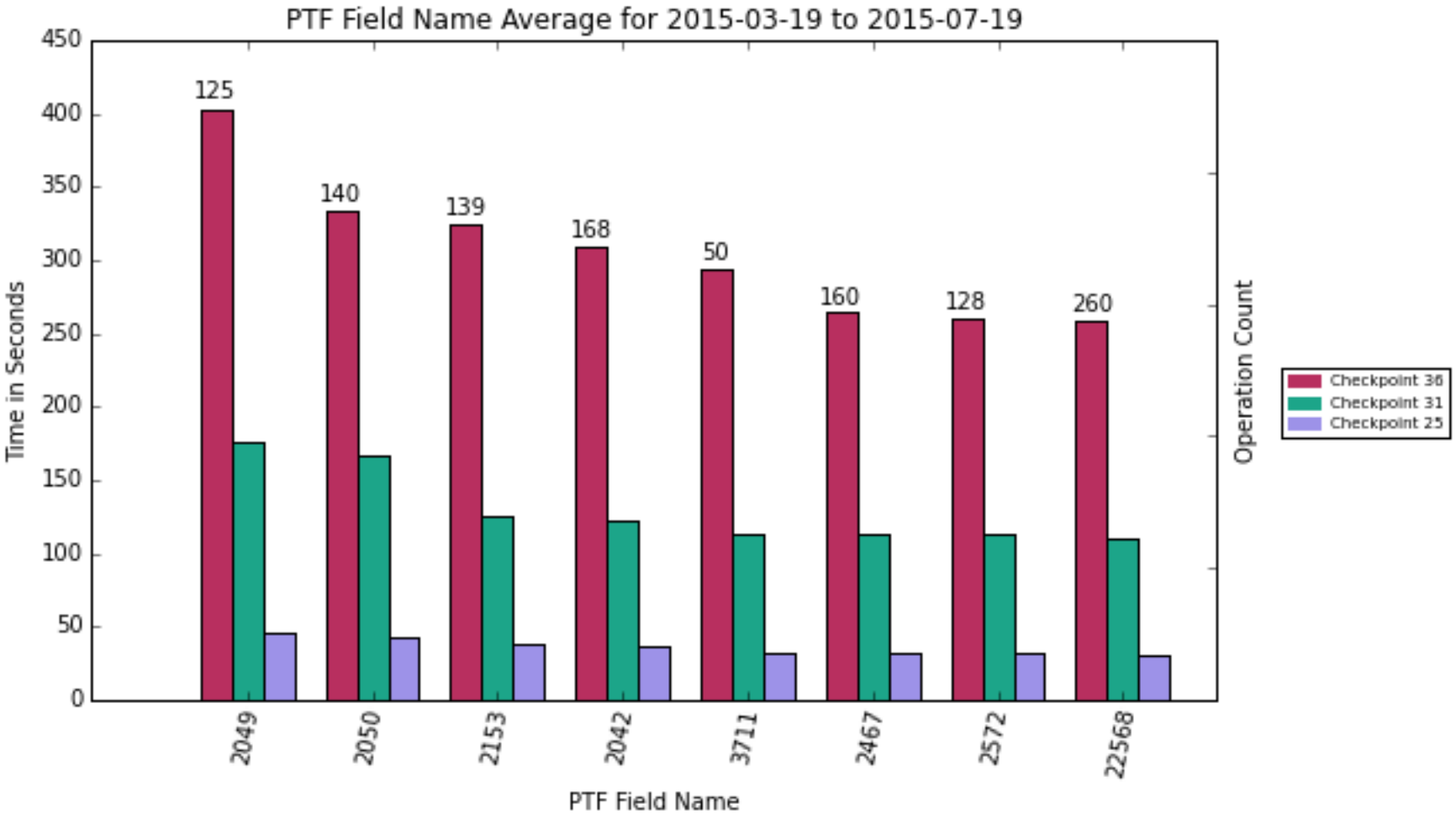
Checkpoints 36, 31, and 25 PTF Fields Average Time



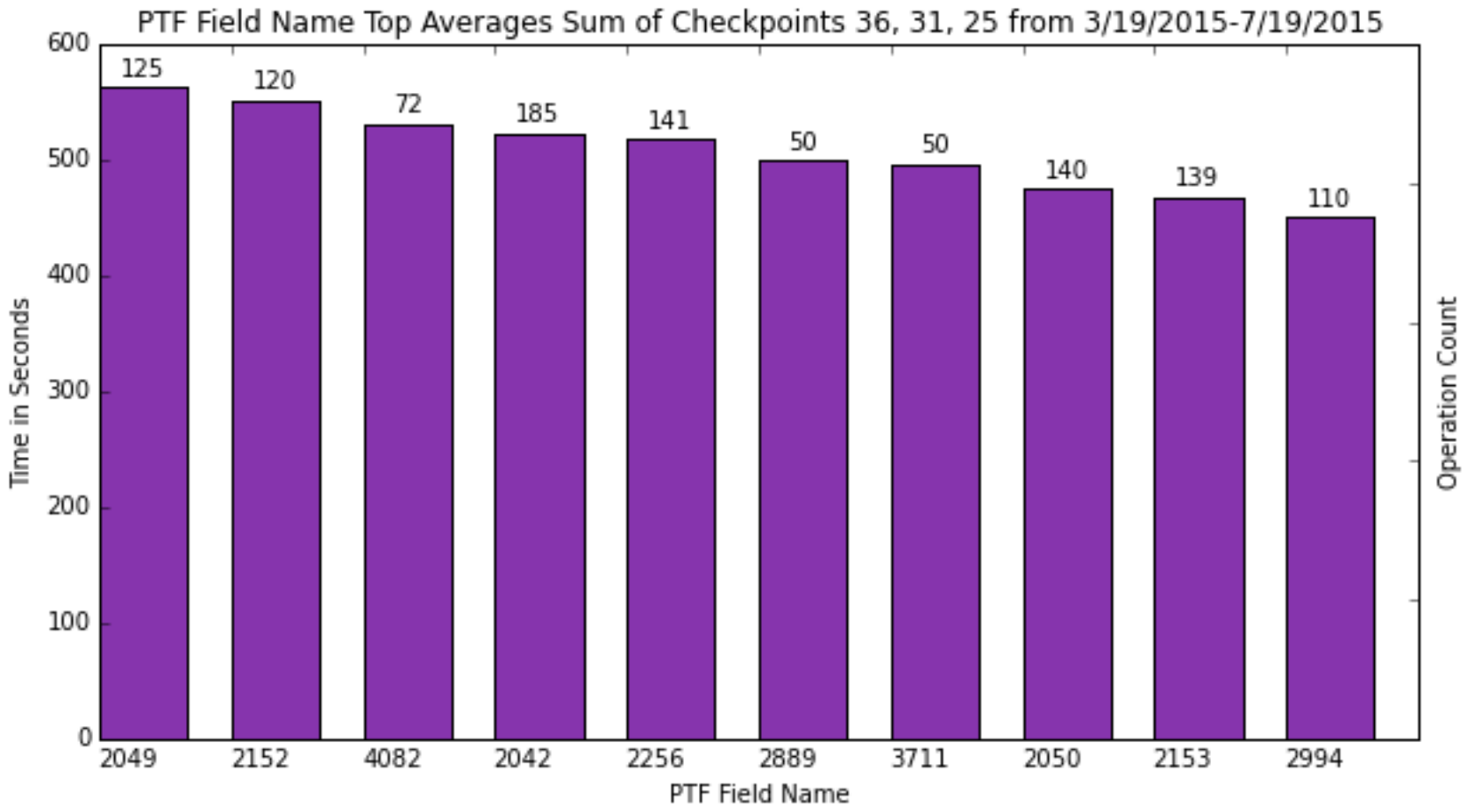




Checkpoints 36, 31, and 25 Shared PTF Fields Average Time

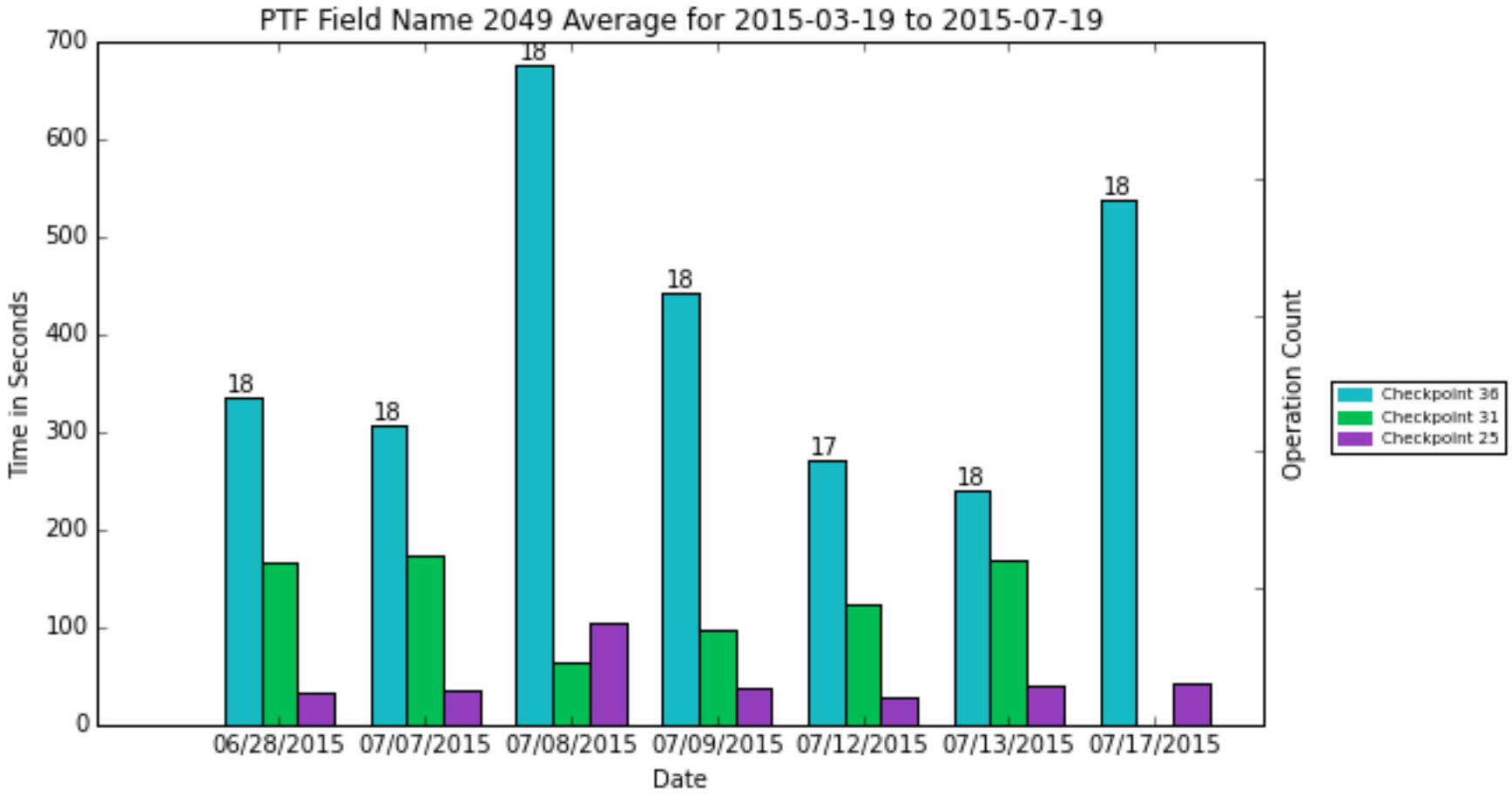


Checkpoints 36, 31, and 25 PTF Fields Top Average Time

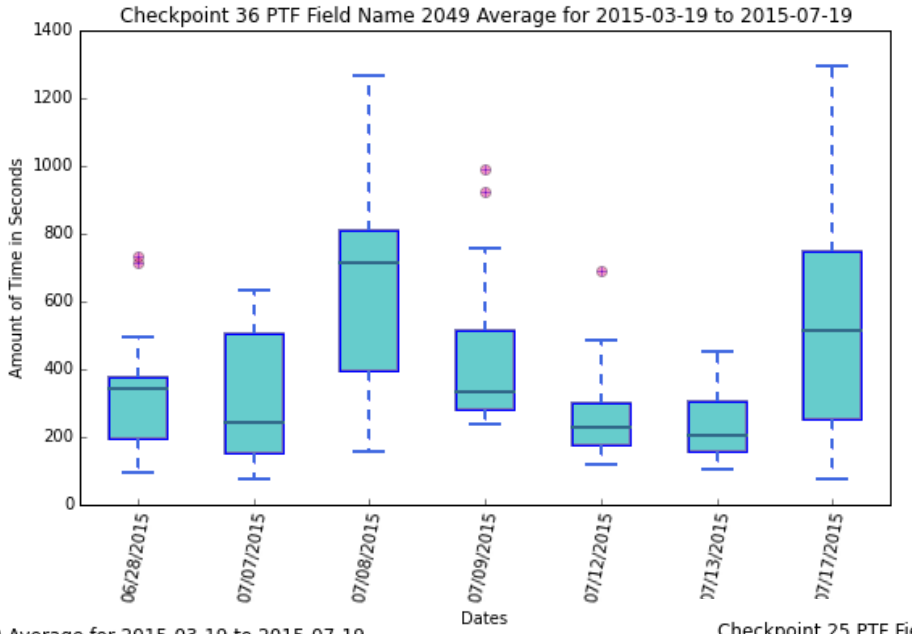


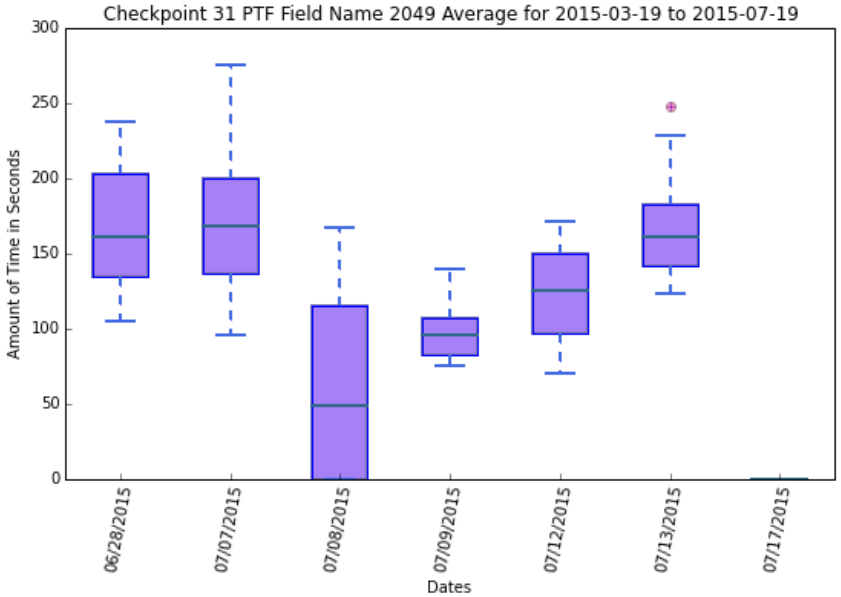
TOP 3 PTF FIELDS WITH TOP AVERAGE TIME ANALYZED: PTF FIELDS 2049, 2152, AND 4082

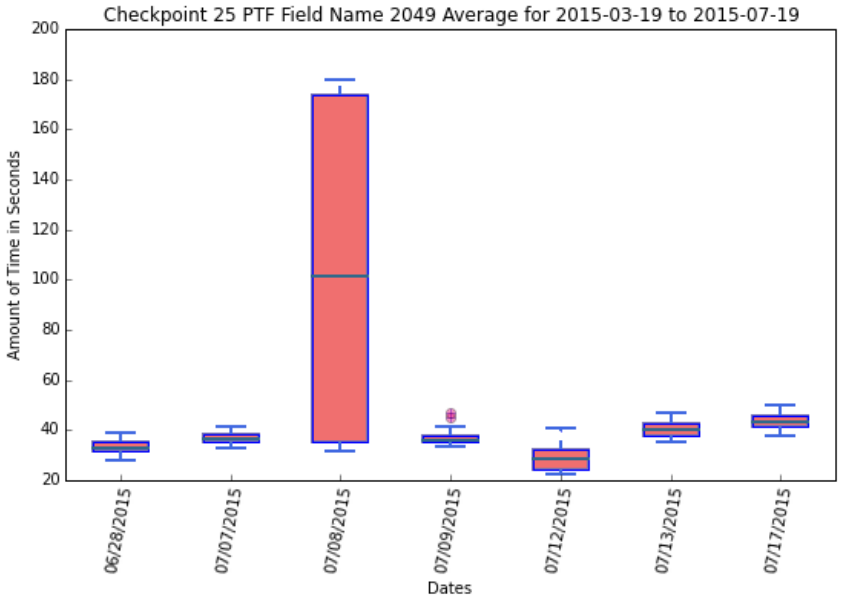
Checkpoints 36, 31, and 25 PTF Field 2049 Average Time



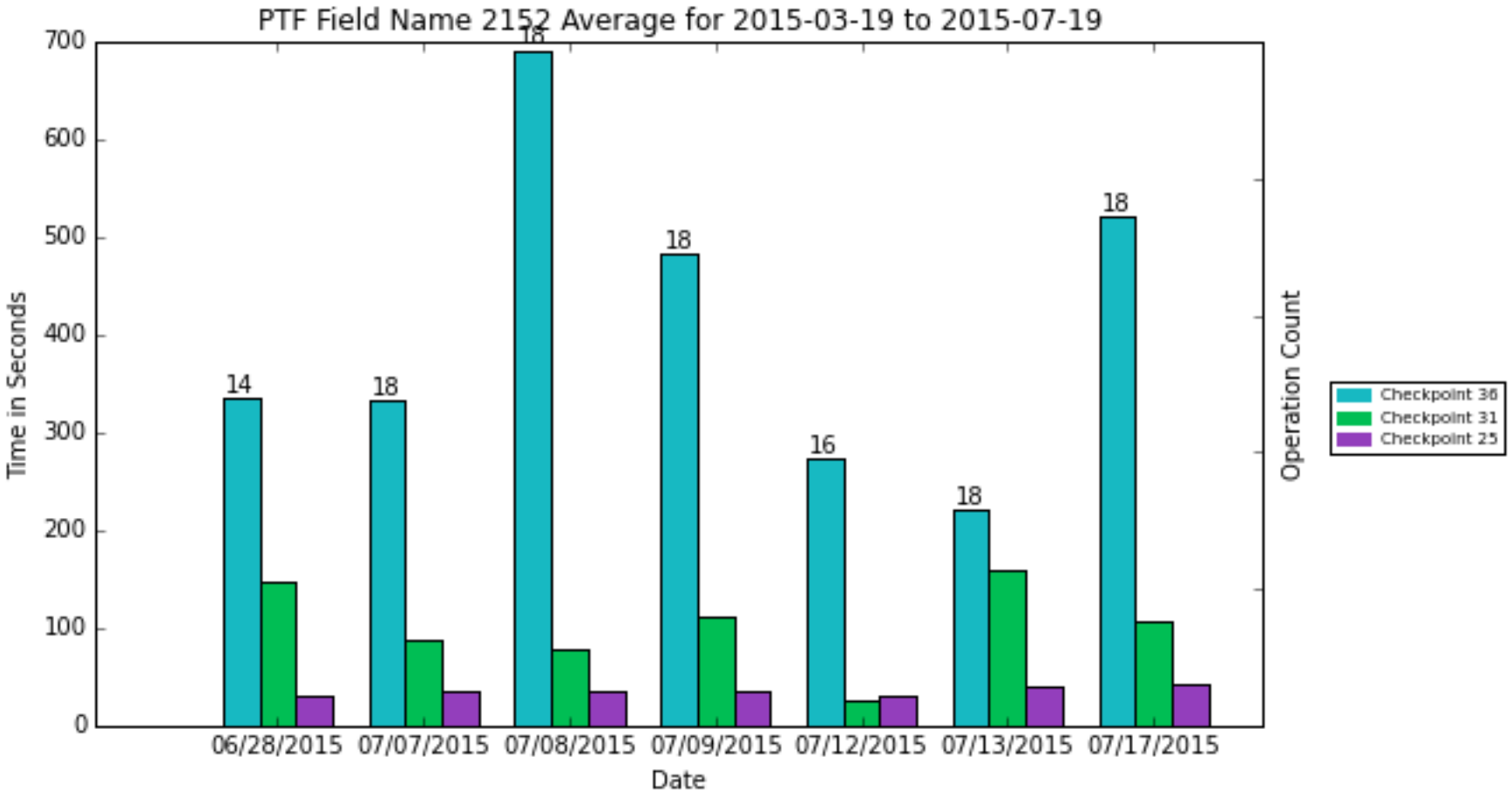
Checkpoints 36, 31, and 25 PTF Field 2049 Average Time



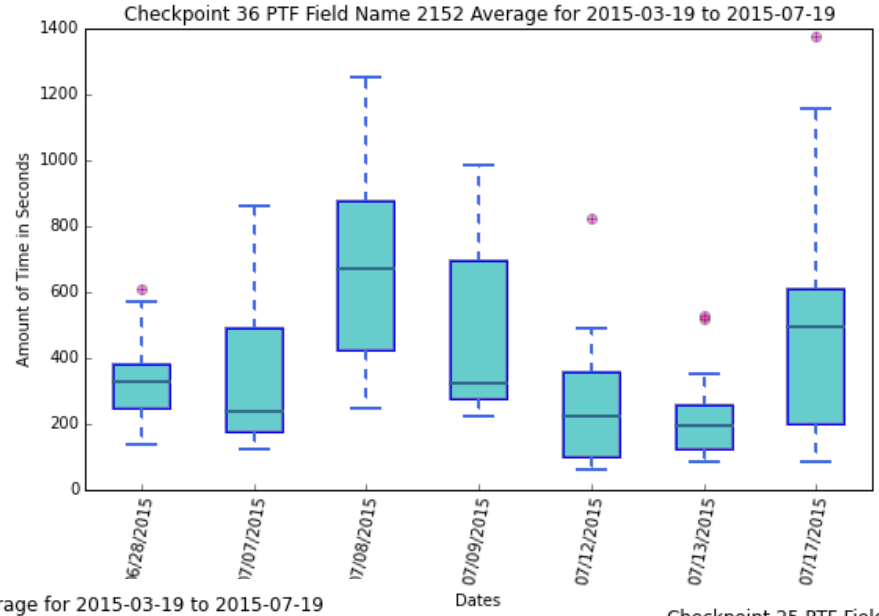


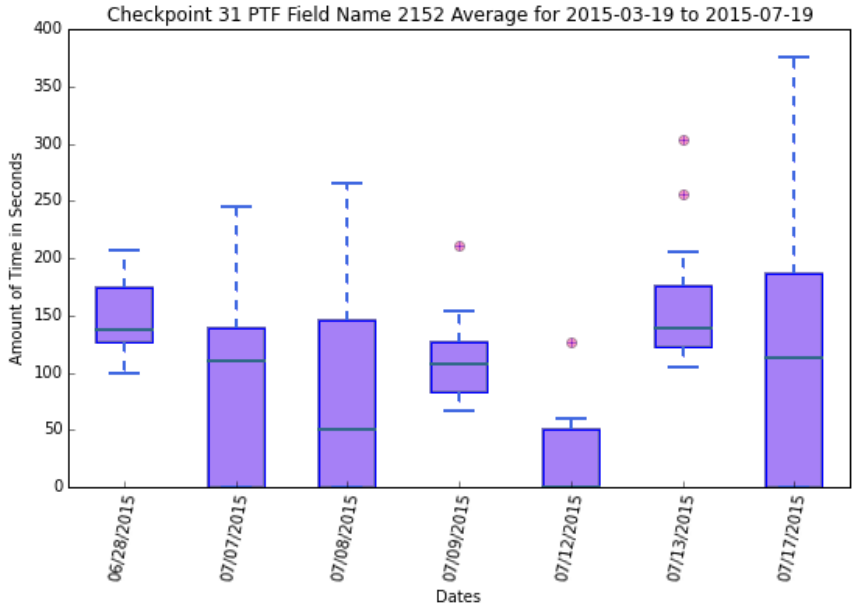


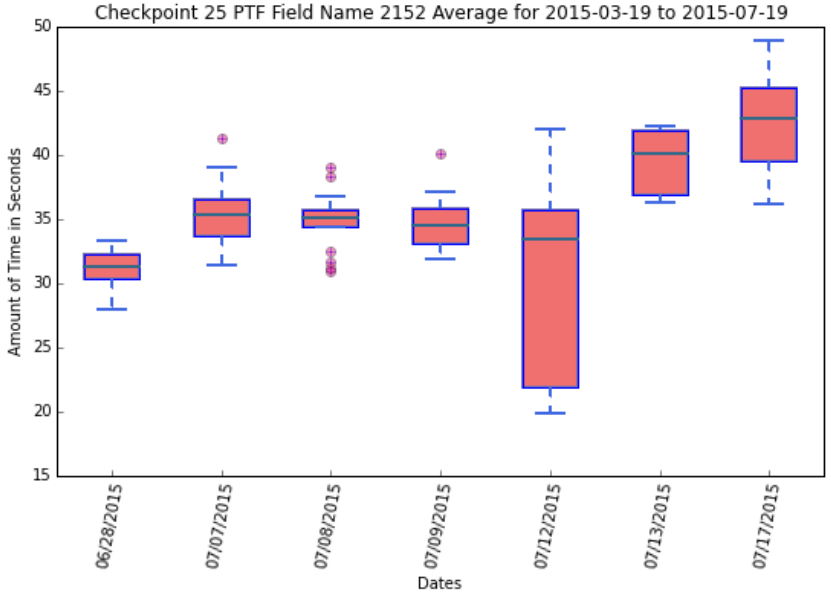
Checkpoints 36, 31, and 25 PTF Field 2152 Average Time



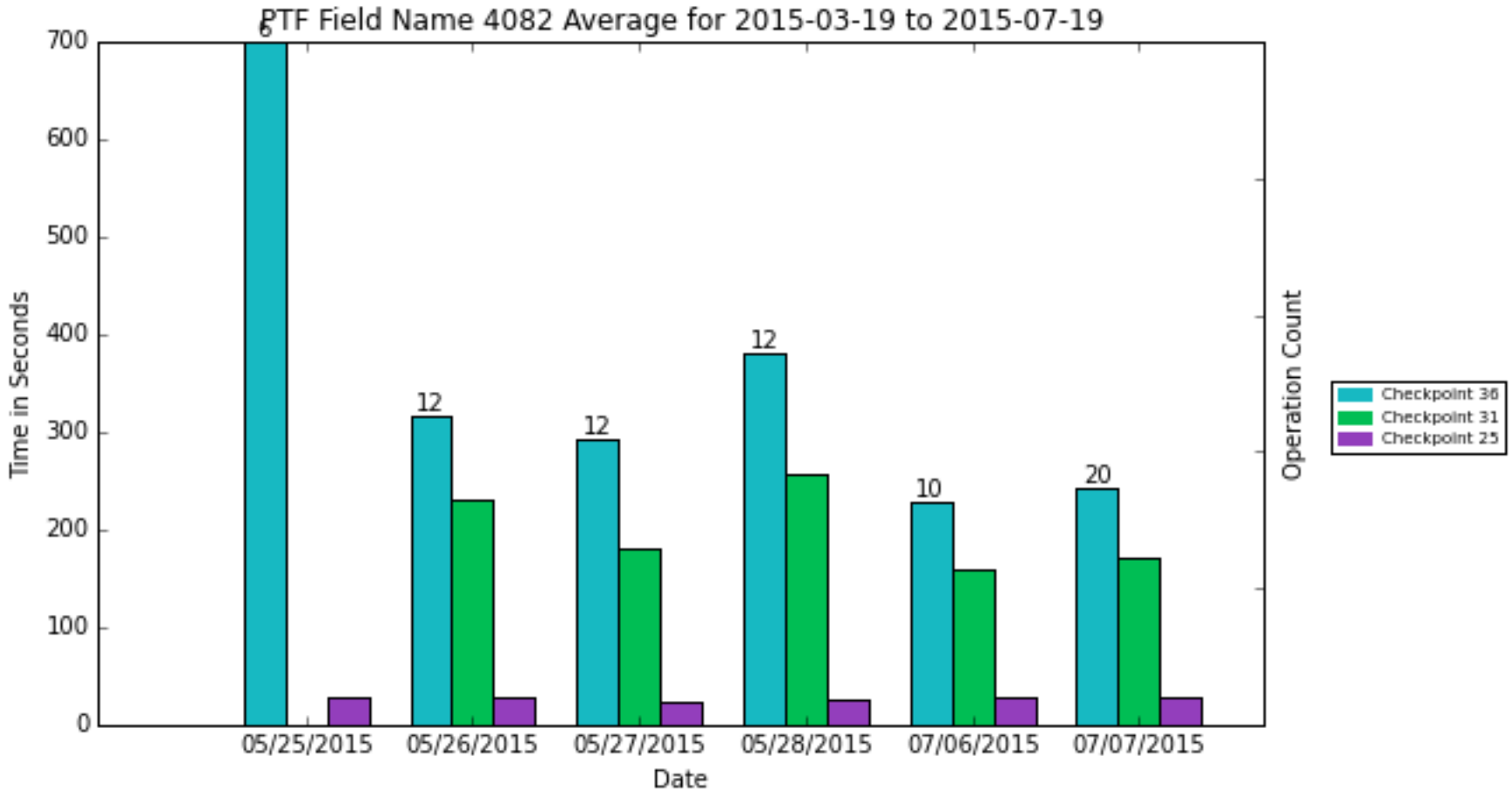
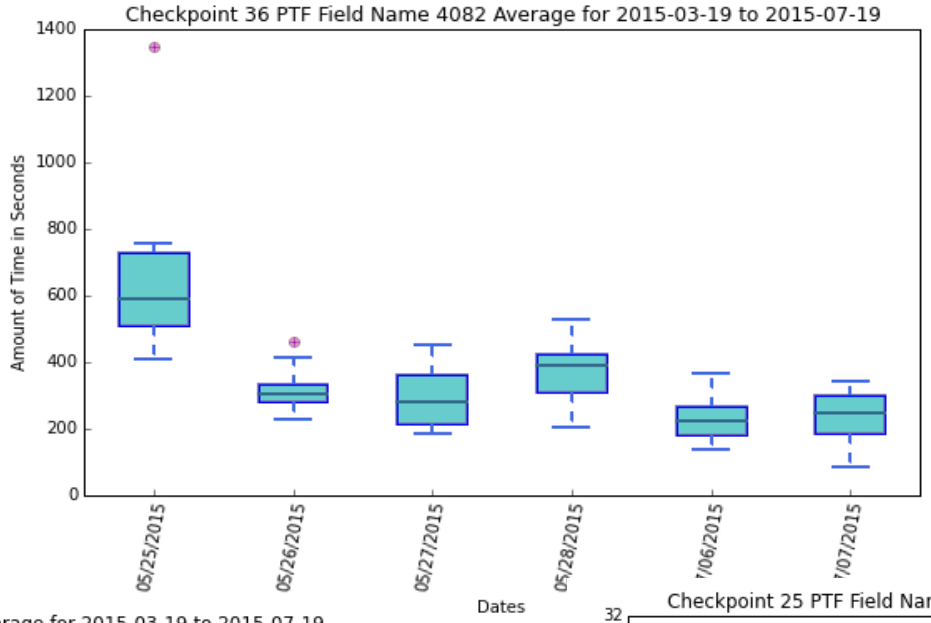
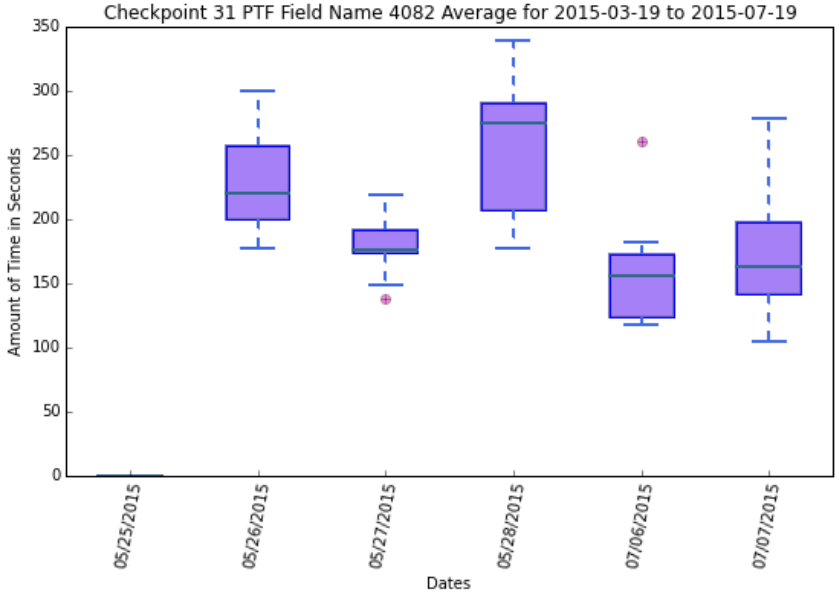
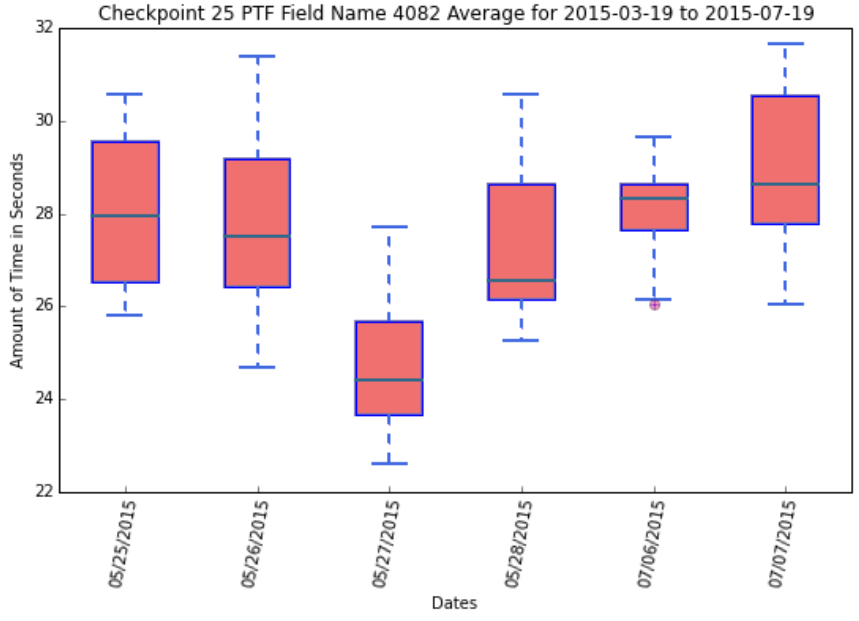
Checkpoints 36, 31, and 25 PTF Field 2152 Average Time







Checkpoints 36, 31, and 25 PTF Field 4082 Average Time



Checkpoints 36, 31, and 25 PTF Field 4082 Average Time