经观察，以下属性为标称属性：'Drive', 'qtr', 'down', 'SideofField', 'ydstogo', 'GoalToGo', 'FirstDown', 'posteam', 'DefensiveTeam', 'PlayAttempted', 'sp', 'Touchdown', 'ExPointResult', 'TwoPointConv', 'DefTwoPoint', 'Safety', 'Onsidekick', 'PuntResult', 'PlayType', 'Passer', 'Passer\_ID', 'PassAttempt', 'PassOutcome', 'PassLength', 'QBHit', 'PassLocation', 'InterceptionThrown', 'Interceptor', 'Rusher', 'Rusher\_ID', 'RushAttempt', 'RunLocation', 'RunGap', 'Receiver', 'Receiver\_ID', 'Reception', 'ReturnResult', 'Returner', 'BlockingPlayer', 'Tackler1', 'Tackler2', 'FieldGoalResult', 'Fumble', 'RecFumbTeam', 'RecFumbPlayer', 'Sack', 'Challenge.Replay', 'ChalReplayResult', 'Accepted.Penalty', 'PenalizedTeam', 'PenaltyType', 'PenalizedPlayer', 'HomeTeam', 'AwayTeam', 'Timeout\_Indicator', 'Timeout\_Team', 'Season'

以下属性为数值属性：'TimeUnder', 'TimeSecs', 'PlayTimeDiff', 'yrdln', 'yrdline100', 'ydsnet', 'Yards.Gained', 'AirYards', 'YardsAfterCatch', 'FieldGoalDistance', 'Penalty.Yards', 'PosTeamScore', 'DefTeamScore', 'ScoreDiff', 'AbsScoreDiff', 'posteam\_timeouts\_pre', 'HomeTimeouts\_Remaining\_Pre','AwayTimeouts\_Remaining\_Pre', 'HomeTimeouts\_Remaining\_Post', 'AwayTimeouts\_Remaining\_Post', 'No\_Score\_Prob', 'Opp\_Field\_Goal\_Prob', 'Opp\_Safety\_Prob', 'Opp\_Touchdown\_Prob', 'Field\_Goal\_Prob', 'Safety\_Prob', 'Touchdown\_Prob', 'ExPoint\_Prob', 'TwoPoint\_Prob', 'ExpPts', 'EPA', 'airEPA', 'yacEPA', 'Home\_WP\_pre', 'Away\_WP\_pre', 'Home\_WP\_post', 'Away\_WP\_post', 'Win\_Prob', 'WPA', 'airWPA', 'yacWPA'

## 数据摘要

1. 标称属性

此处只列出部分数据

|  |  |
| --- | --- |
| value | count |
| 1 | 19261 |
| 7 | 18088 |
| 6 | 17925 |
| 4 | 17875 |
| 5 | 17822 |
| 8 | 17691 |
| 16 | 17525 |
| 17 | 17459 |
| 3 | 17425 |
| 18 | 17339 |
| 10 | 17193 |
| 15 | 17188 |
| 2 | 17084 |
| 9 | 17033 |
| 19 | 16928 |
| 11 | 16898 |
| 14 | 16881 |
| 13 | 16745 |
| 12 | 16724 |
| 20 | 15622 |
| 21 | 13998 |
| 22 | 12071 |
| 23 | 9933 |
| 24 | 7959 |
| 25 | 5842 |
| 26 | 4200 |
| 27 | 2815 |
| 28 | 1658 |
| 29 | 1077 |
| 30 | 675 |
| 31 | 366 |
| 32 | 237 |
| 33 | 97 |
| 34 | 45 |
| 35 | 9 |

其余数据见文件name\_cat\_out\_nfl.txt

1. 数值属性

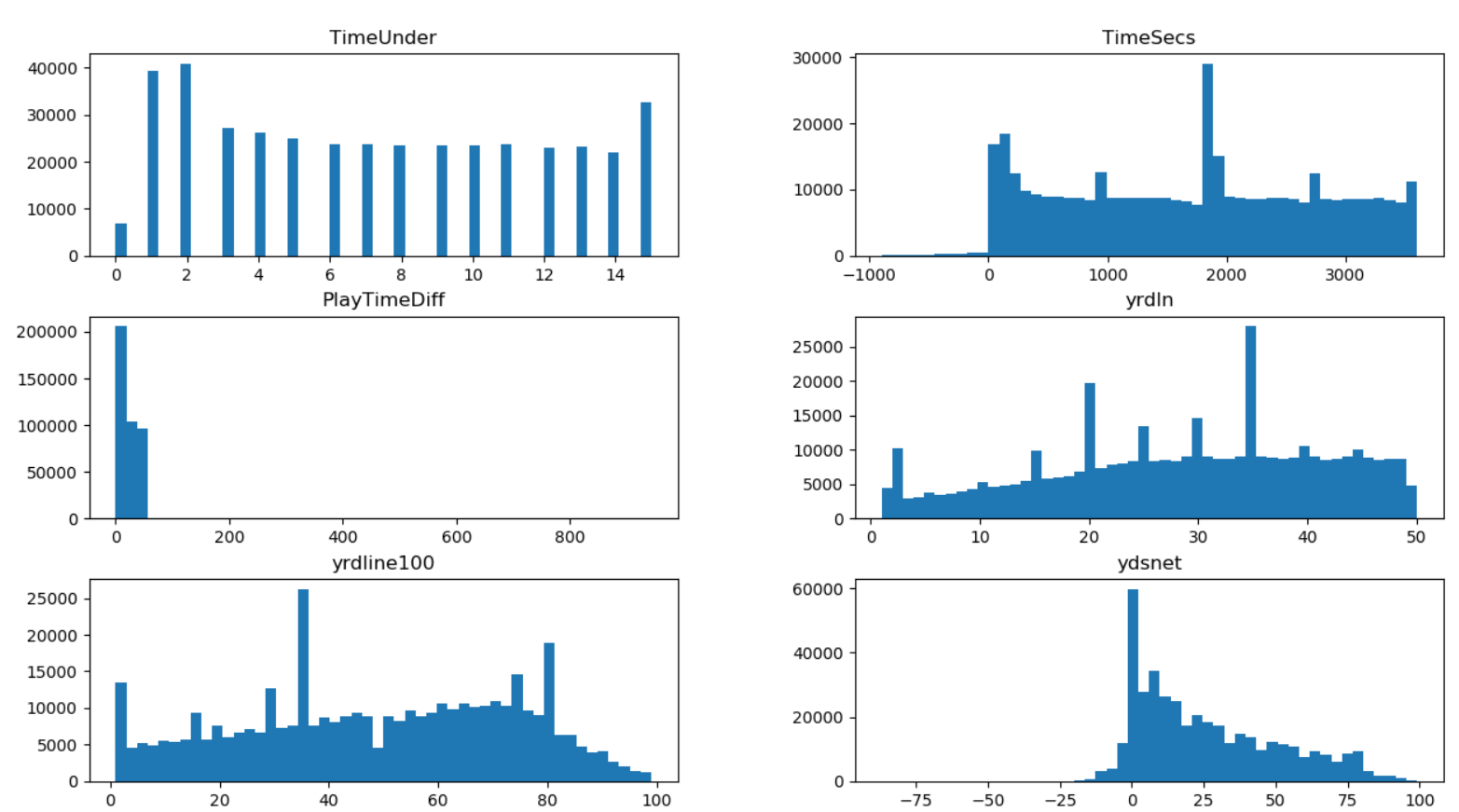
见下表

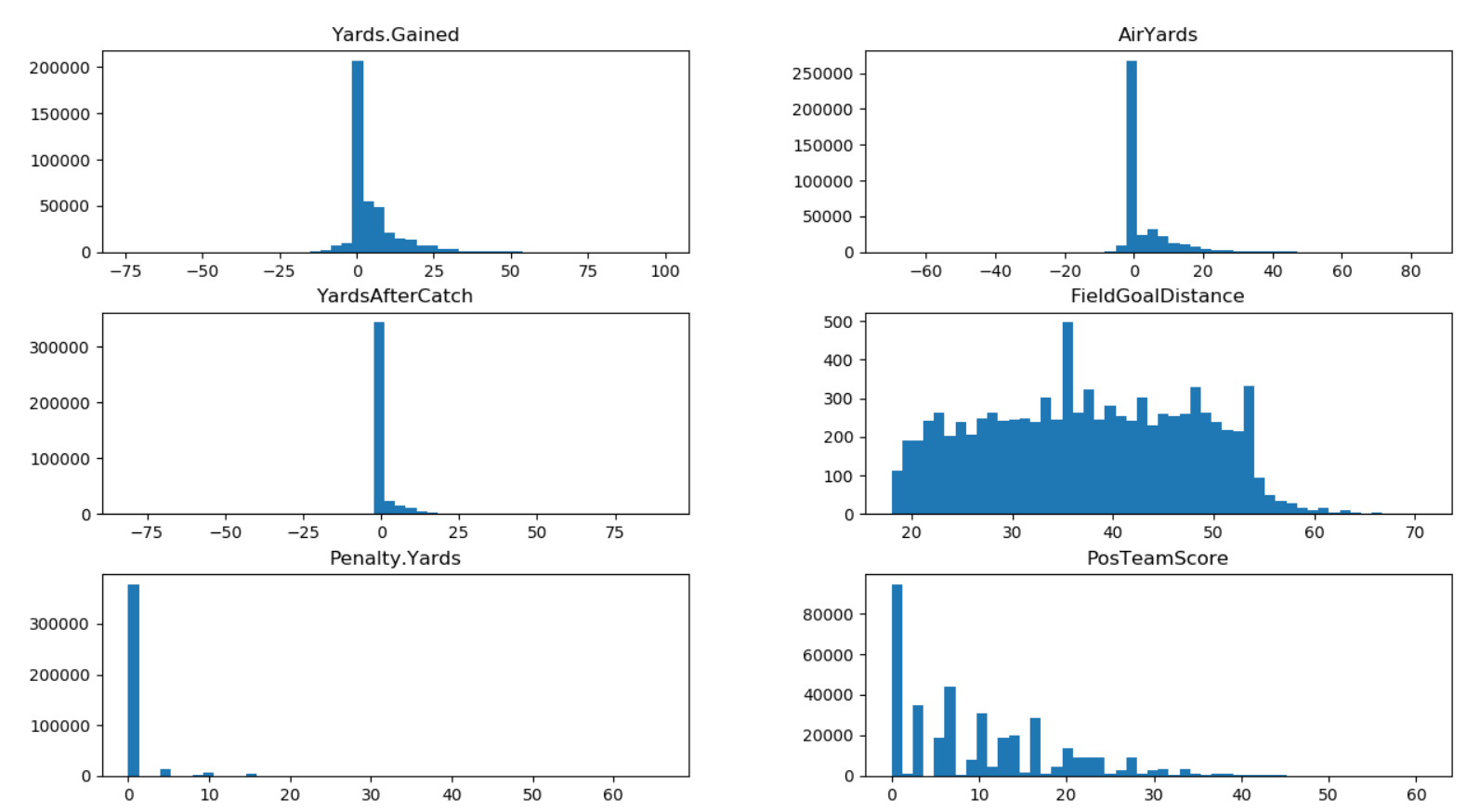


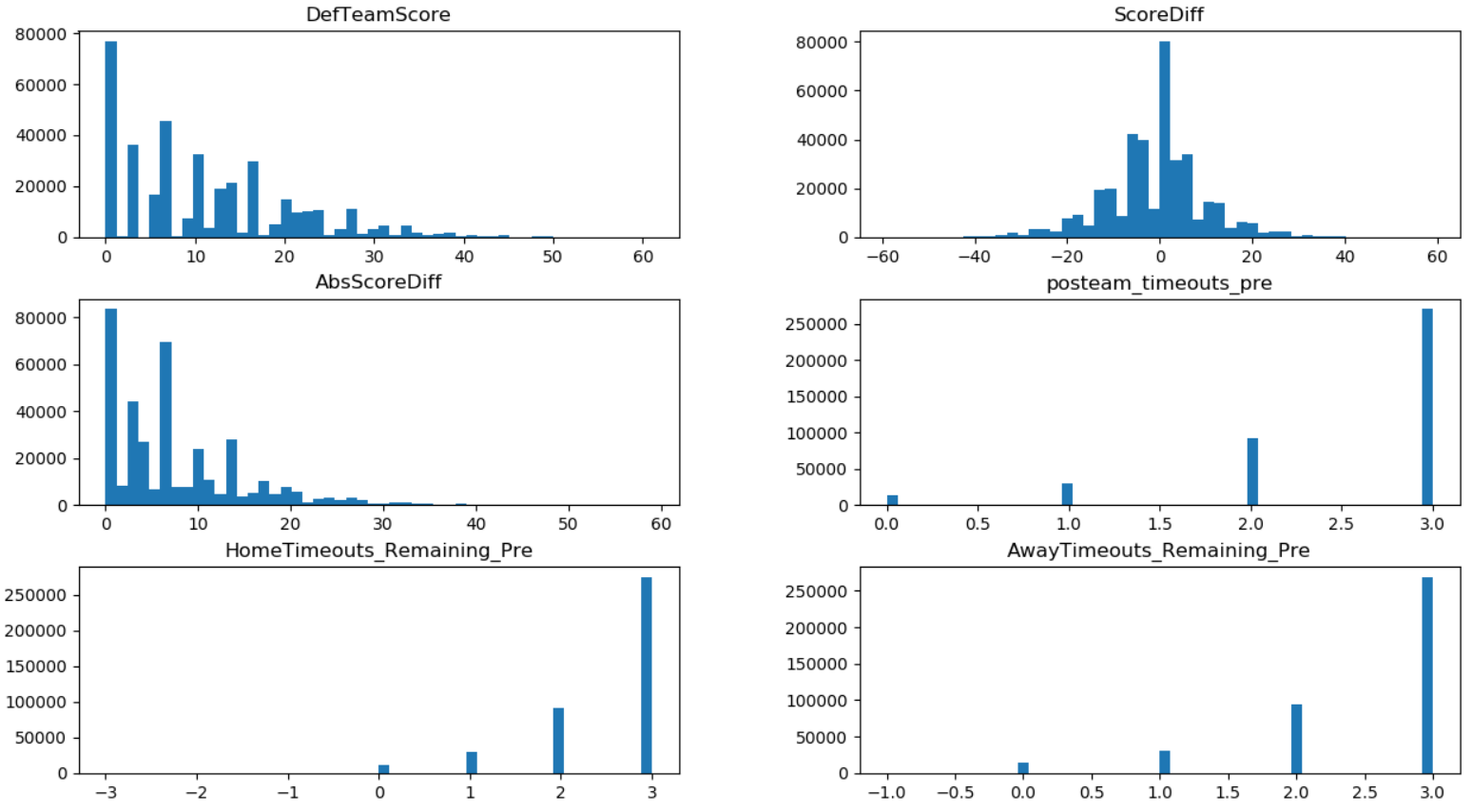
## 数据可视化

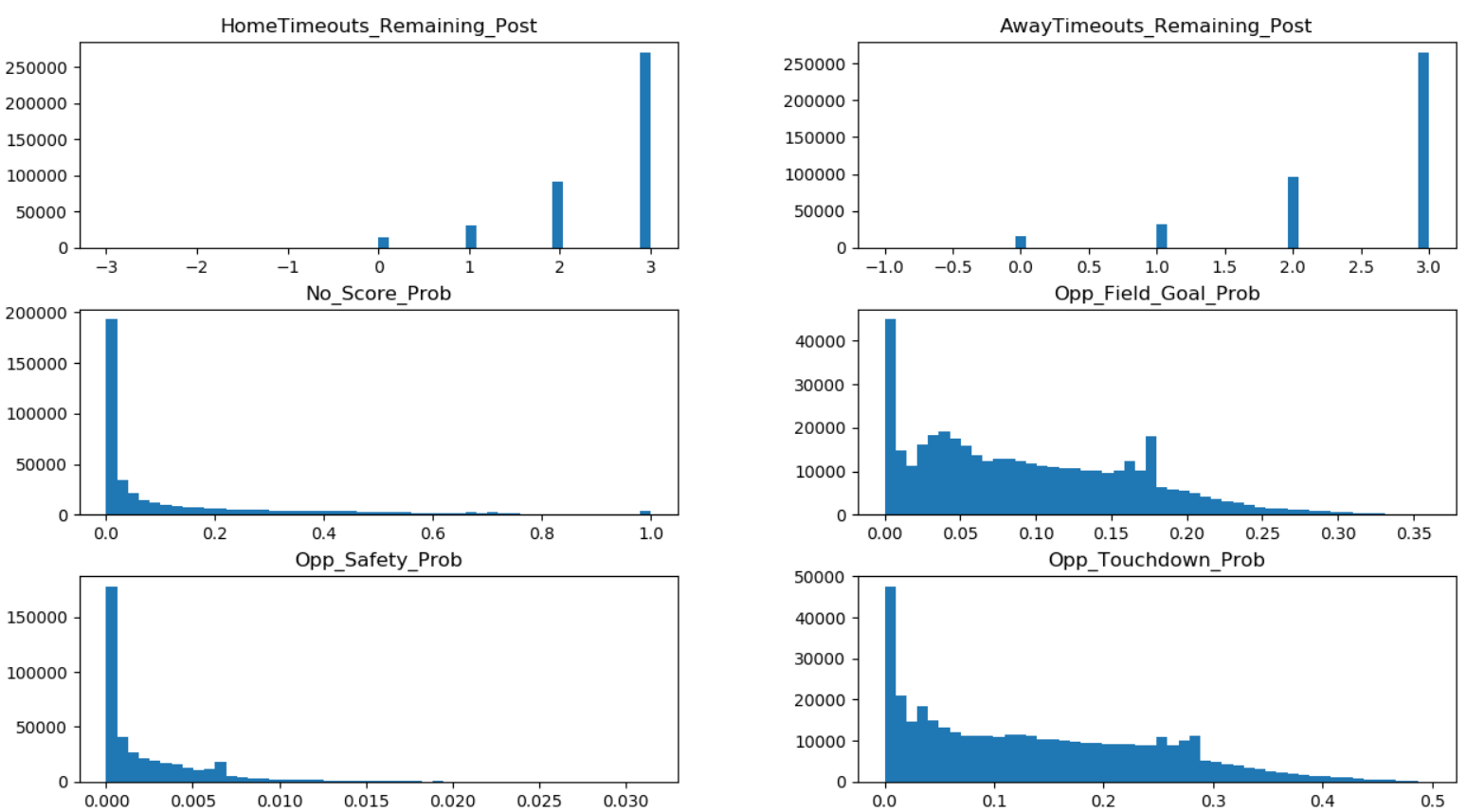
1. 直方图

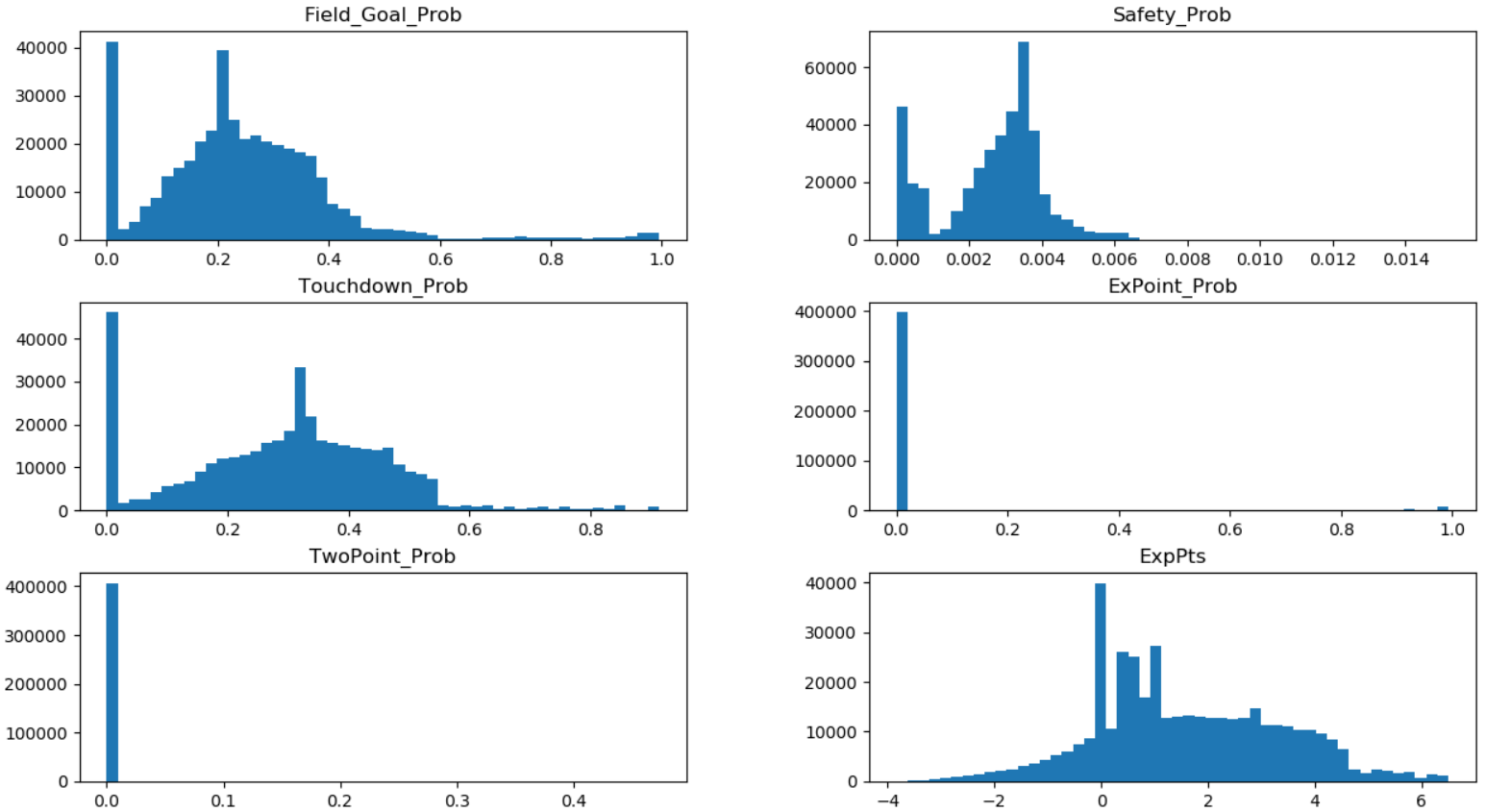
各属性直方图如下

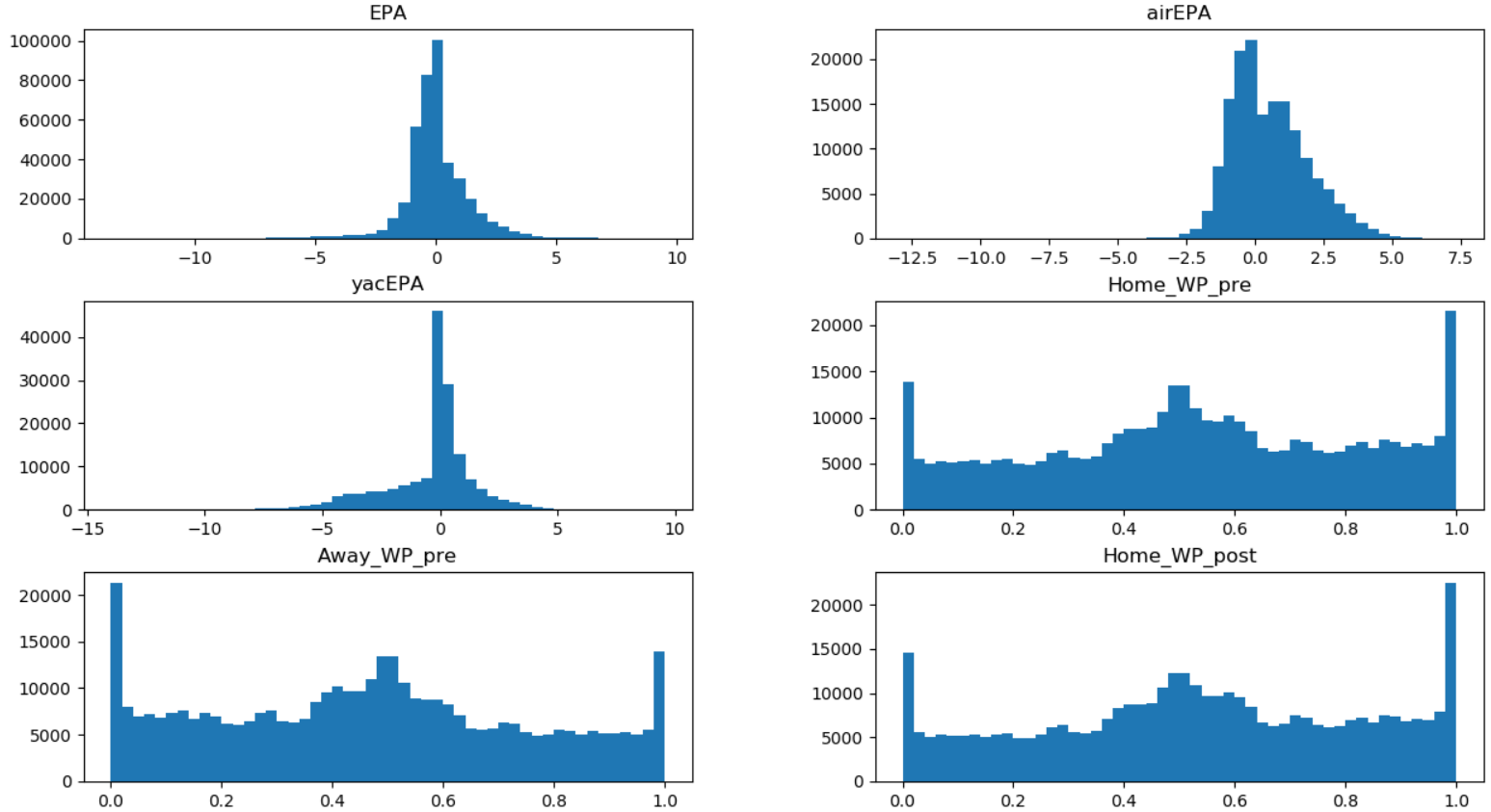


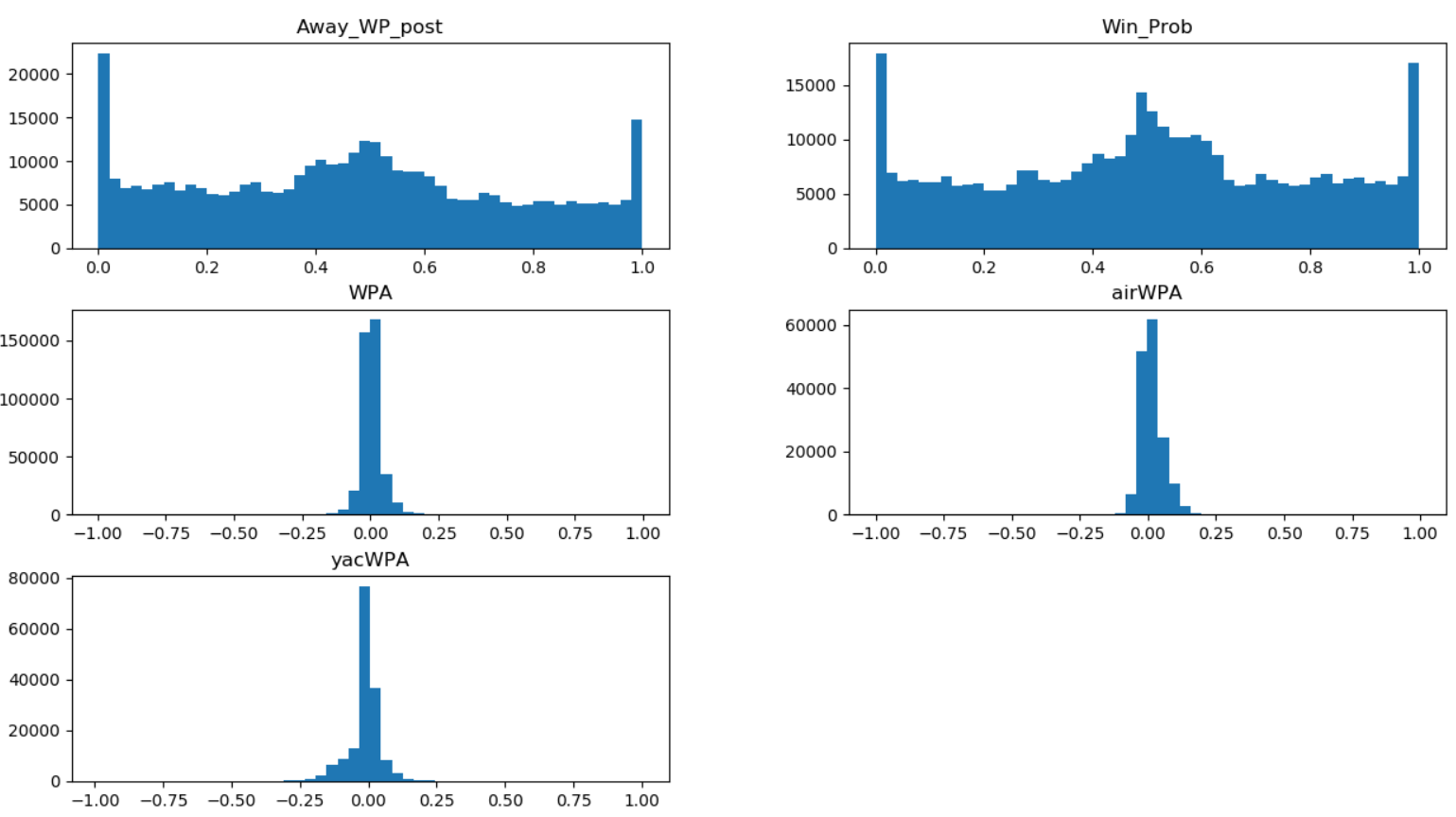






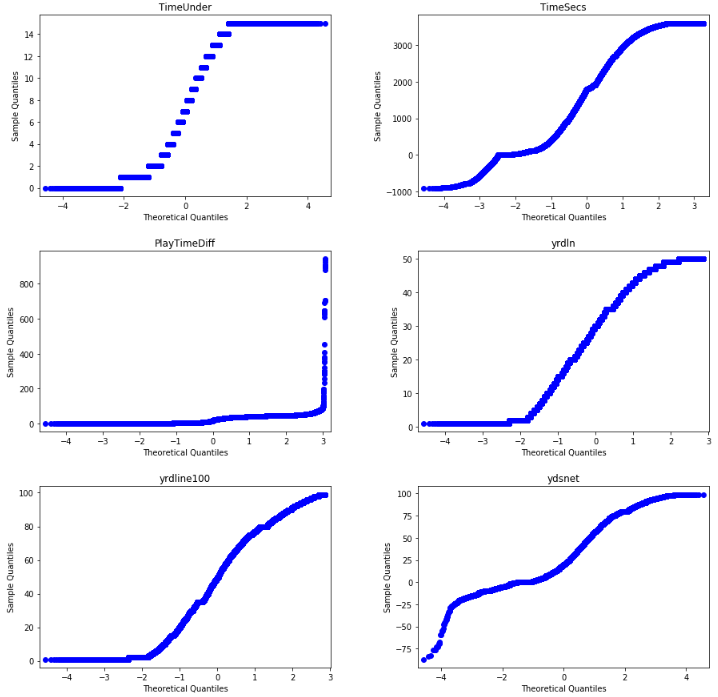


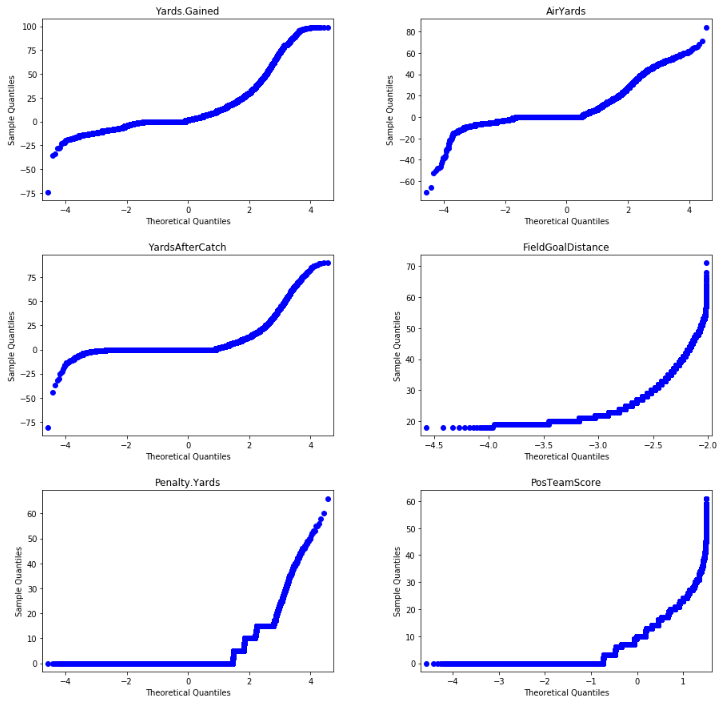


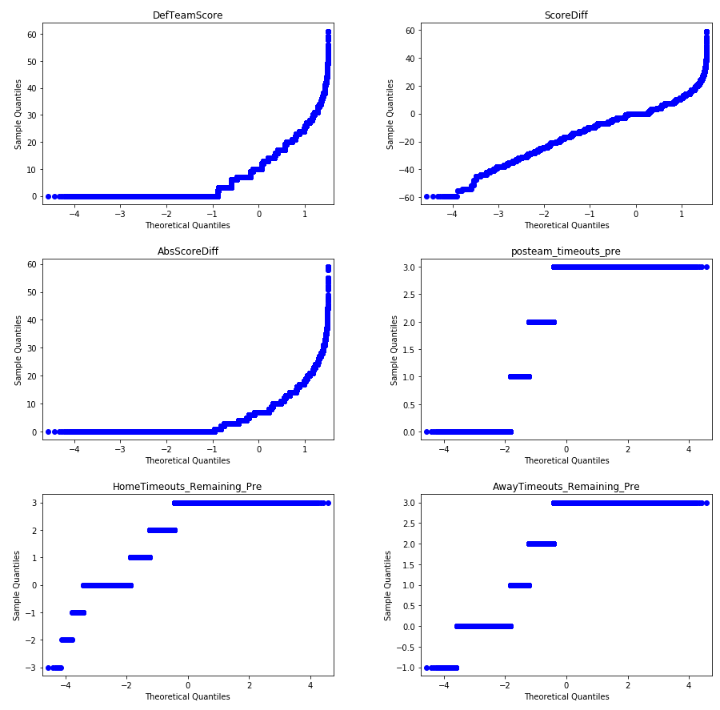


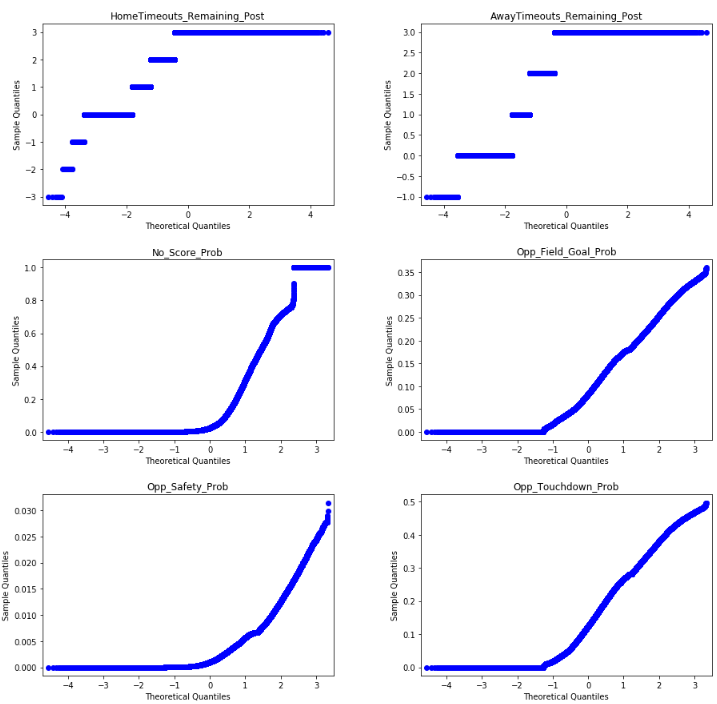
1. Q-Q图

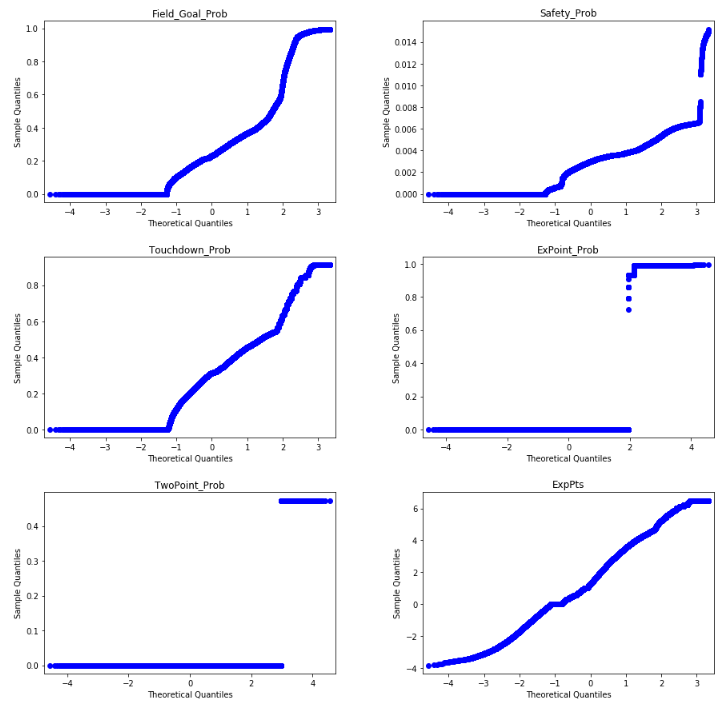
Q-Q图趋于直线的属性接近正态分布

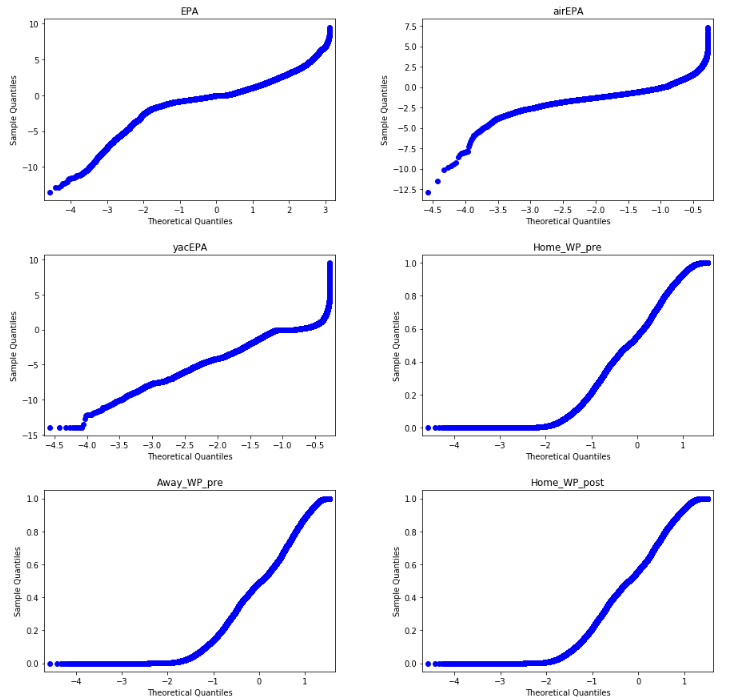




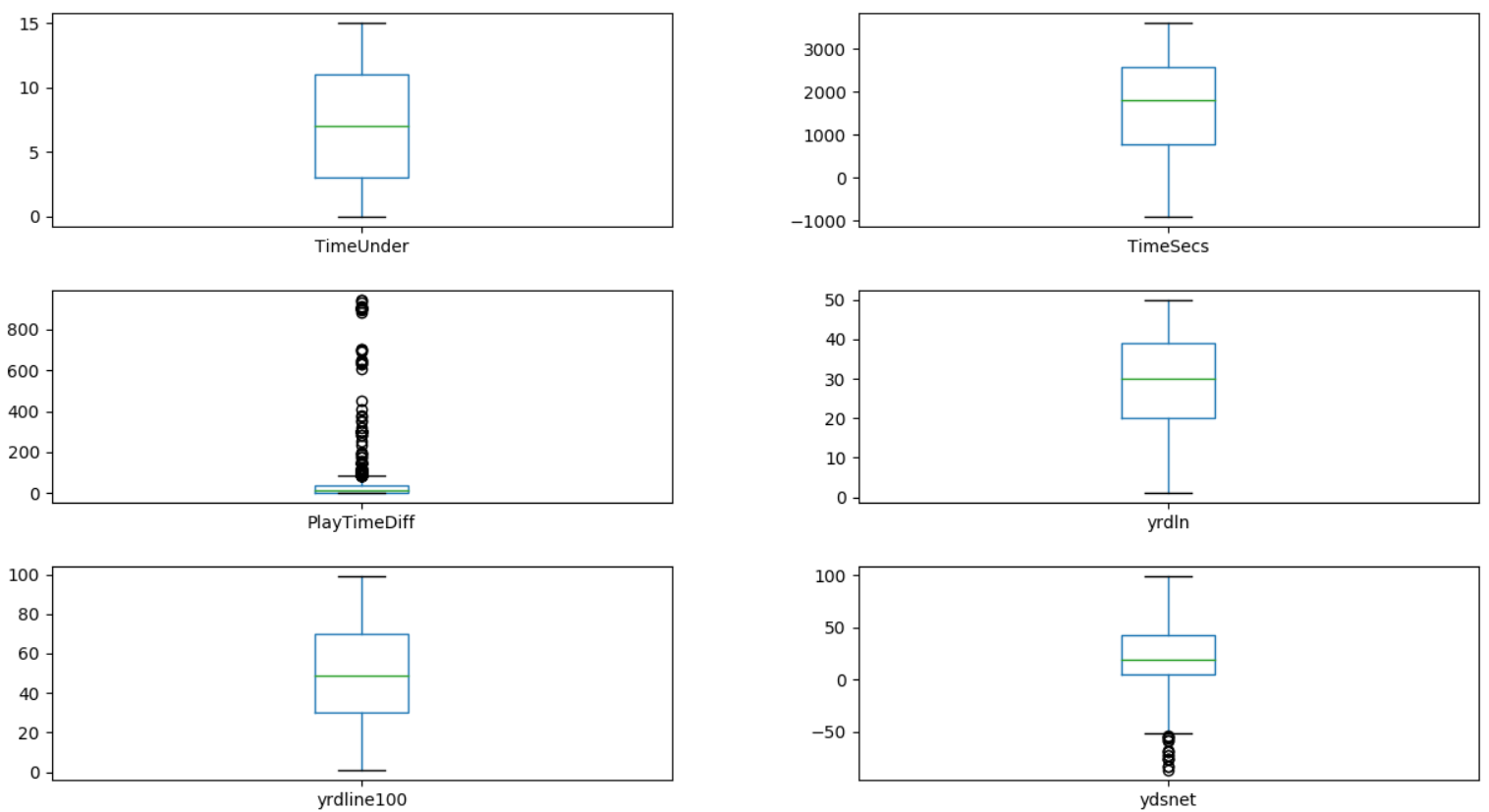


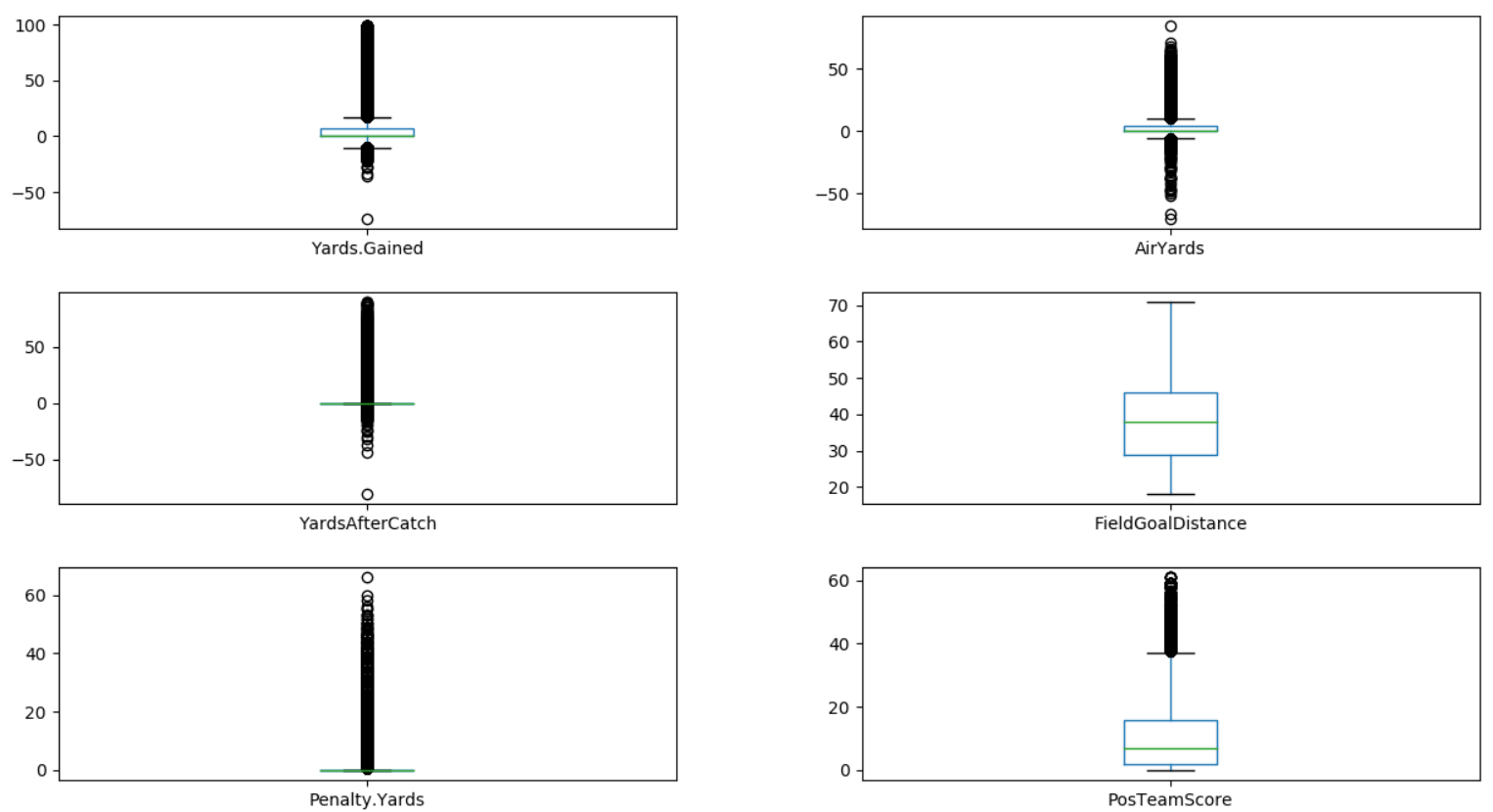


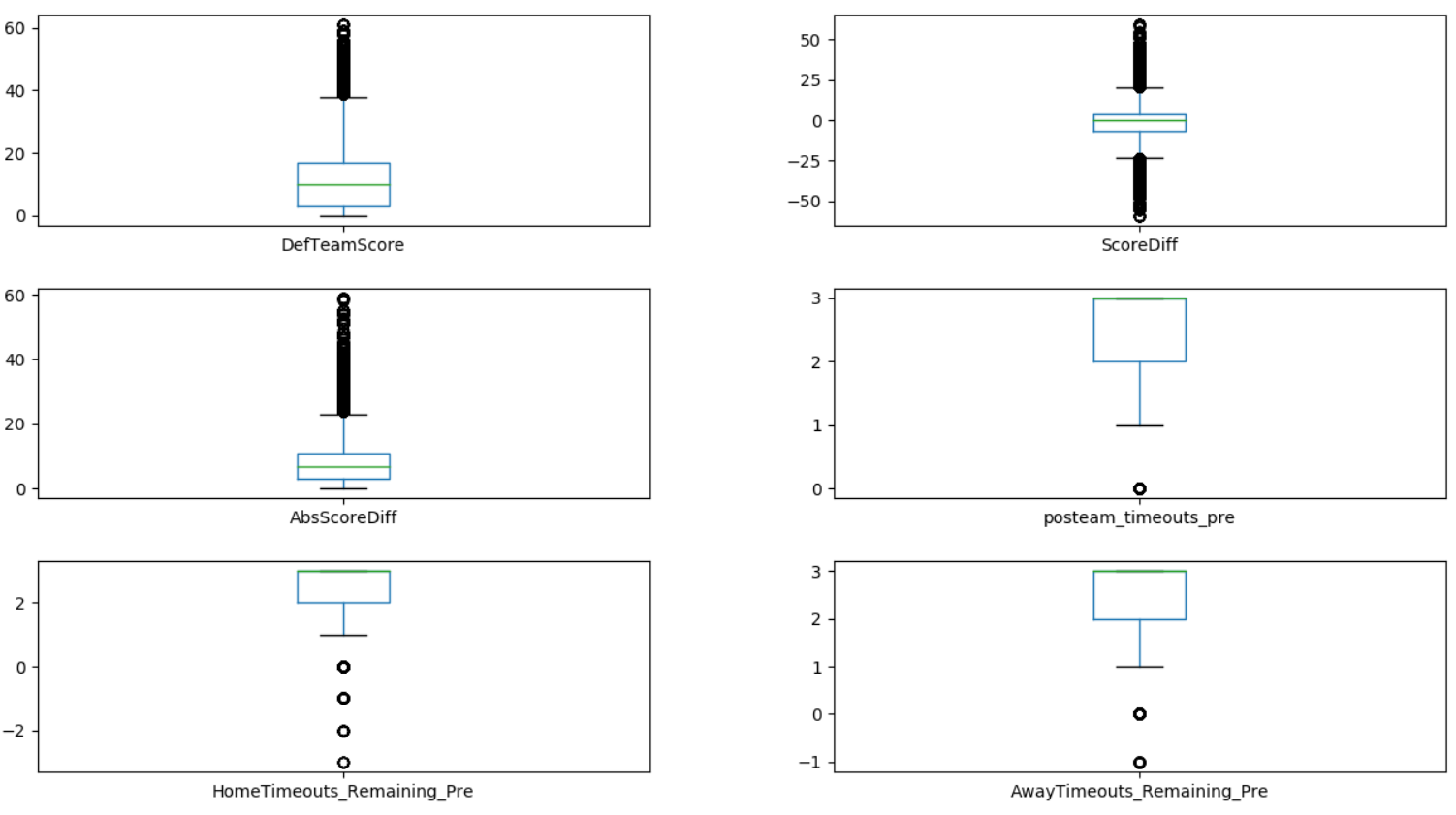


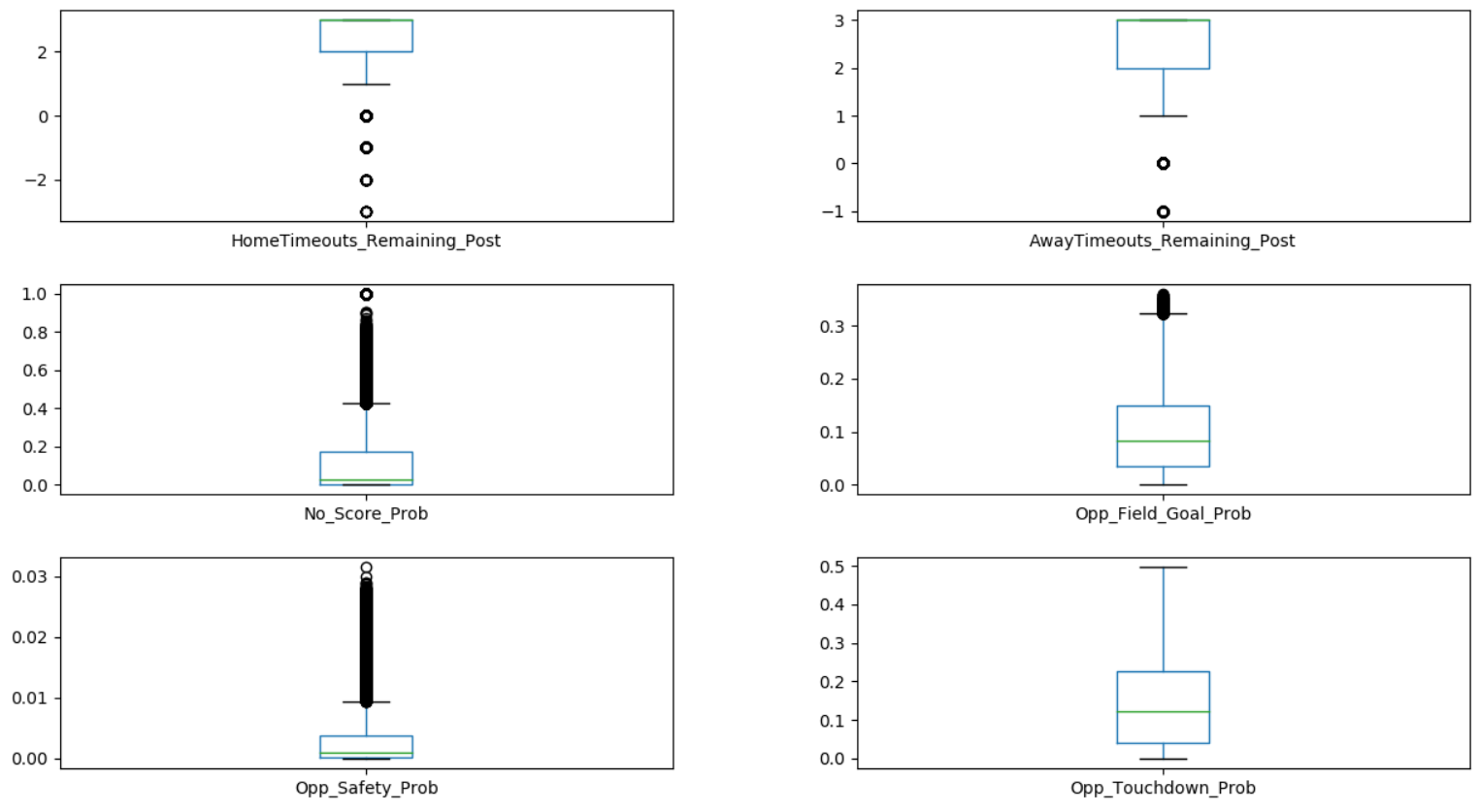


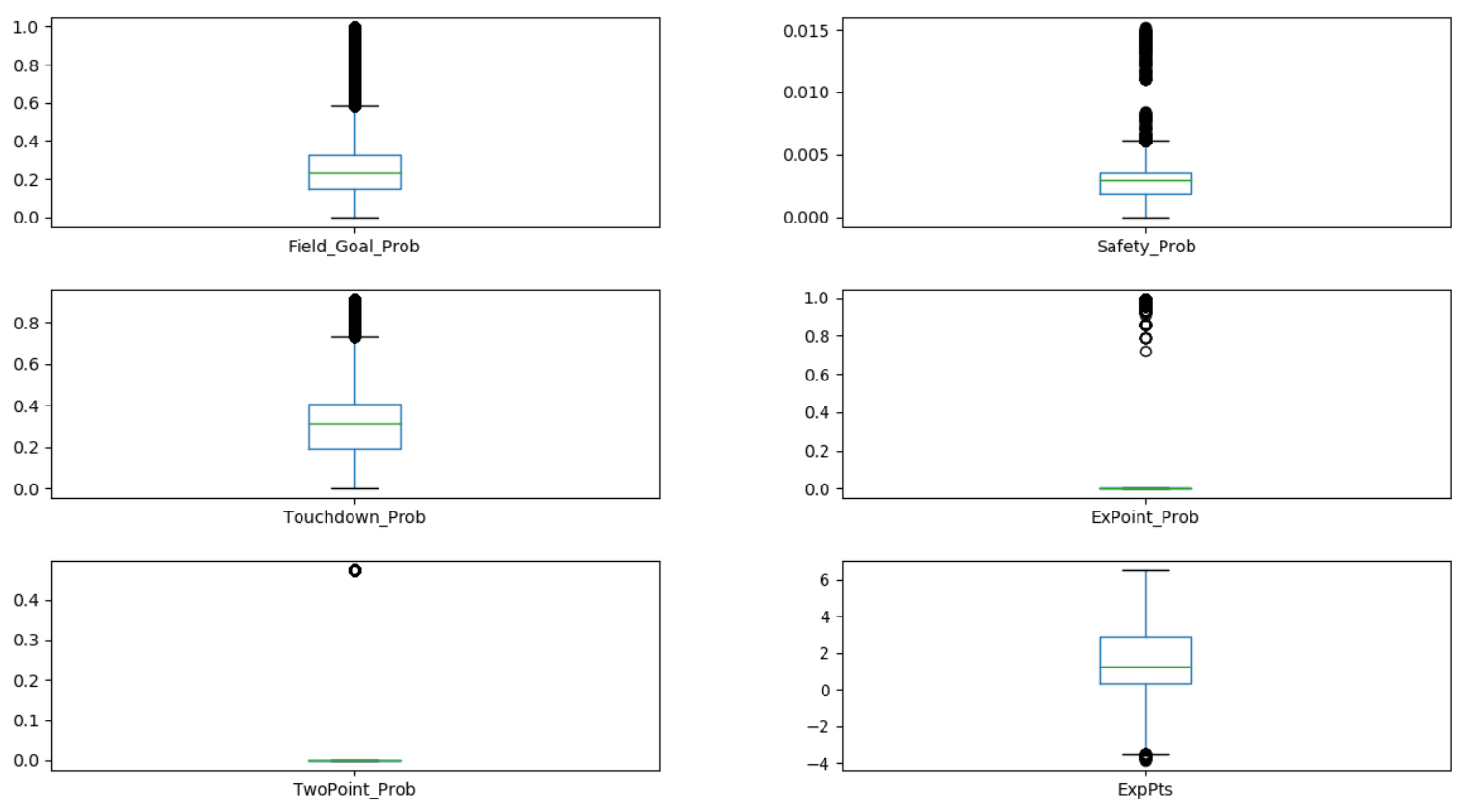
1. 盒图

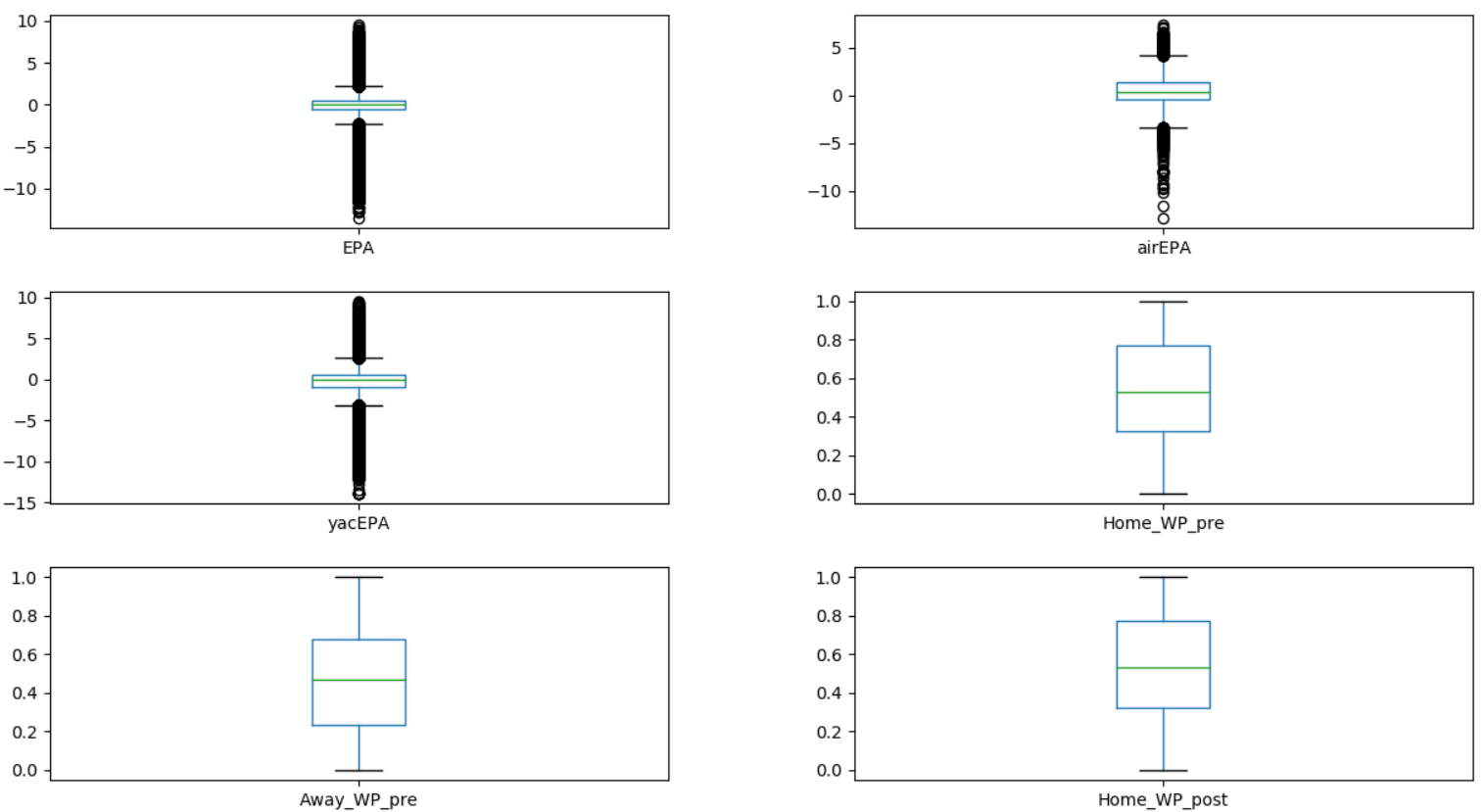


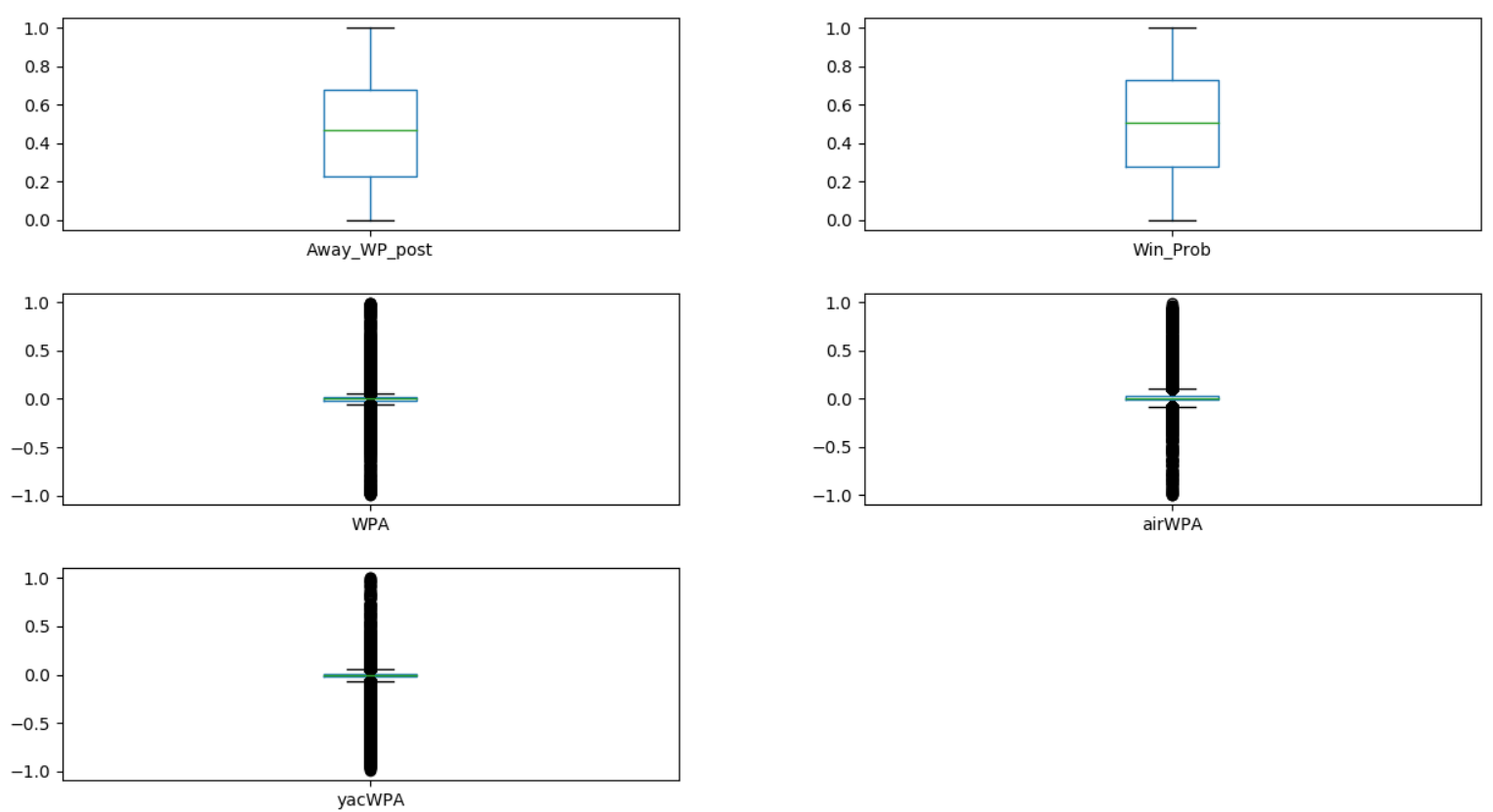








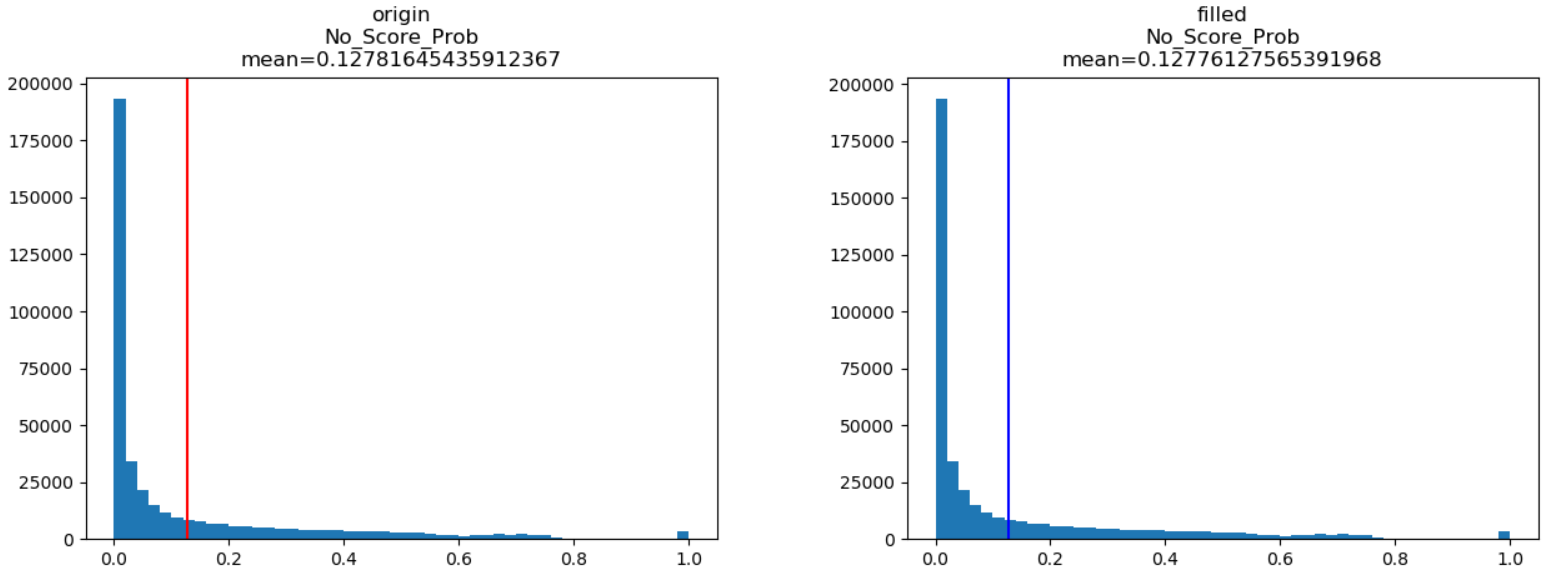


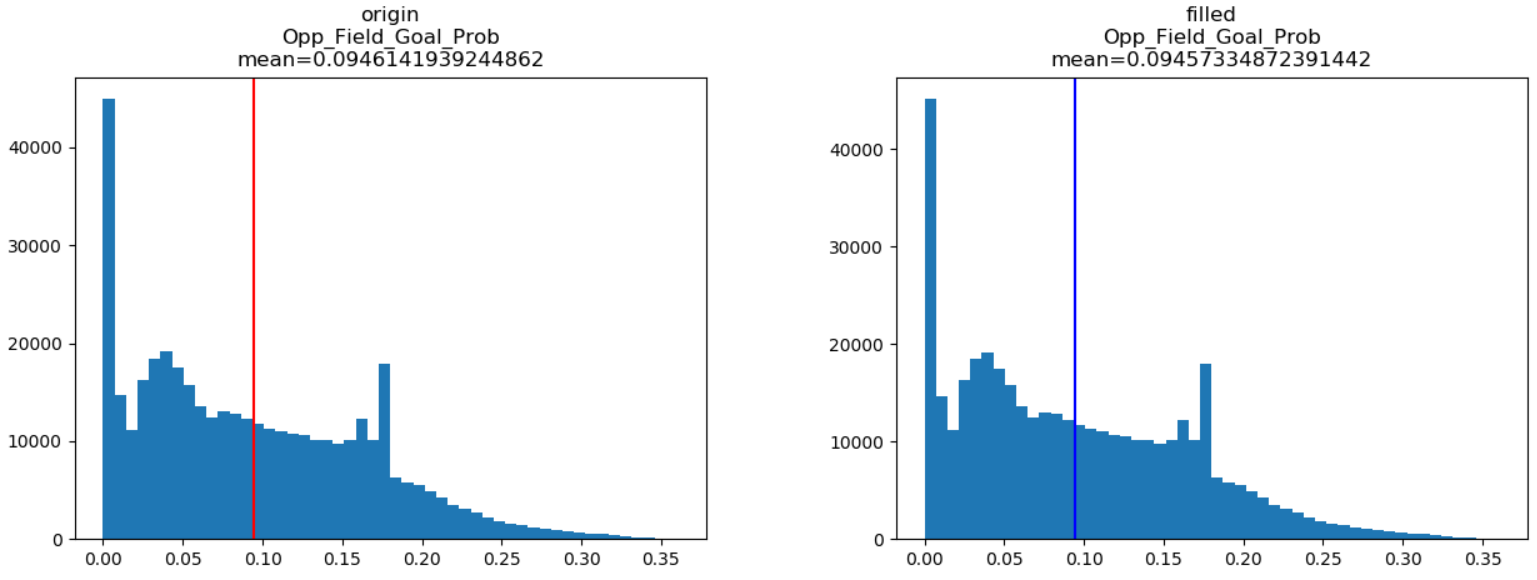


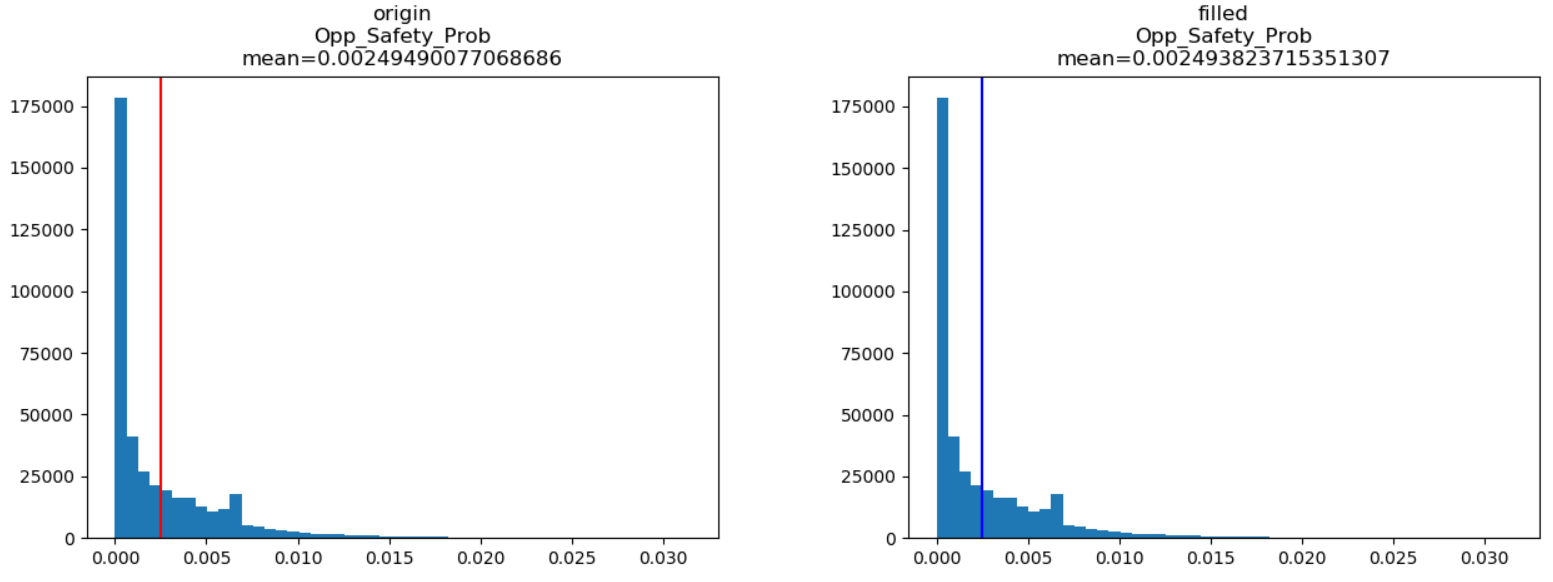
缺失值处理

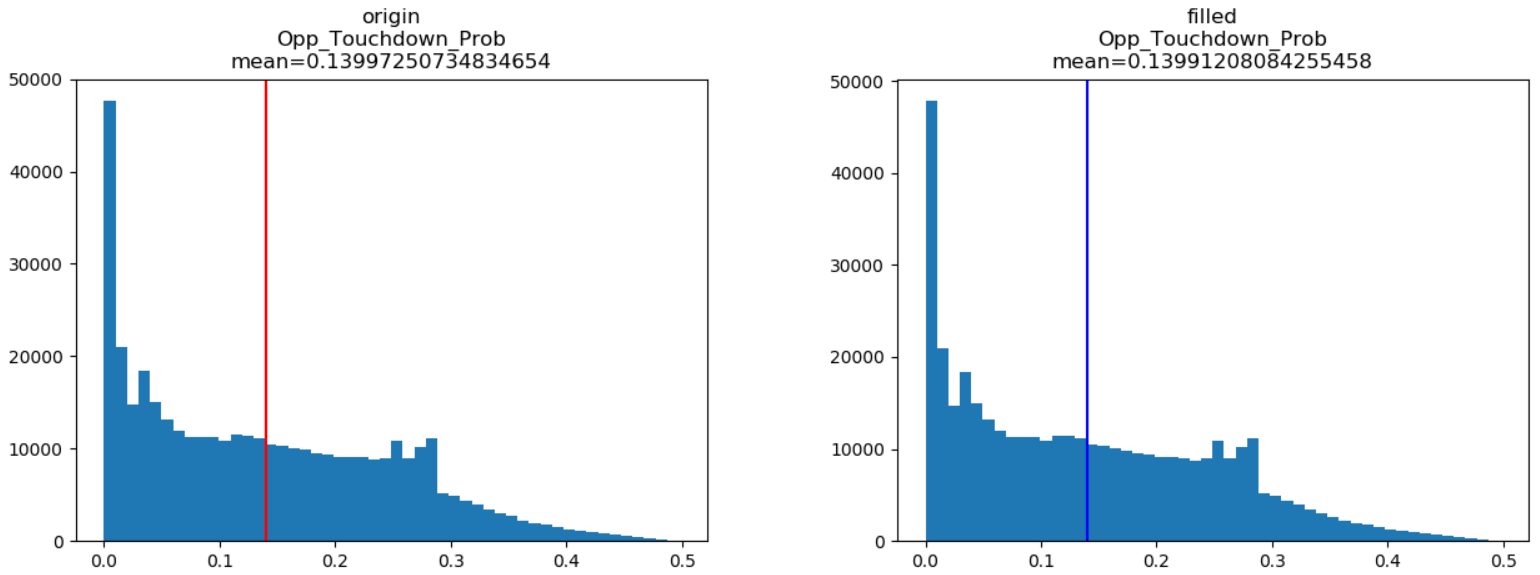
经分析，有填充意义的字段为No\_Score\_Prob，Opp\_Field\_Goal\_Prob，Opp\_Safety\_Prob，Opp\_Touchdown\_Prob，Field\_Goal\_Prob，Safety\_Prob，Touchdown\_Prob，ExPoint\_Prob，TwoPoint\_Prob，ExpPts，EPA，airEPA，yacEPA，Home\_WP\_pre，Away\_WP\_pre，Home\_WP\_post，Away\_WP\_post，Win\_Prob，WPA，airWPA，yacWPA。

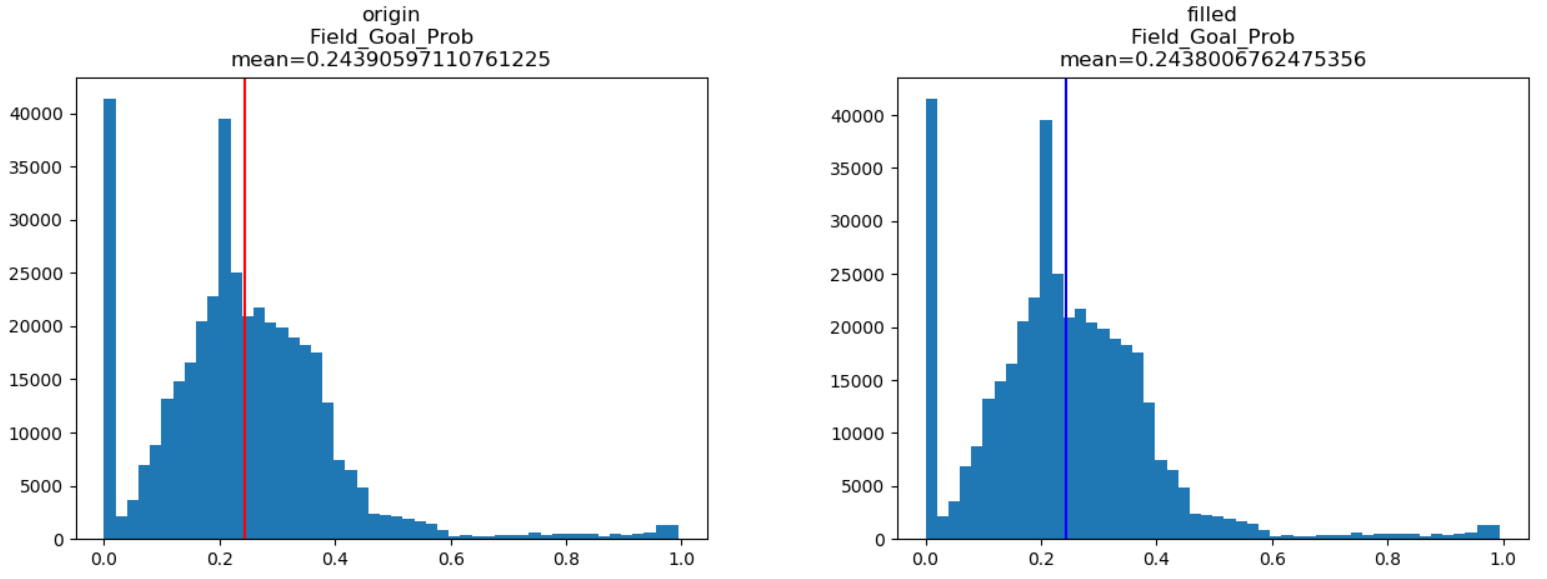
利用最高频率值填充后，效果如下，多数数据在补充之后均值变化不大。

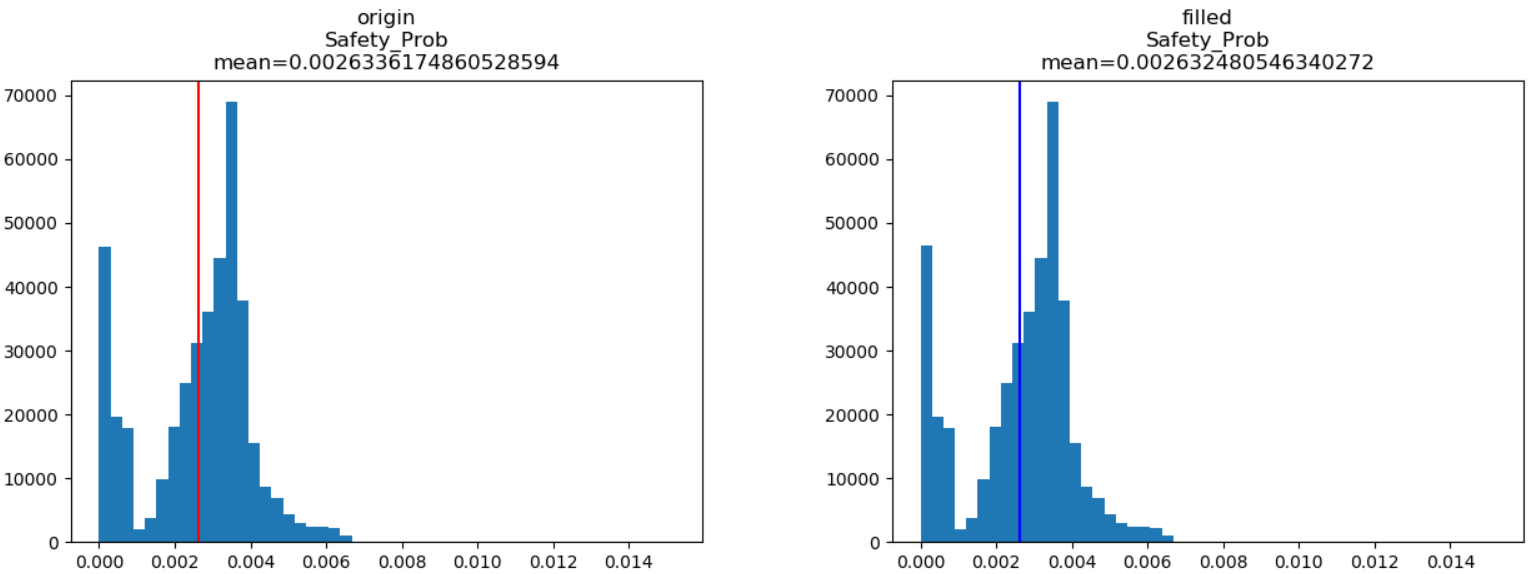


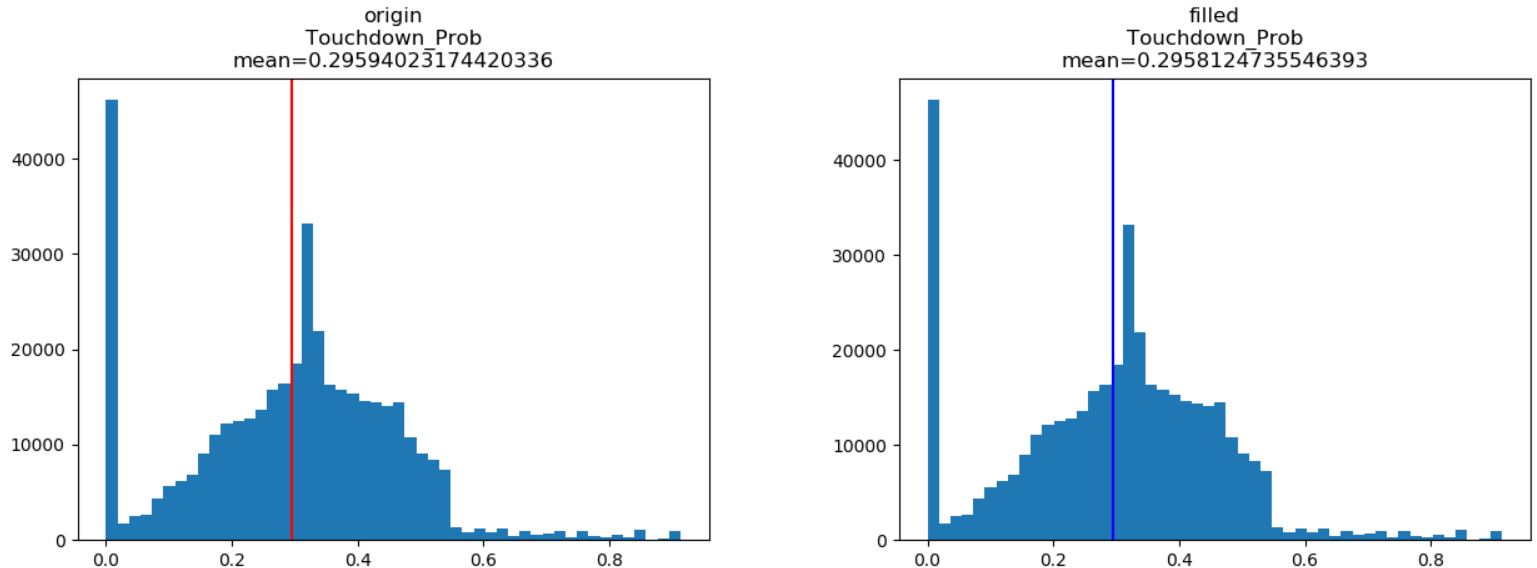


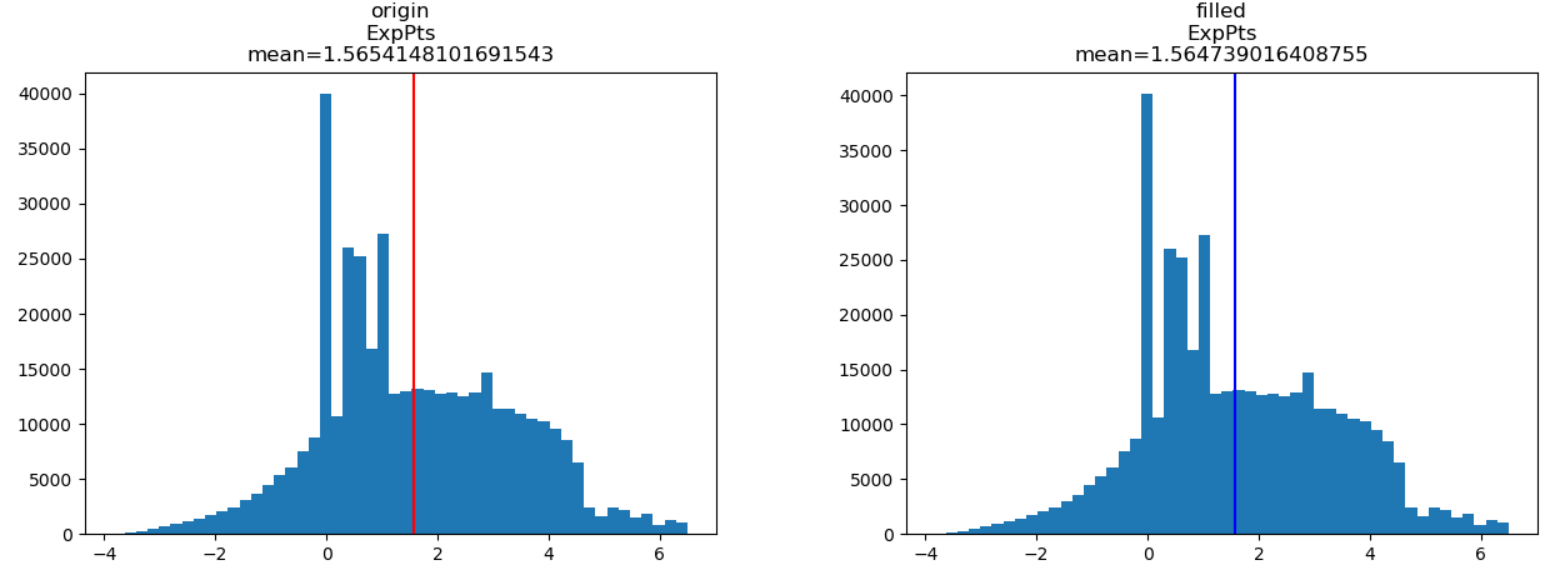


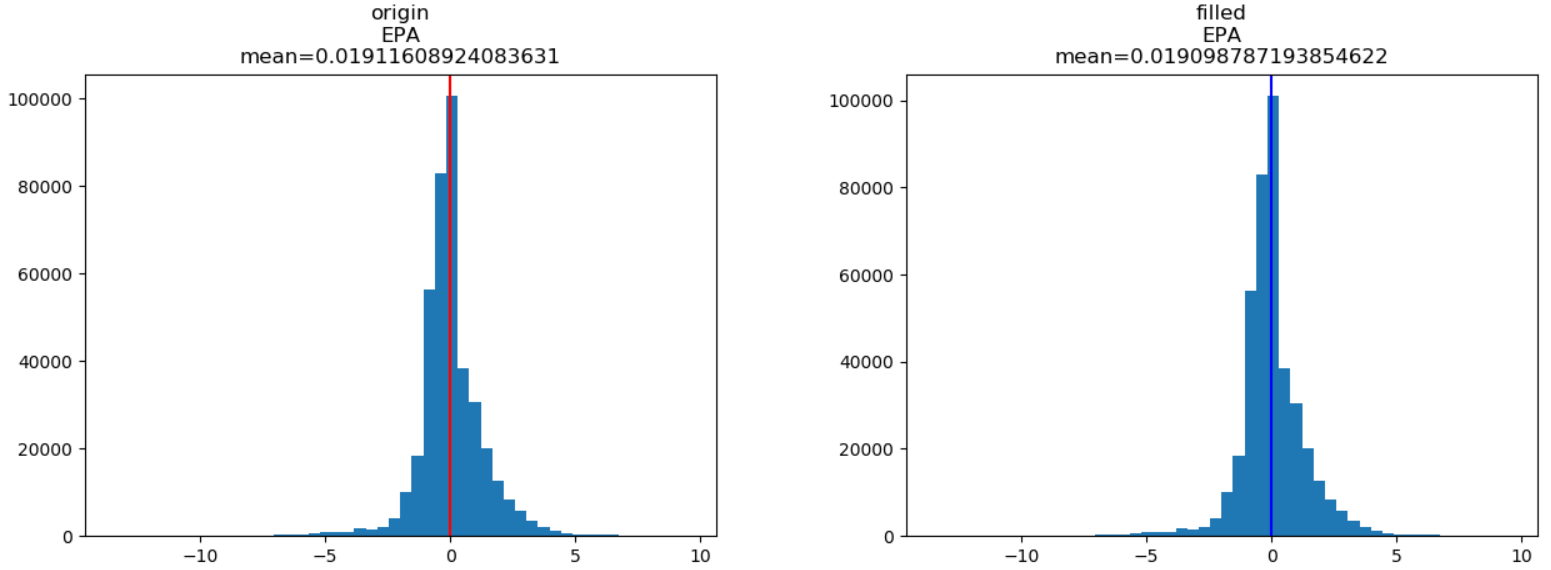












但也有例外，下两图中数据空值太多，所以以此法补充时导致了数据均值的较大变化。

