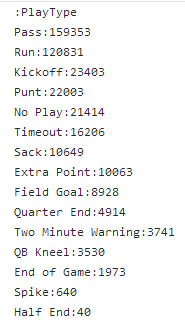
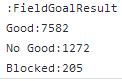
# 作业一 实践报告

#### **3.1 数据可视化和摘要**

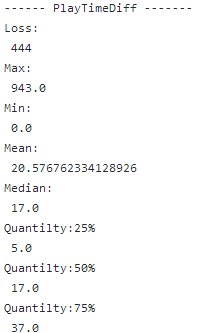
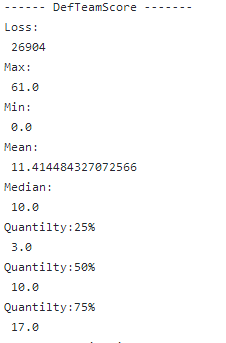
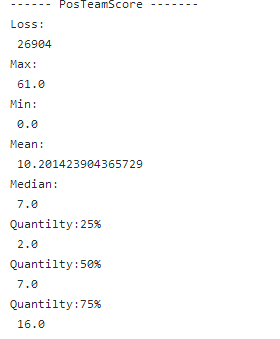
##### **数据摘要**

* 对标称属性，给出每个可能取值的频数，
* 数值属性，给出最大、最小、均值、中位数、四分位数及缺失值的个数。

选取[FieldGoalResult](https://github.com/wh631202311/date-mining-homework/blob/504798ca89b59b605002c9ad7638446babb25b54/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98/3.1%E6%A0%87%E7%A7%B0%E6%95%B0%E6%8D%AE%E5%88%86%E6%9E%90/FieldGoalResult.txt" \o "FieldGoalResult.txt)、[GameID](https://github.com/wh631202311/date-mining-homework/blob/504798ca89b59b605002c9ad7638446babb25b54/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98/3.1%E6%A0%87%E7%A7%B0%E6%95%B0%E6%8D%AE%E5%88%86%E6%9E%90/GameID.txt" \o "GameID.txt)、[PlayType](https://github.com/wh631202311/date-mining-homework/blob/504798ca89b59b605002c9ad7638446babb25b54/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98/3.1%E6%A0%87%E7%A7%B0%E6%95%B0%E6%8D%AE%E5%88%86%E6%9E%90/PlayType.txt" \o "PlayType.txt)分析

（1）标称属性分析结果：

（2）数值属性分析结果：



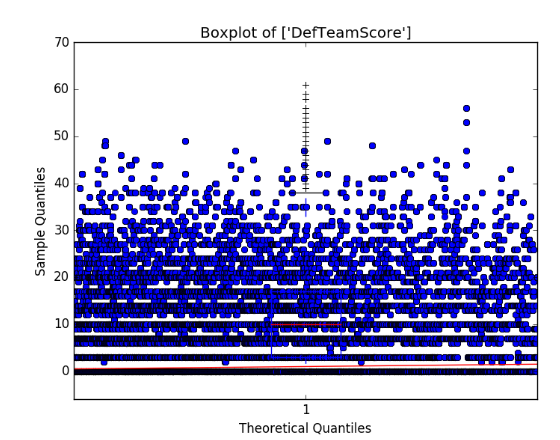
##### **数据的可视化**

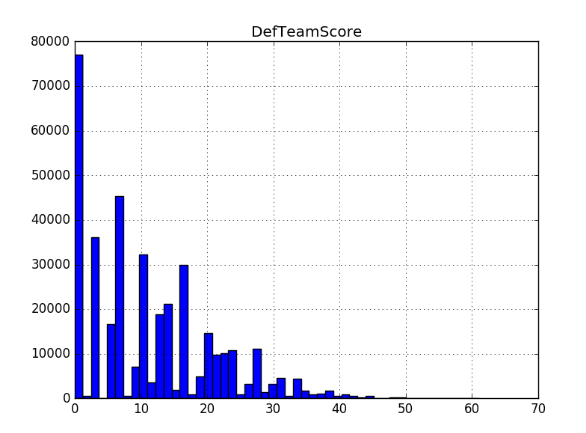
针对数值属性，

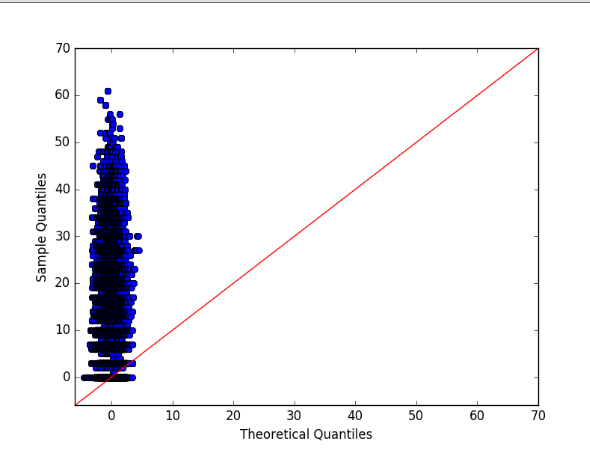
* 绘制直方图，用qq图检验其分布是否为正态分布。
* 绘制盒图，对离群值进行识别

数值属性可视化结果（依次盒图直方图qq图）：

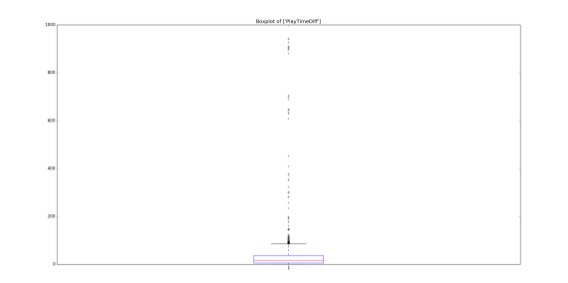
（1）****DefTeamScore****

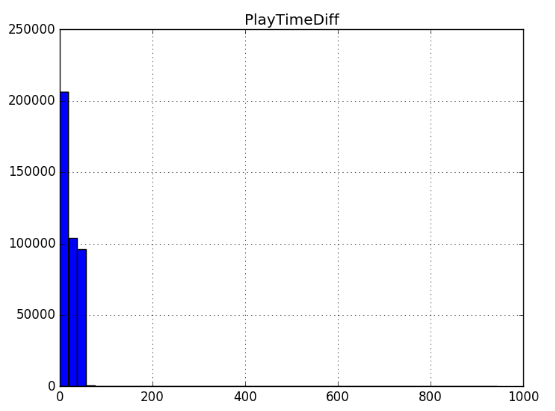


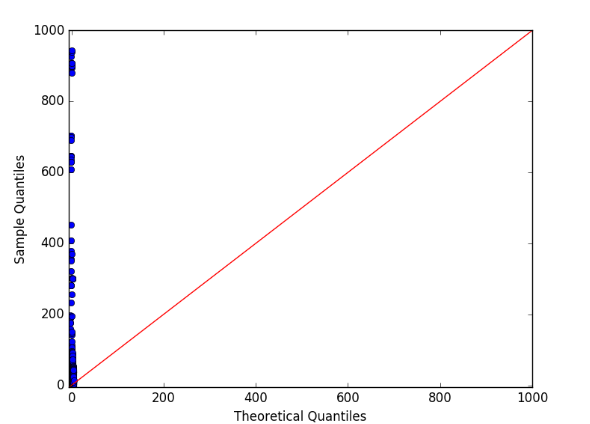




（2）[PlayTimeDiff](https://github.com/wh631202311/date-mining-homework/tree/504798ca89b59b605002c9ad7638446babb25b54/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98/3.3%E5%8F%AF%E8%A7%86%E5%8C%96/PlayTimeDiff)







#### **3.2 数据缺失的处理**

观察数据集中缺失数据，分析其缺失的原因。

可能原因：（1）数据可能被人为因素遗漏

分别使用下列四种策略对缺失值进行处理:

* 将缺失部分剔除
* 用最高频率值来填补缺失值
* 通过属性的相关关系来填补缺失值
* 通过数据对象之间的相似性来填补缺失值

处理后，可视化地对比新旧数据集。

### **4. 提交内容**

* 分析过程的报告
* 分析程序