

近五年 NBA 球队薪资与夺冠的关系

NBA 赛场上一共有 30 支球队，东部和西部各有 15 支。东西部两个分区每个分区常规赛前八的球队可以进入季后赛，争夺当年的总冠军奖杯。本实验的目的是球队的薪资排名是否对球队进入季后赛以及打入东西部决赛争夺总冠军有影响。

首先是要找到近五年每年球队对应的薪资。经过搜索后找到这个网站

<https://hoopshype.com/salaries/2018-2019/>，里面有所需要的信息。在利用 xpath 提取文本的时候出现了文本提取空白的问题。从浏览器中复制的 xpath 路径如下：

`//*[@id="content-container"]/div/div[3]/div[2]/div/div[2]/table/tbody/tr`，但是提取不出来对应的文本。删除到 `//*[@id="content-container"]/div/div[3]/div[2]` 的时候又能提取到信息，说明这里出现了缺失的问题，打开网页的源代码搜索查看后发现源代码缺失了 `div/div[2]` 这段子标签的信息。

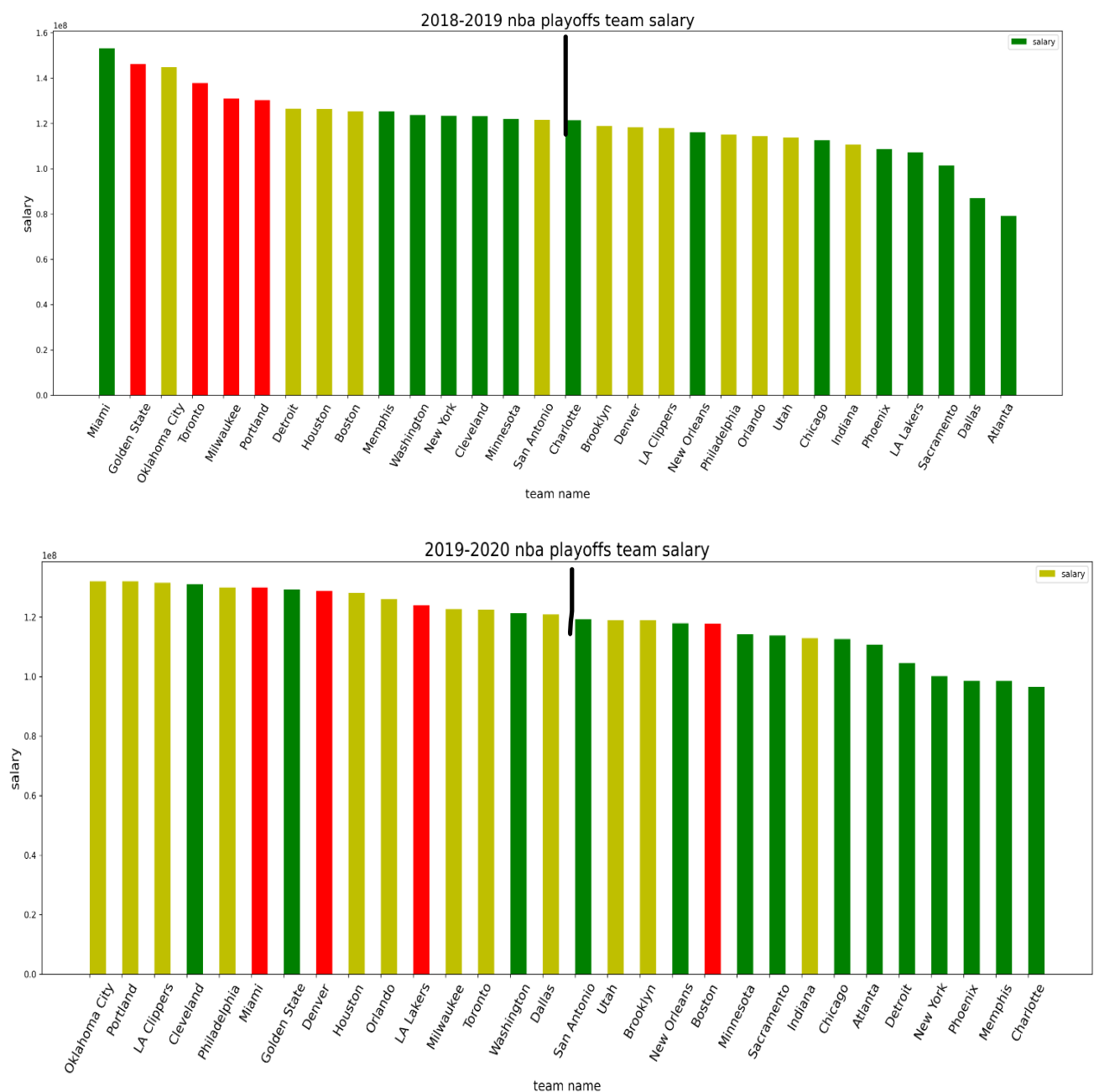
```
<div class="salaries-team-selector-top">⋮</div>
<div class="hh-salaries-ranking">
  <div class="table-wrapper">flex
    <div class="scrollable">⋮</div>
    <div class="pinned">==
      <table class="hh-salaries-ranking-table hh-salaries-table-sortable">
        <thead>⋮</thead>
        <tbody>
          <tr>
            <td class="rank">1.</td>
            <td class="name">⋮</td>
            <td class="hh-salaries-sorted" data-value="153171497" style="display: none;">
              $153,171,497 </td>
            <td class="" data-value="185460477" style="display: none;"> $185,460,477 </td>
          </tr>
          <tr>⋮</tr>
          <tr>⋮</tr>
          <tr>⋮</tr>
          <tr>⋮</tr>
          <tr>⋮</tr>
          <tr>⋮</tr>
        </tbody>
      </table>
    </div>
  </div>
</div>
```

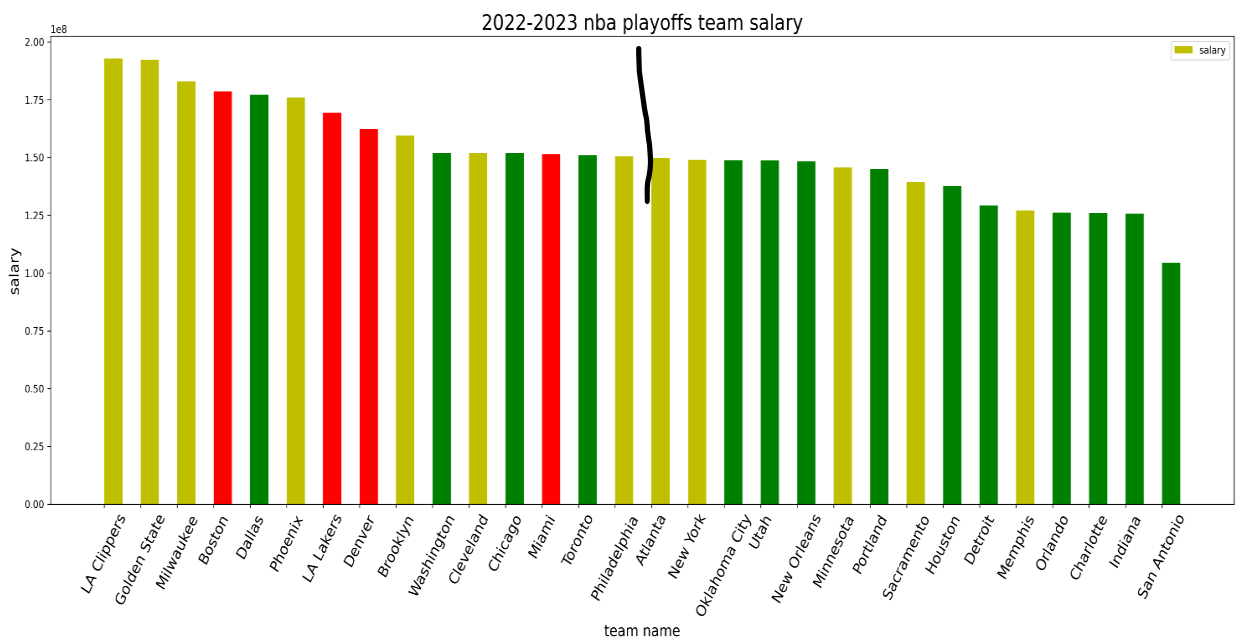
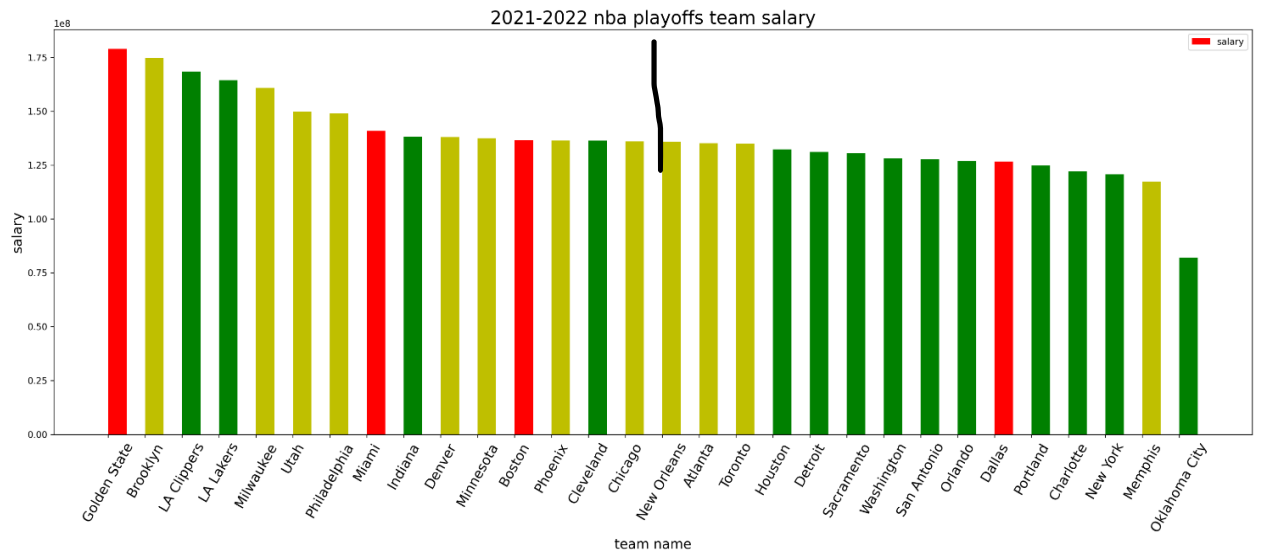
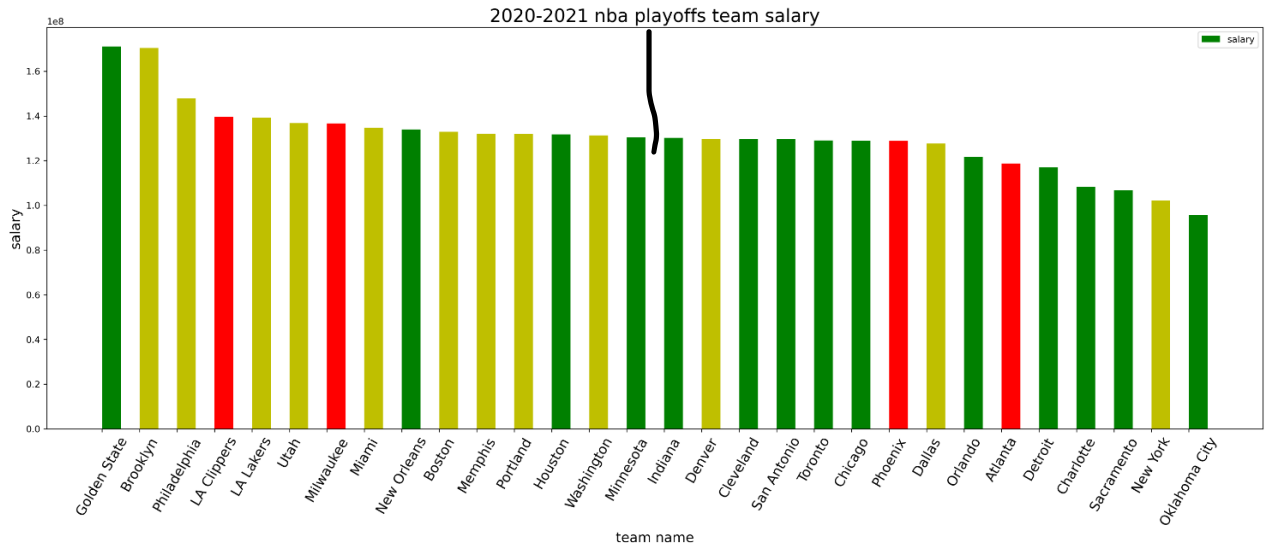
```
</div>
<div class="hh-salaries-ranking">
  <table class="hh-salaries-ranking-table hh-salaries-table-sortable responsive">
    <thead>
      <tr class="table-index">
        <td class="rank"></td>
        <td class="name">Team</td>
        <td class="">2018/19</td>
        <td class="">2018/19(*)</td>
      </tr>
    </thead>
    <tbody>
      <tr>
        <td class="rank">1.</td>
        <td class="name">
          <a href="https://hoopshype.com/salaries/miami_heat/2018-2019/">
            Miami </a>
        </td>
        <td class="" data-value="153171497"></td>
        <td class="" data-value="185460477"></td>
      </tr>
      <tr>
        <td class="rank">2.</td>
        <td class="name">⋮</td>
        <td class="" data-value="⋮"></td>
        <td class="" data-value="⋮"></td>
      </tr>
    </tbody>
  </table>
</div>
```

在删除多余的标签信息后得到正确的 xpath 为：

//*[@id="content-container"]/div/div[3]/div[2]/table/tbody/tr，成功爬取到球队和对应的薪资。由于爬下来的文件被当成一个元素的列表进行储存，并且还有\n\t 符号需要去掉，所以使用强制类型转换的功能将列表转换为字符串之后再用 replace 函数的功能将\n\t 符号去掉得到想要的球队名称及其对应的薪资，最后将其以 csv 文件的格式储存。

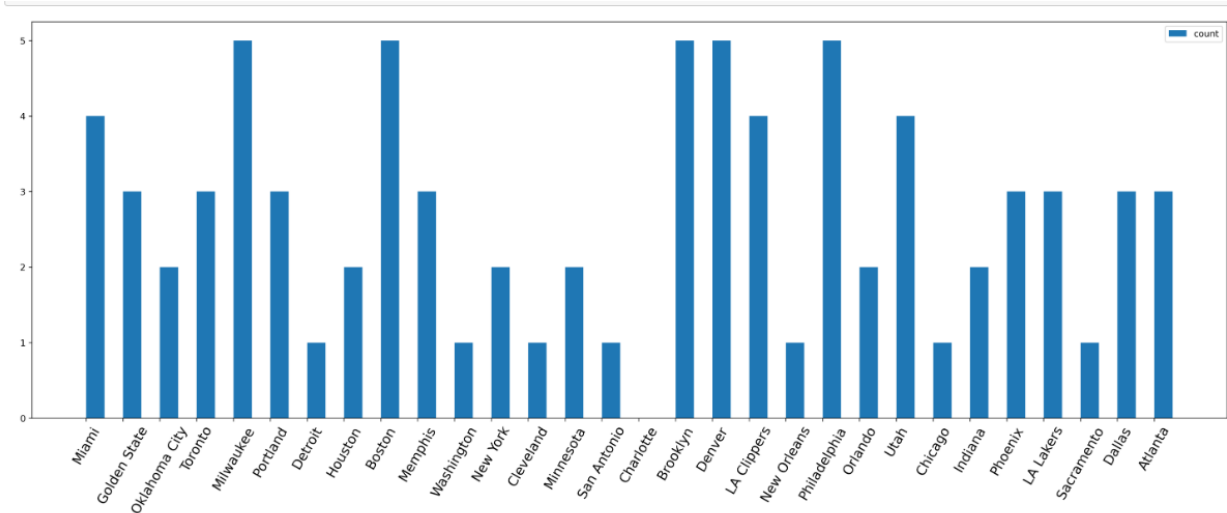
下面是对爬取到的数据进行分析。除了球队对应的薪资以外，我还手动标注了当年哪些球队进入了季后赛以及东西部的决赛，如果是 1 则表示进入，0 表示没有进入。利用 pandas 的 read_csv 函数功能读取文件。再将读取到的文件使用 numpy 将其转换为数组的格式，设置图标的 x 轴坐标为 30 等分。由于球队的名字较长，需要利用 figure 功能将图表的尺寸设置的更大一些，然后再让 x 轴上的标签进行一个角度旋转，就能够看得清每只球队的名字了。在区分球队是否进入季后赛的方法上，我采用了循环赋值的方法，将没有进入季后赛球队的薪资柱子设置为绿色，进入了设置为黄色，进入东西部决赛的设置为了红色来进行更好的区分得到如下的图片。





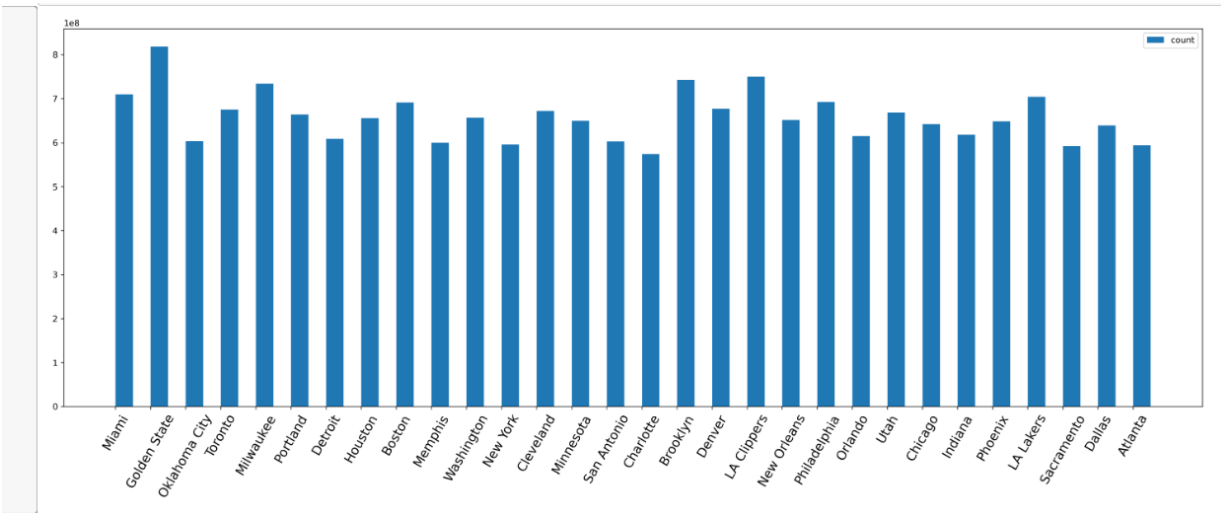
可以从上面的图看到，在黑色线前面为薪资排名前 15 的球队，18-19 年是最多的有 6 支球队没有进入季后赛，而 19-20 年最少，只有 3 只。其余时间都是 4 支球队没有进入季后赛。说明高薪资为球队带来更好的球员以及轮换阵容能让球队在漫长的常规赛中胜出，进入季后赛的概率是比较大的（排除伤病等等意外的干扰，受伤的球员虽然打不了比赛但是薪资是照算的，可能会产生一定的误差）。也可以看到 nba 球队的工资也是一年年的上涨的。再看到进入东西部决赛的球队，我认为能打到这一部分，除了有一部分运气之外更多的还是有实力的。可以看到四支球队有两年都是在前 15 的薪资球队里面，有两年是有一只不在前 15 的球队里面，最后有一年是对半开。说明高薪资带来的巨星能够带领球队在季后赛中走的更远，夺冠的概率也更大。当然也会有爆冷的情况，有两年薪资第一的球队连季后赛都没有进去，这可能就是被伤病所影响，导致球队的战绩不佳。

下面是对五年来进入季后赛球队次数的统计。



由于 NBA 球员的每份合同一般都是签 3 年或者 4 年，通过研究这五年球队进入季后赛的次数，可以看出有些球队签的球员是比较合理的，能够在几年内都打进季后赛。其中东部球队雄鹿和凯尔特人还有热火比较强，他们不但年年打进季后赛（热火差一年），而且还经常打入东部的决赛。说明只要球队还维持当下的阵容大体不变，然后再进行一些补强的话，对总冠军的竞争是很强力的。西部则比较平均，最多的是掘金，每年都进入季后赛且打入西决，说明西部掘金的阵容是比较好的，其他的相差不大，谁都有可能打入西决，说明西部竞争比较强烈。

再看这些球队五年的总薪资。



可以看到勇士队是薪资最高的，高薪资也让勇士在五年内成功打入西决两次。雄鹿热火和凯尔特人相近，都是处于联盟较高的水平，而且这几队也常常打入东决。但是西部薪资结构最好的还是掘金，在薪资处于联盟中上的水平却能年年打入季后赛，可以预见掘金在未来几年仍然是强有力的冠军竞争者。

总结：较高的薪资能为球队带来更好的轮换阵容，让球队在面对伤病干扰的情况下也能有较大概率打进季后赛。而在季后赛，高薪资能为球队带来更强的球星，能够让球队走的更远。同时对平均几年的薪资情况和进入季后赛的次数进行分析，可以得出哪些球队在未来几年仍然是争冠的热点球队，有能力夺得总冠军。