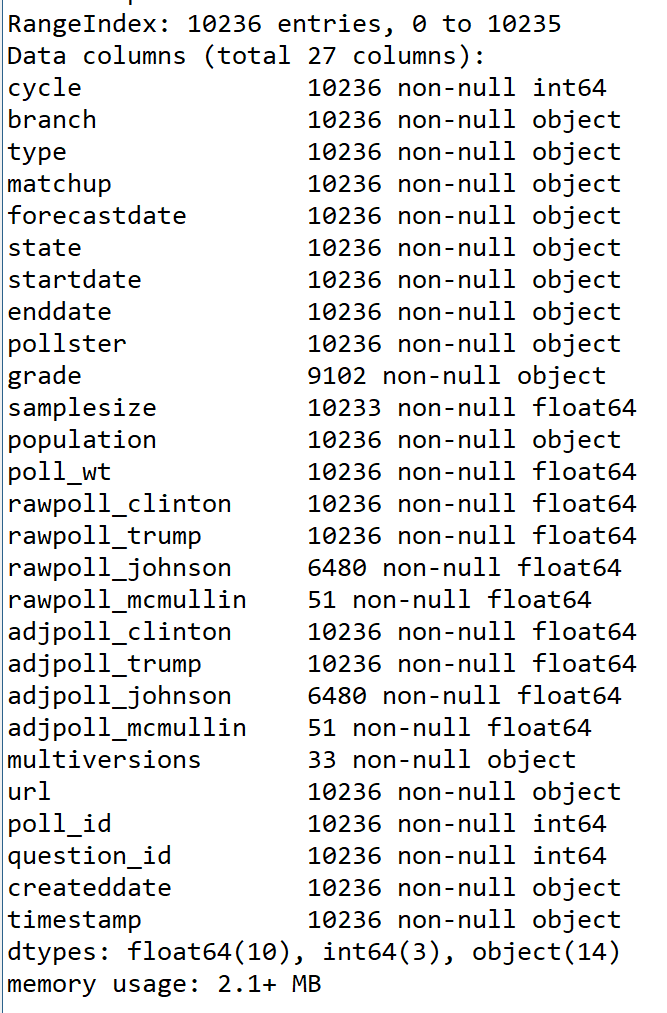
美国大选民意调查趋势分析报告

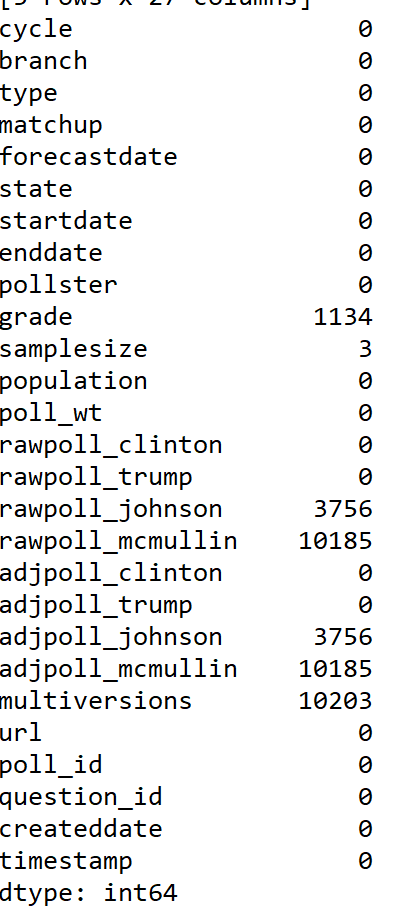
1. 报告背景

2016年美国大选的结果在当时可谓轰动一时，大多数人都认为希拉里最终会赢得大选成为美国首位女总统，然而最终特朗普在这场大选中获胜。在当时人们对此众说纷纭，现在利用python的数据可视化，强大的绘图功能来研究一下这次选举的一些趋势特征。

1. 数据预处理

首先观察一下数据集中所包含的数据。

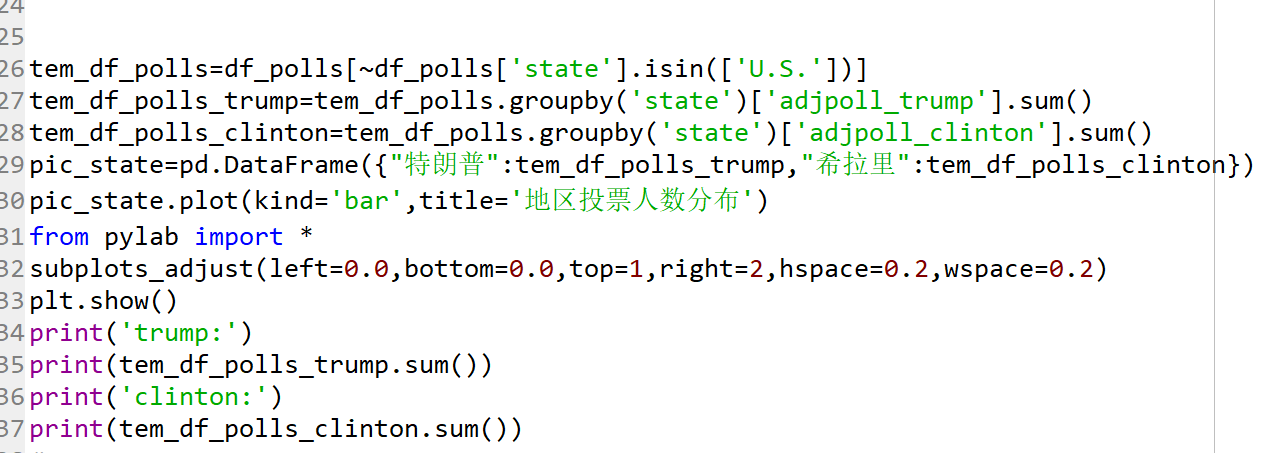


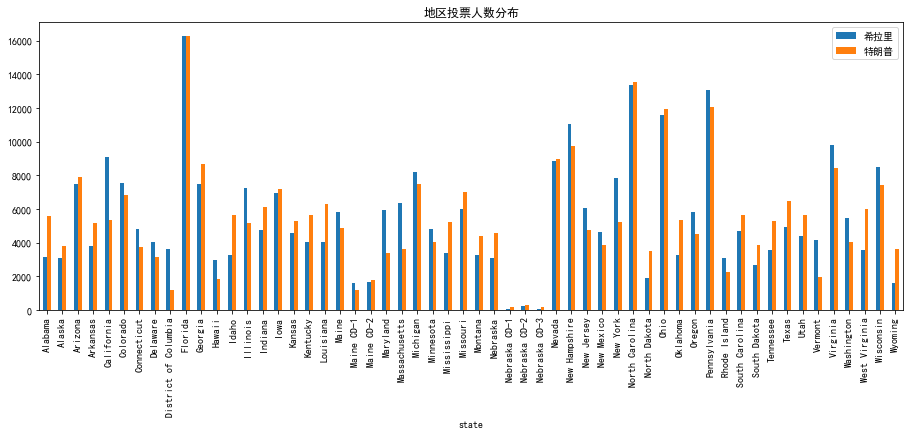


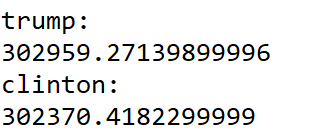
观察值为空的列名，发现johnson和mcmullin的值有很多的缺失，但根据实际结果我们可知，这两位并非此次投票大选的主角，因此不对这些空值进行处理。

1. 绘图分析
2. 地区因素

我们都知道美国有五十个州，而每个州都是一个独立的选区，选区的选民有格给总统候选人们投票。因为每个州的政治倾向可能都不同，所以在分析时考虑地区是非常重要的。



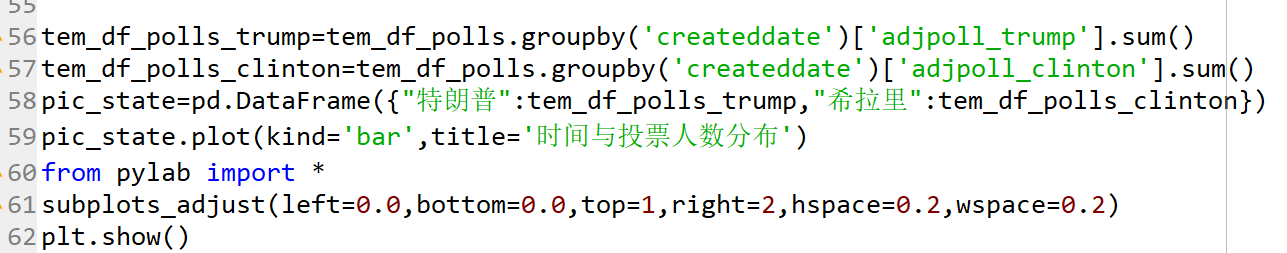


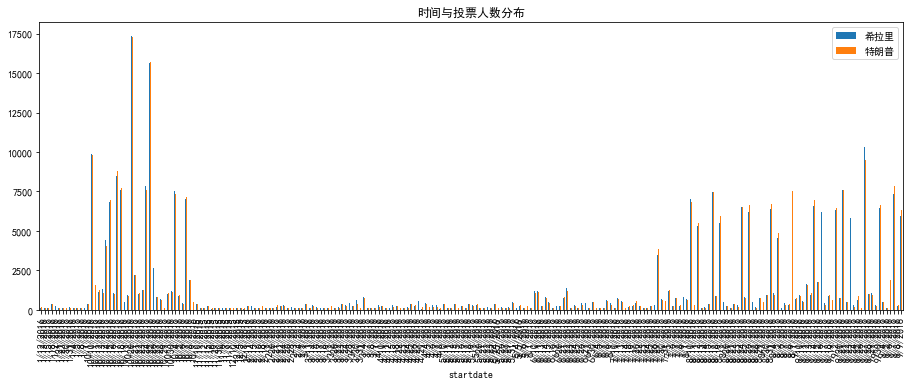


透过上述两幅图我们可以发现，总体的票数大致相符，然而多数地区产生了明显的差异。如东南部的亚拉巴马州，中西部的内布拉斯加州，明显给特朗普投了更多的票；同样的我们也可以看到希拉里在西部的加利福尼亚州，东部的马里兰州有更高的票数。

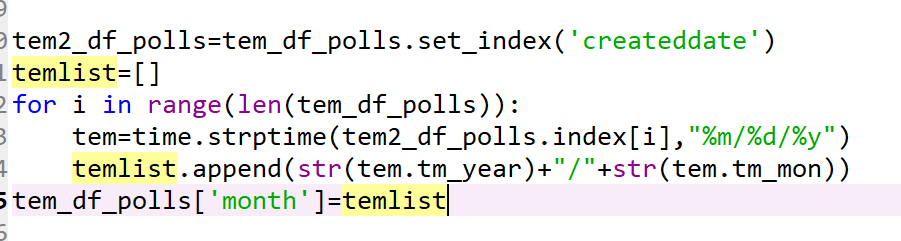
1. 时间因素

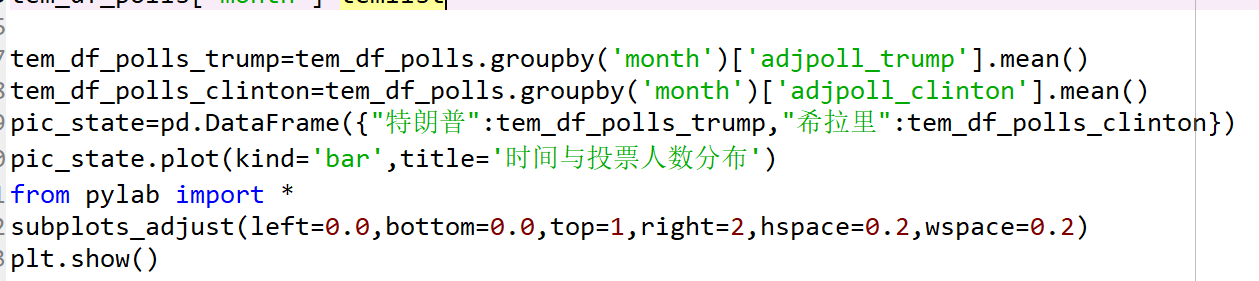
美国的选举是一个长时间持续的过程，在其中候选人们需要出席各式各样的活动，做演讲报告，甚至是公开进行辩论赛。考虑到这点，我们需要将时间也计入考虑因素当中。

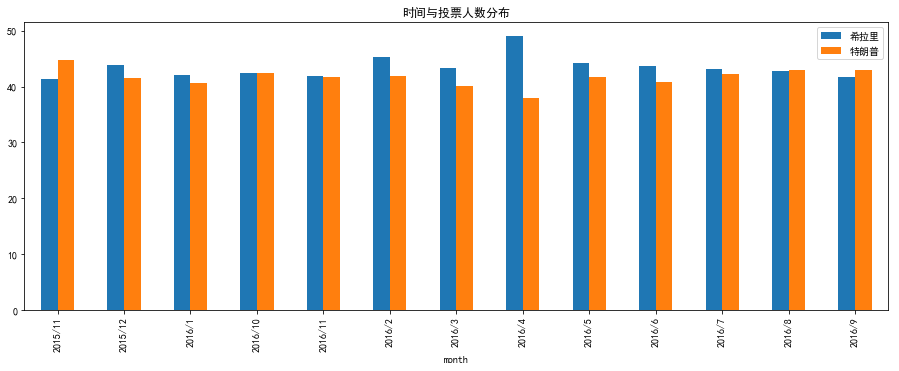




从上图我们观察到，由于数据量过于庞大，时间的分布比较密集，无法直接获得易读的柱形图，需要进一步对时间进行离散化处理。



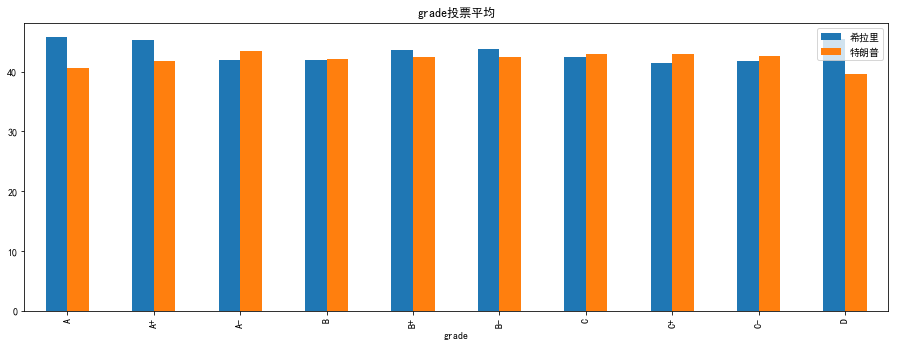


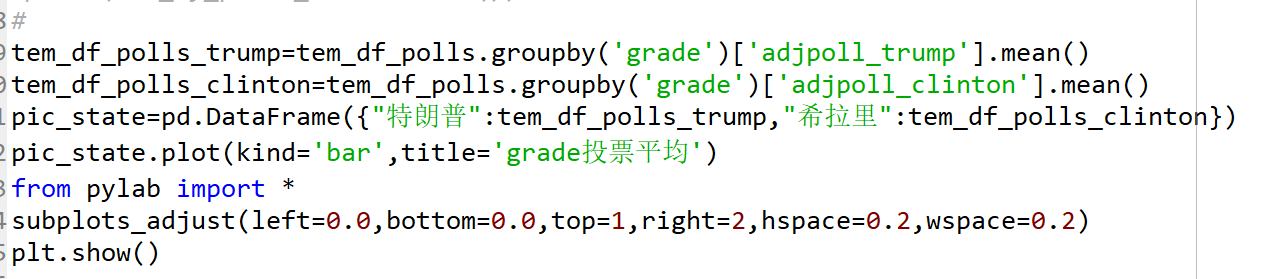


在对时间数据处理过后，我们发现事实上在2015年尾和2016年早期，希拉里一直在这次民意竞选中有一个优势地位，而随着时间的推进，特朗普的数据逐渐向希拉里逼近，甚至在2016年的9月有了明显的反超迹象。

1. 阶级因素

在所获取的数据集中还包含了grade字段，不同社会都有不同的阶级分层，而每个阶级的人都可能会对政治有不同观点。





在此可以发现，中上阶级的民众们普遍看好希拉里，在中下层次两者的情况差不多，但仍然是希拉里要略微领先一点。

1. 小结

在对这次的数据集分析后发现，时间因素尤为重要，在观察了这些数据后，不难明白正是因为当时的竞选过程中希拉里爆出了丑闻，所以才导致了特朗普的横空出世。而特朗普对选民承诺的解决失业问题，也确实让他赢得了中下阶层美国民众的心。