**FIFA 的探索性数据分析及可视化展示报告**

**By 吴挺锋 胡子航**

**一．实验工具及资料：**

**Jupyter notebook （数据清洗和预处理）**

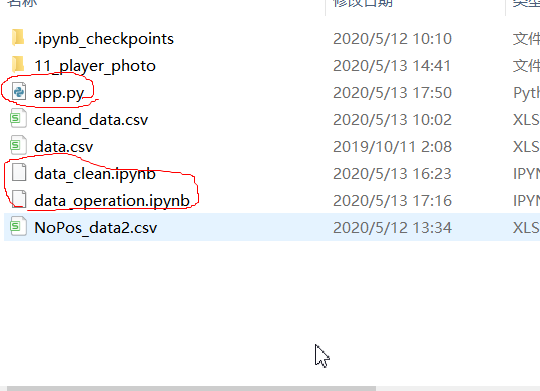
**Plotly （协助在notebook上完成有趣的可视化）**

**Dash （交互式web页面的开发，可视化的移植）**

**需要pip install dash**

**FIFA 19 complete player dataset数据集**

1. **文件目录介绍**



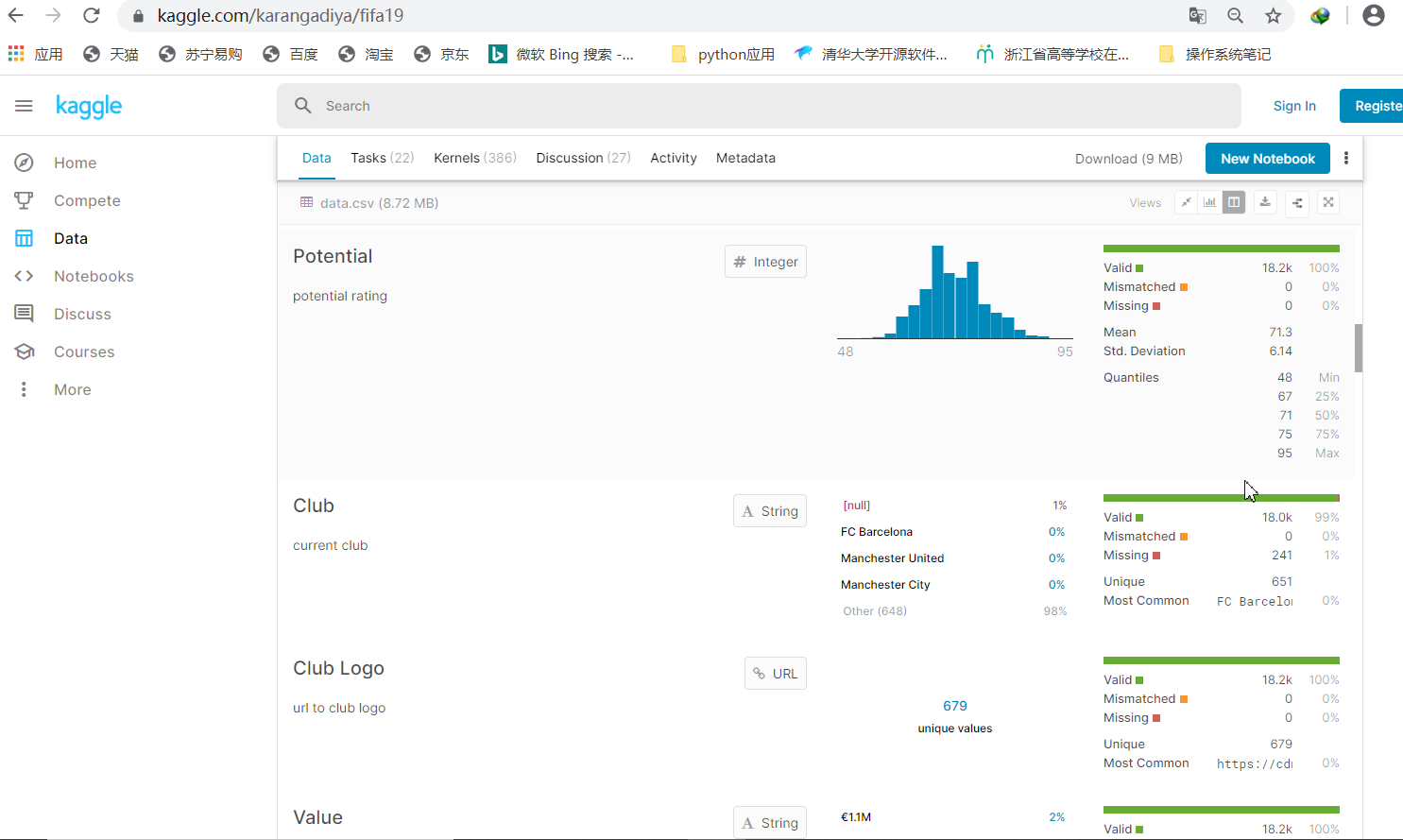
**app.py 是可视化网页的运行文件**

**data\_clean和data\_operation分别是数据清洗和数据可视化**

**（在notebook上运行）**

**三．实验步骤**

**1）首先是观察数据，通过pandas的describe（）方法，我们可以看到各列的缺失值和取值情况。或者我们也可以在kaggles的这个数据集下查看分析**



**由于球员间存在重名情况，俱乐部为空也可说明该球员暂时是自由球员等等。所以去重和填空操作都是不符合逻辑的，并且这类数据在球员能力分析中并不占据分量。**

**2)开始对原数据进行清洗和预处理**

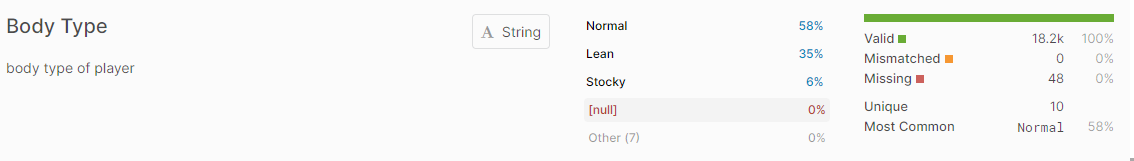


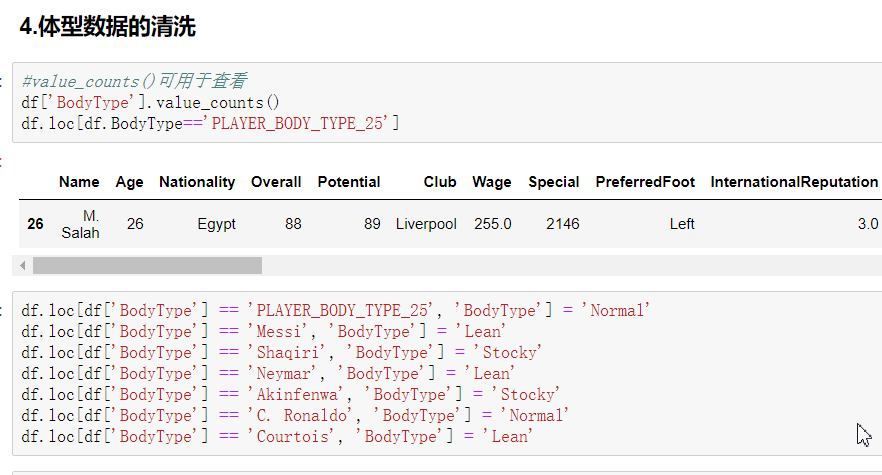
**由于Value\Wage\height等数据的内容存在特殊字符，不方便我们比较，故统一单位**





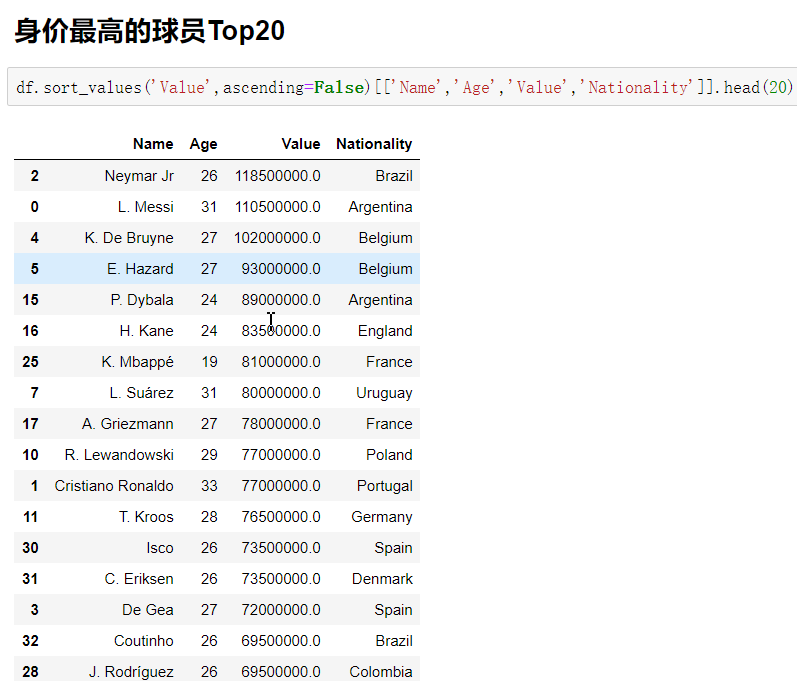
Body Type这一列显然只有三个值，找出出错点进行更正



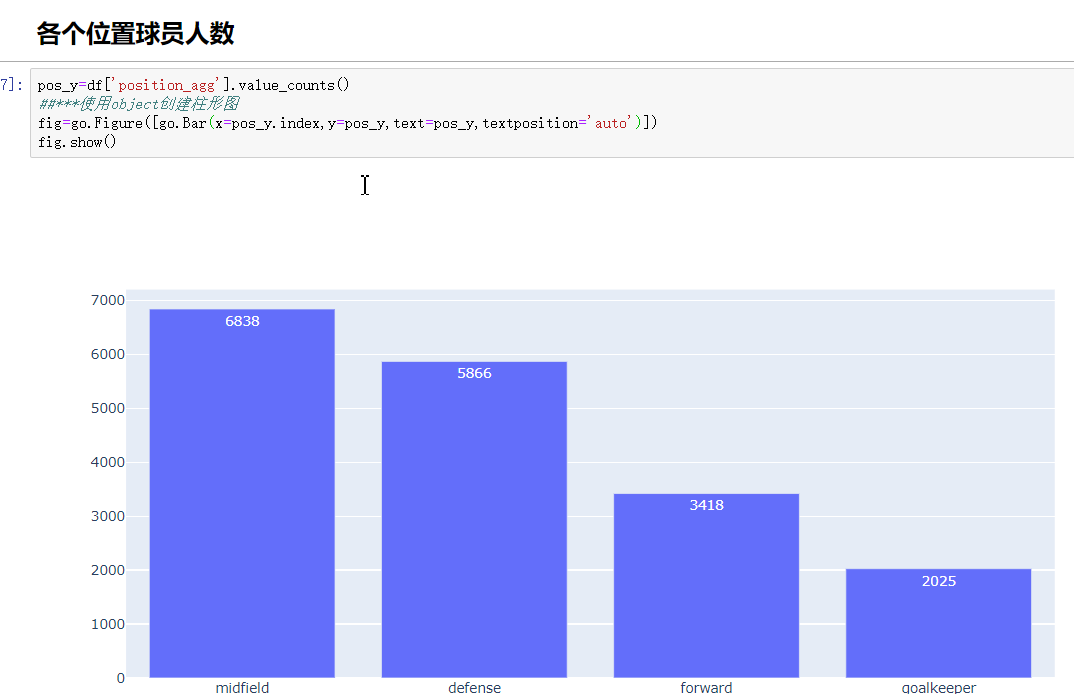


**3）数据处理完毕，做一些可视化工作**

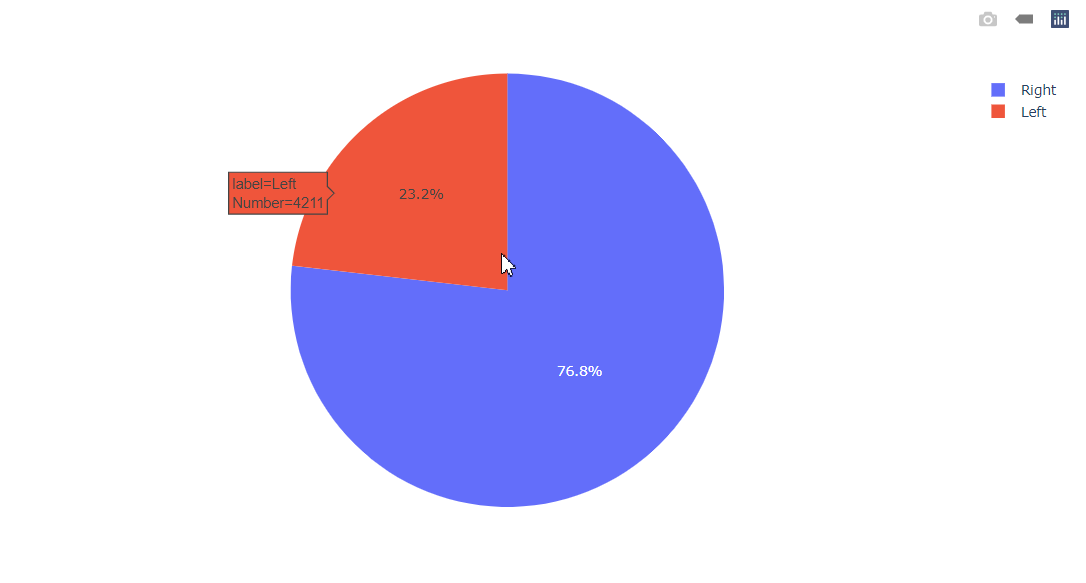
**比如：**



**各位置的球员人数（以下皆借助plotly模块）**



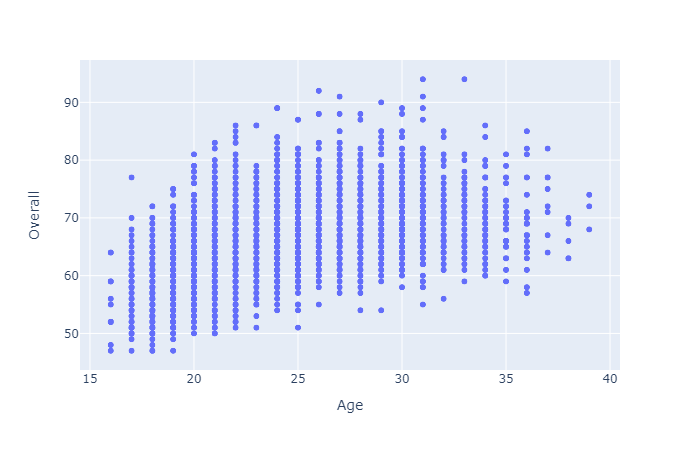
**左右脚的比例等等**



**4）选出最强十一人**

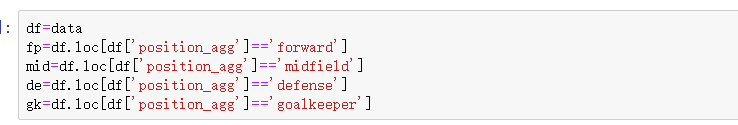
我们的着重任务是找出与球员能力数值高度相关的属性，并通过这些属性的观察来找出我们要的最强球员。怎么样判断两个特征是否相关呢。我们先看一个小问题。

年龄对前锋的评分是否有影响？



从上图我们可以看到，是有的，我们不难发现，存在向下的抛物线的痕迹。

于是乎，我们开始研究那些具体能力值与各位置球员的关系（不同位置的球员分别需要哪些不同的关键属性呢）

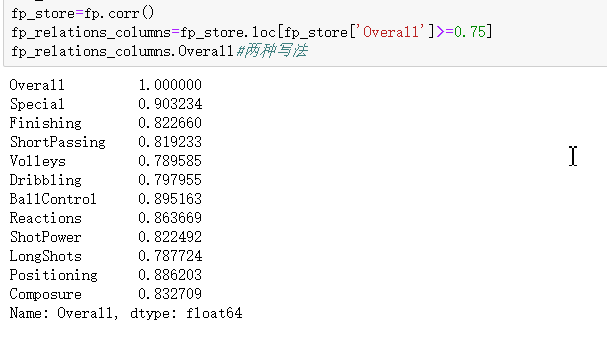


**大方向**

我们将数据分成前锋、中场、后卫以及门将。每一个大方向单独去求出他们的关键属性

比如前锋的关键属性（去计算各列与Overall：评分的相关系数）

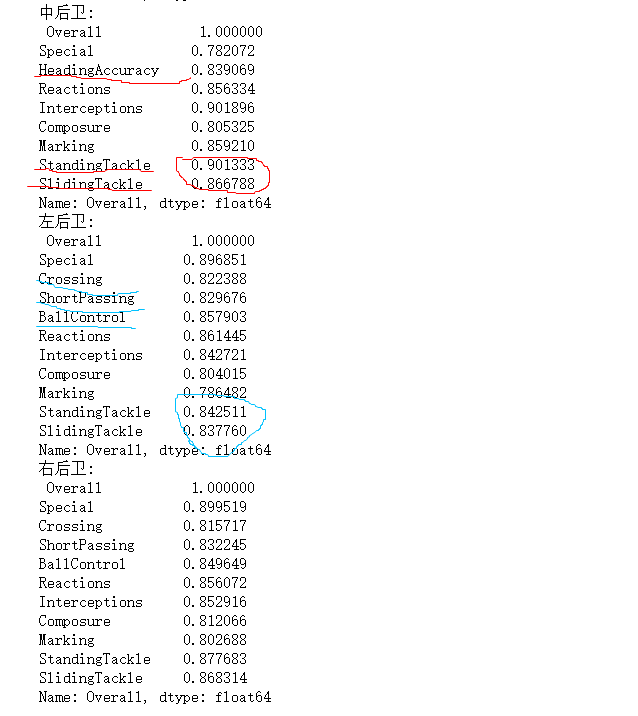
使用dafaframe的内置函数corr（默认是皮尔森相关系数），适用于我们的连续数据。



我们发现Finshing：射术，shortpassing：短传 ，ballcontrol：控球 这些值十分高。也正是说明了他们与Overall存在高度的线性相关关系。因此在选取各位置的最佳球员时尤为关注这些属性。当然这是按照大的位置去分的，如果进入到细分的位置，又会有什么不同呢？

**细分位置**

比如后卫就有：中后卫，左后卫，右后卫等等。他们的关键属性会有什么不同吗



从上图中我们可以发现后卫都有Marking：盯人，Tackle：拦截铲球的关键属性。但是有些不一样，比如中后卫更需要好的头球能力，以及拦截、铲球的能力。

而边后卫在撑起边路进攻时则需要好的传中能力，和拿球（控球）的能力。这就是两者的异同。

以下是我选出的最强十一人名单（挑选规则：在所属位置的关键属性中处在Top级别）

我是按照4132的阵型选出最佳阵容。

前场（一中锋（ST），一影子前锋（CF））

中场（三中场加单后腰）

后卫（两中加两边）

门将

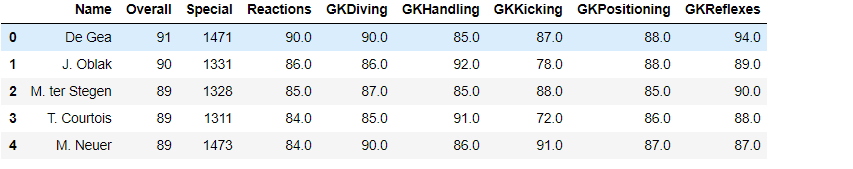
①前锋篇

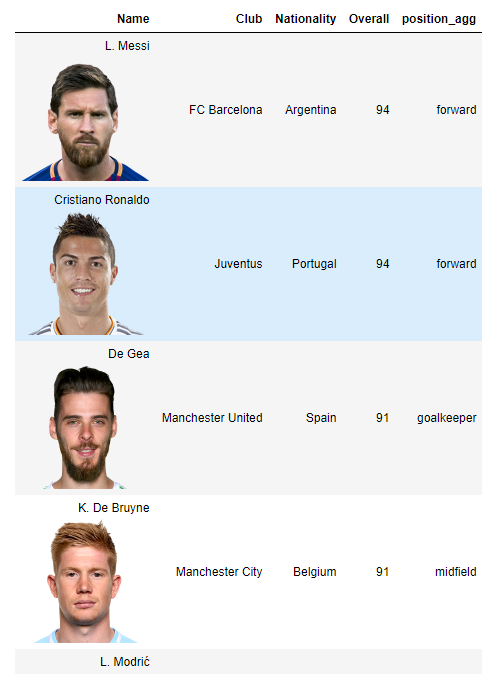
影子前锋的在传统射门的能力要求上，对持球能力（控球、盘带等），传球（传中、短传）数据上更为看重。梅西不论是传球还是持球，以及杀死比赛的能力都处于top，是最佳的策应手。

中锋的选择上是C罗，相比其他中锋，在头球能力毫不逊色的情况下，射术及射门力量等属性更是是Top。

②门将篇

毫无疑问是大卫德赫亚，有三项数据高于第二名，一项持平





③中场篇

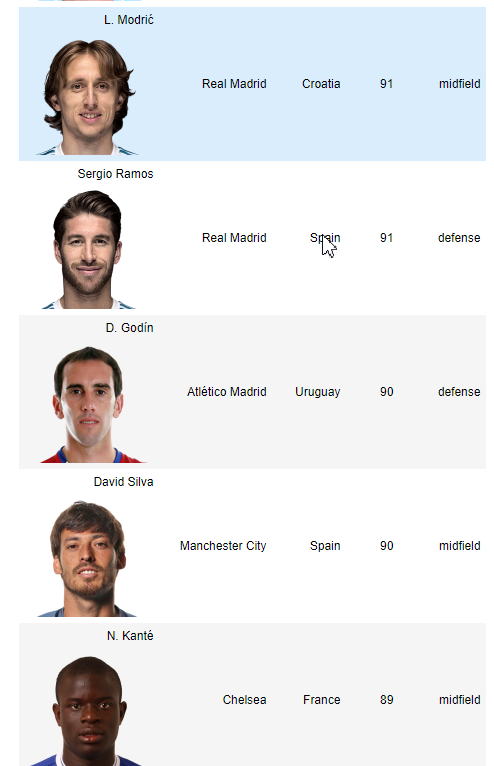
左中场大卫席尔瓦，中场莫德里奇，右中场德布劳内

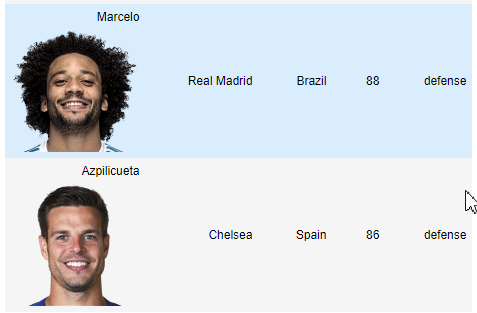
单后腰坎特 持球能力一流，可盘带可过人。攻守兼备

④后卫篇

拉莫斯和戈丁搭档双中卫，他们有最好的拦截和头球能力，盯人、站位和冷静都是顶级

左后卫马塞洛，右后卫阿兹皮利奎塔



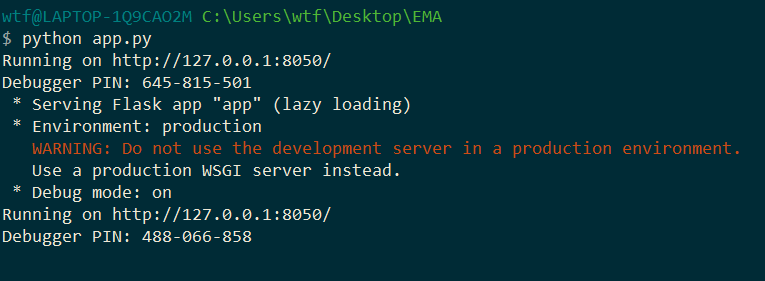


1. **交互式web页面展示（基于Dash）**

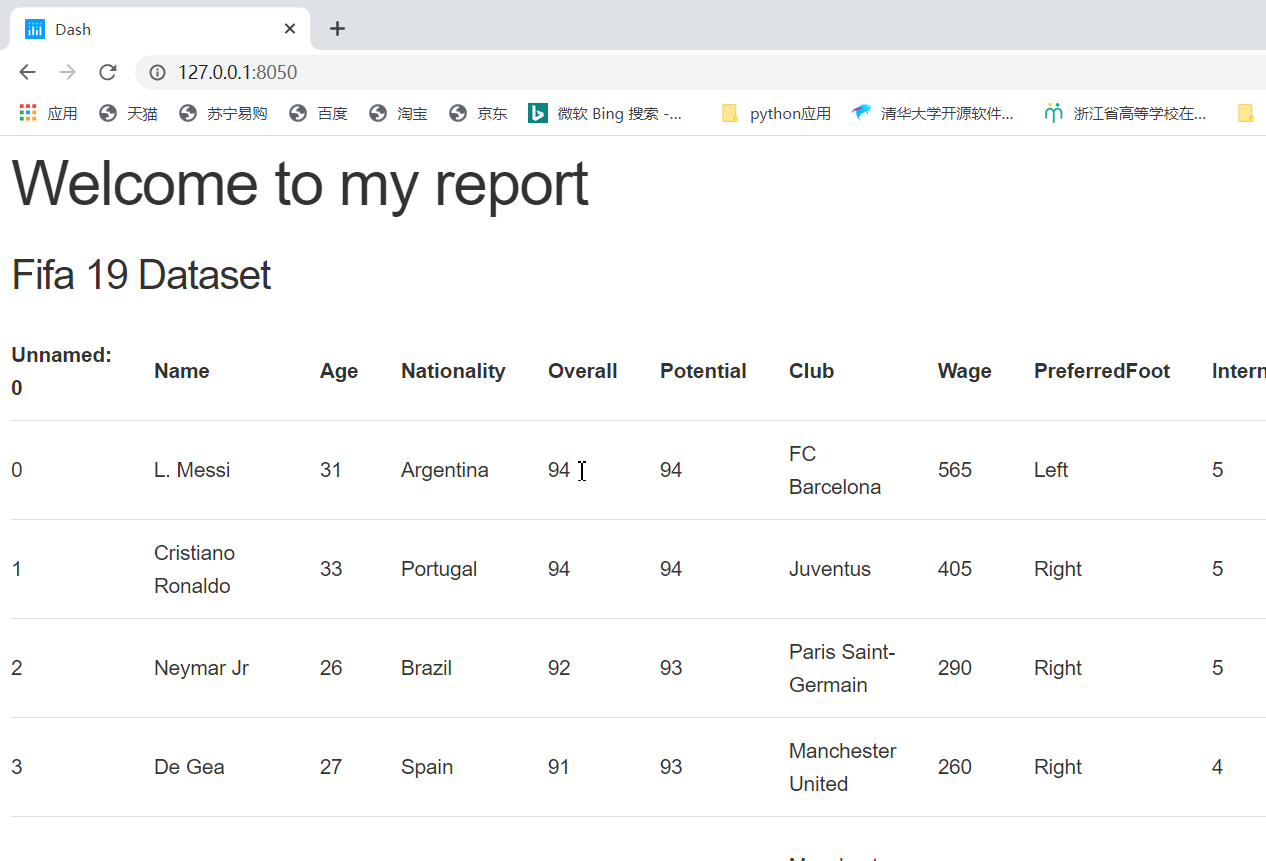
打开方式

①首先pip install dash

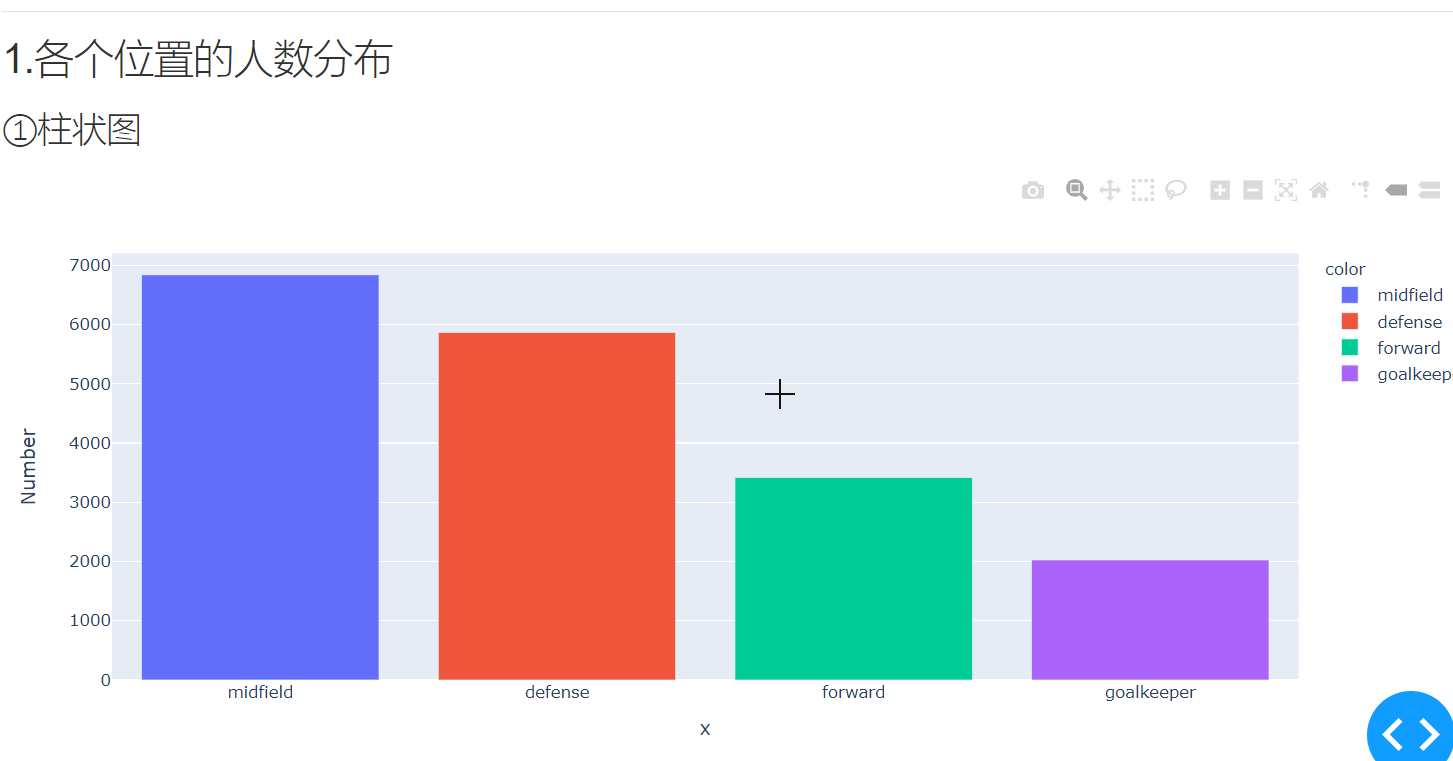
②然后切换到EDA文件目录下python app.py

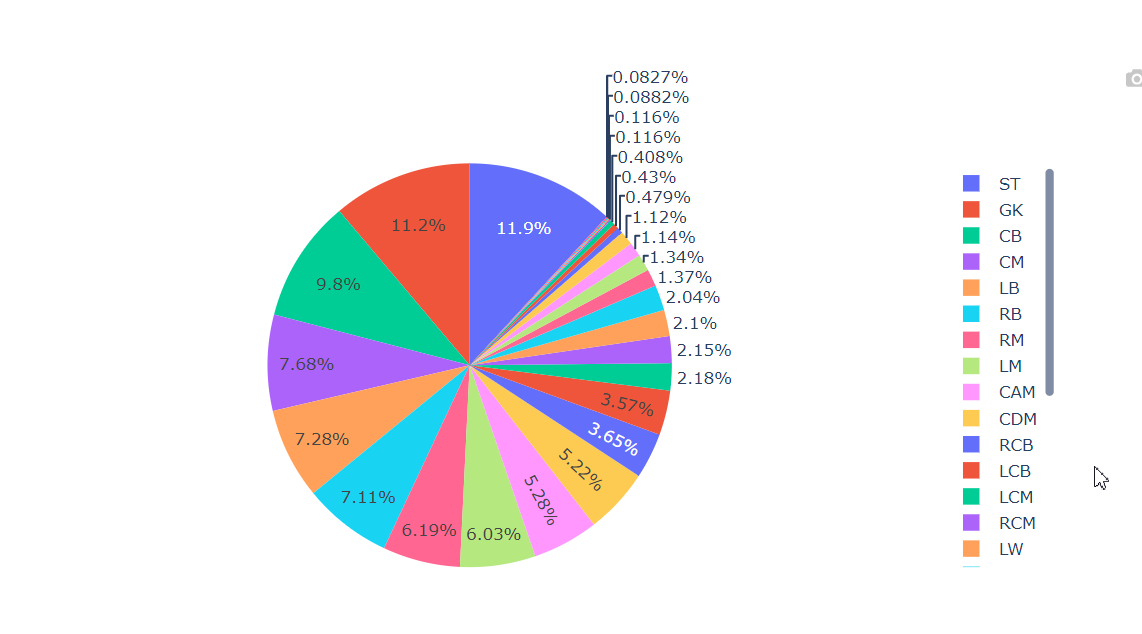


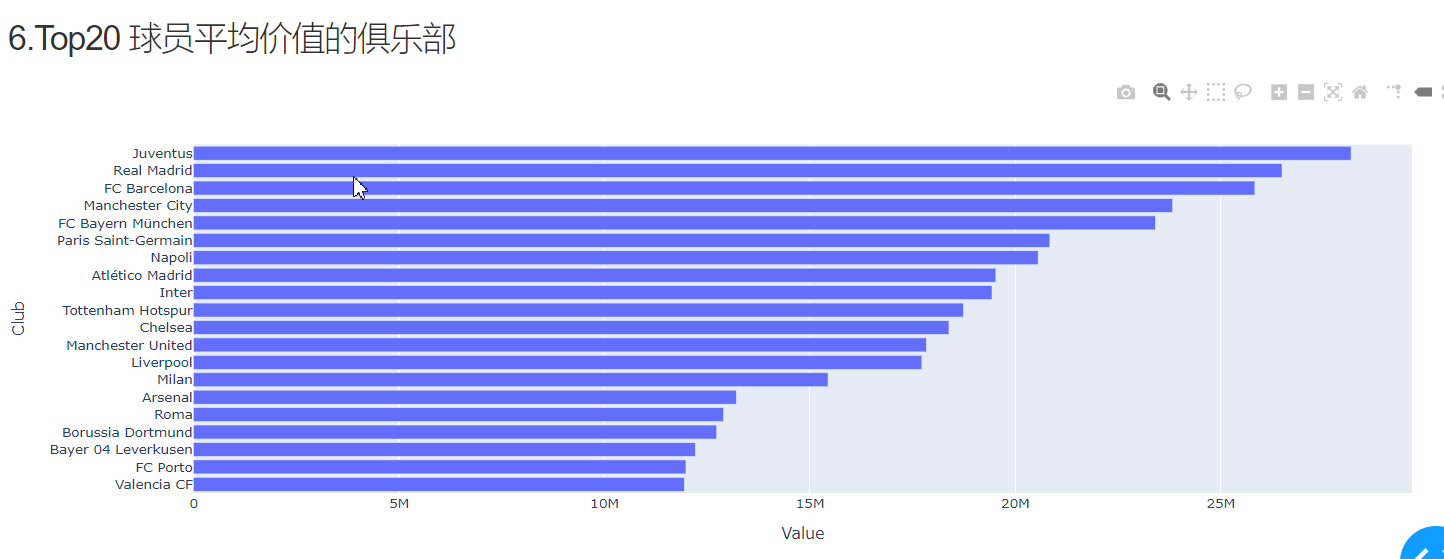
即可进入可视化界面

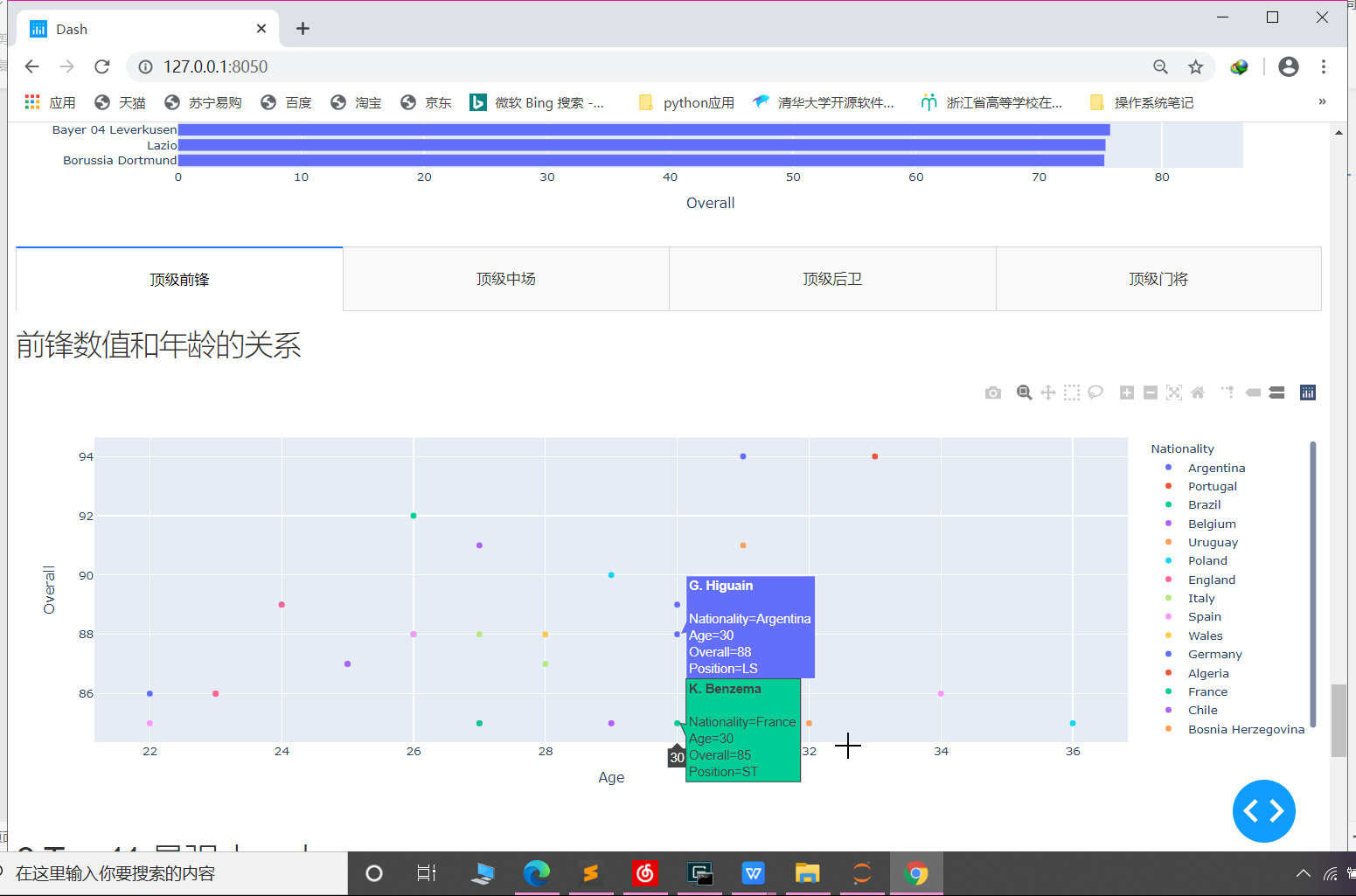


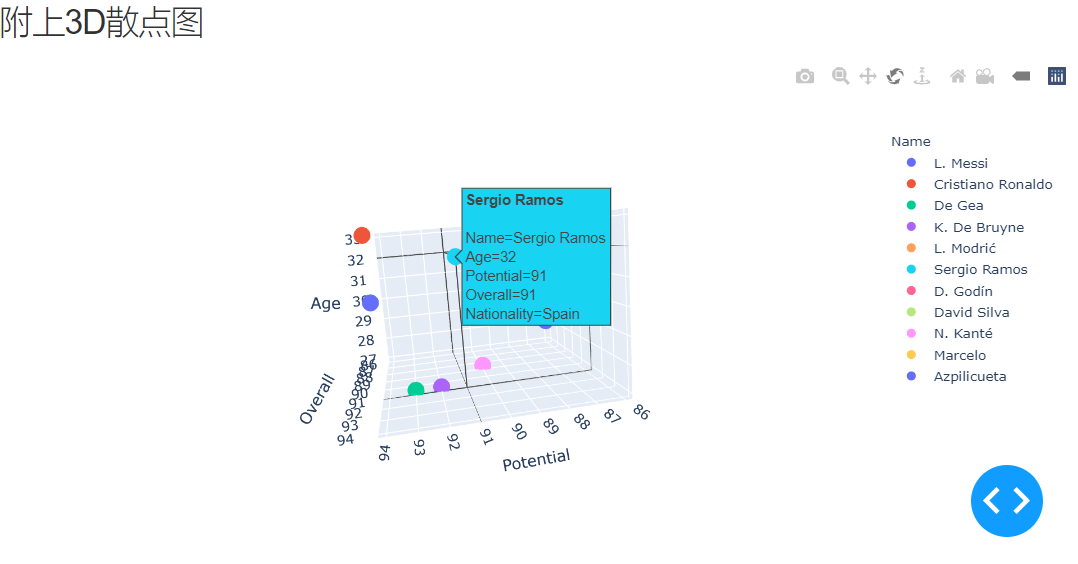
先是简单的表格











<https://dash.plotly.com/>

参考官方文档进行制作