基于中羽的可视化

背景：随着人们对于健康的追求，以及丰富同学们的课余生活，羽毛球运动越来越受大家的喜爱，无论是室内球馆还是室外的空地上都有着大家打羽毛球的声影。而一个名叫“中羽在线”的网站（<https://bbs.badmintoncn.com/>），上面详细的介绍着羽毛球拍、羽毛球鞋、羽毛球运动员等一系列信息。

目标：对“中羽在线”网站进行爬虫，将羽毛球拍的信息等进行整理，方便初学者或者羽毛球爱好者了解相关信息。

成员：周浩、张宇超

项目源码地址：

https://github.com/zhouhaobusy/HomeWork

**项目结构树**：



项目各文件功能解释：

\_\_const.py : 存放一些关于常量和定值的信息如爬虫各字段的 xpath 网站链接等

Data：存放一些爬虫过程对象的序列化文件 （一些pkl文件）

cell/detail\_database.py : 将所有爬取出来的信息生成对应的 Sql 语句

database.sql : database.py 生成的 Sql 语句文件

Front-end : Flask 框架代码和前端页面代码

crawl\_cell.py : 爬取主页面信息 （cell.pkl）

crawl\_detail.py : 爬取物品的详细信息 （detail.pkl）

crawl\_price.py : 爬取各件物品的价格信息 (price.pkl)

Result：存放生成的数据和可视化网页

save\_excel.py 将爬取的数据保存到一个 Excel 表里

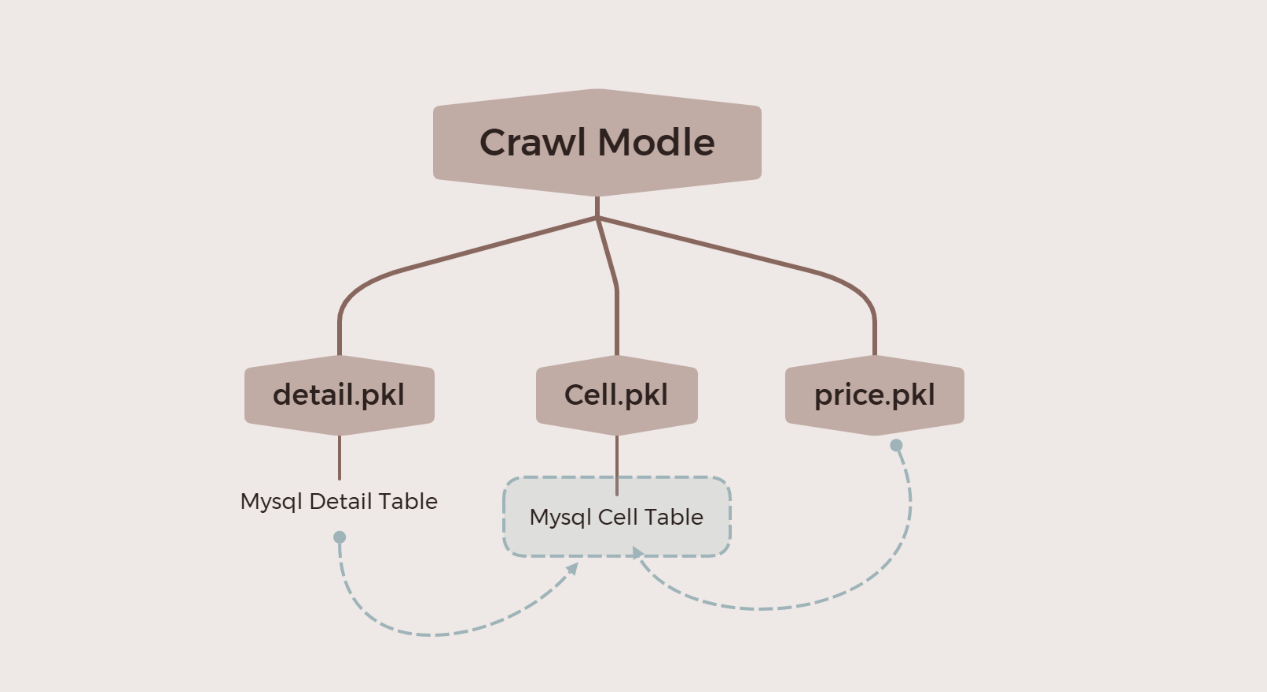
visual.py pyecharts可视化文件

任务内容：

1. 数据爬虫：爬取某个网站的数据，比如当当网的书籍信息、sina网的股票交易信息等等

爬虫使用 xpath 和 selenium 自动化的方式来爬取中羽网站上的各信息,因中羽网站上的数据可能并没有进行审核而直接由用户自主提交,这就导致数据的复杂性高,且存在大量脏数据,这时候我们决定采用多级化数据处理的方式来处理数据

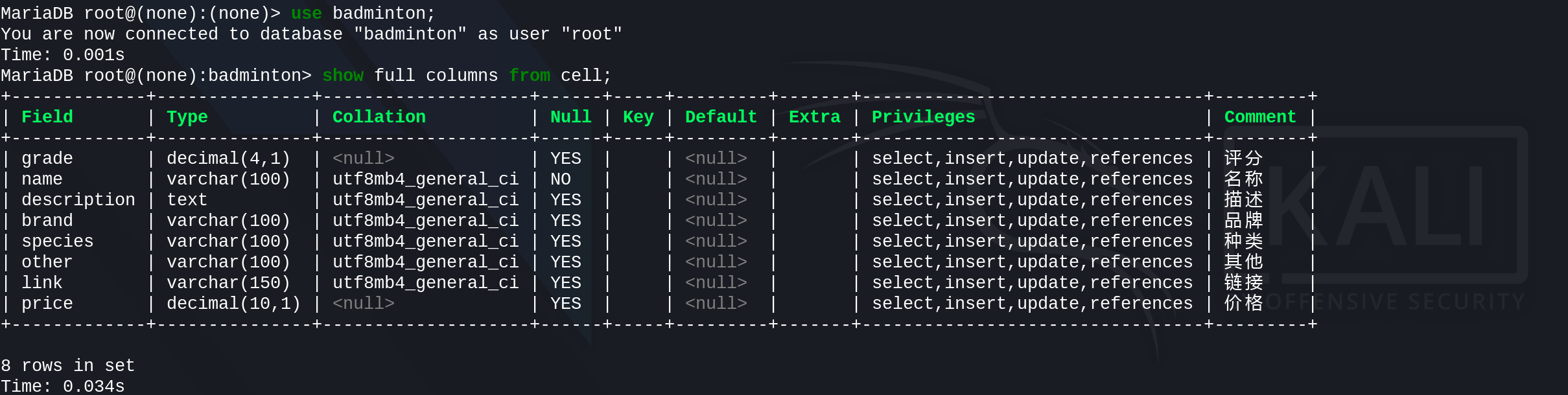
如下图 我们将使用多个爬虫对同一个网站的不同链接点进行爬取,然后将其保存在一个对象里面pkl 然后使用 python 对原始数据进行处理,如将一些脏数据进行替换,将无用数据删除,来达到能导入到mysql的目的.(各字段类型相同)



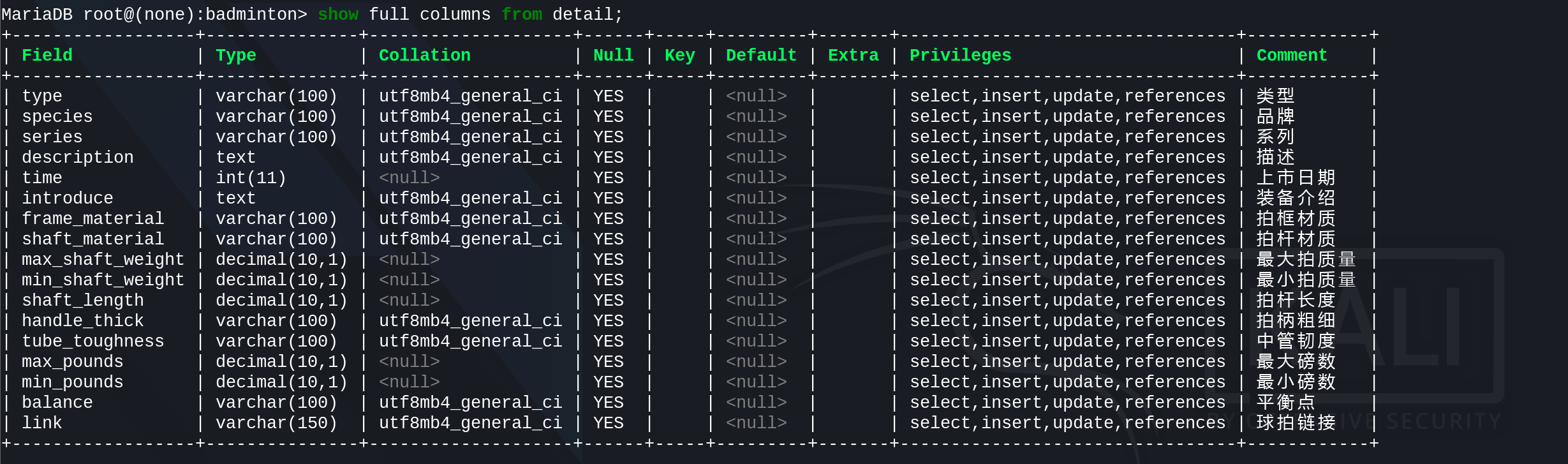
1. 数据库：设计相应的数据库，将爬取的数据存入数据库中

数据库设计

我们对中羽网站的数据设计了两张表,一张为 cell 表,里面存储了所有关于羽毛球运动的物品,如球拍,球鞋…

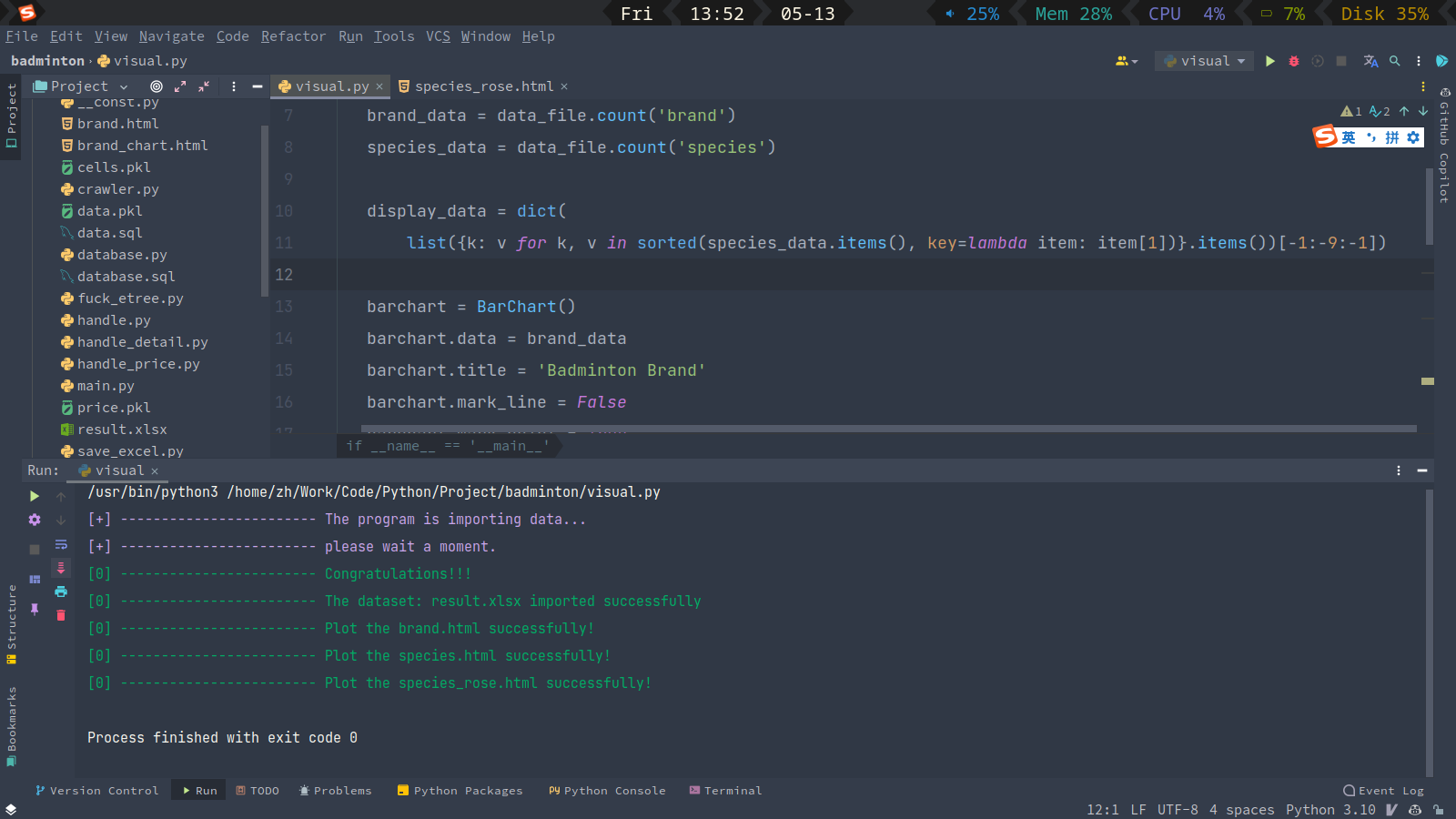


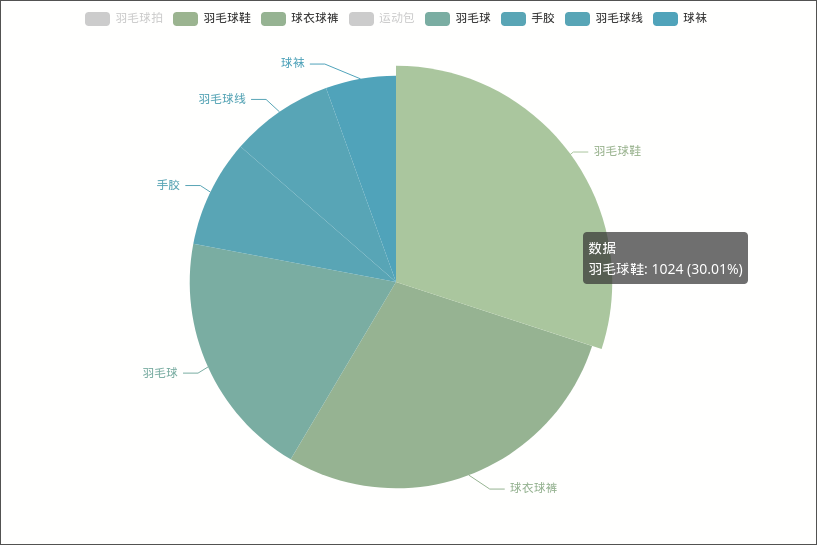
然后因为羽毛球拍的信息相对详细且统一,所有单独设立了一张 detail表 存放的是关于羽毛球拍的详细信息.如下图:

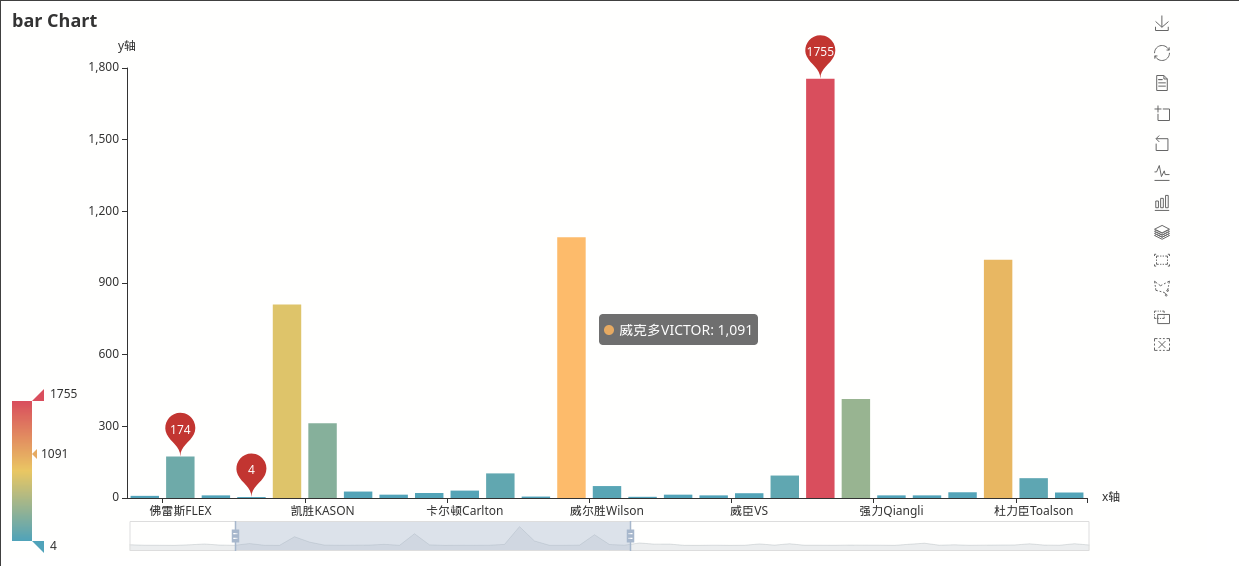


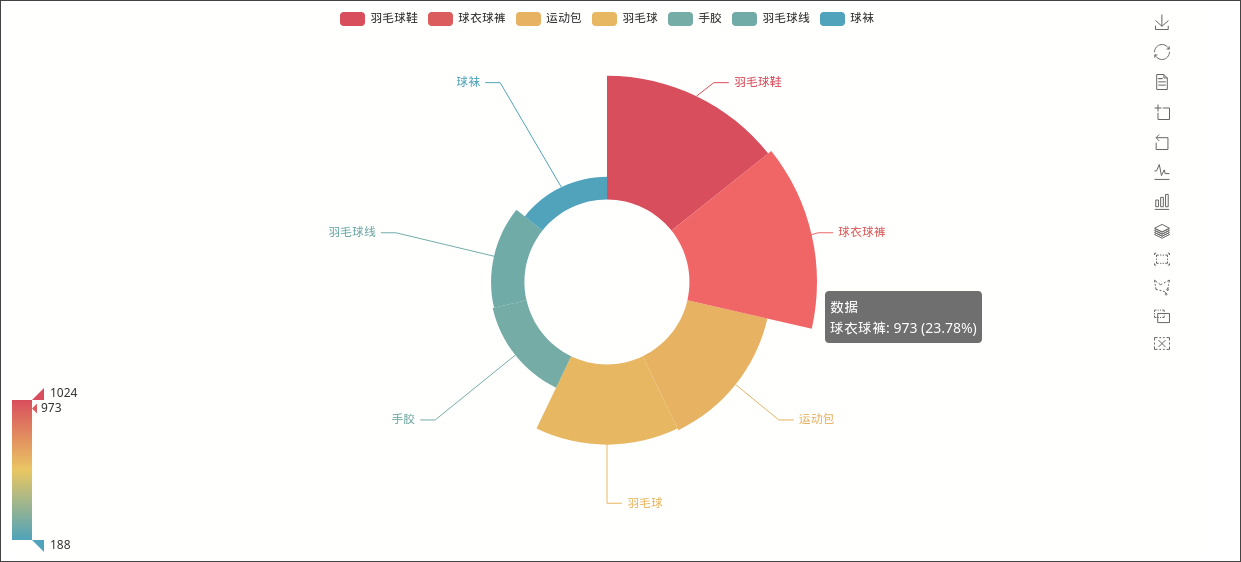
1. 数据分析与可视化

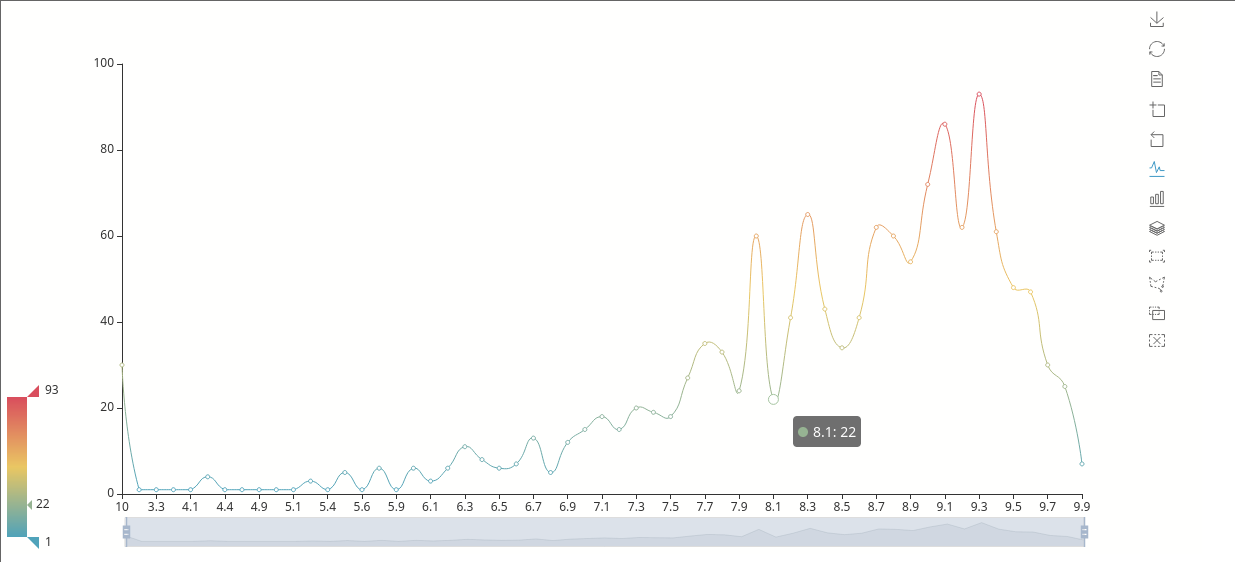
使用 我们自己写的luck 库进行可视化:











使用 flask 进行前端渲染式可视化:

