# **分析报告**

1. 爬虫为spider.py
2. 爬取的原数据存储为legal\_proceedings\_data.csv
3. 经过将数据提取，最终提取的人名或公司名存储为name\_no\_and.txt
4. 总结所有的诉讼结果文档，我将1945-2018这个时间段的诉讼结果文档每年的数目用pyecharts做了分析（文件为：year\_analysis\_line.html），结果如图4.1：



图4.1

从折线图反映出在1945-2001年诉讼结果文档都在增加（我的猜测：在1997香港回归，那时事故比较多，因此处于上涨状态），在2001年时达到高峰，2001-2018整体处于下降趋势（我的猜测：在2001年以后，香港社会秩序相对稳定，法律体系越来越完善，因此诉讼结果文档数目处于下降状态）。

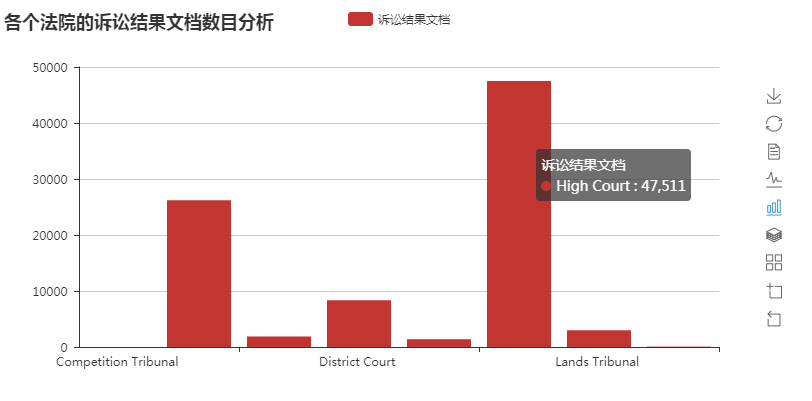
接着总结了各个法院诉讼文件的数据，做了柱状图如图4.2：

图4.2

从图中看出High Court诉讼文档最多，第二多的为[Court of Appeal of the High Court](https://legalref.judiciary.hk/lrs/common/ju/judgment.jsp?L1=CA#H2) ，也就是说一搬的诉讼结果文件都存于高等法院。

5.在不考虑去重的情况下，HKSAR(香港特别行政区（政府）)涉及最多的诉讼案件。