

# QG 工作室暑期实习生两日结

姓名：黄倬熙      组别：数据挖掘      年级：2019      日期：2020-8-11

生活小记	<p>股市方面：今天的股市是真的能让人崩溃的那种，开盘拉升，直到下午两点四十分，也就是收盘前 20 分钟，居然跳水，直接从红的转为绿的，我是真的服，这搞的什么鬼？我的仓位直接-50%，吐了吐了，但是我仍然相信我的预判，八月份必有缩量新高。</p> <p>生活方面：我妈又要出去上课了，留下我一个人在家自己照顾自己，晚上像以前一样做饭给爸妈吃，没办法生活不易啊，他们学习，上班，维持这个家庭真的付出了很多，我也理解。</p> <p>学习方面：开始最终考核咯，要弄一个出租车分析系统，看起来挺有意思嘿嘿嘿，但是看到几个亿的数据其实还是有点害怕的特别是我这高价低配的 MacBook 吃得消吗。</p>
学习 开发 比赛 概要	<p>过滤数据，使用 matplotlib 中的 path 类把数据划分行政区。 尝试使用多种聚类算法寻找高频需求区域</p>
感想收获	<p>我们的考核打算弄点特色，划分行政区来分析每个地区的出租车情况，但当拿到行政区划分边界的经纬度后，我们傻了，怎么判断区域是否在边界内外？我刚开始还想用线性回归来拟合边界线，但是发现那是包围的散点，拟合不是办法，后来继元找到几个博客关于划分的，包括使用Path的办法，但是他们简单尝试，发现不行就放弃了。</p> <p>我本人是比较懒的，能用库绝对不会手搓，看到这个使用Path的办法，我感觉可以啊为什么个个都说不行？划分行政区的也是我的任务，我就按照博客去尝试了。结果发现原来是他们的没有处理过数据，直接读取坐标塞Path了，突然感觉在各位大佬面前没有那么自卑了。</p> <p>使用聚类算法找高频区域，发现使用GMM，KMeans会无脑把地图分成一块块的，使用DBSCAN整个图都是一个颜色，我醉了。</p>
存在问题 (备注)	<p>算法暂时找不到思路，哪个算法都找不到想要的结果</p>