新冠肺炎对北京社会经济影响的分析报告

摘要

本文对疫情期间北京的各项经济影响因素展开数据挖掘，接下来进行描述性分析，分别做出各变量的数据集，之后对数据进行了分类分析和建立模型，从而为我们更理智客观的评价新冠肺炎对北京经济影响起到指导作用。

**关键字：**北京经济 新冠肺炎 大数据

1. 新冠肺炎对北京社会经济影响背景介绍
   1. 新冠肺炎下北京社会经济现状

新型冠状病毒感染肺炎疫情迅速向全国蔓延，举国上下共同抗击疫情，为避免人口大规模流动和聚集，采取了居家隔离、延长春节假期等防控措施。北京作为全国中心城市，社会经济等多方面的影响极具代表性，其交通旅游业、物流、市场股票等多方面受到影响。

* 1. 研究背景

大数据技术发展以来，利用数据可以直观反映经济增长速度，便于分析经济增长影响因素。

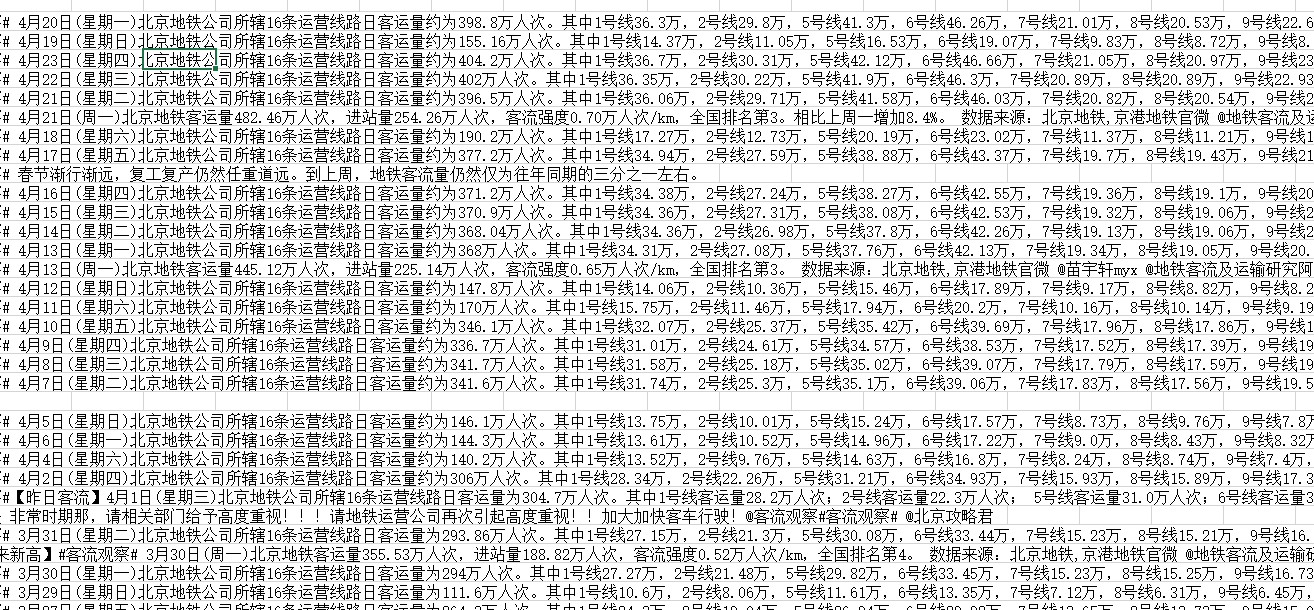
* 1. 研究意义

通过数据挖掘获取地铁人流、居民消费量、股票历史记录和外卖行业业绩的数据，利用大数据方式分析研判，将结果可视化展现，从而发现疫情期间北京物流行业经济受到的影响，以便指导下一步社会经济恢复。

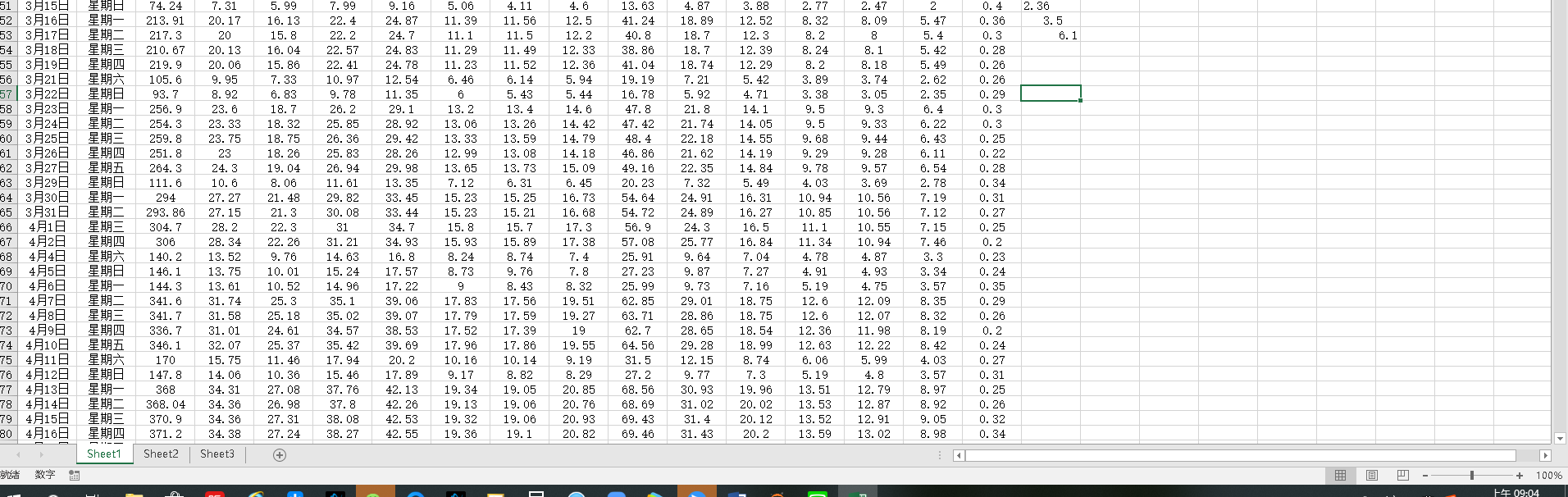
1. 数据说明及清洗

我们的数据主要是从各个网站爬取到的一些与我们研究方向有关数据，其中包括了在疫情期间北京市的地铁人流的情况，以及在期间的市民消费的相关指数，以及部分企业的股票日交易情况，加上对于微博上相关疫情消息的爬取最后做的大家对于疫情的一个情绪分析。

由于我们的地铁人流数据是通过爬取北京地铁的官方微博得到的，其中有大量的重复与无用的信息，最终通过代码筛选过后得到的数据存在部分的缺失。所以这里我们先进行了一次对这部分数据的清洗。

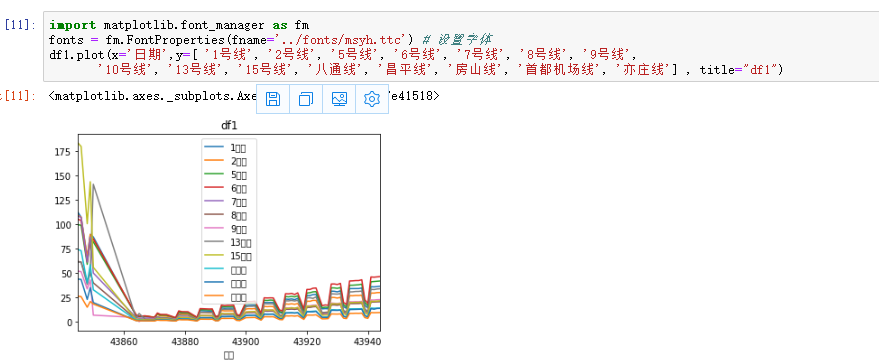


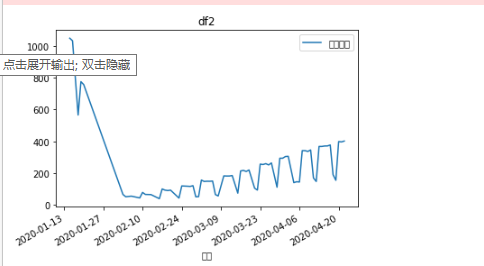
这里是我们初步的到的数据最后整理后化为按照各个线路整理的数据。



可以看到，在亦庄线那里存在部分的缺失，由于这样的缺失对于最后的结果有较大的影响所以我们通过了一次线性回归预测来填补这些缺失，虽然和真实的结果肯定有较大的偏差，但是比起其他的方法，已经可以减小足够的误差，因为这里是一次方案赛，所以就不对这些处理细节进行更多的深入，就假定，我们最后预测的结果便是我们的真实情况，通过北京的地铁人流来说可以看见的是在疫情之前的人流基数为正常北京人流，那么可以看见到了北京疫情公布开始，这里的人流量是急速下降，到复工开始后，有缓慢的提升，在这里我们由于无法获得各个企业详细的复工率，我们组便，拟定用各个线路的人流量来体现北京各区域的经济复苏程度。

最终我们北京地铁人流的数据的结果呈现图

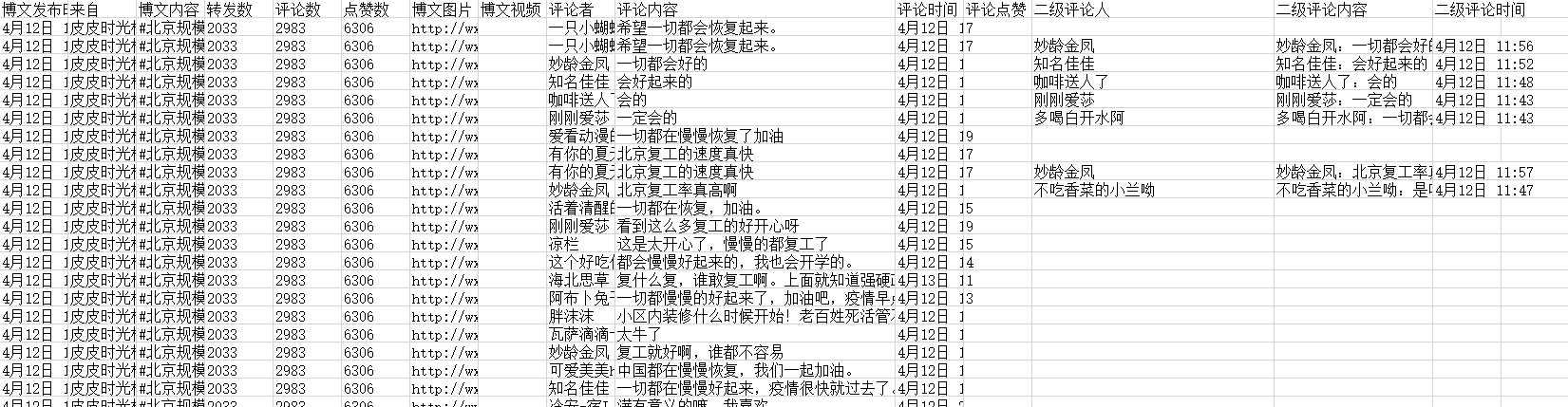




可以看见是从疫情前的暴跌到最后缓慢的恢复，并且在每个周末会有一定的下降，是符合我们的常识的

三. 数据描述性分析

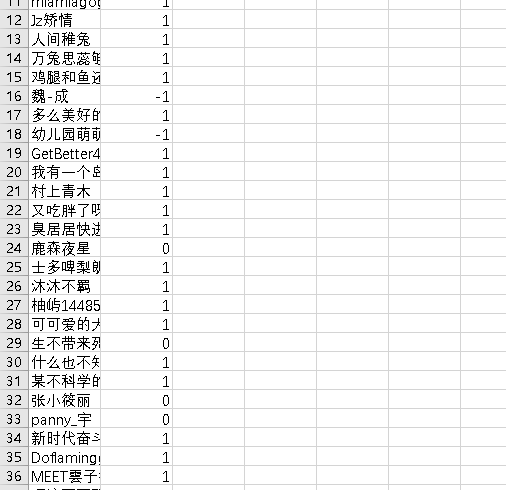
我们对于微博的情绪分析主要是选择了两个话题，一个是复工，一个是恢复上学，得到的结果如下

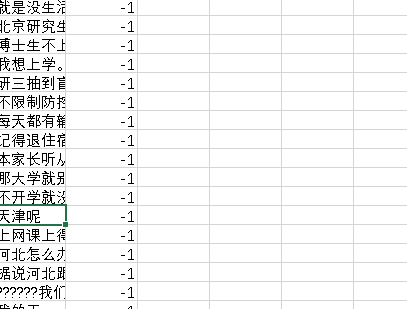


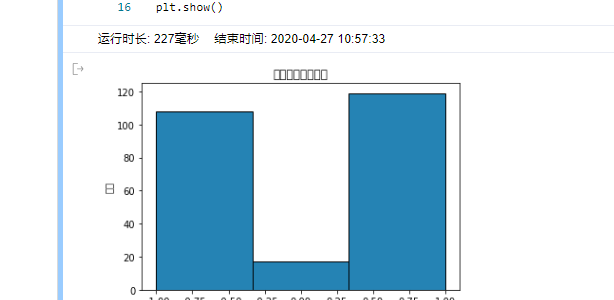


可以看到对于复工的消息大家的情绪是比较积极的，对于恢复上学，大家可以说是喜忧参半，所以考虑到北京疫情的逐渐好转，我们这里没有舍去复工的数据，而是对两个数据进行了加权处理，具体比重会在后面进行处理。

对于这样的数据我们通过建立模型进行了热处理，化为了有效的数据集，其中-1为消极，0为中性，1为积极





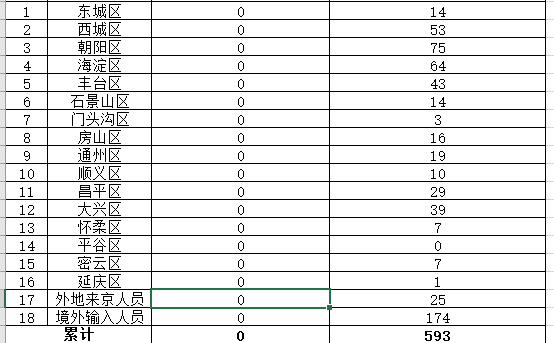


可以看到这个模型的训练效果是很不错的。对于市民的情绪分析也是非常符合我们的分析的。

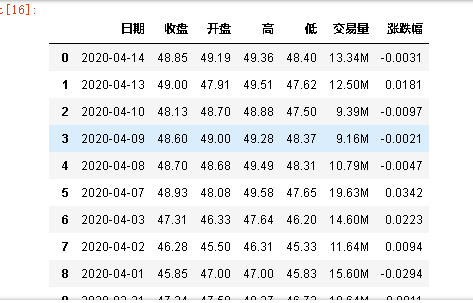
通过这些数据的分析我们可以对北京的市民的恢复情况得到一定的了解，并对此加以分析。

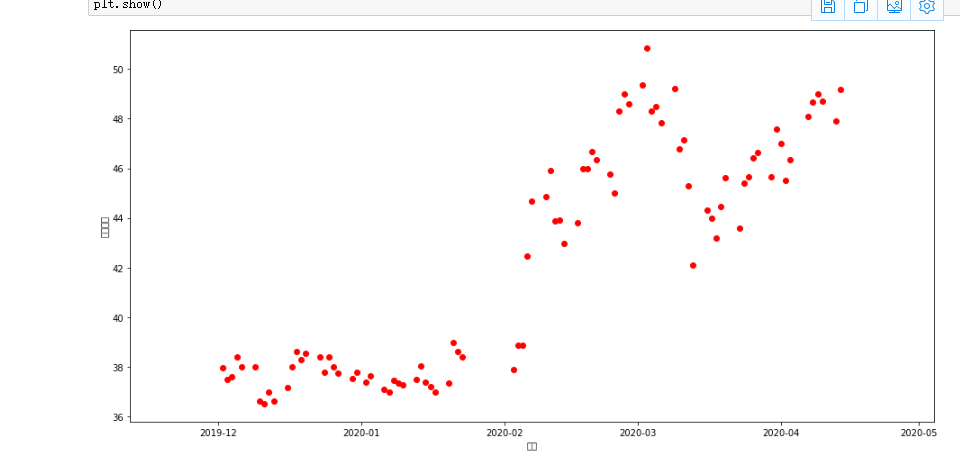
接下来的信息就是北京疫情的详细信息，这个数据一开始是我们从北京卫健委的官网上爬取的，最后大赛提供了政务网，我们就从上面下载了更新的数据。

这个数据的意义在于体现北京的疫情情况，用疫情情况进行分析北京物流行业的影响程度。



以及我们在国家数据网上爬取的北京市分省月度数据和消费数据

最后我们的评定数据为各个物流行业公司的股票波动情况，比如说顺丰，圆通之类的。

可以看见在疫情期间尤其是在二月之前，当时快递行业受到极大的影响，很多是停运的，从这里我们也能看出正是从二月开始顺丰的股价一路向上。