

基于 EDA 方法的 12328 电话数据研究

巩开元

(中国交通通信信息中心, 北京 100011)

摘要: 本文基于某市 12328 交通运输服务监督电话数据, 主要采用了探索性数据分析 (Exploratory Data Analysis, 以下简称 EDA) 方法, 结合某市的交通运输服务行业的实际发展情况, 对某市 2020 年上半年和 2021 年上半年的数据集进行了探索和分析。通过两个时间段的数据对比, 分析新冠疫情对某市的交通运输服务行业产生的影响。同时, 通过数据挖掘, 总结某市的交通运输服务行业存在的规律, 为优化交通运输服务行业管理提供参考。

关键词: 12328 电话; 探索性数据分析; 聚类分析; 交通运输

中图分类号: U491 **文献标识码:** A **文章编号:** 1006—7973 (2021) 11—0134—06

随着中国经济社会的迅速发展, 交通行业的发展无论是对政府还是人民都显得尤为重要。为了进一步使某市的交通运输服务行业平稳顺利的发展, 某市于 2014 年开通了 12328 交通运输服务监督电话。某市 12328 交通运输服务监督电话是各级交通运输主管部门听民声、畅民意、解民忧、汇民智的重要渠道, 是“我为群众办实事”的重要载体, 是广大人民群众表达意愿心声、反映利益诉求、参与行业治理的重要渠道。

为了使 12328 交通运输服务监督电话更好地为人民群众服务, 12328 电话数据分析成为促进交通行业稳定发展、提升交通运输服务水平的重要手段。在过去, 有学者曾对热线电话数据进行研究, 张明岛^[1]等对上海心理健康热线电话心理咨询应用进行了评估, 高文斌^[2]等进行了心理咨询热线在突发性公共卫生事件中的应用分析, 丁小磊^[3]对江苏省 12320 公益电话进行了数据分析, 并阐述了其在突发公共卫生事件中的应用。杨懿轩^[4]等采用了暴露量、因子分析和聚类分析的方法对四川省 12328 电话数据的城市交通问题进行了分析。冯松^[5]等用了文本挖掘技术对江苏某地级城市的 12328 电话数据工单进行了分析。李正银^[6]等对交通服务热线数据信息在行业管理中的应用进行了研究。

本文基于 2020 年 1 月至 6 月以及 2021 年 1 月至 6 月这两个关键性时期某市 12328 交通运输服务监督电话数据, 分析了新冠疫情对某市交通运输行业产生的影响, 为对数据分析结果的应用提出了相关建议。

1 电话数据基本情况

本文从 12328 业务数据表中抽样采集了 2020 年上半年和 2021 年上半年的数据进行分析。2020 年上半年

数据量共计 35 万余条, 2021 年上半年数据量共计 53 万余条, 同比增长 51.43%, 增长主要原因是 2020 年上半年为疫情暴发期间, 百姓的日常出行受限, 话务量较少, 百姓的日常出行不受限, 而 2021 年上半年为疫情常态化期间, 因此话务量较多。如表 1 所示, 本文采用的数据集主要有以下属性: 工单编号、省份编码、业务类型、接听时间、三级业务领域、四级业务领域、答复标记、接听满意度。其中工单编号是字母 JT 加一串数字的序列, 是每一条 12328 电话上报工单独一无二的编码; 省份编码表示 12328 电话服务所在地区省份编码; 业务类型分为三类: 数字 1 代表投诉举报、数字 2 代表信息咨询, 数字 3 代表意见建议; 接听时间的格式为日月年-时分秒; 如图 1 所示, 一级业务领域为城市客运等领域, 二级业务领域为城市交通运行管理等领域, 由于一级业务领域和二级业务领域范围广, 在分析具体业务领域时显得比较宏观, 故在此不做说明, 只考虑三级业务领域和司机业务领域。三级领域包含停车管理、机动车牌照等城市交通运行管理类领域; 四级领域是在三级领域之下细分的类型, 以 02 机动车牌照为例, 其四级领域分为 01 摇号、02 拍卖、03 限行和 99 其他。

表 1 业务数据表

工单编号	JTXXXXXXXXXXXXXXXXXX
省份编码	110000
业务类型	2
接听时间	1/1/2020 09:37:18
三级业务领域	41002
四级业务领域	4100201
答复标记	1
接听满意度	3

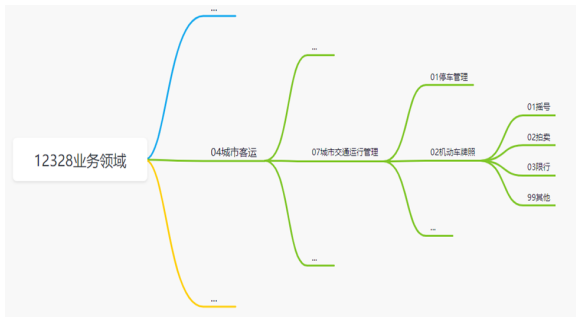


图1 12328 业务领域分级图

2 探索性数据分析 (EDA)

因为在分析数据集前并不清楚本数据集的内在结构，并且考虑到数据集可能会比较杂乱，而探索性数据分析 (EDA) 恰好适合处理这种类型的数据集，所以本文使用探索性数据分析 (EDA) 方法对数据集进行分析。

探索性数据分析 (EDA) 是上世纪六十年代由美国统计学家 John W. Tukey^[7] 所提出的一种对已有的原始数据在尽量少的先验假定下进行探索，通过作图、制表、方程拟合、计算特征量等手段探索数据的结构和规律的一种数据分析方法。

2.1 三级业务领域

表2 2020年上半年各三级业务领域 (SERVICE_FIELD_4) 数据量表格

停车管理	机动车牌照	其他	公共汽电车线路	巡游出租汽车
109866	109570	25492	17711	8774

表3 2021年上半年各三级业务领域 (SERVICE_FIELD_4) 数据量表格

机动车牌照	停车管理	无效电话其他	其他	公共汽电车线路
218885	78881	149899	17391	8617

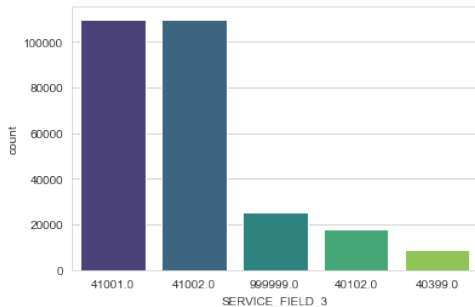


图2 2020年上半年各三级业务领域 (SERVICE_FIELD_3) 数据量柱状图

注：41001 停车管理；41002 机动车牌照；99999 其他；40102 公共汽电车线路；40399 巡游出租车其他；990199 无效电话其他。

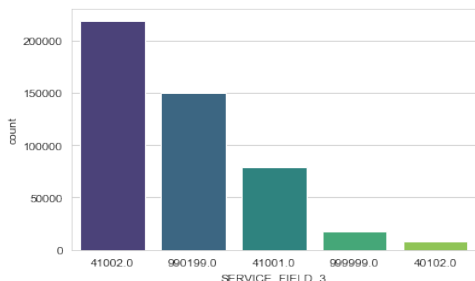


图3 2021年上半年各三级业务领域 (SERVICE_FIELD_3) 数据量柱状图

注：41001 停车管理；41002 机动车牌照；99999 其他；40102 公共汽电车线路；40399 巡游出租车其他；990199 无效电话其他。

如表2和图2，2020年上半年，三级业务领域数据量排名前5的领域依次为停车管理、机动车牌照、其他、公共汽电车线路、巡游出租汽车其他。

如表3和图3，2021年上半年，三级业务领域数据量排名前5的领域依次为机动车牌照、无效电话其他、停车管理、其他、公共汽电车线路。

由此可见，从2020年上半年过渡到2021年上半年，停车管理减少28.20%，公共汽电车线路减少51.35%，机动车牌照增加99.77%。

2.2 四级业务领域

表4 2020年上半年各四级业务领域数据量表格

机动车牌照摇号	机动车牌照限行	机动车牌照其他	路况咨询其他	ETC 收费结算
87060	11862	10648	6643	3537

表5 2021年上半年各四级业务领域数据量表格

机动车牌照摇号	机动车牌照其他	巡游出租车财物遗失	机动车牌照限行	公共汽电车财物遗失
211768	4171	3866	2946	2888

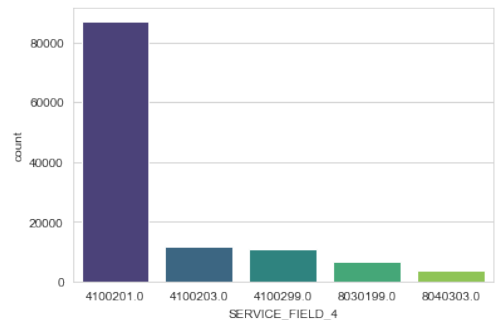


图4 2020年上半年各四级业务领域 (SERVICE_FIELD_4) 数据量柱状图

注：4100201 机动车牌照摇号；4100203 机动车牌照限行；4100299 机动车牌照其他；8030199 路况咨询其他；8040303 ETC 收费结算；4030301 巡游出租车财物遗失；4010801 公共汽电车财物遗失。

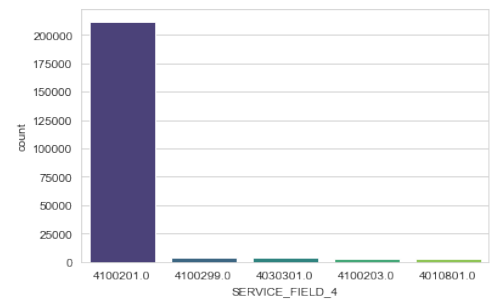


图5 2021年上半年各四级业务领域 (SERVICE_FIELD_4) 数据量柱状图

注：4100201 机动车牌照摇号；4100203 机动车牌照限行；4100299 机动车牌照其他；8030199 路况咨询其他；8040303 ETC 收费结算；4030301 巡游出租车财物遗失；4010801 公共汽电车财物遗失。

如表4和图4，2020年上半年，四级业务领域数据量排名前5的领域依次为机动车牌照摇号、机动车牌照限行、机动车牌照其他、路况咨询其他、ETC 收费结算。

如表5和图5，2021年上半年，四级业务领域数据量排名前5的领域依次为机动车牌照摇号、机动车牌照其他、巡游出租车财物遗失、机动车牌照限行、公共汽电车财物遗失。

由此可见，从2020年上半年过渡到2021年上半年，机动车牌照摇号增加143.24%，机动车牌照限行减少75.16%、机动车牌照其他减少60.83%。

通过对一、二、三、四级业务领域属性的分析，一二级领域主要体现范围较广的业务领域，而三四级业务领域主要体现更具体的业务领域。从三级业务领域来说，百姓最关心的是停车管理问题和机动车牌照的问题，从四级业务领域来说，百姓最主要关注的业务领域是机动车牌照摇号和限行。从现实的角度来分析，因为某市为大型城市，车辆必然比较多，停车场也比较有限，所以停车管理相对较难，因而百姓拨打12328电话的数据量也就比较多；而因为某市对机动车牌照有严格的限号和摇号的政策，所以无论是机动车牌照的限行，还是摇号都是百姓比较关心的问题，百姓拨打12328电话的数据量多也是理所当然。而疫情对业务领域的影响并不明显，但无论如何，机动车牌照的限行还是摇号，关系到民生中的衣食住行中的“行”。

2.3 业务类型

表6 2020年上半年各业务类型数据量表格

投诉举报	信息咨询	意见建议
62511	216474	79447

表7 2021年上半年各业务类型数据量表格

投诉举报	信息咨询	意见建议
44514	288479	200793

如表6和表7所示，从2020年上半年过渡到2021年上半年，投诉举报减少28.79%，信息咨询增加33.26%，意见建议增加152.74%。

通过对业务类型属性的分析，信息咨询和意见建议数据量有较大增幅。而结合2020年上半年和2021年上半年的实际情况而言，2020年上半年百姓出行较少，2021年上半年百姓出行较多，因此，出行增加是导致信息咨询和意见建议的重要因素之一。

2.4 业务类型 & 业务领域

表8 2020年上半年不同业务类型下不同业务领域的的数据量表格

SERVICE_TYPE	SERVICE_FIELD	SERVICE_FIELD_2	SERVICE_FIELD_3	SERVICE_FIELD_4	total
118	2.0	4.0	410.0	41002.0	4100201.0 85142
167	3.0	4.0	410.0	41002.0	4100299.0 10648
166	3.0	4.0	410.0	41002.0	4100203.0 6749
120	2.0	8.0	803.0	80301.0	8030199.0 6643
119	2.0	4.0	410.0	41002.0	4100203.0 5113
102	2.0	4.0	401.0	40108.0	4010801.0 3486
112	2.0	4.0	403.0	40303.0	4030301.0 3088
121	2.0	8.0	804.0	80402.0	8040299.0 2677
124	2.0	8.0	804.0	80403.0	8040303.0 2580
34	1.0	4.0	401.0	40105.0	4010511.0 1874
165	3.0	4.0	410.0	41002.0	4100201.0 1812
76	1.0	4.0	407.0	40701.0	4070103.0 1315
26	1.0	4.0	401.0	40104.0	4010401.0 1252
82	1.0	8.0	804.0	80403.0	8040303.0 944
53	1.0	4.0	403.0	40301.0	4030107.0 816

表9 2021年上半年不同业务类型下不同业务领域的的数据量表格

SERVICE_TYPE	SERVICE_FIELD	SERVICE_FIELD_2	SERVICE_FIELD_3	SERVICE_FIELD_4	total
121	2	4	410	41002.0	4100201.0 209461
164	3	4	410	41002.0	4100299.0 4171
115	2	4	403	40303.0	4030301.0 3866
105	2	4	401	40108.0	4010801.0 2888
163	3	4	410	41002.0	4100203.0 2328
162	3	4	410	41002.0	4100201.0 1959
35	1	4	401	40105.0	4010511.0 1329
27	1	4	401	40104.0	4010401.0 1073
113	2	4	403	40301.0	4030199.0 690
54	1	4	403	40301.0	4030107.0 646
14	1	3	305	30501.0	3050199.0 637
123	2	8	803	80301.0	8030199.0 636
31	1	4	401	40105.0	4010504.0 629
122	2	4	410	41002.0	4100203.0 618
95	2	3	305	30501.0	3050101.0 547

如表8所示，2020年上半年，值得重点关注SERVICE_TYPE为1的数据总量（即业务类型为投诉举报的数据总量），投诉举报数据量最多的业务领域为4010511、4070103和4010401，相对应的业务领域名称为公共汽电车的文明服务、互联网租赁自行车的资金、公共汽电车的运行间隔。

如表9所示，2021年上半年，值得重点关注SERVICE_TYPE为1的数据总量（即业务类型为投诉举报的数据总量），投诉举报数据量最多的业务领域为4010511、4010401和4030107，相对应的业务领域名称为公共汽电车的文明服务、公共汽电车运行时间的运行间隔、巡游出租汽车的文明驾驶。

由此可见，无论2020年上半年还是2021年上半年，公共汽电车的文明服务和公共汽电车运行时间的运行间隔都是百姓投诉的主要方面。

通过对业务类型和业务领域这两个属性的分析，从2020年上半年和2021年上半年来看，百姓投诉的问题主要是公共汽电车的文明服务（司乘人员服务态度等）和公共汽电车的运行间隔（包括发车间隔、等待时间等），而互联网租赁自行车的资金（包括乘客账户资金安全、以及退款时效等）和巡游出租汽车的文明驾驶（包括违章变道、超速、接打电话、发微信、操作手机软件等影响乘客安全与舒适性的驾驶行为）则分别是百姓2020年上半年和2021年上半年投诉较多的问题。

2.5 答复标记

表10 2020年上半年答复标记数据量表格

非即时答复	已即时答复
65505	292927

表11 2021年上半年答复标记数据量表格

非即时答复	已即时答复
56667	477119

如表10和表11所示，2020年上半年和2021年上半年非即时回复减少13.49%，已即时答复增加62.88%。

通过对答复标记数据量的分析，2020年上半年和

2021年上半年，非即时答复数据量变化不大，已即时答复数据量有接近20万的增幅，这说明某市12328即时答复率增加显著，侧面反映出了话务员能解答的问题更多，电话服务质量更好。

2.6 电话服务满意度

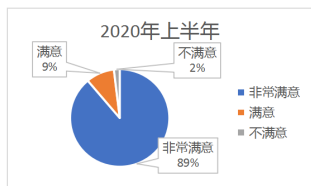


图6 2020年上半年满意度百分比饼图

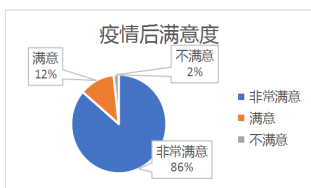


图7 2021年上半年满意度百分比饼图

如图6和图7所示，从2020年上半年过渡到2021年上半年，非常满意的百分比降低3%，满意的百分比提高3%，不满意的百分比保持不变。

通过对电话服务满意度的分析，2020年上半年和2021年上半年，虽然电话数据量增加，但是非常满意和满意的电话数据量并没有太大变化。

2.7 每周 & 每小时热力图 (heatmap)

热力图 (heatmap) 是一种比较常见的可视化手段，因其丰富的色彩变化和生动饱满的信息表达被广泛应用于各种大数据分析场景。

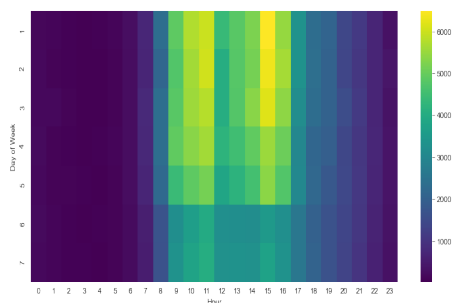


图8 2020年上半年每周 & 每小时热力图

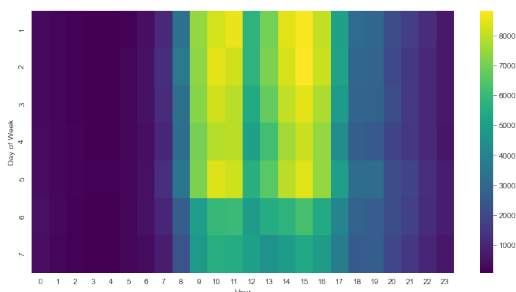


图9 2021年上半年每周 & 每小时热力图

如图8和图9所示，将每星期每天、小时数和相应

的电话数据量建立联系画出热力图，颜色较浅的区域表明电话数据量较多，颜色较深的区域表明电话数据量较少。

2.8 每周 & 每小时聚类图 (clustermap)

聚类图是一种可对数据进行分类的图像。Python中的seaborn模块提供了聚类功能，可以使用多种聚类分析方法对数据进行聚类分析，本文使用了聚类图里由Sokal和Michener提出的UPGMA (Unweighted Pair Group Method with Arithmetic Mean)^[8]聚类算法，以下是算法简介：
$$d(u,v) = \sum_{i,j} \frac{d(u[i],v[j])}{(|u| * |v|)}$$

UPGMA (具有算术均值的非加权对组方法) 是一种凝聚 (自下而上) 层次聚类方法。对于所有的点*i*和点*j*，*l_u*和*l_v*分别是类*u*和类*v*的基数。

聚类图除了可以采用热力图对相关系数进行可视化，还可以根据相关系数对变量进行聚类，发现热力图数据的结构。

在聚类图函数里，不仅使用了热力图，而且还使用了聚类功能，使得绘制出来的热力图具有横向样本和纵向样本的聚类功能。

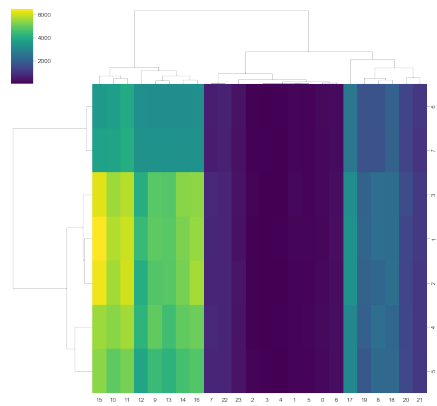


图10 2020年上半年每周 & 每小时聚类图

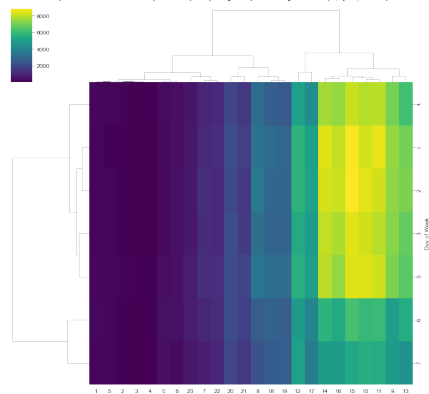


图11 2021年上半年每周 & 每小时聚类图

如图10所示，从横向聚类分析，聚类图将周六周日的的数据量分为一类，周一到周五的数据量分为另一类，更多分类依此类推。从纵向聚类分析，聚类图将9点到

下午 16 点的数据量分为一类，其他时间段的数据量分为另一类，更多分类依此类推。

如图 11 所示，从横向聚类分析，聚类图也是将周六周日的的数据量分为一类，周一到周五的数据量分为另一类，更多分类依此类推。从纵向聚类分析，聚类图将 9 点到下午 17 点的数据量分为一类，其他时间段的数据量分为另一类，更多分类依此类推。

通过对每周与每小时相关的热力图 and 聚类图的分析，2020 年上半年和 2021 年上半年，周一到周四上午的 9 点至下午 16 点，都是电话数据量比较多的时间。2021 年上半年，周五上午 9 点至下午 16 点数据量明显比 2020 年上半年多。

2.9 每月 & 星期热力图 (heatmap)

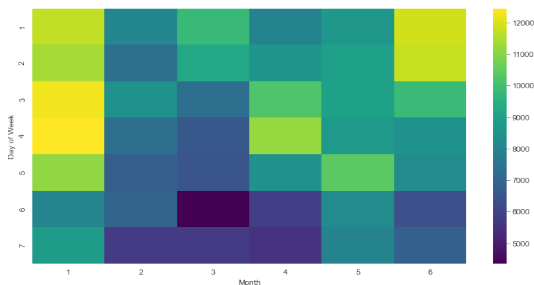


图 12 2020 年上半年每月 & 每星期热力图

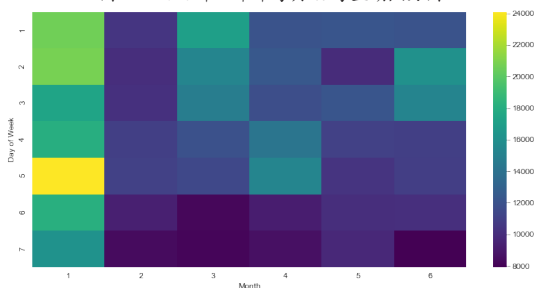


图 13 2021 年上半年每月 & 星期热力图

如图 12 和图 13 所示，将每星期每天、每个月和相应的电话数据量建立联系画出热力图，颜色较浅的区域表明电话数据量较多，颜色较深的区域表明电话数据量较少。

2.10 每月 & 每周聚类图 (clustermap)

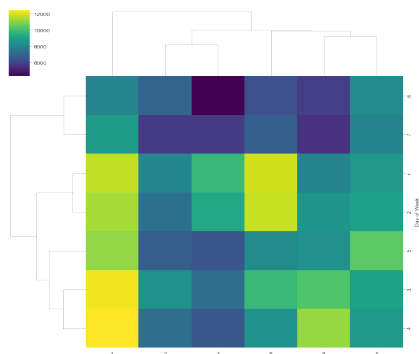


图 14 2020 年上半年每月 & 每周聚类图

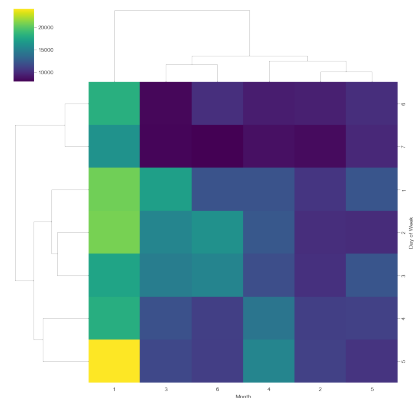


图 15 2021 年上半年每月 & 每周聚类图

如图 14 所示，从横向聚类分析，聚类图将周六周日的的数据量分为一类，周一到周五的数据量分为另一类，更多分类数量依此类推。从纵向聚类分析，聚类图将 1 月份数据量分为一类，其他月份的数据量分为另一类，更多分类依此类推。

如图 15 所示，从横向聚类分析，聚类图也是将周六周日的的数据量分为一类，周一到周五的数据量分为另一类，更多分类依此类推。从纵向聚类分析，聚类图将一月份的数据量分为一类，其他月份的数据量分为另一类，更多分类依此类推。

通过对每周与每月相关的热力图 and 聚类图的分析，2020 年上半年，周一到周五数据量较多，其中一月份周三和周四、六月份周一和周二电话数据量较多，其他时间段的电话数据量相对较少。2021 年上半年，一月份的周五电话数据量较多，周一、周二、周三、周四、周六和周日电话数据量次之，其他时间段的电话数据量相对少。

3 建议

本文所提供的数据分析结果可为相关行业人员提供一定参考，比如数据量多的时候可以多安排一些话务员，数据量少的时候可以适当少安排话务员，减少人力物力成本。又比如为了减少百姓对公共汽电车的文明服务（司乘人员服务态度等）和公共汽电车的运行间隔（包括发车间隔、等待时间等）提出投诉，为群众办实事，可以进一步提升公共汽电车的文明服务，优化完善公共汽电车的运行间隔时间。再比如可根据每星期每天 12328 数据量的多少，适当调整限号政策，一定程度上减轻交通压力。12328 电话监督服务是反映百姓对交通服务满意度晴雨表，从探索性分析的角度不仅可以使相关行业者对交通行业的现状有大概的了解，并且能从数

城市道路排水工程设计分析

金莉莉

(杭州市交通规划设计研究院, 浙江 杭州 310006)

摘要: 水是城市道路形成病害的主要影响因素, 在设计、施工和养护中, 必须重视路基排水设施设计与施工, 把排水工程做好。本文介绍了 S304 余杭小林至塘栖段改扩建工程中排水设施的排水设计的概念和原则, 并对城市道路排水设施的设计进行了简要的分析, 为同仁借鉴参考。

关键词: 城市道路; 排水设施; 排水设计

中图分类号: U417.3 **文献标识码:** A **文章编号:** 1006—7973 (2021) 11—0139—04

1 前言

S304 余杭小林至塘栖段改扩建工程属于杭州城市道路之一, 兼顾环城高速公路通车目的, 我院针对本项目的水文特征, 认真地做好仔细调查分析, 结合地方需求, 确定一套功能完善、排水能力强的排水综合系统。该排水工程设计除符合杭州城市发展和余杭区域总体规划外, 排水专业规划和其它各相关专业规划相协调一致。

2 设计原则

(1) 路基排水设计流量计算: 采用当地 15 年重现期降雨历时内的最大降雨强度; 路面排水设计, 按当地 5 年的重现期降雨历时内的最大降雨强度。

(2) 路基路面排水结合路线设计, 在充分考虑沿线地形、水系、排灌系统的基础上, 进行总体设计, 引入“安全、耐久、节约、和谐”的设计理念, 使之形成

据中挖掘出百姓对交通服务有哪些疑问, 对哪些方面有意见, 并能够结合行业具体情况, 调整相关政策, 做到脚踏实地为百姓办实事, 提升百姓对交通服务行业的满意度。

4 总结及展望

本文所采用的探索性数据分析 (EDA) 方法不仅对 2020 年上半年和 2021 年上半年两个时间段的数据进行了单属性画图分析, 而且将多个属性交叉进行画图分析, 说明了疫情发生后, 12328 电话的大部分属性的数据量都有所减少。同时, 将业务领域数据量与接听时间进行有效结合, 发现了电话数据量随固定时间发生变化的一般性规律, 并进行了聚类分析。但是, 在探索 12328 电话数据过程中仍然面临一些挑战, 比如在分析文本类 12328 电话数据, 用文本提取相关方法统计并对词频进行排序时, 虽然能统计出文本中出现频率最高的关键词, 但是对实际生活产生意义不大。又比如, 通过几种数据属性进行机器学习, 虽然能预测出接听满意度, 但是准确率并不高。无论今后能否优化文本分析以及预测算法, 本文都能为进一步探索分析 12328 电话数据提供必要参考。

参考文献:

- [1] 季建林, 张明岛, 储展明, 等. 上海心理健康热线电话心理咨询应用评估 II: 电话咨询服务的优点及疗效 [J]. 临床精神医学杂志, 1995(5):270-271.
- [2] 高文斌, 陈祉妍. 心理咨询热线在突发性公共卫生事件中的应用分析 [J]. 心理科学进展, 2003(4):400-404.
- [3] 丁小磊. 江苏省 12320 公益电话数据分析及在突发公共卫生事件中的应用 [J]. 江苏预防医学, 2010(1):60-61.
- [4] 杨懿轩, 陆斌, 徐梦清, 等. 基于 12328 电话数据的城市交通问题分析 [J]. 交通运输工程与信息学报, 2019(4):113-121.
- [5] 冯松, 李正银, 孙菲. 地级城市 12328 电话数据分析与成果应用 —— 以出租汽车投诉工单为例 [J]. 现代交通技术, 2021(2):65-67,82.
- [6] 李正银, 夏斯明, 金春良. 交通服务热线数据信息在行业管理中的应用研究 [J]. 科技创新与应用, 2017(32):135-136.
- [7] Tukey, John W. (1977). Exploratory Data Analysis. Pearson. ISBN 978-0201076165.
- [8] Sokal, Michener(1958). "A statistical method for evaluating systematic relationships". University of Kansas Science Bulletin.38:1409-1438.