

2019 CCF 大数据与计算智能大赛
海口市-交通流量时空演变特征可视分析
(CCF Big Data & Computing Intelligence Contest)
答 卷

参赛队名称：西南交通大学-前行者

指导老师：西南交通大学人工智能研究院院长 李天瑞教授

团队成员： 王德贤，西南交通大学，wangdexian2019@outlook.com，队长

杨善雯，西南交通大学，swyang@my.swjtu.edu.cn

刘佳，西南交通大学，xiaoke92@foxmail.com

李浙川，成都理工大学，cdutlzc@gmail.com

李博，成都理工大学，2524722791@qq.com

是否学生队（是或否）： 是

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）：

Excel, Echarts, D3, HighCharts, Python, Java

简介

海口，别称“椰城”，地处热带，热带资源呈现多样性，是一座富有海滨自然旖旎风光的南方滨海城市。海口城市交通主要依赖于四通八达的路面交通。通过海口市出行大数据对海口市的交通进行分析来解决交通中存在的问题，为市民和城市交通管理部分提供建议具有非常重要的意义。我们基于2017年5月1日-10月31日海口市的滴滴平台上的订单相关的数据，结合海口市地理特征、城市特性和人口信息，利用可视分析技术与表达模型，建立了“海口市交通流量可视化系统”平台，该系统包括一个主视图和多个辅助视图，如图1所示。平台从各个方面对2017年半年来海口市交通需求演变进行勾画，包括天订单量，时订单量，点单量实时流动等。其次，我们分析影响出行变化的主要因素，包括天气，特殊时间段和节假日等。接下来，我们分别从题目的三个问题进行逐一分析。

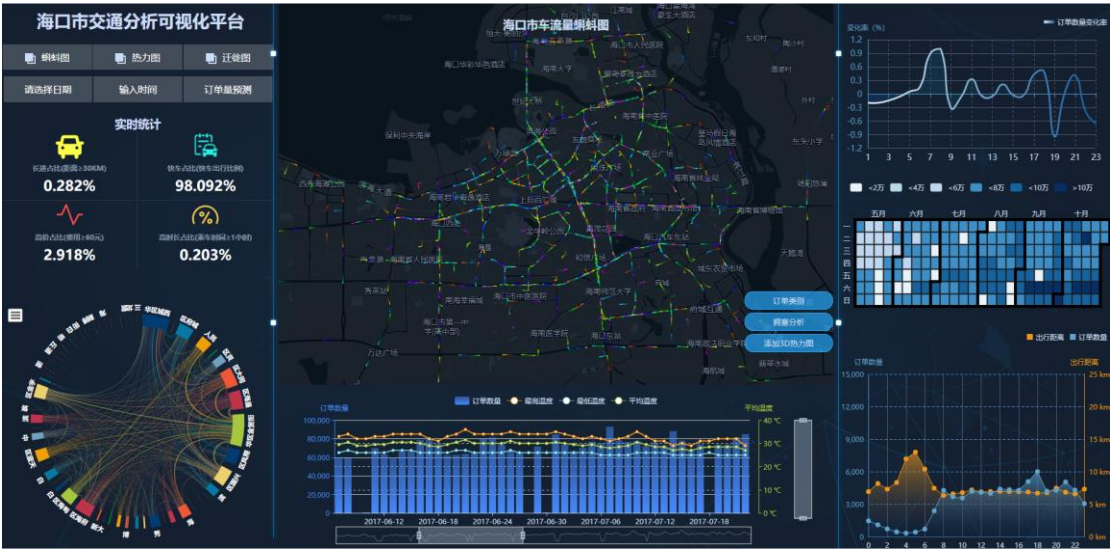


图1 海口市交通流量可视化系统概览图1

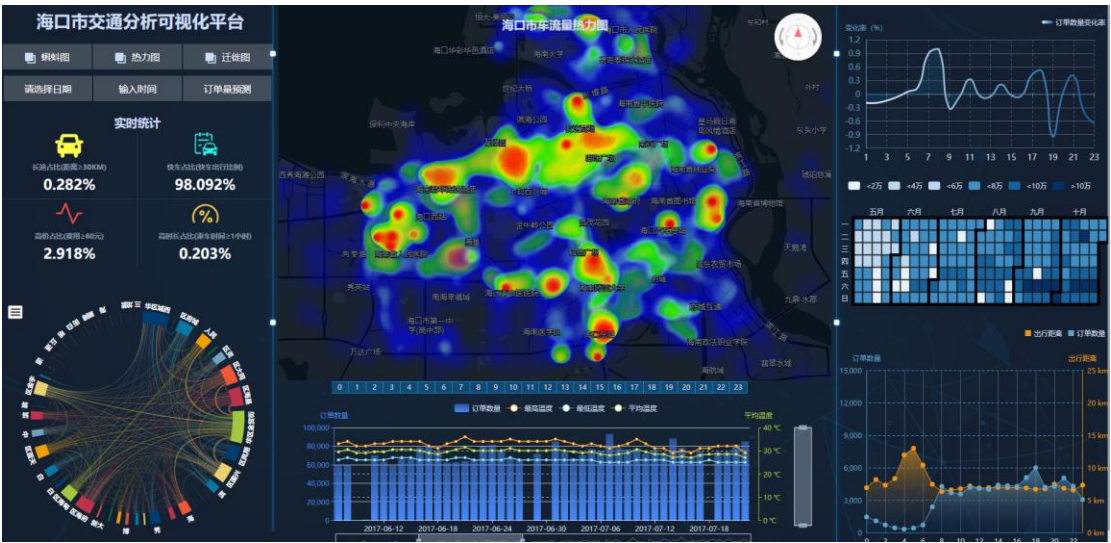


图2 海口市交通流量可视化系统概览图2

问题 1. 请您分析 2017 年 5 月 1 日—10 月 31 日海口市交通需求演化情况，包括但不限于出发区域、到达区域、出发时间、出行距离、出行次数等不同维度的变化；

首先，我们使用日历热力图（图 2-1）展示了 2017 年 5 月 1 日—10 月 31 日海口市的滴滴订单数量变化。日历热力图展示的主要信息包括日期、订单量、天气情况、星期几、平均温度和节假日相关信息（以及有可能的相关事件），其中以订单数量为主要数值展示属性（即图中每个方格的颜色代表订单量）。

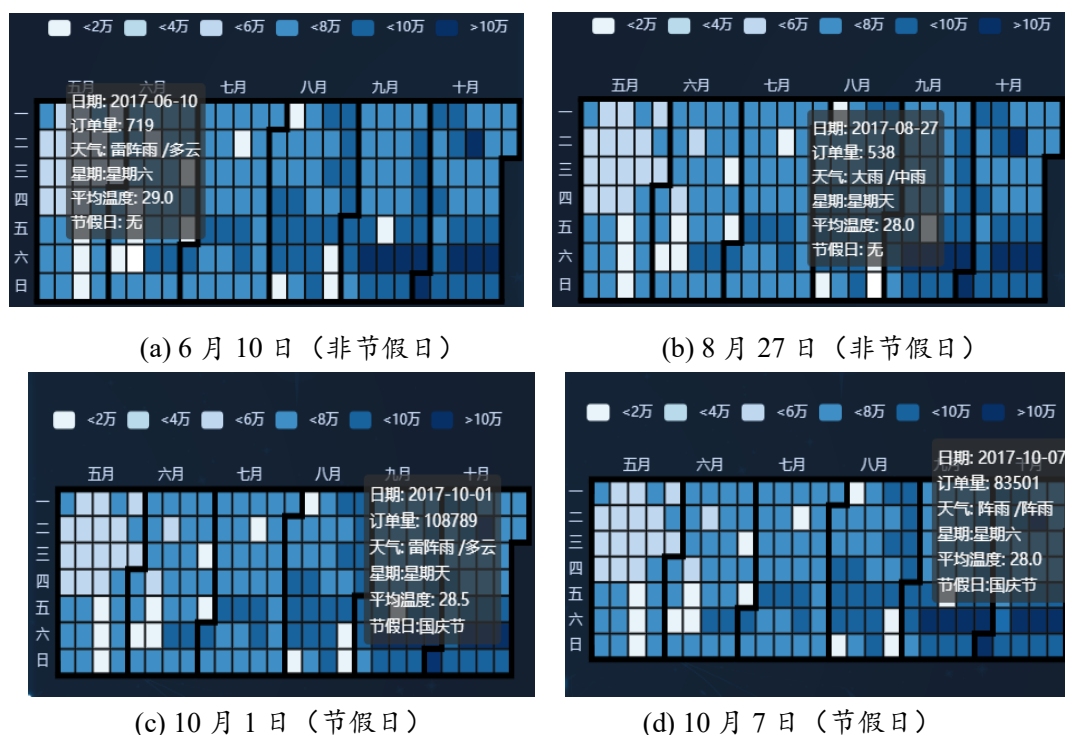


图 2-1 滴滴打车订单信息日历热力图

可以看出 2017 年 5 月 1 日—10 月 31 日；7、8、9、10 月的订单数量明显高于 5、6 月的订单数，这可能的原因包括：

- (1) **高温天气的影响：**使得许多人选择滴滴打车作为出行方式；
- (2) **假期的影响：**由于海口市是国内重要的旅游城市，暑假期间有大量的人到海口市旅游，因此暑期间的订单量增多；
- (3) **季节变化的影响，**9、10 月份各地到海口旅游的人数增加。其次，普遍情况下周六、周日的订单数量明显高于工作日的订单数量，这也说明周末出行的人增多且大量的人使用滴滴打车作为出行方式；
- (4) 最后，对比非节假日（图 2-1(a)-(b)）和节假日（图 2-1(c)-(d)），节假日滴滴出行订单数量远远大于非节假日的订单数量（国庆节期间当天订单数量增加明显，国庆节前一天为本次订单数量最大，主要原可能为国庆节到海口的人数增加），因此节假日会对海口市交通流量演化存很大影响，因此市民在这些特殊日子前可以提前做好出行规划和城市交通可以提前进行管理减轻订单流量的突然增加带来的道路拥堵问题。

然后，我们分别从三个方向来分析海口市的交通需求演化情况，包括出行时间和出

行距离方面，出行区域的订单量方面和每天的出行数据统计方面。

1. 出行时间与出行距离

系统右下方折线图（图 2-2）显示了一天的订单数量与出行距离随时间的变化。针对以上属性，我们对节假日与非节假日之间的关系进行了分析，分析结果如下：

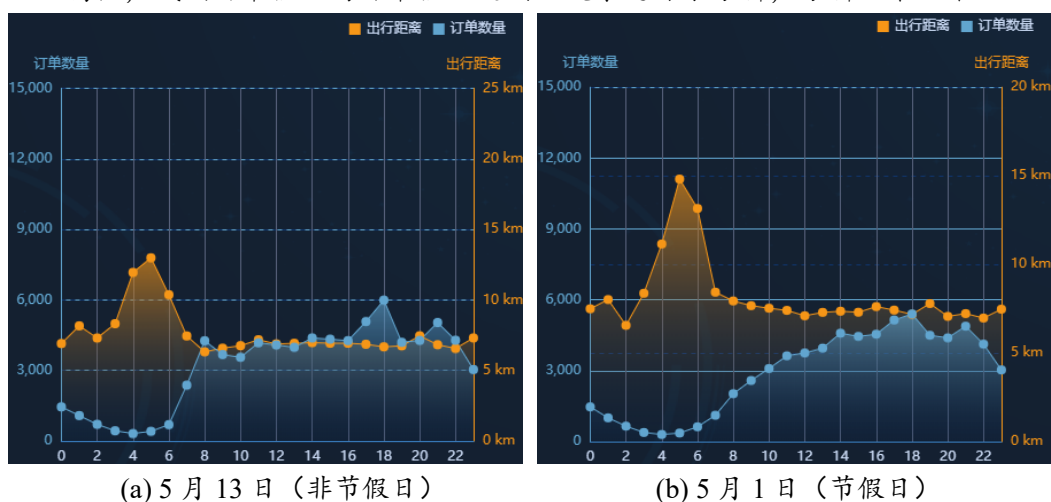


图 2-2 交通流量变化折线图

1.1 节假日与非节假日订单数量的变化：

非节假日

- 1) 订单数量在 4:00—8:00 之间呈现增长趋势，且在 6:00—8:00 之间增加最为迅速，这可能是因为早高峰的出现；
- 2) 订单数量在 8:00—16:00 之间趋于稳定，这可能是因为早高峰的结束；
- 3) 订单数量在 16:00—18:00、20:00—21:00 之间再次呈现增长趋势，这可能是因为晚高峰的出现以及市民夜生活的结束；
- 4) 订单数量在 21:00 以后开始呈现下降趋势，意味着一整天的结束，新的一天开始。

节假日

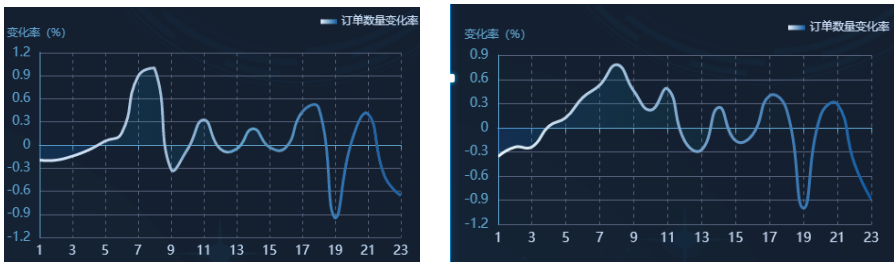
- 5) 订单数量在 4:00—11:00 之间一直呈现增长趋势，这可能是因为假期没有早高峰，市民出行的时间各不相同；
- 6) 订单数量在 11:00—14:00、16:00—18:00 之间呈现缓慢的增长趋势，这可能是因为到了中午饭点（或者部分市民选择午后出行）以及晚饭时间；
- 7) 订单数量在 20:00—21:00 之间呈现增长趋势，可能是因为夜间游玩结束，市民回家（或酒店）；
- 8) 订单数量在 21:00 之后呈现下降趋势，意味着一整天的结束，新的一天开始。

1.2 节假日与非节假日出行距离的关系：

从图 2-3 可以看出，海口市市民滴滴出行距离在节假日和非节假日之间没有很明显的区别：

- 1) 越早出发的市民行程越远，对非节假日而言，由于市民居住地离工作地点较远，需要早点出发；对于节假日而言，由于市民游玩地相较于市民居住地较远，也需要提早出发。
- 2) 不论是节假日还是非节假日，0:00—5:00 的订单数量虽少，但都是长距离出行，这可能是因为在这个时间段内，公共交通还未开放，前往机场、车

站或从机场、车站到达之类的提前预定好的订单，因此建议相关部门提供机场、车站通宵 Bus 等。



(a) 5 月 13 日（非节假日） (b) 10 月 1 日（节假日）
图 2-3 订单变化趋势图

从订单变化率视图可以看出，节假日 5：00—14：00 之间，订单变化率为正，即订单数量呈增长趋势，可能是假期出行的影响；而在 16：00—18：00、20：00—21：00 左右再次呈现增长趋势；非节假日 5：00—9：00、16：00—18：00 之间，订单变化率为正，即出现早高峰和晚高峰，这符合以上通过折线图对节假日和非节假日出行订单的分析结果。

2. 出行区域

主视图（如图 2-4）在地图上使用蝌蚪图、热力图和迁徙图对每个区域的出行订单量及其变化分别进行了展示。通过添加热力图，并进行缩放，可以看出海口市交通流量主要分布于海口东站、海口西站、上邦百汇城、友谊商场、明珠广场、海口汽车东站、海南省人民医院等区域。可以看出，车站（海口东站、海口西站和海口汽车东站），商城（上邦百汇城、友谊商场和明珠广场），医院（海南省人民医院）等地方都是人比较多的地方，而且这方的流动性也是非常大的。因此，对于这些地方，相关部门应该重点关注并提出一些长期有效的建立以减轻交通拥堵。

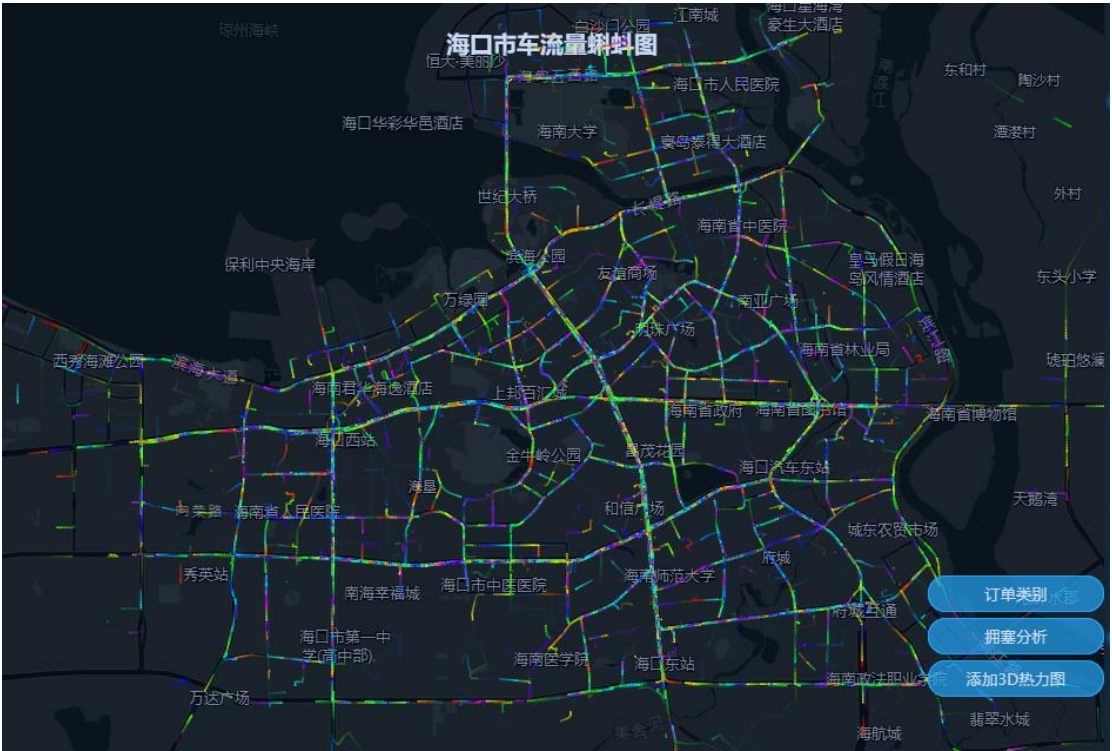


图 2-4 海口市交通动态蝌蚪图

通过添加海口市区域行政区域迁徙图，如图 2-5 所示，可以看出海口市滴滴订单主要分布于龙华区、秀英区、美兰区和琼山区。从图中可以看出，龙华区相关的订单量（龙华区为起点或龙华区为终点的订单）是最大的，因此我们以龙华区作为重点分析区域。



图 2-5 海口市行政区域订单流量迁徙图

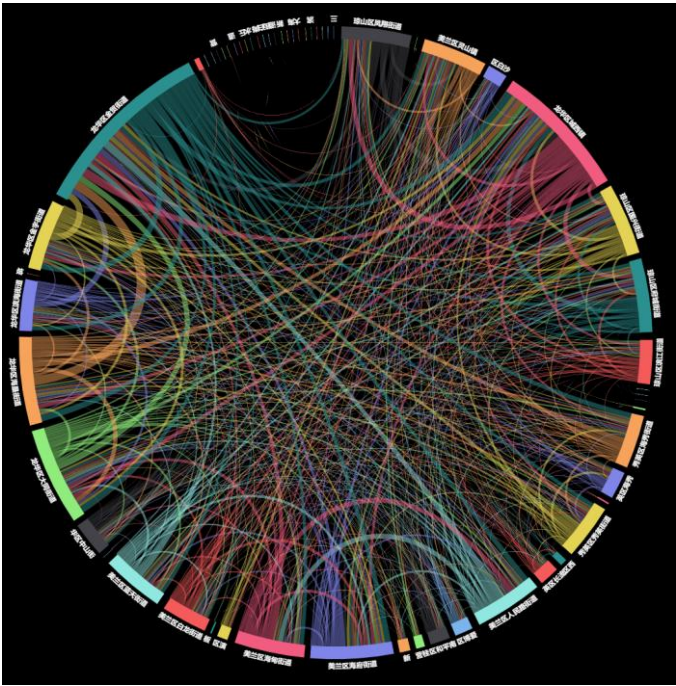


图 2-6 街道滴滴订单和弦图

订单起终点街道和弦关系图展示了海口市不同区域与街道之间的流通交互的关系，从图 2-6 可以看出，海口市交通流量最大的两个区域为龙华区城西镇，其次是龙华区金贸街道，再次是龙华区大同街道、龙华区垦海街道等。由此可见，海口市龙华区的交通流量相较于其他区更大。

3. 出行数据统计

系统通过订单的实时统计如图 2-7 所示，包括长途订单数量、快车数量、高价订单数量以及高时长订单数量。可以看出长途、高价订单数量很少，大部分人选择快车出行、非高价方式出行。



图 2-7 海口市滴滴订单类型实时统计图

问题 2. 请您分析 2017 年 5 月 1 日—10 月 31 日海口市的出行变化最显著的 3 个特征，以及每个特征相关联的主要影响因素，包括但不限于季节、天气、节假日、交通管制等；

根据以上视图的联动展示，可以发现海口市交通出行变化最显著的 4 个特征为：节假日、季节、出行时间段和行政区功能。

1. 节假日、天气、季节以及出行时间段对交通路况的影响：

1.1 长途订单少且多发于早上 5 点；订单数量呈现规律性上升下降，与时间节点密切相关：

从左下方折线图可以看出，早上 5 点左右订单少且路程长，平均超过 10km；此后一天的订单数量无大的波动，但节假日的订单数量普遍高于一般日期，其中节假日平均峰值订单数在 6000 左右，而平时日常平均峰值订单数只在 3000 左右，如下图所示：

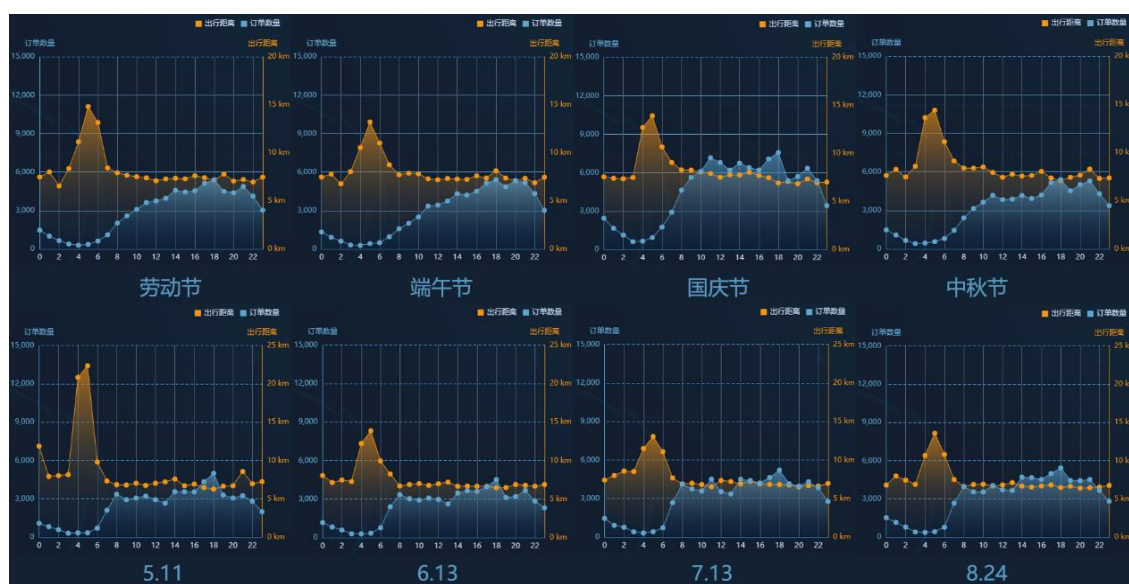


图 3-1 订单数量、出行距离折线图（节假日与平常对比）

劳动节的订单数相较于国庆节更少，是因为在 2017 年 5 月 1 日是劳动节最后一天假期。端午节订单相较于国庆节更少，可能是因为端午节假期比国庆节更短，且中秋节和国庆节假期重合。

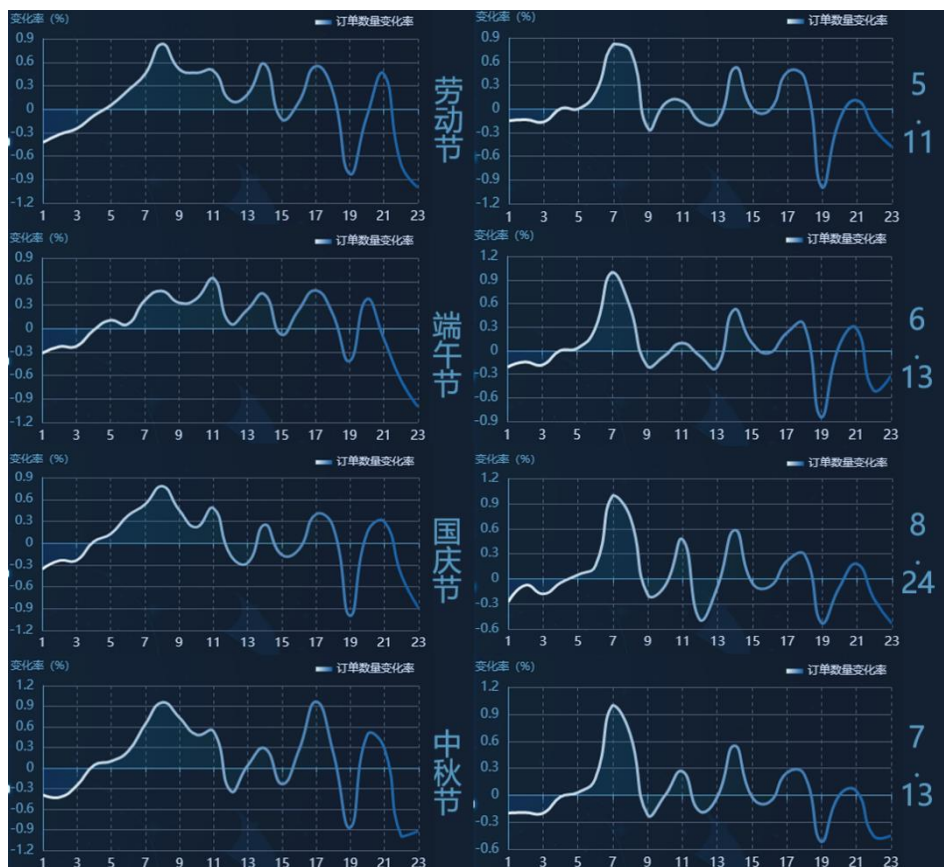


图 3-2 订单变化率曲线图

从订单变化率可以看出，非节假日出行呈现五波峰四波谷。在早高峰时间段（6：00—8：00 左右）、10：00—11：00 左右、13：00—14：00 左右、16：00—18：00 左右以及 20：00—21：00 左右呈现增长趋势；但其增长的程度没有节假日高，且变化更频繁，可能是因为一般工作日人们出行主要是工作和就餐。

节假日的变化基本也遵从呈现五波峰四波谷，不同之处在于其在 5：00—12：00 时间段，订单变化率一直为正，即订单数量呈增长趋势，这是由于假期出行导致与非节假日不同。

这两者最大的差别在于上午节假日的出行订单数一直在增长，而一般日期则只是在上班高峰期和午餐时间呈现增长。

表 3-1 节假日出行订单变化表

时间段	5:00-12:00	12:00-14:00	14:00-15:00	15:00-18:00	18:00-20:00	20:00-21:00	21:00-23:00
变化趋势	增长	增长	下降	增长	下降	增长	下降

表 3-2 非节假日出行订单变化表

时间段	5:00-8:00	8:00-10:00	10:00-14:00	14:00-16:00	16:00-18:00	18:00-20:00	20:00-23:00
变化趋势	增长	下降	先增长后下降	下降	增长	下降	增长

1.2 高温出行少，暑假出行多：

从下方柱状图可以看出海口市市民出行与季节之间的关系，7、8、9、10月的订单数量明显高于5、6月的订单数，这可能是由于暑假和高温的影响。根据温度和出行数据之间的关系，可以看出高温出行的人数较少，随着温度降低，订单的数量逐渐增多。根据温度折线图可以看出5月到10月，海口市平均温度大体介于20℃-30℃之间，比较适合出行，其中从五月到十月有部分天数订单数量出现缺失值，导致缺少了这部分的订单信息。到了假期7、8月，订单整体相比5、6月要多，可以推测是学生放假出行的原因。

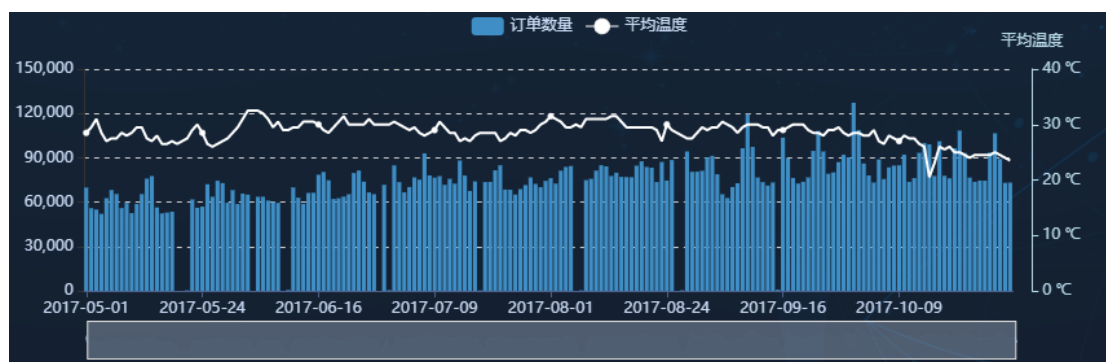


图 3-3 平均温度与订单数量关系图

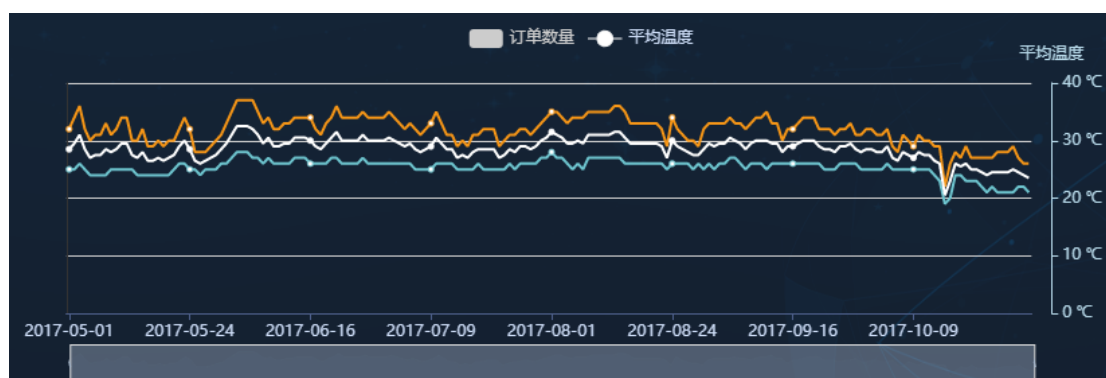


图 3-4 海口市最高温度、最低温度及平均温度关系图

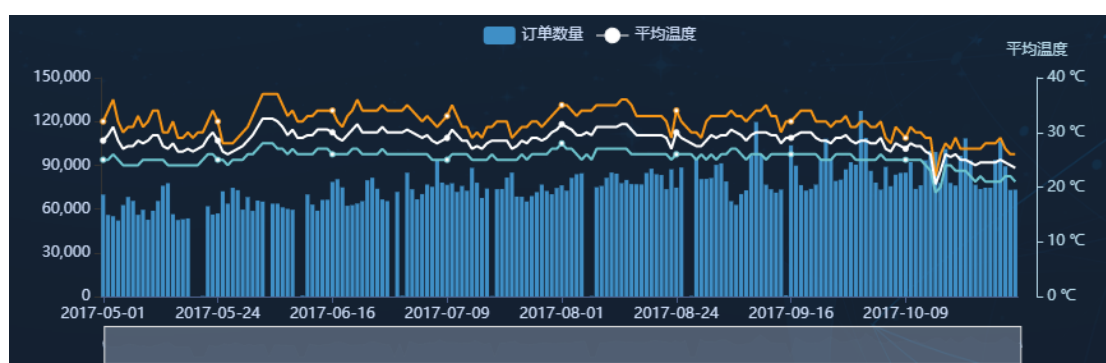


图 3-5 订单数量与温度关系图

1.3 订单密集地呈现点状分布:

(1) 从整体来看:

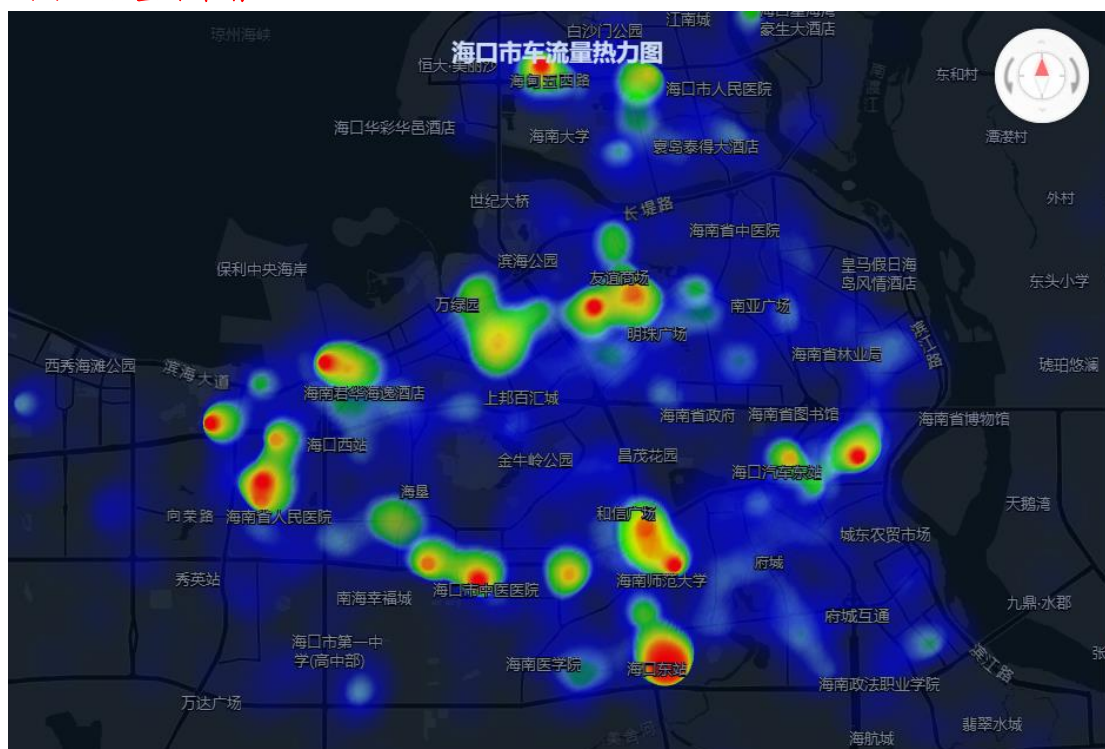


图 3-6 早上 7 点 海口市地图热力图
(红色代表订单数量很多, 蓝色代表订单数量相对较少)

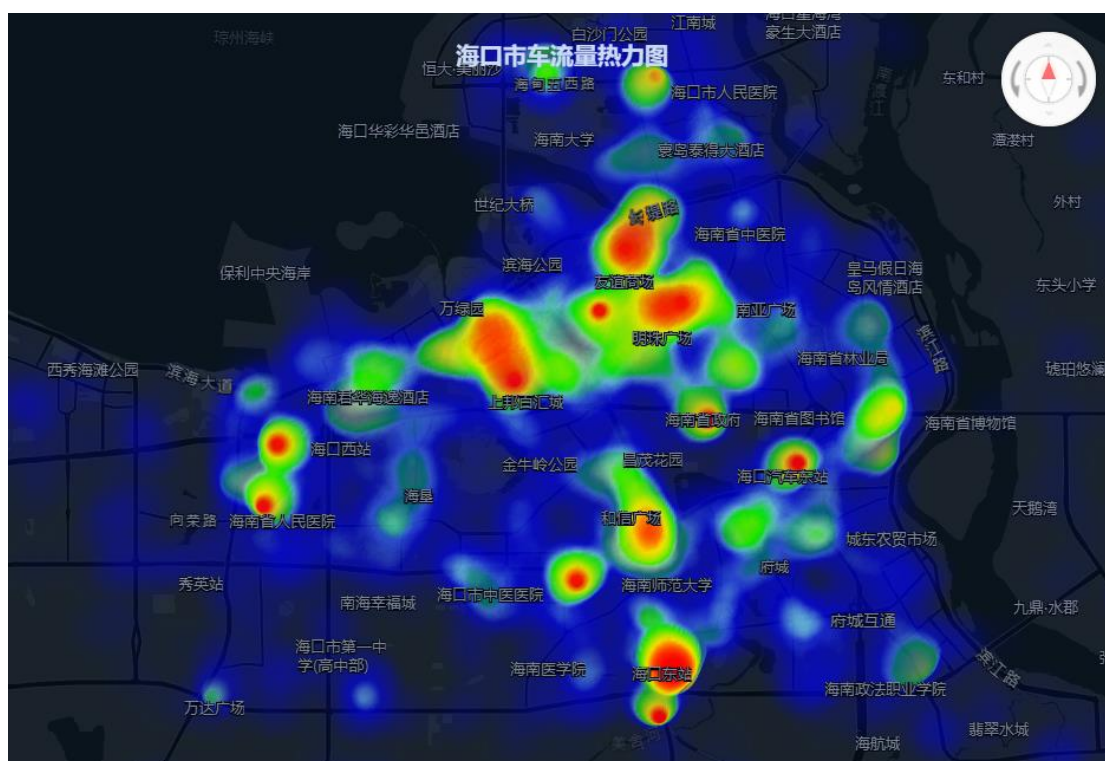


图 3-7 海口市地图热力图 (早上 10 点)

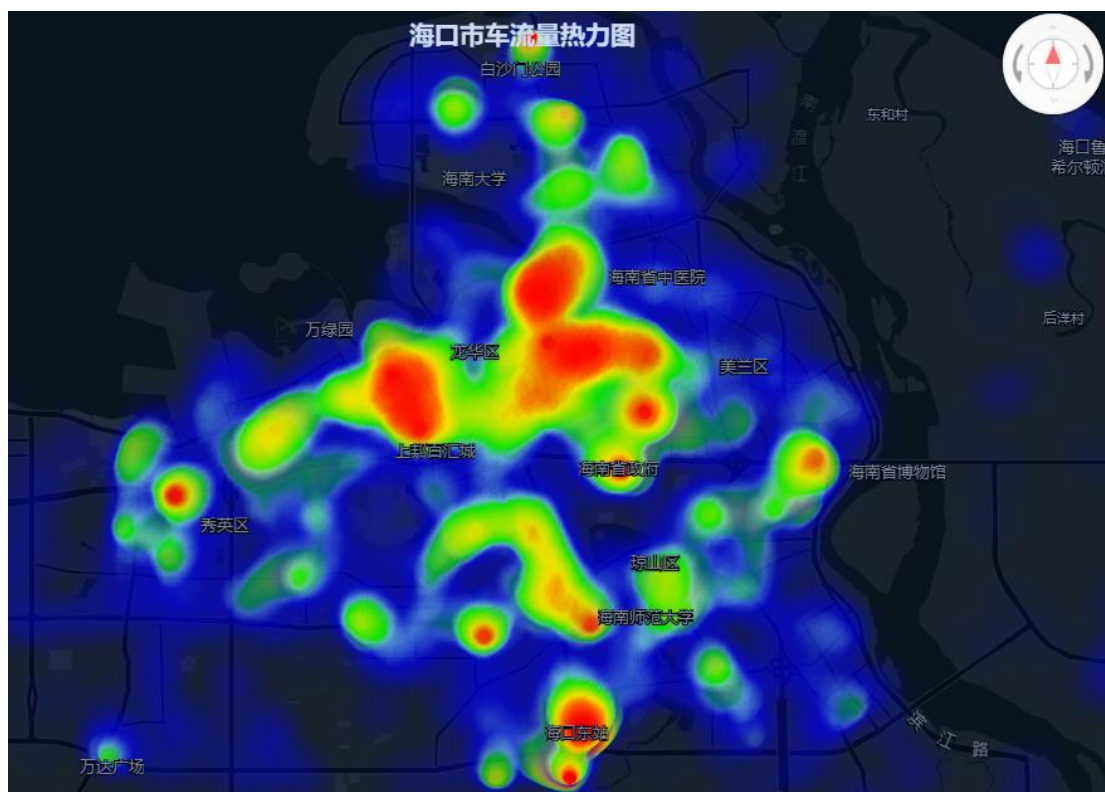


图 3-8 海口市地图热力图（下午 2 点）

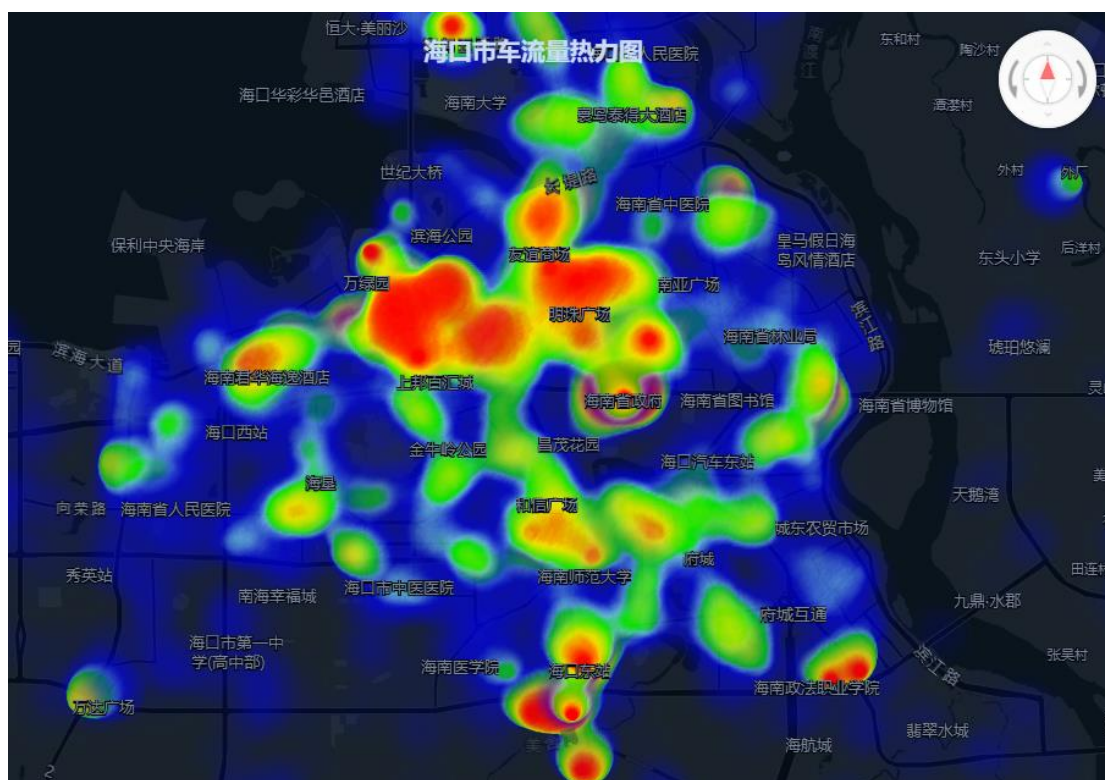


图 3-9 海口市地图热力图（晚上 9 点）

从上方热力图可以看出海口市市民出行与地点之间的关系。通过观察热力图变化可以发现，一天中海南师范大学、海口汽车东站、海口市中医医院、海口市人民医院、海

口美兰国际机场、海南省政府、万达广场、友谊商城、海口骑楼老街等（车站、机场、医院、学校、广场、政府设施等公共设施）是热门到达地；19点至凌晨1点则相反，是热门出发地，特别是广场地区。

(2) 从局部来看：

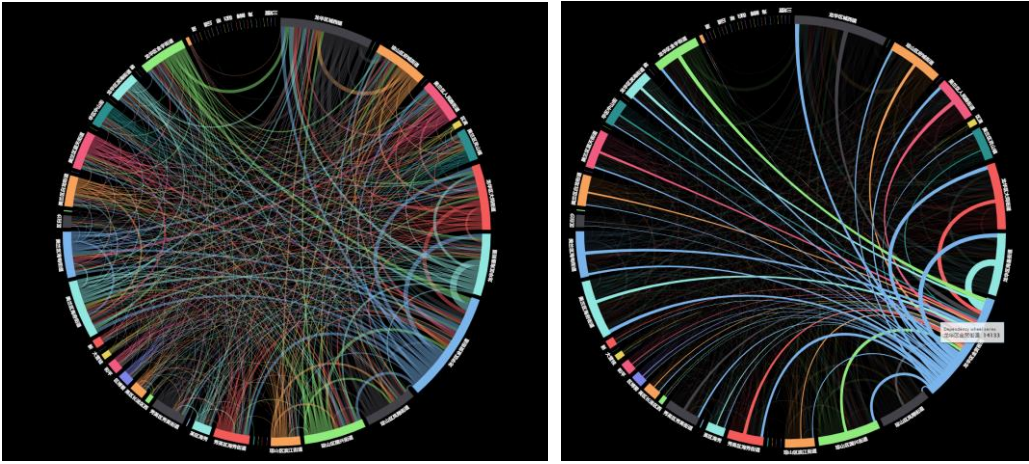


图 3-10 非节假日订单街道订单和弦图

从上方和弦图可以看出海口市市民出行与街道之间的关系，龙华区经贸大道的订单数量(14133)明显高于其他街道的订单数，这可能是由于该街道为上班族较多的地方。其订单数量均匀的分布到各个区域。

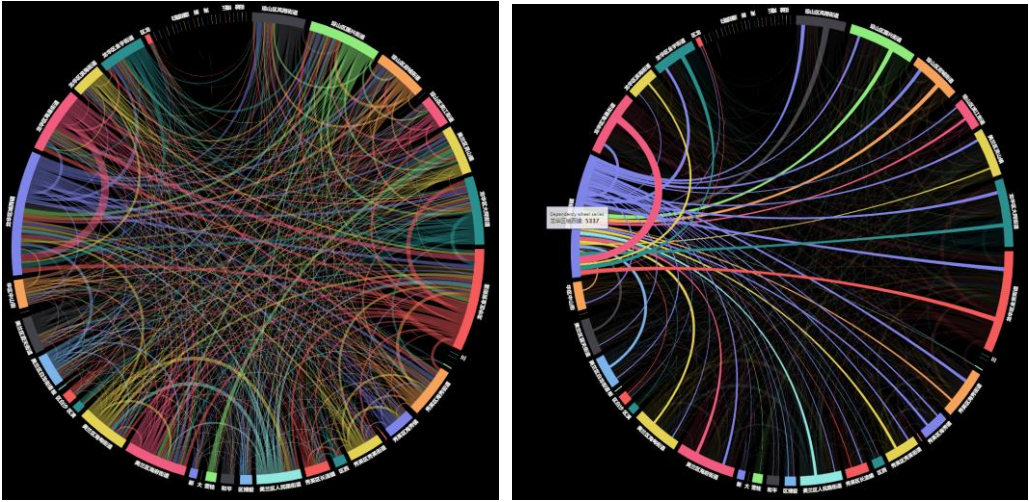


图 3-11 节假日订单街道订单和弦图

从上方和弦图可以看出海口市市民出行与街道之间的关系，龙华区域西镇的订单数量(5337)略高于其他街道的订单数，和龙华区经贸大道的订单数量(4444)相差不大。由于节假日订单分布较为均匀，和非节假日对比明显。

2. 市区行政区之间对交通路况的影响：

根据订单流动迁移图，图 3-12 与 3-13 所示，可以看出龙华区的订单量最多，而其他较多订单量的区域，如美兰区，秀英区等，都与订单量最多的龙华区具有流量边

相连。这些较多订单量的区域通常会聚集到订单量最多的龙华区，而订单量最多的龙华区的流量会分散流入到各个较多订单量的区域。例如秀英区，这是因为秀英区包含海口市车站，美兰区对于当地民众包含了很多工作与学习生活的地点以及旅游可能会去的相关地点，也是订单起点与终点的地方。

再根据图 3-14 行政区订单数量热力图可以清晰地看到海口市西站以及海口东站是大量滴滴订单的发起地点以及到达终点，龙华区为海口市市中心，所以平日民众需要很大的出行需求。根据中午十二点以及晚上八点以及二十三点的地图订单情况热力图可以清晰地看出这一结论。



图 3-12 海口市行政区订单流动迁徙图（非节假日）



图 3-13 海口市行政区订单流动迁徙图（节假日-国庆节）

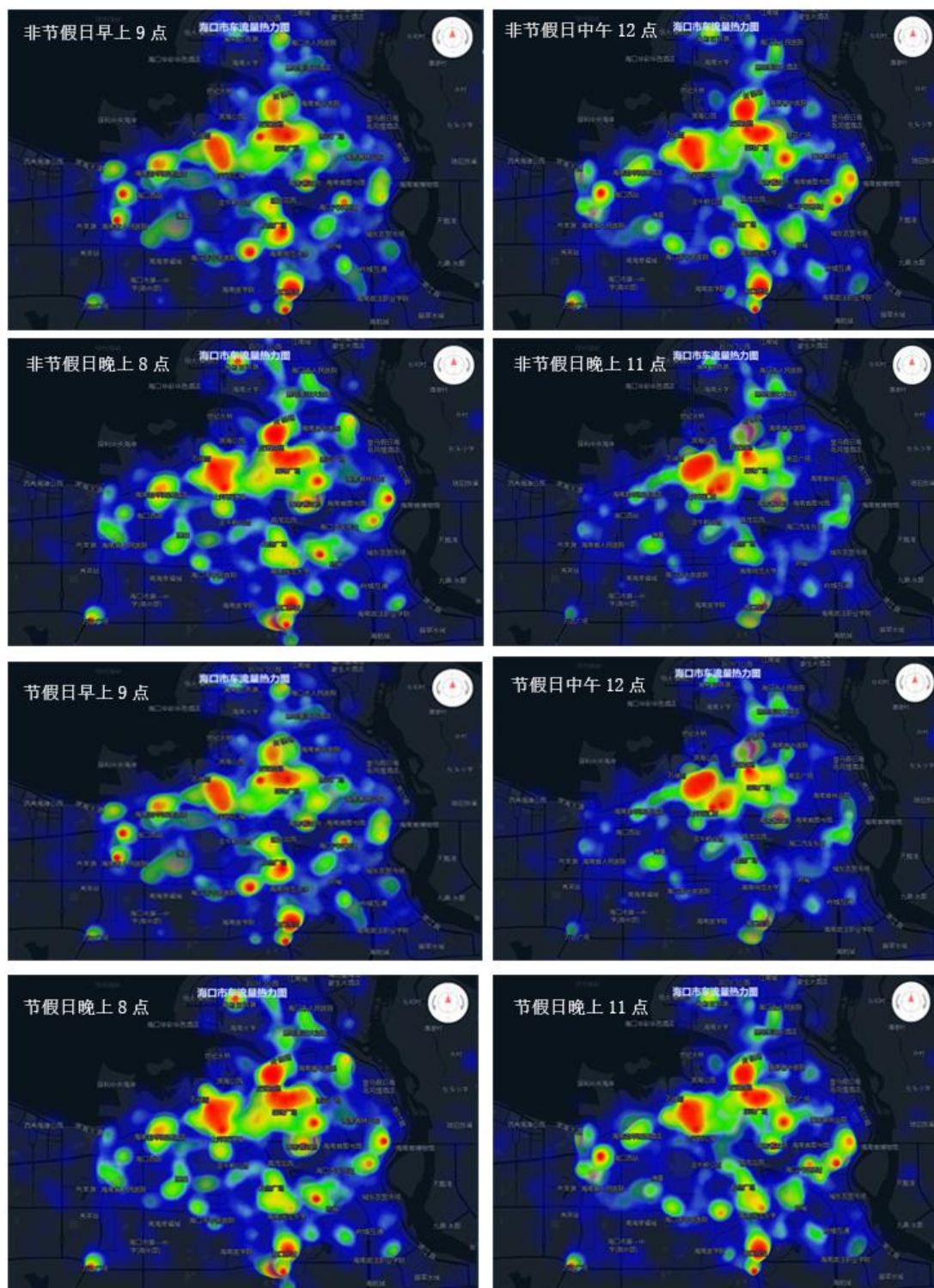


图 3-14 海口市行政区域订单数量热力图

再根据图 3-15 海口市街道路况热力图可以清晰地看到海口市无论是工作日还是节假日大量滴滴订单产生并有可能造成拥堵的区域是龙华区，订单各类型主要涉及的街道是龙昆南路、龙昆北路、南大立交桥、海秀快速路、海秀中路以及国兴大道等。这些街道都是滴滴订单车辆的必经之地。

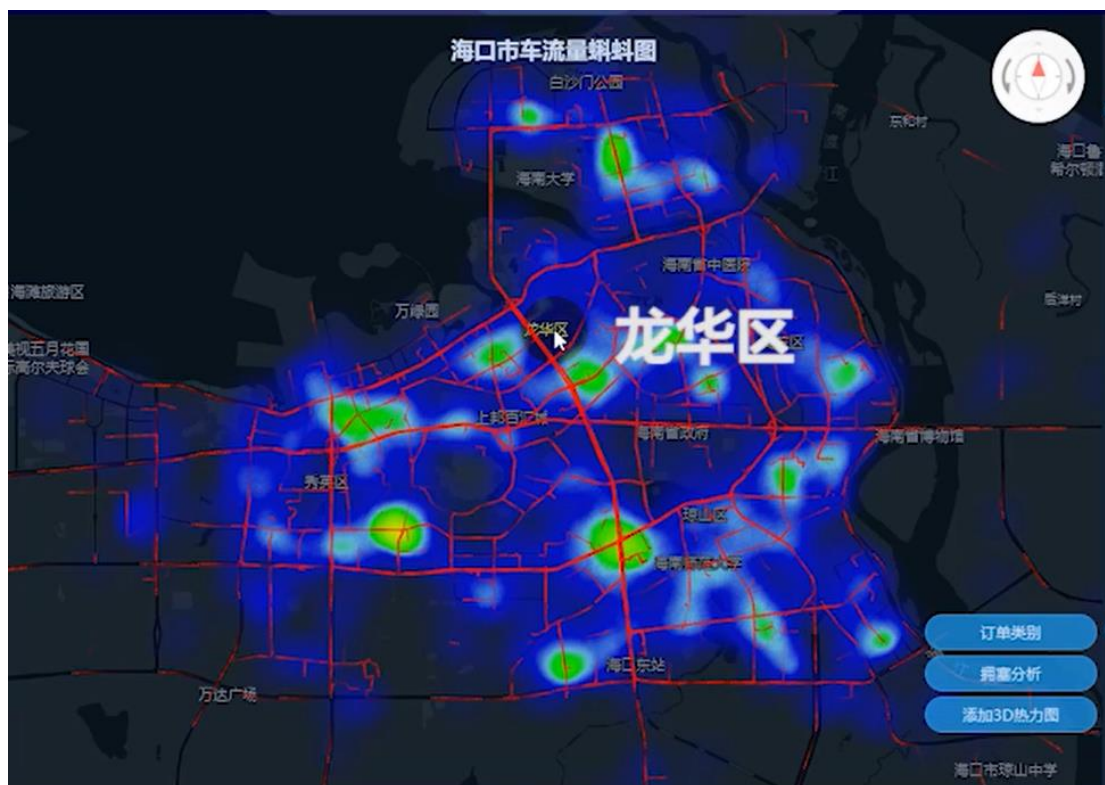


图 3-15 海口市街道路况热力图

3. 请您根据上述数据分析，为市民或城市交通管理部门提供建议，如面向市民的雨天出行建议、面向交通规划部门的公共交通线网规划和站点设置建议等。

根据以上分析，我们发现海口市市民出行受节假日、时间段和季节的影响较大。因此，我们使用深度学习算法来预测未来的订单量，市民和交通管理部门可以根据预测的订单量和天气预报（现有的天气预报已经较为准确）来提前做好出行和交通管理的准备。

算法介绍：考虑到不同时刻的订单量具有很明显的时间序列特征，而长短期记忆网络（LSTM, Long Short-Term Memory）是一种时间循环神经网络，因此我们使用 LSTM 来预测订单量。LSTM 循环单元包含三个门（gate），分别负责遗忘哪些历史信息（Forget gate）、增加哪些历史信息（updating gate）、以及输出门（Output gate）。模型图如图 4.1 所示，公式化表示如下所示：

$$\Gamma_f^{(t)} = \sigma(W_f[a^{(t-1)}, x^{(t)}] + b_f)$$

$$\Gamma_u^{(t)} = \sigma(W_u[a^{(t-1)}, x^{(t)}] + b_u)$$

$$\tilde{c}^{(t)} = \tanh(W_c[a^{(t-1)}, x^{(t)}] + b_c)$$

$$c^{(t)} = \Gamma_f^{(t)} * c^{(t-1)} + \Gamma_u^{(t)} * \tilde{c}^{(t)}$$

$$\Gamma_o^{(t)} = \sigma(W_o[a^{(t-1)}, x^{(t)}] + b_o)$$

$$a^{(t)} = \Gamma_o^{(t)} * \tanh(c^{(t)})$$

σ 和 \tanh 是激活函数， W_f ， W_u ， W_c 和 W_o 是权重矩阵， b_f ， b_u ， b_c 和 b_o 是模型的 biases。在实际训练过程中，我们首先对订单量数据进行归一化，然后通过前两天的数据来预测第三天的数据。训练集和测试集的比例为 7:3。LSTM 层数为 2，隐藏层大小为 4。模型结构为 LSTM 层接全连接层。损失函数使用的是均方误差损失 nn.MSELoss()，

公式为 $Loss = \frac{1}{2n} \sum_x \|y(x) - a^L(x)\|^2$ 。优化算法选择 Adam，学习率为 0.01，训练轮数为 5000 轮。

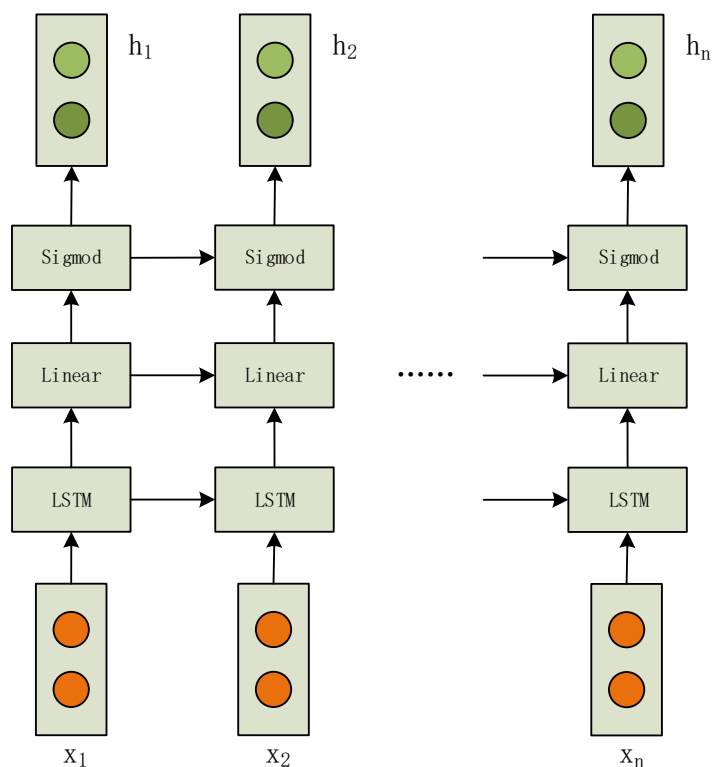


图 4-1 订单量预测模型图

本次的预测分为总体预测和局部预测，通过总体和局部结合，为交管部门及时的交通管控提供有效的参考。

总体预测，通过所有的数据，通过 LSTM 算法模型，预测未来一周订单总量，并预测出订单总量缺失的数据；

通过该算法预测我们可以很清楚地看到模型的预测结果与之前实际订单数量差距很小，可以说明我们使用的算法的有效性，该算法还可详细到行政区以及街道的订单预测，由图 4-2 与图 4-3 可以看出这一结论。

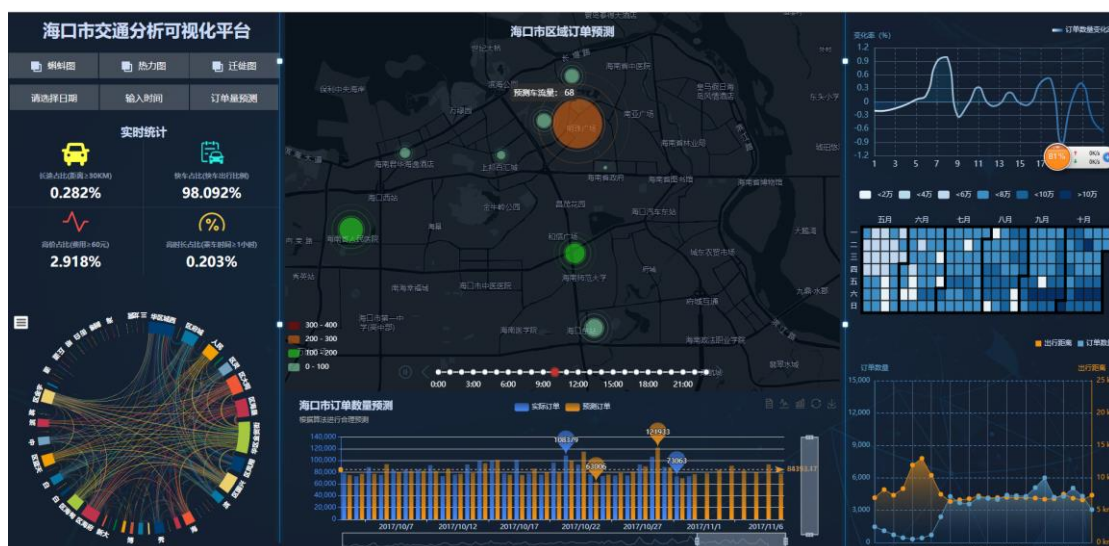


图 4-2 订单预测情况



图 4-3 订单预测整体效果情况

局部预测，通过历史数据，选区海口市 9 个重点区域来对他们的订单进行实时预测，每个区域通过四个坐标来确定，通过坐标关系过滤出 5 个月内在该区域内出现过的所有订单，通过 5 个月的所有订单通过 LSTM 来进行预测，为交通管理提供有力参考。**区域的选区及相应坐标如下图 4-4 所示：**



区域：（车站->动车站和汽车站）

#订单量预测处理

110.342046,19.986007 （左上）

110.350285,19.986733 （右上）

110.342775,19.978707 （左下）

110.348912,19.978304 （右下）



区域：（海南省政府）

#订单量预测处理

110.345992,20.020268 （左上）

110.351249,20.020087 （右上）

110.345927,20.017486 （左下）

110.351613,20.017123 （右下）

下图为 9 个重点区域预测数据变化图 4-5：

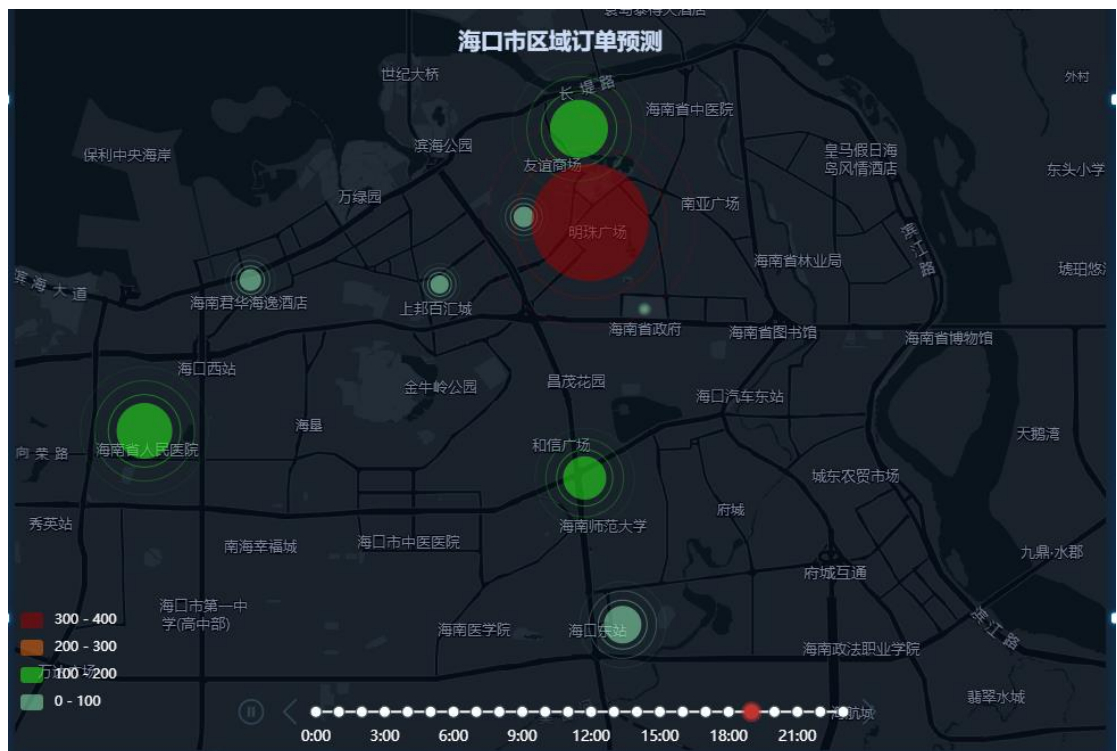


图 4-4 9 个重点区域预测数据变化图

上图为九个城市重点区域预测随时间变化订单变化散点图，散点图越大，表示订单数量越多，当订单量大于 300，就会弹出拥堵警告弹框，如下图 4-5 所示。

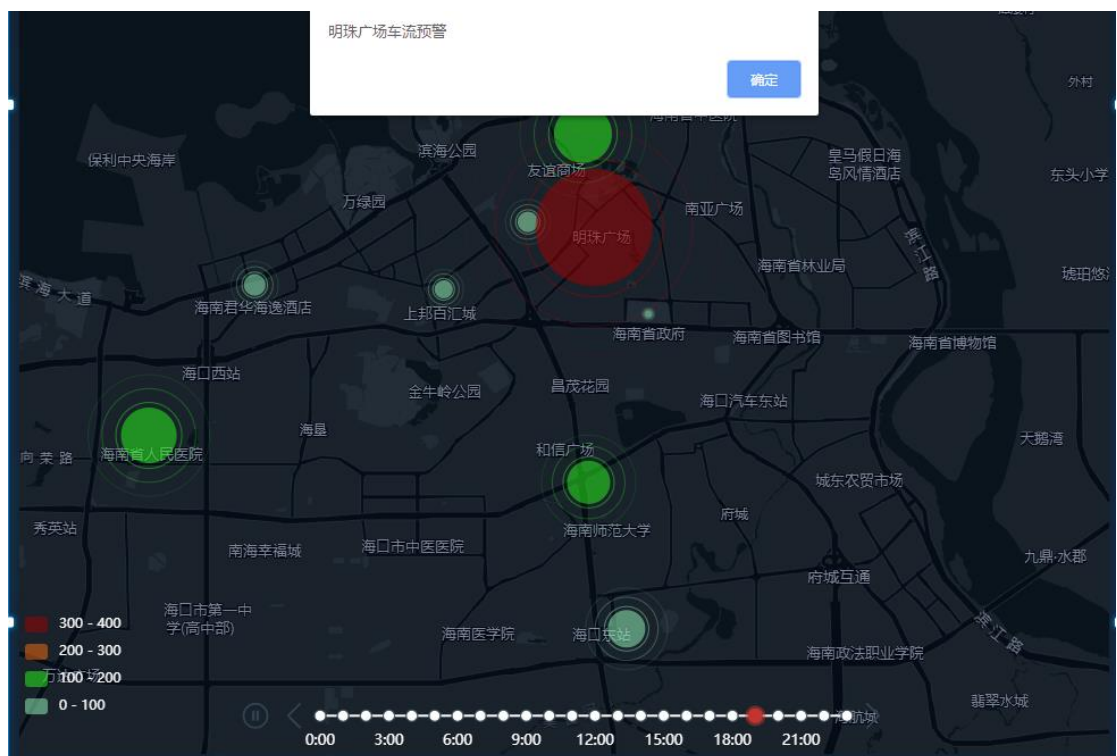


图 4-5 区域订单量大于 300 弹出拥堵预警

为此，针对市民和交通管理部门，我们提出以下建议：

1. 针对市民：

- 1) 高温减少出行，并错过早晚高峰、节假日等；
- 2) 7、8月容易突降暴雨，出行请做好防护措施；
- 3) 雨天打车的人会急剧增多，导致车辆分配不够和道路拥堵等问题，因此雨天市民应提前作好出行规划；
- 4) 提前查看天气预报和出行的交通预测量，安排出行时间。
- 5) 在预测数据可得，早上九点后明珠广场车流较多，一直会持续到晚上 23，市民出行尽量避免该区域。

2. 针对交通管理部门：

- 1) 工作日市民出行具有周期性，可适当在早晚高峰增加公共交通（地铁、公交等）数量；
- 2) 市民在龙华区与琼山区之间出行次数最多，可适当增加这两个行政区之间的长途公共交通；（需要根据迁徙图得出具体的行政区）
- 3) 在公共设施附近要适当增加公交站、地铁站（出入口）；在机场、车站等流量大的地区应增设直达班车、地铁、出租搭乘点等；在广场周边适当增加公交站点或新增地铁线路；
- 4) 提前查看天气预报和出行区域的交通预测量，做好拥堵区域的管理避免拥堵；
- 5) 较长的节假日如国庆等，节假日当天或前一天时增加相应公共车辆，加强交通管理。
- 6) 明珠广场车流较多，一直会持续到晚上 23 点，特别在节假日等特殊日子，在该区域可增加交通管制。