**海口市-交通流量时空演变特征可视分析 答卷**

【在线演示】<http://datavisualization.club:8888/traffic/dashboard>

本答卷在使用传统可视化技术的同时，也尝试使用机器学习回答预设问题、具体使用了PangRank，Betweeness（聚类），Arima时序预测，特征相关性矩阵 等技术。同时本答卷给出租车司机提供一些有用的信息，希望帮助滴滴更好得运营。

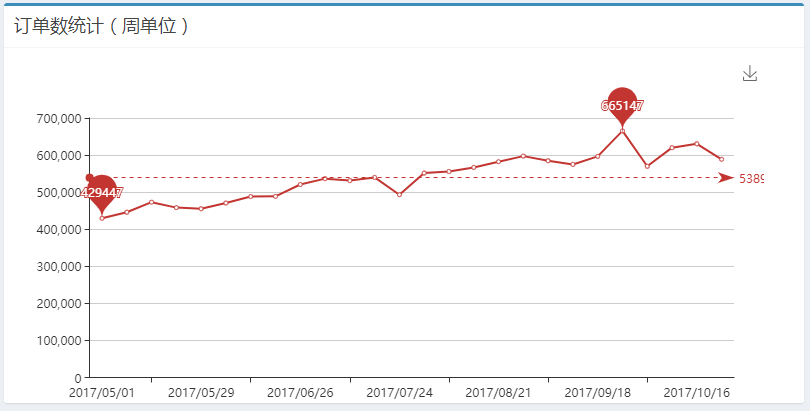
# 1、题目一：

请您分析2017年5月1日-10月31日海口市的交通需求演化情况，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；

**解答：**

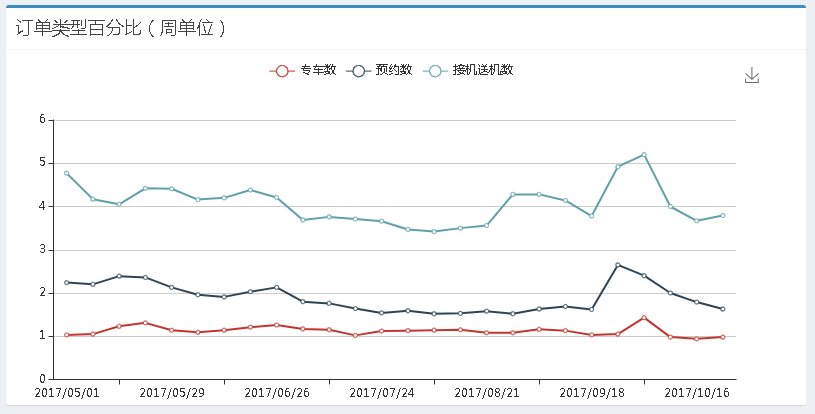
## 1.1概况

**宏观上分析：**首先，我们从宏观上分析了一下订单数量的变化，**发现随着时间的推移，订单量是不断上涨的。**以下是以周为单位的订单数统计，我们可以看到5月1日开始的一周订单量大约为43万单，10月23日开始的一周，大约为60万单，涨幅大约为35%。其中9月25日开始的一周达到峰值为66万单。



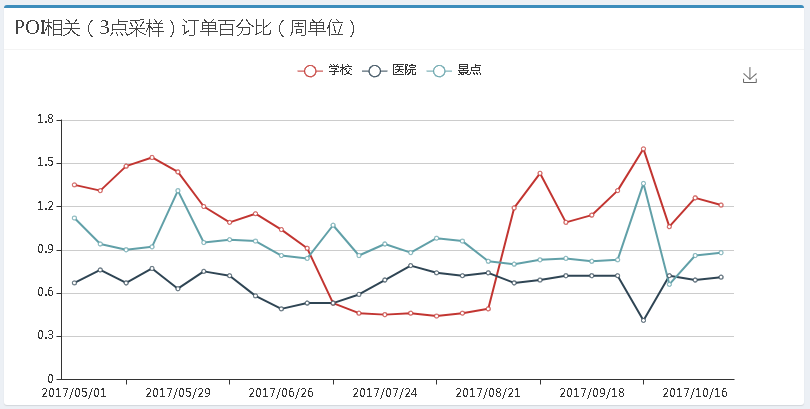
### 1.1.1不同类型的打车变化情况

我们考察了不同类型的打车变化情况。专车数，预约数，接送机数在整体订单中占比是很一致的变化着。考查一下其峰值出现在五一，中秋，国庆这些假期中。

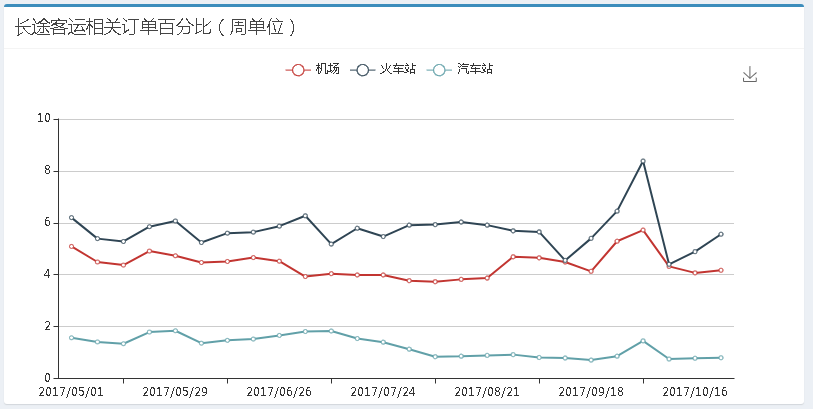


### 1.1.2不同POI的变化

选取了海口市的3家医院，3个学校和3个景点作为基准，考查了订单的变化情况。学校在暑假期间订单量明显减少了，景点的订单比例在假期中明显增多，医院由于订单量变化不大，所以在整体百分比中，假期的时候占比减少了。



机场，火车站，汽车站的订单占比情况和景点的占比情况类似，假期的时候有一个明显的峰值。



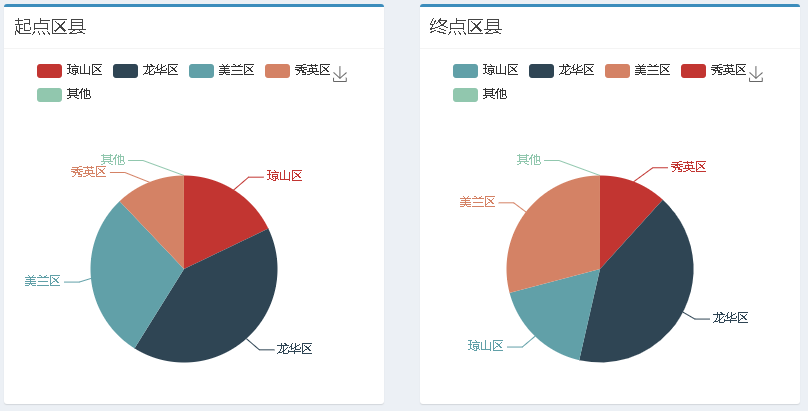
### 1.1.3订单平均距离分析

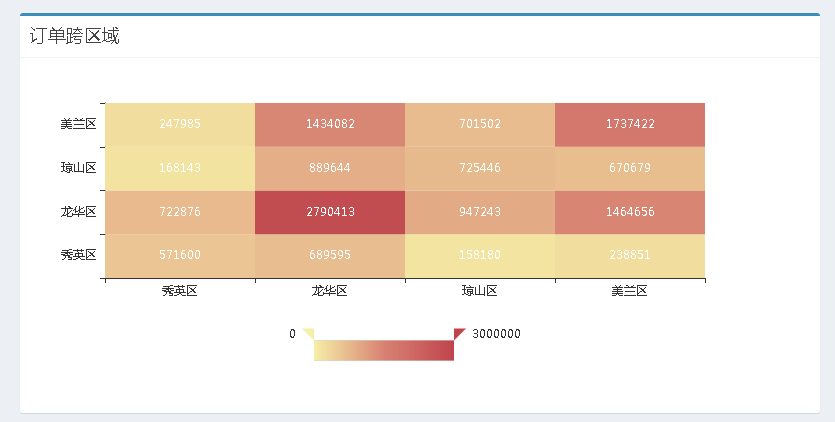
对5-10月所有订单的里程数进行了分析，大致呈现出一种右偏分布，公里数为4的订单数量最多，达到了1874580单。大部分订单的里程数集中在10公里以内。**因此可以在小区域内投放共享单车或者公共交通等，减少短距离的订单量，缓解交通拥挤的同时也在践行绿色环保的生态文明理念。**



### **1.1.4 订单区域分析**

通过对于订单区域分析可以发现，起始点都在龙华区的订单有280万单，占据了整体订单数的20%。这里可以**建议龙华区区内增加公共交通的建设，减少区内的打车量。**

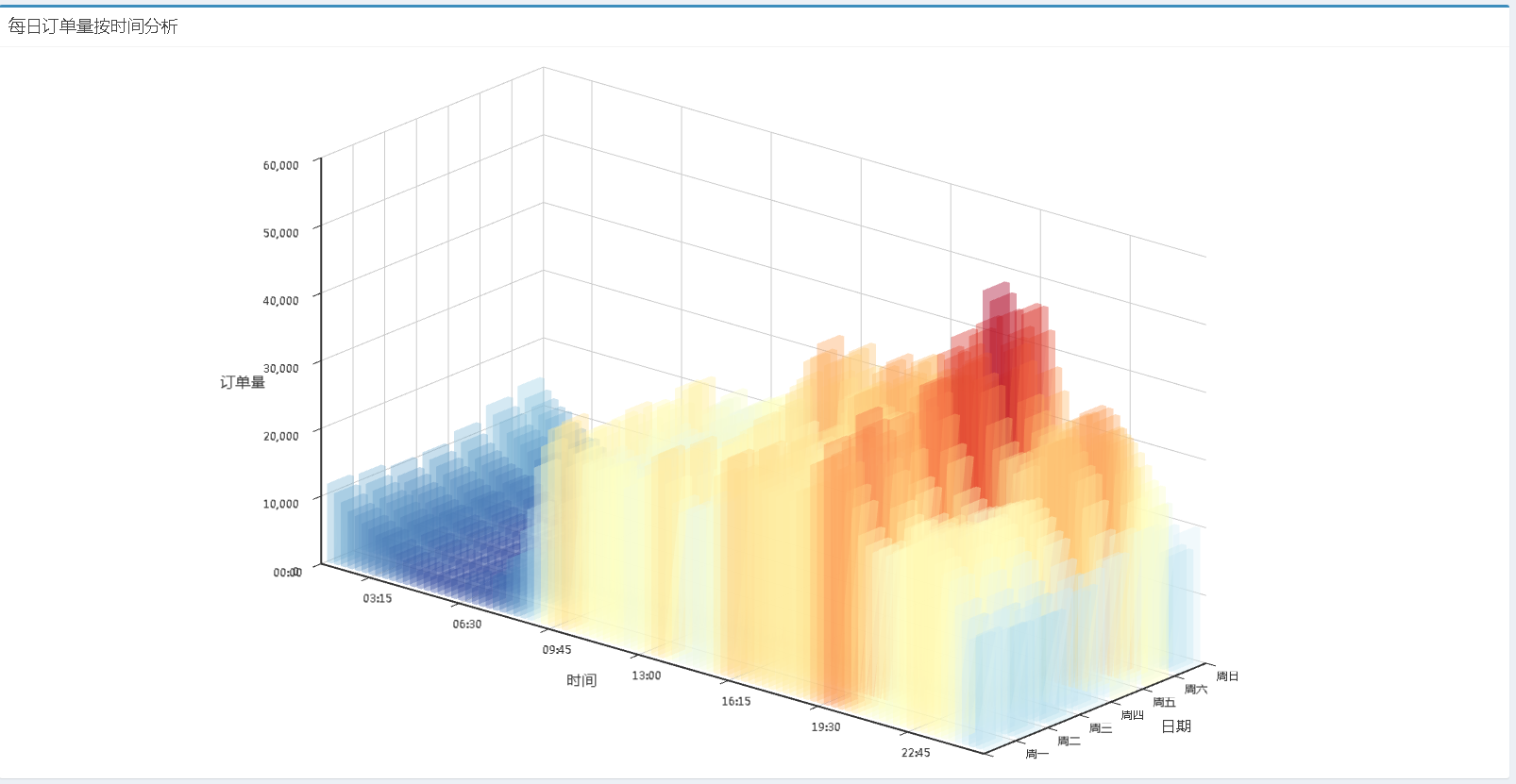




\* 订单终点的行政区是参考起点行政区而来，可能有细微偏差

## 1.2日期分析

对于整体订单在各个时间段，周别上进行了分析。

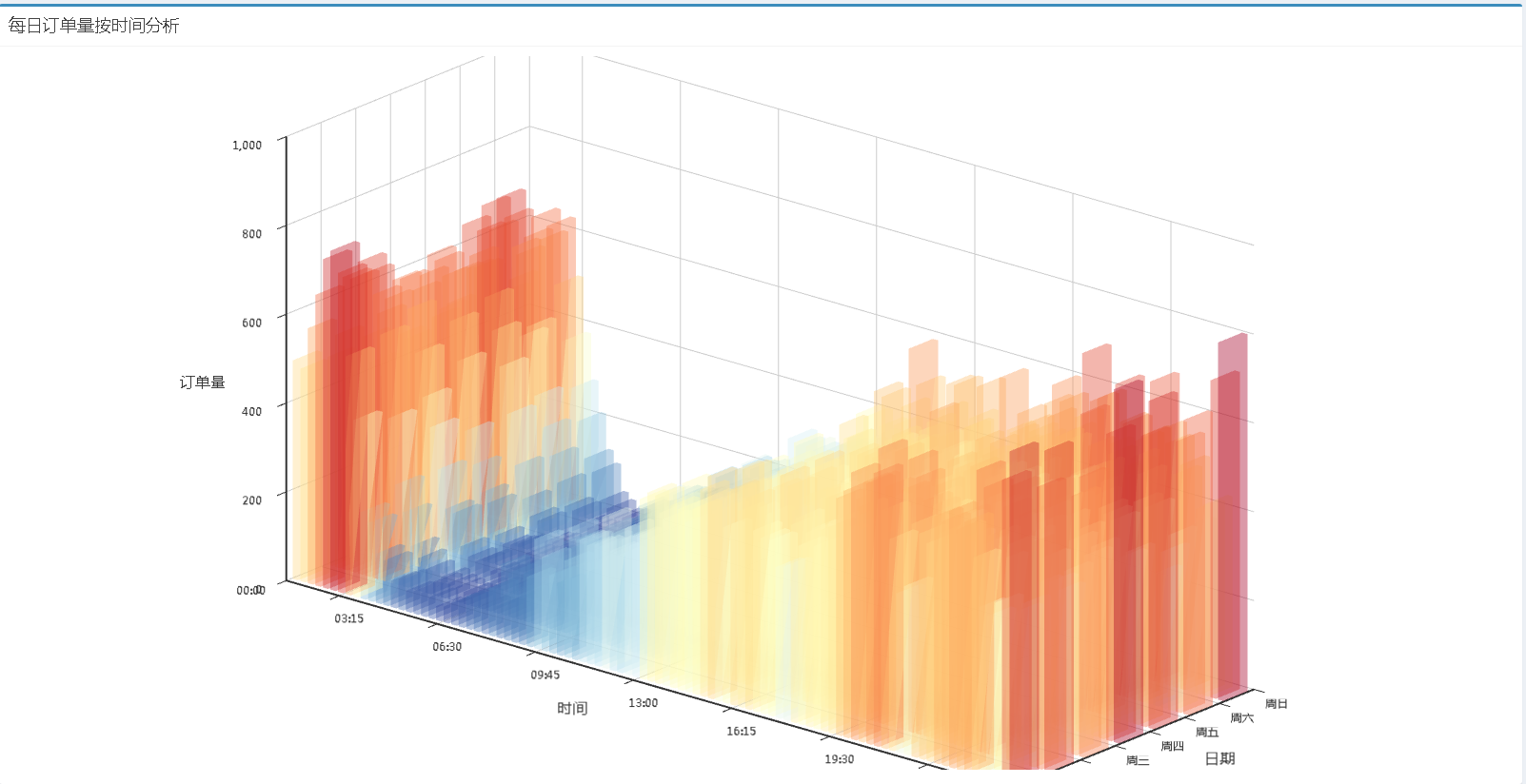


我们可以看到，从时间段来说，**凌晨4点左右是低谷，夜晚8点左右是高峰。**上下班时间的打车量也高于其他时段。从周别上来说，**周五和周六是高峰，周五周六聚会和娱乐是打车的推动因素。**

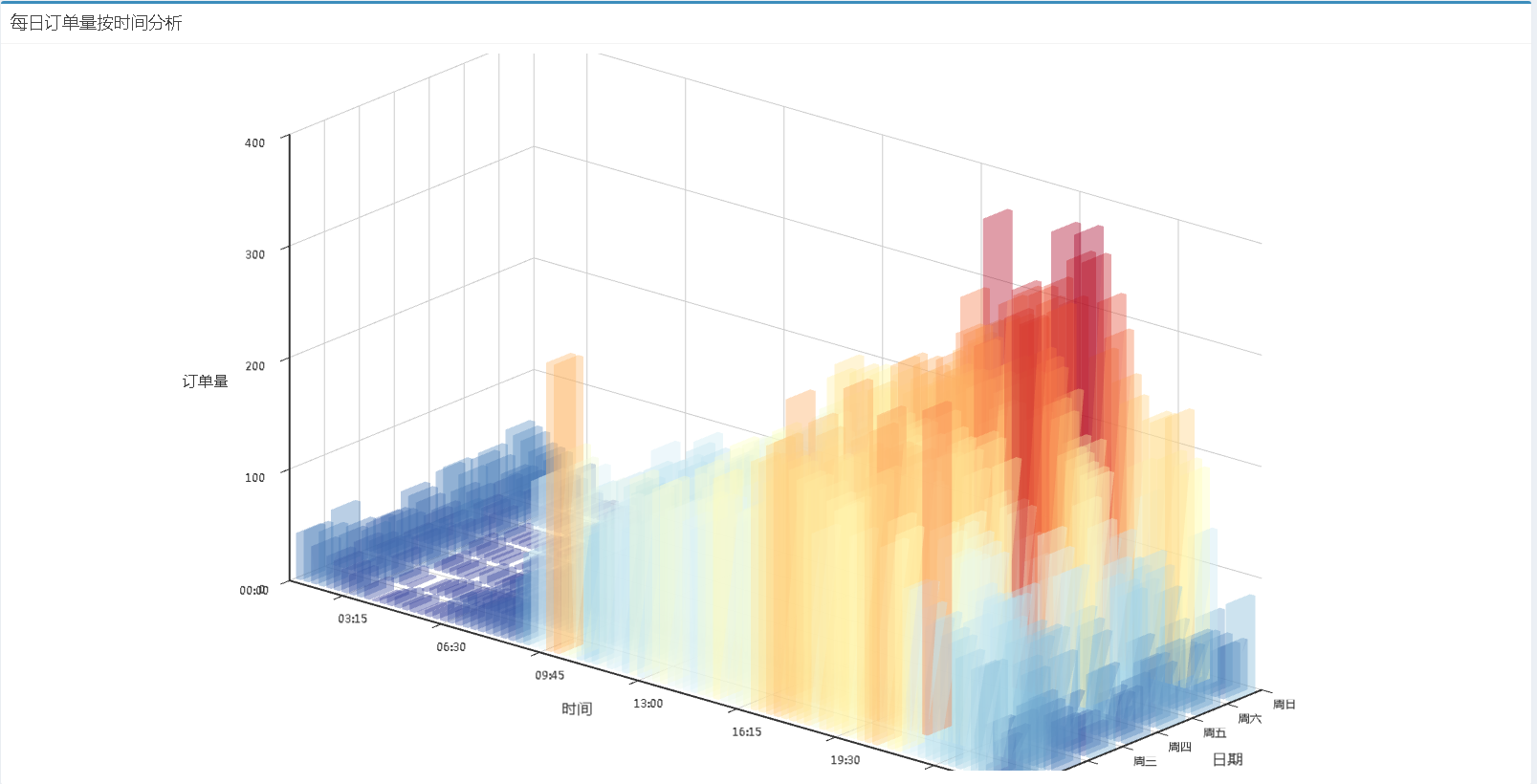
### 1.2.1不同POI的变化

同时我们对于不同的POI场景做了不同的时间段和周别分析

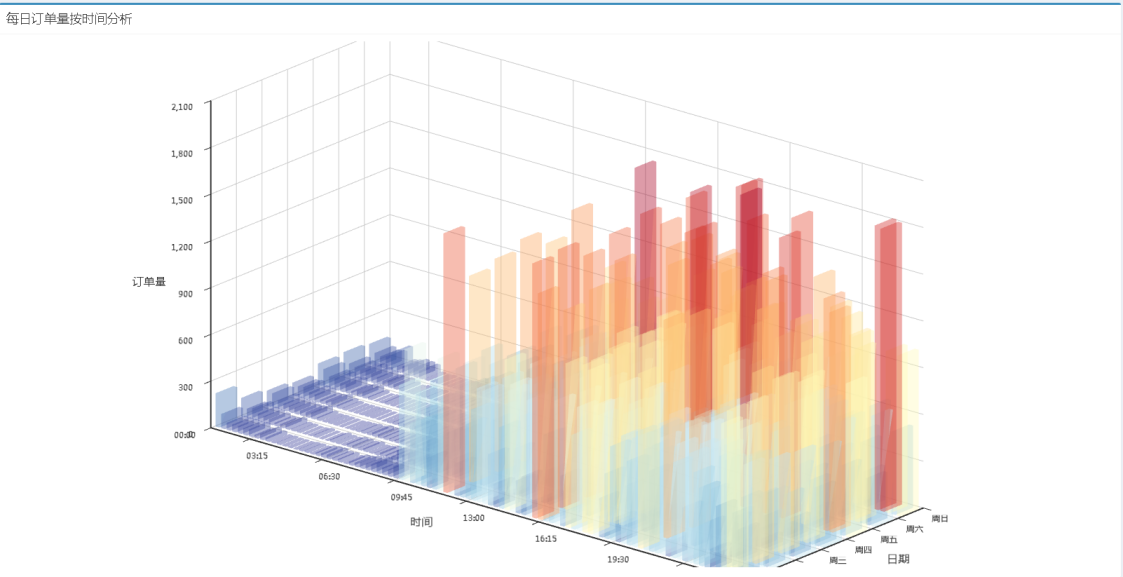
**机场**的状态和航班时间相关，凌晨1-2点也非常繁忙，大致呈现一种早晨7时次日到凌晨1时递增的趋势，大概在晚上12时到次日1时达到了峰值。



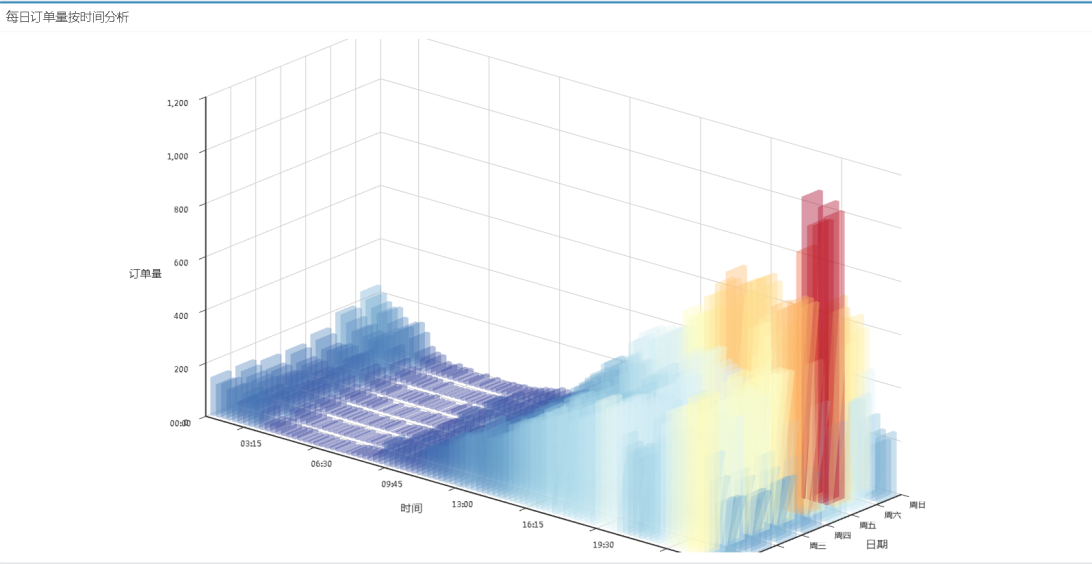
**汽车站**的订单一般集中在下午2点到晚上10点，尤其在周五、周六、周日这三天特别明显。并且订单在11点之后订单开始急剧减少了。特别地，周一早晨，汽车站订单相对于一星期内的其他天是最多的。



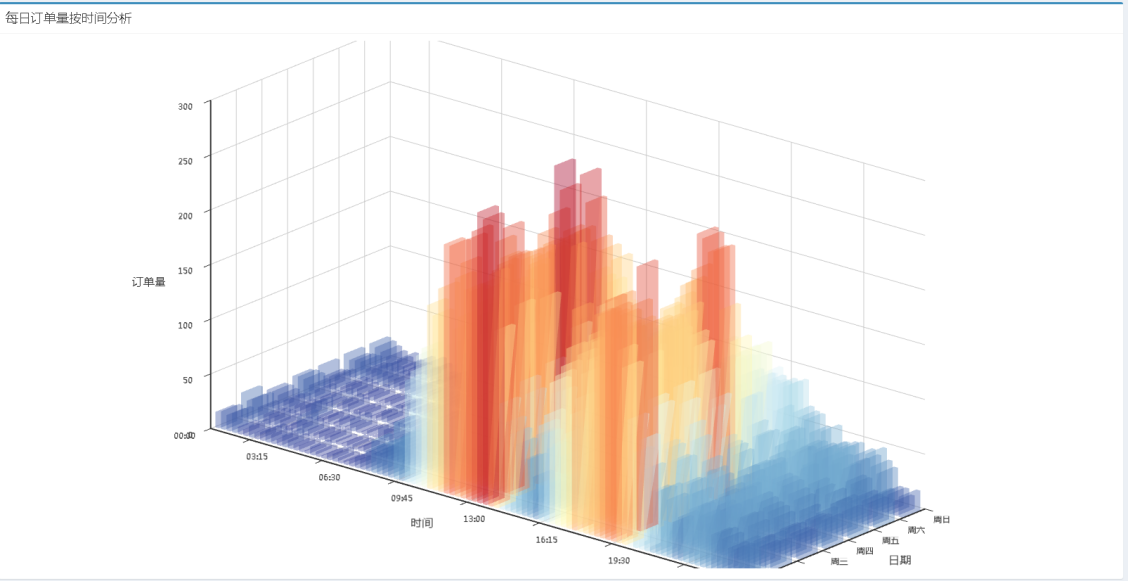
**火车站**的订单主要集中在早上10点到晚上12点，凌晨到早上9点订单数量较少，差距较大。



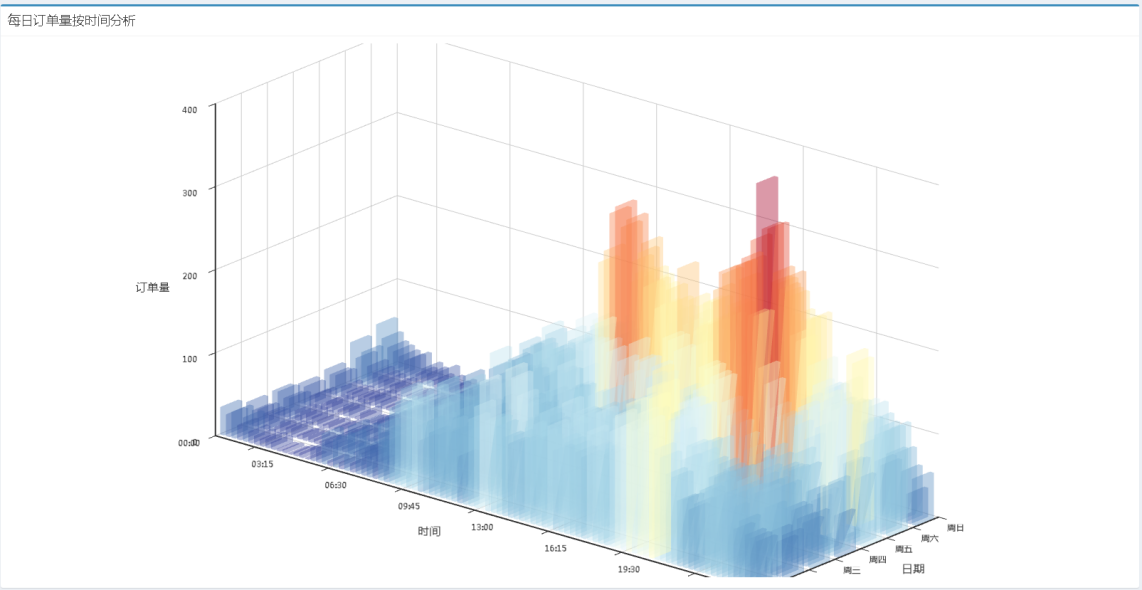
**商圈**则是下午开始到晚上持续递增，一般在周六晚9点到11点出现峰值。



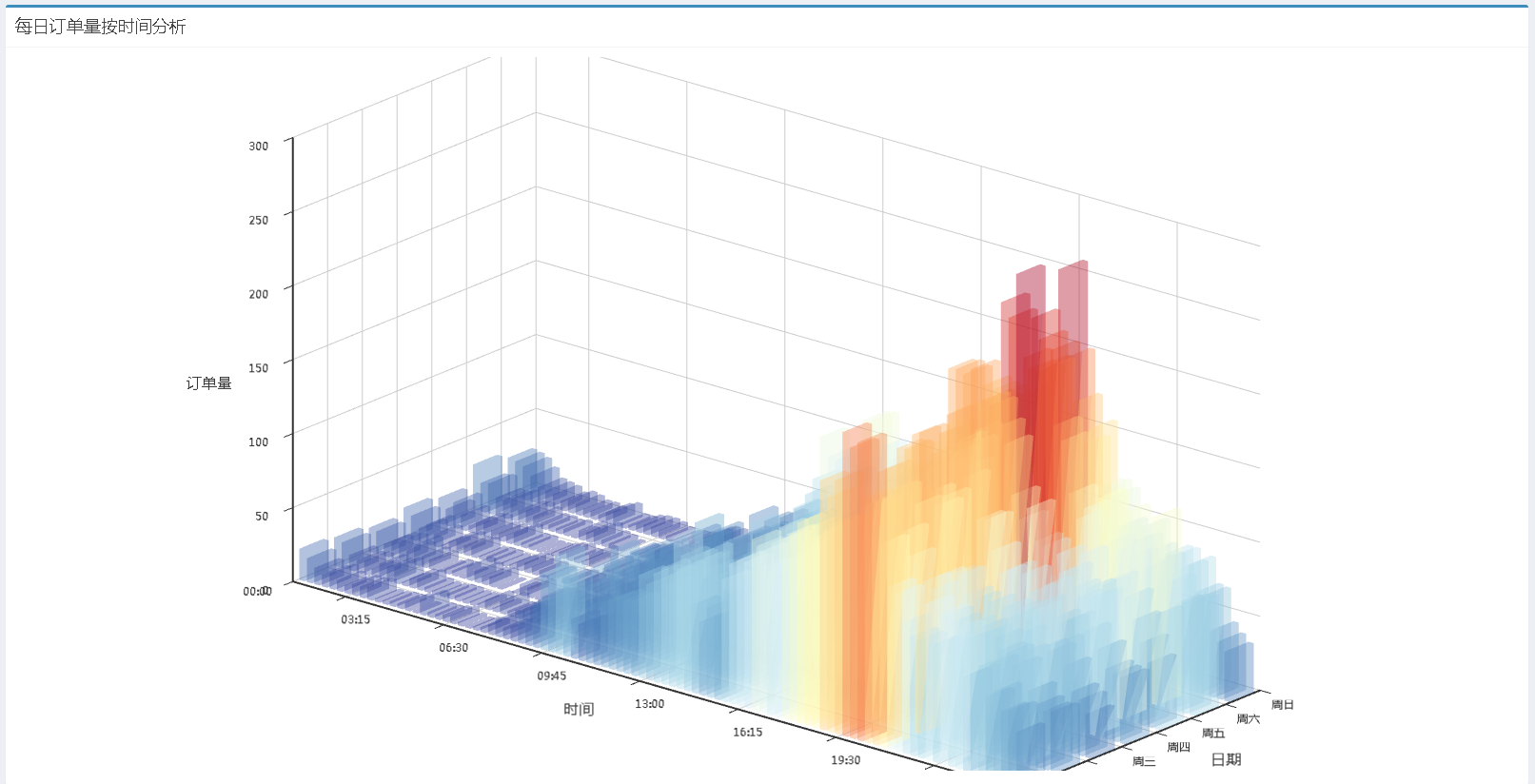
**医院**一般都是9点到21点有订单，周末订单反而减少。



**学校**和医院类似，订单量配合正常的作息时间起伏，在周五和周六下午6点左右出现峰值。

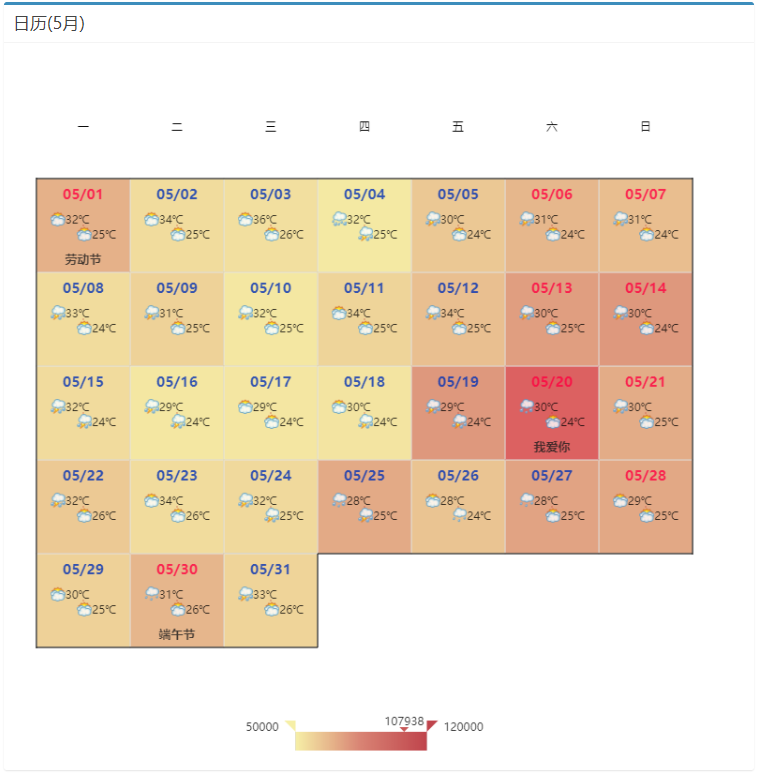


**景点**和游客出行时间相关联。

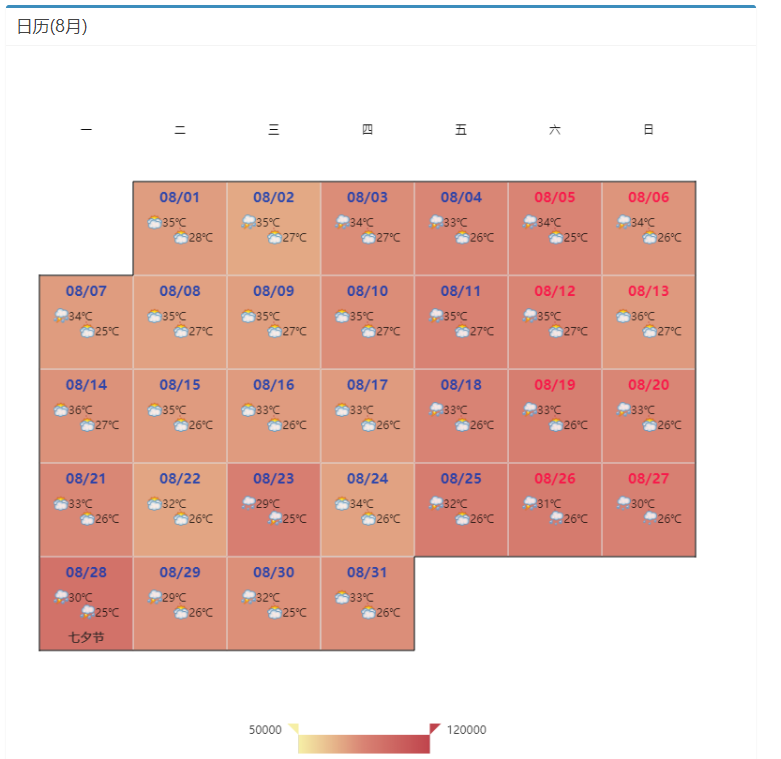


## 1.3日期对于订单数的影响

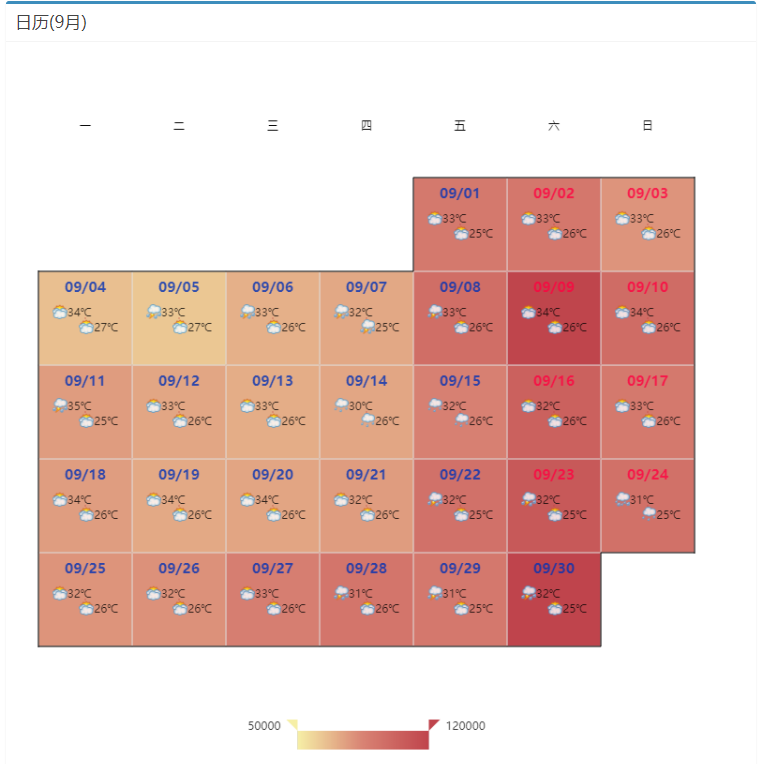
5月份的最高值出现在5月20日。由于周六和520（我爱你）两个因素叠加，出现了5月份的订单峰值。节假日订单量也多于工作日。



和520一样，中国传统的七夕节也是一个打车高峰。同时，8月的恶劣天气稍微增加了一些订单数，但是并没有想象中那么明显。



国庆节的前一天出现了极端最高订单数12万7千单。**从海口出发的游客和进入海口的游客在这天集中出行**。

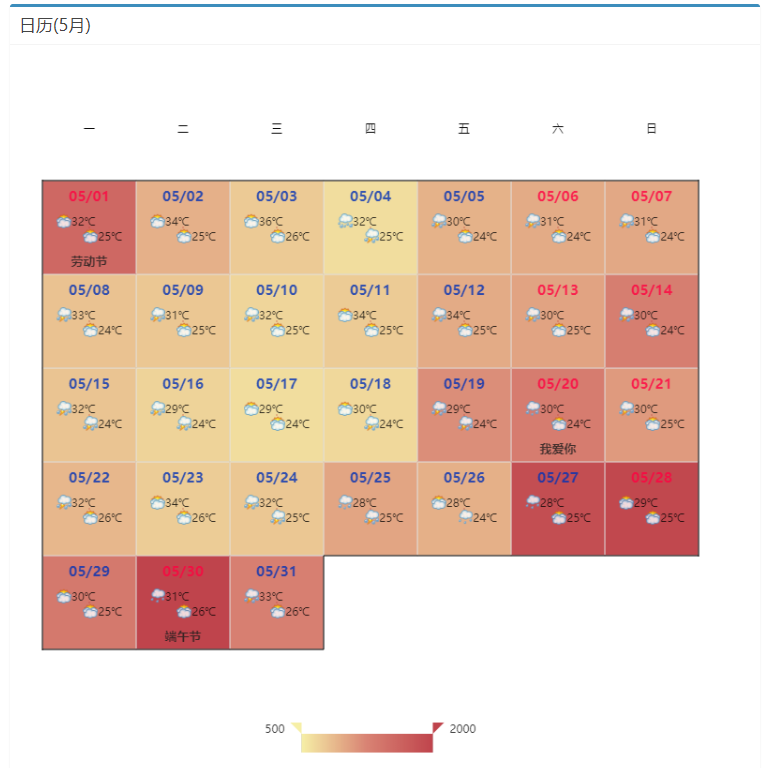


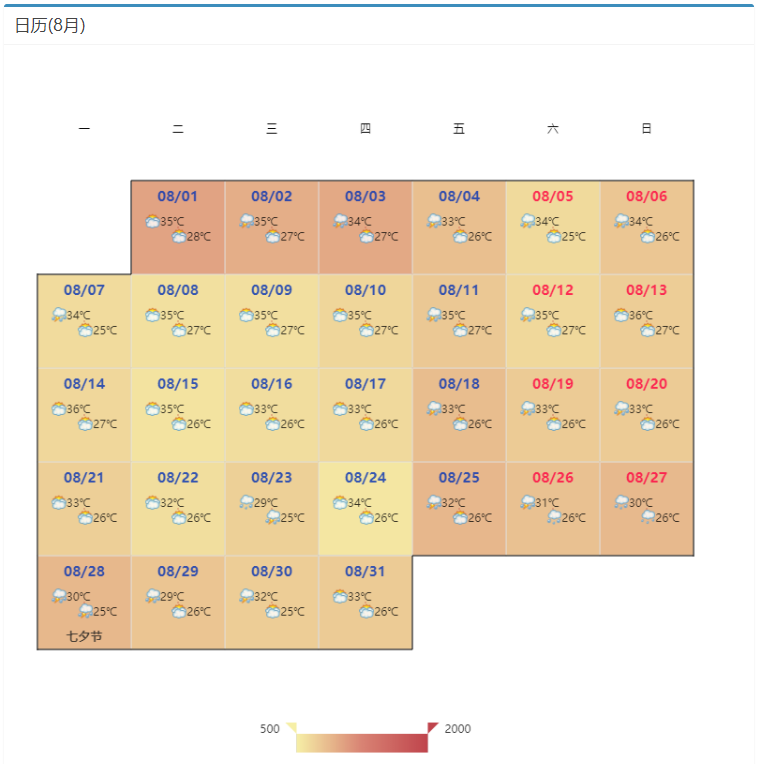
### 1.3.1对于不同POI不同的日期分析

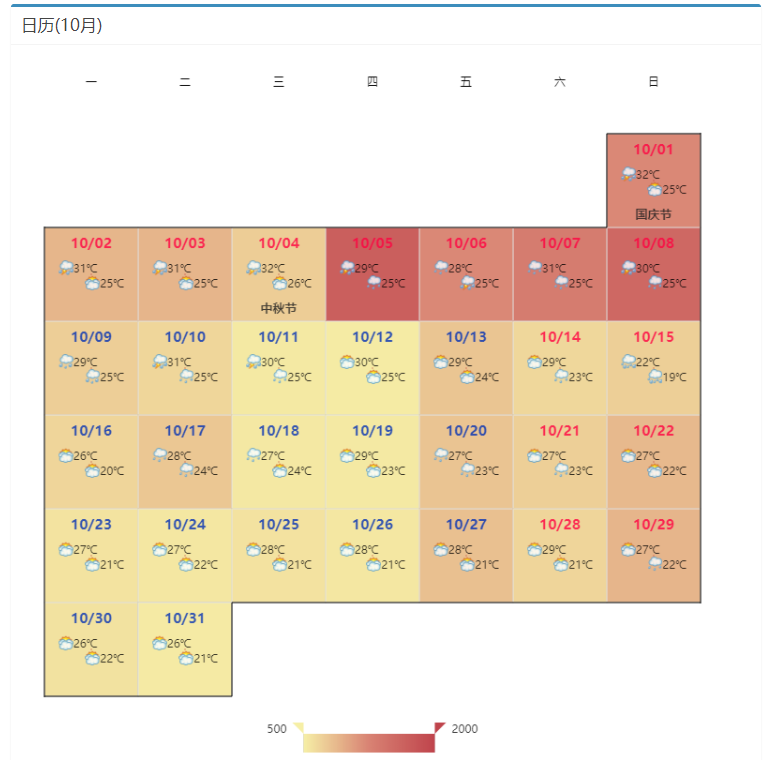
**飞机场**：非常平稳，受到外部因素干扰影响较少。



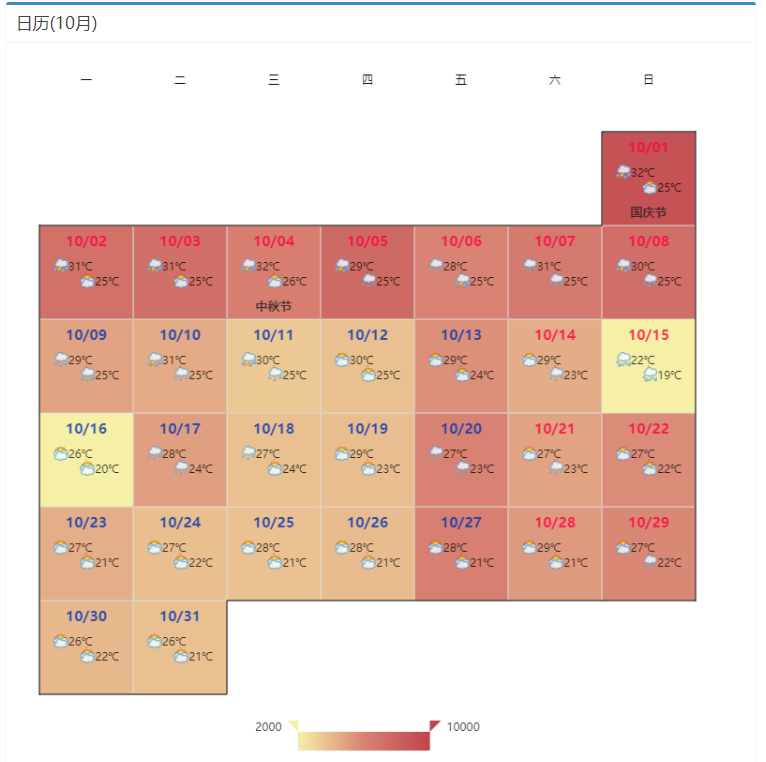
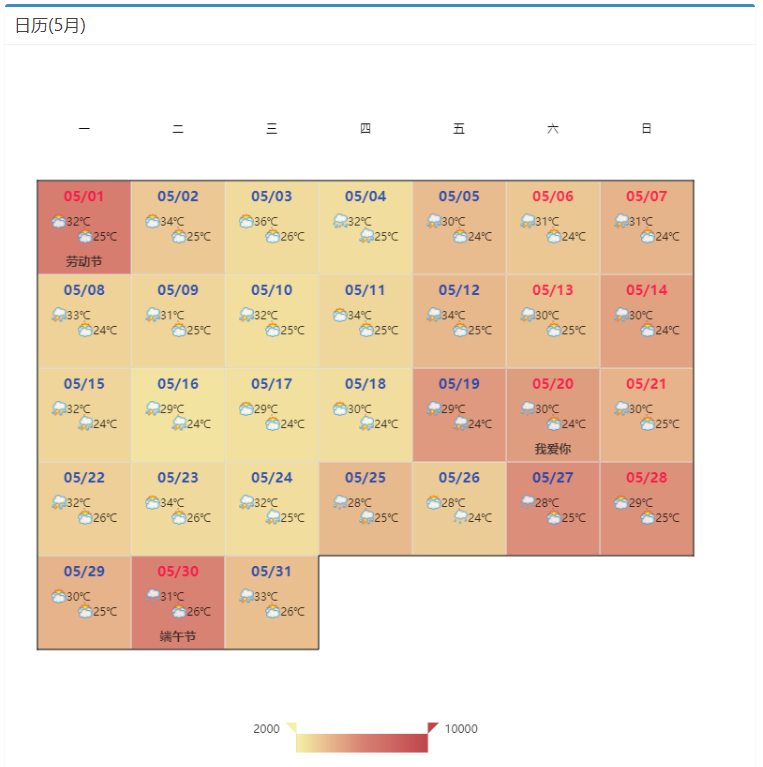
**汽车站**：受到季节（可能和温度相关）影响比较大。



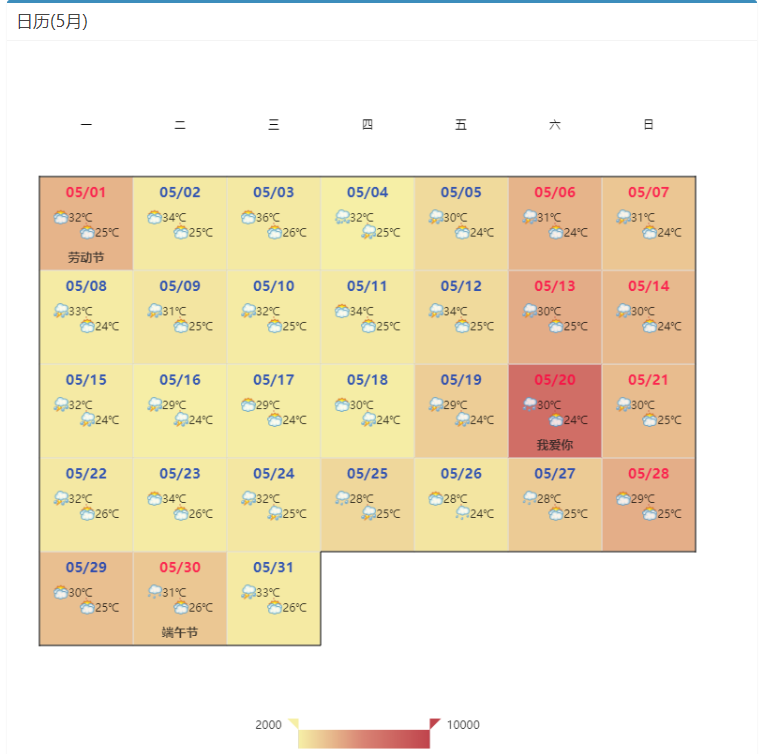


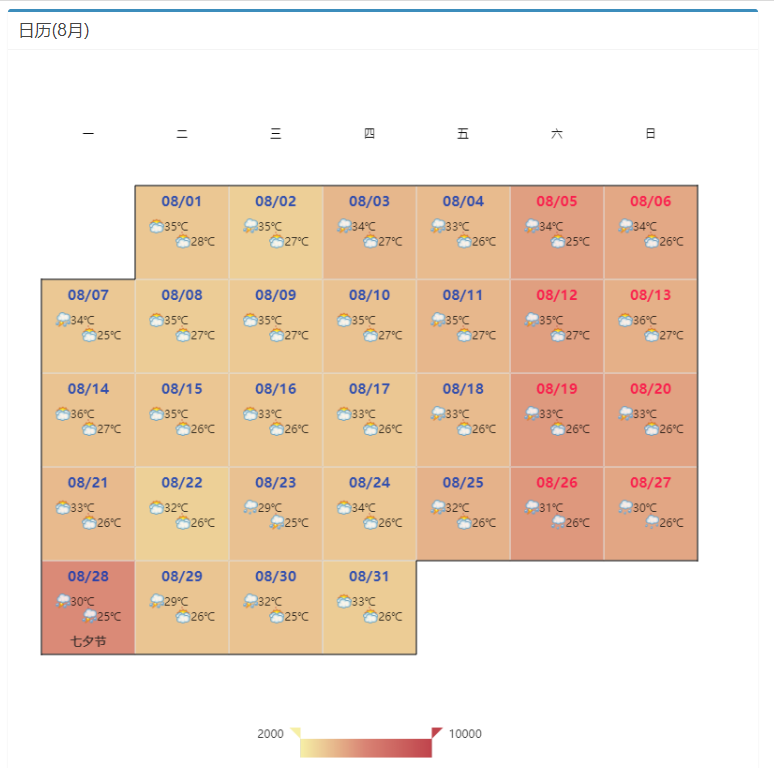


**火车站**：收到节假日影响非常大。

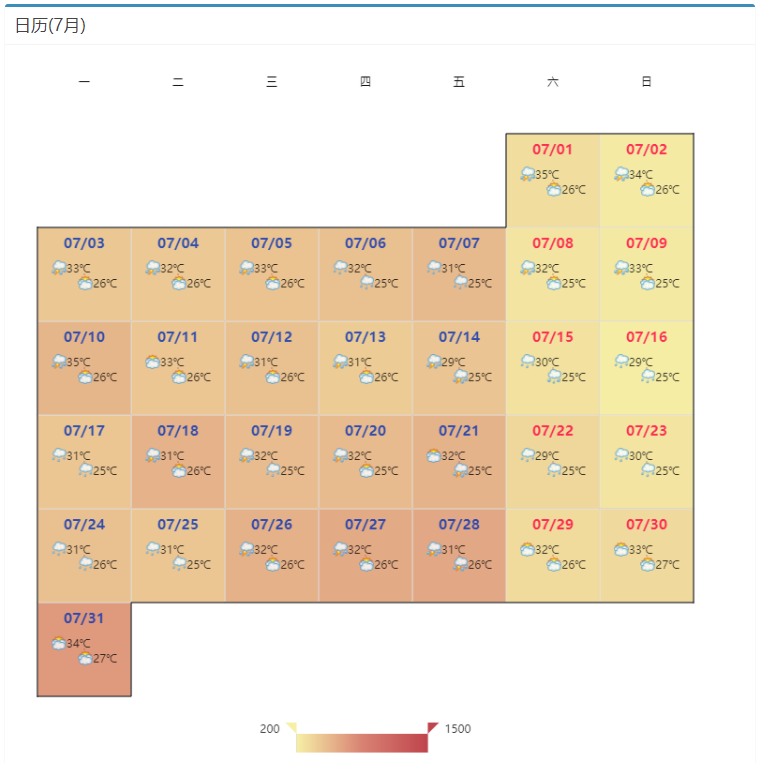


**商圈**：对于520、七夕这样的日子特别敏感。周末人数也较多。



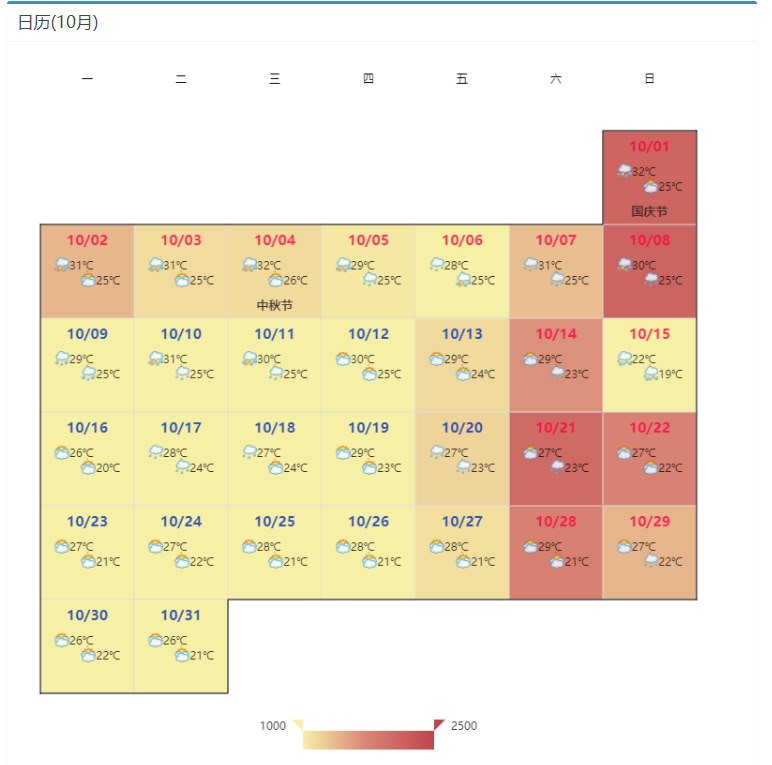


**医院**：周末人比较少，起伏不大。

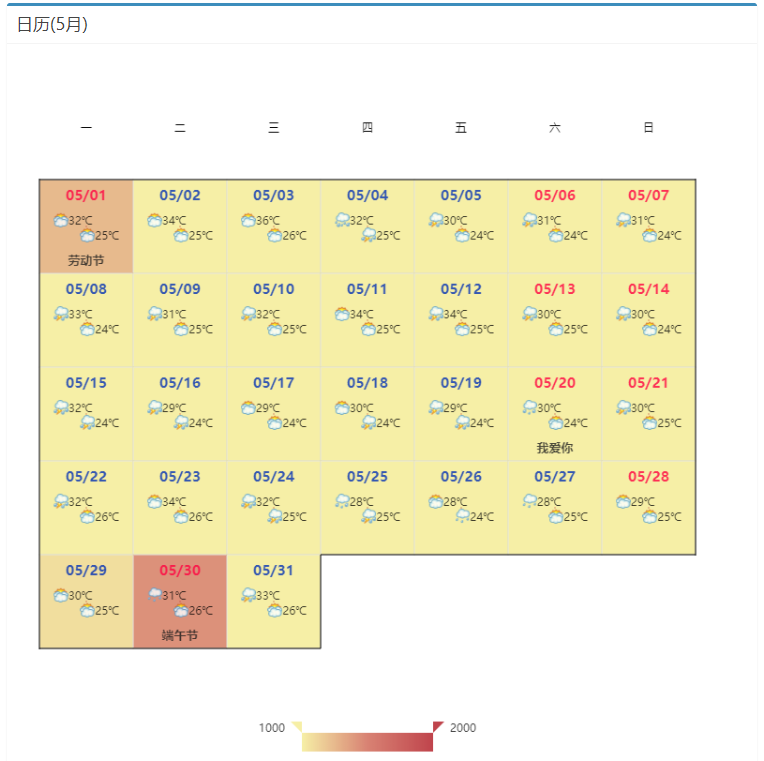


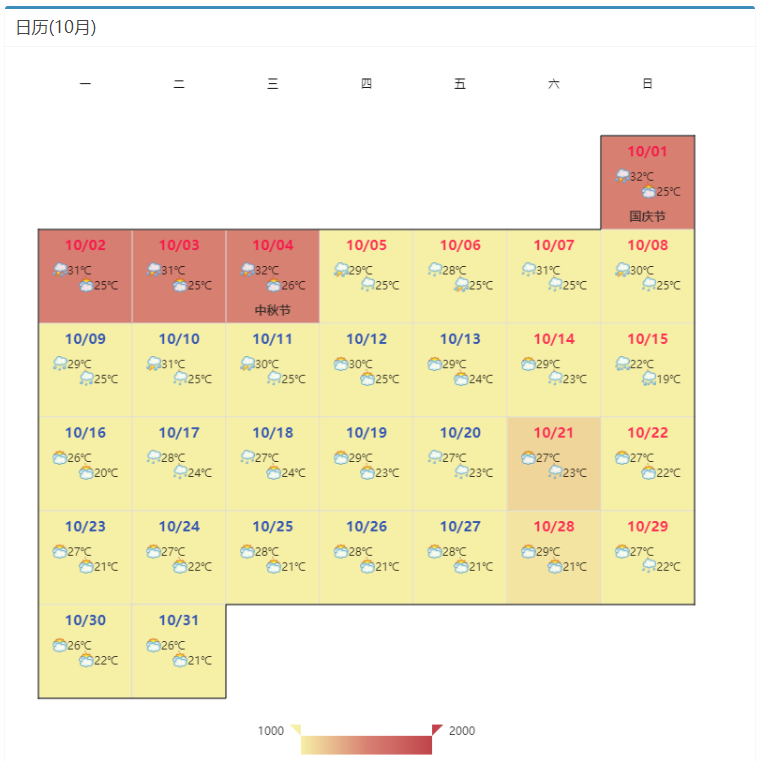
**学校**：520这样的节日敏感，7，8月份假期打车数很少。开学之后节假日用车是高峰。





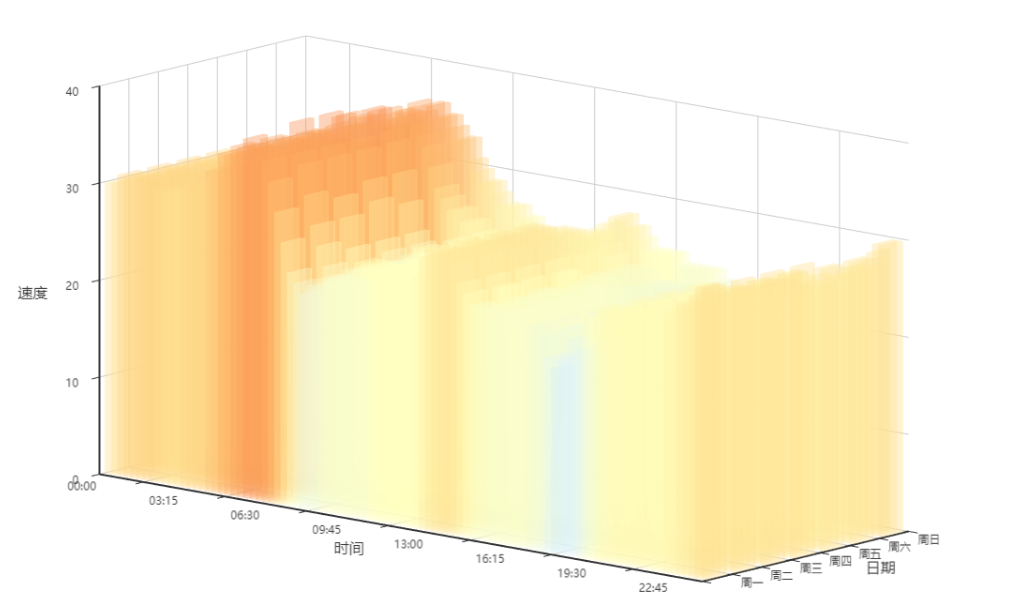
**景点**：国定假日打车数比较多，其他时段很平稳。





## 1.4速度分析

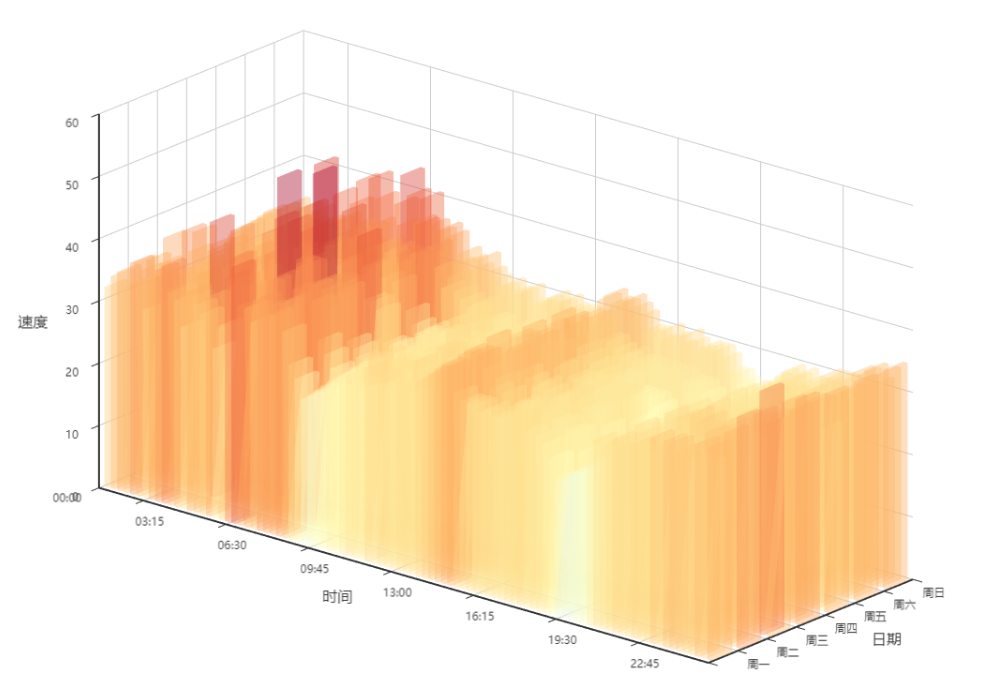
总体来说，速度普遍维持在20—40km/h，受时间段影响较大，基本不受周末或工作日的影响。凌晨车流量小，车速较快，早高峰和晚高峰车速较慢。



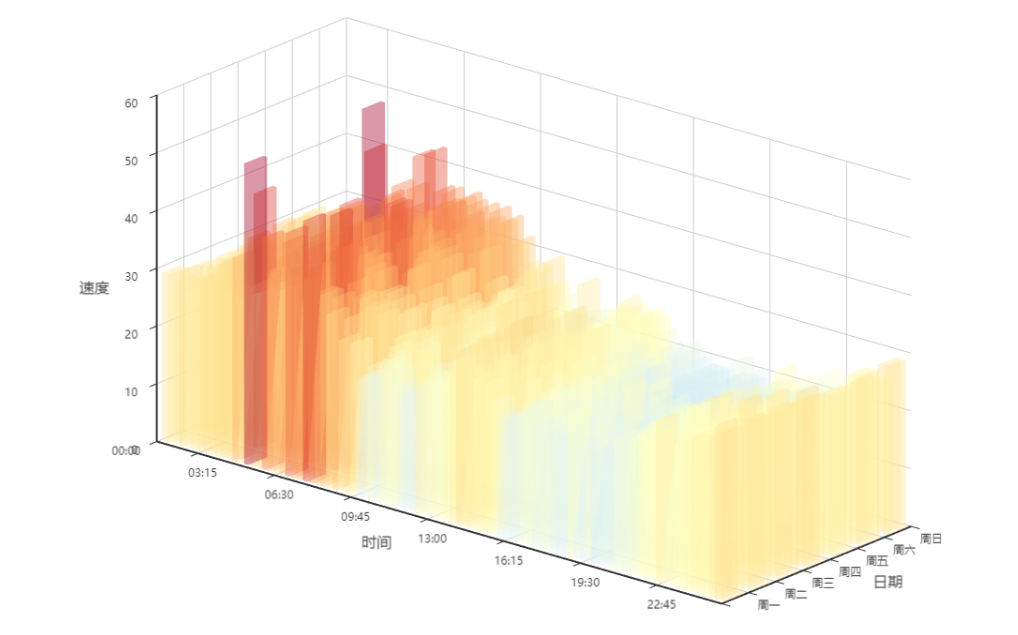
**飞机场**的车速普遍较快，且受时间的影响较小，一天内的车速变化不大，基本为此在40km/h。



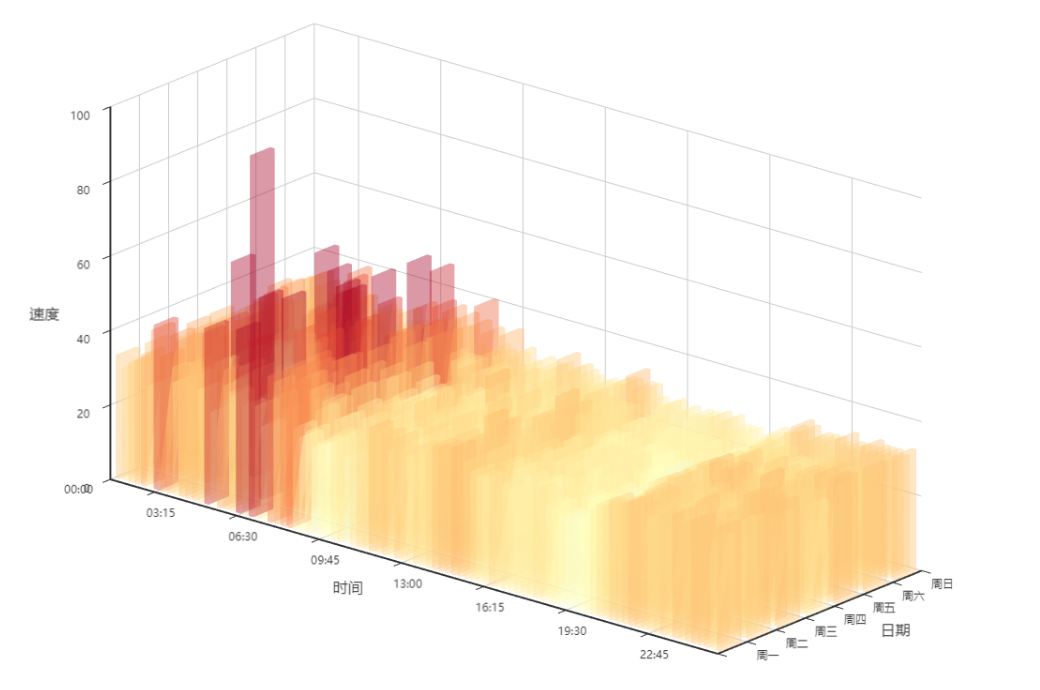
**汽车站**订单的车速在凌晨出现高峰，但是总体速度较慢，受早高峰和晚高峰影响。



**火车站**订单的车速有一个峰值和谷值，高峰出现在周一与周六的凌晨，低谷出现在16:00-19:00，不受工作日及周末影响，一天内车速差异较大。

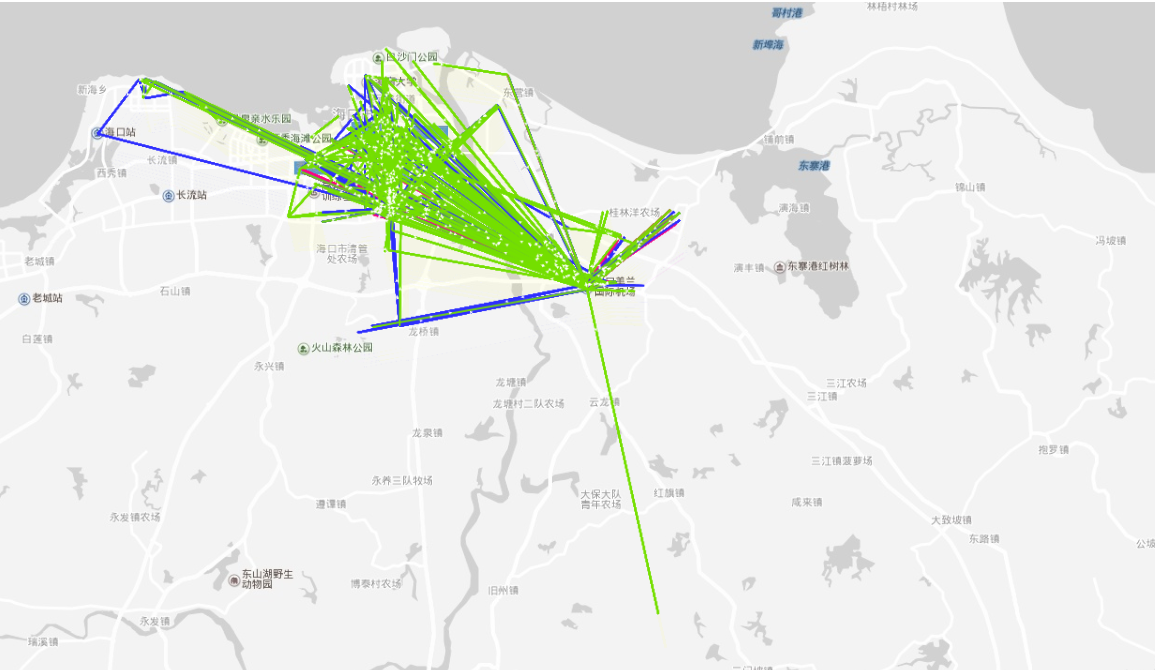


景点订单的车速普遍较低，集中在30km/h左右，可能景区车流量较大，因此车速难以提升。另外，凌晨3:00-6:00这一段时间，景点可能叫空旷，车速较快。



## 1.5轨迹分析

我们分析了一下最多的出发地和目的地轨迹：机场，汽车站和火车站是极其热门的点。同时，也发现了一些意外的热门线路，从机场直接去文昌的人也有不少。其中有些重复率高的线路使用了红色进行标注。



我们也以小时为单位，观察了目的地和出发地的变化，由于是动态画面，将在视频中展示。同样，我们也以周围单位，观察了目的地和出发地的变化，也会在视频中展示。

# 2、题目二：

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，以及与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；

**解答：**

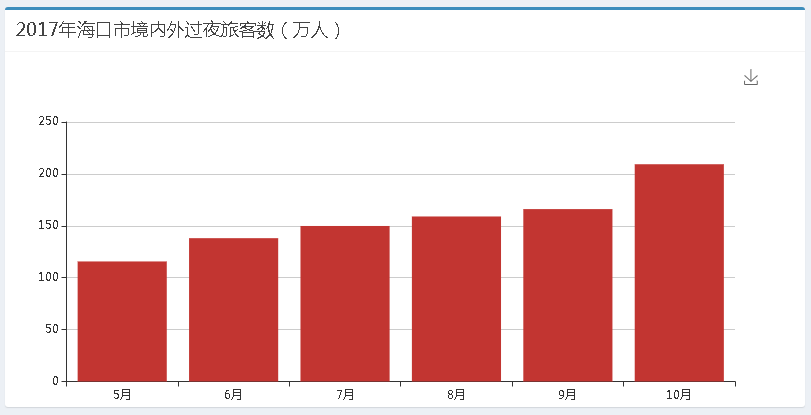
## **2.1特征**：订单数在不断提高。

**影响因素：**

人民生活水平的提高：原本短途公共交通出行的地方，现在可能直接打车了。周末的活动越来越丰富，也提升了打车需求。

**滴滴打车越来越便捷**：APP越来越完善，使用的人数也越来越多。

**海口市的旅游业越来越兴旺：**以下为海口市政府公布的入境过夜旅客数统计：旅客数的增加也带动了订单量的增加。



## 2.2特征：订单量宏观上升的同时，微观有波动

**影响因素：**

**特殊日子：**520（我爱你），七夕节这样需要浪漫的日子，往往就是打车的高峰。

**黄金周：**每一个长假的前夕，也是打车的高峰，数据中的极端最高订单数出现在9月30日。

## 2.3特征：不同的POI变化不同

**影响因素：**

对于学校来说，寒假暑假，入学返校这些都会影响订单的变化。

对于机场，汽车站，火车站来说，假日会影响订单变化。

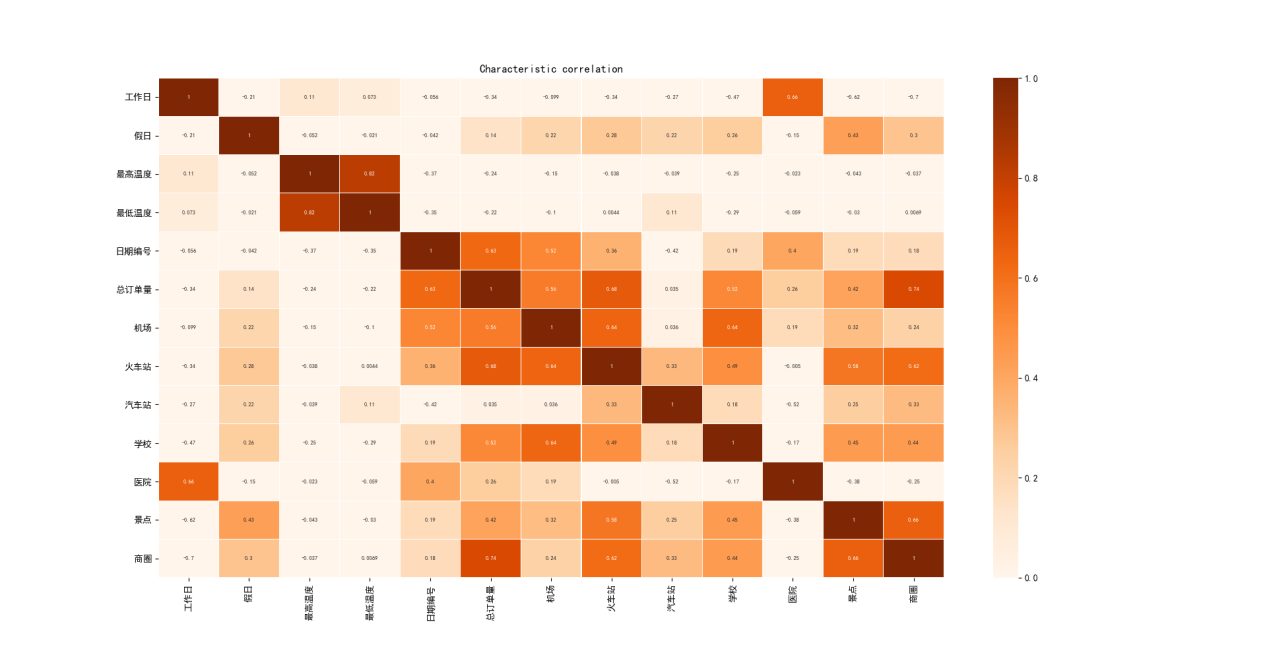
对于商圈来说，周末，特殊的日子会影响订单变化。

## **2.4特征矩阵**

从图中可以看出，由于整体订单量上升，所以日期对于订单的影响是非常巨大的，接下来就是Holiday假日也会产生一定的影响。是否工作日由于订单递增的原因，相关性也变得不重要了。（大图请查看演示用网站）

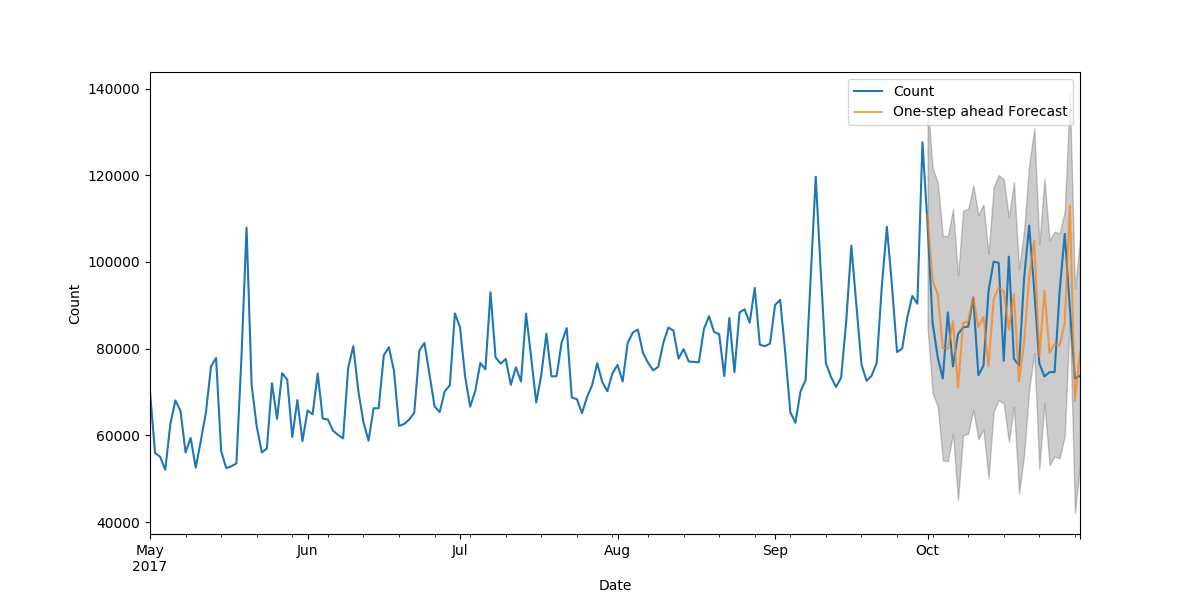
同时我们也观察到订单总量和日期序号（给每个日期一个序号）有着强相关性，进一步佐证了随着时间的推移，**订单量不断上升的趋势**。订单总量也和商圈订单，火车站订单，机场订单有着很强的相关性，**说明娱乐和长途出行的需求和出租车需求是相辅相成的。**

我们还发现，**医院的订单量和工作日**的相关性比较大。

****

## **2.5 Arima预测**

同时我们尝试使用了Arima模型进行订单量按日期的预测，由于订单量和外部数据，特别是节假日相关性较高，**单纯使用Arima进行预测的准确度并不乐观** （可以在演示用网站中查看具体技术指标）



# 3、题目三：

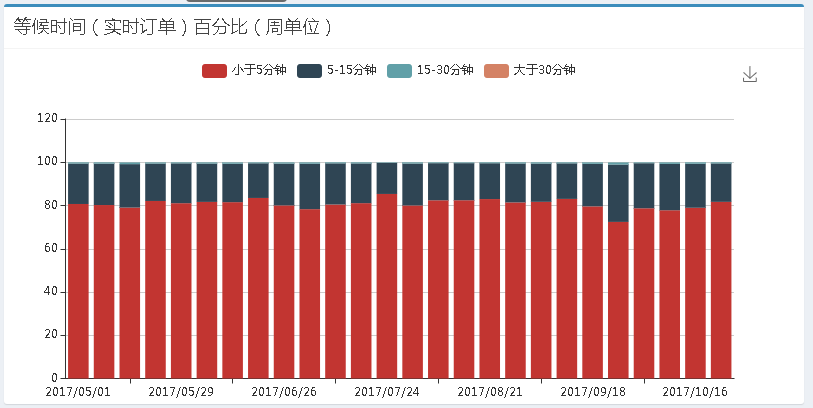
请您根据上述数据分析，为市民或城市交通管理部门提供建议，如面向市民的雨天出行建议、面向交通规划部门的公共交通线网规划和站点设置建议等。

**解答：**

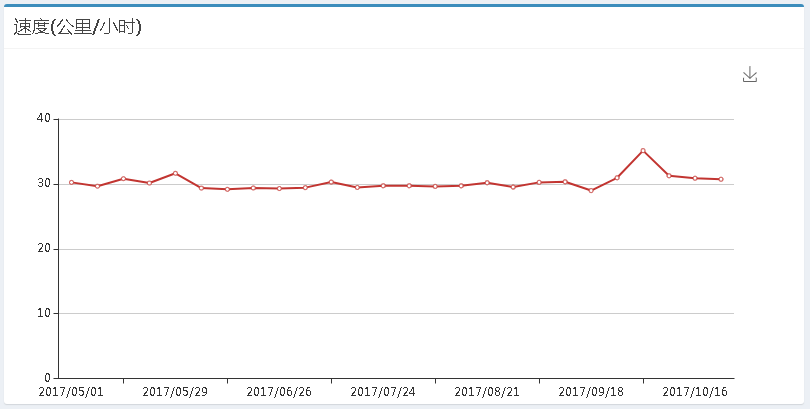
## 3.1面向市民：

在节假日的时候，外地游客较多，尽量避免去海口本地的景点。

海口市的打车还是很方便的，80%的订单，等车时间均在5分钟之内。



海口的路况还可以，时速基本在30KM上下。

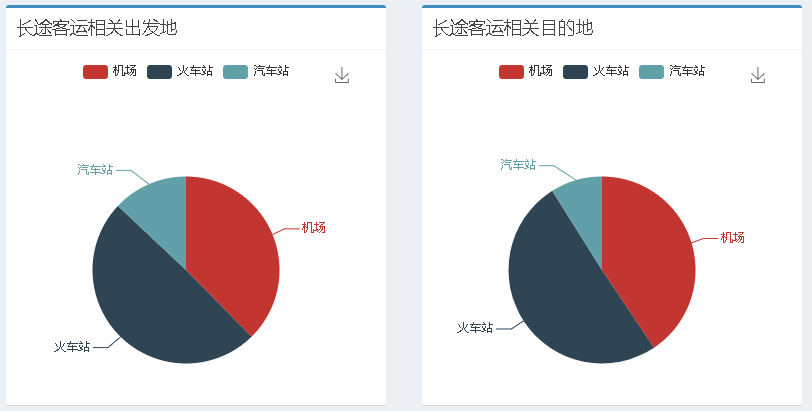


极端天气的日子，打车人可能会增多，减少出行比较好。

## 3.2面向交通规划部门

1.我们通过出发地和目的地的轨迹分析，可以发现很多相同出发地和目的地的订单，这些轨迹可以考虑**增设固定的公共交通**。

2.我们发现，相比汽车站和机场来说，**火车站周围应该设置更多的线路**：



3.通过对于不同时间段的订单量分析，交通出行有很大的**“潮汐”规律，在出行高峰可以增加公共交通的班次，在出行低谷减少班次。**

4.订单中80%左右是10公里以内短途的出行，可以**在人口密集区域增加共享单车的投放**，缓解交通状况。



5.我们对500个热门地点（作为出发地或者目的地），然后发现有些地点虽然靠得很近，但是一些作为出发地很热门，一些作为目的地很热门，这些地方可以进行设施改造，设立出租车上下客排队区，使得作为目的地热门的区域下客的出租车立刻可以接下一单乘客。

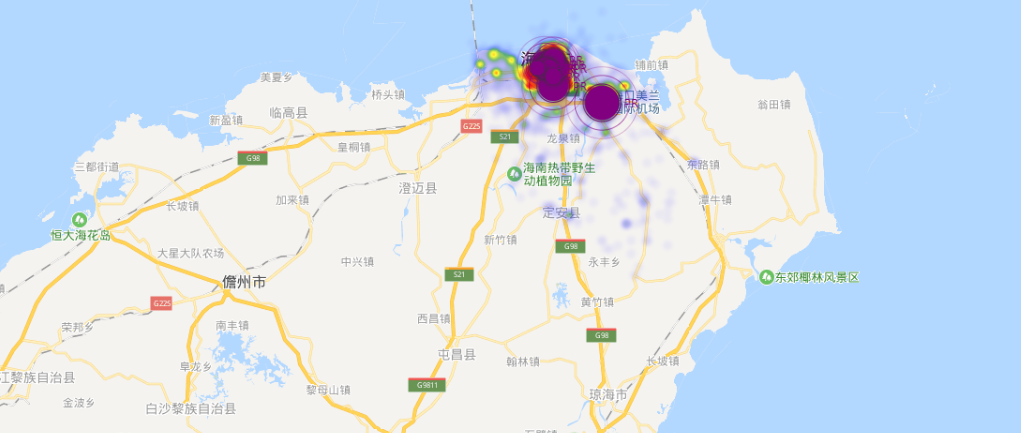


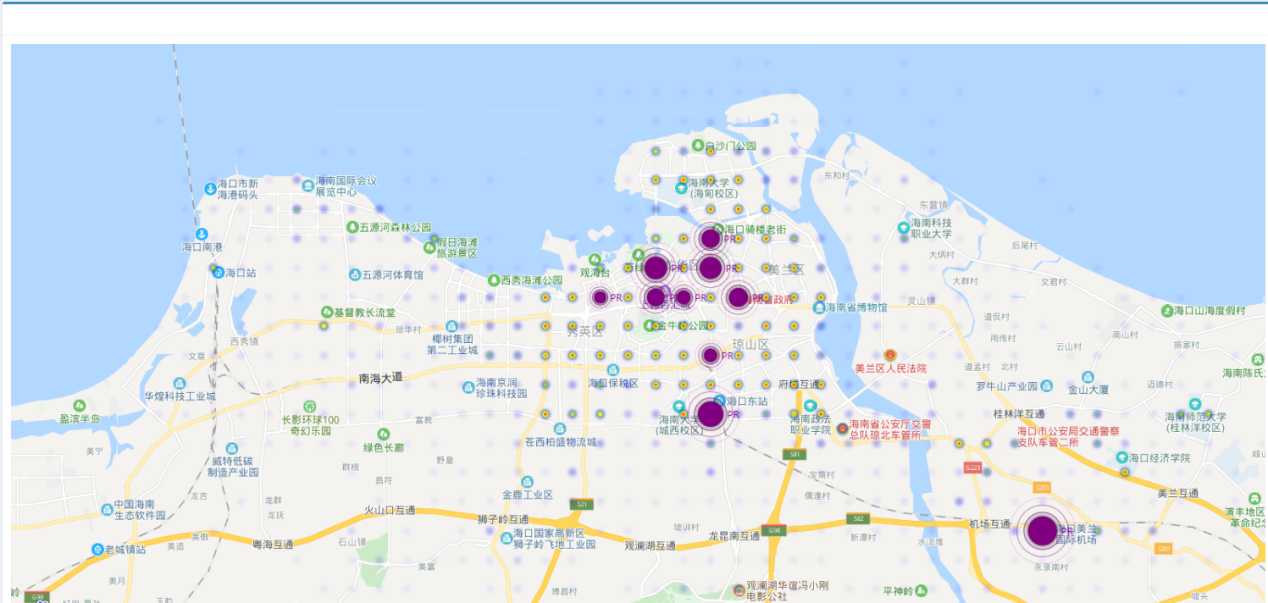


1. **通过pagerank分析，确定交通枢纽的设定。**

由于业务场景中也存在着订单从一个区域流向另一个区域的特点。在以往的分析中，单纯的按区域进行聚合统计，但是忽略了一个比较重要的事实。如果我们的订单流向了一个一个热点区域，那么这个订单，质量就更高。如果流向了一个非热点区域，这个订单质量就更低。使用这个算法，可以更合理的计算热点区域的质量。

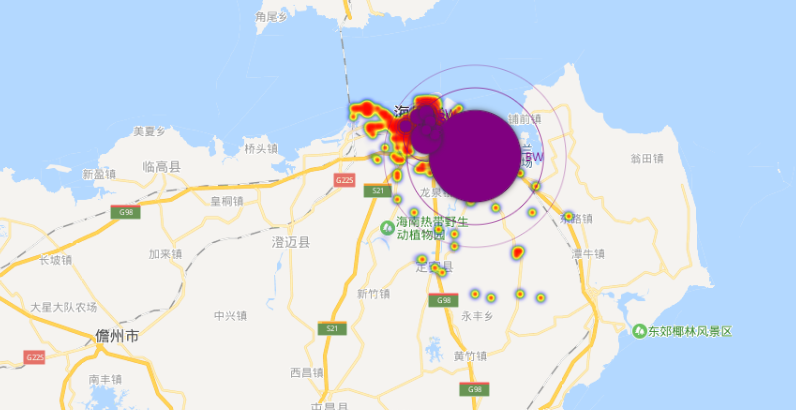
Pagerank的结果指向市中心的热点区域。用于发现城市外围热点区域核心路径点。如下是计算出来的排名靠前的坐标点，因此可以**在这些点设置交通枢纽**。

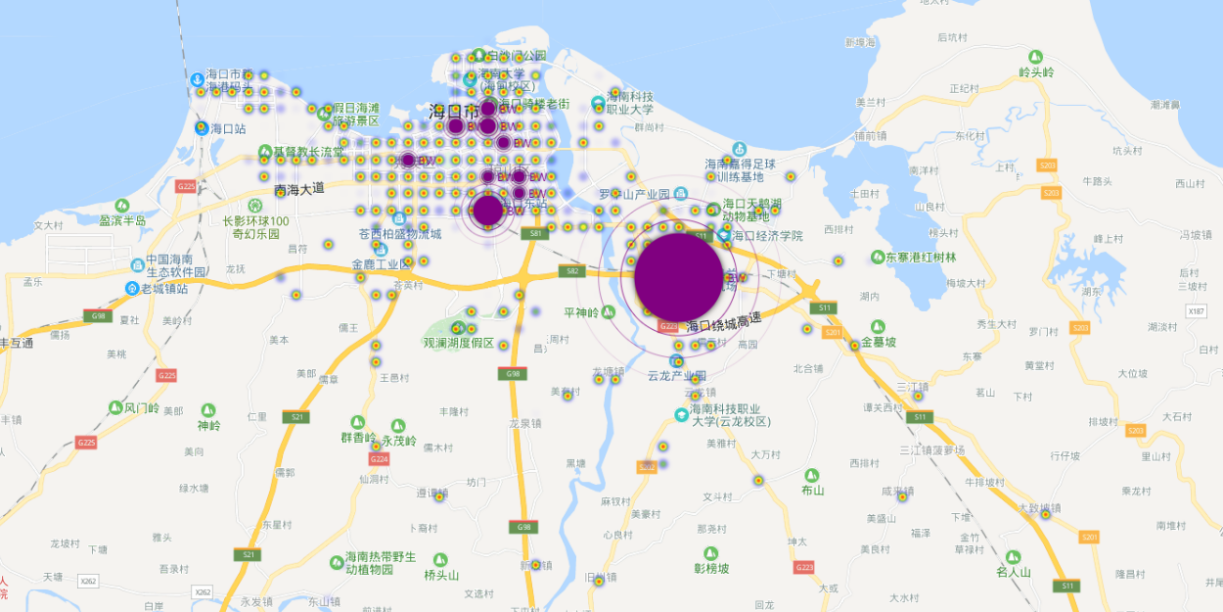




**7.通过betweenness（中介性核心性）分析，增加长途汽车路线。**

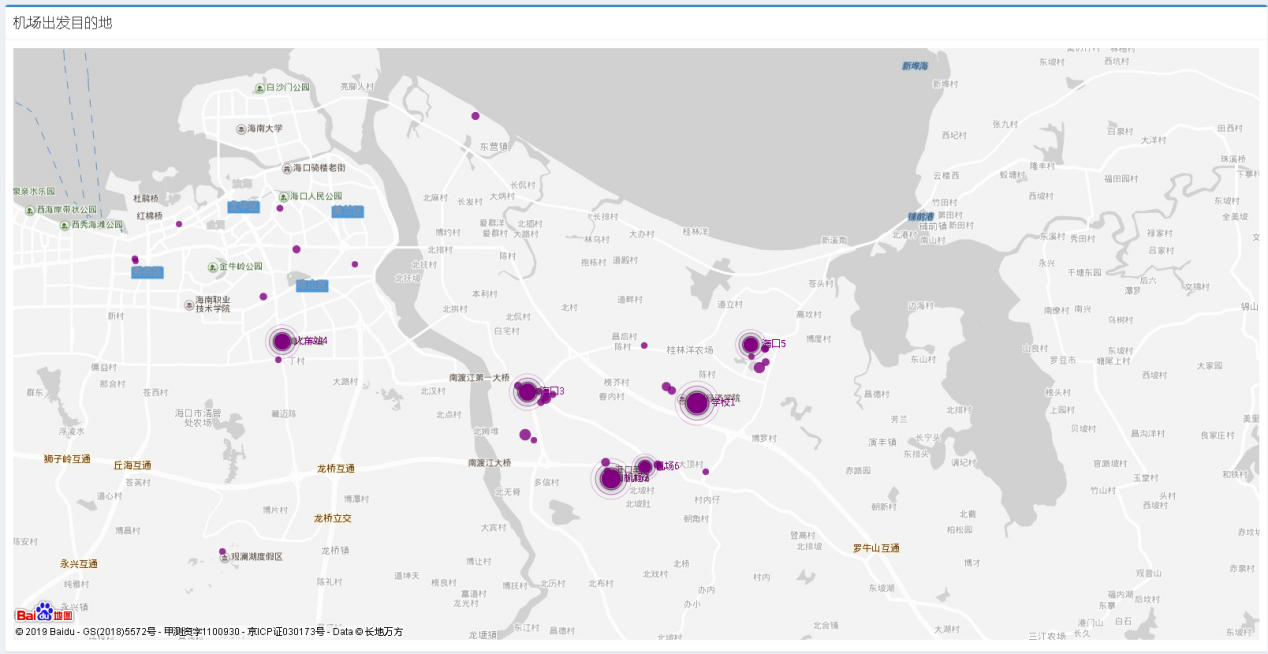
用于发现城市外围热点区域核心路径点。计算出来的排名靠前的坐标点。

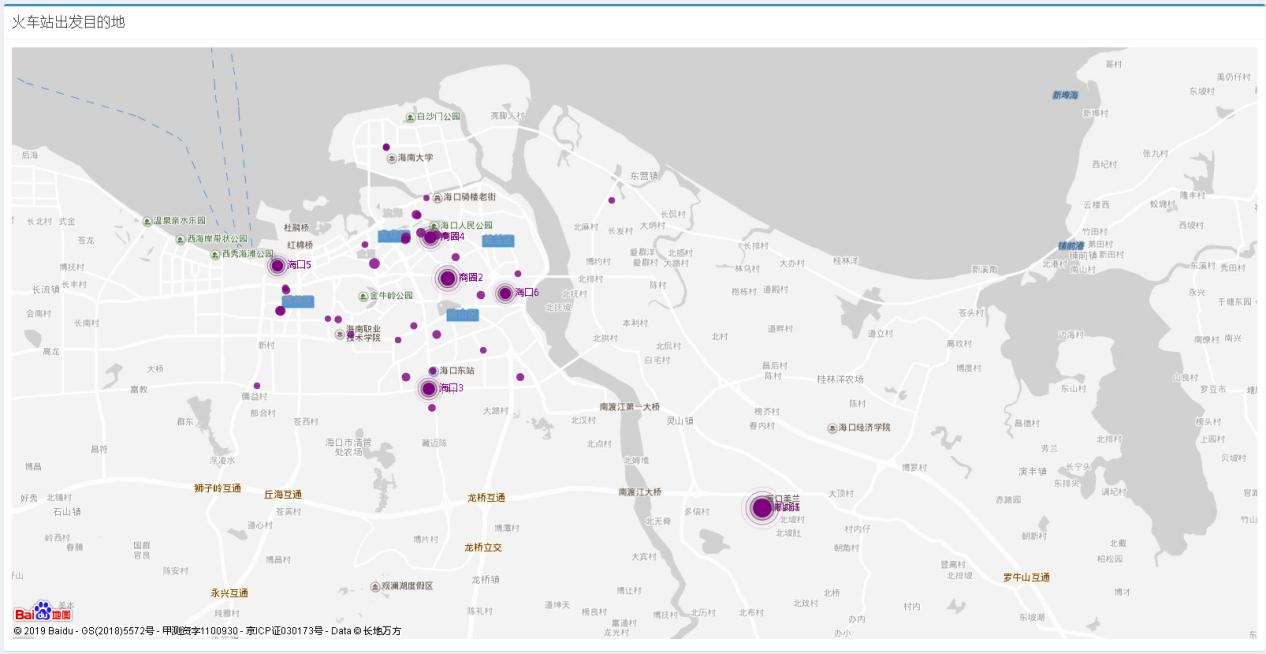




通过观察发现。该点多为城市外围的热点区域的主要交通口。特点为周边有学校、医院、产业园等公共场所。集中于各个城镇周围，符合正常的出行场景。这些点，存在着很多的订单，在城市边缘与中心进行交互，属于外围流量的必经入口。通过该方法，可以有效的为各个城市外围热门区域运营提供支持。我们可以在以上点作为停靠站点。**上面这些地点作为推荐长途汽车或者城际公交的停靠点**。

8.我们这里还分析了从机场和火车站出发的订单的终点，以及去机场和火车站的起点。





通过对于这些地点的整理，可以在这些地点设定机场（火车站）快线的车站。同时也可以**告诉滴滴司机哪些地方比较容易接到去机场的长距离订单**