

Week3 作业

维度	字段名	Defintion
订单维度	order_id	订单号
	order_date	订购订单日期
	order_price	订单金额
用户维度	user_id	用户编号
	User_City	用户所在城市
	user_avgprice	用户消费均价
	user_avgroomnum	用户历史平均入住间数
	user_maxprice	用户消费最高价
	user_minprice	用户消费最低价
	user_ordernum	用户历史订购订单数
酒店/房间	Hotel_City	酒店所在城市
	room_area	房间面积
	roomid	每个售卖房型号
	basicroomid	物理房型号
	hotel_id	酒店号

1.热门出行和热门入住地区/城市的分布是怎样的？

选择什么变量？

思路：如何体现热门？呈现怎样的数据关系？可以选择怎样的图表？

将订单数量在地图上按出行城市和入住城市分组显示；分布关系；地图；

数据选择：出行是指用户所在地区/城市，User_City；入住是指酒店所在地区/城市，Hotel_City；每个城市的订单数量是指将 order_id 按城市分组；同时在出行城市的表中添加了订单金额（order_price）的均值来表示该城市用户的平均出行水平；在入住城市的表中添加

工作表：





结论:

在出行地区中,热门出行城市多为沿海的城市,但是从居民的平均订单金额来看,内陆出行城市要稍高于沿海城市的居民,说明沿海城市居民虽然出行次数多,但平均订单金额不如内陆出行城市;

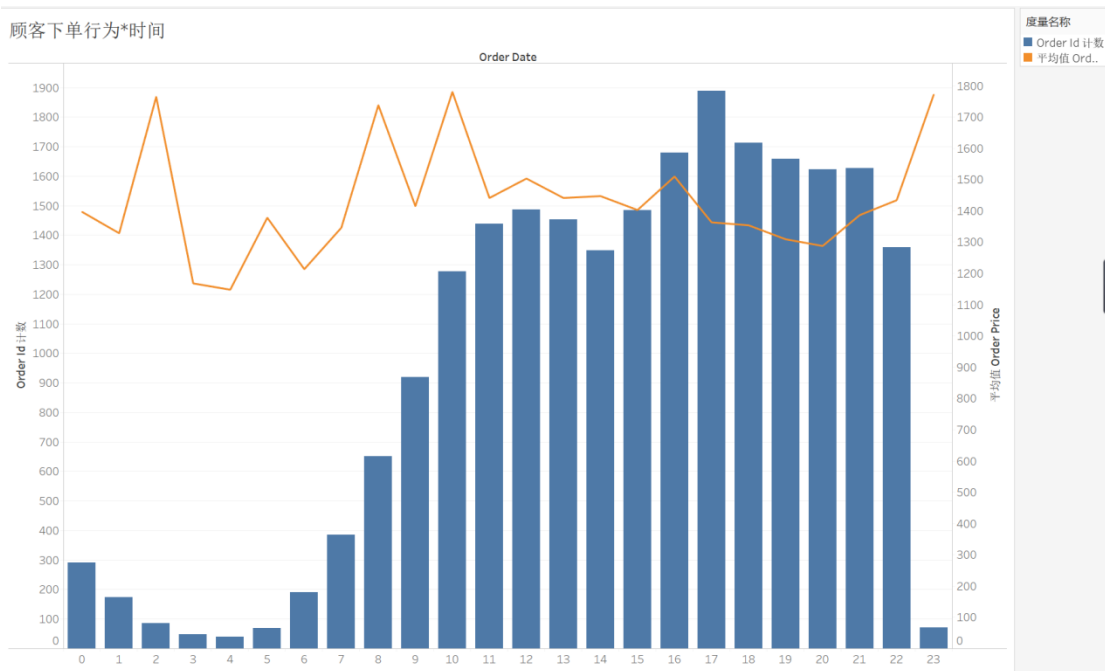
在入住地区中,热门入住地区分布与热门出行城市分布极为相似,说明居民爱旅游的城市同时也是对游客吸引力非常强的城市,同时在热门入住城市中,只有少数城市的平均房间面积排名靠前,绝大多数的房间面积处于中等水平,说明房间面积并不是吸引顾客前往该入住城市的首要原因。

2.顾客下单行为在一天中不同时间段的变化是怎样的?

思路: 如何体现热门? 呈现怎样的数据关系? 可以选择怎样的图表?

将下单数量和金额按时间段分组展示; 数量和金额随时间的变化关系; 柱形图+折线图

数据选择: 将 `order_id` 的合计用来表示下单数量, 将 `order_price` 的均值用来表示订单金额的平均值, 从订单日期中选取小时作为代表时间段的变量
工作表:



结论：从订单数量来看，用户下单时间基本集中在上午 10 点到晚上 10 点的十二个小时之间，但平均订单金额最高的是凌晨 2 点，上午 8 点和 10 点，以及晚上 11 点

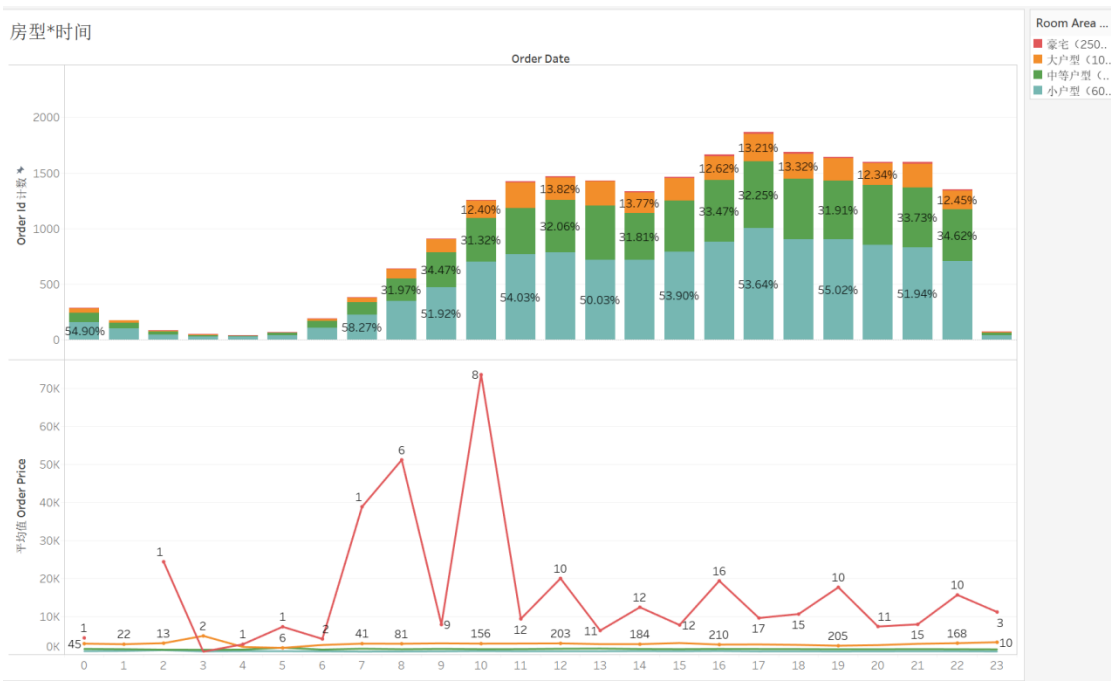
3.不同房型的销售情况在一天中不同时间段的变化是怎样

思路：如何体现热门？呈现怎样的数据关系？可以选择怎样的图表？

将订单数量按房型和时间段分组展示；订单数量金额随着房型和时间段的变化趋势；柱形图+折线图

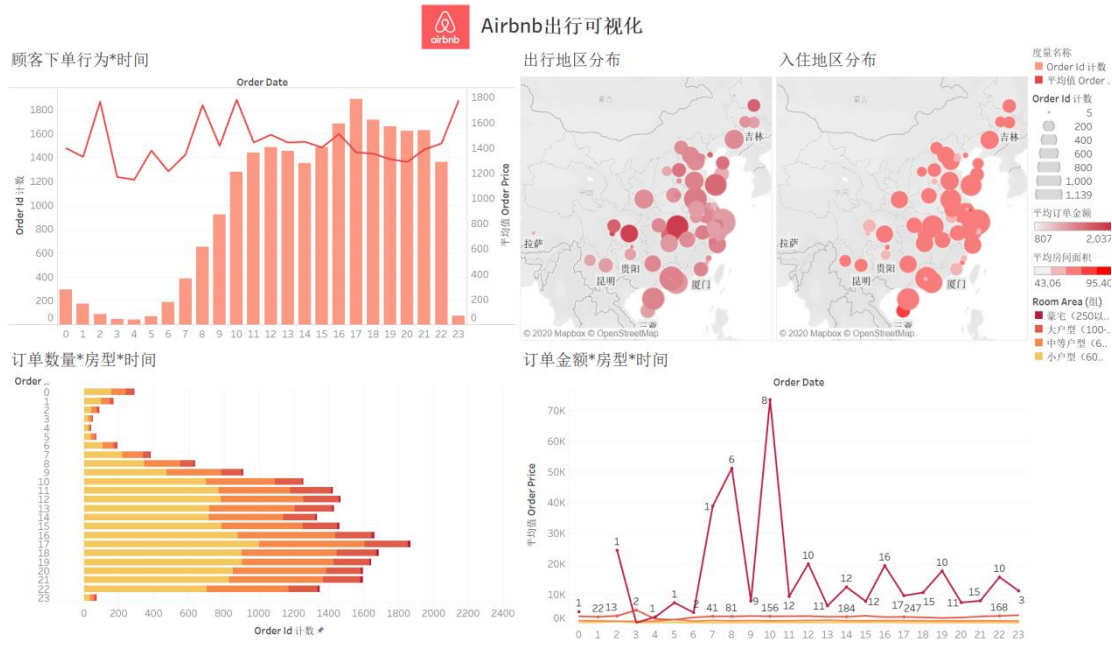
数据选择：订单数量用 `order_id` 的合计来表示；房型按房间面积大小分为了四组（小户型、中等户型、大户型和豪宅，用筛选器剔除了面积为-1 的异常数据）；从订单日期中选取小时作为代表时间段的变量；

工作表：



结论：从图中可以看出，订购小户型（60 以下）的客户数量在每个时间段的下单量都超过半数，其次是中等户型和大户型，下单豪宅的订单虽然占比较少，但是平均订单金额远超其他三种房型

仪表盘：



总结论：

Airbnb 的热门出行和入住地区多数分布在东南沿海的发达城市，但出行城市中部分内陆城市的用户的平均订单金额较高，入住城市的房间面积不是主要吸引游客的因素；用户下单时间基本集中在上午 10 点到晚上 10 点的十二个小时之间，但平均订单金额最高的是凌晨 2 点，上午 8 点和 10 点，以及晚上 11 点；订购小户型的客户数量在每个时间段的下单量都超过半数，其次是中等户型和大户型，下单豪宅的订单虽然占比较少，但是平均订单金额远超其他三种房型；