

数据可视化作业七：可视分析系统的设计和实现

薛春宇 518021910698

Deadline: 2021年4月21日 晚23:00

1. 可视化系统概述

本项目基于 *Flask* 后端框架及 *JS* 前端实现，使用 *D3.js* 工具对各图表的内容、形式，以及图表间的联动配合进行实例化，并在可视化系统的基础上对酒店数据进行分析，得出全面可靠的一系列总结性结论。实现代码见 `./Project`，若出错可直接访问[Github工程文件](#)。

- 为找出酒店淡季、旺季的时间段，本项目使用了数据中的入住率 **Occupancy** 和平均入住时间 **Length of Stay** 作为评估依据，并创建了一个统计量：总居住时间 **Total Time** 作为表征酒店淡、旺季的统计指标。 TT 的计算公式为： $TT = Occ \times LoS$ 。基于上述统计指标，我们将得出酒店淡季、旺季的直接定量表示。具体分析结果见 2.1 节。
- 为找出造成酒店淡季、旺季现象的原因，本项目将上述酒店淡、旺季的定量表示 TT ，同数据中的平均价格 **Price** 和是否有大型活动 **Conventions** 结合，通过绘制热力图的形式进行横、纵向对比。具体分析结果见 2.2 节。
 - 注意：季节因素也会影响酒店的客流量，相关分析将在后续给出。
- 为分析酒店入住人员的特点，本项目采用了以下酒店数据绘制百分比柱状图，并划分为 **Area**、**Gender**、**Age**、**Group** 四个大类：
 - **Area**: *Local*、*USA*、*SA*、*EU*、*MEA*、*ASIA*
 - **Gender**: *Female*、*Male*（其中，*Male* 的比例根据 *Female* 比例间接得出）
 - **Age**: *U20*、*20to35*、*35to55*、*M55*
 - **Group**: *Businessmen*、*Tourists*

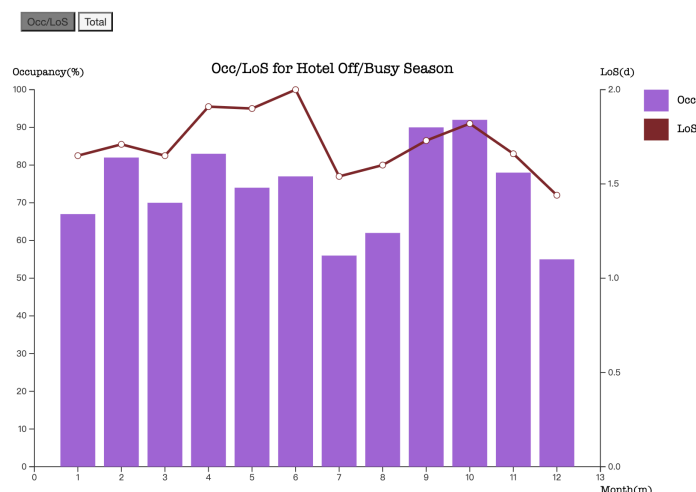
基于上述特点的可视化结果，本项目将结合之前的热力图综合分析，得出酒店住客的特征统计结果。具体分析结果见 2.3 节。

- 为更加清晰地看出同一大类里各子类的相对比例关系，本项目采用由上述百分比柱状图实时控制的饼图的形式，来进行更加直观的表达。（横向对比）具体分析结果见 2.4 节。
 - 为更加清晰地看出某一子类随时间的变化关系，本项目采用了由上述百分比柱状图实时控制的折线图的形式，来进行更加直观的表达，并得出酒店住户相应特征随时间的变化规律，以及哪些因素可能导致了上述变化。（纵向对比）具体分析结果见 2.5 节。
 - 季节因素对酒店影响的分析是一个全面的过程，本项目将在 2.6 节中基于上述全部图表的可视化结果，来对该因素进行全面的分析，以得出哪些因素具有相似的季节性变化规律。
-

2. 数据分析结果

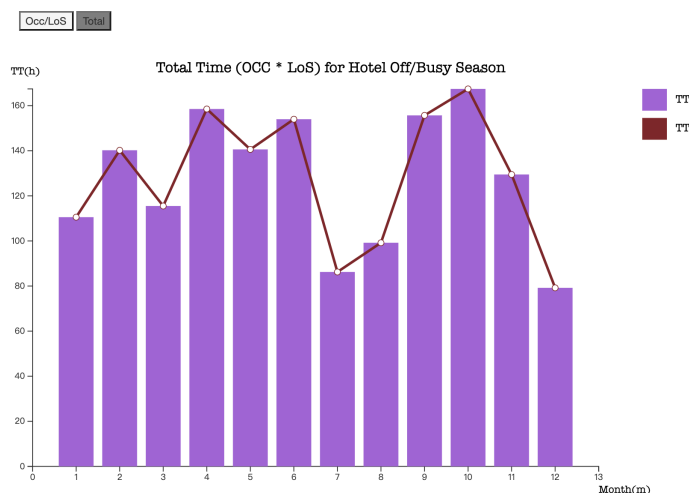
2.1 酒店淡、旺季的定量表示

- 首先，为了在同一张表格里同时表现入住率 *Occupancy* 和平均入住时间 *Length of Stay* 各自随时间的大小变化，我们在本模块的第一个表格内同时绘制柱状图和折线图，并采用两套不同意义的 *y* 轴，来分别指示 *Occ* 和 *LoS* 的相对大小。



其中，柱状图代表入住率 *Occ*，折线图代表平均入住时间 *LoS*；左侧 *y* 轴指示 *Occ* 的相对百分比大小，右侧 *y* 轴指示 *LoS* 的天数大小。

- 然后，通过公式 $TT = Occ \times LoS$ ，我们绘制直接指示酒店淡、旺季的统计指标的柱状图：

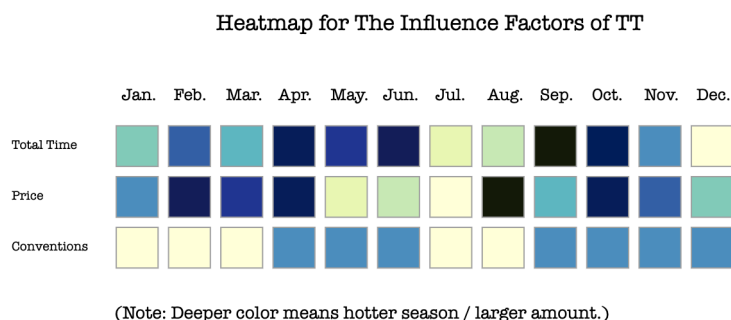


根据图中的折线图或柱状图的相对大小，我们能够对酒店淡、旺季进行定量的判断。

- 结果分析：（酒店淡、旺季的时间段）
 - 入住率 *Occ* 和平均入住时间 *LoS* 在时间上的变化趋势具有一定程度上的相关性，这也更增强了 *TT* 作为统计指标的正确性。然而，两类指标在某些月份上的相对大小仍存在不匹配的现象。
 - 基于 *Occ* 和 *LoS* 计算得到的统计指标 *TT* 则能更好地指示酒店淡、旺季时间段：
 - 旺季：冬末（2月）、春天（4、5月）、夏初（6月）、秋天（9、10、11月）
 - 淡季：冬天（12月、1月）、初春（3月）、夏天（7、8月）

2.2 造成酒店淡、旺季现象的原因（部分）

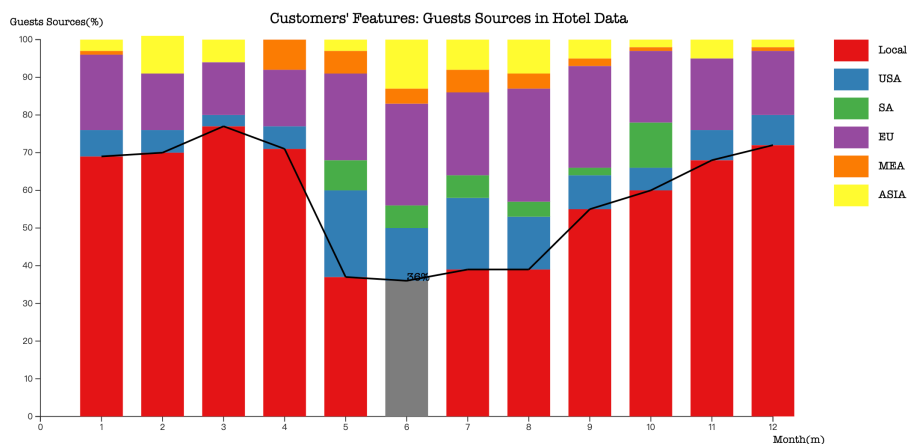
- 基于 2.1 节中计算得到的统计指标 TT ，同数据中的平均价格 $Price$ 和是否有大型活动 $Conventions$ 结合，通过绘制热力图的形式进行横、纵向对比。该热力图会随着用户点击 2.1 节中的 $Total$ 按钮显示出来。



- 结果分析：（造成酒店淡、旺季现象的原因）
 - 平均价格和 TT 存在一定的负相关关系（例如在5、6、8月份中），其原因可能是住户由于房价较低，更倾向于居住在该酒店。
 - 平均价格和 TT 在某些月份也存在一定程度上的正相关关系（例如在2、4、8月份中），其原因可能是由于该期间当地风景较好，酒店为增加营收做出的策略性提价。
 - 是否有大型活动与 TT 存在着明显的正相关关系，在4、5、6、9、10、11月份中，由于存在大型活动，即使这其中某些月份的房价较高，游客量仍维持在较高水平。12月为例外，原因可能是当时的气候过于寒冷。
 - 总体来说：
 - 是否有大型活动与 TT 的正相关性最大，是客流量的主要影响因素；
 - 房价在酒店不同的营收策略下，会分别对客流量产生不同的影响；
 - 季节因素也会对客流量产生一定影响，气候宜人时游客更倾向于前往。

2.3 酒店入住人员特点

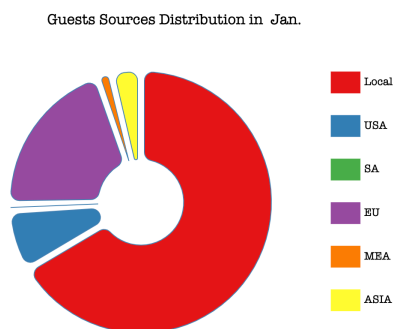
- 为了全面表现出酒店住户的特点，我们分别选择了来源地、性别、年龄和目的四个大类，采用百分比柱状图的形式对住户特点进行多方面的刻画。同时，本模块还添加了辅助折线、比例提示等元素，来帮助用户理解数据。但是，由于该表格内呈现的数据过于繁多，用户仍不能单单以此来获取对数据的直观感受。因此，本项目中将该模块作为一个选择器，通过点选不同的元素，来为之后更详细的数据呈现提供接口。



- 结果分析：（酒店住户的特征，在本模块中仅能较为直观地得出一些宏观的结论）
 - 住户来源地上，本地居民通常占大多数，其次是欧洲、美国等西方国家，亚洲、南非等地区的住户较少。可以推测本酒店位于欧洲或美国临近的地区；
 - 住户性别上，男性占据了大多数，通常为70%左右；
 - 住户年龄上，约一半的住户处于中年（35岁至55岁），青年（20岁至35岁）和老年（大于55岁）住户大约各占四分之一，少年（小于20岁）住户占比很少。可以推测本酒店较为吸引中年人群体；
 - 住户目的上，约8成的住户是为了商业目的，可以推测本酒店附近无特别有名的景点，但地理位置较好，临近商圈。

2.4 同一大类里各子类的横向对比

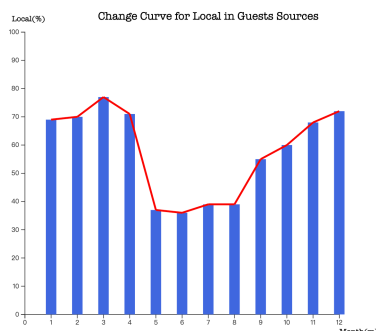
- 为了使用户能够更加直观地获得某一类信息的横向对比结果，本项目实现了一个由 2.3 节中百分比柱状图控制的饼图：



- 注意，本模块所呈现的内容会随用户在百分比柱状图中点选不同的rect元素而改变，需要选择的是大类和月份。本模块自身并不能单独提供有效的结果分析，而是作为 2.3 节中尚有不足的横向对比的补充。

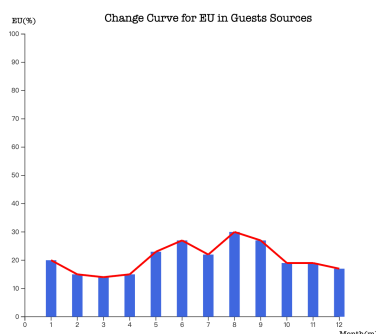
2.5 某一子类随时间的纵向对比

- 为了使用户能够更加直观地获得某一子类随时间的变化趋势，本项目实现了一个由 2.3 节中百分比柱状图控制的折线图：

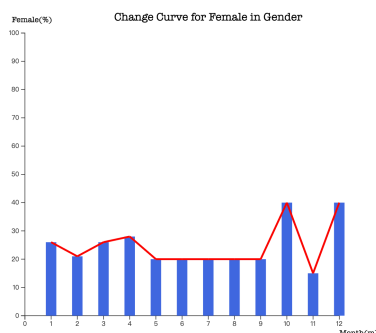


注意，本模块所呈现的内容会随用户在百分比柱状图中点选不同的rect元素而改变，需要选择的是大类和子类。

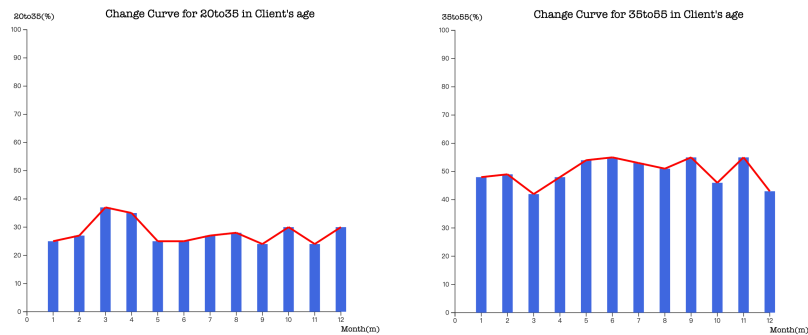
- 通过本模块，我们能够在 2.3 节宏观结论的基础上获得更加细致的结果分析：（酒店住户相应特征随时间的变化规律以及造成变化的因素）
 - 住户来源地上，本地住户倾向于在春天、秋天和冬天入住，而其他地区的住户则倾向于在春末、夏天和初秋入住，二者近似互补。可以推测，其他地区的住户入住目的多为旅游，而本地居民入住目的多为商业，原因是其他地区住户的入住高峰期正好为暑假期间。因此，住户来源地的特征主要受季节（是否处于假期）的影响，房价、活动等因素的影响因子较低。



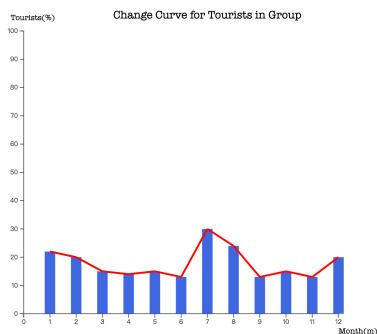
- 住户性别上，男性住户始终占据大头，这可能也与住户目的主要为商业有关。此外，10和12月为女性住户相对比例的最大时间段，再结合热力图上的 TT 和大型活动信息，推测可能该时间段内存在某种对女性更具吸引力的活动。因此，住户性别特征主要受住户目的和大型活动的影响。



- 住户年龄上，青年住户（20岁至35岁）更倾向于在3、4月入住该酒店，而中年住户（35岁至55岁）则更倾向于在其他时间内入住，再结合热力图上的 TT 和大型活动信息，推测4月可能存在某种对青年住户更具有吸引力的非商业活动。因此，住户年龄特征主要受大型活动的影响。



- 住户目的上，带有旅游目的的住户主要集中在7、8月和1月，分析原因可能是因为当前季节正处于暑假/寒假。因此，住户目的特征主要受季节的影响。



2.6 季节因素对酒店的影响

- 住户特征上的影响：
 - 当季节处于气候宜人，或位于寒暑假期间时，以旅游为目的的住户比例将会相对提高；
 - 当季节位于寒暑假时，来自外地的住户比例将会提高；否则，来自本地的住户比例将会提高；
- 酒店淡、旺季（客流量）的影响：
 - 旺季：冬末（2月）、春天（4、5月）、夏初（6月）、秋天（9、10、11月）
 - 淡季：冬天（12月、1月）、初春（3月）、夏天（7、8月）
- 酒店平均房价的影响：当位于春天或秋天时，酒店会适当提高平均房价；当位于冬天或夏天时，酒店会适当降低平均房价；
- 大型活动的影响：当位于春天、夏初、秋天和初冬时，当地更倾向于举办大型活动（该因素又会间接影响到住户特征）。
- 以下因素具有较为明显的相似季节性变化规律：
 - 住户目的和住户来源地；
 - 客流量和大型活动。

2.7 其他结论

- 住户的房间预订方式以直接现场预订为主，其次是通过代理人预订，最后是空勤预订，可以推测该酒店数据的年份较为久远，或当地数字化程度相对滞后。