

出租车订单数据分析

一、数据检验

对原始数据进行判断：

- 订单的距离和费用因为必为正数，故不能有明显异常值（包括负数）。
- 若同一辆出租车有多次订单，则通常情况下每次订单的起始时间应在上一次订单的结束时间之后。

经检查，无异常数据。则挑选起终点都在上海市嘉定区的出租车订单数据进行分析。

二、出租车出行方式的时空特征

2.1 订单生成和完成数的小时变化

为了分析 4 月 18 日上海市嘉定区每个时间段的出租车订单开始和结束的订单数量，做如下折线图：（1_1.py）

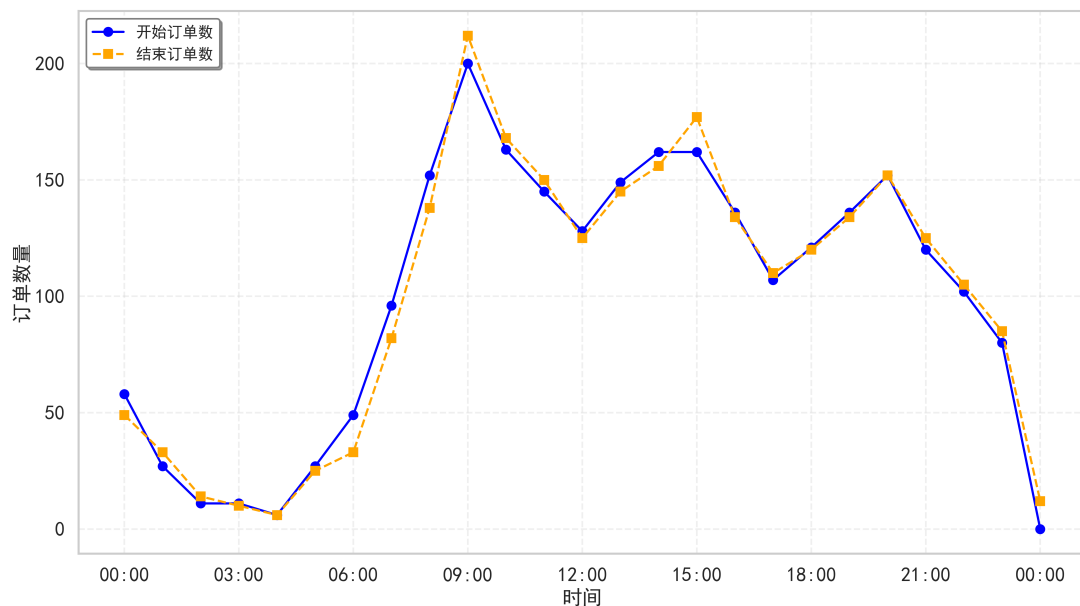


图 1 小时订单数量

- 由图可知在 6 : 00 ~ 9 : 00 和 17 : 00 ~ 20 : 00 期间，订单数量有明显的增长（特别是在早上这个时间段），表明这是一天中的高峰期。早高峰时主要因为有大量的通勤需求（工业园区员工上班），晚高峰时则存在较少的通勤需求（从商业区域或地铁站返回居民区）。
- 凌晨时段 (0 : 00 ~ 5 : 00)，订单数量非常少，说明这段时间是一天中的低谷期，出租车需求较低。因为大多数人此时休息或无出行需求。中午 (11 : 00 ~ 14 : 00) 订单数量略有下降，可能与大多数

居民及出租车司机需要午休有关。

- 开始订单和结束订单的趋势基本一致（斜率同增同减），这表明订单的起始和完成时间之间较为均匀，且大多数订单的持续时间较短。具体来说，订单开始和结束数量曲线的同步反映了嘉定区内短距离订单的比例较高。

2.2 订单时长分布

为研究每次订单的时长分布，用原始数据的订单结束时间减去订单开始时间即可得到每次订单时长。对时长进行描述性分析，如下图所示：（1_2.py）

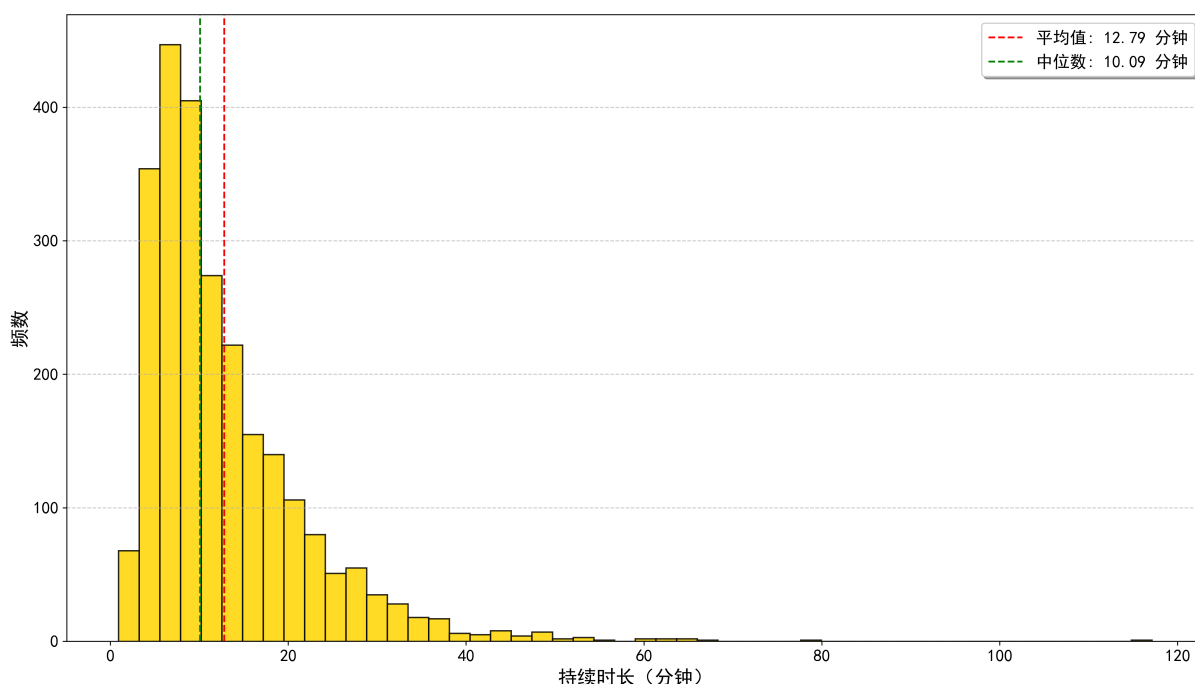


图 2 订单时长分布图

从图中可以看到上海市嘉定区出租车行驶的持续时长分布存在显著的偏态特征，以下是从图中得到的主要信息：

- 订单持续时长的分布呈现出典型的右偏分布，绝大多数的行程时长集中在较短的时间范围内，例如 0-20 分钟。这表明出租车的主要服务是短距离出行需求，这可能与嘉定区的城市功能布局有关，居民的日常通勤距离相对较短。
- 从平均值和中位数的对比可以发现，平均持续时长（约 12.79 分钟）高于中位数（约 10.09 分钟）。这表明少数长时间订单拉高了整体的平均值。尽管如此，这些长时间订单仅占小比例，不是主要的出行类型。
- 长时间订单（持续时长超过 40 分钟的订单）数量明显稀少，这表明嘉定区内的长距离出行需求有限。这可能是因为嘉定区范围内的活动或需求密集地较为集中，长距离需求可能由其他交通方

式满足（如地铁、公交等）。

- 频率分布显示，0-10 分钟的行程占比非常高，可能是周边居民前往商业区、学校、地铁站等短途目的地的需求。相比之下，超过 30 分钟的订单几乎微不足道。
- 总体来看，嘉定区出租车订单时长的分布反映了该地区出行需求的特点，即以短途为主，偶尔有长途需求。这一趋势对优化出租车服务（例如短途定价策略）和调度规划提供了重要参考，同时也显示了区域居民的出行特点和生活习惯。

2.3 订单里程分布

为研究每次订单的里程分布，做出的订单里程分布图如下：（1_3.py）

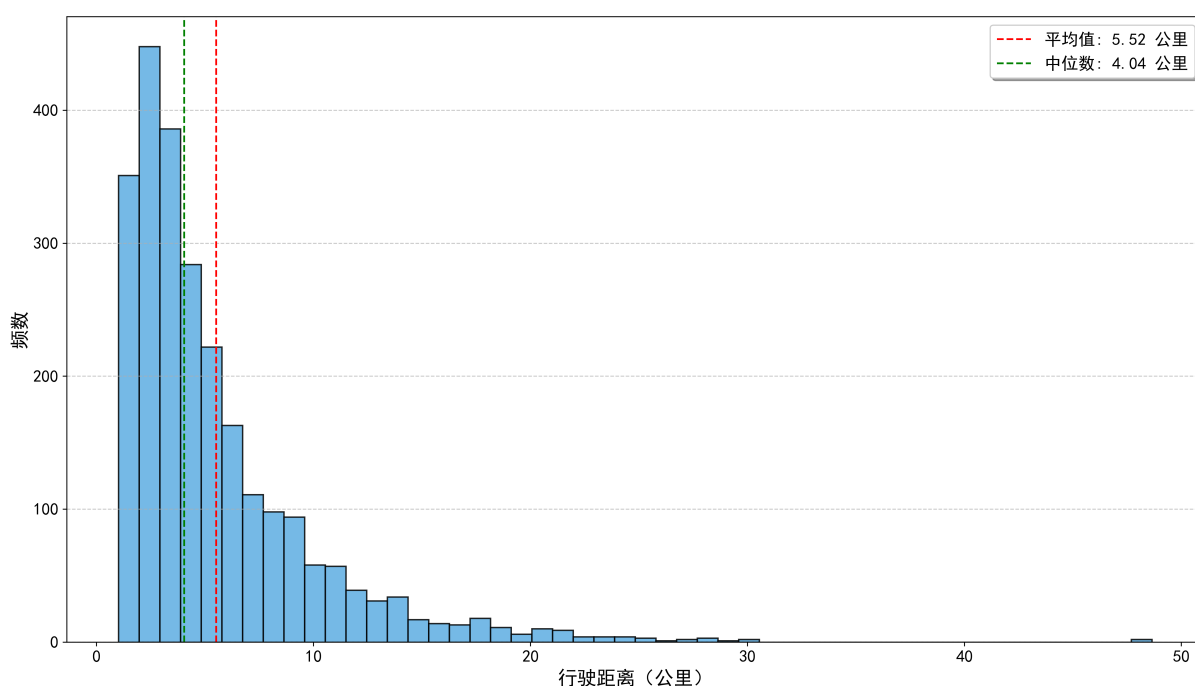


图 3 订单里程分布图

从图中可以看出，嘉定区当天的出租车行驶距离以较短距离为主，呈现出典型的右偏分布特征。大部分订单的行驶距离集中在 0 到 10 公里范围内，其中 4 公里左右的订单数量最多。图中还标注了平均值和中位数的位置，分别为 5.52 公里（红色虚线）和 4.04 公里（绿色虚线）。可以看出，平均值略高于中位数，表明存在少量较长距离的订单拉高了平均值。总体上，这种分布说明短途订单是主要构成，长途订单较为少见。这也与当天的出租车订单时长分布情况相统一。

2.4 所有出租车每天完成的订单数量分布

为了直观地看出当天出租车的接单数量，做如下订单数量分布图：（1_4.py）

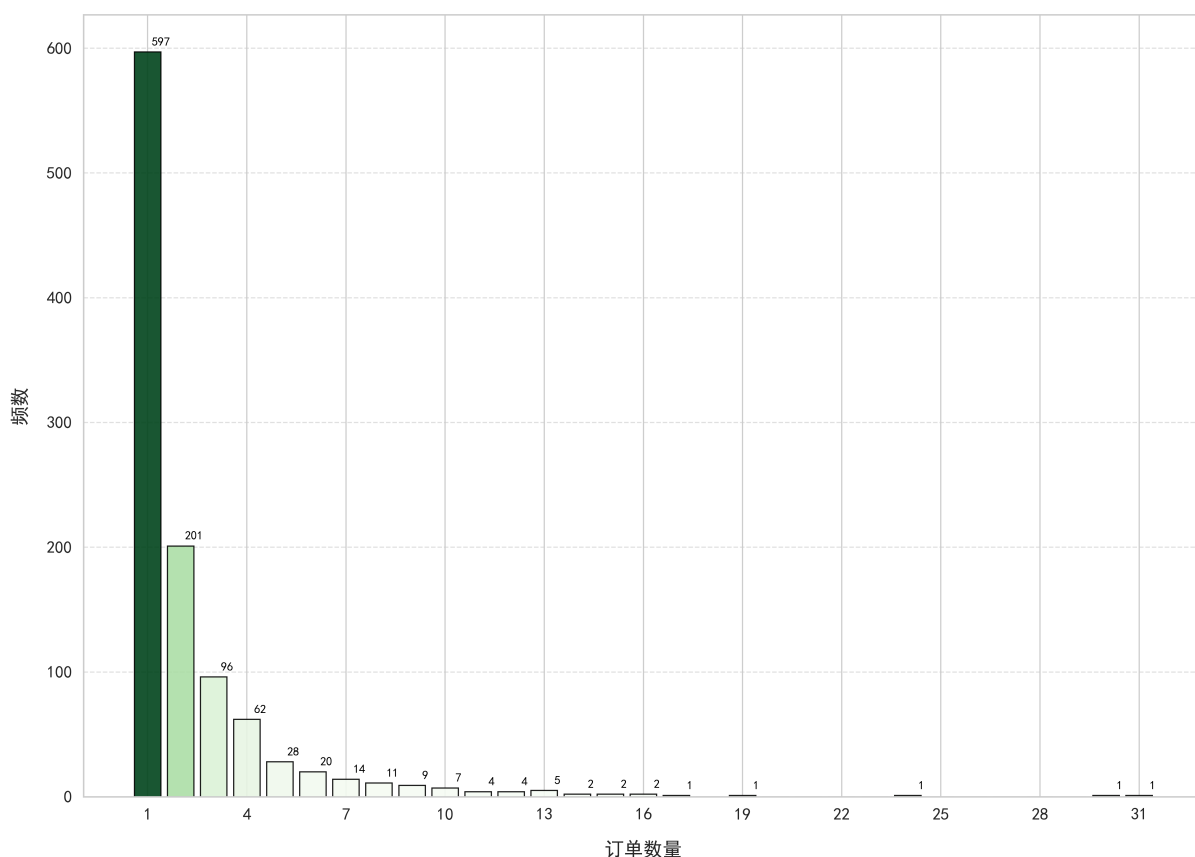


图 4 当天出租车订单数分布情况

图中横坐标表示出租车每天接收的订单数量，纵坐标表示对应订单数量的出租车频数。通过柱状图的颜色梯度变化可以发现，第一根柱子的颜色最深，表示接收订单数量最少的出租车频数最大，随着订单数量的增加，柱状图的颜色逐渐变浅，表明订单数量多的出租车频数明显减少。

从数据分布来看，接收 1 单的出租车数量最多，达到 597 辆，占绝大多数；接收 4 单、7 单的出租车数量分别显著减少，分别为 201 辆和 96 辆。这种分布呈现出典型的长尾特征，即绝大部分出租车完成的订单量较少，仅有少数出租车完成了较多订单。

另外，在接近尾部的数据中，完成订单数量在 16 单及以上的出租车非常少见，甚至完成 22 单、28 单、31 单的车辆仅有 1 辆。这可能反映出在当天的嘉定区运营中，少数出租车由于工作时间更长或接单效率更高，从而完成了更多订单。

这种分布可能与区域出租车的供需特性、乘客的打车选择情况以及司机的接单偏好有关。例如，完成 1 单的车辆可能为短时运营的车辆或接到长时间、长距离的订单，而完成大量订单的车辆可能全天候在岗。总体上，图形反映了嘉定区出租车运营的明显不均衡特性，短途、少量订单构成了主体，而高效接单的车辆为极少数。

三、出租车接客的热点区域

以下分析均针对起点在嘉定区，接单时间在早高峰（7:00—8:00）时间段内的出租车订单数据。

3.1 接客订单数量热力图

本题以 WGS84 地理坐标参考系统为准，绘制出 2018 年 4 月 18 日早高峰时段嘉定区的接单数量热力图如下：（2_1.py）

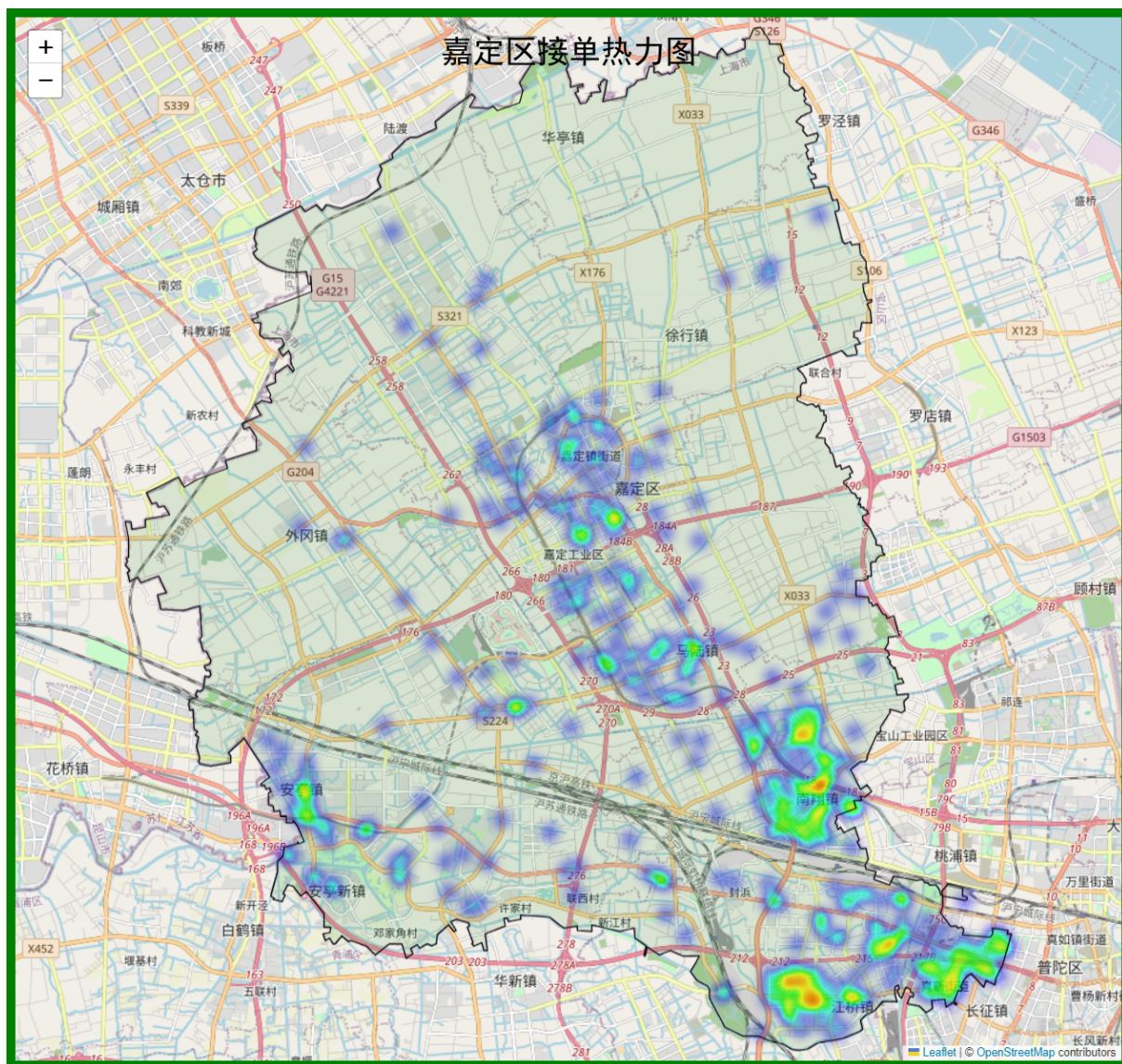


图 5 嘉定区接单热力图

这张热力图中颜色的变化代表了接单量的差异，其中红黄色区域表示接单量高度集中，而蓝绿色区域则显示接单量较为稀疏。从图中可以看出，接单量较高的区域主要集中在金鹤公园附近的生活区、南翔镇、外环高速 S20 东南区等等，这些地方是一些商业中心、居民区密集的区域，或是与交

通枢纽、重要公共设施相关的地段。

另一方面，嘉定区的边缘和相对偏远的地方接单量显著较少，反映了这些区域可能人口密度较低，或者经济和商业活动较少。这些地方的服务需求不足可能也与地理位置偏僻、交通不便有关，导致服务资源的供给和分布可能较少。

整体来看，这张热力图不仅展示了嘉定区内接单数量的分布情况，也为区域内服务资源的优化提供了重要的参考依据。例如，企业（出租车公司）可以针对高需求区域增派资源，提升服务能力，同时对低需求区域进行进一步的市场调查，评估潜在需求或是否存在服务盲点。通过这样的分布特征分析，可以更有效地制定调度策略和资源分配计划，以满足区域内不同时段和地段的服务需求。

3.2 热定区域出租车选站建议

选择出租车停靠点时应综合考虑站点的建设环境、交通条件、乘客出行需求以及与其他交通方式的衔接性。在具体选址时，应考虑出租车停靠点的如可达性、交通流量、周边土地利用性质等，使停靠点尽可能满足不同时间段和空间分布下的乘客需求，同时降低对交通流的干扰 [1]。

本题选择以金鹤公园（图中东南方向的红黄色区域）为中心的接客热点区域。为下图的区域增设一个出租车停靠点（橙色圆形区域），如图所示：

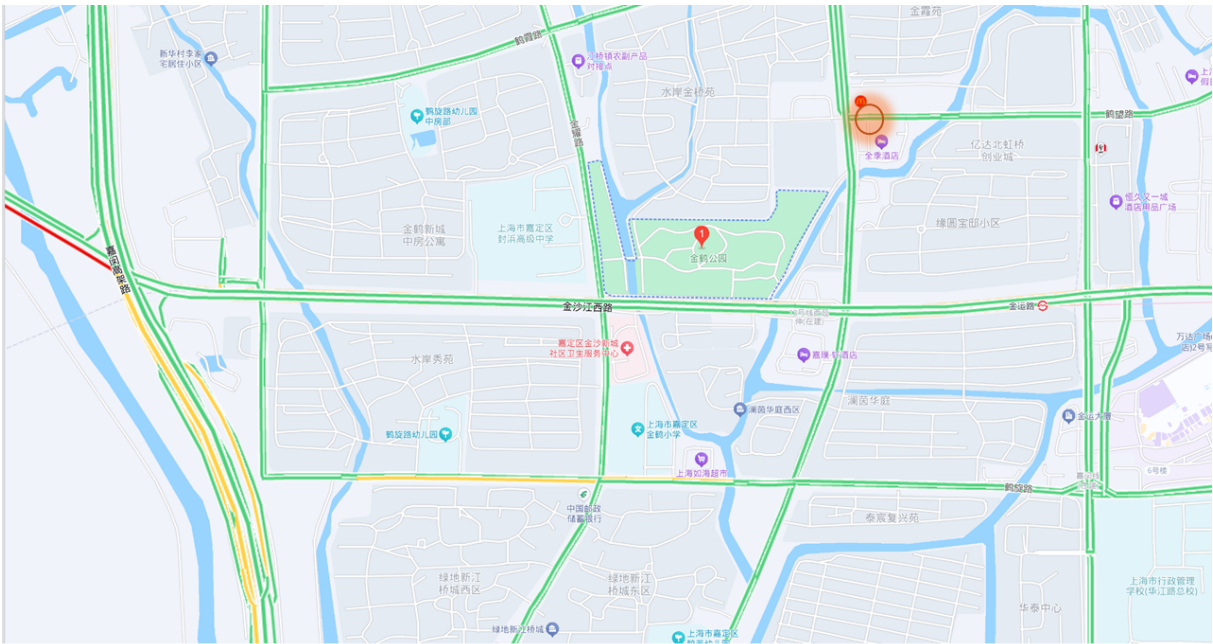


图 6 金鹤公园附近交通图

该图基于百度地图的实时路况，图中绿色线条表示在早高峰区域该道路畅通，黄色线条表示早高峰区域道路较拥挤。选择该停靠点的原因如下：

- 热点区域的匹配性：图中的棕色圆形区域靠近金鹤公园，麦当劳餐厅，全季酒店及一些商业区等

设施，这些区域是典型的高人流量场所。根据上述论文中分析出行热点的方法，类似的商业、休闲和公共设施周围通常是出租车乘客上下车需求集中的地方。

- 道路条件：棕色区域位于绿标的畅通路段，避免了黄色拥堵路段，符合论文中提到的选择交通流畅地段设置停靠点的原则。位置远离主要交叉口，可避免对交通流动的进一步干扰。
- 多交通模式的接驳：该位置邻近步行街区及金霞苑，水岸金桥苑等社区，可以方便周边居民快速接驳出租车出行，特别是在步行范围内的出行者需求较大的地方。临近公交站和餐饮区域的交汇点，具有较高的综合交通需求。
- 步行可达性：根据论文中提到停靠点选择方案，选址需确保服务范围覆盖的步行可达距离合理，通常选址要确保 500 米以内的步行覆盖范围。离该站点 500 米范围内，约有三个大型社区和一个公园。假设人的步行速度为 4.5 公里/小时（约 1.25 米/秒），5 分钟步行的距离约为 375 米，10 分钟约为 750 米，即在较短的时间内就能到达该停靠点。若是骑行，则距离该停靠点更远的小区也能很快到达。

参考文献

- [1] 金雷. 考虑出行需求的出租车停靠站选址及优化技术研究. Master's thesis, 哈尔滨工业大学, 2014.