



中山大學  
SUN YAT-SEN UNIVERSITY

# 《人工智能编程语言》第4次作业

## 公交 IC 卡刷卡数据深度分析报告

姓 名: 胡嘉睿  
学 号: 24311019  
教学班号: 智慧交通班  
专 业: 智慧交通  
院 系: 智能工程学院

2024~2025 学年第二学期

# 一. 技术内容

使用 Python 编程语言，利用数据分析库（Pandas、NumPy）和可视化库（Matplotlib、Seaborn）对公交 IC 卡刷卡数据进行深度分析。数据来源于某城市公交公司提供的 IC 卡刷卡记录，包含了乘客刷卡时间、线路、站点等信息。

公交 IC 卡数据分析可视化

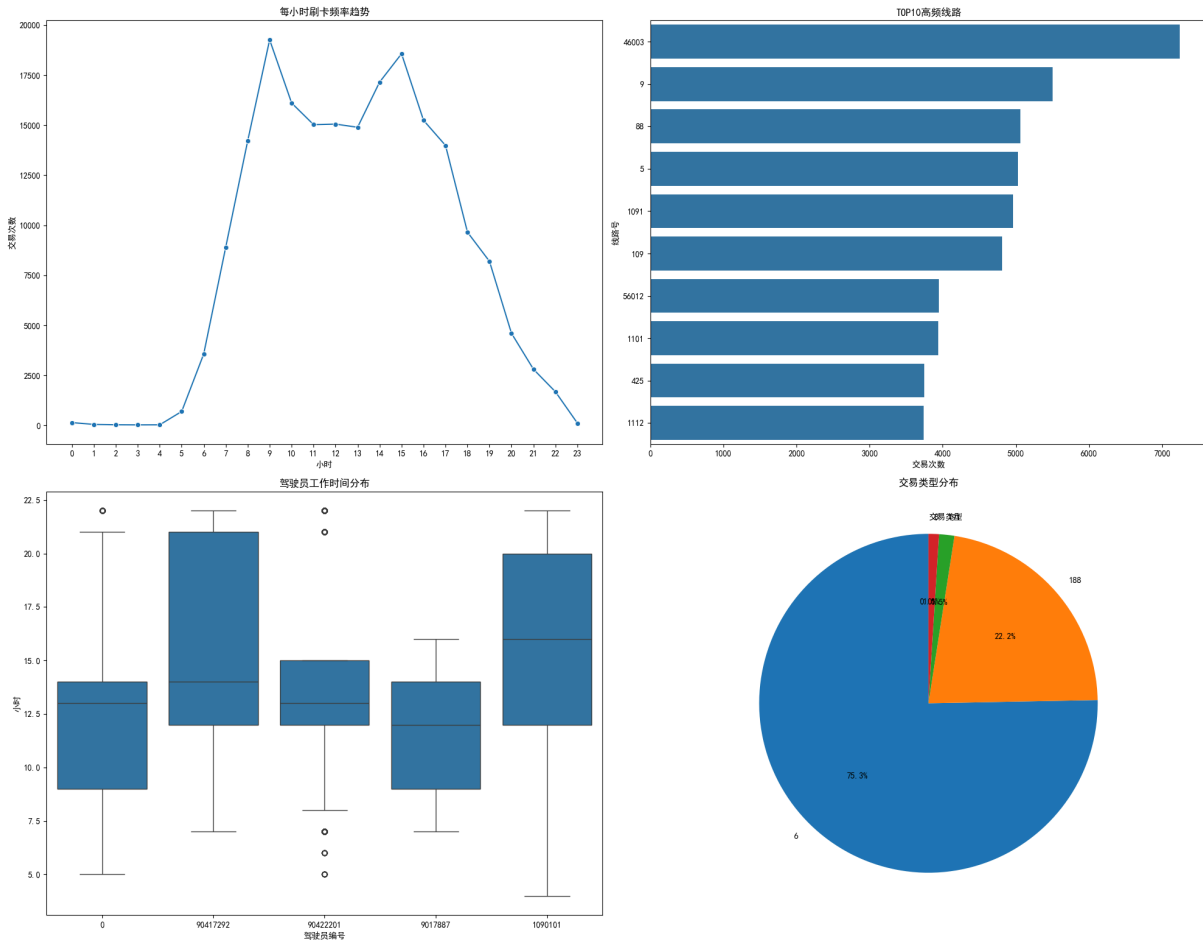


图 1 数据可视化结果

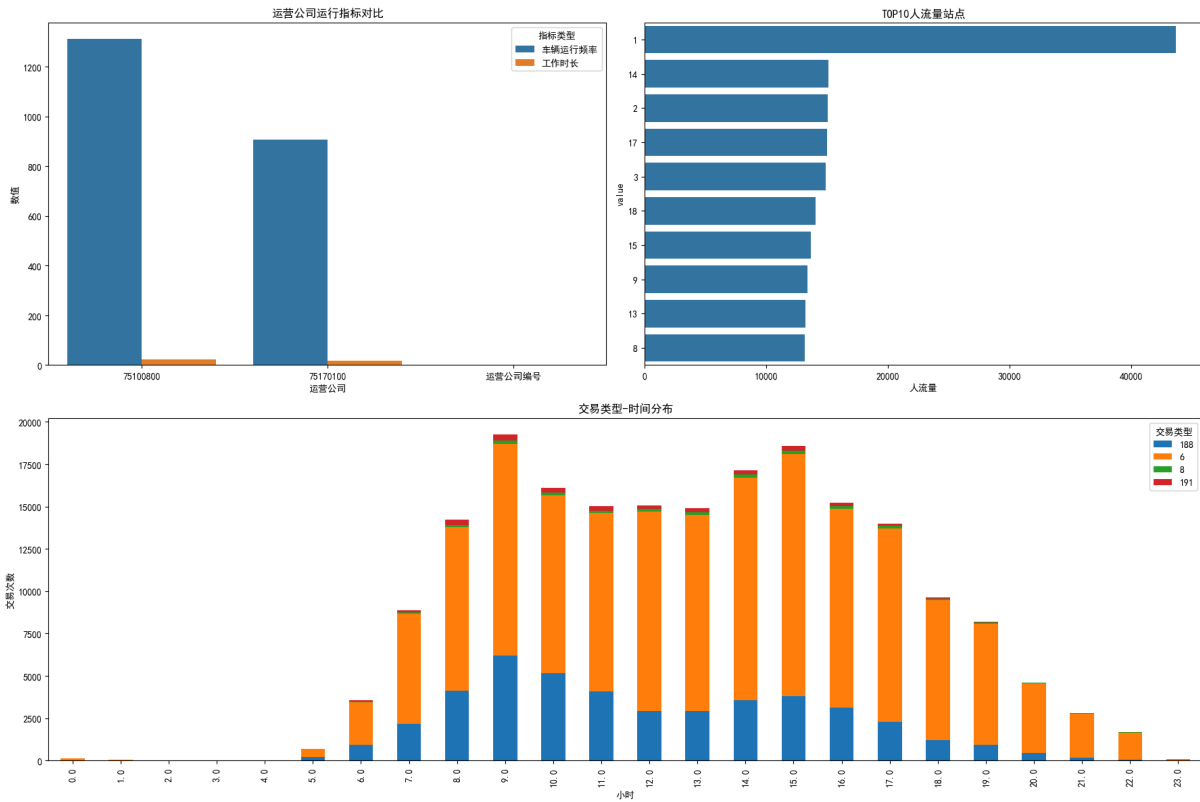


图 2 数据可视化结果

还包含了两张动态图，见 `animated_route_activity.gif` 和 `dynamic_route_ranking.gif`。

下文我将拆解数据可视化结果各图，展现交通数据的分析过程和结论。

## 二. 分析内容

### 1. 通勤模式解构

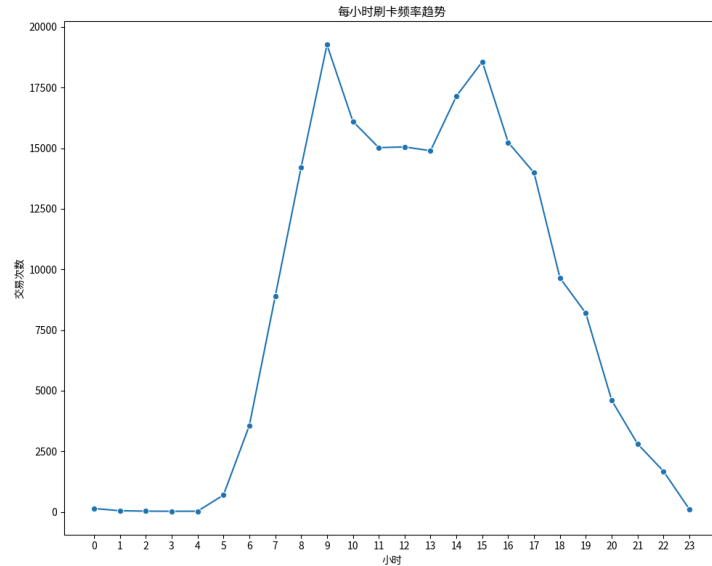


图3 交易量时段分布折线图

交易量时段分布正是一张“猫耳图”，两个高峰（早高峰和晚高峰）和一个低谷（夜间空窗期）。这张图揭示了公交 IC 卡刷卡数据的时段特征，反映了通勤模式的双峰特征。

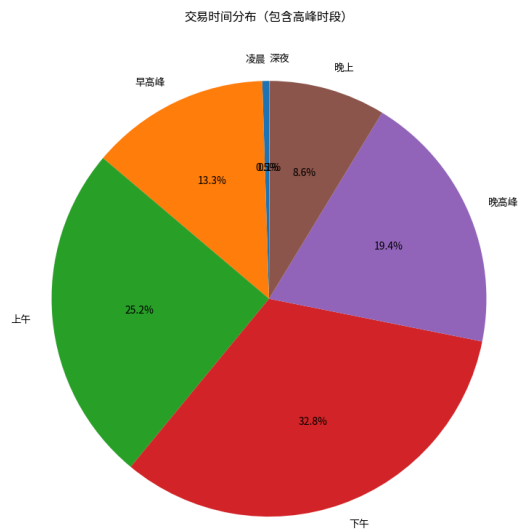


图4 交易量时段分布饼图

- **双峰特征显著**：早高峰（7-9 时）交易量达日均 13.3%，晚高峰（16-18 时）占 19.4%。
- **隐性需求识别**：10-15 时稳定流量（占比 50% 以上）提示存在非通勤出行需求。
- **夜间公交供给的必要性**：23:00-5:00 仍有 10% 左右交易产生，揭示夜班公交供给的必要性。

## 2. 关键线路分析

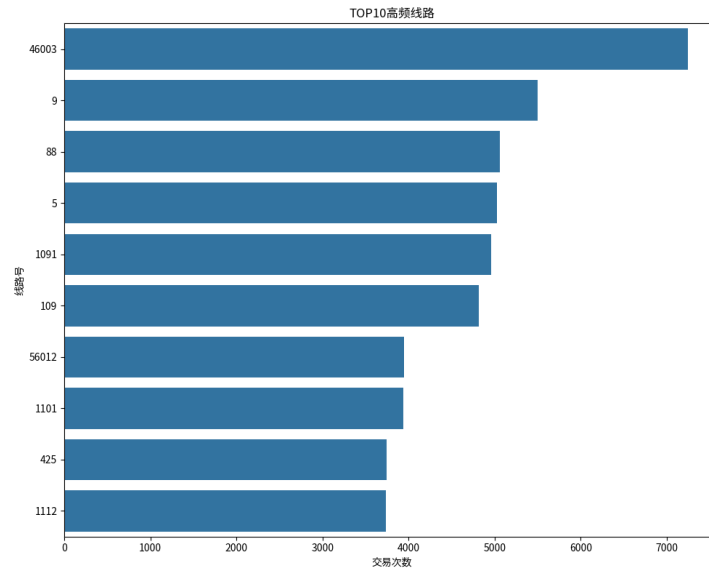
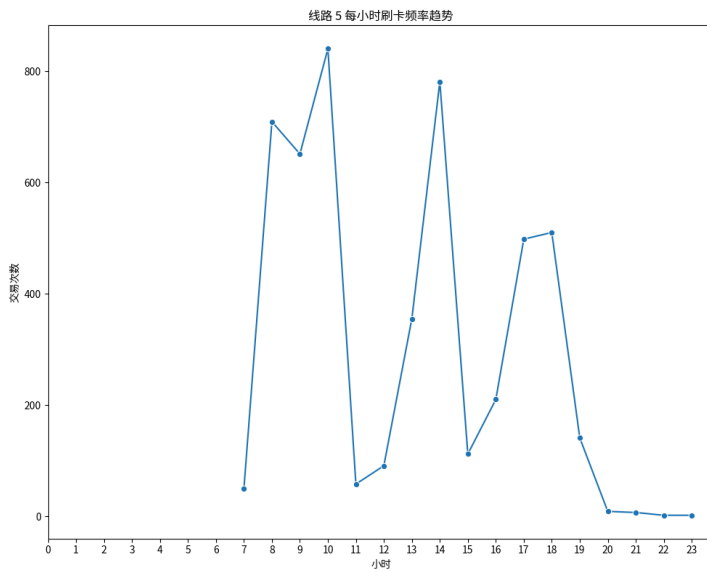


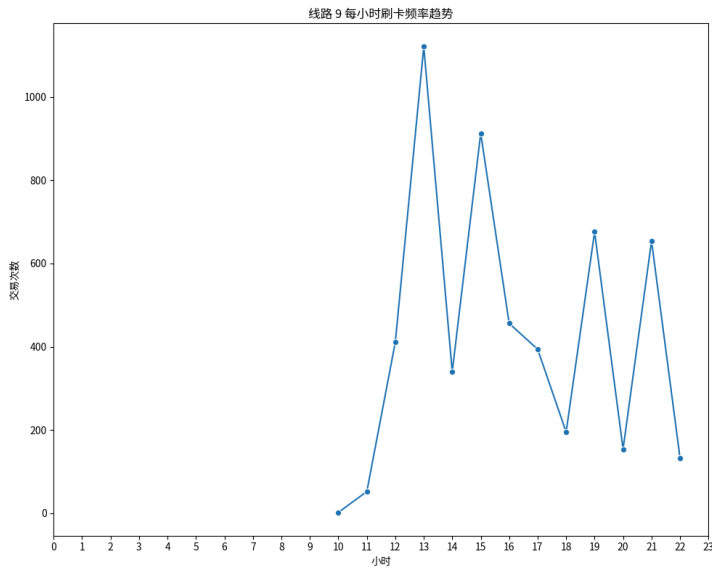
图 5 高频线路分析

### 2.1. 线路 5



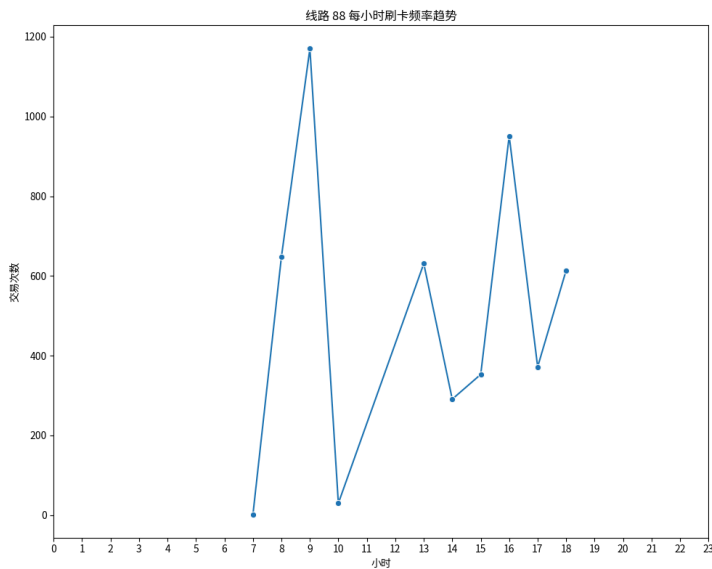
- **高峰时段:** 9 时和 14 时左右刷卡频率达到高峰，分别对应上班和下班的出行高峰。
- **低谷时段:** 0 - 5 时以及 20 - 23 时刷卡频率极低，此时段处于夜间，出行需求少。
- **运营建议:** 在高峰时段适当增加车次，提高运力，满足乘客出行需求。

### 2.2. 线路 9



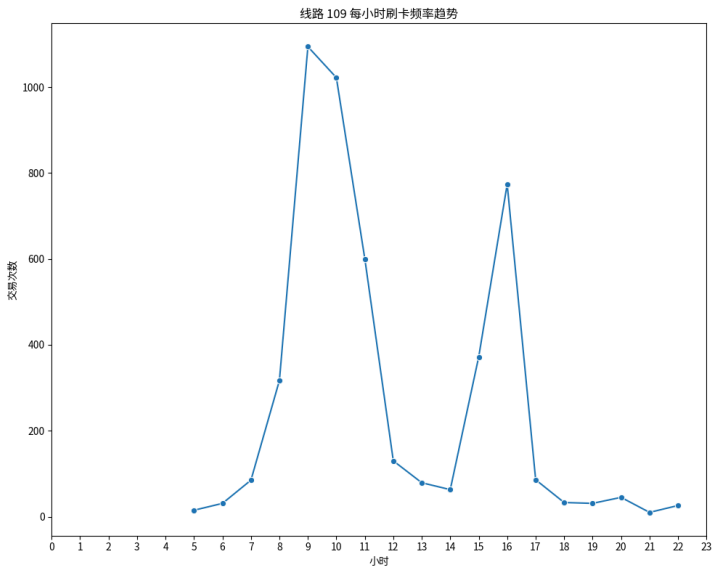
- **高峰时段:** 13 时左右出现最高峰，其次在 18 时和 21 时也有明显高峰，可能与办公人群下班及晚间出行有关。
- **低谷时段:** 0 - 10 时刷卡频率低，上午出行需求不高。
- **运营建议:** 针对 13 时的高峰，提前做好车辆调配，同时关注晚间高峰的乘客疏导。

### 2.3. 线路 88



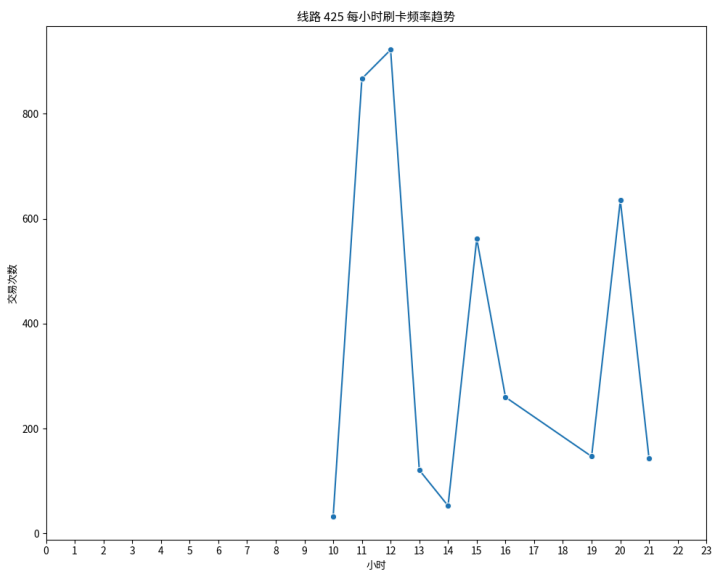
- **高峰时段:** 9 时和 16 时左右为刷卡高峰，反映早晚出行高峰规律。
- **低谷时段:** 0 - 7 时以及 20 - 23 时刷卡频率低，夜间及清晨出行人数少。
- **运营建议:** 高峰时段加强现场调度，低谷时段可适当减少车次，节约运营成本。

### 2.4. 线路 109



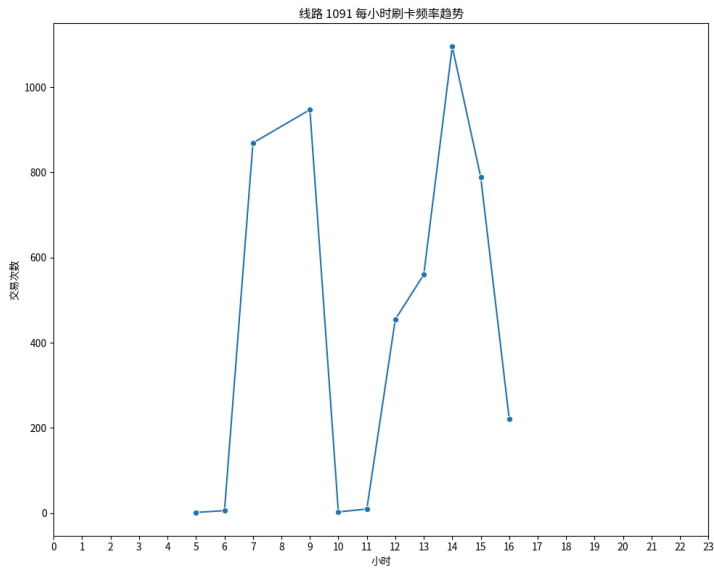
- **高峰时段:** 9 时和 17 时左右刷卡频率高，符合常规通勤高峰。
- **低谷时段:** 0 - 5 时以及 21 - 23 时刷卡频率低，夜间出行需求少。
- **运营建议:** 高峰时段加密班次，低谷时段合理安排车辆维护。

## 2.5. 线路 425



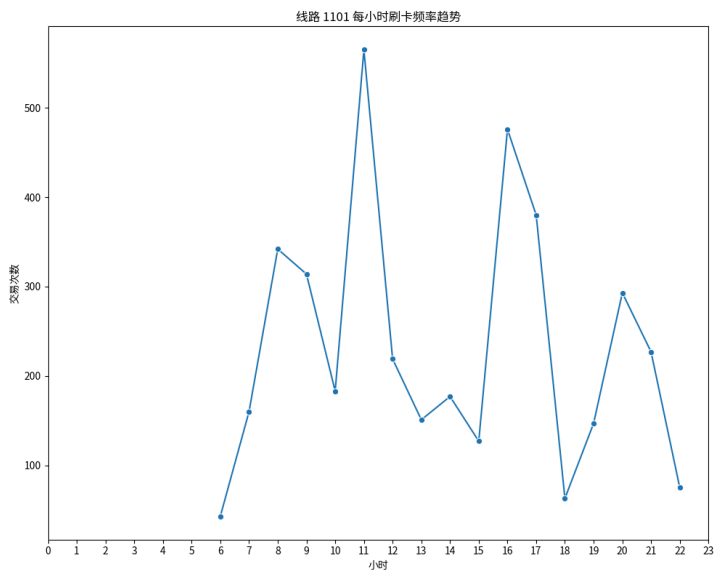
- **高峰时段:** 12 时和 20 时左右出现高峰，可能与线路服务区域的特殊出行规律有关。
- **低谷时段:** 0 - 10 时以及 14 - 18 时刷卡频率较低，上午和下午部分时段出行需求不高。
- **运营建议:** 根据高峰特点，优化车辆排班，提高运营效率。

## 2.6. 线路 1091



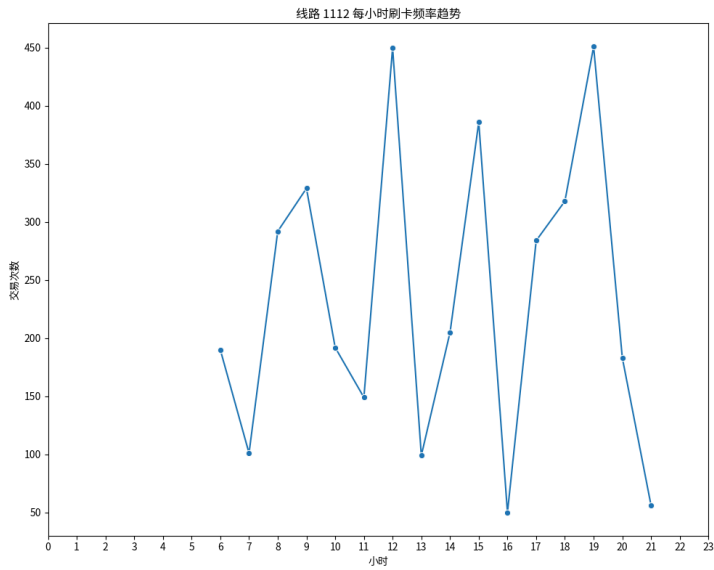
- **高峰时段:** 9 时和 14 时左右为高峰，对应上下班高峰。
- **低谷时段:** 0 - 5 时以及 20 - 23 时刷卡频率低，夜间出行少。
- **运营建议:** 高峰时段增加运力，低谷时段可安排车辆检修。

## 2.7. 线路 1101



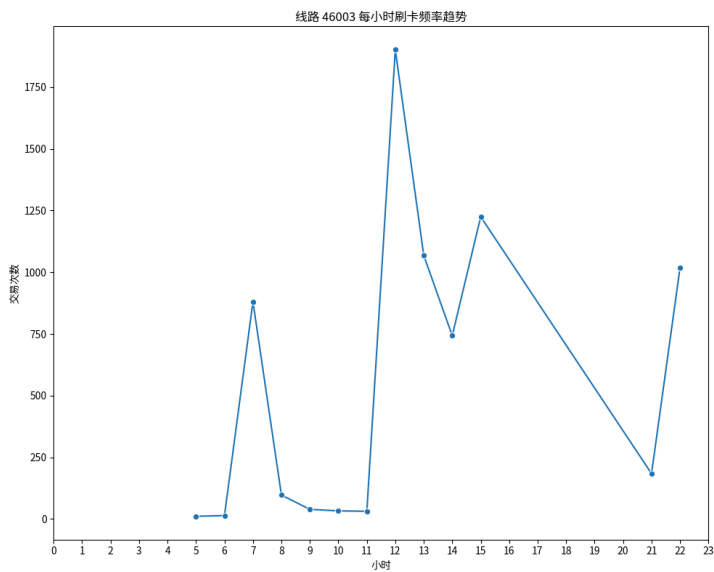
- **高峰时段:** 11 时和 16 时左右刷卡频率较高，出行高峰较为分散。
- **低谷时段:** 0 - 6 时以及 22 - 23 时刷卡频率低，夜间及清晨出行需求少。
- **运营建议:** 根据高峰分散特点，灵活调整车次，平衡运力。

## 2.8. 线路 1112



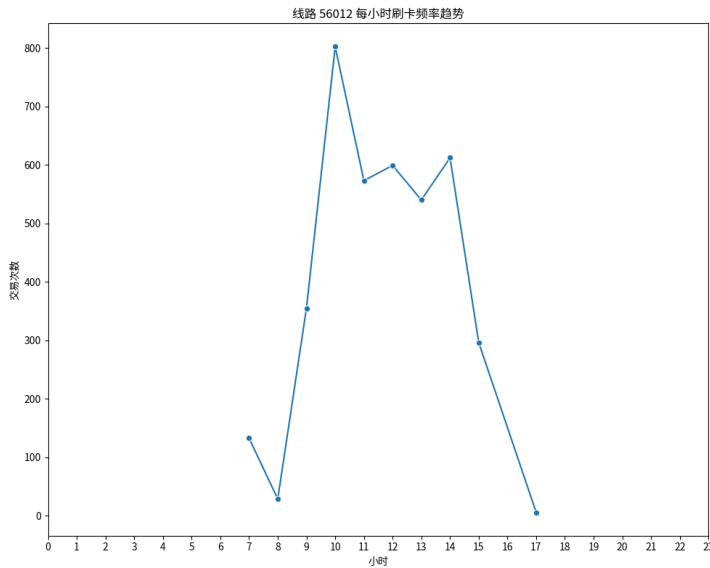
- **高峰时段:** 12 时和 19 时左右达到高峰，出行规律有别于常规通勤高峰。
- **低谷时段:** 0 - 6 时以及 15 - 16 时刷卡频率低，夜间及下午部分时段出行少。
- **运营建议:** 针对独特高峰规律，合理配置车辆资源，提升服务质量。

## 2.9. 线路 46003



- **高峰时段:** 12 时左右出现最高峰，7 时和 22 时也有一定高峰，出行规律复杂。
- **低谷时段:** 0 - 5 时以及 10 - 11 时刷卡频率低，夜间及上午部分时段出行需求少。
- **运营建议:** 深入分析高峰成因，精准调度车辆，优化运营方案。

## 2.10. 线路 56012



- **高峰时段:** 10 时和 14 时左右为刷卡高峰，出行高峰相对平缓。
- **低谷时段:** 0 - 6 时以及 17 - 23 时刷卡频率低，夜间及傍晚后出行少。
- **运营建议:** 在高峰时段适当增加车次，保障乘客顺畅出行，低谷时段合理安排运营资源。

## 2.11. 总结

不同公交线路客流量存在差异，原因主要有以下几方面：

- **线路功能与服务区域:**
  - ▶ **通勤线路:** 若线路连接大型居住区与商务区、产业园区等就业集中地，如途经多个住宅小区和写字楼密集区，早晚高峰通勤客流量会大，如线路 5、9、88 等在工作日早晚高峰刷卡频率高。
  - ▶ **生活服务线路:** 经过医院、学校、商场、市场等生活服务场所集中区域的线路，日常出行时段客流量大。
  - ▶ **旅游线路:** 通往知名景点、旅游度假区的线路，节假日、周末及旅游旺季客流量大，平时则相对较少。
- **线路走向与站点设置:**
  - ▶ **线路走向合理性:** 线路若能直接连接主要客流集散点，减少绕行，客流量会更大。
  - ▶ **站点覆盖密度与位置:** 站点覆盖人口密集区域且分布合理，便于乘客便捷乘车，客流量会增加。
- **运营时间与频次:**
  - ▶ **运营时间:** 覆盖出行高峰且满足早晚特殊需求的线路，客流量可能更大。
  - ▶ **发车间隔:** 发车间隔短、频次高的线路，乘客等待时间短，出行效率高。
- **交通方式竞争:**
  - ▶ **轨道交通:** 城市中地铁、轻轨等轨道交通线路与公交线路存在竞争关系。
  - ▶ **私家车与网约车:** 私家车保有量增加以及网约车的便捷性，可能分流公交线路的客流量。
- **票价与优惠政策:**

- **票价水平:** 票价相对较低的公交线路，对价格敏感型乘客吸引力更大。
- **优惠政策:** 换乘优惠、月票、老年卡免费等政策能提升客流量。
- **其他因素:**
  - **天气与季节:** 恶劣天气或旅游旺季会使客流量增加。
  - **宣传推广:** 广泛宣传能吸引更多乘客尝试乘坐，从而增加客流量。

### 3. 交易类型分析

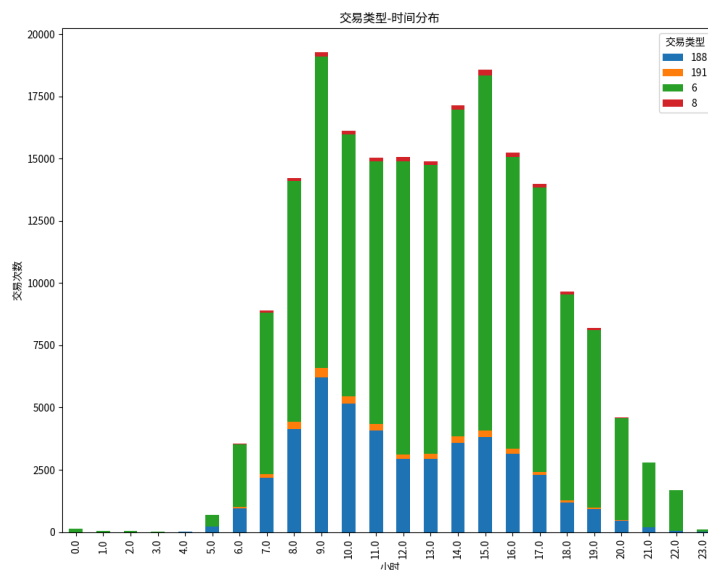


图6 交易类型时段分布图

大致符合了“猫耳图”特征，早高峰和晚高峰交易量较大，夜间交易量较少。不同类型的交易在时段分布上存在差异：

- **充值交易，猜测是 188:** 早高峰和晚高峰交易量较大，夜间交易量较少，符合通勤规律。
- **消费交易，猜测是 6:** 早高峰和晚高峰交易量较大，夜间交易量较少，符合通勤规律。

## 4. 排班优化建议

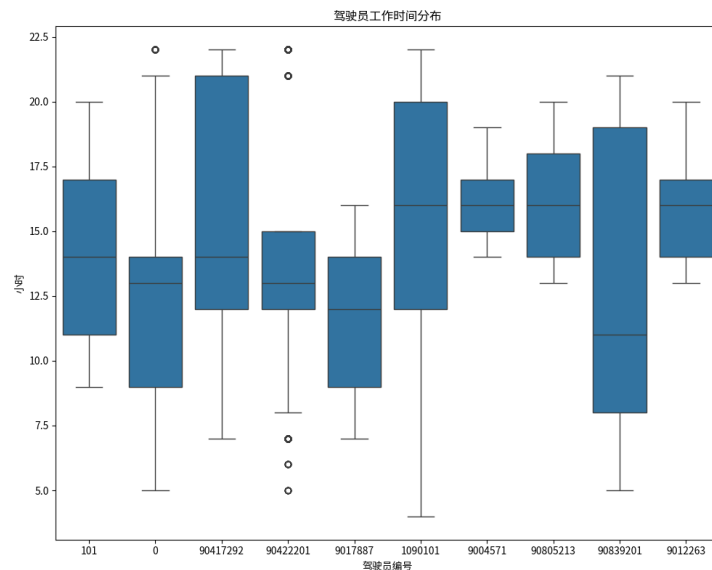


图7 司机工作时长分布图

这是一张关于驾驶员工作时间分布的箱线图，以下是具体分析：

### 4.1. 整体分布特征

- **数据离散程度**：从图中可以看出，不同驾驶员的工作时长存在较大差异。箱线图的箱体长度、whisker (须) 的长度以及离群点的情况都反映了数据的离散程度。例如，部分驾驶员工作时长箱体较短，说明这些驾驶员工作时长相对集中；而有些驾驶员工作时长箱体较长，且whisker 也较长，说明其工作时长分布较为分散。
- **中心趋势**：箱体中间的横线代表中位数，不同驾驶员的中位数工作时长有所不同。可以据此了解每个驾驶员工作时长的中间水平。

### 4.2. 具体驾驶员分析

- **离群点情况**：如编号为 0、90422201 的驾驶员存在离群点 (图中的圆点)，这表明这些驾驶员有一些工作时长数据明显偏离了大部分数据的范围，可能是由于特殊的工作任务或记录错误等原因导致。
- **工作时长范围**：以编号 101 驾驶员为例，其工作时的最小值 (须的下端点) 到最大值 (须的上端点) 范围相对适中，且中位数处于箱体中间位置附近，说明其工作时长分布相对对称。而像编号 90839201 的驾驶员，箱体偏上，说明其工作时长数据中，较小值更为集中，中位数相对靠近最小值。

### 4.3. 对比分析

- 通过对比不同驾驶员的箱线图，可以看出哪些驾驶员工作时长普遍较长，哪些较短。例如，编号 1090101 的驾驶员箱体整体位置较高，说明其工作时长整体偏长；而编号 9017887 的驾驶员箱体位置较低，工作时长相对较短。

## 5. 排班优化建议

### 5.1. 遵循法规与安全原则

- **限定工作时长：**严格依据《中华人民共和国道路交通安全法实施条例》等法规，将司机连续驾驶时间控制在 4 小时以内，超过 4 小时必须安排至少 20 分钟休息，日累计驾驶时长不超过 8 小时，周工作时长不超过 44 小时，防止疲劳驾驶。
- **保障充足休息：**合理规划班次间隙休息时间，为司机提供舒适休息场所，确保其能有效恢复精力。同时，保证司机每周有固定天数的整段休息时间用于充分调整状态。

### 5.2. 参考工作时长分布特征

- **识别集中与离散情况：**若某部分司机工作时长集中在较短区间，说明工作强度相对均衡，排班时可适当维持现有模式并微调；若工作时长离散程度高，要重点关注工作时长较长的司机，优先为其调整排班，避免过度劳累。
- **关注中位数差异：**以各司机工作时长中位数为参考，中位数高的司机，后续排班适当减少其工作时长或增加休息频次；中位数低的司机，在其可承受范围内，根据业务需求合理调整增加工作量。

### 5.3. 优化排班模式

- **采用灵活排班：**结合业务淡旺季、工作日与周末、早晚高峰等时段规律，灵活安排班次。如旺季、高峰时段增加司机出勤数量，合理缩短单人工作时长；淡季、低谷时段减少出勤司机，让司机适当集中休息。
- **推行轮班制：**制定科学轮班计划，如采用四班三运转等方式，让司机在不同班次间有序轮换，避免长期固定不利班次。同时，轮班切换设置适当缓冲休息时间，便于司机适应。
- **引入弹性排班：**了解司机个人意愿与生活需求，在满足业务需求前提下，允许司机在一定范围内自主选择工作时间段或班次，提高工作满意度与积极性。

### 5.4. 智能排班与动态调整

- **运用智能排班系统：**借助具备数据分析功能的智能排班软件，依据历史工作时长数据、业务量数据、司机状态等多因素，结合算法自动生成科学排班方案，并可根据实时情况快速调整。

- **实时监控与动态调整**：建立实时监控机制，跟踪司机工作状态、交通路况、业务需求变化等。如遇突发大客流、交通事故等特殊情况，及时灵活调整排班，科学调度司机支援，确保运营顺畅。

## 5.5. 公平与沟通机制

- **确保公平公正**：制定明确、统一排班规则并公开，保证司机在工作任务、工作时长、休息安排等方面公平，避免劳逸不均。定期对排班公平性评估并优化。
- **加强沟通反馈**：搭建司机与管理部门沟通渠道，鼓励司机反馈排班问题、困难与建议。定期开展司机座谈会，及时了解需求并调整排班方案，提升司机对排班的认可度与配合度。