**🧠 公交站点与延误数据分析思路大纲**

**✅ 第一部分：准备与验证（数据质量与结构理解）**

1. **理解每一列的含义与单位**

• Stop, StopName: 站点唯一ID与名称

• Didok: 地点编号（可能用于地理匹配）

• Platform: 站台编号

• Short\_Name: 简化代号（可能代表线路或系统内简称）

• Average\_Delay\_Seconds: 平均延误（秒）

• Latitude, Longitude: 站点坐标

• On\_Second\_Class, Off\_Second\_Class: 二等舱上下车人数（单位需要确认）

2. **检查数据质量**

• 是否有缺失值（如经纬度、延误）

• 延误是否有异常（如负值极大、极小）

• 经纬度是否都在 Fribourg 或 CH 范围内

• 确认坐标类型（WGS84）

3. **检查与线路关联逻辑**

• Short\_Name 是否表示公交线路？

• 是否可以从其他数据中映射到具体线路？（需结合 PLC 文件）

**✅ 第二部分：基础探索分析**

4. **按站点聚合 (Group by Stop/StopName)**

• 计算：

• 平均延误 (mean)

• 上下客量 (sum(On\_Second\_Class), sum(Off\_Second\_Class))

• 上下客总量、比例、换乘强度等指标

• 识别：

• 延误最严重的站点 Top-N

• 上车/下车量最多的站点（如枢纽候选）

5. **绘制地图热力图或标注图**

• 用 Plotly / Folium / QGIS 可视化：

• 延误热度

• 上下客密度

• 结合地理聚类（KMeans / DBSCAN）分析“热点区域”

**✅ 第三部分：深入分析思路**

6. **结合 PLC 文件分析“上下文信息”**

• PLC 文件 = 每辆车的行驶记录，包含车站、时间等

• 可以验证：

• 延误是在哪些时段出现？

• 哪些线路/方向容易发生延误或 bunching？

• 哪些站点经常出现在线路终点？（有助于 Thread 1 的last-mile设计）

• 分析特征：

• 延误与出发/到达时间的关系

• Door opening/closing 事件识别是否精确（判断真实到站时间）

7. **探查异常值与非周期事件**

• 延误 > 某阈值，或者为负（提前）

• 是否有周期性？节假日 vs 工作日

• 分析异常延误是否集中在某些站点或时间段

**✅ 第四部分：面向项目目标的分析（Thread 1 & 2）**

**🎯 Thread 1 - Last-mile 服务定位**

8. **识别服务不足的站点区域（农村 + 枢纽）**

• 低上下客但高延误

• 距离最近的主干线路较远（可引入空间分析）

• 考虑人口密度/POI数据联动（找潜在需求区域）

9. **结合地图与运营指标挑选 last-mile 候选点**

**🎯 Thread 2 - 拓展/替代固定线路服务**

10. **城市中心区域分析：**

• 高延误/拥堵区域识别

• 公交结队（bunching）现象检测

• 动态需求波动点位标记（用 PLC 推出 Demand Peak）

**✅ 第五部分：进一步发展（可选）**

11. **构建预测模型**

• 用天气/时段/上下客量预测延误

• 延误 → 驱动微公交调度机制（基于需求预测）

12. **和地理空间数据集成（如 geo.admin.ch）**

• 精确定位站点行政区划、交通等级

• 可用于日后扩展进 MaaS 平台路径规划模块

**📌 总结建议（行动路线）**

你可以从以下步骤入手：

1. ✅ 熟悉字段、检查数据质量和单位

2. ✅ 用 groupby 统计延误、上下车量

3. ✅ 做基本地图或表格热力图

4. ✅ 引入 PLC 数据进行时序分析

5. ✅ 针对 Thread 1 挖掘 rural 枢纽服务不足点

6. ✅ 针对 Thread 2 找城市中心延误热点和拥堵替代机会