

GIỚI THIỆU TỔNG QUAN VỀ BỘ DỮ LIỆU

Bộ dữ liệu của cuộc thi bao gồm thông tin về 30 tuyến xe buýt tại Thành phố Hồ Chí Minh cùng nhật ký GPS của các phương tiện tham gia trong tháng 4 năm 2025. Toàn bộ dữ liệu được lưu trữ trong 3 thư mục, phân loại thành các nhóm sau: Thông tin về lộ trình và trạm dừng của từng chuyến, Nhật ký GPS di chuyển của các phương tiện, Biểu đồ chạy xe.

NHẬN ĐỊNH VÀ CHIẾN LƯỢC

Nhận định:

- Dữ liệu GPS có giá trị cao trong việc hiểu hành vi di chuyển, đánh giá hiệu suất mạng lưới và đề xuất cải thiện ngay lập tức cho hệ thống xe buýt.
- Trong gần 9 giờ thi đấu, nhóm tập trung toàn bộ nguồn lực vào thu thập, làm sạch và phân tích dữ liệu GPS nhằm chuyển kết quả thành các giải pháp thực tiễn cho vận hành xe buýt tại Thành phố Hồ Chí Minh.

2 khía cạnh chính:

- Chất lượng trạm dừng — yếu tố thu hút hành khách
- Hành vi lái xe — yếu tố then chốt về an toàn và trải nghiệm

Chất lượng trạm dừng:

- Hệ thống trạm dừng phân bố rộng khắp nhưng mức độ sử dụng không đồng đều. Phân tích mật độ hành khách theo trạm cho phép xác định các điểm có nhu cầu cao, từ đó ưu tiên cải thiện cơ sở hạ tầng, an ninh và tiện ích tại những trạm trọng yếu để tăng sức hấp dẫn của xe buýt.

Hành vi lái xe:

- Tài xế là nhân tố quyết định sự an tâm của hành khách và an toàn giao thông. Những hành vi như lái nhanh, vượt ẩu không chỉ làm giảm trải nghiệm mà còn tiềm ẩn rủi ro va chạm, thiệt hại tài sản và trách nhiệm quản lý. Phân tích GPS cho phép phát hiện các mẫu lái nguy cơ cao và phát triển cảnh báo, báo cáo hoặc chương trình đào tạo nhắm tới cải thiện hành vi lái.

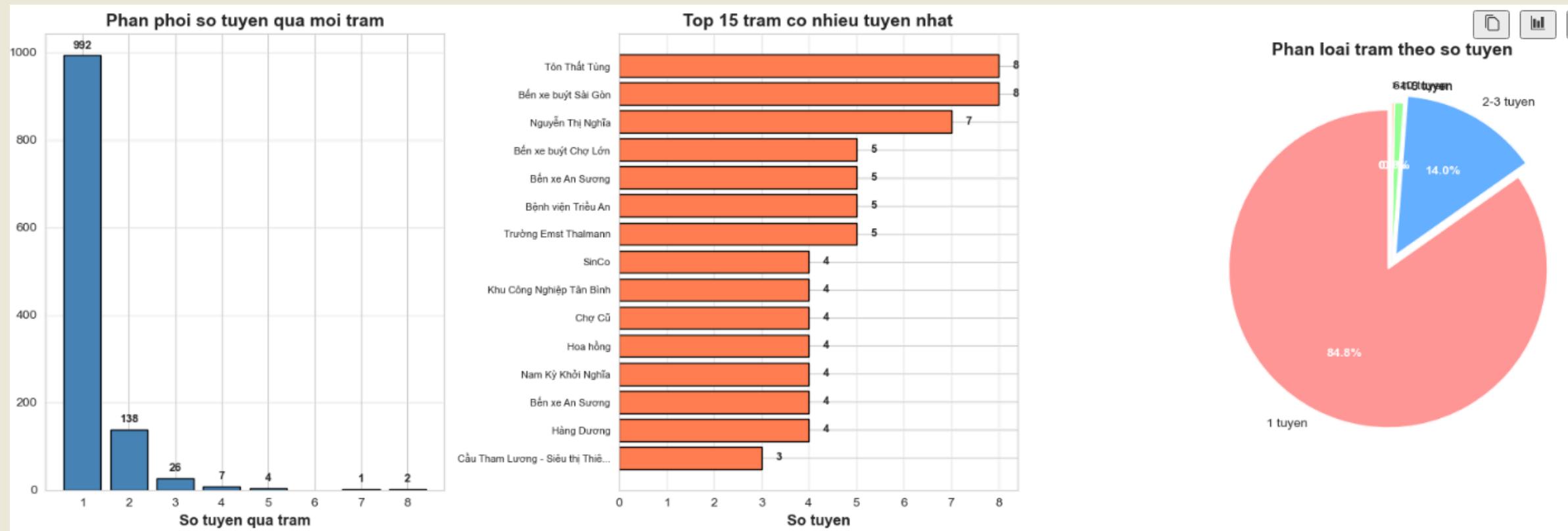
Hướng phân tích:

- Xây dựng dashboard thể hiện tần suất đón trả khách ở các trạm dừng theo từng chuyến.
- Xây dựng dashboard biểu diễn và so sánh thời gian hoàn thành tuyến của từng phương tiện.

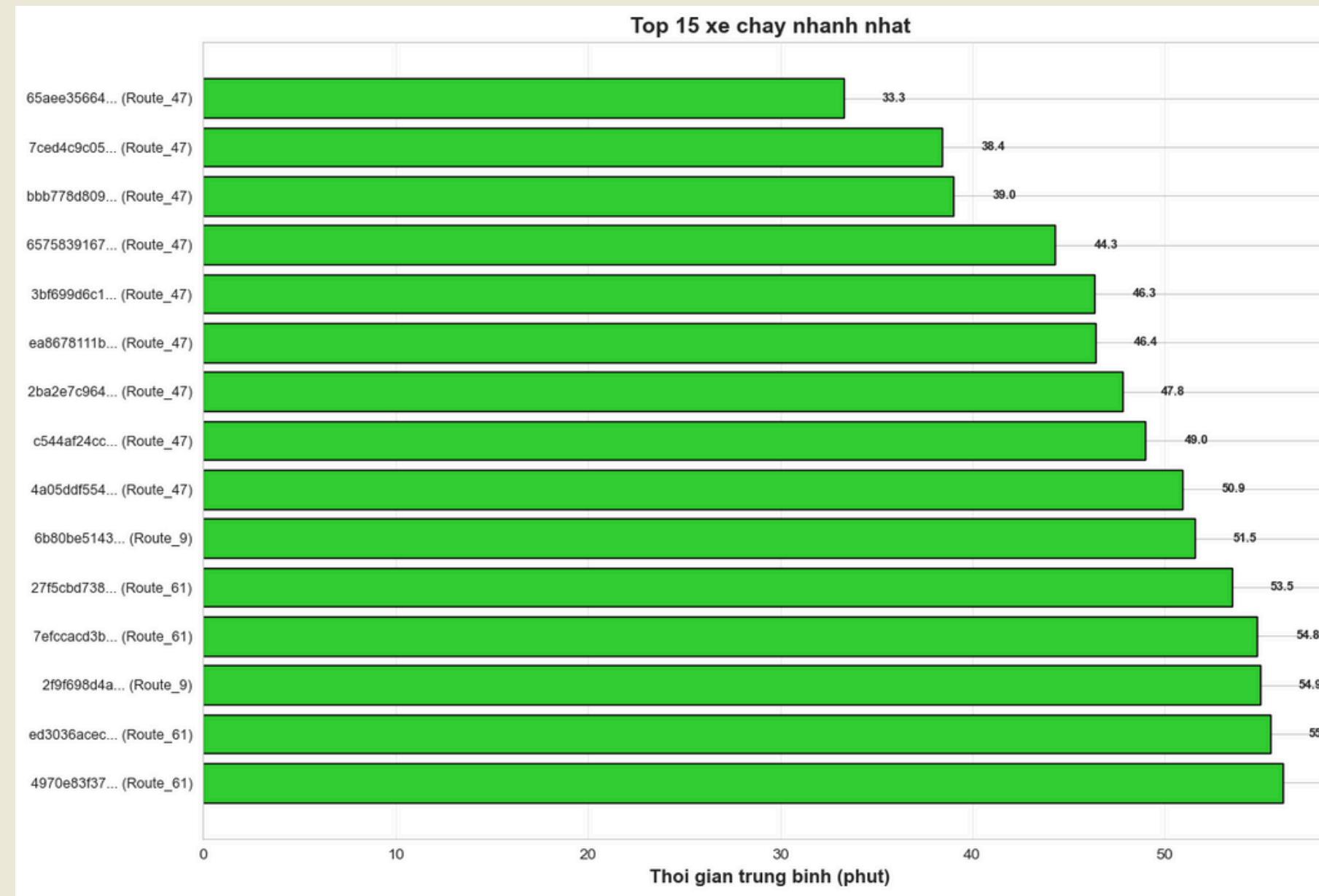
KHÓ KHĂN

- Cần chuyển đổi dữ liệu GPS thô và tình trạng mở cửa để tính toán tần số sử dụng trạm dừng
- Cần chuyển đổi dữ liệu GPS thô và thông tin của các trạm dừng để tính toán thời gian di chuyển trong của từng phương tiện.

PHÂN TÍCH VỀ CÁC CHUYẾN XE BUÝT



- Thực trạng: Mạng lưới rời rạc và "tuyến tính" - **84.8%** số trạm chỉ phục vụ duy nhất 1 tuyến.
- Khả năng trung chuyển trực tiếp tại trạm rất thấp.
- Hệ quả: Áp lực lên trạm cao điểm
- Nhu cầu chuyen tuyến bị dồn vào số ít các đầu mối lớn (Bến xe Sài Gòn, Chợ Lớn, An Sương...).
- Nguy cơ tắc nghẽn cục bộ cao.
- Giải pháp trọng tâm:
- Điều chỉnh lộ trình để tăng các điểm giao cắt dọc tuyến (không chỉ ở bến cuối).
- Mục tiêu: Nâng cao tỷ lệ các trạm có 2-3 tuyến đi qua.



- **Sự thống trị của Tuyến 47 (Route_47):**

- Chiếm 9/15 vị trí dẫn đầu.
- Đây là tuyến có hiệu suất thời gian tối ưu nhất toàn mạng lưới.

Biến động thời gian trên cùng một tuyến:

- Cùng là Route 47, nhưng xe nhanh nhất (65aee...) chỉ mất 33.3 phút.

Trong khi các xe khác cùng tuyến (như 4a05d...) mất tới 50.9 phút.

Chênh lệch: ~17 phút (tương đương nhanh hơn 35%).

Các tuyến tiềm năng khác:

Route 9 và Route 61 cũng xuất hiện trong top, thời gian dao động từ 51 - 56 phút.

2. PHÂN TÍCH TẦN SUẤT ĐÓNG/MỞ CỦA XE BUÝT TẠI CÁC TRẠM NHẰM ĐÁNH GIÁ MỨC ĐỘ HOẠT ĐỘNG VÀ LƯU LƯỢNG HÀNH KHÁCH TẠI TÙNG TRẠM DỪNG Ở TP.HCM

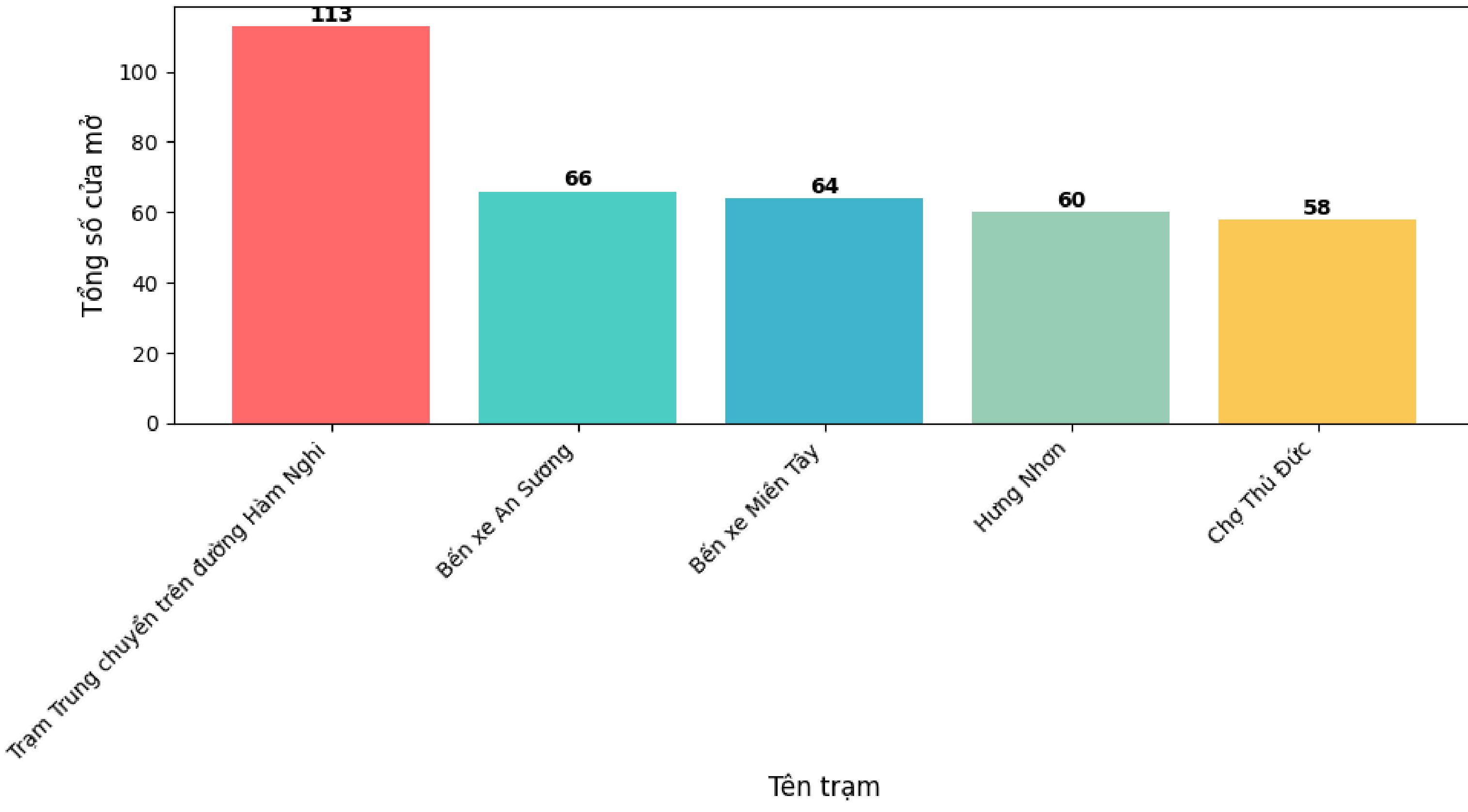
- Trước hết, dữ liệu GPS thô trong folder raw_gps đã được chuyển đổi để đồng bộ với tọa độ chính xác trong folder route.
- Dựa trên dữ liệu này xác định các trạm có mật độ hành khách đông nhất, sử dụng thông tin về sự kiện mở và đóng cửa của từng xe buýt tại từng trạm.

TRONG QUÁ TRÌNH PHÂN TÍCH, CÓ HAI VẤN ĐỀ CHÍNH CẦN LƯU Ý

Sai số khi đồng bộ số
liệu kinh độ, vĩ độ trong
file raw GPS

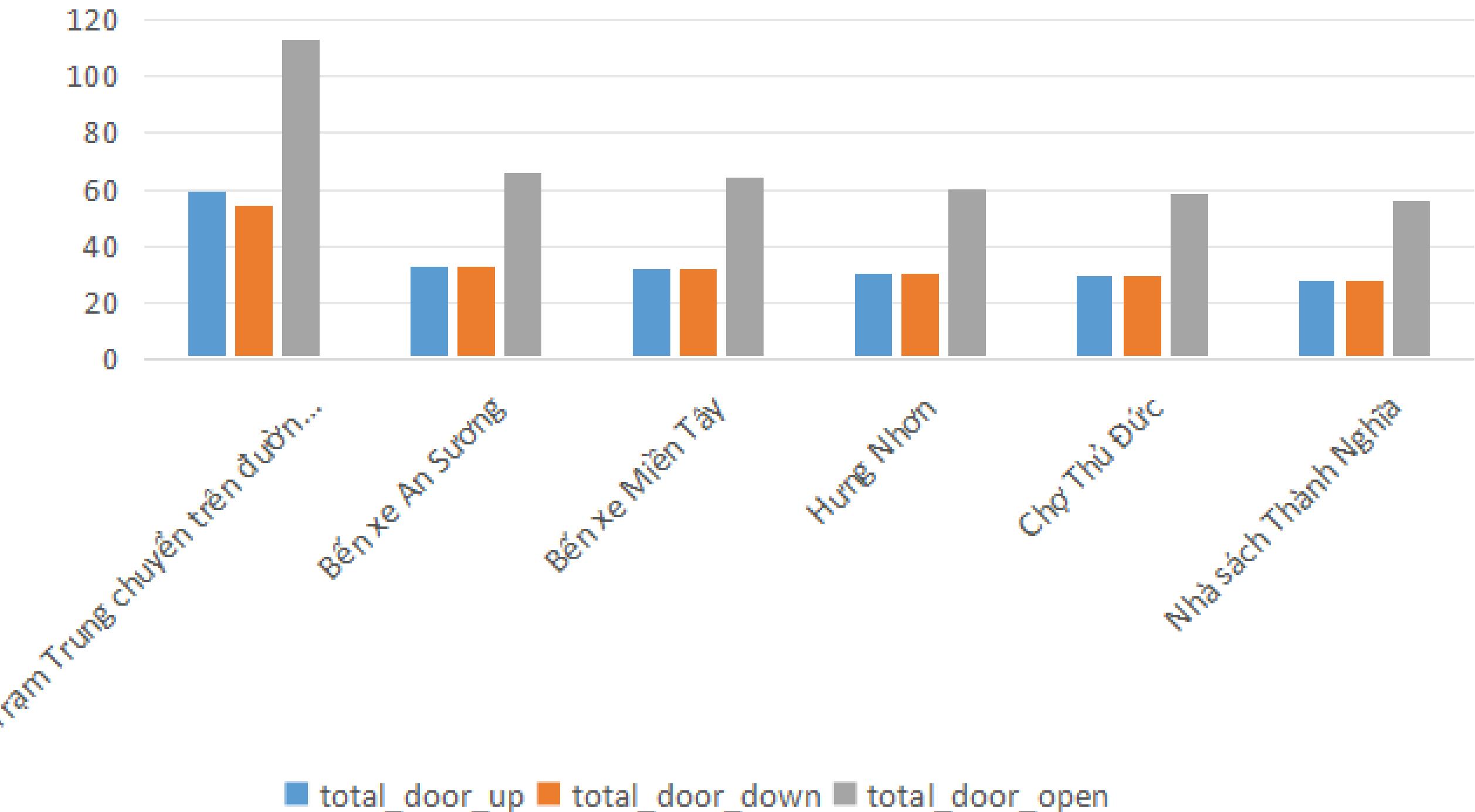
Trùng lắp sự kiện mở cửa: Ví
dụ 1 phút chỉ tính một lần mở
cửa cho cả hai cửa, thay vì
tính theo từng giây dữ liệu,
nhằm tránh phóng đại số
liệu.”

TOP 5 TRẠM CÓ SỐ CỬA MỞ CAO NHẤT



Hình 1: Biểu đồ thể hiện top 5 trạm xe buýt có mật độ hành khách cao nhất

Top 5 mật độ đông nhất tại các trạm xe buýt



HÌNH 2: BIỂU ĐỒ THỂ HIỆN CỤ THỂ TÌNH TRẠNG LÊN XUỐNG CỦA TẠI TRẠM XE BUÝT

HƯỚNG MỞ RỘNG NHÓM ĐỀ XUẤT

- Mật độ sử dụng của từng trạm theo từng chuyến và theo từng khoảng thời gian nhất định.
- Đánh giá các đoạn đường thường gây ùn tắc, chậm trễ dựa vào dữ liệu GPS, từ đó tối ưu hoặc quy hoạch tuyến đường mới cho xe buýt
- Tần suất làm việc của tài xế theo tình trạng giao thông thực tế để cải thiện môi trường làm việc

**CẢM ƠN MỌI NGƯỜI
ĐÃ CHÚ Ý LẮNG
NGHE BÀI THUYẾT
TRÌNH CỦA NHÓM 16**