



XÁC ĐỊNH RỦI DO KHI DI CHUYỂN TRÊN CÁC TUYẾN ĐƯỜNG CỦA HÀ NỘI

A44644 - Nguyễn Đình Nam



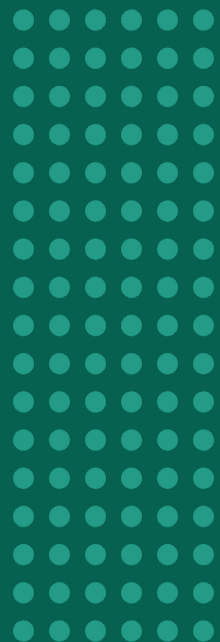
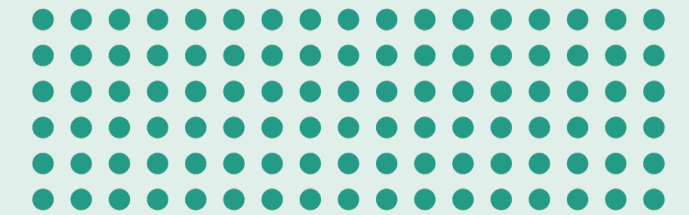
NỘI DUNG

1 Giới thiệu bài toán

2 Thu thập dữ liệu

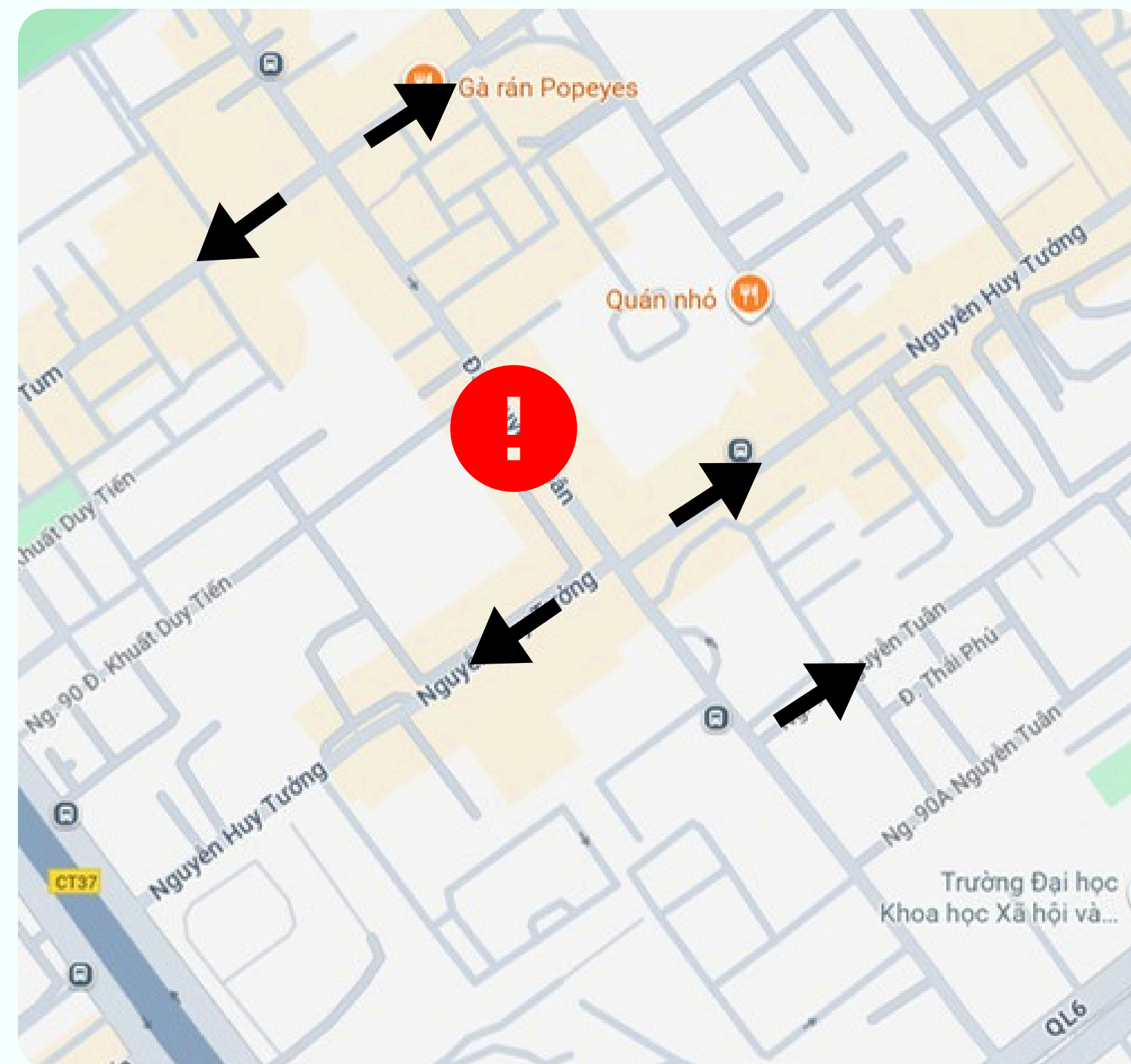
3 Làm sạch dữ liệu

4 Phân tích dữ liệu



VẤN ĐỀ

- Khi di chuyển, nếu gặp một tuyến đường bị ùn tắc, người tham gia giao thông thường chuyển sang các tuyến khác.
 - Việc đổi tuyến **không** đảm bảo rút ngắn thời gian, mà có thể khiến hành trình khó dự đoán hơn.
- Cần đánh giá rủi ro và độ tin cậy của việc di chuyển, không chỉ là tắc hay không tắc.



GIỚI THIỆU BÀI TOÁN

Bài toán: Đánh giá tình trạng và rủi ro di chuyển tại Hà Nội dựa trên dữ liệu thực tế.

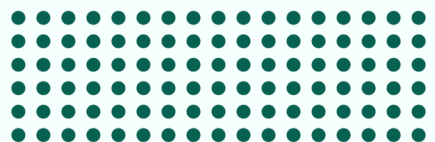
Ý tưởng: Sử dụng dashboard Power BI để phân tích giao thông từ tổng quan đến chi tiết theo thời gian, khu vực và điều kiện mưa.

Mục tiêu:

- Phân tích hiệu suất giao thông
- Nhận diện rủi ro ùn tắc
- Đề xuất khuyến nghị di chuyển

Khán giả chính:

- Cơ quan quản lý giao thông: ra quyết định và ưu tiên cải thiện hạ tầng
- Người tham gia giao thông: lựa chọn thời điểm và tuyến đường an toàn



THU THẬP DỮ LIỆU

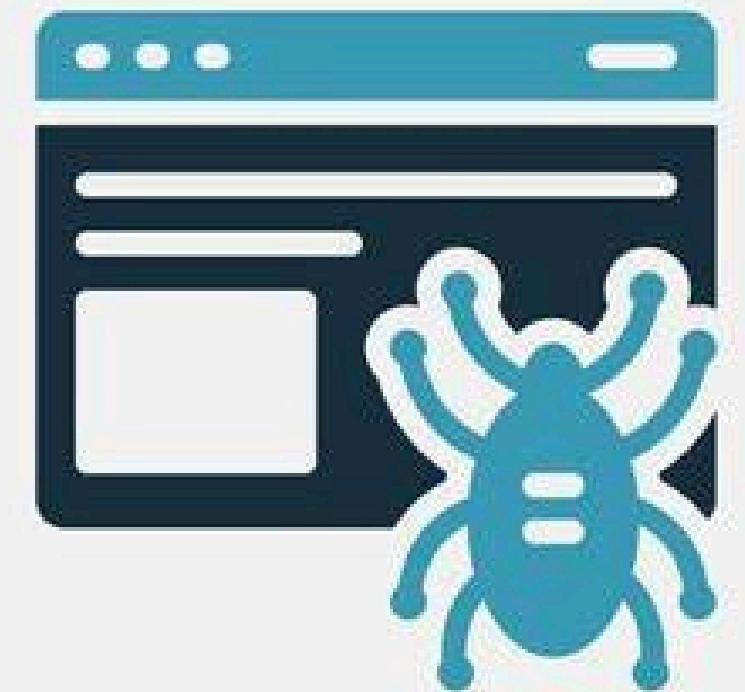
Nguồn: here.com(Giao thông) , www.weatherapi.com (Thời tiết)

Công cụ sử dụng: Selenium (Tự động hóa thu thập dữ liệu từ web),
BeautifulSoup (Phân tích và trích xuất dữ liệu từ HTML)

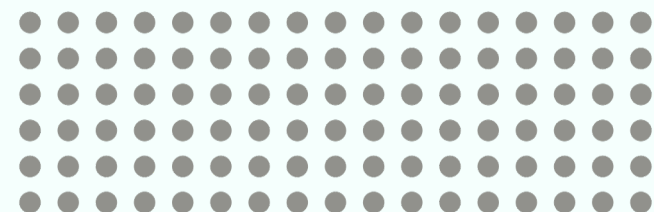
Dữ liệu lấy gồm: Dữ liệu giao thông (dtime, district, Street, speed, speedUncapped, freeFlow, jamFactor, confidence, traversability)
Dữ liệu thời tiết (Time, Date, Temperature, Rainfall ,Cloud Cover, Pressure ,Wind Speed, Description)

Quy trình thu thập:

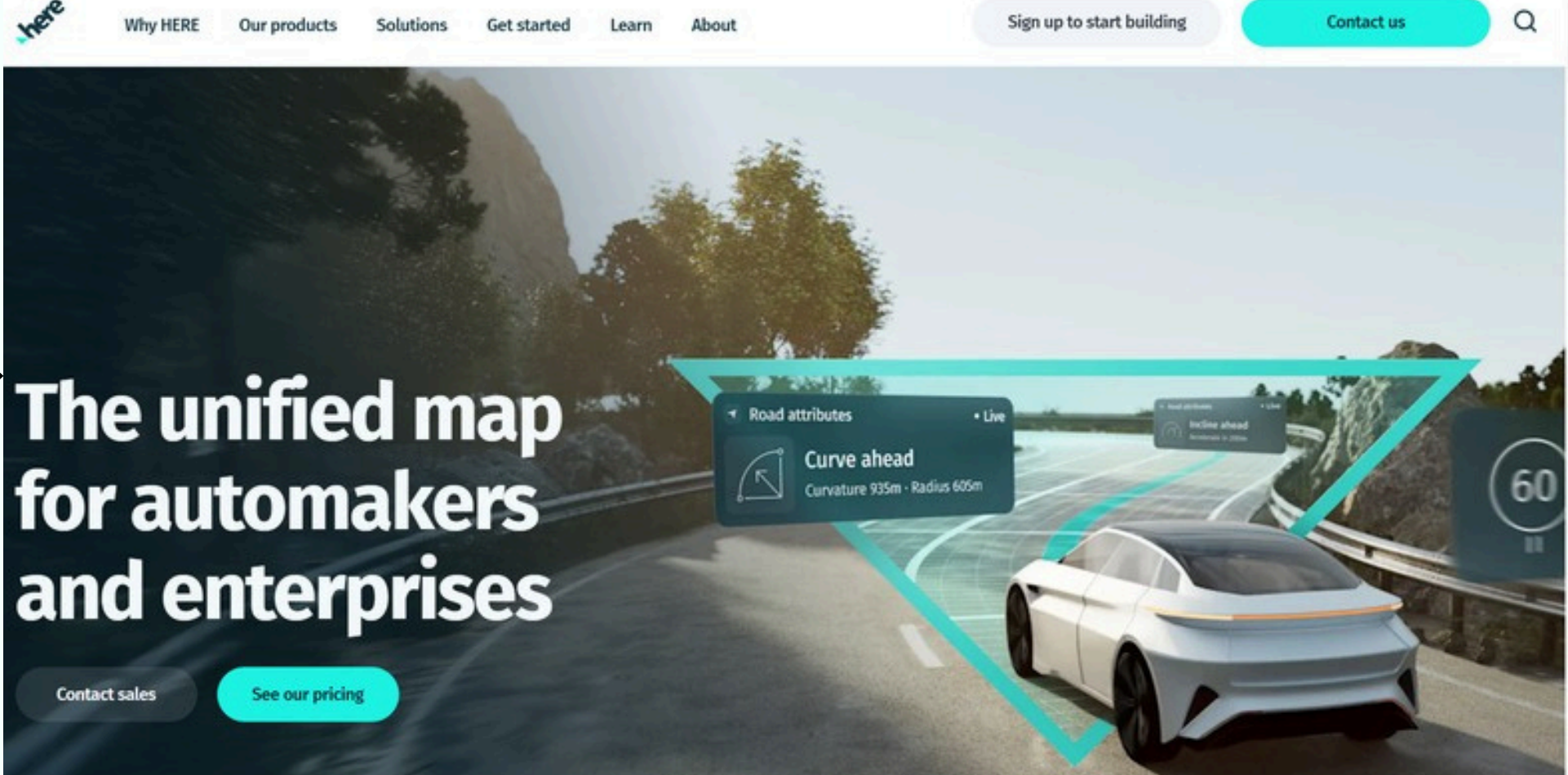
- Thu thập dữ liệu giao thông qua API
- Thu thập dữ liệu thời tiết từ website
- Lưu trữ dữ liệu dạng bảng phục vụ phân tích



WEB CRAWLER

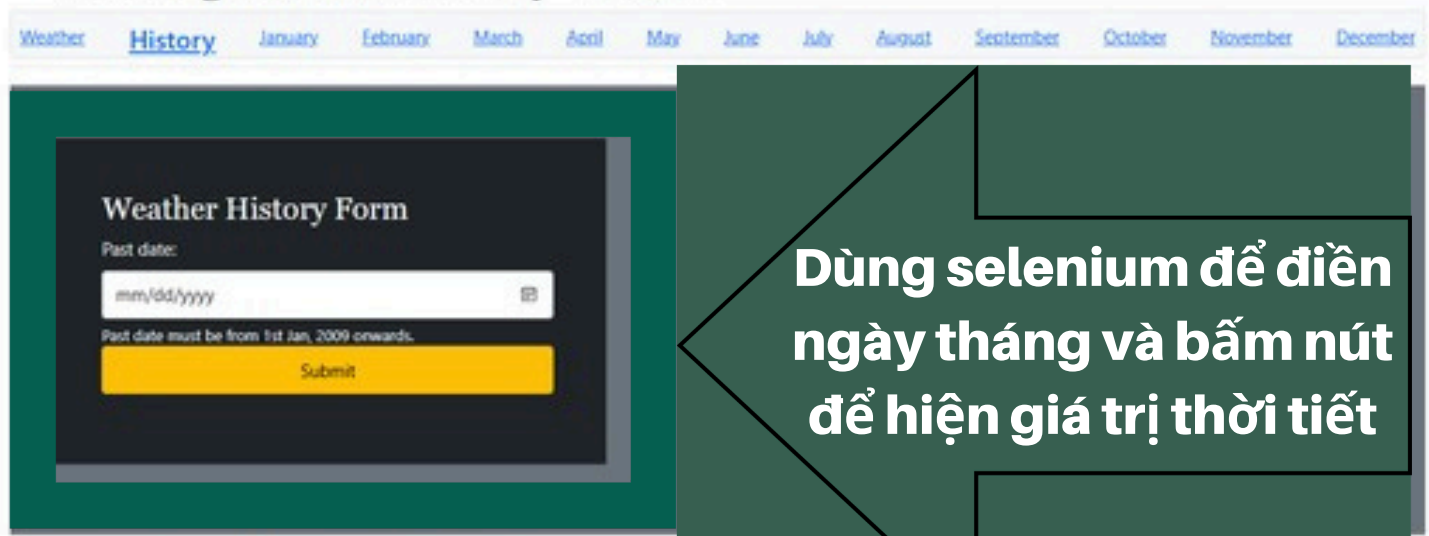


Here cung cấp sẵn api theo thời gian thực dữ liệu giao thông



Giao thông

Thời tiết



Dùng selenium để điền ngày tháng và bấm nút để hiện giá trị thời tiết

Dùng beautiful soup đọc các thẻ để lấy giá trị của bảng

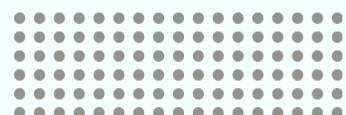
Sunrise: 06:30 AM Sunset: 05:42 PM		Moonrise: 07:57 PM Moonset: 08:16 AM		M	20.7 °c	16.1 °c	17.9 °c	0.18 mm	16.6 kph
	Thu 16 0:00 am	Thu 16 3:00 am	Thu 16 6:00 am	Thu 16 9:00 am	Thu 16 12:00 pm	Thu 16 3:00 pm	Thu 16 6:00 pm	Thu 16 9:00 pm	
Icon									
Temp	16.4°C	16.1°C	16.2°C	18.1°C	20.7°C	19.8°C	18.4°C	18.2°C	
Wind	10.8 kmph	10.8 kmph	9.7 kmph	11.5 kmph	15.5 kmph	16.2 kmph	10.1 kmph	6.8 kmph	
Precip	0.00 mm	0.00 mm	0.00 mm	0.00 mm	0.00 mm	0.02 mm	0.03 mm	0.01 mm	
Cloud	100%	100%	69%	100%	100%	100%	65%	50%	
Humidity	82%	81%	78%	68%	60%	71%	80%	82%	
Pressure	1025 mb	1023 mb	1024 mb	1025 mb	1023 mb	1020 mb	1021 mb	1023 mb	

Kết quả thu thập

Tập dữ liệu: 1 bảng, hơn 1,9 triệu bản ghi (mỗi 10 phút 1 lần), 15 cột

Ý nghĩa mỗi cột:

- **dtime**: Thời điểm ghi nhận giao thông
- **district**: Quận/huyện
- **street**: Tuyến đường
- **speed**: Tốc độ di chuyển thực tế
- **speedUncapped**: Tốc độ thực tế chưa giới hạn
- **freeFlow**: Tốc độ khi đường thông thoáng
- **jamFactor**: Mức độ ùn tắc
- **confidence**: Độ tin cậy dữ liệu
- **traversability**: Khả năng lưu thông của tuyến đường
- **Temperature**: Mức nhiệt đo được tại thời điểm (°C).
- **Rainfall**: Lượng mưa ghi nhận được (mm).
- **Cloud Cover**: Phần trăm hoặc mức độ mây che phủ trên bầu trời.
- **Pressure**: Áp suất khí quyển đo tại thời điểm (hPa).
- **Wind Speed**: Tốc độ gió tại thời điểm (km/h)
- **Description**: Mô tả chi tiết về điều kiện thời tiết (nắng, mưa, nhiều mây, hoặc sương mù).



LÀM SẠCH DỮ LIỆU

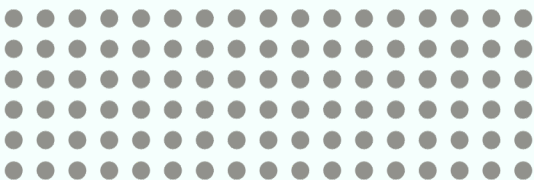
- Loại bỏ dữ liệu không hợp lệ
 - Xóa các cột trống hoàn toàn và bản ghi trùng lặp
 - Loại bỏ các dòng có Street rỗng hoặc giá trị Unknown
- Chất lượng dữ liệu
 - Chỉ giữ các bản ghi có jamFactor và speed hợp lệ (≥ 0)
 - Lấy các dòng có độ tin cậy dữ liệu lớn hơn 70%
- Chuẩn hóa thời gian
 - Chuyển cột thời gian về định dạng datetime
 - Đồng bộ timezone sang giờ Việt Nam



FEATURE ENGINEERING

- Chuẩn hóa thời gian (dtime) để phân tích theo giờ và khung thời gian trong ngày.
- Phân loại điều kiện mưa / không mưa từ dữ liệu thời tiết để đánh giá ảnh hưởng của mưa.
- Tính toán % Speed Reduction, xây dựng các chỉ số đánh giá rủi ro như Avg Speed, Worst Speed (P05), Buffer Time Index, % Time in Jam.

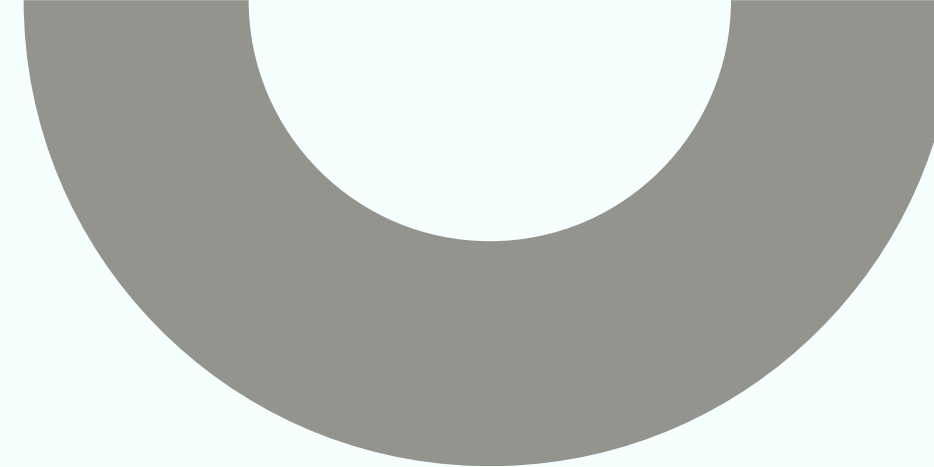
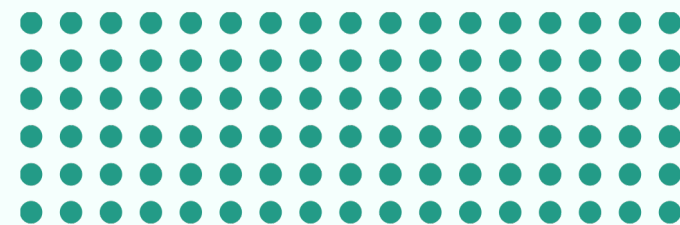
P05 – 5% tốc độ di chuyển chậm nhất	$P05_{speed} = 5\text{th percentile of Speed}$
P95 – 95% của JamFactor	$P95_{jam} = 95\text{th percentile of JamFactor}$
% Time in Jam – tỉ lệ tuyến đường bị tắc	$\%TimeInJam = \frac{\text{số bản ghi có jamFactor} > 5}{\text{Tổng số bản ghi}} \times 100\%$
Buffer Time Index – thời gian cần dự phòng khi di chuyển vào tình huống tệ nhất	$\text{Buffer Time Index} = \frac{T_{95}}{T_{avg}} = \frac{\frac{S}{P05_{speed}}}{\frac{S}{AvgSpeed}} = \frac{AvgSpeed}{P05_{speed}}$
% Speed Reduction - mức suy giảm tốc độ so với khi đường thông thoáng	$\%SpeedReduction = \frac{FreeFlowSpeed - ActualSpeed}{FreeFlowSpeed} \times 100\%$



PHÂN TÍCH DỮ LIỆU



Show dashboard PBI



ĐỊNH HƯỚNG TRONG TƯƠNG LAI

- Mở rộng dữ liệu theo nhiều giai đoạn thời gian để đánh giá xu hướng dài hạn.
- Kết hợp thêm dữ liệu sự cố giao thông và lưu lượng xe nhằm nâng cao độ chính xác phân tích rủi ro.
- Phát triển mô hình dự báo ùn tắc ngắn hạn, hỗ trợ lựa chọn thời điểm và lộ trình di chuyển.
- Tối ưu dashboard để hỗ trợ ra quyết định thời gian thực cho cơ quan quản lý và người dân.

