BANGKOK BUS TRAFFIC ANALYSIS & TRAVEL TIME ESTIMATION

66011610 - Win Thawdar Aung
66011647 - Than Thar Min Htet
66011649 - Thet San Htar
66011609 - Win Lae Mon

Project Overview

Goal

- Analyze Bangkok's traffic patterns to enhance bus service reliability using predictive analytics, focusing on estimating bus travel times and understanding route performance under congestion and recommend the fastest time to travel

Scope

- Combine traffic, bus stop, and route datasets
- Detect congestion patterns and rush-hour trends
- Train ML models to predict speed and waiting time
- Visualize route-level performance interactively

Outcomes

- Travel time estimates
- Route Comparison
- Insights for future route optimization
- Interactive dashboard for route comparison

Problem Statement

Key Problem

- Bangkok faces frequent bus delays due to traffic congestion and inconsistent travel times.
- Passengers experience long and unpredictable waiting times.
- Existing systems lack data-driven forecasting for improving reliability.

Objective

To predict bus travel times using machine learning models trained on traffic and route data, helping planners and passengers make informed decisions.

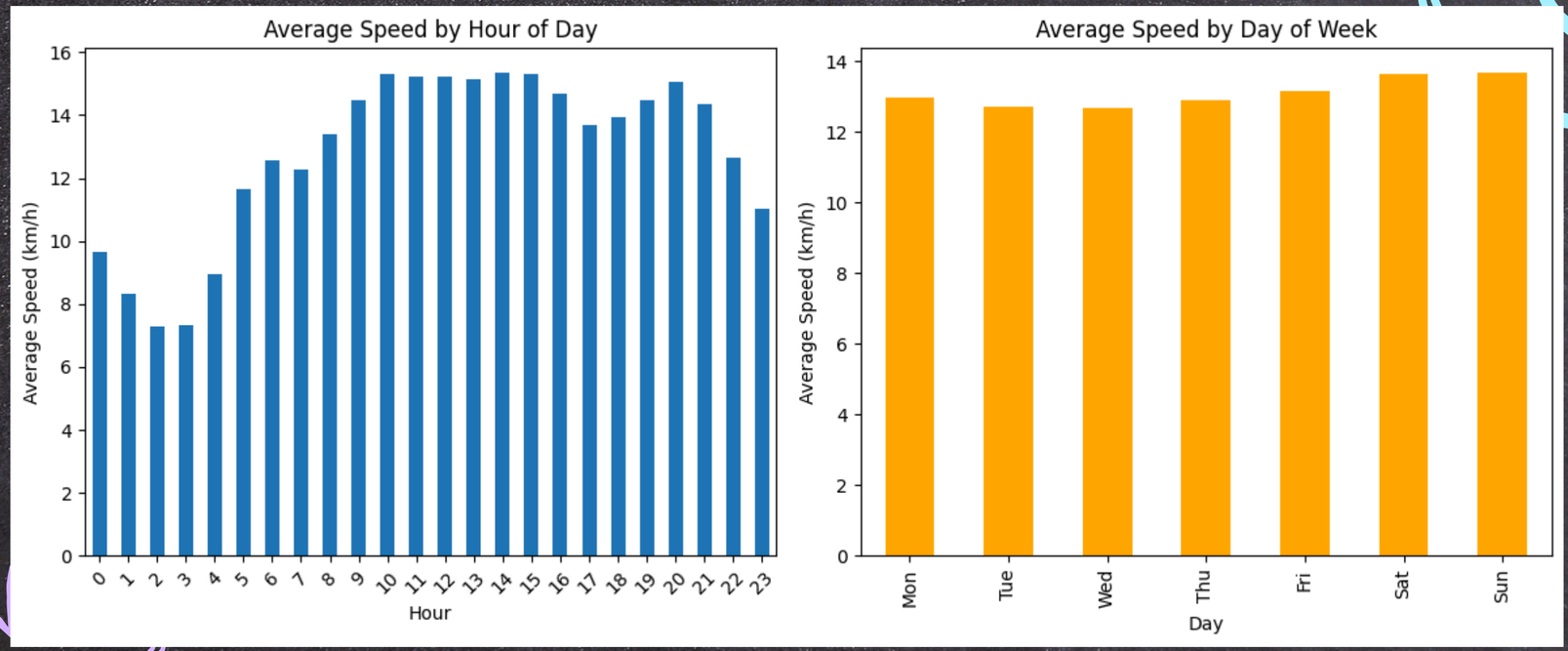
Tools and Technologies

Category	Tools / Libraries
Programming Language	Python
Data Handling & Processing	Pandas, NumPy, JSON, AST, Pathlib, OS
Machine Learning & Modeling	Scikit-learn (RandomForestRegressor,XGBoost,K-Fold, MAE, R^2), Pickle
Congestion Area Analysis	HDBScan, cKDTree
Distance & Spatial Computation	Geopy, BallTree (scikit-learn), Math (radians, cos, sin, asin, sqrt)
Visualization	Matplotlib, Seaborn, Plotly (Express, Graph Objects, Subplots), Folium, Folium Plugins
Dashboard & App Development	Streamlit, streamlit-folium, IPyWidgets, IPython.display
Data Hosting & APIs	Hugging Face Datasets, Requests

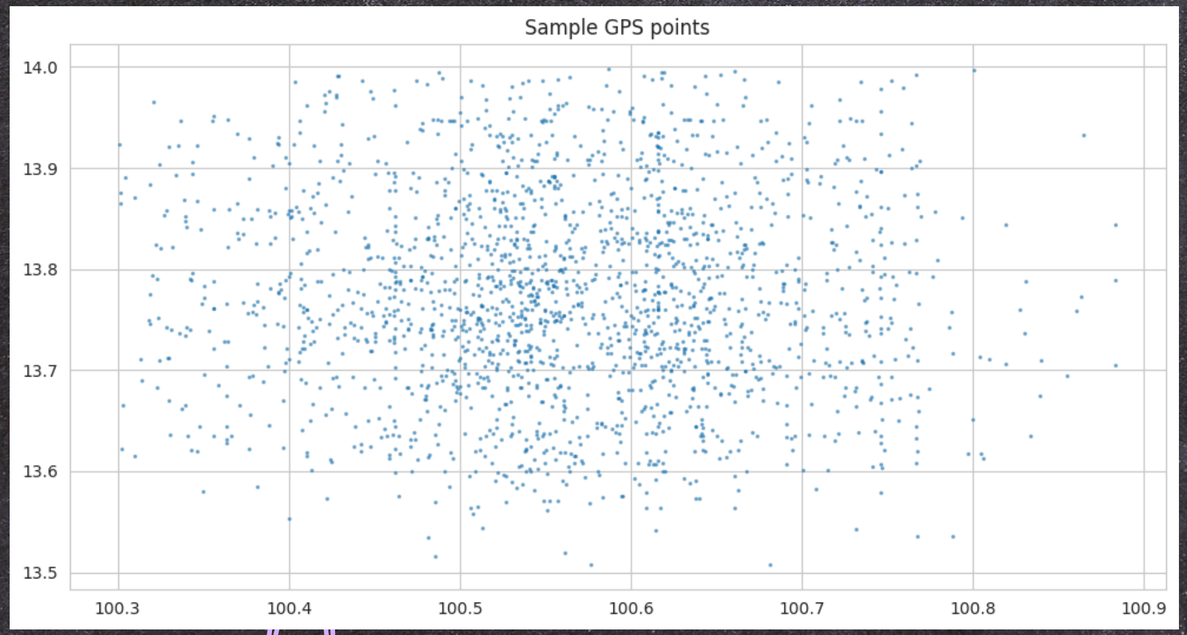
Datasets Used

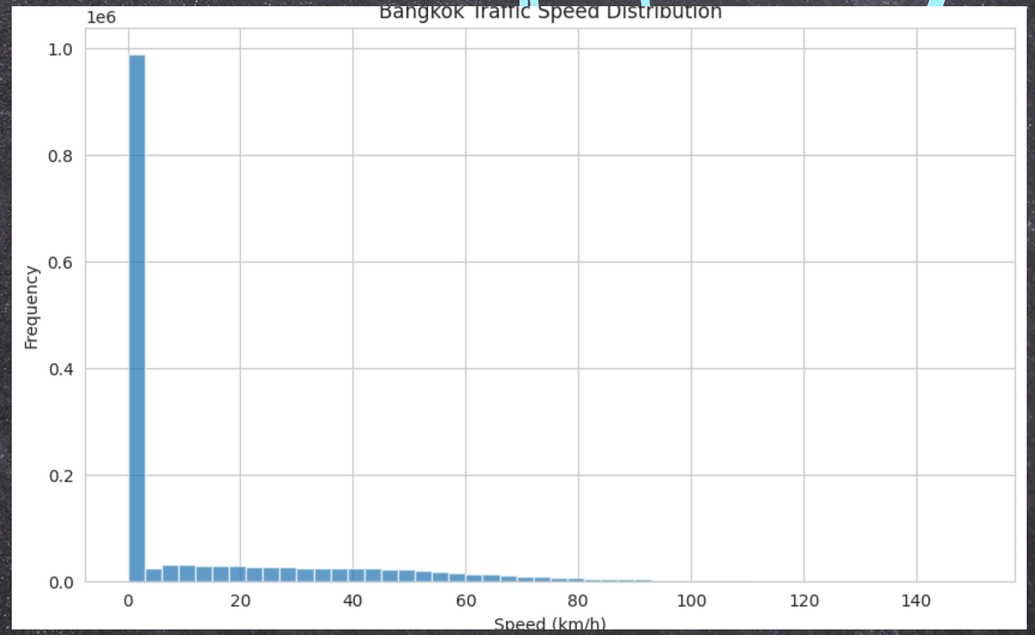
Dataset	Description	Source
traffic.csv	GPS-based traffic speeds, lat/lon, timestamps	https://traffic.longdo.com/opendata
cleaned_bus_routes_file.csv	Bus route coordinates and paths	https://www.openstreetmap.org/#map=6/13.15/101.49
cleaned_bus_stops_file.csv	Bus stop locations with coordinates	

DATA FINDINGS

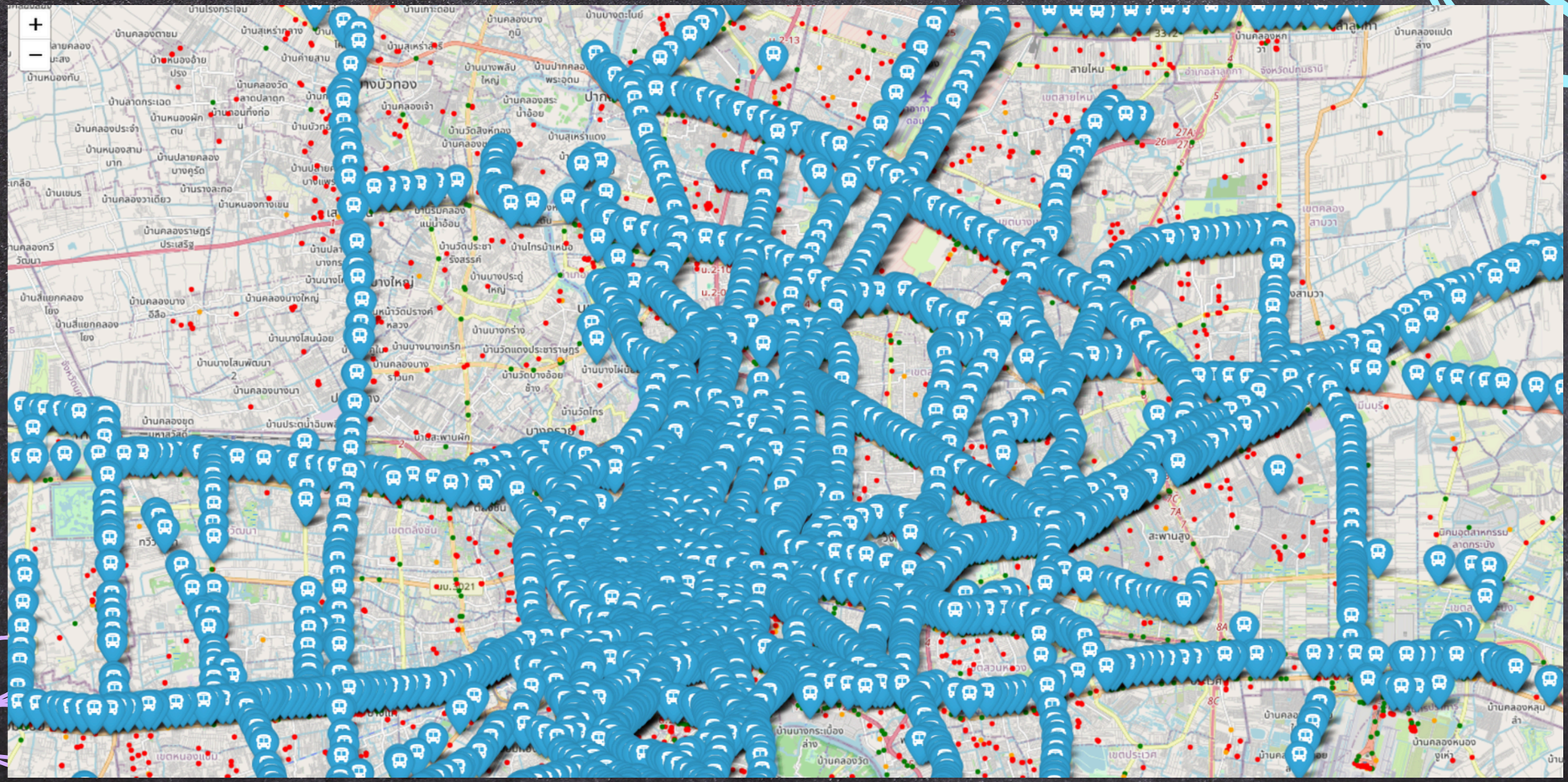


Data findings





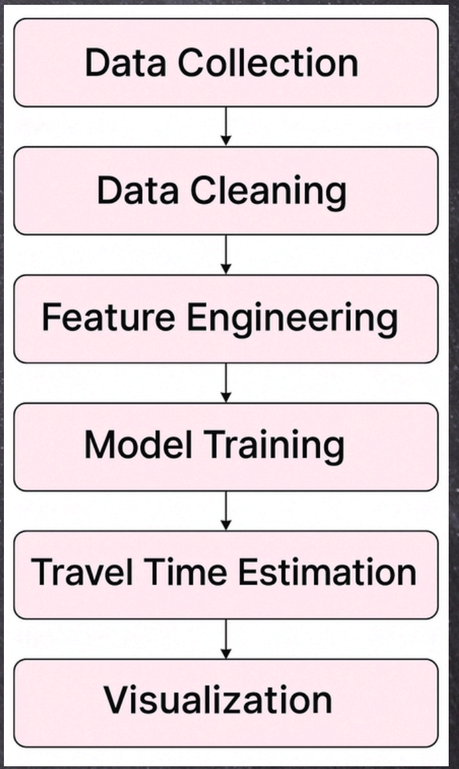
DATA FINDINGS



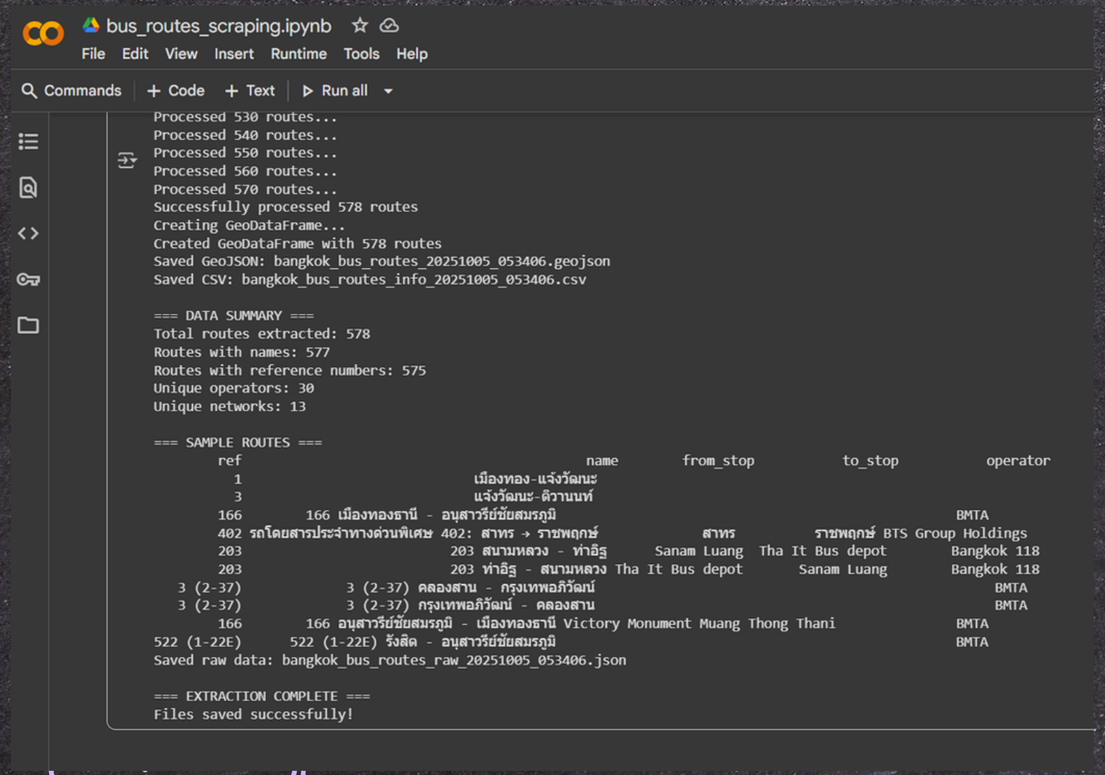
Data Pipeline

End-to-End Pipeline

1. Data Collection - Gather traffic, route, and stop data
2. Data Cleaning & Preprocessing - Handle missing values, remove duplicates, extract hour/day/weekend
3. Feature Engineering - Add time-based, spatial, and congestion-related features
4. Model Training - Train route-level predictive models per route
5. Travel Time Estimation - Predict travel times between stops and recommend the realtime fastest route
6. Visualization - Deploy Streamlit dashboard for interactive exploration



1. Data collection & Data Cleaning



bus_routes_scraping.ipynb

File Edit View Insert Runtime Tools Help

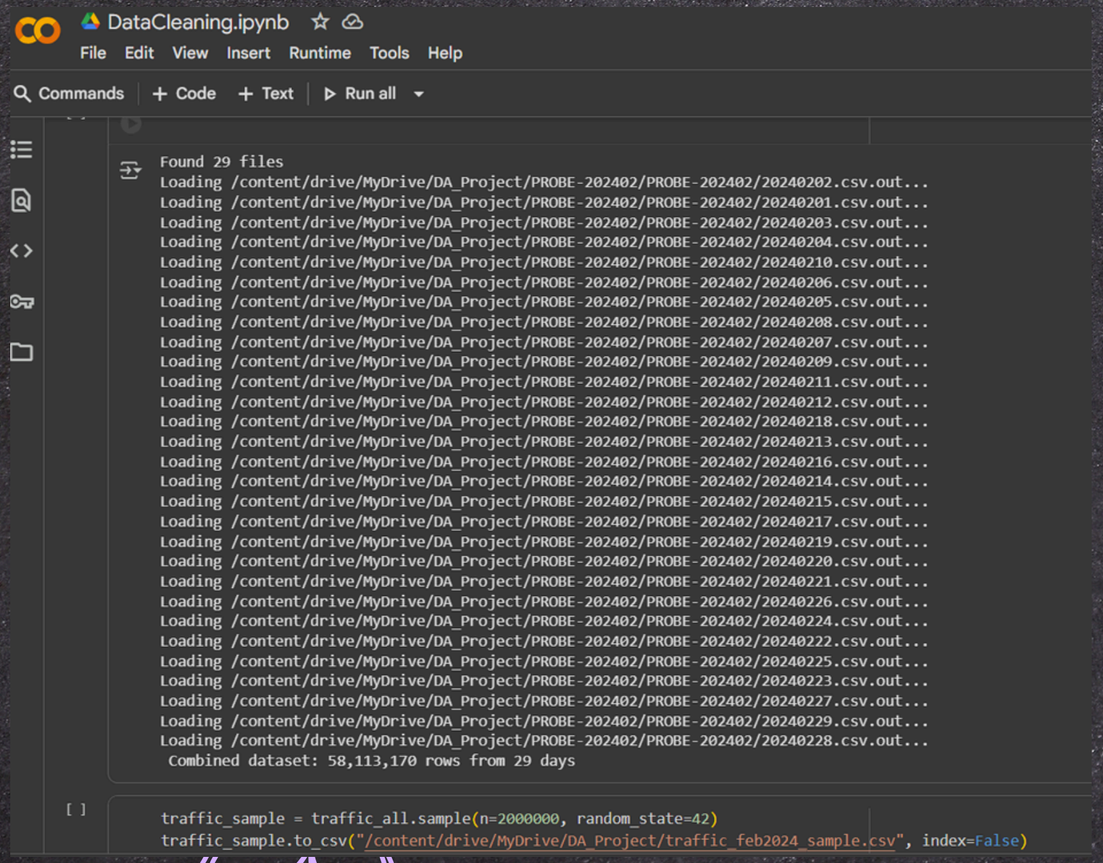
Commands + Code + Text ▷ Run all

```
Processed 530 routes...
Processed 540 routes...
Processed 550 routes...
Processed 560 routes...
Processed 570 routes...
Successfully processed 578 routes
Creating GeoDataFrame...
Created GeoDataFrame with 578 routes
Saved GeoJSON: bangkok_bus_routes_20251005_053406.geojson
Saved CSV: bangkok_bus_routes_info_20251005_053406.csv

=== DATA SUMMARY ===
Total routes extracted: 578
Routes with names: 577
Routes with reference numbers: 575
Unique operators: 30
Unique networks: 13

=== SAMPLE ROUTES ===
       ref                                                name        from_stop          to_stop          operator
         1                                  เมืองทอง-แจ้งวัฒนะ
         3                                  แจ้งวัฒนะ-ดิวานนท์
       166          166 เมืองทองธานี - อนุสาวรีย์ชัยสมรภูมิ                                                BMTA
       402 รถโดยสารประจำทางด่วนพิเศษ 402: สาทร → ราชพฤกษ์             สาทร          ราชพฤกษ์ BTS Group Holdings
       203                              203 สนามหลวง - ท่าอิฐ      Sanam Luang  Tha It Bus depot       Bangkok 118
       203                              203 ท่าอิฐ - สนามหลวง Tha It Bus depot       Sanam Luang       Bangkok 118
 3 (2-37)                3 (2-37) คลองสาน - กรุงเทพอภิวัฒน์                                                BMTA
 3 (2-37)                3 (2-37) กรุงเทพอภิวัฒน์ - คลองสาน                                                BMTA
       166          166 อนุสาวรีย์ชัยสมรภูมิ - เมืองทองธานี Victory Monument Muang Thong Thani            BMTA
522 (1-22E)       522 (1-22E) รังสิต - อนุสาวรีย์ชัยสมรภูมิ                                               BMTA
Saved raw data: bangkok_bus_routes_raw_20251005_053406.json

=== EXTRACTION COMPLETE ===
Files saved successfully!
```

BUS_ROUTES_DATA

- Crawled traffic bus routes data from OpenStreetMap
- Store that data in CSV file with columns route_id,name,ref, from_stop, to_stoproute_type,coordinates
- Remove duplicates and unnecessary columns
- Handle NULL values
  - Replace NULL in from_stop and to_stop with values from route coordinates

1. Data collection & Data Cleaning



TRAFFIC_DATA

- Combined one-month data and randomly chose 2M data to avoid data biasing by using one-day or one-week data
- Removed impossible data(e.g., negative speed)
- Removed unrelated data (e.g., outside Bangkok traffic data)

2. Data Processing & Feature Engineering

BUS_ROUTES_DATA

Spatial Features:

- segment_distance_list
- total_distance_km

Used “Haversine“ Equation to calulate distance between each coordinate

TRAFFIC_DATA

Spatial & Temporal Features:

- hour, day_of_week, is_weekend, is_rush_hour
- lat, lon, distance_from_center
- near_congestion, congestion_severity_encode

Added congestion indicators to improve model accuracy.

Congestion Zone Detection(HDBSCAN)

Purpose - Identify persistent congestion areas to better understand Bangkok's traffic flow and hotspots.

Method

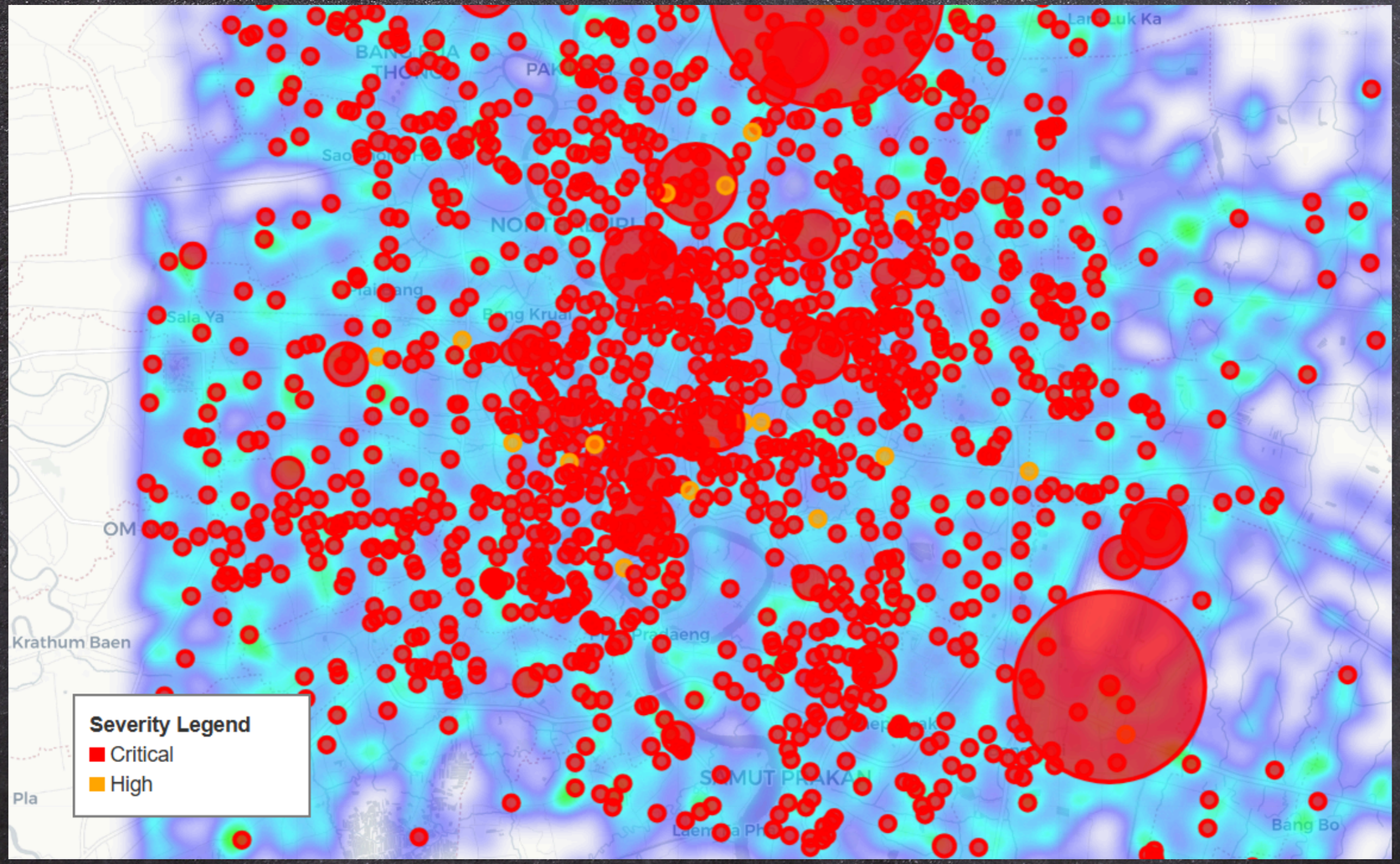
- Filtered slow-speed points (< 15 km/h).
- Applied HDBSCAN clustering to detect dense slow-moving zones.

Generated congestion_zones.csv with: Zone ID, center coordinates, avg speed, cluster size, and severity label

Why HDBSCAN

- No need for predefined cluster count.
- Handles noise and irregular patterns.
- Highlights Critical & High congestion areas.
- Helps in both feature engineering (marking congestion-prone areas) and visualization.

Congestion Zone Detection(HDBSCAN)

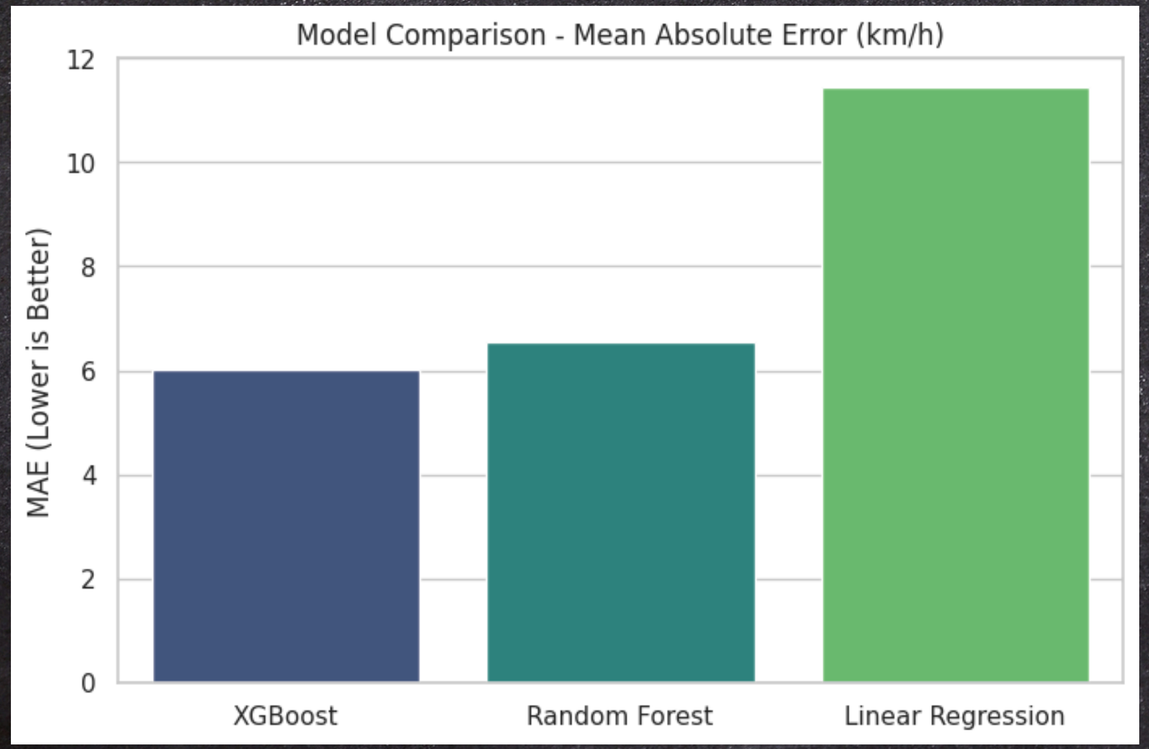


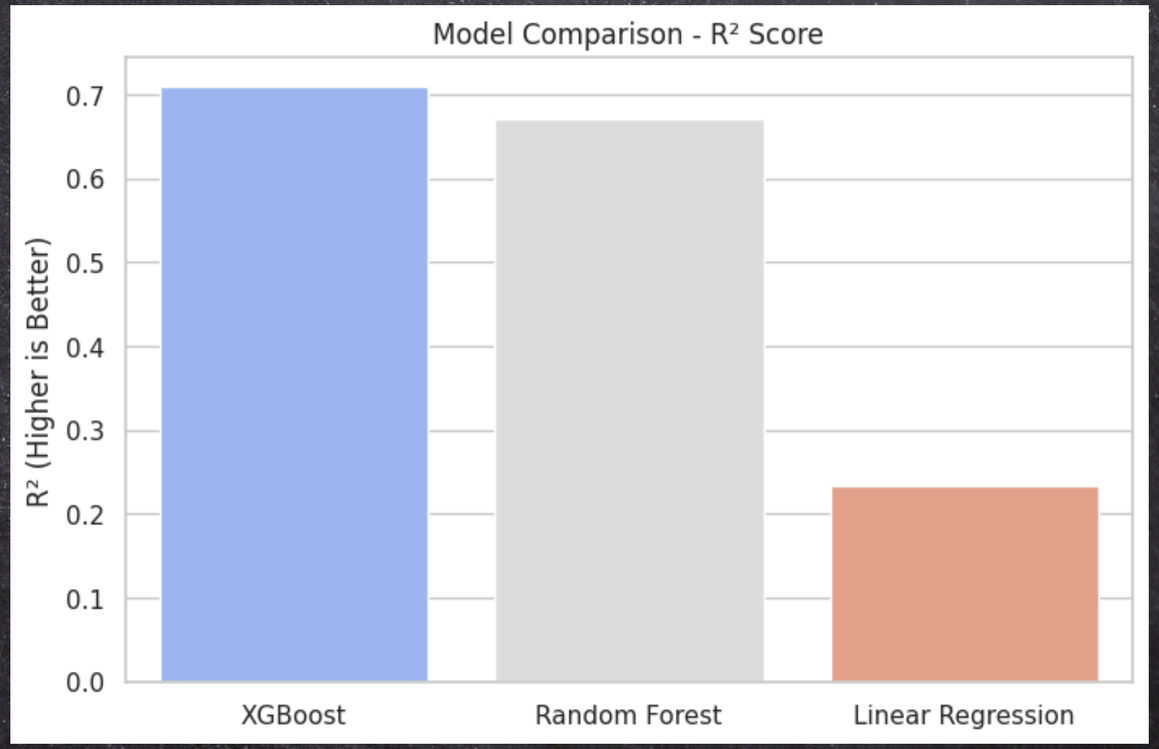
USED TOOLS FOR SMOOTH MAPPING

- Spatial Tree Search
  - cKD Tree Search for better and faster mapping with traffic data and congestion area
  - BallTree to map the stops coordinates and names correctly
- Geodesic distance (Geopy) along with “Haversine” for segment distance calculation.

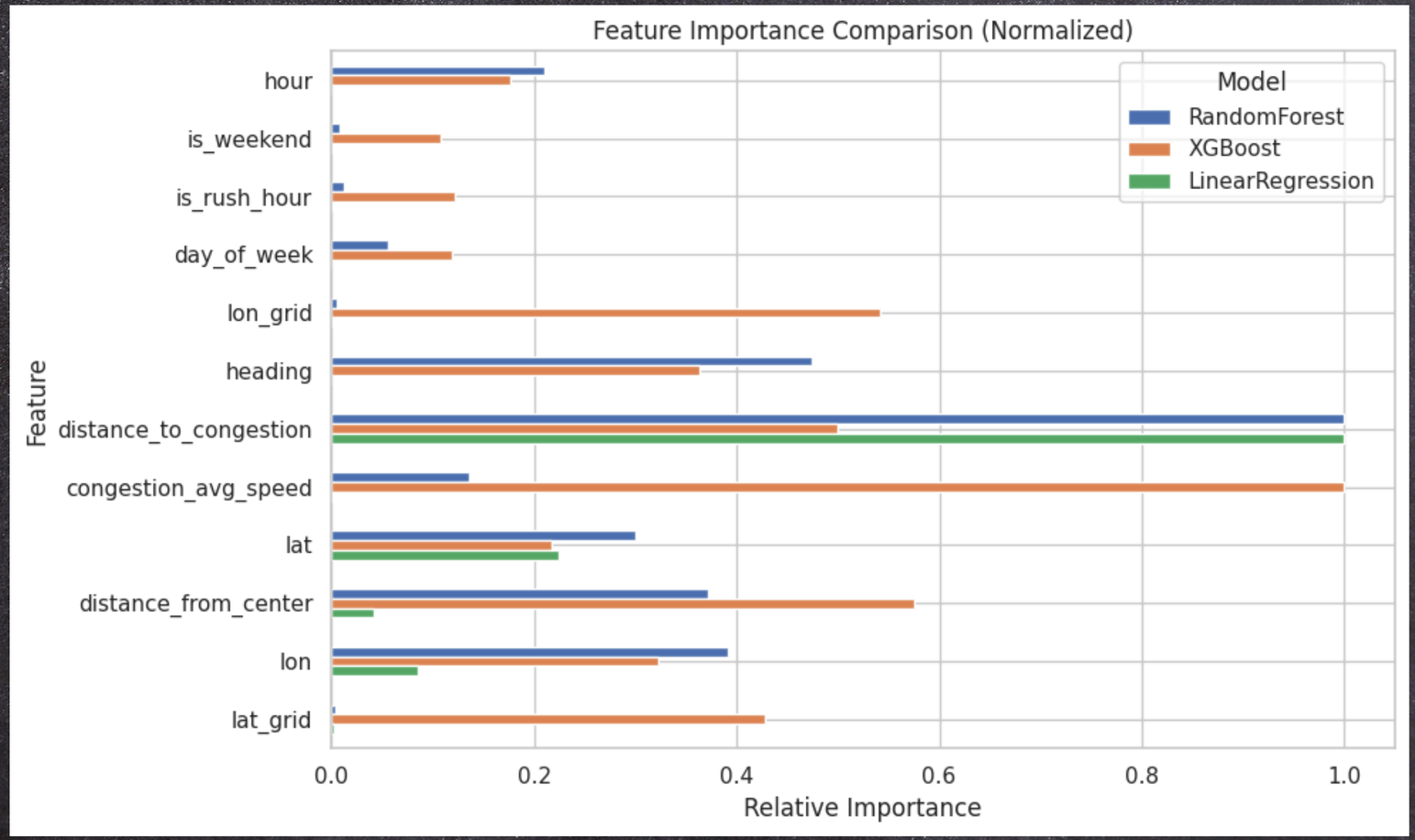
Model Selection

	Model	MAE	RMSE	R^2	CV_MAE (mean)	CV_R^2 (mean)
2	XGBoost	6.011121	10.588736	0.709505	5.981436	0.711815
1	Random Forest	6.550604	11.281783	0.670234	6.526889	0.671460
0	Linear Regression	11.441992	17.207772	0.232817	11.419889	0.232175





Feature Selection



Model Development

Model Used
- XGBoost Regressor (main model per route)

Why XGBoost
- Handle for non-linear traffic data
- High interpretability and consistent accuracy

Training Approach
- 5-Fold Split to compare models performance
- Trained individual models for each bus route (route_models.pkl).
- Features used stored in (feature_columns.pkl).
- Evaluated with Mean Absolute Error (MAE) and R^2 metrics.

Travel Time Estimation

STEPS

1. Compute segment distance using the Haversine formula.

$$\text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1 - \cos(\theta)}{2}$$

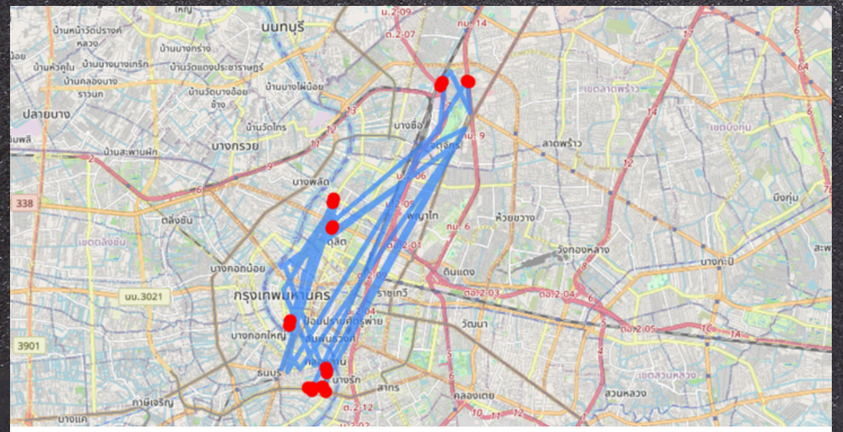
2. Predict bus speed (km/h) for each segment using the trained ML model.
3. Calculate segment travel time:

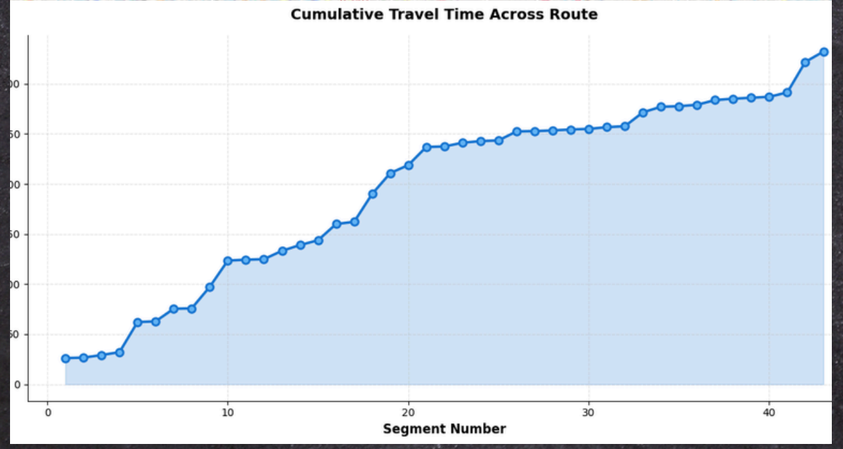
$$t_{segment} = \frac{\text{segment distance}}{\text{predicted speed}} \times 60$$

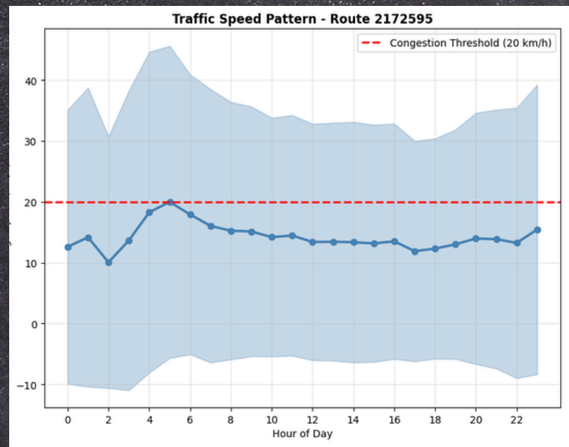
4. Sum all segment times to get total route travel time.
5. Convert total time to minutes and estimate arrival / waiting time.

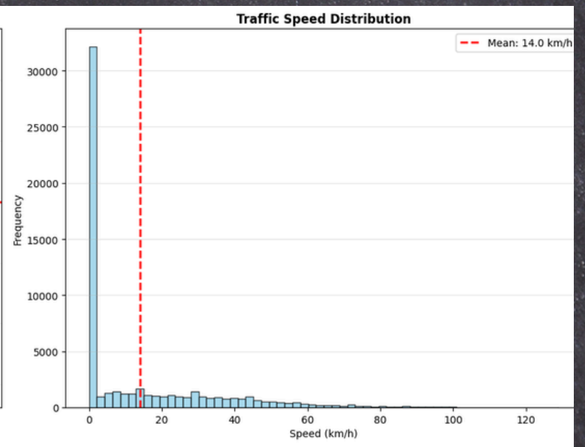
Travel Time Estimation

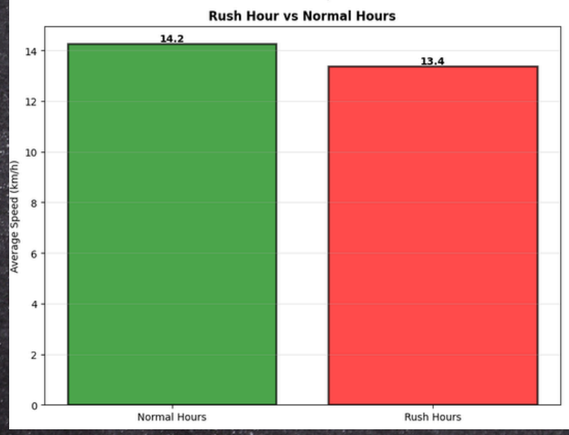
Each Route

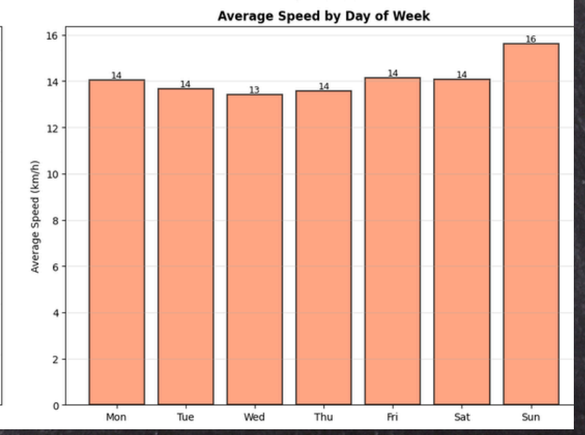












Travel Time Estimation

Each Route

Model Performance Metrics

MEAN ABSOLUTE ERROR	R² SCORE	TRAINING DATA POINTS
7.289 km/h	0.7269 Variance explained	59,420 Traffic records

FEATURES USED

lat , lon , heading , hour , day_of_week , is_weekend , is_rush_hour , lat_grid , lon_grid , distance_from_center , distance_to_congestion

	segment_index	start_stop	end_stop	distance_km	predicted_speed_kmh	segment_travel_time_min	cumulative_travel_time_min	Time/km (min)
0	1	BTS Krungthonburi (Exit 3)	yan phahon yothin	9.639	22.09	26.19	26.19	2.72
1	2	yan phahon yothin	yan phahon yothin	0.150	22.09	0.41	26.59	2.72
2	3	yan phahon yothin	Queen Sirikit Park	0.951	22.09	2.58	29.18	2.72
3	4	Queen Sirikit Park	TMB Headquarter	1.081	22.09	2.94	32.11	2.72
4	5	TMB Headquarter	Krungthonburi Junction	11.059	22.09	30.04	62.16	2.72
5	6	Krungthonburi Junction	Wat Suwan	0.246	22.09	0.67	62.83	2.72
6	7	Wat Suwan	trongkham ban phra athit	4.628	22.09	12.57	75.40	2.72
7	8	trongkham ban phra athit	Phra Athit	0.146	22.09	0.40	75.80	2.72
8	9	Phra Athit	BTS Mochit (Chatuchak Park)	7.920	22.09	21.52	97.31	2.72
9	10	BTS Mochit (Chatuchak Park)	OPP Klongsan Market	9.635	22.09	26.18	123.49	2.72
10	11	OPP Klongsan Market	Somdet Chaopraya Hospital (Latya side)	0.332	22.09	0.90	124.39	2.72
11	12	Somdet Chaopraya Hospital (Latya side)	soi Latya 17	0.190	22.09	0.52	124.91	2.72

Best Route Recommendation

Step

1. Identify nearby routes
   - Check which bus routes have stops or GPS points within 0.5 km of both the origin and destination.
2. Direct route check
   - If both locations are covered by the same route, select it as a direct route.
3. If no direct route found
   - Search for connecting routes (transfer routes).
   - Find potential transfer points where two different routes pass close to each other (within 0.5 km).
4. Combine routes with transfer
   - Merge travel segments from Route A → transfer point → Route B.
5. Calculate total travel time
   - For each possible route or route pair, compute:
   - Total Time=Travel Time (predicted)+Waiting Time

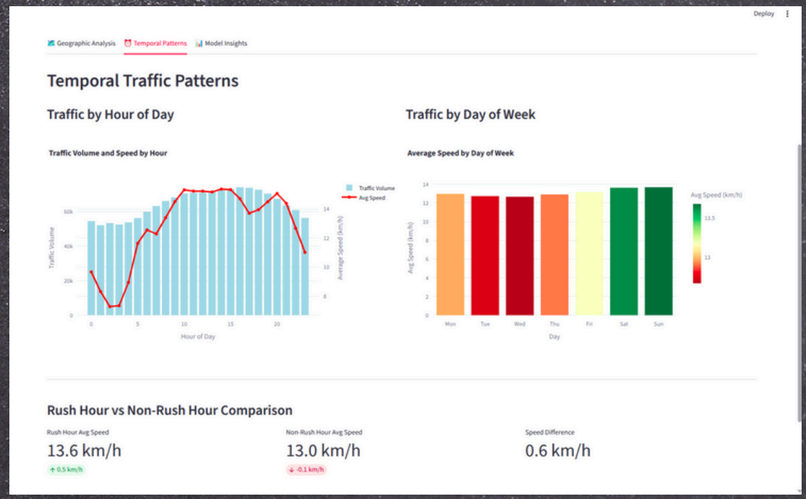
Best Route Recommendation

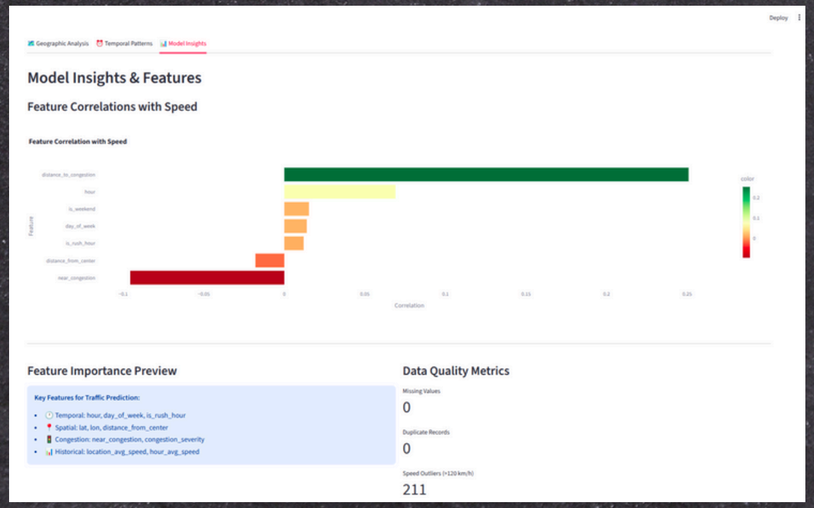
```
Searching for direct routes...
Searching for routes with one transfer...
Found 80 routes near origin
Found 48 journey options:
  - 33 direct routes
  - 15 routes with 1 transfer
```

Top Journey Options

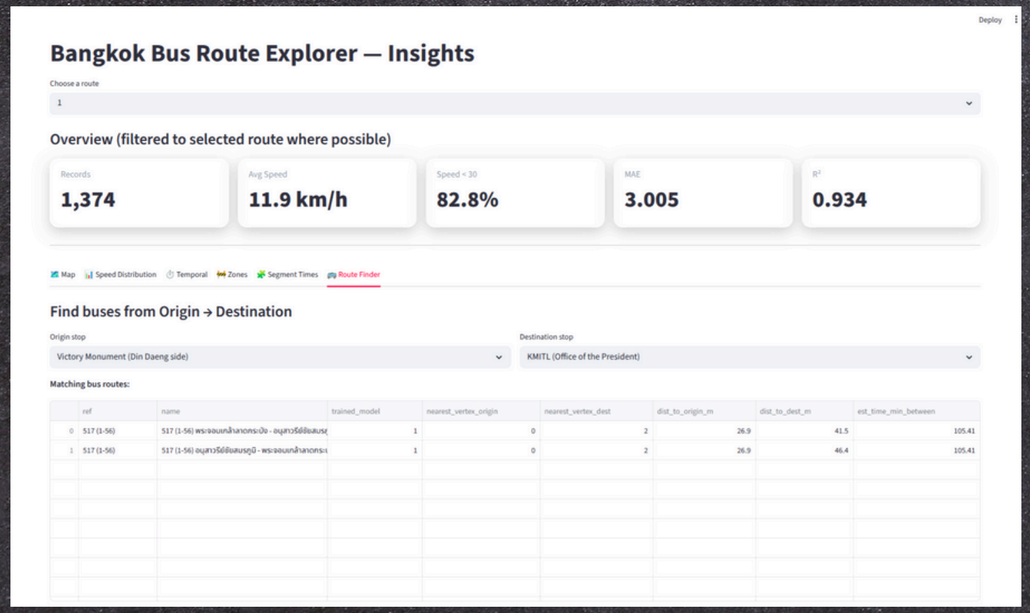
Direct	**Route 17899706** (2-53 วงกลม : อนุสาวรีย์ชัยสมรภูมิ - รัชดาภิเษก (วนซ้าย))	64 min Arrive: 00:11
Direct	**Route 13731714** (A1 หมอชิต 2 - ท่าอากาศยานดอนเมือง)	78 min Arrive: 00:25
Direct	**Route 15533080** (104 (2-16) ปากเกร็ด - หมอชิต 2)	90 min Arrive: 00:37
Direct	**Route 14134302** (136 (3-47) คลองเตย - หมอชิต 2)	146 min Arrive: 01:33
Direct	**Route 15533081** (104 (2-16) หมอชิต 2 - ปากเกร็ด)	150 min Arrive: 01:37

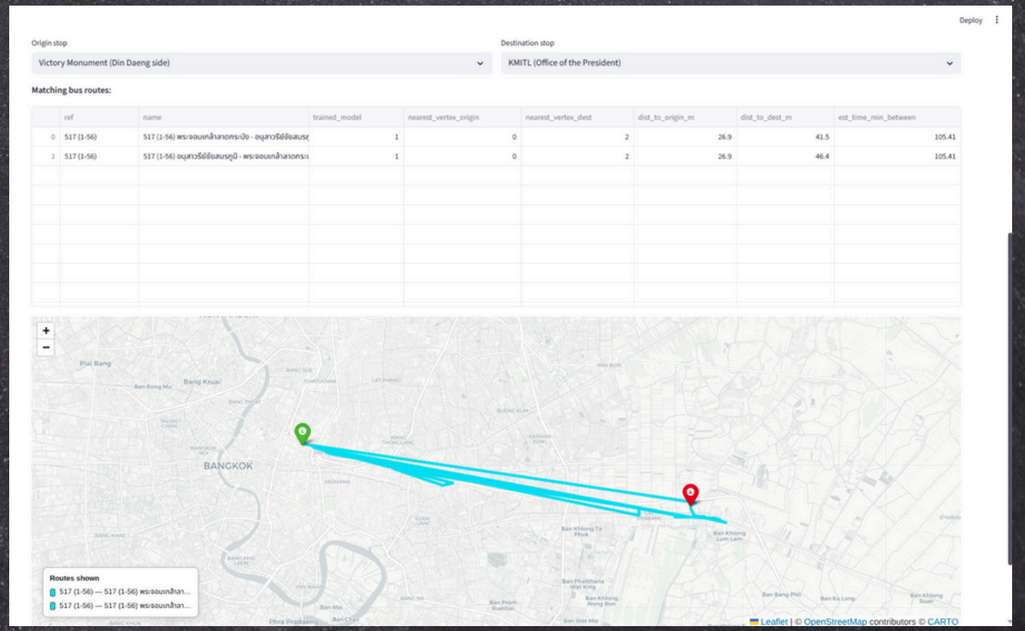
Streamlit Dashboard





Streamlit Dashboard





Key Findings

1

- Clear congestion patterns during rush hours.
- HDBSCAN effectively identified persistent congestion zones.
- Machine learning models gave strong predictive results

2

- Travel times increase significantly under congestion conditions.
- Dashboard enabled clear route and traffic insights
- Findings support smarter route recommentations

Conclusion

- Built a complete predictive pipeline for bus travel-time estimation and bus route recommendation.
- Combined traffic analytics and ML for actionable insights.
- Delivered a user-friendly dashboard for visualization.
- Supports data-driven improvements in Bangkok's public transport.


THANK
YOU