

Introduction

- **Urban traffic congestion** remains a critical challenge for modern cities, leading to increased travel times, emissions, and economic costs
- Traffic conditions are shaped by **complex interactions** between such as traffic demand, network performance, temporal patterns and external factors
- **Traditional approaches** often rely on limited variables or lack interpretability
- **Recent advances** in data-driven methods and Machine Learning (ML) enable:
 - Improved traffic state identification
 - Enhanced prediction accuracy
- However, there is still a **need for interpretable frameworks** that combine:
 - Multiple data sources
 - Robust modeling techniques

Objective & Research Questions

Objective

To identify and predict traffic states on major urban corridors using a multi-stage data-driven framework

Research Questions

- RQ1.** How can traffic states be effectively classified using traffic speed and traffic load data?
- RQ2.** How accurately can ML models predict traffic states?
- RQ3.** What is the role of temporal and weather factors in traffic conditions?

Methodology

Data Collection & Integration

- **Study area:** Four major urban corridors in Athens [2 roads × 2 directions] (Figure 1)

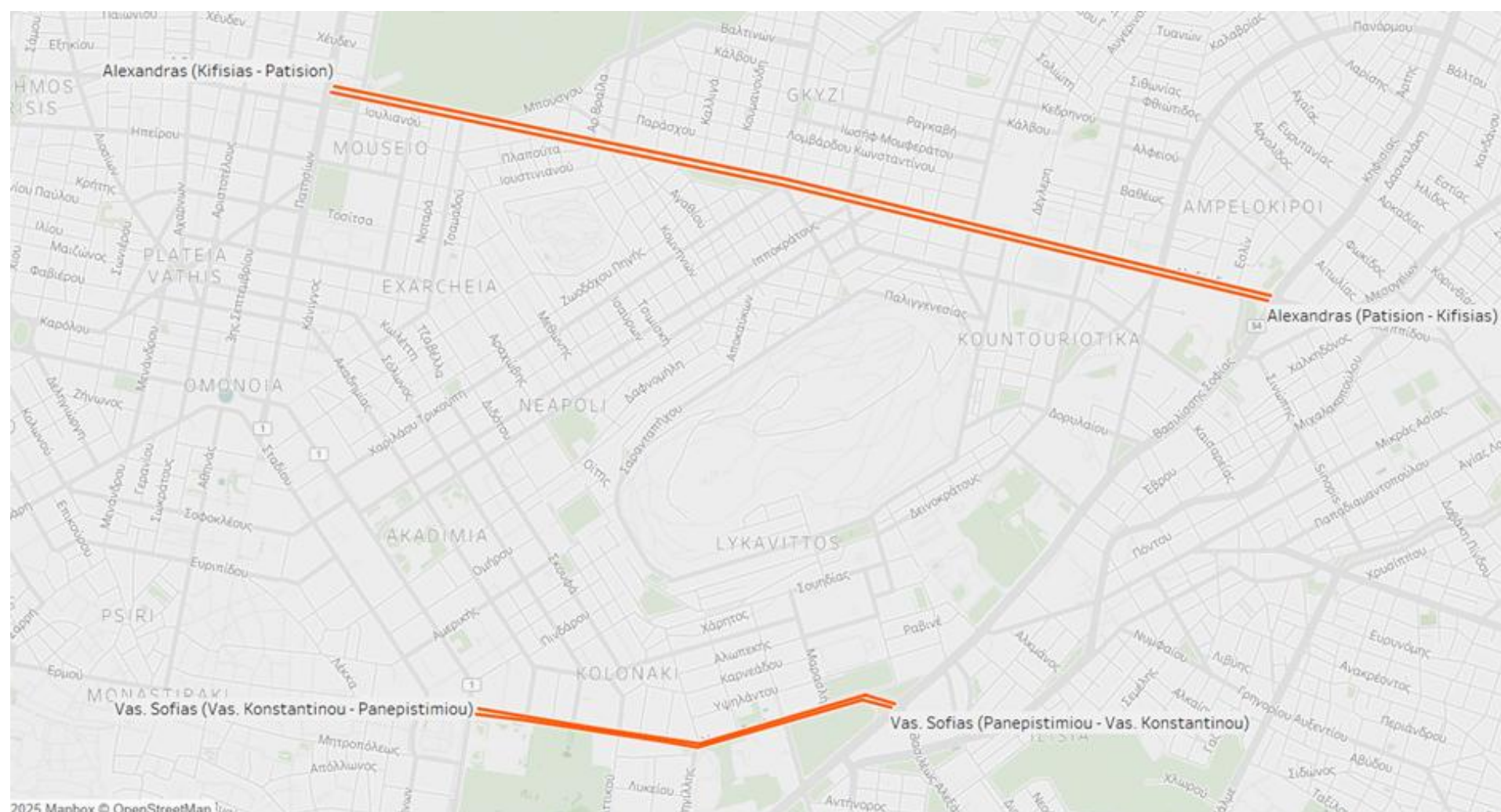


Figure 1: Study Area

- **Multi-source dataset:**
 - Traffic speed & travel time obtained from Google Directions API at high temporal resolution
 - Traffic load (vehicles/hour/lane) collected from loop detectors operated by the Traffic Management Center (TMC)
 - Weather data (temperature, rainfall, snowfall, daylight) retrieved from Open-Meteo
- **Data integration:**
 - All datasets were spatiotemporally matched at the segment-hour level
 - Traffic data were aligned with corresponding weather and temporal variables
- **Dataset characteristics:**
 - Time Period: January – July 2022
 - Resolution: Hourly observations
 - Focus on peak periods during weekdays and weekends

The final dataset included the variables summarized in Table 1.

Table 1: Description of variables

Variable	Description	Data Source
Traffic_Speed	Traffic speed per hour (km/h)	Google Directions API
Traffic_Load	Vehicles/hour/lane	TMC
Weekday	0=weekend; 1=weekday	n/a
Month	1(=January) – 12(=December)	n/a
Day	1(=Monday) – 7(=Sunday)	n/a
Hour	0-23	n/a
is_day	0=night; 1=day	
temperature_2m	Average hourly temperature (°C)	Open-Meteo
rain	Precipitation height (mm)	
snowfall	Snowfall depth (mm)	

Methodology (Contd.)

Analytical Framework

Step 1 – Traffic State Identification

- Applied K-means clustering on:
 - Traffic speed
 - Traffic load
- Optimal number of clusters determined using Silhouette index

Step 2 – Traffic State Prediction

- **Model:** XGBoost (Gradient Boosting)
- **Input variables:** traffic variables, temporal features (hour, weekday) and weather conditions
- **Dataset split:** 80% training / 20% testing
- **Model optimization:** 4-fold cross-validation
- Class imbalance handled using SMOTE

Step 3 – Model Interpretation

- Applied SHAP (Shapley Additive Explanations)
- Enables quantification of feature importance and understanding of model decision mechanisms

Results

Traffic State Identification

- K-means clustering identified **two distinct traffic states** across all road segments:
 - Congested: lower speeds, higher traffic load
 - Less-congested: higher speeds, lower traffic load
- The Silhouette index (0.44–0.65) indicates **acceptable clustering quality** (Figure 2 and Table 2)

Table 2: Descriptive statistics of the identified traffic clusters

Cluster	Cluster Label	Traffic Speed	Traffic Load per lane	Cluster Size	Av. Silhouette Width
Alexandras (from Kifissias to Patision)					
1	less-congested	27.095	361.813	377	0.52
2	congested	15.527	453.579	981	0.65
Alexandras (from Patision to Kifissias)					
1	less-congested	29.576	334.469	364	0.49
2	congested	17.343	406.542	991	0.62
Vas. Sofias (from Panepistimiou to Vas. Konstantinou)					
1	congested	16.091	524.781	821	0.51
2	less-congested	22.339	427.675	534	0.44
Vas. Sofias (from Vas. Konstantinou to Panepistimiou)					
1	less-congested	20.448	216.923	453	0.46
2	congested	12.659	284.094	900	0.44

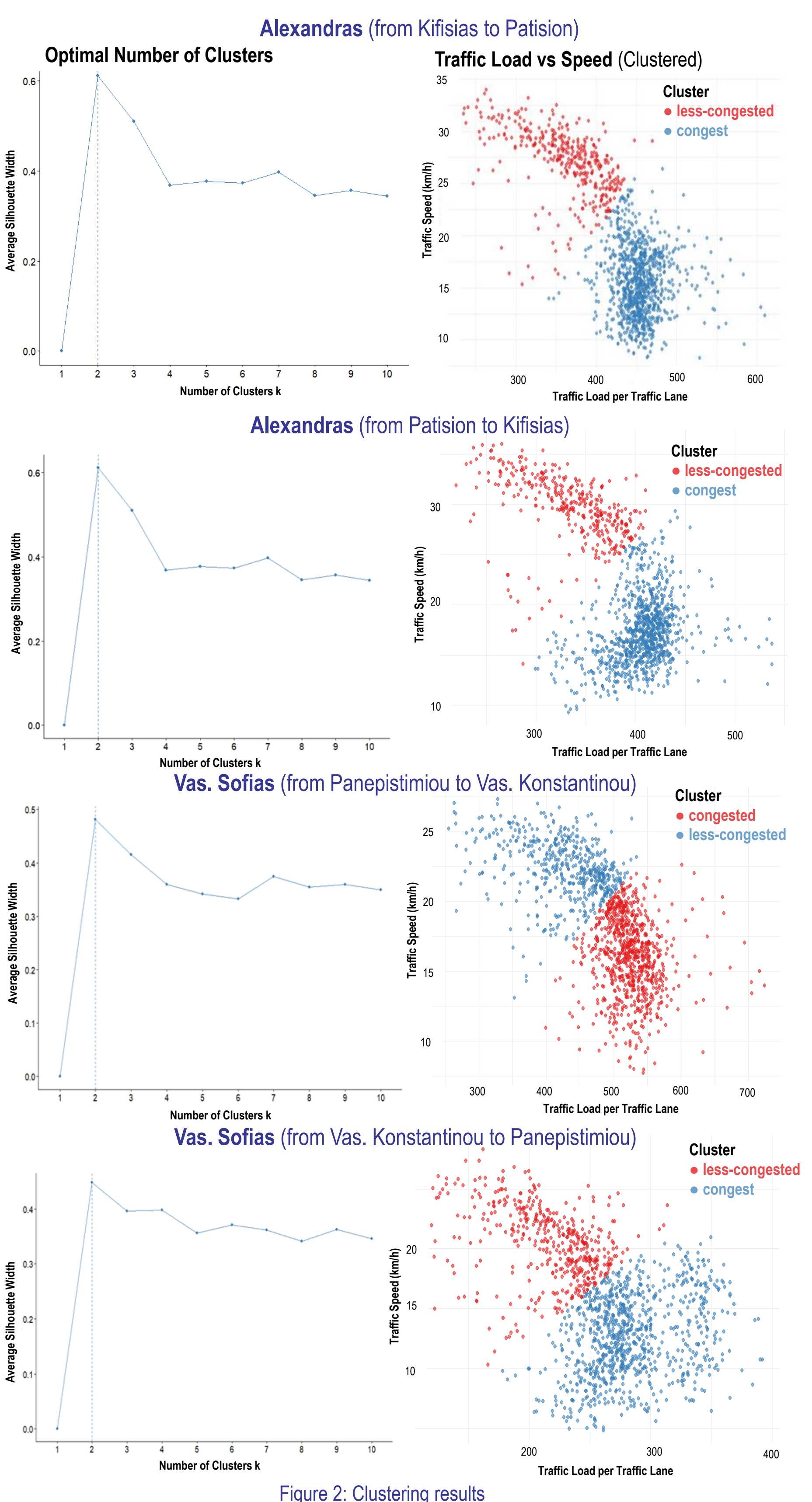


Figure 2: Clustering results

Results (Contd.)

Traffic State Prediction

- XGBoost models achieved **strong predictive performance**
- **Best performance observed on Alexandras Avenue**, while Vasilissis Sofias showed slightly lower but still robust results

Table 3: XGBoost models performance

Metric	Alexandras (from Kifissias to Patision)	Alexandras (from Patision to Kifissias)	Vas. Sofias (from Panepistimiou to Vas. Konstantinou)	Vas. Sofias (from Vas. Konstantinou to Panepistimiou)
Accuracy (Test)	0.923	0.912	0.875	0.812
F1-score (Class 1)	0.949	0.940	0.835	0.867
Balanced Accuracy	0.893	0.879	0.876	0.776
AUC	0.959	0.934	0.947	0.867
Kappa	0.791	0.771	0.735	0.550

Key Drivers of Traffic States (SHAP Analysis)

- **Temporal variables dominate predictions:**
 - Weekday vs. weekend (strongest effect)
 - Hour of day (captures peak demand patterns)
- **Temperature** also significantly influences congestion likelihood
- Rain and snowfall as well as the month play a **secondary role**, with moderate influence on congestion

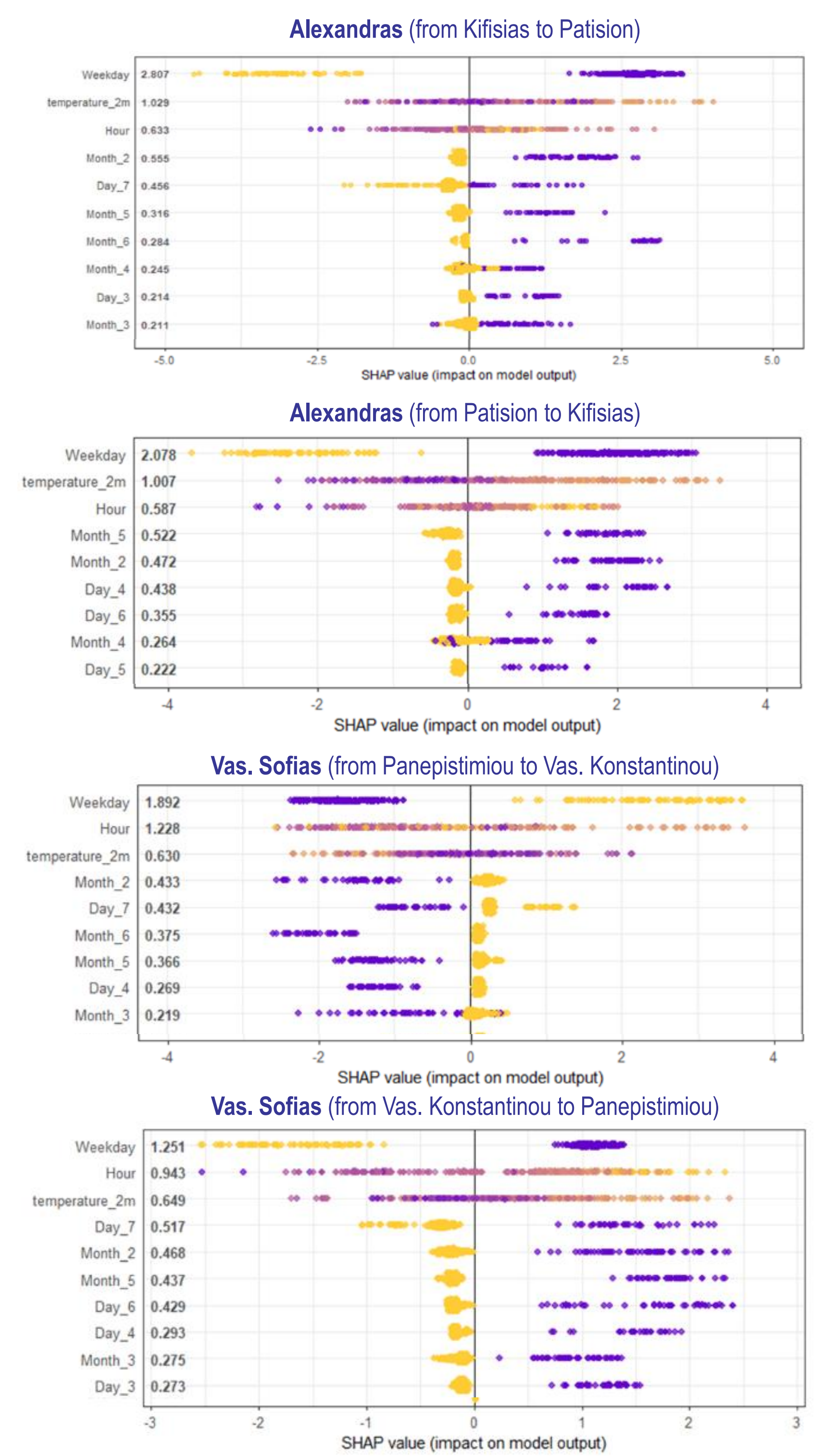


Figure 5: Summary plots of SHAP values

Discussion & Conclusions

- **Two traffic states** (congested / less-congested) effectively describe urban traffic conditions
- **Temporal factors are the main drivers**, while weather plays a secondary role
- The proposed framework achieves **high predictive accuracy (81–92%)**, supporting real-time applications
- The developed approach offers a **scalable tool for traffic management and planning**
- The analysis focuses on peak periods and specific road corridors, **limiting generalizability**
- **Future research** should extend the framework to:
 - Additional road types and networks
 - Off-peak conditions
 - Inclusion of further explanatory variables (e.g., road geometry)

Contact Information

Virginia Petraki

Email: vpetraki@mail.ntua.gr

Web: www.nrso.ntua.gr/vpetraki