Estimating Cycling Mode Share in European Cities

Carol Sobral, Mark Nieuwenhuijsen, Lynn H. Kaack

SEPTEMBER 2025

Mode Share

Definition

Percentage of trips made by each transport mode in a city/region.

Measurement

Travel surveys: households record all [commuting] trips for several days.

Motivation

Cities commonly rely on mode share to assess cycling, but surveys are costly, infrequent and [usually] inconsistent across contexts.

Motivation

Cities commonly rely on mode share to assess cycling, but surveys are costly, infrequent and [usually] inconsistent across contexts.

As a result...

Many European cities lack mode share values, and where they exist, methodological differences make estimates difficult to compare across cities and over time.

Large gaps in the understanding of cycling levels.

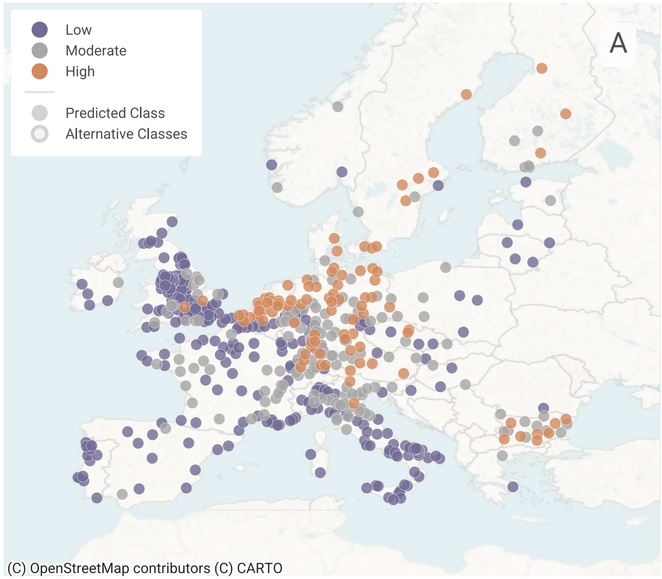
Our Idea

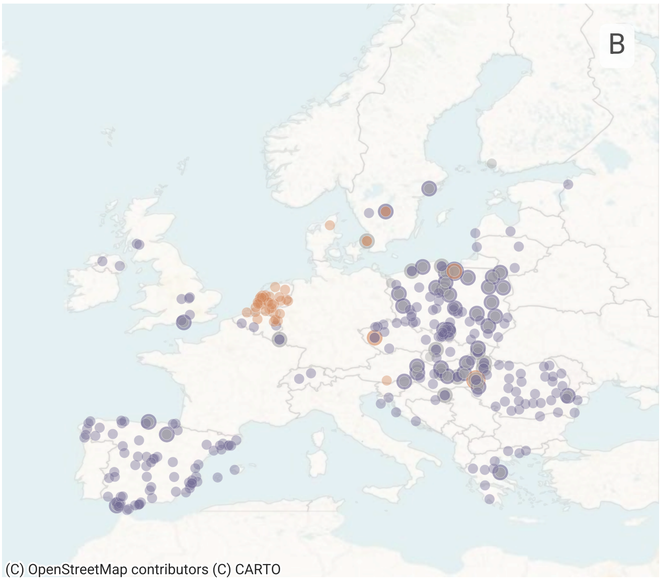
We use ML + open data on city characteristics to estimate cycling adoption for European cities.

Our Idea

We use ML + open data on city characteristics to estimate cycling adoption for European cities.

Cycling levels mapped for 864 cities.





Our Approach

We use cities with available mode share data to learn the relationship between city characteristics and cycling levels, then apply the trained model to predict outcomes for cities without data.

① **Train ML Model**

② **Predict Out-of-Sample**

Our Approach

We use cities with available mode share data to learn the relationship between city characteristics and cycling levels, then apply the trained model to predict outcomes for cities without data.

We trained 7 algorithms and selected the best-performing (XGBoost).



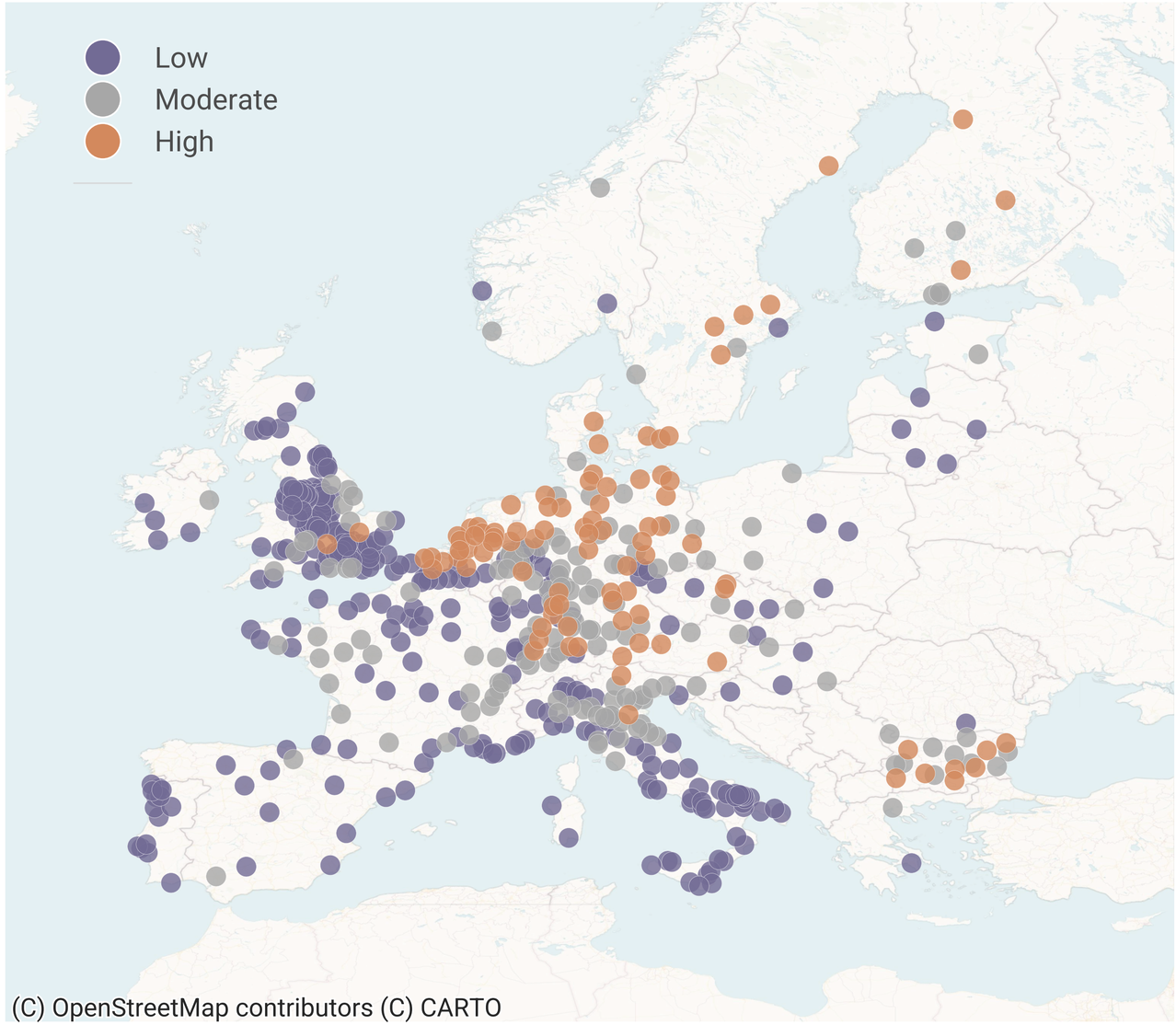
① **Train ML Model**

② **Predict Out-of-Sample**

①

Train ML Model

Mode Share



Cycling mode share data from surveys in **564** European cities, drawn from *Eurostat's Urban Audit* and *The ABC of Mobility (Prieto-Curiel & Ospina)*.

①

Train ML Model

City Characteristics

Street network density, connectivity, cycling infrastructure, rental stations, parking, public transit, bus stops, subway stations, road attributes, speed limits, number of lanes, traffic lights, stop signs, cultural facilities, recreational facilities, libraries, museums, swimming pools, urban space coverage, green space coverage, city area.

Elevation, terrain variability, average temperature (coldest month), average temperature (warmest month), precipitation, snowfall, wind speed, air quality.

Income levels, poverty levels, *employment* by sector, educational attainment, school enrollment, household composition, living arrangements, health outcomes, mortality rates.

Population size, population density, age distribution, gender composition, migration status, nationality, population growth, dependency ratios.

Features selected according to relevance and availability for all cities in the sample -- built and natural environment, socioeconomic, and demographic.

①

Train ML Model

Classification

Classify cities according to *low*, *moderate*, and *high* cycling levels.

0-5% *Low*	*London, Brussels, Milan*
5-15% *Moderate*	*Berlin, Helsinki, Bristol*
> 15% *High*	*Copenhagen, Amsterdam, Basel*

①

Train ML Model

Classification

Mode share is continuous, but inconsistencies across surveys make classification more robust and policy-relevant.

Classify cities into *low*, *moderate*, and *high* cycling levels.

0-5% *Low*	*London, Brussels, Milan*
5-15% *Moderate*	*Berlin, Helsinki, Bristol*
> 15% *High*	*Copenhagen, Amsterdam, Basel*

82.3%

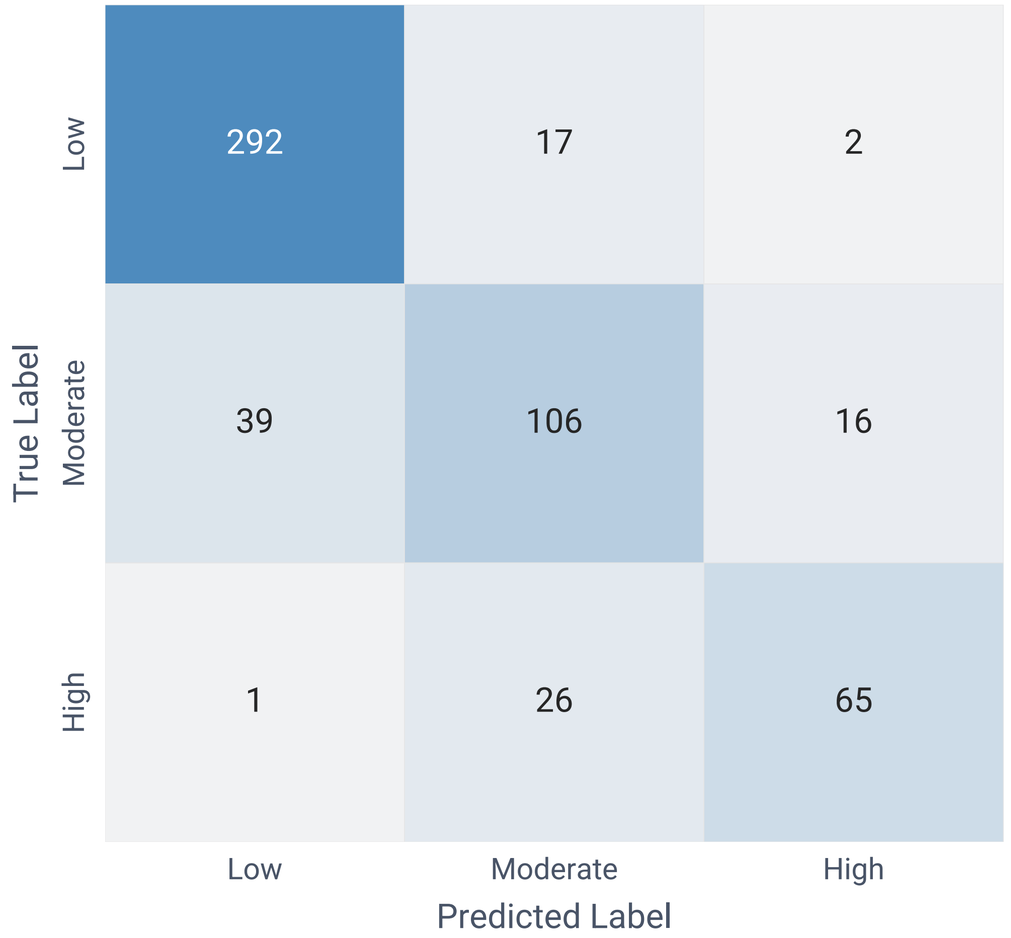
Overall accuracy of the XGBoost model. 8.6 percentage point improvement over linear model.

Macro F1 = 0.788, showing balanced performance across classes.

82.3%

Overall accuracy of the XGBoost model. 8.6 percentage point improvement over linear model.

Macro F1 = 0.788, showing balanced performance across classes.



82.3%

Overall accuracy of the XGBoost model. 8.6 percentage point improvement over linear model.

Macro F1 = 0.788, showing balanced performance across classes.



Our Approach

①

Train ML Model

②

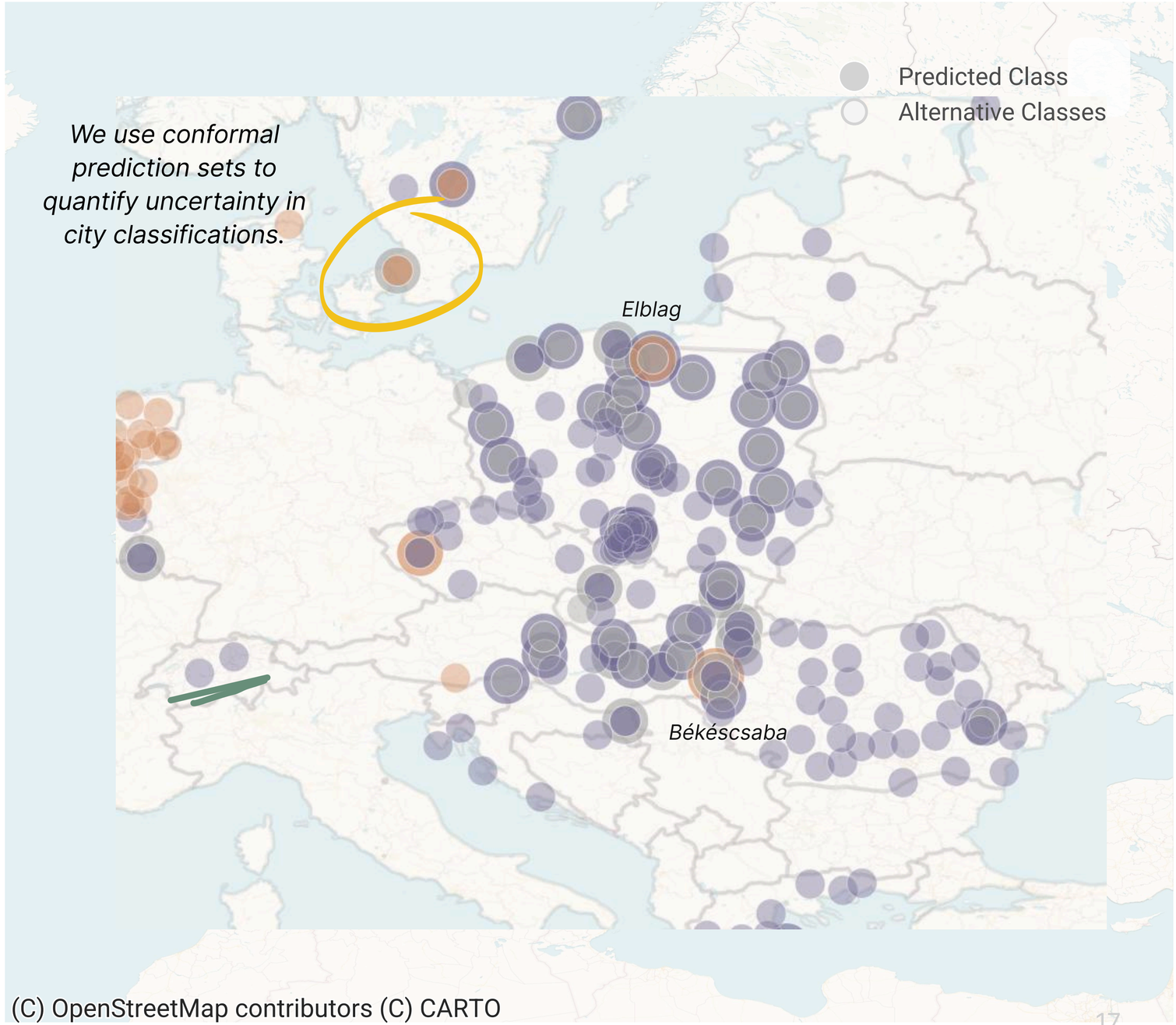
Predict

Apply the trained model to **300** cities without survey data to generate estimates.

②

Predict

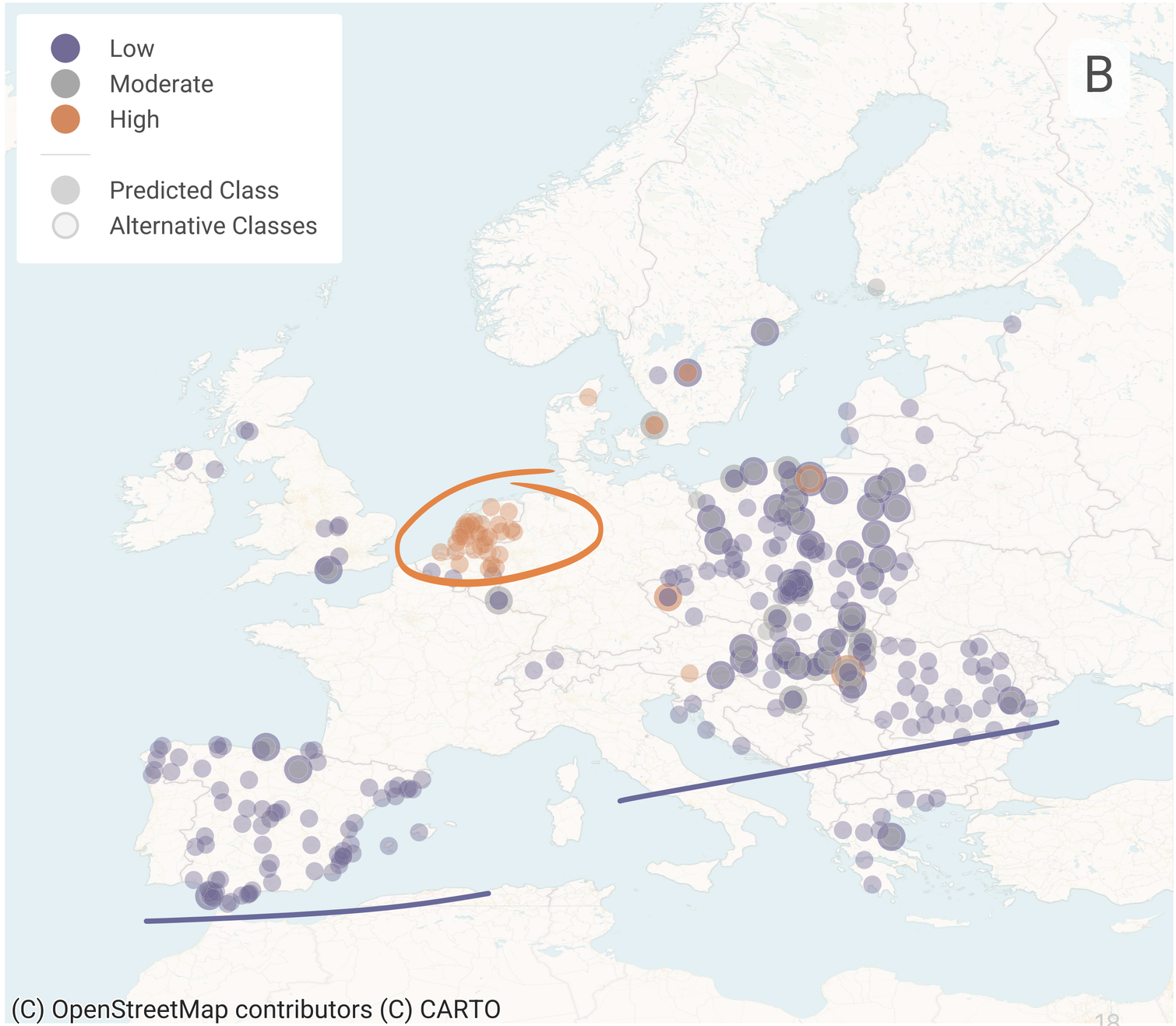
The model is mostly decisive, with 82% of cities assigned a single label. Only *two cases* are highly uncertain, receiving all three possible labels.

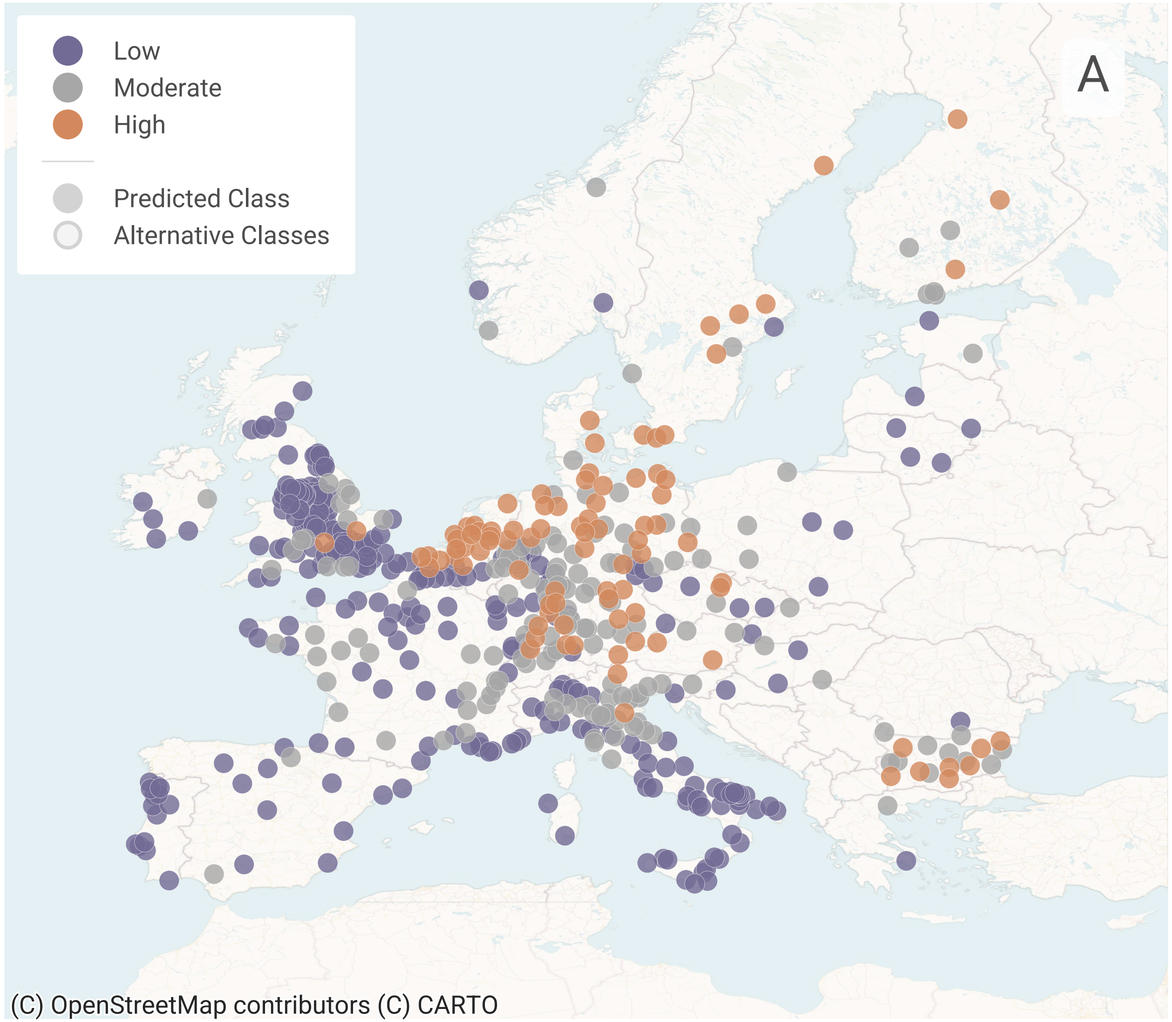


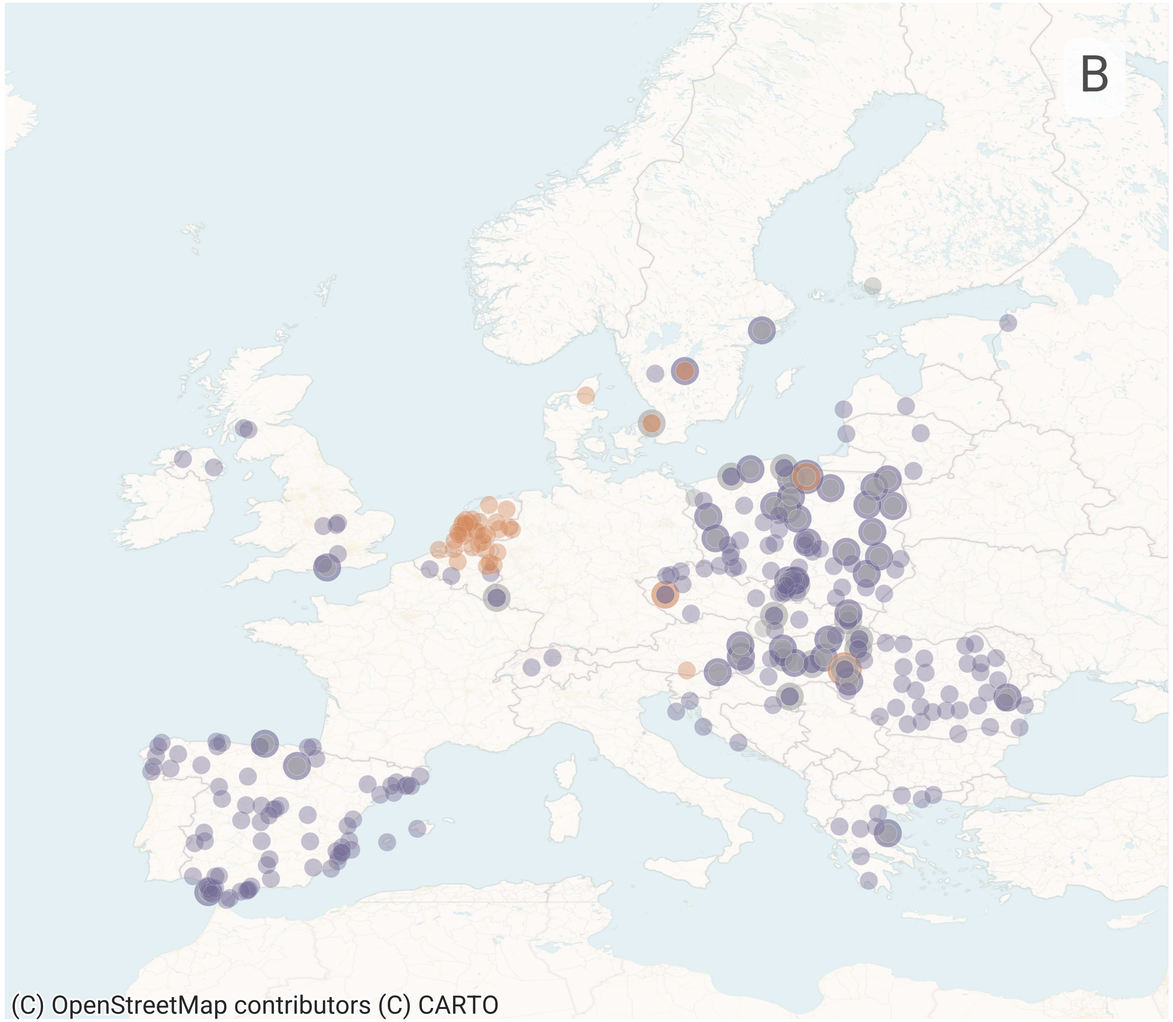
②

Predict

High cycling levels are concentrated in Northwestern Europe, while low levels dominate much of Southern and Eastern Europe.





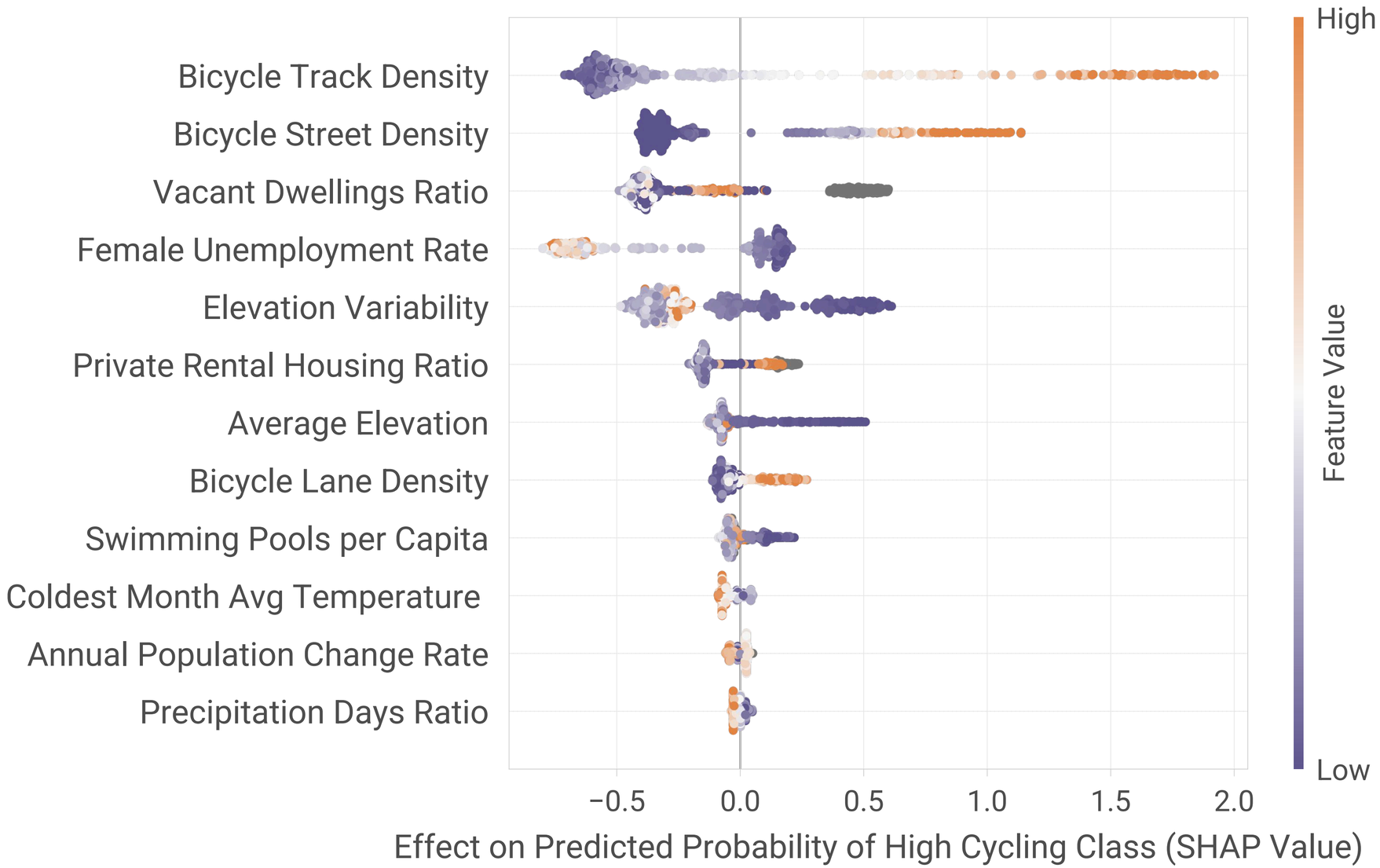


Combined Dataset

Harmonized dataset of 864 European cities with observed and predicted cycling mode share.

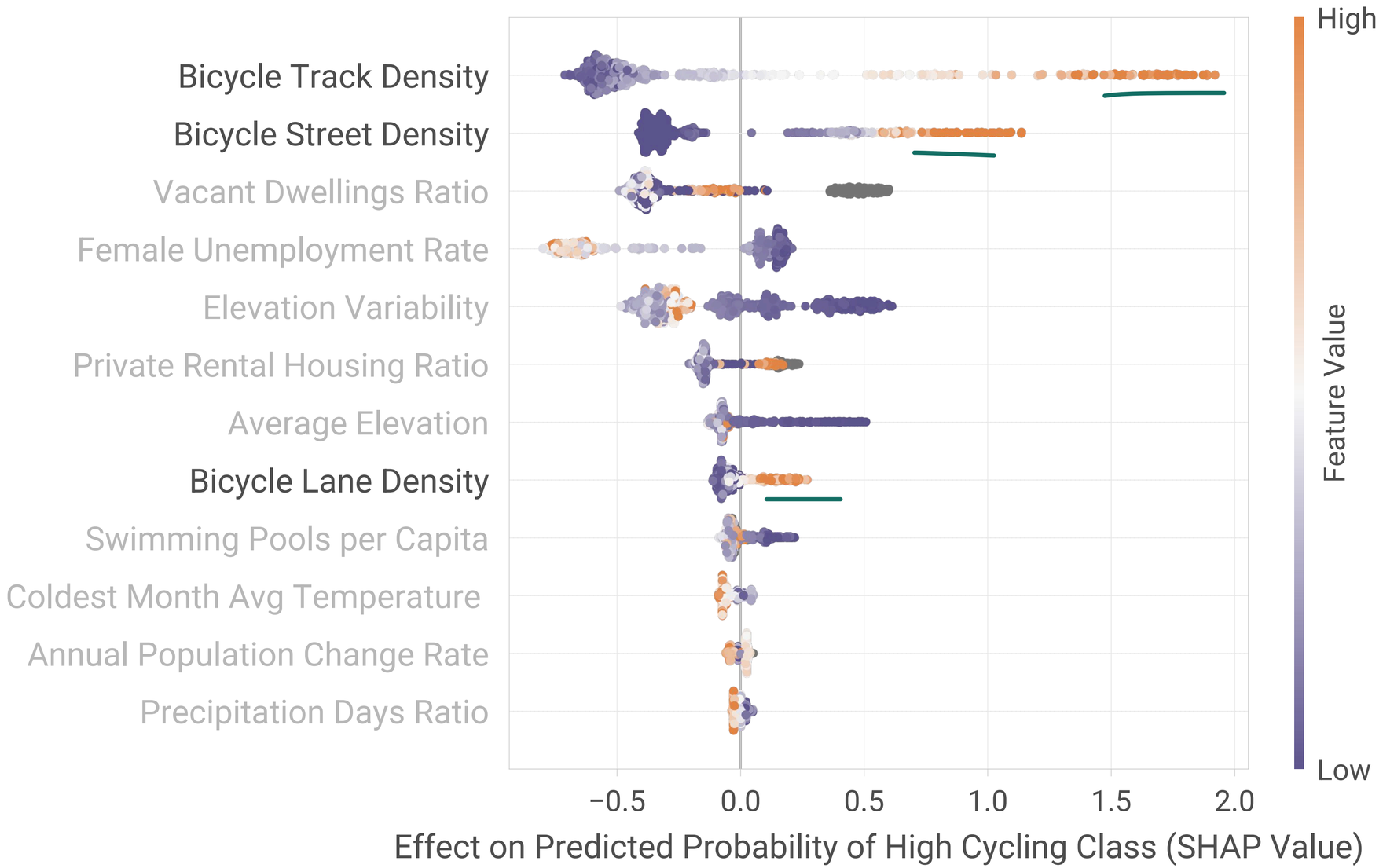
What Predicts Cycling in Europe?

SHAP values



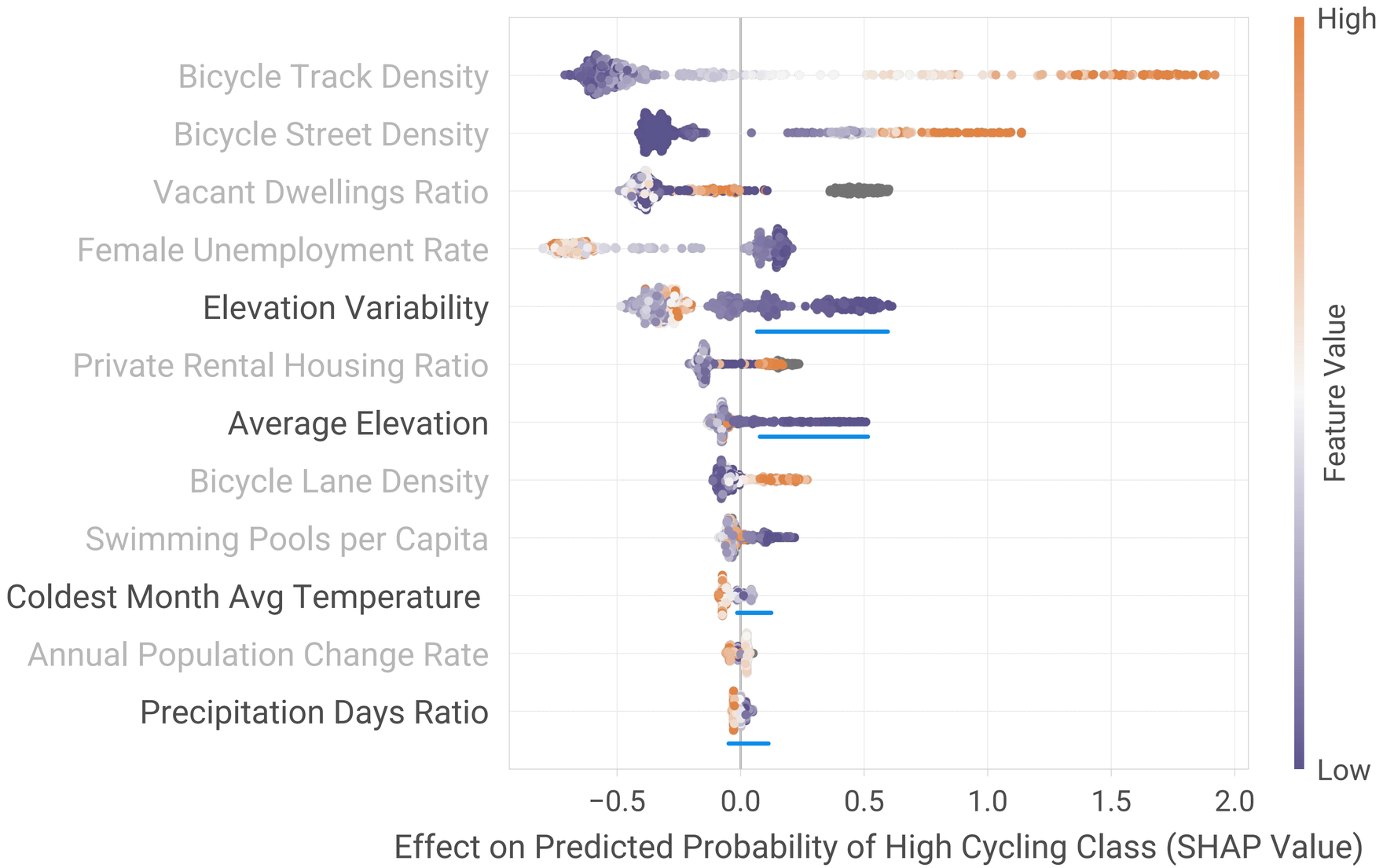
What Predicts Cycling in Europe?

Bike infrastructure



What Predicts Cycling in Europe?

Natural environment



A tale of two cities...

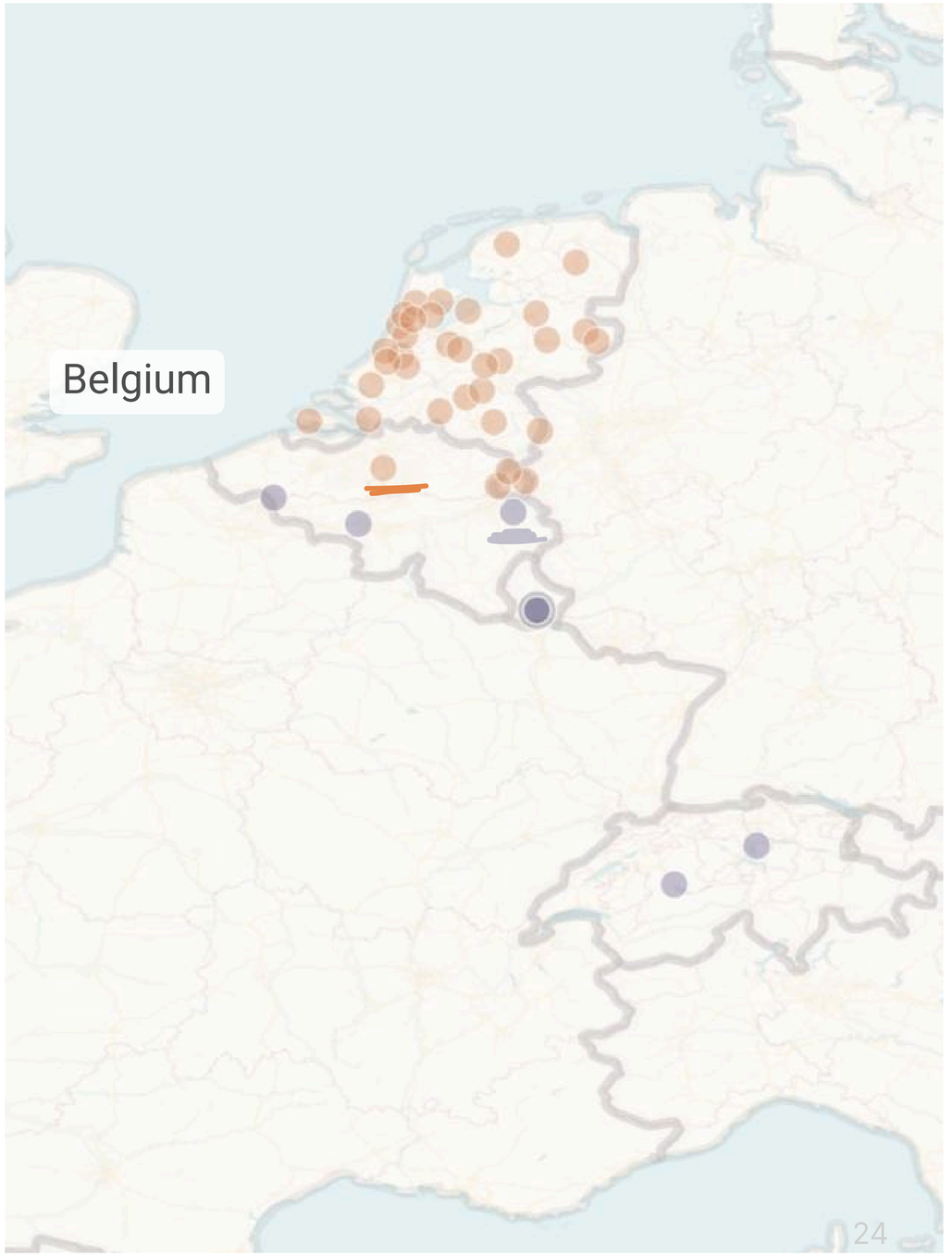
What drives high vs. low classifications?

Belgium has well-documented regional divides between Flanders and Wallonia, and the model reflects this.

A tale of two cities...

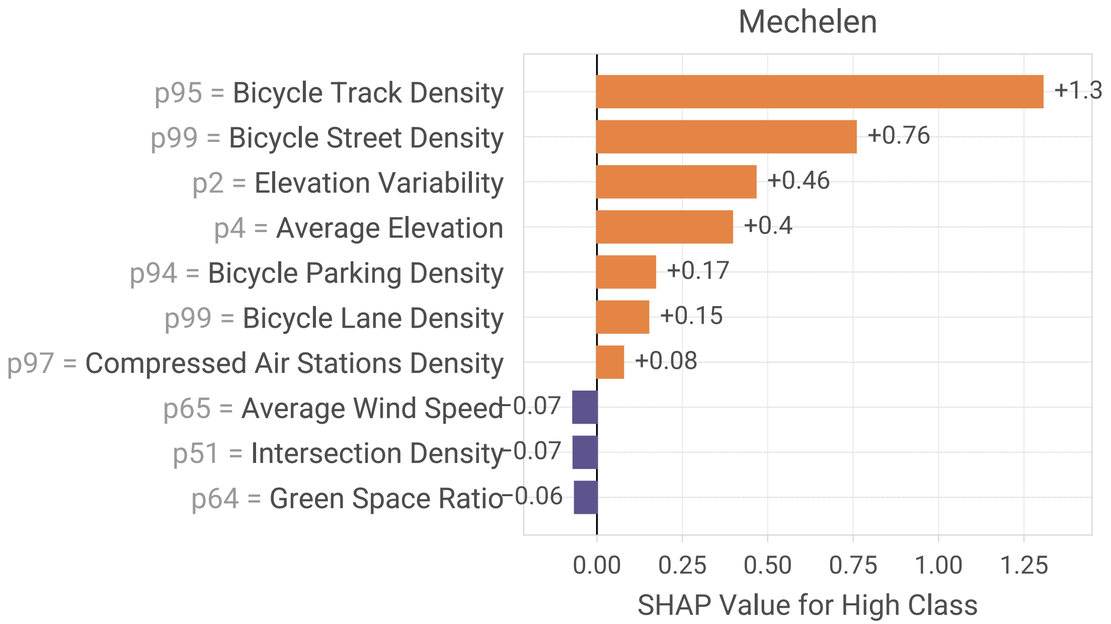
What drives high vs. low classifications?

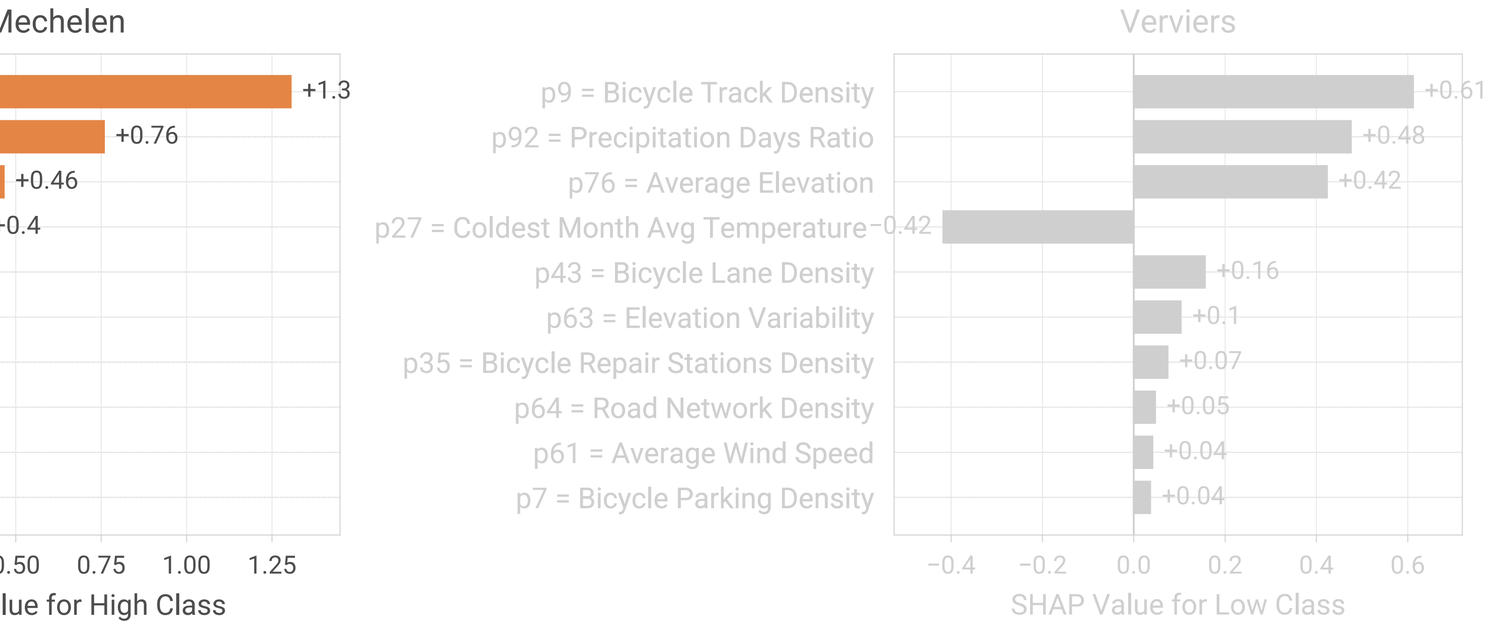
Belgium has well-documented regional divides between Flanders and Wallonia, and the model reflects this. We look at **Mechelen** and **Verviers**.



A tale of two cities...

What drives high vs. low classifications?



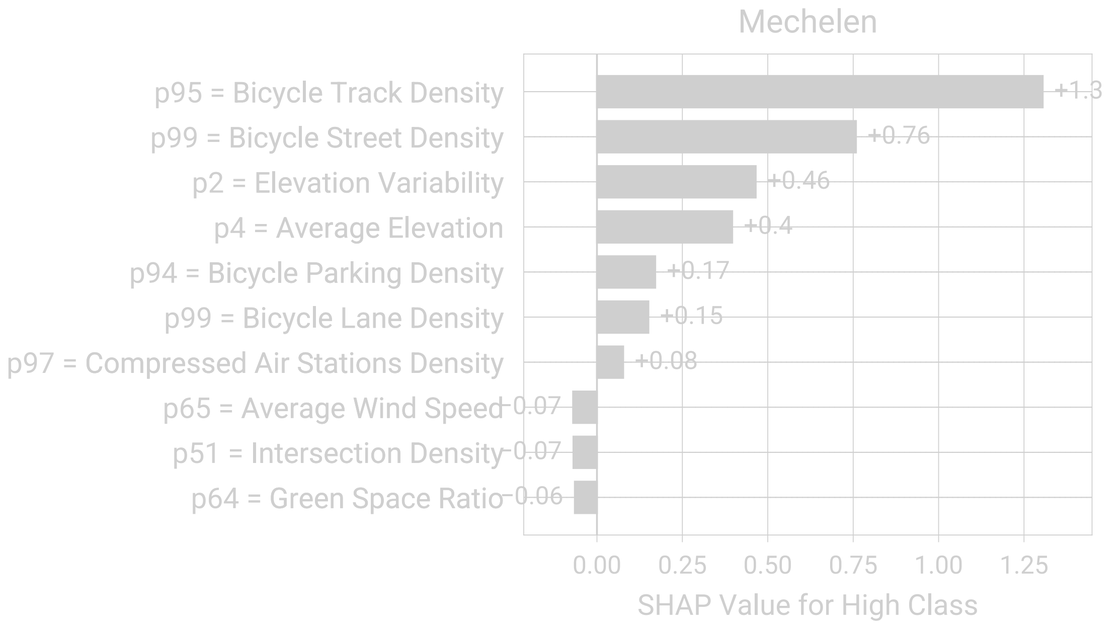


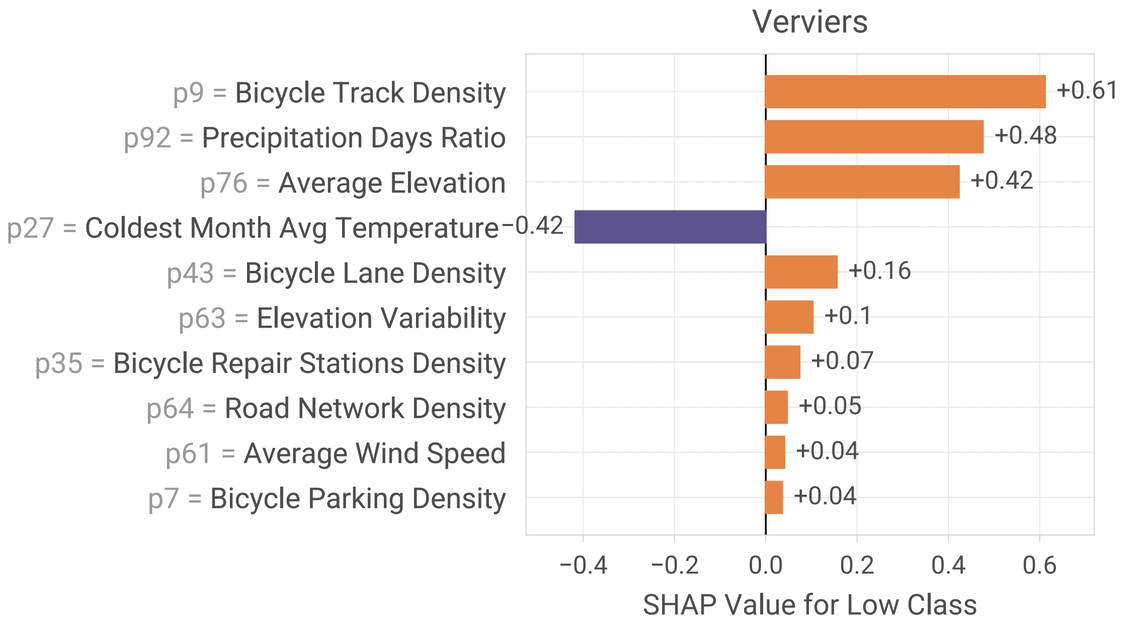
Mechelen is classified as high due to extensive infrastructure and flat terrain.

Verviers is classified as low, reflecting limited infrastructure, steep terrain, and more frequent rain.

A tale of two cities...

What drives high vs. low classifications?





Mechelen is classified as high due to extensive infrastructure and flat terrain.

Verviers is classified as low, reflecting limited infrastructure, steep terrain, and more frequent rain.

What We Learned

- ML + open data can help fill persistent gaps in cycling data.
- Model separates high- vs. low-cycling cities effectively.
- Uncertainty measures show which predictions are reliable.

- Although some European countries lead in cycling, most of the continent is still low-cycling.

Why This Matters

- The estimates provide a starting point that cities can use to set more realistic cycling targets.
- The harmonized dataset enables continent-wide studies on links between cycling, health, air quality, congestion, and emissions.

Carol Sobral
PhD Researcher
sobral@hertie-school.org

Thanks!

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101057131, Climate Action To Advance HeaLthY Societies in Europe (CATALYSE).