**Tracking Sustainable Mobility with Google Trends**

**Antonio Gutiérrez-Lythgoe,** IEDIS, University of Zaragoza

**José Alberto Molina,** IEDIS, University of Zaragoza

**Abstract**

Urban transport is central to net-zero strategies, but tracking public engagement with low-carbon mobility remains challenging. Traditional indicators like surveys or census data lack timeliness and often miss early shifts in perception. This paper addresses that gap by using Google Trends as a proxy for digital interest in sustainable urban mobility across fifteen high-income countries (2018–2025), based on monthly data. We compile a harmonized set of transport-related search terms, classified under the Avoid–Shift–Improve (ASI) framework, and correct for sampling noise via control-anchored imputation. Using principal component analysis (PCA) and K-means clustering, we identify latent attention patterns and regional typologies. The resulting Sustainable Mobility Index (SMI) tracks monthly changes in public attention to proximity-based, low-carbon, and technological mobility solutions. Its structure reflects a behavioral continuum: from car dependence to travel-avoiding behaviors, with electric vehicle interest closer to traditional modes than to genuine modal shift. Results reveal distinct engagement regimes: countries like Germany or the Netherlands show steady interest, while others remain reactive or inconsistent despite similar policy contexts. This is the first study to operationalize the ASI framework using harmonized digital trace data. The SMI provides a scalable, near-real-time tool to benchmark engagement and monitor the behavioral dimension of mobility transitions. Future work will integrate panel models with monthly economic indicators—such as consumer price indices (CPI), unemployment, and energy costs—to assess how structural and regulatory factors shape public interest.

*Keywords*: sustainable mobility, Google Trends, Big Data, machine learning, digital behavior, transportation policy

*JEL Classification*: R41, C53, Q56

**1. Introduction**

Cities are at the forefront of climate action, accounting for 70% of global carbon emissions and nearly two-thirds of energy use (United Nations, 2020). Within this urban equation, transportation plays a pivotal role—contributing around 30% of global energy consumption (IEA, 2020). Yet, even ambitious strategies centered on electrification and efficiency have proven insufficient without a significant reduction in private car use (Winkler et al., 2023). Achieving net-zero mobility thus demands more than cleaner vehicles: it requires a structural rethinking of demand itself, through policies that address not only technology, but also infrastructure, behavior, and institutions (Banister, 2008).

This broader view of mobility invites us to move beyond narrow definitions. While traditionally understood as the movement of people through space and time (Wang et al., 2022), mobility is increasingly seen as a socially embedded capability. Kaufmann et al. (2004, 2014) introduce the concept of *motility*—the capacity to access, interpret, and act upon mobility options—highlighting how opportunities for movement are shaped by socioeconomic conditions, planning decisions, and infrastructure. In this sense, mobility is also about power, inequality, and the ability (or inability) to choose among alternatives (Duranton & Turner, 2011).

Understanding these dynamics requires better tools to observe not just how people move, but how they relate to mobility itself. In recent years, digital data sources—mobile phone records, GPS traces, smartcards, geolocated social media—have transformed mobility research (Batty et al., 2012; Wu & Zhou, 2023). These datasets allow for unprecedented spatial and temporal resolution, revealing fine-grained movement patterns and behavioral routines. However, they often remain silent on motivations, perceptions, and attitudes. Mobility may be observed, but its meaning remains elusive.

To fill that gap, we turn attention to the *perceptual dimension* of mobility transitions: how people express interest in sustainable or unsustainable modes. This raises a key question: can digital traces—such as online search behavior—offer timely and scalable signals of public engagement with mobility choices? Traditional tools like surveys or census data lack the frequency and responsiveness needed to detect early shifts in interest or emerging behavioral trends.

Google Trends (GT) offers a promising alternative. As a high-frequency, geographically disaggregated source of online search data, GT has been used to track public attention across domains ranging from epidemiology (Ginsberg et al., 2009) to economics (Choi & Varian, 2012), well-being (Brodeur et al., 2021), and tourism (Havranek & Zeynalov, 2021). More broadly, GT exemplifies how Big Data can be leveraged to measure issue salience—what people care about—especially where conventional data sources fall short (Mellon, 2013; Einav & Levin, 2014).

This paper contributes to that literature by using Google Trends to track digital interest in sustainable urban mobility across fifteen high-income countries (2018–2025). Guided by the Avoid–Shift–Improve (ASI) framework (Bakker et al., 2014), we classify a harmonized set of transport-related keywords, correct for sampling inconsistencies using control-anchored imputation, and apply unsupervised learning techniques—principal component analysis and K-means clustering—to identify attention patterns over time. The resulting Sustainable Mobility Index (SMI) provides a standardized, cross-national measure of engagement with low-carbon strategies. Its structure captures a behavioral spectrum—from car-dependent routines to proximity-based and travel-reducing behaviors—offering insight into the orientation, not just intensity, of digital interest. By focusing on digitally expressed interest, this study offers a new behavioral lens for monitoring the transition to sustainable mobility. The proposed framework is scalable, replicable, and sensitive to change—making it a valuable complement to more static or costly instruments. It also opens new avenues for understanding how public attention aligns (or lags) policy ambition across contexts.

Our results reveal consistent cross-country differences in digital attention to sustainable mobility. While central European countries such as Austria, Germany, and the Netherlands exhibit sustained and proactive engagement, other groups display more reactive or volatile patterns—often driven by external shocks. The cluster analysis further distinguishes typologies of attention dynamics, highlighting how behavioral engagement varies not only in magnitude, but in temporal consistency. These patterns suggest that digital interest may serve as a leading indicator of public responsiveness to mobility policies.

The remainder of the paper is structured as follows. Section 2 reviews literature on sustainable mobility, behavioral indicators, and digital trace data. Section 3 describes the dataset, including keyword selection and country sample. Section 4 outlines the methodological approach. Section 5 presents empirical results on SMI dynamics and clustering. Section 6 discusses findings in relation to the ASI framework. Section 7 details limitations and directions for future research. Section 8 concludes.

**2. Literature review**

This section reviews four strands of literature that inform the present study: conceptual frameworks for sustainable mobility (ASI), psychological and behavioral models of modal shift, the use of big data in mobility research, and the growing role of Google Trends in social science analysis.

*The Avoid–Shift–Improve Framework*

The concept of sustainable mobility emerged in the early 1990s as a response to the escalating social and environmental costs associated with motorized transport. As an alternative to the dominant “predict and provide” logic—which promoted continuous expansion of road infrastructure—this paradigm emphasized the need to manage travel demand, enhance accessibility, and reduce car dependency (Holden et al., 2019; Foltýnová et al., 2020). One of its most enduring conceptual tools is the Avoid–Shift–Improve (ASI) framework, which organizes strategies for reducing transport-related emissions and energy use into three complementary pillars: *avoiding* unnecessary travel, *shifting* to more sustainable modes, and *improving* vehicle technologies and fuels.

Since its introduction in the late 1990s, the ASI framework has been widely adopted by policymakers, urban planners, and sustainability researchers as a guiding principle for designing low-carbon mobility systems (Turan et al., 2024). Its appeal lies in its systemic orientation: rather than privileging technological innovation alone, it promotes a balanced approach that includes demand reduction and modal substitution. However, recent empirical assessments suggest a persistent misalignment between the framework’s priorities and real-world transport interventions. For instance, Jarre et al. (2024) find that “Avoid” strategies—those aiming to reduce travel demand through urban planning, telecommuting, or behavioral change—are consistently underrepresented in climate and transport policy portfolios. In a comparative analysis of national policy databases, only 6–22% of measures were classified as “Avoid,” compared to 33% focused on technological efficiency and a growing share of “Shift” strategies.

This imbalance is notable given the high potential of demand-side measures. Arnz et al. (2024), using integrated energy modeling, demonstrate that sufficiency-oriented approaches—encompassing both travel avoidance and modal shift—can achieve reductions in energy demand and emissions comparable to those expected from improvements in vehicle efficiency. Moreover, these strategies entail different societal trade-offs: while technological improvements often depend on private investment and consumer adoption, “Avoid” and “Shift” policies typically require public coordination and infrastructural change. As such, the ASI framework continues to offer a valuable lens not only for classifying interventions, but for interrogating the political and distributive dimensions of sustainable mobility.

More than three decades after its initial formulation, ASI remains a cornerstone in sustainable transport thinking. Its relevance is particularly salient in the context of global climate goals and the urgent need to reduce transport emissions beyond what efficiency gains alone can deliver. For this reason, integrating all three pillars—even bold Avoid/Shift strategies alongside technological Improve measures—is increasingly recognized as essential. As the present study explores interest in sustainable mobility across countries, the ASI framework provides a structured and policy-relevant basis for categorizing modal strategies and interpreting digital search behavior through the lens of systemic transition. While the ASI framework defines what types of strategies are available, understanding how individuals adopt them requires turning to behavioral research on motivation and intention.

*From Behavior to Intention in Mobility*

Shifting travel behavior toward sustainability involves not just new infrastructure or technology, but also changes in attitudes, motivations, and intentions. Research on modal shift finds that convenience, cost, and personal norms all influence whether travelers choose alternatives to the private car. Environmental and climate awareness can be a powerful motivator: individuals concerned about climate change tend to drive less and show greater willingness to adopt sustainable modes (Mouratidis & Næss, 2024). Such findings suggest that raising climate awareness may translate into greater intention to use low-carbon transport. Indeed, environmental concern has been linked with increased public transit use habits and reduced car dependence over time (Bouscasse et al., 2018).

However, there is often a gap between pro-environmental intentions and actual behavior. Behavioral scientists note that daily travel is highly habitual and routine; past car use can reinforce itself, making it harder for even well-intentioned individuals to change modes​. In developed countries, strong car-use habits significantly reduce the intention to switch to public transport, even when people acknowledge environmental benefits. This intention–behavior gap means that many who claim willingness to change their travel behavior struggle to do so unless convenient alternatives and supportive policies are in place. Studies measuring willingness to change modes confirm that breaking car habits requires not only personal motivation but also external enablers like improved service, infrastructure, and incentives. For example, surveys in urban areas have identified segments of travelers open to shifting from cars to biking or transit – especially when they are younger or more climate-conscious – but also point out perceived barriers (safety, reliability) that prevent follow-through.

Therefore, transforming mobility demand hinges on both motivational factors (like climate awareness, personal norms) and situational factors (like habit disruption and viable alternatives). Effective interventions combine “soft” measures (education, awareness campaigns tapping into climate concern) with “hard” measures (policy and design changes that make sustainable modes the default)​ (Dasandi et al., 2024).

In sum, closing the gap from behavioral intention to action requires addressing psychological drivers and practical constraints – leveraging people’s growing willingness to change while dismantling the habitual and structural lock-ins that keep them in cars. This gap between intention and behavior underscores the value of complementary indicators—such as search behavior—that can signal public engagement before it manifests in observable mode shifts. The Sustainable Mobility Index (SMI) proposed in this paper aims to capture that latent attention, offering a behavioral early-warning system for mobility transitions.

*Big Data and the Limits of Mobility Insight*

The past decade has seen an explosion in digital trace data – from mobile phones, GPS devices, smart cards, and sensors – which has revolutionized how researchers study mobility. These passively collected big datasets capture travel flows at unprecedented temporal and spatial resolution, enabling analyses that were impractical with traditional travel surveys. For example, anonymized mobile phone records can now track population movements in near real time, revealing patterns of commuting, leisure travel, and even responses to shocks like pandemics​. These data-driven approaches let planners observe large-scale behavior and calibrate high-frequency models, informing policy with fine-grained trends like congestion cycles or shock responses. However, the advantages of digital mobility data come with important limitations. One key issue is representativeness and bias – not everyone’s movements are captured equally. Studies have found that different data sources (e.g. different mobile carriers or apps) can vary significantly in their coverage of travel routes and populations, leading to biased mobility estimates if used alone (Chin et al., 2025). Another limitation is the lack of social context in pure digital traces. While we can observe where and when people travel, we often lack information on who they are (demographics, income) and why a trip is made. This makes it challenging to interpret equity implications or motivations behind observed patterns. Researchers have begun to bridge this gap by combining digital mobility data with other datasets or using clever proxies. For example, studies in urban areas link phone-based movement data with neighborhood socio-economic indicators to uncover mobility inequalities: one study in Shenzhen, China found that lower-income groups and migrant workers travel significantly less (shorter distances and fewer trips) than wealthier residents, reflecting unequal access and mobility opportunities (Pan & He, 2023). researchers must be cautious about inherent biases and missing context; methodological innovations (data fusion, bias correction, privacy-aware surveys) are crucial to ensure that big mobility data lead to accurate and equitable knowledge​ When used carefully, these data can help diagnose issues like accessibility gaps or urban mobility inequalities that were previously hard to quantify, guiding more inclusive and effective transportation planning. Among the most widely used digital indicators of public attention is Google Trends, which has gained traction in multiple disciplines as a high-frequency proxy for issue salience.

*Google Trends in Social Sciences*

Google Trends (GT) data has gained increasing popularity in social sciences, particularly after 2014, reflecting a broader shift toward the use of internet search data as a tool for analyzing public interest and behavior (Hölzl et al., 2025). Hölzl et al. (2025) identify three key constructs that GT data can capture in social science research: issue salience, attitudes, and behavior. *Issue salience* refers to the level of public awareness or attention to a specific topic; *attitudes* reflect the positive or negative sentiment toward an issue, individual, or phenomenon; and *behavior* denotes either the engagement in offline actions or the expression of intention through search behavior. Among these, most GT-based studies focus on issue salience as a proxy for public interest.

A major methodological advantage of GT data lies in its capacity to capture actual search behavior, thereby bypassing some of the limitations of self-reported survey data. Unlike surveys, GT does not rely on respondents' comprehension of questions, memory accuracy, or willingness to provide socially acceptable answers. As such, it avoids cognitive biases and social desirability biases, offering a more unfiltered window into public concerns and interests. For instance, internet search data has been used to study sensitive or stigmatized topics that are often underreported in surveys, including racism (Stephens-Davidowitz, 2014), sexism (Owen & Wei, 2021), and voting behavior (Askitas, 2015a, 2015b; DiGrazia, 2017).

In fields closely related to this study, GT data has been applied to track temporal patterns of interest in climate change (Anderegg & Goldsmith, 2014, Ngheim et al., 2016). For example, Anderegg and Goldsmith (2014) analyzed the popularity of various climate-related search terms, selecting those that best matched the public framing of environmental issues and that achieved the highest relative search volume across permutations. Their study demonstrates how GT data can be aligned with longitudinal survey data to assess issue salience, following methodological strategies outlined by Mellon (2013).

Taken together, these strands of research point to the need for complementary approaches that capture both the behavioral and perceptual dimensions of mobility transitions. Yet while Big Data reveals where and when people move, it says little about their orientation toward change. By focusing on digitally expressed interest—through harmonized search behavior—this paper offers a novel lens to capture the perceptual dimension of sustainable mobility transitions.

**3. Data**

*Google Trends*

Google Trends (GT) provides data on search activity, reflecting the popularity of a given keyword as the "search interest in a particular topic, within a specific location and time frame." Instead of reporting absolute search volumes, GT normalizes data into an index ranging from 0 to 100. This normalization process allows for comparisons over time without being affected by the general increase in search queries since 2004. The index is adjusted daily based on the total number of searches in each region, accounting for seasonal fluctuations. The highest search interest within the selected period is assigned a value of 100, while all other values are scaled relative to this peak. As a result, the GT index facilitates meaningful comparisons across time periods, geographic regions, and multiple search terms, thereby mitigating biases associated with raw search volume data.

However, Google assigns a value of 0 to data points where search volumes fall below a certain undisclosed threshold. According to the company, this measure is intended to protect user anonymity (Google News Initiative, 2024). Consequently, a 0 value may indicate either a complete absence of searches for a term or a search volume too low to be reported. This poses a challenge for analyzing terms with inherently low search frequency, as their time series may contain numerous missing or zero values, limiting their interpretability and consistency in long-term trend analysis. Literature identifies three major challenges when using Google Trends (GT) in social science research: internal validity, reliability, and generalizability (Höltzl et al., 2025).

*Internal validity*

Internal validity refers to whether search volumes accurately reflect real-world behaviors and attitudes or merely indicate temporary interest, which may be influenced by external factors such as media coverage, social trends, or changes in search terminology.

Previous research has explored the connection between online search behavior and urban mobility, suggesting that keyword popularity can serve as a proxy for movement patterns (Kostakos et al., 2013). Their findings indicate that search terms are often semantically relevant to specific locations, reinforcing the idea that online interest may be linked to physical mobility. However, their results have also highlighted limitations in using high-frequency (e.g., hourly) search data, as circadian rhythms (daily activity cycles) can generate strong correlations across all locations, reducing the ability to differentiate mobility trends effectively. Daily aggregated data has proven to be a more reliable approach in this context. While search data is collected at a national scale, there is evidence that pedestrian flows at specific locations exhibit strong correlations only with relevant search queries, suggesting that search behavior is not entirely random but reflects actual movement patterns.

*Reliability and sample variation*

One of the main concerns when using Google Trends (GT) data is measurement reliability. Because Google’s internal algorithms and scaling methods are undisclosed, there is limited transparency regarding how search queries are sampled and normalized. Rather than reporting the full universe of searches, GT applies a sampling procedure with unknown characteristics. As a result, the same query can yield different time series when extracted repeatedly, introducing inconsistencies that may affect the interpretability and reproducibility of empirical models (Choi & Varian, 2012; Cebrián & Doménech, 2024).

To mitigate this, we introduce a high-volume, stable control term— “google”—into each extraction. The control acts as a technical benchmark to stabilize the index returned by GT, reducing internal scaling inconsistencies and allowing for consistent comparison both across extractions and across keywords. We also construct a combined query (e.g., “supermarket near me + google”) to further reduce variation in the target keyword’s trend.

*[Figure 1 about here]*

Figure 1 illustrates the extent of sampling variation using 50 repeated extractions for the query “supermarket near me” in Sweden. Panel (a) displays the raw signal, where large differences across samples are evident. In contrast, panel (b) shows the control term alone, and panel (c) displays the combined query. Both lead to markedly more stable signals, supporting the use of control-anchored queries to reduce variability.

To recover a consistent signal from multiple extractions, we implement a regression-based imputation procedure by Stephens-Davidowitz (2014). For each keyword, we regress its observed values against both the control term and the combined query using repeated samples and impute missing or noisy values based on the estimated relationship. This approach is detailed in Appendix A.

*[Figure 2 about here]*

Figure 2 contrasts the 50 raw extractions with the final imputed series (in red), obtained via regression on the control term. The resulting signal is more stable and robust across time. This imputation procedure is repeated for every keyword and country in the dataset.  
This correction protocol enables the construction of longitudinal and cross-national indicators based on GT, addressing key barriers to its use in comparative mobility research.  
While this method substantially reduces sample variation, several structural sources of error mat remain. These include low search popularity, small population sizes at the regional level, and higher-frequency data extractions (Eichenauer et al., 2022; Cebrián & Doménech, 2024). Eichenauer et al. (2022) show that even monthly GT data can be unreliable in areas with fewer than 10 million inhabitants, and that daily and weekly data exhibit similar instability. However, these limitations are less relevant in our case, as we work with monthly data aggregated at the national level and restrict the analysis to high-income countries with adequate search volumes. Another limitation is GT’s caching mechanism, which returns the same sample if a query is repeated within a 24-hour window. This can delay data collection and affect its freshness for real-time monitoring. To address these issues, prior studies recommend aggregating extractions taken on different days (Carrière-Swallow & Labbé, 2013; D’Amuri & Marcucci, 2017; Eichenauer et al., 2022).

*Comparison of terms and generalization*

A separate limitation of GT concerns the comparison of multiple keywords. Since each query allows a maximum of five terms and is normalized independently, the returned values cannot be directly compared across batches. For example, the trend for “car” extracted in one batch cannot be meaningfully compared to that of “bicycle” extracted in another, as each is scaled relative to the maximum value in its respective set.To address this, we include a shared control term in all keyword groups (as described above), allowing each keyword to be expressed relative to the same benchmark. This adjustment ensures consistent scaling and enables valid comparisons across terms, time periods, and countries. As a result, trends extracted from GT reflect relative salience rather than artifacts of batch-specific normalization. Finally, it is important to acknowledge that GT data may not fully represent the general population. Differences in internet access, digital literacy, and platform usage can introduce selection biases—particularly in countries or regions with lower digital penetration (Höltzl et al., 2025).

*Data collection strategy*

To ensure semantic and cultural relevance, we compiled a country-specific and language-adapted set of keywords reflecting everyday mobility-related search behavior. Our sample includes fifteen high-income countries selected for their adequate levels of urbanization, internet penetration, and data availability. These comprise both European and non-European cases: Spain, France, Germany, Italy, Portugal, the Nordic countries, the United Kingdom, Ireland, the Netherlands, and the United States.

Keywords were translated into contextually appropriate language and organized following the Avoid–Shift–Improve (ASI) framework:

* **Avoid:** minimizing the need for travel (e.g., working from home, online shopping, supermarket/restaurant near me)
* **Shift** (modal shift): changing modes (e.g., bus, bike, taxi, car)
* **Improve** (technological efficiency): increasing vehicle efficiency or reducing emissions (e.g., electric car, hybrid car, car charging, fuel consumption)

This taxonomy captures both intentional behaviors and structural enablers relevant to mobility decisions. To allow cross-country comparability, all localized keywords were mapped to a set of unified semantic categories (e.g., ‘coche eléctrico’, ‘voiture électrique’, ‘electric car’ → electric\_car) using a multilingual dictionary (see Appendix B). This mapping enabled aggregation across languages and ensured conceptual consistency in the PCA step.

**4. Methods**

*Sustainable mobility index (SMI)We focused*

To synthesize mobility-related search interest into a single indicator, we apply Principal Component Analysis (PCA) to the full panel of imputed, standardized, and deseasonalized keyword time series across all countries, using monthly data from 2018 to 2025. This technique reduces the dimensionality of the dataset while retaining the maximum possible variance across terms, allowing us to capture a latent dimension of sustainable mobility interest shared across contexts.

Seasonality is removed using STL decomposition (Seasonal-Trend decomposition based on Loess), a robust filtering technique that separates each time series into trend, seasonal, and residual components. These preprocessing steps ensure that the PCA captures structural behavioral trends rather than differences in scale or recurring seasonal cycles.

*[Figure 3 about here]*

We retain only the first principal component (PC1), which captures the largest share of variance common to all keywords. This component is interpreted as a synthetic indicator of digitally expressed interest in sustainable mobility.

*[Figure 4 about here]*

The contribution of each keyword to PC1 (i.e., the loadings) provides insight into the behavioral structure underlying digital interest. As shown in Figure 5, proximity-based or digital behaviors (e.g., *supermarket near me*, *remote work*) load positively, whereas car-dependent terms (e.g., *car*, *traffic*) tend to load negatively. This pattern confirms that PC1 reflects a gradient from traditional to more sustainable mobility preferences.

*[Figure 5 about here]*

The resulting Sustainable Mobility Index is standardized (mean = 0, standard deviation = 1) and expresses, at each point in time, how a country's search interest deviates from the global average. Higher values indicate relatively stronger interest in sustainable mobility behaviors.

*Clustering Sustainable Mobility Countries*

To identify territorial patterns in digital attention toward sustainable mobility, we apply the K-means clustering algorithm to the country-level trajectories of the Sustainable Mobility Index. K-means is an unsupervised machine learning method widely used in regional analysis for its efficiency and interpretability. The algorithm partitions the data into *k* clusters by minimizing within-cluster variance, assigning each observation to the nearest cluster centroid based on Euclidean distance (Saputra et al., 2020).

A critical step in applying K-means is determining the optimal number of clusters *k*, which is not known a priori. We use the elbow method to guide this decision—a heuristic that plots the explained variance (or reduction in within-cluster sum of squared errors) as a function of *k*. The optimal *k* corresponds to the point at which additional clusters yield marginal improvements in model fit.

A formal analysis of structural and contextual drivers of the Sustainable Mobility Index (SMI) is left for future work. Appendix C outlines a proposed panel data specification that integrates ASI-aligned predictors with country and year fixed effects.

**5. Results**

*Correction of Sampling Variability in Google Trends Data*

Google Trends (GT) data are known to exhibit sampling variability, which can compromise the temporal consistency of extracted time series. As detailed in Section 3, we mitigate this issue by combining control-anchored queries with a regression-based imputation strategy. Figures 1 and 2 illustrate the effectiveness of this approach. In Figure 1a, 50 repeated extractions of the query “supermarket near me” in Sweden display substantial dispersion, suggesting that a significant share of observed variation stems from Google’s internal sampling and scaling routines. In contrast, the inclusion of a high-volume control term (“google”) and the use of a combined query (Figure 1b and 1c) yield much more stable series. Figure 2 further shows that regression-based imputation can recover a coherent and interpretable trend by smoothing erratic fluctuations and preserving seasonal structure.

These results confirm the usefulness of the correction method and highlight some remaining limitations. While comparability and signal stability are improved, residual noise persists for low-volume terms and smaller countries, reinforcing the need for cautious interpretation of short-term trends. From a policy perspective, corrected GT series offer a timely and low-cost proxy for monitoring public attention to sustainable mobility—particularly in contrast to traditional mobility surveys, which are expensive and infrequent. However, their use should be framed as complementary rather than substitutive: residual measurement error, linguistic bias, and digital divide effects still constrain representativeness. Combining GT-based indicators with behavioral data—such as mobility tracking or vehicle registration records—can improve robustness and offer a more complete picture of mobility transitions.

*Interpreting Digital Interest in Sustainable Mobility*

To track digitally expressed interest in sustainable mobility across countries and over time, we construct a composite indicator based on Google Trends data. As detailed in Section 3, we apply Principal Component Analysis (PCA) to the panel of standardized, deseasonalized, and imputed keyword time series. The first principal component (PC1), which captures the largest share of shared variance across terms, is retained and interpreted as the Sustainable Mobility Index (SMI).

This index reflects a latent behavioral dimension: higher values indicate stronger relative interest in sustainable mobility behaviors—such as proximity-based consumption, public transport, or electric vehicles—while lower values reflect a prevalence of car-dependent or traditional mobility preferences.

The structure of PC1 loadings confirms this interpretation (Figure 5). Keywords associated with car use—such as *car*, *traffic*, and *fuel consumption*—exhibit the largest absolute loadings, anchoring one end of the axis. In contrast, proximity-based and “avoid” behaviors—such as *supermarket near me*, *restaurant near me*, and *remote work*—are closest to zero, and thus positively associated with the sustainable mobility dimension. In between, we find intermediate terms reflecting modal shift (*bus*, *bike*, *car sharing*) and technological improvement (*hybrid car*, *electric car*).

This pattern reveals a coherent continuum: from car dependence at one pole to travel avoidance and localized behavior at the other. Notably, even keywords linked to cleaner vehicle technologies (e.g., *electric car*) still load negatively, suggesting that technological “improvements” remain behaviorally closer to car use than to genuine mode shift or travel reduction. Meanwhile, *online shopping* and *parking* occupy a central position, underscoring their dual role—both as alternatives to travel and as facilitators of car-based mobility.

Together, these loadings validate the SMI as a meaningful and interpretable index of digitally expressed sustainable mobility interest, sensitive to both behavioral and structural dimensions of transition.

*[Figure 5 about here]*

*Clustering of SMI Trajectories Across Countries*

To explore cross-national heterogeneity in sustainable mobility interest, we apply K-means clustering to the time series of the Sustainable Mobility Index. The goal is to identify groups of countries that exhibit similar SMI trajectories over time, regardless of differences in magnitude or timing.

*[Figure 6 about here]*

The optimal number of clusters is determined using the elbow method, which evaluates the marginal gain in explained variance as the number of clusters (*k*) increases. Figure 6 reveals a clear inflection point at *k* = 3, suggesting that this configuration captures the underlying structure of the data without overfitting. This choice is further validated by the silhouette analysis (Figure 8a), which shows that the three-cluster solution achieves a reasonable trade-off between internal cohesion and separation across clusters, with consistently positive silhouette scores across groups.

*[Figure 8a and 8b about here]*

Each violin plot in Figure 8a shows the distribution of silhouette coefficients for the three clusters. The mean silhouette score across all observations is 0.046, indicating modest but positive separation. Cluster 2 (green) exhibits the highest internal cohesion, with values concentrated between 0.03 and 0.12 and a median around 0.06. Cluster 1 (orange) shows intermediate performance, with a broader range and some slightly negative values. Cluster 0 (blue) has the lowest scores overall, though still mostly positive. These patterns confirm that the clusters reflect meaningful latent structure, even if boundaries are not sharply defined.

*[Table 1 about here]*

Cluster composition is summarized in Table 1:

* Cluster 0 includes Denmark, Finland, Norway, and Sweden—Nordic countries with moderately low average SMI (mean = –0.328) and relatively high internal variability (std = 2.064).
* Cluster 1 is the largest and most heterogeneous group, comprising Belgium, Spain, France, the United Kingdom, Ireland, Italy, Portugal, and the United States. These countries exhibit the lowest average SMI (mean = –0.431) but a similar degree of dispersion (std = 2.050), reflecting a mix of stable and volatile engagement patterns.
* Cluster 2 includes Austria, Germany, and the Netherlands—central European countries with slightly higher average interest (mean = –0.407) and the lowest variability (std = 1.896), suggesting more consistent engagement with sustainable mobility topics over time.

Figure 8b illustrates the distribution of SMI values by cluster. Cluster 0 is centered well below the global mean (median ≈ –0.5), indicating persistently low digital engagement with sustainable mobility topics. Cluster 1 has a wider spread and a median just below zero, reflecting fluctuating interest levels with occasional spikes. Cluster 2 displays the highest median (≈ +0.5) and a long positive tail, capturing consistently high—and at times exceptionally strong—engagement.

These distinctions point to meaningful behavioral and structural differences among groups. Cluster 2 countries tend to combine earlier policy action, greater digital maturity, and stronger EV adoption. Cluster 0 reflects more reactive trajectories, possibly shaped by recent fuel shocks or delayed policy uptake. Cluster 1 falls in between, with mixed drivers and more heterogeneous patterns of engagement.

*Drivers of Sustainable Mobility Interest*

Panel data specification (outlined in Appendix C)

**6. Discussion and Policy Implications**

*Reliability of Google Trends Indicators*

The analysis in Figures 1 and 2 confirms that raw Google Trends (GT) series contain significant sampling noise, as documented in prior literature. Our correction method—based on a stable control term and regression-based imputation—substantially improves signal reliability and yields three key contributions.

First, it offers a replicable workflow for using GT at monthly resolution, showing that much of the observed variance is algorithmic rather than behavioral. Second, it enhances comparability across countries by anchoring low-volume queries to a globally consistent reference term, which is essential for constructing composite indices like the SMI. Third, it reveals the limits of GT data: residual noise remains in low-search or small-population contexts, suggesting that short-term fluctuations should be interpreted with caution. From a policy perspective, the corrected series provide a more timely proxy for public interest in sustainable mobility than traditional surveys. They could support early-warning systems to evaluate communication or incentive policies, though they should be used in conjunction with behavioral data to fully assess policy impact.

*Interpreting the Sustainable Mobility Index (SMI)*

The Sustainable Mobility Index (SMI) captures a behavioral continuum from car-dependence to proximity-based behaviors. This gradient nuance moves beyond the binary “car vs. green” framing often used in transport policy and helps account for why interest in cleaner vehicle technologies—such as electric or hybrid cars—tends to remain behaviorally closer to car use than to modal shift alternatives. Rather than signaling a categorical transition, the SMI reflects a spectrum of mobility behaviors. For instance, online shopping or parking can both substitute for travel and reinforce car-centric lifestyles—depending on context.

The structure of keyword loadings along PC1 aligns closely with the Avoid–Shift–Improve (ASI) framework, lending empirical support to its behavioral relevance. Proximity-based or digital behaviors (e.g., “supermarket near me”) cluster at one end of the spectrum, while car-related queries (e.g., “traffic”) anchor the other. Modal-shift options (e.g., “bus,” “bike”) and technological improvements (e.g., “hybrid car,” “electric car”) occupy intermediate positions. This pattern suggests that online interest in sustainable mobility is not monolithic; even among environmentally aligned behaviors, meaningful differences persist in how they relate to existing transport paradigms.

Constructing the SMI using a pooled panel across countries facilitates international comparability, allowing for relative levels of interest to be tracked without needing to re-scale the index. This property supports its potential use in cross-country panel models—such as fixed-effects regressions incorporating economic variables or policy shocks—as discussed in Section 5. However, several limitations must be acknowledged. The composition of the keyword set may influence principal component structure over time; future applications would benefit from periodic reassessment to account for evolving terminology or mobility trends. Additionally, cross-linguistic differences and unequal internet access may introduce measurement bias across countries, particularly in lower-penetration or multilingual contexts.

Despite these constraints, the SMI offers a promising proxy for tracking digitally expressed attention to sustainable mobility in near real time. It may complement more costly or infrequent instruments—such as travel surveys or adoption data—by providing continuous signals of public responsiveness. Policymakers can leverage such indicators to monitor reactions to campaigns, evaluate early-stage pilots (e.g., low-emission zones), or anticipate behavioural shifts following fuel price changes or new incentives. Future work should test its predictive capacity formally and integrate it with structural modelling of transport behavior and policy uptake.

*Typologies of Digital Attention: Cluster Analysis and Policy Implications*

The cluster analysis of Sustainable Mobility Index (SMI) trajectories reveals a meaningful, though moderately distinct, segmentation of countries in terms of digitally expressed interest in sustainable mobility. The three-cluster solution—identified via the elbow method and supported by positive silhouette scores (mean = 0.046)—offers a heuristic typology for interpreting public attention dynamics and aligning policy approaches to behavioral context. Although boundaries are not sharply defined, the configuration yields practical insights when treated as indicative rather than definitive.

Cluster 0, comprising Denmark, Finland, Norway, and Sweden, can be characterized as “low-but-volatile.” Despite high performance in EV uptake, cycling infrastructure, and modal shift outcomes, these countries show below-average SMI levels (mean = –0.328) and the highest intra-annual variability. One plausible explanation is that many sustainable practices are already institutionalized, so marginal online search interest becomes episodic—spiking in response to events such as fuel price fluctuations or technology rollouts. If correct, this would suggest that GT-based indicators may understate maturity in advanced contexts. Accordingly, digital attention metrics in Cluster 0 are better interpreted in conjunction with usage-based indicators such as modal split or EV registrations. Communication strategies could be refocused toward next-phase priorities, including freight decarbonization and rural infrastructure gaps.

Cluster 1 includes Belgium, Spain, France, the United Kingdom, Ireland, Italy, Portugal, and the United States—the largest and most heterogeneous group. With the lowest average SMI (–0.431) and the widest standard deviation (≈ 2.05), it is marked by intermittent surges in interest, typically coinciding with salient events such as fuel price shocks, climate protests, or clean transport subsidies. These spikes, however, are often followed by periods of low engagement. This pattern may reflect limited anchoring of sustainable mobility in everyday practices, or a fragmented public discourse. In these cases, continuity instruments—such as stable EV tax incentives, long-term investment in active travel, and permanent low-emission zones—could help translate episodic awareness into sustained behavioral change. Given the group’s internal variability, spatially differentiated strategies (e.g., urban–rural splits) are likely to be more effective than uniform national approaches.

Cluster 2, comprising Austria, Germany, and the Netherlands, stands out for its relatively high median SMI (≈ +0.5) and low volatility (σ ≈ 1.90). These “early-and-steady adopters” appear to maintain consistently elevated digital interest in sustainable mobility topics. This stability may be associated with long-standing investments in public transport, fuel pricing, and EV infrastructure, as well as more cohesive policy communication. For these countries, the policy focus may need to shift from awareness generation to issues of scale and equity. Measures such as road pricing schemes, integrated mobility-as-a-service (MaaS) platforms, or social equity safeguards in low-carbon transitions could build on an already engaged base of public attention.

Across clusters, three broader insights emerge. First, digital infrastructure appears to shape the structure of attention: countries with higher broadband penetration and smartphone use tend to exhibit more stable SMI trajectories, suggesting that virtual and physical transition capacities are mutually reinforcing. Second, external shocks—particularly in Clusters 0 and 1—continue to play a central role in shaping online interest, highlighting the salience-driven nature of digital engagement. Finally, the typology proposed here offers a useful starting point for differentiated policy design: communication and agenda-setting may be more effective in Cluster 1, innovation pilots in Cluster 0, and consolidation strategies in Cluster 2.

While exploratory in nature, these cluster-based distinctions go beyond static country rankings and offer a behavioral lens through which digital interest in mobility transitions can be monitored, interpreted, and eventually modelled. Future research could test whether transitions across clusters—e.g., from episodic to sustained engagement—are associated with structural reforms, pricing instruments, or digital inclusion efforts.

**7. Limitations and Directions for Future Research**

This study offers a novel framework for tracking digitally expressed interest in sustainable mobility across countries and over time. By constructing a composite index based on Google Trends (GT) data and correcting for sampling inconsistencies, we generate a behaviorally meaningful indicator that is scalable, timely, and interpretable. However, several limitations remain, which also open fruitful avenues for future research.

First, the index is sensitive to the composition of the keyword set. While we anchor the selection in the Avoid–Shift–Improve framework and harmonize terms across countries, search vocabulary can evolve over time and may vary by demographic or cultural context. Periodic re-evaluation of term relevance, particularly in response to new technologies or policy shifts, is recommended.

Second, although we correct for internal variability in GT data using control-anchored imputation, structural limitations persist—especially for low-frequency queries or countries with smaller populations. The SMI should therefore be interpreted cautiously in contexts with weak digital penetration or sparse search activity, and ideally complemented by behavioral metrics such as mobility patterns or vehicle registrations.

Third, the current clustering analysis is exploratory and relies on K-means, which assumes spherical clusters and equal variance. Alternative methods (e.g., hierarchical clustering, dynamic time warping, or latent class models) may offer additional insights, particularly in capturing non-linear trajectories or structural transitions in attention over time.

Fourth, the analysis focuses on a set of high-income, mostly European countries, plus the United States. Extending the framework to include countries in Eastern Europe, Latin America, or Asia–Pacific would improve the generalizability of results and allow for testing whether the SMI functions similarly in lower digital maturity environments.

Finally, while the index captures public attention, it does not measure actual behavioral change. Future work could integrate the SMI with high-frequency data on transport use, fuel consumption, or EV adoption to examine whether digital interest anticipates observable shifts in mobility behavior. In this sense, the SMI may serve as an early-warning indicator—particularly useful for evaluating the resonance and impact of public policy measures in near real time.

Future research could also leverage the panel structure of the data to estimate causal models of sustainable mobility interest, using economic, infrastructural, or regulatory variables as predictors. This would allow for stronger inference on the drivers of attention and provide evidence for the effectiveness of specific policy tools under different conditions.

**8. Conclusion**

This working paper presents a first attempt to measure and compare digitally expressed interest in sustainable mobility across fifteen high-income countries, using a composite index derived from corrected Google Trends data. Building on methodological approaches developed in prior work—particularly the use of control terms and regression-based imputation to correct sampling variability—we construct a Sustainable Mobility Index (SMI) using principal component analysis applied to mobility-related search terms organized under the Avoid–Shift–Improve framework.

Our contribution lies in the systematic application and cross-national extension of these methods to the behavioral domain of sustainable mobility. The resulting index provides a temporally consistent and cross-nationally comparable signal of digital attention. Applied to the 2018–2025 period, the analysis reveals clear country differences and stable typologies of engagement. For example, Austria, Germany, and the Netherlands show the highest and most consistent levels of digital interest, while Nordic countries exhibit lower mean values with higher volatility—despite leading the world in EV uptake and modal shift. These patterns suggest that public attention and behavioral adoption are not always aligned, and that digital proxies may capture complementary dimensions of the transition.

The clustering analysis identifies three groups of countries with distinct temporal dynamics, which may reflect differences in policy sequencing, digital maturity, or responsiveness to external shocks. This typology offers a pragmatic lens for tailoring communication, incentives, and infrastructure strategies across national contexts. In addition, the index itself can serve as a low-cost monitoring tool, potentially useful for real-time policy evaluation and public engagement dashboards.

This version is exploratory and does not yet include causal modelling or integration with socioeconomic drivers. Future work will extend the panel dataset, test for policy responsiveness using fixed-effects regression models, and evaluate the predictive validity of the SMI against behavioral outcomes such as vehicle registrations, modal split, or transport emissions.

**References**

Askitas, N. (2015a). Calling the Greek Referendum on the nose with Google Trends. *Available at SSRN 2633443*.

Askitas, N. (2015b). *Predicting the Irish" Gay Marriage" Referendum* (No. 9570). IZA Discussion Papers.

Anderegg, W. R., & Goldsmith, G. R. (2014). Public interest in climate change over the past decade and the effects of the ‘climategate’ media event. *Environmental Research Letters*, *9*(5), 054005.

Arnz, M., Göke, L., Thema, J., Wiese, F., Wulff, N., Kendziorski, M., ... & von Hirschhausen, C. (2024). Avoid, Shift or Improve passenger transport? Impacts on the energy system. *Energy Strategy Reviews*, *52*, 101302.

Banister, D. (2008). The sustainable mobility paradigm. *Transport policy*, *15*(2), 73-80.

Bakker, S., Zuidgeest, M., De Coninck, H., & Huizenga, C. (2014). Transport, development and climate change mitigation: Towards an integrated approach. *Transport Reviews*, *34*(3), 335-355.

Bouscasse, H., Joly, I., & Bonnel, P. (2018). How does environmental concern influence mode choice habits? A mediation analysis. *Transportation research part D: transport and environment*, *59*, 205-222.

Brodeur, A., Clark, A. E., Fleche, S., & Powdthavee, N. (2021). COVID-19, lockdowns and well-being: Evidence from Google Trends. *Journal of public economics*, *193*, 104346.

Carrière‐Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.

Cebrián, E., & Domenech, J. (2024). Addressing Google Trends inconsistencies. *Technological Forecasting and Social Change*, *202*, 123318.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, *88*, 2-9.

Chin, T., Johansson, M. A., Chowdhury, A., Chowdhury, S., Hosan, K., Quader, M. T., ... & Mahmud, A. S. (2025). Bias in mobility datasets drives divergence in modeled outbreak dynamics. *Communications Medicine*, *5*(1), 8.

D’Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, *33*(4), 801-816.

Dasandi, N., Jankin, S., Pantera, D. K., & Romanello, M. (2025). Public engagement with health and climate change around the world: a Google Trends analysis. *The Lancet Planetary Health*, *9*(3), e236-e244.

DiGrazia, J. (2017). Using internet search data to produce state-level measures: The case of tea party mobilization. *Sociological Methods & Research*, *46*(4), 898-925.

Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, *101*(6), 2616-2652

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2022). Obtaining consistent time series from Google Trends. *Economic Inquiry*, *60*(2), 694-705.

Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, *346*(6210), 1243089.

Foltýnová, H. B., Vejchodská, E., Rybová, K., & Květoň, V. (2020). Sustainable urban mobility: One definition, different stakeholders’ opinions. *Transportation research part D: Transport and environment*, *87*, 102465.

Google, 2024. FAQ about Google Trends Data. <https://support.google.com/trends/answer/4365533?hl=en>.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012-1014.

Havranek, T., & Zeynalov, A. (2021). Forecasting tourist arrivals: Google Trends meets mixed-frequency data. *Tourism Economics*, *27*(1), 129-148.

Holden, E., Gilpin, G., & Banister, D. (2019). Sustainable mobility at thirty. *Sustainability*, *11*(7), 1965.

International Energy Agency. (2020). *World energy balances 2020*. IEA.

Jarre, M., Noussan, M., & Campisi, E. (2024). Avoid–Shift–Improve: Are Demand Reduction Strategies Under-Represented in Current Energy Policies?. *Energies*, *17*(19), 4955.

Kaufmann, V., Bergman, M. M., & Joye, D. (2004). Motility: Mobility as capital. *International journal of urban and regional research*, *28*(4), 745-756.

Kaufmann, V. (2014). Mobility as a Tool for Sociology. *Sociologica*, *8*(1), 0-0.

Kostakos, V., Juntunen, T., Goncalves, J., Hosio, S., & Ojala, T. (2013). Where am I? Location archetype keyword extraction from urban mobility patterns. *PloS one*, *8*(5), e63980.

Mellon, J. (2013). Where and when can we use Google Trends to measure issue salience?. *PS: Political Science & Politics*, *46*(2), 280-290.

Mouratidis, K., & Næss, P. (2024). Climate change concern as driver of sustainable mobility and reduced car use. *Transportation Research Part D: Transport and Environment*, *134*, 104345.

Nghiem, L. T., Papworth, S. K., Lim, F. K., & Carrasco, L. R. (2016). Analysis of the capacity of Google Trends to measure interest in conservation topics and the role of online news. *PloS one*, *11*(3), e0152802.

Owen, A. L., & Wei, A. (2021). Sexism, household decisions, and the gender wage gap. *Labour Economics*, *72*, 102062.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, *118*, 26-40.

Pan, Y., & He, S. Y. (2023). An investigation into the impact of the built environment on the travel mobility gap using mobile phone data. *Journal of Transport Geography*, *108*, 103571.

Saputra, D. M., Saputra, D., & Oswari, L. D. (2020, May). Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method. In *Sriwijaya international conference on information technology and its applications (SICONIAN 2019)* (pp. 341-346). Atlantis Press.

Turan, B., Hemmelmayr, V., Larsen, A., & Puchinger, J. (2024). Transition towards sustainable mobility: the role of transport optimization. *Central European Journal of Operations Research*, *32*(2), 435-456.

United Nations Human Settlements Programme. (2020). *World cities report 2020: The value of sustainable urbanization*. UN-Habitat.

Wang, R., Zhang, X., & Li, N. (2022). Zooming into mobility to understand cities: A review of mobility-driven urban studies. *Cities*, *130*, 103939.

Winkler, L., Pearce, D., Nelson, J., & Babacan, O. (2023). The effect of sustainable mobility transition policies on cumulative urban transport emissions and energy demand. *Nature Communications*, *14*(1), 2357.

**Figure 1. Sample variation in Google search data**

Imagen que contiene Gráfico

El contenido generado por IA puede ser incorrecto.

Note: Each line represents one of 50 repeated extractions from Google Trends. Panel (a) shows the raw query, (b) the control term (“google”), and (c) the combined query used for imputation. Values are smoothed using a 3-period moving average

.

**Figure 2. Original and imputed Google Trends values**

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Note: The red dashed line represents the imputed series obtained through regression on control term queries. The gray lines show 50 repeated extractions for the same search term, *“supermarket near me”*, in Sweden. Imputation stabilizes the signal by correcting sample variability. Values are smoothed using a 3-period moving average.

**Figure 3. Z-score normalization and STL-based deseasonalization**

Gráfico

El contenido generado por IA puede ser incorrecto.

Note: This figure compares two preprocessing techniques applied to the Google Trends signal for the query *“supermarket near me”* in Sweden. The dashed black line shows the Z-score normalized series; the solid blue line represents the deseasonalized signal extracted via STL decomposition. Both series are centered around zero, but STL removes seasonal components, highlighting longer-term dynamics.

**Figure 4. Scree plot of principal components**

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Note: The plot shows the eigenvalues of the first 15 principal components based on monthly search volumes for 16 mobility-related keywords. According to the Kaiser criterion (λ > 1), only the first component is retained, explaining most of the total variance.

**Figure 5. Contribution of each keyword to the first principal component**

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Note: The barplot shows the loadings (weights) of each keyword on the first principal component (PC1). Negative loadings correspond primarily to car-centric terms (e.g., *car*, *traffic*), while positive values reflect proximity or digital alternatives (e.g., *supermarket*, *remote work*). PC1 is interpreted as a synthetic indicator of sustainable mobility interest.

**Figure 6. Elbow method for Optimal k number of clusters**

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Note: Figure 1 applies the elbow method to identify the optimal number of clusters for the Sustainable Mobility Index (SMI). We find a slight inflection point appears at k=3, indicating that additional clusters would yield diminishing returns in reducing within-cluster variance.

**Figure 7 – Evolution of Sustainable Mobility Index by Cluster**

Gráfico, Gráfico de líneas, Histograma

El contenido generado por IA puede ser incorrecto.

Note: Figure 7 illustrates the temporal evolution of the SMI across the three identified clusters. Despite common normalization around zero, temporal patterns diverge significantly, especially during major global events such as the COVID-19 outbreak or energy crises.

**Figure 8. Cluster validation and group-specific distribution of the Sustainable Mobility Index (SMI)**

Gráfico

El contenido generado por IA puede ser incorrecto.

Panel (a) shows silhouette coefficients for each country using a k-means clustering solution with *k = 3*. The red dashed line indicates the average silhouette score (0.045), suggesting moderate separation across clusters. Panel (b) displays the distribution of the Sustainable Mobility Index (SMI) by cluster, using violin plots with embedded boxplots. Cluster 0 (blue) groups mostly Nordic countries, cluster 1 (orange) includes Western and Southern countries, and cluster 2 (green) corresponds to Central Europe. The three groups exhibit distinct median levels of interest in sustainable mobility.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Cluster** | **Countries** | **SMI (Median)** | **SMI Std.** |
|  |  |  |  |
| 0 | Denmark (DK), Finland (FI), Norway (NO), Sweden (SE) | -0.328 | 2.064 |
|  |  |  |  |
| 1 | Belgium (BE), Spain (ES), France (FR), United Kingdom (GB), Ireland (IE), Italy (IT), Portugal (PT), United States (US) | -0.431 | 2.050 |
|  |  |  |  |
| 2 | Austria (AT), Germany (DE), Netherlands (NL) | -0.407 | 1.896 |
|  |  |  |  |
|  |  |  |  |

**Table 1. Country composition of the three clusters based on SMI trends**

Note: Countries are grouped according to the result of a k-means clustering algorithm (*k = 3*) applied to the monthly evolution of the Sustainable Mobility Index (SMI) over the 2018–2025 period. Each cluster exhibits distinct dynamics in digital interest in sustainable mobility.

**Appendix A. Data Download Strategy**

Since Google Trends internally normalizes values and limits the number of queries per IP address, we implemented a data extraction strategy inspired by the method proposed by Stephens-Davidowitz (2014):

For each keyword *x*, we downloaded three distinct time series:

1. searches for *x*
2. searches for a control term
3. searches for *x + control term*

This approach enables subsequent imputation of missing values (i.e., values below the privacy threshold) and facilitates comparisons across terms and countries using regression techniques.

To increase the robustness of the time series, we repeated the download process multiple times, incorporating proxies, random delays, and rotating countries to minimize the risk of query blocking. Data was collected at a monthly frequency between January 2018 and January 2025.

**4. Data Organization**

Each extraction is documented with a sample identifier, country, search term, and conceptual label (x, w, xw). The data are stored in .csv files, organized by country and series type (x, w, xw). This structure enables the construction of comparable panel datasets across countries and supports transparent robustness analysis.

**Normalization of Google Trends series**

 Although the raw Google Trends signal is returned on a nominal 0‑to‑100 scale, this scale is *local to each query, country and time‑window*: a value of 100 merely denotes the highest search interest observed within that individual download, and is therefore not comparable across keywords or across countries. To obtain a metric that is commensurate across all series we first standardised every keyword‑country time‑series to zero mean and unit variance (z‑score), after the sample‑bias imputations described in Section 3. This step removes the arbitrary baseline and equalises the within‑series variance.

**Conceptual Mapping and Keyword Unification**

All localized keywords were mapped to a set of common semantic labels (e.g., *"carro elétrico"*, *"voiture électrique"*, *"electric car"* → electric\_car) using a multilingual mapping dictionary. This allowed us to unify conceptually equivalent queries under a single identifier (keyword\_common) and later aggregate multiple synonyms (e.g., *Uber*, *Cabify*) into composite categories (e.g., ride\_hailing).

**Deseasonalization**

Before applying PCA, we removed deterministic monthly seasonality from each country-keyword series using the **STL decomposition** (Seasonal-Trend Loess). The residual component (devoid of seasonal and trend components) was retained as the de-seasonalized signal. This step prevents the PCA from capturing cyclic fluctuations that do not reflect structural change in interest.

**Appendix B: Keywords used by country**

|  |  |  |
| --- | --- | --- |
| **GB** | **ES** | **PT** |
|  |  |  |
| car | coche | carro |
| taxi | taxi | táxi |
| bicycle | bicicleta | bicicleta |
| bus | autobús | autocarro |
| work from home | teletrabajo | trabalho remoto |
| supermarket near me | supermercado cerca | supermercado perto |
| restaurant near me | restaurante cerca | restaurante perto |
| online shopping | compra online | compras online |
| electric car | coche eléctrico | carro elétrico |
| petrol consumption | consumo gasolina | consumo gasolina |
| hybrid car | coche híbrido | carro híbrido |
| parking | aparcamiento | estacionamento |
| traffic | trafico | trânsito |
| car sharing | car sharing | car sharing |
| uber | uber | uber |
|  | cabify | cabify |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **DK** | **NO** | **IT** |
|  |  |  |
| bil | bil | auto |
| taxa | taxi | taxi |
| cykel | sykkel | bicicletta |
| bus | buss | autobus |
| hjemmearbejde | hjemmekontor | telelavoro |
| supermarked i nærheden | butikk i nærheten | supermercato vicino |
| restaurant i nærheden | restaurant i nærheten | ristorante vicino |
| online shopping | netthandel | acquisti online |
| elbil | elbil | auto elettrica |
| brændstofforbrug | bensinforbruk | consumo benzina |
| hybridbil | hybridbil | auto ibrida |
| parkering | parkering | parcheggio |
| trafik | trafikk | traffico |
| delebil | bildeling | car sharing |
| uber | uber | uber |
|  |  | cabify |
|  |  |  |
| **IE** | **SE** | **BE** |
|  |  |  |
| car | bil | auto |
| taxi | taxi | taxi |
| bicycle | cykel | fiets |
| bus | buss | bus |
| remote work | distansarbete | thuiswerken |
| supermarket near me | mataffär nära mig | supermarkt in de buurt |
| restaurant near me | restaurang nära mig | restaurant in de buurt |
| online shopping | online shopping | e-commerce |
| electric car | elbil | elektrische auto |
| fuel consumption | bensinförbrukning | benzine verbruik |
| hybrid car | hybridbil | hybride auto |
| parking | parkering | parking |
| traffic | trafik | trafic |
| car sharing | bilpool | car sharing |
| uber | uber | uber |
|  |  | cabify |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **FI** | **NL** | **AT** |
|  |  |  |
| auto | auto | auto |
| taksi | taxi | taxi |
| pyörä | fiets | fahrrad |
| bussi | bus | bus |
| etätyö | thuiswerken | homeoffice |
| lähikauppa | supermarkt in de buurt | supermarkt in der nähe |
| ravintola lähellä | restaurant in de buurt | restaurant in der nähe |
| verkkokauppa | online winkelen | online einkaufen |
| sähköauto | elektrische auto | elektroauto |
| polttoaineen kulutus | benzineverbruik | benzinverbrauch |
| hybridiauto | hybride auto | hybrid auto |
| pysäköinti | parkeren | parkplatz |
| liikenne | verkeer | verkehr |
| yhteiskäyttöauto | car sharing | car sharing |
| uber | uber | uber |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **FR** | **DE** | **US** |
|  |  |  |
| voiture | auto | car |
| taxi | taxi | taxi |
| vélo | fahrrad | bike |
| autobus | bus | bus |
| télétravail | homeoffice | work from home |
| supermarché  proche | supermarkt in der nähe | grocery store near me |
| restaurant proche | restaurant in der nähe | restaurant near me |
| achat en ligne | online einkaufen | online shopping |
| voiture électrique | elektroauto | electric car |
| consommation essence | benzinverbrauch | gas consumption |
| voiture hybride | hybridauto | hybrid car |
| parking | parkplatz | parking |
| trafic | verkehr | traffic |
| autopartage | carsharing | car sharing |
| uber | uber | uber |
| cabify |  |  |
|  |  |  |

Note: Country abbreviations follow ISO 3166-1 alpha-2 codes.

UK = United Kingdom ES = Spain PT = Portugal DE = Germany FR = France IT = Italy US = United States IE = Ireland SE = Sweden NO = Norway DK = Denmark FI = Finland NL = Netherlands  
AT = Austria BE = Belgium

**Appendix C. Panel data analysis**

To investigate the structural and contextual factors associated with digital interest in sustainable mobility, we estimate a panel data model using the *Sustainable Mobility Index* as the dependent variable. The dataset includes annual observations for a set of countries over the 2018–2023 period.

The model specification takes the following form:

Where the Sustainable Mobility Index for country *i* in year *t*, it​ is a vector of explanatory variables, captures country-specific effects, controls for year-specific shocks, and ​ is the error term.

We consider a set of independent variables drawn from the Avoid–Shift–Improve (ASI) framework:

* **Avoid-related factors**: urban density, average commute times, and telework adoption rates.
* **Shift-related factors**: public transport coverage, cycling infrastructure per capita, and modal split indicators.
* **Improve-related factors**: electric vehicle adoption rates, fuel prices, and carbon intensity of transport.

In addition, we include controls for socioeconomic development (e.g., GDP per capita, education level), digital infrastructure (e.g., internet penetration), and environmental concern (e.g., national climate policy scores or environmental attitudes, where available). Depending on the outcome of the Hausman test, we estimate the model using fixed or random effects. Year fixed effects are included to account for common global shocks (e.g., COVID-19, energy crisis). This modeling approach allows us to examine whether structural, normative, or policy-related factors are systematically associated with a higher (or lower) level of digitally expressed interest in sustainable urban mobility.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table C1. Candidate Explanatory Variables** | | | | | |
| **Description** | **Source** | **Freq** | **Countries** | **Date accessed** | **Notes/Treatment** |
|  |  |  |  |  |  |
| Sustainable mobility index | Google Trends | monthly | DE, FR, GB, IT, PT, US, ES |  |  |
| Consumer Price Index (all units) | OECD | monthly | DE, FR, GB, IT, PT, US, ES | 21/04/2025 | ISO3 → ISO2 conversion, merged by month, no data for USA 01/2025 |
| CPI (Transport) | OECD | monthly | DE, FR, GB, IT, PT, US, ES | 21/04/2025 | ISO3 → ISO2 conversion, merged by month |
| CPI (Energy) | OECD | monthly | DE, FR, GB, IT, PT, ES | 21/04/2025 | ISO3 → ISO2 conversion, merged by month, find value for US |
| CPI (US only) | FRED | monthly | US | 21/04/2025 | Value for US energy CPI |
| Diesel retail price (€/L, including taxes) | EU Energy Oil Bulletin | monthly | DE, FR, GB, IT, PT, ES | 21/04/2025 | Aggregated weekly prices to monthly mean |
| Gasoline (Euro95) retail price (€/L) | EU Energy Oil Bulletin | monthly | DE, FR, GB, IT, PT, ES | 21/04/2025 | Aggregated weekly prices to monthly mean |
| Diesel price (€/L, converted from USD/gal) | FRED | monthly | US | 21/04/2025 | Converted from USD/gal to €/L, fixed FX (0.92) |
| Gasoline price (€/L, converted from USD/gal) | FRED | monthly | US | 21/04/2025 | Converted from USD/gal to €/L, fixed FX (0.92) |
| New electric vehicle registrations (BEV + PHEV) | EAFO | monthly | DE, FR, GB, IT, PT, ES | 21/04/2025 | Sum of BEV + PHEV |
| New electric vehicle registrations (US) | Argonne Lab | monthly | US | 22/04/2025 | Parsed from PDF, merged from cleaned CSV |
| % change in retail/recreation activity vs baseline | Google Mobility Reports | monthly | DE, FR, GB, IT, PT, US, ES | 22/04/2025 | Aggregated daily to monthly mean |
| % change in grocery/pharmacy visits vs baseline | Google Mobility Reports | monthly | DE, FR, GB, IT, PT, US, ES | 22/04/2025 | Aggregated daily to monthly mean |
| % change in grocery/pharmacy visits vs baseline | Google Mobility Reports | monthly | DE, FR, GB, IT, PT, US, ES | 22/04/2025 | Aggregated daily to monthly mean |
| % change in public transport station activity vs baseline | Google Mobility Reports | monthly | DE, FR, GB, IT, PT, US, ES | 22/04/2025 | Proxy for public transport use |
| Monthly unemployment rate (age 15+) | FRED | monthly | DE, FR, GB, IT, PT, US, ES | 22/04/2025 | Seasonally adjusted |
| USD to EUR exchange rate | FRED | monthly | Global | 22/04/2025 | Daily series aggregated to monthly average |
| Harmonized OECD Consumer Confidence Index | OECD | monthly | DE, FR, GB, IT, PT, US, ES | 23/04/2025 | Amplitude adjusted |
| Monthly mean temperature (°C) | Own elaboration | monthly | DE, FR, GB, IT, PT, US, ES | 23/04/2025 | Used as seasonal control |
| Population ages 15-64 (% of total) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Labor force proxy |
|  |  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Source** | **Frequency** | **Countries** | **Date accessed** | **Notes/Treatment** |
|  |  |  |  |  |  |
| Female population (% of total population) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Gender composition |
| Population aged 65 and above (% of total) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Aging population indicator |
| Gini index (0 = perfect equality, 100 = perfect inequality) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Income inequality |
| Gross enrollment ratio, tertiary education (% of population of official age) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Education control |
| CO2 emissions from transport (Mt CO2e, AR5 methodology) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Mobility-related emissions |
| N2O emissions from transport (Mt CO2e, AR5 methodology) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Mobility-related emissions |
| Age dependency ratio (% of working-age population) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Demographic burden |
| Aggregated from World Bank population by age (% by group: 0–14, 15–64, 65+) | World Bank | yearly | DE, FR, GB, IT, PT, US, ES | 25/04/2025 | Simplified broad age groups |
|  |  |  |  |  |  |