

Matching User Preferences and Behavior for Mobility

Silvia Basile^{1,*}, Cristian Consonni²,
Matteo Manca², Ludovico Boratto²

¹ C3.ai, Rome, Italy - [c3.ai](https://www.c3.ai)

² Eurecat - Centre Tecnològic de Catalunya, Big Data & Data Science Unit, Barcelona, Spain - [eurecat.org](https://www.eurecat.org)

* This work was developed while the author was working at Eurecat

MoTiV: *M*obility and *T*ime Value

MoTiV (**M**obility and **T**ime **V**alue) is an EU Horizon 2020 project.

Goals

- Introduce and validate a conceptual framework for the estimation of Value of Travel Time (VTT)
- explore the dynamics of individual preferences, behaviors and lifestyles that influence travel and mobility choices
- address VTT from the perspective of personality, preferences, needs and expectations of the single individual, in contrast with the traditional viewpoint of the economic dimension (time and cost savings).
- contribute to **advance research on VTT** by introducing a conceptual framework for the estimation of VTT at an individual level based

Woorti App

Data have been collected through the *Woorti* app:

- **8 months**, May 1st, 2019 - December, 13th 2019
- Data collection campaigns from **10 European countries**
- App translated in 12 languages
(cat, dut, eng, fin, fre, hrv, ita, nob, por, slo, spa)



Woorty: Data Collection Phases



Onboarding

Gather user travel preferences:

- Generic worthwhileness values
- Specific worthwhileness values

(+ optional demographics data)



Trip recording

Collect data in background when a user starts the trip



Trip validation

Review and validate the trip when it is finished, preserves user privacy.

Scope, Goals and Idea

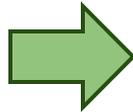
Scope and Goals

- Characterize the perception users have of their mobility.
- Study the alignment between users' preferences and their actual usage of transports.
- Offer intelligent solutions.

Idea



Analyze preferences **before** users' travels, expressed in terms of transport modes.



Monitor the actual **behaviour** of users and perceptions registered **after** their travel.



Contributions

- Build user's profile considering preferences expressed explicitly and the monitored user behavior.
- Cluster user data obtaining an aggregated portrayal of users.
- Matching between the preference-based profiles and those created considering the actual user behavior.
- Characterize user mobility in Europe during an extended period of time.

Value of Travel Time

- **Traveler's perspective:** time and cost savings are not the main factors influencing route and transport mode choices.
- Other factors can have an impact (environmental impact, comfort, weather conditions...)
- Travel time can be *“worth it”* and can be allocated for activities that users find useful, enjoyable or productive.
- **Worthwhile time** is independent of what can be monetized.

Worthwhileness Elements

Characterize the travel time considering three factors called **worthwhileness values**:

- **Productivity**: captures how much the user values the possibility of using travel time to complete some tasks, either personal or work-related
- **Enjoyment** is related to how the travel can be used for fun or relaxing activities
- **Fitness**: measures how much the users value the fact that when traveling they can exercise

Productivity



Pleasure



Fitness



Clustering

User modelings: each user is represented by a vector of 18 elements.

(3 generic worthwhileness values and 15 specific ones (3 for each of the 5 transport categories)).

- Onboarding profile
- Trip profile

Dimensionality reduction and clustering.

- UMAP (Uniform Manifold Approximation and Projection) algorithm for dimensionality reduction (2 dimensions used)
- Hierarchical clustering:
 - 5 clusters
 - Ward's method for linkage criterion and Euclidean distance

Cluster Matching

Cluster distance measures: how to match onboarding and trip profiles.

- **Symmetric difference Distance**

$$d_{\Delta}(B_i, T_j) = |B_i \cup T_j| - |B_i \cap T_j|$$

- **Jaccard Distance**

$$d_J(B_i, T_j) = \frac{|B_i \cup T_j| - |B_i \cap T_j|}{|B_i \cup T_j|}$$

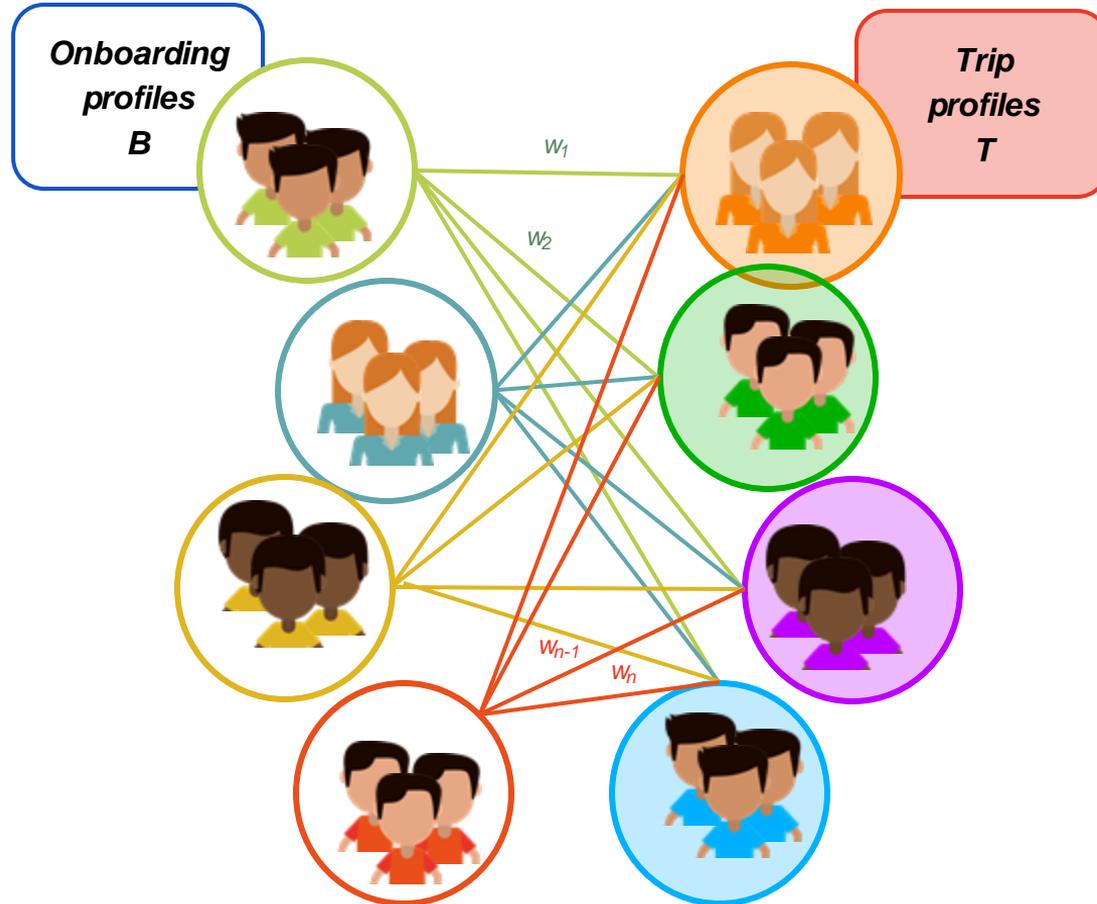
- **Centroid Distance**

$$d_C(B_i, T_j) = \|\mathbf{c}_{\mathcal{B}_i} - \mathbf{c}_{\mathcal{T}_i}\|$$

Matching Clusters

Cluster matching:

- Bipartite graph
- Hungarian algorithm



Cluster Matching: Results

Symmetric Difference Distance

$B \setminus T$	1	2	3	4	5
1	1460	1247	945	1241	1109
2	1186	757	907	735	737
3	947	820	1342	1014	1174
4	1213	526	908	734	734
5	1412	771	811	811	547

Jaccard Distance

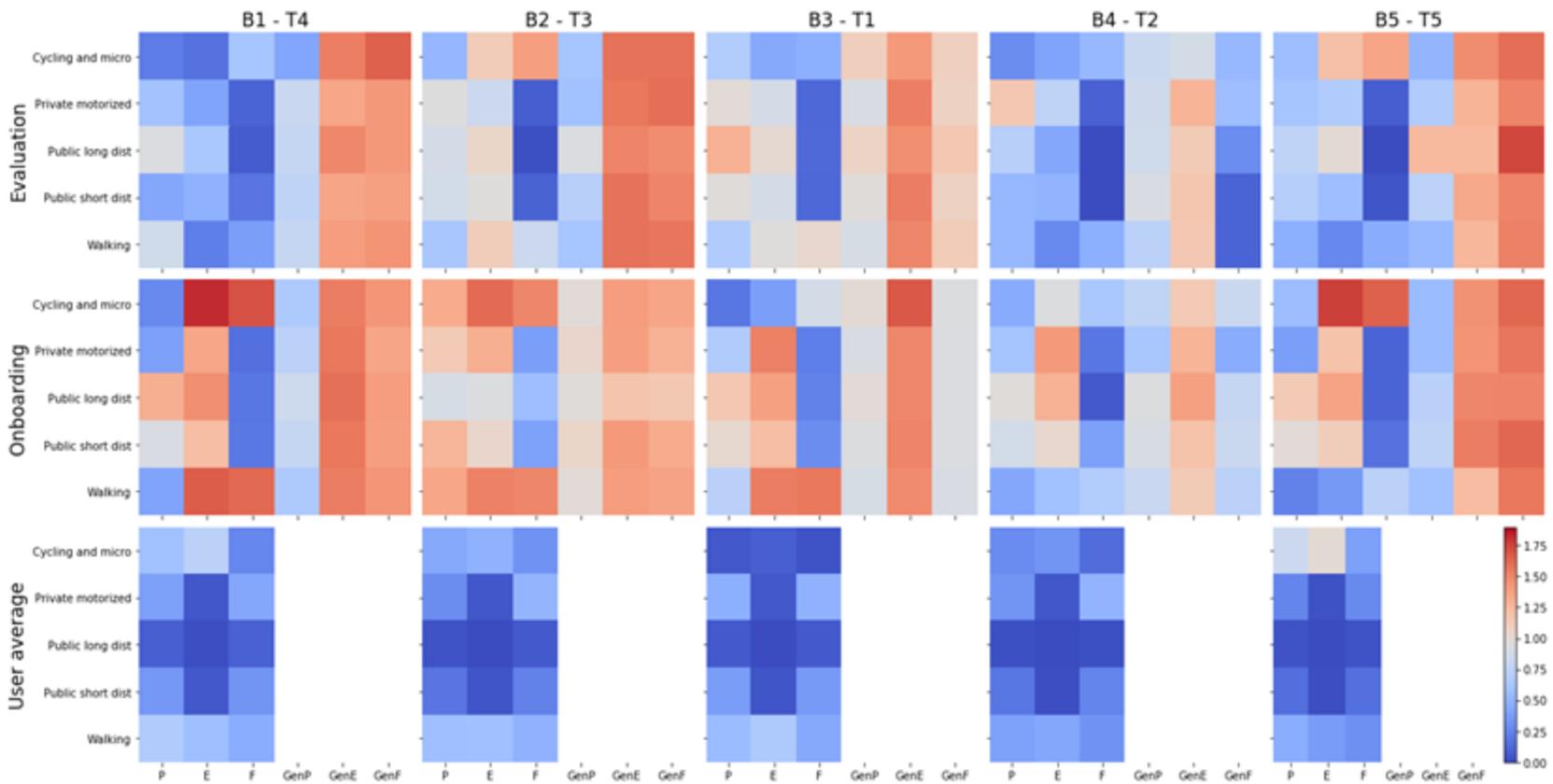
	1	2	3	4	5
1	0.172	0.046	0.266	0.096	0.125
2	0.119	0.033	0.083	0.125	0.081
3	0.318	0.160	0.020	0.112	0.007
4	0.085	0.168	0.050	0.088	0.041
5	0.036	0.029	0.143	0.082	0.232

User Profiles

1. Users with **high** values for fitness and enjoyment while having **low** values for productivity. They are the **active people**, whose preferred modes of transport are comprised in the walking and **running, cycling, and public transport categories**;
2. Users with **medium** values for fitness and productivity but **high** values for enjoyment. They are **active people**, who choose to **walk and cycle, but also use their private car**;
3. Users with medium to **low** values for fitness, **high** enjoyment, and **medium** productivity. These users walk, use their **private car and local public transport**, but they do not cycle or use micromobility modes of transport;
4. Users with medium to **low** values for fitness and productivity, and **medium** values for enjoyment. These users have the overall lowest values in all categories. They mostly use **private cars and public transport**;
5. Users with **high** values for fitness, **medium-to-high** values for enjoyment, and **low** values for productivity. These users use **cycling and new forms of micromobility, together with private cars and public transport**.



Results



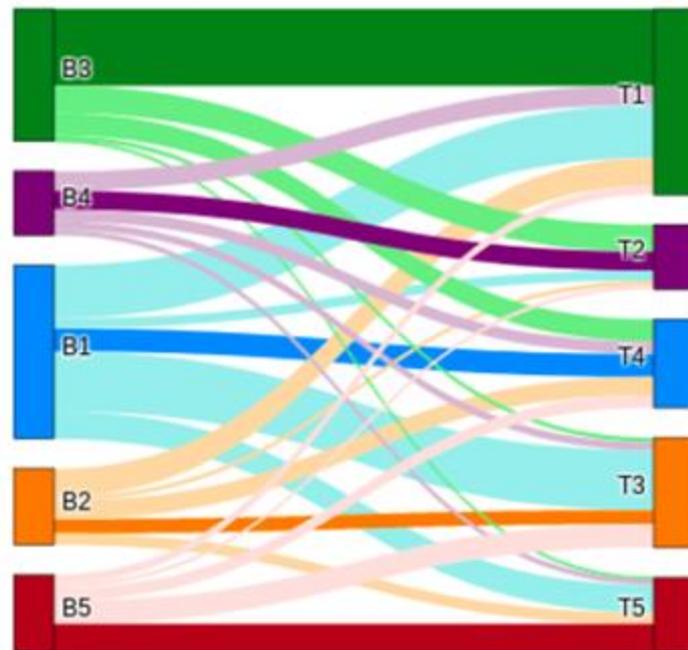
Results: User Migration Between Clusters

Onboarding profiles
 B

Trip profiles
 T

	(B_1, T_4)	(B_2, T_3)	$B_3, T_1)$	(B_4, T_2)	(B_5, T_5)
$B_i \setminus T_j$	865	355	320	262	282
$T_j \setminus B_i$	376	552	627	264	265
$B_i \cap T_j$	132	82	442	106	165

Only 927 out of 3011 users (**31%**) migrate to their matching profile



Conclusions

- User profile analysis
- Users overestimate their preferences for specific mobility modes
- The matching algorithm reveals that many users end up being characterized in a different way with respect to their intentions and their actual behaviour
- General worthwhileness values are even more affected by this difference.
- Productivity receives lower scores

Future works

- Consider additional features related to user profile (gender, age, provenience...)
- Build trip-ranking algorithms

Thanks for your attention



E-mail: silvia.basile@c3.ai

Twitter:  [@silviabasile](https://twitter.com/silviabasile)



Horizon 2020
European Union Funding
for Research & Innovation

This work received the support of the MoTiV project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 825225



31st ACM Conference on Hypertext and Social Media



Backup slides



MoTiV Project Consortiums Partners



University of Žilina (Project Coordinator)
Slovakia



INESC ID
Portugal



Eurecat - Centre Tecnològic de Catalunya
Spain



TIS.pt
Portugal



CoReorient
Finland



European Cyclists' Federation (ECF)
Belgium



routeRANK
Switzerland

Tools and Methodology

- Python  python™



- Pandas, Scipy, Numpy



- Matplotlib , Seaborn , Plotly



On-boarding Data

Name	Description and Admissible values
Generic worthwhileness (F, E, P)	Overall evaluation of how much fitness, enjoyment, and productivity matter for a user's travel experiences ([0-100])
Specific worthwhileness (F, E, P)	Evaluation for each preferred mode of transport of how much fitness, enjoyment, and productivity matter for a user's travel experiences using that mode of transport ([0-100])
Gender	Male, Female, Other
Education level	Basic, High school, or University
Language	Catalan, Dutch, English, Finnish, French, Italian, Norwegian, Portuguese, Slovak, Spanish and Croatian
Age range	16-19, 20-24, 25-29, 30-39, 40-49, 50-64, 65-74, 75+
Marital Status	Divorced, Married, Registered partnership, Single, Widowed
Labour Status	Employed full-time, Employed part-time, Pensioner, Student, Unemployed

Trip/Leg Data

Name	Description and Admissible values
Trip id	Identifier of the trip the leg refers to
User id	Identifier of the who performed the travel leg
Leg duration	Leg travel time expressed in minutes
Leg distance	Leg travel distance expressed in meters
Mode of transport	Mode of transport used; there are 37 different modes of transport
Transport category	Transport category associated to the transport mode; there are 5 different categories (Cycling and emerging micromobility, Private motorized, Public transport short distance, Public transport long distance, Walking and running)
Worthwhileness evaluation	Worthwhileness evaluation of the trip Worthwhileness elements are expressed on a scale from low to high (low, medium, high), while the overall worthwhileness rating for a trip is on a scale from 1 (low) to 5 (high)
Mood evaluation	Mood evaluation for a trip. Mood relates to the overall question "How did you feel about this trip?". The mood rating for a trip is given on a scale from 1 (low) to 5 (high).

User and Transport Categories

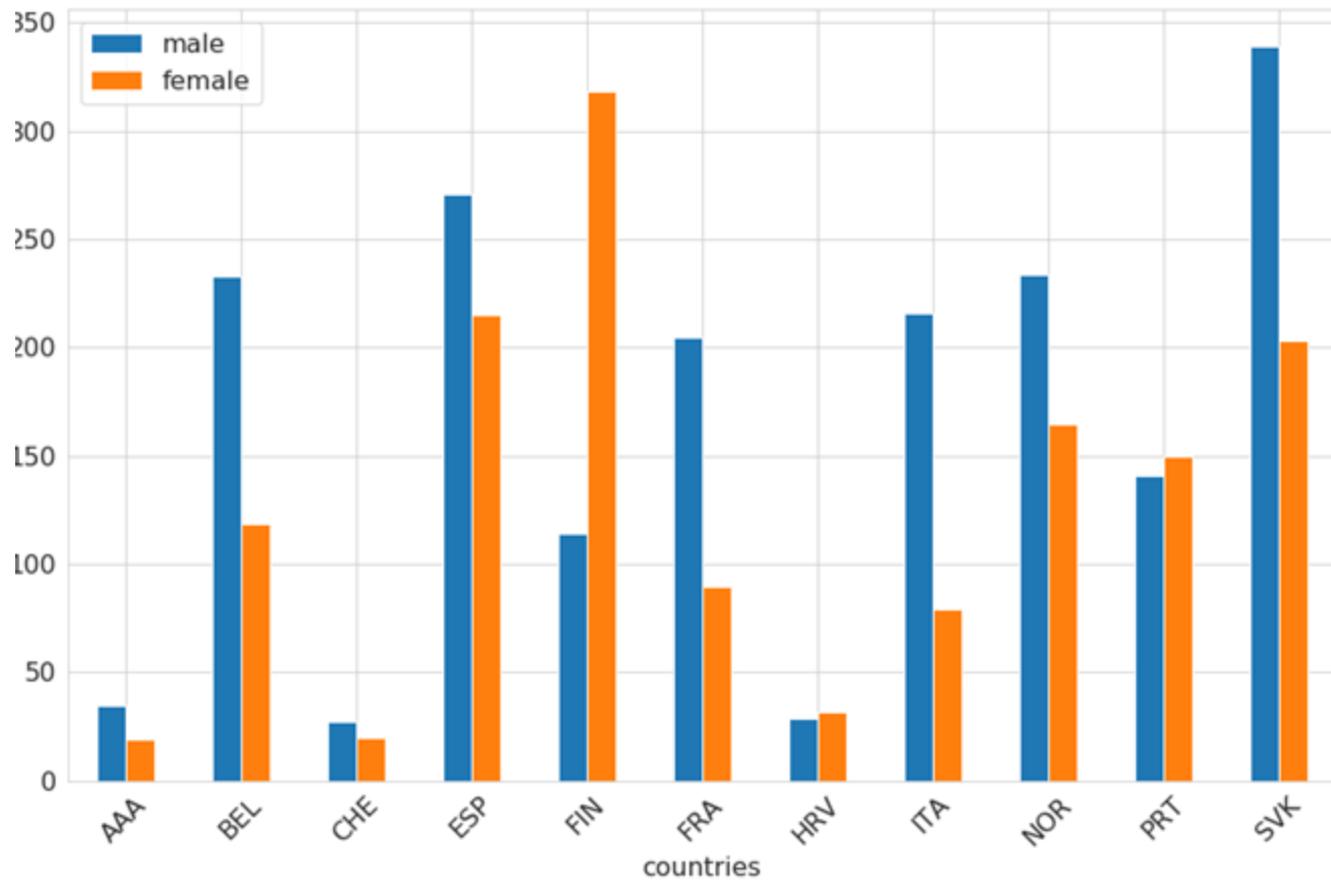
	Users	Trips	Legs
Male	1,844	38,981	98,321
Female	1,410	24,986	60,173
Other	15	131	403
Total	3,269	64,098	158,897

Number of users, number of trips and legs validated by gender

Category	Trips		Users	
	#	<i>f (%)</i>	#	<i>f (%)</i>
Cycling and Micromobility	11,023	21.99	1,317	43.74
Private Motorised	15,003	29.93	1,741	57.82
PT long-distance	481	0.96	189	5.28
PT short-distance	6,097	12.16	1,211	40.22
Walking	17,529	34.96	2,124	70.54
Total	50,133	100	6,582	-

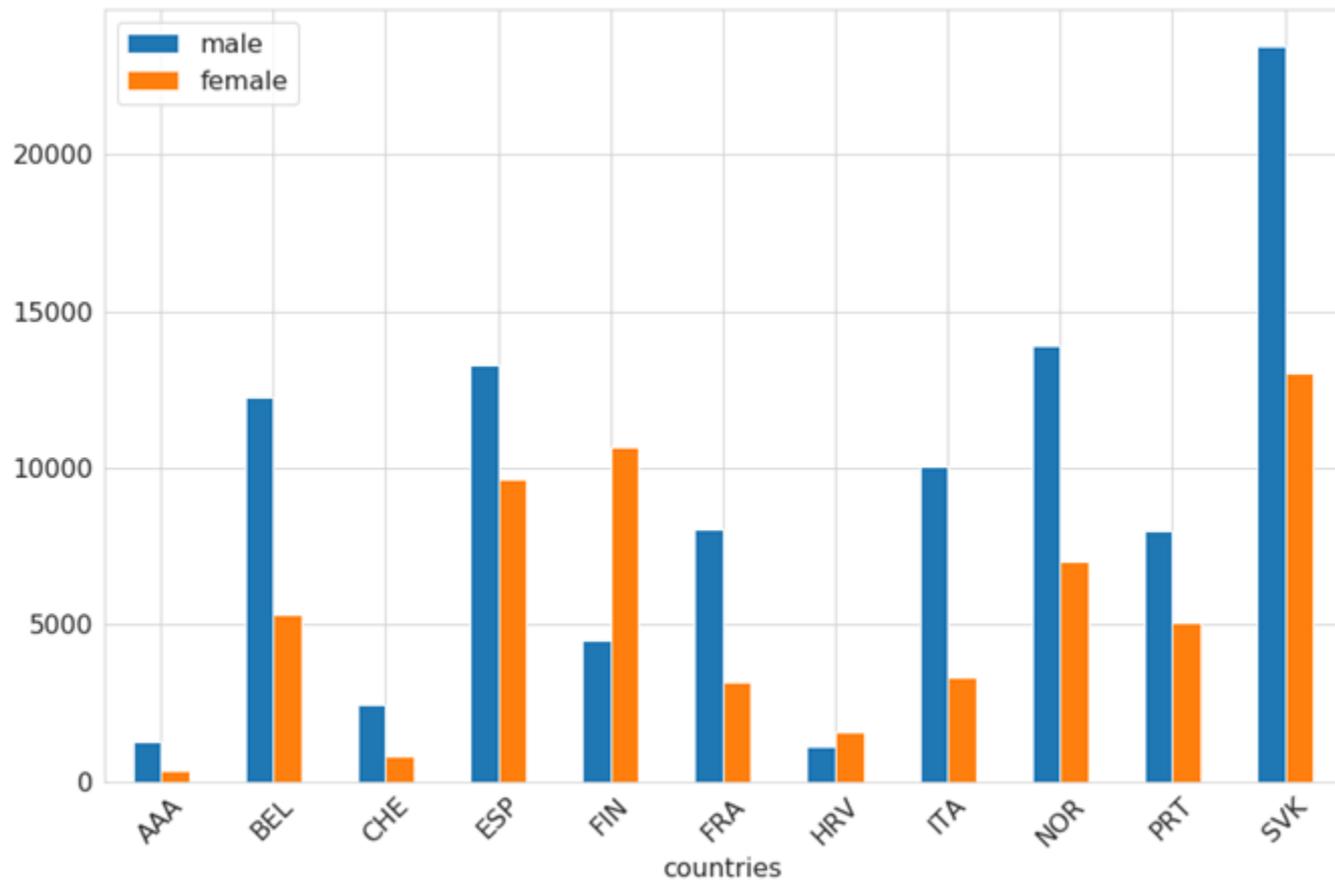
Number (#) and fraction expressed as a percentage (f) of trips by transport category (trips) and users with at least one trip in that category (Users)

Gender Distribution by Country (Users)



Number of users by gender and country: AAA – other, BEL – Belgium, HRV – Croatia, FIN – Finland, FRA – France, ITA – Italy, NOR – Norway, PRT – Portugal, SVK – Slovakia, ESP – Spain)

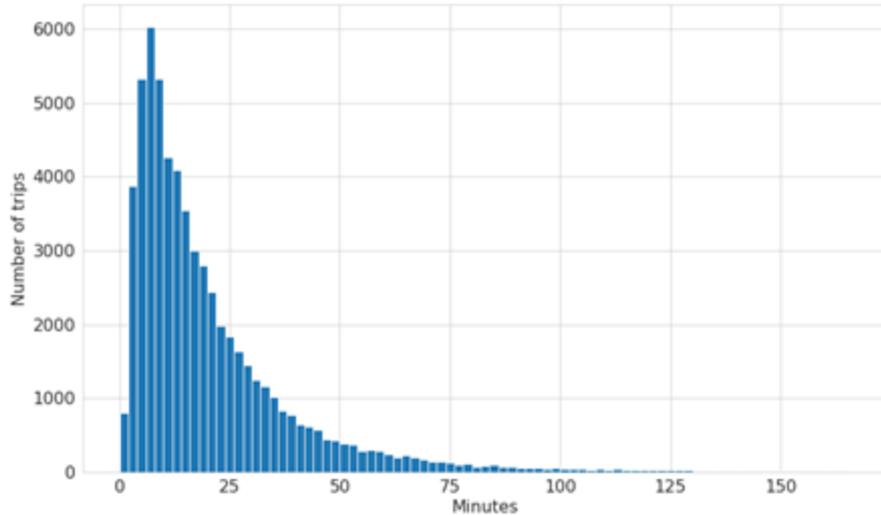
Gender Distribution by Country (Legs)



Number of trip legs by gender and country (AAA – other, BEL – Belgium, HRV – Croatia, FIN – Finland, FRA – France, ITA – Italy, NOR – Norway, PRT – Portugal, SVK – Slovakia, ESP – Spain)

Distribution of Leg Duration and Length

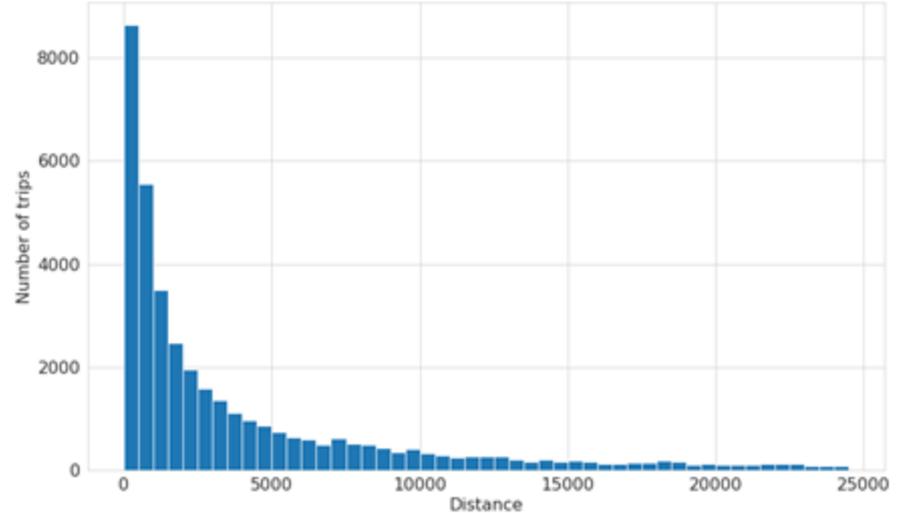
Histogram of total travel time



Distribution of duration of travel time (in minutes) for all trip legs

- Leg median duration: ~14 min
- Leg mean duration: ~21 min

Histogram of total travel distance



Distribution of distance travelled (in meters) for all trip legs

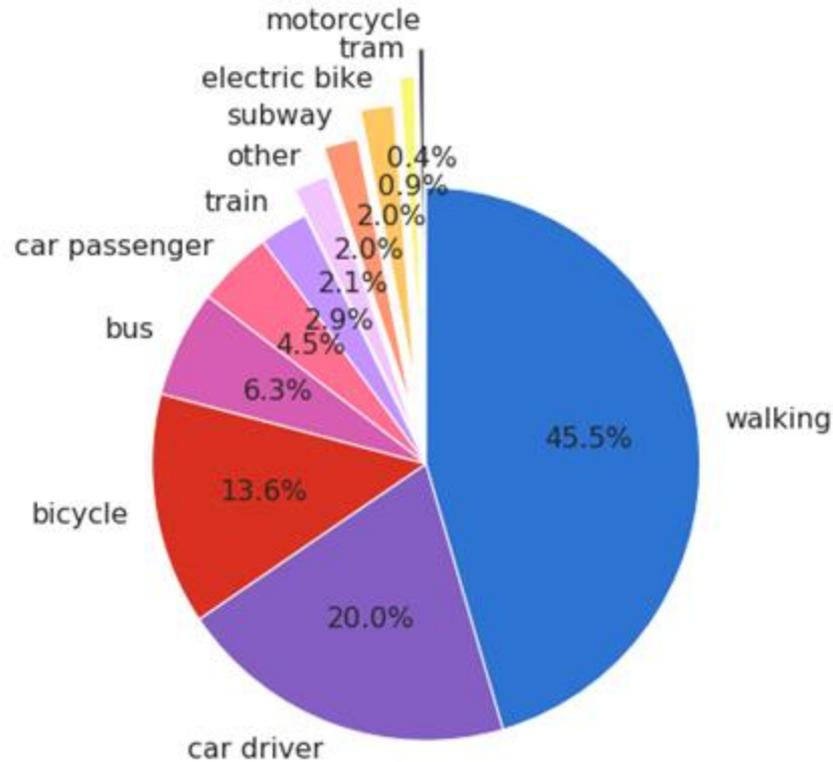
- Leg median length: ~2.1 km
- Leg mean length: ~8.3 km

MoTiV Dataset

- Publication of the anonymized dataset as Open Data on Zenodo
- Publication of the code used for the analysis with an Open License on GitHub

The Zenodo logo consists of the word "zenodo" in a white, lowercase, sans-serif font, centered on a solid blue rectangular background.The GitHub logo features the Octocat icon, a stylized black silhouette of a cat with a white chest, positioned above the word "GitHub" in a bold, black, sans-serif font, all on a light gray rectangular background.

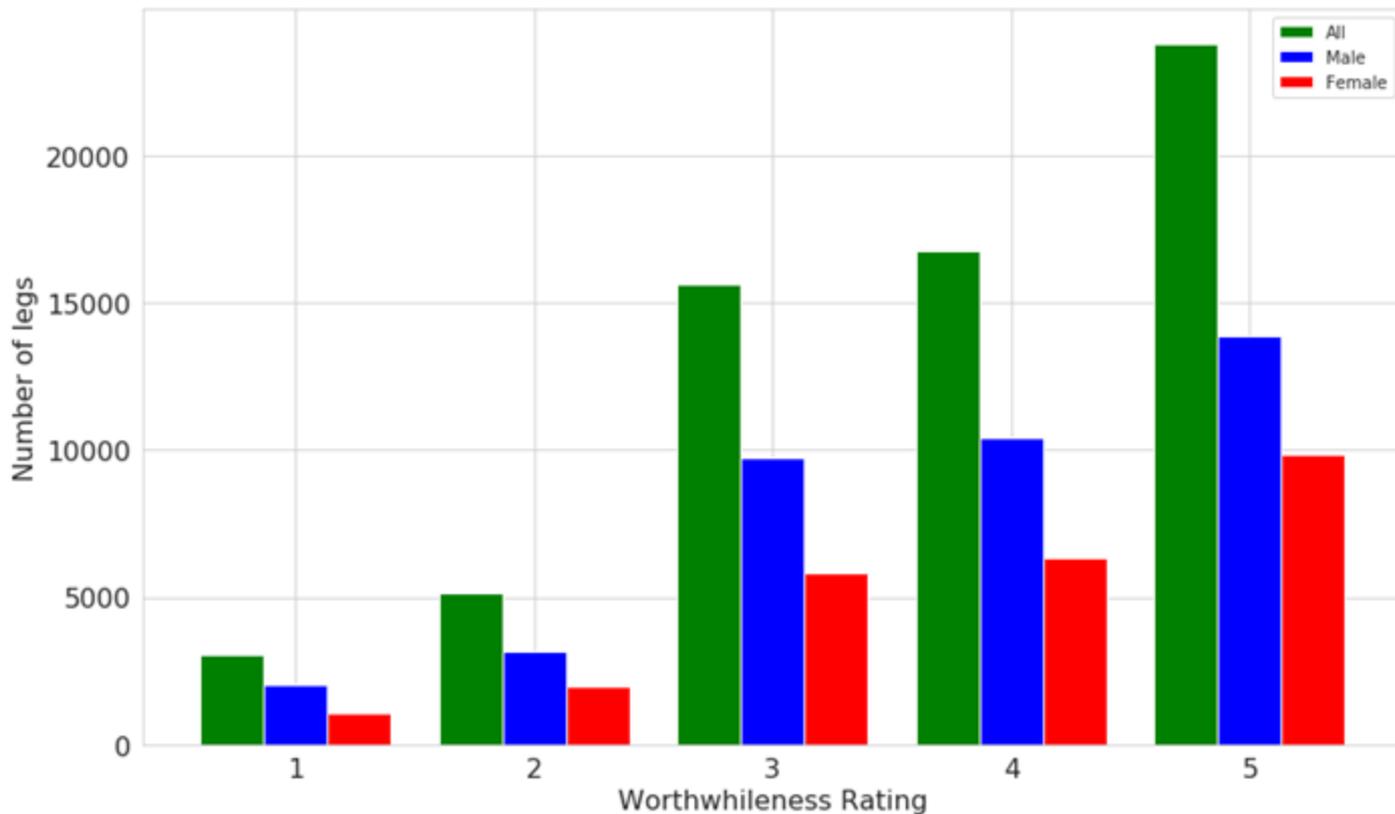
Modal Split



Modal split based on the number of validated trip legs for all transport modes

Distribution of worthwhileness by gender

Distribution of worthwhileness ratings by gender



Distribution of trip legs across worthwhileness ratings by gender