Scraping Urban Mobility

Analysis of Berlin Carsharing

Florian König

Agenda

- Motivation
- Data + where it comes from
- Patterns in demand
- Predicting destinations

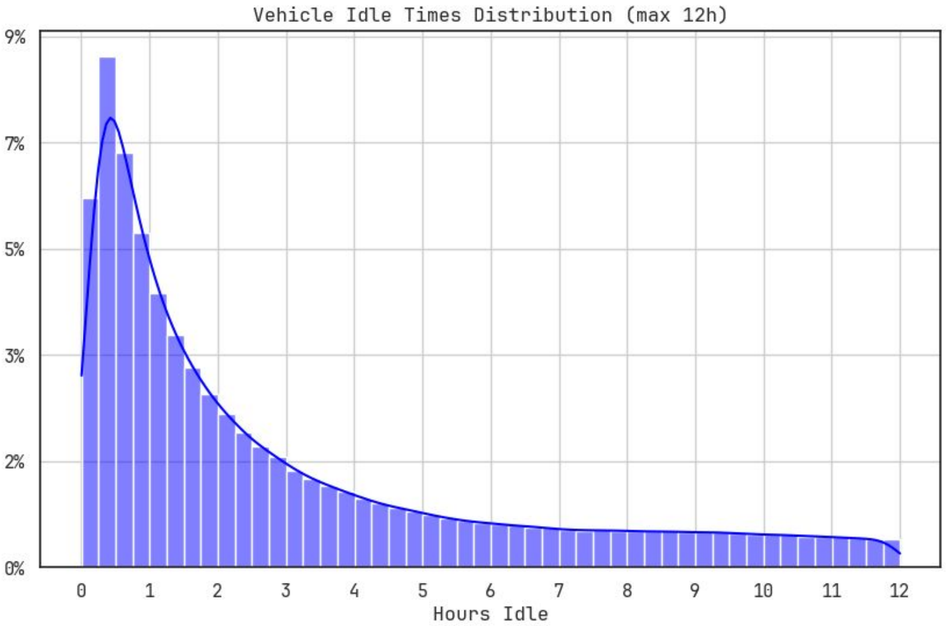
Background

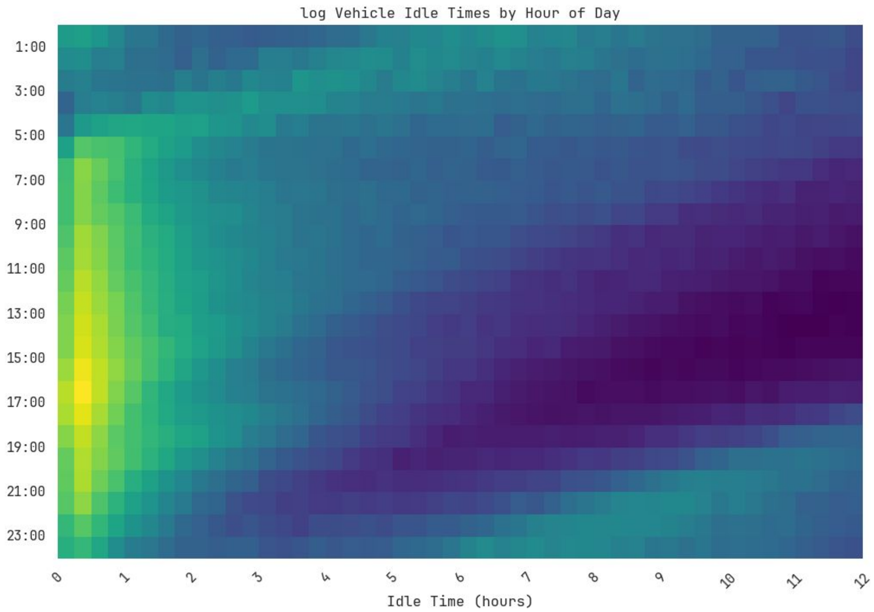
How We Got Here

- We're all nerds
- Human mobility is exciting
- Carsharing contributes to sustainable cities

Background

Idle Vehicles





27min

avg. time per relocation
in Munich

(Weikl, 2016)

~10€*

cost per relocation

* based on average driver income

up to

22.5%

increase in revenue
while improving
fleet balance

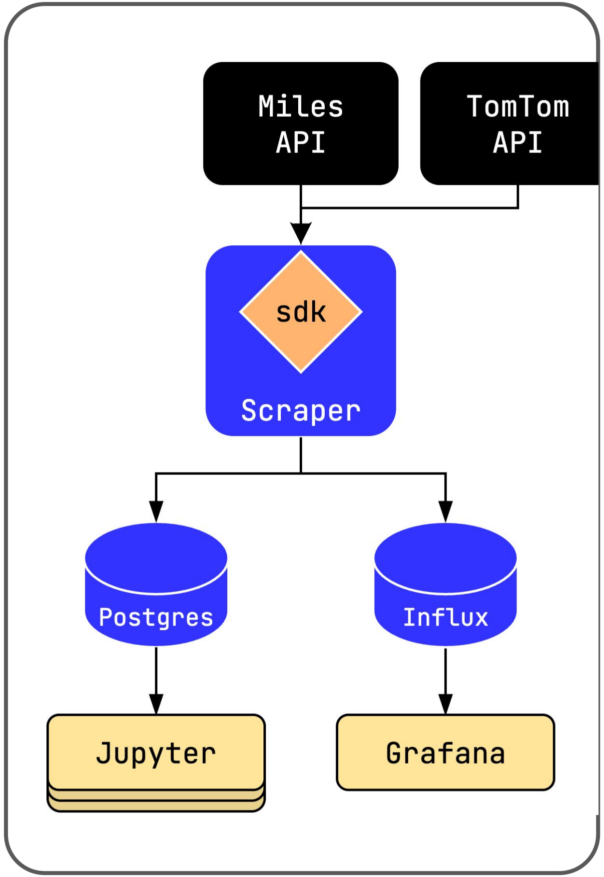
(Wang et al., 2021)

Motivation

Scraping 101

- Scraped map + vehicles

  + data from TomTom, wttr.in
- Up to 2 minute delay
- Clustered in Uber H3
- Observed through Influx



Data

Data Is the New Gold

1.18M trips

5.2M waypoints

6.5K cars

877 vans

1.4M reservations

1.9M discounts

42K POIs

weather + traffic
every 20min

Data

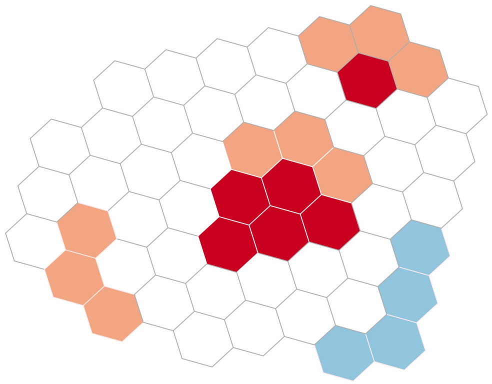
Privacy

- No user identifiers
- Tracking trips in real-time and in history
- Exposing some data is necessary

Demand Prediction

Hypothesis

Areas of **high demand are predictable** by temporal, spatial, and contextual factors.

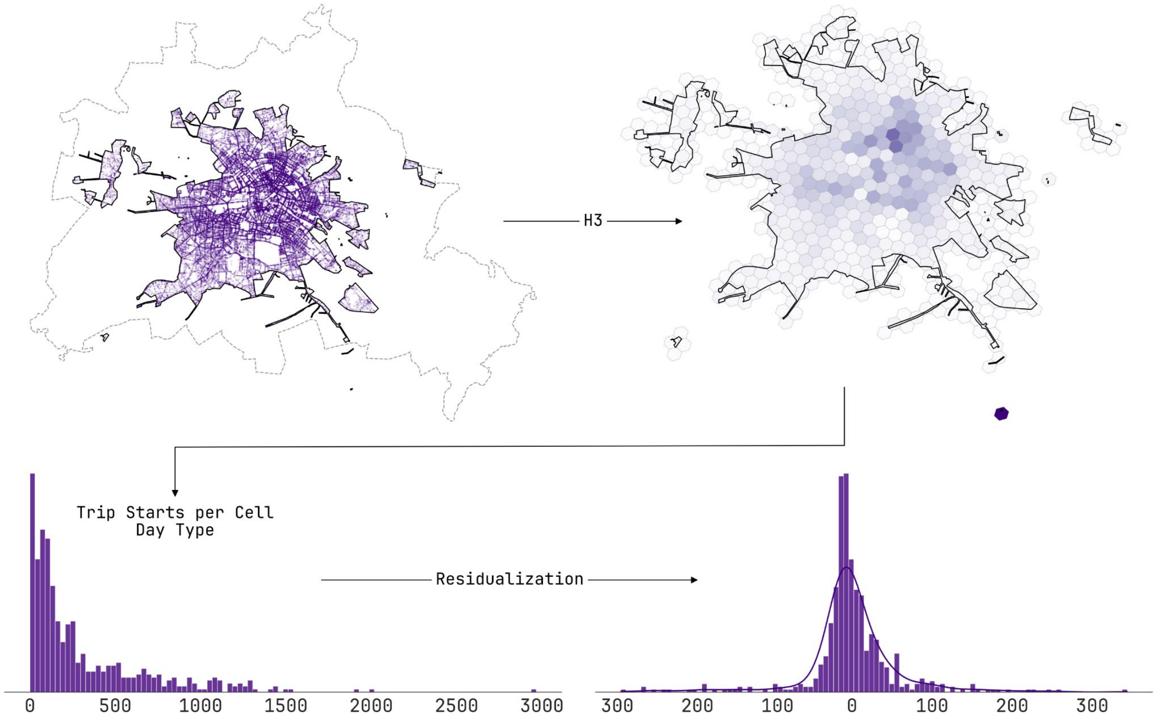


“ carsharing is a complex business and humans can anticipate certain effects [...] before they are reflected in the data ”

(Wagner et al., 2015)

Demand Patterns

Accounting for Urban Density

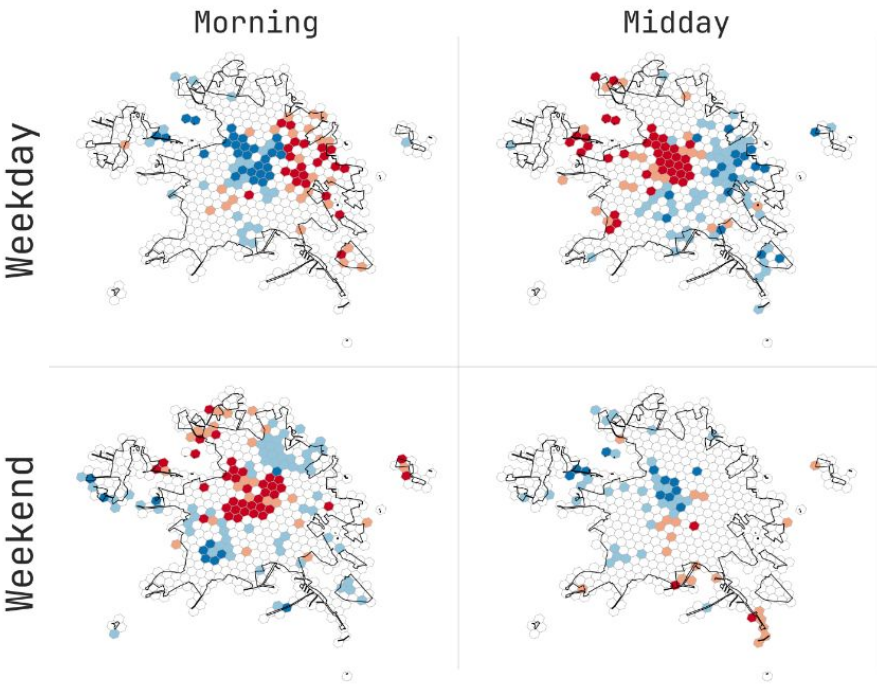


Demand Patterns

Getis-Ord Gi*

“Clusters of hot and cold spots”

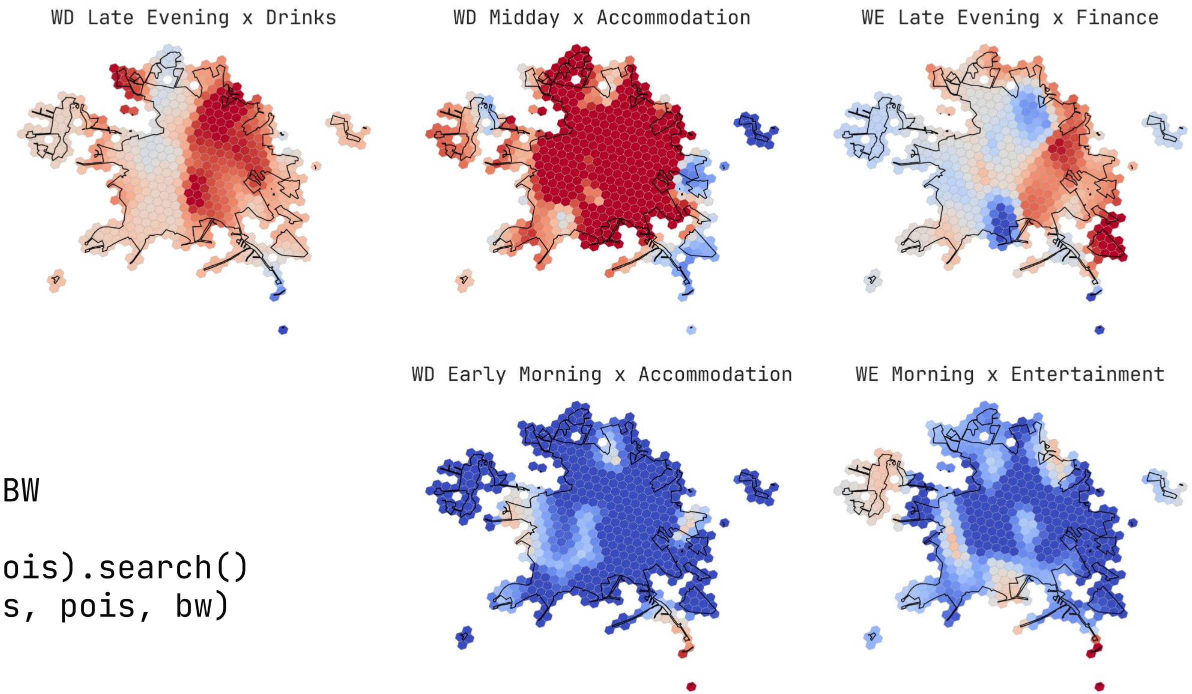
```
from esda import G_Local

g = G_Local(trips, weights, star=True)
```



Demand Patterns

Geographically Weighted Regression

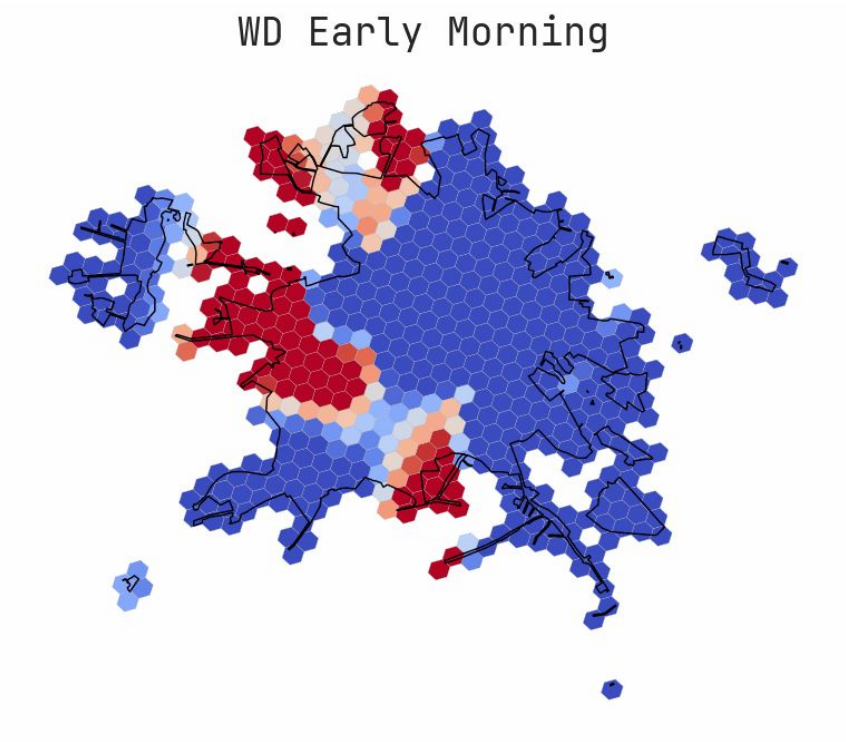
“Regression, but one for each spatial class.”

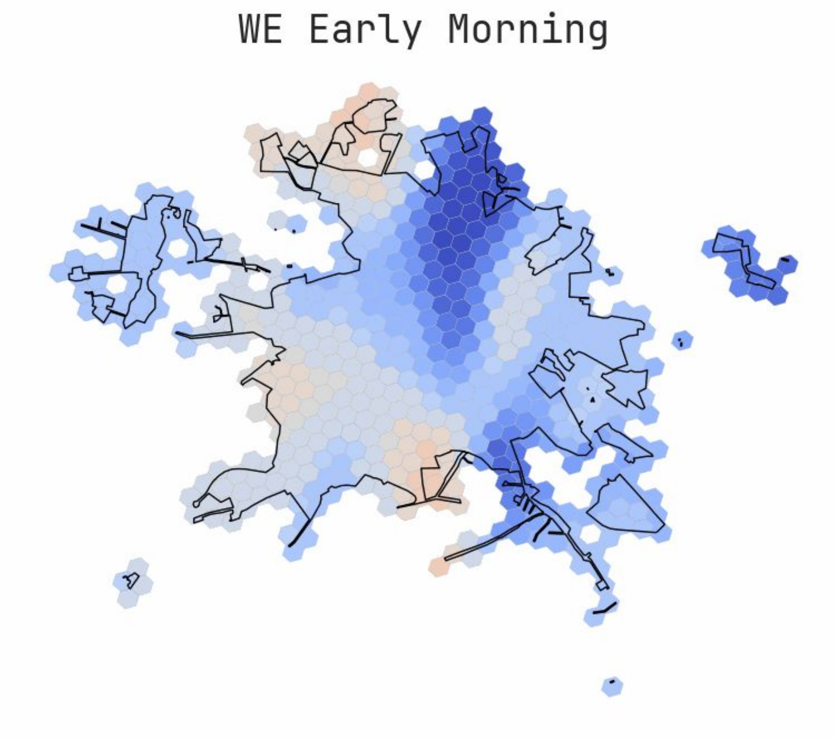



```
from mgwr.gwr import GWR
from mgwr.sel_bw import Sel_BW

bw = Sel_BW(coords, trips, pois).search()
gwr_model = GWR(coords, trips, pois, bw)
```

Demand Patterns

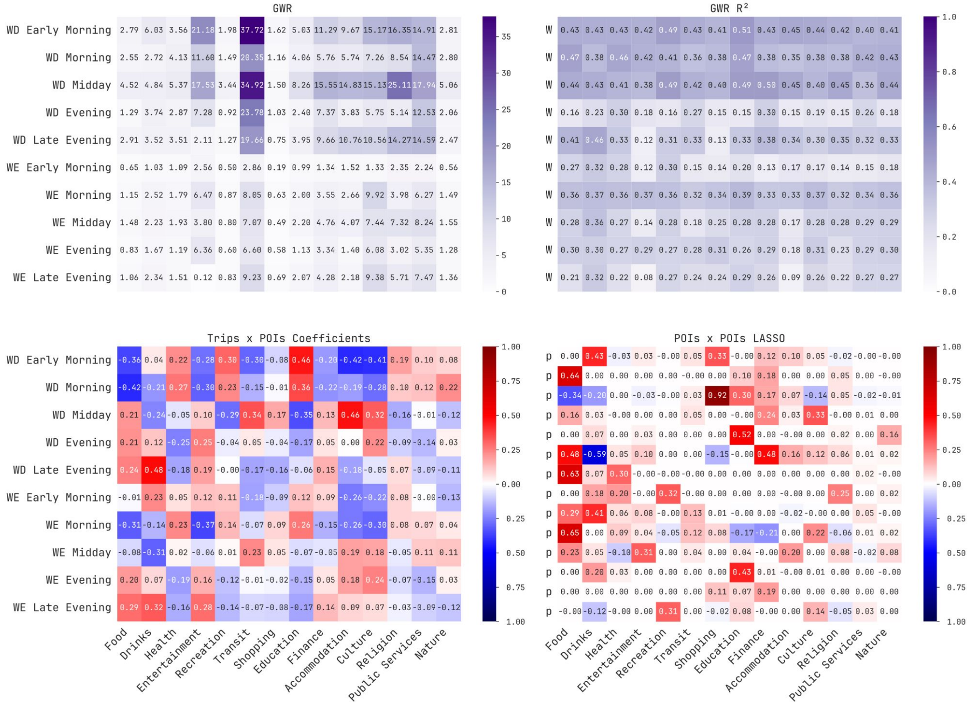
GWR on Transit POIs





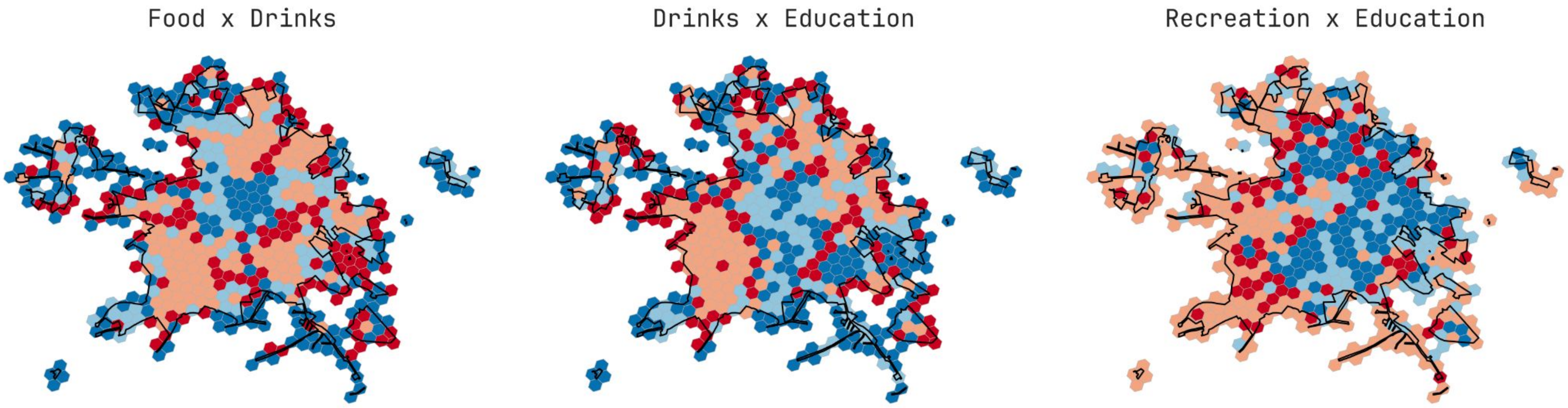
Demand Patterns

POI Colocation



Demand Patterns

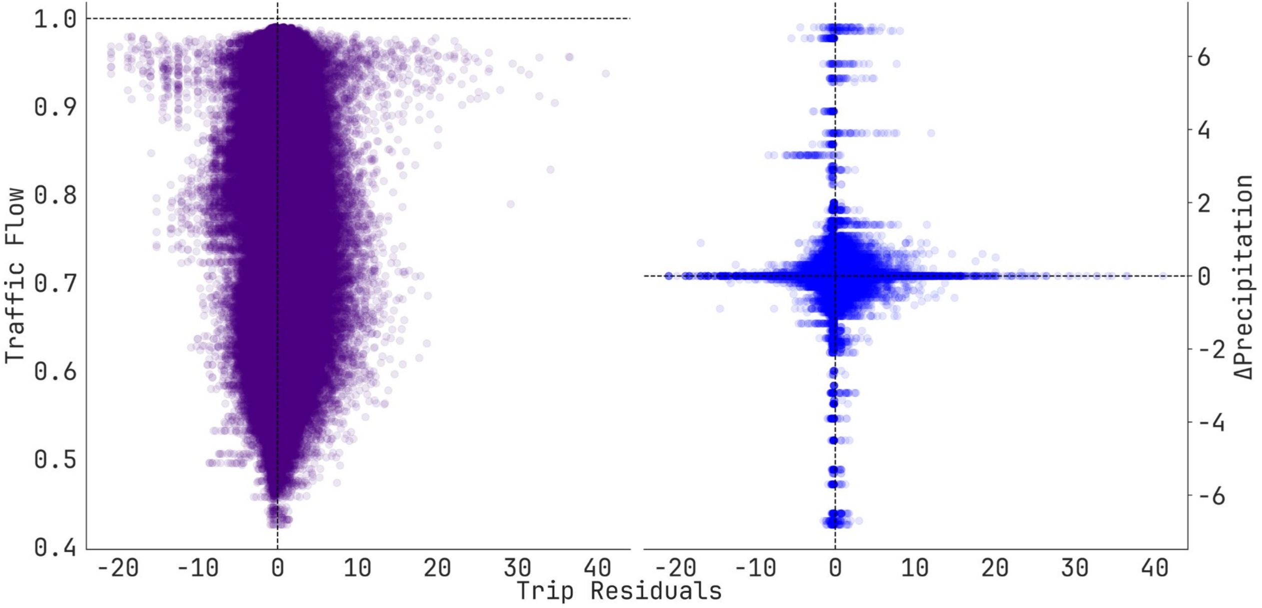
Bivariate Local Moran's I

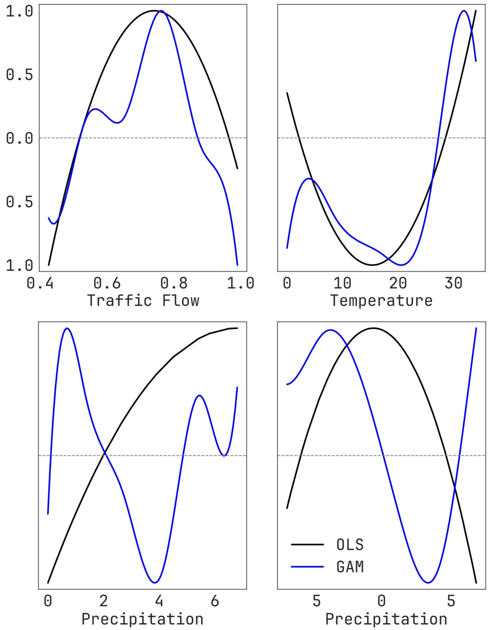



```
from esda.moran import Moran_Local_BV

moran = Moran_Local_BV(base, compare, weights)
```

Demand Patterns

Real-Time Impacts

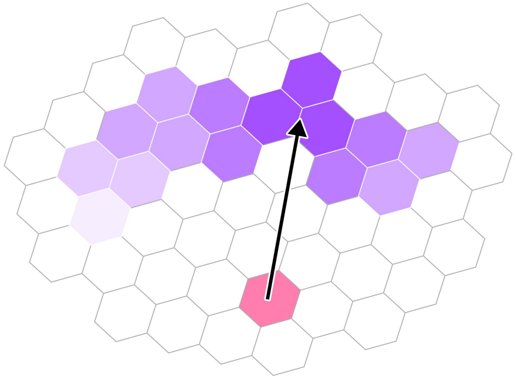




Destination Prediction

Hypothesis

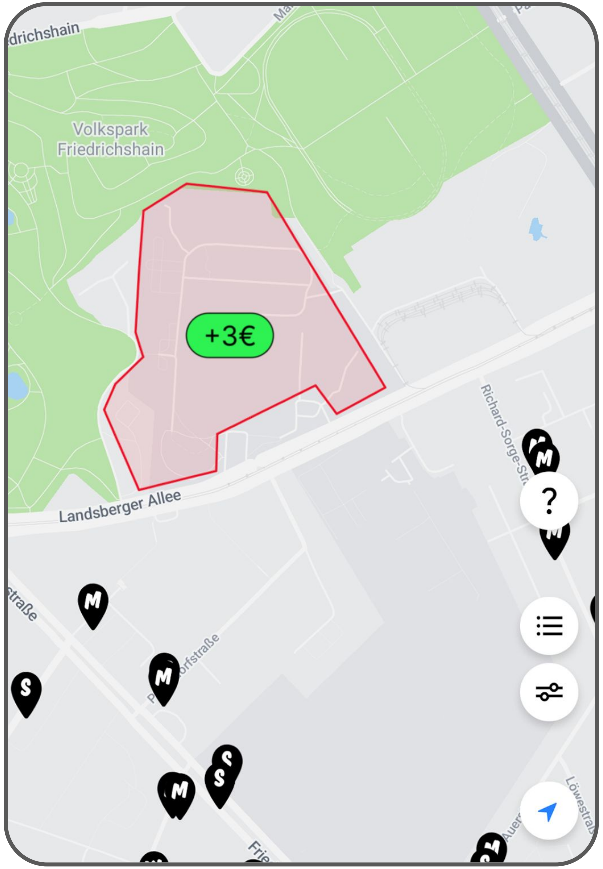
An anonymous **user's destination is predictable**, before they start a ride, by temporal, spatial, and contextual factors.



Why Predict Destinations?

Bonus Zones

+ Common in micromobility
– May not align with user intention
– Feels *cheap*

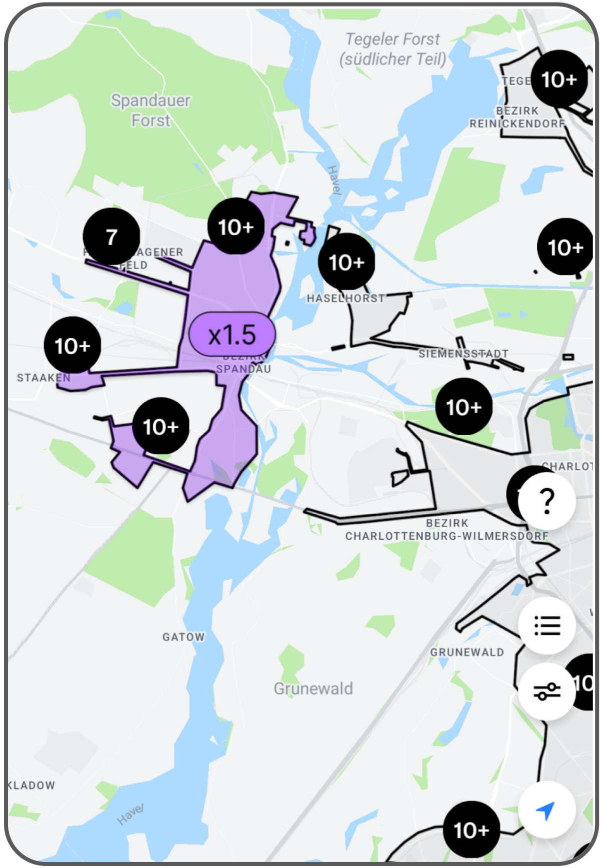


Incentive effects studied by (Wang et al., 2019)

Why Predict Destinations?

Zone-Based Pricing

+ Well researched
– Complex to understand for users

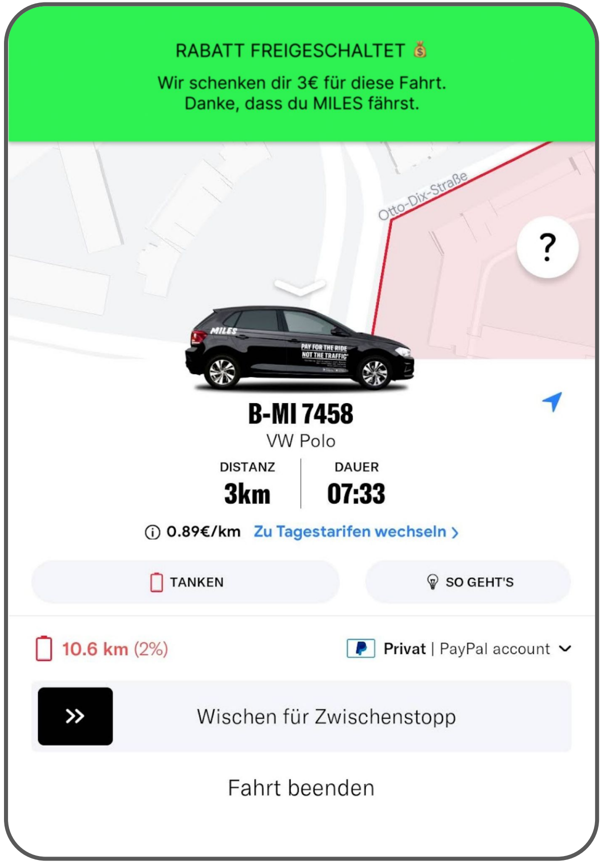


Incentive effects studied by (Lippoldt et al., 2019)

Why Predict Destinations?

Extrapolated Trips

- \+ Literature shows potential
- \+ Casino effect
- − Can't motivate relocation before ride started



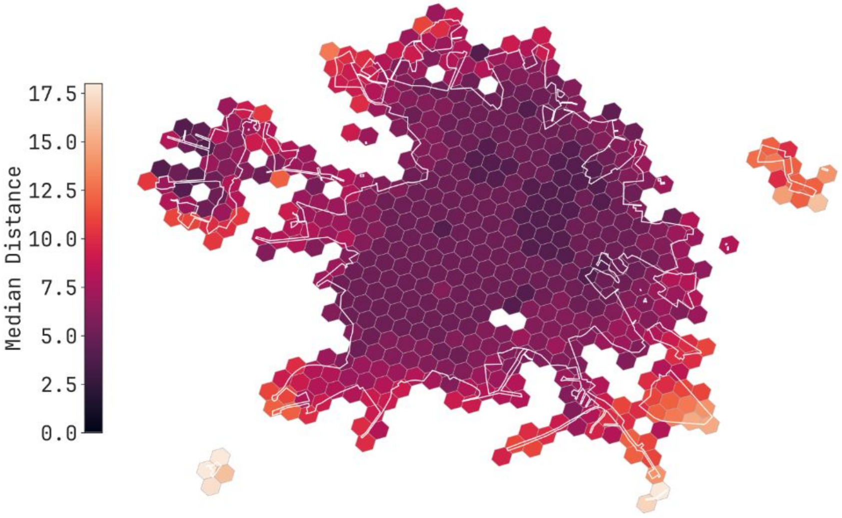
Trip extrapolation studied by (Casabianca et al., 2021)

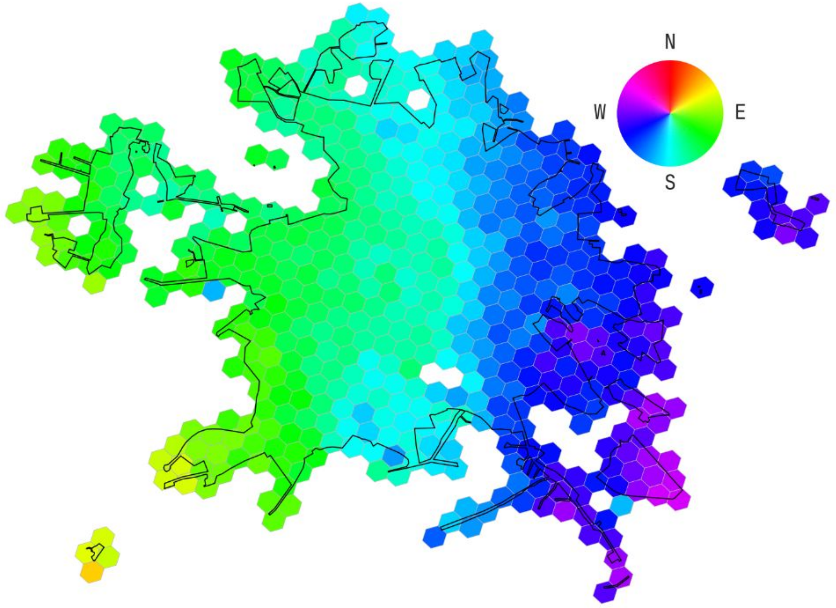
“ predictability are optimal when n = 2, with an accuracy and ” predictability ranging from 70% to 95%.

(Gambs et al., 2012)
on human mobility

Demand Prediction

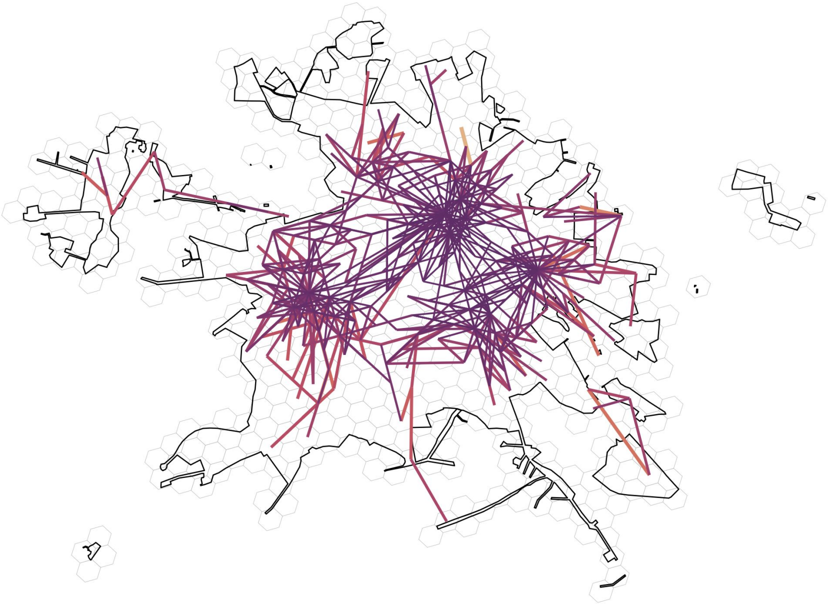
Holistic View

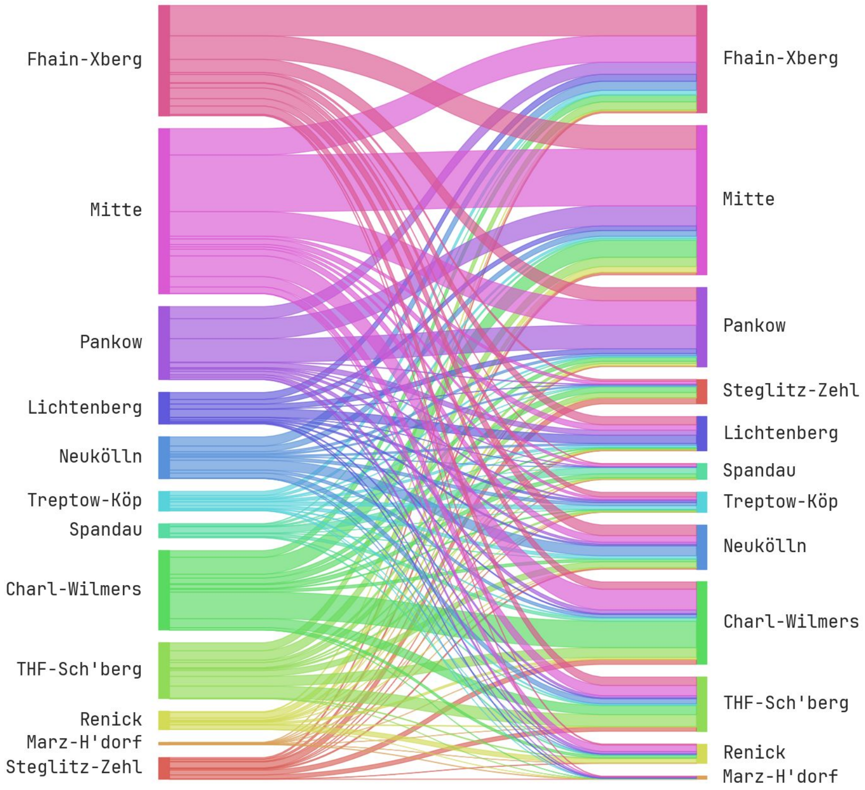




Demand Prediction

Most Common Trips

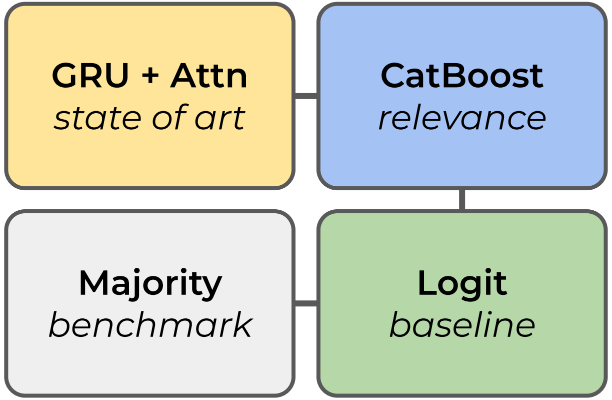




Destination Prediction

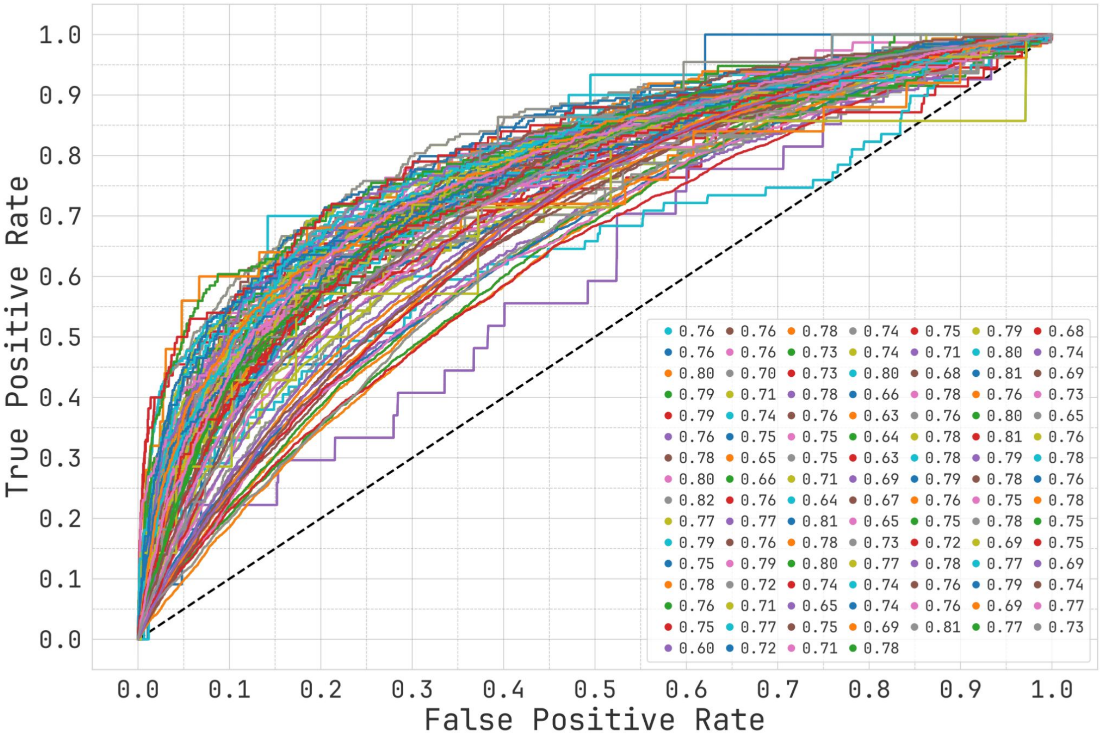
Models

- 109 spatial classes
- 3 models
- Increasing complexity
- Common approaches in literature



Destination Prediction

Logit

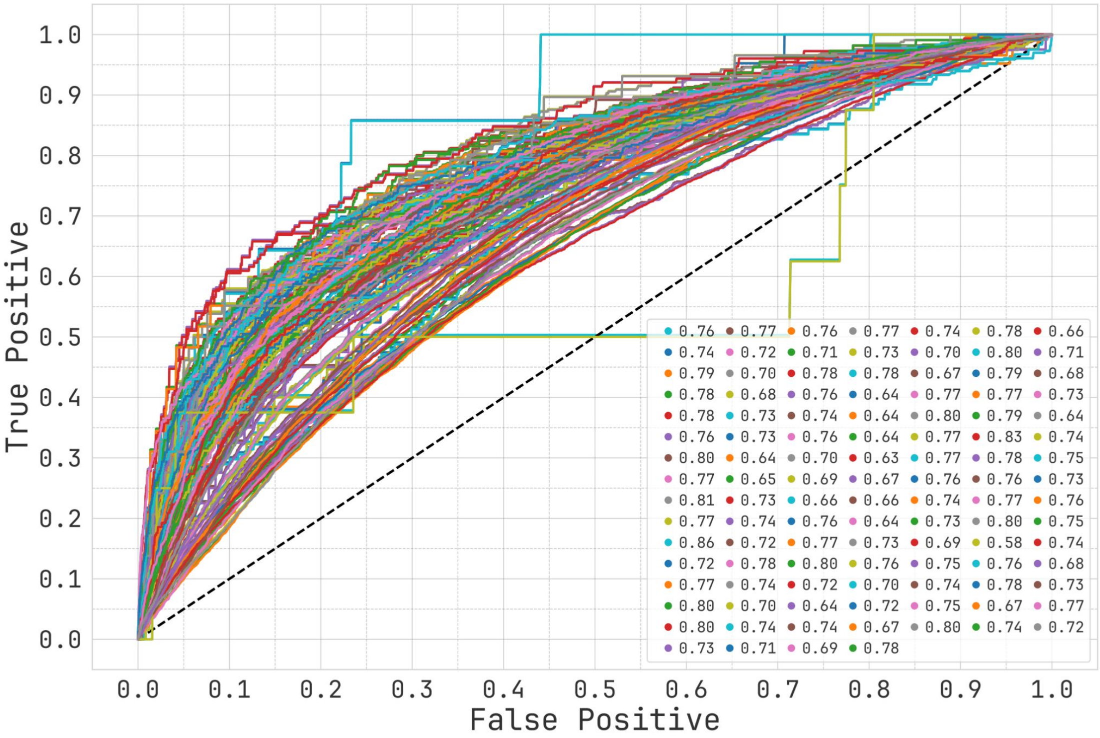


	Top-1	**Top-2**	**Top-3**
Model	8.87%	15.72%	21.65%
Baseline	6.63%	11.93%	16.51%

```python
from sklearn.linear_model \
    import LogisticRegression

logit = LogisticRegression(
   multi_class="multinomial", ...
)
# ohe encode categories in pipe
pipe.fit(train, targets)
```

Destination Prediction

CatBoost

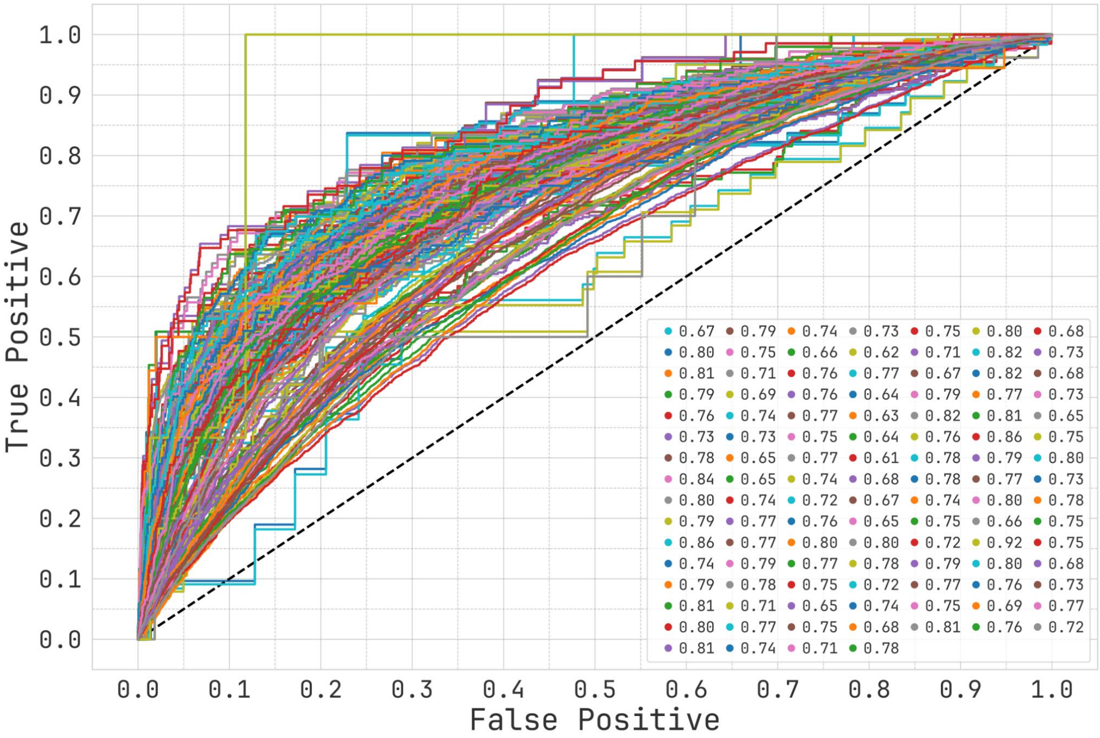


	Top-1	Top-2	Top-3
Model	8.69%	15.57%	21.58%
Baseline	6.63%	11.93%	16.51%

```
from catboost \
    import CatBoostClassifier

cb = CatBoostClassifier(...)
cb.fit(train, eval_set=test)
```

Destination Prediction

GRU + Attention

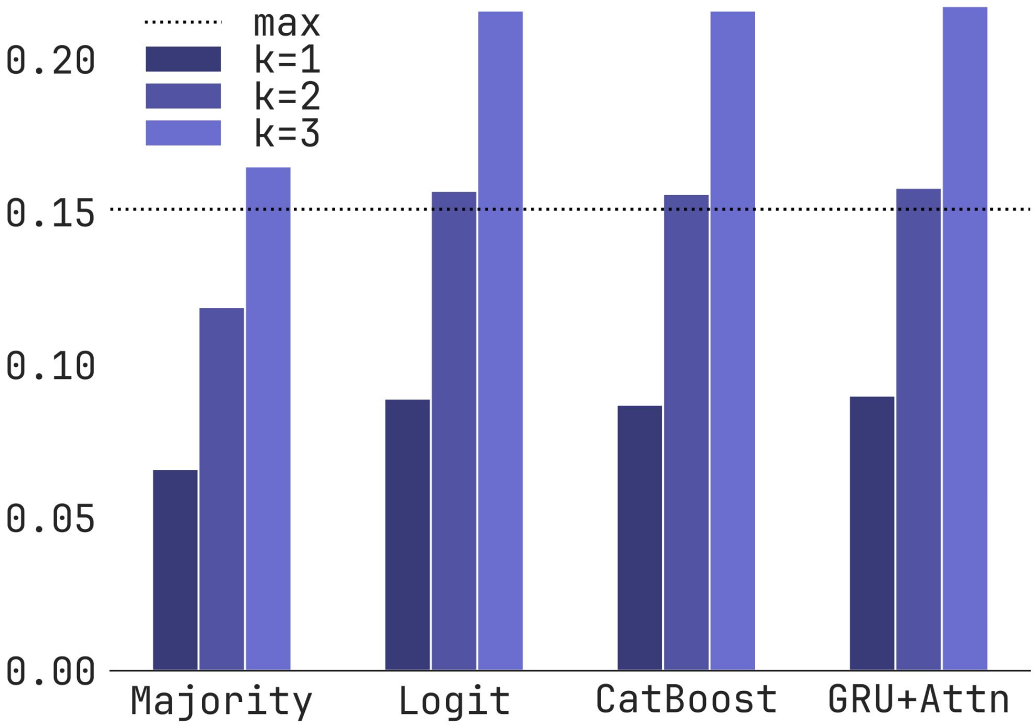


	Top-1	**Top-2**	**Top-3**
Model	8.97%	15.81%	21.75%
Baseline	6.63%	11.93%	16.51%

using torch.nn.GRU

Destination Prediction

Results



Destination Prediction

There's Hope

up to

93%

human movement
predictability on historic
information

(Song et al., 2010)

up to

44.5%

upper bound
just adding user
identifiers

based on Song et al.'s Π^*

Thoughts

koeni.dev