Lifestyle Importance Weighted Carbon Footprint Reduction

Karl Roush
Georgia Institute of Technology, AE 2020

Wells Fargo Campus Analytics Challenge

OVERVIEW

SOLUTION

- Objective
- Initial analysis vs. Machine Learning

PROCESS

- Data extraction
- Analysis (Initial & ML selection)

COMPARISONS

- Initial analysis vs. Machine Learning
- Insights

CONCLUDING REMARKS

- Deployment
- Future recomendations
- *Additional Insights*

Q & A

Objective

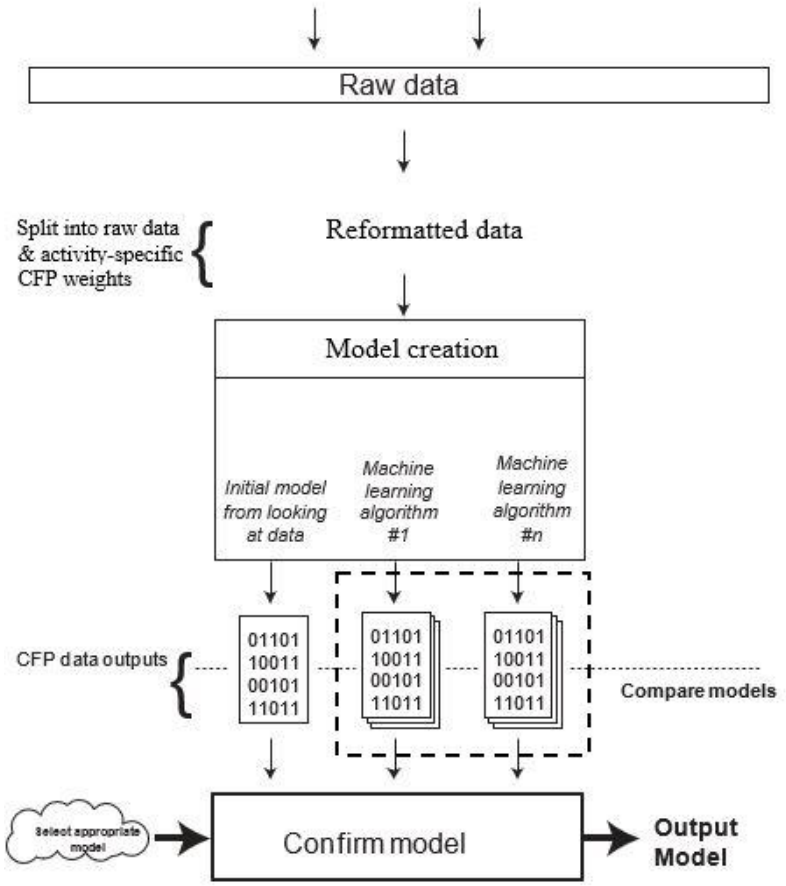
“to create a machine learning algorithm that minimizes carbon footprint for each customer while maintaining their total quality of life”.

- Define Carbon Footprint (**CFP**) as non-dimensional parameter
  - Sum of [Activity duration * weight of activity] for all activities
  - Higher number is worse
- Quality of life (**QoLI**) is correlated to importance
  - Represented as a percentage of importance
  - Preserve this number as much as possible for each activity

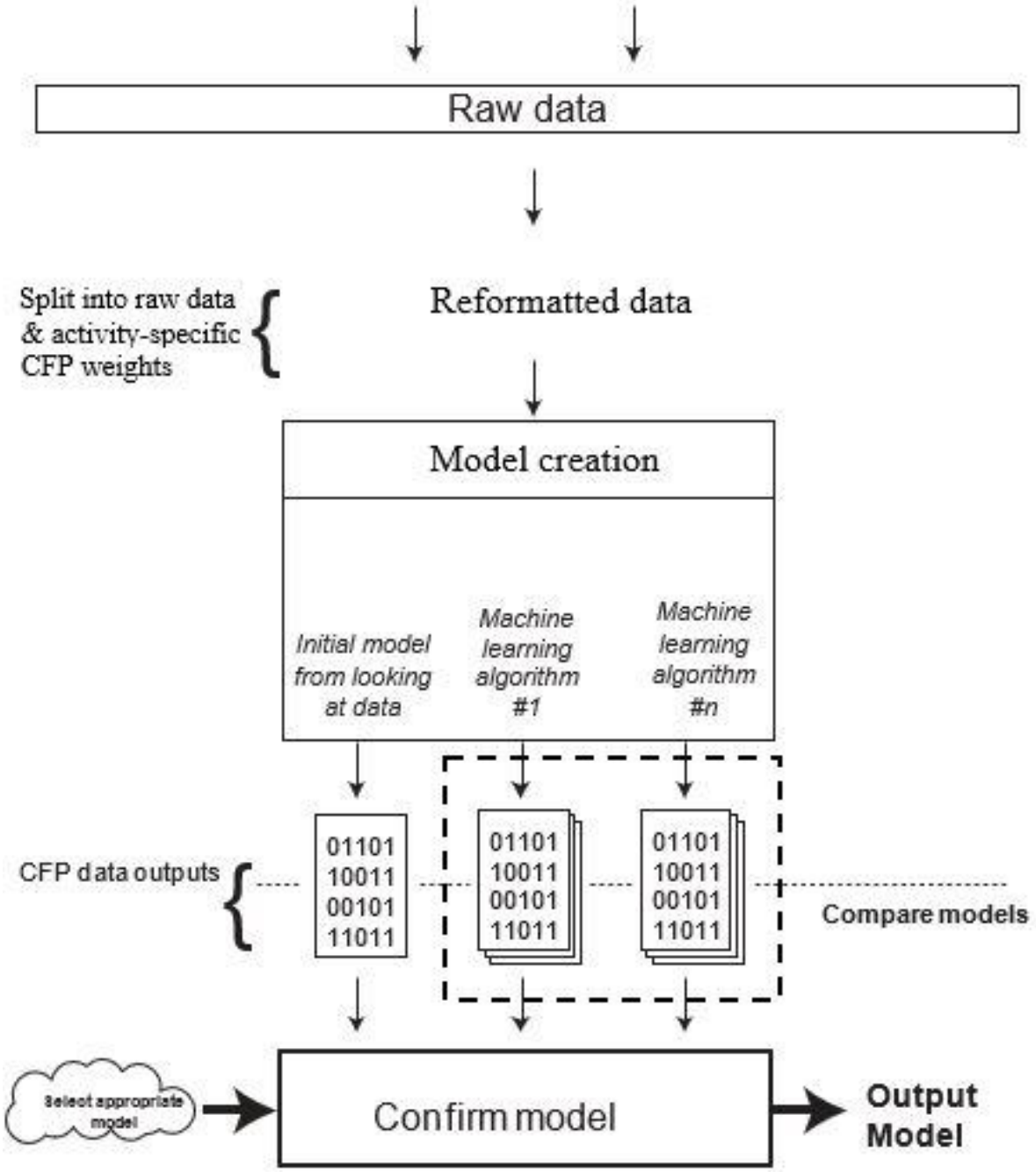
Solution

- Initial analysis produced mathematical model
  - Weights each activity duration by importance
  - Allow for duration to be “in line” with importance
- Machine learning model: SVM
  - Tested 6 different models to subset
  - Compared based on accuracy & stdev
- Results
  - Average CFP reduction of 26.16%
  - ML – model = 98.55% accuracy
  - Underestimate by 0.0145

$$Total\ CFP = \sum_{activities} dur * \frac{importance}{100} * specific\ CFP$$



PROCESS


Raw data
Split into raw data & activity-specific CFP weights
Reformatted data
Model creation
Initial model from looking at data
Machine learning algorithm #1
Machine learning algorithm #n
CFP data outputs
01101 10011 00101 11011
01101 10011 00101 11011
01101 10011 00101 11011
Compare models
Select appropriate model
Confirm model
Output Model

Data extraction

Missing Values

- Created index of QoLI and consumption
- Allows for easier ref. and measures of central tendancy
- Replaced missing values with median value from index
- Replacing with zero makes too many assumptions

QoLI

- Given as number with no context
- Assumed relative percentage
- Converted to decimal

```
93 def activity2num(activity):
94     return {
95         'Household heating => 70F': 1,
96         'Household heating < 70F': 2,
97         'Use of heat pump':3,
98         #...
99         'hazardous or electric items disposed':26,
100        'large items disposed':27,
101    }[activity]
```

Shortened index for referencing activities

Analysis

Initial Analysis

$$Total\ CFP = \sum_{activities} dur * \frac{importance}{100} * specific\ CFP$$

- Need bound since lowest CFP is zero
- Weights each activity duration by importance
- Allow for duration to be "in line" with importance

Machine Learning models

- To be compared based on accuracy & standard deviation on subset
- Tested:
  - Logistic regression
  - Linear discriminant analysis
  - K-nearest neighbors
  - Decision Tree
  - Gaussian Naive Bayes
  - **Support vector machine**

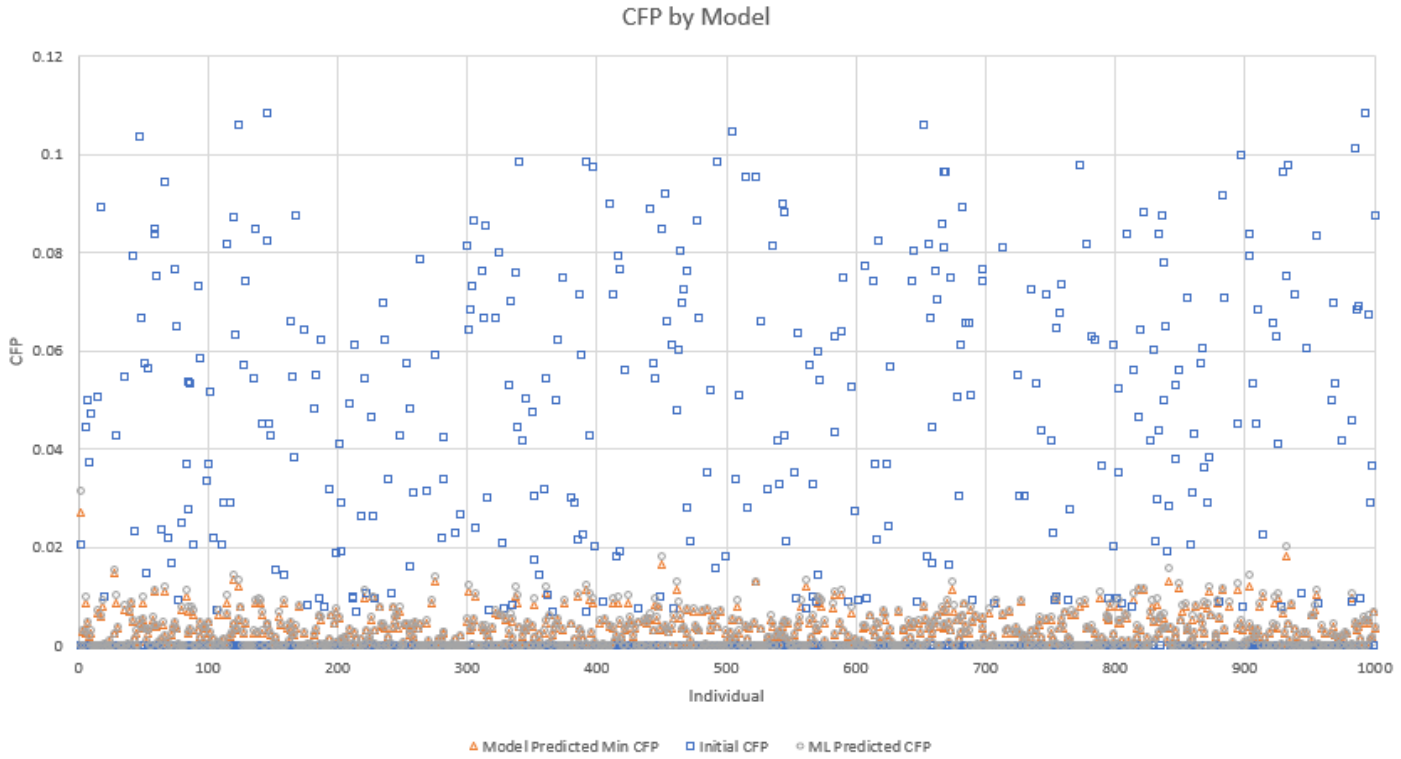
Model	Accuracy	Standard Dev.
LR	0.966667	0.040825
LDA	0.975000	0.038188
KNN	0.983333	0.033333
DT	0.975000	0.038188
GNB	0.975000	0.053359
SVM	0.991667	0.025000

Accuracy and Standard Deviation Results from initial 10-fold cross, by algorithm

Comparisons

ML vs initial model

- Difference (ML-model) between the two output datasets was -1.45%
- Accuracy= 98.55%
- ML often underestimated the minimum CFP for an individual by about 0.0145.

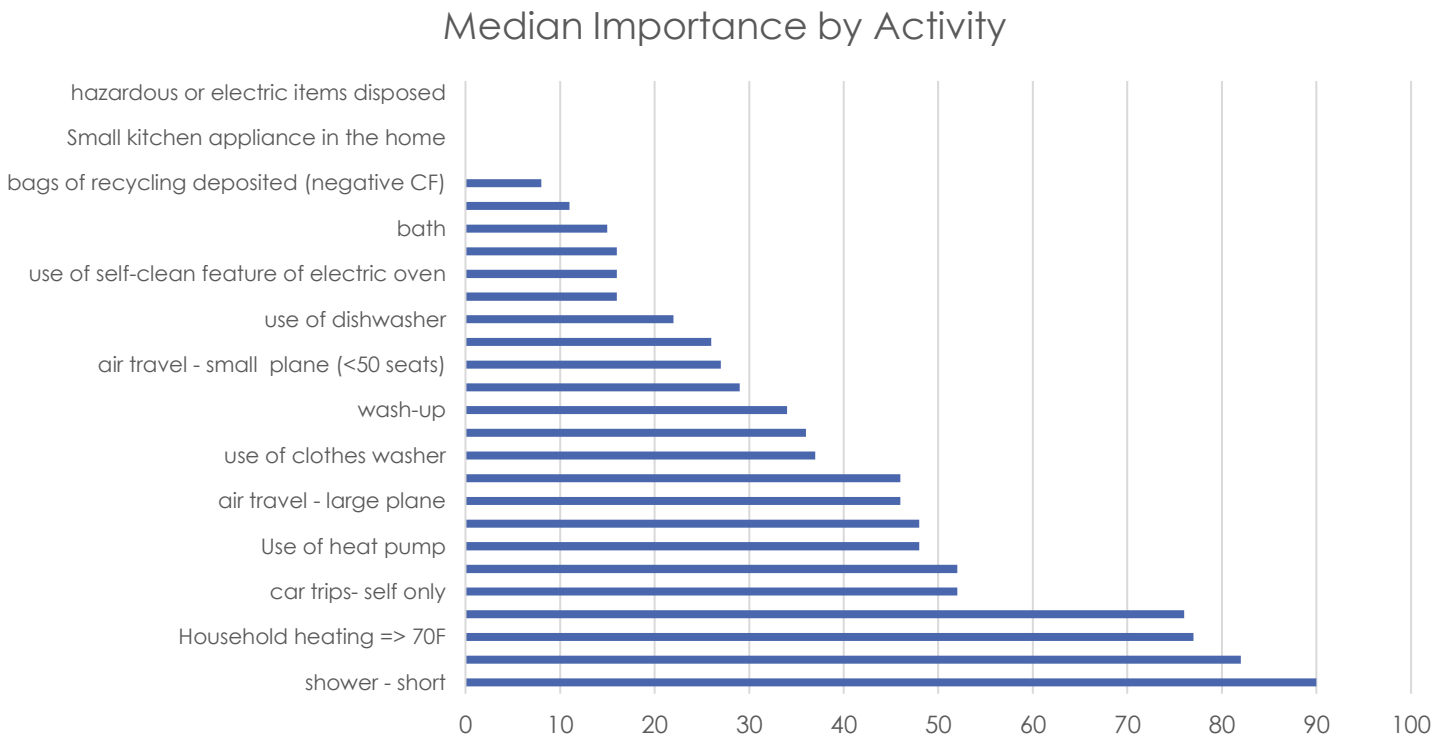


Insights

Activities

- Showering, heating, and transportation are the most valued
- “air travel- small plane” and “oven self-clean” have relatively low importance
  - High CFP impact → prime target for reducing individual’s total CFP

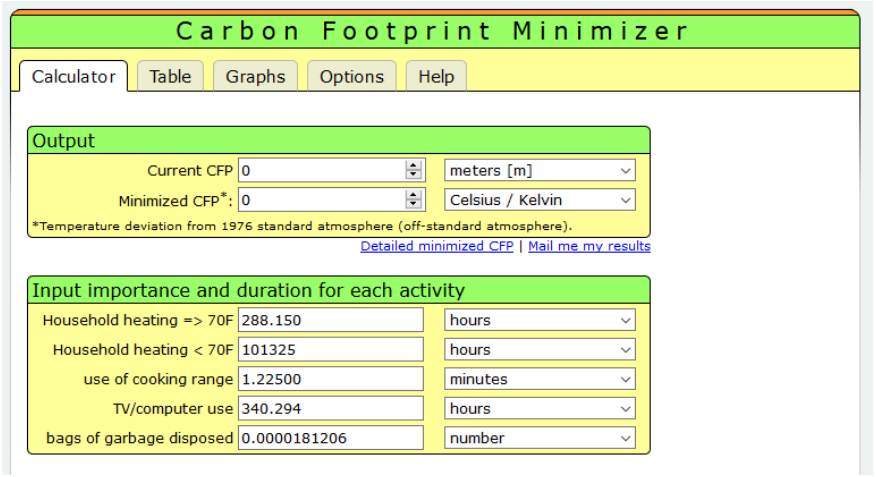
Overall **drastic change between initial CFP and minimized CFP**, average of 26.16%



CONCLUSION

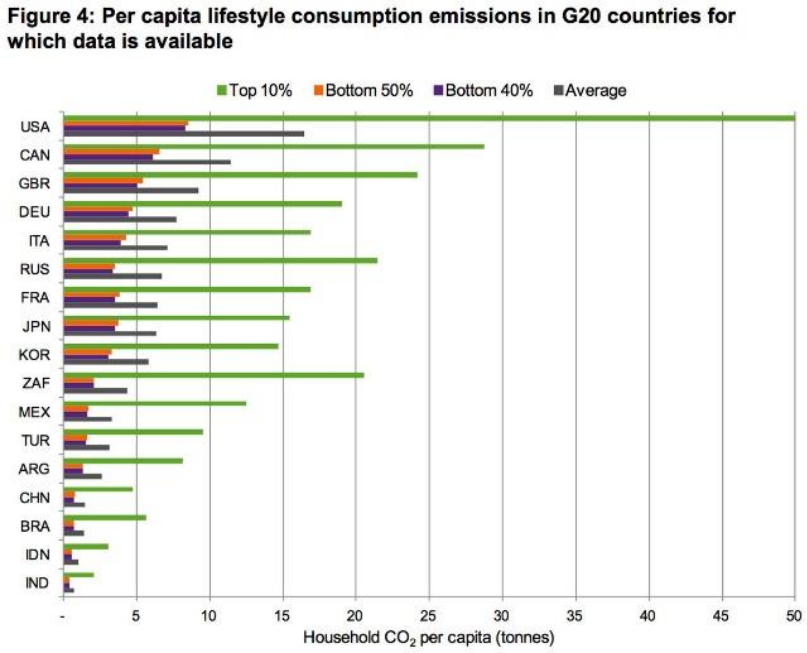
Individual Awareness

- Provide an online calculator
- Gives people an idea of what their CFP actually is



Comparison

- Demonstrate comparisons
- Household water usage example

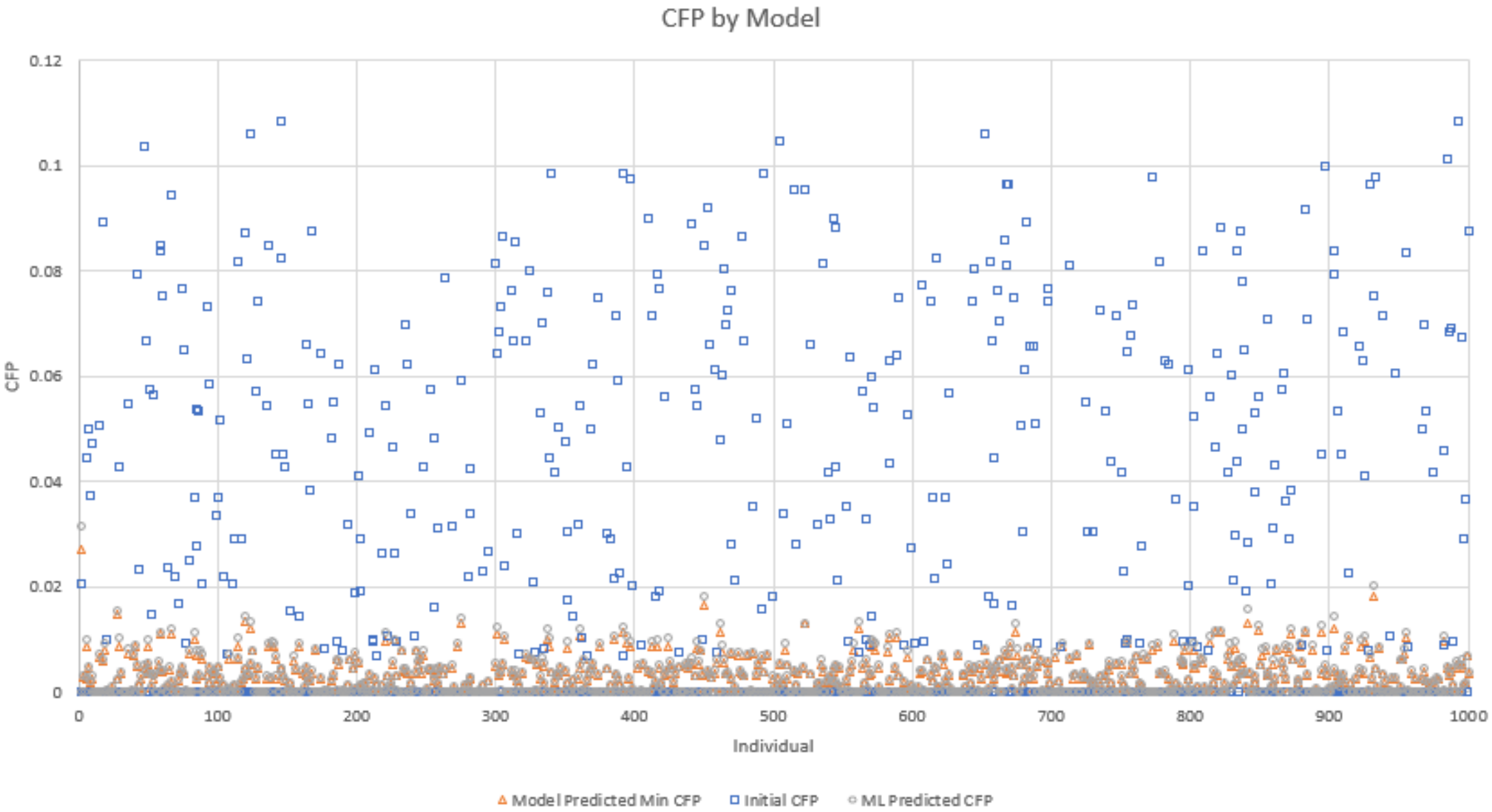


Future recommendations

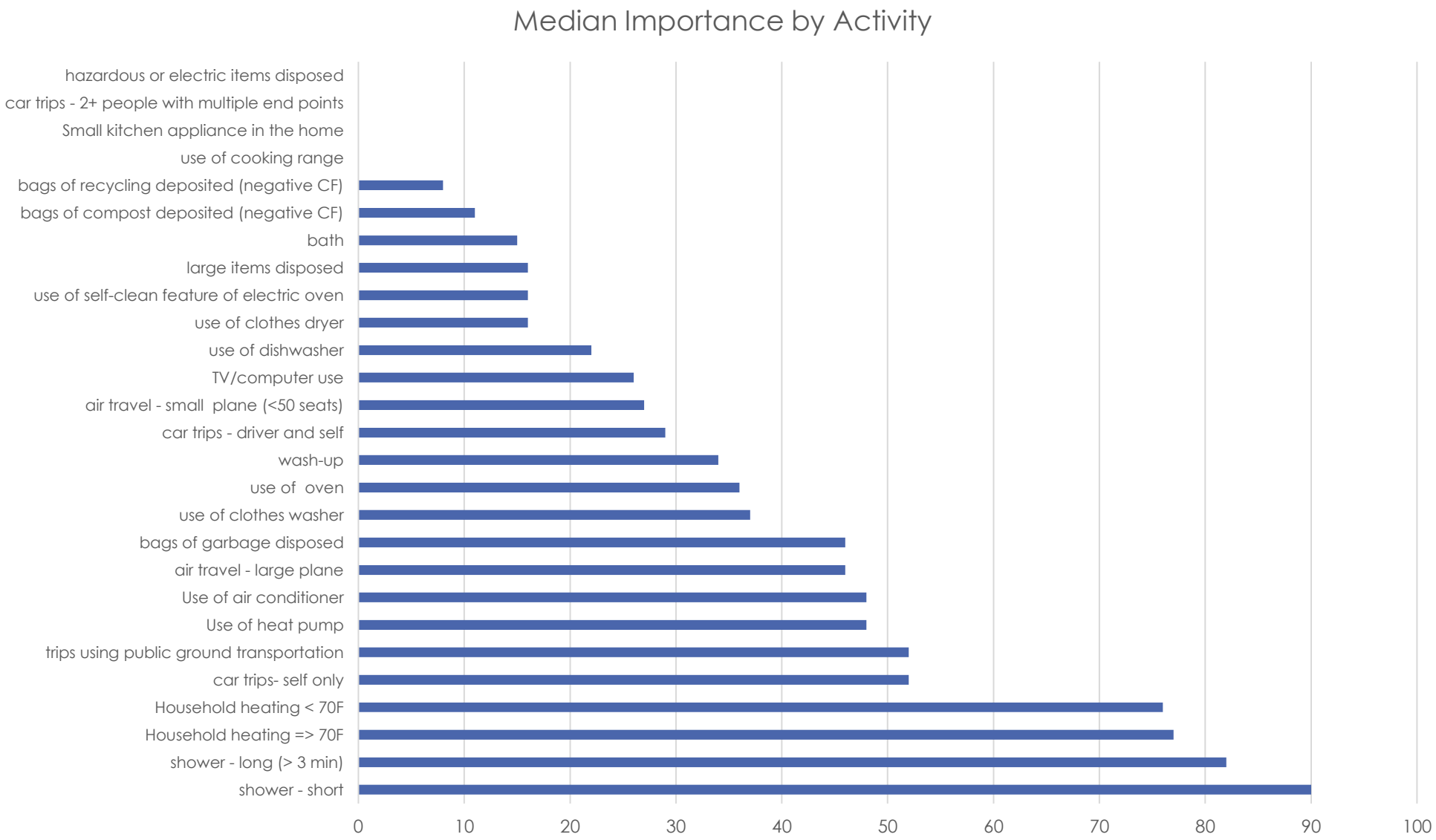
- More custom algorithm, tailored exclusively for this task
- More training data
- “What happened after” analysis

Q & A

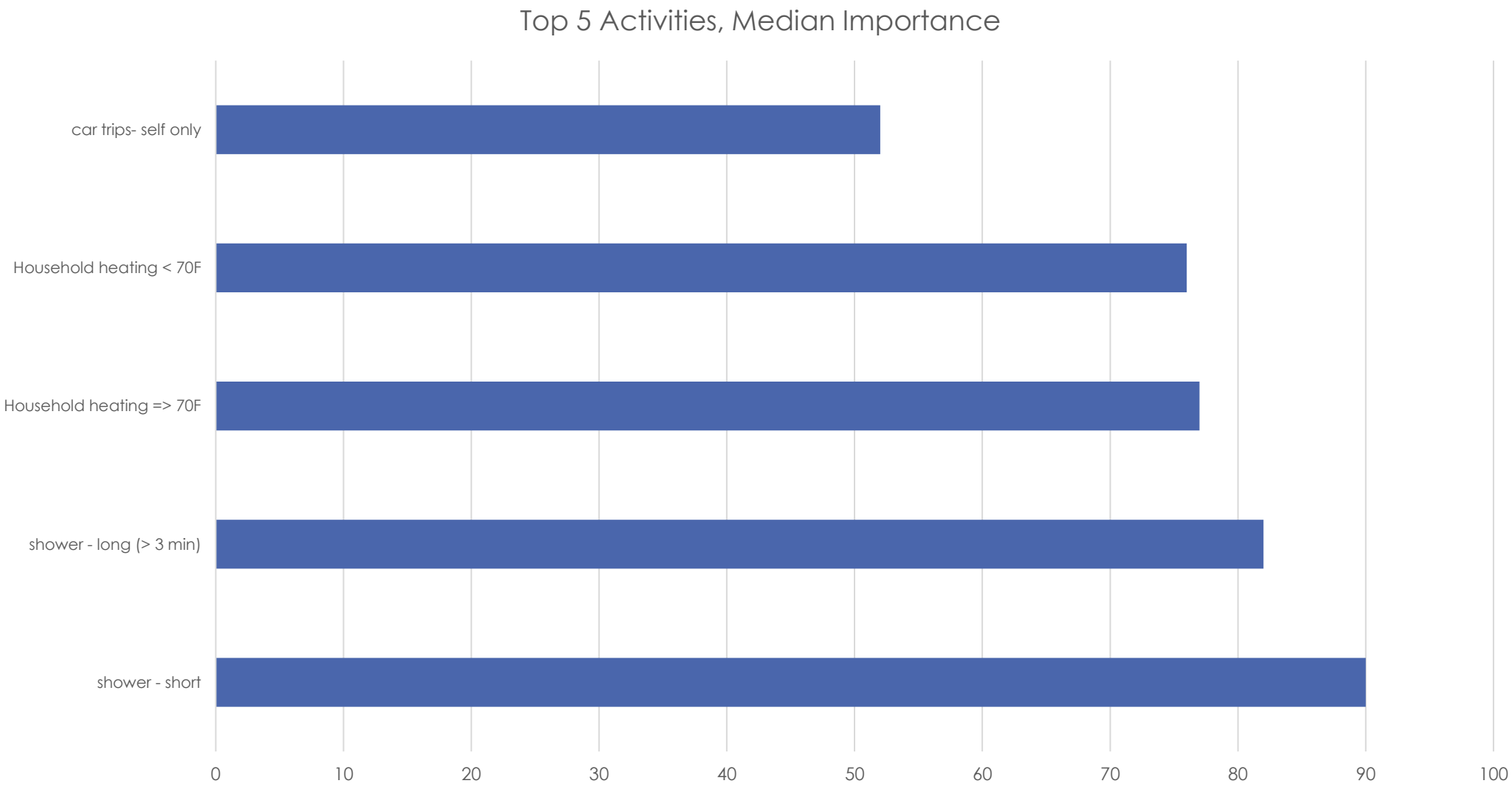
Insights (Continued)



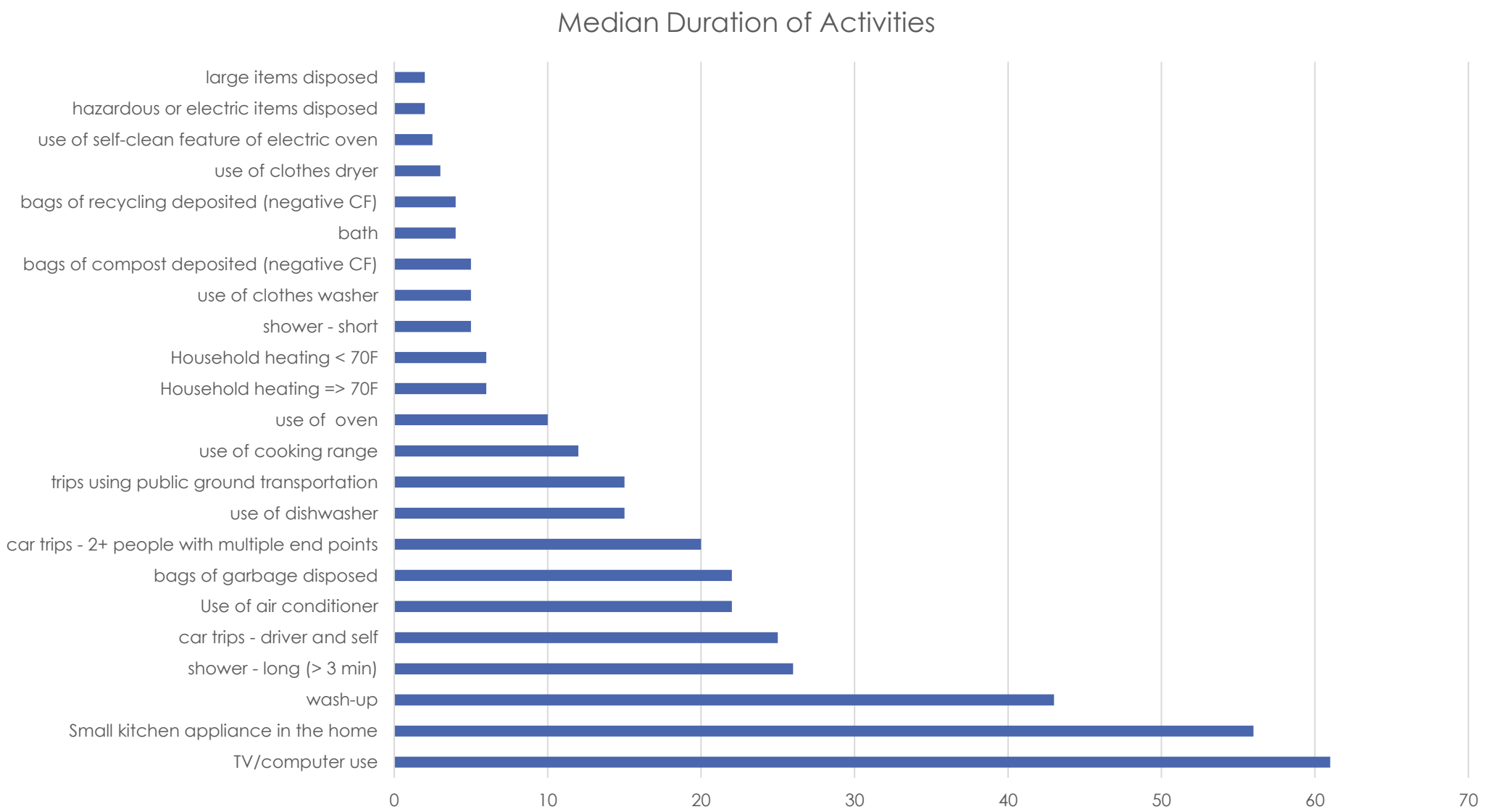
Insights (Continued)



Insights (Continued)



Insights (Continued)



Insights (Continued)