Triangle Regional Model

Spring 2025 User Forum

April 23, 2025

Alexander Yoshizumi, Si Shi

NC Capital Area Metropolitan Planning Organization

TRIANGLE WEST Transportation Planning Organization

GO Triangle

ITRE Institute for Transportation Research and Education

SYSTEMS PLANNING & ANALYSIS

Agenda

- Welcome & Attendance
- Staffing Updates
- TRM Updates
- Transit Ridership Analysis
- Discussion Session #1: AI and Big Data
- Discussion Session #2: The Future of the TRM
- Networking

Welcome & Attendance

Attendance

1) Alexander Yoshizumi

2) NC State University

3) Program Manager

4) ayoshiz@ncsu.edu

5) hiking, biking, cooking, woodworking, basketball, miniature painting, video games

1) Full Name

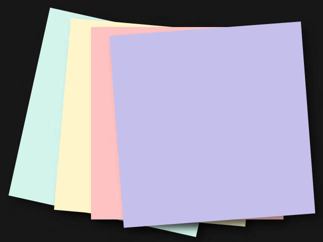
2) Affiliation

3) Title

4) Email

5) Hobbies / Activities

Attendance



Pass the sticky notes down to the end of the table...

...and we will collect them

Staffing Updates

Team Member Updates

Welcome Kshitiz Khanal!

Kshitiz is a planning professional with experience in big data and machine learning applications in transportation, energy, economic development and other planning subdomains. He recently finished his PhD in City and Regional Planning at the University of North Carolina at Chapel Hill. He holds undergraduate and graduate degrees in engineering and energy systems planning from Kathmandu University and has previously worked in research, teaching, consulting, and engineering roles in Nepal and the US.

TRM Updates

TRMG2 V1.3.2

- TRMG2 v1.3.2 remains the latest official model.
- Updates have shifted focus to TRMG2 v2 as we approach the MTP submission deadline towards the end of calendar year 2025.

Communities of Concern

- The Communities of Concern reporting tool was completed in December.
- The tool can use MPO Community-of-Concern definitions or derive Communities of Concern from the population synthesis data.
  - If users derive Communities of Concern from population synthesis, TRMG2 can define the proportion of population that qualifies as Communities of Concern.
- The tool produces measures like job accessibility by mode, by Community of Concern, and by all TAZs.
- The tool produces per capita metrics such as average delay and hours of travel.

Other Performance Measures

- The model now outputs performance measures relevant to MTP development with every run.

- We added more reporting for transit summaries.

Household Travel Survey

- The 2024 Household Travel Survey was completed last calendar year.
- The unweighted dataset has been reviewed by our TRM Technical Team, and we expect the weighted dataset and report will be complete by the end of June.
- We collected university-specific travel data which will inform enhancements to the TRMG2 v3 model redevelopment.
- We will be working on a travel trends analysis in early FY2026.

K-12 School Forecasting

- We are constructing a K12 forecasting sub-model to more realistically estimate and assign K12 travel in future-year scenarios.
- The sub-model will be operationalized using a mixed modeling approach that combines expert-informed, rule-based processes with stochastic and/or machine learning models.
- We are aiming to complete the public, non-charter portion of the sub-model by the end of June 2025.
- We expect that the initial version of the model will be written in Python.

Fiscal Year 2026 Road Map

- We are working to revise and expand the model documentation with an emphasis on better diagramming of model processes towards improved traceability.

G2v3

- Planning enhancements for how transit is handled within the model:
  - (a) more uniform route coding across LB and BRT.
  - (b) more sophisticated treatment of PNR and KNR in Chapel Hill.
- A revised university model informed by the 2024 university household travel survey response segment.

Transit Ridership Analysis

Institute for Transportation Research and Education,
North Carolina State University
April 2025

Transit Analysis

Based on 2023 Transit On Board Survey

Table of Contents

Methodology

- Aggregated to zone level
- Random forest model predicts ridership:
  - By boardings, alightings, and boardings + alightings
  - By time of day
  - By purpose
  - By time of day and purpose

What's Random Forest?

- Random forest, as the name suggested, is a collection of decision trees that works together to reach a singular prediction.
  - Decision trees can be viewed as a series of questions. These questions make up the decision nodes in the tree, along which the data are split.
  - While decision trees are commonly used in machine learning, they can be prone to problems, such as bias and overfitting. However, when multiple decision trees form an ensemble in the random forest algorithm, their prediction accuracy increases, particularly when the individual trees are uncorrelated with each other.
- Key benefits:
  - It can handle both regression and classification tasks.
  - Easy to determine feature importance.
  - Reduced risk of overfitting.

Evaluate Random Forest Model

- Mean Minimal Depth
  - In decision trees, variables are selected at each split point based on how well they partition the data. A variable appearing near the root of the tree plays a more critical role in explaining the target variable. For each decision tree, the minimal depth is the level at which a variable first appears as a split point. Mean minimal depth is then the average minimal depth calculated across all trees within a random forest analysis.
- Times a Root
  - The number of times this variable serves as the root variable (the initial split). Related to mean min depth.
- MSE Increase
  - The reduction in model performance (increase in prediction error) when a specific feature's values are shuffled.
- Node Purity Increase
  - Node purity refers to how well a feature splits the data into homogeneous groups at each decision tree node. In regression tasks, purity is often measured by the reduction in variance when a feature is used for splitting.

Explanatory Variables

Icon	Category	Variable	Definition
	Accessibility	access_transit	Transit accessibility metric from accessibility model
	Accessibility	access_walk	Walk accessibility metric from accessibility model
	Land use	Walkability	The probability of walk trips (does not consider skims)
	Land use	walk_attr_dens	Density of GS attraction
	Land use	GSIndex	Land use diversity. 0 means the TAZ only has one thing.
	Land use	Density	Population and employment density
	Land use	EmpDensity	Total employment density
	Land use	IndEmpDensity	Industrial employment density
	Demographic	LargeHHPct	Percentage of household with more than 1 person
	Demographic	HHSize1Pct	Percentage of household with 1 person
	Demographic	senior_pct	Percentage of senior population
	Demographic	HiIncomePct	Percent of high-income household
	Demographic	poverty_pct	Percent of households that meet poverty threshold
	Demographic	Median_Inc	Median income of the TAZ
	Auto ownership	v0_pct	Percentage of household with 0 vehicle
	Auto ownership	vi_pct	Percentage of household with insufficient vehicles
	Parking	isPark	Maximum of all parking cost
	Parking	ParkCostO	Parking cost for other trips
	Parking	ParkCostU	Parking cost for university trips
	Parking	ParkCostW	Parking cost for work trips

Results Interpretation

Overall

First total ridership is used as the response variable. The total ridership only includes home-based work and home-based other trips in this analysis. The home-based university trip is not considered as it is highly concentrated near the university campus.

To understand the relationship between the demographic and geographic explanatory variables and total ridership, a total of 500 trees are built to improve accuracy and reduce overfitting. Each tree is built using a random subset of data and features.

Mean Min Depth

- A **low mean minimal depth** indicates that the feature is frequently chosen for splitting early in the trees, which typically suggests high importance.
- Top 5 variables:
  - Access transit
  - Walk attraction density
  - Density
  - Walkability
  - GS Index

Variable	Mean Min Depth	Top 5 Indicator
access_transit	2.9	
walk_attr_dens	3.0	
Density	3.2	
Walkability	3.4	
GSIndex	3.5	
access_walk	3.6	
EmpDensity	3.7	
Median_Inc	4.0	
vi_pct	4.2	
poverty_pct	4.3	
HiIncomePct	4.3	
HHSize1Pct	4.5	
LargeHHPct	4.6	
v0_pct	4.8	
IndEmpDensity	5.2	
senior_pct	5.3	
ParkCostU	5.3	
isPark	6.7	
ParkCostW	6.7	
ParkCostO	7.9	

Accessibility | Land Use | Demographic

Auto Ownership | Parking

Times a Root

- A **high value in times a root** suggests that a variable is highly influential in splitting the dataset into subsets that reduce variance (in regression) or impurity (in classification) right at the beginning of the tree.
- Top 5 variables:
  - Access transit
  - Percentage of senior population
  - Walkability
  - Access Walk
  - Percentage of poverty households

Accessibility | Land Use | Demographic
Auto Ownership | Parking

Variable	Time a Root	Top 5 Indicator
access_transit	92	
senior_pct	57	
Walkability	52	
access_walk	47	
poverty_pct	41	
v0_pct	40	
Density	34	
ParkCostU	28	
GSIndex	21	
ParkCostW	19	
Median_Inc	16	
ParkCostO	15	
walk_attr_dens	11	
vi_pct	8	
HHSize1Pct	6	
isPark	5	
HiIncomePct	5	
EmpDensity	2	
LargeHHPct	1	
IndEmpDensity	0	

MSE Increase

- A **high MSE increase** indicates that large error is introduced to the prediction when the variable is permuted.
- Top 5 variables:
  - Walkability
  - Walk attraction density
  - Density
  - Employment density
  - Parking cost for other trips

Variable	MSE Increase	Top 5 Indicator
Walkability	62,326	
walk_attr_dens	55,386	
Density	36,299	
EmpDensity	22,059	
ParkCostO	13,302	
LargeHHPct	12,348	
HHSize1Pct	7,558	
access_transit	7,149	
vi_pct	6,100	
senior_pct	5,859	
GSIndex	5,371	
access_walk	4,665	
HiIncomePct	4,663	
isPark	2,358	
ParkCostW	2,351	
poverty_pct	1,552	
ParkCostU	712	
v0_pct	(157)	
IndEmpDensity	(585)	
Median_Inc	(724)	

Accessibility | Land Use | Demographic

Auto Ownership | Parking

Node purity increase

- A **high node purity increase** suggests greater reduction in variance when the variable is used for splitting.
- Top 5 variables:
  - Access transit
  - GS index
  - Walk attraction density
  - Walkability
  - Access walk

Variable	Node Purity Increase	Top 5 Indicator
access_transit	4,950,338	
GSIndex	3,060,985	
walk_attr_dens	3,044,184	
Walkability	2,757,487	
access_walk	4,236,014	
Density	1,670,235	
EmpDensity	1,376,284	
HHSize1Pct	946,507	
ParkCostO	842,253	
v0_pct	1,366,042	
vi_pct	1,135,257	
HiIncomePct	1,593,155	
LargeHHPct	322,721	
senior_pct	828,570	
poverty_pct	1,225,568	
Median_Inc	1,297,692	
IndEmpDensity	1,529,718	
isPark	1,386,043	
ParkCostW	3,422,692	
ParkCostU	3,276,418	

Accessibility | Land Use | Demographic
Auto Ownership | Parking

Importance Ranking summary

Icon	Variable	Mean min depth	MSE increase	Node purity increase	Times a root
	access_transit	1	12	2	1
	access_walk	6	13	5	4
	Walkability	5	1	4	3
	walk_attr_dens	2	2	4	6
	GSIndex	3	11	2	11
	Density	5	4	5	4
	EmpDensity	6	5	7	6
	IndEmpDensity	14	19	16	17
	LargeHHPct	13	5	11	15
	HHSize1Pct	10	6	11	11
	senior_pct	16	12	13	19
	HiIncomePct	11	13	14	15
	poverty_pct	12	16	17	18
	Median_Inc	8	19	15	12
	v0_pct	13	17	10	13
	vi_pct	10	9	11	16
	isPark	18	12	17	7
	ParkCostO	20	6	12	9
	ParkCostU	17	17	20	13
	ParkCostW	18	12	16	6

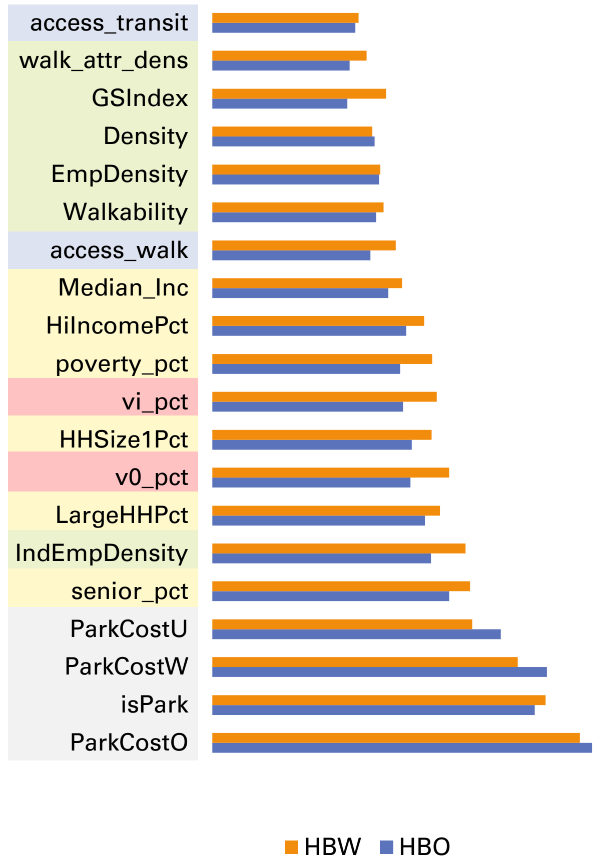
The table on the left summarize the importance ranking of each variable across four (4) metrics. Overall, accessibility variables are highly correlated with ridership, but do not contribute much to the prediction accuracy. Land use variables also significantly impact ridership and can improve prediction performance. On the other side, demographic and auto ownership variables are less important in this zonal level analysis, potentially because the mode choice decision is made based on these variable at a personal level only.

Comparison by Trip Purpose

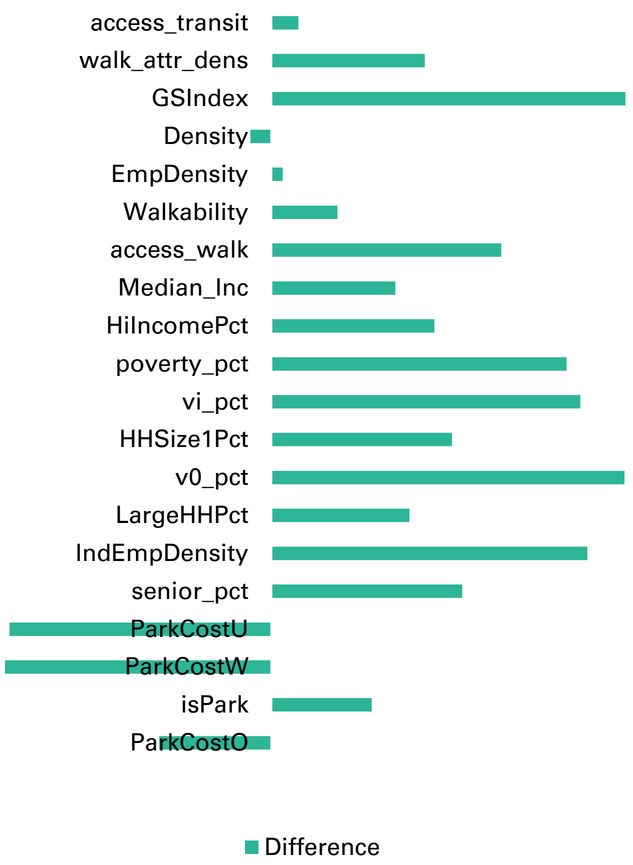
In this section, we split total ridership by trip purposes, focusing on home-based work and home-based other trips only. The goal is to understand how each explanatory variable impacts work and non work trips differently.

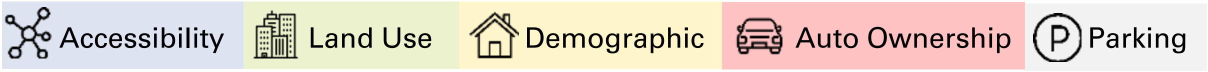
Mean Min Depth

Mean Minimal Depth



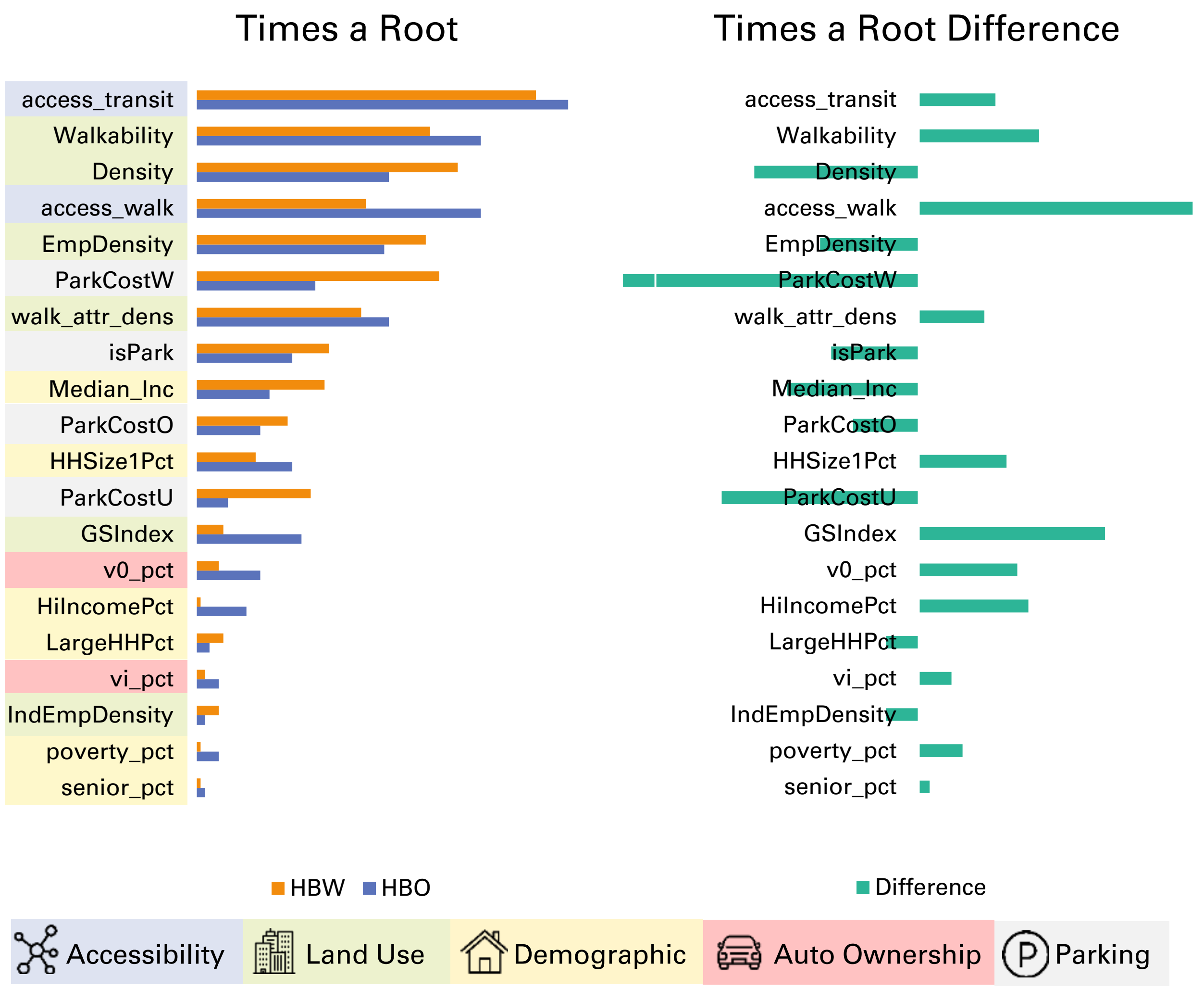
Mean Minimal Depth Difference





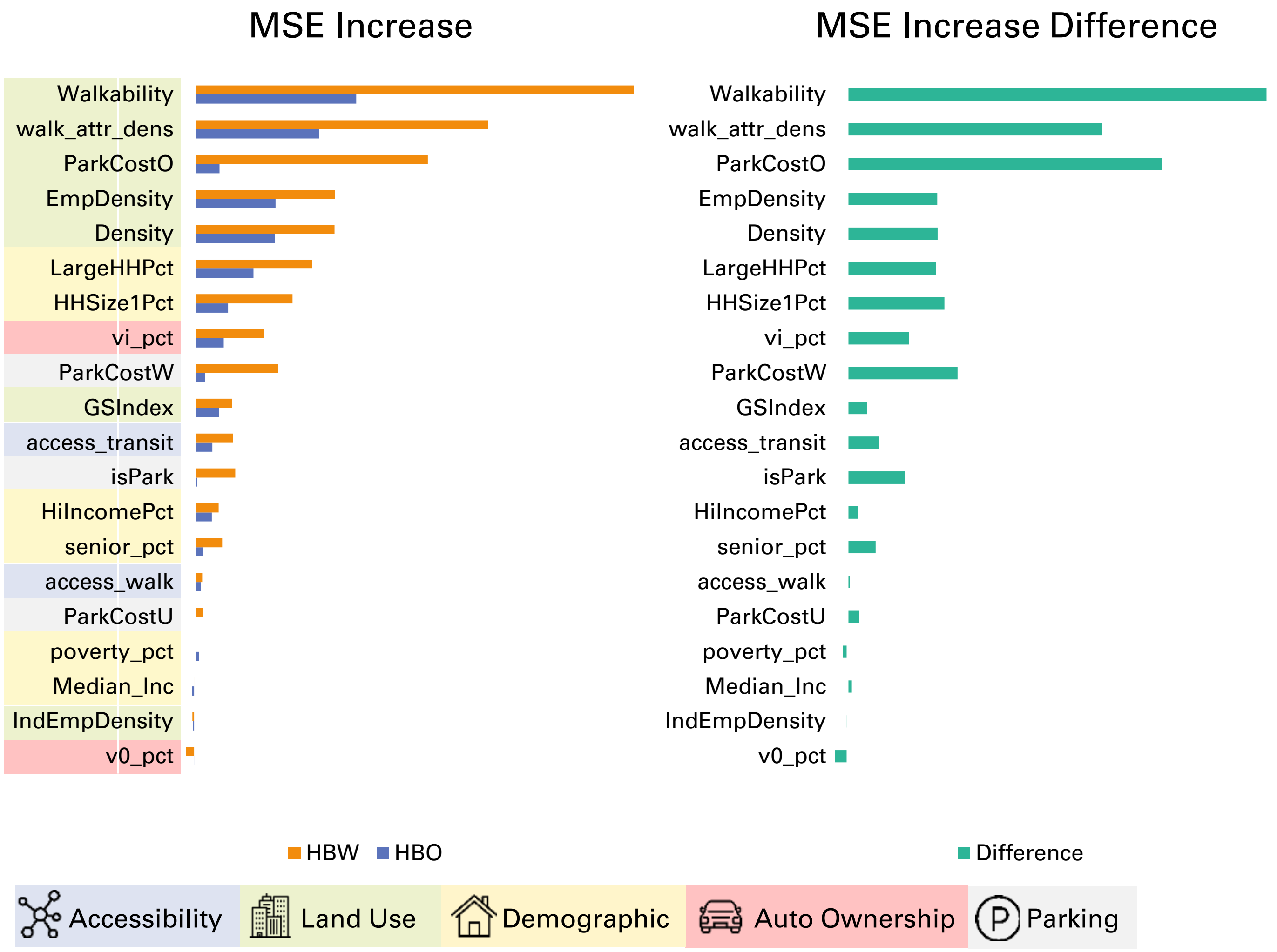
The two charts on the left summarize the mean minimal depth by trip purpose and the difference between them. It suggests that parking cost variables are generally more important for HBW while GS index, auto ownership, income are more important for HBO trips.

Times a Root



Similarly, parking cost variables more frequently serve as the root variable in the home-based work (HBW) trip model, highlighting their importance for predicting work-related ridership. Conversely, walk environment variables such as walk accessibility, GS index, and walkability play a more significant role in predicting non-work trips.

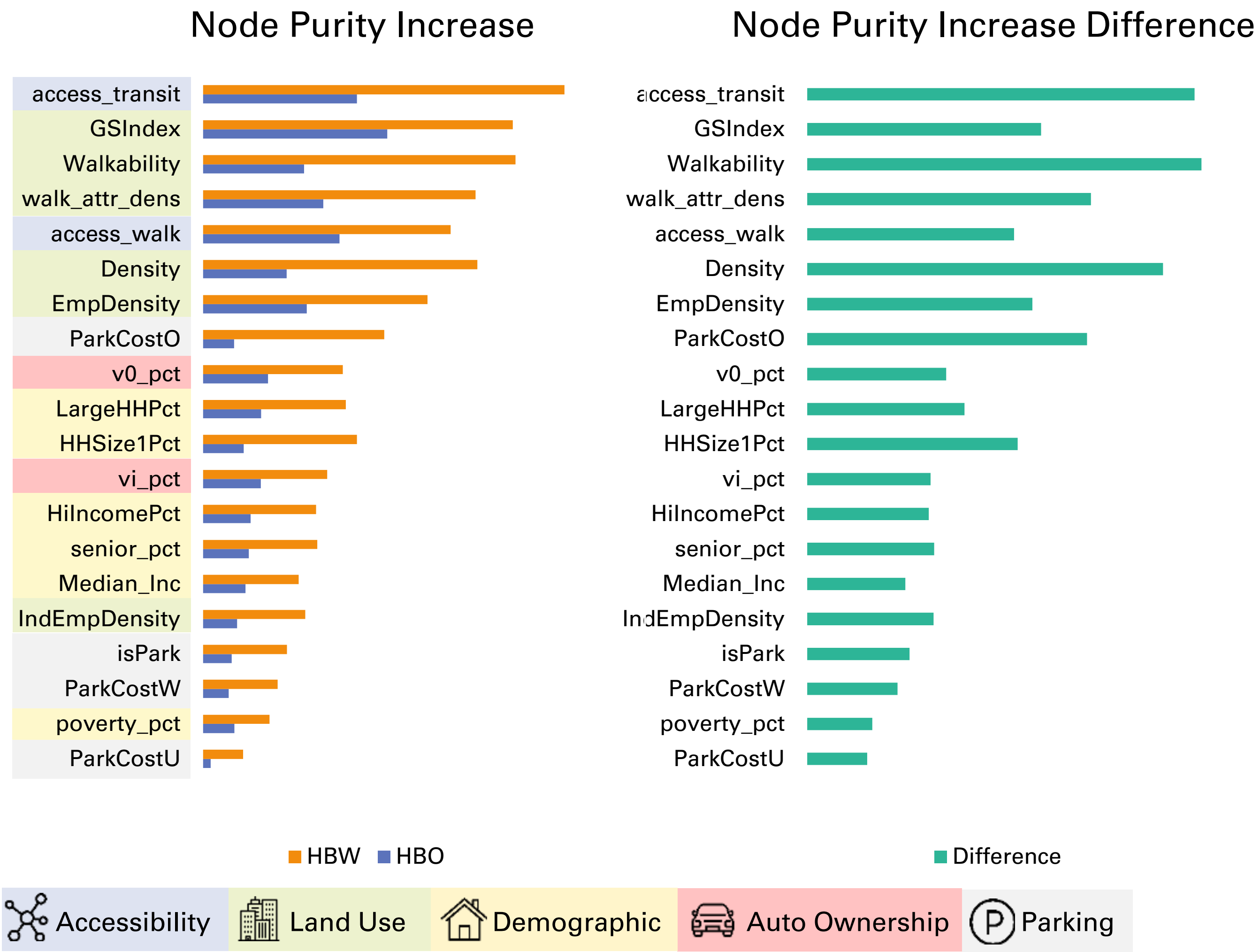
MSE Increase



Across most variables, the MSE increase for the HBW model is much larger than that for HBO, indicating that work trips are more sensitive to differences in these variables at the zone level. Parking variables still exhibit significantly higher importance (higher MSE increase) for predicting HBW trips. The substantial difference indicates that parking costs are crucial factors when forecasting work-related trips compared to non-work trips.

Walk environment factors – walkability, walk attraction density, and density - show a much higher influence on HBW trips. While they are important for both types of trips, the larger values suggest they are especially critical in predicting trips to work locations.

Node Purity Increase



The consistently higher node purity increase for variables in the HBW model compared to the HBO model indicates that HBW trips tend to have more regular and predictable patterns compared to other purpose trips. It also suggests that the current set of explanatory variables may have clearer, more significant relationships with HBW ridership.

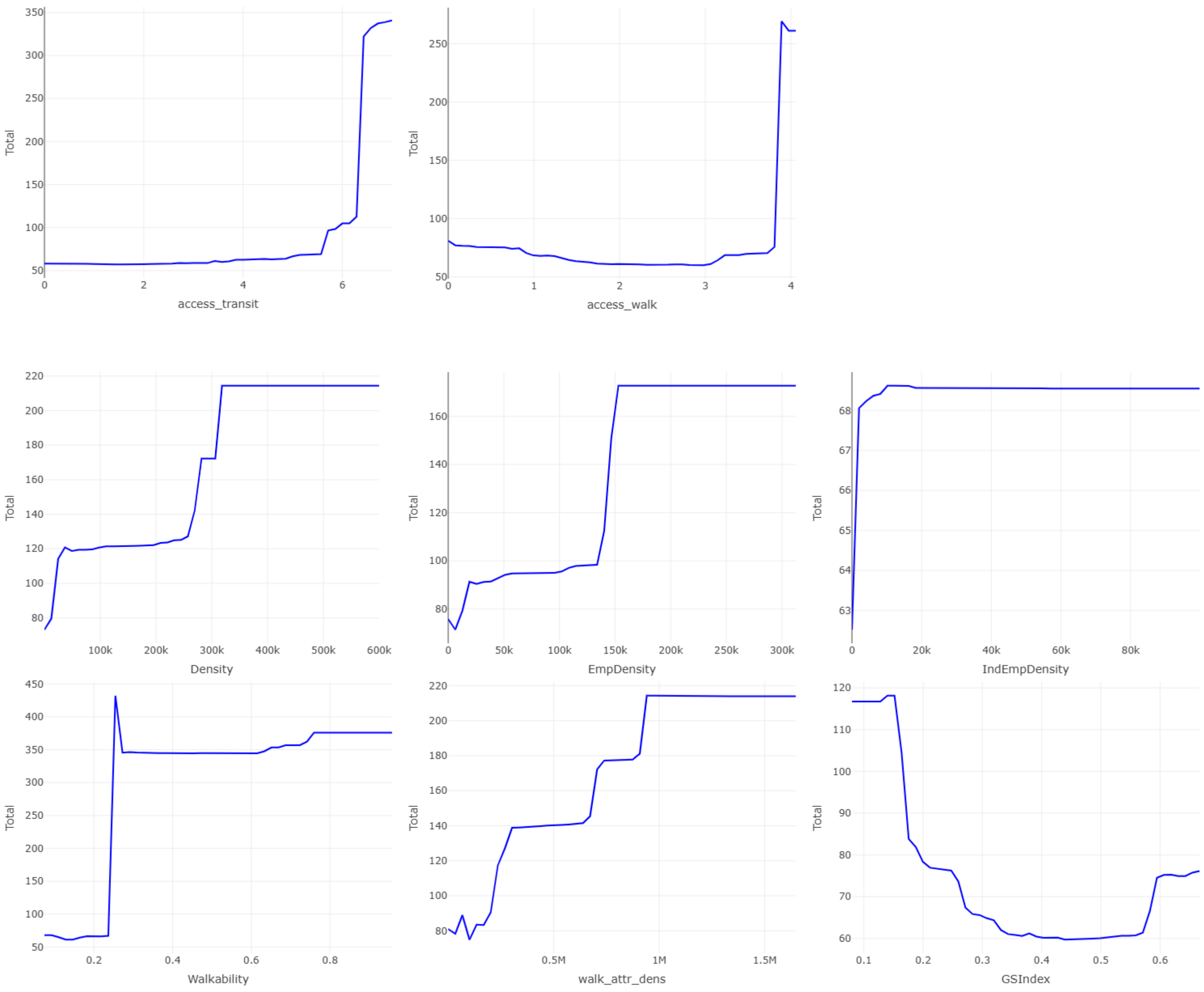
Variables such as accessibilities, walkability, and densities have significantly higher node purity increases in the HBW model, indicating that these variables are more important for predicting ridership for home-based work trips compared to non-work trips.

Partial Dependence Plots

Lastly, we examined how each variable interacts with predicted ridership. The random forest model only highlights the strength of the relationship between total ridership and each variable. The partial dependence plots provide deeper insights by illustrating how changes in each variable impact ridership.

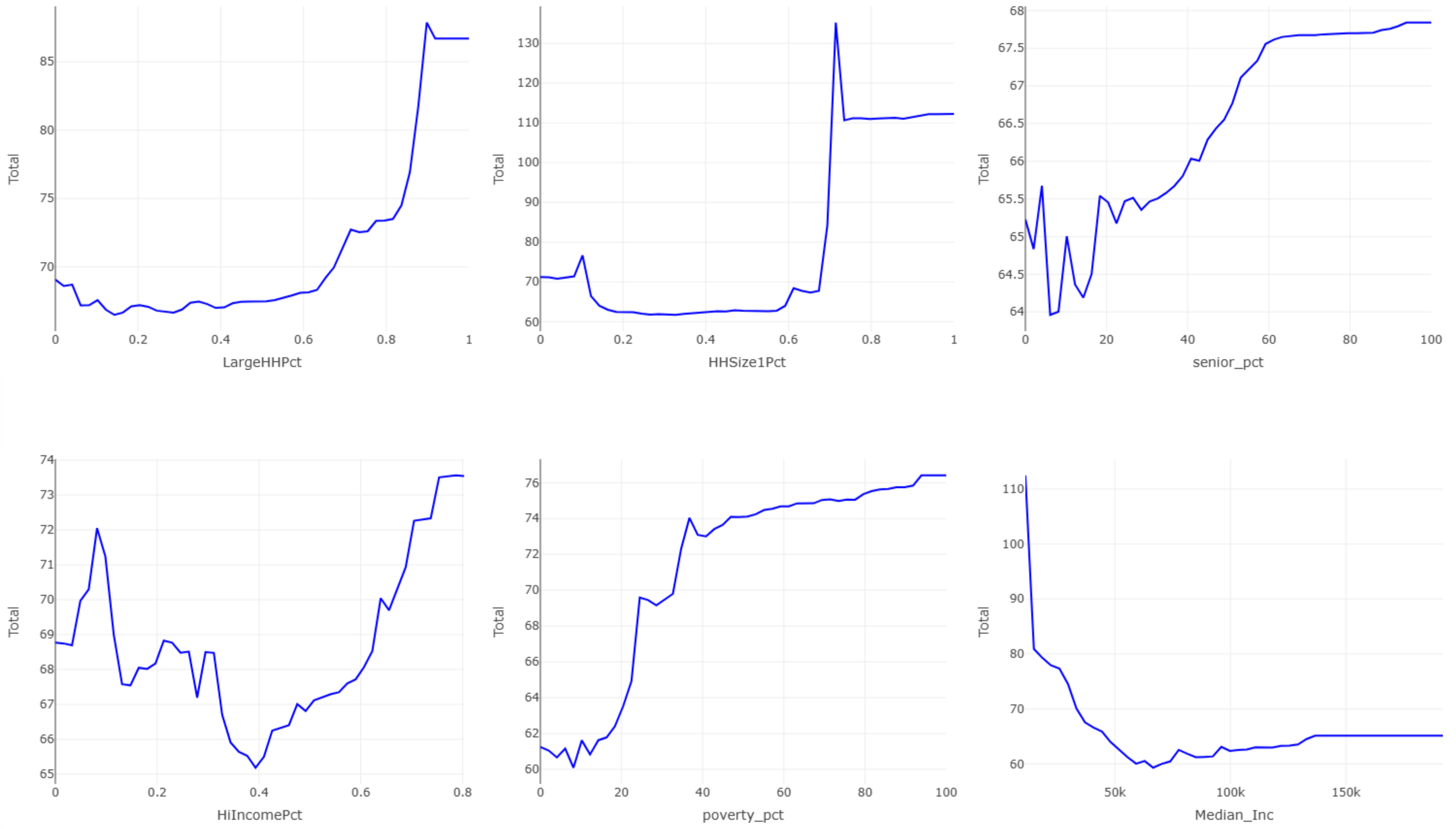
Accessibility

Land Use



As anticipated, areas with high transit ridership typically exhibit strong accessibility. Additionally, rising land use densities are associated with increased ridership. Interestingly, greater land use diversity, as measured by the GS index, does not consistently result in higher ridership. This may be because locations with high concentrations of offices and employment also tend to support substantial ridership, regardless of land use diversity.

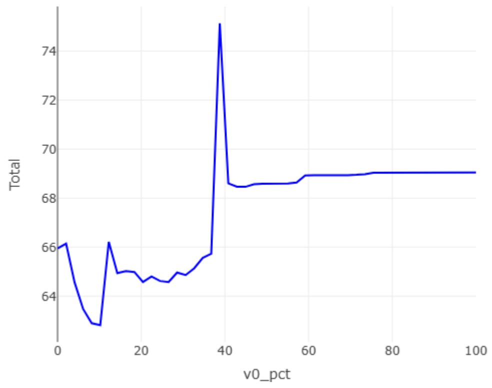
Demographic

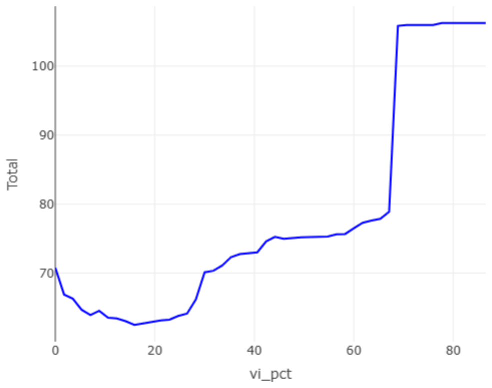


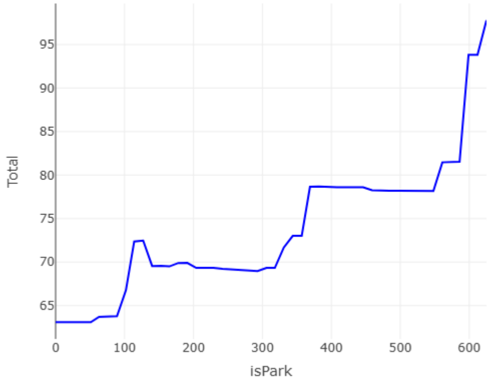
TAZs with high senior percentage and more low-income households are typically associated with high ridership based on the random forest model. Given this is a zone level analysis, TAZ income level does not relate to actual transit rider income. Results with the large household percentage and 1-person household percentage seem counterintuitive to each other. This may indicate that we are less likely to see ridership in zones with more moderate sized households.

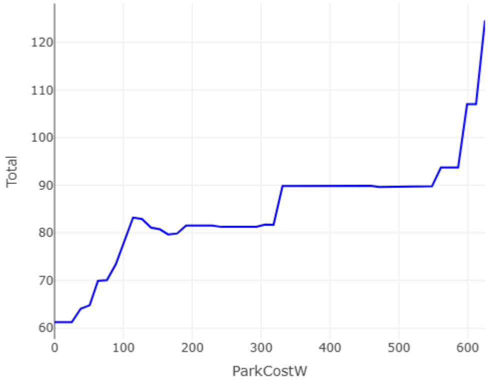
Auto Ownership

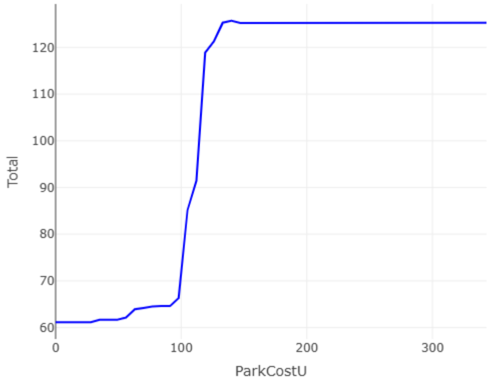
Parking

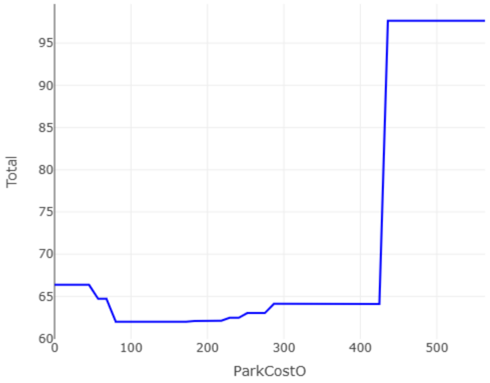












As expected, TAZ with high concentration of zero vehicle and vehicle insufficient households typically supports higher ridership. Transit is also more preferred where the parking cost is not free.

Utilizing random forest model gives us another way to predict transit ridership other than traditional travel demand modeling and STOPS modeling. We examine the interactions between ridership and accessibility, land use, auto ownership, demographic, and parking cost variables at aggregated zonal level. The results indicate that accessibility and land use variables have significant importance in predicting transit ridership at zonal level. This analysis also highlights the importance of further analysis at personal level as ridership is much driven by personal choices. A future analysis is recommended to incorporate transit service data as well as demographic and socio-economic data at personal level.

Summary

Short Break

Discussion Session #1: AI and Big Data

AI and Big Data | Proficiency / Expertise

~~A) 2 (disagree)~~

~~A) 1 (agree)~~

A) Not sure

A) I feel proficient with machine learning methods.

B) I feel proficient with using LLMs and other forms of generative AI tools.

C) I feel proficient in applying big data and location-based data (e.g., Streetlight, etc.) to transportation modeling questions.

…

Strongly Disagree			Neither Agree nor Disagree			Strongly Agree
-3	-2	-1	0	+1	+2	+3

AI and Big Data | Confidence / Trust

...

D) I have confidence in the accuracy and results of machine learning methods.

E) I have confidence in the accuracy and results of LLMs and generative AI tools.

F) I have confidence in the accuracy of big data and location-based data providers.

...

Strongly Disagree			Neither Agree nor Disagree			Strongly Agree
-3	-2	-1	0	+1	+2	+3

AI and Big Data | Personal Practice

...

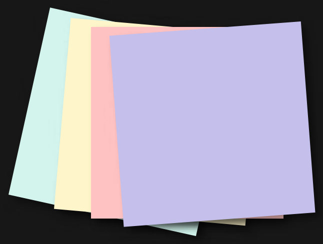
G) I embrace the use of machine learning methods.

H) I embrace the use of LLMs and generative AI tools.

I) I embrace the use of big data and aggregated location-based data.

Strongly Disagree			Neither Agree nor Disagree			Strongly Agree
-3	-2	-1	0	+1	+2	+3

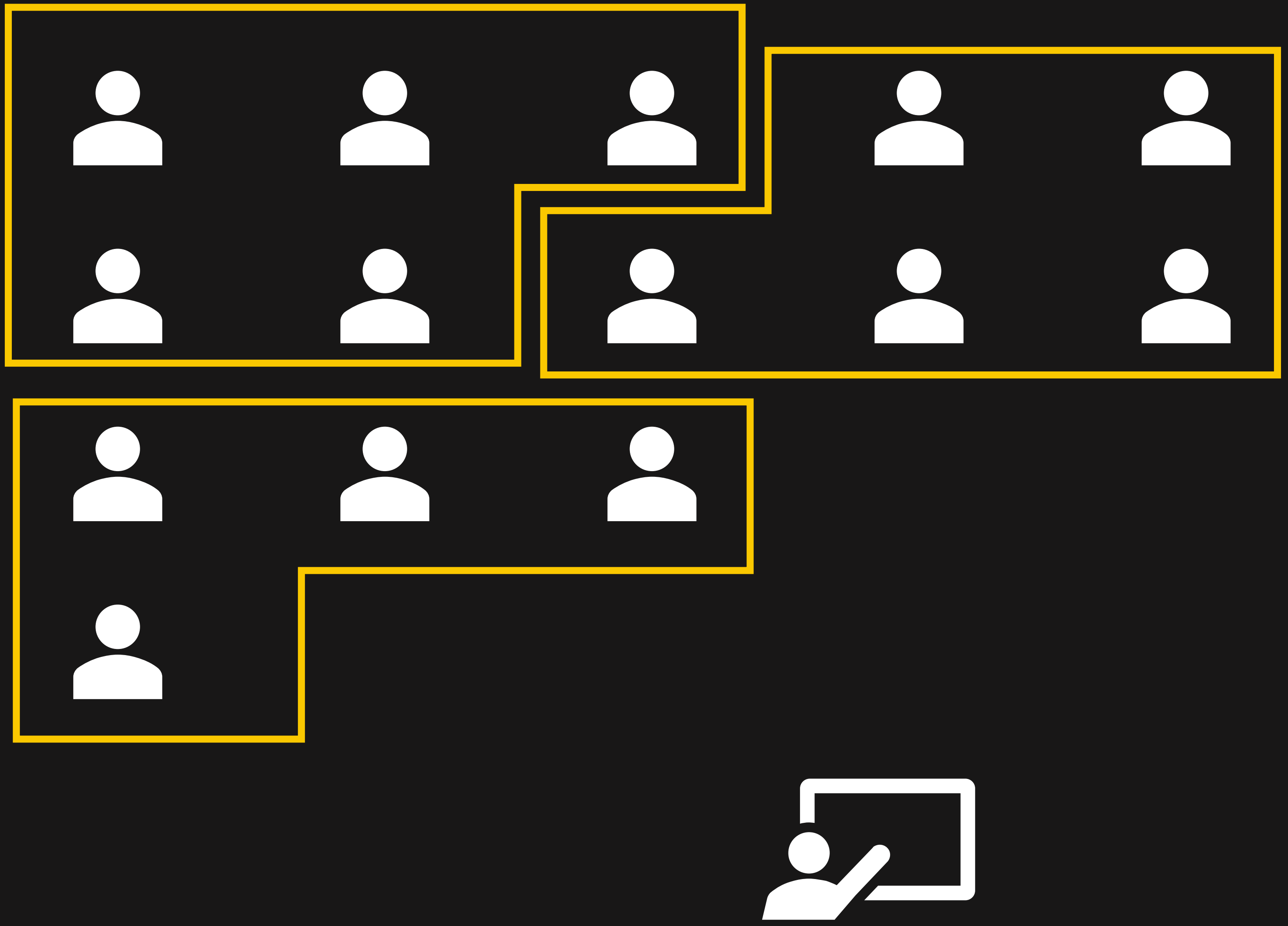
AI and Big Data



Pass the sticky notes down to the end of the table...

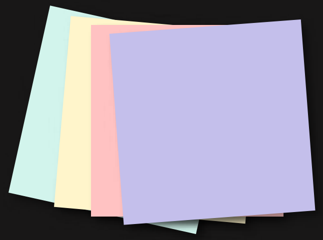
...and we will collect them

Form Groups Based on Proximity



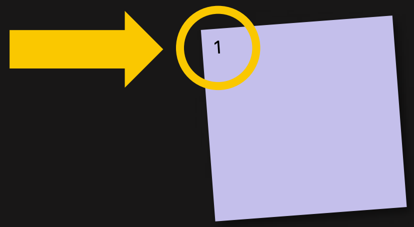
3 to 5 members
per group

AI and Big Data | Group Questions



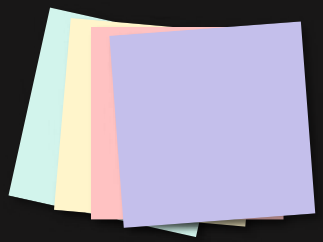
Use as many sticky notes as you like to answer these questions. You may answer as a group or independently.

1. How are you using AI (any form) in your work, and how do you plan to use it in the future?
2. How are you using big data in your work, and how do you plan to use it in the future?
3. What are some current limitations that you have encountered with big data?
4. Do you follow any guidance or protocols when it comes to the use of AI? If so, can you provide details?
5. Are there applications of AI or big data that you find exciting? What are they?
6. Are there applications of AI or big data that concern you? What are they?
7. What location-based data providers have you worked with? Please list them.



Please be sure to label the corner!

AI and Big Data



Pass the sticky notes down to the end of the table...

...and we will collect them

Discussion Session #2: The Future of the TRM

Discussion Session #2

Prompt 1

Prompt 2

Prompt 3

Prompt 4

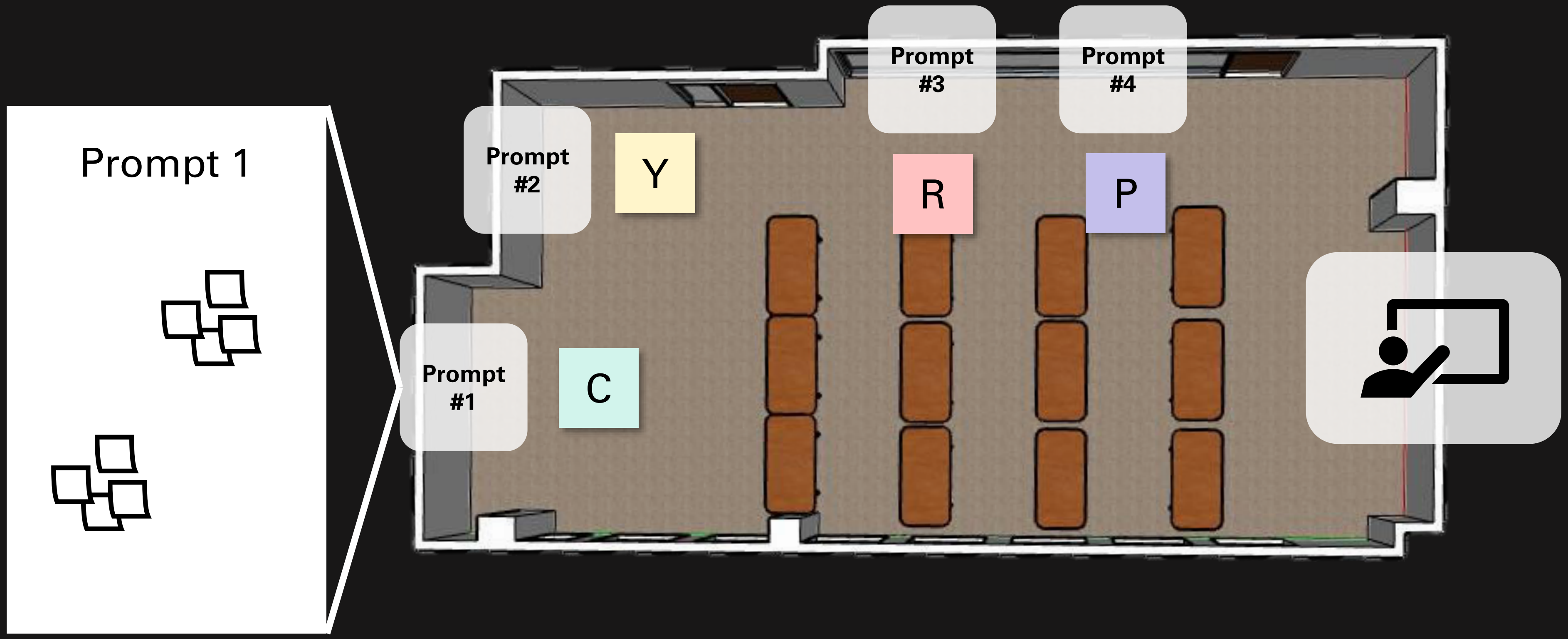
What are some short-term (accomplishable in 1-year) improvements to the TRM that you think would be most impactful?

What longer-term goals (3-5 year horizon) do you think the TRM modeling team should work towards?

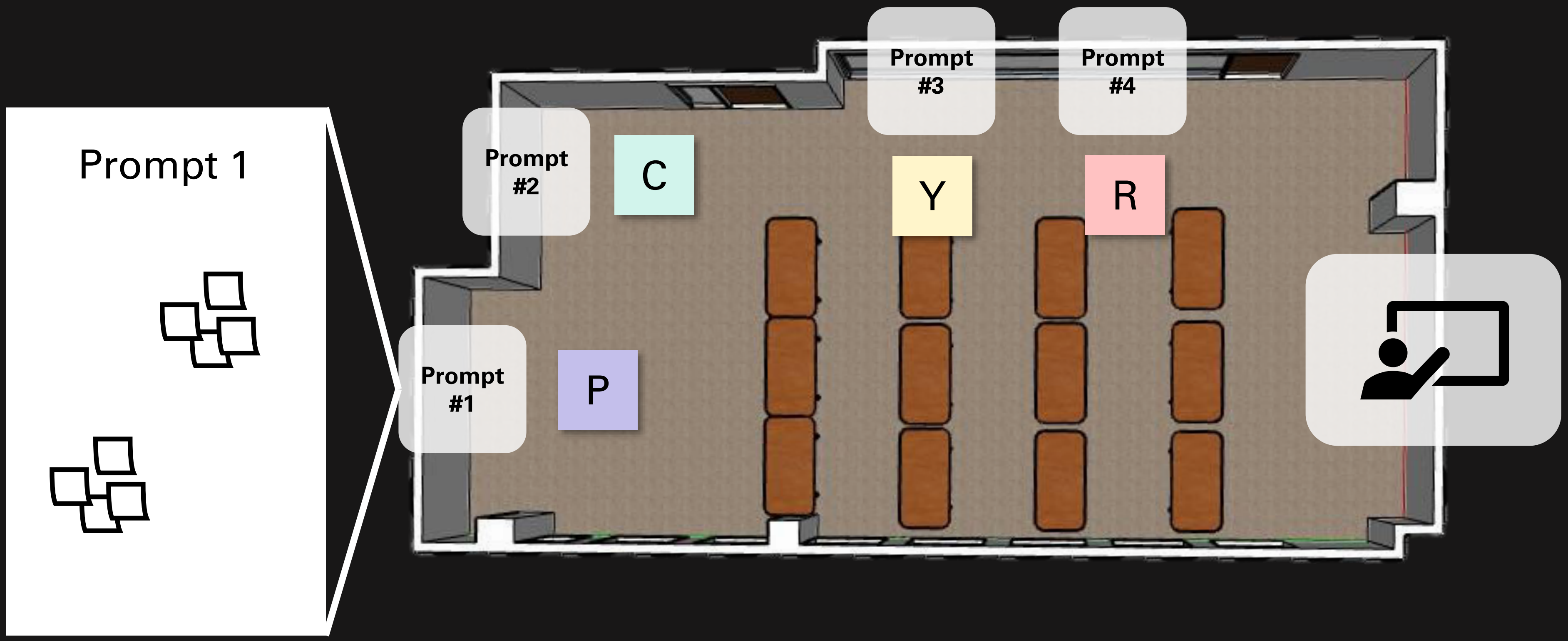
What do you see as the future of travel demand modeling (in terms of methods, new approaches, or key topics/issues)?

Do you have any thoughts or recommendations on ways to improve the model user experience?

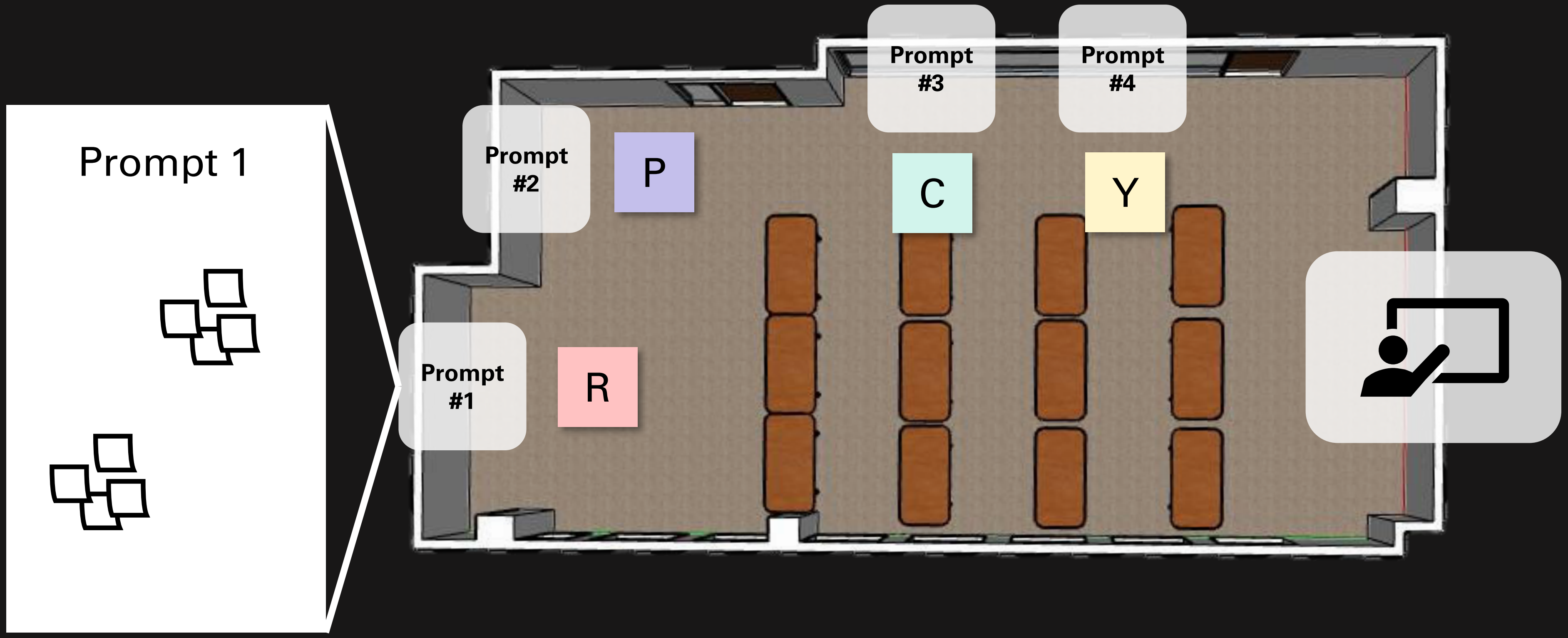
TRM | Round 1



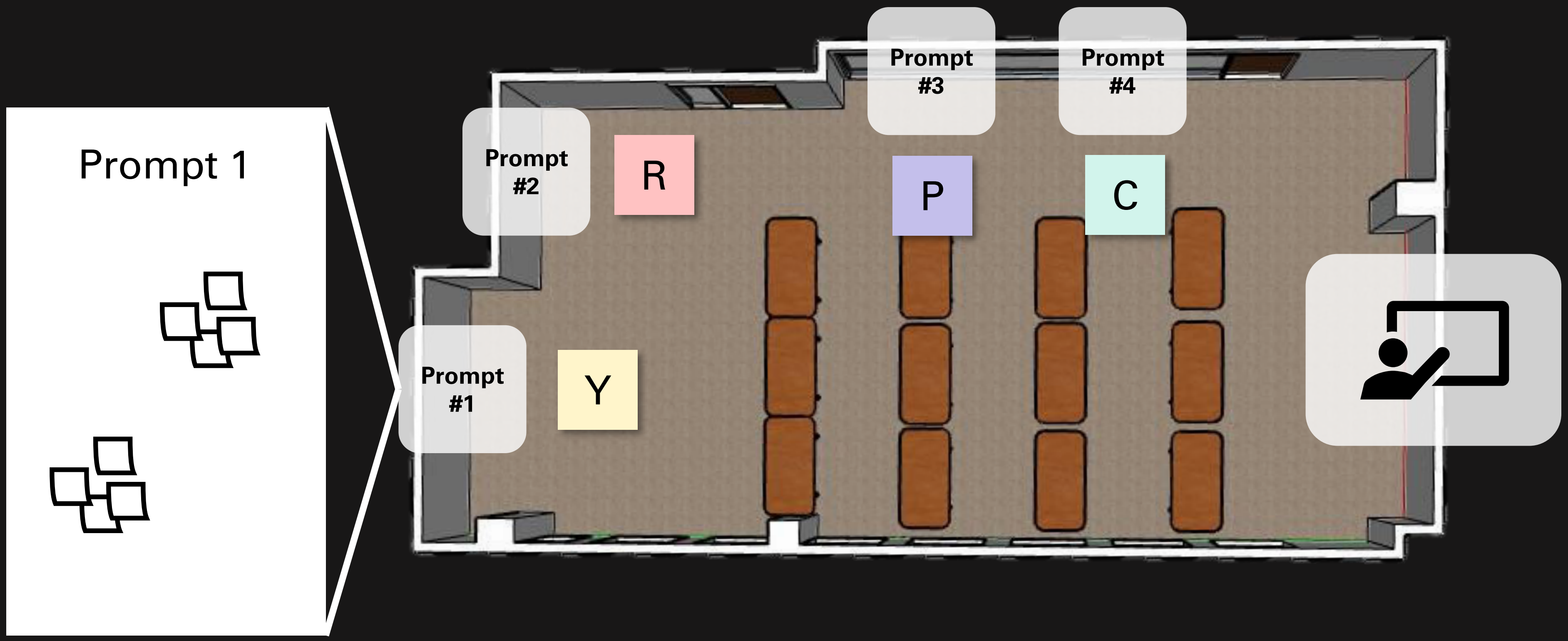
TRM | Round 2



TRM | Round 3



TRM Round 4



Report Back

Did any ideas stand out to you?

Are there any ideas that you are excited about?

Networking

ITRE
Institute for Transportation Research and Education
SYSTEMS PLANNING & ANALYSIS