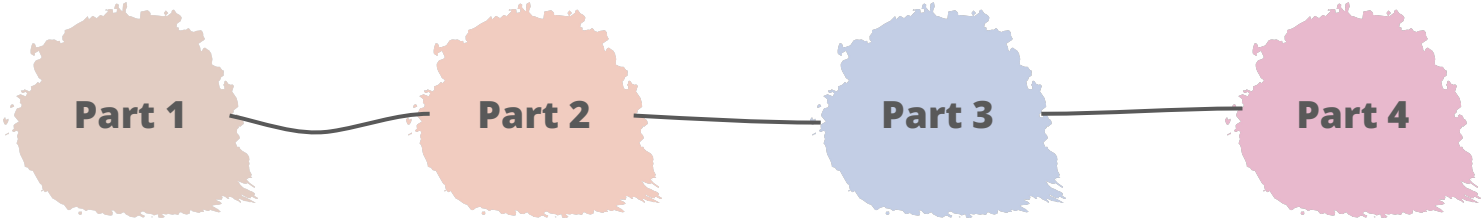
Better programs, better parks

Improving programs and visits in Philadelphia Parks & Rec with SafeGraph data

Jeff Stern, Hui Tian, Lan Xiao, Shaun Zheng

April 26 2022 Update

Table of Contents



Introduction

Use Case

Huff Model

Dashboard

https://tinyurl.com/pprWow

01

Introduction

About Philadelphia Parks and Rec (PPR)

01 Introduction

Philadelphia Parks and Rec **(PPR)** has **524 facilities** that host around **4,000 programs** annually throughout the city

FIND YOUR PATH

PHILADELPHIA PARKS & RECREATION

01 Introduction

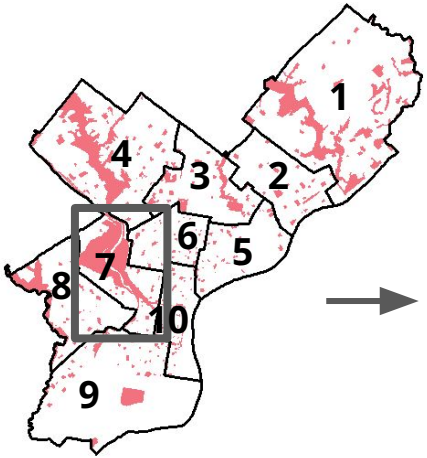
The objective:

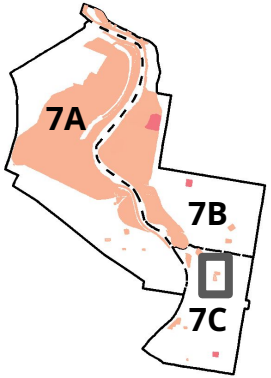
To help PPR schedule program activities by understanding visits

- How many visits 
- Where visit comes from 
- When visit happens 
- …

01

PPR Program Hierarchy





After School
Athletic
Camp
Cultural
Educational
Environmental
Others

10 Districts

Service Areas

Facilities (500+)

Program

02

Use Case

Find a dynamic relationship between programs & visitors

02

How does a PPR official schedule programs?

Business as usual

Refer to historical schedules and even count garbage

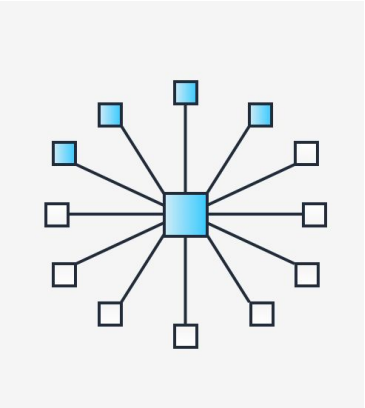
?

We propose!

Use SafeGraph data to allocate programs effectively

What is SafeGraph data?

- **In a simple way**



SafeGraph Device panels to collect users' location data

E.g

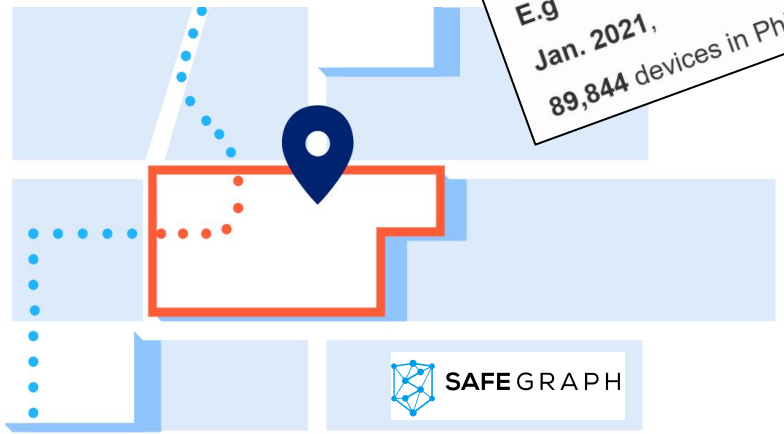
Jan. 2021,

Put 89,844 devices in Philadelphia,

e.g. 74 devices in cbg 421010117001

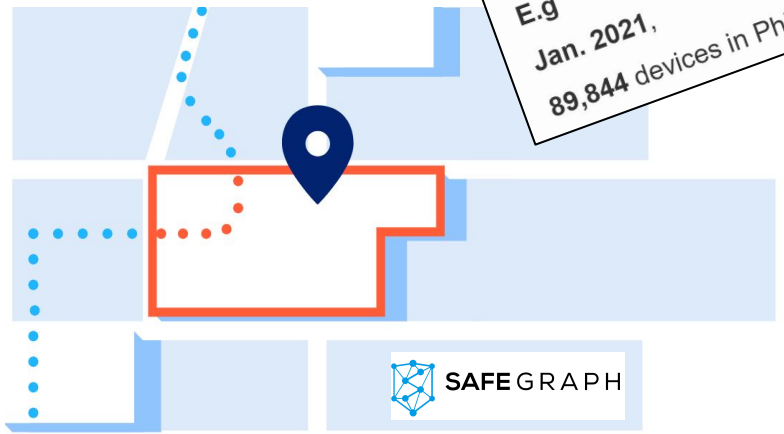
Collect daily footprints and **aggregate it into Pattern data**

SafeGraph provides with rich data about people's travel patterns - (in device panel)



Location Name	Design School
Date Range	01/01/2021 - 01/05/2021
Visit Counts	**10**
Visitor Home CBGs	**"cbg1"**: 5, **"cbg2"**: 3, **"cbg3"**: 2
Bucketed Dwell Times	**"<5"**: 2, **"5-10"**: 0, **"11-20"**: 8
Visits by day	**"d1"**: 2, **"d2"**: 1, **"d3"**: 3, **"d4"**: 2, **"d5"**: 2
Popularity by hour	**"h1"**: 0, **"h2"**: 1,….., **"h24"**: 0

SafeGraph provides with rich data about people's travel patterns - (in device panel)



E.g
Jan. 2021,
89,844 devices in Philadelphia,

Location Name	Design School
Date Range	01/01/2021 - 01/05/2021
Visit Counts	**10**
Visitor Home CBGs	**"cbg1"**: 5, **"cbg2"**: 3, **"cbg3"**: 2
	': 2
Popularity by hour	**"h1"**: 0, **"h2"**: 1,….., **"h24"**: 0

Need to be scaled based on **number of the population and the safegraph devices** in each census block group

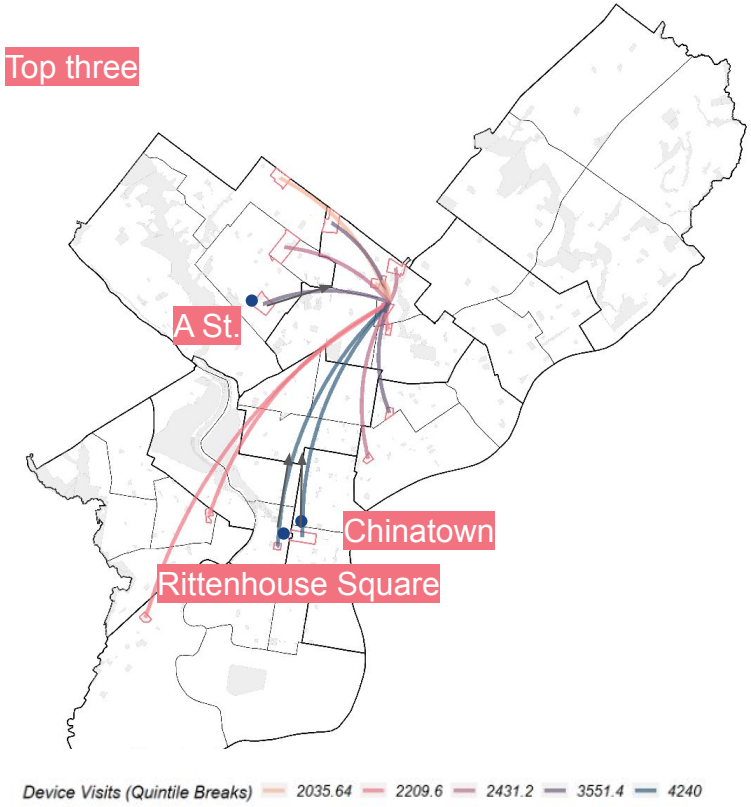
Application1:

For each facility, where do people come from?

Facility: Charles J Ziehler Playground

Application1:

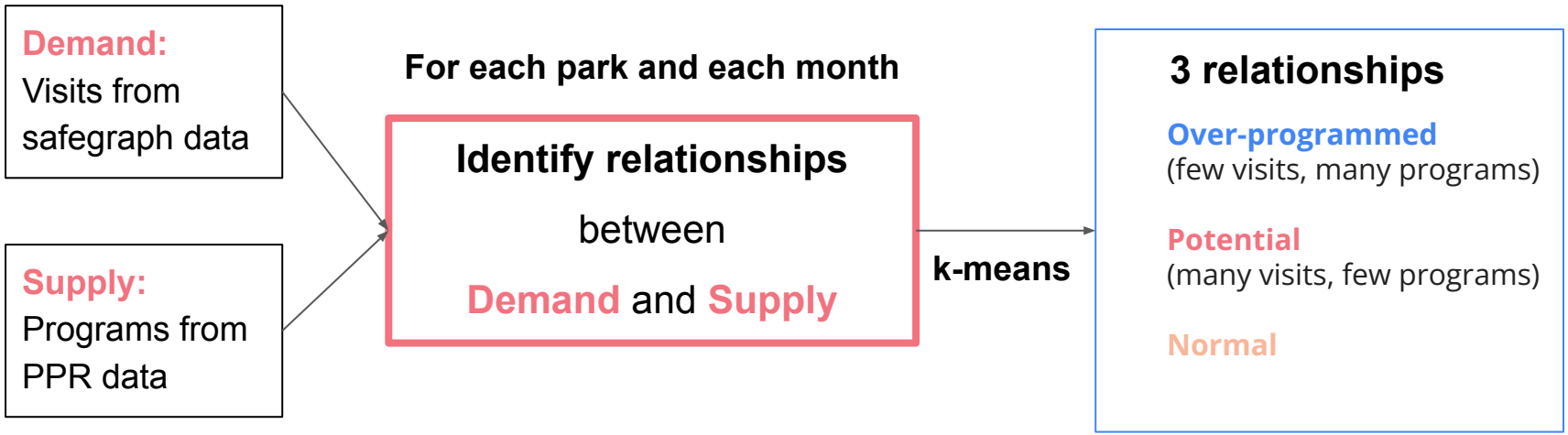
For each facility, where do people come from?



Facility: Charles J Ziehler Playground

Application2:

Inspect current program allocations

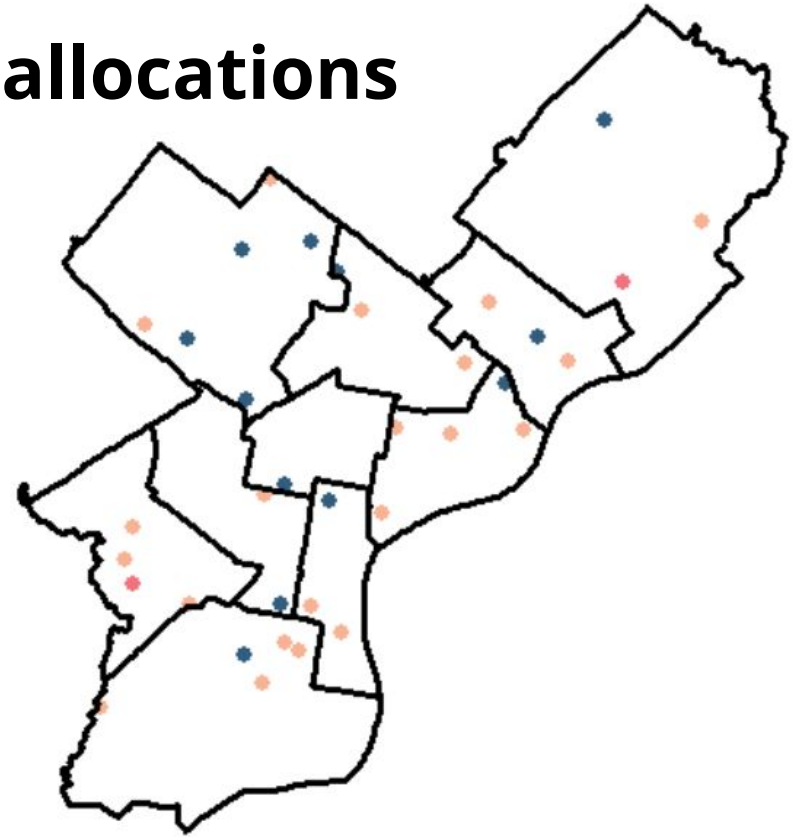


Application2:

Inspect current resource allocations

3 relationships

- **Over-programmed** 35 place~month
  (few visits, many programs)
- **Potential** 12 places~month
  (many visits, few programs)
- **Normal** 333 places~month



Relationship Distribution in October

Case of Potential Location: Albert Christy Rec Center

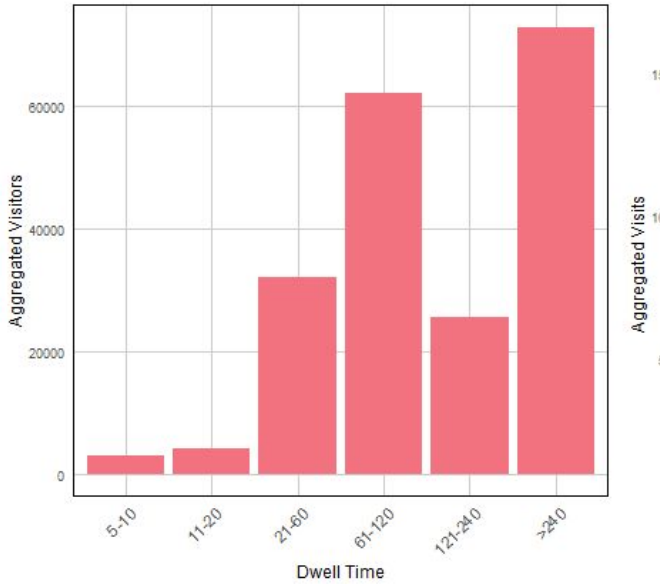
Potential Site for 2021

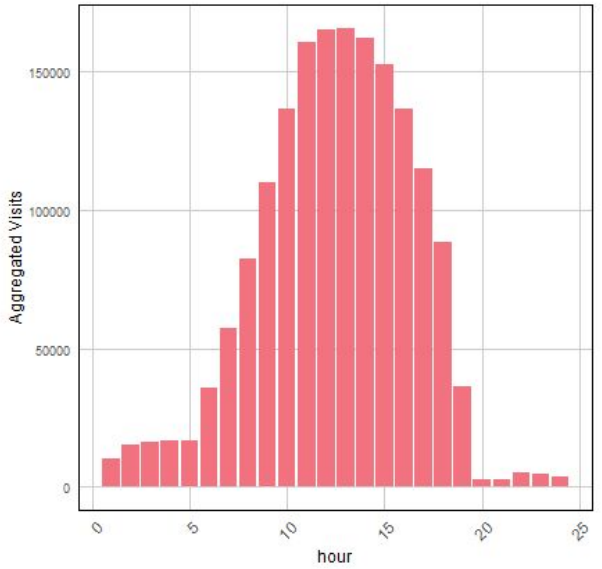
Demand: High Visits
(1000+ visitors on a single day)

Supply: Small 33 Programs
(Soccer & Afterschool Activity)

Case of Potential Location: Albert Christy Rec Center

Most visitors stay more than 1 hour, or even 4 hours. The rush hour is 11am - 3 pm.

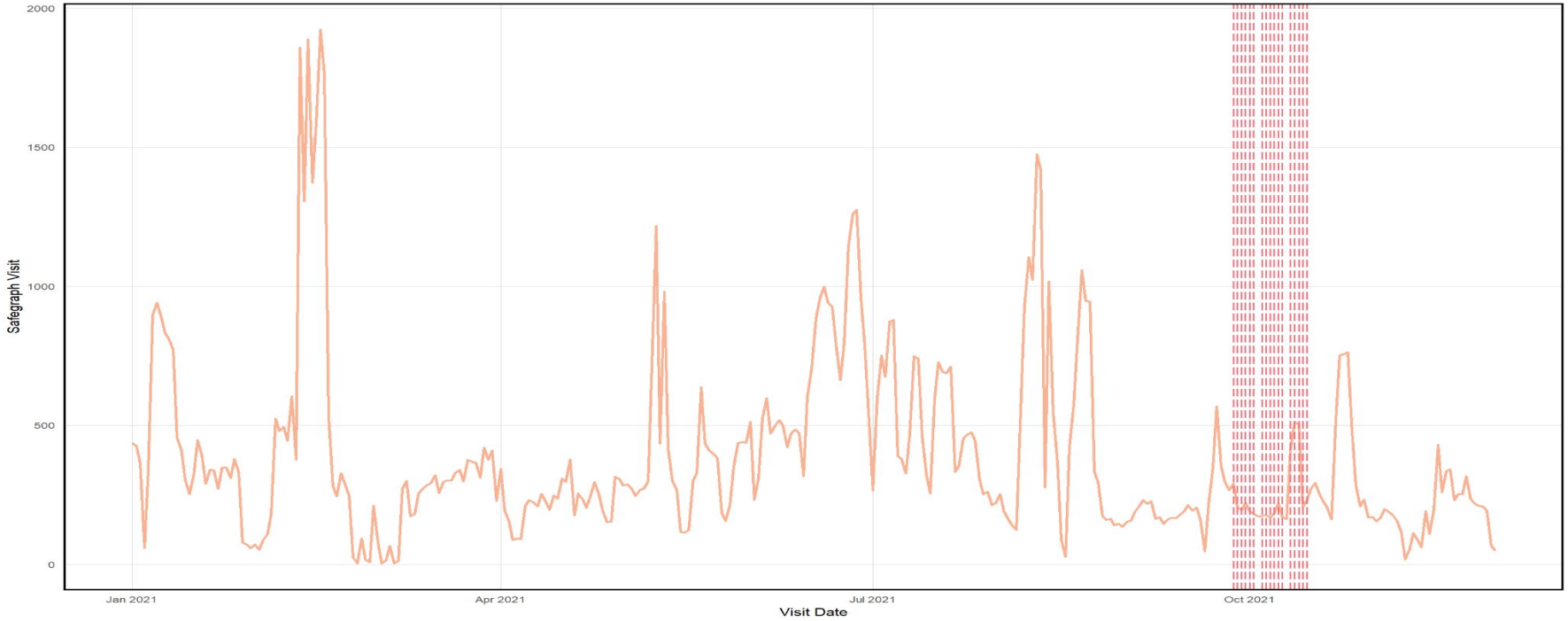




Case of Potential Location: Albert Christy Rec Center

Visits recorded by safegraph

Program

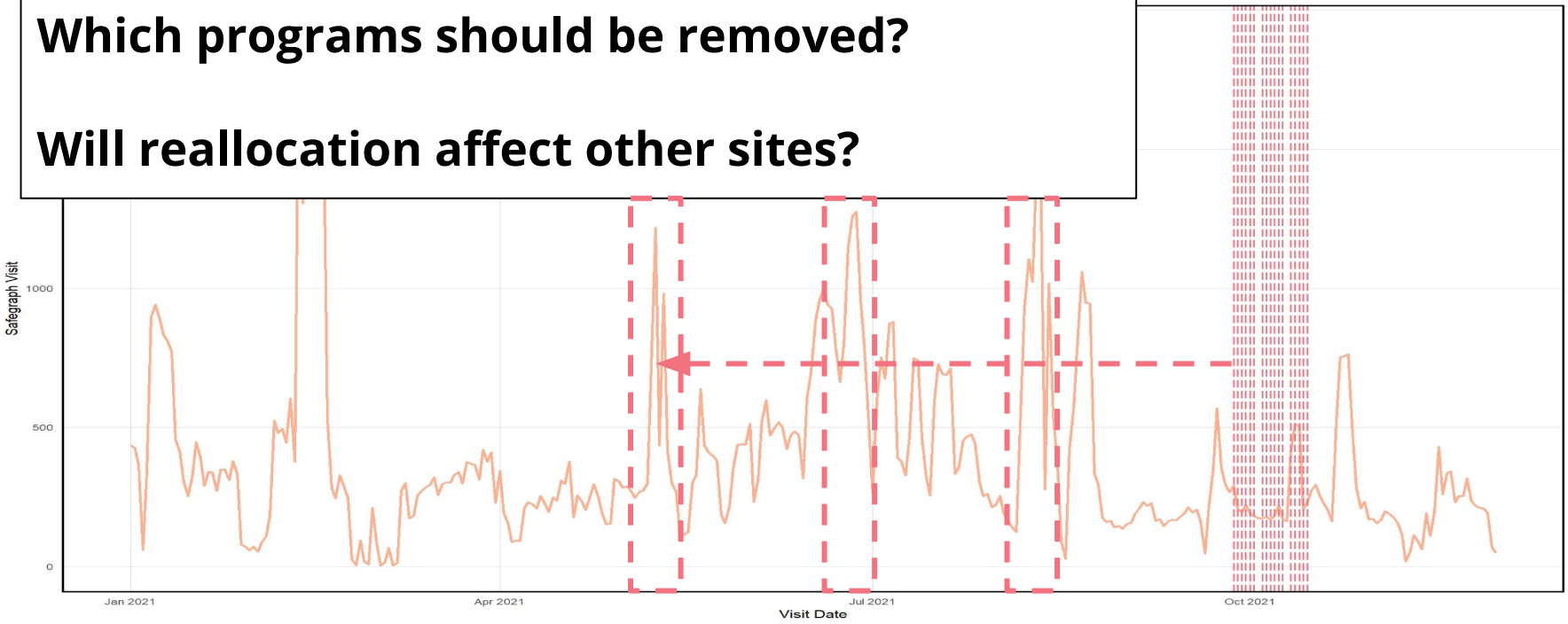


Case of Potential Location: Albert Christy Rec Center

How many additional programs are needed?

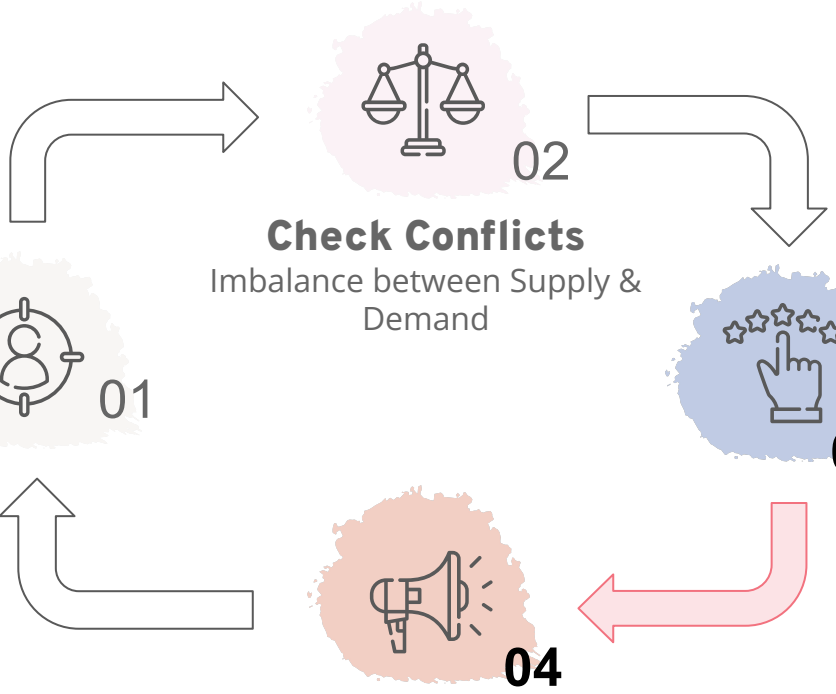
Which programs should be removed?

Will reallocation affect other sites?



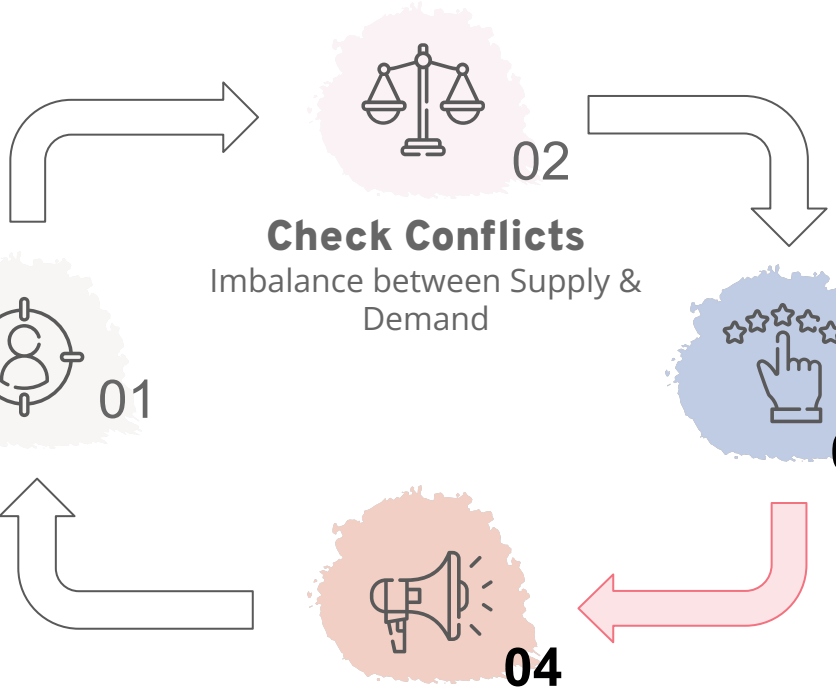
02

Use Case Diagram



02

Use Case Diagram



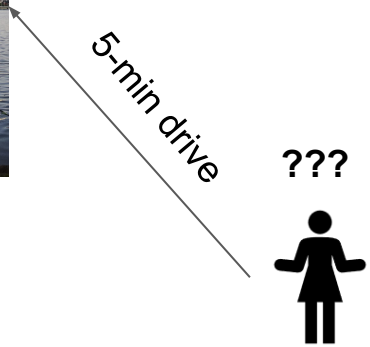
03

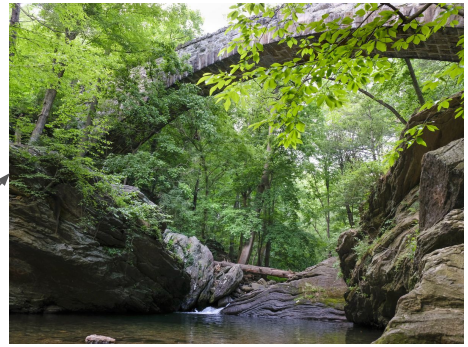
Huff Model
With Neighboring Effect

What is Huff Model?

Huff model is a tool for predicting the **probability** of a consumer **visiting a site,** as a function of the distance, its attractiveness etc.

Schuylkill river park
Area: 14.72 Acre
Program: Schuylkill River Sojourn
…





Wissahickon valley park
Area: 2,064.12 Acre
Trail Mile: 82,428
Program: none
…

03

Model Usage

Model Input

- Attractiveness of Parks
- Neighboring Attractiveness
- Distance
- Route Visit Frequency Level

Huff Model

Model Output

The predicted probability of **a park visit from each census block group**

Final Result

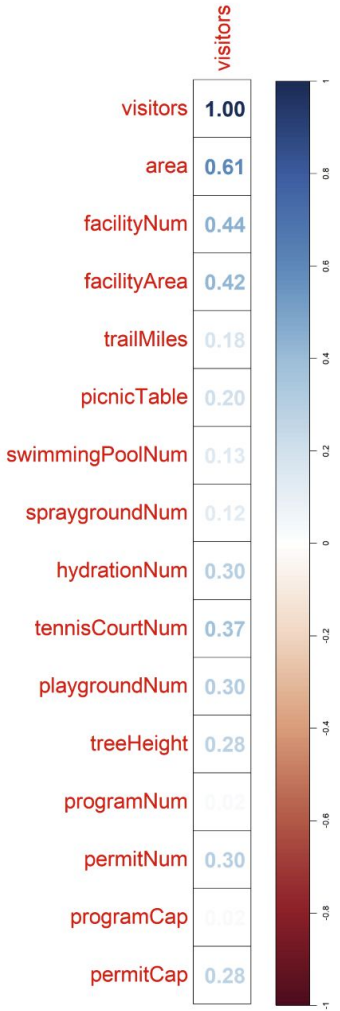
- **Market Areas**
- The predicted # of visitors

03

Model Input

- **Attractiveness of Parks**
- Neighboring Attractiveness
- Distance
- Route Visit Frequency Level

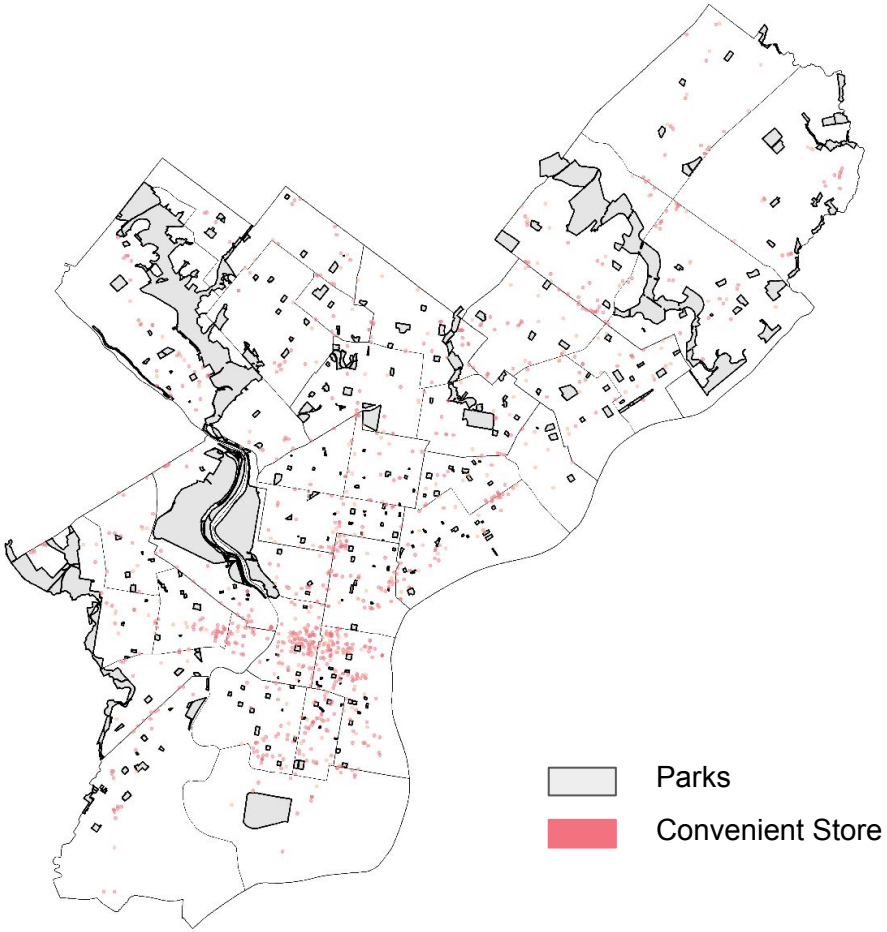
- Park Area
- Facility Areage
- Trail Mileage
- Average Tree Height
- # of Facilities
- # of Picnic Table
- # of Swimming Pool
- # of Sprayground
- # of Hydration
- # of Tennis Court
- # of Playground
- # of Program
- # of Permit



03

Model Input

- Attractiveness of Parks
- **Neighboring Attractiveness**
- Distance
- Route Visit Frequency Level



03

Model Input

- Attractiveness of Parks
- Neighboring Attractiveness
- **Distance**
- Route Visit Frequency Level

Park 1

Census block group **1**	5-min-drive
Census block group **2**	3-min-drive
Census block group **3**	11-min-drive
Census block group ...	...
Census block group **1336**	20-min-drive

... ...

... ...

... ...

... ...

Park 273

Census block group **1**	4-min-drive
Census block group **2**	6-min-drive
Census block group **3**	1-min-drive
Census block group ...	...
Census block group **1336**	32-min-drive

03

Model Input

- Attractiveness of Parks
- Neighboring Attractiveness
- Distance
- **Route Visit Frequency Level**

Probability	Frequency Level
0.5 ~ 1	Super High
0.1 ~ 0.5	High
0.03 ~ 0.1	Medium
0.015 ~ 0.03	Mid-Low
0.009 ~ 0.015	Low
0 ~ 0.009	Super Low

03

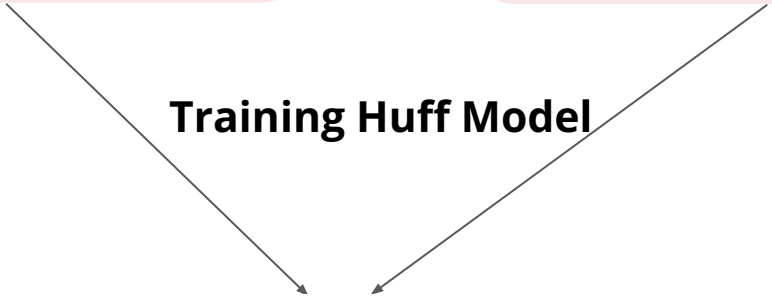
Huff Model Training

Input - Predictors

- Attractiveness of Parks
- Neighboring Attractiveness
- Distance
- Route Visit Frequency Level

Output - Actual Target Value

SafeGraph records: possibilities of a visit from each census block group to each park

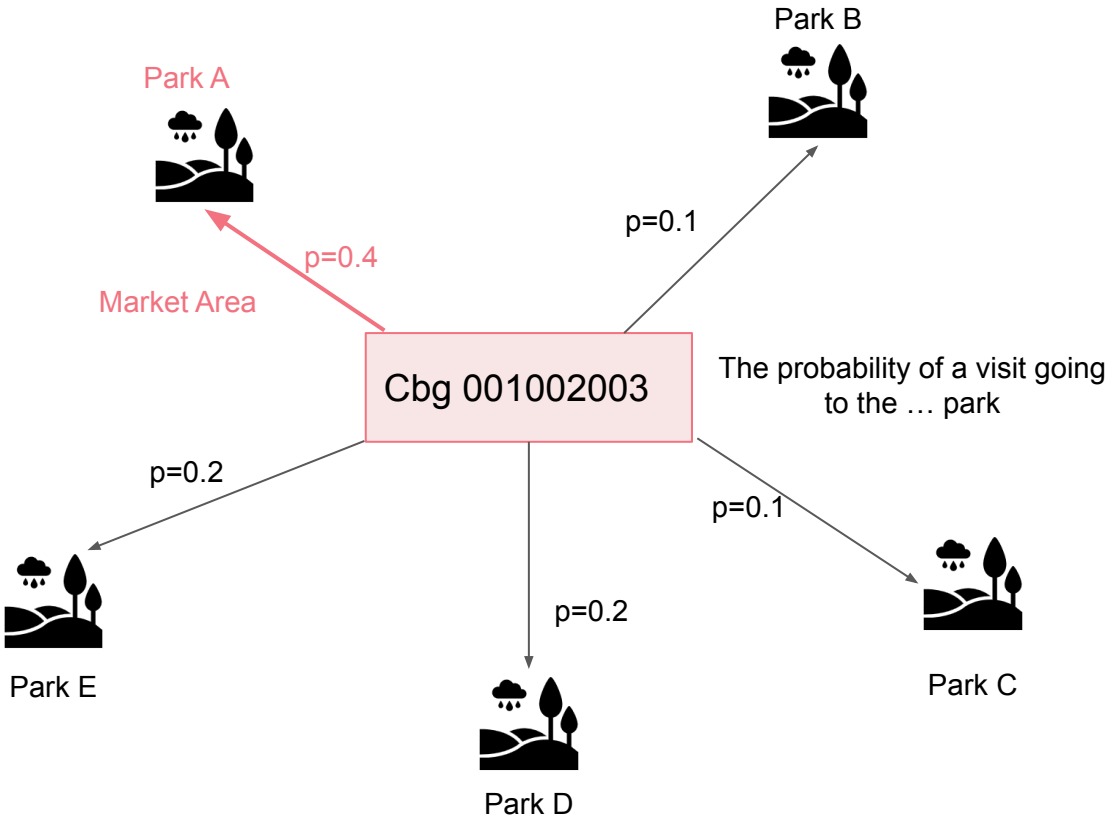


Training Huff Model

When **change the Attractiveness (program numbers)** of Parks, models give us **predicted probability,** Calculate **Market area** and number of **predicted visits**

03

Model Output

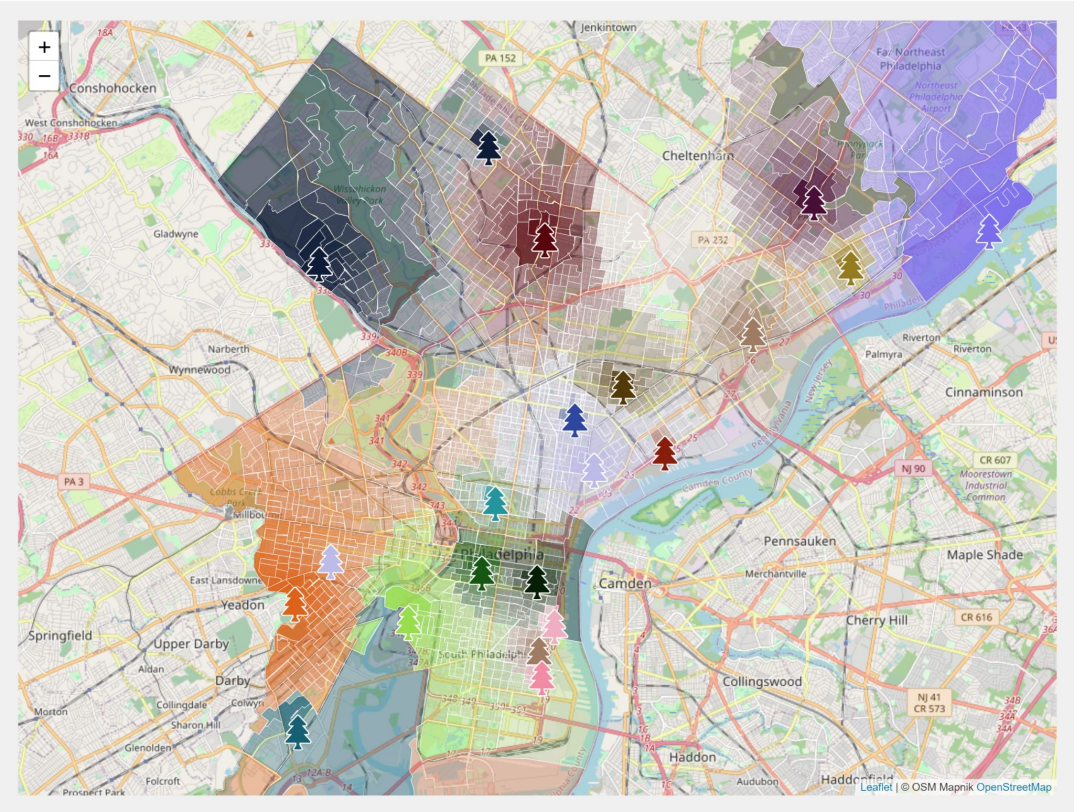


Final Result

- **Market Areas**
- The predicted # of visitors

03

Model Output



Final Result

- **Market Areas**
- The predicted # of visitors

03

Model Application

Albert Christy Rec Center

Identified potential sites,
expect to assign more programs here

↓

Adjust Program Numbers

↓

Attractiveness Change

↓

Probability Prediction Change

↓

Output New Visits Prediction

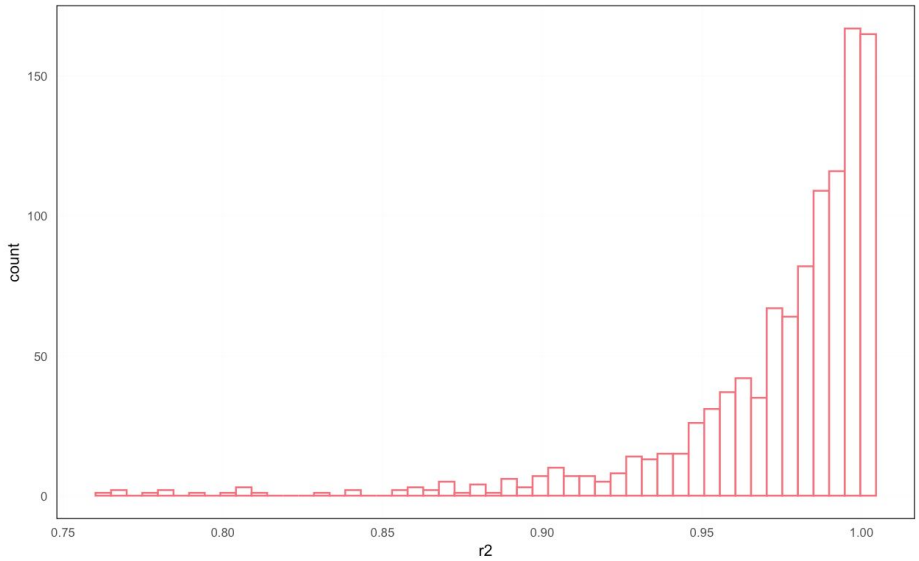
Market Area Prediction

Till supply matches demand

03

Model Scores

R-squared Distribution Histogram
for Optimized Model



1000+ Models

(One per census block group)

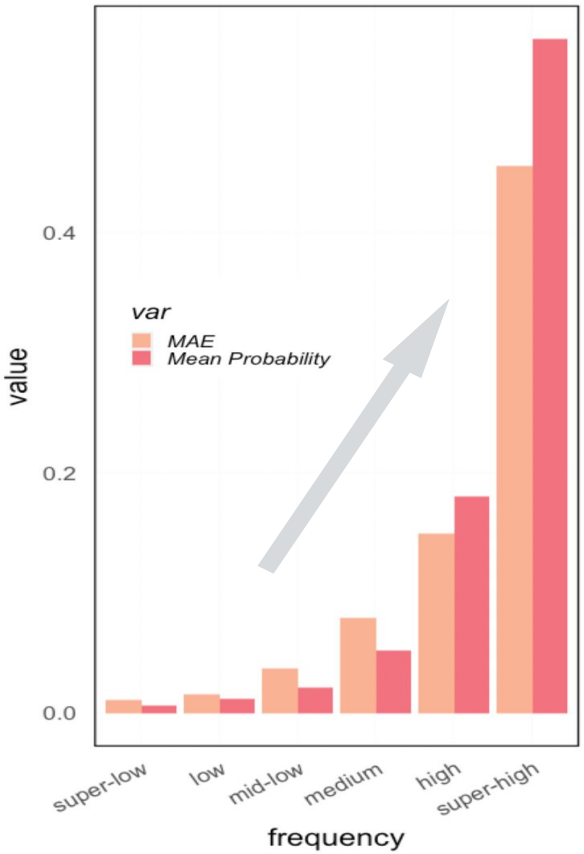
Average R^2 : 97%

(The percentage of the variance in the target value has been explained)

03

Model Accuracy among Frequency Levels

Route Visit Frequency **VS** Error



(Mean Absolute Error)

Average **MAE** **of visit probability:** **0.028**

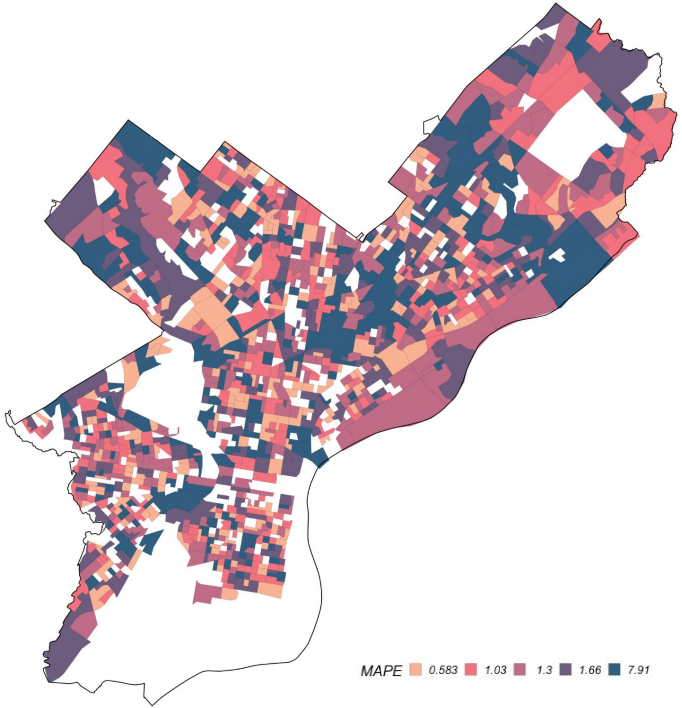
MAE Increases with Visit Probability

Models are more accurate for frequently visited routes

Model Accuracy in Different Aggregation

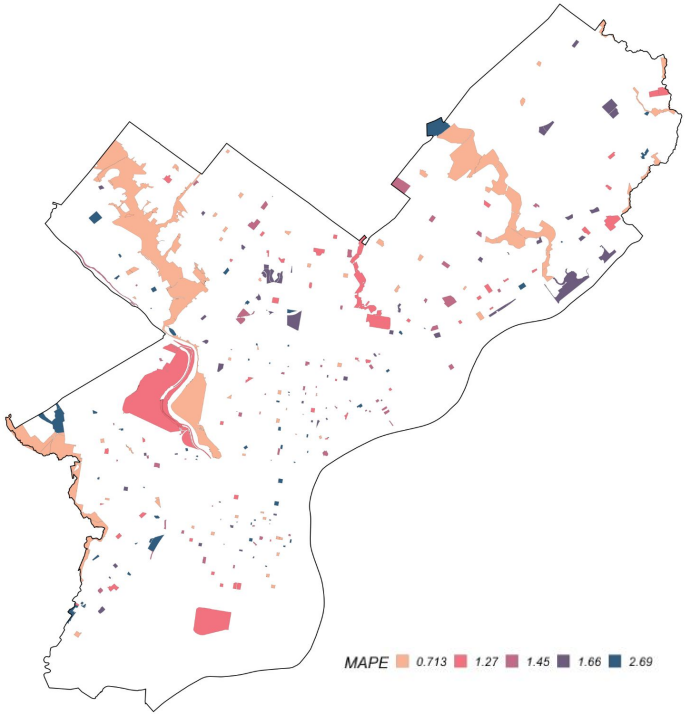
Models are more accurate for denser residential block groups

MAPE Distribution Map for Census Block Groups



Models are more accurate for large parks and parks in central city

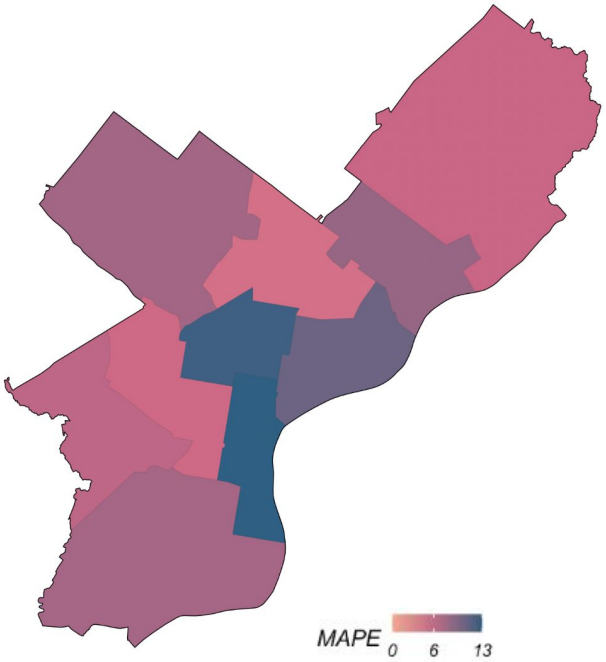
MAPE Distribution Map for Parks



03

Model Generalizability

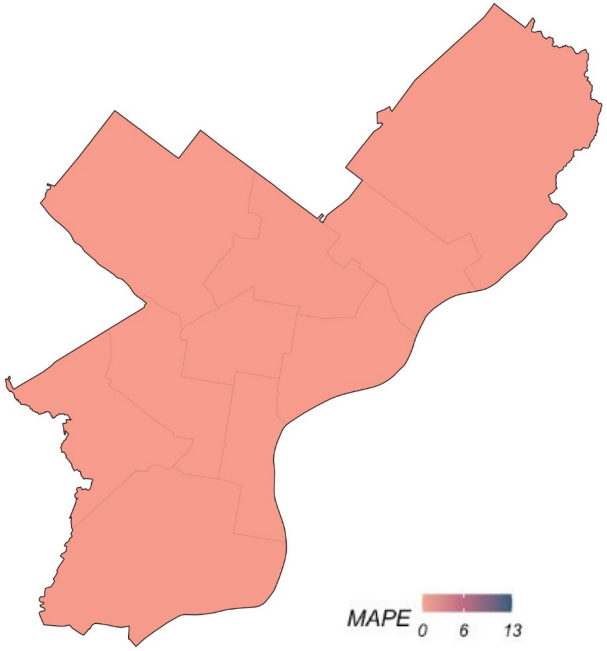
MAPE from **Spatial Cross-Validation**
with **before-Optimized Model** for PPR Districts



Take **route visit frequency level**
into consideration

Model become
more generalizable

MAPE from **Spatial Cross-Validation**
with **Optimized Model** for PPR Districts



04

Dashboard

Wireframe and UX Design

https://tinyurl.com/pprWow

THANKS

Questions?

Appendix

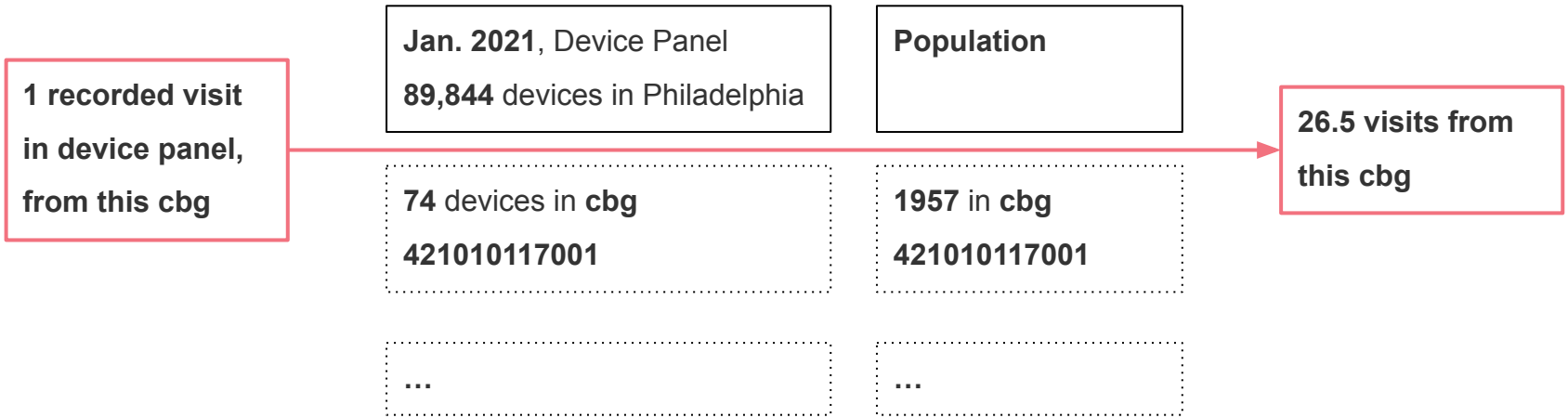
Normalized Data

$$\text{Quotient} = \frac{\text{population}}{\text{Sg Residing Devices}}$$

For each cbg

$$\text{Scaled Visits} = \text{Sg Visits} * \text{Quotient}$$

SafeGraph still needs **Normalization**



How might we program and allocate staff differently?

Identify Conflicts between **Demand** and **Supply**

(Device Panel) **Visits** vs **Activities**

Facilities are divided into **three groups**

Table 4.1 Mean values of clusters for conflicts

cluster	group	visitors	programs	size
1	normal	2647.048	3.936937	333
2	potential	28475.018	3.250000	12
3	over-programmed	3241.220	66.514286	35