TRAFFIC CRUISING

DATA SCIENCE FOR SOCIAL GOOD @ UW

FINAL PRESENTATION | 08.17.17

BRETT BEJCEK

DSSG FELLOW

ORYSYA STUS

DSSG FELLOW

MIKE VLAH

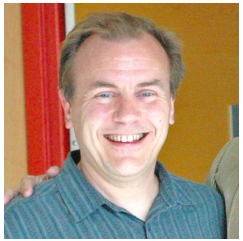
DSSG FELLOW

ANAMOL PUNDLE

DSSG FELLOW

VALENTINA STANEVA

DATA SCIENTIST



VAUGHN IVERSON

DATA SCIENTIST

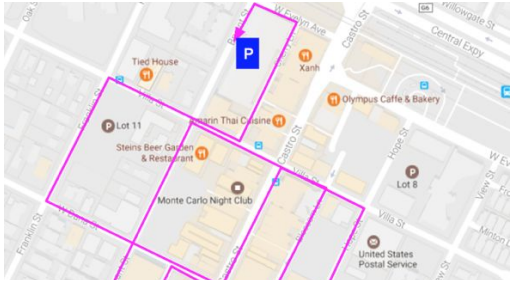
STEVE BARHAM

PROJECT LEAD

LARGE AMOUNT OF CONGESTION CAUSED BY TRAFFIC CRUISING

LOOKING FOR PARKING

DEADHEADING VEHICLES FOR HIRE

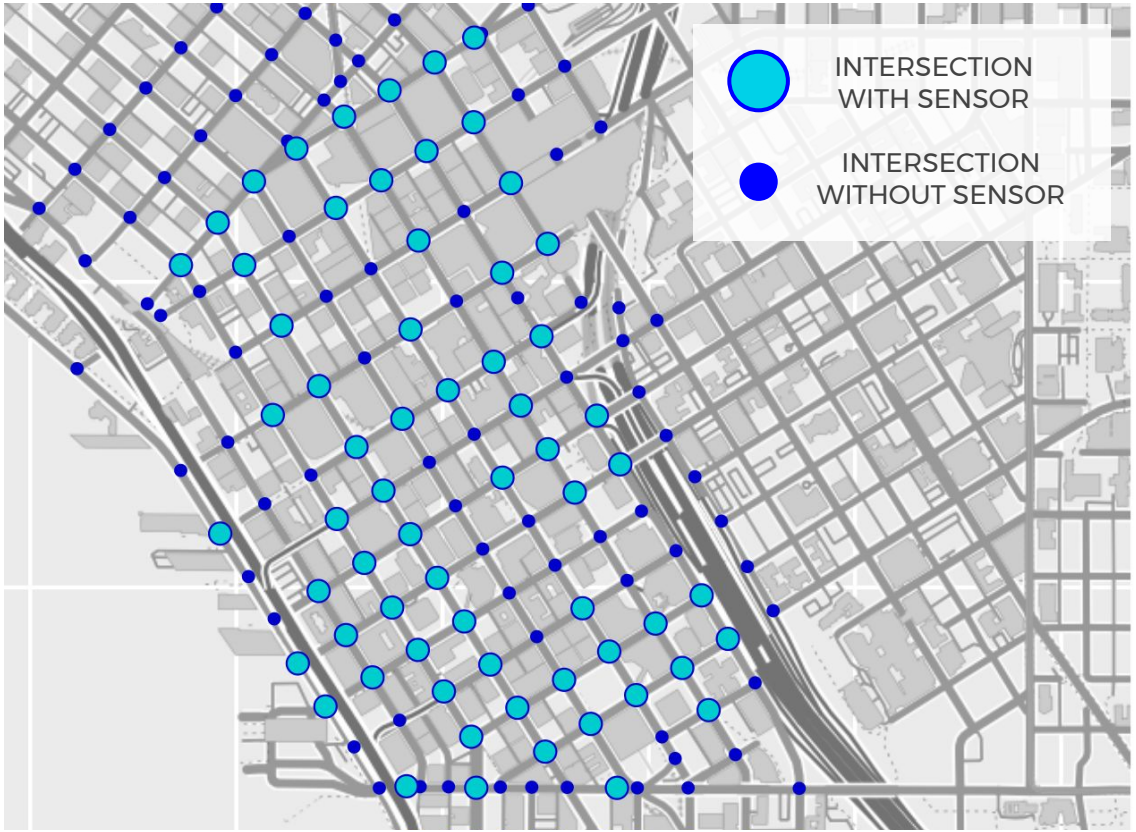


GOOGLE MAPS

10 MILES 7-13 MILES

SCHALLER CONSULTING

PROBLEM



HASHED MAC	TIME	SENSOR	STRENGTH
KD98SDK8AH	8:32:01	276105	-52
8DJSKDLSX0	8:32:01	276102	-55
439WOA09A	8:32:01	265402	-75
777AJDKAL8	8:32:05	293010	-50
QKSJ239A99	8:32:07	251040	-45
DQWPPOA09	8:32:10	265402	-49
KD98SDK8AH	8:32:11	265302	-54

PROBLEM

TECHNICAL CHALLENGES

INCOMPLETE GRID

SENSORS ONLY COVER

37%

OF THE GRID.

BIG DATA

SENSORS PRODUCE

200K

OBSERVATIONS / HR.

SENSOR DETECTION

SENSORS DETECT

38%

OF DEVICES W/ WIFI ON.

PRIVACY AND DATA GOVERNANCE

INEXACT LOCATION

AGGREGATED VIEWS

DATA IS ANONYMIZED

RAW DATA NOT RETAINED

DEPLOYABLE BEHIND TRANSPORTATION DATA COLLABORATIVE

OBJECTIVES

1

REPURPOSE SENSOR NETWORK

2

DIFFERENTIATE TYPES OF TRAFFIC CRUISING

3

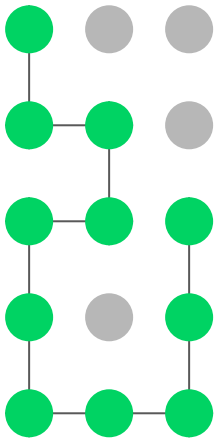
VISUALIZE WITH A HEAT MAP

4

SCALE UP AND DEPLOY

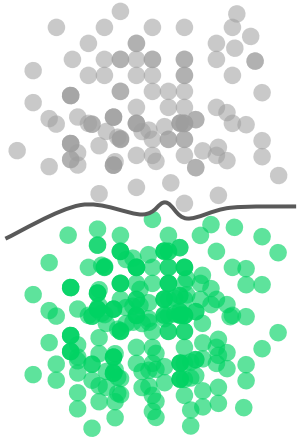
ENCRYPTED HASH	TIME	SENSOR
KD98SDK8AHD8X	8:32:00	276105
8DJSKDLSX0WKK	8:32:01	276102
DQWPPOA09DSD	8:32:01	265402
KDOSPALDKSX03	8:32:01	265302
DJKFOSPHSHAH8	8:32:01	265101
0138DSJCVBNNAA	8:32:01	273777
0DLPZZZZ98A999	8:32:02	265000
NVMCSKDJALKSSS	8:32:02	265103
QALSJDKJHD0000	8:32:02	265434
DQWPPOA09DSD	8:32:02	273099
0138DSJCVBNNAA	8:32:02	273010
0DLPZZZZ98A898	8:32:02	265001
ASKDPPPPQLSLAA	8:32:03	265000

ANONYMOUS SENSOR READINGS

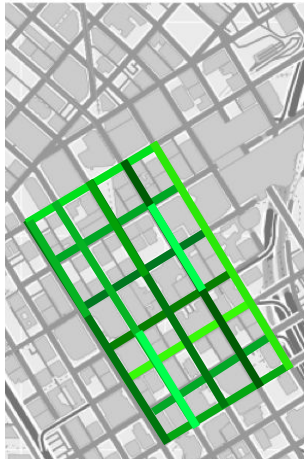


ESTIMATED PATHS

METADATA

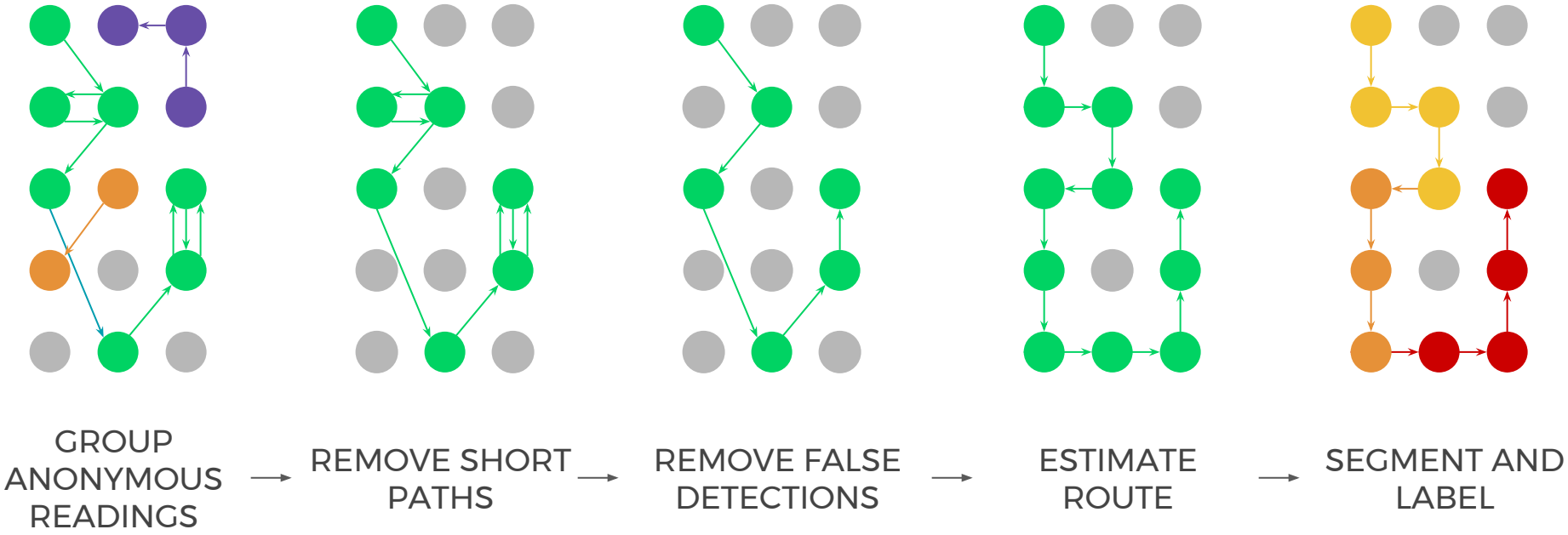


LABELED PATHS



AGGREGATE HEATMAP

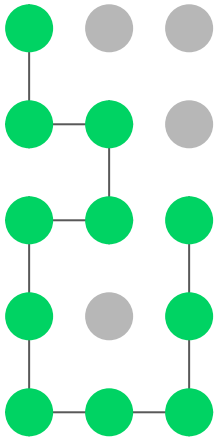
APPROACH OVERVIEW


GROUP ANONYMOUS READINGS
REMOVE SHORT PATHS
REMOVE FALSE DETECTIONS
ESTIMATE ROUTE
SEGMENT AND LABEL

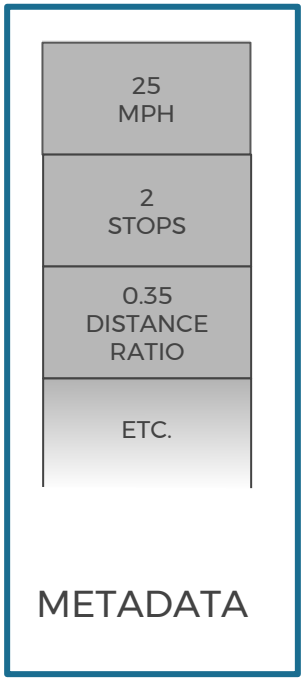
PIPELINE

ENCRYPTED HASH	TIME	SENSOR
KD98SDK8AHD8X	8:32:00	276105
8DJSKDLSX0WKK	8:32:01	276102
DQWPPOA09DSD	8:32:01	265402
KDOSPALDKSX03	8:32:01	265302
DJKFOSPHSHAH8	8:32:01	265101
0138DSJCVBNNAA	8:32:01	273777
0DLPZZZZ98A999	8:32:02	265000
NVMCSKDJALKSSS	8:32:02	265103
QALSJDKJHD0000	8:32:02	265434
DQWPPOA09DSD	8:32:02	273099
0138DSJCVBNNAA	8:32:02	273010
0DLPZZZZ98A898	8:32:02	265001
ASKDPPPPQLSLAA	8:32:03	265000

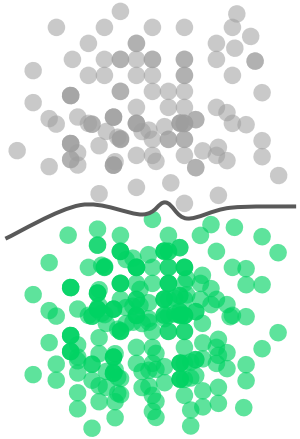
ANONYMOUS SENSOR READINGS



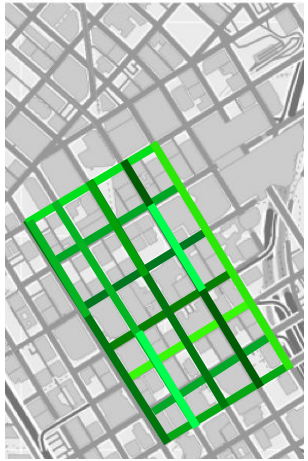
ESTIMATED PATHS



METADATA



LABELED PATHS

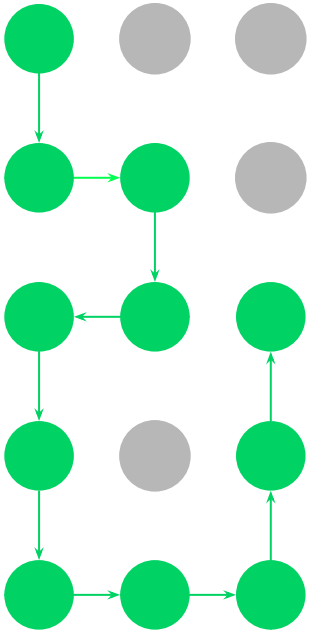


AGGREGATE HEATMAP

APPROACH OVERVIEW

METADATA COLLECTION

PATH

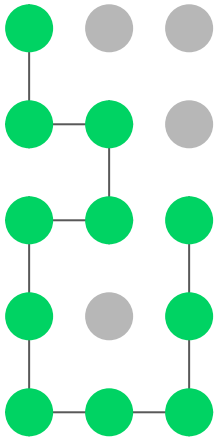


FEATURES

- NUMBER OF TIMES PATH CROSSED
- AVERAGE SPEED
- MAX SPEED
- STANDARD DEVIATION OF SPEED
- NUMBER OF LONG STOPS
- PERCENTAGE OF TIME DRIVING
- PERCENTAGE OF TIME WALKING

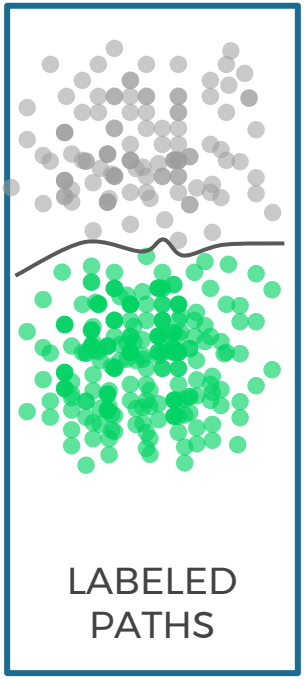
ENCRYPTED HASH	TIME	SENSOR
KD98SDK8AHD8X	8:32:00	276105
8DJSKDLSX0WKK	8:32:01	276102
DQWPPOA09DSD	8:32:01	265402
KDOSPALDKSX03	8:32:01	265302
DJKFOSPHSHAH8	8:32:01	265101
0138DSJCVBNNAA	8:32:01	273777
0DLPZZZZ98A999	8:32:02	265000
NVMCSKDJALKSSS	8:32:02	265103
QALSJDKJHD0000	8:32:02	265434
DQWPPOA09DSD	8:32:02	273099
0138DSJCVBNNAA	8:32:02	273010
0DLPZZZZ98A898	8:32:02	265001
ASKDPPPPQLSLAA	8:32:03	265000

ANONYMOUS SENSOR READINGS

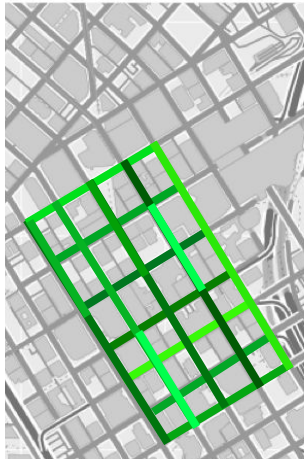


ESTIMATED PATHS

METADATA



LABELED PATHS



AGGREGATE HEATMAP

APPROACH OVERVIEW

DEFINING CRUISING

PROBLEM

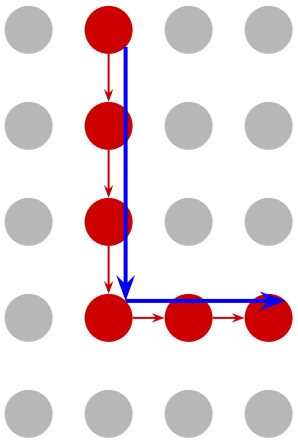
1. NO STANDARD DEFINITION OF CRUISING
2. CURRENTLY NO GROUND TRUTH

SOLUTION

1. MULTI-STEP CLASSIFICATION
2. LABEL SUBSET OF EXTREME CASES
3. MACHINE LEARNING TO IDENTIFY CRUISING INDICATORS

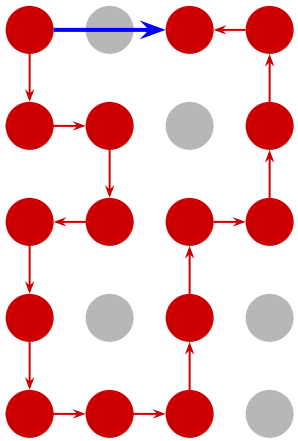
LABELING

USING DISTANCE RATIO



5 / 5 = 1.0

PROBABLY NOT CRUISING

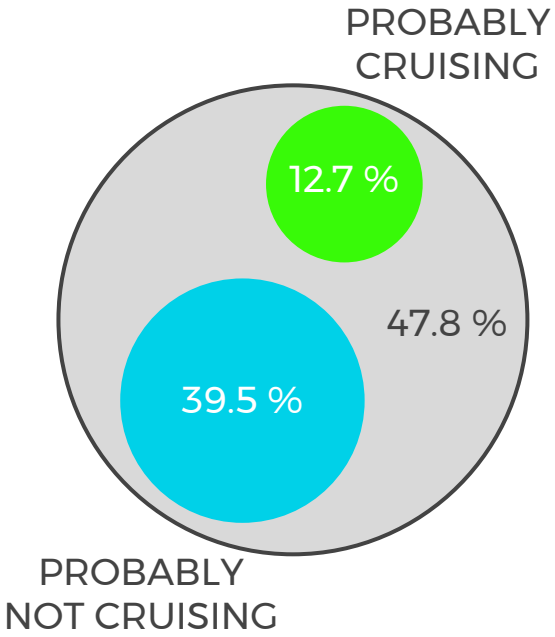


14 / 2 = 7.0

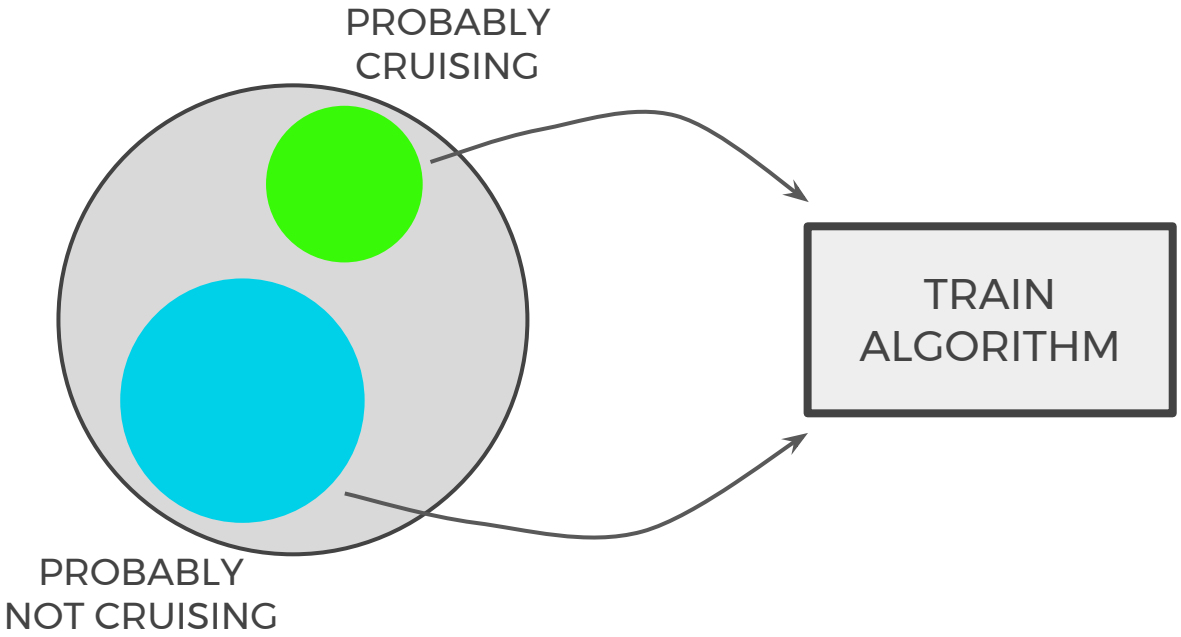
PROBABLY CRUISING

MULTI-STEP CLASSIFICATION

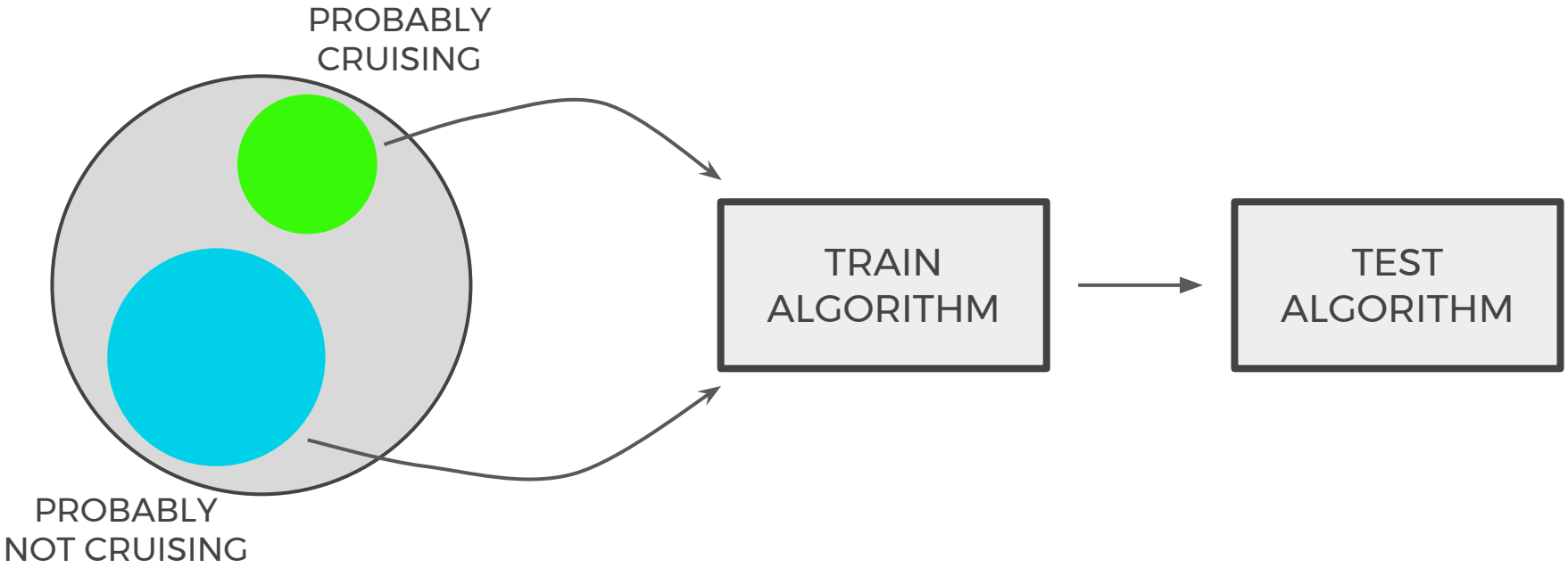
MULTI-STEP CLASSIFICATION



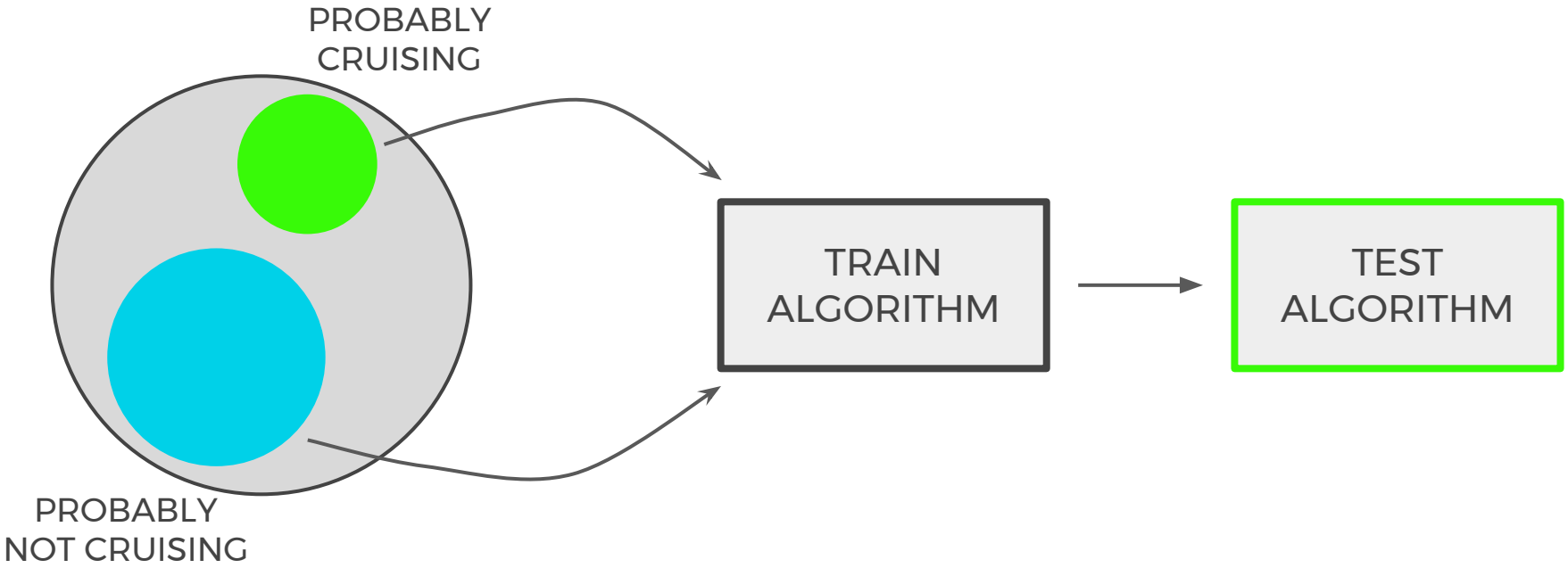
MULTI-STEP CLASSIFICATION



MULTI-STEP CLASSIFICATION



MULTI-STEP CLASSIFICATION



MULTI-STEP CLASSIFICATION

TESTING ALGORITHM

DECISION TREE	LOGISTIC REGRESSION	GRADIENT BOOSTING
ACCURACY: 96.6%	ACCURACY: 91.4%	ACCURACY: 97.1%

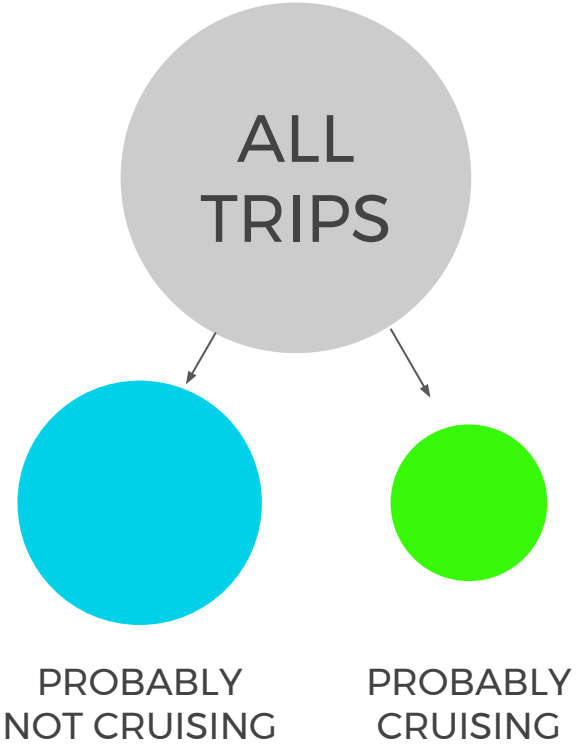
MULTI-STEP CLASSIFICATION

TESTING ALGORITHM

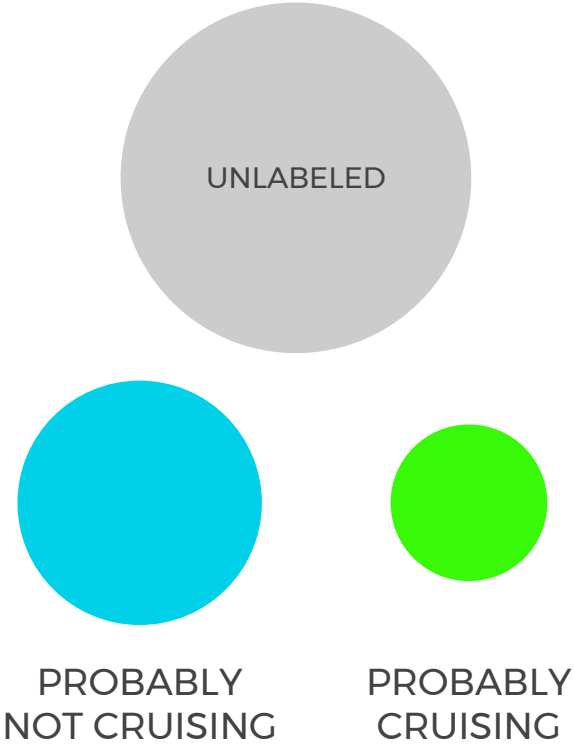
DECISION TREE	LOGISTIC REGRESSION	GRADIENT BOOSTING
ACCURACY: 96.6%	ACCURACY: 91.4%	ACCURACY: 97.1%

LABELING ALL DATA

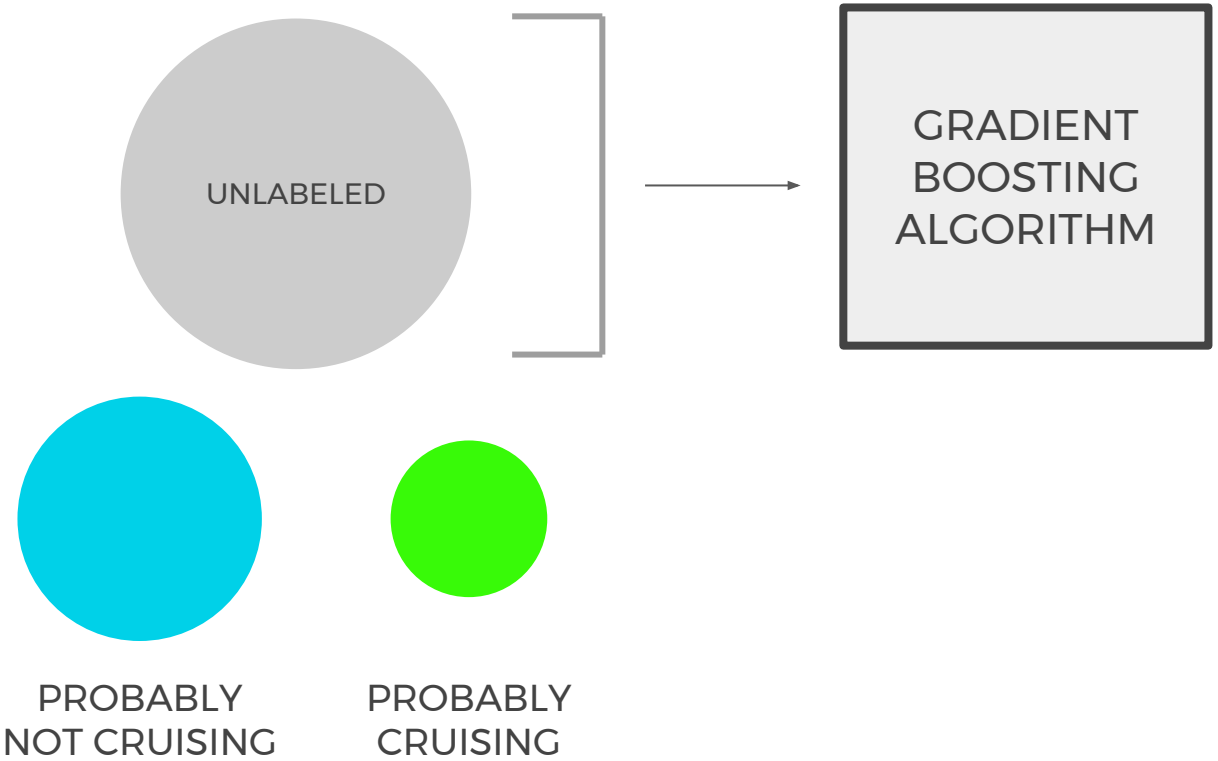
LABELING ALL DATA



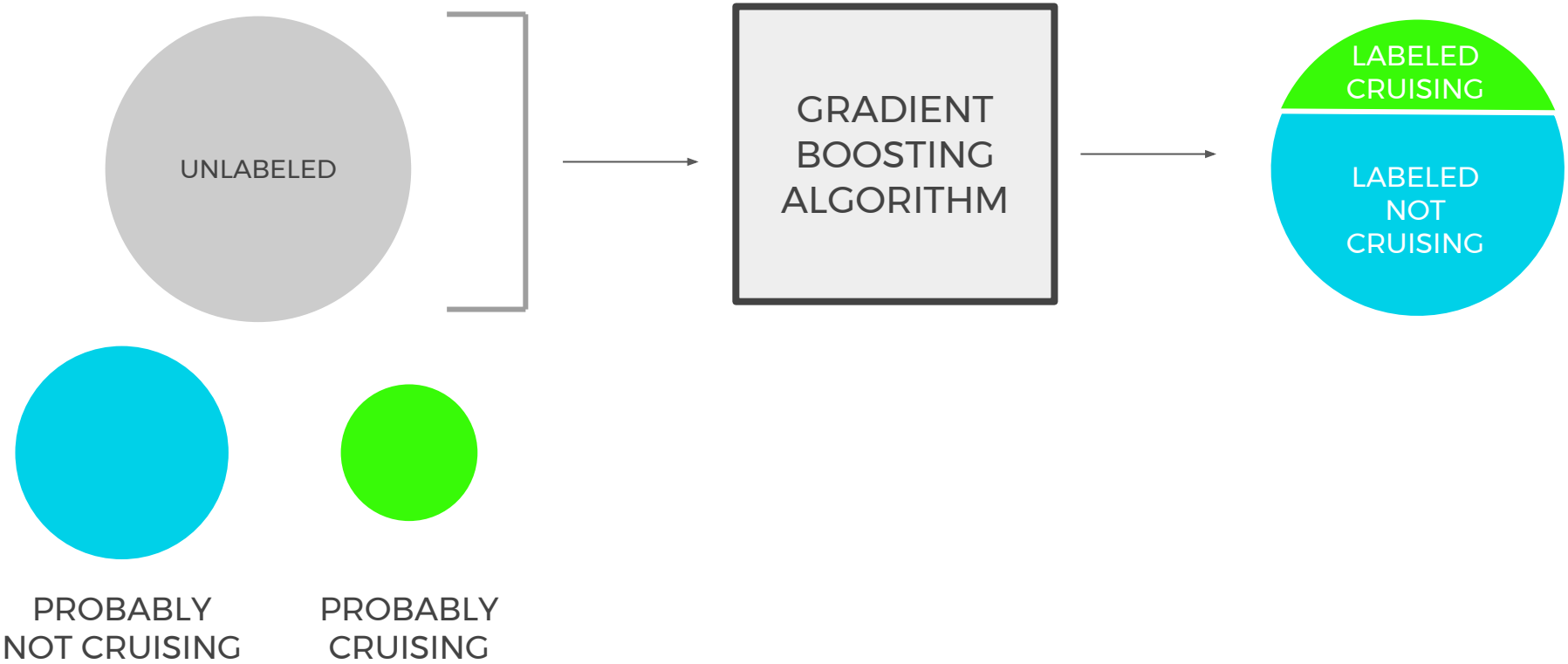
LABELING ALL DATA



LABELING ALL DATA



LABELING ALL DATA

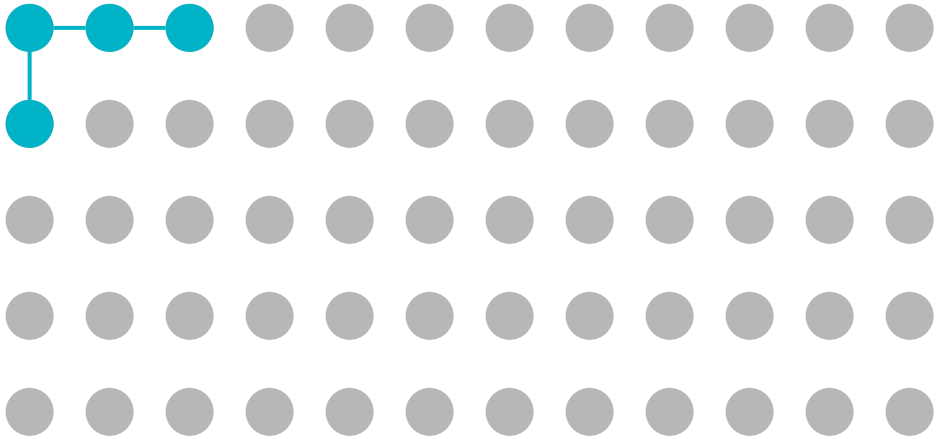


IDENTIFYING FOR-HIRE VEHICLES

IDENTIFYING FOR-HIRE VEHICLES

FOR-HIRE VEHICLE EXAMPLE

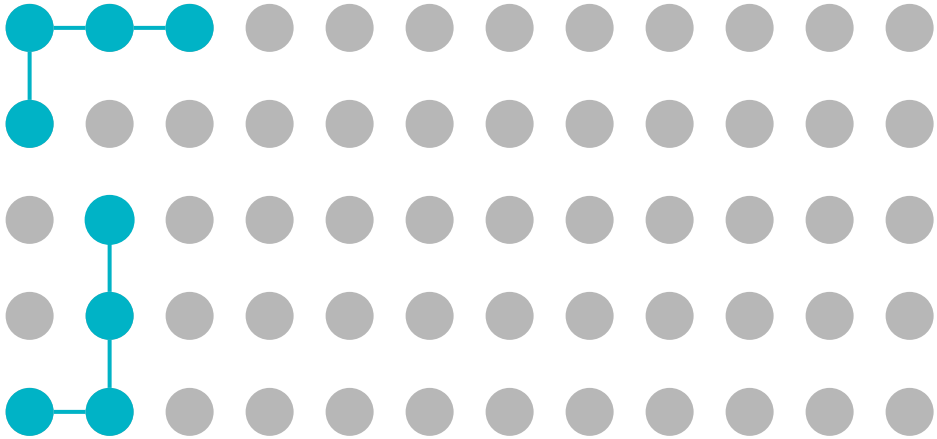
4 LARGE GAPS IN READ TIMES (5 TRIPS)



IDENTIFYING FOR-HIRE VEHICLES

FOR-HIRE VEHICLE EXAMPLE

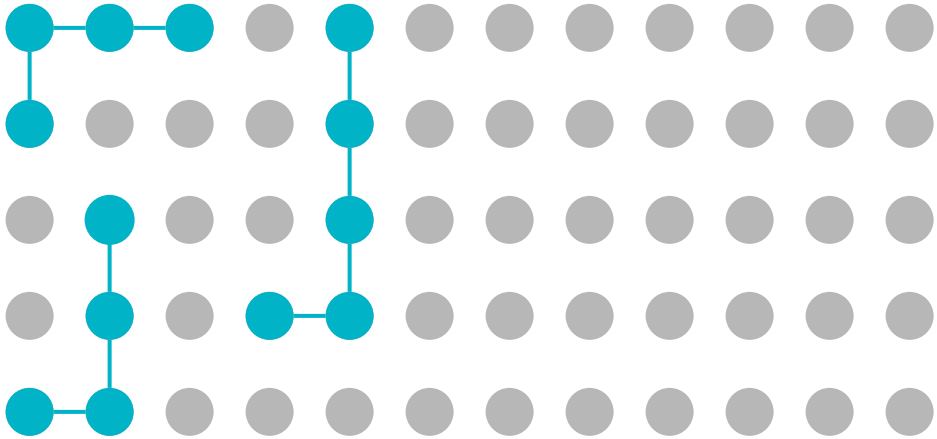
4 LARGE GAPS IN READ TIMES (5 TRIPS)



IDENTIFYING FOR-HIRE VEHICLES

FOR-HIRE VEHICLE EXAMPLE

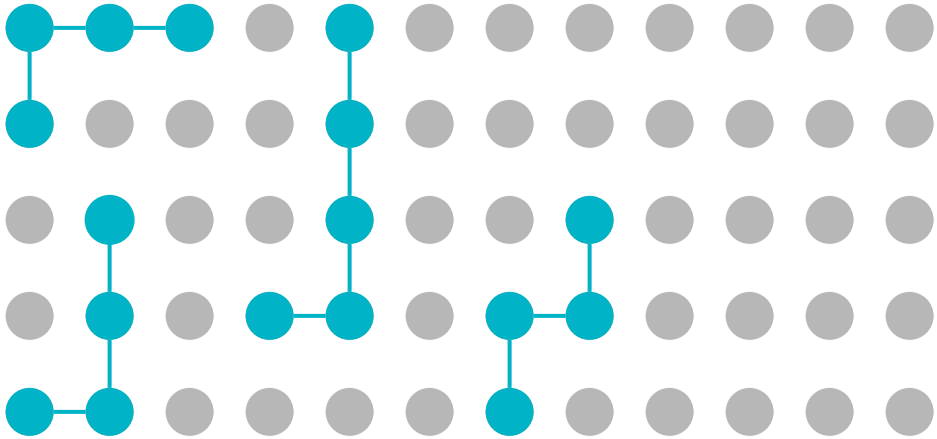
4 LARGE GAPS IN READ TIMES (5 TRIPS)



IDENTIFYING FOR-HIRE VEHICLES

FOR-HIRE VEHICLE EXAMPLE

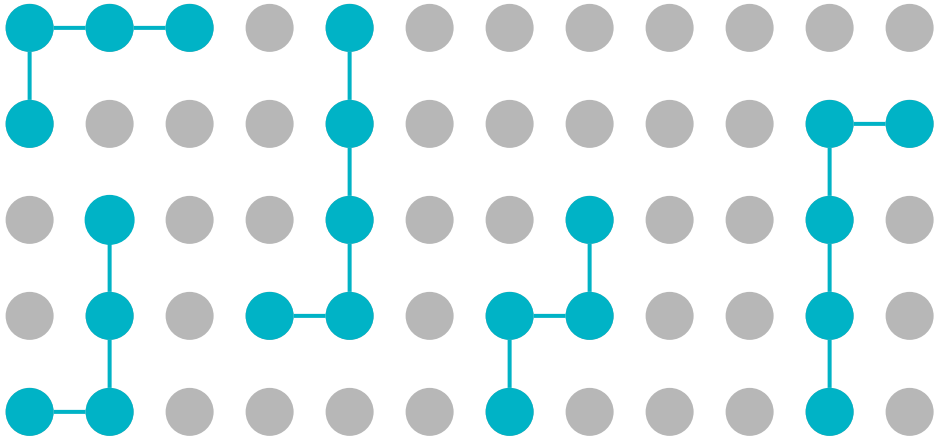
4 LARGE GAPS IN READ TIMES (5 TRIPS)



IDENTIFYING FOR-HIRE VEHICLES

FOR-HIRE VEHICLE EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)



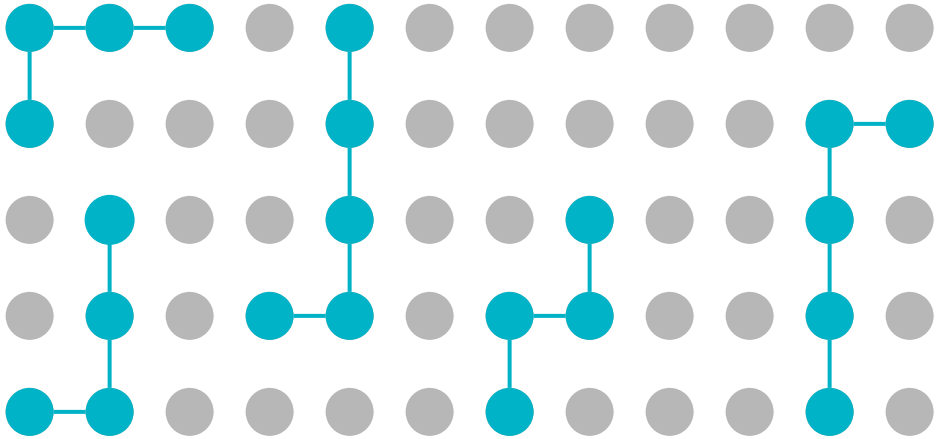
IDENTIFYING FOR-HIRE VEHICLES

FOR-HIRE VEHICLE EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

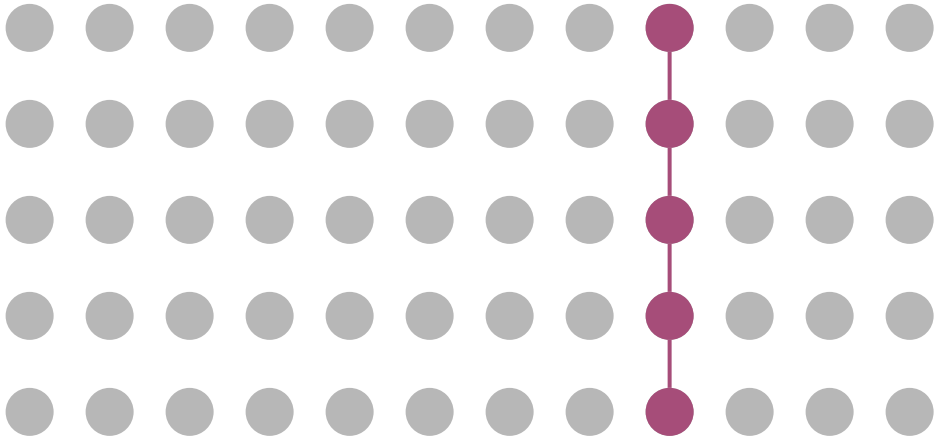
UNIQUE SENSORS / TOTAL READS =

22 / 22 = 1.0 [HIGH DISPERSION]



IDENTIFYING FOR-HIRE VEHICLES

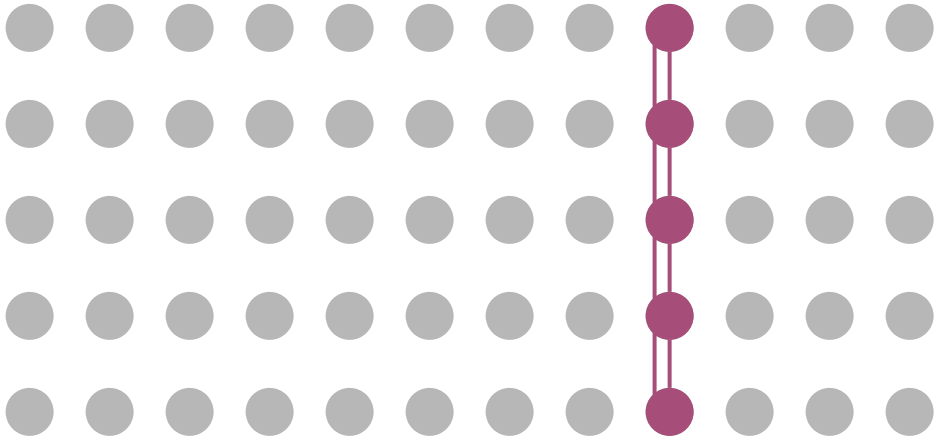
IDENTIFYING FOR-HIRE VEHICLES



BUS EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

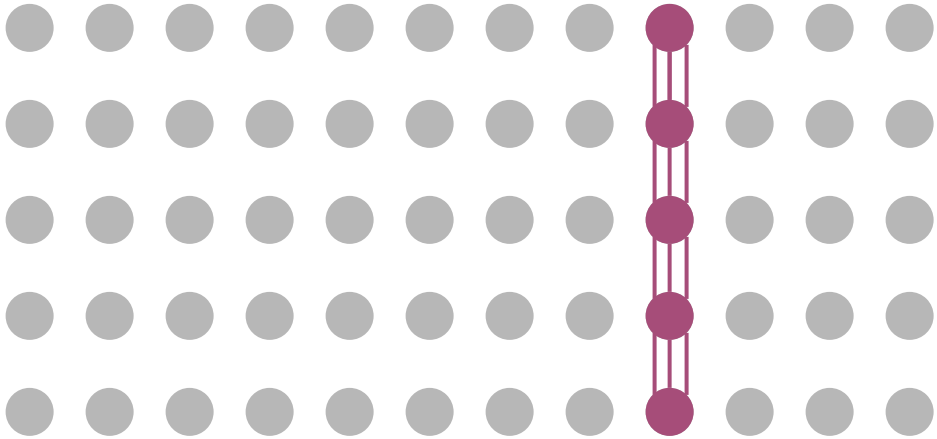
IDENTIFYING FOR-HIRE VEHICLES



BUS EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

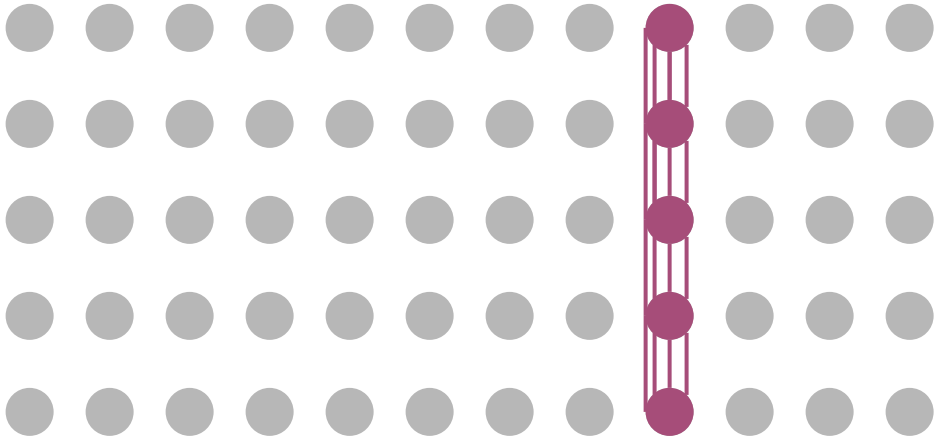
IDENTIFYING FOR-HIRE VEHICLES



BUS EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

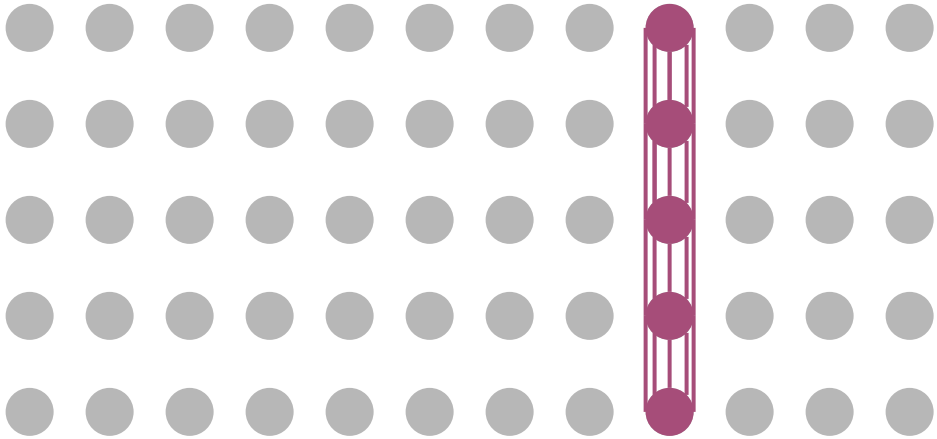
IDENTIFYING FOR-HIRE VEHICLES



BUS EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

IDENTIFYING FOR-HIRE VEHICLES



BUS EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

UNIQUE SENSORS / TOTAL READS =

5 / 25 = 0.2 [LOW DISPERSION]

IDENTIFYING FOR-HIRE VEHICLES

FOR-HIRE VEHICLE EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

UNIQUE SENSORS / TOTAL READS =

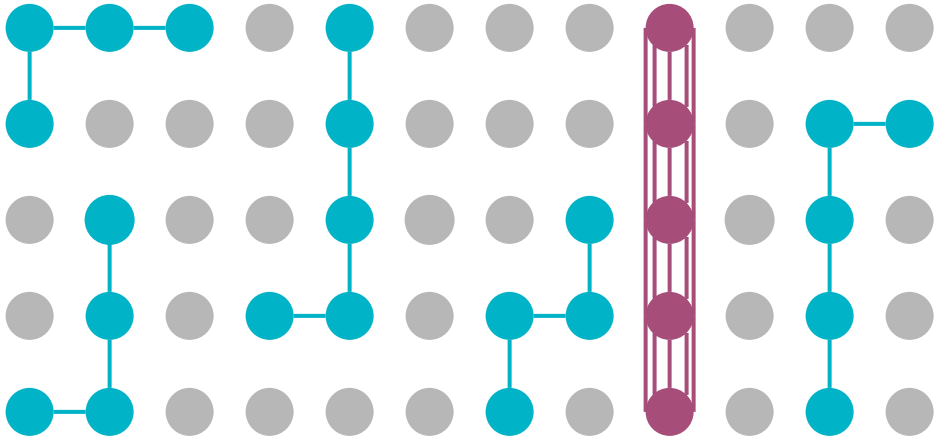
22 / 22 = 1.0 [HIGH DISPERSION]

BUS EXAMPLE

4 LARGE GAPS IN READ TIMES (5 TRIPS)

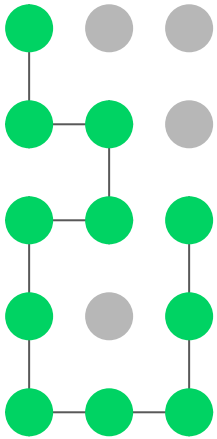
UNIQUE SENSORS / TOTAL READS =

5 / 25 = 0.2 [LOW DISPERSION]



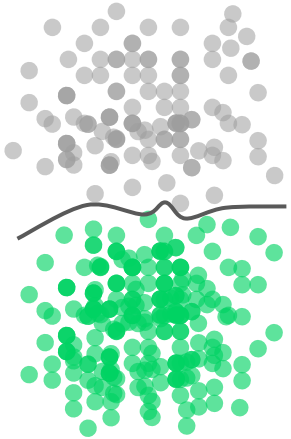
ENCRYPTED HASH	TIME	SENSOR
KD98SDK8AHD8X	8:32:00	276105
8DJSKDLSX0WKK	8:32:01	276102
DQWPPOA09DSD	8:32:01	265402
KDOSPALDKSX03	8:32:01	265302
DJKFOSPHSHAH8	8:32:01	265101
0138DSJCVBNNAA	8:32:01	273777
0DLPZZZZ98A999	8:32:02	265000
NVMCSKDJALKSSS	8:32:02	265103
QALSJDKJHD0000	8:32:02	265434
DQWPPOA09DSD	8:32:02	273099
0138DSJCVBNNAA	8:32:02	273010
0DLPZZZZ98A898	8:32:02	265001
ASKDPPPPQLSLAA	8:32:03	265000

ANONYMOUS SENSOR READINGS →

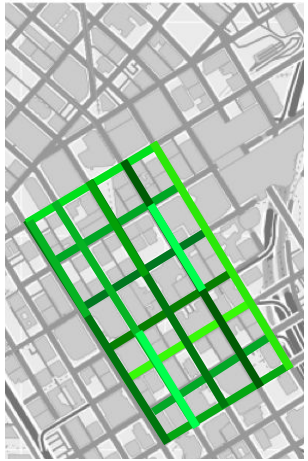


ESTIMATED PATHS →

METADATA →

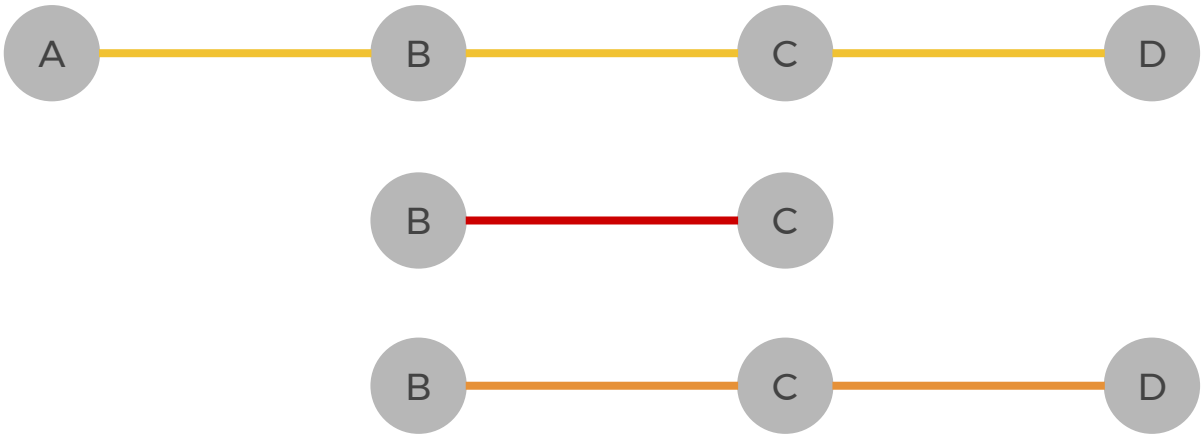


LABELED PATHS →



AGGREGATE HEATMAP

APPROACH OVERVIEW



PATH ONE

+

PATH TWO

+

PATH THREE

AGGREGATED COUNTS

AGGREGATION

CRUISING IN DOWNTOWN SEATTLE

As part of The Data Science for Social Good Program at The University of Washington, the Traffic Cruising Team has produced a heatmap to identify cruising in the downtown Seattle area.

SELECT TYPE OF CRUISING

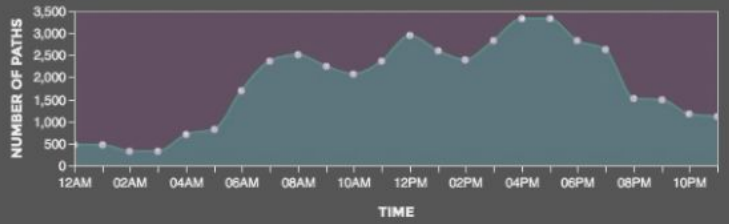
PARKING

SELECT DAY

MON TUE WED THR FRI SAT SUN

SELECT TIME

ALL DAY MORNING MIDDAY EVENING



CRUISING IN DOWNTOWN SEATTLE

As part of The Data Science for Social Good Program at The University of Washington, the Traffic Cruising Team has produced a heatmap to identify cruising in the downtown Seattle area.

SELECT TYPE OF CRUISING

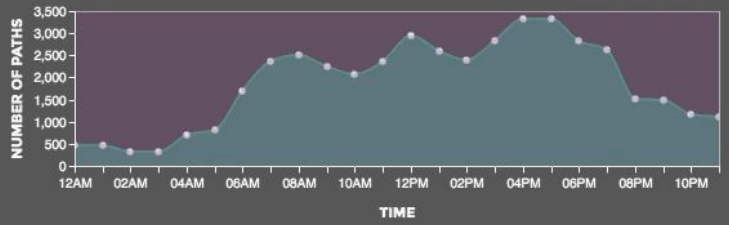
PARKING | PASS-THRU VEHICLES

SELECT DAY

MON | TUE | WED | THR | FRI | SAT | SUN

SELECT TIME

ALL DAY | MORNING | MIDDAY | EVENING



CONCLUSION

EXISTING TRAFFIC SENSORS CAN BE USED TO IDENTIFY CRUISING

- YEAR-ROUND OBSERVATION OF CRUISING PATTERNS
- CITY OF SEATTLE BENEFITS
- SAME SYSTEM CAN BE APPLIED TO OTHER CITIES

CONCLUSION

EXISTING TRAFFIC SENSORS CAN BE USED TO IDENTIFY CRUISING

YEAR-ROUND OBSERVATION OF CRUISING PATTERNS

CITY OF SEATTLE BENEFITS

SAME SYSTEM CAN BE APPLIED TO OTHER CITIES

CONCLUSION

EXISTING TRAFFIC SENSORS CAN BE USED TO IDENTIFY CRUISING

- YEAR-ROUND OBSERVATION OF CRUISING PATTERNS
- CITY OF SEATTLE BENEFITS
- SAME SYSTEM CAN BE APPLIED TO OTHER CITIES

CONCLUSION

EXISTING TRAFFIC SENSORS CAN BE USED TO IDENTIFY CRUISING

YEAR-ROUND OBSERVATION OF CRUISING PATTERNS

CITY OF SEATTLE BENEFITS

SAME SYSTEM CAN BE APPLIED TO OTHER CITIES

Microsoft

Urban@UW

UNIVERSITY of WASHINGTON
eScience Institute

CASCADIA URBAN
ANALYTICS COOPERATIVE

GORDON AND BETTY
MOORE
FOUNDATION

Washington Research
FOUNDATION

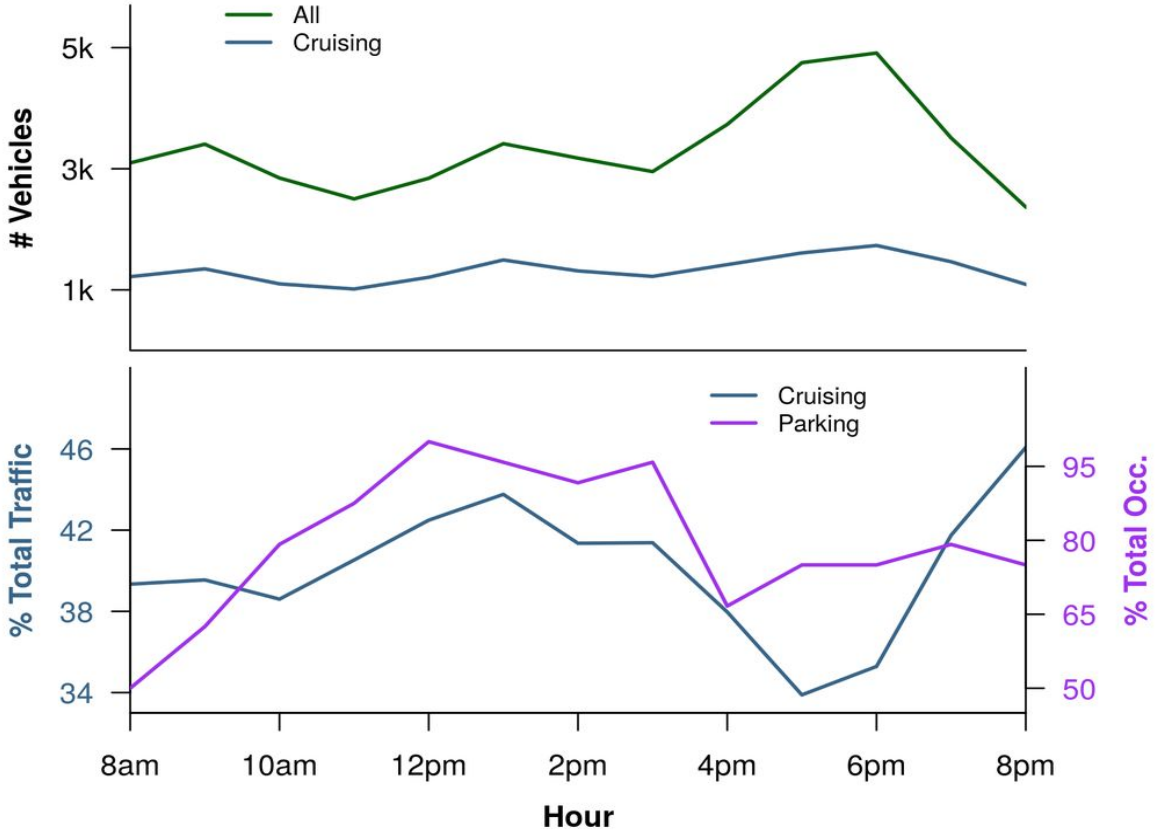
Seattle
Department of
Transportation

ALFRED P. SLOAN FOUNDATION
1934

THANK YOU

QUESTIONS?

APPENDIX


All
Cruising
5k
3k
1k
Vehicles
Cruising
Parking
46
42
38
34
% Total Traffic
95
80
65
50
% Total Occ.
8am
10am
12pm
2pm
4pm
6pm
8pm
Hour

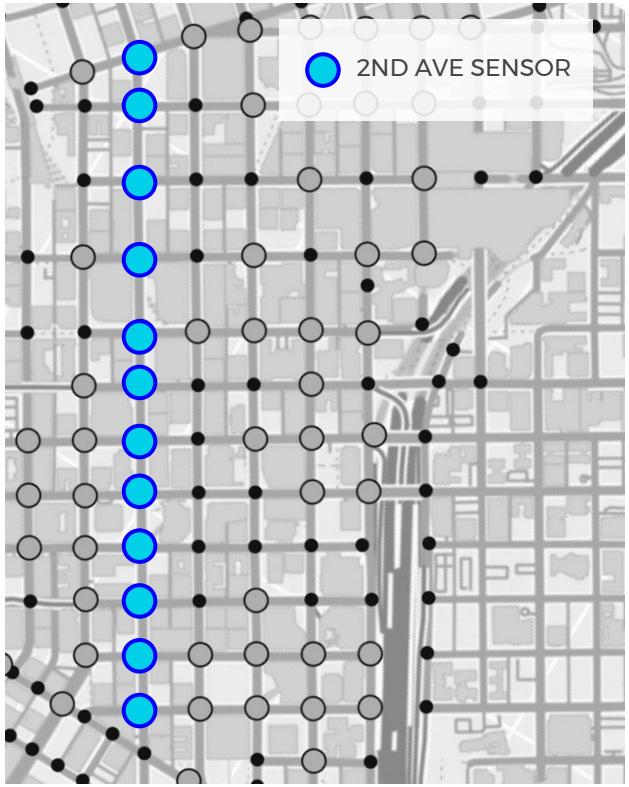
VALIDATION

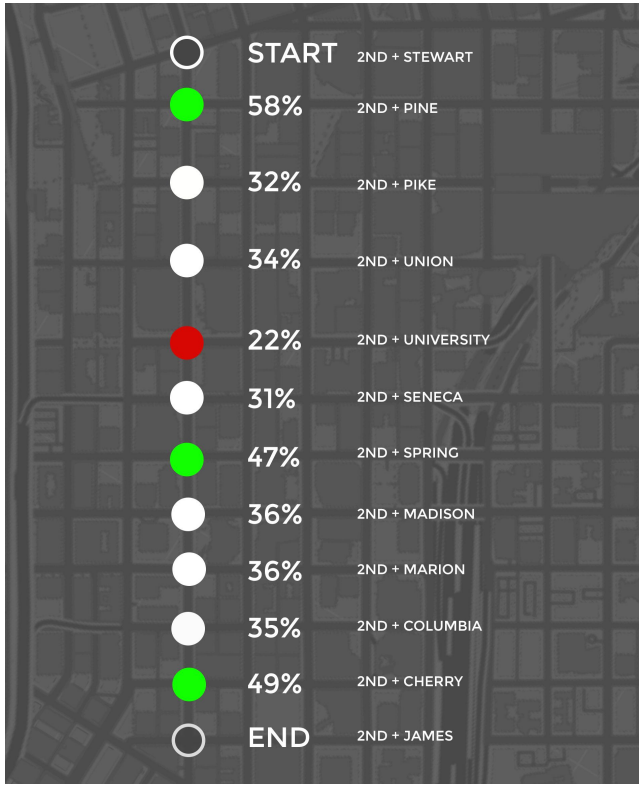
CLASSIFICATION

MODEL	ACCURACY	AUC_ROC	PRECISION	RECALL	F1_SCORE
DECISION TREE	0.966 +/- 0.000	0.933 +/- 0.000	0.993 +/- 0.000	0.868 +/- 0.000	0.926 +/- 0.000
LOGISTIC REGRESSION	0.925 +/- 0.001	0.870 +/ 0.001	0.921 +/- 0.001	0.762 +/- 0.001	0.834 +/- 0.001
GRADIENT BOOSTING CLASSIFIER	0.967 +/- 0.000	0.935 +/- 0.000	0.994 +/- 0.000	0.871 +/- 0.000	0.928 +/- 0.000

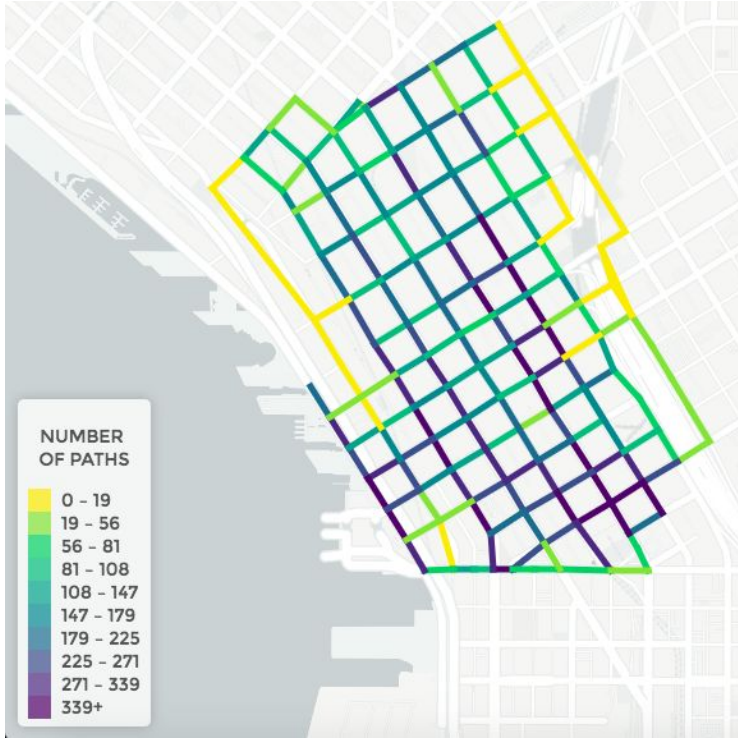
FEATURE	IMP
NUMBER OF TIMES PATH CROSSED	0.3442
AVERAGE SPEED	0.1191
PERCENTAGE OF TIME DRIVING	0.1694
HOUR	0.0101
MAXIMUM SPEED	0.0442
STD DEV OF SPEED	0.0556
STOPS 0 - 2 MIN	0.0323
STOPS 10 - 15 MIN	0.0126
STOPS 15+ MIN	0.0377
TIME PERCENTAGE CHILLING	0.0906
TIME PERCENTAGE WALKING	0.0351

SENSOR DETECTION RATE

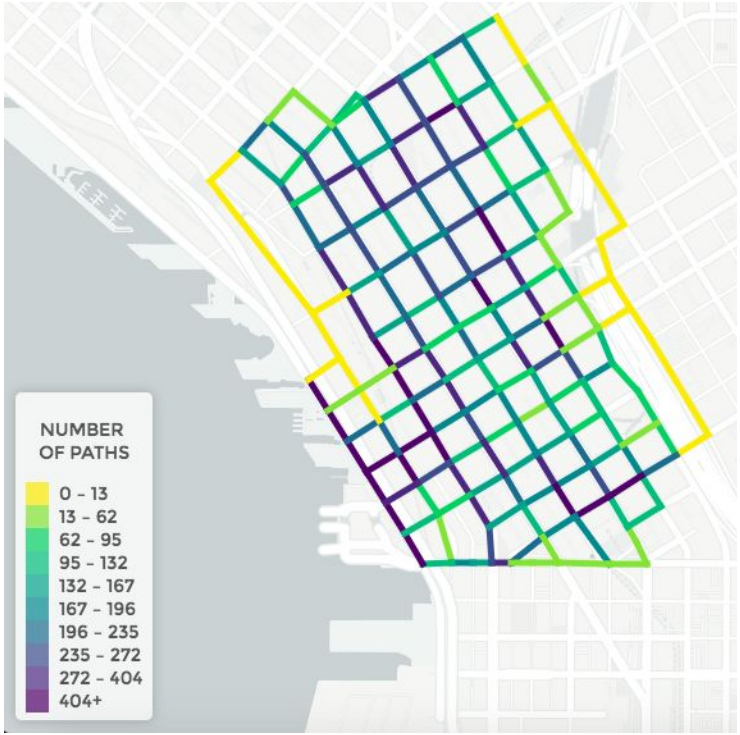




PARKING | TUE | MORNING

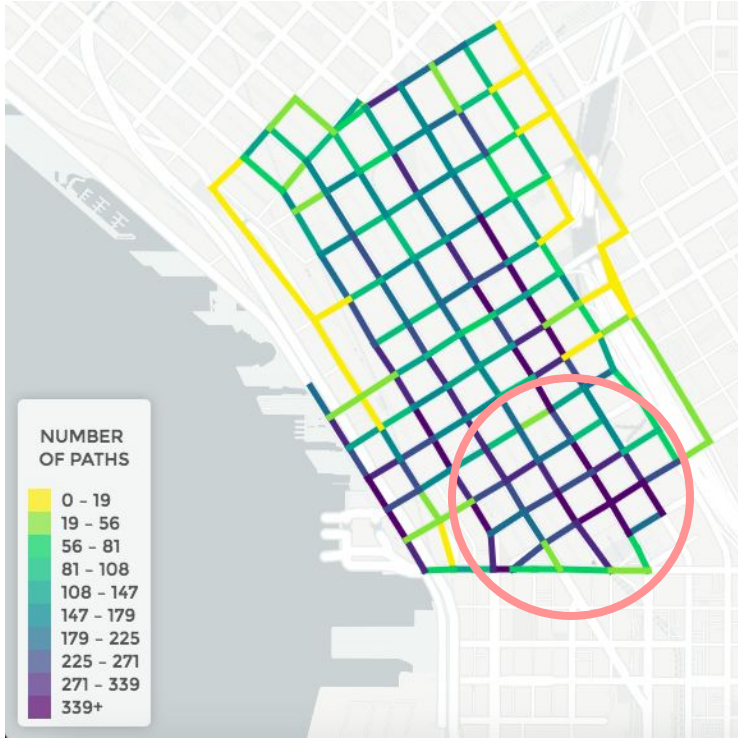


PARKING | SAT | EVENING

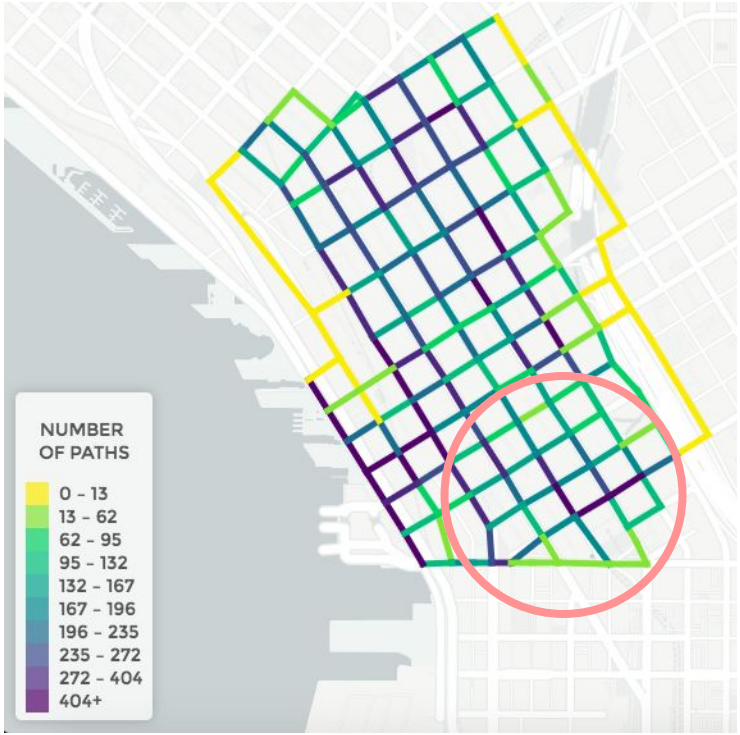


RESULTS

PARKING | TUE | MORNING



PARKING | SAT | EVENING



RESULTS

PARKING | TUE | MORNING

PARKING | SAT | EVENING

RESULTS