EXTRACTING THE MOST FROM URBAN DATA: MULTI-PERSPECTIVE ANALYSIS OF NYC TAXI TRIPS

Anil Yazici, Ph.D.

Assistant Professor, Civil Engineering, Stony Brook University, NY

Ground Transportation Technology Symposium: Big Data and Innovative Solutions for Safe, Efficient and Sustainable Mobility

November 19th 2014

An Analogy…

Neo's eyes light up as he steps closer to the screens that seem alive with a constant flow of data.

                    NEO
        Is that...?

                    CYPHER
        The Matrix?   Yeah.

Neo stares at the endlessly shifting river of information, bizarre codes and equations flowing across the face of the monitor.

                    NEO
        Do you always look at it encoded?

                    CYPHER
        Have to. The image translators
        sort of work for the construct
        programs but there's way too much
        information to decode the Matrix.
        You get used to it, though. Your
        brain does the translating. I
        don't even see the code. All I
        see is blonde, brunette, and
        redhead. You want a drink?

TLC Yellow Cab Data

- GPS dataset with more than 370 million taxi trips covering the period from January 1, 2009 to November 28, 2010.
- Each trip record includes:
  - trip origin and destination
  - the time of pick-up and drop-off
  - the number of passengers
  - trip fare
  - trip distance
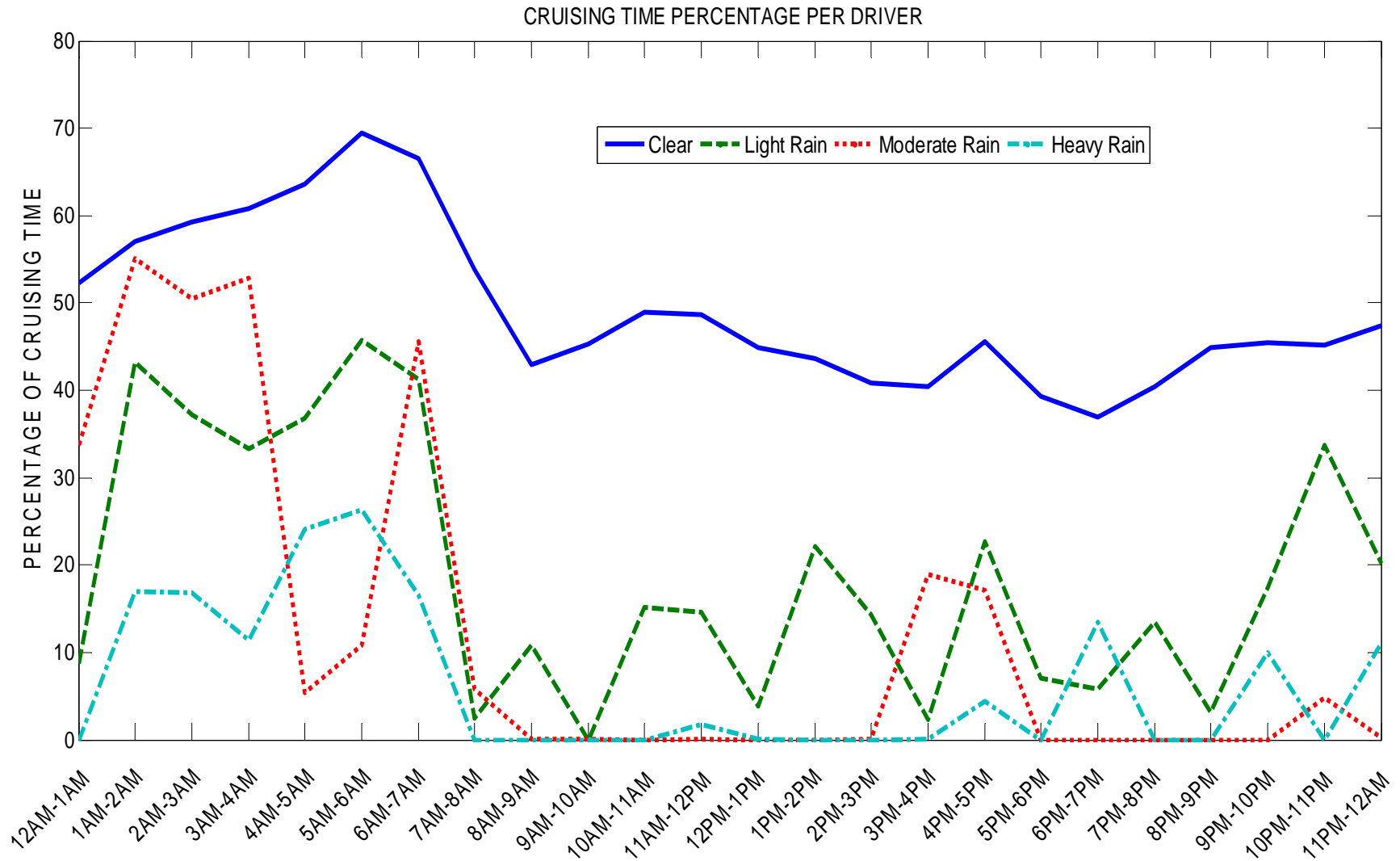 - occupancy
  - …

How to make sense of taxi data?

- Taxi as a taxi
→ Taxi industry, a multi billion $ business

- Taxis as probe vehicles
→ travel time patterns

E-hail? Less empty cruising?
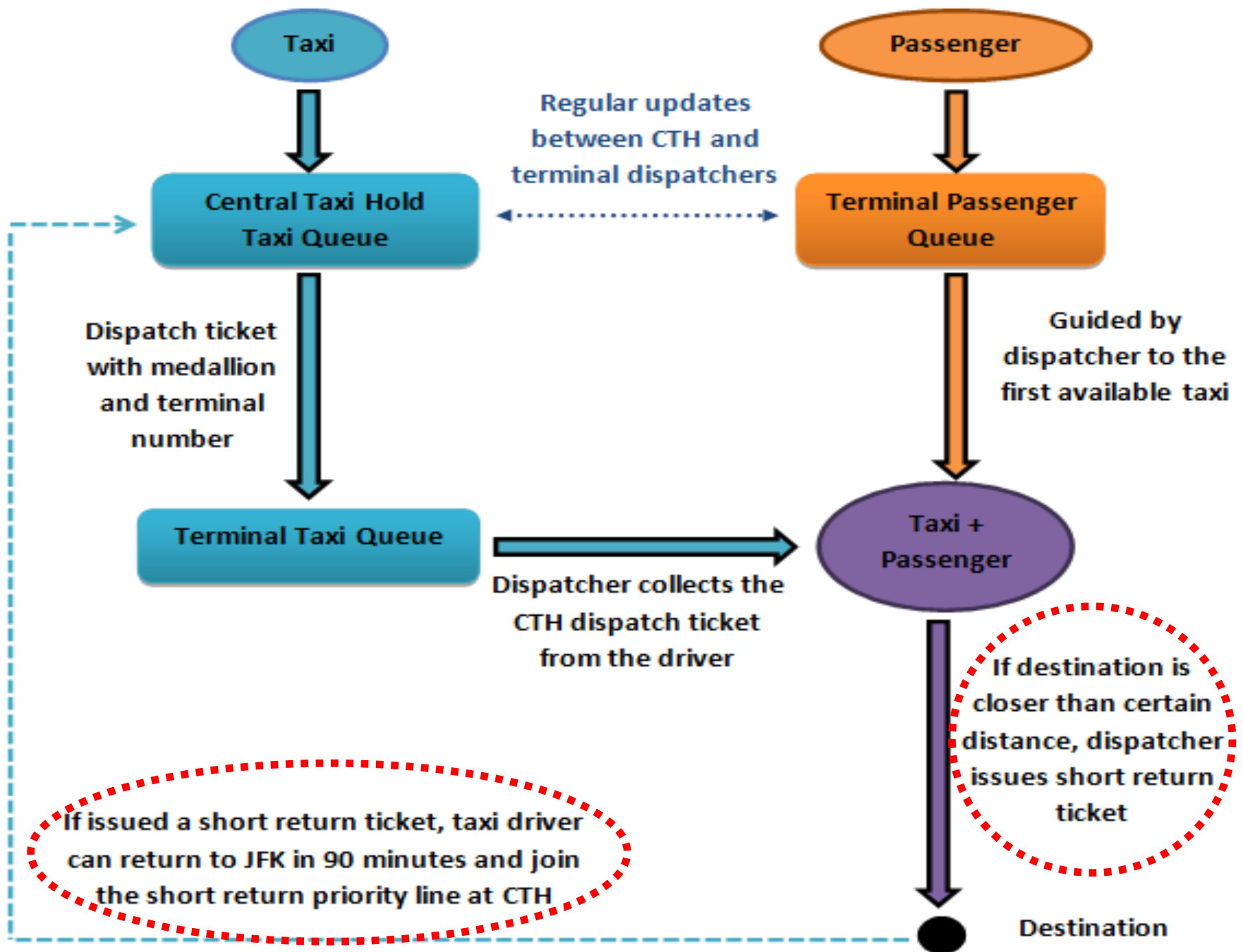
Percentage of Cruising Times
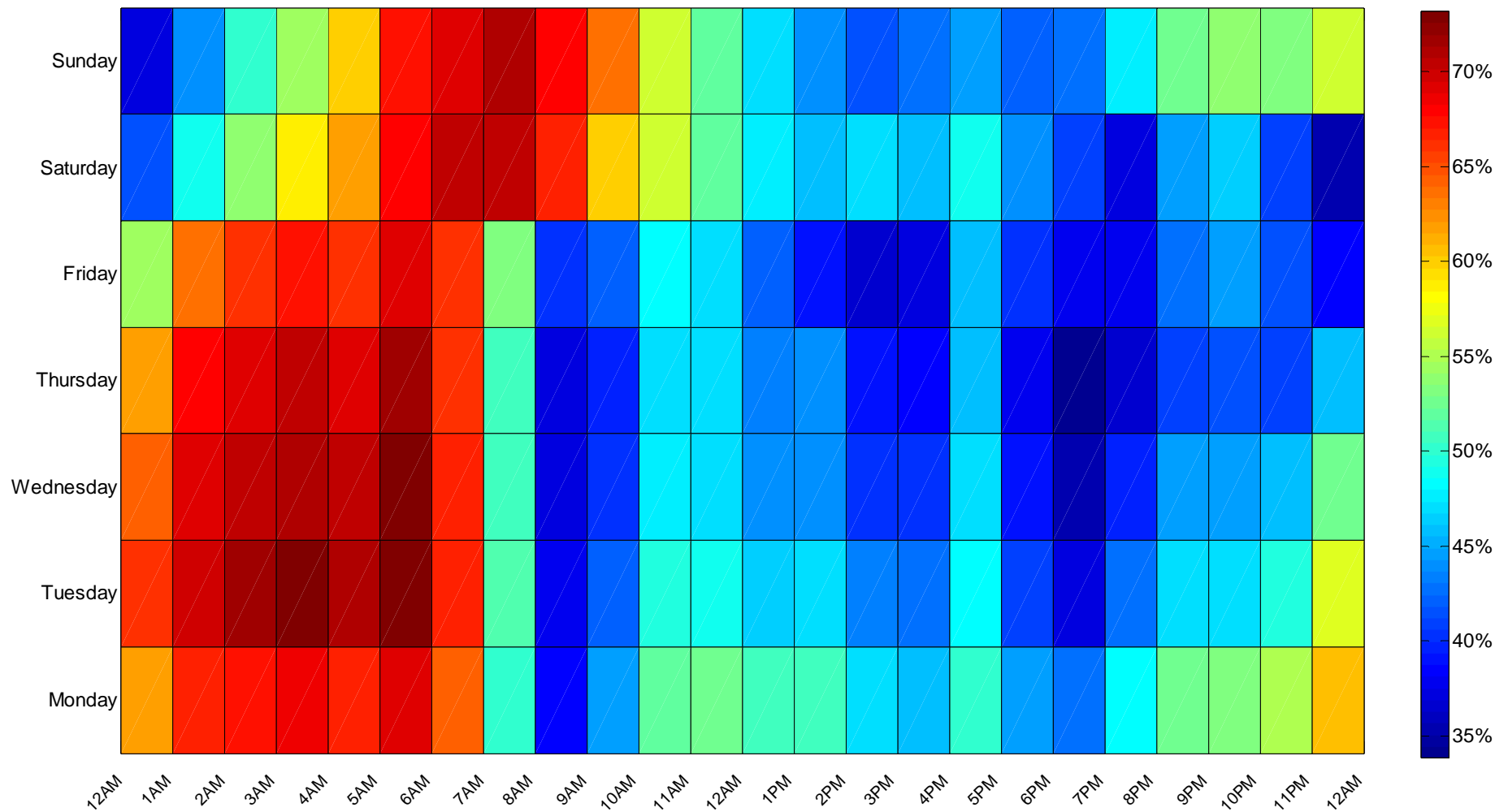
Cruising for Passenger Under Different Weather Conditions

CRUISING TIME PERCENTAGE PER DRIVER

Taxi Shortage at the Airports

- In New York City (NYC), taxis carry 35% and 23% of all airport bound passengers to LaGuardia (LGA) and John F. Kennedy airports respectively
- Long passenger queues for taxis at JFK
- Airport accessibility may affect travelers' airport decision among competing airports in a region → would be a problem for airlines using a particular airport as a hub
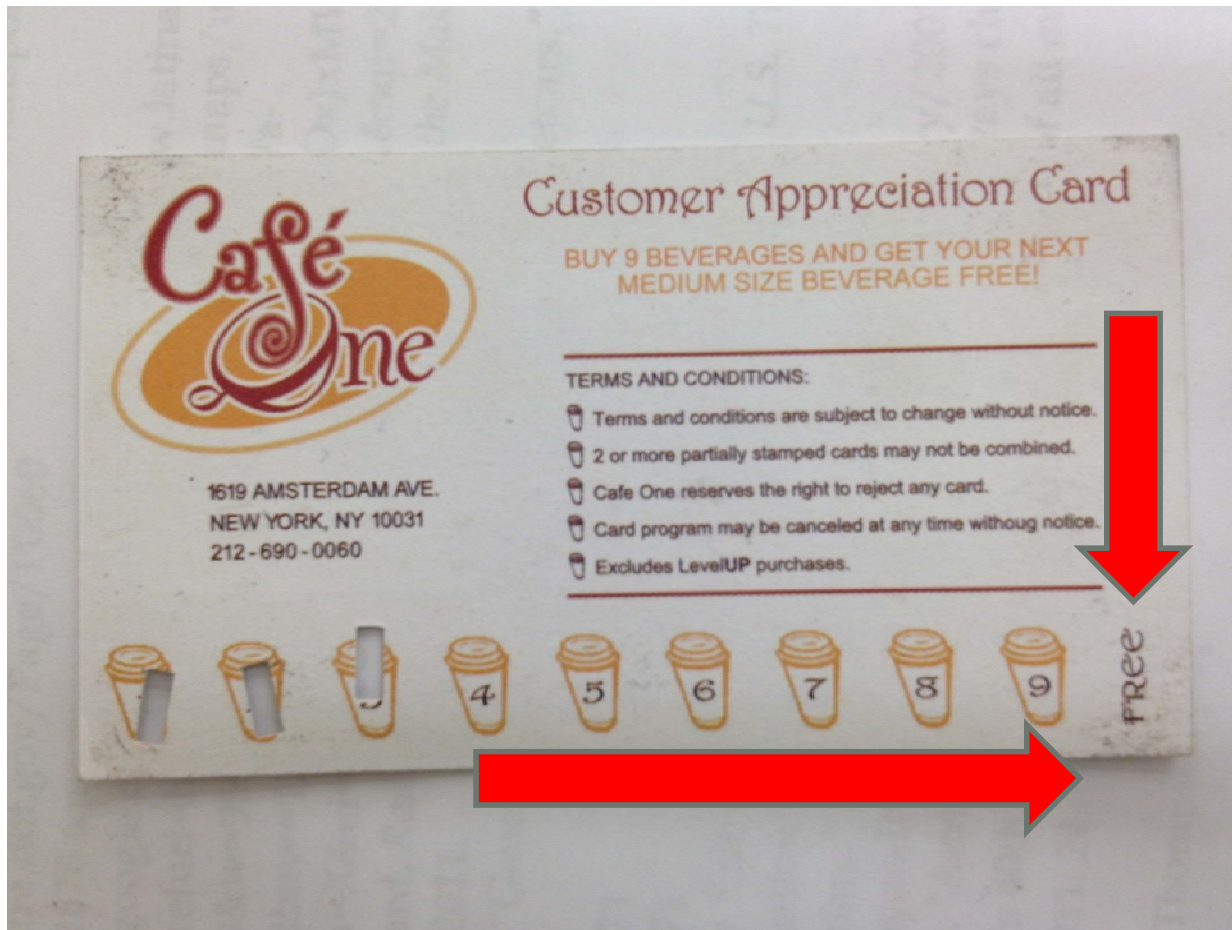
Taxi

Passenger

Regular updates between CTH and terminal dispatchers

Central Taxi Hold Taxi Queue

Terminal Passenger Queue

Dispatch ticket with medallion and terminal number

Guided by dispatcher to the first available taxi

Terminal Taxi Queue

Taxi + Passenger

Dispatcher collects the CTH dispatch ticket from the driver

If destination is closer than certain distance, dispatcher issues short return ticket

If issued a short return ticket, taxi driver can return to JFK in 90 minutes and join the short return priority line at CTH

Destination

Taxi Driver Airport Pick-up Decision Model

- Decision: *To go* or *not to go* to JFK for passenger pick-up
  - Decision is made at the end of each trip, i.e. dropping off a passenger
  - Binary variable → Appropriate for logistic regression

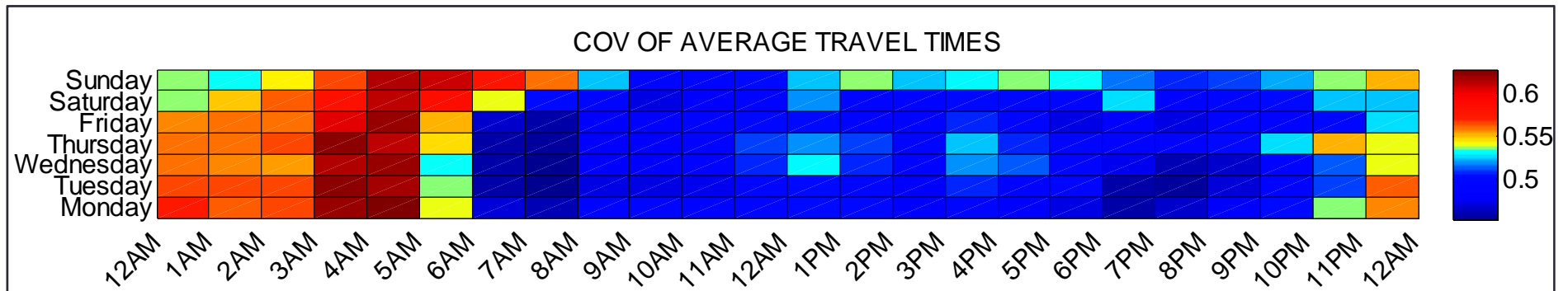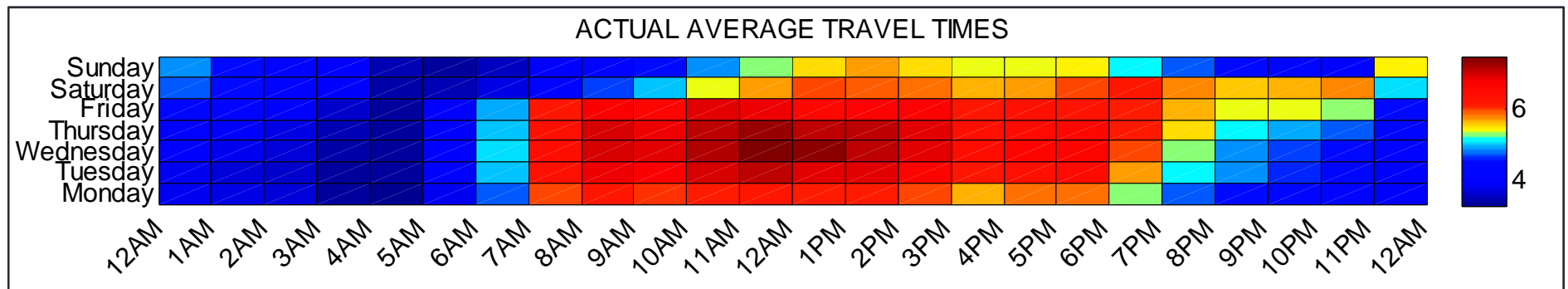# POSSIBLE POLICIES FOR IMPROVED AIRPORT GROUND ACCESS

- Time-of-Day, Day-of-Week, Location, Weather affect the probability of airport pick up
  - → Not much to suggest
- Still good to know the drivers' decision mechanism
  - We know the condition during which the drivers are less likely to go JFK → Incentives?
- Fare increase?
  - Impact of fare increase is found to be very marginal
  - Instead of fare increase, efficient dispatching can achieve more
- Drivers with short return ticket are ~43 times more likely to make airport pickups
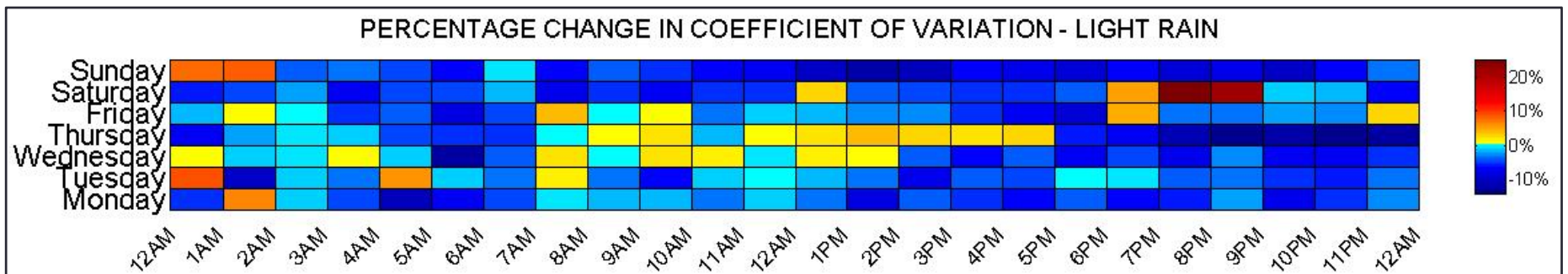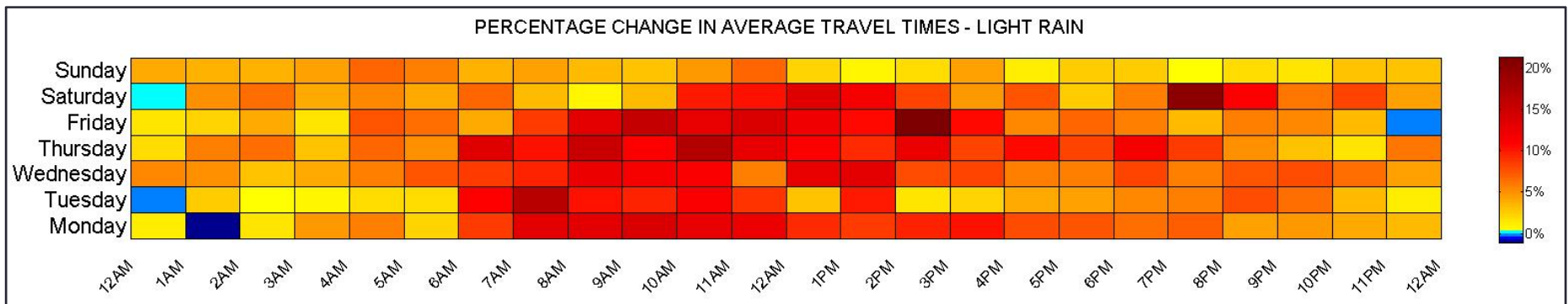 - More short return tickets? But how?

Short Return Punch Card?

Taxis as Probe Vehicles: Travel Time, Congestion, Delay…

- Value of travel time (VOT), Value of travel time reliability (VOR)
- → Toll pricing, congestion pricing
- → Economic evaluation of transport projects

- Actual *vs.* Perceived travel time
- → Objective *vs.* Subjective travel time
- → Travel decisions; time, mode…

→ Taxi GPS data with almost 400 million records provides good basis for objective travel times

Travel Time Patterns in NYC

Weather Impacts

PERCENTAGE CHANGE IN AVERAGE TRAVEL TIMES - LIGHT RAIN

PERCENTAGE CHANGE IN COEFFICIENT OF VARIATION - LIGHT RAIN

# More fields, more to extract: Traffic Delay at NYC Urban Network

- **IDEA:** Employ existing taxis as probe vehicles and exploit the taxi fare structure to quantify the urban delay.

❖ Taxi trip fare in NYC is the summation of fixed charges (opening fare and surcharges) and a certain amount for each "additional unit."

❖ Additional Unit: *"one-fifth of a mile, when the taxicab is traveling at 6 miles an hour or more;* ***or 60 seconds when not in motion or traveling at less than 6 miles per hour****. The taximeter shall combine fractional measures of distance and time in accruing a unit of fare."*

→Additional cost for each trip= Recorded Fare - Hypothetical fare (based solely on fixed charges and trip distance)

→Analysis unit is "$ per mile" → distance neutral

Extra Fare 24/7

Average Additional Fare, $ per mile per trip
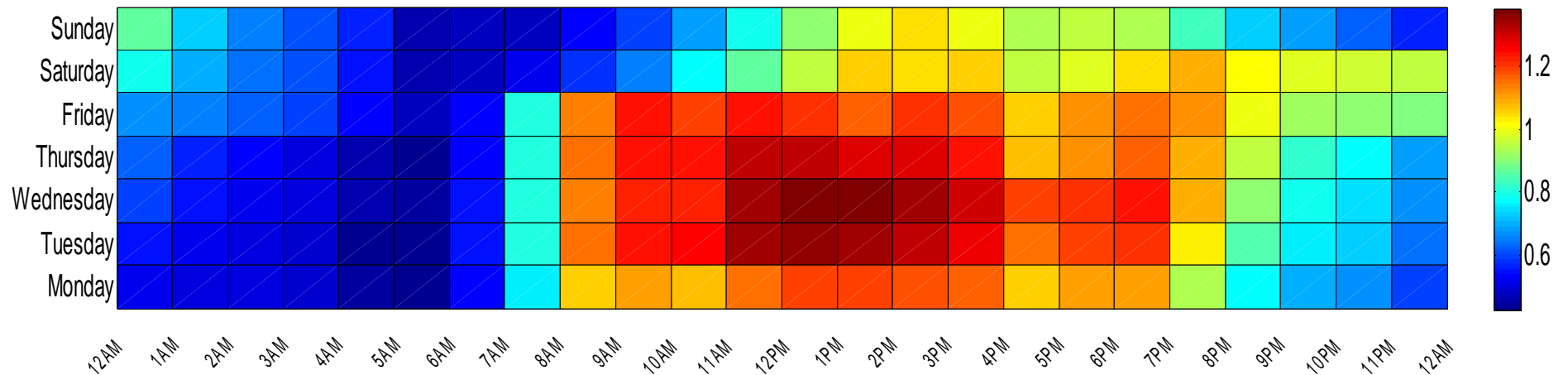

Average Travel Time, minutes per mile
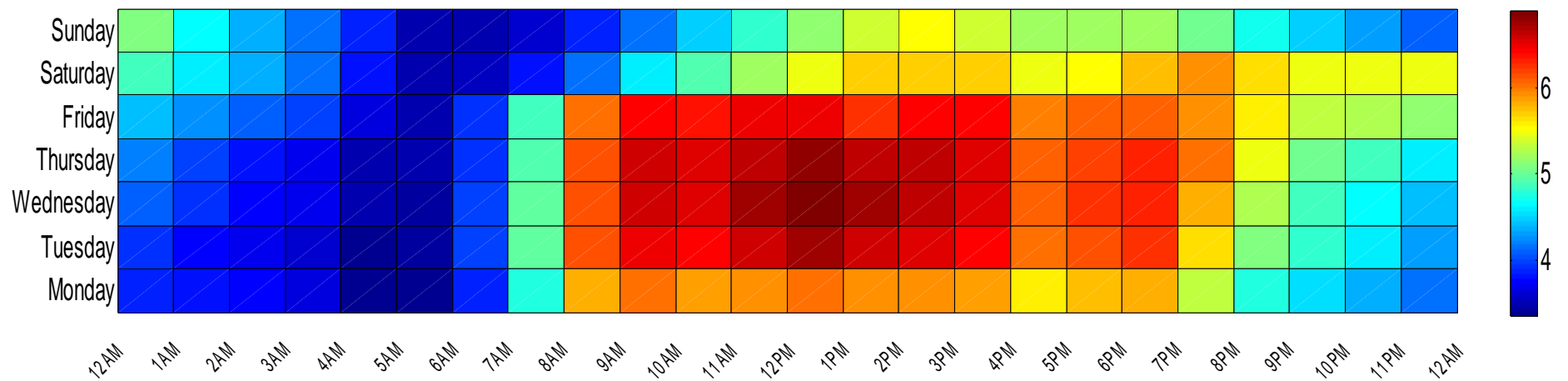
Causes of Delay

❖ <u>Assumption</u>: Additional fare between 5AM-6AM ≈ signal delay only

❖ <u>Literature</u>: Peak-hour signal delay ≈ 1.57 x off-peak signal delay

➢ Recurrent congestion delay =

Additional fare under clear weather – signal delay

➢ Weather related delay =

Total additional fare under rain - Additional fare under clear weather

Causes of Urban Delay
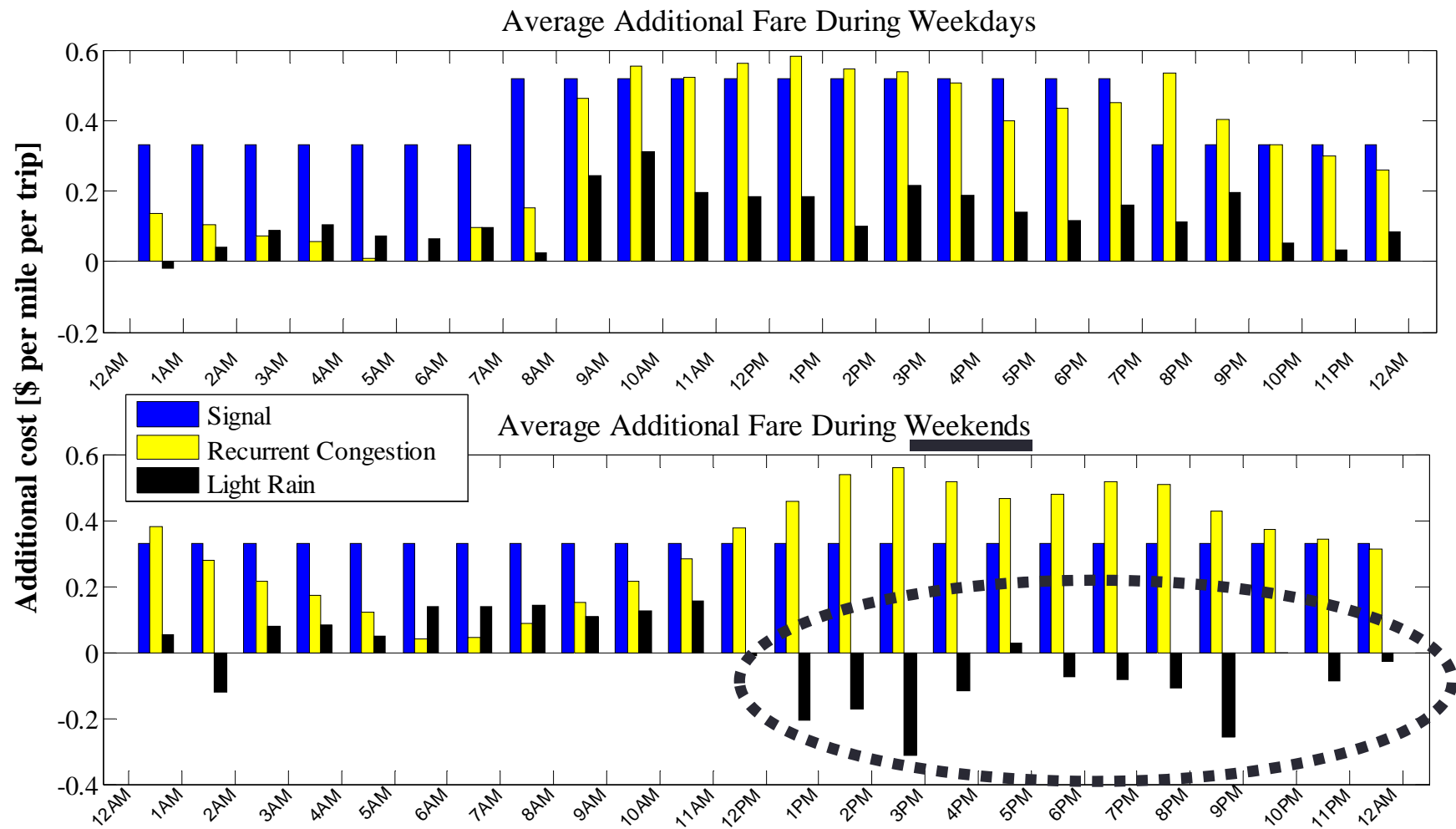
Thank you!
Q&A

Anil.Yazici@stonybrook.edu