Yiming AI
Haozhe CHENG
Xiangyi CHEN
Tianyang HUANG
TADAOUT Omar

ARTEFACT
VALUE BY DATA

&

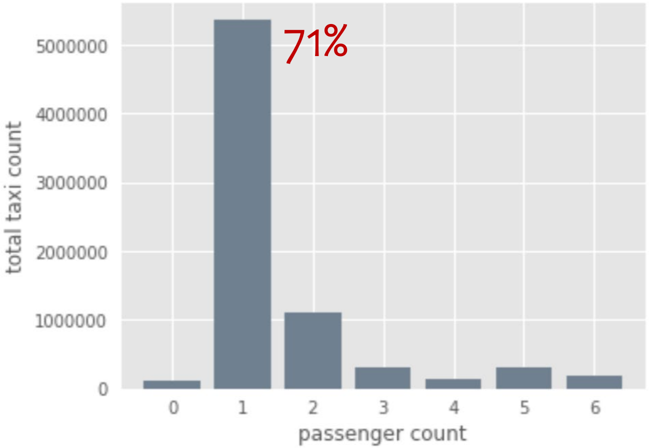
SHANGHAI JIAO TONG UNIVERSITY
1896

Carpooling Study
Based on NYC Yellow Taxi Dataset

Contents

Introduction: idea and concept

A glimpse during EDA



Low travelling efficiency

Dilemma of traditional taxis

UBER

lyft

DiDi
More than a journey

BlaBlaCar

Objectives

Analysis on the *potential* of carpooling.

Development of a feasible carpooling algorithm *strategy*.

Analysis on *benefits* of carpooling.

Foundation of the research

Explanatory Data Analysis

EDA: Data Preparation

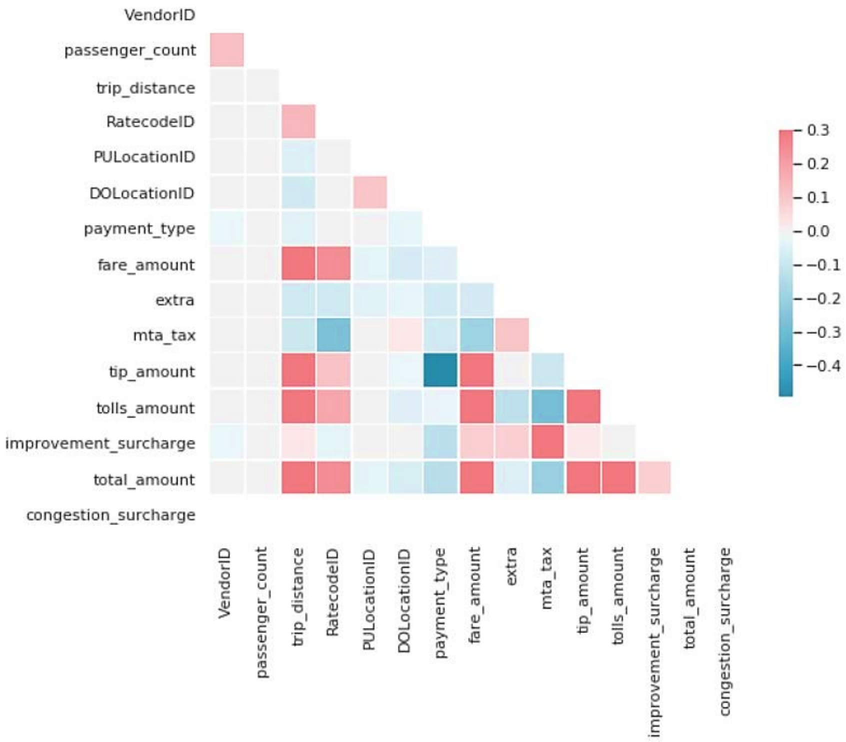
To simplify the calculations and facilitate subsequent research, we've made the following preparations:

1. Filling in *missing values*

2. *Adding columns:* trip_duration avg_ speed

3. *Data cleaning:*
   - Elimination of extreme values
   - Elimination of unreasonable values

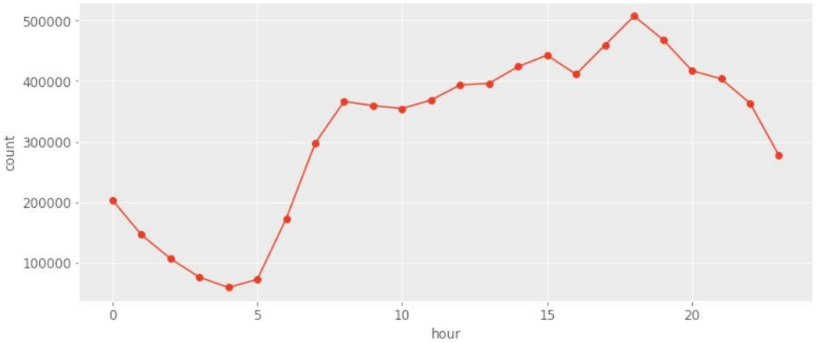
EDA: Data Preparation

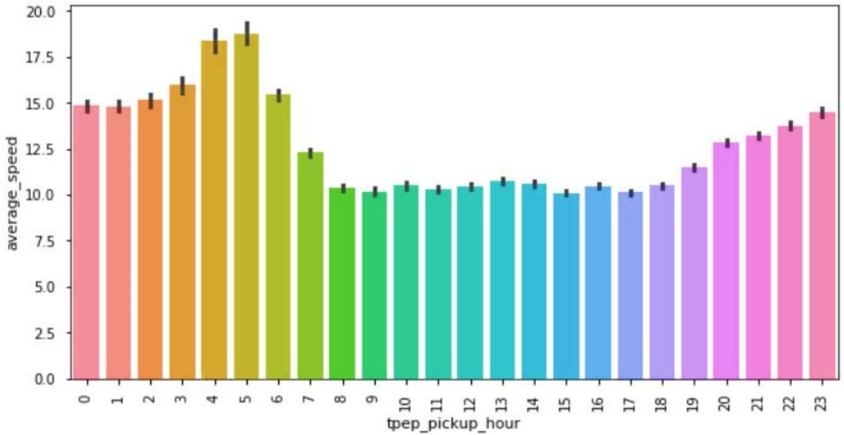


Correlation among all the variables

With analysis and comparison of the correlations among all the variables, we selected those *highly correlated* variables and considered them of analytical value for subsequent research.

EDA: Distribution of different hours





The *number* of taxis

The *average speed*

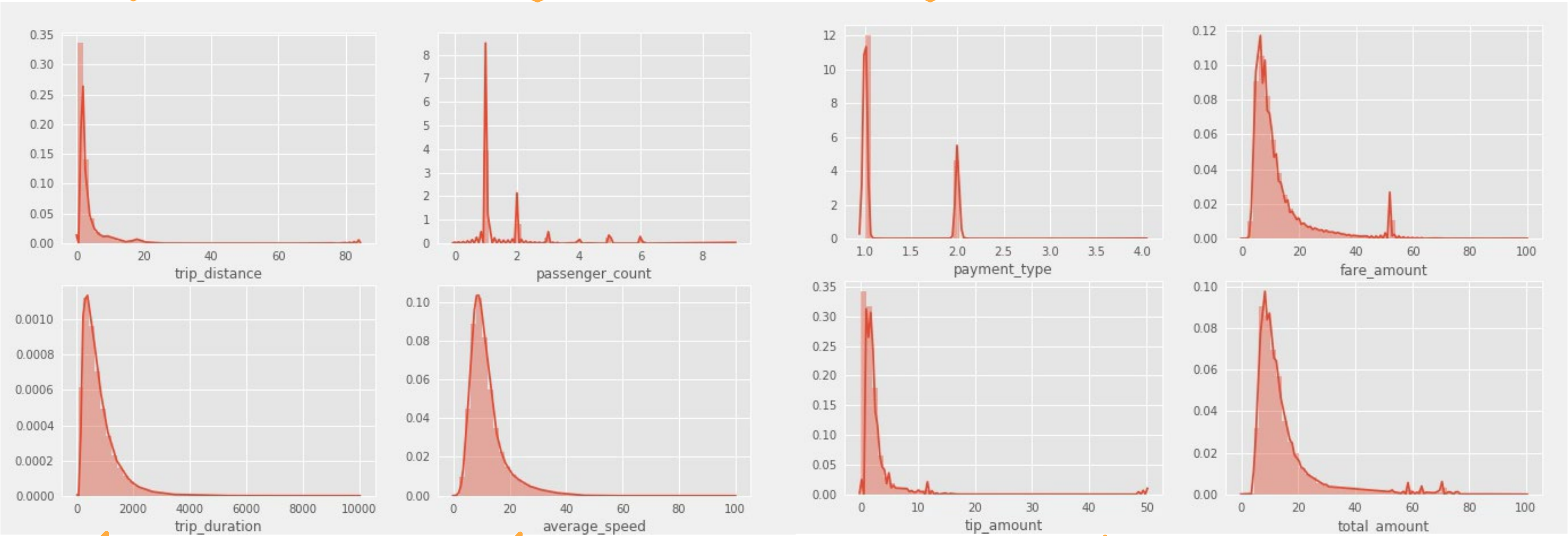
i.e. the *worse* the *congestion*.

EDA: Distribution of other variables

Trip distance

Passenger count

Payment type

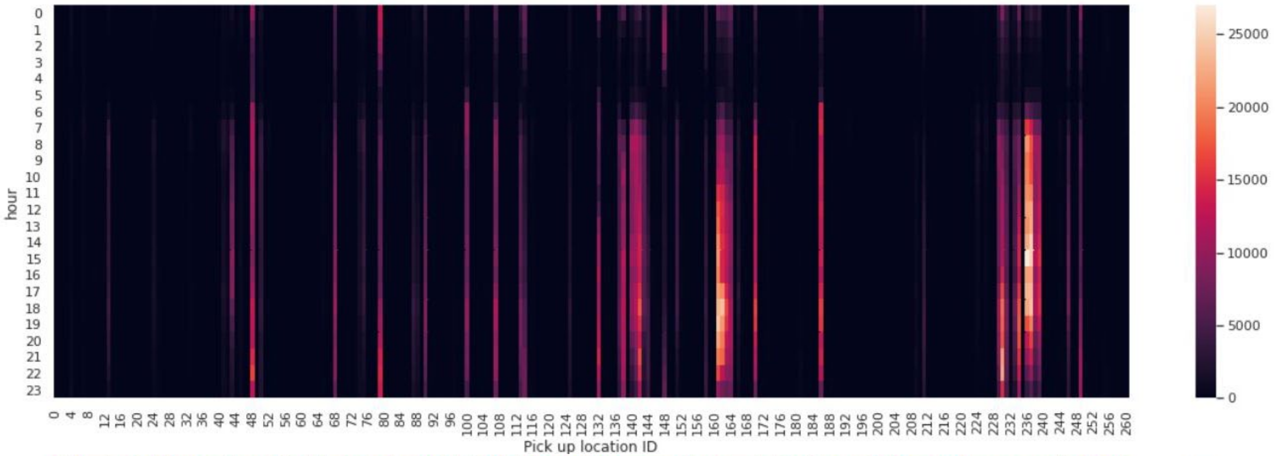


Trip duration

Average speed

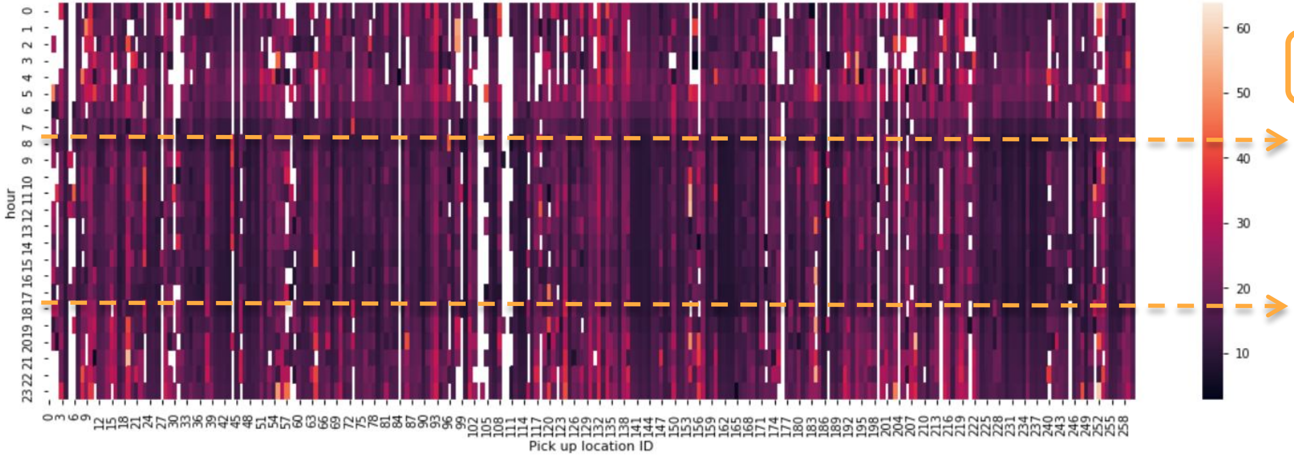
Various amount

EDA: Distribution of taxi count and average speed in different locations



Heat map of taxi count

The distribution of moments and Pickup location IDs is quite ***uneven.***



Heat map of average speed

Commuting

Morning peak

Evening peak

During the commuting, the average speed is obviously low.

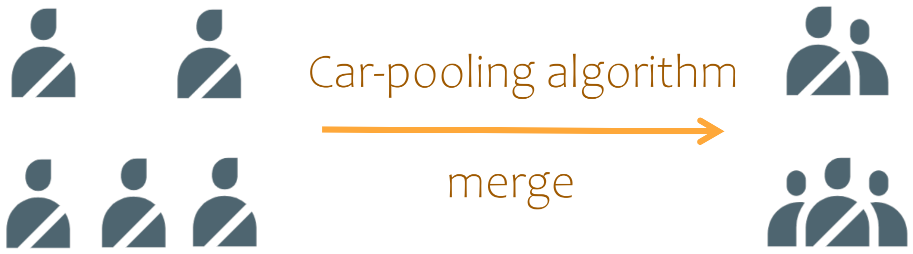
Model and algorithm for identifying and analyzing

Car-pooling strategies

Introduction: idea and concept

Taxi car-pooling potentials & general car-pool criteria:

1. Single passenger
2. Same pick-up place
3. Same drop-off place
4. Same departure time



General assumptions of the model

1. All potential car-poolers are *single passengers*.

2. All passengers with a potential car-pool partner are *willing* to car-pool.

3. Each car-pool represents a *merge* of several trips that *retains* the trip distance and trip duration of *the first trip being merged*.

The car-pool strategy algorithm

Separate time zones

Example: pick-up time 01:36 am

Hour: 01	Minute:36

zone 0 zone 1 zone 2 zone 3

Time zone = hour x 3 + int (minute ÷ 20)

Car-poolers arrangement

ALL TRIPS IN THE DATASET

- 0 Non car-pooler
- 1 "Hailer" car-pooler
- 2 "Ride-taker" car-pooler

Group 2333

Identify and group potential car-pool trips

DATA

CAR-POOL CRITERIA

1. Passenger count = 1
2. Same pick-up Location ID
3. Same drop-off Location ID
4. Same time zone for pick-up time

len(Group)≥2

Group 1
PU ID = 1
DO ID = 1
Time-zone = 0

⋮

Group 11638
PU ID = 263
DO ID = 263
Time-zone = 72

Apply the car-pool strategy

New data after car-pooling

- 0 Non car-pooler
- 1 "Hailer" car-pooler
- ~~2 "Ride-taker" car-pooler~~

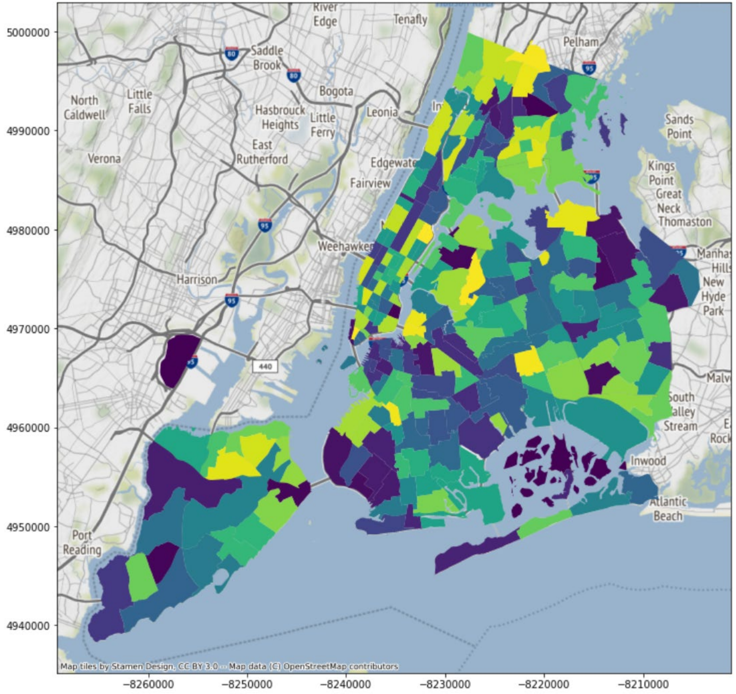
Carpooling's role in

Congestion Reduction

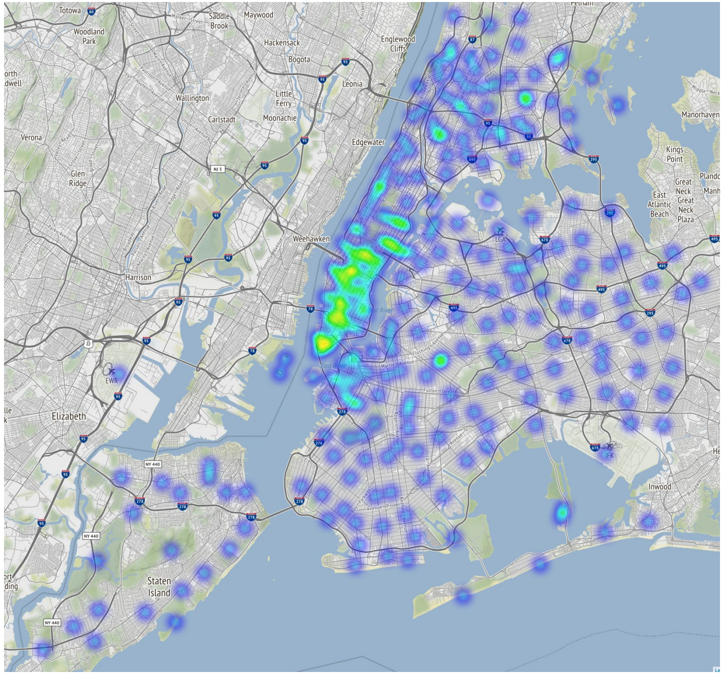
Geographical Approach

Problem Simplification

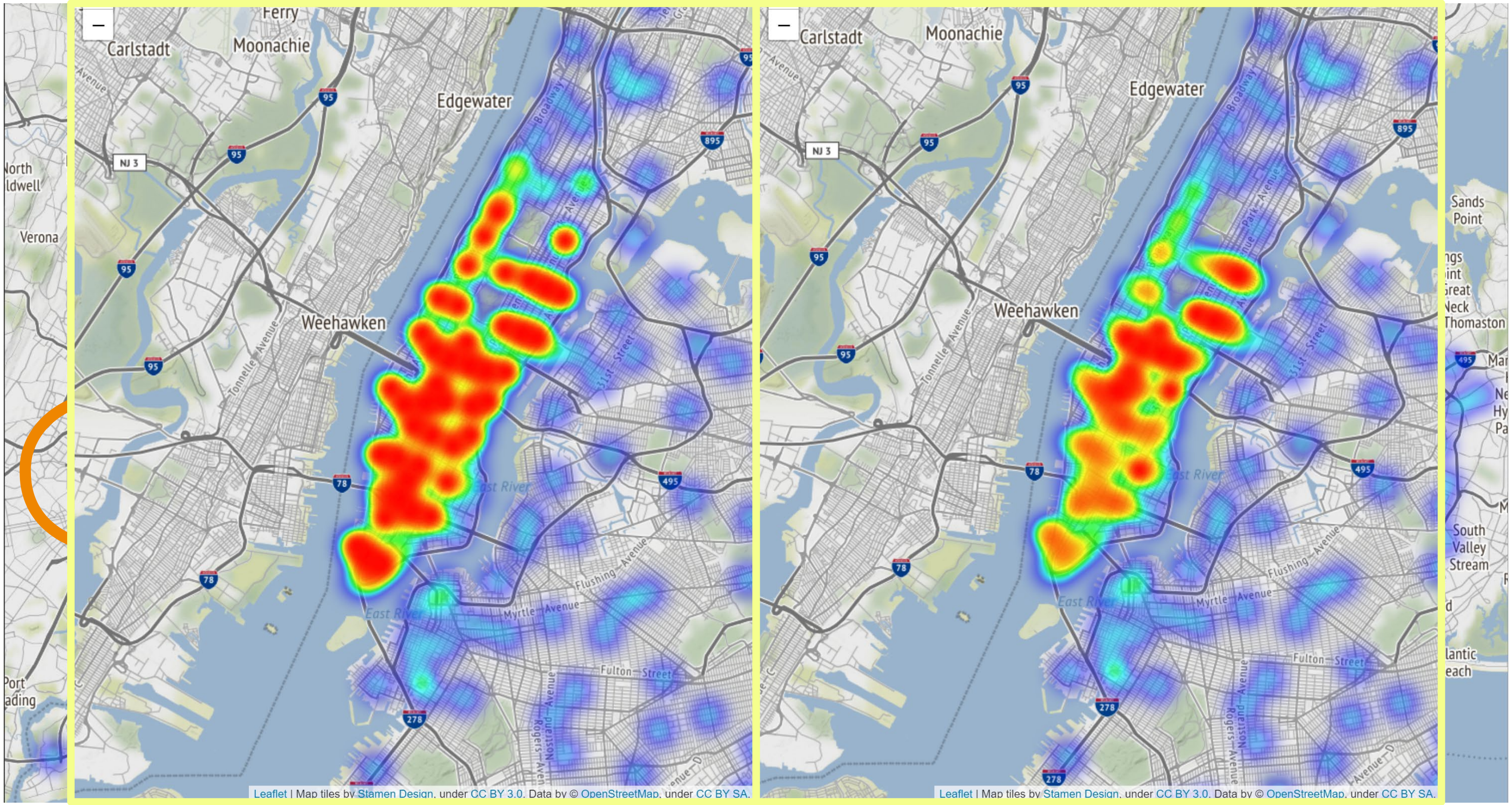
ID zones



Hotspots representing corresponding ID zones

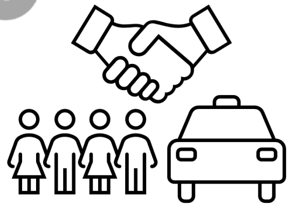


Result: Decentralization of demand



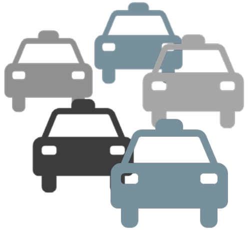
Quantifying Congestion

1



Car pooling

2

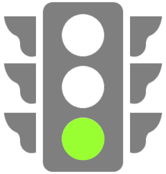


Number of cars drops

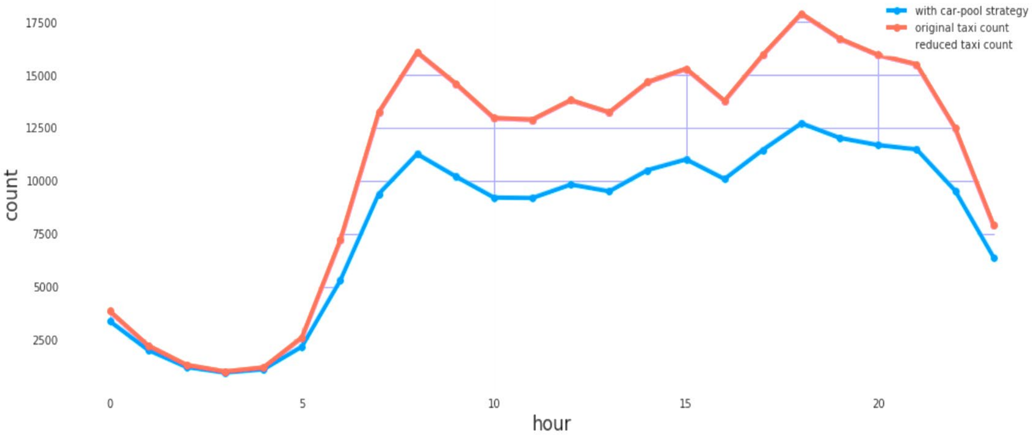
3

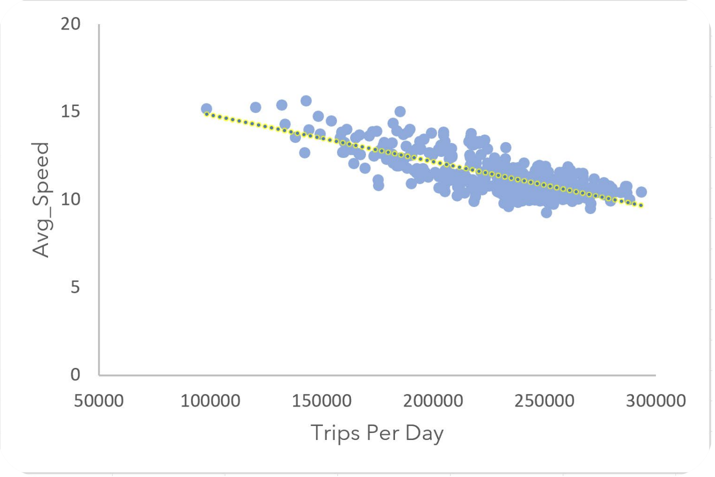
Average Speed increases

4



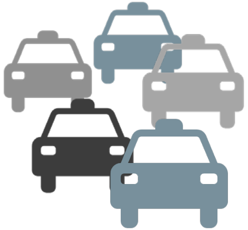
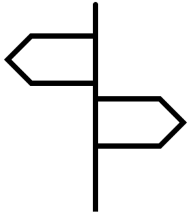
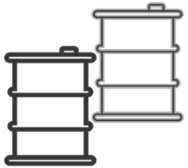
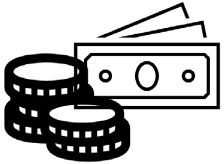
Congestion reduces





Quantifying Carpooling

Interval: 10 min Car-pooler: 2

				
Total Trips	Mileage(mile) per day	Fuel Saved(L) per day	Cost saved($) per passenger	Avg_speed(mile/hr)
-13.1%	-61436	+6426	+5.28	+1.4

Quantifying Carpooling

Interval	Car-pooler	Total Trips	Mileage(mile)	Fuel Saved(L)	Cost saved($)	Waiting time(min)
10 min	2	-13.1%	-61436	6426	5.28	+4’51”
10 min	3	-17.2%	-79858	8353	7.25	-
20 min	2	-18.0%	-89898	9409	5.51	+5’44”
20 min	3	-23.8%	-117897	12333	7.55	+7’31”
20 min	2	-20.7%	-107078	11201	6.71	+6’42”
30 min	3	-27.4%	-140824	14731	7.73	-
			Per day	Per day	Per trip	

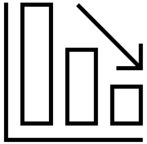
≈ 30,000 Barrels

Carpooling’s role in Competitiveness

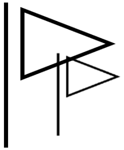
New Fare Calculation Method

K(%)

I. K(%) decrease of original cost



II. Decrease of demand



III. Reallocation of Idle Drivers

Keff(%)

IV. Keff(%) boost in income

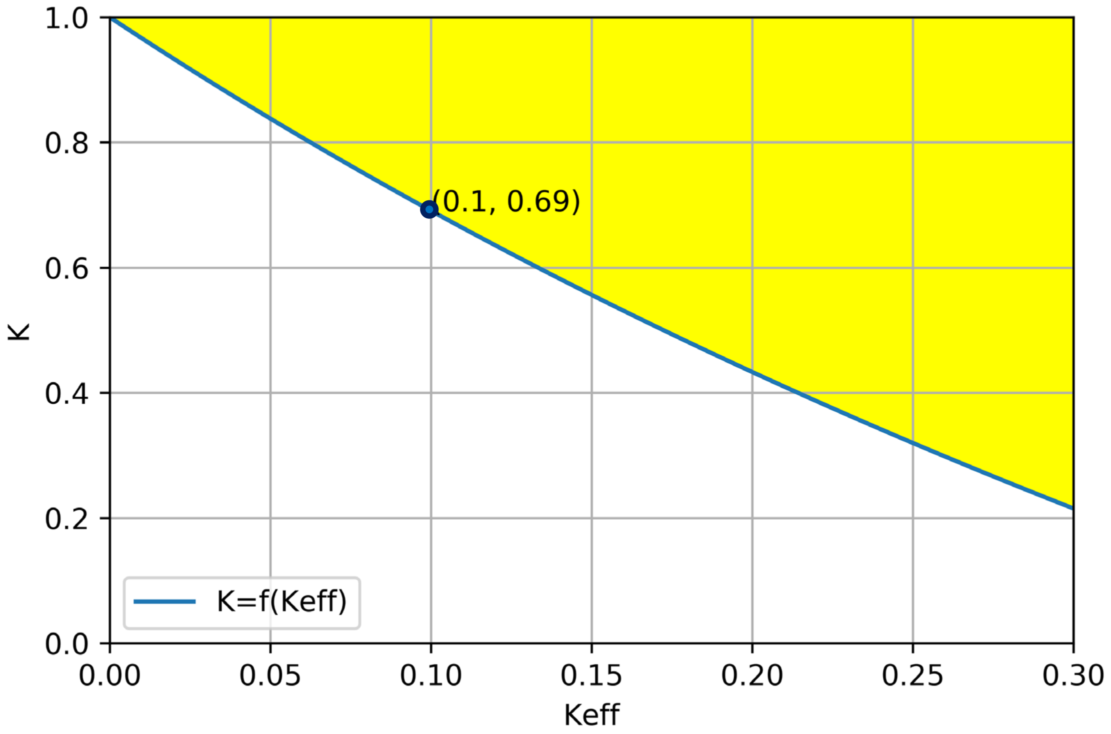
New Fare Calculation Method : Inequation

$$\begin{cases} (1+K_{eff})\left(\sum_{\text{those who car pool}} K \times \text{original fare} + \sum_{\text{those who don't car pool}} \text{original fare}\right) \geq \sum_{\text{everyone}} \text{original fare} \\ 0 \leq K \leq 1 \end{cases}$$

The taxi drivers should earn no less than their original income

The passengers should not pay more than how much they originally pay.

New Fare Calculation Method : Solution



Conclusion: Advantages and Limits

30,000 Barrels saved per year

Lower maintenance cost & longer life span of taxis

Average speed +1.4 miles/hr

Smarter and more flexible

Higher income for drivers

A more attractive pricing strategy

Longer Waiting time +5’

Heavier Burden On Privacy Protection

Higher dependency on smart phones

END

THANK YOU