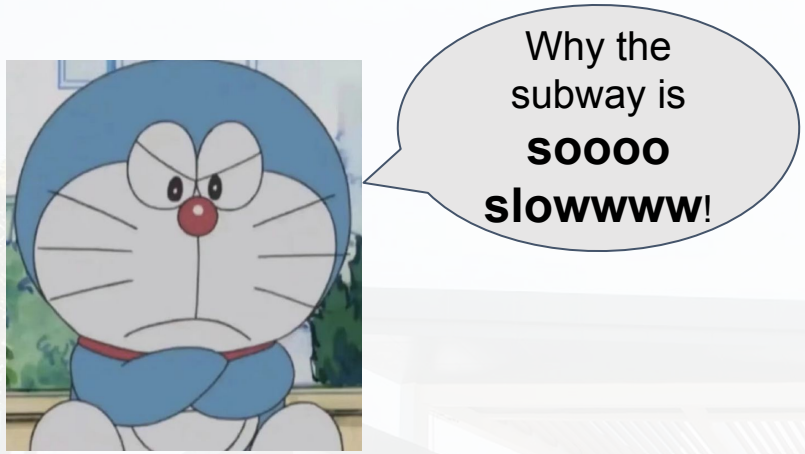
Data Science Practice :

Shenzhen Metro System Optimization

SUSTech
Southern University of Science and Technology

REPORTERS: Guo Cheng, He Qijia, Wang Lu, Yan Lingmin, Zhou Jiachen

Motivation & Background



- Shenzhen's metro system faces significant challenges during peak hours, including overcrowding and long waiting times.
- Line 5 is the main line running east to west, with a daily passenger volume of 1.0262 million, making it the busiest metro line in Shenzhen.
- **Optimization approach**: Through algorithm optimization, certain trains will be designated to skip specific stations during peak hours to reduce overall commuting time across the network.

SUSTech Southern University of Science and Technology

CATALOGUE

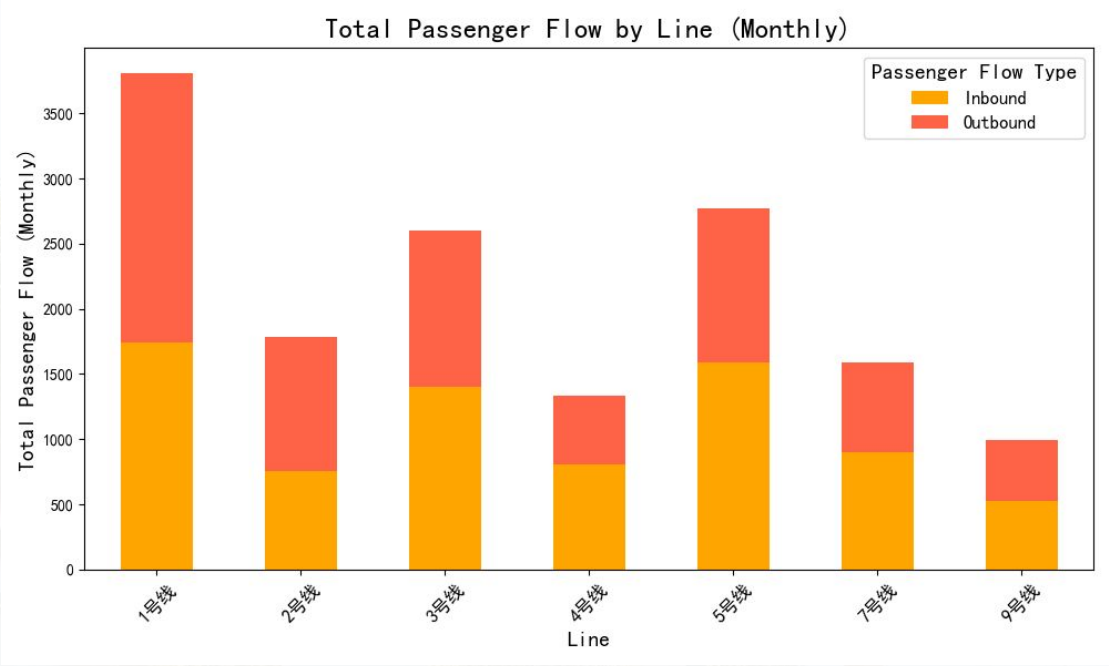
SUSTech Southern University of Science and Technology

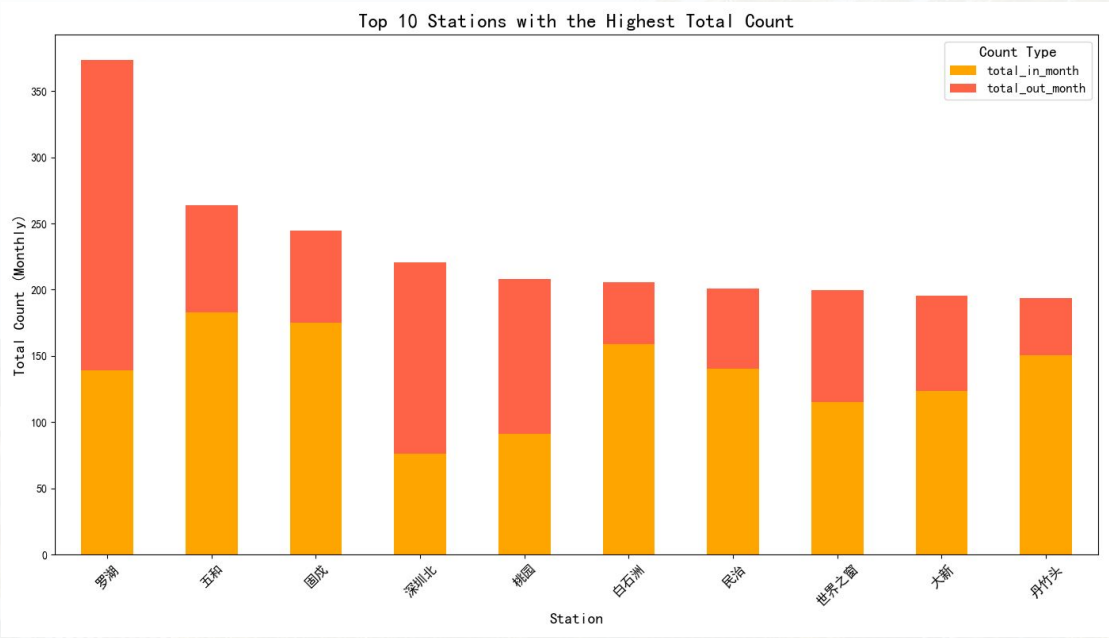
01

EXPLORATORY DATA ANALYSIS

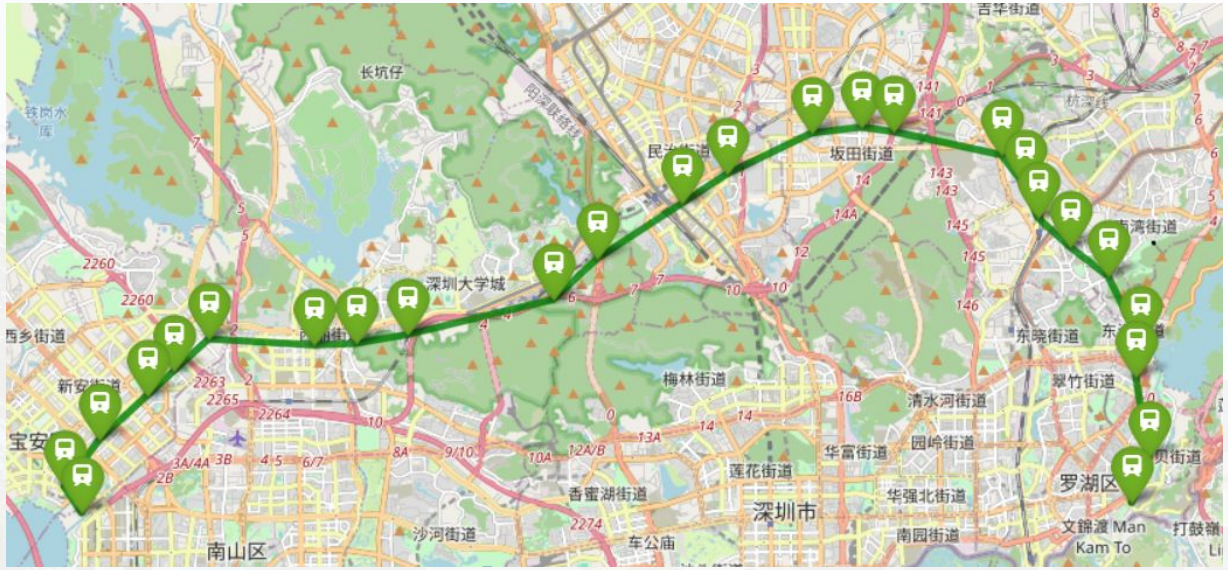
Data introduction and visualization.

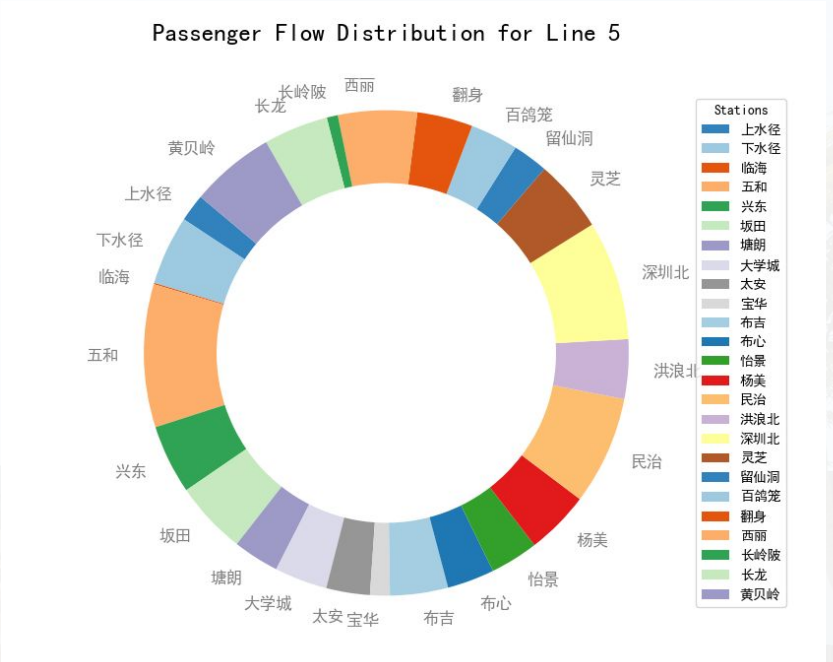
Passengers flow of different stations & lines are different





- Total passenger flow ranking for different lines: line 1 > line 5 > line 3
- Total passenger flow ranking for different stations: Luohu > Wuhe > Gushu





- Shenzhen Metro Line 5 follows an M-shaped route from southeast to northwest to southwest with a total length of 47.393 kilometers and 34 stations.

- The three busiest stations on Line 5 are: Wuhe, Shenzhenbei Station, and Minzhi Station.

SUSTech Southern University of Science and Technology

02

ALGORITHM DESIGN

Greedy algorithm and local search.

Model Assumptions

- The time saved remains constant for each station if a station is skipped.
- The distance between each neighboring station is constant.
- Subway crashes are not considered since we are optimizing scenarios with two subway railroads.
- There is a fixed capacity for each subway. This allows us to discuss situations where too many people are entering the subway.

Problem Formulation

parameters	pure simulated	Shenzhen data
station_interval	2	1.5
skipped_saved	1.5	1.2
station_N	20	25
train_N	28	27
passengers_max	110	75
N	20,000	10,000

Table 1: parameter table

ID	start	end	time
0	2	7	110
1	9	18	23
2	16	13	152
...	...	...	...
N-1	16	10	152

Table 2: input passengers flow

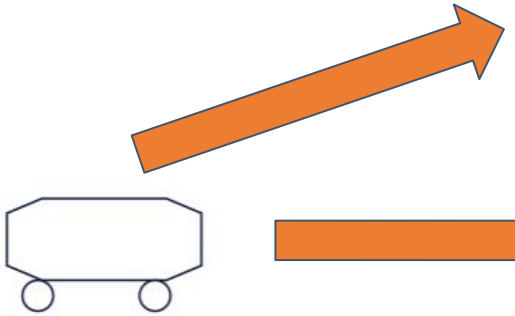
station_0	station_1	station_2	...	2*station_N
0	1	1	...	1
1	0	1	...	1
1	0	1	...	0
...	...	...	...	...
1	1	0	...	1

Table 3: output subway schedule

Greedy Algorithm

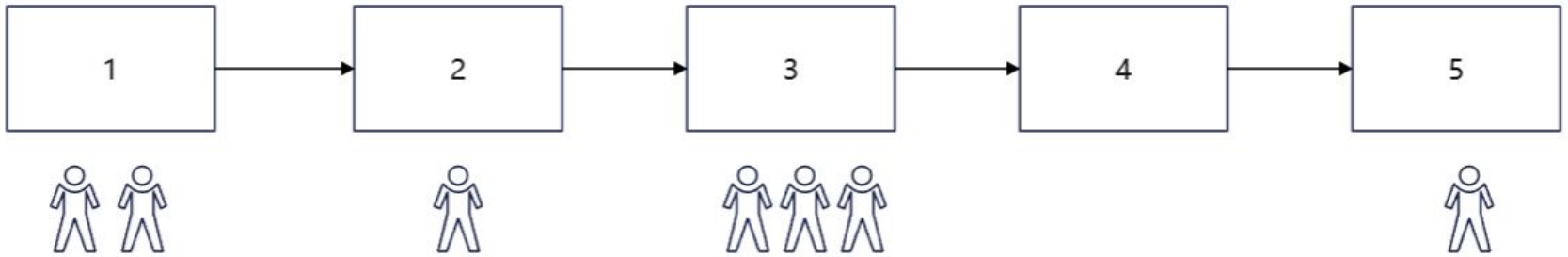
ID	start	end	time
0	2	7	110
1	9	18	23
2	16	13	152
...	...	...	...
N-1	16	10	152

Skip?



Stop √

ID	start	end	time
0	2	7	110
1	9	18	23
2	16	13	152
...	...	...	...
N-1	16	10	152

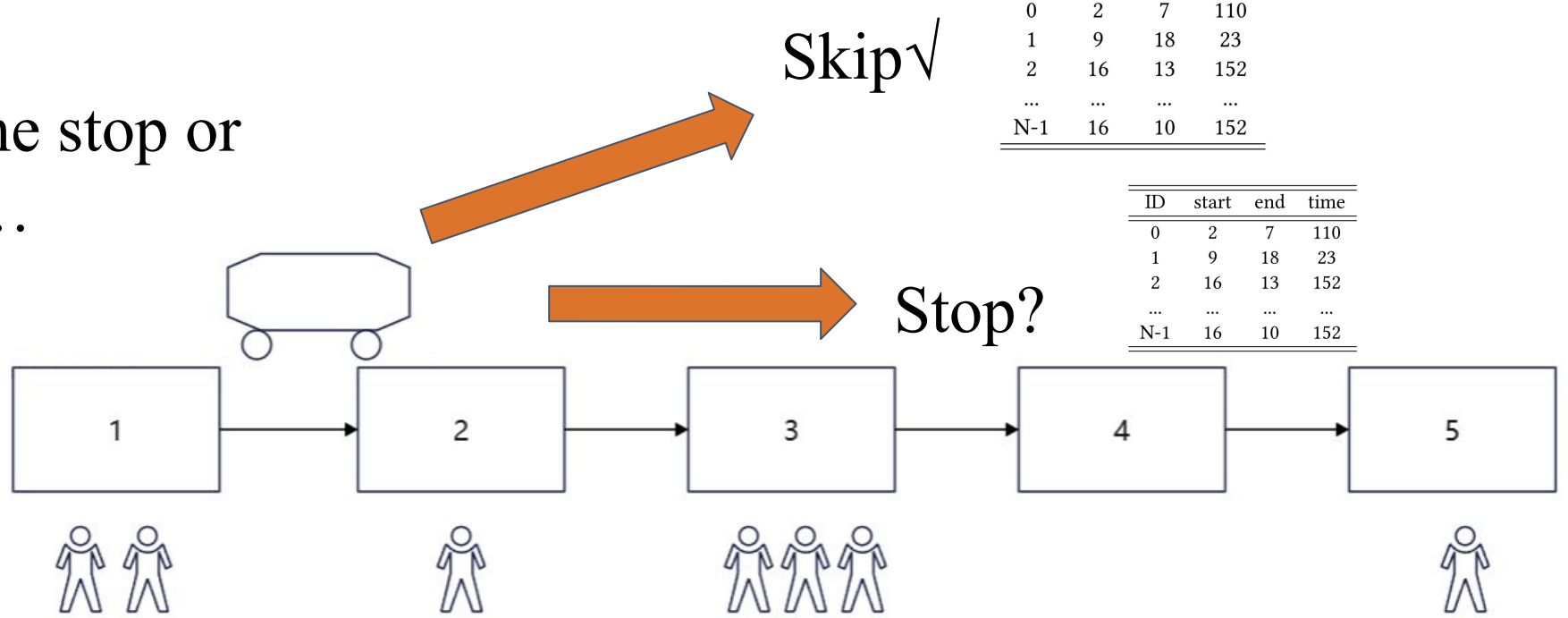


Initial Stage	stop	stop	stop	stop	stop
Improved Stage..	stop				

Greedy Algorithm

Keep asking the stop or skip question…

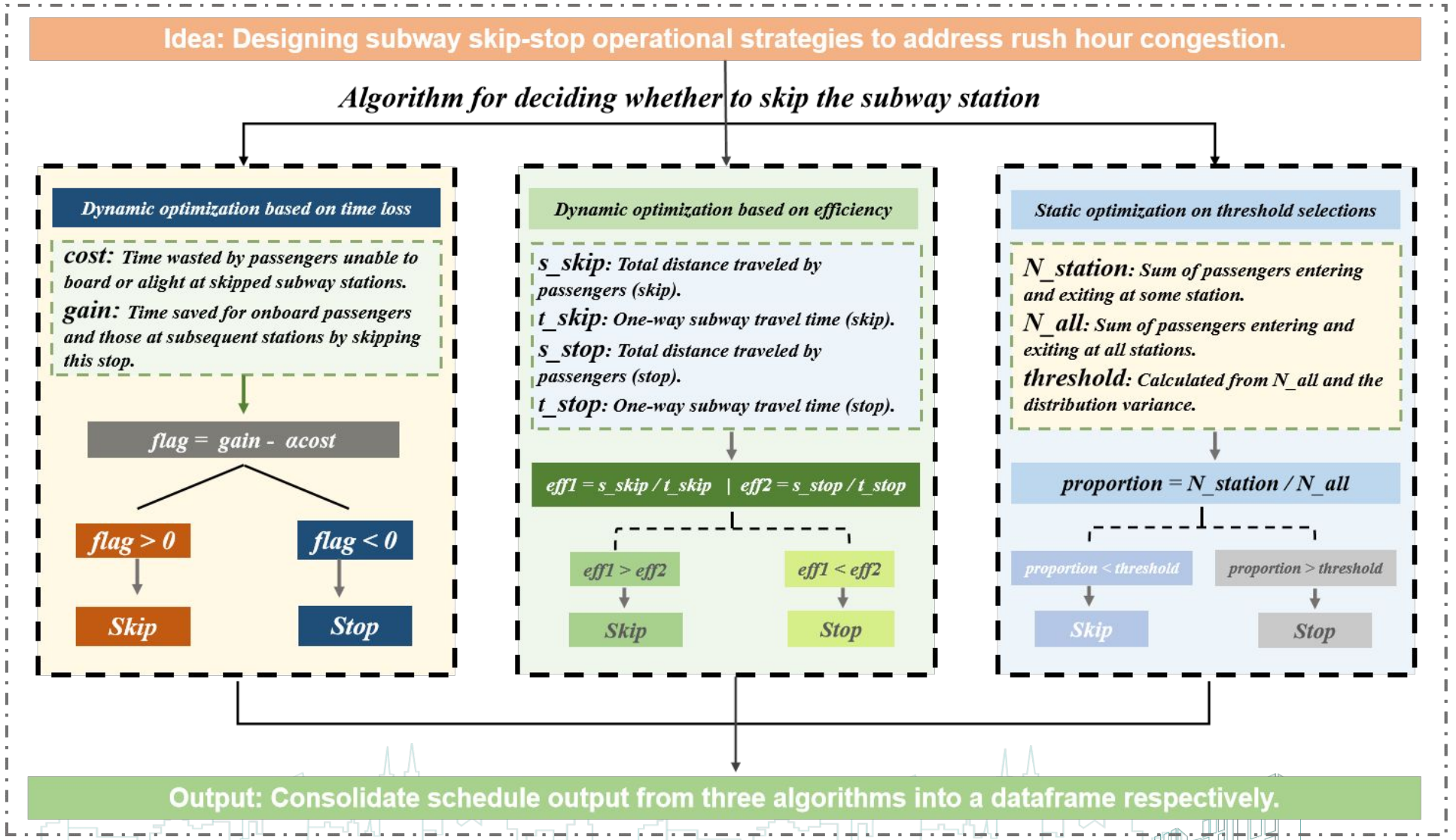
ID	start	end	time
0	2	7	110
1	9	18	23
2	16	13	152
...	...	...	...
N-1	16	10	152



ID	start	end	time
0	2	7	110
1	9	18	23
2	16	13	152
...	...	...	...
N-1	16	10	152

Initial Stage	stop	stop	stop	stop	stop
Improved Stage..	stop	skip			
Improved Stage	stop	skip	stop	skip	stop

Greedy Algorithm



ID	start	end	time
0	2	7	110
1	9	18	23
2	16	13	152
...	...	...	...
N-1	16	10	152

Table 2: input passengers flow

Implementation of Local Search

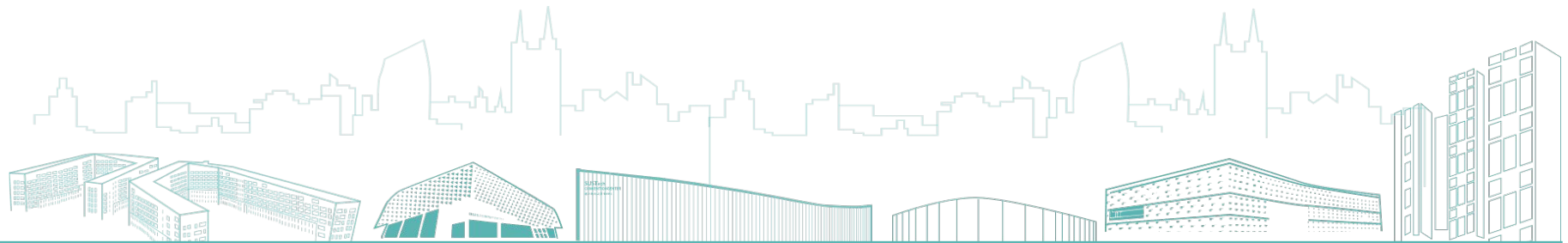
- **Optimization Objective**: Minimize total passenger commuting time.

- **Process**:
  - Evaluate neighbors of the current schedule.
  - If a better neighbor is found, update the current schedule.
  - If not, retain the current schedule.

station_0	station_1	station_2	...	2*station_N
0	1	1	...	1
1	0	1	...	1
1	0	1	...	0
...	...	...	...	...
1	1	0	...	1

Table 3: output subway schedule

- **Neighbors Definition**: Based on Hamming distance of the schedule matrix.
  - Maximal neighbourhood distance set to 3.



Local Search

Why Local Search?

- **Complexity of Advanced Methods**:Other algorithms require extensive computation.
  - Example: Calculating total waiting time for 100,000 passengers takes ~2.5 minutes.
  - Running 1000 iterations could take 41.7 hours.

- **Computational Constraints**: The schedule matrix for the morning peak contains over 2,500 elements.
  - 1,000 iterations are not enough to reach global optimal

- **Efficiency of Local Search**:
  - Local Search provides a relative good feasible solution without excessive computational time.

SUSTech Southern University of Science and Technology

03

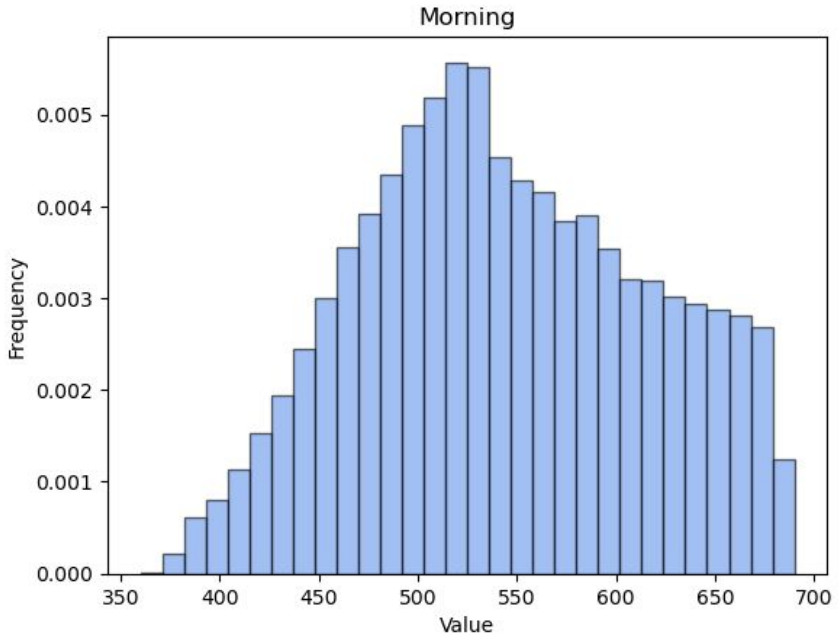
EXPERIMENT

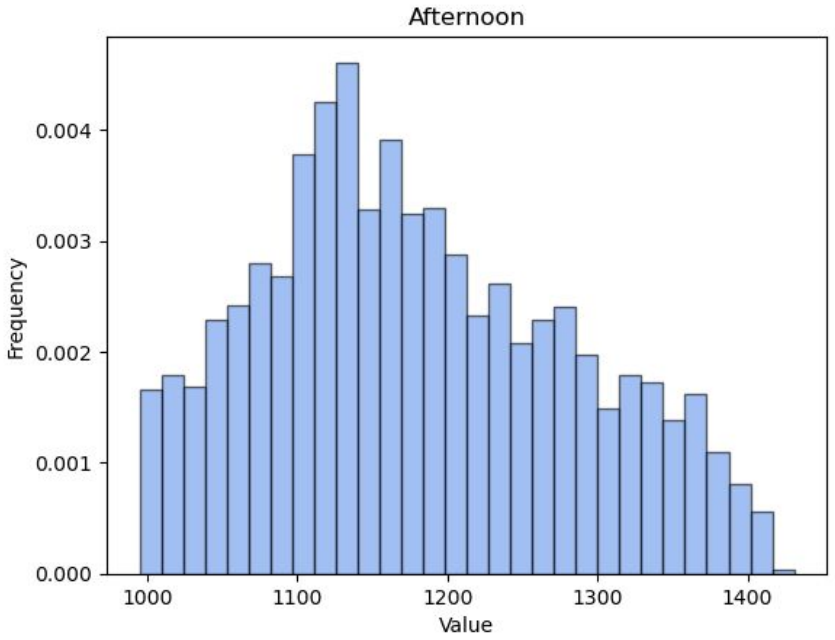
Experiment conduction and results.

Experiment

- **Simulated Data Generation**
  - Based on Real Data
  - Probability Generating Data
- **Experiment Results and Performance Comparison**
  - Commuting Time Distribution
  - Local Search (searching for optimal)
  - Analysis of Commuting Time Improvement
  - Algorithms' Feature Analysis
  - Algorithm Sensitivity Analysis

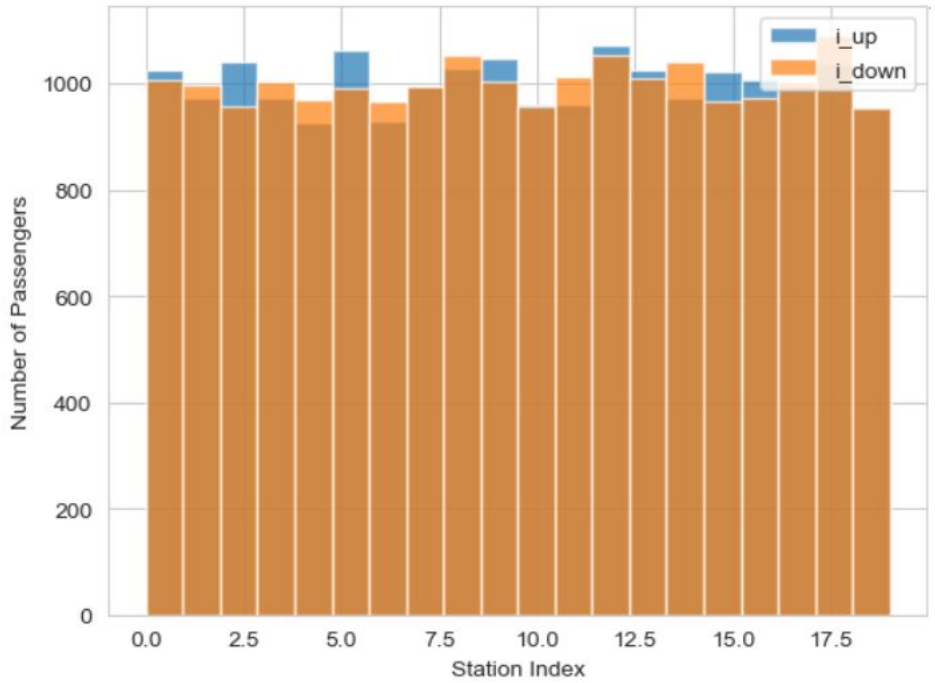
Based on Real Data

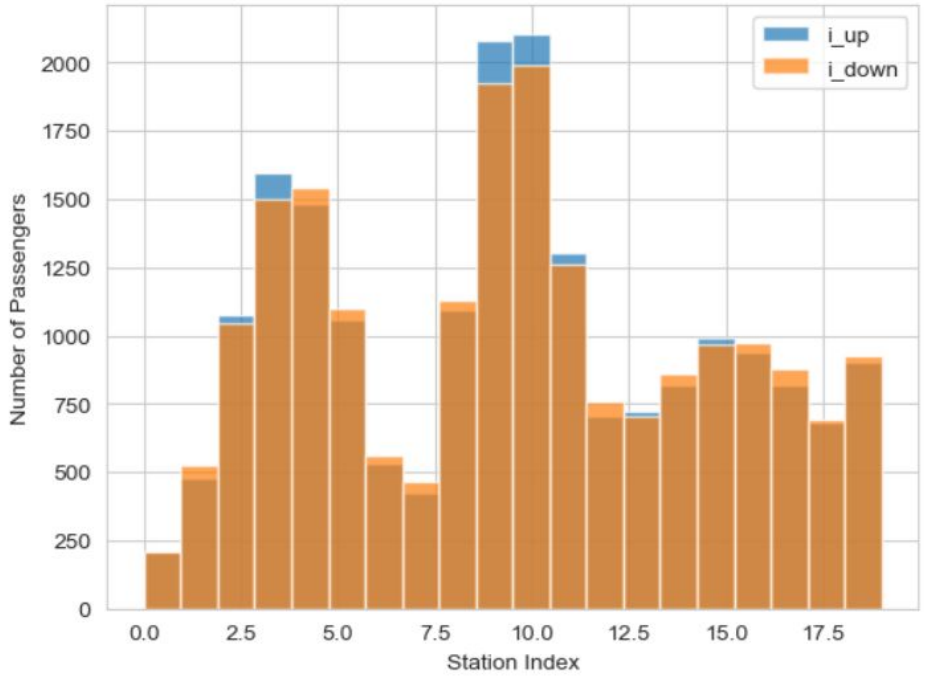




- Generate the simulated data based on the distribution of real data.

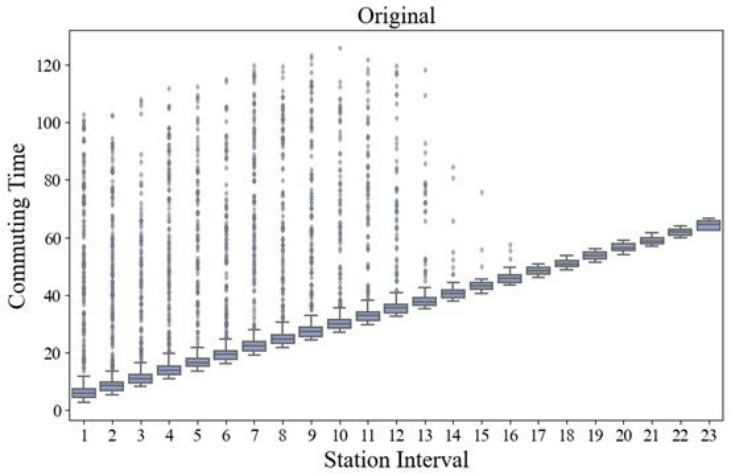
Probability Generating Data





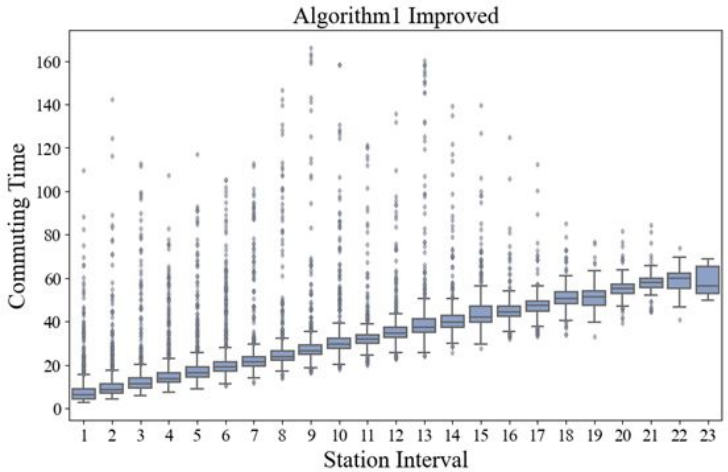
- We used uniform distribution and normal distribution to first generate a simulation Data

Commuting Time Distribution ~ Time Interval

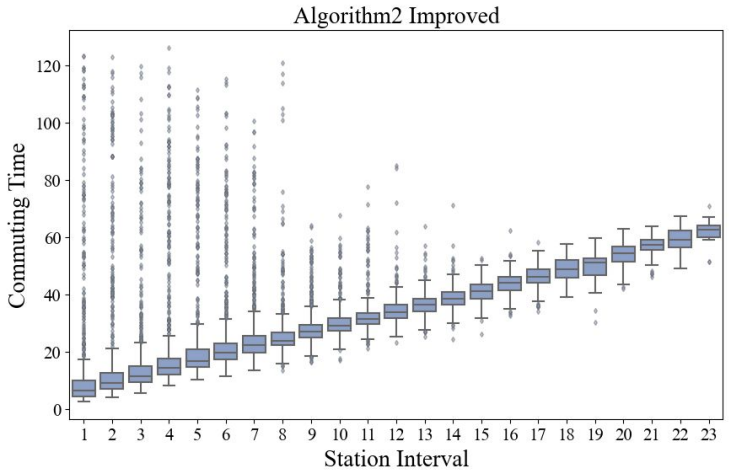


- Larger distance means longer commuting time
- Almost Linear
- Outliers are those who meet the train at morning peak

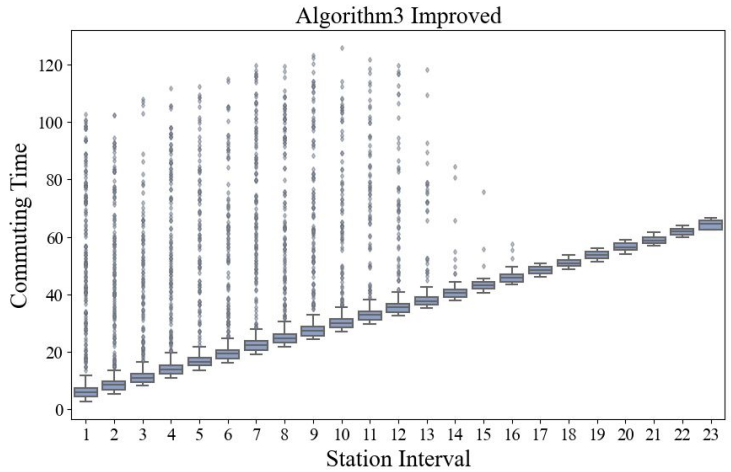
NOT OBVIOUS…



Algorithm 1

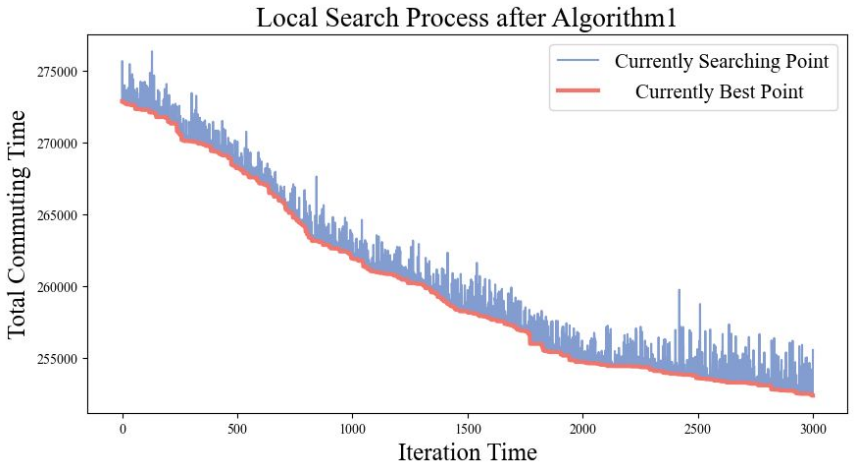


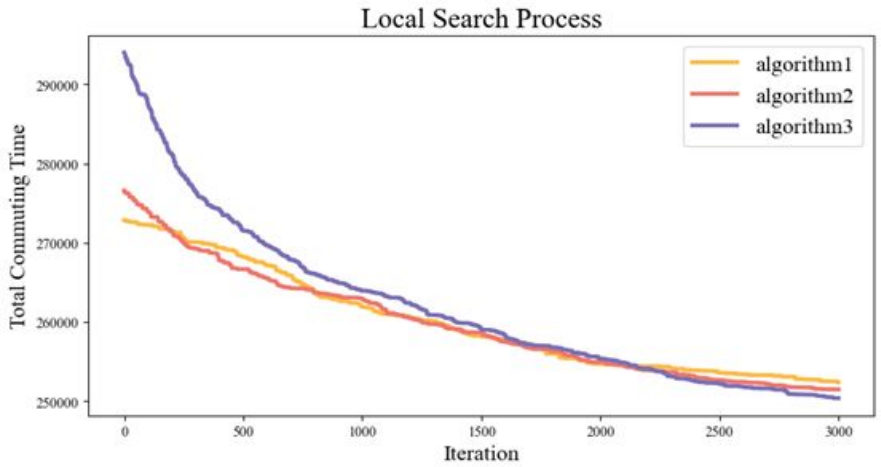
Algorithm 2

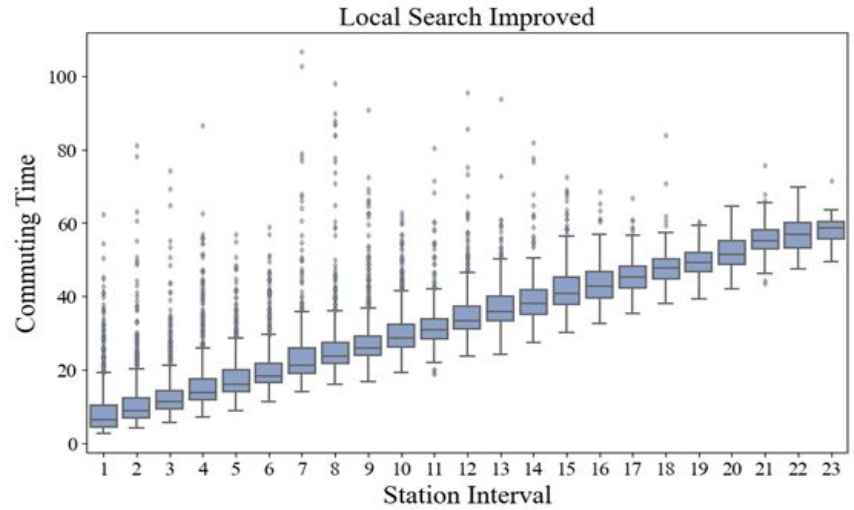


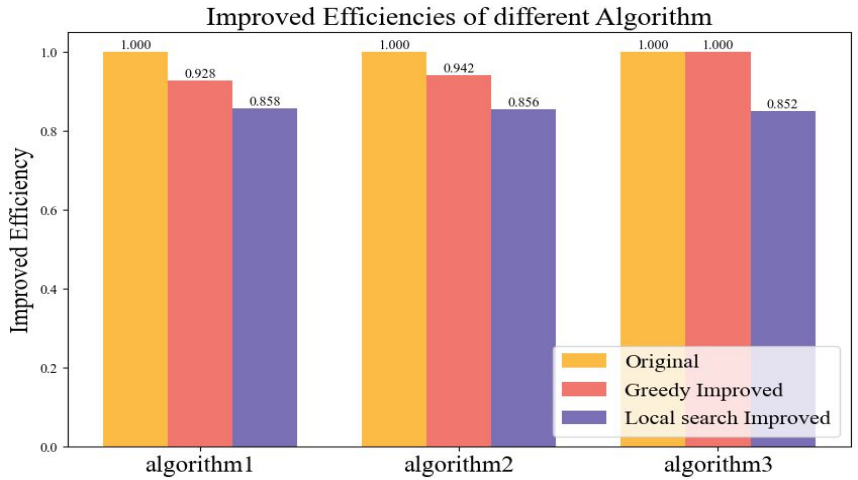
Algorithm 3

Local Search (searching for optimal)



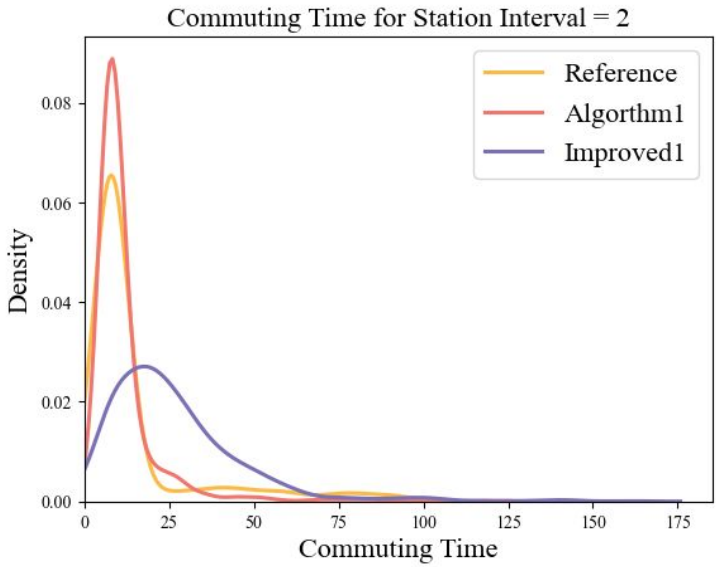


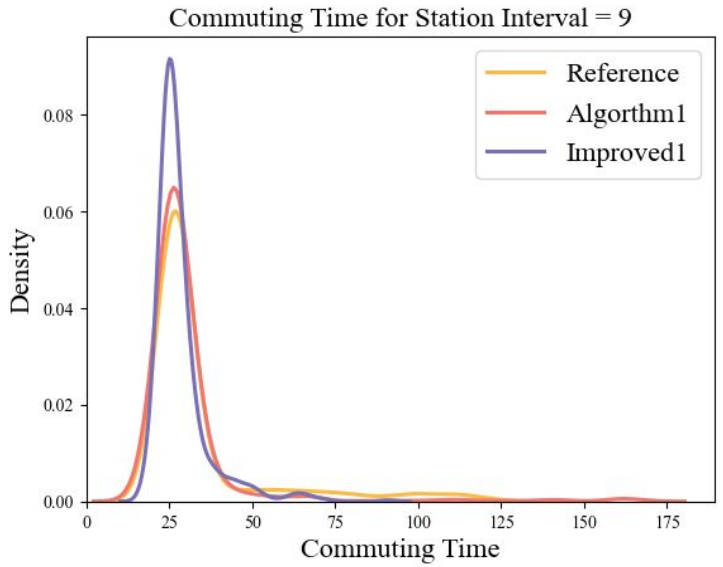


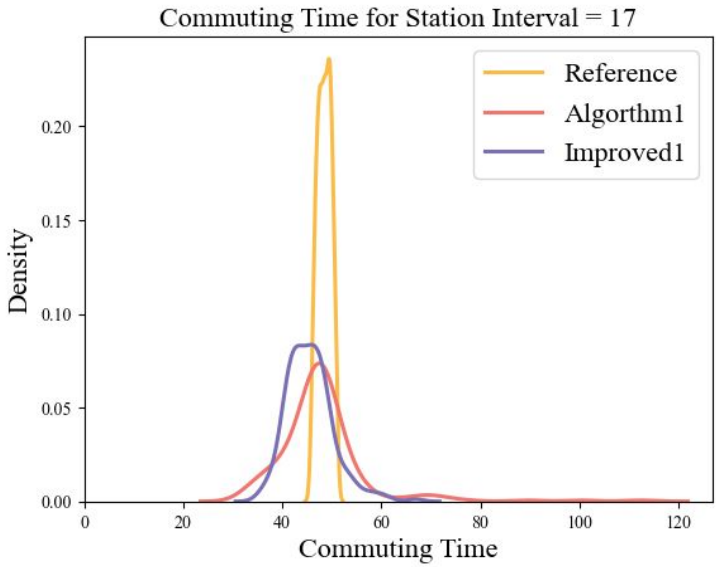


- Local search consistently elevates the total saved commuting time to approximately 15%.

Analysis of Commuting Time Improvement







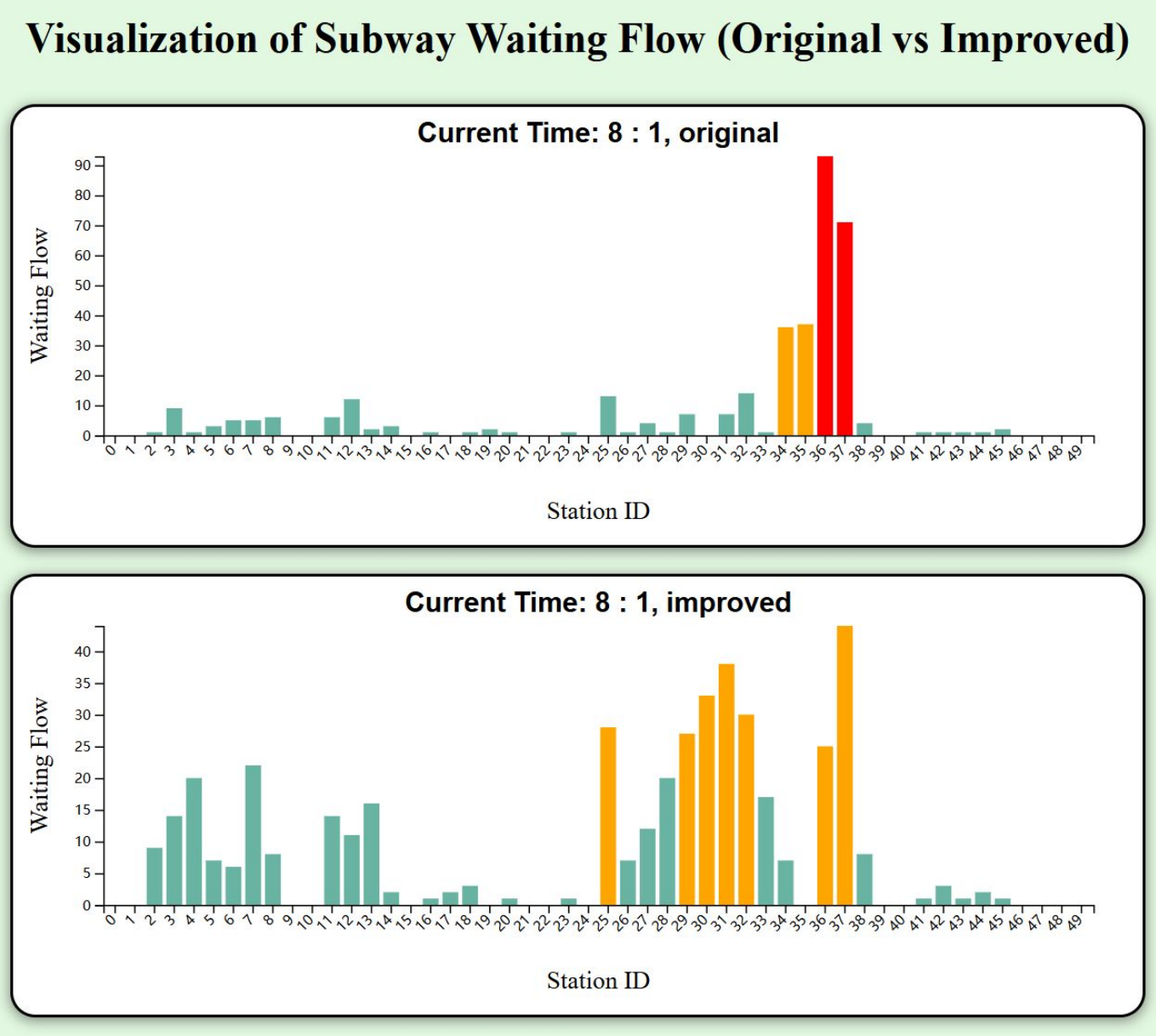
Station Interval = 2

Station Interval = 9

Station Interval = 17

- MORE TIME IS SAVED AS STATION INTERVAL INCREASES!

A short display of the performance visualization



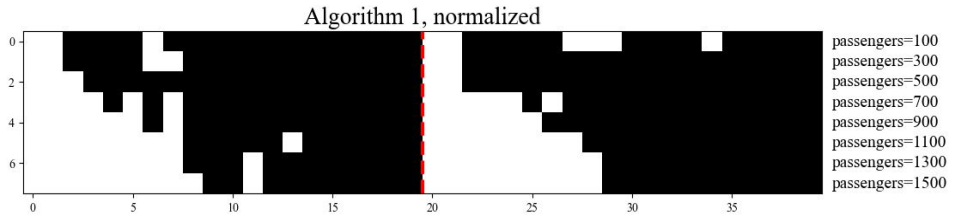
SUSTech Southern University of Science and Technology

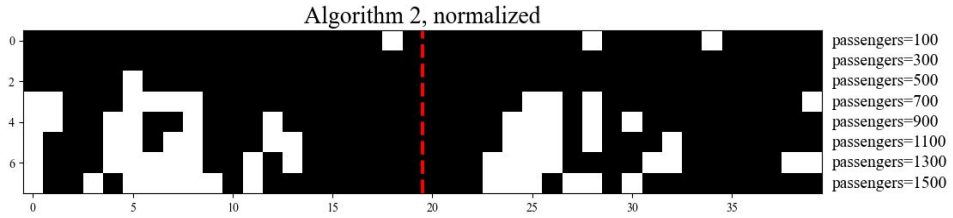
Algorithms' Feature Analysis

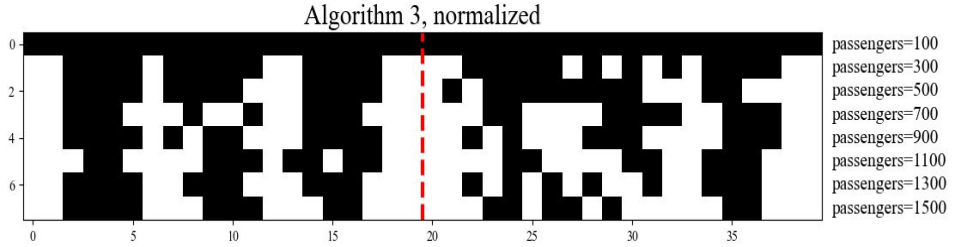
Algorithm 1

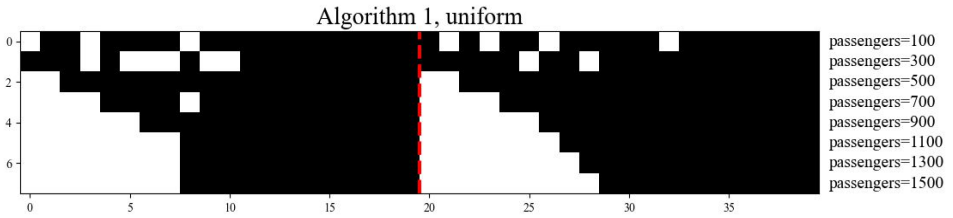
Algorithm 2

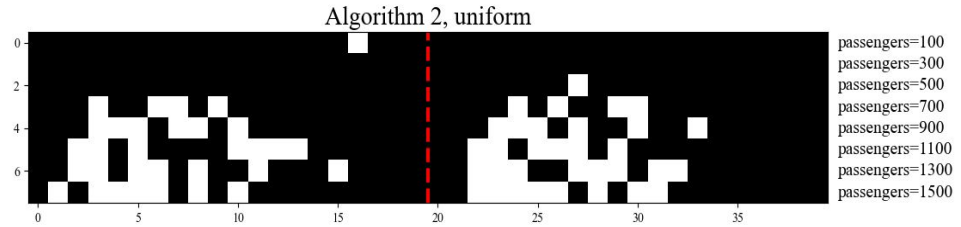
Algorithm 3

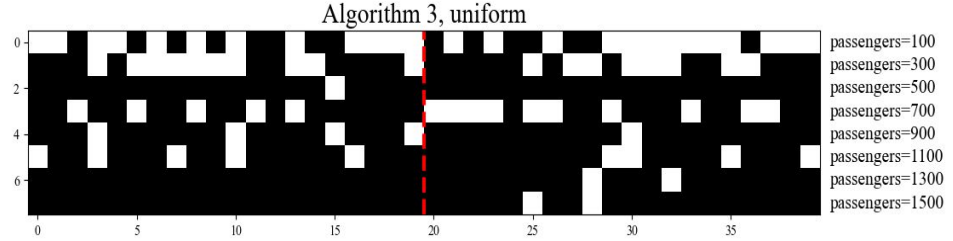












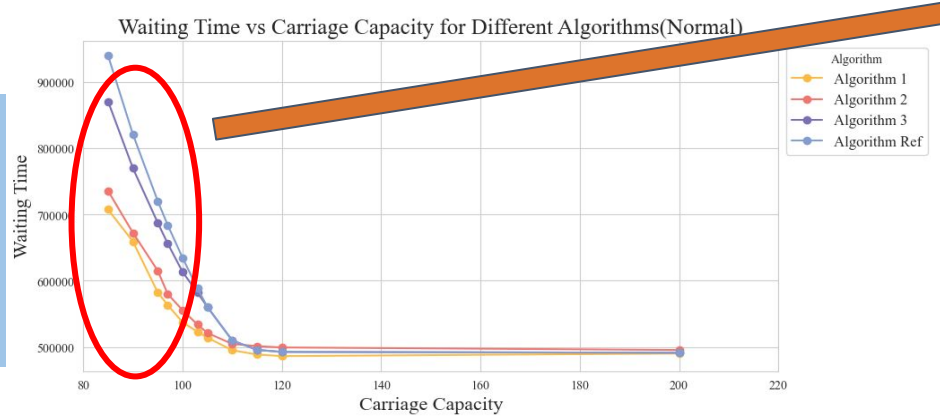
Normal distributed data

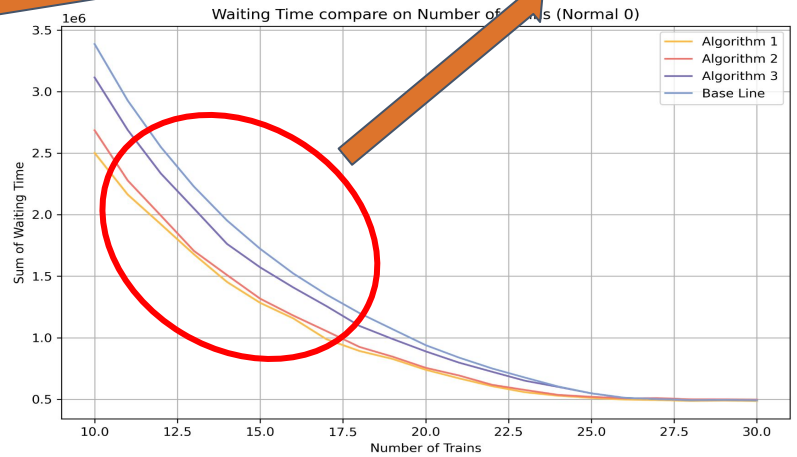
Uniform distributed data

Algorithms' Sensitivity Analysis

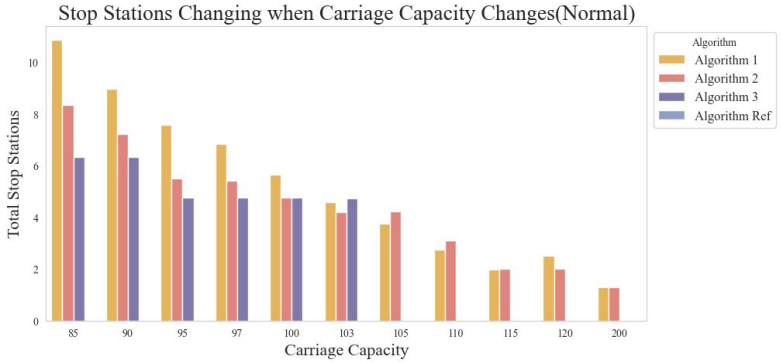
Know the risk!

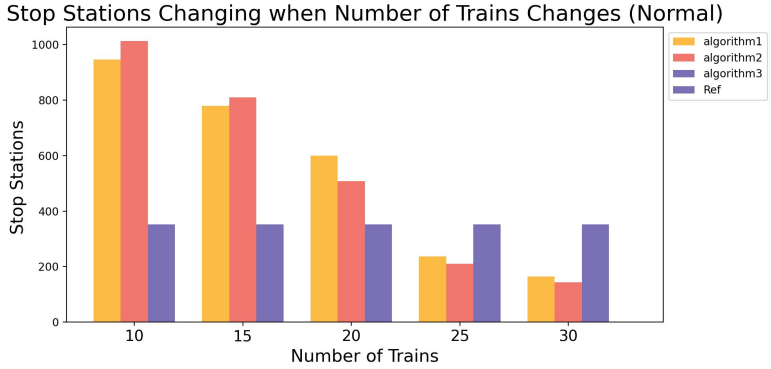
Total Commuting Time





\# Skipped Stations





Carriage Capacity

\# Trains

- IMPROVEMENTS ARE GREATER WHEN THE SYSTEM IS CROWDED

SUSTech Southern University of Science and Technology

SUSTech Southern University of Science and Technology

04

CONCLUSION

Strategy in real life and further dissemination.

Conclusion

- **Model application results**
  - An average decrease in travel time of approximately 15%.

- **Effectiveness of dynamic optimization strategies**
  - Effectively balancing passenger loads across different stations not only alleviates congestion at key nodes but also ensures a smoother and more reliable commuting experience.
  - Example: Passenger congestion at key interchange stations such as Shenzhenbei and Minzhi has decreased by 20%.

- **Sensitivity analysis**
  - The adjustability of parameters allows for customization of the model to various subway lines in Shenzhen, and it may also be applicable to other cities facing similar urban transportation obstacles.

SUSTech
Southern University of Science and Technology

THANKS FOR WATCHING!