MGSC 404 – Final Project

Bike Repositioning Optimization

Using Multi-Stop Truck Routing Model

Tiphaine Levan

Marc Pujol

Ernest Bragg

William Fecteau

Laura Pantaleon

McGill

McGill | BIXI

SUMMARY

1. Data Processing
2. Problem Definition & Formulation
3. Model Parameters & Assumptions
4. Model Implementation & Solution Approach
5. Findings
6. Sensitivity Analysis & Managerial Insights
7. Limitations & Conclusion

Data Exploration & Analysis

1. Data Cleaning

Step 1: Merge Q3 and Q4 Dataset

Step 2: Remove Duplicates

Step 3: Check Format and Data Type
- Decomposed 'Date' and 'Time' related variables
- Added 'day_of_the_week' *e.g. 'Monday', 'Tuesday', etc.*

Step 4: Remove Client's associated variables

Step 5: Remove outliers: using Interquartile Ranges

2. Exploration Highlights

- Found high imbalance between station pickups & drop-offs.
- Identified peak usage hours, especially during commuting times.

- Key insight: Imbalance requires targeted repositioning.

Final 'Trips' Dataset (head)

	trip_id	bikeid	tripduration	from_station_id
0	16734065	1411	349	216

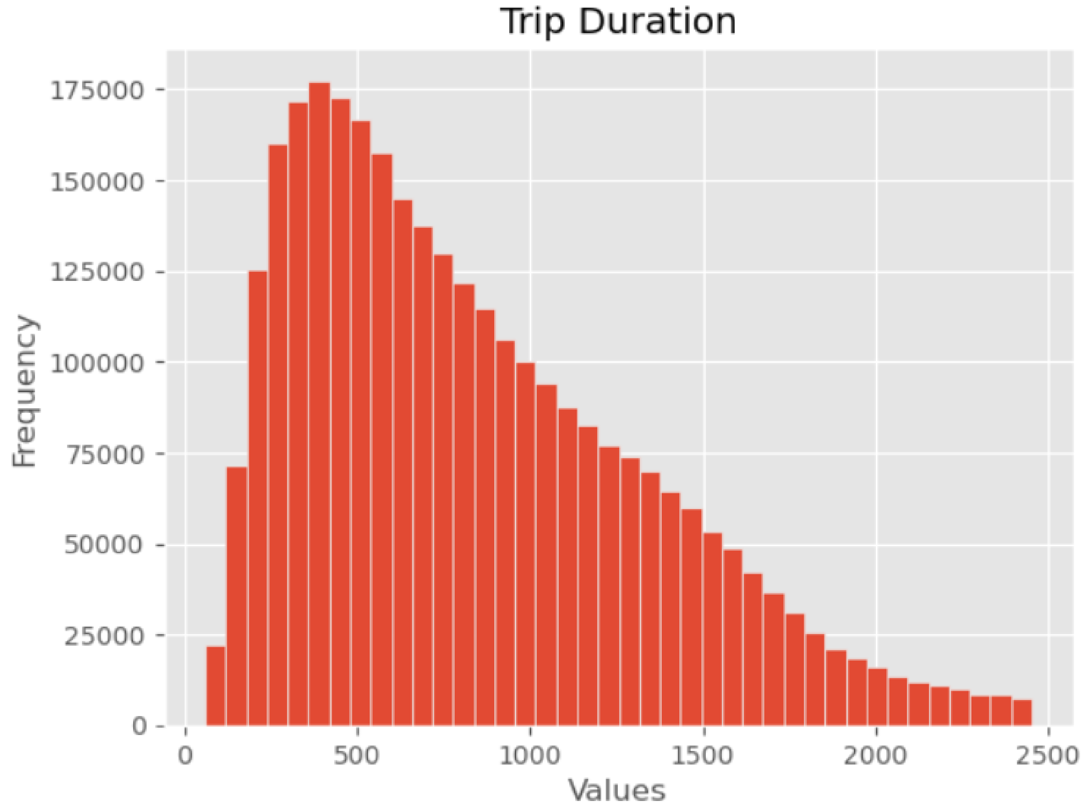
from_station_name	to_station_id	to_station_name	start_date_hour_minute
California Ave & Division St	259	California Ave & Francis Pl	2017-09-30 23:59:58

start_date	start_hour_minute	start_hour	end_date_hour_minute	end_date
2017-09-30	23:59:58	23	2017-10-01 00:05:47	2017-10-01

end_hour_minute	end_hour	day_of_week	day_of_week_num
00:05:47	0	Saturday	5

3053574 rows × 17 columns

1. *Average Bike Journey*



Key Insights:

Highly skewed to the right: trips have relatively short durations, with fewer trips taking longer periods.

Peak trip duration: between 300 and 600 seconds (5 to 10 min)

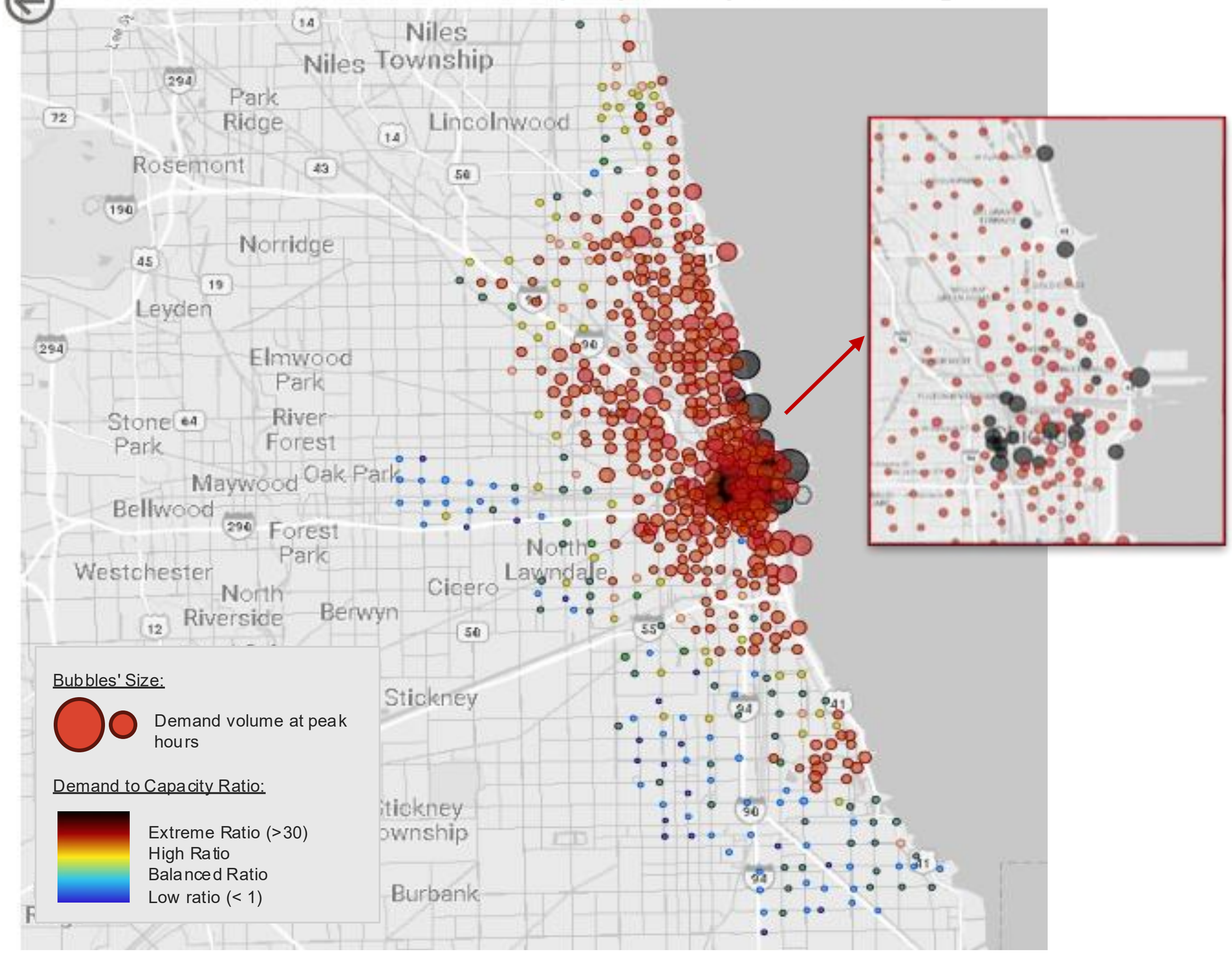
Longer trips: very few lasting beyond 1800 seconds (30 min)

Model Implications:

- Quick commutes or errands between well connected areas.
- Pricing structure might discourage extended use.
- Longer rides might be associated with recreation

1. Busiest and Idle Stations

Peak Hour Demand and Peak Demand to Capacity Ratio by latitude and longitude



Key Insights:

Busiest Stations: Downtown Chicago, with less than 1km radius to each other.

Iddle stations: Suburban areas (North Lawndale) have more docks than necessary.

Station 287: 56.8 ratio, peak demand is 1534

Station 192, 35, 91, 77: prime candidates for the model

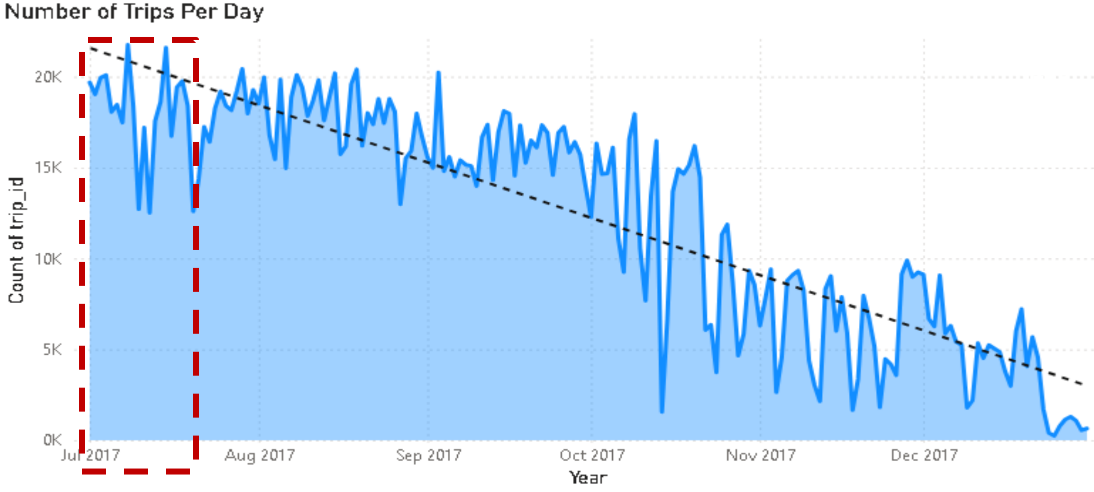
Model Implications:

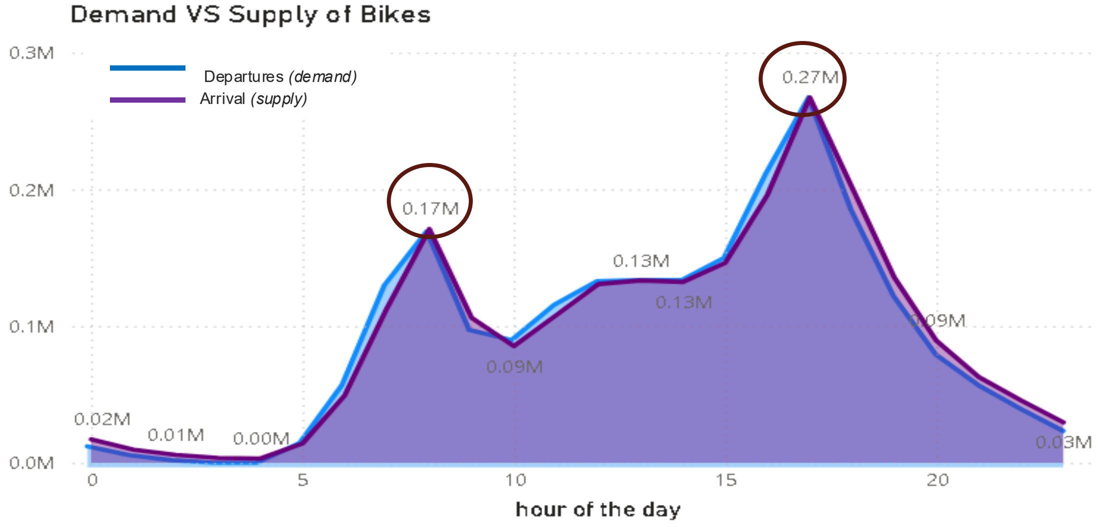
- Bike reallocation is more feasible in downtown core.
- Demand in suburban significantly lower.
- Geographical clustering of high demand.

= **High need for strategic bike repositioning**.

Data Exploration & Analysis

1. Evaluate Period of Activity





Key Insights:

Seasonality: Peak traffic occurs in July

Time of the day: Morning (7am and 8 am), Evening (6pm to 6pm)

Departures Vs. Arrivals: similar trends

Model Implications:

Because of the large data scope, we reduced our analysis to:

- Month of July
- Selecting the following time window:
  - Morning : 7-9 AM
  - Evening: 4-6pm

Data Exploration & Analysis

2. Evaluate Period of Activity

start_hour	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday	Total
0	2066	1395	3260	4070	1474	1340	1391	14996
1	1023	789	2256	2645	718	686	600	8717
2	534	415	1361	1725	369	371	319	5094
3	364	265	767	982	315	202	280	3175
4	498	427	440	658	422	467	459	3371
5	3003	2782	763	827	3226	3385	3229	17215
6	10061	10095	2028	1747	10992	12126	11722	58761
7	23364	23375	4348	3632	23987	26846	24785	130337
8	29166	29766	8662	7288	30677	32436	30981	168976
9	14588	14315	14187	13057	14193	14171	13842	98353
10	11251	11694	19888	18519	10159	10305	9317	91133
11	14995	14878	25110	23201	12910	13380	11802	116276
12	18269	17645	28151	25440	14836	14762	14053	133156
13	18744	17667	28836	25470	14706	14763	13856	134042
14	18901	17415	29289	25254	14283	14876	13981	133999
15	23343	19714	28317	26175	17418	17841	16944	149752
16	33662	32761	26207	25191	30428	31884	30744	210877
17	35792	48044	23445	22138	45329	46038	44155	264941
18	23404	31758	20514	19370	29539	30769	28982	184336
19	15074	20362	14886	14749	18325	20160	19120	122676
20	9862	12616	10332	9881	12053	13558	12391	80693
21	7216	8867	8052	7230	8938	9515	9011	58829
22	5902	5164	6902	5034	6393	6429	5967	41791
23	4482	2856	5685	2893	3803	2998	3275	25992
Total	325564	345055	313686	287176	325493	339308	321206	2257488

Key Insights:

Weekday: Peak traffic concentration at 5pm, lower but high concentration traffic 7-8am: work commutes.

Weekend: Spread traffic with concentration during days hours (10am-4pm): recreational.

Model Implications:

Final model time frame

	Model 1	Model 2
Month	July	July
Day	Monday	Wednesday
Hour	7-9AM	4-6 PM

Problem Definition & Formulation

1. Objective

$$\min \sum_{t \in T} \sum_{u=1}^{U} \left(F\, y_{u,t} + c_d\, D_{u,t} \right) + \sum_{i \in S} p_s\, r_i^s + \sum_{i \in D} p_d\, r_i^d.$$

Model Type: Mixed-Integer Programming (MIIP)

Determine **how many bikes** should be repositioned between stations during peak demand periods to **minimize operational costs** while maintaining service quality.

2. Decision Variables

Binary Variables

Whether a truck visits a station, uses an arc, next station.

- $y_{u,t} \in \{0,1\}$ ▶ Equals 1 if truck *u* is active in time period *t*, and 0 otherwise.
- $x_{u,i,s,t} \in \{0,1\}$ ▶ Equals 1 if truck *u* visits station *i* as its *s*-th stop in period *t*, and 0 otherwise
- $z_{u,i,j,s,t} \in \{0,1\}$ ▶ Equals 1 if truck *u* travels from station *i* to station *j* immediately after stop *s* in period t, for *i =/= j*.

Continuous Variables

- $p_{u,i,s,t} \in \mathbb{Z}_{\geq 0}$ ▶ Number of bikes **picked** up by truck *u* at station *i* during stop *s* in period *t*
- $d_{u,i,s,t} \in \mathbb{Z}_{\geq 0}$ ▶ Number of bikes **dropped** off by truck *u* at station *i* during stop s in period *t*.
- $\ell_{u,s,t} \in \mathbb{Z}_{\geq 0}$ ▶ **Load** (number of bikes) carried by truck u after stop s in period t, with $\ell_{u,s,t} \leq C$
- $D_{u,t} \geq 0$ ▶ **Total distance** traveled by truck *u* in period *t*
- $r_i^s \in \mathbb{Z}_{\geq 0}$ ▶ For **surplus** stations ($i \in S$), represents the number of surplus bikes not moved.
- $r_i^d \in \mathbb{Z}_{\geq 0}$ ▶ For **deficit** stations ($i \in D$), represents the number of deficit bikes not fulfilled.

McGill | BIXI

3. Constraints Formulation

(1) Truck Activation Constraint

$$\sum_{i=1}^{N}\sum_{s=0}^{S-1} x_{u,i,s,t} \le S_{\max}\, y_{u,t}, \quad \forall u,\ t \in T$$

If truck *u* is not activated in time period *t* (*y u,t = 0*), then it cannot visit any station (*x* must be 0). If truck is activated, it can make *Smax* stops.

(2) Load Management Constraint

$$\ell_{u,0,t} = \sum_{i=1}^{N}\left(p_{u,i,0,t} - d_{u,i,0,t}\right), \quad \forall u,\ t \in T,$$

Defines the truck's load after the initial stop (stop 0) = picked up - dropped off.

$$\ell_{u,s,t} = \ell_{u,s-1,t} + \sum_{i=1}^{N}\left(p_{u,i,s,t} - d_{u,i,s,t}\right) \quad s = 1,\ldots,S-1,\ \forall u,\ t \in T,$$

Updates the load for each subsequent stop by adding the net change (pickups minus drop-offs) from the current stop to the load from the previous stop.

$$\ell_{u,s,t} \le C, \quad \forall u,\ s,\ t \in T.$$

Lad carried by the truck never exceeds its capacity *C*.

(3) Visit Operation-Linkage

$$p_{u,i,s,t} \le C\, x_{u,i,s,t}, \quad \forall u,\ i,\ s,\ t \in T,$$
$$d_{u,i,s,t} \le C\, x_{u,i,s,t}, \quad \forall u,\ i,\ s,\ t \in T,$$

If the truck does not visit station *i* at stop *s*, then no bikes can be picked up or dropped off there.

$$\sum_{i=1}^{N} x_{u,i,s,t} \le y_{u,t}, \quad \forall u,\ s,\ t \in T.$$

A truck can only perform an operation at a stop if it is active in that time period.

(4) Route Continuity Constraint

$$\sum_{i=1}^{N} x_{u,i,s,t} \ge \sum_{i=1}^{N} x_{u,i,s+1,t}, \quad \forall u,\ s = 0,\ldots,S-2,\ t \in T,$$

If a truck makes a stop at position *s + 1*, then it must have also made a stop at position *s.*

$$\sum_{\substack{j=1 \\ j\neq i}}^{N} z_{u,i,j,s,t} = x_{u,i,s,t}, \quad \forall u,\ i,\ s = 0,\ldots,S-2,\ t \in T,$$

If the truck visits station *i* at stop *s*, then it must depart from that station to some other station immediately after.

$$\sum_{\substack{i=1 \\ i\neq j}}^{N} z_{u,i,j,s,t} = x_{u,j,s+1,t}, \quad \forall u,\ j,\ s = 0,\ldots,S-2,\ t \in T.$$

If the truck arrives at station *j* as the *(s + 1)-th* stop, it must have come directly from some other station at the previous stop.

3. Parameters & Notation

Indices and Sets

$u \in \{1, \ldots, 20\}$: Trucks

$i, j \in \{1, \ldots, 20\}$: Stations

$s \in \{0, 1, \ldots, S-1\}$: Stop indices along a truck's route (with S being the maximum number of stops per truck).

$t \in T$: Time periods, where

$$T = \{\text{Mon}, \text{Wed}\}.$$

Notation and Parameters

$C = 40$: Truck capacity.

$S_{\max} = 5$: Maximum number of stops per truck.

$F = 50\$$: Fixed cost per truck used.

$c_d = 3\$$: Transportation cost per kilometer.

$D_{\max} = 50km$: Maximum distance a truck can travel in a time period.

$p_s = 50\$$: Surplus penalty (for bikes left over at stations with a surplus).

$p_d = 50\$$: Deficit penalty (for unmet bike demand at stations with a deficit).

d_{ij}: Distance (in km) between station i and station j.

Δ_i: Inventory imbalance at station i, defined as

$$\Delta_i = \text{Target Bikes}_i - \text{Current Bikes}_i.$$

Define:

- $\mathcal{S} = \{i : \Delta_i < 0\}$ (stations with surplus bikes),
- $\mathcal{D} = \{i : \Delta_i > 0\}$ (stations with bike deficits).

4. Model Assumptions

1. Each Stations starts at 50% capacity.
2. Trucks begin at their first assigned station, and not a depot (might underestimate initial travel cost).
3. Operation window limited to x2 two hours.
4. Distance capped at 50km, but truck speed and traffic not modeled.
5. Fixed cost based on public data.
6. Maximum 20 stations exist for each model.

Limitations & Conclusion

1. Simplifying Assumptions

Assumes each station starts at half capacity and trucks begin at their first stop rather than a central depot, which simplifies the model but may underestimate initial travel and inventory dynamics.

2. Lack of Time and Traffic Modeling

The model uses a 50 km distance cap but doesn't incorporate travel time, truck speed, or congestion—factors that could impact route feasibility during peak hours.

3. Omission of Real-World Constraints

Does not account for loading/unloading time, weather, special events, or physical limitations at stations that may restrict truck operations.

4. Data and Scope Limitations

Relies on assumed cost parameters and limits analysis to 20 stations and two time periods. Computational limits required accepting near-optimal solutions

Our findings: Model results

McGill | BIXI

Wednesday (4-6pm)	Monday (4-6pm)
5 trucks used with a total cost of $**582.71**	**4** trucks used with a total cost of $**228.27**

Cost breakdown (Wednesday)

$250 truck cost	**$32.71** transport	**$300** penalties

→ High imbalance, more trucks used, higher deficit penalties.

Cost breakdown (Monday)

200 $ truck cost	$ **28.27** transport

→ More efficient, lower total costs, no unmet demand

Severe demand Imbalance **VS** **Natural demand alignment**

More trucks ≠ enough to meet demand

Highlights need for **aggressive redistribution & prioritization**

Supports **proactive late-night rebalancing**

Transportation cost is key driver → **optimize routing**

Sensitivity Analysis & Managerial Insights

Our sensitivity analysis confirms these insights

MIP model means traditional sensitivity analysis is not applicable because

- *"what-if" testing on **fixed truck cost** & **deficit penalty***
- ***Other parameters had no significant insights***

Managerial Insights

Wednesday

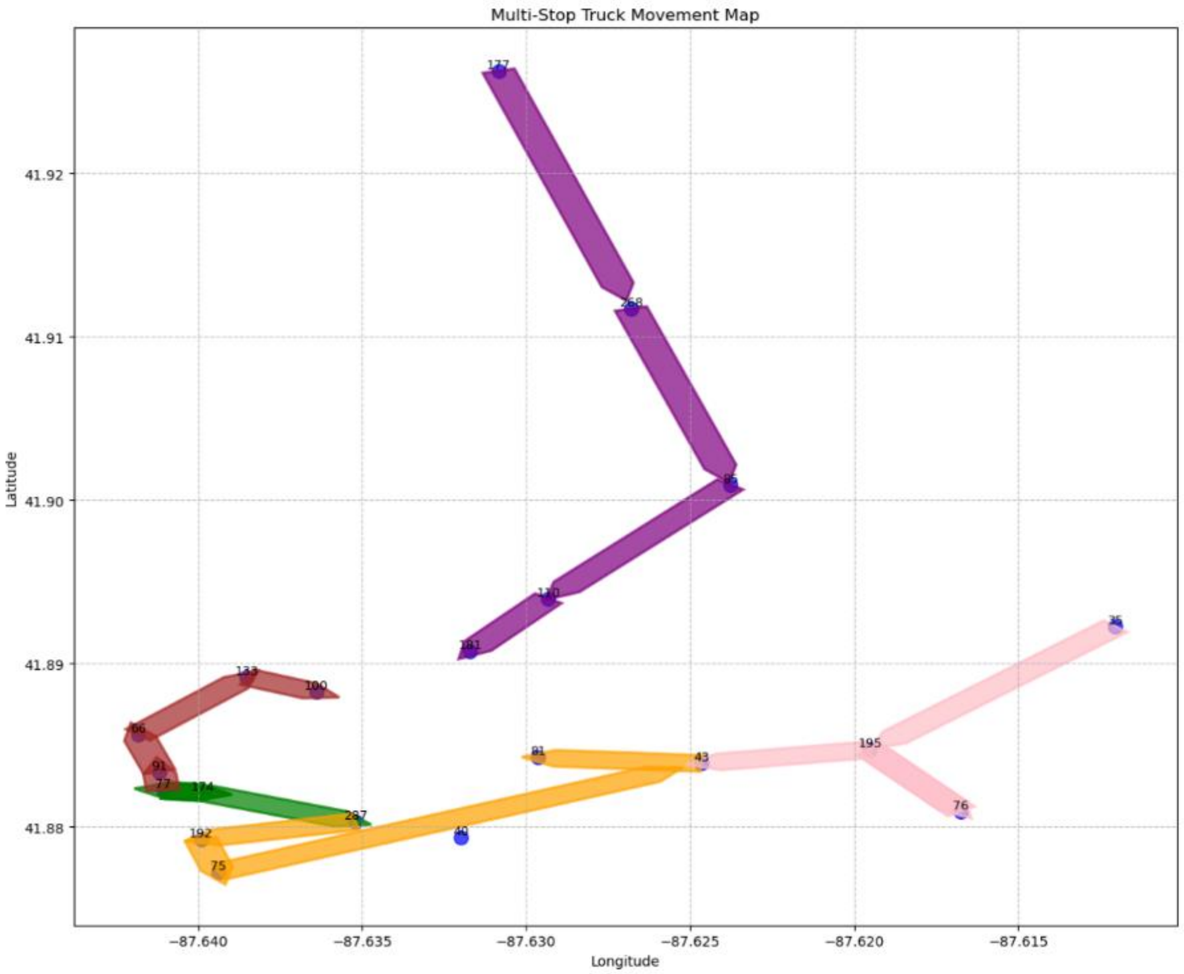
63%
Increase in fixed cost

+ 300$
Penalty

For High-demand periods:

- Use tiered prioritization for core stations (287, 192, 35)
- Explore just-in-time servicing

Mutli-Stops Trucks Movement: Wednesday



Sensitivity Analysis & Managerial Insights

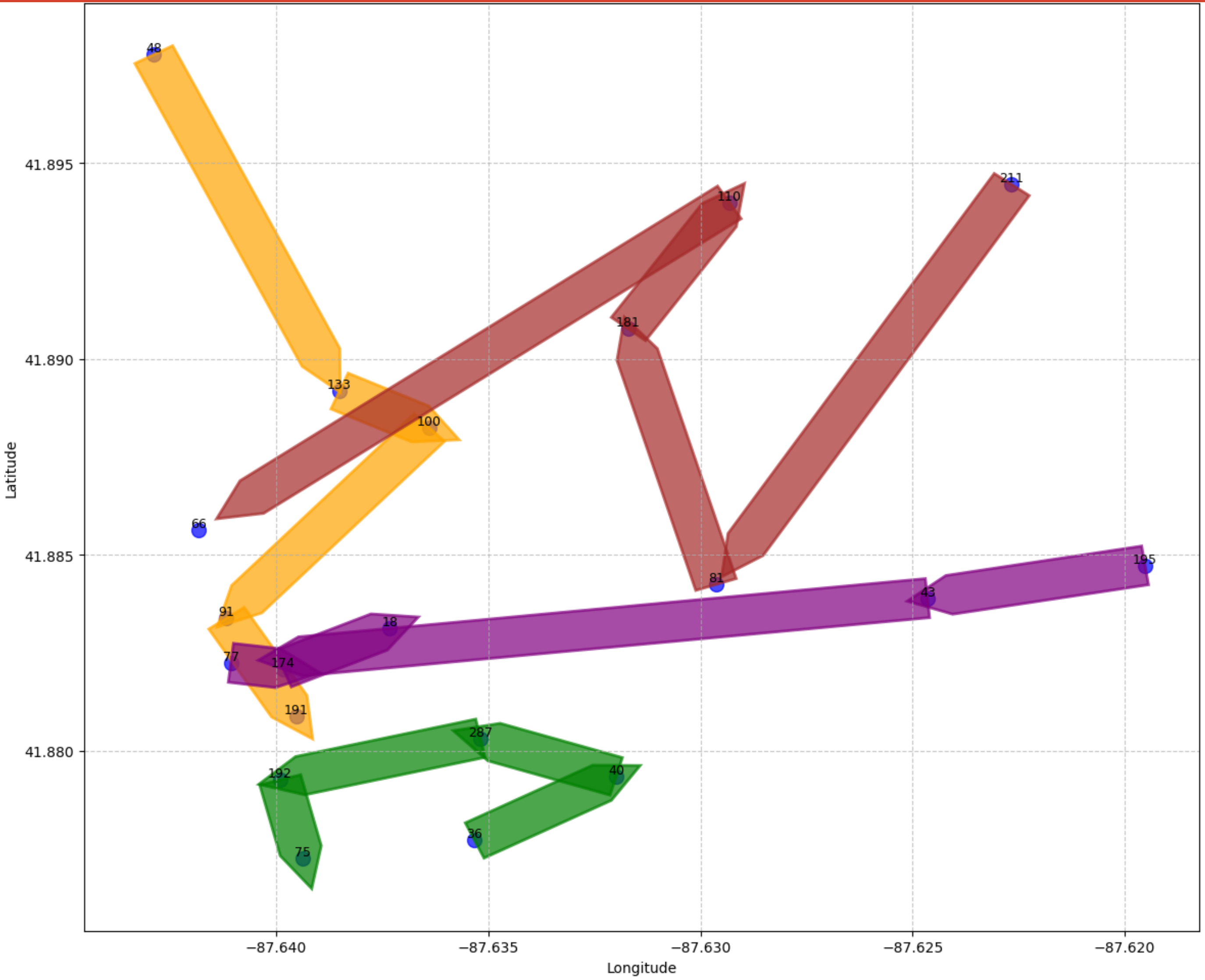
Managerial Insights

Monday

No significant increases in fixed costs or penalties

For low-demand periods:

- ➢ Minimal trucks and late-night rebalancing is cost-effective
- ➢ Suitable for commuter corridors (e.g., Station 91)



Mutli-Stops Trucks Movement: Monday

Takeaways / Conclusions

Deficit Penalties Drive Priorities

On Wednesday, high deficit penalties push the model to avoid stockouts, even if transport costs increase.

Cost Trade-offs Are Contextual

No penalties were incurred on Monday, suggesting that focusing on transport and fixed costs is more impactful.

Policy Use

Municipal bike-share programs can use this model to budget incentives for surplus/deficit management or decide when it's economically justifiable to increase fleet size.

Decision Support Tool

Rather than replacing judgment, the model serves as a decision-support system. Managers can simulate various cost structures (e.g., fuel hikes, labor costs) and quickly visualize operational impacts.

Final Takeaway

The current output serves as an illustrative example of how the model operates. However, the model can be applied to any day of the week and any time window, offering flexibility for broader use.

Thank you!

Q&A