

Government of New York City

School Bus Delay Analytics

Leveraging Data to Uncover Delay Insights and Draft Recommendations to Improve Network Reliability

AN6003 – Analytics Strategy

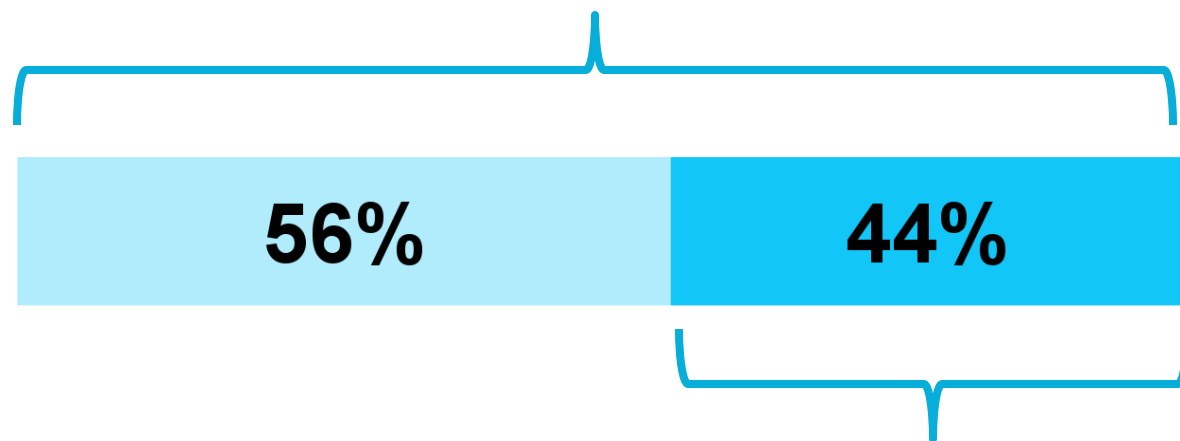
Team 6: Liu Chang, Xu Anlan, Chen Zixuan, Li Ang, Arif Farhan Bukhori, Simon Eppig

New York City's students rely heavily on a reliable and efficient school bus network



150,000

children riding school buses



66,000

children with disabilities or in temporary housing

For many students, school buses are the only way to school.

This makes school buses an essential piece in giving underprivileged children access to education.

Source: Parra, D. (2025). *A New School Year is Here—And So Are School Bus Delays*. CityLimits. <https://citylimits.org/a-new-school-year-is-here-and-so-are-school-bus-delays/>

School bus delays remain a persistent threat to the network's reliability



80,000

Delays in the 2023/24
school year

20%

Increase in bus complaints
compared to last year

“Each year, as the school year begins, [...] our phones will ring off the hook with calls from families whose children can’t get to school due to school bus service failures”

- Randi Levine, policy director of
Advocates for Children of New York

A comprehensive data analytics approach is required to uncover delay causes and develop actionable recommendations.

Source: Parra, D. (2025). *A New School Year is Here—And So Are School Bus Delays*. CityLimits. <https://citylimits.org/a-new-school-year-is-here-and-so-are-school-bus-delays/>

Official city data is used after extensive data cleaning and preparation



“Bus Breakdown and Delays”

2015-2025

School years

Includes each reported breakdown and delay incident on the city's school bus network

~750,000

Records

21

Variables

Non-essential columns removed

Inconsistencies / duplicates removed

Missing/wrong values removed

Formats standardised

~580,000



Records

16

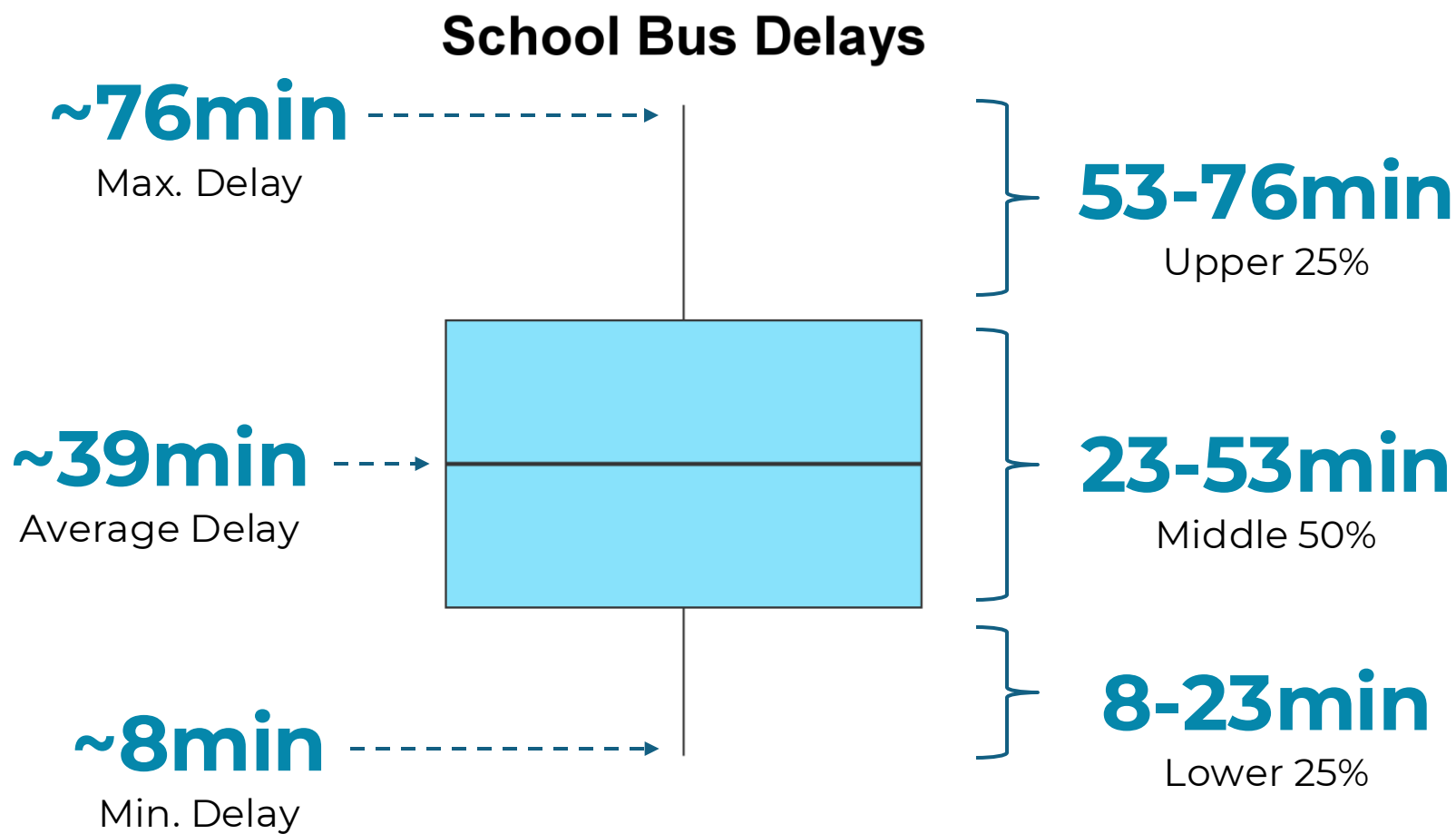
Variables

Data Quality issues persist, requiring targeted measures to ensure consistency



Issues	Examples	Implication	Recommendation
 Errors in manual data input	<div>Y & M TRANSIT CORP (B2192)</div> <div>Y & M TRANSIT CORP (B2192)</div> <div>Y & M TRANSIT CORP. (B2192)</div>	<div>1 Extensive standardisation and data cleaning required</div>	<div>Introduce Input Format Restrictions:</div> <div>1 Only allow single numerical inputs for continuous variables</div>
 Inconsistent Data Formats	<div>15m20-30m1:30h</div> <div>↓ ↓ ↓</div> <div>152590</div>	<div>2 Data loss due to errors that cannot be easily fixed</div>	<div>2 Introduce drop-down menu for categorical variables</div>

Reported incidents lead to an average delay of over half an hour



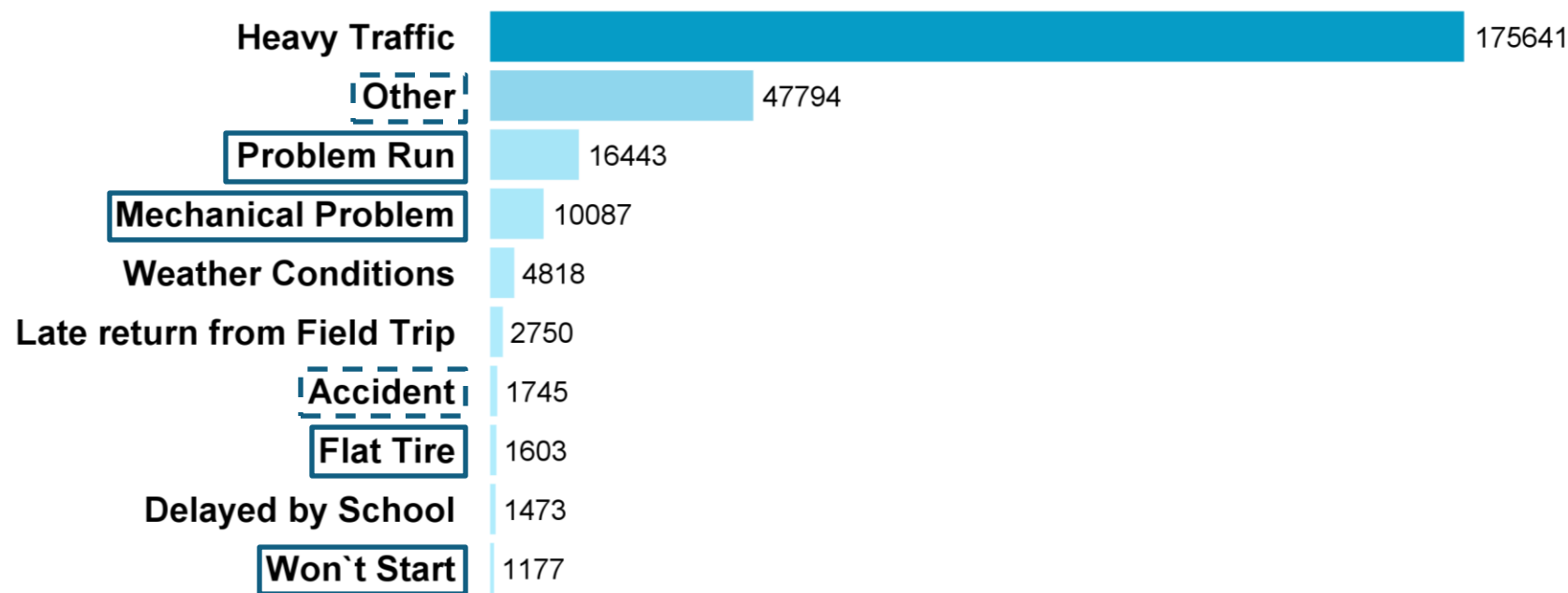
Most reported bus delays are severe enough to lead to students missing classes, with extreme delays of over an hour being not rare.

Delays might lead to educational disadvantages for underprivileged students who are dependent on school buses.

Traffic is the most common delay cause, but bus operators also hold responsibility



Frequency of Delay Causes



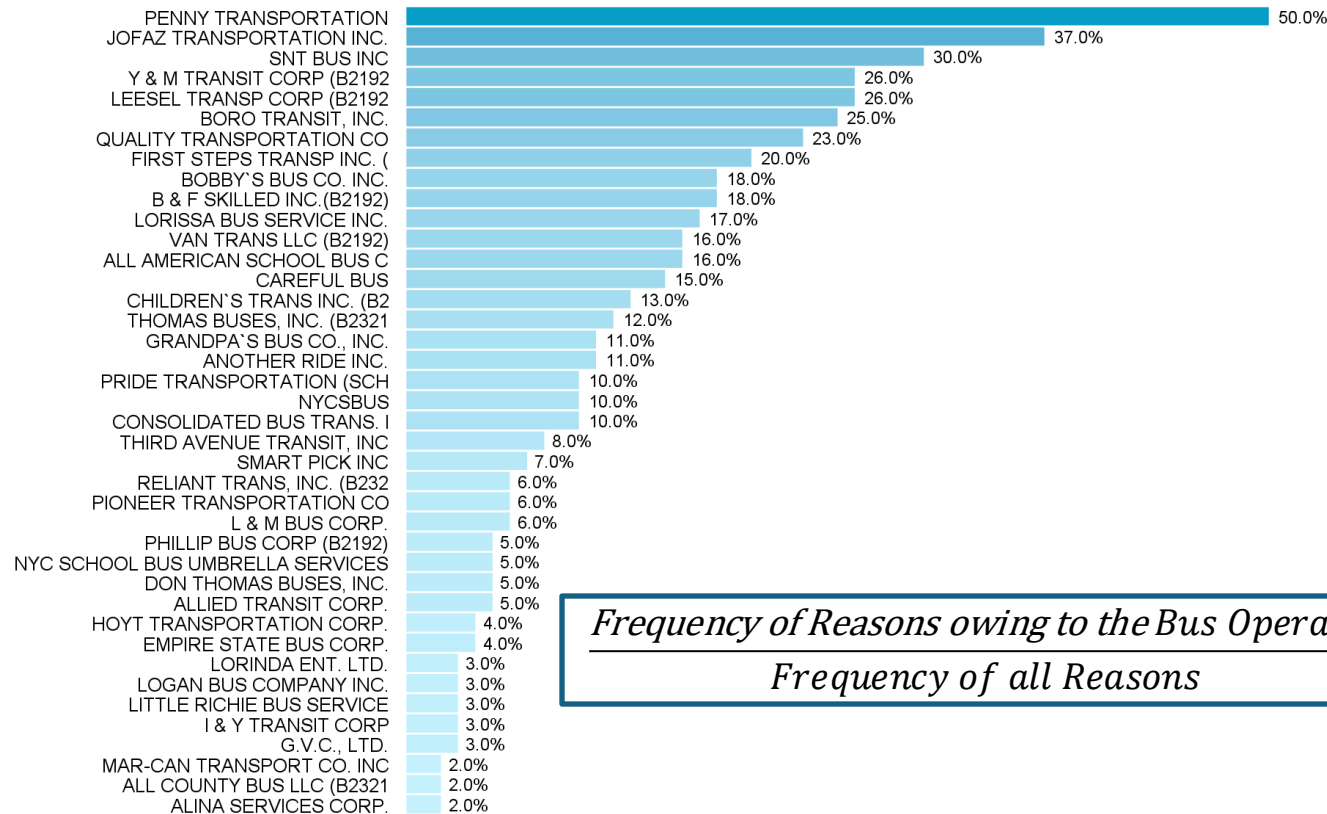
Traffic is the most dominant cause for delays, hinting towards potential for improvement through better bus routing.

Several other top delay causes are fully or partly in the responsibility of bus operators. Further investigation into operator performance is required.

Bus operators differ greatly in Operation Error Index, showing varying performance



Operators by Operation Error Index



Frequency of Reasons owing to the Bus Operator

Frequency of all Reasons

Operation Error Index shows the percentage of delay causes that are within operator control.

The strong differences by operator support the assumption that bus operator are an important factor to reduce delays.

Operator performance alone does not explain delays – deeper analytics required



Potential Delay Factors:



**Bus
Operator**



**Day and
Time**



Bus Route



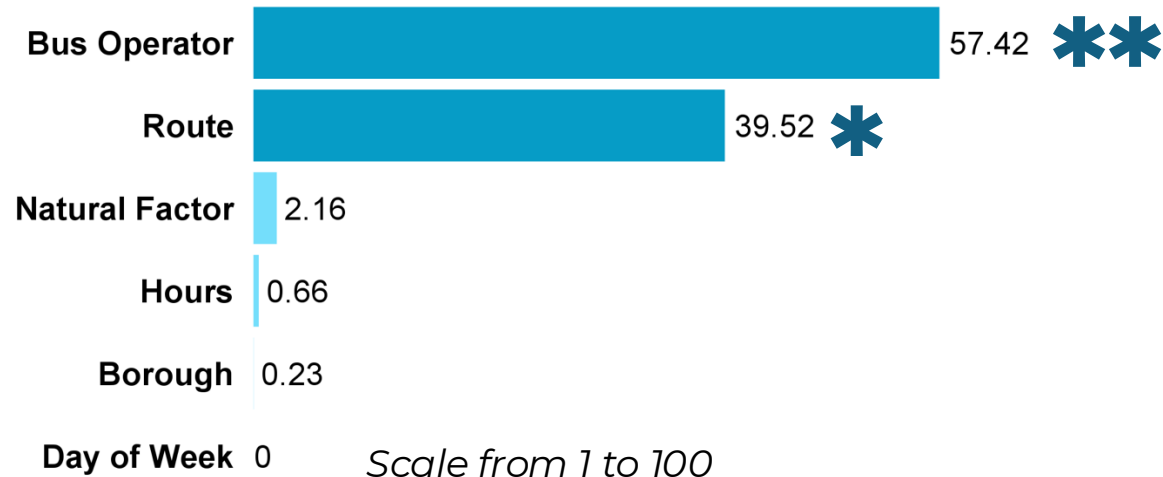
Borough

A variety of factors could – directly or indirectly – cause bus delays. Only more advanced data analytics and statistical modelling can uncover which factors are significant and relevant.

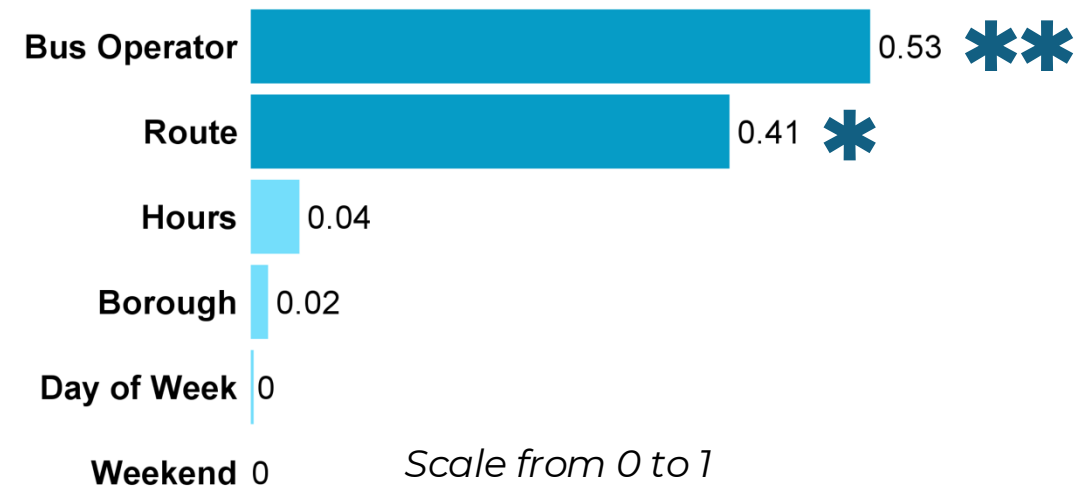
Route Number and Bus Operator are the critical factors that determine delays



CART Variable Importance



Random Forest Feature Importance



* Significance

Y = Delay in Minutes

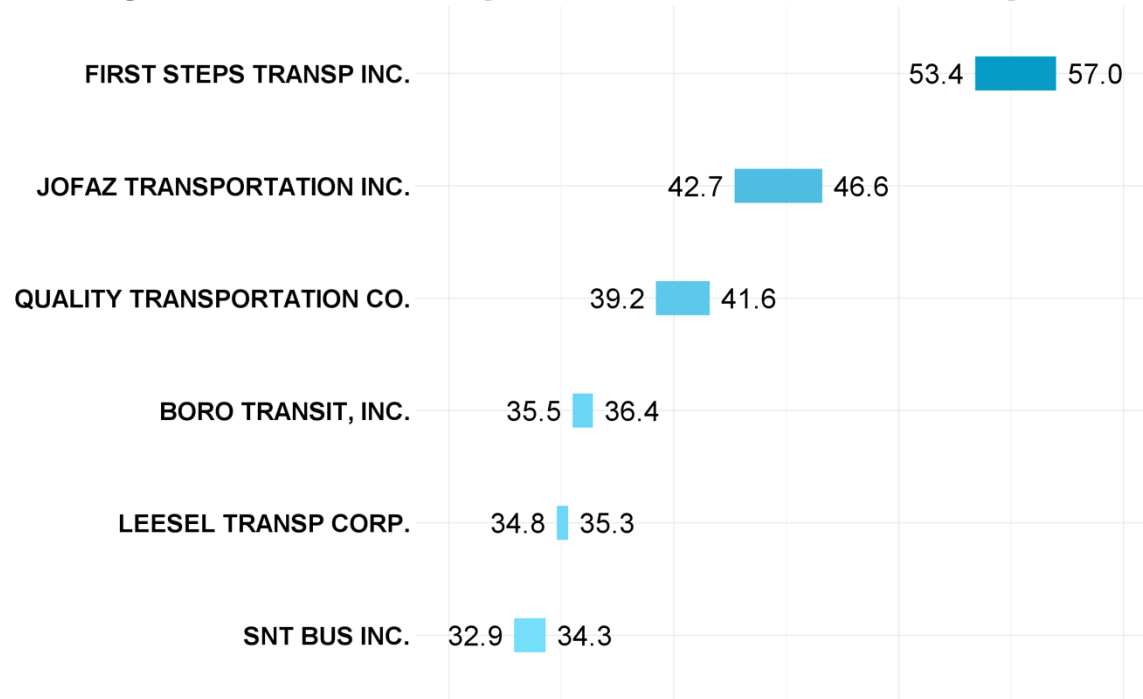
Two statistical models independently confirm that bus operator and route are not only significant in determining delays, but are also the only relevant factors.

For internal issues, we can identify six problematic school bus operators

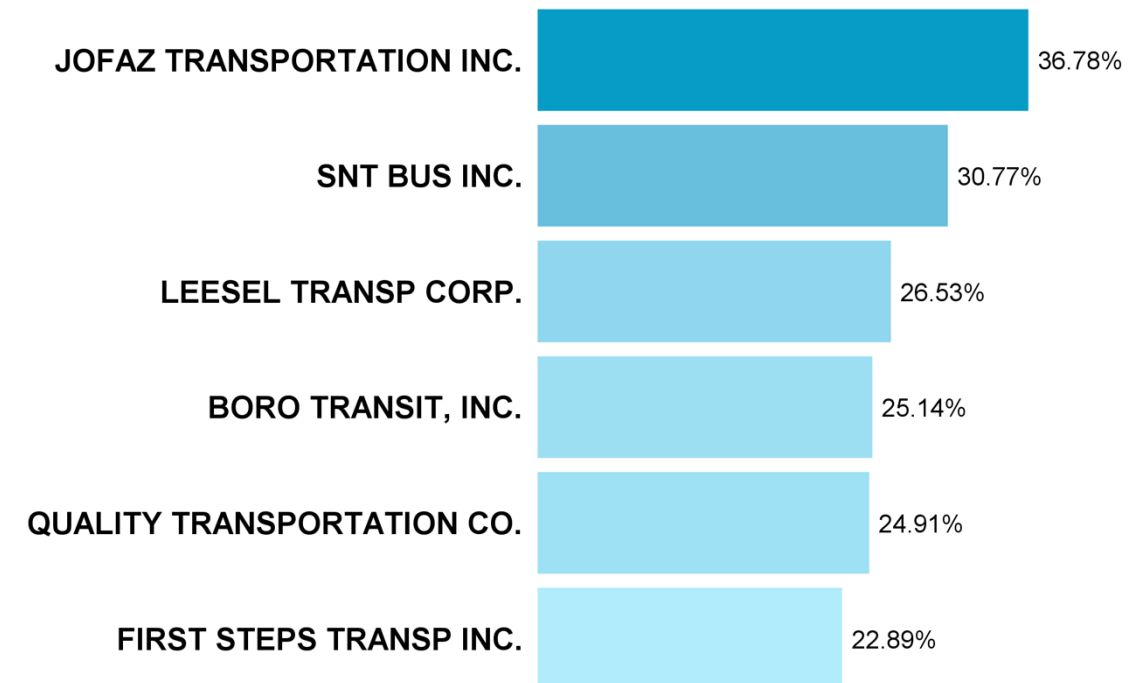


Bootstrap Results:

Delay Interval for Top Six Problematic Companies



Percentage of Operational Issues by Operator



Problematic operators can be grouped into three categories based on delay causes



BORO TRANSIT, INC.

Mean Delay Range: 35.5 – 36.4 mins

Top Failure Types: **Mechanical Problem:** 66.3% – 69.2%

Problem Run: 13.8% – 16.0%

Late Return: 13.5% – 15.6%

LEESEL TRANSP CORP

Mean Delay Range: 34.8 – 35.4 mins

Top Failure Types: **Problem Run:** 95.5% – 96.3%

Mechanical Problem: 2.2% – 2.7%

Flat Tire: 1.1% – 1.6%

QUALITY TRANSPORTATION CO

Mean Delay Range: 39.3 – 41.6 mins

Top Failure Types: **Mechanical Problem:** 41.3% – 49.0%

Late Return: 26.6% – 33.7%

Problem Run: 11.0% – 16.1%

Categories:

Primarily
Mechanical
Problems

FIRST STEPS

JOFAZ

BORO TRANSIT

SNT BUS

Upgrade bus
fleet or replace
with more
modern
operators

Primarily
Operational
Problems

LEESEL

Induce
management
change or
replace with
more efficient
operators

Mixed
Problem
Types

QUALITY TR.

Find individual
problem
causes or
replace with
overall better
operator

Paring four regression models enables us to evaluate bus operators holistically



Linear Regression A (Efficiency): Negative coefficient/lower is better

$$\text{Delay Time} = \beta_0 + \beta_1 \cdot \text{Bus Company Name} + \beta_2 \cdot \text{Boro} + \beta_3 \cdot \text{Reason} + \beta_4 \cdot \text{Run Type} + \varepsilon$$

Linear Regression B (Responsiveness): Negative coefficient/lower is better

$$\text{Time Gap} = \beta_0 + \beta_1 \cdot \text{Bus Company Name} + \beta_2 \cdot \text{Boro} + \beta_3 \cdot \text{Reason} + \beta_4 \cdot \text{Run Type} + \varepsilon$$

Logistics Regression C (Reliability): Severe_Delay OR, OR < 1/lower is better

$$\text{Severe Delay} = \beta_0 + \beta_1 \cdot \text{Bus Company Name} + \beta_2 \cdot \text{Boro} + \beta_3 \cdot \text{Reason} + \beta_4 \cdot \text{Run Type} + \varepsilon$$

Logistics Regression D (Compliance): Notification_Success OR, OR > 1/higher is better

$$\text{Notification Success} = \beta_0 + \beta_1 \cdot \text{Bus Company Name} + \beta_2 \cdot \text{Boro} + \beta_3 \cdot \text{Reason} + \beta_4 \cdot \text{Run Type} + \varepsilon$$

Severe Delay = delay time > 30 min; Notification Success = Both notified School and Parents

Operational Performance Index provides a clearer insight into overall performance



Weighted Factors

Efficiency

40%

Responsiveness

30%

Reliability

20%

Compliance

10%

Z-standardization

$$\text{OPI} = \text{Efficiency} * 40\% + \text{Responsiveness} * 30 + \text{Reliability} * 20\% + \text{Compliance} * 10\%$$

OPI for Problematic Companies



Operators with OPI < 70 will not have their contracts renewed, while other companies will receive warning and offer to collectively work on individual issues.

Certain bus routes show persistent strong delays, requiring immediate action



Bootstrap Results: Impact by external factors (ex.High Frequency on busy morning)

FIRST STEPS	BORO TRANSIT	SNT BUS	JOFAZ	LEESEL	QUALITY TR.
Q863	Y508	M1491	K1304	N063	Q2962
M136	K276	M1127	K1842	N343	Q9885
M166	Q995	M1066	L683	N559	X9766
Q863	M130	M9175	K107	N534	M8200
Q868	K9216	M9062	K9842	N294	Q2987
Q196	K2245	K9110	K132	N062	Q2623

Regression Results: Systematic delay factor (ex.Route design and operation problem)

Some bus operators are responsible for high delays – induce change or replace



Key Insights and Recommendations from Data Analytics:



Delay Causes:

While traffic is the most common delay cause, operator-related causes remain important with bus operator being the most significant variable in determining delays



Operator Differences:

There is a clear set of bus operators which disproportionally contribute to delays and should be targeted through directed measures



Operator Consequences:

Three operators with lowest OPI should not have their contracts renewed, while other three problematic operators need to adopt changes to improve delays

Multiple routes cause problems – adjust bus routing or address traffic issues



Key Insights and Recommendations from Data Analytics:



Delay Causes:

Traffic, the most common delay cause, is directly related to routing and bus route is a significant factor in determining delays



Route Differences:

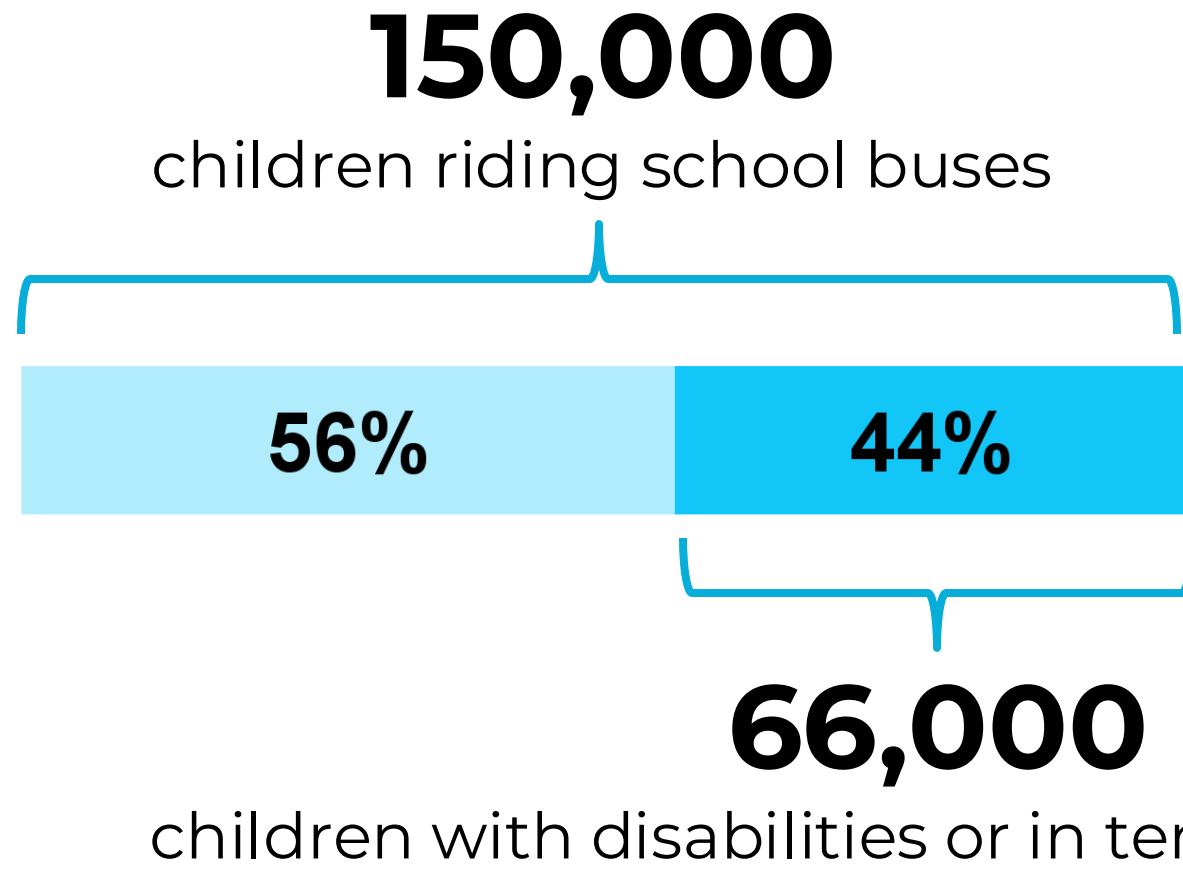
There is a clear set of routes which consistently lead to high delays and require further analysis to identify segments that delay buses through traffic or road conditions



Route Consequences:

The city administration should try to re-route buses to avoid problematic route segments and adjust for more time in school bus schedules where needed

Reducing delays means serving students better, ensuring educational perspectives



Reducing bus delays is not merely about improving business KPIs –

It is about bringing underprivileged students to school and giving them a perspective.

Results might not be immediate, but small steps need to be taken to secure a bright future for New York City's children.

Source: Parra, D. (2025). *A New School Year is Here—And So Are School Bus Delays*. CityLimits. <https://citylimits.org/a-new-school-year-is-here-and-so-are-school-bus-delays/>