



Management

SCHOOL BUS BREAKDOWNS IN NEW YORK CITY

Elizabeth Borseti ¹, Paul D. Berger ²

¹ Bentley University, MSMA Program, Waltham, MA., USA

² Bentley University, Marketing Department, Waltham, MA., USA

Abstract

In this paper, we examine the dataset representing bus breakdowns and delays in the New York public school system. We analyze several measures involving the companies involved in delays, the season/date of the delays, the causes of the delays and other measures. We have several conclusions and recommendations.

Keywords: School Buses; New York City; Delays; Accidents; Parental Notification.

Cite This Article: Elizabeth Borseti, and Paul D. Berger. (2020). “SCHOOL BUS BREAKDOWNS IN NEW YORK CITY.” *International Journal of Research - Granthaalayah*, 8(1), 336-349. <https://doi.org/10.5281/zenodo.3673435>.

1. Introduction

We analyze the data set, Bus Breakdowns and Delays, from the open source data site Kaggle. The data is a collection of information from school bus vendors in the New York City school systems. The bus companies record delays in real-time on their bus route, which are then used by a centralized customer service system to notify parents about delays. The data are were accessed in November 2019.

This data set was chosen primarily to improve the New York OPT. The company website is lacking and OPT’s Google rating is 1.3 out of five stars. Insights from these data could lead to changes that would improve their brand and offer solutions for improvement.

Through analysis, the following questions were addressed:

- Which bus companies are responsible for the highest percentage of breakdowns and delays?
- What type of bus company-caused issues commonly come from the same contractor?
- Which companies are better at notifying parents, and which are worse?
- Does notifying correlate with bus breakdown/delay time?
- Does the *type* of breakdown/delay correlate with breakdown/delay time?
- Is there a seasonality effect for weather-caused breakdowns or delays?

From this analysis and the literature review, recommendations will be made for further analysis and to the bus companies, schools, and OPT as to how to improve operations of their school bus system.

Inadequate school bus functionality is not a new concept to the boroughs of New York. Between 2015 and 2019, the number of breakdowns and delays increased by 73%. In the 2018-2019 school year alone, there were 9,488 breakdowns in New York City (Sanders, 2019). Even though this is such a large problem, and the raw data from the New York City Department of Public Transportation is publicly available, there has not been much publicized research into solving the crisis.

The earliest article on improving the school bus system in New York City is from 1994, where researchers Braca, Bramel, Posner, and Simchi-Levi created a computerized algorithm to project improvements of school bus routes in the city (Braca et al., 1995). While many bus routes use algorithms and trackers in 2019, this early test in 1994 exemplifies that even 25 years ago, there was significant room for improvement in the New York City school bus system.

Braca, Bramel, Posner, and Simchi-Levi's research serves as a precursor to similar, more modern studies. For example, inner-city schools in Boston were experiencing similar breakdowns and delays in the 2010s. Boston spent \$2,000 per student per year on transportation costs and yet were still inconsistent (Grossman, 2019). A team of MIT doctoral students took on this challenge and, in 2017, implemented an algorithm called *Quantum* to improve the bus routes. Testing *Quantum* during the summer of 2017, the school system eliminated 50 buses from the routes and reduced carbon emissions by 20,000 pounds per day, resulting in \$5 million dollars in savings during the 2017-2018 school-year (Grossman, 2019).

Similar findings to the Boston experiment were found in studies in Rome, Italy in 2017 (Brinchi et al, 2017) and New York City in 2016. Although both studies were on public transportation buses, not school buses, the general conclusion was that with sufficient data and analysis, it is possible, especially with regard to route planning, to make improvements to bus service planning capabilities (Hanft et al., 2016).

The data discussed here focus specifically on school bus delays in New York City and boroughs of the tri-state area and the root causes. Although the literature reviewed does not touch on this topic specifically, the studies show different perspectives on similar topics and potential ways to mitigate the problems the New York school buses are experiencing. Ultimately, the goal of this data analysis is to identify the causes of the bus delays and offer potential solutions. The literature above was reviewed because it shares this common goal. Some of these cases, including Sanders' spotlight on New York City school buses causing delays, especially for special-needs students, and Austin's research on reducing school bus emissions to improve students' academic performance and health (Austin et al., 2019), point out the obvious positive implications of improving the system. While others, such as MIT's *Quantum* mapping algorithm for Boston school buses, and the New York City algorithm for mapping public transportation in 1994, offer solutions. This paper will dive more deeply into New York's data more specifically, taking into consideration this prior research when analyzing the data.

2. Materials and Methods

The data set represents all reported breakdowns and delays for buses that transport public school students to and from a New York City-based school, beginning with the 2015-2016 school year through March 2019 (not the complete 2018-2019 school year).

These data are broken up into several different variables, including:

- The date and time of when the breakdown or delay occurred
- The date and time when the breakdown or delay was reported to OPT, the school bus contractor
- The date and time when the breakdown or delay was reported to parents
- The borough where the breakdown or delay occurred (or if it occurred outside New York City, a geographic indicator of the location of the breakdown or delay)
- The name of the bus company that was responsible for that bus
- The length of time it took to resolve the breakdown or delay
- Number of children on board the bus when it broke down or was delayed
- And, the reason given for the breakdown or delay.

Based on an initial look at the data, we can see that there are several clear observations from our data set. First, the number of breakdowns and delays has risen dramatically since 2015-16, as seen in Figure 1. There was a 50% increase in breakdowns and delays from 2015-16 to 2017-18, and as of March 2019, there have already been more breakdowns and delays in 2018-19 than 2017-18.

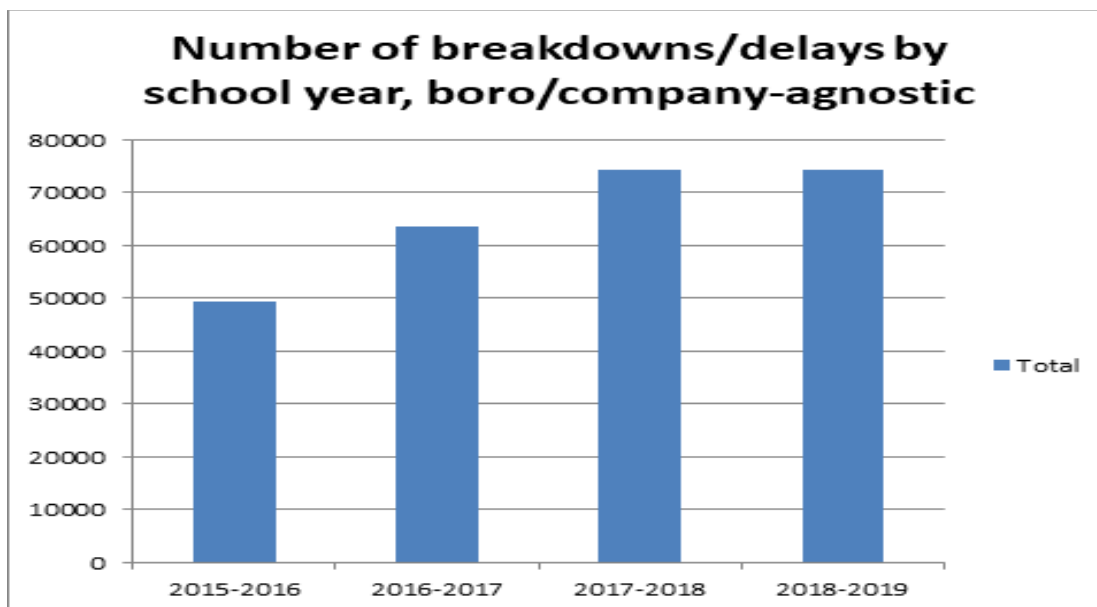


Figure 1

Second, there are significantly more breakdowns and delays in Manhattan, the Bronx, Brooklyn, and Queens out of the ten regions in this data set, seen in Figure 2 below. Combined, there are not as many breakdowns and delays in Westchester, Staten Island, Rockland County, New Jersey, Nassau County and Connecticut as there are in Manhattan.

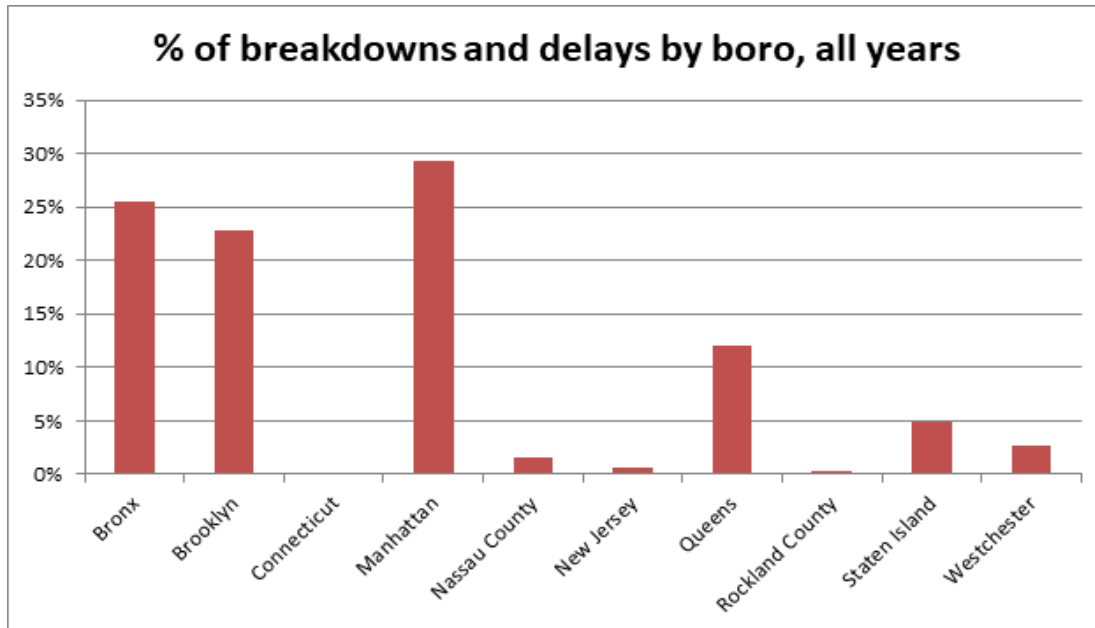


Figure 2

To prepare the data for proper analysis, the data needed to be cleaned. Because the entries were made by different people from different bus companies, all of whom used different methods, many of the data points were formatted differently.

For example, a bus company could be inputted under different names, such as “Montauk Student Trans” and “Montauk Student Trans LLC.” Through cleaning, each company name was standardized to avoid confusion and produce clear results.

A similar problem occurred with the “Length of delay” column, which included the same length of time inputted in different formats (i.e., “1 hour,” “1 hr,” “60,” “60 mins,”). This column was standardized into number of minutes format for all entries under length of delay.

To analyze the data, a mixture of Tableau and Excel was used. In Tableau, analysis was focused on seeking trends and similarities between bus companies, seasons, and reporting errors. Smart columns and maps were created to detect patterns and “top offenders” – companies with the most or longest breakdowns and delays. Using Tableau was necessary here to account for the highly qualitative aspect of the data, and particularly helpful with detecting and displaying these patterns in a way that is relatively difficult to deal with in Excel or SPSS. For example, with Tableau, seasonality and geographical effects were routinely identified.

Excel was then used to analyze data quantitatively and test the effects on the time of delay/breakdown. Correlations, simple regressions, and multiple regressions were performed to test the relationships among number of students on the bus, type of delay/breakdown, notifying schools of delay/breakdown, and notifying OPT of delay/breakdown on the length (in minutes) of the delay/breakdown. The goal was to test if any of these factors correlated with shorter or longer breakdowns. Building off the Tableau analysis, breakdowns/delays because of weather was examined further.

3. Results and Discussion

To best answer the goal questions outlined in the introduction, Tableau was used for a holistic view of the breakdowns and delays. First, the frequency and locations of incidents were mapped (Exhibit 1). While not representative of the total geographical area of each New York City borough, using zip codes in place of “Boro” allowed for a general indication of incident locations. Manhattan, Brooklyn, and the Bronx, in that order, had the most breakdowns and delays.

The highest number of breakdowns and delays by company was then compiled into a top-ten offenders list (Exhibit 2). Leesel Transp Corp (48,701), GVC LTD. (24,618), and Pioneer Transportation Co. (24,208) had the most. These data were then compared with the amount of schools serviced (Exhibit 3), and, interestingly, Leesel Transp Corp and GVC LTD. were not in this top three (although they are in the top ten).

Next, a new top ten was created, emphasizing factors where individual bus companies may be at fault for breakdowns and delays. Exhibit 4 displays accident and mechanical issue data. While it is nearly impossible to identify the root cause of an accident or delay, there is a probability that the bus itself can be a cause of mechanical issues. These issues can often be dangerous, which means safety and brand reputation implications.

As shown in Exhibit 4, the ten companies with the highest accident ratios have ratios between 2.2% and 4.1%. This suggests that for every 100 delay records, an accident is the reason for a delay about 2-4 times. Mechanical problems are much more prevalent, with the top ten companies having reported mechanical problems as the reason for a delay between 13% and 28.1% of the times. These data points also display a much larger range than in the accident data.

Most interestingly, five bus companies, different than the overall breakdown and delay offenders, appear in both of the top-ten lists for accident and mechanical issue ratios. These companies are Boro Transit, Quality Transportation, Consolidated Bus Transportation, Jofaz Transportation, and SNT Bus Inc, highlighted in orange in Exhibit 4.

Further emphasizing SNT Bus Inc., this company holds the second-highest accident ratio and the highest mechanical issue ratio. These numbers are distressing and should be addressed. Through this analysis, it is possible to conclude that there is something intrinsic to these five companies that is causing the statistics to be so disturbing.

Based on this analysis, we decided to look further into other potential harmful practices. This flowed into a new top ten, emphasizing the lowest frequency-reporting individual bus companies that experienced incidents to their clients (i.e., parents and schools). While it is impossible for a company to never experience an incident, it is protocol and ethical that companies report the event to the schools and parents. This would give their clients time to respond accordingly and adjust for the disruption. The failure to do so leads to anger and resentment by the parents (which can be seen on OPT’s Google business page) as well as the schools, therefore leading to significant brand reputation implications.

Exhibit 5 shows the top ten worst reporting companies both by notification of parents and notifications of schools when an incident occurs. For example, in the first column of Pioneer Transportation Company, of all records of incidents for that company, parents were not notified 99.76% of the time. One can argue that this is a shocking revelation.

What is just as shocking is that, again, five specific bus companies appear in the top 10 for both not reporting incidents to parents and schools. These companies are Lorissa Bus Service Inc., Quality Transportation Co., All American School Bus Corp., Safe Coach Inc., and Mar-Can Transport Co. Inc., highlighted in orange in Exhibit 5.

The seasons, and more specifically the weather changes that come with the seasons in New York City, affect the number and length of delays and breakdowns. The original hypothesis was that there would be a rise in the number of breakdowns and delays, as well as length of breakdowns and delays, in the winter months (December-March). This could likely be attributed to snowstorms and snow traffic, as well as cautious driving when it is darker earlier in the afternoon.

First, number of breakdowns and delays were analyzed by month; the three months with the most delays and breakdowns were, in order, January, October, and March (Exhibit 6). Studying length of delays by month (Exhibit 7) indicated that the delays are longer in the winter months. This partially confirms the seasonality hypothesis, given that snow is standard in January and March in New York. February and December are not represented in the top three months. It can be determined that this is likely related to the school calendar, since students have time off in December and February. When exploring these data on a per school day basis, it is clear that December and February have a high number of delays, as well as lengthier delays.

Looking into length of delay further, Excel was used to determine correlations between time in minutes of delays/breakdowns, and potential contributing factors. First, number of students on the bus, notifying schools, and notifying OPT was reviewed. There were no strong correlations among these factors (Exhibit 9), so simple and multiple regressions were run to create a model predicting length of delay in minutes. Although the P-value for each independent variable was low enough to be considered a significant factor, the R-squared of 0.00023 was very low, meaning the model was not useful in predicting length of breakdown/delay (Exhibit 10). The results of the linear regression with number of students on bus as the independent variable and length of delay as the dependent variable also shows that this was not a good predictor (Exhibit 11).

The next hypothesis to test was whether the reason for the delay affects the time of delay. This was a reasonable variable to test considering that bus drivers had ten options to choose from (won't start and no answer were not tested due to lack of answers in these categories). Holding accident constant as an independent variable, delayed by school, flat tire, heavy traffic, late return from field trip, mechanical problem, other, and weather conditions were tested against the dependent variable length of delay/breakdown. Relatively surprisingly, the R-squared also came out very low indicating that this model was not a good predictor of length of delay/breakdown (Exhibit 12).

Circling back to the Tableau testing on seasonality effect, one last test was done on the affect of weather conditions on length of delay. Because the Tableau analysis indicated that winter months had a higher amount of delays/breakdowns, it was assumed weather would increase time of

delay/breakdown. This was not the case, with the test again resulting in too low of an R-squared value to be significant (Exhibit 13).

4. Conclusions and Recommendations

It was unexpected to not find correlations for any factors affecting time of breakdown/ delay, but overall the need for improvement in the New York school bus system was very clear.

One recommendation is to further address the companies that are under-reporting to OPT's clients, schools and parents. These companies are Lorissa Bus Service Inc., Quality Transportation Co., All American School Bus Corp., Safe Coach Inc., and Mar-Can Transport Co. Inc. These companies do not report between 65% and 95% of incidents. There is a concerning gap between the top five and other companies in the data set, who report almost every incident.

Further research is recommended into the subject of why companies and bus drivers are not reporting. Though reporting does not directly correlate with length of delay, it is still alarming to parents who do not know where their kids are, and damages the education of the students who are consistently late. This is clearly a problem, evident in the articles reviewed from *Brookings* and *The Daily News*.

The next recommendation is for the system to address the five companies that appear in both the top ten worst offenders of traffic accidents and mechanical issues. These companies are Boro Transit, Quality Transportation, Consolidated Bus Transportation, Jofaz Transportation, and SNT Bus Inc. As some of these incidents are likely at least partially due to the negligence of the bus company, it is clear that there may be some factors internal to these companies that lead to potential lack of safety awareness and implementation of necessary safety protocols.

In terms of seasonality, there is a spike in incidents during January, February, March and October, as seen in Exhibit 6. It is probable that this spike is due to dangerous winter weather conditions and increased darkness during driving hours, and it is recommended to analyze this issue further with the proper data. Time of incident and time of report is indicated in the data analyzed in this case, but the actual weather conditions, possibly from a different database, could provide more insight.

Overall, the data analyzed did not paint a perfectly clear picture into how to improve the system, but did give significant insight into how to fix certain problems. A solution like MIT's *Quantum* algorithm may help, but further analysis into the bus companies, routes, and potentially the actual bus drivers, would provide useful insight.

Appendix

Incidents by Zip Code

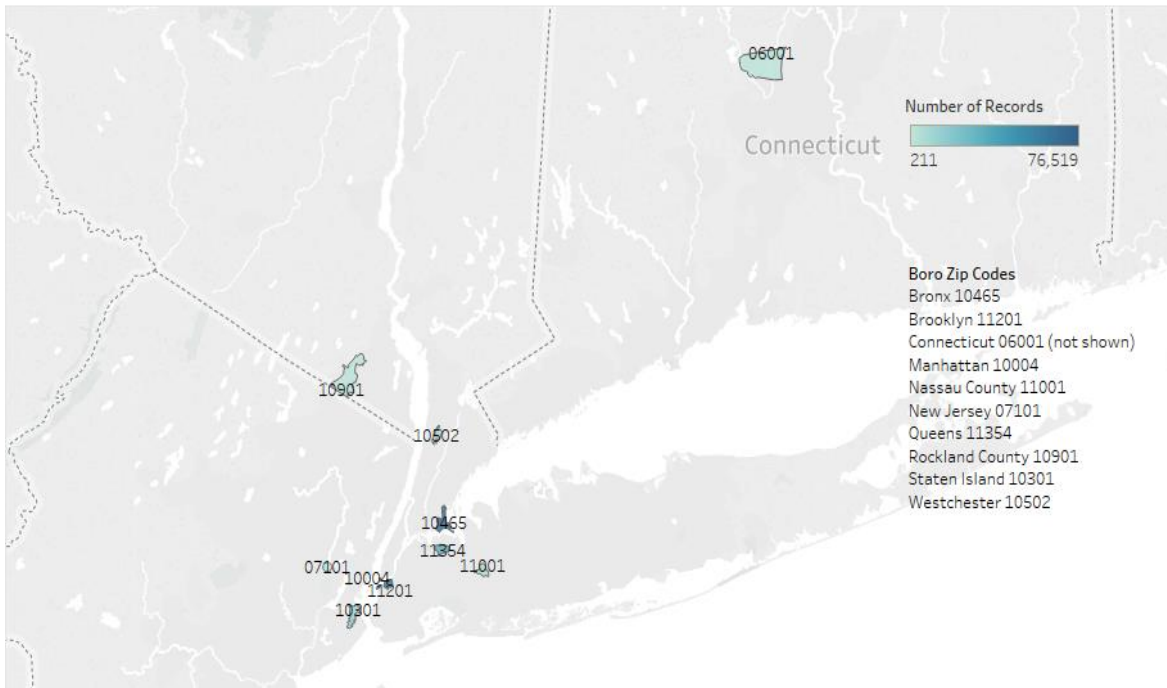


Exhibit 1: Incidents by Zip Code

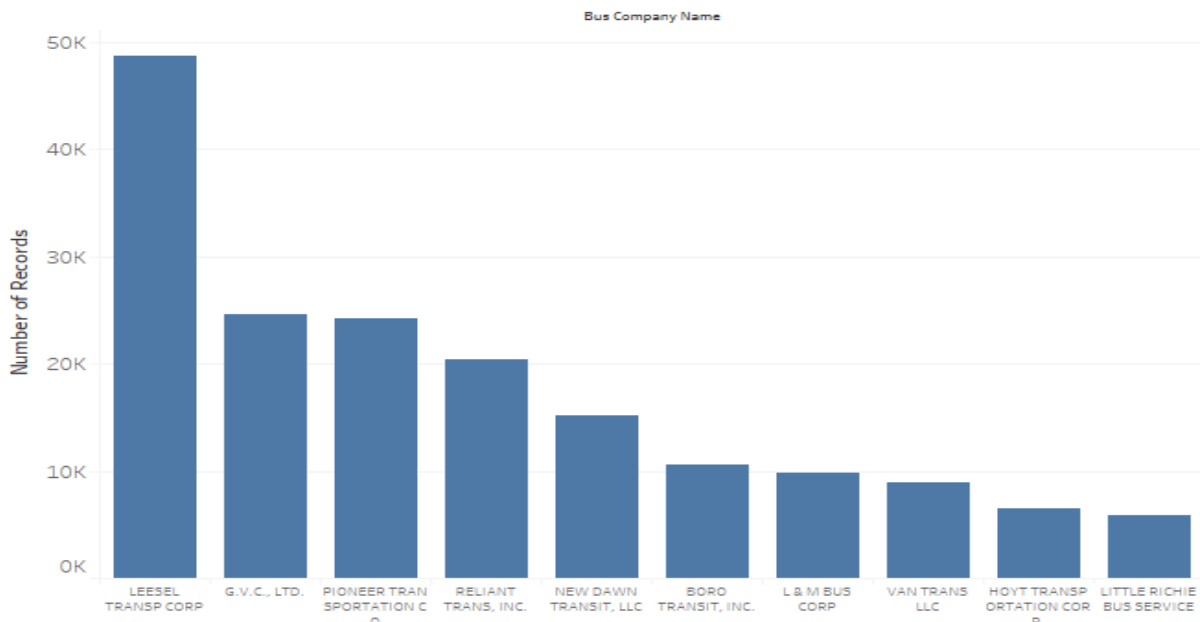


Exhibit 2: Number of Incidents by Company

Number of Schools Serviced by Company

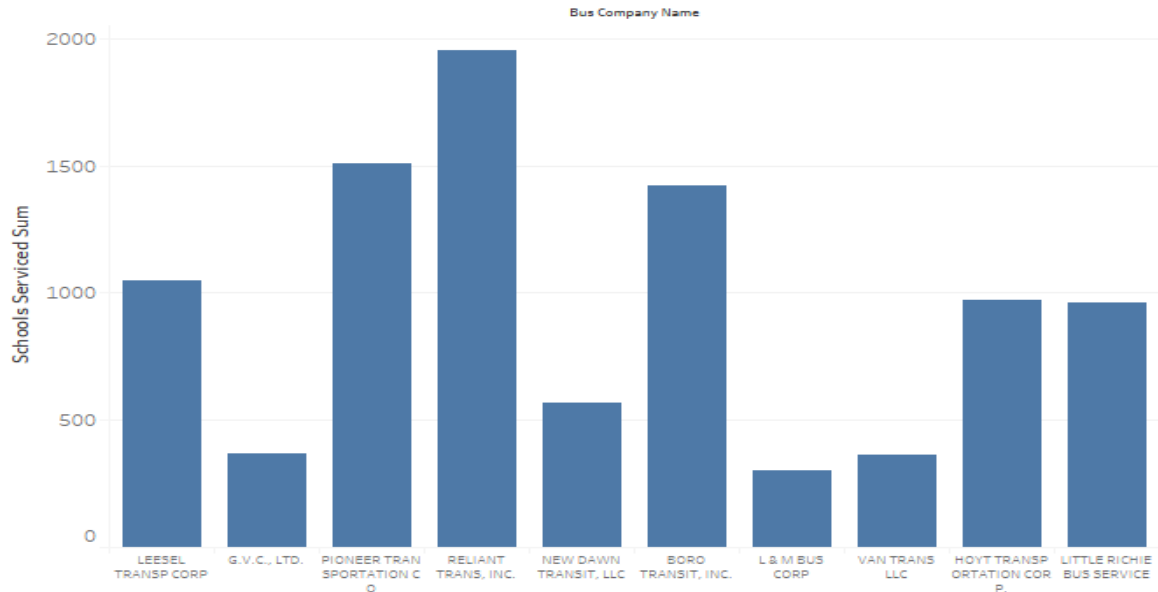
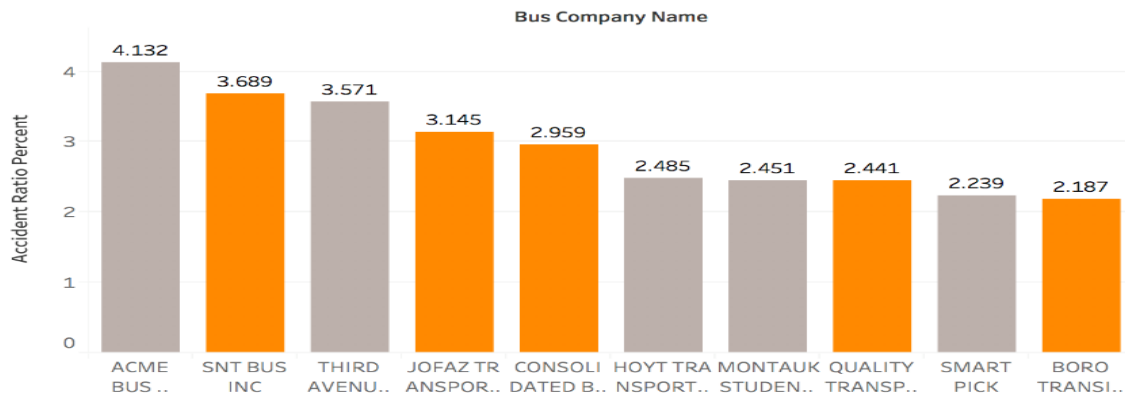


Exhibit 3: Number of Schools Serviced by Company

Top 10 Accident Frequency by Bus Company



Top 10 Mechanical Issue Frequency by Bus Company

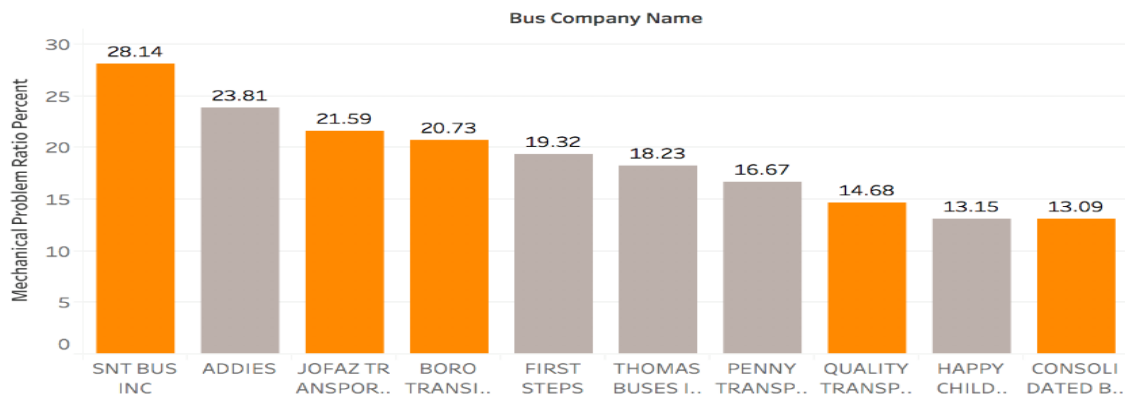
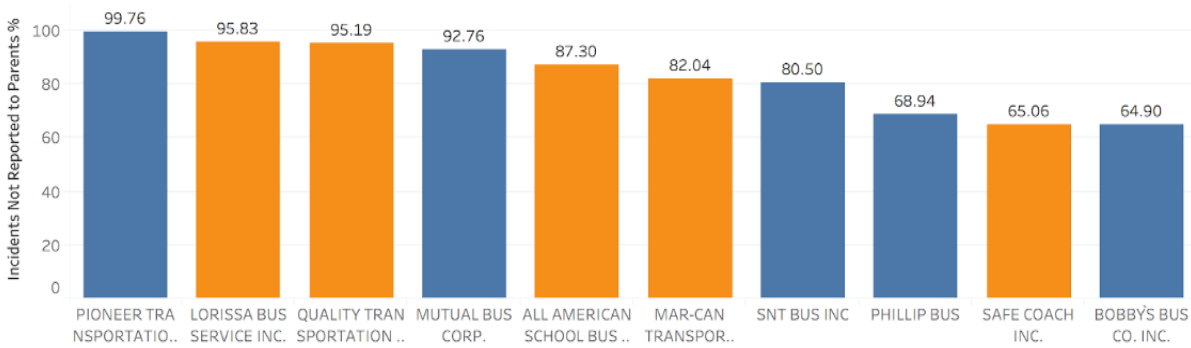


Exhibit 4: Dangerous Incidents by Bus Company - Top 10

Top 10 Has not Notified Parent



Top 10 Has not Notified Schools

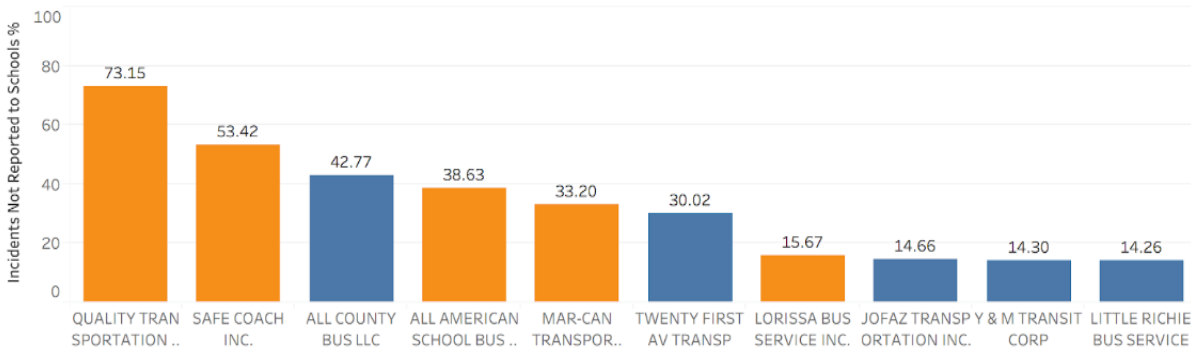


Exhibit 5: Non-Reporting Companies - Top Ten

Number of delays and breakdowns by month (excluding summer months)

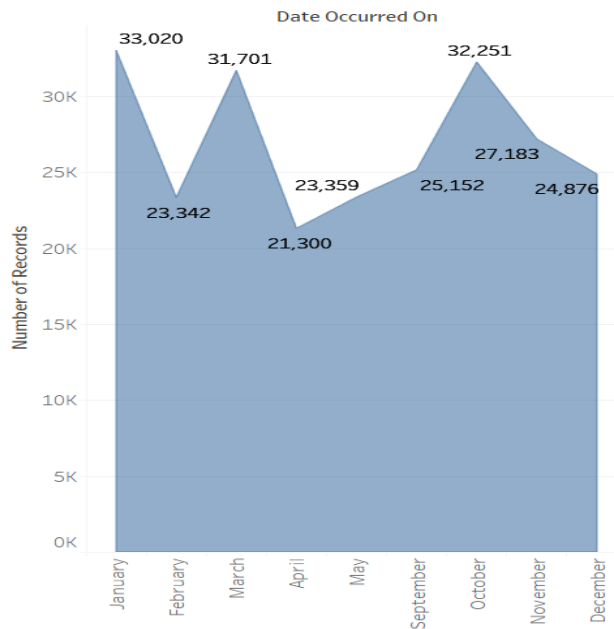


Exhibit 6: Breakdowns and Delays by Month

Average length of delay by month
 (excluding summer months)

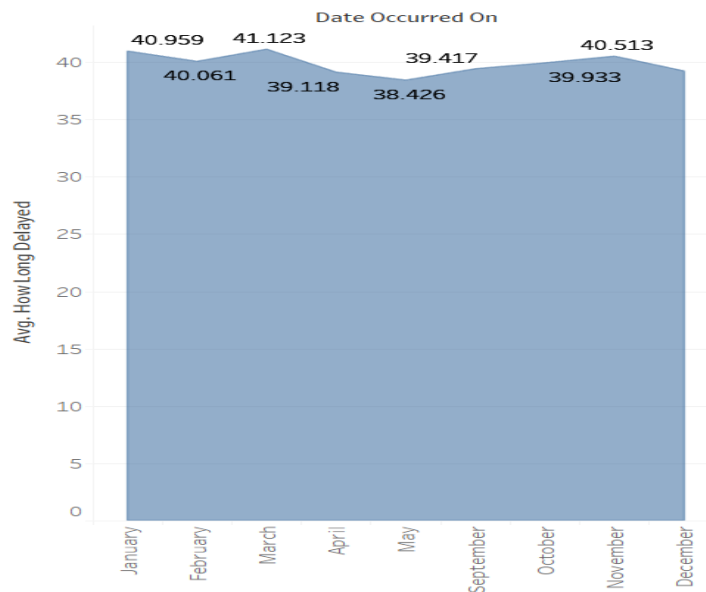


Exhibit 7: Length of Delay by Month

Number of delays in October by week, YOY

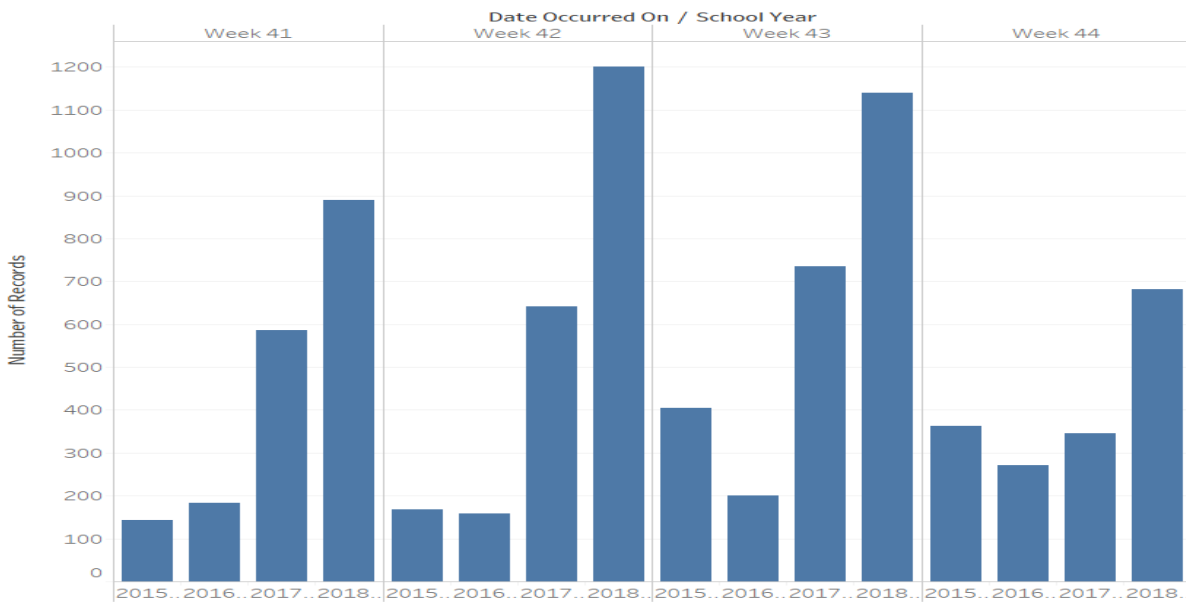


Exhibit 8: Number of Delays in October by Week, Year Over Year

Exhibit 9: correlation between potential factors affecting length of delay/breakdown

CORRELATION			
	<i># of Students on Bus</i>	<i>Notified schools?</i>	<i>notified opt?</i>
<i># of Students on Bus</i>	1		
<i>Notified schools?</i>	0.0017	1	
<i>notified opt?</i>	0.0090	0.4502	1

Exhibit 10: multiple regression on length of delay as dependent variable

Summary Output Multiple Regression								
Regression Statistics								
Multiple R	0.0153							
R Square	0.0002							
Adjusted R Square	0.0002							
Standard Error	29.83							
Observations	261,716							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sign. F</i>			
Regression	3	54282	18094	20.3267	3.64085E-13			
Residual	261712	232965163	890.159					
Total	261715	233019446						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	40.1	0.213	188.2	0	39.6	40.6	39.6	40.6
# of St'ts on Bus	-0.002	0.0009	-2.4	0.016	-0.0039	-0.0004	-0.0039	-0.0004
Notified schools?	-0.494	0.223	-2.219	0.026	-0.931	-0.058	-0.931	-0.058
Notified opt?	1.006	0.137	7.323	2.43E-13	0.736	1.275	0.736	1.275

Exhibit 11: linear regression of number of students with length of delay as dependent variable

Summary Output Students On Bus								
Regression Statistics								
Multiple R	0.0046							
R Square	2.104 E-05							
Adjusted R Square	1.722 E-05							
Standard Error	29.84							
Observations	261,716							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sign. F</i>			
Regression	1	4903.	4903.	5.507	0.0189			
Residual	261,714	233014542	890.34					
Total	261,715	233019446						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	39.842	0.058	682	0	39.728	39.956	39.728	39.956
# of Students on Bus	-0.002	0.001	-2.347	0.019	-0.0038	-0.0003	-0.0038	-0.0003

Exhibit 12: multiple regression testing reason against length of delay/breakdown as dependent variable

Summary Output with Accident As Constant								
Regression Statistics								
Multiple R	0.142							
R Square	0.020							
Adjusted R Square	0.020							
Standard Error	29.536							
Observations	261,716							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sign F</i>			

Regression	10	471099	471099	540.01	0			
Residual	261705	2.28E+08	872.388					
Total	261715	2.33E+08						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	49.650	0.352	140.89	0	48.959	50.341	-130.63	229.93
delayed by school	-19.351	0.753905	-25.67	3.96 E-145	-20.8294	-17.87	-405.03	366.32
flat tire	0.190	0.559	0.340	0.734	-0.906	1.287	-285.96	286.35
heavy traffic	-12.156	0.359	-33.85	1.30 E-250	-12.861	-11.452	-195.88	171.57
late return from field trip	-7.862	0.527	-14.92	2.299E-50	-8.895	-6.830	-277.32	261.60
mechanical problem	-1.553	0.432	-3.5944	0.0003	-2.400	-0.706	-222.62	219.51
other	-6.315	0.385	-16.38	2.67E-60	-7.070	-5.559	-203.50	190.86
problem run	5.017	0.570	8.80	1.31E-18	3.901	6.135	-286.52	296.56
weather conditions	-7.810	0.507	-15.406	1.57E-53	-8.804	-6.817	-267.166	251.545

Exhibit 13: linear regression on weather delay against length of delay/breakdown

Summary Output Weather Conditions								
Regression Statistics								
Multiple R	0.0107							
R Square	0.0001							
Adjusted R Square	0.0001							
Standard Error	29.837							
Observations	261,716							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sign F</i>			
Regression	1	27058.45	27058.	30.394	3.53E-08			
Residual	261714	2.33E+08	890.25					
Total	261715	2.33E+08						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	39.783	0.059	673.51	0	39.668	39.899	39.6679	39.899
weather conditions	2.055	0.373	5.51	3.52E-08	1.324	2.787	1.324	2.787

References

[1] Austin, W., Heutel, G., & Kreisman, D. Fixing school buses is an effective (and cheap) way to improve students’ health and academic performance. *Brookings*. 2019, April 22. Retrieved from <https://www.brookings.edu/blog/brown-center-chalkboard/2019/04/21/fixing-school-buses-is-an-effective-and-cheap-way-to-improve-students-health-and-academic-performance/>

[2] Braca, J., Bramel, J., Posner, B. A computerized approach to the New York City school bus routing problem. *IIE Transactions*. (1997). Retrieved from <https://doi.org/10.1023/A:1018526202990>

[3] Brinchi, S., Comi, A., Nozzolo, A., & Verhini, R. Bus travel time variability: some experimental evidences. *Transportation procedia*, (1997). 27, 101 – 108. <https://www.sciencedirect.com/science/article/pii/S2352146517309699>

[4] City of New York. *NY bus breakdowns and delays*. (2019). Retrieved from <https://www.kaggle.com/new-york-city/ny-bus-breakdown-and-delays>

- [5] Grossman, D. How an algorithm made the buses in Boston better. *Popular mechanics*. (2019, August 13). Retrieved from <https://www.popularmechanics.com/technology/infrastructure/a28689713/algorithm-boston-buses/>
- [6] Hanft, J., Iyer, S., Levine, B., & Reddy, A. Transforming bus service planning using integrated electronic data sources at NYC transit. *Journal of public transportation*, (2016). 19. Retrieved from <https://scholarcommons.usf.edu/jpt/vol19/iss2/6/>
- [7] The New York City Department of Transportation. NYC department of education. (2019). Retrieved from <https://www.schools.nyc.gov/Offices/Transportation/default.html>
- [8] Sanders, A. NYC school bus delays and breakdowns increased 73% in four years, with special education kids most impacted. *Daily News*. (2019, August 25). Retrieved from <https://www.nydailynews.com/news/politics/ny-nyc-school-bus-delays-breakdowns-increase-special-education-student-20190825-zvvx47czbngejlkqw7va2binga-story.html>
- [9] School calendar 2018- 2019. (2019). NYC department of education. Retrieved from <https://www.schools.nyc.gov/docs/default-source/default-document-library/school-calendar-2018-2019>

*Corresponding author.
E-mail address: pberger@ Bentley.edu