Final Project Report

EM-624

Fall 2018

Title: NY City School-Bus Breakdown and Delay

Name: Ekta D Solanki

Guided by: Professor Carlo Lipizzi

Table of Content:

Overview of Research Question

Students throughout New York City rely on a fleet of thousands of school buses to arrive safely at school in the morning, and at home each evening. But when school buses are delayed, break down, or become stuck in traffic, a host of logistical and safety concerns result. Students miss important class time and may be delayed for extensive periods while the source of the delay is resolved.

This data analysis will provide information that is adding up to delays and breakdowns of school bus system and things that needs to be consider for mitigating further delays. Like, due to which reasons , how much average delay time because of each reason, borough where it's happening the most, Bus companies that are most responsible, most reliable vendors, Buses that need to be replaced etc.

Dataset Description:

URL: https://www.kaggle.com/new-york-city/ny-bus-breakdown-and-delays

Source of Data: NYC department of Education

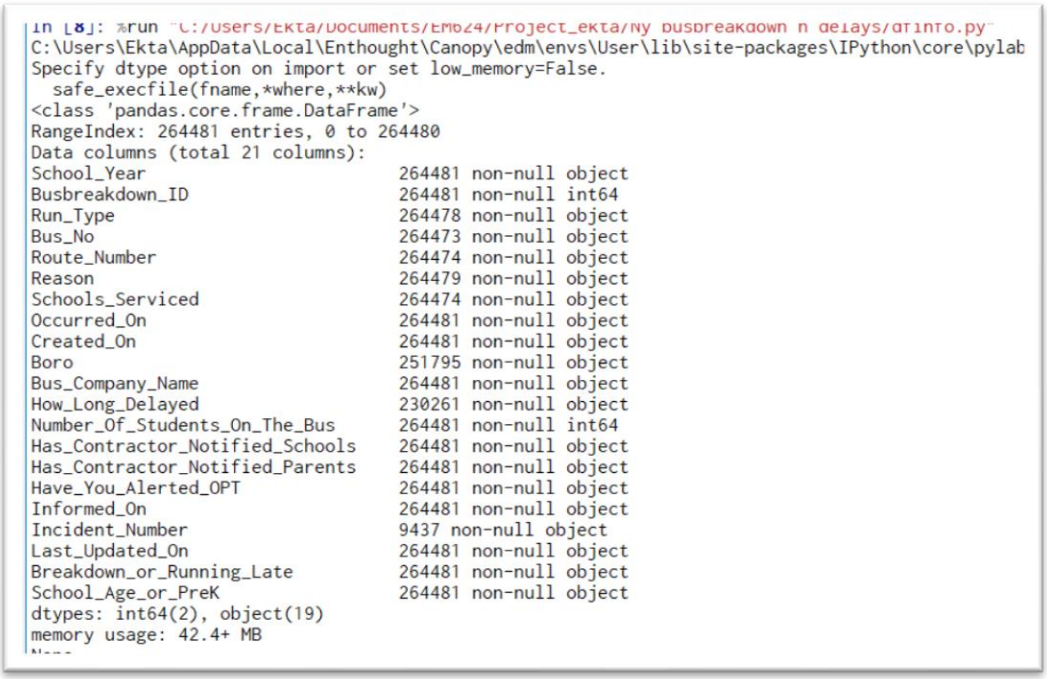
The Department of Education's Office of School Support Services (OSSS) and Office of Pupil Transportation (OPT) are providing public data in cooperation with Department of Information Technology and Telecommunications (DoITT) and the Mayor's Office of Data Analytics Public Law

"NY city Bus Breakdown and Delay" is a dataset created by a system named Bus Breakdown and Delay. This system collects data from school bus vendors operating out in field in real time. Whenever the school bus is running late or break down due to any reason, the bus driver reports it to Office of Pupil Transportation (***OPT***) which is handles by NY Department of education. OPT then uses this system to inform the parents who calls them to know why the school bus is late. All information in the system is entered by school bus vendor staff.

The Office of Pupil Transportation administers school bus service to New York City schools for students attending both public and non-public schools. OPT does not own any school buses or employ any bus drivers; all school bus service is contracted with various school bus companies(Vendors).

The Bus Breakdowns and Delays dataset is supplied to the Open Data website on a daily basis. OPT does not provide these datasets for three months (July, August and September) because: a. most public schools are closed for the summer and the data available represents a small subset of the rest of the school year and b. data during the summer and beginning of the fall term is extremely volatile with frequent changes coming from systems outside of OPT.

Following are the details of the dataset after running df.info() in python.



```
In [8]: %run "C:/Users/Ekta/Documents/EM624/Project_Ekta/Ny Busbreakdown n delays/df1nfo.py"
C:\Users\Ekta\AppData\Local\Enthought\Canopy\edm\envs\User\lib\site-packages\IPython\core\pylab
Specify dtype option on import or set low_memory=False.
  safe_execfile(fname,*where,**kw)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 264481 entries, 0 to 264480
Data columns (total 21 columns):
School_Year                         264481 non-null object
Busbreakdown_ID                     264481 non-null int64
Run_Type                            264478 non-null object
Bus_No                              264473 non-null object
Route_Number                        264474 non-null object
Reason                              264479 non-null object
Schools_Serviced                    264474 non-null object
Occurred_On                         264481 non-null object
Created_On                          264481 non-null object
Boro                                251795 non-null object
Bus_Company_Name                    264481 non-null object
How_Long_Delayed                    230261 non-null object
Number_Of_Students_On_The_Bus       264481 non-null int64
Has_Contractor_Notified_Schools     264481 non-null object
Has_Contractor_Notified_Parents     264481 non-null object
Have_You_Alerted_OPT                264481 non-null object
Informed_On                         264481 non-null object
Incident_Number                     9437 non-null object
Last_Updated_On                     264481 non-null object
Breakdown_or_Running_Late           264481 non-null object
School_Age_or_PreK                  264481 non-null object
dtypes: int64(2), object(19)
memory usage: 42.4+ MB
None
```

As we can see, the dataset consists of 21 Columns, with 264481 Rows.

Almost all the columns have enough data to perform analysis.

Types of Data Items:

Columns:

1. School_Year

Indicates the school year the record refers to. The DOE school year starts in September every year.

2. Busbreakdown_ID

Unique ID of each record.

3. Run Type

Designates whether a breakdown or delay occurred on a specific category of busing service. The categories are: • General Ed AM Run - stop-to-school service in the morning with pick-ups at bus stops and drops-offs at school(s). • General Ed PM Run - stop-to-school service in the afternoon with pick-ups at school(s) and drop-offs at bus stops. • **General Ed Field Trip** - refers to any

mid-day bus trip that does not involve at least one student with an IEP for busing. For OPT, the term "field trip" is a contract term referring to any mid-day trip, including work programs as well as traditional field trips to the museum.

• **Special Ed AM Run** – curb-to-curb service in the morning with pick-ups at residences and drop-offs at school(s). • Special Ed PM Run - curb-to-curb service in the afternoon with pick-ups at school(s) and drop-offs at residences.

• **Special Ed Field Trip** - refers to any mid-day bus trip that includes at least one student with an IEP for busing. For OPT, the term "field trip" is a contract term referring to any mid-day trip, including work programs as well as traditional field trips to the museum.

• **Pre-K/EI** - Pre-Kindergarten or Early Intervention program curb-to-curb run. Pre-K/EI service has different service levels and contract terms from OPT's school aged service. OPT provides curb-to-curb busing to Pre-K students with an Individualized Education Program (IEP) for busing and selected children in the Dept. of Health and Mental Hygiene's Early Intervention program.

• **Project Read PM Run** – curb-to-curb busing service for special education students enrolled in late-day programs. As a rule, bus service is only available to schools or sites that dismiss students before 4:30 PM. There are a small number of approved programs that dismiss students after 4:30 PM.

4. Bus_No

The bus number is assigned by the bus vendor. The numbers are not unique identifiers and may be repeated across vendors. For example, many bus vendors may have a Bus number1. Each bus vendor may have their own numbering convention.

5. Route_Number

This refers to the unique identifier four (1 alpha + 3 numeric) character route numbers indicate curb-to-curb service while five (1 alpha + 4 numeric) indicates stop-to-school service. Pre-K/EI routes may have any value in the Route Number field; bus vendors are not required to follow any naming convention for those routes.

6. Reason

Reason for delay as entered by staff employed by reporting bus vendor. User chooses from the following categories:

• **Accident** - to be selected if any accident that delays the bus is recorded. The bus can be delayed by an accident that the bus itself is not actually involved in.

• **Delayed by School** - to be selected by reporting bus vendors when schools have delays ushering students to or from the bus.

• **Flat Tire** - to be selected by reporting bus vendors for any flat tire that requires the bus to stop and wait for assistance

• **Heavy Traffic** - to be selected by the reporting bus vendor when heavy traffic conditions delay the scheduled service.

• **Mechanical Problem** - to be selected by reporting bus vendors for any type of mechanical problem other than 'Flat Tire' or 'Won't Start'

• **Other** - to be selected by the reporting bus vendor when the delay cannot be classified within the available categories

• **Problem Run** - Some routes may be late to a destination because the bus vendor believes the route has too many stops or is too long. There is a process for bus vendors to submit these Problem Runs to OPT staff for research and resolution. While that work is being done, bus vendors can record routes as being delayed by selecting the Problem Run category.

• **Weather Conditions** - to be selected by the reporting bus vendor when weather conditions delay the scheduled service.

• **Won't Start** - to be selected by the reporting bus vendor when a bus that has already left the bus yard won't start. There is usually a delay while a replacement bus is dispatched to the out-of-service bus's location

7. Schools_Serviced

OPT Codes of all transportation sites on the route. If there is more than one site, each site code will be separated by a comma. If the incident occurred on a bus used for Pre-K/EI service, the code will have one alpha, three numeric and sometimes one additional alpha character. If the incident occurred on a bus used for school-aged service, the code will have five text formatted numerals and may include a leading zero.

8. Occurred_On
Time/date the incident occurred, as entered by the staff employed by the reporting bus vendor

9. Created_On
Time/date the record was created in the OPT Breakdown and Delay system.

10. Boro
Borough, county or state in which the delay occurred, as entered by the staff employed by the reporting bus vendor.

11. Bus_Company_Name
Bus vendor name of the reporting bus vendor.

12. How_Long_Delayed
Length of delay as estimated by the staff employed by the reporting bus vendor. OPT does not systematically monitor the contents of this field in real time.

13. Number_Of_Students_On_The_Bus
Number of students on the bus at the time of the incident as estimated by the staff employed by the reporting bus vendor. OPT does not systematically monitor the contents of this field in real time.

14. Has_Contractor_Notified_Schools
Indicator status as reported by the staff employed by the reporting bus vendor. OPT does not systematically monitor the contents of this field in real time.

15. Has_Contractor_Notified_Parents
Indicator status as reported by the staff employed by the reporting bus vendor. OPT does not systematically monitor the contents of this field in real time.

16. Have_You_Alerted_OPT
Indicator status as reported by the staff employed by the reporting bus vendor. OPT does not systematically monitor the contents of this field in real time.

17. Informed_On

Date on which the school, parents or OPT was notified, as reported by the bus vendor.

18. Incident_Number

Some reports of bus breakdowns or delays originate from calls to the OPT Customer Service line who records incidents. When this happens, the record will have the Incident reference number.

19. Last_Updated_On

Time/date the record was last edited in the OPT Breakdown and Delay system.

20. Breakdown_or_Running_Late

Indicator status as reported by the staff employed by the reporting bus vendor. OPT does not edit this field. Designates whether a bus has broken down (and requires another vehicle to be dispatched to finish the route) or is delayed (and may not require another vehicle).

21. School_Age_or_PreK

Indicator of whether the route serves the school-age or Pre-K/EI population. These two busing types have very different contract terms. OPT does not perform route planning for Pre-K service and doesn't assign route numbers.

Data Preparation:

Some of the data in columns are inconsistent but useful so we cannot omit that. In order to use that data efficiently we need to clean and refine the data.

I have filtered this data in more meaningful way to add the quality to the analysis.

Like **"Bus_Company_Name"** column, It has inconsistency in data of name of company/Vendor. The same company name was written in different ways. I have manually replaced them in the Excel using "Find and Replace" (Can be done in python also but it is much easier in Excel)

1. LEESEL TRANSPORTATION CORP as
   - LEESEL TRANSPORTATION CORP (B2192)
   - LEESEL TRANSP CORP (B2192)
   - LEESEL TRANSP CORP (B2192
2. RELIANT TRANS, INC as
   - RELIANT TRANS, INC. (B232
   - RELIANT TRANS, INC. (B2321)
3. NEW DAWN TRANSIT, LLC as
   - NEW DAWN TRANSIT, LLC (B2
   - NEW DAWN TRANSIT, LLC (B2321)
4. G.V.C., LTD. as
   - G.V.C. LTD. (B2192)
5. PIONEER TRANSPORTATION CO as
   - PIONEER TRANSPORTATION CORP and many more.

"How_Long_Delayed" Column data is alphanumeric, in minutes and hour unit, to run the analysis is crucial to have a single unit of all the data. In this column, the units are used as Min/MIN/min/MINUTE/Hour/Hr/hr . So, I replaced 1 Hour with 60 minutes manually in excel using Find and Relace (Can be done in python also but it is much easier in Excel).

Also, I made a new Column named **“Filtered_Time”** in Excel using,

{=TEXTJOIN("",TRUE,IFERROR(MID(A1,ROW(INDIRECT("1:100")),2)+0,""))}

This column has only number as minutes, so we can run the mean() on the column.

Few of the columns are unrequired due to the kind of data it contains,

Like Breakdown_ID, Incident_Number (which is a unique id of each record),Route_Number (Inconsistent and unuseful for this analysis), School_Serviced (Provide no meaningfulness to the analysis), Occurred_On and Created_On(It gives date and time of breakdown and delay, which is not that necessary because we have Run_Type column which will give us information about time), Last_Updated_On.

Result:

Total number of buses got delayed or breakdown from school year 2015 to 2018

```
Total 63185 buses faced delay/breakdown in school year 2015-16
Total 83141 buses faced delay/breakdown in school year 2016-17
Total 89433 buses faced delay/breakdown in school year 2017-18
Total 28722 buses faced delay/breakdown in school year 2018-19
```

It is also important to know what the reasons are for the schools bus delays.
Top reasons for the delay/Breakdown
Reasons due to which school bus got delayed are as per shown in pie chart below.
We can see maximum times the delay occurred is due to Heavy Traffic.

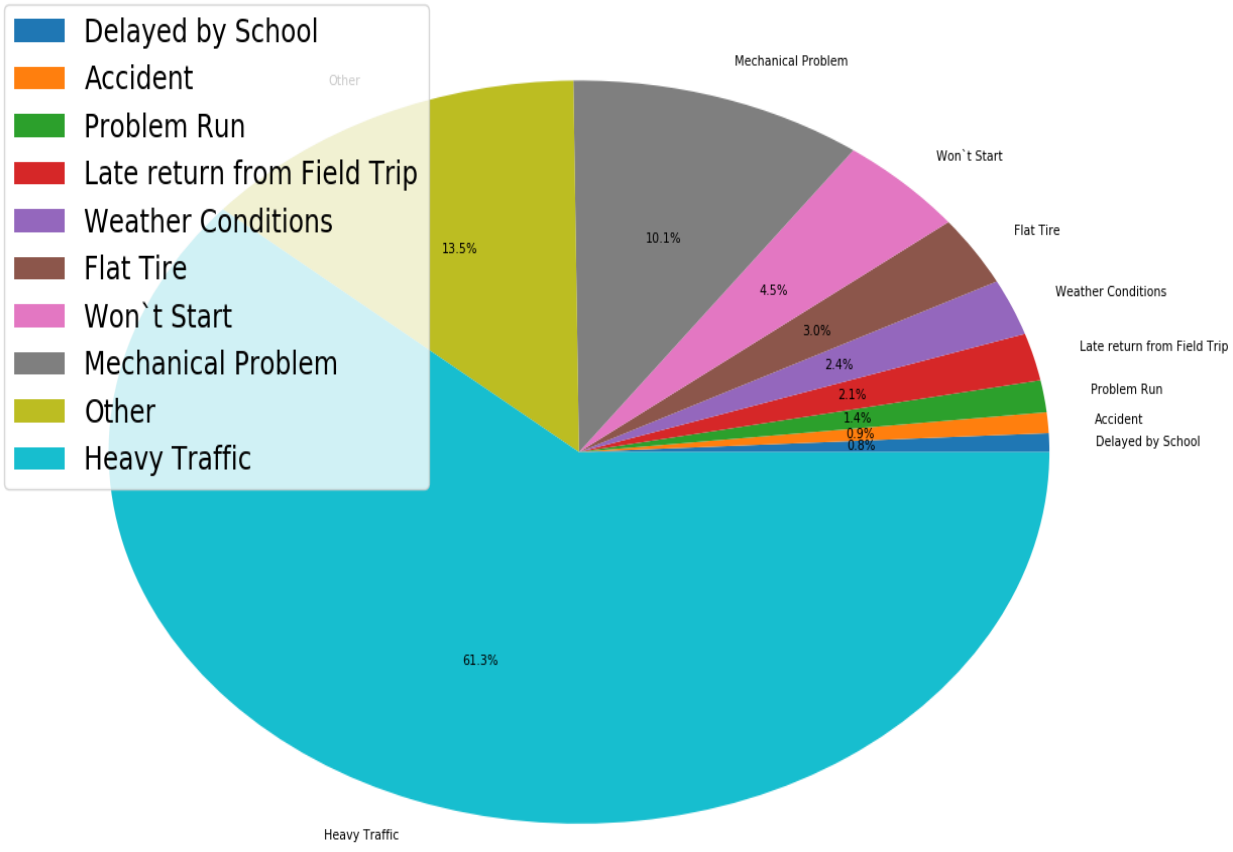


Figure: Percentage of Reason for delay

62.3% of delays are due to heavy traffic. Followed by 13.5% due to other reasons (bus driver choses “other” when the reason is none of above), and 3rd one is mechanical problem.

Taking in to consideration the mechanical problem, let’s see the top bus companies who’s buses are facing the most delays and break downs.

	||BUS COMPANY||	||NUMBER OF INCIDENTS||
1	LEESEL TRANSPORTATION CORP	33795
2	RELIANT TRANS, INC.	25847
3	G.V.C., LTD.	23608
4	PIONEER TRANSPORTATION CORP	23011
5	NEW DAWN TRANSIT, LLC	14737
6	BORO TRANSIT, INC.	11165
7	LITTLE RICHIE BUS SERVICE	10310
8	VAN TRANS LLC (B2192)	8207
9	LOGAN BUS COMPANY INC.	6684
10	SNT BUS INC	6624

And the most reliable companies to have least number of delays and breakdown are,

	||BUS COMPANY||	||NUMBER OF INCIDENTS||
1	gvc	5
2	SELBY TRANSPORTATION CORP (B2192)	3
3	R & C TRANSIT, INC	3
4	DON THOMAS BUSES	2
5	L&M Bus Corp.	1
6	SMART PICK INC	1
7	alina	1
8	FORTUNA BUS COMPANY	1
9	phillip bus service	1
10	1967	1

.

Lets take a look at average breakdown and delay time caused by each reason,

Reason	Breakdown_or_Running_Late	Time taken
Accident	Breakdown	NaN
	Running Late	38.489535
Delayed by School	Breakdown	NaN
	Running Late	22.547334
Flat Tire	Breakdown	31.666667
	Running Late	33.787477
Heavy Traffic	Breakdown	NaN
	Running Late	25.745570
Late return from Field Trip	Breakdown	NaN
	Running Late	29.533383
Mechanical Problem	Breakdown	32.857143
	Running Late	34.234117
Other	Breakdown	30.000000
	Running Late	29.760972
Problem Run	Breakdown	NaN
	Running Late	31.497854
Weather Conditions	Breakdown	NaN
	Running Late	30.791188
Won`t Start	Breakdown	NaN
	Running Late	33.138956

We can notice that Mechanical problem, Flat tire and Other are the only Reasons to have breakdown time.

We can see that the maximum average delay time is 34.87 minute which is also due to “Heavy Traffic”. More regulation of traffic is required on School Bus routes and certain areas.

Area where most of the time bus delay or breakdown happen are,

	||BOROUGH||	||INCIDENTS||
1	Bronx	67084
2	Manhattan	62119
3	Brooklyn	58876
4	Queens	36444
5	Staten Island	13614
6	Westchester	6928
7	Nassau County	3768
8	New Jersey	1511
9	Rockland County	850
10	All Boroughs	403
11	Connecticut	198

Bronx, Manhattan, Brooklyn and queens are the most prominent area where delays happen. So, combining the reasons and area where most delay and breakdown happen needs to be analyzed. Like methods to curb the traffic situation in these particular areas.

Also knowing during which kind of run-type the delays and breakdown are happening can help us manage the issue better. Below is the run-type and incident number table

	Run_Type
Special Ed AM Run	152786
Special Ed PM Run	34711
General Ed AM Run	33464
Pre-K/EI	33199
General Ed PM Run	7982
Special Ed Field Trip	940
General Ed Field Trip	866
Project Read PM Run	443
Project Read AM Run	85
Project Read Field Trip	2

Most number of delays are during special Ed AM Run, which is morning pick up at resident and drop-off at school. So, most delays are at morning 6 AM.

Bus_Number column can give us idea about which buses are causing delays and breakdown more frequently. By knowing this we can get the idea of which buses needs to be replaced and which bus needs a proper servicing.

Below is the Graph of Bus number vs Number of incidents.

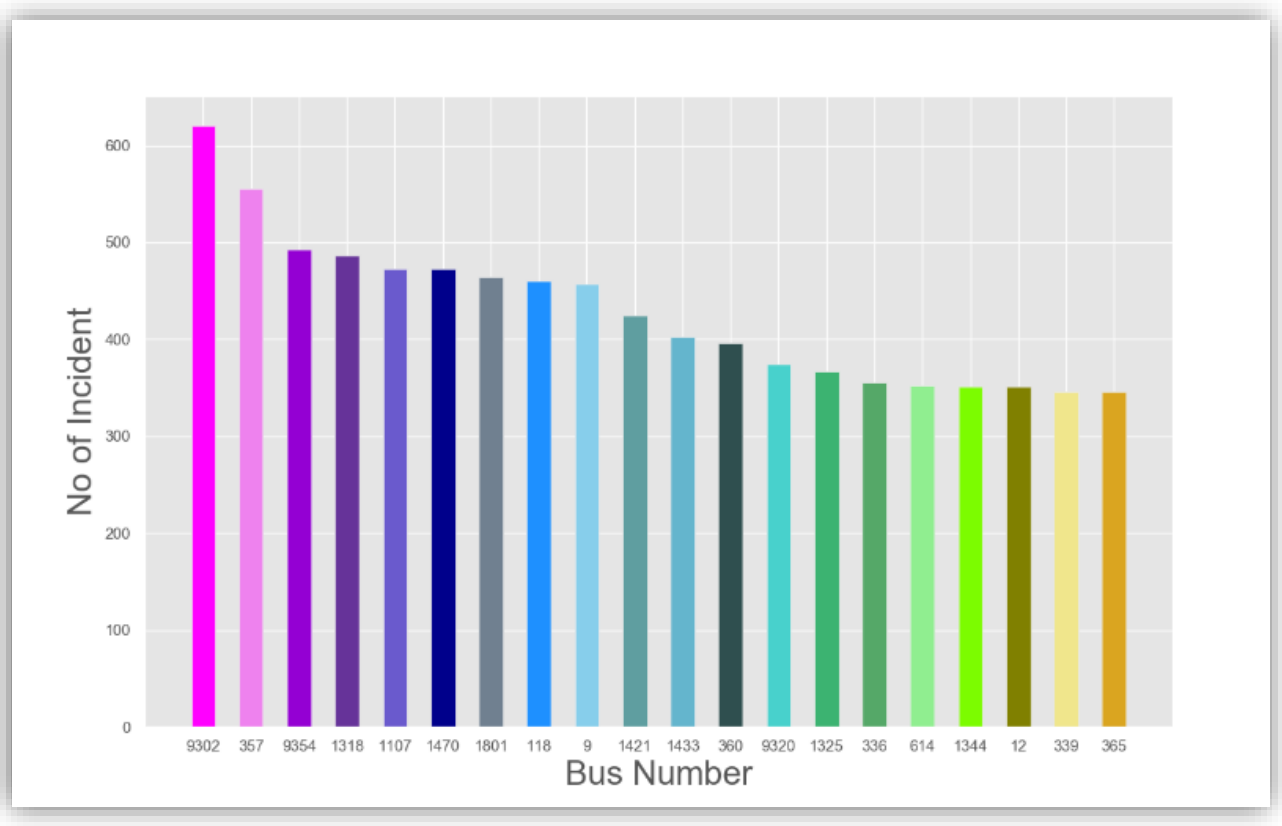


Figure: No of incident of delay-breakdown and Bus number

How many kids where there on bus when these breakdown and delays happened is important issue. Whether on the way to school or returning home, children having to sit in one place without proper observation/monitor is hard to handle situation.

```
Total 690058 were School age kids and 182675 were Pre KG kids during the break down and delay.
```

Total number of times the Bus driver notified OPT (Office of pupil Transportation) ,

	OPTNotified
No	189960
Yes	74521

Total number of times the Bus driver notified the Parents,

	ParentsNotified
Yes	196234
No	68247

Total number of times the Bus driver notified School,

	SchoolNotified
Yes	243128
No	21353

So out of 264481 incidents the bus driver notified school 243128 time, OPT 189960 times and Parents 196234 times. Bus driver need to be more active on informing OPT. (the parents will be notified by school anyways)

Conclusion:

From this analysis we came to know about the greatest number of delays are caused by Traffic, mechanical breakdown and others, Particular buses that needs to be replaced, Particular Bus companies/Vendors that needs to be rechecked, Average delay time due to various reasons, Awareness of school Bus driver in reporting the incident to the school, OPT and parents, Borough that are facing the most school bus delays and breakdowns.

Resorting all those issues we can create a more efficient and susceptible School Bus system.